Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΑΝΑΛΥΣΗ ΑΝΑΛΟΓΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ COPULAS

Βασιλική Κ. Λιλιτάκη

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 Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Πειραιάς

Ιούλιος 2016

Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΑΝΑΛΥΣΗ ΑΝΑΛΟΓΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ COPULAS

Βασιλική Κ. Λιλιτάκη

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 Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Πειραιάς

Ιούλιος 2016

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' αριθμ. …… συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών σπουδών στην Αναλογιστική Επιστήμη και Διοικητική κινδύνου

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής: Κούτρας Μάρκος (Επιβλέπων)
- 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 γνωμών του συγγραφέα.

U N I V E R S I T Y O F P I R A E U S

**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
ACTUARIAL SCIENCE AND RISK MANAGEMENT**

ANALYSIS OF ACTUARIAL DATA USING COPULAS IN

By

Vasiliki K. Lilitaki

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 Actuarial Science and Risk Management

Piraeus, Greece

July 2016

Στην οικογένειά μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά όσους συνετέλεσαν στην ολοκλήρωση της παρούσας διπλωματικής εργασίας. Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή κ. Κούτρα Μάρκο για τις πολύτιμες συμβουλές του, για τη γνώση και την καθοδήγηση που μου προσέφερε όχι μόνο κατά τη συγγραφή της παρούσας διπλωματικής εργασίας, αλλά και καθ' όλη τη διάρκεια των μεταπτυχιακών σπουδών μου. Επίσης, θα ήθελα να ευχαριστήσω και τα υπόλοιπα μέλη της τριμελούς επιτροπής, τον Καθηγητή κ. Ηλιόπουλο Γεώργιο και τον Επίκουρο Καθηγητή κ. Μπούτσικα Μιχαήλ για τη συμμετοχή τους στην εξεταστική επιτροπή.

Τέλος, θα ήθελα να ευχαριστήσω θερμά την οικογένειά μου που με στηρίζει όλα τα χρόνια της ακαδημαϊκής μου πορείας, τους φίλους μου και τους συμφοιτητές μου για την ηθική τους συμπαράσταση και υποστήριξη.

Πειραιάς, Ιούλιος 2016

Βασιλική Κ. Λιλιτάκη

Περίληψη

Η βιβλιογραφία σχετικά με τις ιδιότητες και τις εφαρμογές των Copulas, σε πολλούς επιστημονικούς κλάδους, αναπτύσσεται ραγδαία τα τελευταία χρόνια. Ειδικά στην αναλογιστική επιστήμη και στη διοικητική κινδύνου, τα Copulas αποτελούν σημαντικό εργαλείο, αφού τα δεδομένα που χρησιμοποιούνται παρουσιάζουν εξάρτηση. Στην εργασία αυτή θα κάνουμε μια εισαγωγή στην έννοια των Copulas, δίνοντας τους βασικούς ορισμούς και τις ιδιότητες τους. Στη συνέχεια, θα περιγράψουμε τη σχέση τους με τα μέτρα εξάρτησης και θα παρουσιάσουμε δύο από τις βασικότερες οικογένειες των Copulas που έχουν εφαρμογή στον αναλογισμό, τα Archimedean Copulas και τα Compound Copulas. Επίσης θα καλυφθούν θέματα στατιστικής συμπερασματολογίας και θα δοθούν οι αλγόριθμοι παραγωγής δεδομένων για τις παραπάνω οικογένειες. Τέλος, θα παρουσιάσουμε μια εφαρμογή των Copulas στον αναλογισμό.

Abstract

The literature on the properties and applications of Copulas has been developing rapidly in recent years in many scientific fields. Especially in actuarial science and risk management, Copulas are an important tool, since the data we deal with present dependence. In this dissertation we will make an introduction to the concept of Copulas, giving their basic definitions and properties. Then we will describe their relationships to measures of dependence and we will introduce two of the most important families of Copulas which are applicable to actuarial science, the Archimedean and Compound Copulas. Moreover, we will cover topics of statistical inference and we describe algorithms for generating data from these families. Finally, we present an application of Copulas in actuarial science.

Πίνακας περιεχομένων

Κατάλογος Πινάκων xvii

Κατάλογος Σχημάτων xviii

ΚΕΦΑΛΑΙΟ 1 1

COPULAS: ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ 1

ΚΕΦΑΛΑΙΟ 2 5

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΝΝΟΙΑ ΤΩΝ COPULAS 5

2.1 Εισαγωγή 5

2.2 Εισαγωγικοί Ορισμοί 5

2.3 Ορισμός και Ιδιότητες του Copula 7

2.4 Φράγματα Fréchet-Hoeffding για τα Copulas 13

2.5 Θεώρημα του Sklar 18

2.6 Copula Επιβίωσης (Survival Copulas) 24

2.7 Πολυδιάστατα Copulas (Multivariate Copulas) 27

ΚΕΦΑΛΑΙΟ 3 31

ΕΞΑΡΤΗΣΗ-ΣΥΣΧΕΤΙΣΗ ΤΩΝ COPULAS 31

3.1 Εισαγωγή 31

3.2 Συντελεστές εξάρτησης-συσχέτισης 31

3.2.1 Επιθυμητές ιδιότητες των συντελεστών εξάρτησης-συσχέτισης 31

3.2.2 Συντελεστής γραμμικής συσχέτισης του Pearson 32

3.2.3 Συντελεστές συσχέτισης τάξης (rank correlation) 34

3.2.3.1 Συντελεστής συσχέτισης Tau του Kendall 36

3.2.3.2 Συντελεστής συσχέτισης Rho του Spearman 37

3.2.4 Συντελεστής εξάρτησης ουράς (tail dependence) 39

ΚΕΦΑΛΑΙΟ 4 41

ΕΙΔΙΚΕΣ ΟΙΚΟΓΕΝΕΙΕΣ ΤΩΝ COPULAS 41

4.1 Εισαγωγή 41

4.2 Archimedean Copulas 41

4.2.1 Οικογένεια Gumbel 45

4.2.2 Οικογένεια Clayton 49

4.2.3 Οικογένεια Frank 53

4.3 Compound Copulas 57
4.3.1 Μέθοδος των Marshall και Olkin (1988) 60
ΚΕΦΑΛΑΙΟ 5 64
ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΠΡΟΣΟΜΟΙΩΣΗΣ 64
5.1 Εισαγωγή 64
5.2 Μέθοδοι εκτίμησης των Copulas 64
5.2.1 Παραμετρική μέθοδος εκτίμησης ενός Copula 64
5.2.2 Ημιπαραμετρική μέθοδος εκτίμησης ενός Copula 67
5.2.3 Μη παραμετρική μέθοδος εκτίμησης ενός Copula 68
5.3 Αξιολόγηση και επιλογή του κατάλληλου Archimedean Copula 69
5.3.1 Κριτήρια επιλογής 69
5.4 Αλγόριθμοι Προσομοίωσης 71
5.4.1 Αλγόριθμος παραγωγής δεδομένων από τα Archimedean Copulas 71
5.4.2 Αλγόριθμος παραγωγής δεδομένων από τα Compound Copulas 74
ΚΕΦΑΛΑΙΟ 6 77
ΕΦΑΡΜΟΓΗ ΤΩΝ COPULAS ΣΤΟΝ ΑΝΑΛΟΓΙΣΜΟ 77
6.1 Εισαγωγή 77
6.2 Εφαρμογή των Copulas με δεδομένα τις απώλειες-ζημιές και τα έξοδα μιας ασφαλιστικής εταιρίας 77
6.2.1 Προσαρμογή του Copula στα δεδομένα 80
6.2.2 Τιμολόγηση συμβολαίου αντασφάλειας 86
ΒΙΒΛΙΟΓΡΑΦΙΑ 89

Κατάλογος Πινάκων

1 Archimedean Copulas και οι γεννήτορές τους……..56

2 Archimedean Copulas και τα μέτρα εξάρτησής-συσχέτισης τους……57

3 Archimedean Copulas και Μετασχηματισμός Laplace……62

4 Συνοπτικός Πίνακας Περιγραφικών Στατιστικών Στοιχείων για τις Απώλειες και τα Έξοδα……....78

5 Εκτιμήσεις Παραμέτρων με Gumbel Copula και Περιθώριες Pareto……...85

6 Προσομοίωση με Βάση τα Αντασφάλιστρα…….88

Κατάλογος Σχημάτων

1.1 Διάγραμμα άνω φράγματος ενός δισδιάστατου Copula…….. 15

1.2 Διάγραμμα κάτω φράγματος ενός δισδιάστατου Copula…… 15

1.3 Διάγραμμα του Product Copula……16

1.4 Contour Diagram: Άνω φράγμα ενός δισδιάστατου Copula……16

1.5 Contour Diagram: Κάτω φράγμα ενός δισδιάστατου Copula……. 17

1.6 Contour Diagram του Product Copula……. 17

4.1 Διάγραμμα της σ.κ του Gumbel Copula για $\alpha = 2$ …….……...……. 46

4.2 Διάγραμμα της σ.π του Gumbel Copula για $\alpha = 2$ …….……...……. 47

4.3 Contour Diagram της σ.κ του Gumbel Copula για $\alpha = 2$ ……..…….....47

4.4 Διάγραμμα της σ.κ του Clayton Copula για $\alpha = 4$ ……..……... 50

4.5 Διάγραμμα της σ.π του Clayton Copula για $\alpha = 4$ ……..……... 51

4.6 Contour Diagram της σ.κ του Clayton Copula για $\alpha = 4$ ……...……... 51

4.7 Διάγραμμα της σ.κ του Frank Copula για $\alpha = \frac{3}{2}$ ……53

4.8 Διάγραμμα της σ.π του Frank Copula για $\alpha = \frac{3}{2}$ ……54

4.9 Contour Diagram της σ.κ του Frank Copula για $\alpha = \frac{3}{2}$ ……...54

6.1 Scatterplot της απώλειας έναντι του ALAE……..79

6.2 Γράφημα Προσαρμοσμένης Συνάρτησης Κατανομής του ALAE……..……..81

6.3 Γράφημα Προσαρμοσμένης Συνάρτησης Κατανομής της απώλειας…..……..…..81

6.4 Q-Q plots των παραμετρικών και μη παραμετρικών εκτιμήσεων της $K(z)$ ……...82

ΚΕΦΑΛΑΙΟ 1

COPULAS: ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Τις τελευταίες δεκαετίες όλο και περισσότεροι επιστήμονες από διαφορετικούς επιστημονικούς κλάδους ασχολούνται με τα Copulas διότι συνήθως τα δεδομένα που μελετώνται σε αυτούς είναι πολυδιάστατα και παρουσιάζουν συσχέτιση. Η ιστορία τους μπορεί να μην είναι εκτενής καθώς είναι ένα σχετικά πρόσφατο αντικείμενο στην επιστήμη της στατιστικής, αλλά η βιβλιογραφία σχετικά με τις στατιστικές ιδιότητες και εφαρμογές τους αναπτύσσεται ραγδαία τα τελευταία χρόνια. Αρκετοί ήταν οι επιστήμονες που ασχολήθηκαν κατά καιρούς με το αντικείμενο αυτό, ή πιο συγκεκριμένα με τη μελέτη πολυμεταβλητών κατανομών με δεδομένες περιθώριες κατανομές.

Η έννοια της συνάρτησης που χαρακτηρίζει τη δομή εξάρτησης μεταξύ πολλών τυχαίων μεταβλητών προέρχεται από το έργο του Hoeffding στις αρχές της δεκαετίας του 40. Ο Hoeffding ασχολήθηκε κυρίως με δισδιάστατες τυποποιημένες κατανομές και συνεισέφερε στη δημιουργία ορίων για τα Copulas. Επίσης, ο Fréchet συντέλεσε και αυτός στη δημιουργία ορίων. Με την πρωτοποριακή εργασία του Hoeffding και αργότερα του Fréchet έχει δημιουργηθεί το λεγόμενο Fréchet-Hoeffding όριο, που είναι ένα άνω όριο σε όλες τις διαστάσεις των συνήθη Copulas. Ανεξάρτητα από αυτό, και άλλοι συγγραφείς εισήγαγαν συναφείς έννοιες, ωστόσο η πρώτη εμφάνιση του όρου Copula, καθώς και το όνομα του, οφείλονται στον Abe Sklar, ο οποίος το χρησιμοποίησε το 1959 για να περιγράψει και να αποδείξει ένα θεώρημα που πλέον φέρει το όνομα του, σε μια επιστολή του προς τον Fréchet.

Επομένως, ο Abe Sklar είναι εκείνος που δημιούργησε μια καινούργια κλάση συναρτήσεων, και με το θεώρημά του θεμελίωσε τις βάσεις για τη δημιουργία της θεωρίας των Copulas. Πώς όμως σκέφτηκε να δώσει αυτήν την ονομασία; Τι μπορεί να τον ενέπνευσε; Η απάντηση είναι σχετικά απλή. Ας σημειώσουμε κατ’ αρχάς ότι στη Λατινική Γλώσσα το Copula σημαίνει δεσμός. Στην ουσία είναι ένας όρος της γραμματικής που χρησιμοποιείται σε μία πρόταση για να ενώσει το υποκείμενο με το κατηγόρημα. Η επιλογή λοιπόν της ονομασίας αυτής δεν ήταν τυχαία, καθώς και εδώ υπάρχει ένα είδος δεσμού

μεταξύ των περιθώριων κατανομών και της από κοινού συνάρτησης κατανομής. Έτσι ο όρος Copula στην ελληνική γλώσσα έχει μεταφραστεί και διαδοθεί ως σύζευξη.

Μια σύζευξη, λοιπόν, είναι μια συνάρτηση που συνδέει μια πολυδιάστατη συνάρτηση κατανομής με τις περιθώριες μονοδιάστατες συναρτήσεις κατανομών της. Πρόσφατα έχει γίνει ένα πολύ σημαντικό στατιστικό και μαθηματικό εργαλείο, αφού αντικατοπτρίζει τη σχέση εξάρτησης μεταξύ τυχαίων μεταβλητών και έχει εφαρμογή σε τομείς όπως η αναλογιστική επιστήμη (actuarial science), η διοικητική κινδύνου (risk management), τα χρηματοοικονομικά (finance), η ανάλυση χαρτοφυλακίου (portfolio analysis), κ.τ.λ. Τη δεκαετία του 90 έγινε ευρεία διάδοση της θεωρίας των συζεύξεων, με συγγραφή διαφόρων βιβλίων στο συγκεκριμένο αντικείμενο, και τώρα πια χρησιμοποιούνται ευρέως από όλους σχεδόν τους επιστήμονες.

Παρά το γεγονός ότι οι συζεύξεις έχουν εφαρμογή σε πολλούς τομείς, η εφαρμογή τους στον αναλογιστικό τομέα έχει γίνει εξαιρετικά δημοφιλής στις μέρες μας. Γενικά η αναλογιστική επιστήμη ασχολείται κυρίως με τη μελέτη των συνεπειών γεγονότων που εμπεριέχουν κινδύνους και αβεβαιότητα. Η κατανόηση των βασικών αρχών της αναλογιστικής επιστήμης καθιστά ικανούς τους αναλογιστές να αναπτύξουν μοντέλα, για αυτά τα γεγονότα, καθώς και τεχνικές για την επίλυση των αντίστοιχων πρακτικών προβλημάτων. Έτσι με σκοπό να αποκτήσουν γνώσεις σχετικά με μελλοντικές δυνατότητες, βασίζονται στη παρατήρηση που αποκτήθηκε μέσα από την προγενέστερη εμπειρία. Χρησιμοποιούν αυτές τις παρατηρήσεις και την εμπειρία για την κατασκευή, την επικύρωση και εφαρμογή κατάλληλων μοντέλων, και ενσωματώνουν συνεχώς πρόσθετες παρατηρήσεις και γνώσεις σε αυτά.

Πιο συγκεκριμένα, ένας από τους βασικούς στόχους των αναλογιστών είναι να αναπτύξουν μαθηματικά μοντέλα, τα οποία θα αντιπροσωπεύουν με μεγάλη ακρίβεια διάφορα είδη απωλειών και θα είναι χρήσιμα για την πρόβλεψη τους. Τα μοντέλα αυτά που αφορούν χρόνους ζωής ατόμων και συνταξιοδότηση, οι αναλογιστές τα περιγράφουν ως μοντέλα επιβίωσης (survival models) και συχνά θεωρούνται ως υποσύνολο των περισσότερων γενικών μοντέλων απώλειας. Επιπλέον, δίνουν έμφαση στις ασφάλειες υγείας και στις ασφάλειες ατυχήματος και ενδιαφέρονται για τη μοντελοποίηση και πρόβλεψη των θανάτων. Έτσι, έρχονται σε επαφή με διάφορους τύπους απωλειών που έχουν διαφορετική συχνότητα και βαρύτητα. Αυτές οι απώλειες, ακολουθούν τις λεγόμενες κατανομές με βαριά ουρά (long-tail distributions), οι οποίες χαρακτηρίζονται από μεγάλες και σπάνιες απαιτήσεις

που είναι αρκετά σοβαρές και δεν μπορούν να αγνοηθούν από τη διαδικασία ταξινόμησης και αποθεματοποίησης των ζημιών.

Έτσι, γίνεται και η εισαγωγή των συζεύξεων στον αναλογισμό, ως ένα πολύ σημαντικό εργαλείο για τη μοντελοποίηση και τη μελέτη της δομής εξάρτησης μεταξύ τυχαίων μεταβλητών, καθώς επίσης και της συσχέτισης διαφόρων διαστάσεων ενός αποτελέσματος. Με τη βοήθεια τους, η πλειοψηφία των αναλογιστικών προβλημάτων που περιλαμβάνουν δύο η και περισσότερες τυχαίες μεταβλητές μπορούν να αναλυθούν και να μοντελοποιηθούν.

Η θεωρία των συζεύξεων, αποτελεί μια απλή μεθοδολογία μοντελοποίησης της δομής εξάρτησης ομαδοποιημένων παρατηρήσεων χωρίς να περιορίζεται σε πολύπλοκες υποθέσεις. Εκτός από τη μαθηματική τους απλότητα, έχουν τη δυνατότητα να κατασκευάσουν μια ποικιλία δομών εξάρτησης που βασίζονται στα υπάρχοντα παραμετρικά μοντέλα των περιθώριων κατανομών. Ξεχωρίζουν την περιθώρια συμπεριφορά των μεταβλητών από τη δομή εξάρτησης τους, με τη χρήση των συναρτήσεων κατανομών, και είναι εξαιρετικά ευέλικτα και αναλυτικά εργαλεία που βρίσκουν εφαρμογή σε αναλογιστικά ζητήματα, όπως είναι η εκτίμηση κινδύνων και η διαχείριση χαρτοφυλακίων.

Ένα σημαντικό θέμα στις αναλογιστικές πρακτικές είναι αυτό της μοντελοποίησης, με πολύ αποτελεσματικό τρόπο, της εξάρτησης των χρόνων ζωής πολλών ατόμων. Η ανάπτυξη και η εύρεση αποτελεσματικών και αξιόπιστων μοντέλων για το πρόβλημα αυτό έχει γίνει απολύτως αναγκαία. Επίσης, συχνά ενδιαφέρον παρουσιάζει η εξέταση της από κοινού εξέλιξης της θνησιμότητας ομάδων που αποτελούνται από περισσότερο από ένα άτομα. Μια τέτοια ομάδα θα μπορούσε να είναι, για παράδειγμα, ο σύζυγος και η σύζυγος, μια οικογένεια με παιδιά, ή δίδυμα. Στην πράξη είναι λογικό να θεωρήσει κανείς ότι υπάρχει εξάρτηση της θνησιμότητας σε ζεύγη ή ομάδες ατόμων. Για το λόγο αυτό η στατιστική ανάλυση των μοντέλων θνησιμότητας, για παράδειγμα των παντρεμένων ζευγαριών, γίνεται συχνά αφού πρώτα διενεργηθεί ο λεγόμενος έλεγχος του συνδρόμου ‘‘broken heart’’.

Η πρώτη εφαρμογή των συζεύξεων στα από κοινού μοντέλα ζωής προέκυψε έμμεσα μέσα από το έργο του Clayton (1978) σε μελέτη του για δισδιάστατους πίνακες ζωής πατέρων και γιων. Ο Clayton επισήμανε επίσης την ερμηνεία του μοντέλου, για τις τυχαίες επιδράσεις, που στη συνέχεια αναπτύχθηκε από τον Oakes (1982). Εκτός από τις μελέτες του Clayton και Oakes, και διάφορα άλλα έργα έχουν διερευνήσει τη χρήση των μοντέλων συζεύξεων για τη μελέτη της συμπεριφοράς των πολλών ζωών. Ο Hougaard (1992), για παράδειγμα, ανέλυσε την από κοινού επιβίωση των Δανών δίδυμων που γεννήθηκαν μεταξύ 1881 και 1930. Εν

συνεχεία, ο Wang (1997), πρότεινε τις συζεύξεις για τη μοντελοποίηση συνολικών κατανομών απώλειας που συσχετίζονται με ασφαλιστήρια συμβόλαια. Επίσης οι Frees και Valdez (1998), και Klugman και Parsa (1999), χρησιμοποίησαν τις συζεύξεις για να μοντελοποιήσουν διμεταβλητά δεδομένα ασφαλιστικών αποζημιώσεων. Τέλος, ο Embrechts (1999), έκανε μια περιεκτική επισκόπηση της εφαρμογής των συζεύξεων.

Αν και μέχρι σήμερα, η θεωρία των συζεύξεων είναι καλά εδραιωμένη, δέχεται πολλή κριτική. Ένας από τους λόγους που συμβαίνει αυτό είναι η αλόγιστη χρήση τους σε διάφορους επιστημονικούς κλάδους. Το σίγουρο όμως είναι, πως η θεωρία των συζεύξεων βοήθησε κατά πολύ τους αναλογιστές, και όλους τους υπόλοιπους επιστήμονες, και συνεχίζει να είναι ένα άκρως ενδιαφέρον αντικείμενο μελέτης.

ΚΕΦΑΛΑΙΟ 2

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΝΝΟΙΑ ΤΩΝ COPULAS

2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα δώσουμε έναν ακριβή ορισμό για τα Copulas και θα παρουσιάσουμε κάποιες βασικές ιδιότητες τους, αφού πρώτα εξετάσουμε κάποιους εισαγωγικούς ορισμούς (grounded, 2-αύξουσα συνάρτηση), οι οποίοι είναι απαραίτητοι για την κατανόηση τους. Έπειτα, θα παρουσιάσουμε τα φράγματα Fréchet-Hoeffding, το θεώρημα του Sklar και θα ορίσουμε τα survival Copulas. Τέλος, παρότι θα επικεντρωθούμε στη δισδιάστατη περίπτωση, θα παρουσιάσουμε και μια γενίκευση των Copulas στις n διαστάσεις.

2.2 Εισαγωγικοί Ορισμοί

Έστω ότι έχουμε δύο τυχαίες μεταβλητές X, Y με συναρτήσεις κατανομής F και G αντίστοιχα, όπου $F(x) = P(X \le x)$, $G(y) = P(Y \le y)$. Η από κοινού συνάρτηση κατανομής τους είναι: $H(x, y) = P(X \le x, Y \le y)$. Έπειτα θα συμβολίσουμε με $\overline{R}$ το επεκτεταμένο σύνολο των πραγματικών αριθμών, το οποίο ορίζεται ως: $\overline{R} = [-\infty, \infty]$ και $I = [0,1]$. Σκοπός μας είναι να ορίσουμε τον H-όγκο, την 2-αύξουσα συνάρτηση και την grounded συνάρτηση.

Ορισμός 2.2.1: Έστω ότι S_1, S_2 είναι υποσύνολα του $\overline{R}$ και H είναι μια δισδιάστατη πραγματική συνάρτηση με πεδίο ορισμού $DomH = S_1 \times S_2$. 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 δίνεται από τον παρακάτω τύπο:

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

Ας ορίσουμε τώρα την πρώτης τάξης διαφορά της H πάνω στο ορθογώνιο παραλληλόγραμμο B ως:

$$\Delta_{x_1}^{x_2} H(x,y) = H(x_2,y) - H(x_1,y) \text{ και } \Delta_{y_1}^{y_2} H(x,y) = H(x,y_2) - H(x,y_1),$$

με τον τελεστή διαφορών:

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

Τότε, ο H-όγκος ενός ορθογώνιου παραλληλογράμμου B είναι η δεύτερης τάξης διαφορά της H πάνω στο B:

$$V_H(B) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x,y).$$

Ορισμός 2.2.2: Μία δισδιάστατη πραγματική συνάρτηση H καλείται 2-αύξουσα αν $V_H(B) \geq 0$ για όλα τα ορθογώνια παραλληλόγραμμα B των οποίων οι κορυφές ανήκουν στο $DomH = S_1 \times S_2$.

Καλό θα ήταν να επισημάνουμε πως ακόμα και όταν ισχύει ότι η $H \geq 0$, δεν συνεπάγεται ότι ο H-όγκος είναι μεγαλύτερος η ίσος του μηδενός. Υπάρχουν περιπτώσεις που παρότι η H δεν είναι θετική σε όλο το πεδίο ορισμού της, είναι 2-αύξουσα και αντίστοιχα περιπτώσεις που ενώ είναι θετική σε όλο το πεδίο ορισμού της, δεν είναι 2-αύξουσα.

Ορισμός 2.2.3: Ας υποθέσουμε ότι a_1, a_2 είναι τα ελάχιστα στοιχεία των συνόλων $S_1 \times S_2$, αντίστοιχα. Τότε η από κοινού συνάρτηση H με πεδίο ορισμού το $S_1 \times S_2$ ονομάζεται grounded, αν $H(x,a_2) = 0 = H(a_1,y)$.

Παρακάτω ακολουθεί ένα πόρισμα, το οποίο αφορά 2-αύξουσες και grounded συναρτήσεις με περιθώριες.

Πόρισμα 2.2.4: Έστω ότι S_1, S_2 είναι υποσύνολα του $\overline{R}$ και H είναι μια grounded και 2-αύξουσα συνάρτηση με πεδίο ορισμού $DomH = S_1 \times S_2$. Έστω επίσης (x_1, y_1) και (x_2, y_2) σημεία που ανήκουν στο $S_1 \times S_2$.

Τότε ισχύει ότι:

$$\left|H(x_2, y_2) - H(x_1, y_1)\right| \le \left|F(x_2) - F(x_1)\right| + \left|G(y_2) - G(y_1)\right|.$$

Ακόμη, θα δώσουμε τον ορισμό της αντίστροφης συνάρτησης κατανομής.

Ορισμός 2.2.5: Έστω F μια συνάρτηση κατανομής. Τότε ορίζεται ως αντίστροφη της F κάθε συνάρτηση $F^{(-1)}$ με $DomF = I = [0,1]$ για την οποία ισχύει:

$$F^{(-1)}(t) = \inf\{x \mid F(x) \ge t\} = \sup\{x \mid F(x) \le t\}.$$

Ενώ γενικά, για τις συναρτήσεις κατανομής ισχύουν τα παρακάτω.

Ορισμός 2.2.6: Η συνάρτηση κατανομής είναι μια συνάρτηση F με πεδίο ορισμού το $\overline{R}$, η οποία είναι αύξουσα και έχει τις παρακάτω οριακές ιδιότητες:

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

Ορισμός 2.2.7: Η από κοινού συνάρτηση κατανομής είναι μια συνάρτηση H με πεδίο ορισμού το $\overline{R^2}$, η οποία είναι 2-αύξουσα και έχει τις παρακάτω οριακές ιδιότητες:

$$H(x,-\infty) = H(-\infty, y) = 0 \text{ και } H(\infty,\infty) = 1.$$

2.3 Ορισμός και Ιδιότητες του Copula

Στη παράγραφο αυτή θα δώσουμε το μαθηματικό ορισμό των Copulas και στη συνέχεια κάποιες βασικές ιδιότητες τους, αφού πρώτα ορίσουμε το Subcopula.

Ορισμός 2.3.1: 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 $u \in S_1$ και $v \in S_2$ ισχύει ότι:

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

Σημειώνουμε ότι για κάθε $(u,v) \in DomC'$, ισχύει ότι $0 \le C'(u,v) \le 1$, έτσι ώστε $R(C')$ να είναι και αυτό υποσύνολο του I.

Το δισδιάστατο Copula, που θα μας απασχολήσει περισσότερο, είναι μια γενίκευση του δισδιάστατου Subcopula με τη διαφορά ότι το πεδίο ορισμού του είναι το I^2. Όπως θα δούμε στον παρακάτω ορισμό, οι πιο σημαντικές ιδιότητες τους είναι κοινές.

Ορισμός 2.3.2: Ένα δισδιάστατο Copula (2-Copula) είναι μια συνάρτηση $C: I^2 \to I$, η οποία έχει τις παρακάτω 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 $u_1 \le u_2$ και $v_1 \le v_2$ ισχύει ότι,

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

Συμπεραίνουμε λοιπόν πως η C είναι μια 2-αύξουσα και grounded συνάρτηση που παίρνει τιμές στο $[0,1]$. Grounded συνάρτηση είναι αφού για την ελάχιστη τιμή της μιας περιθώριας και για όλες τις τιμές της άλλης η από κοινού συνάρτηση μηδενίζεται, όπου ελάχιστη τιμή $u = v = 0$ και 2-αύξουσα αφού ο C-όγκος είναι μεγαλύτερος ή ίσος από το μηδέν. Ουσιαστικά αν και η C δεν έχει οριστεί ως από κοινού συνάρτηση κατανομής, είναι όπως συμπεραίνουμε από τον παραπάνω ορισμό αλλά και όπως θα δούμε και παρακάτω στο θεώρημα του Sklar. Επομένως αν η C είναι από κοινού συνάρτηση κατανομής με τυποποιημένες ομοιόμορφες τυχαίες μεταβλητές U,V ισχύει ότι:

$$C(u,v) = P(U \le u, V \le v)$$

Για το λόγο αυτό υπάρχει και η ιδιότητα $C(u,1)=u$, $C(1,v)=v$ που εξασφαλίζει πως αν η $C(u,v)$ είναι η από κοινού συνάρτηση κατανομής μιας δισδιάστατης τυχαίας μεταβλητής τότε οι συναρτήσεις $C(u,1)=u$, $C(1,v)=v$ είναι οι περιθώριες συναρτήσεις κατανομής των U,V αντίστοιχα.

Ένα παράδειγμα για το παραπάνω ορισμό, είναι το ακόλουθο:

Παράδειγμα: Δίνεται η συνάρτηση $C_\theta(u,v)=uv\left[1+\theta(1-u)(1-v)\right]$ με $(u,v)\in[0,1]^2$ και $-1\le\theta\le 1$. Θα αποδείξουμε ότι το C είναι Copula.

Αρχικά έχουμε ότι:

$$\mathrm{C}_\theta(u,v)=uv\left[1+\theta(1-u)(1-v)\right]$$

και θέλουμε να αποδείξουμε ότι:

$$0\le C_\theta(u,v)\le 1\Leftrightarrow$$

$$0\le uv[1+\theta(1-u)(1-v)]\le 1$$

Είναι γνωστό ότι:

$$-1\le\theta\le 1\Leftrightarrow$$

$$-(1-u)(1-v)\le\theta(1-u)(1-v)\le(1-u)(1-v)\Leftrightarrow$$

$$1-(1-u)(1-v)\le 1+\theta(1-u)(1-v)\le 1+(1-u)(1-v)\Leftrightarrow$$

$$uv\left[1-(1-u)(1-v)\right]\le uv\left[1+\theta(1-u)(1-v)\right]\le uv\left[1+(1-u)(1-v)\right]$$

Αρχικά, για το αριστερό μέλος της ανίσωσης, θέλουμε να αποδείξουμε ότι:

$$C_\theta(u,v)=uv\left[1+\theta(1-u)(1-v)\right]\ge 0$$

Για κάθε $(u,v)\in[0,1]^2$ ισχύει ότι:

$$(1-u)(1-v)\le 1\Leftrightarrow$$

$$1-(1-u)(1-v)\ge 0\Leftrightarrow$$

$$uv\left[1-(1-u)(1-v)\right]\ge 0$$

Επομένως,

$$0 \le uv\left[1-(1-u)(1-v)\right] \le uv\left[1+\theta(1-u)(1-v)\right] \Leftrightarrow$$

$$uv\left[1+\theta(1-u)(1-v)\right] \ge 0$$

Έπειτα, για το δεξί μέλος της ανίσωσης, θέλουμε να αποδείξουμε ότι:

$$uv\left[1+(1-u)(1-v)\right] \le 1 \Leftrightarrow$$
$$uv\left[1+(1-u)(1-v)\right]-1 \le 0$$

Έστω τώρα η συνάρτηση:

$$f(u) = uv\left[1+(1-u)(1-v)\right]-1$$

Τότε,

$$f'(u) = v\left[1+(1-u)(1-v)\right]+uv\left[-(1-v)\right] = v\left[1+(1-v)(1-2u)\right]$$

Είναι γνωστό ότι:

$$0 \le u \le 1 \Leftrightarrow$$
$$-2 \le -2u \le 0 \Leftrightarrow$$
$$-1 \le 1-2u \le 1 \Leftrightarrow$$
$$1-(1-v) \le 1+(1-v)(1-2u) \le 1+1-v \Leftrightarrow$$

$$v \le 1+(1-v)(1-2u) \le 2-v$$

Αφού, $0 \le v \le 1$ τότε η $f'(u) > 0 \Rightarrow f \uparrow$.

Λαμβάνοντας υπόψιν ότι:

$$0 \le u \le 1 \text{ και } f \uparrow$$

συμπεραίνουμε πως:

$$f(0) \le f(u) \le f(1) \Leftrightarrow -1 \le f(u) \le v - 1 \le 0 .$$

Άρα αποδείξαμε ακόμα πως:

$$uv\left[1 + (1-u)(1-v)\right] - 1 \le 0 \Leftrightarrow$$
$$uv\left[1 + (1-u)(1-v)\right] \le 1$$

Επομένως,

$$uv\left[1 + \theta(1-u)(1-v)\right] \le uv\left[1 + (1-u)(1-v)\right] \le 1 \Leftrightarrow$$
$$uv\left[1 + \theta(1-u)(1-v)\right] \le 1$$

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 $u_1 \le u_2$ και $v_1 \le v_2$ ο C-όγκος υπολογίζεται από τον τύπο:

$$V_c\left([u_1,u_2]\times[v_1,v_2]\right) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) =$$
$$u_2v_2\left[1+\theta(1-u_2)(1-v_2)\right] - u_2v_1\left[1+\theta(1-u_2)(1-v_1)\right] - u_1v_2\left[1+\theta(1-u_1)(1-v_2)\right] + u_1v_1\left[1+\theta(1-u_1)(1-v_1)\right] =$$

$$u_2v_2 + u_2v_2\theta(1-u_2)(1-v_2) - u_2v_1 - u_2v_1\theta(1-u_2)(1-v_1) - u_1v_2 - u_1v_2\theta(1-u_1)(1-v_2) + u_1v_1 + u_1v_1\theta(1-u_1)(1-v_1) =$$

$$u_2(v_2 - v_1) - u_1(v_2 - v_1) + \theta u_2(1-u_2)\left[v_2(1-v_2) - v_1(1-v_1)\right] - \theta u_1(1-u_1)\left[v_2(1-v_2) - v_1(1-v_1)\right] =$$

$$(v_2 - v_1)(u_2 - u_1) + \theta u_2(1-u_2)\left[v_2 - v_1 - v_2^2 + v_1^2\right] - \theta u_1(1-u_1)\left[v_2 - v_1 - v_2^2 + v_1^2\right] =$$

$$(v_2 - v_1)(u_2 - u_1) + \theta u_2(1-u_2)\left[v_2 - v_1 - (v_2 - v_1)(v_2 + v_1)\right] - \theta u_1(1-u_1)\left[v_2 - v_1 - (v_2 - v_1)(v_2 + v_1)\right] =$$

$$(v_2 - v_1)(u_2 - u_1) + \theta u_2(1-u_2)(v_2 - v_1)(1 - v_1 - v_2) - \theta u_1(1-u_1)(v_2 - v_1)(1 - v_1 - v_2) =$$

$$(v_2 - v_1)(u_2 - u_1) + \theta(v_2 - v_1)(1 - v_1 - v_2)\left[u_2(1-u_2) - u_1(1-u_1)\right] =$$

$$(v_2 - v_1)(u_2 - u_1) + \theta(v_2 - v_1)(1 - v_1 - v_2)(u_2 - u_1 - u_2^2 + u_1^2) =$$

$$(v_2 - v_1)(u_2 - u_1) + \theta(v_2 - v_1)(1 - v_1 - v_2)(u_2 - u_1)(1 - u_1 - u_2)$$

Έτσι, καταλήγουμε στη σχέση:

$$(v_2 - v_1)(u_2 - u_1)\left[1 + \theta(1 - v_1 - v_2)(1 - u_1 - u_2)\right]$$

Επειδή $(1 - v_1 - v_2)(1 - u_1 - u_2) \in [-1,1]$ για όλα τα $u_1, u_2, v_1, v_2 \in [0,1]$ τότε ο όγκος $V_c([u_1, u_2] \times [v_1, v_2]) \geq 0$ αν και μόνο αν το $\theta \in [-1,1]$, πράγμα το οποίο ισχύει από την εκφώνηση.

Επομένως, το C_θ είναι ένα Copula, σύμφωνα με τον παραπάνω ορισμό, το οποίο ανήκει στην οικογένεια που είναι γνωστή ως οικογένεια Farlie-Gumbel-Morgenstern(F-G-M).

2.4 Φράγματα Fréchet-Hoeffding για τα Copulas

Όπως ήδη έχουμε αναφέρει στο Κεφάλαιο 1, πρώτος ο Hoeffding(1940,1941) και έπειτα ο Fréchet(1951) ασχολήθηκαν με τη δημιουργία ορίων για τα Copulas. Έτσι προς τιμήν του έργου και των δύο, δόθηκε στα φράγματα αυτά η ονομασία Fréchet-Hoeffding. Το παρακάτω θεώρημα έχει αποδειχθεί για τα Subcopulas αλλά ισχύει και για τα Copulas, αφού κάθε Copula είναι ένα Subcopula.

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

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

Η απόδειξη του θεωρήματος αυτού είναι εύκολη αν σκεφτεί κανείς ότι:

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

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

Επομένως συνεπάγεται πως:

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

Επιπλέον, ξέρουμε πως:

$V_c([u,1]\times[v,1]) \ge 0$ αφού η C είναι 2-αύξουσα από τον ορισμό της.

$$V_c([u,1]\times[v,1]) \ge 0$$

$$C(1,1) - C(1,v) - C(u,1) + C(u,v) \ge 0 \Rightarrow$$

$$1 - v - u + C(u,v) \ge 0 \Rightarrow$$

$$C(u,v) \ge u + v - 1$$

Αφού,

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

Συνεπώς,

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

Τώρα, θα ορίσουμε ως κάτω φράγμα το $C^{-}(u,v) = \max(u+v-1,0)$ και ως άνω φράγμα το $C^{+}(u,v) = \min(u,v)$. Δηλαδή έχουμε:

$$C^{-}(u,v) = \max(u+v-1,0) \leq C(u,v) \leq \min(u,v) = C^{+}(u,v)$$

Στη δισδιάστατη περίπτωση, έχει αποδειχθεί πως τα άνω και κάτω όρια είναι Copula και αυτά.

Από το παραπάνω θεώρημα προκύπτει πως σε περίπτωση που τα u,v πάρουν τη μέγιστη τιμή τους, τότε το άνω όριο είναι ίσο με 1. Αντίστοιχα, αν πάρουν την ελάχιστη τιμή τους το κάτω όριο είναι ίσο με 0, δηλαδή $R(C) = [0,1]$. Συνεπώς, τα όρια αυτά είναι σημαντικά, διότι όταν επιτυγχάνονται βρισκόμαστε σε μία οριακή κατάσταση.

Παρακάτω, στο θεώρημα του Sklar, θα αναφερθούμε στα φράγματα Fréchet-Hoeffding για τις από κοινού συναρτήσεις κατανομής.

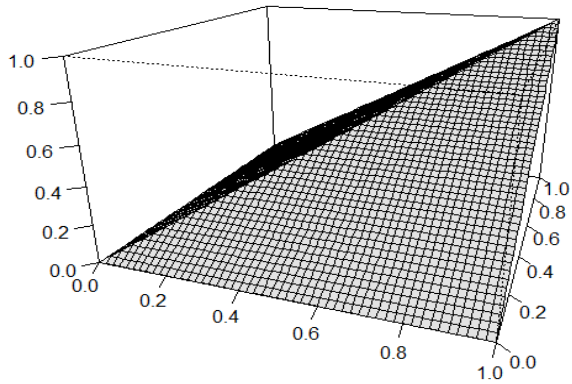
Άλλο ένα σημαντικό Copula, είναι το Product Copula (Copula ανεξαρτησίας) το οποίο συμβολίζεται με:

$$C^{\perp}(u,v) = uv$$

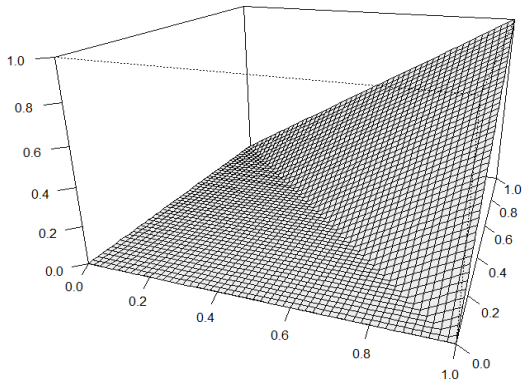
και συνδέεται με την ανεξαρτησία 2 τυχαίων μεταβλητών, αφού πρώτα έχουμε θεωρήσει τις αντίστοιχες συναρτήσεις κατανομής.

Ας δούμε τώρα τα διαγράμματα του άνω και κάτω ορίου, καθώς επίσης και τα Contour Diagrams, τα οποία απεικονίζουν μονοδιάστατες καμπύλες για τις οποίες θεωρούμε πως τα Copula είναι σταθερά.

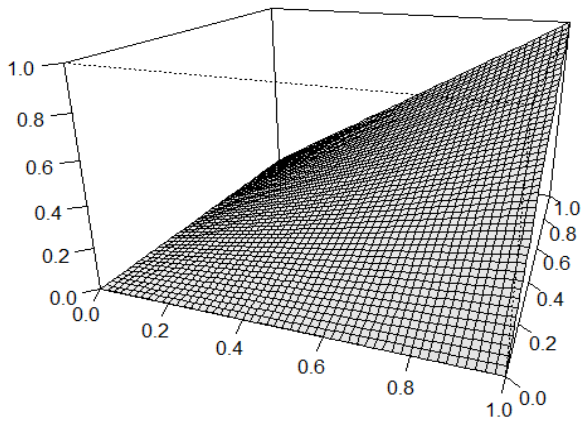
Σχήμα 1.1: Διάγραμμα άνω φράγματος ενός δισδιάστατου Copula.



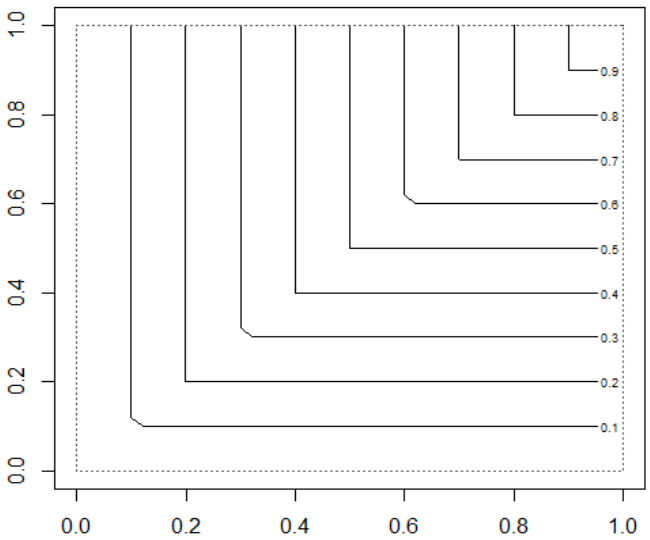
Σχήμα 1.2: Διάγραμμα κάτω φράγματος ενός δισδιάστατου Copula



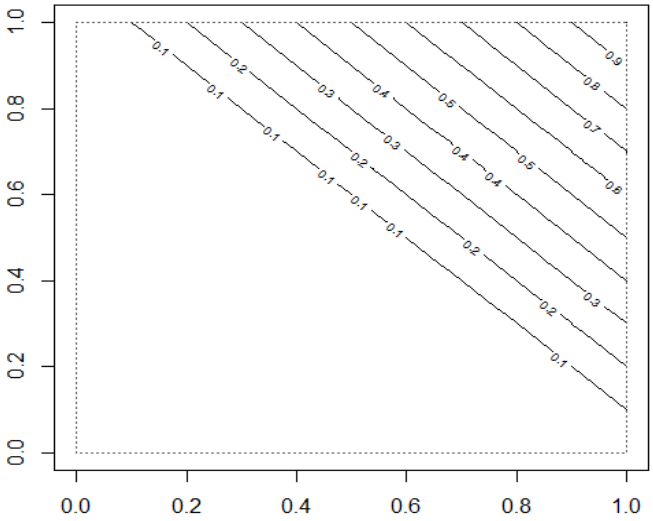
Σχήμα 1.3: Διάγραμμα του Product Copula



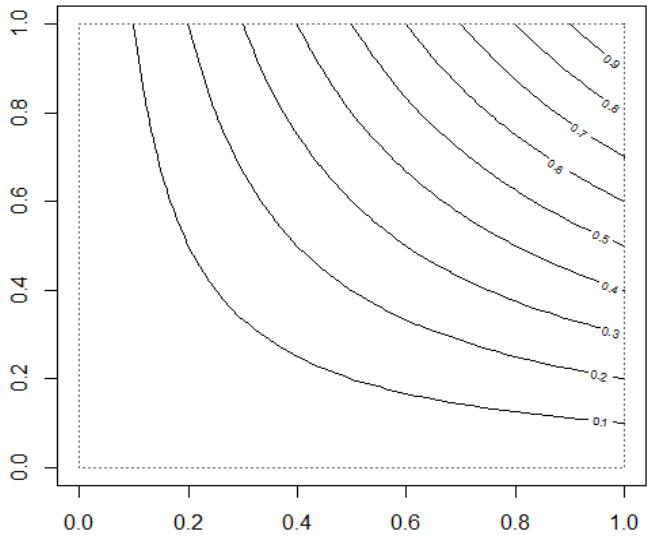
Σχήμα 1.4: Contour Diagram: Άνω φράγμα ενός δισδιάστατου Copula



Σχήμα 1.5: Contour Diagram: Κάτω φράγμα ενός δισδιάστατου Copula



Σχήμα 1.6: Contour Diagram του Product Copula



2.5 Θεώρημα του Sklar

Στην ενότητα αυτή θα παρουσιάσουμε το θεώρημα του Sklar, το οποίο αποτελεί τη βάση για τη θεωρία των Copulas αλλά και πολλών, αν όχι όλων, εφαρμογών της στη στατιστική. Το θεώρημα αυτό διευκρινίζει το ρόλο που έπαιξαν τα Copulas στις σχέσεις μεταξύ των πολυμεταβλητών συναρτήσεων κατανομής και των μονομεταβλητών περιθώριων τους.

Θεώρημα του Sklar:

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

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

Αν οι F,G είναι συνεχείς τότε και το C είναι μοναδικό, αλλιώς το C είναι μοναδικά ορισμένο στο $R(F)\times R(G)$.

Αντίστροφα, αν το C είναι Copula και F,G είναι συναρτήσεις κατανομής, τότε η συνάρτηση H, που δίνεται παραπάνω, είναι η από κοινού συνάρτηση κατανομής των τυχαίων μεταβλητών X,Y με περιθώριες συναρτήσεις κατανομής F,G αντίστοιχα.

Επομένως, το θεώρημα αυτό περιγράφει τη σύνδεση που υπάρχει μεταξύ της από κοινού συνάρτησης κατανομής και των περιθώριων κατανομών της, ορίζοντας το Copula ως από κοινού συνάρτηση κατανομής και συνδέοντας το με τις περιθώριες.

Ας δούμε τώρα το αντίστροφο του θεωρήματος αυτού, το οποίο λέει πως μπορούμε να κατασκευάσουμε ένα Copula εάν έχουμε την από κοινού συνάρτηση κατανομής και τις περιθώριες.

Ορισμός 2.5.1: Έστω ότι H είναι η από κοινού συνάρτηση κατανομής δύο συνεχών τυχαίων μεταβλητών X,Y με περιθώριες συναρτήσεις κατανομής F,G αντιστοίχως 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:

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

Φράγματα Fréchet-Hoeffding για τις από κοινού συναρτήσεις κατανομής:

Ως συνέπεια του θεωρήματος του Sklar, όταν H είναι η από κοινού συνάρτηση κατανομής δύο τυχαίων μεταβλητών X,Y με περιθώριες F,G αντιστοίχως, τότε για όλα τα $x,y \in \overline{R}$ ισχύει ότι:

$$\max\left(F(x)+G(y)-1,0\right) \le H(x,y) \le \min\left(F(x),G(y)\right).$$

Στη συνέχεια θα αναφέρουμε ένα θεώρημα το οποίο μας δείχνει πως τα Copula παραμένουν αναλλοίωτα σε γνησίως αύξοντες μετασχηματισμούς των X,Y, αλλά και τη συμπεριφορά τους στους υπόλοιπους γνησίως μονότονους μετασχηματισμούς.

Θεώρημα: Έστω X,Y δύο συνεχείς τυχαίες μεταβλητές με Copula $C_{X,Y}$ και α,β γνησίως μονότονες συναρτήσεις ορισμένες στο $R(X),R(Y)$ αντιστοίχως. Τότε ισχύει:

1. Εάν α,β είναι γνησίως αύξουσες συναρτήσεις, τότε
$$C_{\alpha(X),\beta(Y)}(u,v) = C_{X,Y}(u,v)$$
2. Εάν α γνησίως αύξουσα συνάρτηση και β γνησίως φθίνουσα συνάρτηση, τότε
$$C_{\alpha(X),\beta(Y)}(u,v) = u - C_{X,Y}(u,1-v)$$
3. Εάν α γνησίως φθίνουσα συνάρτηση και β γνησίως αύξουσα συνάρτηση, τότε
$$C_{\alpha(X),\beta(Y)}(u,v) = v - C_{X,Y}(1-u,v)$$
4. Εάν α,β γνησίως φθίνουσες συναρτήσεις, τότε
$$C_{\alpha(X),\beta(Y)}(u,v) = u+v-1+C_{X,Y}(1-u,1-v)$$.

Κλείνοντας την ενότητα αυτή, θα παρουσιάσουμε τρία παραδείγματα κατασκευής Copula, γνωρίζοντας την από κοινού συνάρτηση κατανομής και τις περιθώριες.

Παράδειγμα 1: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής H, η οποία δίνεται από τον τύπο:

$$H(x,y)=\begin{cases}\dfrac{(x+1)(e^{y}-1)}{x+2e^{y}-1}, & (x,y)\in[-1,1]\times[0,\infty],\\ 1-e^{-y}, & (x,y)\in(1,\infty]\times[0,\infty],\\ 0, & \text{αλλιώς.}\end{cases}$$

και F, G περιθώριες συναρτήσεις κατανομών που δίνονται από τους τύπους:

$$F(x)=\begin{cases}0, & x<1,\\ \dfrac{(x+1)}{2}, & x\in[-1,1],\\ 1, & x>1.\end{cases} \qquad \text{και} \qquad G(y)=\begin{cases}0, & y<0,\\ 1-e^{-y}, & y\geq 0.\end{cases}$$

Οι γενικευμένες αντίστροφες των F, G, οι οποίες είναι και γνησίως αύξουσες, είναι: $F^{(-1)}(u)=2u-1$ και $G^{(-1)}(v)=-\ln(1-v)$ για κάθε $u,v\in I$, αφού:

$$F(x)=u$$

Αν και μόνο αν,

$$\frac{(x+1)}{2}=u$$

Άρα,

$$x=2u-1.$$

Οπότε, $F^{(-1)}(u)=2u-1$. Όμοια, για την $G^{(-1)}(v)=-\ln(1-v)$.

Αφού $R(F)=R(G)=I \Rightarrow DomC=R(F)\times R(G)=I^{2}$

Έτσι, από τον παραπάνω ορισμό προκύπτει:

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

Παράδειγμα 2: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής τη δισδιάστατη λογιστική κατανομή του Gumbel (Gumbel 1961), η οποία δίνεται από τον παρακάτω τύπο:

$$H(x,y) = (1+e^{-x}+e^{-y})^{-1} \text{ για κάθε } x, y \in \overline{R}.$$

Οι X, Y έχουν μονοδιάστατες λογιστικές κατανομές αφού:

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

Οι γενικευμένες αντίστροφες των F, G, οι οποίες είναι και γνησίως αύξουσες, είναι:

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

$$F(x) = u$$

Αν και μόνο αν,

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

Άρα,

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

Οπότε, $F^{(-1)}(u) = \ln u - \ln(1-u)$. Όμοια, για την $G^{(-1)}(v) = \ln v - \ln(1-v)$.

Αφού $R(F) = R(G) = I \Rightarrow DomC = R(F) \times RG) = I^2$

Έτσι, από τον παραπάνω ορισμό προκύπτει:

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

Κατασκευάσαμε λοιπόν το Copula του Παραδείγματος 1, καθώς επίσης και το Copula του Παραδείγματος 2, και τελικά καταλήξαμε στο συμπέρασμα ότι είναι κοινά. Επίσης, το συγκεκριμένο Copula τυχαίνει να ανήκει σε μία μεγάλη οικογένεια των Copulas, γνωστή ως Archimedean Copulas, στην οποία θα αναφερθούμε παρακάτω.

Παράδειγμα 3: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής τη δισδιάστατη εκθετική κατανομή του Gumbel (Gumbel 1960α), η οποία δίνεται από τον παρακάτω 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 $[0,1]$.

Οι X, Y έχουν μονοδιάστατες εκθετικές κατανομές αφού:

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

Οι γενικευμένες αντίστροφες των F, G, οι οποίες είναι μονότονες συναρτήσεις, είναι:

$F^{(-1)}(u) = -\log(1-u)$ και $G^{(-1)}(v) = -\log(1-v)$ για κάθε $u, v \in I$, αφού:

$$F(x) = u$$

Αν και μόνο αν,

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

Άρα,

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

Οπότε, $F^{(-1)}(u) = -\ln(1-u)$. Όμοια, για την $G^{(-1)}(v) = -\ln(1-v)$.

Αφού $R(F) = R(G) = I \Rightarrow DomC = R(F) \times R(G) = I^2$.

Έτσι, από τον παραπάνω ορισμό προκύπτει:

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

2.6 Copula Επιβίωσης (Survival Copulas)

Στην ενότητα αυτή θα αναφερθούμε στα Copula επιβίωσης, τα οποία έχουν μεγάλο ενδιαφέρον και έχουν χρησιμοποιηθεί σε πολλές αναλογιστικές εφαρμογές. Ο λόγος είναι ότι οι τυχαίες μεταβλητές αναπαριστούν τους χρόνους ζωής μονάδων, ανθρώπων, ζευγαριών που ανήκουν σε κάποιο πληθυσμό. Έτσι όταν μια τυχαία μεταβλητή εκφράζει τον χρόνο ζωής μονάδων, τότε ορίζεται στο $[0,\infty]$. Η πιθανότητα λοιπόν να επιβιώσει μια μονάδα πέραν κάποιου χρονικού σημείου x, εκφράζεται από την παρακάτω συνάρτηση επιβίωσης:

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

όπου η F είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής X.

Ωστόσο, στη δισδιάστατη περίπτωση θα θεωρήσουμε πως οι χρόνοι ζωής μπορούν να πάρουν τιμές στο $\overline{R}$, παρότι ορίζονται στο $[0,\infty]$. Έτσι για ένα ζεύγος τυχαίων μεταβλητών, X, Y, με από κοινού συνάρτηση κατανομής H, η από κοινού συνάρτηση επιβίωσης ορίζεται ως:

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

και εκφράζει την πιθανότητα και οι δύο μονάδες να έχουν επιβιώσει πέραν κάποιων χρονικών σημείων x, y, αντίστοιχα.

Οι περιθώριες της $\overline{H}$ είναι οι παρακάτω:

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

Ας υποθέσουμε τώρα πως C είναι το copula των X, Y. Τότε έχουμε:

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

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

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

θα έχουμε:

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

Συμπεραίνουμε λοιπόν, πως το $\hat{C}$ είναι Copula και συγκεκριμένα είναι το Copula επιβίωσης των X, Y, το οποίο συνδέει την από κοινού συνάρτηση επιβίωσης $\overline{H}$ με τις περιθώριες συναρτήσεις επιβίωσης, με τρόπο ανάλογο που ένα Copula συνδέει την από κοινού συνάρτηση κατανομής με τις περιθώριες του. Θα πρέπει όμως να προσέξουμε το εξής. Το Copula επιβίωσης $\hat{C}$ δεν είναι η από κοινού συνάρτηση επιβίωσης $\overline{C}$ δύο τυχαίων μεταβλητών που κατανέμονται ομοιόμορφα στο $(0,1)$ και με από κοινού συνάρτηση κατανομής το Copula C. Αυτό συμβαίνει γιατί:

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

Ακολουθεί ένα παράδειγμα υπολογισμού του Copula επιβίωσης και της από κοινού συνάρτησης επιβίωσης.

Παράδειγμα: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής τη δισδιάστατη εκθετική κατανομή του Gumbel (Gumbel 1960α), η οποία δίνεται από τον παρακάτω 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 $[0,1]$.

Από το Παράδειγμα 3, της προηγούμενης ενότητας, είχαμε αποδείξει ότι:

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

Το Copula επιβίωσης τώρα υπολογίζεται από τον παρακάτω τύπο:

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

Επομένως έχουμε:

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

Καταλήξαμε λοιπόν σε έναν πιο απλό τύπο Copula.

Η από κοινού συνάρτηση επιβίωσης υπολογίζεται από τη σχέση:

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

δηλαδή,

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

2.7 Πολυδιάστατα Copulas (Multivariate Copulas)

Στην ενότητα αυτή θα επεκτείνουμε τα αποτελέσματα που δόθηκαν στις προηγούμενες παραγράφους, για την πολυδιάστατη περίπτωση, δηλαδή για την περίπτωση που έχουμε διάσταση $n \geq 3$. Θα παρουσιάσουμε, αναφορικά, τους περισσότερους από τους ορισμούς και τα θεωρήματα στην πολυδιάστατη περίπτωση, παραλείποντας τις αποδείξεις τους.

Ας ξεκινήσουμε λοιπόν από τους ορισμούς του H-όγκου, της 2-αύξουσας συνάρτησης και της grounded συνάρτησης για την πολυδιάστατη περίπτωση.

Ορισμός 2.7.1: Έστω $S_1, S_2, ..., S_n$ μη κενά υποσύνολα του $\overline{R}$ και H μια n-διάστατη πραγματική συνάρτηση με πεδίο ορισμού το $DomH = S_1 \times S_2 \times ... \times S_n$. Έστω ακόμη $B = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$ με $a \leq b$ ένα n-διάστατο παραλληλεπίπεδο (n-box) με τις κορυφές του στο $DomH$. Τότε, ο H-όγκος του B δίνεται από τον παρακάτω τύπο:

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

όπου το άθροισμα υπολογίζεται για όλες τις κορυφές $\mathbf{c} = (c_1, c_2, ..., c_n)$ του B ενώ το $\operatorname{sgn}(\mathbf{c})$ από:

$$\operatorname{sgn}(\mathbf{c}) = \begin{cases} 1, & \text{αν } c_k = a_k, k \text{ άρτιος} \\ -1, & \text{αν } c_k = a_k, k \text{ περιττός} \end{cases}$$

Εναλλακτικά, ο H-όγκος ενός n-διάστατου παραλληλεπιπέδου B είναι η n-οστής τάξης διαφορά της H πάνω στο B:

$$V_H(B) = \Delta_a^b H(t) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} ... \Delta_{a_1}^{b_1} H(t)$$

Ορισμός 2.7.2: Μία 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 $DomH = S_1 \times S_2 ... \times S_n$.

Ορισμός 2.7.3: Έστω H μία n-διάστατη πραγματική συνάρτηση με $DomH = S_1 \times S_2 \times ... \times S_n$, όπου κάθε S_k έχει ένα ελάχιστο σημείο. Τότε η H ονομάζεται grounded όταν $H(\mathbf{t}) = 0$ για όλα τα $\mathbf{t} \in DomH = S_1 \times S_2 \times ... S_n$, τέτοια ώστε τουλάχιστον ένα από τα t_k ισούται με την ελάχιστη τιμή του.

Ορισμός 2.7.4: Μία n-διάστατη συνάρτηση κατανομής είναι μία συνάρτηση H με πεδίο ορισμού το $\overline{R}^n = (\overline{R} \times \overline{R} \times ... \times \overline{R})$ έτσι ώστε:

1. Η H είναι n-αύξουσα
2. $H(\mathbf{t}) = 0$ για όλα τα $\mathbf{t} \in \overline{R}^n$, έτσι ώστε $t_k = -\infty$ για τουλάχιστον ένα k και $H(\infty, \infty, ..., \infty) = 1$.

Συνεχίζουμε με τον ορισμό ενός n-διάστατου Copula, δίνοντας επίσης τα φράγματα Fréchet-Hoeffding και το θεώρημα του Sklar.

Ορισμός 2.7.5: Ένα n-διάστατο Copula (n-Copula) είναι μια συνάρτηση $C: I^n \to I$, η οποία έχει τις παρακάτω ιδιότητες:

1. Για κάθε $\mathbf{u} \in I^n$,

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

αν τουλάχιστον μία συνιστώσα του $\mathbf{u}$ είναι μηδενική (grounded συνάρτηση)

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

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

3. Για κάθε $a, b \in I^n$ με $a \le b$ ισχύει ότι:

$$V_C([a,b]) \ge 0, \text{ δηλαδή είναι } n\text{-αύξουσα.}$$

Τέλος, αποδεικνύεται εύκολα πως για οποιοδήποτε n-Copula, $n \geq 3$, κάθε k-περιθώρια της C είναι ένα k-Copula, $2 \leq k \leq n$.

Φράγματα Fréchet-Hoeffding για n-διάστατα Copulas:

Οι επεκτάσεις των φραγμάτων που είδαμε στην δισδιάστατη περίπτωση είναι:

Το άνω όριο: $C^{+}(\mathbf{u}) = \min(u_1, u_2, ..., u_n)$

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

Το Product Copula: $C^{\perp}(\mathbf{u}) = u_1 u_2 ... u_n$

Παρότι στη δισδιάστατη περίπτωση υπάρχουν φράγματα τα οποία αποδεικνύεται πως είναι και τα ίδια Copula, στις n-διαστάσεις το κάτω όριο (C^{-}) δεν είναι Copula, για διάσταση $n > 2$. Ωστόσο η αριστερή ανισότητα χρησιμοποιείται, αφού αποδεικνύεται πως για $n \geq 3$ και για κάθε $\mathbf{u} \in I^n$, υπάρχει ένα Copula C τέτοιο ώστε:

$$C(\mathbf{u}) = C^{-\,n}(\mathbf{u})$$

Τέλος, για το άνω φράγμα και το Product Copula ισχύει το παρακάτω θεώρημα:

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

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

Θεώρημα του Sklar για n-διάστατα Copulas:

Έστω ότι H είναι μία n-διάστατη συνάρτηση κατανομής με περιθώριες $F_1, F_2, ..., F_n$. Τότε υπάρχει ένα n-Copula τέτοιο ώστε για κάθε $\mathbf{x} \in \overline{R}^n$, ισχύει:

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

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

Αντίστροφα, αν το C είναι ένα n-Copula και $F_1, F_2, ..., F_n$ είναι συναρτήσεις κατανομής, τότε η συνάρτηση H, που δίνεται παραπάνω, είναι μια n-διάστατη συνάρτηση κατανομής με περιθώριες τις $F_1, F_2, ..., F_n$.

Επομένως, το θεώρημα αυτό περιγράφει τη σύνδεση που υπάρχει μεταξύ της από κοινού συνάρτησης κατανομής και των περιθώριων κατανομών της, ορίζοντας το Copula ως από κοινού συνάρτηση κατανομής και συνδέοντας το με τις περιθώριες.

ΚΕΦΑΛΑΙΟ 3

ΕΞΑΡΤΗΣΗ-ΣΥΣΧΕΤΙΣΗ ΤΩΝ COPULAS

3.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με τα μέτρα εξάρτησης-συσχέτισης (measures of association) των Copulas. Συγκεκριμένα θα παρουσιάσουμε κάποιες επιθυμητές ιδιότητες των συντελεστών εξάρτησης και θα δώσουμε τους γενικούς ορισμούς του μέτρου συμφωνίας (measure of concordance). Έπειτα θα παρουσιάσουμε το συντελεστή γραμμικής συσχέτισης του Pearson, το συντελεστή συσχέτισης τάξης Rho του Spearman και Tau του Kendall καθώς επίσης και το συντελεστή εξάρτησης ουρών.

3.2 Συντελεστές εξάρτησης-συσχέτισης

Στην ενότητα αυτή θα παρουσιάσουμε κάποιες έννοιες εξάρτησης ή συσχέτισης, όπως: γραμμική συσχέτιση (linear correlation), συμφωνία (concordance), εξάρτηση ουράς (tail dependence) καθώς επίσης και κάποιες ιδιότητες που θέλουμε να έχουν οι συντελεστές εξάρτησης-συσχέτισης. Έπειτα θα παρουσιάσουμε κάποιους από αυτούς, που συνδέονται με τις παραπάνω έννοιες.

3.2.1 Επιθυμητές ιδιότητες των συντελεστών εξάρτησης-συσχέτισης

Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F, G αντίστοιχα και από κοινού συνάρτηση κατανομής H. Τότε θα λέμε ότι είναι εξαρτημένες ή συσχετισμένες, αν δεν είναι ανεξάρτητες, δηλαδή όταν ισχύει ότι:

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

Στη δισδιάστατη περίπτωση, για κάθε συντελεστή εξάρτησης $\delta(X,Y)$ επιθυμούμε τις ακόλουθες ιδιότητες:

1. $\delta(X,Y)=\delta(Y,X)$ (συμμετρία).
2. $-1\le\delta(X,Y)\le 1$ (κανονικότητα).
3. $\delta(X,Y)=1\Leftrightarrow(X,Y)$ comonotonic.

   Δηλαδή, τα X,Y έχουν τέλεια θετική εξάρτηση όταν ο συντελεστής παίρνει τη μέγιστη τιμή του. Αντίστοιχα,
4. $\delta(X,Y)=-1\Leftrightarrow(X,Y)$ countermonotonic.

   Δηλαδή όταν τα X,Y έχουν τέλεια αρνητική εξάρτηση, όταν ο συντελεστής παίρνει την ελάχιστη τιμή του.
5. Για ένα γνησίως μονότονο μετασχηματισμό $T:R\to R$ του X ισχύει ότι:

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

Ένα τέτοιο μέτρο εξάρτησης είναι ο συντελεστής γραμμικής συσχέτισης Pearson, που θα δούμε παρακάτω.

3.2.2 Συντελεστής γραμμικής συσχέτισης του Pearson

Έστω X,Y δύο τυχαίες μεταβλητές με συναρτήσεις κατανομής F,G αντίστοιχα και από κοινού συνάρτηση κατανομής H. Τότε ο συντελεστής γραμμικής συσχέτισης του Pearson για τα X,Y δίνεται από το τύπο:

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

όπου $\sigma_X\sigma_Y$ οι τυπικές αποκλίσεις των X,Y αντίστοιχα.

Σύμφωνα με την έκφραση του Hoeffding (1940) για τη συνδιακύμανση ισχύει γενικά ότι:

$$Cov(X,Y) = \iint_D \left[F(x,y) - F_1(x)F_2(y)\right]dxdy \text{ , όπου } D = DomF_1 \times DomF_2$$

Επομένως ο συντελεστής γραμμικής συσχέτισης του Pearson μπορεί να γραφεί:

$$\rho(X,Y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[H(x,y) - F(x)G(y)\right]dxdy}{\sigma_X \sigma_Y}.$$

Για $F(x) = u, G(y) = v$ και $H(x,y) = C\big(F(x), G(y)\big) = C(u,v)$:

$$\rho(X,Y) = \frac{\int_0^1\int_0^1\left[C(u,v) - uv\right]dF^{(-1)}(u)dG^{(-1)}(v)}{\sigma_X \sigma_Y}$$

Είναι φανερό πως ο συγκεκριμένος συντελεστής αφορά γραμμικές και μόνο συσχετίσεις των τυχαίων μεταβλητών και ικανοποιεί μόνο δύο από τις παραπάνω επιθυμητές ιδιότητες, αυτήν της συμμετρίας και της κανονικότητας.

Αντιθέτως, τα σημεία στα οποία υστερεί ο συντελεστής αυτός είναι αρκετά. Πρώτον, οι τιμές του εξαρτώνται από τις περιθώριες συναρτήσεις κατανομής, επομένως αυτές πρέπει να είναι γνωστές. Δεύτερον δεν μας δίνει πληροφορίες σχετικά με τη δομή εξάρτησης των τυχαίων μεταβλητών, εκτός από κάποιες ειδικές περιπτώσεις της κανονικής κατανομής. Δηλαδή υπάρχει περίπτωση η μηδενική συσχέτιση να μη συνεπάγεται πάντα με την ανεξαρτησία. Επιπλέον, ο συντελεστής συσχέτισης του Pearson δεν ορίζεται για κάποιες κατανομές με βαριές ουρές (heavy-tail distributions) για τις οποίες δεν υπάρχουν ροπές δεύτερης τάξης όπως πχ της δισδιάστατης t-Student κατανομής με βαθμούς ελευθερίας 1 ή 2. Τέλος, ο $\rho(X,Y)$ παραμένει αναλλοίωτος σε γραμμικούς και μόνο μετασχηματισμούς.

Λόγω των περιορισμών αυτών έχουν προταθεί και άλλοι συντελεστές, οι λεγόμενοι συντελεστές συσχέτισης τάξης, οι οποίοι αναλύονται στη συνέχεια.

3.2.3 Συντελεστές συσχέτισης τάξης (rank correlation)

Στη συγκεκριμένη κατηγορία ανήκουν οι συντελεστές Rho του Spearman και Tau του Kendall όπου και οι δύο βασίζονται στην έννοια της συμφωνίας (concordance) και ασυμφωνίας (discordance) των τυχαίων μεταβλητών. Ένα πλεονέκτημα που έχουν και οι δύο συντελεστές σε σύγκριση με το συντελεστή γραμμικής συσχέτισης του Pearson, ρ_{XY}, είναι πως δεν εξαρτώνται από τις περιθώριες κατανομές. Έτσι λοιπόν στο σημείο αυτό θα δώσουμε τον ορισμό της διάταξης συμφωνίας δύο Copulas στη δισδιάστατη περίπτωση.

Ορισμός: Έστω ότι έχουμε δύο διαφορετικά Copula C_1, C_2. Τότε θα λέμε ότι το C_1 είναι μικρότερο από το C_2 ($C_1 \prec C_2$) όταν ισχύει η παρακάτω ανισότητα

$$C_1(u,v) \le C_2(u,v) \text{ για κάθε } (u,v) \in I^2$$

Επίσης, αν X, Y είναι δύο τυχαίες μεταβλητές με συναρτήσεις κατανομής F, G αντίστοιχα. Τότε για $u = F(x)$ και $v = G(y)$ η ανισότητα γίνεται

$$C_1\big(F(x)\big) \le C_2\big(G(y)\big)$$

Δηλαδή συμπεραίνουμε πως όταν ισχύει η παραπάνω ανισότητα, οι τυχαίες μεταβλητές είναι περισσότερο συσχετισμένες μέσω του Copula C_2. Ακόμη είναι φανερό πως το κάτω φράγμα Fréchet-Hoeffding, C^-, είναι μικρότερο από κάθε Copula και αντίστοιχα το άνω φράγμα Fréchet-Hoeffding, C^+, είναι το μεγαλύτερο.

Συμφωνία-Ασυμφωνία (Concordance-Discordance)

Γενικά, δύο τυχαίες μεταβλητές X, Y λέμε ότι είναι σε συμφωνία, όταν μεγάλες (μικρές) τιμές της μίας τείνουν να συσχετιστούν με μεγάλες (μικρές) τιμές της άλλης, ενώ λέμε ότι είναι σε ασυμφωνία, όταν μεγάλες (μικρές) τιμές της μίας τείνουν να συσχετιστούν με μικρές (μεγάλες) τιμές της άλλης. Για να είμαστε πιο ακριβείς μπορούμε να δώσουμε τον παρακάτω ορισμό.

Ορισμός: Έστω X,Y δύο συνεχείς τυχαίες μεταβλητές και $(x_1,y_1),(x_2,y_2)$ δύο ζεύγη των παρατηρούμενων τιμών τους. Τότε θα έχουμε συμφωνία των $(x_1,y_1),(x_2,y_2)$ όταν:

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

Ομοίως, θα έχουμε ασυμφωνία των $(x_1,y_1),(x_2,y_2)$ όταν:

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

Εναλλακτικά, τα $(x_1,y_1),(x_2,y_2)$ βρίσκονται σε συμφωνία όταν:

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

Και αντίστοιχα σε ασυμφωνία όταν:

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

Μέτρο συμφωνίας (Measure of Concordance)

Ορισμός: Έστω X,Y δύο συνεχείς τυχαίες μεταβλητές, M ένα μέτρο συσχέτισης τους και C το Copula τους. Τότε θα λέμε πως το M είναι μέτρο συμφωνίας των X,Y και θα το συμβολίζουμε με $M_{X,Y} = M_C$ εάν ικανοποιεί τις παρακάτω ιδιότητες:

1. Το $M_{X,Y}$ ορίζεται για κάθε ζεύγος X,Y συνεχών τυχαίων μεταβλητών.
2. $M_{X,Y} \in [-1,1]$, με $M_{X,X} = 1$ και $M_{X,-X} = -1$.
3. $M_{X,Y} = M_{Y,X}$.
4. Αν οι X,Y είναι ανεξάρτητες τότε $M_{X,Y} = 0$.
5. $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$.
6. Αν $\{(X_n,Y_n)\}$ είναι μία ακολουθία συνεχών τυχαίων μεταβλητών με Copulas C_n και αν η C_n συγκλίνει σημειακά στη C, δηλαδή $\lim_{n\to+\infty} C_n(u,v) = C(u,v)$ για κάθε $(u,v) \in I^2$, τότε $\lim_{n\to+\infty} M_{C_n} = M_C$.
7. Αν C_1 και C_2 είναι Copulas με $C_1 \prec C_2$ τότε $M_{C_1} \le M_{C_2}$.

Στη συνέχεια θα παρουσιάσουμε τους συντελεστές συσχέτισης τάξης Rho του Spearman και Tau του Kendall οι οποίοι όπως έχει αποδειχθεί είναι μέτρα συμφωνίας.

3.2.3.1 Συντελεστής συσχέτισης Tau του Kendall

Ο συντελεστής συσχέτισης Tau του Kendall, που συμβολίζεται με τ, εισήχθη για πρώτη φορά από τον Fechner περίπου το 1900 και ανακαλύφθηκε εκ νέου από τον Kendall το 1938. Ο συντελεστής αυτός βασίζεται στην έννοια της συμφωνίας και ασυμφωνίας, συνεπώς μπορούμε να δώσουμε τον ακόλουθο ορισμό.

Ορισμός: Έστω (X_1, Y_1) και (X_2, Y_2) δύο ανεξάρτητα και ισόνομα τυχαία διανύσματα με από κοινού συνάρτηση κατανομής H και X_1, X_2 έχουν συνάρτηση κατανομής F, ενώ τα Y_1, Y_2 έχουν συνάρτηση κατανομής G. Τότε ο συντελεστής συσχέτισης τ του Kendall ισούται με την πιθανότητα της συμφωνίας μείον την πιθανότητα της ασυμφωνίας:

$$\tau(X,Y) = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right]$$

Στη συνέχεια, θα δώσουμε έναν ακόμη ισοδύναμο ορισμό για το συντελεστή συσχέτισης τ του Kendall, όταν έχουμε το Copula C.

Ορισμός: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F, G αντίστοιχα και με Copula C. Τότε ο συντελεστής συσχέτισης τ του Kendall ορίζεται ως:

$$\tau(X,Y) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1.$$

Ο παραπάνω συντελεστής εκτός από τις ιδιότητες του μέτρου συμφωνίας, ικανοποιεί και τις επιθυμητές ιδιότητες που θέλουμε να έχουν οι συντελεστές συσχέτισης αφού:

1. $\tau_{X,Y} = \tau_{Y,X}$ (συμμετρία).
2. $-1 \le \tau(X,Y) \le 1$.
3. $\tau(X,Y) = 1 \Leftrightarrow C = C^+ \Leftrightarrow$ τέλεια θετική συσχέτιση.
4. $\tau(X,Y) = -1 \Leftrightarrow C = C^- \Leftrightarrow$ τέλεια αρνητική συσχέτιση.

5. το $\tau(X,Y)$ παραμένει αναλλοίωτο σε γνησίως μονότονους μετασχηματισμούς των X,Y.

3.2.3.2 Συντελεστής συσχέτισης Rho του Spearman

Ο συντελεστής συσχέτισης Rho του Spearman, που συμβολίζεται με ρ_S, προτάθηκε και πήρε το όνομα του από τον Charles Spearman το 1904. Ουσιαστικά, όπως και ο συντελεστής συσχέτισης του Kendall, έτσι και αυτός βασίζεται στην έννοια της συμφωνίας και ασυμφωνίας. Έτσι μπορούμε να δώσουμε τον παρακάτω ορισμό.

Ορισμός: Έστω $(X_1,Y_1),(X_2,Y_2)$ και (X_3,Y_3) τρία ανεξάρτητα τυχαία διανύσματα με 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 ο συντελεστής συσχέτισης ρ_S του Spearman ορίζεται να είναι ανάλογος με την πιθανότητα της συμφωνίας μείον την πιθανότητα της ασυμφωνίας για τα διανύσματα (X_1,Y_1) και (X_2,Y_3), τα οποία έχουν κοινές περιθώριες και το ένα διάνυσμα έχει από κοινού συνάρτηση κατανομής την H ενώ το άλλο έχει ανεξάρτητες συνιστώσες:

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

Σημειώνουμε εδώ, πως στον παραπάνω τύπο θα μπορούσαμε να έχουμε χρησιμοποιήσει εξίσου και το διάνυσμα (X_3,Y_2). Επίσης, ενώ το διάνυσμα (X_1,Y_1) έχει από κοινού συνάρτηση κατανομής την $H(x,y)$ το διάνυσμα (X_2,Y_3) έχει από κοινού συνάρτηση κατανομής το $F(x)G(y)$, αφού X_2,Y_3 είναι ανεξάρτητα. Τέλος, εφόσον ο συντελεστής ρ_S του Spearman είναι μέτρο συμφωνίας, πρέπει να ικανοποιεί τη δεύτερη ιδιότητα των μέτρων συμφωνίας, δηλαδή να ανήκει στο διάστημα $[-1,1]$. Για τα λόγο αυτό υπάρχει και ο συντελεστής 3 στον παραπάνω τύπο.

Έπειτα, θα δώσουμε έναν ακόμη ισοδύναμο ορισμό για το συντελεστή συσχέτισης ρ_S του Spearman, όταν έχουμε το Copula C.

Ορισμός: Έστω X,Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F,G αντίστοιχα και με Copula C. Τότε ο συντελεστής συσχέτισης ρ_S του Spearman ορίζεται ως:

$$\rho_S(X,Y) = 12\int_0^1\int_0^1 [C(u,v) - uv]dudv$$

Ακριβώς όπως ο συντελεστής συσχέτισης τ του Kendall, έτσι και ο ρ_S του Spearman, ικανοποιεί εκτός από τις ιδιότητες του μέτρου συμφωνίας και τις επιθυμητές ιδιότητες των συντελεστών συσχέτισης. Δηλαδή:

1. το ρ_S είναι συμμετρικό.
2. $-1 \le \rho_S \le 1$.
3. $\rho_S = 1 \Leftrightarrow C = C^+ \Leftrightarrow$ τέλεια θετική συσχέτιση.
4. $\rho_S = -1 \Leftrightarrow C = C^- \Leftrightarrow$ τέλεια αρνητική συσχέτιση.
5. το ρ_S παραμένει αναλλοίωτο σε γνησίως μονότονους μετασχηματισμούς των X,Y.

Παρά το γεγονός ότι και οι δύο συντελεστές τάξης μετρούν την πιθανότητα συμφωνίας μεταξύ δύο τυχαίων μεταβλητών, με Copula C, οι τιμές τους συχνά διαφέρουν κατά πολύ. Σύμφωνα λοιπόν με τους Durbin και Stuart (1951), οι συντελεστές αυτοί συνδέονται μεταξύ τους με κάποιες ανισοτικές σχέσεις.

Θεώρημα: Έστω X,Y δύο συνεχείς τυχαίες μεταβλητές και τ,ρ_S, οι συντελεστές συσχέτισης Tau του Kendall και Rho του Spearman, αντίστοιχα. Τότε ισχύουν οι παρακάτω ανισότητες:

$$-1 \le 3\tau - 2\rho_S \le 1$$

$$\frac{3\tau - 1}{2} \le \rho_S \le \frac{1 + 2\tau - \tau^2}{2}, \quad \text{για } \tau \ge 0$$

και

$$\frac{\tau^2+2\tau-1}{2} \le \rho_S \le \frac{1+3\tau}{2}, \quad \text{για } \tau \le 0 .$$

Συνοψίζοντας, οι παραπάνω συντελεστές συσχέτισης τάξης, υπερτερούν σε σύγκριση με το συντελεστή γραμμικής συσχέτισης του Pearson εφόσον ικανοποιούν όλες τις επιθυμητές ιδιότητες και δεν εξαρτώνται από τις περιθώριες συναρτήσεις κατανομής, αλλά μόνο από το Copula C. Επίσης παραμένουν αναλλοίωτοι σε γνησίως μονότονους μετασχηματισμούς των τυχαίων μεταβλητών και όχι μόνο σε γραμμικούς. Παρ 'όλα αυτά, λόγω του ότι είναι συντελεστές συσχέτισης και όχι εξάρτησης, δεν μας παρέχουν όλες τις δυνατές πληροφορίες για τη δομή εξάρτησης και υπάρχουν περιπτώσεις που λόγω της πολυπλοκότητας των τύπων τους, δεν είναι εύκολος ο υπολογισμός τους.

3.2.4 Συντελεστής εξάρτησης ουράς (tail dependence)

Ο συντελεστής εξάρτησης ουράς, είναι ένα μέτρο εξάρτησης που αφορά τις ουρές των κατανομών και μετρά την εξάρτηση των μεταβλητών που παίρνουν τιμές στο επάνω-δεξιά (upper-right quadrant) και στο κάτω-αριστερά (lower-left quadrant) τεταρτημόριο μιας δισδιάστατης κατανομής. Ουσιαστικά, ο συντελεστής αυτός είναι χρήσιμος όταν θέλουμε να μελετήσουμε την εξάρτηση μεταξύ ακραίων τιμών. Στη περίπτωση λοιπόν που έχουμε δύο συνεχείς τυχαίες μεταβλητές X,Y, το μέτρο εξάρτησης ουράς θα είναι συνάρτηση του Copula και έτσι θα παραμένει αναλλοίωτο σε γνησίως αύξοντες μετασχηματισμούς των τυχαίων μεταβλητών X,Y.

Ορισμός: Έστω X,Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F,G αντίστοιχα. Ο συντελεστής εξάρτησης της άνω ουράς (upper tail dependence), λ_U είναι το όριο, της δεσμευμένης πιθανότητας, ότι το Y είναι μεγαλύτερο του $100u-$οστού ποσοστημορίου της G, δεδομένου ότι το X είναι μεγαλύτερο του $100u-$οστού ποσοστημορίου της F, καθώς το $u \to 1$. Δηλαδή:

$$\lambda_U = \lim_{u \to 1^-} P\left(Y > G^{(-1)}(u) \mid X > F^{(-1)}(u)\right)$$

με την προϋπόθεση ότι το όριο λ_U υπάρχει.

Εάν το όριο $\lambda_U \in (0,1]$, τότε θα λέμε ότι τα X, Y είναι ασυμπτωτικά εξαρτημένα στην άνω ουρά, ενώ αν το όριο $\lambda_U = 0$, τότε θα λέμε ότι τα X, Y είναι ασυμπτωτικά ανεξάρτητα στην άνω ουρά.

Εάν υποθέσουμε ότι C είναι το Copula των X, Y ,ισχύει ότι:

$$\lambda_U = \lim_{u \to 1^-} \frac{\overline{C}(u,u)}{1-u} = \lim_{u \to 1^-} P[U_1 > u \mid U_2 > u].$$

Τότε εάν το $\lambda_U \in (0,1]$, θα λέμε ότι το Copula C παρουσιάζει άνω εξάρτηση ουράς, ενώ αν το όριο $\lambda_U = 0$, τότε θα λέμε ότι το Copula C παρουσιάζει άνω ανεξαρτησία ουράς. Επίσης, το λ_U χρησιμοποιείται αρκετά στις αναλογιστικές εφαρμογές τις θεωρίας ακραίων τιμών, για το χειρισμό της πιθανότητας του ότι ένα ακραίο γεγονός εξαρτάται από ένα άλλο εξίσου ακραίο.

Αντίστοιχα, ο συντελεστής εξάρτησης της κάτω ουράς (lower tail dependence), λ_L είναι το όριο, της δεσμευμένης πιθανότητας, ότι το Y είναι μικρότερο ή ίσο του $100u-$οστού ποσοστημορίου της G, δεδομένου ότι το X είναι μικρότερο ή ίσο του $100u-$οστού ποσοστημορίου της F, καθώς το $u \to 0$. Δηλαδή:

$$\lambda_L = \lim_{u \to 0^+} P\left(Y \le G^{(-1)}(u) \mid X \le F^{(-1)}(u)\right)$$

με την προϋπόθεση ότι το όριο λ_L υπάρχει.

Τώρα εάν υποθέσουμε ότι C είναι το Copula των X, Y ,ισχύει ότι:

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u,u)}{u} = \lim_{u \to 0^+} P[U_1 \le u \mid U_2 \le u]$$

Τότε εάν το $\lambda_L \in (0,1]$, θα λέμε ότι το Copula C παρουσιάζει κάτω εξάρτηση ουράς, ενώ αν το όριο $\lambda_L = 0$, τότε θα λέμε ότι το Copula C παρουσιάζει κάτω ανεξαρτησία ουράς.
Συμπεραίνουμε λοιπόν, πως και οι δύο παραπάνω συντελεστές είναι μη παραμετρικοί εφόσον εξαρτώνται μόνο από το Copula των X, Y και όχι από τις περιθώριες κατανομές.

ΚΕΦΑΛΑΙΟ 4

ΕΙΔΙΚΕΣ ΟΙΚΟΓΕΝΕΙΕΣ ΤΩΝ COPULAS

4.1 Εισαγωγή

Όπως είδαμε και παραπάνω, στο θεώρημα του Sklar, τα Copulas αποτελούν σημαντικά εργαλεία για τη μελέτη πολυμεταβλητών κατανομών και τη δημιουργία των από κοινού συναρτήσεων κατανομής τους. Στο κεφάλαιο αυτό, θα παρουσιάσουμε δύο ειδικές οικογένειες των Copulas, που έχουν εφαρμογή στον αναλογισμό και στη διοικητική κινδύνου. Αυτές είναι: τα Archimedean Copulas (διμεταβλητή περίπτωση) και τα Compound Copulas.

4.2 Archimedean Copulas

Τα Archimedean Copulas είναι μία από τις πιο σημαντικές οικογένειες των Copulas, που έχουν εφαρμογή σε πολλούς τομείς και πιο συγκεκριμένα στην αναλογιστική επιστήμη και στη διοικητική κινδύνου. Για παράδειγμα, σε πολλές ασφαλιστικές εφαρμογές υπάρχει ισχυρή εξάρτηση μεταξύ μεγάλων ζημιών (πχ μεγάλο ποσό αποζημίωσης), που μπορεί να μοντελοποιηθεί από την οικογένεια αυτή. Έχουν συνήθως κλειστό τύπο και δεν προέρχονται από τις πολυμεταβλητές συναρτήσεις κατανομών που παίρνουμε χρησιμοποιώντας το θεώρημα του Sklar. Επίσης, παρουσιάζουν μεγάλο ενδιαφέρον και χρησιμοποιούνται ευρέως αφού είναι εύκολο να κατασκευαστούν και υπάρχουν αρκετές οικογένειες Copulas που προέρχονται από τα αυτά, κάποιες από τις οποίες θα δούμε παρακάτω.

Παρότι το Archimedean Copula είναι συνάρτηση πολλών μεταβλητών, θα εξετάζουμε κάθε φορά μία μονομεταβλητή συνάρτηση την οποία θα συμβολίζουμε με ϕ. Η κατασκευή τους λοιπόν, βασίζεται στη μονομεταβλητή συνάρτηση ϕ την οποία ονομάζουμε γεννήτρια συνάρτηση του Copula C_ϕ.

Στη συνέχεια, θα ορίσουμε την ψευδοαντίστροφη συνάρτηση της ϕ και έπειτα τη συνάρτηση του Archimedean Copula.

Ορισμός: Έστω ϕ μία συνεχής και γνησίως φθίνουσα συνάρτηση, όπου $\phi:[0,1]\to[0,\infty]$ με $\phi(1)=0$. Τότε η ψευδοαντίστροφη συνάρτηση της ϕ είναι η συνάρτηση $\phi^{[-1]}:[0,\infty]\to[0,1]$ και ορίζεται ως εξής:

$$\phi^{[-1]}(t)=\begin{cases}\phi^{-1}, & 0\le t\le\phi(0)\\ 0, & \phi(0)\le t\le\infty\end{cases}$$

Σημειώνουμε εδώ ότι η συνάρτηση $\phi^{[-1]}$ είναι συνεχής και φθίνουσα στο διάστημα $[0,\infty]$ και γνησίως φθίνουσα στο $[0,\phi(0)]$. Επιπλέον, για τη συνάρτηση $\phi^{[-1]}$ ισχύουν τα παρακάτω:

1. $\phi^{[-1]}\big(\phi(u)\big)=u$, στο διάστημα $[0,1]$.
2. $\phi\big(\phi^{[-1]}(t)\big)=\begin{cases}t, & 0\le t\le\phi(0)\\ \phi(0), & \phi(0)\le t\le\infty\end{cases}=\min\big(t,\phi(0)\big)$.
3. Αν $\phi(0)=\infty$ τότε έχουμε ότι $\phi^{[-1]}=\phi^{-1}$.

Θεώρημα: Έστω ϕ μία συνεχής και γνησίως φθίνουσα συνάρτηση, όπου $\phi:[0,1]\to[0,\infty]$, ώστε $\phi(1)=0$ και $\phi^{[-1]}$ η ψευδοαντίστροφη της ϕ. Τότε, η συνάρτηση $C:[0,1]^2\to[0,1]$ που δίνεται από τον τύπο:

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

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

Η απόδειξη του θεωρήματος αυτού βρίσκεται στο βιβλίο του Nelsen και το χρησιμοποιούμε για να ορίσουμε τα Archimedean Copulas.

Επομένως, τα Copulas που έχουν την παραπάνω μορφή ονομάζονται Archimedean Copulas και η συνάρτηση ϕ καλείται γεννήτορας τους. Αν $\phi(0)=\infty$, τότε έχουμε Archimedean Copula με αυστηρό γεννήτορα ϕ. Στην περίπτωση αυτή, $\phi^{[-1]}=\phi^{-1}$, και το Copula $C(u,v)=\phi^{-1}\big(\phi(u)+\phi(v)\big)$ θα ονομάζεται αυστηρό Archimedean Copula.

Παράδειγμα: Έστω οι συναρτήσεις $\phi(t) = -\log t$ και $\phi(t) = 1 - t$ για κάθε $t \in [0,1]$. Θα αποδείξουμε πως τα Archimedean Copulas είναι το Product Copula (Copula ανεξαρτησίας) και το κάτω φράγμα Fréchet-Hoeffding, αντίστοιχα.
Αρχικά, για τη συνάρτηση $\phi(t) = -\log t$, η ψευδοαντίστροφη συνάρτηση είναι:

$$\phi^{[-1]}(t) = \phi^{-1}(t) = e^{-t}$$

και ισχύει ότι $\phi(0) = \infty$, άρα έχουμε αυστηρό γεννήτορα ϕ. Το Copula τώρα που παράγεται από αυτόν είναι:

$$C(u,v) = \phi^{[-1]}\big(\phi(u) + \phi(v)\big) = e^{[-(-\log u - \log v)]} = uv = C^{\perp}(u,v).$$

Τώρα, για τη συνάρτηση $\phi(t) = 1 - t$, η ψευδοαντίστροφη συνάρτηση είναι:

$$\phi^{[-1]}(t) = \begin{cases} 1-t, & 0 \le t \le 1 \\ 0, & t > 1 \end{cases} = \max(1-t, 0)$$

Έτσι το Copula που παράγεται είναι:

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

Παρακάτω, θα δώσουμε δύο θεωρήματα που μπορούν να χρησιμοποιηθούν για να δούμε αν το Copula πλησιάζει κάποιο από τα φράγματα, στην περίπτωση που η παράμετρος του τείνει προς κάποιο άκρο του εύρους τους.

Θεώρημα 1: Έστω C_α είναι ένα Archimedean Copula με παράμετρο $\alpha \in \mathrm{A}$ και γεννήτορα ϕ_α που ανήκει στο σύνολο όλων των γεννητόρων Φ και είναι διαφορίσιμος. Τότε το $C = \lim C_\alpha$ είναι ένα Archimedean Copula αν και μόνο αν υπάρχει συνάρτηση $\phi \in \Phi$ τέτοια ώστε για όλα τα $s, t \in (0,1)$ να ισχύει:

$$\lim \frac{\phi_\alpha(s)}{\phi'_\alpha(t)} = \frac{\phi(s)}{\phi'(t)},$$

όπου το όριο αυτό, είναι το κατάλληλο μονόπλευρο όριο καθώς το α τείνει σε κάποιο από τα δύο άκρα του A.

Θεώρημα 2: Έστω C_α είναι ένα Archimedean Copula με παράμετρο $\alpha \in \mathrm{A}$ και γεννήτορα ϕ_α που ανήκει στο σύνολο όλων των γεννητόρων Φ και είναι διαφορίσιμος. Τότε το $\lim C_\alpha(u,v) = C^+(u,v)$ αν και μόνο αν ισχύει:

$$\lim \frac{\phi_\alpha(t)}{\phi'_\alpha(t)} = 0 \text{, για } t \in (0,1),$$

όπου το όριο αυτό, είναι το κατάλληλο μονόπλευρο όριο καθώς το α τείνει σε κάποιο από τα δύο άκρα του A.

Θα κλείσουμε την ενότητα αυτή δίνοντας δύο πορίσματα, που θα χρησιμοποιήσουμε παρακάτω, σχετικά με τους συντελεστές συσχέτισης για τα Archimedean Copulas, που έχουν αποδειχθεί.

Πόρισμα 1: Έστω X,Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής ένα Archimedean Copula C, που παράγεται από την $\phi \in \Phi$. Τότε ο συντελεστής συσχέτισης tau του Kendall, τ, για τα X,Y δίνεται από τον τύπο:

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

Πόρισμα 2: Έστω X,Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής ένα Archimedean Copula C, που παράγεται από την $\phi \in \Phi$. Τότε οι συντελεστές εξάρτησης της άνω και της κάτω ουράς δίνονται από τους παρακάτω τύπους:

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

και

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

Στη συνέχεια θα αναφερθούμε στις τρείς πιο συνήθεις οικογένειες των Archimedean Copulas και στις ιδιότητες τους.

4.2.1 Οικογένεια Gumbel

Η οικογένεια Gumbel, γνωστή και ως Gumbel-Hougaard, παίρνει τη μορφή:

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

όπου $\alpha \in [1,\infty)$. Ο γεννήτορας ϕ_α της οικογένειας Gumbel, είναι μια συνεχής και γνησίως φθίνουσα συνάρτηση από $[0,1] \to [0,\infty]$ με $\phi(1) = 0$, η οποία ορίζεται ως εξής:

$$\phi_\alpha(t) = (-\ln t)^a.$$

Ο γεννήτορας αυτός είναι αυστηρός, αφού $\phi_\alpha(0) = \infty$. Επιπλέον έχουμε ότι: $C_1 = C^\perp$ και $C_\infty = C^+$. Αυτό είναι εύκολο να αποδειχθεί, χρησιμοποιώντας τα παραπάνω θεωρήματα.
Πιο συγκεκριμένα, για το Copula ανεξαρτησίας γνωρίζουμε ότι : $\phi(t) = -\ln t$.
Επομένως,

$$\frac{\phi(s)}{\phi'(t)} = \frac{-\ln s}{-\frac{1}{t}} = t\ln s$$

και

$$\lim_{\alpha\to 1^+} \frac{\phi_\alpha(s)}{\phi'_\alpha(t)} = \lim_{\alpha\to 1^+} \frac{(-\ln s)^\alpha}{-\frac{\alpha}{t}(-\ln t)^{\alpha-1}} = \lim_{\alpha\to 1^+} \frac{t(-\ln s)^\alpha}{-\alpha(-\ln t)^{\alpha-1}} = t\ln s$$

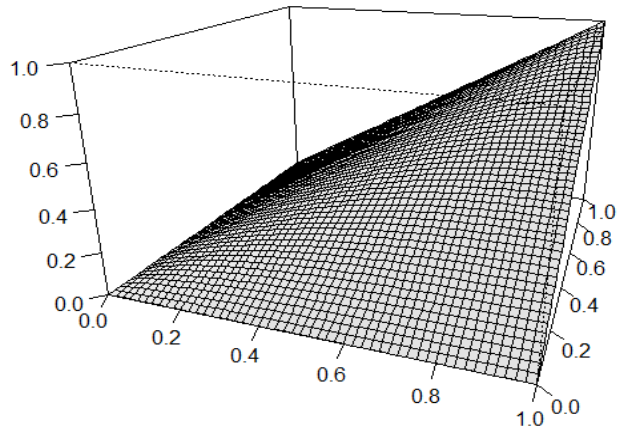
Εφόσον, $\lim_{\alpha \to 1^+} \frac{\phi_\alpha(s)}{\phi'_\alpha(t)} = \frac{\phi(s)}{\phi'(t)} = t \ln s$ τότε $C_1 = C^\perp$.

Αντίστοιχα, για το $C_\infty = C^+$ έχουμε:

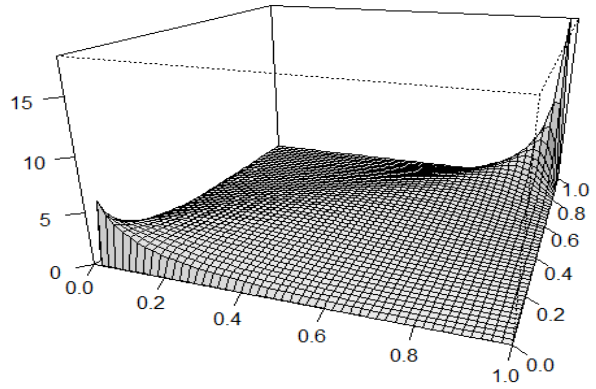
$$\lim_{\alpha \to \infty} \frac{\phi_\alpha(t)}{\phi'_\alpha(t)} = \lim_{\alpha \to \infty} \frac{-t - \ln t}{\alpha} = 0.$$

Στη συνέχεια θα δούμε το διάγραμμα της συνάρτησης κατανομής και της συνάρτησης πυκνότητας του Gumbel Copula για $\alpha = 2$.

Σχήμα 4.1: Διάγραμμα της σ.κ του Gumbel Copula για $\alpha = 2$.

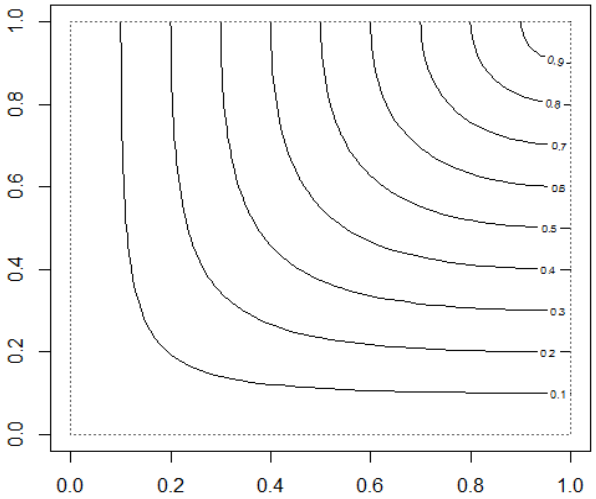


Σχήμα 4.2: Διάγραμμα της σ.π του Gumbel Copula για $\alpha = 2$.



Επίσης, το Contour Diagram της συνάρτησης κατανομής του Gumbel Copula με παράμετρο $\alpha = 2$, είναι:

Σχήμα 4.3: Contour Diagram της σ.κ του Gumbel Copula για $\alpha = 2$.



Οι συντελεστές εξάρτησης-συσχέτισης, συνοψίζουν πληροφορίες για το Copula σχετικά με την εξάρτηση ή συσχέτιση που υπάρχει μεταξύ των τυχαίων μεταβλητών. Στη συγκεκριμένη περίπτωση είναι φανερό πως το Gumbel Copula, λόγω του περιορισμένου παραμετρικού χώρου ($\alpha \geq 1$), επιτρέπει μόνο μη αρνητικές συσχετίσεις. Παράλληλα, παρουσιάζει ισχυρή εξάρτηση στην άνω ουρά σε αντίθεση με την κάτω που είναι αδύναμη.

Είναι ενδιαφέρον επομένως να παρουσιάσουμε τους συντελεστές συσχέτισης-εξάρτησης, για το συγκεκριμένο Copula. Κάνοντας χρήση λοιπόν των παραπάνω πορισμάτων, έχουμε τα εξής:

Αρχικά, για το συντελεστή συσχέτισης Tau του Kendall, τ , για το Gumbel Copula ισχύει ότι:

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt = 1 + 4\int_0^1 \frac{t \ln t}{\alpha} dt = 1 + \frac{4}{\alpha}\left([\frac{t^2}{2} \ln t]_0^1 - [\frac{t^2}{4}]_0^1 \right) = 1 - \frac{1}{\alpha} = \frac{\alpha - 1}{\alpha}$$

Ο γεννήτορας $\phi_\alpha(t) = (-\ln t)^a$ της οικογένειας Gumbel είναι αυστηρός, επομένως η αντίστροφη συνάρτηση του είναι: $\phi_\alpha^{-1}(t) = \exp(-t^{\frac{1}{\alpha}})$.

Για το συντελεστή εξάρτησης άνω ουράς έχουμε:

$$\lambda_U = 2 - \lim_{s \to 0^+} \frac{1 - \phi^{[-1]}(2s)}{1 - \phi^{[-1]}(s)} = 2 - \lim_{s \to 0^+} \frac{1 - \exp(-(2s)^{\frac{1}{\alpha}})}{1 - \exp((-s^{\frac{1}{\alpha}}))} \overset{(\frac{0}{0})}{=} 2 - \lim_{s \to 0^+} \frac{[1 - \exp(-(2s)^{\frac{1}{\alpha}})]'}{[1 - \exp((-s^{\frac{1}{\alpha}}))]'} = 2 - 2^{\frac{1}{\alpha}}$$

ενώ για το συντελεστή εξάρτησης κάτω ουράς έχουμε:

$$\lambda_L = \lim_{s \to \infty} \frac{\phi^{[-1]}(2s)}{\phi^{[-1]}(s)} = \lim_{s \to \infty} \frac{\exp(-(2s)^{\frac{1}{\alpha}})}{\exp((-s^{\frac{1}{\alpha}}))} \overset{(\frac{0}{0})}{=} \lim_{s \to \infty} \frac{[\exp(-(2s)^{\frac{1}{\alpha}})]'}{[\exp((-s^{\frac{1}{\alpha}}))]'} = \lim_{s \to \infty} 2\exp((-s^{\frac{1}{\alpha}}))2^{\frac{1}{\alpha}-1} = 0.$$

4.2.2 Οικογένεια Clayton

Η οικογένεια Clayton, γνωστή και ως Clayton,Cook-Johnson, παίρνει τη μορφή:

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

όπου $\alpha \in [-1,\infty) \setminus \{0\}$. Ο γεννήτορας ϕ_α της οικογένειας Clayton, είναι:

$$\phi_\alpha(t) = \frac{t^{-\alpha} - 1}{\alpha}.$$

Ωστόσο, το Clayton Copula χρησιμοποιείται συνήθως για $\alpha > 0$. Τότε, ο παραπάνω γεννήτορας είναι αυστηρός και η οικογένεια Clayton παίρνει τη μορφή:

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

Επιπλέον έχουμε ότι: $C_{-1} = C^-$, $C_\infty = C^+$ και $C_0 = C^\perp$. Για το άνω και κάτω όριο η απόδειξη είναι απλή, καθώς:

Για το κάτω φράγμα γνωρίζουμε ότι: $\phi(t) = 1 - t$.

Επομένως,

$$\frac{\phi(s)}{\phi'(t)} = \frac{1-s}{-1} = s - 1$$

και

$$\lim_{\alpha \to -1^+} \frac{\phi_\alpha(s)}{\phi'_\alpha(t)} = \lim_{\alpha \to -1^+} \frac{\frac{s^{-\alpha} - 1}{\alpha}}{-s^{-(\alpha+1)}} = \lim_{\alpha \to -1^+} -\frac{s^{\alpha+1}}{\alpha}(s^{-\alpha} - 1) = s - 1$$

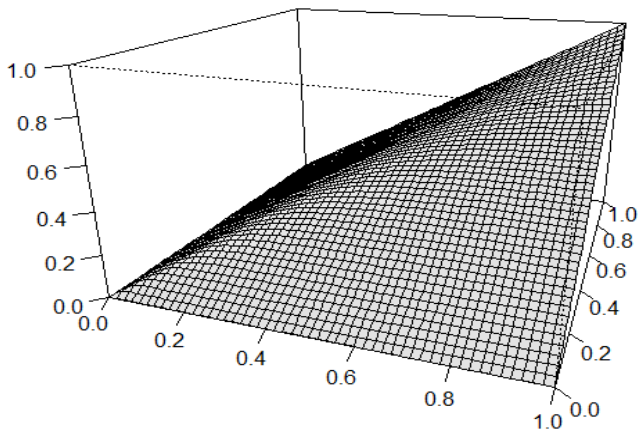
Εφόσον, $\lim_{\alpha \to -1^+} \frac{\phi_\alpha(s)}{\phi'_\alpha(t)} = \frac{\phi(s)}{\phi'(t)} = s-1$ τότε $C_{-1} = C^-$.

Αντίστοιχα, για το $C_\infty = C^+$ έχουμε:

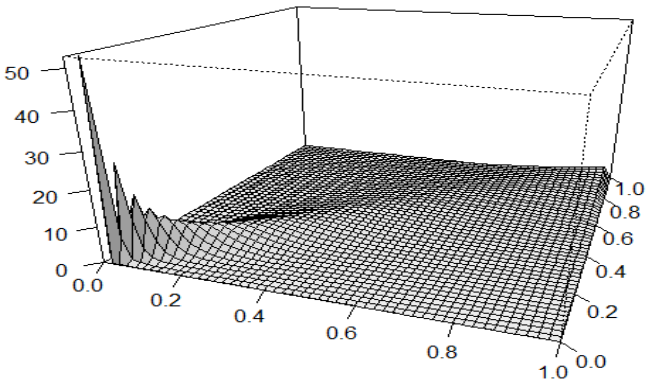
$$\lim_{\alpha \to \infty} \frac{\phi_\alpha(t)}{\phi'_\alpha(t)} = \lim_{\alpha \to \infty} \frac{\frac{t^{-\alpha}-1}{\alpha}}{-t^{-(\alpha+1)}} = -\frac{t^{\alpha+1}}{\alpha}(t^{-\alpha}-1) = -\frac{t-t^{\alpha+1}}{\alpha} = 0.$$

Παρακάτω, θα δούμε το διάγραμμα της συνάρτησης κατανομής και της συνάρτησης πυκνότητας του Clayton Copula για $\alpha = 4$.

Σχήμα 4.4: Διάγραμμα της σ.κ του Clayton Copula για $\alpha = 4$.

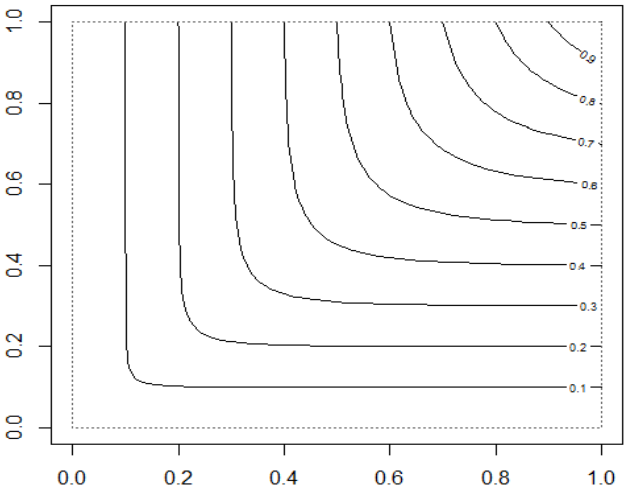


Σχήμα 4.5: Διάγραμμα της σ.π του Clayton Copula για $\alpha = 4$.



Το Contour Diagram της συνάρτησης κατανομής του Clayton Copula με παράμετρο $\alpha = 4$ είναι το παρακάτω.

Σχήμα 4.6: Contour Diagram της σ.κ του Clayton Copula για $\alpha = 4$.



Όπως και το Gumbel Copula, έτσι και το Clayton Copula, λόγω του περιορισμένου παραμετρικού χώρου, επιτρέπει μόνο μη αρνητικές συσχετίσεις. Παράλληλα, παρουσιάζει ισχυρή εξάρτηση κάτω ουράς και σχετικά αδύναμη εξάρτηση άνω ουράς, όπως θα δούμε παρακάτω. Για το λόγο αυτό έχει χρησιμοποιηθεί για τη μελέτη συσχετισμένων κινδύνων, όπως για παράδειγμα το σύνδρομο ‘‘broken heart’’, όπου οι ηλικίες θανάτου των παντρεμένων ζευγαριών συνήθως σχετίζονται.

Αρχικά, ο συντελεστής συσχέτισης Tau του Kendall, τ , για το Clayton Copula είναι:

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt = 1 + 4\int_0^1 -\frac{t - t^{\alpha+1}}{\alpha}dt = 1 + \frac{4}{\alpha}\int_0^1 (t^{\alpha+1} - t)dt = 1 + \frac{4}{\alpha}\left([\frac{t^{\alpha+2}}{\alpha+2}]_0^1 - [\frac{t^2}{2}]_0^1\right) = \frac{\alpha}{\alpha+2}$$

Για $\alpha > 0$, ο γεννήτορας $\phi_\alpha(t) = \frac{t^{-\alpha} - 1}{\alpha}$ είναι αυστηρός, και η οικογένεια Clayton παίρνει τη μορφή: $C_\alpha(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}}$. Τότε $\phi_\alpha^{-1}(t) = (1 + \alpha t)^{-\frac{1}{\alpha}}$.

Επομένως, για το συντελεστή εξάρτησης άνω ουράς έχουμε:

$$\lambda_U = 2 - \lim_{s\to 0^+} \frac{1 - \phi^{[-1]}(2s)}{1 - \phi^{[-1]}(s)} = 2 - \lim_{s\to 0^+} \frac{1 - (1+2\alpha s)^{-\frac{1}{\alpha}}}{1 - (1+\alpha s)^{-\frac{1}{\alpha}}} \overset{(\frac{0}{0})}{=} 2 - \lim_{s\to 0^+} \frac{[1 - (1+2\alpha s)^{-\frac{1}{\alpha}}]'}{[1 - (1+\alpha s)^{-\frac{1}{\alpha}}]'} = 2 - 2 = 0$$

ενώ για το συντελεστή εξάρτησης κάτω ουράς έχουμε:

$$\lambda_L = \lim_{s\to\infty} \frac{\phi^{[-1]}(2s)}{\phi^{[-1]}(s)} = \lim_{s\to\infty} \frac{(1+2\alpha s)^{-\frac{1}{\alpha}}}{(1+\alpha s)^{-\frac{1}{\alpha}}} \overset{(\frac{\infty}{\infty})}{=} \lim_{s\to\infty} \frac{(1+2\alpha s)^{-\frac{1}{\alpha}}}{(1+\alpha s)^{-\frac{1}{\alpha}}} = 2^{-\frac{1}{\alpha}}.$$

4.2.3 Οικογένεια Frank

Η οικογένεια Frank παίρνει τη μορφή:

$$C_\alpha(u,v) = -\frac{1}{\alpha}\ln(1+\frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1}),$$

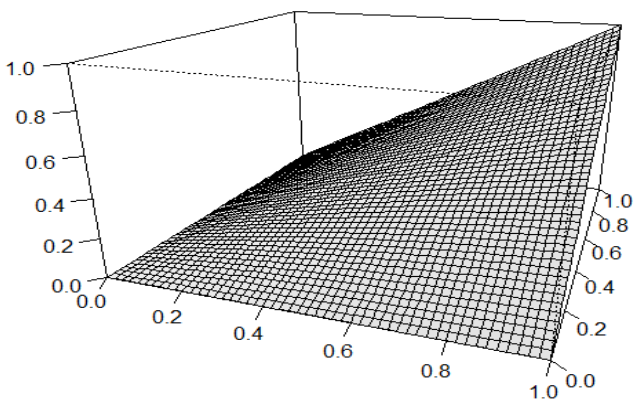
όπου $\alpha \in R\setminus\{0\}$. Ο αυστηρός γεννήτορας, ϕ_α, της οικογένειας Frank είναι:

$$\phi_\alpha(t) = -\ln\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}$$

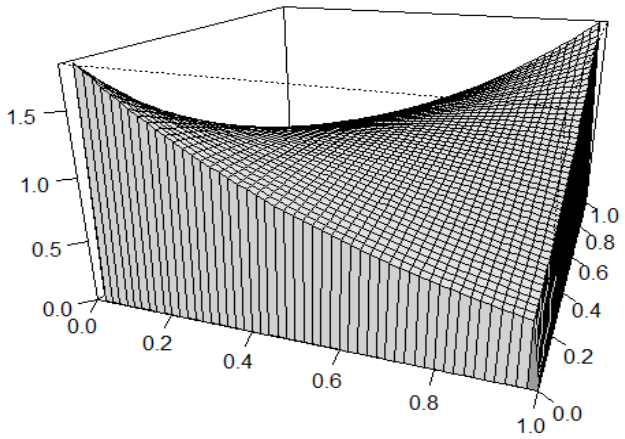
Το Frank Copula είναι ένα συμμετρικό Copula, που ικανοποιεί την ακτινική συμμετρία, $C = \hat{C}$. Επιπλέον έχουμε ότι: $C_{-\infty} = C^-$, $C_\infty = C^+$ και $C_0 = C^\perp$.

Ακολουθούν τα διαγράμματα της συνάρτησης κατανομής και συνάρτησης πυκνότητας του Frank Copula, καθώς επίσης και το Contour Diagram της συνάρτησης κατανομής του με παράμετρο $\alpha = \frac{3}{2}$.

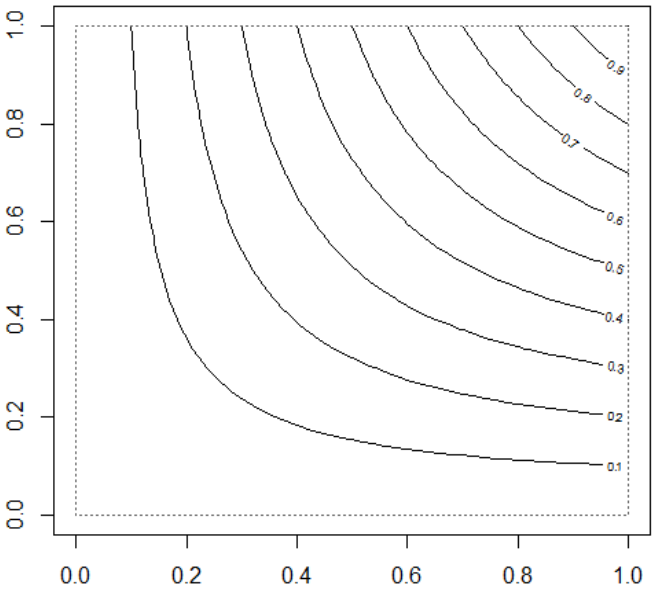
Σχήμα 4.7: Διάγραμμα της σ.κ του Frank Copula για $\alpha = \frac{3}{2}$.



Σχήμα 4.8: Διάγραμμα της σ.π του Frank Copula για $\alpha = \frac{3}{2}$.



Σχήμα 4.9: Contour Diagram της σ.κ του Frank Copula για $\alpha = \frac{3}{2}$.



Σε αντίθεση με το Gumbel και Clayton Copula, το Frank Copula επιτρέπει και αρνητική συσχέτιση, εφόσον ο συντελεστής μπορεί να πάρει οποιαδήποτε τιμή στο $(-\infty,0)\cup(0,\infty)$. Κατά συνέπεια, μπορεί θεωρητικά να χρησιμοποιηθεί για τη μοντελοποίηση αποτελεσμάτων με ισχυρή θετικά ή αρνητική συσχέτιση. Ωστόσο, όπως θα δούμε παρακάτω, η εξάρτηση των ουρών του είναι σχετικά ασθενής και η ισχυρότερη συσχέτιση είναι στο κέντρο της κατανομής, γεγονός που υποδηλώνει ότι το Frank Copula είναι το πλέον κατάλληλο για δεδομένα που εμφανίζουν ασθενή εξάρτηση ουράς.

Ο συντελεστής συσχέτισης Tau του Kendall, τ, παίρνει τη μορφή:

$$\tau = 1 - \frac{4}{\alpha}(1 - D_1(\alpha)),$$

όπου $D_1(a)$ είναι η συνάρτηση Debye, που έχει γενικό τύπο:

$$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^k - 1}dt, \text{ για } k = 1,2$$

Επίσης, για τη συνάρτηση Debye ισχύει:

$$D_k(-x) = D_k(x) + \frac{kx}{k+1}.$$

Ο γεννήτορας $\phi_\alpha(t) = -\ln\frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$ της οικογένειας Frank είναι αυστηρός, επομένως η αντίστροφη συνάρτηση του είναι: $\phi_\alpha^{-1}(t) = -\frac{1}{\alpha}\ln[1 - e^{-t}(1 - e^{-\alpha})]$.

Για το συντελεστή εξάρτησης άνω ουράς έχουμε:

$$\lambda_U = 2 - \lim_{s\to 0^+}\frac{1-\phi^{[-1]}(2s)}{1-\phi^{[-1]}(s)} = 2 - \lim_{s\to 0^+}\frac{1+\frac{1}{\alpha}\ln[1-e^{-2s}(1-e^{-\alpha})]}{1+\frac{1}{\alpha}\ln[1-e^{-s}(1-e^{-\alpha})]}\left(\tfrac{0}{0}\right) = 2 - \lim_{s\to 0^+}\frac{[1+\frac{1}{\alpha}\ln[1-e^{-2s}(1-e^{-\alpha})]]'}{[1+\frac{1}{\alpha}\ln[1-e^{-s}(1-e^{-\alpha})]]'} = 0$$

ενώ για το συντελεστή εξάρτησης κάτω ουράς έχουμε:

$$\lambda_L = \lim_{s\to\infty} \frac{-\frac{1}{\alpha}\ln[1-e^{-2s}(1-e^{-\alpha})]}{-\frac{1}{\alpha}\ln[1-e^{-s}(1-e^{-\alpha})]} \overset{(\frac{0}{0})}{=} \lim_{s\to\infty} \frac{[-\frac{1}{\alpha}\ln[1-e^{-2s}(1-e^{-\alpha})]]'}{[-\frac{1}{\alpha}\ln[1-e^{-s}(1-e^{-\alpha})]]'} = \lim_{s\to\infty} \frac{-2e^{-s}(1-e^{-s}(1-e^{-\alpha}))}{(1-e^{-2s}(1-e^{-\alpha}))} = 0.$$

Στη συνέχεια, ακολουθούν δύο πίνακες, που περιέχουν συνοπτικά τα αποτελέσματα της ανάλυσης των Archimedean Copulas.

Πίνακας 1

Archimedean Copulas και οι γεννήτορές τους

Οικογένεια	Γεννήτορας $\phi_\alpha(t)$	Παραμετρικός Χώρος (α)	Copula $C_\alpha(u,v)$
Gumbel	$(-\ln t)^a$	$\alpha \in [1,\infty)$	$\exp(-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{\frac{1}{\alpha}})$
Clayton	$\frac{t^{-\alpha}-1}{\alpha}$	$\alpha \in [-1,\infty)\setminus\{0\}$	$\max([u^{-\alpha}+v^{-\alpha}-1]^{-\frac{1}{\alpha}},0)$
Frank	$-\ln\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}$	$\alpha \in R\setminus\{0\}$	$-\frac{1}{\alpha}\ln(1+\frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1})$

Πίνακας 2

Archimedean Copulas και τα μέτρα εξάρτησής-συσχέτισης τους

Οικογένεια	Copula $C_\alpha(u,v)$	Kendall's τ	λ_U	λ_L
Gumbel	$\exp(-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{\frac{1}{\alpha}})$	$\frac{\alpha-1}{\alpha}$	$2-2^{\frac{1}{\alpha}}$	0
Clayton	$\max([u^{-\alpha}+v^{-\alpha}-1]^{-\frac{1}{\alpha}},0)$	$\frac{\alpha}{\alpha+2}$	0	$2^{-\frac{1}{\alpha}}$
Frank	$-\frac{1}{\alpha}\ln(1+\frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1})$	$1-\frac{4}{\alpha}(1-D_1(\alpha))$ όπου, $D_k(x)=\frac{k}{x^k}\int_0^x \frac{t^k}{e^k-1}dt$	0	0

4.3 Compound Copulas

Τα Compound Copulas προέρχονται, σύμφωνα με τους Marshall και Olkin (1988), από μια γενική μέθοδο κατασκευής Copulas, τη λεγόμενη μέθοδο σύνθεσης-μίξης (compounding method). Είναι εύκολο να καταλάβουμε τη χρησιμότητα της μεθόδου αυτής, αν σκεφτούμε πως οι σύνθετες-μεμιγμένες κατανομές (compound distributions), χρησιμοποιούνται εκτεταμένα στην κατηγοριοποίηση των κινδύνων (risk classification) και ιδιαίτερα στη θεωρία αξιοπιστίας χαρτοφυλακίου. Επίσης, οι Marshall και Olkin (1988) έδειξαν πως με τη μέθοδο της σύνθεσης-μίξης (compounding method), μπορούν να δημιουργηθούν αρκετές σημαντικές οικογένειες Copulas.

Στη συνέχεια, θα δώσουμε ένα γενικό ορισμό για τις σύνθετες-μεμιγμένες κατανομές.

Ορισμός: Για κάθε παραμετρική οικογένεια κατανομών $F(\cdot|\theta)$, μπορούμε να θεωρήσουμε την παράμετρο θ ως την τιμή μίας τυχαίας μεταβλητής Θ με κατανομή G. Τότε η $F(\cdot|\theta)$ είναι μια δεσμευμένη κατανομή δοθέντος $\Theta = \theta$ και η αντίστοιχη μη δεσμευμένη κατανομή:

$$H(x) = \int F(x|\theta)dG(\theta)$$

είναι μια σύνθετη-μεμιγμένη κατανομή (compound distribution).
Αντίστοιχα, για ένα δείγμα $X_1, X_2, ..., X_n$ ισχύει:

$$H(\underset{\sim}{x}) = \int \Pi F_i(x_i|\theta)dG(\theta),$$

όπου η G και κάθε F_i είναι μια μονομεταβλητή συνάρτηση κατανομής.

Στο παράδειγμα που ακολουθεί, θα δούμε πως οι σύνθετες-μεμιγμένες κατανομές μπορούν να χρησιμοποιηθούν για την παραγωγή Copulas υψηλού ενδιαφέροντος.

Παράδειγμα: Έστω ότι έχουμε ένα χαρτοφυλάκιο X ζημιών και το γ αποτελεί την κατηγοριοποίηση κινδύνου. Τότε, δοθέντος της παραμέτρου γ, η τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή με παράμετρο γ:

$$P(X \le x|\gamma) = 1 - e^{-\gamma x}.$$

Εάν η παράμετρος γ ακολουθεί την κατανομή Γάμμα με παραμέτρους (α, λ), τότε:

$$\begin{aligned}
F(x) &= \int_0^\infty P(X \le x|\gamma)\frac{\lambda^\alpha}{\Gamma(\alpha)}\gamma^{\alpha-1}e^{-\lambda\gamma}\,d\gamma \\
&= \int_0^\infty (1-e^{-\gamma x})\frac{\lambda^\alpha}{\Gamma(\alpha)}\gamma^{\alpha-1}e^{-\lambda\gamma}\,d\gamma \\
&= 1-\int_0^\infty e^{-\gamma x}\frac{\lambda^\alpha}{\Gamma(\alpha)}\gamma^{\alpha-1}e^{-\lambda\gamma}\,d\gamma \\
&= 1-(1+\frac{x}{\lambda})^{-\alpha}
\end{aligned}$$

Επομένως, η $F(x)$ ακολουθεί την κατανομή Pareto με παραμέτρους (α,λ). Αντίστοιχα, η περιθώρια κατανομή της είναι η $f(x)=\dfrac{\alpha\lambda^{\alpha}}{(\lambda+x)^{\alpha+1}}$.

Αν υποθέσουμε τώρα ότι, δοθέντος της κατηγορίας κινδύνου γ, τα X_1, X_2 είναι ανεξάρτητα και ισόνομα κατανεμημένα και προέρχονται από την ίδια κατηγορία κινδύνου. Τότε είναι φανερό πως το γ προκαλεί εξάρτηση. Έτσι, η από κοινού συνάρτηση κατανομής είναι:

$$\begin{aligned}
F(x_1,x_2) &= 1-P(X_1>x_1)-P(X_2>x_2)+P(X_1>x_1,X_2>x_2)\\
&= 1-(1+\frac{x_1}{\lambda})^{-\alpha}-(1+\frac{x_2}{\lambda})^{-\alpha}+(1+\frac{x_1+x_2}{\lambda})^{-\alpha}\\
&= F_1(x_1)+F_2(x_2)-1+\left[(1-F_1(x_1))^{-\frac{1}{\alpha}}+(1-F_2(x_2))^{-\frac{1}{\alpha}}-1\right]^{-\alpha}
\end{aligned}$$

Επομένως, για $u=F_1(x_1)$ και $v=F_2(x_2)$, μπορούμε να ορίσουμε τη συνάρτηση του Copula ως:

$$C(u,v)=u+v-1+\left[(1-u)^{-\frac{1}{\alpha}}+(1-v)^{-\frac{1}{\alpha}}-1\right]^{-\alpha}$$

και να εκφράσουμε τη διμεταβλητή συνάρτηση κατανομής ως:

$$H(x_1,x_2)=C(F_1(x_1),F_2(x_2)).$$

Αντίστοιχα, αν αντί για συνάρτηση κατανομής είχαμε συνάρτηση επιβίωσης με θετικές τυχαίες μεταβλητές χρόνων ζωής, τότε θα μπορούσαμε να θεωρήσουμε το Copula επιβίωσης:

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

Και να εκφράσουμε την από κοινού συνάρτηση επιβίωσης:

$$\overline{C}(u,v)=P(U_1>u,U_2>v)=\hat{C}(1-u,1-v).$$

Το παράδειγμα αυτό, αποτελεί μια ειδική περίπτωση μιας γενικής μεθόδου κατασκευής Copulas, σύμφωνα με τους Marshall και Olkin, την οποία θα αναλύσουμε στη συνέχεια.

4.3.1 Μέθοδος των Marshall και Olkin (1988)

Είναι μία μέθοδος κατασκευής Copulas, που περιλαμβάνει το μετασχηματισμό Laplace και την αντίστροφη συνάρτηση του. Στο σημείο αυτό, ας υπενθυμίσουμε ότι ο μετασχηματισμός Laplace της θετικής τυχαίας μεταβλητής γ ορίζεται ως:

$$\tau(s) = E_{\gamma}(e^{-s\gamma}) = \int_0^{\infty} e^{-st} dF_{\gamma}(t)$$

όπου F_{γ} είναι η συνάρτηση κατανομής της γ. Επίσης, είναι η ροπογεννήτρια συνάρτηση (moment generating function, mgf), που εκτιμάται στο $-s$. Γνωρίζοντας λοιπόν την $\tau(s)$, μπορούμε να προσδιορίσουμε την κατανομή.
Η μέθοδος των Marshall και Olkin για την κατασκευή Copulas, περιγράφεται ως εξής:

Έστω X_i μια τυχαία μεταβλητή. Δοθέντος μίας θετικής λανθάνουσας μεταβλητής γ_i, η δεσμευμένη συνάρτηση κατανομής της τυχαίας μεταβλητής, ορίζεται ως εξής:

$$H_i(x \mid \gamma_i) = H_i(x)^{\gamma_i}$$

όπου $H_i(\cdot)$ είναι μια βασική συνάρτηση κατανομής για $i = 1,2,...,n$.

Οι Marshall και Olkin, θεώρησαν πολυμεταβλητές συναρτήσεις κατανομής, της μορφής:

$$F(x_1, x_2,..., x_n) = E\left[K(H_1(x_1)^{\gamma_1}, H_2(x_2)^{\gamma_2},..., H_n(x_n)^{\gamma_n}\right]$$

όπου K είναι μια συνάρτηση κατανομής με ομοιόμορφες περιθώριες και η μέση τιμή υπολογίζεται για τα $\gamma_1, \gamma_2,..., \gamma_n$.

Ας θεωρήσουμε τώρα, πως όλες οι λανθάνουσες μεταβλητές είναι ίσες μεταξύ τους. Τότε, χρησιμοποιώντας τη συνάρτηση κατανομής που αντιστοιχεί σε ανεξάρτητες περιθώριες, οι Marshall και Olkin (1988) έδειξαν ότι:

$$F(x_1, x_2, ..., x_n) = E(H_1(x_1)^{\gamma} ... H_n(x_n)^{\gamma}) = \tau(\tau^{-1}(F_1(x_1)) + ... + \tau^{-1}(F_n(x_n)))$$

όπου F_i είναι η $i-$οστή περιθώρια κατανομή της F και $\tau(\cdot)$ ο μετασχηματισμός Laplace της γ.

Από την παραπάνω εξίσωση μπορούμε να διακρίνουμε πως η αντίστροφη συνάρτηση τ^{-1} χρησιμεύει ως γεννήτρια συνάρτηση για ένα Archimedean Copula. Αυτό έχει ως αποτέλεσμα, η πολυμεταβλητή συνάρτηση κατανομής να μπορεί να δοθεί από τη μορφή ενός τέτοιου Copula. Για να γίνει πιο κατανοητό αυτό, θα παρουσιάσουμε ένα θεώρημα που βρίσκεται στο βιβλίο του Joe (1997).

Θεώρημα: Έστω M μια μονομεταβλητή συνάρτηση κατανομής μιας θετικής τυχαίας μεταβλητής και τ, ο μετασχηματισμός Laplace της. Τότε, θεωρώντας μια μονομεταβλητή συνάρτηση κατανομής F, υπάρχει μια μοναδική συνάρτηση κατανομής G, τέτοια ώστε:

$$F(x) = \int_0^{\infty} G^a(x) dM(a) = \tau(-\ln G(x))$$

έτσι ώστε: $G = \exp\{-\tau^{-1}(F)\}$.

Αντίστοιχα, υπάρχει παρόμοια σχέση για τις συναρτήσεις επιβίωσης:

$$\overline{F}(x) = \int_0^{\infty} \overline{H}^a(x) dM(a) = \tau(-\ln \overline{H}(x))$$

έτσι ώστε: $\overline{H} = \exp\{-\tau^{-1}(\overline{F})\}$.

Στη συνέχεια, αν θεωρήσουμε τη διμεταβλητή συνάρτηση κατανομής $F(F_1, F_2)$. Για $j = 1,2$ έχουμε $G_j = \exp\{-\tau^{-1}(F_j)\}$. Τότε θα ισχύει:

$$F(F_1, F_2) = \int_0^{\infty} G_1^a G_2^a dM(a) = \tau(-\ln G_1 - \ln G_2) = \tau(\tau^{-1}(F_1) + \tau^{-1}(F_2)).$$

Είναι φανερό λοιπόν πως η αντίστροφη συνάρτηση ενός μετασχηματισμού Laplace αποτελεί ένα σημαντικό τύπο γεννήτριας για τα Archimedean Copulas τα οποία είναι της μορφής:

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

Συνδέοντας τώρα το παραπάνω θεώρημα με τη μέθοδο των Marshall και Olkin, έχουμε:

$$\tau[-\ln H_i(x)] = E\exp\{-[-\ln H_i(x)]\gamma\} = F_i(x)$$

έτσι ώστε:

$$H_i(x) = \exp\{-\tau^{-1}[F_i(x)]\}.$$

Με τον τρόπο αυτό, χρησιμοποιώντας την περιθώρια κατανομή και την κατανομή της λανθάνουσας τυχαίας μεταβλητής, μπορούμε να προσδιορίσουμε τη βασική συνάρτηση κατανομής $H_i(x)$.

Στη συνέχεια, ακολουθεί ένας πίνακας, ο οποίος παρέχει τη συνάρτηση του μετασχηματισμού Laplace, για τους αντίστροφους γεννήτορες των Gumbel, Clayton και Frank Copulas.

Πίνακας 3

Archimedean Copulas και Μετασχηματισμός Laplace

Οικογένεια	Γεννήτορας $\phi_\alpha(t)$	Μετασχηματισμός Laplace $\tau(s) = \phi^{-1}(s)$	Παραμετρικός Χώρος (α)	Κατανομή μετασχηματισμού Laplace
Gumbel	$(-\ln t)^a$	$\exp(-s^{\frac{1}{\alpha}})$	$\alpha \in [1,\infty)$	Θετική σταθερά
Clayton	$t^{-\alpha} - 1$	$(1+s)^{-\frac{1}{\alpha}}$	$\alpha \in (1,\infty)$	Γάμμα
Frank	$-\ln\dfrac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$	$-\dfrac{1}{\alpha}\ln[1 + e^s(e^{-\alpha} - 1)]$	$\alpha \in R \setminus \{0\}$	Κατανομή Λογαριθμικής σειράς θετικών ακέραιων μεταβλητών

Γενικότερα, ο μετασχηματισμός Laplace έχει καλά ορισμένες αντίστροφες. Παραπάνω, ορίσαμε ως $\tau(s)$ το μετασχηματισμό Laplace μιας θετικής τυχαίας μεταβλητής γ. Επομένως, για το Gumbel Copula, η τυχαία μεταβλητή γ, είναι μια $1/\alpha$ θετική σταθερά, για το Clayton Copula, είναι μια $\text{Gamma}(1,1/\alpha)$ και για το Frank Copula είναι μια τυχαία μεταβλητή λογαριθμικής σειράς, που ορίζεται από όλους τους φυσικούς αριθμούς.

Σύμφωνα με τους Marshall και Olkin, η μέθοδος αυτή έχει εφαρμογή στις περιπτώσεις που έχουμε χρόνους ζωής. Αυτό μπορεί να γίνει, ορίζοντας τις συναρτήσεις κατανομής ως συναρτήσεις επιβίωσης. Για τα πολυμεταβλητά μοντέλα ευπάθειας (frailty models) λοιπόν, υποθέτουμε ότι n ζωές $T_1, T_2, ...T_n$, είναι ανεξάρτητες δοθέντος γ, με $P(T_i > t \mid \gamma) = B_i(t)^{\gamma}$. Τότε για την από κοινού πολυμεταβλητή συνάρτηση επιβίωσης ισχύει ότι:

$$P(T_1 > t_1, ..., T_n > t_n) = E\{B_1(t_1)...B_n(t_n)\}^{\gamma} = \tau(\tau^{-1}\{S_1(x_1)\} + ... + \tau^{-1}\{S_n(x_n)\}).$$

Ακόμα, ισχύει:

$$\tau[-\ln B_i(t)] = S_i(t) = 1 - F_i(t)$$

και

$$B_i(t) = \exp\{-\tau^{-1}[S_i(t)]\}.$$

ΚΕΦΑΛΑΙΟ 5

ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΠΡΟΣΟΜΟΙΩΣΗΣ

5.1 Εισαγωγή

Στο κεφάλαιο αυτό αρχικά θα παρουσιάσουμε τρόπους για την εκτίμηση των Copulas και πιο συγκεκριμένα θα δώσουμε την παραμετρική, ημιπαραμετρική και μη παραμετρική εκτίμηση. Θα συνεχίσουμε παρουσιάζοντας μια μέθοδο για την επιλογή του καλύτερου Archimedean Copula και στο τέλος θα δώσουμε αλγόριθμους για την παραγωγή δεδομένων από τα Archimedean Copulas και τα Compound Copulas.

5.2 Μέθοδοι εκτίμησης των Copulas

Στην ενότητα αυτή, θα παρουσιάσουμε τρείς βασικές μεθόδους εκτίμησης των Copulas, την παραμετρική την ημιπαραμετρική και τη μη παραμετρική μέθοδο. Στην παραμετρική μέθοδο εκτίμησης ανήκει η μέθοδος μεγίστης πιθανοφάνειας (Maximum Likelihood Method-ML) και η μέθοδος συμπερασματολογίας των συναρτήσεων των περιθώριων (Method of Inference Functions for Margins-IFM), στην ημιπαραμετρική μέθοδο εκτίμησης ανήκει η κανονική μέθοδος μεγίστης πιθανοφάνειας (Canonical Maximum Likelihood Method-CML) και τέλος στη μη παραμετρική μέθοδο εκτίμησης θα εισάγουμε το Εμπειρικό Copula.

5.2.1 Παραμετρική μέθοδος εκτίμησης ενός Copula

1. Μέθοδος Μεγίστης Πιθανοφάνειας

Έστω $X_1, X_2, ..., X_n$ συνεχείς τυχαίες μεταβλητές με περιθώριες συναρτήσεις κατανομής $F_1, F_2, ..., F_n$ και περιθώριες συναρτήσεις πυκνότητας πιθανότητας $f_1, f_2, ..., f_n$, αντίστοιχα. Τότε, η από κοινού πυκνότητα τους είναι:

$$f(x_1, x_2, ..., x_n) = c\left(F_1(x_1), F_2(x_2), ..., F_n(x_n)\right)\prod_{i=1}^{n} f_i(x_i),$$

όπου c είναι η πυκνότητα του Copula και ισούται με:

$$c(F_1(x_1), F_2(x_2), ..., F_n(x_n)) = \frac{\partial^n C(F_1(x_1), F_2(x_2), ..., F_n(x_n))}{\partial F_1(x_1)...\partial F_n(x_n)}.$$

Η λογαριθμική συνάρτηση πιθανοφάνειας (log-likelihood function) είναι:

$$L = \sum_{j=1}^{d} \log f(x_1^j, ..., x_n^j),$$

που αποτελείται από ένα δείγμα με d παρατηρούμενα τυχαία διανύσματα $\mathbf{x}$.

Στόχος μας είναι η εκτίμηση των παραμέτρων των n περιθώριων αλλά και του Copula. Έστω λοιπόν $a_1, a_2, ..., a_n$ οι παράμετροι των n περιθώριων και θ η παράμετρος του Copula. Τότε, η λογαριθμική συνάρτηση πιθανοφάνειας μπορεί να γραφεί ως:

$$L(a_1, ..., a_n, \theta) = L_C(a_1, ..., a_n, \theta) + \sum_{i=1}^{n} L_i(a_i),$$

όπου

$$L_C(a_1, ..., a_n, \theta) = \sum_{j=1}^{d} \log c\left(F_1(x_1^j; a_1), ..., F_n(x_n^j; a_n); \theta\right)$$

είναι η συνεισφορά στην πιθανοφάνεια της δομής εξάρτησης των δεδομένων από το Copula C και

$$L_i(a_i) = \sum_{j=1}^{d} \log f_i(x_i^j; a_i)$$

είναι η συνεισφορά στην πιθανοφάνεια από την κάθε περιθώρια.

Επομένως,

$$L(a_1,...,a_n,\theta)=\sum_{j=1}^{d}\log c\left(F_1(x_1^j;a_1),...,F_n(x_n^j;a_n);\theta\right)+\sum_{i=1}^{n}\sum_{j=1}^{d}\log f_i(x_i^j;a_i).$$

Ο εκτιμητής μέγιστης πιθανοφάνειας-ΕΜΠ (Maximum Likelihood Estimator-MLE)- $(\hat{a}_1,...,\hat{a}_n,\hat{\theta})$ των παραμέτρων $(a_1,...,a_n,\theta)$ είναι αυτός που μεγιστοποιεί τη λογαριθμική συνάρτηση πιθανοφάνειας. Έτσι, έχουμε:

$$(\hat{a}_1,...,\hat{a}_n,\hat{\theta})=\arg\max L(a_1,...,a_n,\theta).$$

Κάτω από συνθήκες ομαλότητας, το σημείο που μεγιστοποιείται η πιθανοφάνεια προκύπτει από τις λύσεις των παρακάτω εξισώσεων:

$$(\frac{\partial L}{\partial a_1},...,\frac{\partial L}{\partial a_n},\frac{\partial L}{\partial \theta})=0.$$

Επίσης, κάτω από συνθήκες ομαλότητας, για τον εκτιμητή μεγίστης πιθανοφάνειας ισχύει η ασυμπτωτική κανονικότητα και ο πίνακας διακυμάνσεων είναι ο αντίστροφος του πίνακα πληροφορίας του Fisher.

2. Μέθοδος συμπερασματολογίας των συναρτήσεων των περιθώριων

Η μέθοδος μεγίστης πιθανοφάνειας, που είδαμε παραπάνω, παρουσιάζει υπολογιστική δυσκολία, ειδικά σε περιπτώσεις πολλών διαστάσεων. Αυτό συμβαίνει διότι είναι απαραίτητο να εκτιμηθούν από κοινού οι παράμετροι των περιθώριων κατανομών και οι παράμετροι της δομής εξάρτησης μέσω του Copula. Για το λόγο αυτό έχει προταθεί μια άλλη μέθοδος, η μέθοδος συμπερασματολογίας των συναρτήσεων των περιθώριων, λεγόμενη ως IFM (Method of Inference Functions for Margins), η οποία αποτελείται από τα παρακάτω δύο βήματα:

1. Αρχικά εκτιμούμε τις παραμέτρους a_i, για $i=1,...,n$ των περιθώριων κατανομών F_i με τη μέθοδο μεγίστης πιθανοφάνειας:

$$\hat{a}_i^{IFM}=\arg\max_{a_i} L_i(a_i)=\arg\max_{a_i}\sum_{j=1}^{d}\log f_i(x_i^j;a_i)$$

2. και στη συνέχεια εκτιμούμε την παράμετρο θ του Copula C, δεδομένου των εκτιμήσεων των παραμέτρων των περιθώριων που υπολογίσαμε στο παραπάνω βήμα:

$$\hat{\theta}^{IFM} = \arg\max_{\theta} L_C(\hat{a}_1^{IFM},...,\hat{a}_n^{IFM},\theta) = \arg\max_{\theta} \sum_{j=1}^{d} \log c\left(F_1(x_1^j;\hat{a}_1^{IFM}),...,F_n(x_n^j;\hat{a}_n^{IFM});\theta\right).$$

Σύμφωνα με τα παραπάνω βήματα, ο IFM εκτιμητής $(\hat{a}_1^{IFM},...,\hat{a}_n^{IFM},\hat{\theta}^{IFM})$ προκύπτει κάτω από συνθήκες ομαλότητας από τις λύσεις των παρακάτω εξισώσεων:

$$(\frac{\partial L_1}{\partial a_1},...,\frac{\partial L_n}{\partial a_n},\frac{\partial L}{\partial \theta}) = 0.$$

Η μέθοδος IFM προτιμάται περισσότερο από αυτή του ML διότι υπάρχει μεγαλύτερη ευκολία στους υπολογισμούς. Αυτό συμβαίνει διότι οι παράμετροι των περιθώριων κατανομών εκτιμώνται ξεχωριστά από τις παραμέτρους του Copula. Παρόλα αυτά για την εκτίμηση του IFM εκτιμητή, χρησιμοποιούμε και την ML μέθοδο. Το κοινό τους σημείο είναι η ασυμπτωτική κανονικότητα. Ακριβώς όπως για τον εκτιμητή μεγίστης πιθανοφάνειας, έτσι και για τον IFM εκτιμητή, κάτω από συνθήκες ομαλότητας ισχύει η ασυμπτωτική κανονικότητα με τη μόνη διαφορά ότι ο πίνακας διακυμάνσεων είναι ο αντίστροφος του πίνακα Godambe.

5.2.2 Ημιπαραμετρική μέθοδος εκτίμησης ενός Copula

Στις παραπάνω παραμετρικές μεθόδους εκτίμησης, θεωρούσαμε γνωστές τις περιθώριες συναρτήσεις κατανομών αλλά και το Copula. Αυτό βέβαια δεν είναι πάντα εύκολο. Για τις περιπτώσεις λοιπόν, που δε γνωρίζουμε τις περιθώριες αλλά μόνο το Copula, χρησιμοποιούμε την κανονική μέθοδο μεγίστης πιθανοφάνειας (Canonical Maximum Likelihood Method), τη λεγόμενη CML μέθοδο, σύμφωνα με την οποία οι παράμετροι του Copula μπορούν να εκτιμηθούν χωρίς τον προσδιορισμό των περιθώριων.

Η μέθοδος αυτή αποτελείται από τα παρακάτω δύο βήματα:

1. Αρχικά, δεδομένου ότι οι περιθώριες F_i για $i = 1,...,n$ δεν είναι γνωστές, τις εκτιμούμε μη παραμετρικά χρησιμοποιώντας τις εμπειρικές συναρτήσεις κατανομών $\hat{F}_i$, οι οποίες έχουν γενικό τύπο:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_A(X_i \le x),$$

όπου $I_A(X_i \le x)$ είναι η δείκτρια συνάρτηση.

2. Έπειτα, εκτιμούμε με τη μέθοδο μεγίστης πιθανοφάνειας την παράμετρο θ του Copula. Έτσι για ένα δείγμα $(X_1^j,...,X_n^j)$, όπου $j=1,...,d$, ο εκτιμητής της παραμέτρου θ του Copula είναι:

$$\hat{\theta}^{CML} = \arg\max_{\theta} \sum_{j=1}^{d} \log c(\hat{F}_1(x_1^j),...,\hat{F}_n(x_n^j);\theta).$$

Αποδεικνύεται επίσης πως ο $\hat{\theta}^{CML}$, κάτω από συνθήκες ομαλότητας, είναι συνεπής και ασυμπτωτικά κανονικός εκτιμητής.

5.2.3 Μη παραμετρική μέθοδος εκτίμησης ενός Copula

Στις παραπάνω μεθόδους εκτίμησης, ένα κοινό σημείο είναι πως το Copula είναι γνωστό. Έτσι είναι και πιο εύκολη η εκτίμηση του. Στη μη παραμετρική μέθοδο εκτίμησης, θεωρούμε πως το Copula δεν είναι γνωστό όπως ακριβώς και οι περιθώριες κατανομές. Για το λόγο αυτό θα παρουσιάσουμε το εμπειρικό Copula (Empirical Copula), και θα το εκτιμήσουμε σύμφωνα με τα δεδομένα που θα έχουμε.

Το εμπειρικό Copula εισάχθηκε από τον Deheuvels το 1979, ο οποίος έδωσε και τον παρακάτω ορισμό:

Ορισμός: Έστω, $\{x_1^{(t)},...,x_n^{(t)}\}$ είναι τα order statistics και $\{r_1^t,...,r_n^t\}$ είναι τα rank statistics ενός δείγματος $(x_1,...,x_n)$, όπου $t=1,...,T$. Τότε, οποιοδήποτε Copula με πεδίο ορισμού: $L=\{(\frac{t_1}{T},...,\frac{t_N}{T}):1\le n\le N, t_n=0,...,T\}\}$ και συνάρτηση:

$$\hat{C}(\frac{t_1}{T},...,\frac{t_N}{T}) = \frac{1}{T}\sum_{t=1}^{T}\prod_{n=1}^{N} I(r_n^t \le t_n)$$

είναι ένα εμπειρικό Copula.

Η πυκνότητα ενός εμπειρικού Copula δίνεται από τον τύπο:

$$\hat{c}(\frac{t_1}{T},...,\frac{t_N}{T}) = \sum_{i_1=1}^{2}...\sum_{i_n=1}^{2}(-1)^{i_1+...+i_N}\hat{C}(\frac{t_1-i_1+1}{T},...,\frac{t_N-i_N+1}{T}).$$

5.3 Αξιολόγηση και επιλογή του κατάλληλου Archimedean Copula

Σε εμπειρικές εφαρμογές των Copulas, αυτό που ουσιαστικά χρειάζεται να ξέρουμε είναι πόσο καλά ταιριάζει το μοντέλο με τα διαθέσιμα δεδομένα. Επίσης είναι ιδιαίτερα χρήσιμο να γνωρίζουμε αν υπάρχουν δείκτες που δεν έχουν προσδιοριστεί ακριβώς. Για παράδειγμα, υπάρχει περίπτωση τα δεδομένα να εμφανίζουν τον τύπο της εξάρτησης ουράς, που το επιλεγμένο Copula δε μπορεί να εντοπίσει. Ακόμη, στην περίπτωση που δοκιμαστούν διάφορα Copulas, θα πρέπει να είμαστε στη θέση να αναγνωρίσομε ποιο ταιριάζει καλύτερα στα δεδομένα μας.

5.3.1 Κριτήρια επιλογής

Διαφορετικές συναρτήσεις Copulas εμφανίζουν διαφορετικές μορφές εξάρτησης. Ωστόσο, αν ένας ερευνητής θέλει να διερευνήσει τη δομή εξάρτησης, μπορεί να υπολογίσει διαφορετικά Copulas και να επιλέξει αυτό που ταιριάζει καλύτερα στα δεδομένα. Το πρώτο βήμα για την εκτίμηση του Copula είναι ο προσδιορισμός και η εκτίμηση των μονομεταβλητών περιθώριων κατανομών. Το δεύτερο βήμα, που ενδεχομένως είναι πιο δύσκολο, απαιτεί τις προδιαγραφές του Copula.

Η πρώτη προσέγγιση για την επιλογή του μοντέλου οφείλεται στους Genest και Rivest (1993), οι οποίοι δημιούργησαν μια μέθοδο η οποία περιγράφει τον τρόπο επιλογής του Copula που ταιριάζει καλύτερα στα δεδομένα μας.

Υποθέτουμε λοιπόν, ότι διαθέτουμε ένα τυχαίο δείγμα διμεταβλητών παρατηρήσεων (X_{1i}, X_{2i}) για $i = 1,...,n$, των οποίων η συνάρτηση κατανομής F αντιστοιχεί σε ένα Archimedean Copula C_ϕ. Στόχος μας είναι ο προσδιορισμός της γεννήτριας συνάρτησης ϕ.

Για το λόγο αυτό, οι Genest και Rivest, εισήγαγαν μια νέα τυχαία μεταβλητή $Z_i = F(X_{1i}, X_{2i})$, που έχει συνάρτηση κατανομής $K(z) = P(Z_i \le z)$.

Αυτή η συνάρτηση κατανομής, σχετίζεται με τη γεννήτρια συνάρτηση ενός Archimedean Copula σύμφωνα με τον παρακάτω τύπο:

$$K(z) = z - \frac{\phi(z)}{\phi'(z)}.$$

Για τον προσδιορισμός της γεννήτριας συνάρτησης ϕ, ακολουθούμε τα παρακάτω βήματα:

1. Εκτιμούμε το συντελεστή συσχέτισης του Kendall

$$\tau_n = \binom{n}{2}^{-1} \sum_{i<j} sign[(X_{1i} - X_{1j})(X_{2i} - X_{2j})].$$

2. Κατασκευάζουμε μια μη παραμετρική εκτίμηση της συνάρτησης K, η οποία δίνεται από τον παρακάτω τύπο:

$$\hat{K}(z) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i \le z),$$

όπου $Z_i = \frac{1}{n-1} \sum_{i=1}^{n} I(x_{1j} < x_{1i}, x_{2j} < x_{2i})$ για $i = 1,...,n$

3. Κατασκευάζουμε μια παραμετρική εκτίμηση της συνάρτησης K χρησιμοποιώντας τη σχέση:

$$K_\phi(z) = z - \frac{\phi(z)}{\phi'(z)}.$$

Οι γεννήτριες συναρτήσεις πολύ γνωστών Archimedean Copulas, βρίσκονται συγκεντρωμένες στον Πίνακα 1 του κεφαλαίου 3.

Επιλέγοντας λοιπόν μια συγκεκριμένη γεννήτρια συνάρτηση, υπολογίζουμε την εκτίμηση της παραμέτρου α, έστω $\hat{\alpha}$. Χρησιμοποιώντας αυτήν την εκτίμηση, μπορούμε να εκτιμήσουμε και την $\phi(z)$, έστω $\hat{\phi}(z)$ και τέλος να εκτιμήσουμε παραμετρικά και την $K_\phi(z)$, έστω $\hat{K}(z)$.

Στη συνέχεια, επαναλαμβάνουμε το τελευταίο βήμα για διάφορες επιλογές γεννήτριας συνάρτησης. Τέλος, συγκρίνοντας κάθε μια παραμετρική εκτίμηση με τη μη παραμετρική του βήματος 2, καταλήγουμε στην επιλογή του κατάλληλου Archimedean Copula που ταιριάζει καλύτερα στα δεδομένα μας.

Ουσιαστικά, στόχος μας είναι η επιλογή εκείνης την γεννήτριας συνάρτησης όπου η παραμετρική εκτίμηση πλησιάζει περισσότερο τη μη παραμετρική. Αυτό, μπορούμε να το δούμε είτε γραφικά χρησιμοποιώντας ένα Q-Q plot μεταξύ της συνάρτησης $K_{\phi}(z)$ και της εκτίμησης της $\hat{K}(z)$, είτε ελαχιστοποιώντας την απόσταση μεταξύ τους.

5.4 Αλγόριθμοι Προσομοίωσης

Οι αναλογιστές, συνήθως ασχολούνται με πολύπλοκες και μη γραμμικές συναρτήσεις, όπως είναι οι παρούσες αξίες τυχαίων μεταβλητών. Έτσι η προσομοίωση αποτελεί για αυτούς ένα ευρέως χρησιμοποιούμενο εργαλείο, που τους βοηθά στο να προσομοιώσουν τα δεδομένα σε κάποια κατανομή και στη συνέχεια να αναλύσουν τα αποτελέσματα αυτών. Για το λόγο αυτό, η κατασκευή των Copulas μας επιτρέπει να προσομοιώνουμε εύκολα τα αποτελέσματα μιας πολυμεταβλητής κατανομής. Στην ενότητα αυτή, θα παρουσιάσουμε δύο αλγόριθμους, έναν για τη παραγωγή δεδομένων από τα Archimedean Copulas και έναν για τη παραγωγή δεδομένων από τα Compound Copulas.

5.4.1 Αλγόριθμος παραγωγής δεδομένων από τα Archimedean Copulas

Η μέθοδος προσομοίωσης των Archimedean Copulas, προέρχεται από μια γενική μέθοδο προσομοίωσης ενός επιλεγμένου Copula, η οποία διαμορφώνεται χρησιμοποιώντας τη δεσμευμένη προσέγγιση. Στόχος μας είναι να κατασκευάσουμε έναν αλγόριθμο, με σκοπό να παράγουμε n ανεξάρτητες τυχαίες μεταβλητές $(U_1, U_2, ..., U_n) \in [0,1]$ και έχουν από κοινού συνάρτηση κατανομής C.

Έτσι, οι Genest και Rivest (1986) και Genest (1987), εισήγαγαν την ιδέα της προσομοίωσης των συναρτήσεων κατανομής των $(U_1, U_2, ..., U_n)$, προσομοιώνοντας τη δεσμευμένη κατανομή U_k δοθέντος των $U_1, U_2, ..., U_{k-1}$ για $k = 2, ..., n$.

Επομένως, στη δισδιάστατη περίπτωση έχουμε:

$$C(u_1 \mid u_2) = P(U_1 \le u_1 \mid U_2 = u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2}$$

Και για $k = 2$,

$$C_2(u_2 \mid u_1) = \frac{\phi^{-1(1)}\left(\phi(u_1) + \phi(u_2)\right)}{\phi^{-1(1)}\left(\phi(u_1)\right)}$$

Εφόσον ισχύει γενικά ότι:

$$C_k(u_k \mid u_1, ..., u_{k-1}) = \frac{\phi^{-1(k-1)}\left(\phi(u_1) + \phi(u_2) + ... + \phi(u_k)\right)}{\phi^{-1(k-1)}\left(\phi(u_1) + \phi(u_2) + ... + \phi(u_{k-1})\right)}.$$

Ο αλγόριθμος που θα χρησιμοποιήσουμε είναι ο εξής:

1. Παράγουμε δύο ανεξάρτητες τυχαίες μεταβλητές $(v_1, v_2) \sim U(0,1)$.
2. Θέτουμε $u_1 = v_1$.
3. Υπολογίζουμε το $v_2 = C_2(u_2 \mid u_1)$.
4. Υπολογίζουμε το $u_2 = C_2^{-1}(v_2 \mid v_1)$.
5. Παίρνουμε τις (u_1, u_2) τυχαίες μεταβλητές.

Παρακάτω, ακολουθούν οι αλγόριθμοι προσομοίωσης του Clayton και Frank Copula, για τη διμεταβλητή περίπτωση.

a. Clayton Copula:

Υπενθυμίζουμε ότι για το Clayton Copula ισχύουν τα παρακάτω:

$$\phi_\alpha(t) = t^{-\alpha} - 1 \text{ και } \phi_\alpha^{-1}(t) = (1+t)^{-\frac{1}{\alpha}}.$$

Επομένως,

$$\phi_{\alpha}^{-1(1)}(t) = -\frac{1}{\alpha}(1+t)^{-\frac{1}{\alpha}-1}.$$

Ο αλγόριθμος που παράγει (u_1, u_2) τυχαίες μεταβλητές από το Clayton Copula είναι:

1. Παράγουμε δύο ανεξάρτητες τυχαίες μεταβλητές $(v_1, v_2) \sim U(0,1)$.
2. Θέτουμε $u_1 = v_1$.
3. Υπολογίζουμε το $v_2 = C_2(u_2 \mid u_1) = \dfrac{\phi^{-1(1)}\left(\phi(u_1) + \phi(u_2)\right)}{\phi^{-1(1)}(\phi(u_1))}$.

   έτσι, $v_2 = \left(\dfrac{u_1^{-\alpha} + u_2^{-\alpha} - 1}{u_1^{-\alpha}}\right)^{-\frac{1}{a}-1}$.
4. Τελικά, $u_2 = \left[v_1^{\alpha}(v_2^{-\frac{\alpha}{\alpha+1}} - 1) + 1\right]^{-\frac{1}{\alpha}}$.

b. Frank Copula:

Υπενθυμίζουμε ότι για το Frank Copula ισχύουν τα παρακάτω:

$$\phi_{\alpha}(t) = \ln\frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1} \text{ και } \phi_{\alpha}^{-1}(t) = -\frac{1}{\alpha}\ln[1 + e^{t}(e^{-\alpha} - 1)]$$

Επομένως,

$$\phi^{-1(1)}(t) = -\frac{1}{\alpha}\frac{e^{t}(e^{-\alpha} - 1)}{1 + e^{t}(e^{-\alpha} - 1)}.$$

Ο αλγόριθμος που παράγει (u_1, u_2) τυχαίες μεταβλητές από το Frank Copula είναι:

1. Παράγουμε δύο ανεξάρτητες τυχαίες μεταβλητές $(v_1, v_2) \sim U(0,1)$.
2. Θέτουμε $u_1 = v_1$.
3. Υπολογίζουμε το $v_2 = C_2(u_2 \mid u_1) = \dfrac{\phi^{-1(1)}\left(\phi(u_1) + \phi(u_2)\right)}{\phi^{-1(1)}(\phi(u_1))}$.

έτσι, $v_2 = \dfrac{e^{-\alpha u_1}(e^{-\alpha u_2}-1)}{(e^{-\alpha}-1)+(e^{-\alpha u_1}-1)(e^{-\alpha u_2}-1)}$.

4. Τελικά, $u_2 = -\dfrac{1}{\alpha}\ln\left(1+\dfrac{v_2(1-e^{-\alpha})}{v_2(e^{-\alpha v_1}-1)-e^{-\alpha v_1})}\right)$.

Για το Gumbel Copula, ο υπολογισμός του $v_2 = C_2(u_2 \mid u_1)$ δεν είναι εύκολος, αφού απαιτεί μια επαναληπτική λύση. Για το λόγο αυτό, θα παρουσιάσουμε έναν εναλλακτικό αλγόριθμο, που προτάθηκε από τους Marshall και Olkin (1988), για τις σύνθετες μεθόδους κατασκευής των Copulas.

5.4.2 Αλγόριθμος παραγωγής δεδομένων από τα Compound Copulas

Για να παράγουμε $X_1, X_2, ..., X_n$ τυχαίες μεταβλητές, έχοντας γνωστές συναρτήσεις κατανομής,

$$F(x_1, x_2, ..., x_n) = E_\gamma(H_1(x_1)^\gamma ... H_n(x_n)^\gamma) = \tau(\tau^{-1}(F_1(x_1)) + ... + \tau^{-1}(F_n(x_n))),$$

οι Frees και Valdez (1998) πρότειναν τον ακόλουθο αλγόριθμο:

1. Παράγουμε μια λανθάνουσα τυχαία μεταβλητή γ, η οποία έχει μετασχηματισμό Laplace τ.
2. Ανεξάρτητα από το προηγούμενο βήμα, παράγουμε $U_1, U_2, ..., U_n$ ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την $U(0,1)$.
3. Για $k = 1,2,...,n$ υπολογίζουμε το $X_k = F_k^{-1}(U_{*k})$,

   όπου $U_{*k} = \tau(-\dfrac{1}{\gamma}\ln U_k)$.
4. Παίρνουμε το X_k.

Ο αλγόριθμος αυτός, αν και είναι σχετικά απλός για τα περισσότερα Copulas που παράγονται από τη μέθοδο της σύνθεσης-μίξης (compounding method), απαιτεί την παραγωγή μιας επιπλέον τυχαίας μεταβλητής.

Αυτό είναι και το κύριο μειονέκτημα του, αφού υπολογιστικά είναι πιο απλός σε σύγκριση με τον αλγόριθμο παραγωγής δεδομένων από τα Archimedean Copulas.

Παρακάτω θα παρουσιάσουμε τους αλγόριθμους προσομοίωσης του Gumbel και Clayton Copula.

a. Gumbel Copula

Υπενθυμίζουμε ότι για το Gumbel Copula ισχύουν τα παρακάτω:

$$\phi_\alpha(t) = (-\ln t)^a \text{ και } \tau(s) = \phi^{-1}(s) = \exp(-s^{\frac{1}{\alpha}}).$$

Ο αλγόριθμος που παράγει $X_1, X_2, ..., X_n$ τυχαίες μεταβλητές από το Gumbel Copula είναι:

1. Παράγουμε μια λανθάνουσα τυχαία μεταβλητή γ, με παράμετρο $\frac{1}{\alpha}$.
2. Ανεξάρτητα από το προηγούμενο βήμα, παράγουμε $U_1, U_2, ..., U_n$ ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την $U(0,1)$.
3. Για $k = 1,2,...,n$ υπολογίζουμε το $X_k = F_k^{-1}(U_{*k})$,

   όπου $U_{*k} = \tau(-\frac{1}{\gamma}\ln U_k) = \exp\left\{-(-\frac{1}{\gamma}\ln U_k)^{\frac{1}{\alpha}}\right\}$.

b. Clayton Copula

Υπενθυμίζουμε ότι για το Clayton Copula ισχύουν τα παρακάτω:

$$\phi_\alpha(t) = t^{-\alpha} - 1 \text{ και } \tau(s) = \phi^{-1}(s) = (1+s)^{-\frac{1}{\alpha}},$$

ο οποίος είναι ο μετασχηματισμός Laplace της κατανομής $Gamma(1, \frac{1}{a})$.

Ο αλγόριθμος που παράγει $X_1, X_2, ..., X_n$ τυχαίες μεταβλητές από το Clayton Copula είναι:

1. Παράγουμε μια λανθάνουσα τυχαία μεταβλητή γ η οποία είναι η $Gamma(1, \frac{1}{\alpha})$.
2. Ανεξάρτητα από το προηγούμενο βήμα, παράγουμε $U_1, U_2, ..., U_n$ ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την $U(0,1)$.
3. Για $k = 1,2,...,n$ υπολογίζουμε το $X_k = F_k^{-1}(U_{*k})$,

   όπου $U_{*k} = \tau(-\frac{1}{\gamma}\ln U_k) = [1 + (-\frac{1}{\gamma}\ln U_k)]^{-\frac{1}{\alpha}}$.

Σε αντίθεση με το Gumbel Copula και το Clayton Copula που μπορούν να προσομοιωθούν εύκολα με το συγκεκριμένο αλγόριθμο, για το Frank Copula η προσομοίωση δεν είναι εύκολη διότι το γ είναι μια λογαριθμική σειρά θετικών ακέραιων μεταβλητών.

ΚΕΦΑΛΑΙΟ 6

ΕΦΑΡΜΟΓΗ ΤΩΝ COPULAS ΣΤΟΝ ΑΝΑΛΟΓΙΣΜΟ

6.1 Εισαγωγή

Στο κεφάλαιο αυτό, θα παρουσιάσουμε μια εφαρμογή, η οποία αφορά την προσαρμογή των Copulas στις απώλειες και τα έξοδα μιας ασφαλιστικής εταιρίας, καθώς επίσης και στην τιμολόγηση ενός συμβολαίου αντασφάλειας.

6.2 Εφαρμογή των Copulas με δεδομένα τις απώλειες-ζημιές και τα έξοδα μιας ασφαλιστικής εταιρίας

Στην ενότητα αυτή, θα περιγράψουμε μεθόδους προσαρμογής των Copulas σε αποζημιώσεις απαιτήσεων μιας ασφαλιστικής εταιρίας και θα περιγράψουμε τη χρησιμότητά τους, τιμολογώντας ένα συμβόλαιο αντασφάλειας και εκτιμώντας τις δαπάνες για προκαθορισμένες απώλειες. Ουσιαστικά, θα παρουσιάσουμε την εφαρμογή των Fees και Valdez (1998), οι οποίοι χρησιμοποίησαν 1500 γενικές απαιτήσεις αστικής ευθύνης που επιλέχθηκαν τυχαία και δόθηκαν από το Γραφείο Ασφαλιστικών Υπηρεσιών (Insurance Services Office). Όσον αναφορά τις συγκεκριμένες γενικές απαιτήσεις αστικής ευθύνης, η ασφαλιστική εταιρία έχει αναλάβει να προστατεύσει τον ασφαλισμένο της από ορισμένους κινδύνους που αναγράφονται στο ασφαλιστήριο συμβόλαιό του.

Η εφαρμογή μας περιλαμβάνει δύο τυχαίες μεταβλητές X_1, X_2. Στη μεταβλητή X_1 απεικονίζονται οι απώλειες (losses), δηλαδή οι αποζημιώσεις ενώ στη μεταβλητή X_2 απεικονίζονται τα έξοδα, τα οποία αναλογούν σε συγκεκριμένες απαιτήσεις (allocated loss adjustment expense-ALAE). Τέτοια έξοδα μπορεί να είναι αμοιβές δικηγόρων, εμπειρογνωμόνων και γενικότερα όλων των ερευνών που πραγματοποιούνται για να διερευνηθούν οι απαιτήσεις.

Βασικός μας στόχος είναι να περιγράψουμε την από κοινού κατανομή των απωλειών και των εξόδων. Το βασικότερο πρόβλημα που αντιμετωπίζουμε στην εκτίμηση της από κοινού συνάρτησης κατανομής των απωλειών και των εξόδων, είναι η πολυπλοκότητα της. Αυτό οφείλεται στην παρουσία λογοκρισίας (censoring), που είναι ένα σύνηθες χαρακτηριστικό των δεδομένων απώλειας. Στην περίπτωση που εξετάζουμε, για τις μεταβλητές X_1, X_2, έχουμε ένα όριο συμβολαίου (policy limit), το οποίο αντικατοπτρίζει το μέγιστο ύψος απαίτησης. Με την παρουσία λοιπόν του ορίου συμβολαίου, η μεταβλητή απώλειας X_1 λογοκρίνεται επειδή το ποσό της απαίτησης δεν μπορεί να υπερβαίνει το αναφερόμενο όριο συμβολαίου. Στις περιπτώσεις για τις οποίες το όριο συμβολαίου ήταν άγνωστο, υποθέσαμε ότι δεν υπήρχε κανένα όριο και τις συμπεριλάβαμε στην έρευνα.

Παρακάτω, ακολουθεί ένας συνοπτικός πίνακας με απλά περιγραφικά στατιστικά στοιχεία για τις απώλειες και τα έξοδα 1500 ασφαλιστηρίων συμβολαίων.

Πίνακας 4

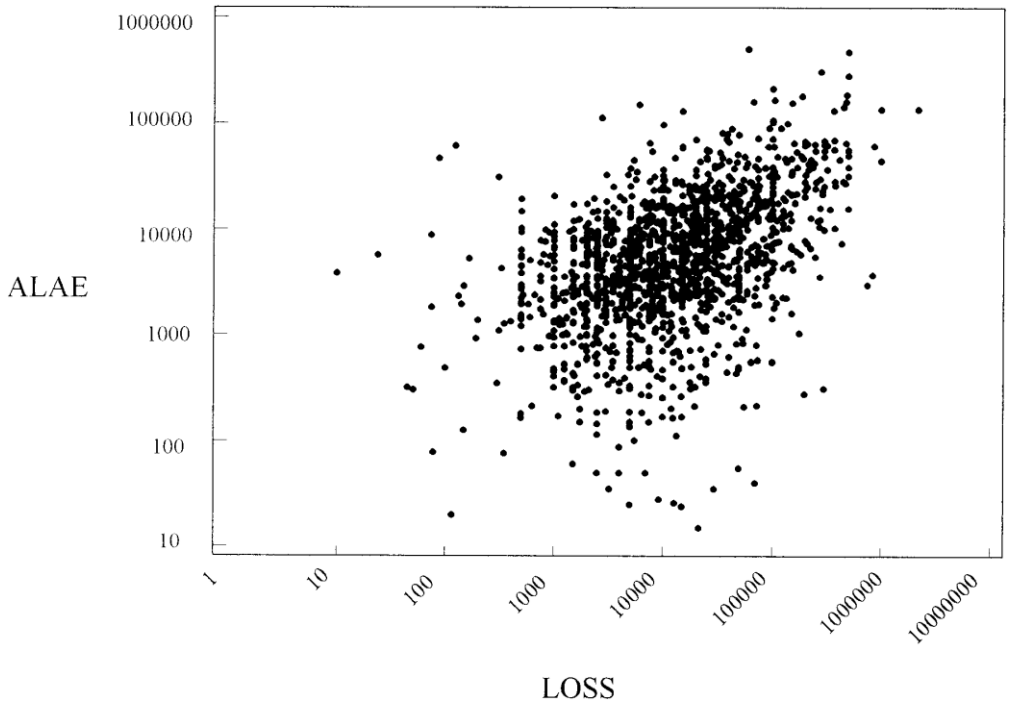
Συνοπτικός Πίνακας Περιγραφικών Στατιστικών Στοιχείων για τις Απώλειες και τα Έξοδα

	ALAE (X_2)	Loss (X_1)	Policy Limit	Loss (Uncensored)	Loss (Censored)
Number	1,500	1,500	1,352	1,466	34
Mean	12,588	41,208	559,098	37,110	217,491
Median	5,471	12,000	500,000	11,048	100,000
Standard Deviation	28,146	102,748	418,649	92,513	258,205
Minimum	15	10	5,000	10	5,000
Maximum	501,863	2,173,595	7,500,000	2,173,595	1,000,000
25th Percentile	2,333	4,000	300,000	3,750	50,000
75th percentile	12,577	35,000	1,000,000	32,000	300,000

Χρησιμοποιώντας τα δεδομένα των μεταβλητών X_1, X_2, μπορούμε εύκολα να υπολογίσουμε τον συντελεστή συσχέτισής τους, ο οποίος ισούται με 0.41. Επίσης, βλέπουμε πως αν και μόνο 34 από τα 1500 συμβόλαια έχουν απαιτήσεις που είναι ίσες με το όριο του συμβολαίου και τα οποία θεωρούνται λογοκριμένα, η μέση απώλεια τους είναι 217,491, κάτι που σημαίνει ότι είναι πολύ υψηλότερη από τον αντίστοιχο μέσο των μη λογοκριμένων απαιτήσεων.

Έχοντας σαν δεδομένο τον ανωτέρω πίνακα, καθώς και το συντελεστή συσχέτισης των X_1, X_2, μια πρώτη εκτίμηση της εξάρτησης-συσχέτισης των δύο τυχαίων μεταβλητών, αποτελεί το διάγραμμα διασπορών (Scatterplot) που ακολουθεί.

Σχήμα 6.1: Scatterplot της απώλειας έναντι του ALAE.



Στο παραπάνω διάγραμμα διασποράς και οι δυο μεταβλητές βρίσκονται σε λογαριθμική κλίμακα, περιορίζοντας έτσι τις ακραίες τιμές. Το γράφημα αυτό λοιπόν, σε συνδυασμό με τον αντίστοιχο συντελεστή συσχέτισης, φανερώνει μια ισχυρή σχέση μεταξύ ζημιών και εξόδων.

Στη συνέχεια, ακολουθεί η διαδικασία προσαρμογής ενός Copula στις μεταβλητές X_1, X_2.

6.2.1 Προσαρμογή του Copula στα δεδομένα

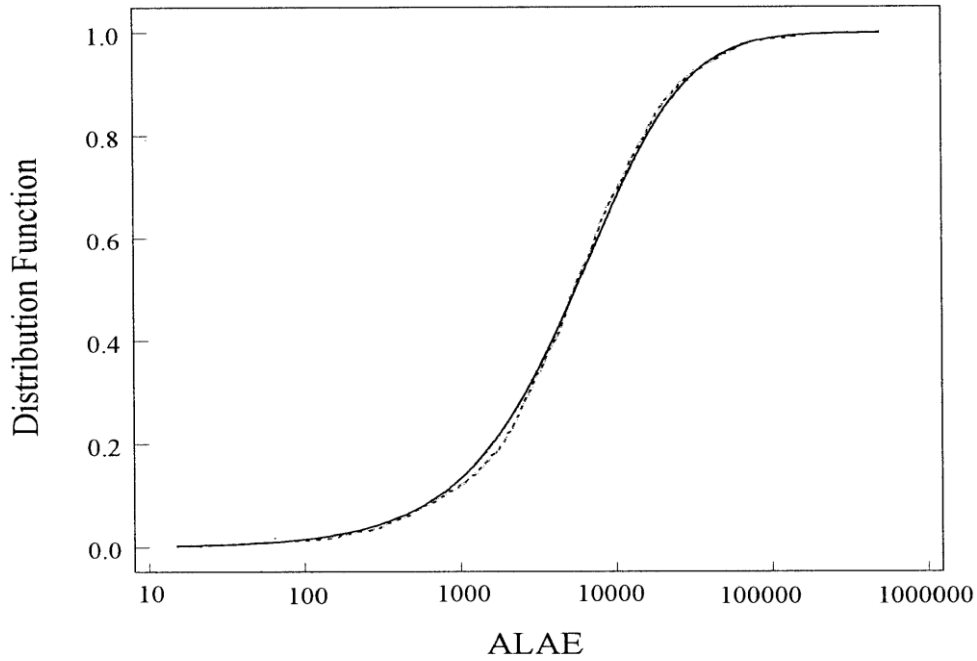
Προτού προχωρήσουμε στην προσαρμογή του μοντέλου μας, θα πρέπει να ορίσουμε τις κατάλληλες περιθώριες συναρτήσεις. Στην προκειμένη, για την προσαρμογή των μονομεταβλητών περιθωρίων, θα χρησιμοποιήσουμε την κατανομή Pareto με παραμέτρους λ και θ, με συνάρτηση κατανομής:

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x} \right)^{\theta}.$$

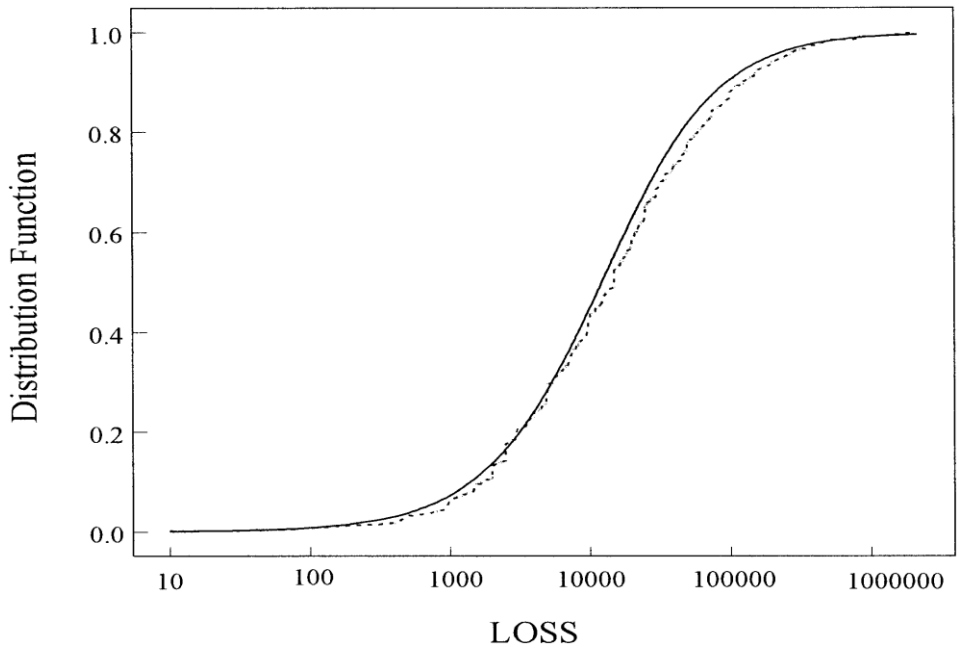
Η επιλογή της κατανομής αυτής, βασίζεται στη βαριά της ουρά. Ένας τρόπος για να ελέγξουμε την προσαρμογή των περιθώριων κατανομών, είναι η γραφική σύγκριση της προσαρμοσμένης συνάρτησης κατανομής έναντι της εμπειρικής συνάρτησης κατανομής τους.

Στα παρακάτω διαγράμματα λοιπόν, η διακεκομμένη καμπύλη απεικονίζει την εμπειρική συνάρτηση κατανομής και η συνεχόμενη καμπύλη απεικονίζει την προσαρμοσμένη συνάρτηση κατανομής χρησιμοποιώντας την κατανομή Pareto. Λόγω της λογοκρισίας, για τη μεταβλητή απώλειας (loss), χρησιμοποιήσαμε την εμπειρική συνάρτηση κατανομής Kaplan-Meier.

Σχήμα 6.2: Γράφημα Προσαρμοσμένης Συνάρτησης Κατανομής του ALAE.



Σχήμα 6.3: Γράφημα Προσαρμοσμένης Συνάρτησης Κατανομής της απώλειας.

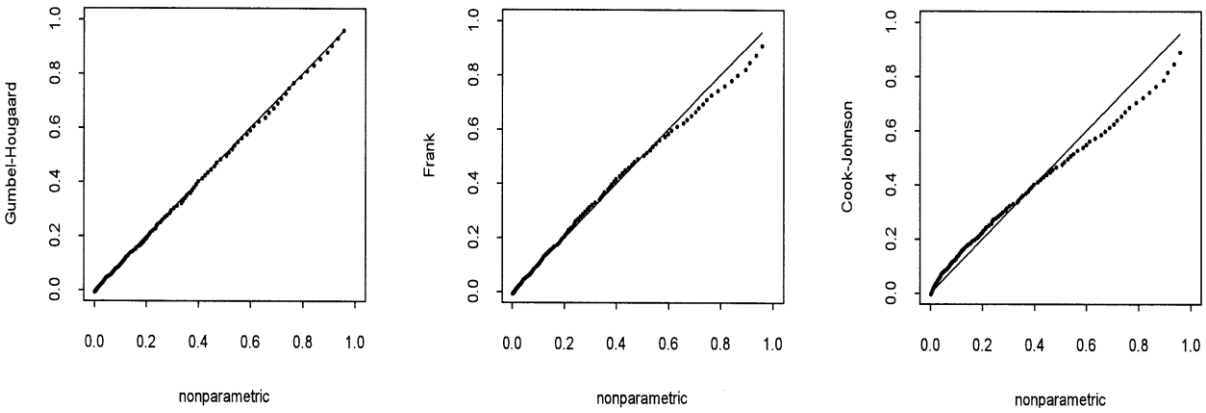


Στο σημείο αυτό, εφόσον ορίσαμε τις περιθώριες συναρτήσεις, μπορούμε να προσδιορίσουμε ποιο είναι το καταλληλότερο Copula για την εφαρμογή αυτή.

Για τον προσδιορισμό του καταλληλότερου Copula θα χρησιμοποιούμε τη μέθοδο των Genest και Rivest (1993), που αναπτύξαμε στην Ενότητα 5.3.1, η οποία περιγράφει τον τρόπο επιλογής του Archimedean Copula που ταιριάζει καλύτερα στα δεδομένα μας, αγνοώντας τις λογοκριμένες απαιτήσεις. Στόχος μας είναι η επιλογή εκείνης την γεννήτριας συνάρτησης για την οποία η παραμετρική εκτίμηση πλησιάζει περισσότερο τη μη παραμετρική. Όπως ήδη έχουμε προαναφέρει, μπορούμε να εργαστούμε γραφικά χρησιμοποιώντας ένα Q-Q plot για τη σύγκριση της συνάρτησης $K_{\phi}(z)$ με την εκτίμηση της $\hat{K}(z)$.

Στη συνέχεια, ακολουθούν τα Q-Q plots, χρησιμοποιώντας τρείς από τις κυριότερες κλάσεις των Archimedean Copulas, που αναλύσαμε στο Κεφάλαιο 4, οι οποίες δεν είναι άλλες από το Gumbel Copula, το Frank Copula και τέλος το Clayton Copula.

Σχήμα 6.4: Q-Q plots των παραμετρικών και μη παραμετρικών εκτιμήσεων της $K(z)$.



Σύμφωνα με τα παραπάνω Q-Q plots, το Copula που προσαρμόζεται καλύτερα στα δεδομένα μας είναι το Gumbel, αφού υπάρχει στενή συμφωνία μεταξύ των παραμετρικών και μη παραμετρικών ποσοστημορίων. Ουσιαστικά, οι διακεκομμένες γραμμές αντιστοιχούν στα ποσοστημόρια της παραμετρικής και μη παραμετρικής εκτίμησης του γεννήτορα ϕ, ενώ οι συνεχόμενες γραμμές αντιστοιχούν στην περίπτωση που τα ποσοστημόρια είναι ίσα. Στο σημείο αυτό, αξίζει να σημειωθεί πως τα ποσοστημόρια που βασίζονται στο Frank Copula είναι επίσης πολύ κοντά στα μη παραμετρικά ποσοστημόρια για μικρές τιμές απωλειών, ωστόσο υπάρχουν μεγάλες αποκλίσεις στα ποσοστημόρια, που αντιστοιχούν σε υψηλές απώλειες και έξοδα.

Εν συνεχεία, θα εκτιμήσουμε τις παραμέτρους του μοντέλου, χρησιμοποιώντας την μέθοδο μεγίστης πιθανοφάνειας. Εκτός από τις μεταβλητές X_1, X_2, οι οποίες έχουν περιθώριες $\mathrm{Pa}(\lambda_1, \theta_1)$ και $\mathrm{Pa}(\lambda_2, \theta_2)$ αντίστοιχα, θα εισάγουμε ένα δείκτη λογοκρισίας (δ), ο οποίος ορίζεται ως εξής:

- $\delta = 1$, αν η απαίτηση είναι λογοκριμένη,
- $\delta = 0$, αν η απαίτηση δεν είναι λογοκριμένη.

Κατά την ανάπτυξη της εξίσωσης πιθανοφάνειας, έγινε χρήση των ακόλουθων μερικών παραγώγων:

$$F_1(x_1, x_2) = \frac{\partial F(x_1, x_2)}{\partial x_1}, \; F_2(x_1, x_2) = \frac{\partial F(x_1, x_2)}{\partial x_2}$$

και

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}.$$

Ομοίως, οι πρώτες μερικές παράγωγοι για το Copula είναι:

$$C_1 = \frac{\partial C(F_1(x_1), F_2(x_2))}{\partial F_1(x_1)}, \; C_2 = \frac{\partial C(F_1(x_1), F_2(x_2))}{\partial F_2(x_2)}$$

ενώ η δεύτερη μικτή μερική παράγωγος είναι:

$$C_{12} = \frac{\partial^2 C(F_1(x_1), F_2(x_2))}{\partial F_1(x_1) F_2(x_2)}.$$

Συνολικά, έχουμε να εκτιμήσουμε ένα σύνολο πέντε παραμέτρων, δύο για κάθε περιθώρια και μία για την παράμετρο εξάρτησης του Copula. Τώρα, για να αναπτύξουμε την πιθανοφάνεια, θα πρέπει να κάνουμε μία διάκριση μεταξύ των λογοκριμένων και μη λογοκριμένων περιπτώσεων.

Στην περίπτωση που η μεταβλητή απώλειας είναι μη λογοκριμένη, δηλαδή $\delta = 0$, τότε η συνεισφορά στη συνάρτηση πιθανοφάνειας είναι:

$$f(x_1,x_2) = f_1(x_1) f_2(x_2) C_{12}\left[F_1(x_1), F_2(x_2)\right]. \qquad (4.1)$$

Ενώ, στην περίπτωση που η μεταβλητή απώλειας είναι λογοκριμένη, δηλαδή $\delta = 1$ η από κοινού πιθανότητα που μας ενδιαφέρει θα δίνεται από τον τύπο:

$$\Pr(X_1 \geq x_1, X_2 \leq x_2) = F_2(x_2) - F(x_1, x_2).$$

και συνεπώς, η αντίστοιχη συνεισφορά στην πιθανοφάνεια είναι ίση με:

$$f_2(x_2) - F_2(x_1, x_2) = f_2(x_2)\{1 - C_2[F_1(x_1), F_2(x_2)]\}. \qquad (4.2)$$

Σύμφωνα με τις σχέσεις (4.1) και (4.2), η λογαριθμική συνάρτηση πιθανοφάνειας είναι:

$$\log L(x_1, x_2, \delta) = (1-\delta)\log f(x_1, x_2) + \delta\{\log f_2(x_2) + \log[1 - C_2(F_1(x_1), F_2(x_2))]\}. \quad (4.3)$$

Οι εκτιμήσεις των παραμέτρων καθορίζονται μεγιστοποιώντας την παραπάνω λογαριθμική συνάρτηση πιθανοφάνειας για το σύνολο των δεδομένων:

$$\sum_{i=1}^{n} \log L(x_{1i}, x_{2i}, \delta).$$

Για τη μεγιστοποίηση θα μπορούσε να χρησιμοποιηθεί μια αριθμητική επαναληπτική μέθοδος, για παράδειγμα η διαδικασία IML του πακέτου SAS.

Στο σημείο αυτό, θα δώσουμε τις μερικές παραγώγους που χρησιμοποιήθηκαν για την εκτίμηση της πιθανοφάνειας στην εξίσωση (4.3), εφαρμόζοντας το Gumbel Copula, όπως αυτό δίνεται στον Πίνακα 1.

Αρχικά, εκφράζουμε εκ νέου το Gumbel Copula ως εξής:

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

Η μερική παράγωγος της παραπάνω ποσότητας, ως προς u είναι:

$$\frac{(-\ln C(u,v))^{\alpha-1}}{C(u,v)}\frac{\partial C(u,v)}{\partial u}=\frac{(-\ln u)^{\alpha-1}}{u}\Leftrightarrow$$

$$\frac{\partial C(u,v)}{\partial u}=\left(\frac{\ln u}{\ln C(u,v)}\right)^{\alpha-1}\frac{C(u,v)}{u}$$

και αντίστοιχα, η μερική παράγωγος ως προς v είναι:

$$\frac{\partial C(u,v)}{\partial v}=\left(\frac{\ln v}{\ln C(u,v)}\right)^{\alpha-1}\frac{C(u,v)}{v}.$$

Τέλος, η δεύτερη μερική παράγωγος του Gumbel Copula βρίσκεται ίση με:

$$\frac{\partial^2 C(u,v)}{\partial u\partial v}=\frac{1}{C(u,v)}\frac{\partial C(u,v)}{\partial u}\frac{\partial C(u,v)}{\partial v}\left[(\alpha-1)(-\ln C(u,v))^{-1}+1\right].$$

Στη συνέχεια, ακολουθεί ένας συνοπτικός πίνακας των αποτελεσμάτων της εκτίμησης μεγίστης πιθανοφάνειας, χρησιμοποιώντας το Gumbel Copula με περιθώριες της Pareto.

Πίνακας 5

Εκτιμήσεις Παραμέτρων με Gumbel Copula και Περιθώριες Pareto

		Bivariate Distribution		Univariate Distribution	
	Parameter	Estimate	Standard Error	Estimate	Standard Error
Loss (X_1)	λ_1	14,036	1,298	14,453	1,397
	θ_1	1.122	0.062	1.135	0.066
ALAE (X_2)	λ_2	14,219	1,426	15,133	1,633
	θ_2	2.118	0.153	2.223	0.175
Dependence	α	1.453	0.034	N/A	N/A

Βλέπουμε λοιπόν, πως στη διμεταβλητή περίπτωση οι εκτιμήσεις είναι πιο ακριβείς, εν αντιθέσει με τη μονομεταβλητή. Αυτό συμβαίνει διότι, τα τυπικά σφάλματα είναι μικρότερα στη διμεταβλητή περίπτωση, παρότι οι εκτιμήσεις είναι σχεδόν ίδιες. Επίσης, η εκτίμηση της παραμέτρου εξάρτησης $\hat{\alpha}$ είναι 1.453, πράγμα που σημαίνει ότι οι ζημίες και τα έξοδα δεν είναι ανεξάρτητα.

Μπορούμε ακόμη να μετατρέψουμε την παράμετρο εξάρτησης σε ένα πιο οικείο μέτρο συσχέτισης. Αυτό μπορεί να είναι, ο συντελεστής συσχέτισης τ του Kendall, ο οποίος είναι ίσος με: $\tau = \frac{\alpha - 1}{\alpha} = 0.31$. Οι Fees και Valdez (1998), κατέληξαν σε αντίστοιχο προσεγγιστικό αποτέλεσμα για τον συντελεστή συσχέτισης ρ_S του Spearman. Τέλος, ένα 95% διάστημα εμπιστοσύνης για την παράμετρο εξάρτησης α, είναι το επόμενο:

$$\hat{\alpha} \pm 1.96 SE(\hat{\alpha}) = 1.453 \pm 1.96(0.034) = (1.386, 1.520),$$

ενώ το αντίστοιχο 95% διάστημα εμπιστοσύνης για το συντελεστή συσχέτισης του Kendall και του Spearman είναι το $(0.28, 0.34)$.

6.2.2 Τιμολόγηση συμβολαίου αντασφάλειας

Αν υποθέσουμε ότι έχουμε ένα συμβόλαιο αντασφάλισης, με όριο συμβολαίου L και όριο ιδίας κράτησης του ασφαλιστή R. Τότε μπορούμε να εξετάσουμε την κατανομή οποιασδήποτε συνάρτησης των X_1 και X_2, για παράδειγμα της $g(X_1, X_2)$, έχοντας ως δεδομένη την από κοινού συνάρτηση κατανομής των (X_1, X_2). Στο συμβόλαιο αυτό, ο αντασφαλιστής είναι υποχρεωμένος να πληρώσει τη ζημιά που ξεπερνά το όριο ιδίας κράτησης, δηλαδή παρατηρεί ζημιές για μια περικομμένη από κάτω και μετατοπισμένη κατανομή ζημιών.

Στη συνέχεια, το μέγεθος ζημιάς της αντασφάλισης, υποθέτοντας έναν αναλογικό επιμερισμό των εξόδων, είναι:

$$g(X_1, X_2) = \begin{cases} 0 & \text{αν } X_1 < R \\ X_1 - R + \dfrac{X_1 - R}{X_1} X_2 & \text{αν } R \le X_1 < L, \\ L - R + \dfrac{L - R}{L} X_2 & \text{αν } X_1 \ge L \end{cases}$$

όπου L είναι το όριο του συμβολαίου και R το όριο ιδίας κράτησης για τις X_1 και X_2.

Η αναμενόμενη πληρωμή-ζημιά της αντασφάλισης $Eg(X_1, X_2)$, μπορεί να υπολογιστεί όπως ήδη αναφέραμε στην Ενότητα 5.4, μέσω της προσομοίωσης, η οποία είναι ένα απλούστερο και ευρέως χρησιμοποιούμενο αριθμητικό εργαλείο αξιολόγησης.

Για το Gumbel Copula, η διαδικασία και τα βήματα της προσομοίωσης, περιγράφονται στον Αλγόριθμο 5.4.2, όπου επαναλαμβάνονται πολλές φορές. Για τη διαδικασία, χρειάζεται και η αντίστροφη των περιθώριων συναρτήσεων κατανομής.

Έτσι, στην περίπτωση της Pareto, έχουμε:

$$F^{-1}(x) = \lambda[(1-x)^{-1/\theta} - 1].$$

Έστω $nsim$ (number of simulations) ο αριθμός των προσομοιώσεων που εκτελούνται με σκοπό να δημιουργήσουμε την αλληλουχία του δείγματος (X_{1i}, X_{2i}), για $i = 1,...,nsim$. Τότε, η εκτιμώμενη αξία για την αναμενόμενη καταβολή του αντασφαλιστή δίνεται από την έκφραση:

$$\hat{g}^*(\mathrm{L}, \mathrm{R}) = \frac{1}{\mathrm{nsim}} \sum_{i=1}^{nsim} g(X_{1i}, X_{2i}),$$

με τυπικό σφάλμα

$$se(\hat{g}^*(\mathrm{L}, \mathrm{R})) = \sqrt{\frac{\frac{1}{nsim} \sum_{i=1}^{nsim} g(X_{1i}, X_{2i})^2 - \hat{g}^*(\mathrm{L}, \mathrm{R})^2}{nsim}}.$$

Στη συγκεκριμένη περίπτωση, το δείγμα για το οποίο πραγματοποιήθηκε η προσομοίωση είναι $nsim = 100.000$ και τα αποτελέσματά της παρουσιάζονται στον παρακάτω πίνακα.

Πίνακας 6

Προσομοίωση με Βάση τα Αντασφάλιστρα

	Ratio of Insurer's Retention to Policy Limit (R/L)				
Policy Limit (L)	0.00	0.25	0.50	0.75	0.95
10,000	15,636 (640)	11,232 (480)	7,220 (320)	3,498 (160)	684 (32)
100,000	34,264 (655)	17,965 (493)	10,003(328)	4,425 (164)	819 (33)
500,000	49,367 (733)	17,457 (544)	9,234 (359)	4,007 (179)	739 (36)
1,000,000	55,683 (818)	16,762 (597)	8,740 (390)	3,716 (193)	672 (38)

Τα αποτελέσματα του παραπάνω πίνακα παρέχουν τα προσαρμοσμένα ασφάλιστρα που ο αντασφαλιστής έχει αξιολογήσει για την κάλυψη των δαπανών των ζημιών και εξόδων, σύμφωνα με διάφορα όρια συμβολαίων και τις αναλογίες του ορίου ιδίας κράτησης του ασφαλιστή προς το όριο του συμβολαίου. Για παράδειγμα, το όριο ιδίας κράτησης για όριο συμβολαίου 500,000 και αναλογία $R/L = 0.5$, είναι 250.000. Για το ποσό αυτό, σύμφωνα με τον αντασφαλιστή, το προσαρμοσμένο ασφάλιστρο είναι ίσο με 9,234. Τα τυπικά σφάλματα της προσομοίωσης είναι στις παρενθέσεις. Στον Πίνακα 4, το μέσο όριο συμβολαίου είναι 559,098 και το άθροισμα των μέσων απωλειών και εξόδων είναι 53,796. Στην περίπτωση που δεν υπάρχει αντασφάλιση, από τα αποτελέσματα του Πίνακα 6 βλέπουμε ένα προσαρμοσμένο ασφάλιστρο 49,367, με τυπικό σφάλμα 733 και όριο συμβολαίου 500,000. Δηλαδή, με ένα σχεδόν ίδιο όριο συμβολαίου, το προσαρμοσμένο ασφάλιστρο είναι μικρότερο από το άθροισμα των μέσων απωλειών και εξόδων. Επιπλέον, παρατηρούμε πως όταν δεν έχουμε αντασφάλιση, έχουμε υψηλότερο ασφάλιστρο για μεγαλύτερα όρια συμβολαίου ενώ παράλληλα έχουμε χαμηλότερο ασφάλιστρο όταν η αναλογία (R/L) είναι υψηλότερη για συγκεκριμένο όριο συμβολαίου.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Allaben M., Diamantoukos C., Dicke A., Gutterman S., Klugman S., Lord R., Luckner W., Miccolis R. and Tan J., (2008), *"Principles Underlying Actuarial Science"*, Society of Actuaries.

2. Arnold H., (2006), *"Dependence Modelling via the Copula Method"*, Quantitative Risk Management Group, CSIRO Mathematical and Information Sciences.

3. Balakrishnan N., and Lai C.-D., (2009), *"Continuous Bivariate Distributions"*, 2nd edition, Springer, New York.

4. Bouye 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).

5. Cherubini U., Luciano E., and Vecchiato W., (2004), "*Copula methods in Finance*", John Wiley & Sons, New York.

6. Durrleman V., Nikeghbali A., and Roncalli T., (2000), *"Which Copula is the right one?"*, Groupe de Recherche Operationnelle, Credit Lyonnais, France.

7. Embrechts P., Lindskog F., and McNeil A., (2001), *"Modelling dependence with Copulas and applications to Risk Management"*, Department of Mathematics, Zurich.

8. Embrechts P., McNeil A., and Straumann D., (1999), *"Correlation and dependence in risk management: Properties and Pitfalls"*, Department of Mathematics, Zurich.

9. Frees E. W., and Valdez E. A., (1997), *"Understanding Relationships Using Copulas"*, 32nd Actuarial Research Conference, University of Calgary, Canada.

10. Genest C., (1987), *"Frank's Family of Bivariate Distributions"*, Biometrika, Vol 74,549–555.

11. Genest C., Ghoudi K., Rivest L.-P, (1995), *"A Semi-Parametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions"*, Biometrika, Vol 82,543-552.

12. Genest C., and Mackay J., (1986), *"The Joy of Copulas: Bivariate Distributions with Uniform Marginals"*, The American Statistician, Vol. 40,280–283.

13. Genest C., and Rivest L.-P., (1993), *"Statistical Inference Procedures for Bivariate Archimedean Copulas"*, Journal of the American Statistical Association, Vol. 88, 1034-1043.

14. Jaworski P., Durante F., Hardle W., and Rychlik T., (2010), *"Copula Theory and its Applications"*, Proceedings of the Workshop held in Warsaw, 25-26 September 2009, Springer.

15. Joe H., (1997), "*Multivariate Models and Dependence Concepts"*, Monographs on Statistics and Applied Probability, Chapman and Hall/CRC.

16. Joe H., Xu J. J., (1996), "*The Estimation Method of Inference Functions for Margins for Multivariate Models*", Department of Statistics, University of British Columbia.

17. Klugman S. A., Parsa R., (1999), "*Fitting Bivariate Loss Distributions with Copulas*", Insurance: Mathematics and Economics Vol 24,139-148.

18. Klugman S. A., Panjer H., Willmot G., (2004), "*Loss Models: From Data to Decisions*", 2nd edition, John Wiley & Sons, Inc., Hoboken, NJ.

19. Marshall A.W., and Olkin I., (1988), *"Families of Multivariate Distributions"*, Journal of the American Statistical Association Vol 83,834–841.

20. Marshall A.W. and Olkin I., (1988), "*Multivariate Distributions Generated from Mixtures of Convolution and Product Families*", Department of Statistics, Stanford University, California.

21. McNeil A. J., Frey R. and Embrechts P., (2005), "*Quantitative Risk Management: Concepts, Techniques and Tools*", Princeton University Press, Princeton and Oxford.

22. Melchiori M. R., (2006), *"Tools for sampling Multivariate Archimedean Copulas",* Universidad Nacional del Litoral, Santa Fe – Argentina.

23. Nelsen R. B., (2003), "*Properties and applications of copulas: A brief survey*", Department of Mathematical Sciences, Lewis & Clark College.

24. Nelsen, R. B., (2006), "*An Introduction to Copulas"*, 2nd edition, Springer, New York.

25. Quesada-Molina J. J., (2003), *"What are copulas?"*, Departamento de Mathematica Aplicada, Universidad de Granada.

26. Trivedi P. K., and Zimmer D. M., (2005), *"Copula Modelling: An Introduction for Practitioners"*, Foundations and Trends in Econometrics Vol 1,1-111.

27. Yan J., (2007), "*Enjoy the Joy of Copulas: With a package copula"*, Journal of Statistical Software, University of Connecticut.