ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

Μεταπτυχιακό πρόγραμμα σπουδών στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Στατιστικές μέθοδοι μοντελοποίησης εξαρτημένων κινδύνων

Δουβής Σπυρίδων

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

Πανεπιστήμιο Πειραιώς

Ιούλιος 2020

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από την 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
- Αναπληρωτής Καθηγητής Χατζηκωνσταντινίδης Ευστάθιος
- Αναπληρωτής Καθηγητής Πολίτης Κωνσταντίνος

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

UNIVERSITY OF PIRAEUS

Department of Statistics and Insurance Science

Postgraduate Program in Actuarial Science and Risk Management

Statistical methods for modelling dependent risks

Douvis Spyridon

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

University of Piraeus

July 2020

Περιεχόμενα

Σε δύο ξεχωριστούς ανθρώπους,

την Παναγιώτα και τον Γιάννη

Περίληψη

Πολλές φορές όταν ενδιαφερόμαστε να μελετήσουμε την από κοινού συμπεριφορά δύο ή περισσότερων τυχαίων μεταβλητών ξεκινάμε με την υπόθεση ότι δεν υπάρχει εξάρτηση μεταξύ τους. Αυτή η υπόθεση όμως σπάνια επιβεβαιώνεται στην πράξη. Στην παρούσα εργασία θα αναφερθούμε στην έννοια του κινδύνου και σε κάποια από τα σημαντικότερα μέτρα του. Στη συνέχεια θα κάνουμε μια εισαγωγή στην θεωρία των Copulas και θα μιλήσουμε για έννοιες όπως η συσχέτιση και η εξάρτηση τυχαίων μεταβλητών. Τέλος, με τη χρήση όσων θα αναπτυχθούν θεωρητικά και της γλώσσας προγραμματισμού R, θα προχωρήσουμε στην κατασκευή ενός μαθηματικού μοντέλου το οποίο θα περιγράφει την από κοινού συμπεριφορά δύο τυχαίων μεταβλητών.

Summary

Sometimes when we are interested in studying the behavior of two or more random variables, we adopt the assumption that there is no dependence between them. However, this hypothesis is rarely confirmed in practice. In this paper, we will deal with the concept of risk and some of its important measures. In addition, we will provide an introduction to the Copulas theory and discuss concepts such as correlation and dependence of random variables. Finally, using the theoretical results presented and the R programming language, we will proceed to the construction of a mathematical model that describes the dependence structure of two random variables.

1. Εισαγωγή

Ως κίνδυνος (risk) μπορεί να χαρακτηριστεί ένα γεγονός το οποίο όταν επέλθει επιφέρει ανεπιθύμητες συνέπειες. Πολλές φορές στην καθημερινότητά μας βρισκόμαστε εκτεθειμένοι σε διάφορες πηγές κινδύνου είτε σε πλήθος μεμονωμένων κινδύνων. Ένα σημαντικό ερώτημα είναι αν οι μεμονωμένοι κίνδυνοι επηρεάζουν ο ένας τον άλλον ή όχι. Αυτό μπορεί να αφορά στην εξάρτηση που μπορεί να υπάρχει ως προς τον χρόνο επέλευσής τους ή τη σφοδρότητα που θα έχουν όταν επέλθουν. Συνεπώς, γίνεται αντιληπτό ότι υπάρχει ανάγκη να κατανοήσουμε πως μπορούμε να περιγράψουμε και να μοντελοποιήσουμε τη δομή εξάρτησης των κινδύνων.

Είναι ξεκάθαρο πως, αν οι κίνδυνοι επηρεάζουν ο ένας τον άλλο με τρόπο ώστε να επέρχονται την ίδια χρονική στιγμή και συνεπώς να αυξάνουν τη σφοδρότητα του συνολικού κινδύνου, βρισκόμαστε αντιμέτωποι με μία κατάσταση πολύ πιο επικίνδυνη απ' ότι θα ήταν αν υπήρχε ανεξαρτησία μεταξύ τους. Για παράδειγμα, ας σκεφτούμε μία παραθαλάσσια περιοχή η οποία μπορεί να πληγεί από έναν σεισμό. Ο σεισμός πιθανόν να προκαλέσει ζημιές σε κάποια κτίρια, αλλά ταυτόχρονα μπορεί να προκαλέσει ένα τσουνάμι το οποίο θα έχει επίσης καταστροφικές συνέπειες. Με άλλα λόγια σε αυτή την περίπτωση φαίνεται να υπάρχει μία ισχυρά θετική εξάρτηση μεταξύ της πρόκλησης ενός τσουνάμι μετά την επέλευση ενός σεισμού. Φυσικά αυτό δε σημαίνει ότι οι δύο κίνδυνοι θα επέρχονται πάντα μαζί καθώς είναι προφανές ότι κάθε σεισμός δεν προκαλεί τσουνάμι. Όμως, παρά το γεγονός ότι οι κίνδυνοι δεν είναι απαραίτητο να επέλθουν ταυτόχρονα, αυτό μπορεί να συμβεί, και έχει παρατηρηθεί ότι σε πολλές περιπτώσεις υπάρχει η τάση να συμβαίνει κάτι τέτοιο. Αυτό οδηγεί στην ανάγκη να κατανοήσουμε και να ποσοτικοποιήσουμε αυτή την τάση.

Στόχος της παρούσας εργασίας είναι να παρουσιαστεί πως μπορούμε να κατανοήσουμε και να μοντελοποιήσουμε τη στατιστική εξάρτηση μεταξύ διαφορετικών κινδύνων. Υπάρχουν δύο βασικές προσεγγίσεις. Η πρώτη έχει να κάνει με εκτίμηση των μεμονωμένων παραγόντων κινδύνου (risk factors) μέσω ενός κατάλληλου μέτρου κινδύνου (risk measure) και στη συνέχεια απλώς να συνυπολογίσουμε τις διαφορετικές τιμές τους. Η δεύτερη είναι να συνδυάσουμε τους κινδύνους μέσω ενός πολυδιάστατου μοντέλου και στη συνέχεια να εκτιμήσουμε την μορφή εξάρτησής τους μέσω αυτού.

2. Μέτρα Κινδύνου

2.1 Ορισμός του Μέτρου Κινδύνου

Κάθε κίνδυνος (risk) μπορεί να εκφραστεί ως μία μη αρνητική τυχαία μεταβλητή. Η μέτρηση του κινδύνου ισοδυναμεί με μία αντιστοίχιση ρ ανάμεσα στο χώρο των τυχαίων μεταβλητών και του συνόλου των μη αρνητικών πραγματικών αριθμών $\mathbb{R}^+$. Ο πραγματικός αριθμός που δηλώνει ένα γενικό μέτρο κινδύνου (risk measure) που σχετίζεται με τον κίνδυνο X θα συμβολίζεται εφεξής με $\rho[X]$. Δηλαδή, ένα μέτρο κινδύνου δεν είναι τίποτα περισσότερο από μία συνάρτηση που αντιστοιχεί σε έναν κίνδυνο έναν μη αρνητικό πραγματικό αριθμό.

Κανένα μέτρο κινδύνου δεν μπορεί να αποδώσει τη συνολική εικόνα του κινδύνου, αλλά το καθένα από αυτά μπορεί να επικεντρωθεί σε μια συγκεκριμένη πτυχή του. Για να γίνει πιο κατανοητό αυτό θα μπορούσε να γίνει ένας παραλληλισμός με τη στατιστική, όπου τα διάφορα χαρακτηριστικά των κατανομών έχουν διαφορετική ερμηνεία και χρησιμότητα. Για παράδειγμα, η μέση τιμή μετράει την κεντρική τάση, η διακύμανση την εξάπλωση των παρατηρήσεων γύρω από τη μέση τιμή, ενώ η κυρτότητα αποδίδει την ασυμμετρία.

Ως μέτρο κινδύνου θα μπορούσε να οριστεί μία απεικόνιση ρ ενός κινδύνου X σε έναν μη αρνητικό παραγματικό αριθμό $\rho[X]$, ενεδεχομένως άπειρο, που αντιπροσωπεύει την επιπλέον αξία που πρέπει να προστεθεί στον X έτσι ώστε να τον καταστήσει αποδεκτό.

Η βασική ιδέα είναι ότι η ρ ποσοτικοποιεί την επικινδυνότητα του X και μάλιστα μεγάλες τιμές του $\rho[X]$ υποδεικνύουν ότι το X είναι πιο επικίνδυνο. Για παράδειγμα, αν X είναι ο κίνδυνος να προκληθεί μία οικονομική απώλεια σε ένα χαρτοφυλάκιο μέσα σε μία χρονική περίοδο, τότε $\rho[X]$ είναι το ποσό του κεφαλαίου που πρέπει να προστεθεί επιπλέον (buffer) στο χαρτοφυλάκιο έτσι ώστε αυτό να γίνει αποδεκτό (από την άποψη ότι είναι επαρκώς προστατευμένο έναντι του κινδύνου) από έναν εσωτερικό ή εξωτερικό ελεγκτή κινδύνου. Σε αυτή την περίπτωση, το $\rho[X]$ ποσοτικοποιεί τον κεφαλαιακό κίνδυνο του χαρτοφυλακίου. Τέτοια μέτρα κινδύνου χρησιμοποιούνται για τον προσδιορισμό προβλέψεων και κεφαλαιακών απαιτήσεων έτσι ώστε να αποφευχθεί η χρεοκοπία.

2.2 Επιθυμητές Ιδιότητες

Ανάμεσα στο πλήθος των στατιστικών χαρακτηριστικών που μπορούν να συνδεθούν με την έννοια του κινδύνου (μέση τιμή, διακύμανση, συσχέτιση κ.λ.π.) λιγοστά είναι αυτά που μπορούν να χαρακτηριστούν ως αποδεκτά μέτρα κινδύνου. Στην παράγραφο αυτή θα παραθέσουμε κάποιες επιθυμητές ιδιότητες που θα θέλαμε να πληροί ένα μέτρο κινδύνου.

i. ***Μη υπερβάλλον φορτίο (Non-excessive loading)***

Για ένα μέτρο κινδύνου που έχει τη συγκεκριμένη ιδιότητα θα πρέπει να ικανοποιείται η ανισότητα:

$$\rho[X] \leq max[X], \text{ για κάθε τυχαία μεταβλητή } X.$$

Η παραπάνω ιδιότητα εκφράζει το λογικό συμπέρασμα ότι είναι άνευ ουσίας να διαθέτουμε επιπλέον κεφάλαιο το οποίο θα είναι μεγαλύτερο από την μέγιστη δυνατή ζημιά.

ii. ***Μη αρνητικό φορτίο (Non-negative loading)***

Για να έχει ένα μέτρο κινδύνου τη συγκεκριμένη ιδιότητα θα πρέπει να ισχύει:

$$\rho[X] \geq E[X], \text{ για κάθε τυχαία μεταβλητή } X.$$

Δηλαδή, η ελάχιστη κεφαλαιακή απαίτηση θα πρέπει να ξεπερνάει την μέση αναμενόμενη ζημία, καθώς σε διαφορετική περίπτωση η χρεοκοπία είναι βέβαιη (λαμβάνοντας υπόψιν τον νόμο των μεγάλων αριθμών).

iii. ***Μετατόπιση (Translativity)***

Αν ένα μέτρο κινδύνου έχει την ιδιότητα της μετατόπισης θα πρέπει να ικανοποιείται η ισότητα:

$$\rho[X + c] = \rho[X], \text{ για κάθε τυχαία μεταβλητή } X \text{ και σταθερά } c.$$

Το μέτρο κινδύνου ορίζεται όπως είδαμε, ως μία συνάρτηση η οποία αναπαριστά το επιπλέον ποσό κεφαλαίου το οποίο θα πρέπει να έχει διαθέσιμο ο κάτοχος μίας επικίνδυνης θέσης, έτσι ώστε αυτή να γίνεται αποδεκτή από κάποιον ελεγκτή. Συνεπώς, οποιαδήποτε αύξηση των υποχρεώσεων (liabilities) κατά ένα σταθερό ποσό c θα πρέπει να αντισταθμίζεται με ισόποση αύξηση επιπλέον κεφαλαίου.

iv. Σταθερότητα (Constancy)

Για οποιαδήποτε σταθερά c θα πρέπει να ισχύει:

$$\rho[c] = c.$$

Για παράδειγμα, προκειμένου να ανταπεξέλθει ένας ασφαλιστής σε μία ζημιά ύψους c αρκεί να έχει ισόποσο κεφάλαιο ως απόθεμα.

v. Υποπροσθετικότητα (Subadditivity)

Για να χαρακτηριστεί ένα μέτρο κινδύνου ως υποπροσθετικό θα πρέπει, για κάθε τυχαία μεταβλητή X και Y να ισχύει η παρακάτω ανισότητα:

$$\rho[X + Y] \leq \rho[X] + \rho[Y]$$

Η λογική πίσω από αυτή την ιδιότητα έγκειται στην πεποίθηση ότι η συγχώνευση περιθώριων κινδύνων δεν δημιουργεί επιπλέον κίνδυνο. Η υποπροσθετικότητα εκφράζει την ιδέα ότι ο κίνδυνος μπορεί να μειωθεί μέσω της διαφοροποίησης (diversification). Η διαφοροποίηση είναι μία στρατηγική διαχείρισης κινδύνου η οποία προβλέπει ότι η σύσταση ενός χαρτοφυλακίου θα πρέπει να αποτελείται από διαφορετικά περιουσιακά στοιχεία. Η λογική πίσω από αυτό είναι ότι κατά την επέλευση ενός κινδύνου θα πληγεί μόνο ένα μέρος του χαρτοφυλακίου και με αυτό τον τρόπο θα περιοριστεί η ζημία. Όταν ισχύει η ισότητα θα μιλάμε για προσθετικότητα. Σε αυτή την περίπτωση, η δομή εξάρτησης μεταξύ των X και Y καθορίζεται ως προσθετικότητα για ανεξάρτητους κινδύνους ή προσθετικότητα για συμμονοτονικούς κινδύνους (comonotonic risks).

Η παραπάνω ιδιότητα, όταν έχουμε n το πλήθος, μεμονωμένους κινδύνους X_i, $i = 1, 2,...,n$ οδηγεί στην ανισότητα:

$$\rho[\sum_{i=1}^{n} X_i] \leq \sum_{i=1}^{n} \rho[X_i] \Leftrightarrow$$

$$\sum_{i=1}^{n} \rho[X_i] - \rho[\sum_{i=1}^{n} X_i] \geq 0$$

Η παραπάνω διαφορά χαρακτηρίζεται και ως επίδραση της διαφοροποίησης (diversification effect) και είναι πάντα θετική για μέτρα κινδύνου που έχουν την υποπροσθετική ιδιότητα.

vi. ***Συμμονοτονική προσθετικότητα (Comonotonic additivity)***

Έστω X, Y συμμονοτονικές τυχαίες μεταβλητές. Τότε:

$$\rho[X + Y] \;=\; \rho[X] \;+\; \rho[Y]$$

Σε αυτό το σημείο θα πρέπει να πούμε λίγα λόγια για την έννοια της συμμονοτονικότητας. Ένα σύνολο $A \subseteq \mathbb{R}^n$ ονομάζεται συμμονοτονικό αν για οποιαδήποτε $\tilde{X}$ και $\tilde{Y}$ στο A ισχύει,

$$\tilde{X} \leq \tilde{Y} \text{ ή } \tilde{Y} \leq \tilde{X}.$$

Για να ισχύει η ανισότητα $\tilde{X} \leq \tilde{Y}$ θα πρέπει $X_i \leq Y_i$ για κάθε $i = 1, 2, \ldots, n$ (προφανώς το αντίστροφο θα πρέπει να ιχύει για την ανισότητα $\tilde{Y} \leq \tilde{X}$).

Ένα τυχαίο διάνυσμα $\tilde{X} = (X_1, X_2, \ldots, X_n)$ θα χαρακτηρίζεται ως συμμονοτονικό αν οι παρακάτω συνθήκες είναι ισοδύναμες:

i. Το διάνυσμα $\tilde{X}$ έχει συμμονοτονικό στήριγμα (support)

ii. Για κάθε $\tilde{X} = (X_1, X_2, \ldots, X_n)$ ισχύει:

$$F_{\tilde{X}}(\tilde{x}) = min\{F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\}$$

iii. Για $U \sim Uni(0,1)$, θα έχουμε:

$$\tilde{X} = (F_{X_1}^{-1}(U), F_{X_2}^{-1}(U), \ldots, F_{X_n}^{-1}(U))$$

iv. Υπάρχει τυχαία μεταβλητή Z και μη-φθίνουες συναρτήσεις f_i, $i = 1, 2, \ldots, n$ τέτοιες ώστε:

$$\tilde{X} = (f_1(Z), f_2(Z), \ldots, f_n(Z))$$

(Η απόδειξη για την ισοδυναμία των τεσσάρων συνθηκών υπάρχει στο επιστημονικό άρθρο των Dhaene, Denuit, Goovaerts and Kaas, 2005).

Οι συμμονοτονικές τυχαίες μεταβλητές έχουν τη μεγαλύτερη δυνατή θετική εξάρτηση.

Η ανισότητα που ίσχυε στην προηγούμενη ιδιότητα μετατρέπεται σε ισότητα, υπό την λογική πως αν λάβουμε υπόψιν συμμονοτονικούς κινδύνους μαζί δεν μειώνεται η επικινδυνότητα της κατάστασης. Οι συμμονοτονικοί κίνδυνοι μπορούν να θεωρηθούν ως στοιχήματα τα οποία δεν μπορούνε να αντισταθμιστούν (hedge) μεταξύ τους.

vii. Θετική Ομοιογένεια (Positive Homogeneity)

Για να ικανοποιείται η ιδιότητα αυτή θα πρέπει, για κάθε τυχαία μεταβλητή X και σταθερά c να ισχύει:

$$\rho(cX) = c\,\rho(X)$$

Η ιδιότητα αυτή συνδέεται συνήθως με ανεξαρτησία. Η θετική ομοιογένεια είναι στενά συνδεδεμένη με την συμμονοτονική προσθετικότητα, καθώς αν θεωρήσουμε έναν ακέραιο c, τότε από την ιδιότητα της θετικής ομοιογένειας θα ισχύει ότι:

$$\rho(cX) = \rho[X + X + \ldots + X] = \rho[X] + \rho[X] + \ldots + \rho[X] = c\,\rho(X)$$

viii. Μονοτονία (Monotonicity)

Θα λέμε ότι ένα μέτρο πιθανότητας έχει την ιδιότητα της μονοτονίας αν για κάθε τυχαία μεταβλητή X, Y ισχύει η σχέση:

$$Pr[X \leq Y] = 1 \Rightarrow \rho(X) \leq \rho(Y)$$

Η ιδιότητα της μονοτονίας προκύπτει με πολύ φυσικό τρόπο καθώς εκφράζει το γεγονός ότι το ύψος του κεφαλαίου που απαιτείται ως μέτρο ασφαλείας στην περίπτωση επέλευσης ζημίας X είναι πάντα μικρότερο από το αντίστοιχο της Y, όταν το Y πάντα υπερβαίνει το X.

ix. Συνέχεια ως προς τη σύγκλιση κατά κατανομή (Continuity with respect to convergence in distribution)

Έστω μία ακολουθία κινδύνων $\{X_n, n = 1,2,\ldots\}$ τέτοια ώστε $X_n \rightarrow_d X$ και $n \rightarrow \infty$, για την οποία:

$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, σε κάθε σημείο συνέχειας x της F_X. Τότε:

$$\lim_{n\to\infty} \rho[X_n] = \rho(X)$$

x. Αντικειμενικότητα (Objectivity)

Το μέτρο $\rho[X]$ εξαρτάται από το X μόνο μέσω της συνάρτησης κατανομής F_X του X. Το γεγονός αυτό εξασφαλίζει ότι η F_X περιέχει όλη την πληροφορία που χρειάζεται ώστε να προσδιοριστεί η επικινδυνότητα του X. Η ιδιότητα αυτή καλείται και νόμος του αναλλοίωτου και εκφράζεται μέσω της σχέσης:

$$X = dY \Rightarrow \rho[X] = \rho[Y]$$

Η ιδιότητα είναι ιδιαίτερα σημαντική για τις εφαρμογές που βρίσκουν τα μέτρα κινδύνου, καθώς είναι απαραίτητη προϋπόθεση να μπορούν να εκτιμηθούν από εμπειρικά δεδομένα. Συνεπώς η έλλειψή της από κάποιο μέτρο κινδύνου, καθιστά το μέτρο μη ασφαλές ώστε να χρησιμοποιηθεί σε οικονομικά μοντέλα.

2.3 Αξία σε Κίνδυνο (Value at Risk)

Τα τελευταία χρόνια υπάρχει ένα συνεχώς αυξανόμενο ενδιαφέρον στα ποσοστιαία σημεία (quantiles) των συναρτήσεων κατανομής. Το άνω ποσοστιαίο σημείο αποτελεί το σημείο της κατανομής για το οποίο το α% των παρατηρήσεων είναι μεγαλύτερες ή ίσες από αυτό και το υπόλοιπο (1-α)% των παρατηρήσεων είναι μικρότερες ή ίσες από αυτό (το αντίστροφο ισχύει για το κάτω ποσοστιαίο σημείο). Για το λόγο αυτό έχει βρει χρήση στο χώρο της διαχείρισης κινδύνου μέσω της έννοιας της αξίας σε κίνδυνο (Value at Risk ή VaR). Η έννοια αυτή εισήχθηκε στην προσπάθεια να δοθεί απάντηση στο ερώτημα "πόσο περιμένουμε να χάσουμε σε διάρκεια μιας ημέρας, εβδομάδας ή έτους κ.λπ. με δεδομένη πιθανότητα". Στη σημερινή εποχή, το VaR είναι ένα ευρέως αποδεκτό και δημοφιλές μέτρο κινδύνου ιδιαίτερα στις χρηματοοικονομικές και ασφαλιστικές επιχειρήσεις, καθώς οι εποπτικές αρχές το αποδέχονται ως τον βασικό δείκτη για την οριοθέτηση των κεφαλαίων που απαιτείται να έχει μια εταιρεία ώστε να καθίσταται φερέγγυα από την έκθεσή της στον κίνδυνο αγοράς.

Δεδομένου ενός κινδύνου X, με αθροιστική συνάρτηση κατανομής $F_X(x)$ και ενός επιπέδου πιθανότητας $\alpha \in (0,1)$, η αξία σε κίνδυνο (η οποία θα συμβολίζεται με VaR) ορίζεται ως:

$$VaR[X;\alpha] = F_X^{-1}(\alpha), \text{ όπου}$$

$$F_X^{-1}(\alpha) = inf\{x : F_X(x) \geq \alpha\}, \text{ είναι}$$

η αντίστροφη συνάρτηση της συνάρτησης κατανομής F_X.

Διαισθητικά, το VaR δηλώνει εκείνο το επίπεδο το οποίο δε θα ξεπεραστεί στο $100 \cdot \alpha\%$ των περιπτώσεων. Για παράδειγμα, αν το VaR σε επίπεδο 95% παίρνει μια τιμή C, τότε ο κίνδυνος παρουσιάζεται στο 5% των περιπτώσεων όπου θα ξεπεραστεί αυτό το φράγμα C. Το VaR έχει ενσωματωθεί στους ευρωπαϊκούς κανονισμούς Βασιλεία ΙΙ (Basel II) και Φερεγγυότητα ΙΙ (Solvency II), οι οποίοι υπαγορεύουν ένα σύνολο εποπτικών κανόνων με βάση τους οποίους οι τράπεζες και οι ασφαλιστικές επιχειρήσεις αντίστοιχα, οφείλουν να διασφαλίζουν την ύπαρξη επιπλέον κεφαλαίων για να καλύψουν πιθανές μελλοντικές απώλειες που μπορεί να προκύψουν από την επιχειρηματική τους δραστηριότητα.

Το VaR ως μέτρο κινδύνου έχει τις ιδιότητες που αναφέρθηκαν στην προηγούμενη παράγραφο εκτός από αυτήν της υποπροσθετικότητας (με εξαίρεση κάποιες ειδικές περιπτώσεις όπου τα X_i ακολουθούν την πολυδιάστατη κανονική κατανομή). Συνεπώς ισχύει γενικά ότι, το VaR του αθροίσματος μπορεί να είναι μεγαλύτερο από το άθροισμα των επιμέρους VaR. Σε μία τέτοια περίπτωση η διαφοροποίηση θα οδηγήσει σε υπερεκτίμηση του κινδύνου.

Ενδιαφέρον παρουσιάζει να δούμε πως το VaR σαν μέτρο κινδύνου βρίσκει εφαρμογή στον κλάδο των ασφαλιστικών επιχειρήσεων καθώς θεωρείται κατάλληλο για τον προσδιορισμό των κεφαλαιακών τους απαιτήσεων. Στις ασφαλιστικές επιχειρήσεις ο παραγωγικός κύκλος λειτουργεί αντίστροφα από το σύνηθες, υπό την έννοια ότι τα ασφάλιστρα πληρώνονται από τον πελάτη (αντισυμβαλλόμενο) πριν η αποζημίωση πληρωθεί από την ασφαλιστική. Ένα χαρτοφυλάκιο μπορεί να αντιμετωπίσει πρόβλημα αν η ζημία του, έστω X, είναι θετική, γιατί οι υποχρεώσεις προς τους ασφαλισμένους δε θα μπορούν να καλυφθούν πλήρως στην περίπτωση αυτή. Για την προστασία των ασφαλισμένων η εποπτική αρχή επιβάλλει την κεφαλαιακή απαίτηση φερεγγυότητας $\rho[X]$. Αυτό σημαίνει ότι η εποπτική αρχή απαιτεί από τις ασφαλιστικές επιχειρήσεις, να έχουν επιπλέον διαθέσιμο κεφάλαιο (δηλαδή πλεόνασμα των στοιχείων του ενεργητικού έναντι των στοιχείων του παθητικού) τουλάχιστον ίσο με το $\rho[X]$. Το κεφάλαιο αυτό λειτουργεί ως ένα μέτρο ασφαλείας έναντι του κινδύνου που υπάρχει, τα κεφάλαια, τα αποθέματα μαζί με τα κέρδη από τις επενδύσεις να μην είναι επαρκή για να καλύψουν τις υποχρεώσεις της επιχείρησης προς τους ασφαλισμένους. Το $\rho[X]$ θα πρέπει να επιλεγεί με τρόπο ώστε να είναι κανείς αρκετά σίγουρος ότι το γεγονός $X > \rho[X]$ δε θα συμβεί.

Ας υποθέσουμε ότι έχουμε ένα χαρτοφυλάκιο και μία ζημιά X. Ο επόπτης θέλει η κεφαλαιακή απαίτηση φερεγγυότητας που σχετίζεται με τη ζημιά X να είναι αρκετά μεγάλη ώστε να διασφαλίσει ότι η χρεοκοπία (εδώ ο όρος χρεοκοπία δεν χρησιμοποιείται με την απόλυτη έννοια του όρου, αλλά περισσότερο περιγράφει μία κατάσταση η οποία θα είναι ιδιαίτερα ζημιογόνα για την επιχείρηση) είναι ικανοποιητικά μικρή. Για την επίτευξη αυτού του σκοπού οι επόπτες μετράνε τον κίνδυνο χρεοκοπίας ως $\mathrm{E}[(X - \rho[X])_+]$. Η διαδικασία για τον προσδιορισμό των κεφαλαιακών απαιτήσεων απαιτεί δύο διαφορετικά μέτρα κινδύνου: ένα για να προσδιοριστεί το κεφάλαιο φερεγγυότητας και της $\mathrm{E}[(X - \rho[X])_+]$ για να μετρήσουμε τον κίνδυνο χρεοκοπίας. Είναι ξεκάθαρο ότι θέλουμε η ποσότητα $\mathrm{E}[(X - \rho[X])_+]$ να είναι ικανοποιητικά μικρή. Επίσης, όσο μεγαλύτερο το κεφάλαιο, τόσο καλύτερα από την άποψη της ελαχιστοποίησης της $\mathrm{E}[(X - \rho[X])_+]$. Από την άλλη η διακράτηση κεφαλαίων έχει κόστος για μία εταιρεία. Το κόστος προκύπτει από το γεγονός ότι όταν μία εταιρεία υποχρεούται να έχει διαθέσιμα κεφάλαια ως δικλίδα ασφαλείας για την περίπτωση επέλευσης ενός καταστροφικού (οικονομικά) γεγονότος, δεν μπορεί να επενδύσει τα κεφάλαια αυτά ούτως ώστε να αποκομίσει από αυτά κάποια απόδοση. Συνεπώς οδηγούμαστε στο ότι η κεφαλαιακή απαίτηση ρ μπορεί να προσδιοριστεί ως η λύση του ακόλουθου προβλήματος ελαχιστοποίησης ως προς $\rho[X]$:

$$min\{\, E[(X - \rho[X])_+] + \rho[X]\, \varepsilon\}, 0 < \varepsilon < 1,$$

το οποίο ισορροπεί ανάμεσα στα δύο αντικρουόμενα κριτήρια του ελάχιστου υπολειπόμενου κινδύνου και του χαμηλού κόστους κεφαλαίου. Το ε μπορεί να ερμηνευτεί ως ένας συντελεστής που καθορίζει τον βαθμό στον οποίο το κόστος κεφαλαίου λαμβάνεται υπόψιν. Η εποπτική αρχή θα αποφασίσει αν το ε θα είναι καθορισμένο περισσότερο με βάση το προφίλ της εταιρείας ή με βάση το προφίλ του κινδύνου. Αν $\varepsilon = 0$ το κόστος κεφαλαίου δεν λαμβάνεται υπόψιν, όμως όσο αυξάνεται η τιμή του ε τόσο αυξάνεται και η βαρύτητα που δίνεται στο κόστος κεφαλαίου και συνεπώς επηρεάζεται η βέλτιστη λύση του προβλήματος. Όπως αποδείχτηκε το μικρότερο δυνατό κεφάλαιο το οποίο είναι λύση της παραπάνω σχέσης είναι το:

$$\rho[X] \; = \; VaR[X \, ; \, 1 - \varepsilon].$$

Το παραπάνω αποτέλεσμα είναι ο λόγος για τον οποίο το VaR χρησιμοποιήθηκε για τον προσδιορισμό των κεφαλαιακών απαιτήσεων φερεγγυότητας. Αυτό που πρέπει να παρατηρήσουμε είναι ότι το VaR δε χρησιμοποιείται για τη μέτρηση του κινδύνου αλλά για τον προσδιορισμό της βέλτιστης κεφαλαιακής απαίτησης. Ο κίνδυνος που πρέπει να μετρηθεί είναι ο κίνδυνος χρεοκοπίας και το μέτρο που χρησιμοποιούμε για τον σκοπό αυτό είναι το $E[(X - \rho[X])_+]$. Επιλέον, παρά το γεγονός ότι το VaR έχει ευρεία χρήση κυρίως σε προβλήματα που παρουσιάζονται στον χρηματοπιστωτικό και ασφαλιστικό τομέα, θα πρέπει να αναφερθεί ένα βασικό μειονέκτημα του.

Το VaR δεν είναι πάντα συναφές (coherent) μέτρο κινδύνου. Ένα μέτρο κινδύνου χαρακτηρίζεται ως συναφές αν έχει τις ιδιότητες της θετικής ομοιογένειας, της υποπροσθετικότητας, της μονοτονίας και μετατόπιση. Με άλλα λόγια, θα πρέπει το μέτρο κινδύνου του αθροίσματος δύο κινδύνων να είναι πάντα μικρότερο από το άθροισμα των μέτρων κινδύνου. Για παράδειγμα, οι τράπεζες εκτιμούν το VaR για κάθε ένα από τα χαρτοφυλάκια τους και στη συνέχεια με βάση κάθε ένα από αυτά εκτιμούν τον συνολικό κίνδυνο στον οποίο είναι εκτεθειμένη η τράπεζα. Αυτό όμως μπορεί να οδηγήσει σε σημαντική υποεκτίμηση της πραγματικής τιμής του VaR.

2.4 Αναμενόμενη Απώλεια (Expected Shortfall)

Στην προηγούμενη παράγραφο είδαμε ότι το VaR δημιουργήθηκε στην προσπάθεια να δώσουμε απάντηση στο ερώτημα "πόσο περιμένουμε να χάσουμε σε διάρκεια μιας ημέρας, εβδομάδας ή έτους κ.λπ. με δεδομένη πιθανότητα". Όμως το VaR με ένα προκαθορισμένο επίπεδο πιθανότητας α δεν δίνει καμία πληροφορία σχετική με το βάρος της δεξιάς ουράς της συνάρτησης κατανομής. Το γεγονός αυτό αξίζει να σημειωθεί καθώς δεν μας ενδιαφέρει μόνο η συχνότητα της χρεοκοπίας αλλά και η σφοδρότητα της. Μας ενδιαφέρει δηλαδή να απαντήσουμε και στο ερώτημα " πόσο άσχημες θα είναι οι συνέπειες όταν συμβεί το ανεπιθύμητο γεγονός"; Για το λόγο αυτό συχνά χρησιμοποιείται ένα άλλο μέτρο κινδύνου το οποίο καλείται Αναμενόμενη Απώλεια (το οποίο εφεξής θα αναφέρεται ως Expected Shortfall ή ES).

Το Expected Shortfall μιας τυχαίας μεταβλητής X με $VaR[X\ ;\ \alpha]$ ορίζεται ως:

$$ES_X(\alpha) = \frac{1}{1-F_X(VaR[X\ ;\ \alpha])} \int_{VaR[X\ ;\ \alpha]}^{\infty} x dF_X(x),$$

όπου F_X η συνάρτηση κατανομής της X.

Όταν η X είναι συνεχής ισχύει $F_X(VaR[X\ ;\ \alpha]) = \alpha$, συνεπώς για $0 < \alpha < 1$,

$$ES_X(\alpha) = \frac{1}{1-\alpha} \int_{\alpha}^{1} VaR[X\ ;\ u]\, du$$

Το ES μπορεί να είναι μεγαλύτερο από το VaR για ίδιο επίπεδο πιθανότητας α. Μπορούμε να το σκεφτούμε ως το άθροισμα του άνω ποσοστιαίου σημείου $VaR[X\ ;\ \alpha]$ και αναμενόμενης υπερβάλουσας ζημίας (expected excess loss). Το συγκεκριμένο μέτρο πιθανότητας είναι συναφές και πολύ απλά είναι η μαθηματική αποτύπωση της " μέσης ζημίας στο χειρότερο $100(1-\alpha)\%$ των περιπτώσεων". Αν θεωρήσουμε ως $\nu = VaR[X\ ;\ \alpha]$ ένα κατώφλι που οριοθετεί το ποιές ζημίες μπορούν να χαρακτηριστούν ως επικίνδυνες σε ένα επίπεδο εμπιστοσύνης α, τότε αυτό το μέτρο κινδύνου παρέχει ένα επίπεδο ασφάλειας έναντι της μέσης τιμής των ζημιών που ξεπερνούν το κατώφλι ν.

Συγκριτικά με το VaR θα μπορούσαμε να πούμε ότι, ενώ το VaR λαμβάνει υπόψιν του το γεγονός εμφάνισης μιας μεγάλης ζημίας, δεν λαμβάνει υπόψιν του το ύψος της ζημίας. Υπό αυτή την έννοια το Expected Shortfall είναι προτιμότερο καθώς μετράει τη συνολική μέση ζημία που ξεπερνάει το VaR. Αν για παράδειγμα, μία τράπεζα κρατήσει επιπλέον κεφάλαια ίσα με το VaR του χαρτοφυλακίου της, η χρεοκοπία είναι βέβαιη στην περίπτωση που επέλθει μία ζημία που ξεπεράσει το όριο του VaR. Σε αντίθεση, αν χρησιμοποιηθεί το ES για τον προσδιορισμό των κεφαλαίων, θα έχει κρατήσει αρκετά κεφάλαια ώστε να αντιμετωπίσει κατά μέσο όρο ένα τέτοιο γεγονός.

3. Copulas

Κεντρικό ρόλο στη διαχείριση κινδύνου παίζει η κατασκευή μοντέλων τα οποία περιγράφουν την τυχαιότητα η οποία υπάρχει σε διάφορες καταστάσεις. Για πολύ καιρό τα στατιστικά μοντέλα βασίζονταν σε απλουστευμένες υποθέσεις, με την πολυδιάστατη κανονική κατανομή να παίζει κυρίαρχο ρόλο. Αυτό συνέβαινε γιατί η συγκεκριμένη κατανομή είναι σχετικά απλή στη χρήση καθώς μπορούμε να περιγράψουμε τη σχέση μεταξύ δύο τυχαίων μεταβλητών γνωρίζοντας μόνο τις περιθώριες κατανομές και τον συντελεστή συσχέτισης. Με την πάροδο των χρόνων όμως άρχισε να γίνεται αντιληπτό τόσο στους επιστήμονες όσο και στους επαγγελματίες της αγοράς, ότι σε πολλές περιπτώσεις η πολυδιάστατη κανονική κατανομή δεν παρείχε επαρκή προσέγγιση της εξάρτησης σε πολλά είδη δεδομένων. Έτσι, άρχισαν να γίνονται προσπάθειες ώστε να κατασκευαστούν κι άλλες πολυδιάστατες κατανομές οι οποίες όμως προκύπτουν κυρίως ως άμεση επέκταση των μονοδιάστατων κατανομών. Αυτού του είδους οι κατανομές έχουν τα μειονεκτήματα ότι (i) χρειαζόμαστε διαφορετικές οικογένειες για τις περιθώριες κατανομές και (ii) τα μέτρα συσχέτισης συχνά εμφανίζονται στις περιθώριες κατανομές. Θα παραθέσουμε ένα παράδειγμα για να γίνει αυτό περισσότερο κατανοητό.

Ας υποθέσουμε ότι θα θέλαμε να κατασκευάσουμε ένα μοντέλο το οποίο θα περιγράφει δύο υπολειπόμενους χρόνους ζωής οι οποίοι μπορεί να επηρεαστούν από κάποιο καταστροφικό γεγονός. Συνεπώς θα θέλαμε να εξετάσουμε αν υπάρχει κάποια εξάρτηση μεταξύ τους. Έστω Y_1, Y_2 δύο ανεξάρτητες τυχαίες μεταβλητές που περιγράφουν τους χρόνους ζωής με συναρτήσεις κατανομής H_1 και H_2 αντίστοιχα. Υποθέτουμε ακόμα ότι υπάρχει μία τυχαία μεταβλητή $Z \sim Exp(\lambda)$ η οποία αναπαριστά το χρόνο έως να επέλθει το καταστροφικό γεγονός. Αφού οι δύο χρόνοι ζωής υπόκεινται στο ίδιο καταστροφικό γεγονός τότε θα μπορούσαμε να θεωρήσουμε τις τυχαίες μεταβλητές $X_1 = \min\{Y_1, Z\}$ και $X_2 = min\{Y_2, Z\}$ οι οποίες περιγράφουν την ηλικία στη οποία η κάθε ζωή θα τελειώσει. Θεωρούμε επίσης ότι $Y_1 \sim Exp(\lambda_1)$ και $Y_2 \sim Exp(\lambda_2)$. Αν $\bar{F}_1, \bar{F}_2$ οι συναρτήσεις ουρών (ή επιβίωσης) των X_1, X_2 τότε:

$$\bar{F}_1(x_1) = \mathrm{P}[X_1 > x_1] = \mathrm{P}[Y_1 > x_1]\ \mathrm{P}[Z > x_1] = e^{-(\lambda+\lambda_1)x_1}$$

$$\bar{F}_2(x_2) = \mathrm{P}[X_2 > x_2] = \mathrm{P}[Y_2 > x_2]\ \mathrm{P}[Z > x_2] = e^{-(\lambda+\lambda_2)x_2}$$

οπότε $X_1 \sim \mathrm{Exp}(\lambda + \lambda_1)$ και $X_2 \sim \mathrm{Exp}(\lambda + \lambda_2)$. Η από κοινού συνάρτηση ουράς των X_1, X_2 θα δίνεται από τη σχέση:

$$\begin{aligned}\bar{F}_X(x_1, x_2) &= \mathrm{P}(X_1 > x_1, X_2 > x_2) = \mathrm{P}[Y_1 > x_1, Y_2 > x_2, Z > \max\{x_1, x_2\}] \\ &= e^{-\lambda_1 x_1}\ e^{-\lambda_2 x_2}\ e^{-\lambda\, max\{x_1, x_2\}} \\ &= \bar{F}_1(x_1)\ \bar{F}_2(x_2)\ \min\{e^{\lambda x_1}, e^{\lambda x_2}\}\end{aligned}$$

Τελικά, η από κοινού συνάρτηση κατανομής θα είναι:

$$F_X(x_1, x_2) = F_1(x_1) + F_2(x_2) - 1 + \bar{F}_1(x_1)\,\bar{F}_2(x_2)\,\min\{e^{\lambda x_1}, e^{\lambda x_2}\}.$$

Στο παραπάνω παράδειγμα κατασκευάσαμε μία από κοινού συνάρτηση κατανομής για δύο τυχαίες μεταβλητές, με στόχο να την χρησιμοποιήσουμε σαν μοντέλο που περιγράφει τη συμπεριφορά δύο υπολειπόμενων χρόνων ζωής. Μέσω του παραδείγματος αυτού μπορούμε να δούμε τα μειονεκτήματα στα οποία αναφερθήκαμε προηγουμένως. Αρχικά, για να κατασκευάσουμε μία δισδιάστατη συνάρτηση κατανομής με εκθετικές περιθώριες, έπρεπε να υποθέσουμε ότι οι X_1, X_2, Z είναι εκθετικά κατανεμημένες. Αυτού του είδους οι κατασκευές όμως δεν μπορούν να γίνουν εύκολα σε άλλες κατανομές. Επιπλέον, η εξάρτηση προκαλείται από τον κοινό παράγοντα Z, ο οποίος εμπλέκεται στις X_1, X_2. Το πόσο ισχυρή είναι η εξάρτηση καθορίζεται από την τιμή της παραμέτρου λ, η οποία όμως εμφανίζεται τόσο στις περιθώριες όσο και στην από κοινού συνάρτηση κατανομής, με αποτέλεσμα να είναι δύσκολη η ερμηνεία της. Η κατασκευή πολυδιάστατων κατανομών με τη χρήση των συναρτήσεων copula ξεπερνάει τα προβλήματα αυτά και είναι πιο ευέλικτη, καθώς δεν υπάρχει κάποιος περιορισμός για τις περιθώριες κατανομές. Μπορούμε να επιλέξουμε ως περιθώριες τις κατανομές που θεωρούμε ότι ταιριάζουν καλύτερα στα δεδομένα μας και στη συνέχεια μέσω ενός κατάλληλου copula να κατασκευάσουμε την από κοινού συνάρτηση κατανομής. Ένα μειονέκτημα των συναρτήσεων αυτών είναι ότι δεν είναι το ίδιο αποτελεσματικές όταν έχουμε διακριτές και όχι συνεχείς τυχαίες μεταβλητές. Για το λόγο αυτό θα ασχοληθούμε με τη συνεχή περίπτωση.

3.1 Δισδιάστατα Copulas

Ένα δισδιάστατο copula είναι μια συνάρτηση C: $[0,1]^2 \rightarrow [0,1]$ η οποία είναι μη-φθίνουσα, δεξιά συνεχής και ικανοποιεί τις παρακάτω ιδιότητες:

i. $\lim\limits_{\substack{u_1 \to 0 \\ u_2 \to 0}} C(u_1,u_2) = 0$

ii. $\lim\limits_{u_1 \to 1} C(u_1,u_2) = u_2$ *και* $\lim\limits_{u_2 \to 1} C(u_1,u_2) = u_1$

iii. Ισχύει η ανισότητα:

$$C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) + C(u_1, u_2) \geq 0, \text{ για } u_1 \leq v_1 \text{ και } u_2 \leq v_2$$

Τα δισδιάστατα copulas είναι συναρτήσεις που μπορούν να αναπαρασταθούν γραφικά στον μοναδιαίο κύβο I^3. Οι Fréchet και Hoeffding έδωσαν φράγματα ανάμεσα στα οποία περιέχονται τα copulas. Δηλαδή, για κάθε $(u, v) \in I^2$ η συνάρτηση C θα ικανοποιεί την παρακάτω σχέση:

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

Αποδεικνύεται μάλιστα, ότι στη δισδιάστατη περίπτωση τα παραπάνω φράγματα είναι copulas.

Η μελέτη των copulas για τη δημιουργία στατιστικών μοντέλων έγινε εντονότερη τα τελευταία χρόνια και ένας από τους λόγους είναι η διατύπωση του θεωρήματος που ακολουθεί, από τον Αμερικάνο μαθηματικό Abe Sklar, το οποίο φέρει και το όνομά του.

Έστω H μία από κοινού συνάρτηση κατανομής με συνεχής περιθώριες συναρτήσεις κατανομής F και G. Τότε υπάρχει copula C τέτοιο ώστε για κάθε $x, y \in \overline{\mathbb{R}}$ να ισχύει:

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

Αντίστροφα, αν C copula και F, G συναρτήσεις κατανομής, τότε η συνάρτηση H που ορίστηκε παραπάνω είναι μια δισδιάστατη συνάρτηση κατανομής με περιθώριες τις F, G. Επιπλέον, αν F, G συνεχείς το C είναι μοναδικό.

Μία ενδιαφέρουσα, εναλλακτική έκφραση της παραπάνω σχέσης όταν έχουμε συνεχείς περιθώριες είναι και η εξής:

$$C(\boldsymbol{u}) = H(F^{-1}(u_1), G^{-1}(u_2)), \boldsymbol{u} \in [0,1]^2$$

Το θεώρημα του Sklar είναι μεγάλης σημασίας καθώς μέσω αυτού ορίστηκε το copula ως από κοινού συνάρτηση κατανομής. Επιπλέον, η δομή εξάρτησης περιγράφεται εξ' ολοκλήρου από το C και διαχωρίζεται από τις περιθώριες F, G, καθώς η επιλογή τους δεν επηρεάζει την επιλογή του copula.

Θα ήταν χρήσιμο να δούμε ένα παράδειγμα για να γίνει πιο κατανοητή η σημασία του θεωρήματος του Sklar.

Έστω Y, Z δύο ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συναρτήσεις κατανομής F. Έστω επίσης $X_1 = min(Y, Z)$ και $X_2 = max(Y, Z)$ με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα. Τότε:

$$\mathrm{P}(X_1 \leq x_1, X_2 \leq x_2) = 2\,F(\min\{x_1, x_2\}\,F(x_2) - F(\min\{x_1, x_2\})^2$$

Οι δύο περιθώριες κατανομές δίνονται από τους τύπους:

i. $F_1(x) = 2\,F(x)$ - $F(x)^2$
ii. $F_2(x) = F(x)^2$

Όμως από το θεώρημα του Sklar γνωρίζουμε ότι $F_X(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ οπότε μπορούμε να συνδέσουμε τα μέλη ως και να πάρουμε την παρακάτω έκφραση:

$$C(u_1, u_2) = 2 \min\{1 - \sqrt{1 - u_1}, \sqrt{u_2}\} \sqrt{u_2} - \min\{1 - \sqrt{1 - u_1}, \sqrt{u_2}\}^2$$

Όπως είναι φανερό μέσω της συνάρτησης copula καταφέραμε να συνδέσουμε με σχετικά απλό τρόπο τις περιθώριες κατανομές και να κατασκευάσουμε την από κοινού συνάρτηση κατανομής. Μερικά απλά, αλλά βασικά copulas αναφέρονται παρακάτω.

Copula Ανεξαρτησίας (Independence Copula)

Θεωρούμε ανεξάρτητες τυχαίες μεταβλητές X_1 και X_2 με συναρτήσεις κατανομής F_1και F_2 αντίστοιχα. Η από κοινού συνάρτηση κατανομής δίνεται από τον τύπο $F_X(x) = F_1(x_1)\, F_2(x_2)$ και το copula θα είναι:

$$C_I(u_1,u_2) = u_1 u_2 \text{ , } u \in [0,1]^2$$

Το παραπάνω ονομάζεται copula ανεξαρτησίας. Αν X_1 και X_2 τυχαίες μεταβλητές με συνάρτηση κατανομής $F_X(x_1,x_2) = C(F_1(x_1), F_2(x_2))$, θα είναι ανεξάρτητες αν και μόνο αν $C \equiv C_I$.

Copula Άνω Φράγματος (Fréchet Upper Bound Copula)

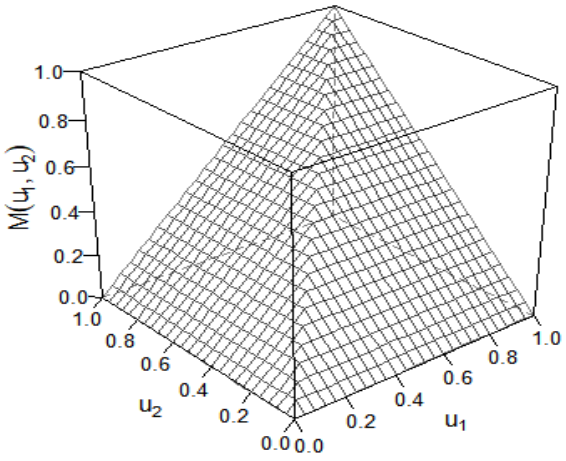
Το Copula άνω φράγματος δίνεται από τον τύπο:

$$C_U(u_1,u_2) = min\{u_1, u_2\} \text{ , } \boldsymbol{u} \in [0,1]^2$$

Όπως φαίνεται στο παρακάτω σχήμα το συγκεκριμένο copula αντιστοιχεί γραφικά σε μονάδα μάζας κατανεμημένη πάνω από την κύρια διαγώνιο $u_1 = u_2$ του μοναδιαίου τετραγώνου. Αν X_1 και X_2 τυχαίες μεταβλητές με συνάρτηση κατανομής $F_X(x_1,x_2) = C(F_1(x_1), F_2(x_2))$, τότε η X_2 θα είναι μη φθίνουσα ως προς την X_1 αν και μόνο αν $C \equiv C_U$.

Μπορούμε να κάνουμε μία γραφική αναπαράσταση χρησιμοποιώντας τον παρακάτω κώδικα στην R:

```
library(copula)
n.grid <- 26
u <- seq(0, 1, length.out = n.grid)
grid <- expand.grid("u[1]" = u, "u[2]" = u)
M <- function(u) apply(u, 1, min)
x.M <- cbind(grid, "M(u[1],u[2])" = M(grid))
wireframe2(x.M)
```



Σχήμα 3.1.1.: Άνω φράγμα δισδιάστατου Copula

Copula Κάτω Φράγματος (Fréchet Lower Bound Copula)

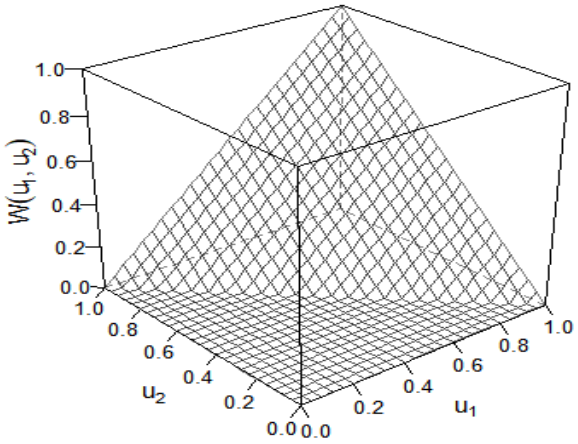
Το Copula κάτω φράγματος δίνεται από τον τύπο:

$$C_L(u_1,u_2) = max\{0\ ,\ u_1+u_2\text{-}1\}\ ,\ \boldsymbol{u} \in [0,1]^2$$

Όπως φαίνεται στο παρακάτω σχήμα το συγκεκριμένο copula αντιστοιχεί γραφικά σε μονάδα μάζας κατανεμημένη πάνω από τη διαγώνιο $u_1 = 1 - u_2$ του μοναδιαίου τετραγώνου. Αν X_1 και X_2 τυχαίες μεταβλητές με συνάρτηση κατανομής $F_X(x_1,x_2) = C(F_1(x_1), F_2(x_2))$, τότε η X_2 θα είναι μη αύξουσα ως προς την X_1 αν και μόνο αν $C \equiv C_L$.

Μπορούμε να κάνουμε μία γραφική αναπαράσταση χρησιμοποιώντας τον παρακάτω κώδικα στην R:

```
library(copula)
n.grid <- 26
u <- seq(0, 1, length.out = n.grid)
grid <- expand.grid("u[1]" = u, "u[2]" = u)
W <- function(u) pmax(0, rowSums(u)-1)
x.W <- cbind(grid, "W(u[1],u[2])" = W(grid))
wireframe2(x.W)
```



Ενδιαφέρον παρουσιάζει η συμπεριφορά των copulas σε γνησίως μονότονους μετασχηματισμούς των τυχαίων μεταβλητών. Αν *X, Y* συνεχείς τυχαίες μεταβλητές με copula C_{XY} και πεδία τιμών $R(X), R(Y)$ αντίστοιχα και α, b γνησώς μονότονες συναρτήσεις θα ισχύουν τα παρακάτω:

- Για a, b γνησίως αύξουσες

$$C_{a(X),b(Y)}(u_1, u_2) = C_{XY}(u_1, u_2)$$

- Για a, b γνησίως φθίνουσες

$$C_{a(X),b(Y)}(u_1, u_2) = u_1 + u_2 - 1 + C_{XY}(1 - u_1, 1 - u_2)$$

- Για a γνησίως αύξουσα και b γνησίως φθίνουσα

$$C_{a(X),b(Y)}(u_1, u_2) = u_1 - C_{XY}(u_1, 1 - u_2)$$

- Για a γνησίως φθίνουσα και b γνησίως αύξουσα

$$C_{a(X),b(Y)}(u_1, u_2) = u_2 - C_{XY}(1 - u_1, u_2)$$

Όπως φαίνεται από τις παραπάνω σχέσεις το copula παραμένει αναλλοίωτο σε γνησίως αύξοντες μετασχηματισμούς, ενώ στους υπόλοιπους γνησίως μονότονους μετασχηματισμούς προκύπτει μέσω τύπων που είναι σχετικά απλοί στον υπολογισμό. Επίσης, όπως μπορούμε να διακρίνουμε τα δεξιά μέλη των ισοτήτων και στις τέσσερις περιπτώσεις είναι ανεξάρτητα των α και b, που σημαίνει ότι η επιλογή τους δεν παίζει κανέναν ρόλο.

Αφού όπως είδαμε μέσω του θεωρήματος του Sklar καταφέραμε να ορίσουμε το copula σαν από κοινού συνάρτηση κατανομής, θα ήταν λογικό να μας απασχολήσει και η σχέση του με την συνάρτηση πυκνότητας. Θα δούμε ότι η πυκνότητα του copula περιέχει όλη την πληροφορία για την εξάρτηση των X_i. Αν $\tilde{X} = (X_1, X_2)$ συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 και συναρτήσεις πυκνότητας f_1, f_2 αντίστοιχα, η από κοινού συνάρτηση πυκνότητας των X_1, X_2 δίνεται από τον τύπο:

$$f_X(x) = f_1(x_1)\, f_2(x_2)\, c(F_1(x_1), F_2(x_2)),\ \mathrm{x} \in \mathbb{R}^2$$

Όπου η συνάρτηση πυκνότητας του copula c θα δίνεται από τη σχέση:

$$c(u_1,u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1,u_2),\ u \in [0,1]^2$$

όπου η παράγωγος $\frac{\partial^2}{\partial u_1 \partial u_2} C(u_1,u_2)$ υπάρχει σχεδόν παντού στο $[0,1]^2$.

Η παράγωγος του copula C, όπου υπάρχει, μπορεί να ερμηνευτεί σαν ένα τοπικό μέτρο εξάρτησης. Πιο συγκεκριμένα, στο σημείο x, το μέρος $f_1(x_1)\, f_2(x_2)$ του γινομένου συνδέεται με την ανεξαρτησία και το υπόλοιπο μέρος, $c(F_1(x_1), F_2(x_2))$ με την εξάρτηση. Αν δηλαδή, τα X_i είναι ανεξάρτητα τότε $c \equiv 1$ και η από κοινού συνάρτηση πυκνότητας απλοποιείται στο γινόμενο $f_1(x_1)\, f_2(x_2)$.

3.2 Πολυδιάστατα Copulas

Στην παράγραφο αυτή θα κάνουμε μια μικρή αναφορά στη γενίκευση των copulas σε n-διαστάσεις και κυρίως στη μορφή που παίρνουν τα φράγματα Fréchet-Hoeffding και το θεώρημα του Sklar. Επεκτείνοντας τον ορισμό του copula σε n-διαστάσεις θα μπορούσαμε να πούμε ότι:

Ένα πολυδιάστατο copula είναι μια συνάρτηση C με πεδίο ορισμού $[0,1]^n$ τέτοια ώστε:

i. Για κάθε $\mathrm{u} \in [0,1]^n$

$$C(u) = 0$$

αν έστω ένα στοιχείο του u είναι μηδενικό.

ii. Αν όλα τα στοιχεία του **u** είναι ίσα με τη μονάδα εκτός ενός το οποίο ισούται με u_k, τότε:

$$C(u) = u_k$$

Αυτό σημαίνει ότι ορίζονται οι περιθώριες.

iii. Για κάθε $a, b \in \mathrm{I}^n$ με $a \leq b$ ισχύει

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

όπου V_C είναι ο C-όγκος του $[a, b]$.

Θα πρέπει στο σημείο αυτό να εξηγήσουμε την έννοια του όγκου ώστε να γίνει κατανοητή η ιδιότητα iii. Έστω $S_1,\ldots,S_n$ μη κενά υποσύνολα του $\bar{R}$, όπου $\bar{R}$ η εκτεταμένη πραγματική ευθεία $[-\infty,\infty]$. Έστω H συνάρτηση n μεταβλητών τέτοια ώστε Dom $H = S_1 \times \ldots \times S_n$ και για $\boldsymbol{a} \leq \boldsymbol{b}$ ($a_k \leq b_k$,για κάθε k), έστω $\mathrm{B} = [\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times \ldots \times [a_n, b_n]$ είναι ένα n-κουτί του οποίου οι πλευρές περιέχονται στο πεδίο ορισμού Dom H. Τότε ο H-όγκος του B δίνεται από τον τύπο:

$$V_H(B) = \sum sgn(c)H(c),$$

$$\text{όπου } \operatorname{sgn}(c) = \begin{cases} 1, & c_k = a_k \, για \; k \; άρτιο \\ -1, & c_k = a_k \, για \; k \; περιττό \end{cases}$$

Ισοδύναμα, ο H-όγκος ενός n-κουτιού $B = [a,b]$ είναι n-τάξης διαφορά του H πάνω στο B δηλαδή,

$$V_H(B) = \Delta_a^b H(t) = \Delta_{a_n}^{b_n} \ldots . \Delta_{a_1}^{b_1} H(t), \text{ όπου}$$

$$\Delta_{a_k}^{b_k} H(t) = H(t_1, \ldots t_{k-1}, b_k, t_{k+1}, \ldots t_n) - H(t_1, \ldots . t_{k-1}, a_k, t_{k+1}, \ldots t_n)$$

Στην προηγούμενη παράγραφο είδαμε ότι πολύ σημαντικό ρόλο στην θεωρία των copulas παίζει το θεώρημα του Sklar καθώς μας επιτρέπει να εκφράσουμε τις εν λόγω συναρτήσεις σαν συναρτήσεις κατανομής. Η γενίκευση του θεωρήματος σε n-διαστάσεις δίνεται παρακάτω.

Έστω H μια n-διάστατη συνάρτηση κατανομής με περιθώριες $F_1, \ldots, F_n$. Τότε υπάρχει ένα n-copula C τέτοιο ώστε, για κάθε $\boldsymbol{x}$ στο $\overline{R^n}$,

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

Αν οι $F_1, \ldots, F_n$ είναι συνεχείς, τότε η συνάρτηση C είναι μοναδική, σε διαφορετική περίπτωση η C είναι μοναδικά ορισμένη επάνω στο σύνολο $Ran\ F_1 \times \ldots \times Ran\ F_n$ (όπου $Ran\ F_i$, το πεδίο τιμών της F_i, $i = 1, \ldots n$) (βλ. Nelsen Roger, 2006).

Αντίστροφα, αν η συνάρτηση C είναι ένα n-copula και $F_1, \ldots, F_n$ είναι συναρτήσεις κατανομής, τότε η συνάρτηση H είναι μια n-διάστατη συνάρτηση κατανομής με περιθώριες $F_1, \ldots, F_n$.

Όπως είναι επόμενο, φαίνεται πως αν έχουμε n το πλήθος περιθώριες αρκεί να κάνουμε κατάλληλη επιλογή του copula και μπορούμε να κατασκευάσουμε την n-διάστατη από κοινού συνάρτης κατανομής.

Όπως έχουμε αναφέρει στη δισδιάστατη περίπτωση κάθε copula φράσσεται από τις συναρτήσεις M και W. Γενικεύοντας τα φράγματα Fréchet – Hoeffding σε n διαστάσεις ισχύει ότι, αν C copula τότε για κάθε $u \in [0,1]^n$ θα έχουμε ότι:

$$W^n(\tilde{u}) \leq C(\tilde{u}) \leq M^n(\tilde{u}),$$

όπου οι συναρτήσεις M'', Π'', W'' είναι ορισμένες στο σύνολο $[0,1]^n$ ως εξής:

$$M^n(\tilde{u}) = min(u_1, \ldots, u_n)$$

$$\Pi^n(\tilde{u}) = u_1,\ldots,u_n$$

$$W^n(\tilde{u}) = \max(u_1+\ldots+ u_n - n +1 ,0)$$

Επιπλέον, οι συναρτήσεις M^n και Π^n είναι copulas για κάθε $n \geq 2$, ενώ η συνάρτηση W^n δεν είναι copula για οποιαδήποτε $n \geq 3$, καθώς:

Αν θεωρήσουμε τον n-κύβο $[1/2, 1]^n \subset [0,1]^n$. Τότε:

$V_{w^n}([1/2, 1]^n) = max(1 +\ldots.+ 1 - n +1 ,0) - n \max(1/2 + 1 +\ldots.+1 - n +1, 0) + \binom{n}{2} max(1/2 + 1/2 + 1 +\ldots.+1 - n +1, 0) +\ldots. + max(1/2 + 1/2 - n +1, 0) = 1 - n/2 + 0 +\ldots.+ 0 < 0.$

Συνεπώς η συνάρτηση W^n δεν είναι copula καθώς για $n \geq 3$ δεν ικανοποιείται η ιδιότητα (iii) του ορισμού. Παρά το γεγονός όμως ότι η συνάρτηση W^n δεν είναι copula για $n \geq 3$, συνεχίζει να αποτελεί το καλύτερο δυνατό κάτω φράγμα. Αυτό συμβαίνει γιατί μπορεί να αποδειχτεί ότι για κάθε $n \geq 3$ και για κάποιο u στο $[0,1]^n$ υπάρχει ένα n-copula τέτοιο ώστε:

$$C(u) = W^n(u)$$

Συνοπτικά θα μπορούσαμε να πούμε ότι το Fréchet – Hoeffding κάτω φράγμα W^2 είναι μικρότερο από κάθε δισδιάστατο copula και κάθε n-διάστατο copula είναι μικρότερο από το Fréchet – Hoeffding άνω φράγμα M^n. Αυτή η μερική διάταξη του συνόλου των copulas καλείται διάταξη συμφωνίας. Λέμε μερική γιατί δεν είναι κάθε ζεύγος copulas συγκρίσιμο μέσω της διάταξης αυτής, όμως πολλές σημαντικές παραμετρικές οικογένειες copulas είναι πλήρως διατεταγμένες.

4. Συσχέτιση – Εξάρτηση – Ανεξαρτησία

Η πρώτη ερώτηση που θα πρέπει να απαντηθεί, είναι πότε ακριβώς υπάρχει εξάρτηση μεταξύ κινδύνων; Μια αρχική προσέγγιση είναι να εξετάσουμε την έννοια της εξάρτησης μέσω της έννοιας της ανεξαρτησίας. Όπως είναι γνωστό από τη θεωρία πιθανοτήτων δύο τυχαίες μεταβλητές X, Y (που για εμάς θα αντιπροσωπεύουν κινδύνους) θα λέμε ότι είναι ανεξάρτητες αν η από κοινού συνάρτηση κατανομής τους ισούται με το γινόμενο των δύο περιθώριων συναρτήσεων κατανομής των δύο τυχαίων μεταβλητών, δηλαδή αν ισχύει η ισότητα:

$$\mathrm{P}(X \leq x, Y \leq y) = \mathrm{P}(X \leq x)\ \mathrm{P}(Y \leq y), \text{ για κάθε } x, y \in \mathbb{R}.$$

Διαφορετικά, θα μπορούσαμε να πούμε ότι η δεσμευμένη συνάρτηση κατανομής της Y δεδομένης της X δεν εξαρτάται από τη X αλλά ισούται με τη συνάρτηση κατανομής της τυχαίας μεταβλητής Y. Δηλαδή,

$$\mathrm{P}(Y \leq y \mid X \leq x) = \mathrm{P}(X \leq x, Y \leq y) \,/\, \mathrm{P}(X \leq x) = \mathrm{P}(Y \leq y), \text{ για κάθε } x, y \in \mathbb{R}.$$

Η ανάγκη να οριστεί η εξάρτηση δύο τυχαίων μεταβλητών έμμεσα, μέσω του ορισμού της ανεξαρτησίας υποδηλώνει ότι υπάρχουν πολλοί τρόποι με τους οποίους οι κίνδυνοι μπορεί να είναι εξαρτημένοι μεταξύ τους. Μία οποιαδήποτε απλή ποσοτικοποίηση της εξάρτησης, όπως για παράδειγμα η εξαγωγή ενός αριθμού από την από κοινού συνάρτηση κατανομής δύο τυχαίων μεταβλητών, αντικατοπτρίζει μία μόνο πτυχή της εξάρτησης ή την περιγράφει πλήρως μόνο σε πολύ ειδικές περιπτώσεις. Στην παράγραφο αυτή θα αναφέρουμε τις επιθυμητές ιδιότητες για τα μέτρα εξάρτησης καθώς και κάποια από τα σημαντικότερα εξ' αυτών.

4.1 Επιθυμητές Ιδιότητες των Μέτρων Εξάρτησης

Πρίν καθορίσουμε τις εν λόγω ιδιότητες θα πρέπει να ορίσουμε την έννοια της συμμονοτονικότητας δύο τυχαίων μεταβλητών.

Για κάθε copula ισχύει:

$$max\{x_1 + \ldots. + \ x_n + 1 - n\ , 0\ \} \leq C(x_1, \ldots., x_n) \leq min\{x_1, \ldots., x_n\}$$

το οποίο προκύπτει από το γεγονός ότι κάθε copula είναι μια συνάρτηση κατανομής ενός τυχαίου διανύσματος $\widetilde{U} = (U_1, \ldots., U_t)$ με $U_i \sim U(0,1)$. Στην περίπτωση που ισχύει $n = 2$ τα φράγματα C_l και C_u είναι επίσης copulas καθώς αν $U \sim U(0,1)$ τότε:

$$C_l(x_1, x_2) = \mathrm{P}[U \leq x_1\ ,\ 1 - U \leq x_2]$$

$$C_u(x_1,x_2) = \mathrm{P}[U \leq x_1 , U \leq x_2]$$

έτσι ώστε C_l και C_u είναι οι δισδιάστατες συναρτήσεις κατανομής των διανυσμάτων (U , 1 - U) και (U , U) αντίστοιχα. Η κατανομή του(U , 1 - U) έχει όλη τη μάζα της στη διαγώνιο ανάμεσα στο (0,1) και (1,0) ενώ του (U, U) στη διαγώνιο ανάμεσα στο (0,0) και (1,1). Στις περιπτώσεις αυτές λέμε ότι C_l και C_u περιγράφουν τέλεια θετική και τέλεια αρνητική εξάρτηση αντίστοιχα. Αυτό συνοψίζεται στον παρακάτω ορισμό.

Αν το (X, Y) έχει copula C_u, τότε οι τυχαίες μεταβλητές X, Y θα λέμε ότι είναι συμμονοντονικές (comonotonic), ενώ αν έχουν copula C_l θα λέμε ότι είναι αντιστρόφως μονοτονικές (countermonotonic).

Ένα μέτρο εξάρτησης συνοψίζει τη δομή εξάρτησης δύο τυχαίων μεταβλητών σε έναν αριθμό. Αυτό που θα θέλαμε είναι να καθορίσουμε ένα σύνολο ιδιοτήτων που θα ήταν επιθυμητό να πληροί κάθε μέτρο. Έστω δ ένα μέτρο εξάρτησης το οποίο αντιστοιχεί έναν πραγματικό αριθμό σε ένα ζευγάρι πραγματικών τυχαίων μεταβλητών X και Y. Ιδανικά, θα θέλαμε ένα μέτρο εξάρτησης να έχει τις παρακάτω ιδιότητες:

i. $\delta(X, Y) = \delta(Y, X)$ (συμμετρία)

ii. $-1 \leq \delta(X, Y) \leq 1$ (κανονικότητα)

iii. $\delta(X, Y) = 1 \Leftrightarrow X, Y$ είναι συμμονοτονικά

$\delta(X, Y) = -1 \Leftrightarrow X, Y$ είναι αντιστρόφως μονοτονικά

iv. Αν T: $\mathbb{R} \rightarrow \mathbb{R}$ είναι μία γνησίως μονότονη συνάρτηση στο πεδίο τιμών της X, τότε:

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

Προφανώς οι παραπάνω ιδιότητες αποτελούν μια επιλογή από μία λίστα ιδιοτήτων η οποία μπορεί να διαφοροποιείται ή να επεκταθεί. Για παράδειγμα θα μπορούσε να μας ενδιαφέρει η ιδιότητα:

v. Αν $\delta(X, Y) = 0 \Leftrightarrow X, Y$ ανεξάρτητες

η οποία όμως έρχεται σε αντίθεση με την ιδιότητα iv, καθώς αποδεικνύεται ότι δεν υπάρχει μέτρο εξάρτησης (συσχέτισης) που να ικανοποιεί τις iv και v. Αν θέλαμε να απαιτήσουμε την ιδιότητα v, θα έπρεπε αναζητήσουμε μέτρα τα οποία αντιστοιχούν μόνο θετικές τιμές σε ζεύγη τυχαίων μεταβλητών.

Αν περιοριστούμε μόνο σε συνεχείς τυχαίες μεταβλητές τότε υπάρχουν μέτρα που ικανοποιούν και τις πέντε ιδιότητες αλλά είναι περισσότερο θεωρητικού παρά πρακτικού ενδιαφέροντος.

4.2 Συντελεστής γραμμικής συσχέτισης του Pearson

Στην παράγραφο αυτή θα παρουσιάσουμε την γραμμική συσχέτιση (linear correlation) σαν ένα μέτρο γραμμικής εξάρτησης (linear dependence). Έστω ένα ζευγάρι τυχαίων μεταβλητών X, Y, όπου καθεμιά από αυτές αντιπροσωπεύει έναν κίνδυνο και για τις οποίες θεωρούμε ότι $E(X^2) < \infty$, δηλαδή έχουν πεπερασμένη ροπή δεύτερης τάξης. Τότε, όπως είναι γνωστό από τη θεωρία πιθανοτήτων η ποσότητα

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

καλείται συντελεστής γραμμικής συσχέτισης του Pearson.

Υπενθυμίζουμε ότι η ποσότητα $\text{cov}(X,Y) = E(XY) - E(X)\,E(Y)$ καλείται συνδιακύμανση των τυχαίων μεταβλητών X, Y και $\text{Var}(X)$, $\text{Var}(Y)$ είναι οι διακυμάνσεις των X, Y αντίστοιχα. Η συνδιακύμανση περιέχει πληροφορία για την εξάρτηση των τυχαίων μεταβλητών X, Y, όμως επηρεάζεται από τη διακύμανσή τους. Για να ξεπεράσουμε αυτό το πρόβλημα και να πάρουμε ένα αποτέλεσμα που μετράει μόνο τις πτυχές της εξάρτησης, κανονικοποιούμε τη συνδιακύμανση διαιρώντας τη με το γινόμενο των τυπικών αποκλίσεων των εμπλεκόμενων τυχαίων μεταβλητών. Ισχύει επίσης ότι:

$$-1 \leq \rho(X,Y) \leq 1$$

και

$$\text{αν } X, Y \text{ είναι ανεξάρτητες τότε } \text{Cov}(X,Y) = 0 \text{ και άρα } \rho(X,Y) = 0.$$

Η συσχέτιση είναι ένα μέτρο γραμμικής εξάρτησης. Πιο συγκεκριμένα αν έχουμε τέλεια γραμμική εξάρτηση τότε και $\rho(X,Y) = \pm 1$.

Ο συντελεστής συσχέτισης χρησιμοποιείται ευρέως και αυτό μπορεί να αποδοθεί σε κάποιους λόγους, μερικούς από τους οποίους θα αναφέρουμε παρακάτω:

- Είναι σχετικά απλό να υπολογιστεί. Για πολλές δισδιάστατες κατανομές είναι εύκολο να υπολογιστούν οι ροπές δεύτερης τάξης άρα να υπολογιστεί ο συντελεστής συσχέτισης, σε αντίθεση με άλλα μέτρα εξάρτησης που έχουν πιο πολύπλοκους υπολογισμούς.

- Η διακύμανση και η συνδιακύμανση είναι εύκολο να υπολογιστούν κάτω από γραμμικούς μετασχηματισμούς αφού έχουμε ότι:

$$\mathrm{Cov}(A\boldsymbol{X} + \boldsymbol{a}, B\boldsymbol{Y} + \boldsymbol{\beta}) = A\ \mathrm{Cov}[\boldsymbol{X}, \boldsymbol{Y}]B^t.$$

Μία ειδική περίπτωση είναι η παρακάτω σχέση μεταξύ διακύμανσης και συνδιακύμανσης ενός τυχαίου διανύσματος. Για κάθε γραμμικό συνδυασμό των $\boldsymbol{a}^t\tilde{X}$, με $\boldsymbol{a} \in \mathbb{R}^n$,

$$\mathrm{Var}(\boldsymbol{a}^t\boldsymbol{X}) = \boldsymbol{a}^t\mathrm{Cov}[\boldsymbol{X}]a$$

Συνεπώς η διακύμανση οποιουδήποτε γραμμικού συνδυασμού καθορίζεται πλήρως από τις ανά ζεύγη συνδιακυμάνσεις μεταξύ των στοιχείων του $\boldsymbol{X}$.

- Αποτελεί το φυσικό μέτρο εξάρτησης για πολυδιάστατες κανονικές κατανομές και γενικά για σφαιρικές και ελλειπτικές κατανομές.

Στη συνέχεια θα αναφέρουμε μερικά απο τα μειονεκτήματα που παρουσιάζει ο γραμμικός συντελεστής συσχέτισης.

- Οι διακυμάνσεις των τυχαίων μεταβλητών $\boldsymbol{X, Y}$ πρέπει να είναι πεπερασμένες, καθώς σε διαφορετική περίπτωση η γραμμική συσχέτιση δεν ορίζεται. Αυτό δεν είναι ιδανικό για ένα μέτρο εξάρτησης και δημιουργεί προβλήματα όταν θέλουμε να εξετάσουμε κατανομές με βαριές ουρές. Για παράδειγμα, η συνδιακύμανση και η συσχέτιση μεταξύ δύο στοιχείων ενός τυχαίου διανύσματος που ακολουθούν την δισδιάστατη $\boldsymbol{t_\nu}$ κατανομή, δεν ορίζεται για $\boldsymbol{\nu} \leq 2$.

- Η ανεξαρτησία δύο τυχαίων μεταβλητών υποδηλώνει ότι είναι ασυσχέτιστες (δηλαδή η γραμμική συσχέτιση είναι ίση με μηδέν) αλλά γενικά το αντίστροφο δεν ισχύει. Ένα απλό παράδειγμα που παρά την ισχυρή εξάρτηση ο συντελεστής μηδενίζεται είναι να θεωρήσουμε την τυχαία μεταβλητή $X \sim \mathrm{N}(0,1)$ και $Y = X^2$, καθώς η τρίτη ροπή της τυπικής κανονικής κατανομής ισούται με μηδέν. Μόνο στην περίπτωση της πολυδιάστατης κανονικής κατανομής μπορούμε να ερμηνεύσουμε την μη ύπαρξη συσχέτισης ως απουσία εξάρτησης.

- Η γραμμική συσχέτιση δεν παραμένει αναλλοίωτη κάτω από μη-γραμμικούς γνησίως αύξοντες μετασχηματισμούς $\boldsymbol{T}$: $\mathbb{R} \rightarrow \mathbb{R}$. Δηλαδή για δύο τυχαίες μεταβλητές ισχύει γενικά:

$$\rho(T(X), T(Y)) \neq \rho(X, Y).$$

Επιπλέον, όπως γίνεται κατανοητό αυτό το μέτρο εξάρτησης έχει μόνο τις δύο πρώτες ιδιότητες (συμμετρία, κανονικότητα) από αυτές που αναφέρθηκαν στην προηγούμενη παράγραφο.
Η συσχέτιση είναι απλώς ένα μέτρο στοχαστικής εξάρτησης ανάμεσα σε πολλά άλλα. Είναι το μέτρο το οποίο μπορεί να χρησιμοποιηθεί όταν κινούμαστε στο χώρο των πολυδιάστατων κατανομών και κυρίως των σφαιρικών και ελλειπτικών κατανομών. Όμως στον χρηματοπιστωτικό και ασφαλιστικό χώρο οι περισσότερες κατανομές που συναντάμε στα προβλήματα με τα οποία ερχόμαστε αντιμέτωποι, σπάνια

ανήκουν σ'αυτή την κλάση, οπότε η χρήση της συσχέτισης δεν μπορεί να τεκμηριωθεί ικανοποιητικά. Για το λόγο αυτό υπάρχει ανάγκη να χρησιμοποιηθούν άλλα μέτρα εξάρτησης που περιγράφουν ακριβέστερα τη δομή της εξάρτησης μεταξύ των τυχαίων μεταβλητών.

4.3 Συντελεστές συσχέτισης τάξης

Με τον όρο απόδοση τάξης (ranking) θα αναφερόμαστε σε έναν μετασχηματισμό των δεδομένων κατά τον οποίο οι αριθμητικές τιμές αντικαθίστανται από τις τάξεις τους. Για παράδειγμα, αν έχουμε τα αριθμητικά δεδομένα 2,9,7,5 τότε οι αντίστοιχες τάξεις θα είναι 1,4,3,2. Η τάξη δηλαδή δηλώνει τη σχετική θέση που θα είχε μια παρατήρηση στην περίπτωση που τα δεδομένα ήταν διατεταγμένα κατά αύξουσα σειρά. Προφανώς, δύο ή περισσότερες παρατηρήσεις θα μπορούσαν να έχουν ίδια τάξη στην περίπτωση που μέσα στο δείγμα παρατηρηθεί μία τιμή περισσότερες από μια φορές. Αντιμετωπίζουμε αυτή την κατάσταση, την οποία καλούμε "δεσμούς" (ties), με έναν απλό τρόπο τον οποίο θα εξηγήσουμε μέσω του ακόλουθου παραδείγματος. Ας υποθέσουμε ότι το δείγμα μας αποτελείται από τις παρατηρήσεις 2,7,9,7,5. Τότε στις ίδιες παρατηρήσεις αποδίδουμε σαν τάξη τον μέσο όρο τους και το δείγμα θα μετατρεπόταν σε 1, 3.5 , 5 , 3.5 ,2. Αντίστοιχα θα πράτταμε στην περίπτωση που είχαμε τρείς η περισσότερες ισοπαλίες.

Για να αποκτήσουμε μια εικόνα του ρόλου που παίζει η αντικατάσταση των τιμών ενός δείγματος από τις αντίστοιχες τάξεις τους, θα φέρουμε ένα παράδειγμα το οποίο περιέχεται στην εργασία των Genest και Favre (2007) το οποίο κάνει χρήση της ιδιότητας που έχουν τα copulas να παραμένουν αμετάβλητα σε γνησίως αύξοντες μετασχηματισμούς των τυχαίων μεταβλητών. Ο παρακάτω πίνακας περιέχει δύο τυχαία δείγματα μεγέθους $n = 6$ τα οποία έχουν παραχθεί με προσομοίωση από την τυπική κανονική κατανομή $N(0,1)$. Οι συγκεκριμένες τιμές χρησιμοποιούνται και στην εργασία που προαναφέραμε, όμως αν θέλει κανείς να παράγει τυχαίες τιμές που προέρχονται από την τυπική κανονική κατανομή, θα μπορούσε να χρησιμοποιήσει την εντολή $rnorm(n, 0,1)$ στο στατιστικό πρόγραμμα R, όπου n εισάγουμε το πλήθος των τιμών που θέλουμε να παράγουμε.

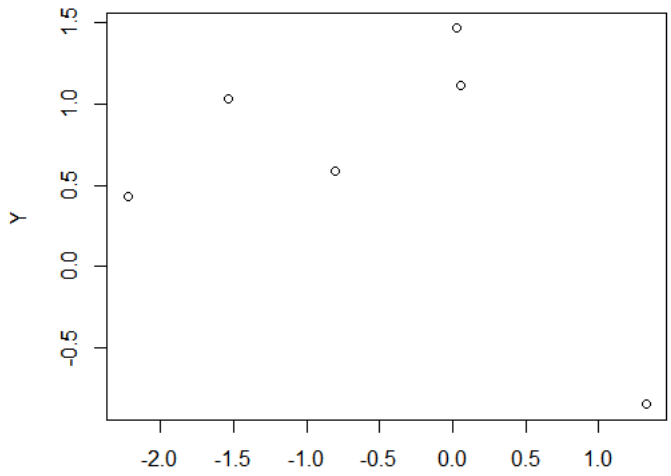
i	1	2	3	4	5	6
X_i	-2,224	-1,538	-0,807	0,024	0,052	1,324
Y_i	0,431	1,035	0,586	1,465	1,115	-0,847

Πίνακας 4.3.1.: Learning data (Genest C., Ghoudi K. & Rivest L. – P., 1995)

Το διάγραμμα διασποράς (scatter plot) που προκύπτει από τα ζεύγη τιμών (X_i, Y_i) μπορεί να μας δώσει μια εικόνα για την δομή εξάρτησης που μπορεί να υπάρχει μεταξύ δύο τυχαίων μεταβλητών. Για το συγκεκριμένο δείγμα το διάγραμμα θα έχει την παρακάτω μορφή.

Ο κώδικας που χρησιμοποιήθηκε στην R για την κατασκευή του σχήματος 4.3.2 είναι ο ακόλουθος:

```
X <- c(-2.224,-1.538,-0.807,0.024,0.052,1.324)
Y <- c(0.431,1.035,0.586,1.465,1.115,-0.847)
plot(X, Y, xlab = "X", ylab = "Y")
```



Σχήμα 4.3.2.: Scatter plot των ζευγών (X_i, Y_i)

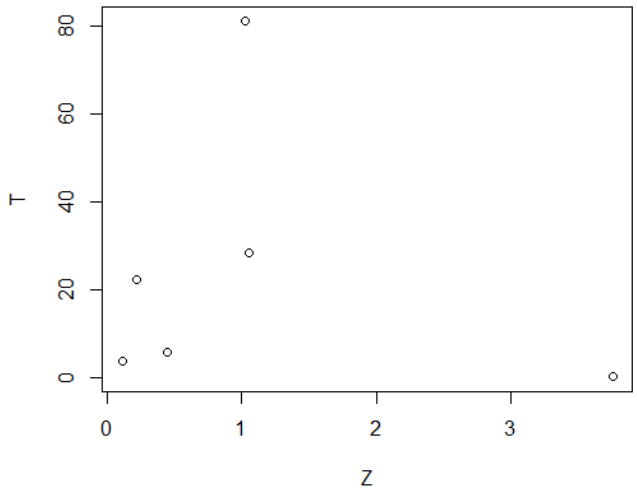
Αν θεωρήσουμε τους μετασχηματισμούς,

$$Z_i = e^{X_i} \text{ και } T_i = e^{3Y_i}, 1 \leq i \leq 6$$

παρατηρούμε ότι το διάγραμμα διασποράς των μετασχηματισμένων τυχαίων μεταβλητών είναι προφανώς πολύ διαφορετικό από το προηγούμενο, όπως φαίνεται παρακάτω.

Ο κώδικας που χρησιμοποιήθηκε στην R για την κατασκευή του σχήματος 4.3.2 είναι ο ακόλουθος:

```
Z <- exp(X)
T <- exp(3*Y)
plot(Z,T,xlab="Z",ylab="T")
```



Σχήμα 4.3.3.: Scatter plot των ζευγών $(Z_i, T_i) = (e^{X_i}, e^{3Y_i})$

Τα διαγράμματα αυτά αναπαριστούν μια "θολή" εικόνα της δομής εξάρτησης που υπάρχει μεταξύ των ζευγών (X, Y) και των ζευγών (Z, T) η οποία όμως περιγράφεται από το ίδιο copula και στις δύο περιπτώσεις καθώς, αν θεωρήσουμε δύο αύξοντες μετασχηματισμούς φ, ψ με αντίστροφες φ^{-1} και ψ^{-1} θα ισχύει:

$$H^*(z,t) = C^*\{F^*(z), G^*(t)\},$$

το οποίο προκύπτει από το θεώρημα του Sklar. Οι περιθώριες συναρτήσεις κατανομής των Z, T προκύπτουν από τις σχέσεις:

$$F^*(z) = P(Z \leq z) = P\{X \leq \varphi^{-1}(z)\} = F\{\varphi^{-1}(z)\}$$

και

$$G^*(t) = P(T \leq t) = P\{Y \leq \psi^{-1}(t)\} = G\{\psi^{-1}(t)\}$$

Συνεπώς,

$$\begin{aligned} H^*(z,t) &= P(Z \leq z, T \leq t) \\ &= P\left(X \leq \varphi^{-1}(z), Y \leq \psi^{-1}(t)\right) \\ &= H\left(\varphi^{-1}(z), \psi^{-1}(t)\right) \\ &= C[F\{\varphi^{-1}(z)\}, G\{\psi^{-1}(t)\}] \\ &= C\{F^*(z), G^*(t)\}, \end{aligned}$$

για κάθε $z, t \in \mathbb{R}$. Αποδεικνύεται δηλαδή η ιδιότητα του αναλλοίωτου. Αφού η δομή εξάρτησης περιγράφεται από το ίδιο copula, θα πρέπει να είναι δυνατό να απεικονιστεί αυτό σχηματικά και να παίρνουμε την ίδια εικόνα είτε έχουμε το ζευγάρι (X, Y) είτε το (Z, T). Αν αντικαταστήσουμε τις τιμές των X_i και των Y_i με R_i και S_i που θα αντιστοιχούν στις τάξεις τους θα προκύψει ο παρακάτω πίνακας.

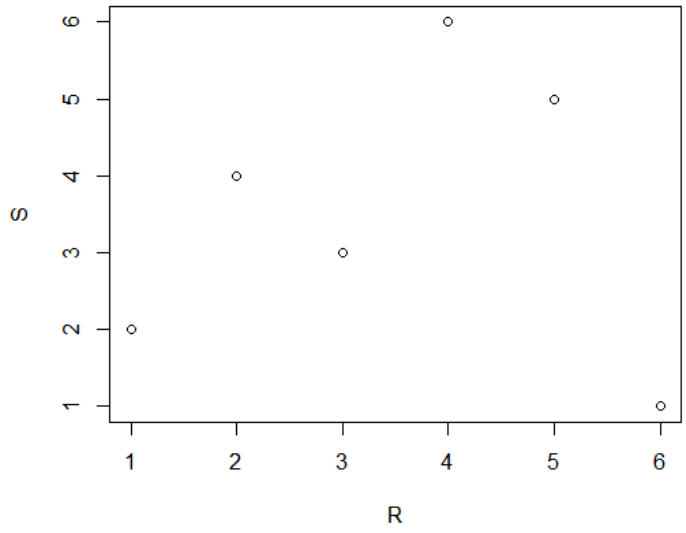
i	1	2	3	4	5	6
Ri	1	2	3	4	5	6
Si	2	4	3	6	5	1

Σχήμα 4.3.4.: Τα μετασχηματισμένα με βάση τις τάξεις δεδομένων του πίνακα 4.3.1. (Genest C., Ghoudi K. & Rivest L. – P., 1995)

Ο ίδιος ακριβώς πίνακας θα προκύψει και με αντικατάσταση των Z_i, T_i από τα αντίστοιχα R_i^* και S_i^* και έτσι θα έχουμε και στις δύο περιπτώσεις την απεικόνιση που ακολουθεί.

Ο κώδικας που χρησιμοποιήθηκε στην R για την κατασκευή του σχήματος 4.3.2 είναι ο ακόλουθος:

```
S <- rank(X)
R <- rank(Y)
plot(S,R,xlab="R",ylab="S")
```



Σχήμα 4.3.5.: Scatter plot του μετασχηματισμού (R_i, S_i) του ζεύγους (X_i, Y_i) και του μετασχηματισμού (R_i^*, S_i^*)του ζεύγους (Z_i, T_i)

Ένα ενδιαφέρον στοιχείο είναι πως αν αναπροσαρμόσουμε τους άξονες πολλαπλασιάζοντας με την ποσότητα $\frac{1}{n+1}$ τα σημεία του διαγράμματος θα βρίσκονται μέσα στο σύνολο $[0,1]^2$, το οποίο αποτελεί το πεδίο ορισμού αυτού που ονομάζουμε εμπειρικό copula (empirical copula) και δίνεται από τον τύπο:

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} 1(A)$$

όπου, 1(A) η δείκτρια συνάρτηση του συνόλου A και A =($\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v$). Αποδεικνύεται ότι για κάθε ζεύγος (u,v) το $C_n(u,v)$ είναι ένας εκτιμητής βασισμένος στις τάξεις (rank-based estimator) της άγνωστης ποσότητας $C(u,v)$.

Οι συντελεστές Tau του Kendall και Rho του Spearman είναι οι πιο γνωστοί συντελεστές συσχέτισης τάξης, βασίζονται στην έννοια της συμφωνίας τυχαίων μεταβλητών, και για αυτούς θα μιλήσουμε στη συνέχεια.

4.3.1 Rho του Spearman

Ο συντελεστής Rho του Spearman είναι ένας από τους πιο διαδεδομένους συντελεστές συσχέτισης τάξης. Πρίν παρουσιάσουμε τον εκτιμητή του εν λόγω συντελεστή αλλά και την έκφρασή του στον πληθυσμό, έχει ενδιαφέρον να δούμε ότι μπορεί να αποτυπωθεί μέσω ενός μετασχηματισμού του συντελεστή γραμμικής συσχέτισης του Pearson. Συγκεκριμένα, αν X, Y είναι συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 και από κοινού συνάρτηση κατανομής F, ο συντελεστής δίνεται από τη σχέση:

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

Παρακάτω θα εξηγήσουμε γιατί αυτός ο συντελεστής χαρακτηρίζεται ως συντελεστής τάξης. Όπως είναι γνωστό με F^{-1} συμβολίζουμε την γενικευμένη αντίστροφης μια συνάρτησης κατανομής F, η οποία είναι η συνήθης αντίστροφη συνάρτηση της F, αν η F είναι γνησίως αύξουσα. Η $F_1(X)$ (τα αντίστοιχα θα ισχύουν και για την $F_2(Y)$) είναι μια τυχαία μεταβλητή η οποία παίρνει τιμές στο [0,1]. Επιπλέον, αν η F_1 είναι συνεχής θα ισχύει ότι:

$$P[F_1(X) \leq x] = P[X \leq F_1^{-1}(x)] = F_1(F_1^{-1}(x)) = x, \text{ για } x \in [0,1]$$

συνεπώς, η $F_1(X)$ είναι μία ομοιόμορφα κατανεμημένη τυχαία μεταβλητή στο διάστημα [0,1]. Άρα ο συντελεστής ρ_S μετράει τη συσχέτιση μεταξύ δύο ομοιόμορφα κατανεμημένων τυχαίων μεταβλητών γεγονός που σημαίνει ότι οι αρχικές τιμές των X και Y είναι άνευ σημασίας.

Ακολουθώντας τη λογική με βάση την οποία έχει κατασκευαστεί ο συντελεστής συσχέτισης του Pearson, μια ιδέα θα ήταν να υπολογίσουμε τη συσχέτιση μεταξύ των ζευγών (R_i, S_i) των τάξεων ή

ισοδύναμα μεταξύ των $(\frac{R_i}{n+1}, \frac{S_i}{n+1})$ τα οποία αποτελούν το στήριγμα (support) του C_n. Έτσι οδηγούμαστε στην παρακάτω έκφραση:

$$\bar{\rho}_S = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}} \in [-1, 1]$$

όπου

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{n+1}{2} = \frac{1}{n}\sum_{i=1}^{n} S_i = \bar{S}.$$

Μία πιο απλουστευμένη μορφή του συντελεστή είναι:

$$\bar{\rho}_S = \frac{12}{n\,(n+1)(n-1)}\sum_{i=1}^{n} R_i S_i - 3\,\frac{n+1}{n-1}$$

Θεωρητικά ο συντελεστής του Spearman είναι ανώτερος από του Pearson για τους εξής λόγους:

i. $E(\bar{\rho}_S) = \pm 1$ αν και μόνο αν οι τυχαίες μεταβλητές X, Y είναι συναρτησιακά εξαρτημένες (functionally dependent), δηλαδή το χαρακτηριστικό τους copula είναι ένα από τα φράγματα Frechet – Hoeffding, M ή Π. Σε αντίθεση, για τον συντελεστή του Pearson ισχύει ότι $E(\bar{\rho}) = \pm 1$ αν και μόνο αν οι X, Y είναι γραμμικές συναρτήσεις η μία της άλλης, γεγονός το οποίο είναι πολύ περισσότερο περιοριστικό.
ii. Ο $\bar{\rho}_S$ εκτιμά μία πληθυσμιακή παράμετρο η οποία είναι πάντοτε καλά ορισμένη σε αντίθεση με τον συντελεστή του Pearson ο οποίος δεν υπάρχει πάντα, όπως για παράδειγμα σε κάποιες κατανομές που έχουν βαριές ουρές (π.χ. Cauchy).

Αποδεικνύεται ότι ο $\bar{\rho}_S$ είναι ένας ασυμπτωτικός εκτιμητής της πληθυσμιακής ποσότητας:

$$\rho_S = 12 \int_0^1 \int_0^1 C(u,v)dudv - 3 = 12\,E(UV) - 3$$

Πιο αναλυτικά, αν (X_1,Y_1), (X_2,Y_2), (X_3,Y_3) είναι τρία ανεξάρτητα ζεύγη τυχαίων μεταβλητών με από κοινού συνάρτηση κατανομής H (της οποίας οι περιθώριες είναι F,G) τότε, ο συντελεστής συσχέτισης τάξης του Spearman ορίζεται ως:

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

Δηλαδή, ο συντελεστής είναι ανάλογος της πιθανότητας συμφωνίας μείον την πιθανότητα ασυμφωνίας των δύο τυχαίων διανυσμάτων (X_1,Y_1), (X_2,Y_3). Πρέπει να σημειωθεί ότι ενώ η από κοινού συνάρτηση κατανομής του (X_1,Y_1) είναι H(x,y), του (X_2,Y_3) είναι $F(x)G(y)$ καθώς οι X_2,Y_3 είναι ανεξάρτητες.

4.3.2 Tau του Kendall

Ο συντελεστής συσχέτισης του Kendall είναι ένα μη παραμετρικό μέτρο συσχέτισης το οποίο χρησιμοποιεί ζεύγη παρατηρήσεων τυχαίων μεταβλητών και εξετάζει κατά πόσο τα ζεύγη αυτά βρίσκονται σε συμφωνία (concordance) ή ασυμφωνία (discordance). Διαισθητικά, θα μπορούσαμε να πούμε ότι ένα ζευγάρι τυχαίων μεταβλητών θα είναι σε συμφωνία αν "μεγάλες" τιμές της μίας τείνουν να σχετίζονται με "μεγάλες" τιμές της άλλης και "μικρές" τιμές της μίας με "μικρές" τιμές της άλλης. Συγκεκριμένα, αν θεωρήσουμε δύο ζεύγη παρατηρήσεων (x_1,y_1) και (x_2,y_2) από το τυχαίο διάνυσμα (X, Y) και θα λέμε ότι:

Τα ζεύγη (x_1,y_1) και (x_2,y_2) βρίσκονται σε συμφωνία αν ισχύει

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

Τα ζεύγη (x_1,y_1) και (x_2,y_2) βρίσκονται σε ασυμφωνία αν ισχύει

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

Η εμπειρική μορφή του εκτιμητή δίνεται από τη σχέση:

$$\bar{\tau}_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n\,(n-1)}\,P_n - 1$$

όπου P_n το σύνολο των ζευγών που βρίσκονται σε συμφωνία και Q_n το σύνολο των ζευγών που βρίσκονται σε ασυμφωνία. Η τιμή του $\bar{\tau}_n$ δηλαδή, αντικατοπτρίζει το πόσο περισσότερα (ή λιγότερα) είναι τα ζεύγη που βρίσκονται σε συμφωνία σε σχέση με αυτά που βρίσκονται σε ασυμφωνία.

Ο συντελεστής του Kendall είναι ένας συντελεστής τάξης και αυτό μπορεί να το συμπεράνει κανείς από το γεγονός ότι $(X_i - X_j)(Y_i - Y_j) > 0$ αν και μόνο αν $(R_i - R_j)(S_i - S_j) > 0$.

Όπως ο συντελεστής Rho του Spearman έτσι και ο Tau του Kendall έχει κάποια σύνδεση με το copula C_n. Για να δούμε ποια είναι αυτή θα ξεκινήσουμε ορίζοντας τη συνάρτηση,

$$I_{ij} = \begin{cases} 1, \alpha\nu\ X_j < X_i, Y_j < Y_i \\ 0, \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases}$$

για $i \neq j$ και με $I_{ii} = 1$ για κάθε $i \in \{1,\ldots,n\}$. Παρατηρούμε ότι,

$$P_n = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}(I_{ij} + I_{ji}) = \sum_{i=1}^{n}\sum_{j\neq i} I_{ij} = -n + \sum_{i=1}^{n}\sum_{j=1}^{n} I_{ij}$$

όπου, $I_{ij} + I_{ji} = 1$ αν και μόνο αν τα ζευγάρια (X_i, Y_i) και (X_j, Y_j) βρίσκονται σε συμφωνία.

Επιπλέον,

$$W_i = \frac{1}{n}\sum_{j=1}^{n} I_{ij}$$

έτσι ώστε αν $\overline{W} = (W_1 + \ldots + W_n)/n$, τότε:

$$P_n = -n + n^2\overline{W}$$

και

$$\bar{\tau}_n = \frac{4}{n\,(n-1)}\,\overline{W} - \frac{n+3}{n-1}.$$

Η σύνδεση με το C_n προκύπτει από το γεγονός ότι εξ' ορισμού,

$$W_i = C_n\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)$$

συνεπώς,

$$\overline{W} = \int_0^1 \int_0^1 C_n(u,v)\,dC_n(u,v).$$

Χρησιμοποιώντας τη σχέση $\bar{\tau}_n = \frac{4}{n\,(n-1)}\,\overline{W} - \frac{n+3}{n-1}$ και υπό τη συνθήκη ότι $C_n \to C$ όταν $n \to \infty$, καταλήγουμε ότι ο $\bar{\tau}_n$ είναι ασυμπτωτικός εκτιμητής του συντελεστή στον πληθυσμό που ορίζεται ως:

$$\tau = 4\int_0^1 \int_0^1 C(u,v)c(u,v)dudv - 1 = 4E\big(C(U,V)\big) - 1$$

Με άλλα λόγια, έστω (X_1,Y_1), (X_2,Y_2) δύο ανεξάρτητα ζεύγη τυχαίων διανυσμάτων με από κοινού συνάρτηση κατανομής F. Τότε ο συντελεστής συσχέτισης τάξης του Kendall θα δίνεται από τον τύπο:

$$\begin{aligned}\tau(X,Y) &= \mathrm{P}[(X_1 - X_2)(Y_1 - Y_2)] > 0] - \mathrm{P}[(X_1 - X_2)(Y_1 - Y_2)] < 0] \\ &= \mathrm{P}[\text{συμφωνίας}] - \mathrm{P}[\text{ασυμφωνίας}]\end{aligned}$$

Συνεπώς, ο συντελεστής μπορεί να εκφραστεί μέσω της αναμενόμενης τιμής του $C(U,V)$, όπου C η από κοινού συνάρτηση κατανομής των τυχαίων μεταβλητών U,V οι οποίες είναι ομοιόμορφα κατανεμημένες στο (0,1).

4.3.3 Ιδιότητες των συντελεστών Tau και Rho

Στην παράγραφο αυτή θα παρουσιαστούν οι ιδιότητες των συντελεστών Tau και Rho και θα αναφερθούν κάποια πλεονεκτήματα και μειονεκτήματά τους. Για δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1 και F_2, από κοινού συνάρτηση κατανομής F και copula C ισχύουν τα εξής:

i. $\rho_S(X, Y) = \rho_S(Y, X)$ και $\rho_\tau(X, Y) = \rho_\tau(Y, X)$

ii. Αν X, Y ανεξάρτητες τότε $\rho_S(X, Y) = \rho_\tau(X, Y) = 0$

iii. $-1 \leq \rho_S(X, Y) \leq 1$ και $-1 \leq \rho_\tau(X, Y) \leq 1$

iv. Αν $T: \mathbb{R} \to \mathbb{R}$ είναι μία συνάρτηση γνησίως μονότονη στο πεδίο τιμών της X, οι ρ_S και ρ_τ ικανοποιούν την ιδιότητα (iv) που αναφέρθηκε στην Παράγραφο 4.1

v. $\rho_S(X, Y) = \rho_\tau(X, Y) = 1 \Leftrightarrow C = C_u \Leftrightarrow Y = T(X)$ με T αύξουσα

vi. $\rho_S(X, Y) = \rho_\tau(X, Y) = -1 \Leftrightarrow C = C_l \Leftrightarrow Y = T(X)$ με T φθίνουσα

Κάνοντας έναν απολογισμό των όσων αναφέραμε για τους δύο συντελεστές θα μπορούσαμε να πούμε τα εξής:

Είναι φανερό ότι ικανοποιούνται και οι τέσσερις ιδιότητες που αναφέρθηκαν στην παράγραφο 4.1. Όσο αφορά την ιδιότητα v, υπάρχουν παραδείγματα από την οικογένεια των σφαιρικών κατανομών που παρά το γεγονός ότι δίνουν συντελεστή συσχέτισης τάξης ίσο με το μηδέν, υπάρχει εξάρτηση μεταξύ των τυχαίων μεταβλητών. Τα βασικά πλεονεκτήματα των συντελεστών συσχέτισης τάξης σε σύγκριση με αυτόν της γραμμικής συσχέτισης είναι ότι παραμένουν αναλλοίωτα κάτω από μονότονους μετασχηματισμούς. Το βασικό τους μειονέκτημα σε σχέση με τον συντελεστή γραμμικής συσχέτισης είναι ότι δεν είναι μέτρα βασισμένα στις ροπές, αλλά πιο πολύπλοκα. Συνεπώς, όσο αφορά τους υπολογισμούς, υπάρχουν περιπτώσεις όπου υπολογίζονται πιο εύκολα από τον συντελεστή του Pearson και περιπτώσεις που υπολογίζονται δυσκολότερα. Για παράδειγμα, αν εργαζόμαστε με πολυδιάστατη κανονική κατανομή ή t-κατανομές, ο υπολογισμός του γραμμικού συντελεστή είναι ευκολότερος καθώς μπορούν σχετικά εύκολα να βρεθούν οι ροπές πρώτης και δεύτερης τάξης. Αν εργαζόμαστε με πολυδιάστατες κατανομές που έχουν κλειστούς τύπους copula, όπως για παράδειγμα η Gumbel τότε ο υπολογισμός των συντελεστών τάξης ίσως είναι πιο εύκολος χρησιμοποιώντας τους τύπους που τους συνδέουν με τα copulas.

4.4 Συντελεστής εξάρτησης ουρών

Με τον όρο εξάρτηση ουρών αναφερόμαστε στο θέμα της εξάρτησης στο πάνω δεξιά και το κάτω αριστερά τεταρτημόριο μιας δισδιάστατης κατανομής. Είναι μία έννοια η οποία σχετίζεται με τη μελέτη της εξάρτησης μεταξύ ακραίων τιμών.

Ας σκεφτούμε ότι μία τυχαία μεταβλητή παίρνει μία τιμή της, η οποία έχει πολύ μικρή πιθανότητα εμφάνισης δεδομένου ότι μία άλλη τυχαία μεταβλητή έχει πάρει μία τιμή της, η οποία είχε επίσης πολύ μικρή πιθανότητα εμφάνισης. Πρακτικά ενδιαφερόμαστε να εξετάσουμε αν η πραγματοποίηση ενός ακραίου γεγονότος (δηλαδή, ενός γεγονότος που πραγματοποιείται με μικρή πιθανότητα) που περιγράφεται από μία τυχαία μεταβλητή μπορεί να οδηγήσει στην πραγματοποίηση ενός άλλου ακραίου γεγονότος που περιγράφεται από μία άλλη τυχαία μεταβλητή. Στην περίπτωση που έχουμε συνεχείς τυχαίες μεταβλητές ο συντελεστής εξάρτησης ουρών είναι μία συνάρτηση copula και συνεπώς παραμένει αναλλοίωτος σε γνησίως αύξοντες μετασχηματισμούς αυτών.

Αν έχουμε δύο τυχαίες μεταβλητές X, Y με συναρτήσεις κατανομής F_1, F_2 ο συντελεστής εξάρτησης άνω ουράς (coefficient of upper tail dependence) θα δίνεται από τη σχέση:

$$\lambda_U = \lim_{a \to 1} \mathrm{P}[Y > {F_2}^{-1}(\alpha) \mid X > {F_1}^{-1}(\alpha)]$$

δηλαδή, ο συντελεστής εξάρτησης άνω ουράς είναι το όριο (αν υπάρχει) της δεσμευμένης πιθανότητας η Υ να είναι μεγαλύτερη από το $100 \cdot \alpha$ ποσοστημόριο της F_2 δεδομένου ότι η Χ είναι μεγαλύτερη από το $100 \cdot \alpha$ ποσοστημόριο της F_1, όταν το α πλησιάζει στο 1. Αντίστοιχα, ο συντελεστής εξάρτησης κάτω ουράς (coefficient of lower tail dependence) θα δίνεται από τη σχέση:

$$\lambda_L = \lim_{a \to 0} \mathrm{P}[Y \leq {F_2}^{-1}(\alpha) \mid X \leq {F_1}^{-1}(\alpha)]$$

δηλαδή, θα είναι το όριο (αν υπάρχει) της δεσμευμένης πιθανότητας η Y να είναι μικρότερη ή ίση από το $100 \cdot \alpha$ ποσοστημόριο της F_2 δεδομένου ότι η X είναι μικρότερη ή ίση από το $100 \cdot \alpha$ ποσοστημόριο της F_1, όταν το α πλησιάζει στο 0.

Οι συντελεστές εξάρτησης ουρών είναι μη παραμετρικοί καθώς δεν εξαρτώνται από τις περιθώριες κατανομές αλλά από το copula των τυχαίων μεταβλητών X, Y. Με τη χρήση των copulas θα μπορούσαμε να τους γράψουμε στη μορφή:

$$\lambda_U = \lim_{a \to 1^-} \frac{\bar{C}(\alpha,\alpha)}{1-a}$$

και

$$\lambda_L = \lim_{a \to 0^+} \frac{C(\alpha,\alpha)}{\alpha}$$

Όταν οι συντελεστές παίρνουν την τιμή μηδέν σημαίνει ότι έχουμε ανεξαρτησία στην άνω (λ_U) ή την κάτω (λ_L) ουρά. Συνεπώς, μπορεί να υπάρχει εξάρτηση μεταξύ των τυχαίων μεταβλητών αλλά όσο αφορά τη σχέση τους στις ουρές, το αν θα πάρει η μία κάποια ακραία τιμή δεν εξαρτάται από το αν θα πάρει ακραία τιμή η άλλη.

4.5 Τεταρτημοριακή Εξάρτηση (Quadrant Dependence)

Θα λέμε ότι δύο τυχαίες μεταβλητές X, Y έχουν θετική τεταρτημοριακή εξάρτηση (Positive Quadrant Dependence ή PQD) αν για κάθε $(x, y) \in \mathbb{R}^2$ ισχύει η σχέση:

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

Με λίγα λόγια, όταν δύο τυχαίες μεταβλητές έχουν αυτού του είδους την εξάρτηση, σημαίνει ότι είναι πιθανότερο να είναι από κοινού μικρότερες από κάποια τιμή από ότι θα ήταν στην περίπτωση της ανεξαρτησίας. Κατά κάποιο τρόπο εξετάζουμε αν είναι περισσότερο πιθανό δύο τυχαίες μεταβλητές να είναι εξαρτημένες από το να είναι ανεξάρτητες.

Αντίστοιχα, θα λέμε ότι έχουν αρνητική τεταρτημοριακή εξάρτηση (Negative Quadrant Dependence ή NQD) αν για κάθε $(x, y) \in \mathbb{R}^2$ ισχύει η σχέση:

$$P(X \leq x, Y \leq y) < P(X \leq x)P(Y \leq y).$$

Για δύο συνεχείς τυχαίες μεταβλητές X, Y με από κοινού συνάρτηση κατανομής H και περιθώριες F, G οι παραπάνω σχέσεις εκφράζονται και ως:

$H(x, y) \geq F(x)G(y)$, για κάθε $(x, y) \in \mathbb{R}^2$ στην περίπτωση που έχουμε PQD

και

$H(x, y) < F(x)G(y)$, για κάθε $(x, y) \in \mathbb{R}^2$ στην περίπτωση που έχουμε NQD

Με τη χρήση των συναρτήσεων copula οι αντίστοιχες ιδιότητες απαιτούν τις συνθήκες:

$C(u, v) \geq uv$, για κάθε $(u, v) \in [0,1]^2$ στην περίπτωση που έχουμε PQD

και

$C(u,v) < uv$, για κάθε $(u,v) \in [0,1]^2$ στην περίπτωση που έχουμε NQD.

Πρακτικά, η συνθήκη που απαιτείται στην περίπτωση του PQD, είναι ότι το copula των X, Y θα πρέπει να είναι μεγαλύτερο από το copula ανεξαρτησίας (αντίστροφα για το NQD).

5. Οικογένειες των Copulas

5.1 Οικογένεια Farlie - Gumbel - Mortgenstern

Τα Copulas που ανήκουν στην οικογένεια Farlie - Gumbel - Mortgenstern έχουν τη μορφή:

$$C(u,v) = uv + \theta uv(1-u)(1-v), \theta \in [0,1].$$

Η μορφή εξάρτησης των τυχαίων μεταβλητών που ακολουθούν αυτή την κατανομή εξαρτάται από την παράμετρο θ και συγκεκριμένα για θ = 0 οι τυχαίες μεταβλητές είναι ανεξάρτητες, ενώ για θ > 0 και θ < 0 είναι παρουσιάζουν θετική ή αρνητική εξάρτηση αντίστοιχα.

Τα copulas που ανήκουν σε αυτή την οικογένεια δεν παρουσιάζουν εξάρτηση στις ουρές καθώς οι συντελεστές εξάρτησης άνω και κάτω ουράς $\lambda_U = \lambda_L = 0$.

Ο συντελεστής του Kendall δίνεται από τον τύπο:

$$\tau = \frac{2\theta}{9}.$$

5.2 Αρχιμήδεια Copulas

Τα αρχιμήδεια copulas είναι από τις πιο σημαντικές οικογένειες των copulas. Έχουν αρκετές εφαρμογές στην στατιστική αλλά και στην αναλογιστική επιστήμη για τους εξής λόγους:

i. Είναι εύκολο να κατασκευαστούν
ii. Υπάρχει μεγάλη ποικιλία copulas που ανήκουν σε αυτή την κλάση
iii. Τα μέλη αυτής της οικογένειας έχουν πολλές καλές ιδιότητες

Τα Αρχιμήδεια copulas ικανοποιούν τη συνθήκη $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$ ή ισοδύναμα,

$$\varphi(H(x,y)) = \varphi(F(x)) + \varphi(G(y))$$

Δεδομένου ότι ενδιαφερόμαστε για τις εκφράσεις που μπορούν να χρησιμοποιηθούν για την κατασκευή των copulas, θα πρέπει να λύσουμε τη σχέση $\varphi(\mathrm{C}(u,v)) = \varphi(u) + \varphi(v)$. Άρα είναι απαραίτητο να βρεθεί μία κατάλληλα ορισμένη «αντίστροφη» $\varphi^{[-1]}$ τέτοια ώστε να ισχύει:

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

Έστω ότι η φ είναι μία συνεχής γνησίως φθίνουσα συνάρτηση από το [0,1] στο [0,∞] τέτοια ώστε $\varphi(1) = 0$. Ορίζουμε ως φευδοαντίστροφη συνάρτηση της φ με πεδίο ορισμού [0,∞] και πεδίο τιμών [0,1], τη συνάρτηση που δίνεται από τον τύπο:

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

Η $\varphi^{[-1]}$ είναι μία συνεχής συνάρτηση, φθίνουσα στο [0,∞] και γνησίως φθίνουσα στο $[0, \varphi(0)]$.
Επιπλέον, γαι την ψευδοαντίστροφη της φ ισχύουν τα εξής:

i. $\varphi^{[-1]}(\varphi(u)) = u$ στο [0,1]

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

iii. Αν $\varphi(0) = \infty$ τότε $\varphi^{[-1]}(t) = \varphi^{-1}(t)$

Έστω ότι η φ είναι μία συνεχής, γνησίως φθίνουσα συνάρτηση από το [0,1] στο [0,∞], τέτοια ώστε $\varphi(1) = 0$ και έστω ότι η $\varphi^{[-1]}$ είναι η φευδοαντίστροφη της φ. Τότε η συνάρτηση C από το $[0,1]^2$ στο [0,1] που ορίζεται από τον τύπο:

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

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

Η συνάρτηση φ ονομάζεται γεννήτορας του copula. Αν $\varphi(0) = \infty$ τότε λέμε ότι η φ είναι ένας αυστηρός γεννήτορας και το

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

είναι ένα αρχιμήδειο copula το οποίο χαρακτηρίζουμε ως αυστηρό.

Ας δούμε για παράδειγμα, τα copulas που παράγονται αν χρησιμοποιήσουμε για γεννήτορες τις συναρτήσεις $\varphi(t) = 1 - t$ και $\varphi(t) = -\log t$, με $t \in [0,1]$.

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

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

Συνεπώς, $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \max(1 - (1 - u) - (1 - v), 0) = \max(u + v - 1, 0) =$, $W(u, v)$. Προκύπτει δηλαδή το κάτω φράγμα Fréchet – Hoeffding.

- Για τον γεννήτορα $\varphi(t) = -\log t$ η ψευδοαντίστροφη είναι:

$$\varphi^{[-1]}(t) = \varphi^{-1}(t) = e^{-t} \text{ (επειδή } \varphi(0) = \infty)$$

Συνεπώς, $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) = e^{-(-\log u - \log v))} = uv = \Pi(u, v)$. Προκύπτει δηλαδή το copula ανεξαρτησίας.

Για τα αρχιμήδεια copulas, ο συντελεστής Tau του Kendall μπορεί να υπολογιστεί μέσω του γεννήτορα του copula από τον τύπο:

$$\tau = 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du + 1,$$

Όσον αναφορά εξάρτηση στις ουρές, ο συντελεστής εξάρτησης της άνω ουράς δίνεται από τον τύπο:

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

και ο συντελεστής εξάρτησης της κάτω ουράς δίνεται από τον τύπο:

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

Στην πράξη μπορούμε να προσαρμόσουμε στα δεδομένα τα μοντέλα που βασίζονται σε copula με τη μέθοδο της μέγιστης πιθανοφάνειας ή με την εκτίμηση της παραμέτρου που καθορίζει τον βαθμό εξάρτησης, με τη βοήθεια της σχέσης μεταξύ του συντελεστή Kendall και του γεννήτορα φ του αρχιμήδειου copula που χρησιμοποιείται για την κατασκευή του μοντέλου.

5.2.1 Οικογένεια Frank

Τα Copulas που ανήκουν στην οικογένεια Frank έχουν τη μορφή:

$$C_\theta(u, v) = -\frac{1}{\theta} log(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}), \theta \in \mathbb{R} \{0\},$$

και ο γεννήτορας δίνεται από τη σχέση:

$$\varphi(x) = -log(\frac{e^{-\theta x} - 1}{e^{-\theta} - 1})$$

Για τις ακραίες περιπτώσεις της παραμέτρου θ οδηγούμαστε στα παρακάτω αποτελέσματα:

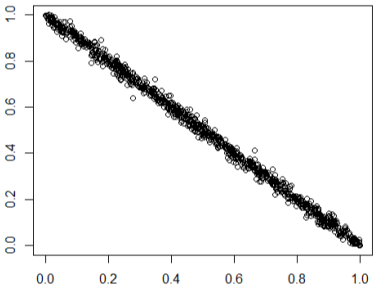
i. $\lim_{\theta \to -\infty} C_\theta = W$

ii. $\lim_{\theta \to +\infty} C_\theta = M$

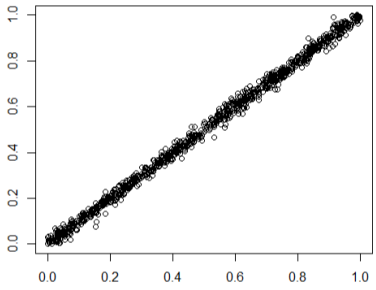
iii. $\lim_{\theta \to 0} C_\theta = \Pi$

Χρησιμοποιώντας τον παρακάτω κώδικά στην R μπορούμε να δούμε σχηματικά τις τρείς περιπτώσεις.

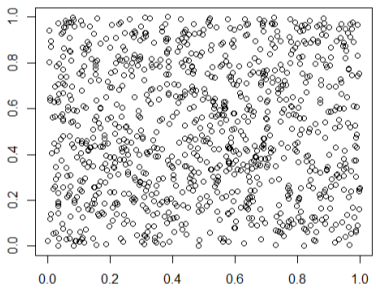
```
frank.cop <- frankCopula(log(10^-45), dim = 2)
plot(rCopula(1000, frank.cop))

frank.cop <- frankCopula(log(10^45), dim = 2)
plot(rCopula(1000, frank.cop))

frank.cop <- frankCopula(0, dim = 2)
plot(rCopula(1000, frank.cop))
```



Σχήμα 5.2.1.1: 1000 τυχαία σημεία από το Frank Copula με θ= $ln10^{-45}$



Σχήμα 5.2.1.2: 1000 τυχαία σημεία από το Frank Copula με θ= $ln10^{45}$



Σχήμα 5.2.1.3.: 1000 τυχαία σημεία από το Frank Copula με θ= 0

Κάθε ένα από τα παραπάνω σχήματα απεικονίζει 1000 τυχαία σημεία που προέρχονται από το Frank Copula με το θ να παίρνει τις τιμές $ln10^{-45}, ln10^{45}$ και 0 αντίστοιχα (από αριστερά προς τα δεξιά). Στις δύο πρώτες περιπτώσεις που η παράμετρος θ παίρνει πολύ μικρή (στην πρώτη) ή πολύ μεγάλη τιμή (στη δεύτερη) φαίνεται να υπάρχει ισχυρή εξάρτηση, ενώ στην περίπτωση που το θ πλησιάζει οριακά το μηδέν φαίνεται να υπάρχει ανεξαρτησία.

Τα copulas που ανήκουν στην οικογένεια Frank έχουν συντελεστές εξάρτησης ουρών ίσες με το μηδέν που σημαίνει ότι ενώ μπορεί να υπάρχει εξάρτηση μεταξύ δύο τυχαίων μεταβλητών, αυτή δε θα αφορά τις ακραίες τιμές.

Ο συντελεστής του Kendall δίνεται από τη σχέση:

$$\tau_{C_\theta} = 1 - 4\theta^{-1}(1 - D_1(\theta)),$$

όπου η $D_1(\theta)$ ονομάζεται συνάρτηση του Debye και ορίζεται ως:

$$D_1(\theta) = \theta^{-1} \int_0^\infty \frac{t}{e^t - 1} dt.$$

5.2.2 Οικογένεια Gumbel

Τα Copulas που ανήκουν στην οικογένεια Gumbel έχουν τη μορφή:

$$C_\theta(u, v) = e^{-[(-logu)^\theta + (-logv)^\theta]^{\frac{1}{\theta}}}, \theta \in [1, \infty)$$

και ο γεννήτορας δίνεται από τη σχέση:

$$\varphi(x) = (-log(x))^\theta$$

Για τις ακραίες περιπτώσεις της παραμέτρου θ οδηγούμαστε στα παρακάτω αποτελέσματα:

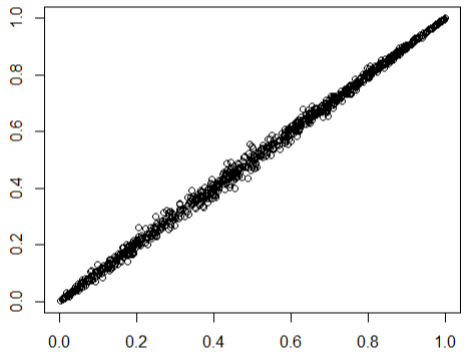
i. $\lim_{\theta \to +\infty} C_\theta = M$

ii. $\lim_{\theta \to 1} C_\theta = \Pi$

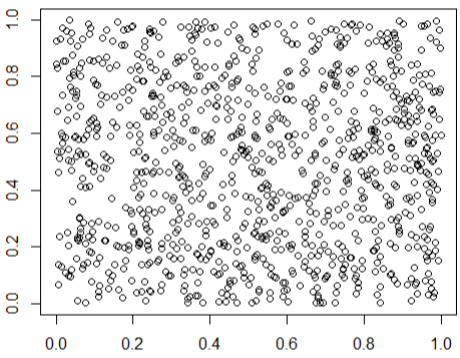
Χρησιμοποιώντας τον παρακάτω κώδικά στην R μπορούμε να δούμε σχηματικά τις δύο περιπτώσεις.

```
gumbel.cop <- gumbelCopula(log(10^15), dim = 2)
plot(rCopula(1000, gumbel.cop))
```

```
gumbel.cop <- gumbelCopula(1, dim = 2)
plot(rCopula(1000, gumbel.cop))
```



Σχήμα 5.2.1.4: 1000 τυχαία σημεία από το Gumbel Copula με θ= 10^{15}



Σχήμα 5.2.1.5: 1000 τυχαία σημεία από το Gumbel Copula με θ= 1

Τα δύο σχήματα απεικονίζουν 1000 τυχαία σημεία που προέρχονται από το Gumbel Copula με το θ να παίρνει τις τιμές 10^{15} και 1 αντίστοιχα (από αριστερά προς τα δεξιά). Στην πρώτη περίπτωση που η παράμετρος θ παίρνει πολύ μεγάλη τιμή φαίνεται να υπάρχει ισχυρή εξάρτηση, ενώ στην περίπτωση που το θ πλησιάζει οριακά το 1 φαίνεται να υπάρχει ανεξαρτησία, κάτι το οποίο περιμέναμε με βάση τις οριακές περιπτώσεις της παραμέτρου θ.

Τα copulas της οικογένειας Gumbel χρησιμοποιούνται συνήθως για να μοντελοποιήσουμε περιπτώσεις όπου υπάρχει υψηλός βαθμός εξάρτησης στη δεξιά ουρά. Ο συντελεστής εξάρτησης δεξιάς ουράς δίνεται από τη σχέση:

$$\lambda_U = 2 - 2^{\frac{1}{\theta}}.$$

Αντίθετα, εμφανίζουν μηδενική εξάρτηση στην κάτω ουρά και ισχύει ότι $\lambda_L = 0$. Επίσης, η σχέση μεταξύ της παραμέτρου θ του copula και του συντελεστή του Kendall δίνεται από τον τύπο:

$$\tau_{C_\theta}^{G} = \frac{\theta-1}{\theta}$$

5.2.3 Οικογένεια Clayton

Τα Copulas που ανήκουν στην οικογένεια Clayton έχουν τη μορφή:

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

και ο γεννήτορας δίνεται από τη σχέση:

$$\varphi(x) = \frac{1}{\theta}(x^{-\theta} - 1)$$

Για τις ακραίες περιπτώσεις της παραμέτρου θ οδηγούμαστε στα παρακάτω αποτελέσματα:

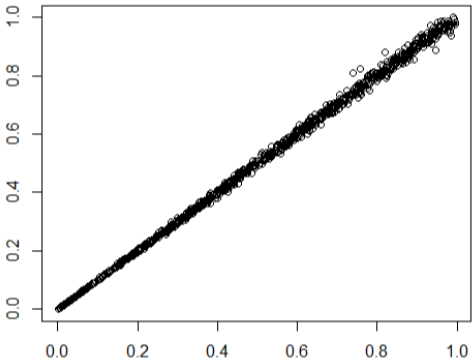
i. $\lim_{\theta \to -1} C_\theta = W$

ii. $\lim_{\theta \to +\infty} C_\theta = M$

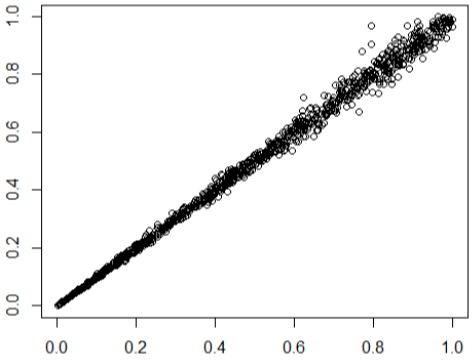
iii. $\lim_{\theta \to 0} C_\theta = \Pi$

Χρησιμοποιώντας τον παρακάτω κώδικά στην R μπορούμε να δούμε σχηματικά πώς φαίνεται να υπάρχει καλύτερη εξάρτηση όταν αυξάνεται η τιμή της παραμέτρου θ.

```
clayton.cop <- claytonCopula(100, dim = 2)
plot(rCopula(1000, clayton.cop))

clayton.cop <- claytonCopula(50, dim = 2)
plot(rCopula(1000, clayton.cop))
```



Σχήμα 5.2.3.1: 1000 τυχαία σημεία από το Clayton Copula με θ= 100



Σχήμα 5.2.3.2: 1000 τυχαία σημεία από το Clayton Copula με θ= 50

Τα Clayton copulas είναι συνήθως κατάλληλα για την περιγραφή της δομής εξάρτησης τυχαίων μεταβλητών που τείνουν να παίρνουν ταυτόχρονα μικρές τιμές καθώς η εξάρτηση είναι ισχυρή στην κάτω ουρά της κατανομής. Ο συντελεστής εξάρτησης κάτω ουράς για $\theta \geq 0$, δίνεται από τον τύπο:

$$\lambda_L = 2^{-\frac{1}{\theta}}$$

ενώ ο συντελεστής εξάρτησης της άνω ουράς είναι ίσος με μηδέν. Η σχέση μεταξύ της παραμέτρου θ του copula και του συντελεστή του Kendall δίνεται από τον τύπο:

$$\tau_{C_\varphi} = \frac{\theta}{\theta + 2}.$$

6. Μέθοδοι Εκτίμησης και Ελέγχου

Σε αυτή την ενότητα θα παρουσιαστούν τρεις μέθοδοι εκτίμησης. Αρχικά, θα δούμε μία παραμετρική μέθοδος εκτίμησης, στη συνέχεια μία ημιπαραμετρική μέθοδο εκτίμησης και τέλος μία μη παραμετρική μέθοδο εκτίμηση των Copulas, η οποία στηρίζεται μόνο στα δεδομένα.

6.1 Παραμετρική Μέθοδος Εκτίμησης ενός Copula

Στην συγκεκριμένη ενότητα θα ασχοληθούμε με την περίπτωση που ένα Copula είναι απόλυτα συνεχές. Όπως έχει αναφερθεί, ορίζεται η πυκνότητα c του Copula. Άρα, έστω $X_1, \dots, X_d$ συνεχείς τυχαίες μεταβλητές που έχουν περιθώριες συναρτήσεις κατανομής $F_1, \dots, F_d$ και περιθώριες συναρτήσεις πυκνότητας $f_1, \dots, f_d$ αντίστοιχα. Τότε η από κοινού πυκνότητα είναι της μορφής:

$$f(x_1, \dots, x_d) = c(F_1, \dots, F_d) \prod_{i=1}^{d} f_i(x_i) ,$$

όπου

$$c(F_1, \dots, F_d) = \frac{\partial^n C(F_1, \dots, F_d)}{\partial F_1 \dots \partial F_d}.$$

Για την εκτίμηση της παραμέτρου του Copula θα μπορούσαμε να μεγιστοποιήσουμε τη συνάρτηση της πιθανοφάνειας. Σε αυτή την περίπτωση όμως πρέπει να λάβουμε υπόψιν μας ότι έχουμε και περιθώριες συναρτήσεις άρα θα έχουμε d + 1 παραμέτρους προς εκτίμηση. Έστω $(a_1, \dots a_d, \theta)$, όπου $a_1, \dots a_d$ είναι οι παράμετροι των d περιθωρίων και θ είναι η παράμετρος του Copula. Ακόμα, έστω ότι έχουμε ένα δείγμα με n τυχαία διανύσματα $\boldsymbol{x_1}, \dots, \boldsymbol{x_n}$. Συνήθως χρησιμοποιούμε τον φυσικό λογάριθμο της συνάρτησης της πιθανοφάνειας γιατί δεν επηρεάζει στη μεγιστοποίηση της συνάρτησης και είναι πιο εύκολες οι πράξεις. Επομένως, έχουμε ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι:

$$L = \sum_{j=1}^{n} \log(f(x_1{}^j, \dots, x_d{}^j).$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας μπορεί να γραφτεί και στην μορφή:

$$L = L_C + \sum_{i=1}^{d} L_i,$$

όπου

$$L_C = \sum_{j=1}^{n} c \log f(x_1{}^j, \dots, x_d{}^j)$$

είναι η συνεισφορά στην πιθανοφάνεια της δομής εξάρτησης των δεδομένων μέσω του Copula C και

$$L_i = \sum_{j=1}^{n} \log f_i(x_i{}^j)$$

είναι οι συνεισφορές στην πιθανοφάνεια από την κάθε περιθώρια.

Με τη μέθοδο της μέγιστης πιθανοφάνειας θα εκτιμούσαμε ταυτόχρονα όλες τις d + 1 παραμέτρους, δηλαδή θα βρίσκαμε ποιος είναι ο εκτιμητής $(\hat{a}_1, \dots, \hat{a}_d, \hat{\theta})$, ο οποίος θα προέκυπτε από τις λύσεις των εξισώσεων:

$$\left(\frac{\partial L}{\partial a_1}, \dots, \frac{\partial L}{\partial a_d}, \frac{\partial L}{\partial \theta}\right) = \mathbf{0}$$

Όμως, δεν χρησιμοποιείται αυτή η μέθοδος αλλά μία άλλη με τα εξής βήματα:

i. Εκτιμάμε τις παραμέτρους κάθε μιας περιθώριας ξεχωριστά με τη μέθοδο της μέγιστης πιθανοφάνειας και
ii. Εκτιμάμε την παράμετρο του Copula με τη μέθοδο της μέγιστης πιθανοφάνειας, αφού πρώτα εισάγουμε τις εκτιμήσεις των παραμέτρων που υπολογίσαμε στο προηγούμενο βήμα

Η μέθοδος που περιγράφηκε παραπάνω ονομάζεται Method of Inference Functions for Margin (IFM). Ο λόγος που χρησιμοποιείται αυτή η μέθοδος και όχι η μέθοδος της πιθανοφάνειας (ML) είναι διότι έχει πιο εύκολους υπολογισμούς καθώς γενικά είναι πιο δύσκολο να γίνεται μια βελτιστοποίηση πολλών παραμέτρων ταυτοχρόνως παρά πολλές βελτιστοποιήσεις με λιγότερες παραμέτρους κάθε φορά.

Άρα με την μέθοδο IFM ο εκτιμητής είναι της μορφής:

$$(\tilde{a}_1, \dots, \tilde{a}_d, \tilde{\theta}),$$

ο οποίος θα προέκυπτε από τις λύσεις των εξισώσεων:

$$\left(\frac{\partial L_1}{\partial a_1}, \dots, \frac{\partial L_d}{\partial a_d}, \frac{\partial L}{\partial \theta}\right) = \mathbf{0}$$

Ο εκτιμητής αυτός θεωρείται αρκετά ικανοποιητικός σε σχέση με αυτό που θα προέκυπτε από τη μέθοδο της μέγιστης πιθανοφάνειας. Ο IFM εκτιμητής ουσιαστικά χρησιμοποιεί την ML μέθοδο και επομένως θα έχει κάποιες κοινές ιδιότητες.

6.2 Ημιπαραμετρική Μέθοδος Εκτίμηση ενός Copula

Στην προηγούμενη υποενότητα είδαμε μία παραμετρική μέθοδο εκτίμησης ενός Copula, την IFM, θεωρώντας ότι γνωρίζαμε τις περιθώριες συναρτήσεις κατανομών. Σε πολλές περιπτώσεις είναι δύσκολο να ξέρουμε ποιες είναι οι περιθώριες συναρτήσεις κατανομών διότι τα δεδομένα μπορεί να έχουν για παράδειγμα βαριές ουρές και ασυμμετρίες και επομένως να είναι δύσκολο να βρούμε ποιες είναι οι

κατανομές αυτές. Σε αυτήν την περίπτωση μπορούμε να χρησιμοποιήσουμε την παρακάτω μέθοδο που αποτελείται από τα εξής δύο βήματα:

i. Θεωρούμε ότι οι περιθώριες δεν είναι γνωστές, οπότε θα πρέπει να επιλέξουμε μία εκτίμηση τους. Συνήθως, επιλέγουμε τις εμπειρικές συναρτήσεις κατανομών, οι οποίες δίνονται από τον τύπο:

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

όπου $1(X_i \leq x)$ είναι η δείκτρια συνάρτηση.

ii. Εκτιμούμε με τη μέθοδο της μέγιστης πιθανοφάνειας την παράμετρο του Copula

Αυτή η μέθοδος ονομάζεται Canonical Maximum Likelihood (CML). Τις περισσότερες φορές έχουμε αρκετά δεδομένα για να χρησιμοποιήσουμε τις εμπειρικές κατανομές, δηλαδή να πάρουμε μια μη παραμετρική εκτίμηση των περιθωρίων, αλλά δεν έχουμε αρκετή πληροφορία για την μη παραμετρική εκτίμηση της δομής εξάρτησης. Άρα, έστω $(X_{1k}, \dots, X_{pk})$, όπου k = 1, 2, …, n ένα δείγμα μεγέθους n. Ο τύπος της ψευδοσυνάρτησης της πιθανοφάνειας είναι:

$$L(\theta) = \sum_{k=1}^{n} \log(c_\theta(F_{1k}(x_{1k}), \dots, F_{pk}(x_{pk}))),$$

όπου F_{in} είναι $\frac{n}{n+1}\times$ την περιθώρια εμπειρική συνάρτηση κατανομής της i τυχαίας μεταβλητής.

Ο εκτιμητής της παραμέτρου θ του Copula είναι:

$$\hat{\theta}_n = argmax_\theta L(\theta) = argmax_\theta \sum_{k=1}^{n} \log(c_\theta(F_{1k}(x_{1k}), \dots, F_{pk}(x_{pk})).$$

Ο $\hat{\theta}_n$ αποδεικνύεται πως κάτω από συνηθισμένες συνθήκες ομαλότητος είναι συνεπής και ασυμπτωτικά κανονικός εκτιμητής.

Αν θέλαμε να εφαρμόσουμε την παραπάνω μέθοδο σε δοσμένα δεδομένα για δισδιάστατες μεταβλητές, τότε θα γράφαμε τον λογάριθμο της ψευδοσυνάρτησης πιθανοφάνειας συναρτήσει των τάξεων των δύο τυχαίων μεταβλητών ως εξής:

$$L(\theta) = \sum_{i=1}^{n} \log\{c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)\}\,,$$

όπου τα R_i, S_i είναι οι τάξεις των X_i, Y_i. Τότε ο CML εκτιμητής θα προέκυπτε ως ρίζα της εξίσωσης:

$$l(\theta) = \frac{\partial L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta}\, c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)}{c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)} = 0.$$

Μια άλλη ημιπαραμετρική μέθοδος εκτίμησης ενός Copula μπορεί να θεωρηθεί η μέθοδος στην οποία χρησιμοποιούμε τους εκτιμητές των συντελεστών συσχέτισης για να εκτιμήσουμε την παράμετρο εξάρτησης του Copula. Αυτή η μέθοδος δεν μπορεί να θεωρηθεί παραμετρική διότι δεν θα χρησιμοποιήσουμε τις περιθώριες συναρτήσεις παρά μόνο τις τάξεις των δεδομένων, αλλά ούτε και μη παραμετρική διότι για να συνδέσουμε τα τ, ρ με το θ πρέπει να έχουμε διαλέξει μια συγκεκριμένη οικογένεια των Copulas.

Άρα, αν υπάρχει μια σχέση της μορφής θ = g(τ) που να συνδέει την παράμετρο εξάρτησης του Copula με το τ του Kendall, τότε ένας λογικός εκτιμητής θα ήταν:

$$\tilde{\theta}_n = g(\tau_n).$$

Επίσης, αν υπάρχει μια σχέση της μορφής θ = h(ρ) που να συνδέει την παράμετρο εξάρτησης του Copula με το ρ του Spearman, τότε ένας λογικός εκτιμητής θα ήταν:

$$\tilde{\theta}_n = h(\rho_n).$$

i. Για το τ του Kendall και $n \to \infty$ έχουμε ότι:

$$\sqrt{n}(\tau_n - \tau) \sim N(0, \sigma^2),$$

όπου σ^2 είναι η διασπορά της ασυμπτωτικής κανονικής κατανομής του εκτιμητή. Έστω ότι η g είναι μια παραγωγίσιμη συνάρτηση με $g'(\tau) \neq 0$. Τότε σύμφωνα με τη μέθοδο Δέλτα έχουμε:

$$\sqrt{n}\big(g(\tau_n) - g(\tau)\big) \sim N(0, \sigma^2 (g'(\tau))^2).$$

Ένας συνεπής εκτιμητής για το σ^2 είναι ο $16S^2$ όπου:

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (W_i + \widetilde{W_i} - 2\widetilde{W})^2,$$

$$W_i = \frac{1}{n} \sum_{j=1}^{n} I_{ij} = \frac{1}{n} \#\{j : X_j \leq X_i, Y_j \leq Y_i\},$$

$$\widetilde{W}_i = \frac{1}{n}\sum_{j=1}^{n} I_{ij} = \frac{1}{n}\,\#\{j: X_i \le X_j, Y_i \le Y_j\}$$

και

$$\widetilde{W} = \frac{1}{n}\sum_{i=1}^{n} W_i$$

Από θεώρημα για θ = g(τ), $\tilde{\theta}_n = g'(\tau_n)$ και $16S^2 \xrightarrow{p} \sigma^2$ έχουμε:

$$\sqrt{n}\frac{(\tilde{\theta}_n - \theta)}{4S} = \frac{\frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\sigma g'(\tau)} \xrightarrow{d} N(0,1)}{\frac{4Sg'(\tau)}{\sigma g'(\tau)} \xrightarrow{p} 1} \xrightarrow{d} N(0,1)$$

Άρα, για $n \to \infty$ ισχύει ότι $\tilde{\theta}_n \sim N(\theta, \frac{1}{n}(4Sg'(\tau))^2)$.

Ένα ασυμπτωτικό διάστημα εμπιστοσύνης $100(1-\alpha)\%$ για το θ είναι: $\tilde{\theta}_n \pm z_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}4S|g'(\tau_n)|$.

ii. Για το ρ του Spearman και $n \to \infty$ έχουμε ότι:

$$\sqrt{n}(\rho_n - \rho) \sim N(0, \sigma^2),$$

όπου σ^2 είναι η διασπορά της ασυμπτωτικής κανονικής κατανομής του εκτιμητή. . Έστω ότι η h είναι μια παραγωγίσιμη συνάρτηση με $h'(\tau) \neq 0$. Τότε σύμφωνα με τη μέθοδο Δέλτα έχουμε:

$$\sqrt{n}\big(h(\rho_n) - h(\rho)\big) \sim N(0, \sigma^2(h'(\rho))^2).$$

Ένας συνεπής εκτιμητής για το σ^2 είναι ο ${\sigma_n}^2 = 144(-9{A_n}^2 + B_n + 2C_n + 2D_n + 2E_n)$, όπου

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{n+1}\,\frac{S_i}{n+1},$$

$$B_n = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{n+1}^{2}\,\frac{S_i}{n+1}^{2},$$

$$C_n = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{R_i}{n+1}\,\frac{S_i}{n+1}\,1\big(R_k \le R_i, S_k \le S_j\big) + \frac{1}{4} - A_n,$$

$$D_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{S_i}{n+1}\,\frac{S_j}{n+1}\max\left(\frac{R_i}{n+1}\,\frac{R_j}{n+1}\right)$$

και

$$E_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{n+1}\,\frac{R_j}{n+1}\max\left(\frac{S_i}{n+1}\,\frac{S_j}{n+1}\right).$$

Από θεώρημα για θ = h(τ), $\tilde{\theta}_n = h'(\rho_n)$ και ${\sigma_n}^2 \xrightarrow{p} \sigma^2$ έχουμε:

$$\sqrt{n}\frac{(\tilde{\theta}_n - \theta)}{\sigma_n} = \frac{\frac{\sqrt{n}\left(\tilde{\theta}_n - \theta\right)}{\sigma h'(\rho)} \xrightarrow{d} N(0,1)}{\frac{\sigma_n h^{'}(\rho)}{\sigma h^{'}(\rho)} \xrightarrow{p} 1} \xrightarrow{d} N(0,1).$$

Άρα, για $n \to \infty$ ισχύει ότι $\tilde{\theta}_n \sim N(\theta, \frac{1}{n}(\sigma_n h'(\rho))^2)$.

Ένα ασυμπτωτικό διάστημα εμπιστοσύνης $100(1-\alpha)\%$ για το θ είναι: $\tilde{\theta}_n \pm z_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sigma_n|h'(\rho_n)|$.

6.3 Μη Παραμετρική Μέθοδος Εκτίμησης ενός Copula

Στην παραπάνω υποενότητα ενώ ξεκινήσαμε να δώσουμε μια μη παραμετρική εκτίμηση των περιθωρίων συναρτήσεων κατανομής, δώσαμε μία ημιπαραμετρική μέθοδο εκτίμησης των Copulas γιατί η εκτίμηση της παραμέτρου του Copula προϋποθέτει τη γνώση του Copula. Σε αυτή την υποενότητα θα υποθέσουμε ότι δεν γνωρίζουμε ούτε το Copula και άρα θα το εκτιμήσουμε σύμφωνα με τα δεδομένα κάθε φορά. Επομένως, η μη παραμετρική μέθοδος των Copulas με την οποία θα ασχοληθούμε είναι το Εμπειρικό Copula.

Το Εμπειρικό Copula είναι οποιοδήποτε Copula με πεδίο ορισμού το $L = \{\left(\frac{t_1}{T}, \dots, \frac{t_N}{T}\right) : 1 \le n \le N, t_n = 0, \dots., T\}$ για το οποίο ισχύει:

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

όπου r_n είναι η νιοστή διατεταγμένη τάξη ενός δείγματος $(x_1, \dots, x_n)$ μεγέθους Τ.

Η πυκνότητα του Εμπειρικού Copula δίνεται από την σχέση:

$$\hat{C}\left(\frac{t_1}{T}, \dots, \frac{t_N}{T}\right) = \sum_{i_1=1}^{2} \dots \sum_{i_n=1}^{2} (-1)^{\sum_{j=1}^{n} i_j} \hat{C}(\frac{t_1 - i_1 + 1}{T}, \dots, \frac{t_n - i_n + 1}{T},$$

Στην δισδιάστατη περίπτωση έστω X, Y τυχαίες μεταβλητές και $\{(x_k, y_n)\}_{k=1}^{n}$ τυχαίο δείγμα n παρατηρήσεων. Τότε το Εμπειρικό Copula δίνεται από τον τύπο:

$$\hat{C}(u, v) = \frac{1}{n} \sum_{i=1}^{n} 1(U_i \leq u, V_i \leq v),$$

όπου (u,v) είναι η τιμή του διανύσματος $(\mathrm{U}, \mathrm{V}) = (\hat{F}(X), \hat{G}(Y))$ και τα U_i, V_i είναι οι τάξεις των παρατηρήσεων του i διανύσματος προς το πλήθος των παρατηρήσεων συν 1.

6.4 Έλεγχοι Καλής Προσαρμογής

Σε αυτή την ενότητα θα αναφερθούμε σε κάποιους ελέγχους καλής προσαρμογής για τα Copula. Οι έλεγχοι που θα κάνουμε είναι της μορφής:

$$H_0: C \in C_\Theta \text{ έναντι της } H_1: C \notin C_\Theta,$$

όπου $C_\Theta = \{C_\theta : \theta \in \Theta\}$ είναι κάποια από τις γνωστές κλάσεις των Copulas.

Για παράδειγμα, στον παραπάνω έλεγχο, δεν έχει νόημα να ασχοληθούμε με τις παραμέτρους των περιθωρίων διότι θα καταλήγαμε σε έναν πιο αυστηρό έλεγχο ο οποίους θα περιλάμβανε και τον έλεγχο προσαρμογής των περιθωρίων. Γι' αυτό οι έλεγχοι που θα παρουσιαστούν παρακάτω θα βασίζονται στις τάξεις των παρατηρήσεων προς τα πλήθος των παρατηρήσεων συν 1.

Στην δισδιάστατη περίπτωση, έχουμε τις συνεχείς τυχαίες μεταβλητές X, Y, δείγμα μεγέθους n και πραγματοποιούμε τους μετασχηματισμούς $U_i = \frac{R_i}{n+1}$ και $V_i = \frac{S_i}{n+1}$, όπου R_i, S_i είναι οι τάξεις της i παρατήρησης των X καιY αντίστοιχα. Άρα, μπορούμε να συμπεράνουμε ότι το Εμπειρικό Copula είναι το Copula C του παραπάνω ελέγχου. Αξίζει να οριστεί μια απόσταση μεταξύ του Εμπειρικού Copula και του παραμετρικού Copula του οποίου έχουμε εκτιμήσει την παράμετρο από μία ημιπαραμετρική μέθοδο. Αυτή η απόσταση θα είναι της μορφής:

$$C_n = \sqrt{n}(C_n - C_{\theta n}),$$

όπου $\hat{\theta}_n$ είναι ο εκτιμητής της παραμέτρου του Copula.

Δύο γνωστές στατιστικές συναρτήσεις είναι οι:

$$S_n = \int_0^1 \int_0^1 C(u,v)^2 dC_n(u,v)$$

και

$$T_n = sup_{(u,v)\in[0,1]^2} |C_n(u,v).$$

Όταν οι παραπάνω στατιστικές συναρτήσεις έχουμε μεγάλες τιμές τότε οδηγούν στην απόρριψη της μηδενικής υπόθεσης.

Για να αποφευχθεί η ύπαρξη ακραίων τιμών, τα ζευγάρια στα οποία συνιστάται να χρησιμοποιείται είναι αυτά στα οποία ισχύει:

$$|\lambda_i| \leq 4(\frac{1}{n-1} - \frac{1}{2})^2$$

Ένα άλλο εργαλείο για την γραφική απεικόνιση βασισμένο στις τάξεις των παρατηρήσεων είναι το k - plot. Πιο συγκεκριμένα, η τεχνική βασίζεται στην απεικόνιση των ζευγών $(W_{i:n}, H_{(i)})$ για $i \in \{1, \dots, n\}$, όπου

$$H_{(1)} < \cdots < H_{(n)}$$

Είναι οι διατεταγμένες στατιστικές συναρτήσεις που προέρχονται από τις $H_1, \dots, H_n$ που χρησιμοποιήθηκαν στα διαγράμματα Chi – plot. Όσον αναφορά το $W_{i:n}$ είναι η αναμενόμενη τιμή i της στατιστικής συνάρτησης από ένα τυχαίο δείγμα μεγέθους n από μία τυχαία μεταβλητή W = C(U, V) = H(X, Y) υπό τη μηδενική υπόθεση της ανεξαρτησίας μεταξύ των U και V (ή μεταξύ των X και Y, το οποίο είναι το ίδιο). Αυτό δίνεται από τον τύπο:

$$W_{i:n} = n \binom{n-1}{i-1} \int_0^1 w k_0(w) \{K_0(w)\}^{i-1} \{1 - K_0(w)\}^{n-1} dw,$$

Όπου

$$K_0(w) = P(UV \leq w) = \int_0^1 P(U \leq \frac{w}{v})\, dv = \int_0^w 1 dv + \int_w^1 \frac{w}{v}\, dv = w - wlog(w)$$

και k_0 είναι η αντίστοιχη πυκνότητα.

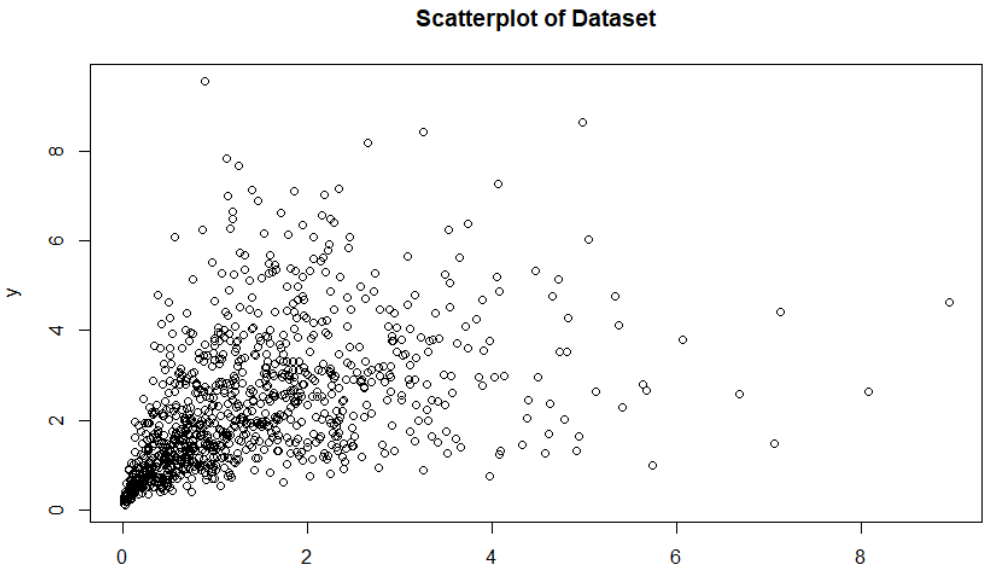
7. Κατασκευή Μοντέλου

Στο παρόν κεφάλαιο θα χρησιμοποιήσουμε τη θεωρία των copulas και κάποιες στατιστικές μεθοδολογίες οι οποίες αναφέρθηκαν σε προηγούμενα κεφάλαια με στόχο την κατασκευή ενός μοντέλου που θα περιγράφει την από κοινού συμπεριφορά δύο εξαρτημένων τυχαίων μεταβλητών. Καθώς δεν υπήρχε η δυνατότητα εύρεσης πραγματικών δεδομένων παρήχθησαν 1000 παρατηρήσεις από το Clayton copula με παράμετρο 1,65 με περιθώριες τις κατανομές Gamma (1,35 , 1,04) και Gamma (2,25 , 1,02) που παρουσιάζονται στη συνέχεια έτσι ώστε να παρουσιαστεί η διαδικασία κατασκευής ενός μοντέλου. Η επιλογή των παραμέτρων έγινε δοκιμαστικά ώστε να υπάρχει γραφικά ένας διαφαινόμενος βαθμός συσχέτισης. Συνεπώς για την κατασκευή του μοντέλου θα χρησιμοποιήσουμε ένα πλήθος χιλίων δισδιάστατων παρατηρήσεων (x_i, y_i), οι οποίες προέρχονται από τις κατανομές Gamma, οι τιμές των οποίων έχουν παραχθεί με τη βοήθεια της γλώσσας R, στην οποία θα δουλέψουμε το μοντέλο.

Στον παρακάτω πίνακα δίνουμε μερικά βασικά περιγραφικά στοιχεία για τα δεδομένα μας.

	X	Y
Min.	0,01396	0,11400
1st Quantile	0,52248	1,12900
Median	1,09064	1,92800
Mean	1,39897	2,28400
3rd Quantile	1,94408	3,04800
Max.	8,94470	9,54700

Ενδιαφέρον παρουσιάζει το διάγγραμμα διασποράς (scatterplot) των δύο τυχαίων μεταβλητών το οποίο παραθέτουμε στη συνέχεια.



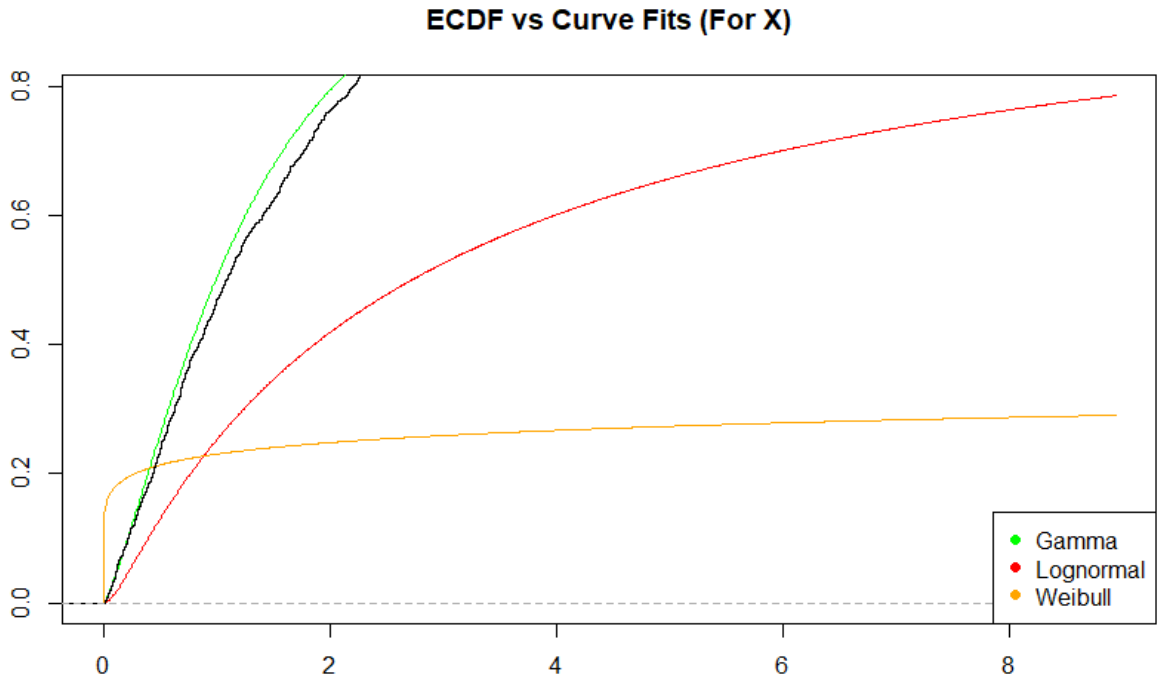
Σχήμα 7.1: Διάγραμμα διασποράς των τ.μ. X,Y

Όπως παρατηρούμε τα δεδομένα μας φαίνεται να παρουσιάζουν θετική συσχέτιση, η οποία είναι πιο ισχυρή στην κάτω ουρά. Ο συντελεστής tau του Kendall παίρνει την τιμή 0,45 κάτι που σημαίνει ότι η συσχέτιση δεν είναι ιδιαίτερα ισχυρή. Μία ισχυρότερη συσχέτιση θα έδινε μεγαλύτερη στατιστική σημασία στα αποτελέσματα παρόλα αυτά η μεθοδολογία που θα ακολουθήσουμε θα ήταν η ίδια και σε δεδομένα με μεγαλύτερο βαθμό συσχέτισης.

Αρχικά, θα προσπαθήσουμε να κάνουμε εκτιμήσεις για τις περιθώριες κατανομές. Θα προσπαθήσουμε δηλαδή να βρούμε από ποιά κατανομή προέρχονται τα δεδομένα που αφορούν τόσο την τυχαία μεταβλητή X όσο και την Y. Θα δοκιμάσουμε τρεις διαφορετικές κατανομές (LogNormal, Gamma, Weibull) και θα δούμε ποιά από αυτές προσαρμόζεται καλύτερα στα δεδομένα μας και για ποιές παραμέτρους. Θα κάνουμε αρχικά μία εκτίμηση των παραμέτρων των κατανομών με τη μέθοδο μέγιστης πιθανοφάνειας. Στη συνέχεια θα απεικονίσουμε γραφικά τις κατανομές αυτές σε ένα κοινό διάγραμμα και σε σύγκριση με την εμπειρική συνάρτηση κατανομής που θα προέλθει από τα δεδομένα μας, όπως βλέπουμε στα παρακάτω σχήματα που προκύπτουν μέσω του ακόλουθου κώδικα.

```
# Estimate distribution parameters for X

# Log normal
# function that calculates negative log-likelihood
x_nLL_LN = function(mu, sigma) -sum(stats::dlnorm(data$x, mu, sigma, log = TRUE))
# estimation of the parameters
x_fit_LN = mle(x_nLL_LN, start = list(mu = 10, sigma = 1), method = "L-BFGS-B", nobs = length(data$x), lower = c(1,0.01))
```

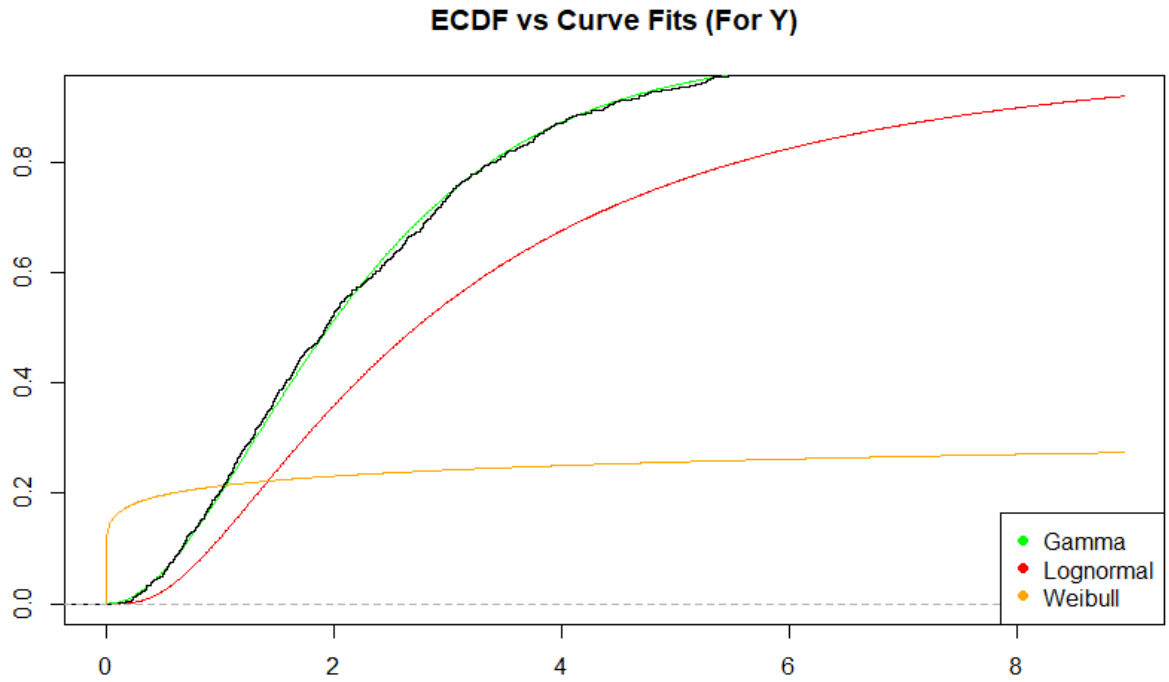
```
# Gamma
# function that calculates negative log-likelihood
x_nLL_G = function(alpha, theta) -sum(stats::dgamma(data$x, alpha, scale = theta, log = TRUE));
# estimation of the parameters
x_fit_G = mle(x_nLL_G, start = list(alpha = 60, theta = 1), method = "L-BFGS-B", nobs = length(data$x), lower = c(0.01,0.01));

# Weibull
x_nLL_WB <- function(shape, scale) -sum(stats::dweibull(data$x, shape, scale, log = TRUE))
# Same as above but for the Weibull distribution.
x_fit_WB <- mle(x_nLL_WB, start = list(shape = 1, scale = 50000), method = "L-BFGS-B", nobs = length(data$x), lower = c(0.1,100))

# Draw the Graph

x = seq(0,max(data$x),max(data$x)/1000) # This defines the x-axis range for the following graph to encompas all x data.

plot(x,plnorm(x,coef(x_fit_LN)[1], coef(x_fit_LN)[2]),type="l",col="red", main="ECDF vs Curve Fits")
lines(x,pgamma(x,coef(x_fit_G)[1], coef(x_fit_G)[2]),type="l",col="green")
lines(x,pweibull(x,coef(x_fit_WB)[1], coef(x_fit_WB)[2]),type="l",col="orange")
plot(ecdf(data$x),add=TRUE)
legend("bottomright", legend=c('Gamma','Lognormal',"Weibull"), col=c('green','red',"orange"),pch = 16)
```



Σχήμα 7.2: Διάγραμμα κατανομών Gamma, Lognormal, Weibull σε σύγκριση με την εμπειρική συνάρτηση κατανομής για την τ.μ. X

```
#  Estimate distribution parameters for Y

# Log normal
# function that calculates negative log-likelihood
y_nLL_LN = function(mu, sigma) -sum(stats::dlnorm(data$y, mu, sigma, log = TRUE))
# estimation of the parameters
y_fit_LN = mle(y_nLL_LN, start = list(mu = 10, sigma = 1), method = "L-BFGS-B", nobs = length(data$y), lower = c(1,0.01))



# Gamma
# function that calculates negative log-likelihood
y_nLL_G = function(alpha, theta) -sum(stats::dgamma(data$y, alpha, scale = theta, log = TRUE));
# estimation of the parameters
y_fit_G = mle(y_nLL_G, start = list(alpha = 60, theta = 1), method = "L-BFGS-B", nobs = length(data$y), lower = c(0.01,0.01));

# Weibull
```

```
#function that calculates negative log-likelihood
y_nLL_WB <- function(shape, scale) -sum(stats::dweibull(data$y, shape, scale, log = TRUE))
# estimation of the parameters
y_fit_WB <- mle(y_nLL_WB, start = list(shape = 1, scale = 50000), method = "L-BFGS-B", nobs = length(data$y), lower = c(0.1,100))

# Graph
y = seq(0,max(data$x),max(data$x)/1000) # This defines the x-axis range for the following graph to encompas all x data.
plot(y,plnorm(y,coef(y_fit_LN)[1], coef(y_fit_LN)[2]),type="l",col="red", main="ECDF vs Curve Fits")
lines(y,pgamma(y, shape = coef(y_fit_G)[1],scale = coef(y_fit_G)[2]),type="l",col="green")
lines(y,pweibull(y,coef(y_fit_WB)[1], coef(y_fit_WB)[2]),type="l",col="orange")
plot(ecdf(data$y),add=TRUE)
legend("bottomright", legend=c('Gamma','Lognormal',"Weibull"), col=c('green','red',"orange"),pch = 16)
```



Σχήμα 7.3: Διάγραμμα κατανομών Gamma, Lognormal, Weibull σε σύγκριση με την εμπειρική συνάρτηση κατανομής για την τ.μ. Υ

Παρατηρούμε ότι και στις δύο περιπτώσεις η κατανομή που φαίνεται να έχει την καλύτερη προσαρμογή είναι η Gamma, με τις παραμέτρους που εκτιμήθηκαν με τη μέθοδο μέγιστης πιθανοφάνειας (MLE) να δίνονται στον παρακάτω πίνακα.

	X	Y
alpha	1,347128	2,245659
theta	1,038486	1,016987

Για την τυχαία μεταβλητή X φαίνεται να υπάρχει κάποια απόκλιση από την εμπειρική συνάρτηση κατανομής, ενώ η Y φαίνεται να έχει πολύ καλή προσαρμογή. Για να επιβεβαιώσουμε το κατά πόσο αυτό που φαίνεται οπτικά είναι στατιστικά σημαντικό και κατ'επέκταση μπορούμε να θεωρήσουμε ότι οι τιμές των τυχαίων μεταβλητών θα μπορούσαν να προέρχονται από κατανομές Gamma με τις παραπάνω παραμέτρους θα εφαρμόσουμε τον έλεγχο Cramer – Von Mises, ο οποίος ουσιαστικά βασίζεται στον υπολογισμό του χωρίου που δημιουργείται ανάμεσα στην εκτιμώμενη συνάρτηση κατανομής και την εμπειρική. Τα αποτελέσματα του ελέγχου για τις δύο τυχαίες μεταβλητές συνοψίζονται στον πίνακα που ακολουθεί.

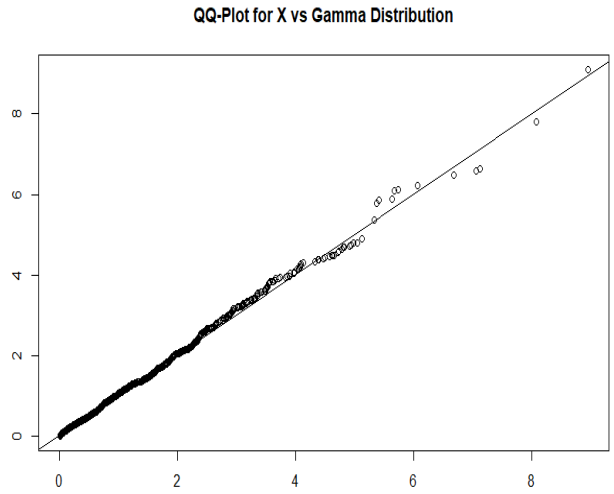
```
#  Goodness of Fit Test (Cramer von Mises)

cvm.test(data$x, "pgamma", shape = coef(x_fit_G)[1], scale = coef(x_fit_G)[2])
cvm.test(data$y, "pgamma", shape = coef(y_fit_G)[1], scale = coef(y_fit_G)[2])
```

Τυχ. Μεταβλητή	Τιμή Στατιστικής Συνάρτησης	p-value
X	0,072051	0,7392
Y	0,086388	0,6559

Όπως φαίνεται η τιμή του p-value και στις δύο περιπτώσεις υποδεικνύει ότι δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση συνεπώς δεχόμαστε ότι τα δύο δείγματα προέρχονται από κατανομές Gamma με τις παραμέτρους που εκτιμήσαμε.

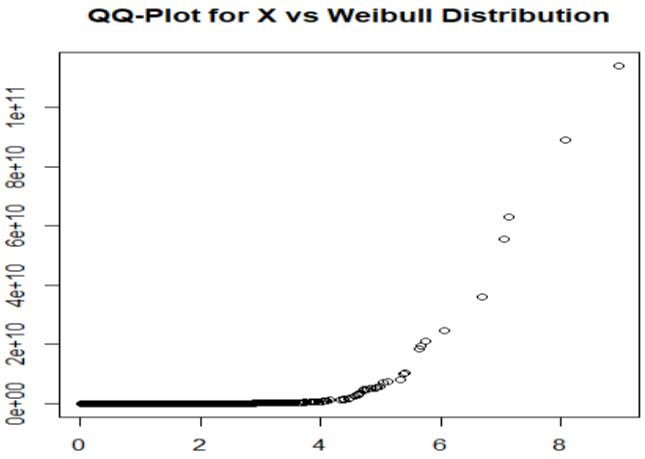
Ένας εναλλακτικός τρόπος ο οποίος θα μπορούσε να μας δώσει μία αρχική ιδέα για την κατανομή των περιθώριων είναι και το διάγραμμα QQ-Plot με το οποίο κάνουμε μία σύγκριση των ποσοστημορίων των δεδομένων μας με τα ποσοστημόρια των κατανομών που επιλέγουμε να εξετάσουμε. Όπως διακρίνουμε στα παρακάτω σχήματα και με αυτόν τον τρόπο θα ήταν λογικό να συμπεράνουμε ότι τόσο η κατανομή που ακολουθεί η X όσο και η Y είναι Gamma.

```
# Gamma
qqplot(data$x, rgamma(1000, shape=coef(x_fit_G)[1], scale = coef(x_fit_G)[2]), xlab='', ylab='', main = 'QQ-Plot for x vs Gamma Distribution')
abline(0,1)
```



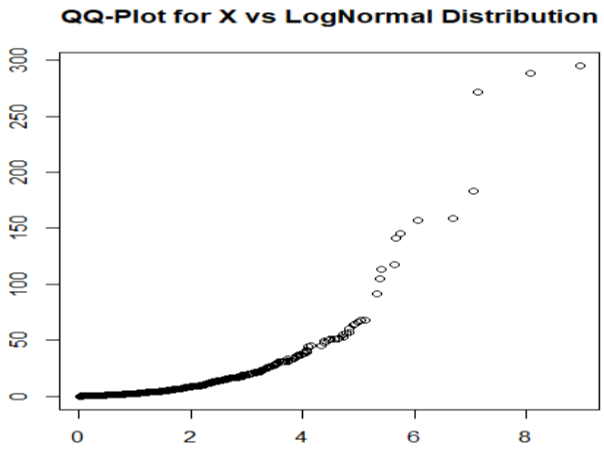
Σχήμα 7.4: QQ – plot για την τ.μ. X

```
# Weibull
qqplot(data$x, rweibull(1000, shape=coef(x_fit_WB)[1], scale = coef(x_fit_WB)[2]), xlab='', ylab='',
main = 'QQ-Plot for x vs Weibull Distribution')
```



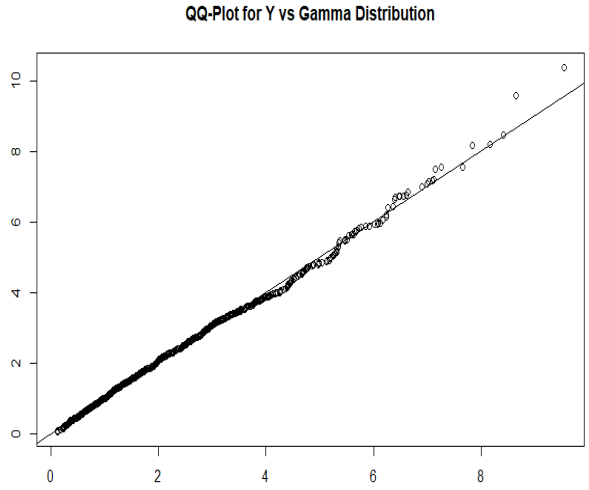
Σχήμα 7.5: QQ – plot για την τ.μ. X

```
#LogNormal
qqplot(data$x, rlnorm(1000, meanlog=coef(x_fit_LN)[1], sdlog = coef(x_fit_LN)[2]), xlab='',
ylab='',  main = 'QQ-Plot for x vs LogNormal Distribution')
```



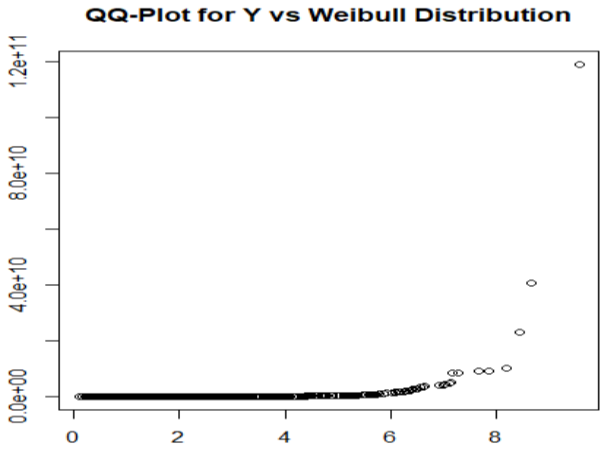
Σχήμα 7.6: QQ – plot για την τ.μ. X

```
# Gamma
qqplot(data$y, rgamma(1000, shape=coef(y_fit_G)[1], scale = coef(y_fit_G)[2]), xlab='', ylab='',
main = 'QQ-Plot for y vs Gamma Distribution')
abline(0,1)
```



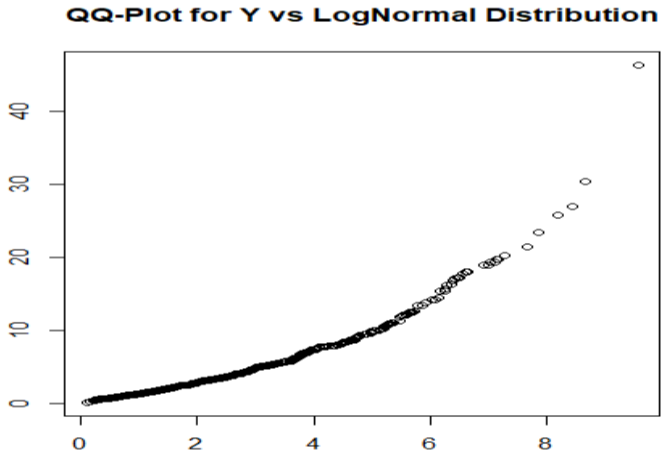
Σχήμα 7.7: QQ – plot για την τ.μ. Y

```
#Weibull
qqplot(data$y, rweibull(1000, shape=coef(y_fit_WB)[1], scale = coef(y_fit_WB)[2]), xlab='', ylab='', main =
'QQ-Plot for Y vs Weibull Distribution')
```



Σχήμα 7.8: QQ – plot για την τ.μ. Υ

#Lognormal
qqplot(data$y, rlnorm(1000, meanlog=coef(y_fit_LN)[1], sdlog = coef(y_fit_LN)[2]), xlab='', ylab='', main = 'QQ-Plot for Y vs LogNormal Distribution')



Σχήμα 7.9: QQ – plot για την τ.μ. Υ

Όπως μπορούμε να παρατηρήσουμε και από τα QQ-Plots η κατανομή Gamma φαίνεται να έχει την καλύτερη προσαρμογή. Αρκετά καλή προσαρμογή εμφανίζει και η κατανομή Lognormal η οποία αρχίζει να αποκλίνει όσο πλησιάζουμε προς τη δεξιά ουρά. Τέλος, η κατανομή Weibull δείχνει να μην προσαρμόζεται καλά στα δεδομένα και σε σύγκριση με τις προαναφερθείσες είναι σαφώς αυτή με την λιγότερο καλή προσαρμογή.

Το επόμενο βήμα μετά την εκτίμηση των περιθώριων κατανομών είναι η επιλογή του κατάλληλου Copula που πρσαρμόζεται καλύτερα στα δεδομένα μας. Στην παρούσα εργασία θα εξετάσουμε την προσαρμογή τριών Copulas που ανήκουν στην κλάση των Αρχιμήδειων και συγκεκριμένα τα Gumbel, Clayton και Frank. Αφού γίνει εκτίμηση των παραμέτρων για κάθε μία από αυτές τις οικογένειες θα εφαρμόσουμε τον έλεγχο Cramer – Von Mises για να διαπιστώσουμε αν και κατά πόσο κάποιο από τα προαναφερθέντα Copulas έχει καλή προσαρμογή για τις εκτιμώμενες παραμέτρους.

Αρχικά θα κάνουμε εκτίμηση των παραμέτρων με τη μέθοδο CML η οποία χρησιμοποιείται όταν δεν γνωρίζουμε τις περιθώριες κατανομές και όπως αναφέρθηκε στην παράγραφο 6.2 βασίζεται στα ranks των παρατηρήσεων διαιρεμένα με το πλήθος τους συν ένα. Τα αποτελέσματα συνοψίζονται στον παρακάτω πίνακα.

```
# CML Method

# transform our data to pseudo observations so we can feed the copula
cmlPseudodata = pobs(data);

# Clayton
cml_clayton <- claytonCopula(dim = 2);
cml_clayton_fit <- fitCopula(cml_clayton, cmlPseudodata, method = 'mpl');
cml_clayton_alpha = coef(cml_clayton_fit); ## round it to 1.65
cml_clayton_tau = tau(claytonCopula(param = 1.65, dim = 2));
summary(cml_clayton_fit)

# Frank
cml_frank <- frankCopula(dim = 2);
cml_frank_fit <- fitCopula(cml_frank, cmlPseudodata, method = 'mpl');
cml_frank_alpha = coef(cml_frank_fit); ## round it to 4.88
cml_frank_tau = tau(frankCopula(param = 4.88, dim = 2));
summary(cml_frank_fit)

# Gubmel
cml_gumbel <- gumbelCopula(dim = 2);
```

```
cml_gumbel_fit <- fitCopula(cml_gumbel, cmlPseudodata, method = 'mpl');
cml_gumbel_alpha = coef(cml_gumbel_fit); ## round it to 1.60
cml_gumbel_tau = tau(gumbelCopula(param = 1.60, dim = 2));
summary(cml_gumbel_fit)
```

Copula	Estimator	Std. Error
Gumbel	1,605	0,035
Clayton	1,647	0,106
Frank	4,881	0,262

Στη συνέχεια θα εφαρμόσουμε τον έλεγχο Cramer – Von Mises για να εξετάσουμε ποιό από τα παραπάνω Copulas έχει καλύτερη προσαρμογή.

```
gofCopula(claytonCopula(coef(ifm_clayton_fit),dim=2),data)
gofCopula(gumbelCopula(coef(ifm_gumbel_fit), dim=2, data)
gofCopula(frankCopula(coef(ifm_frank_fit), dim=2), data)
```

Copula	Τιμή Στατιστικής Συνάρτησης	p-value
Gumbel	0,6523	0,0005
Clayton	0,014269	0,6578
Frank	0,277	0,0005

Σύμφωνα με τον παραπάνω πίνακα στον οποίο συνοψίζονται τα αποτελέσματα του τεστ σε επίπεδο σημαντικότητας α = 0,05, το Copula που έχει την καλύτερη προσαρμογή στα δεδομένα μας είναι το Clayton Copula με παράμετρο $\tilde{\theta}$ = 1,647. Παρατηρώντας τις τιμές των στατιστικών συναρτήσεων είναι ένα αποτέλεσμα που φαίνεται λογικό καθώς στην ουσία η τιμή της στατιστικής συνάρτησης είναι ένα μέτρο απόστασης μεταξύ του πραγματικού Copula και των προς εξέταση οικογενειών. Είναι φανερό λοιπόν πως τη μικρότερη τιμή έχει το Clayton Copula με 0,014269. Τα p-values επιβεβαιώνουν το αποτέλεσμα καθώς το Gumbel και το Frank έχουν τιμές μικρότερες του 0,05 που είναι το επίπεδο σημαντικότητας και επομένως θα πρέπει να απορρίψουμε την υπόθεση ότι το δείγμα μας προέρχεται από μία εκ των δύο αυτών οικογενειών.

Έχει αξία όμως να δούμε ποιές θα ήταν οι εκτιμώμενες τιμές των παραμέτρων αν δεν εφαρμόζαμε την μέθοδο CML αλλά την IFM. Με την μέθοδο αυτή θα εφαρμόσουμε τις περιθώριες κατανομές (Gamma) που εκτιμήσαμε αρχικά και αντί των (X, Y), με X ~ F, Y ~ G, θα χρησιμοποιήσουμε τα (U, V), όπου U = F(X) ~ U(0,1) και V = G(Y) ~ U(0,1). Τα αποτελέσματα συνοψίζονται παρακάτω.

```
# IFM Method
```

```
# transform our data to observations ~ U(0,1) so we can feed the copula
x_pgamma = pgamma(data$x, shape = coef(x_fit_G)[1], scale = coef(x_fit_G)[2]);
y_pgamma =  pgamma(data$y, shape = coef(y_fit_G)[1], scale = coef(y_fit_G)[2]);
ifmData = cbind(x_pgamma, y_pgamma);

# Clayton
ifm_clayton <- claytonCopula(dim = 2);
ifm_clayton_fit <- fitCopula(ifm_clayton, ifmData, method = 'ml');
ifm_clayton_alpha = coef(ifm_clayton_fit); ## round it to 1.63
ifm_clayton_tau = tau(claytonCopula(param = 1.63, dim = 2));
summary(ifm_clayton_fit)

# Frank
ifm_frank <- frankCopula(dim = 2);
ifm_frank_fit <- fitCopula(ifm_frank, ifmData, method = 'ml');
ifm_frank_alpha = coef(ifm_frank_fit); ## round it to 4.82
ifm_frank_tau = tau(frankCopula(param = 4.82, dim = 2));
summary(ifm_frank_fit)
# Gubmel
ifm_gumbel <- gumbelCopula(dim = 2);
ifm_gumbel_fit <- fitCopula(ifm_gumbel, ifmData, method = 'ml');
ifm_gumbel_alpha = coef(ifm_gumbel_fit); ## round it to 1.61
ifm_gumbel_tau = tau(gumbelCopula(param = 1.61, dim = 2));
summary(ifm_gumbel_fit)
```

Copula	**Estimator**	**Std. Error**
Gumbel	1,608	0,040
Clayton	1,630	0,077
Frank	4,823	0,227

Παρατηρούμε ότι οι τιμές των εκτιμητών δεν έχουν μεγάλες διαφορές σε σχέση με την προηγούμενη μέθοδο. Μια λογική εξήγηση είναι ότι η εκτίμηση των περιθώριων είναι αρκετά καλή συνεπώς θα έχουμε και καλές εκτιμήσεις για τις παραμέτρους των Copulas. Μάλιστα οι εκτιμήσεις των παραμέτρων για το Clayton και το Frank παρουσιάζουν μικρότερο τυπικό σφάλμα ενώ το Gumbel λίγο μεγαλύτερο. Μία εξήγηση θα μπορούσε να είναι ότι όπως είδαμε η προσαρμογή στις δεξιές ουρές των περιθώριων δεν ήταν τόσο καλή και συνήθως η Gumbel προσαρμόζεται καλύτερα σε δεδομένα που παρουσιάζουν υψηλότερη εξάρτηση στη δεξιά ουρά.

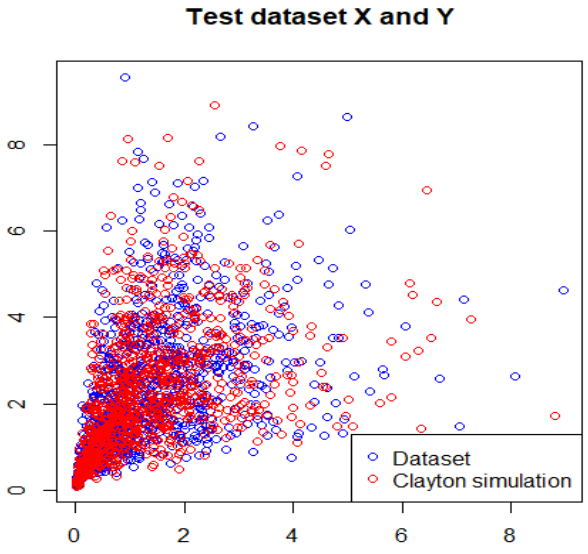
Η μέθοδος IFM παρουσιάζει το μειονέκτημα ότι βασίζεται στις περιθώριες κατανομές. Συνεπώς αν κάνουμε λανθασμένη εκτίμηση για την κατανομή τους θα κάνουμε και λανθασμένη εκτίμηση της

παραμέτρου του Copula. Αντίθετα η CML βασίζεται μόνο στις παρατηρήσεις και όταν έχουμε μεγάλο δείγμα είναι και αυτή που χρησιμοποιείται συνήθως ώστε να αποφευχθεί η πιθανότηατα λανθασμένης εκτίμησης της κατανομής των περιθώριων.

Ένα τελευταίο βήμα που μπορούμε να ακολουθήσουμε έτσι ώστε να έχουμε μία εικόνα για το Copula που επιλέχθηκε είναι να παράγουμε χίλιες τιμές (δηλαδή το ίδιο πλήθος με το δείγμα μας) από το Clayton Copula με παράμετρο $\tilde{\theta}$ = 1,63 και να τις απεικονίσουμε σε ένα κοινό διάγραμμα με το αρχικό μας δείγμα. Η εικόνα που προκύπτει είναι η ακόλουθη.

```
# Clayton
# clayton distribution
clayton_dist <- mvdc(claytonCopula(param = 1.63, dim = 2), margins = c("gamma","gamma"), paramMargins = list(list(shape = coef(x_fit_G)[1], scale = coef(x_fit_G)[2]), list(shape = coef(y_fit_G)[1], scale = coef(y_fit_G)[2])))
# clayton distribution simulation for n=1000
clayton_sim <- rMvdc(1000, clayton_dist)
```

```
# plot of actual data and clayton simulation
plot(data$x, data$y, main = 'Test dataset X and Y', xlab = "", ylab = "", col = "blue")
points(clayton_sim1[,1], clayton_sim1[,2], col = 'red')
legend("bottomright", legend=c('Dataset','Clayton simulation'),col=c('blue','red'), pch = 1)
```



Σχήμα 7.10: Κοινό διάγραμμα διασποράς του δείγματος και τυχαίων τιμών από το εκτιμηθέν Copula

Όπως μπορούμε να διακρίνουμε οι τιμές που παίρνουμε από το μοντέλο είναι αρκετά κοντά στις τιμές που έχουν τα δεδομένα μας. Ένας υψηλότερος βαθμός συσχέτισης θα εξασφάλιζε μια καλύτερη ίσως προσαρμογή αλλά φαίνεται πως το μοντέλο παρουσιάζει ένα ικανοποιητικό βαθμό πρόβλεψης.

8. Βιβλιογραφία

[1] Artzner Philippe, Delbaen Freddy, Eber Jean – Marc & Heath David, (1998), "Coherent Measures of Risk", *Mathematical Finance*, **9**, (3), 203 - 228

[2] Balakrishnan N. & Lai Chin-Diew, (2008), "Continuous Bivariate Distributions," Second edition, Springer New York

[3] Brahim Brahimi , Fatah Benatia & Djabrane Yahia, (2017), "Copula conditional tail expectation for multivariate financial risks", *Arab Journal of Mathematical Science*, **24**, (1), 82 – 100

[4] Denuit M., DhaeneJ. , Goovaerts M. & Kaas R., (2005), "*Actuarial Theory for Dependent Risks*", Jon Wiley & Sons, Ltd

[5] Dhaene Jan, Denuit Michel, Goovaerts Marc, Kaas Rob (2002), "The Concept of Comonotonicity in Actuarial Science and Finance: Theory", *Insurance: Mathematics & Economics*, **31**, (1), 3 – 33

[6] Embrechts Paul, Lindskog Filip & McNeil Alexander, (2001), "*Modelling Dependence with Copulas and Applications to Risk Management*". Department of Mathematics, ETHZ, CH-8092, Zurich, Switzerland

[7] Embrechts Paul, Mcneil Alexander & Straumann Daniel, (2002), "*Correlation and Dependency in Risk Management: Properties and Pitfalls*", Department of Mathematics, ETHZ, Zurich

[8] Genest Christian & Favre Anne – Catherine, (2007), "Everything you always wanted to know about Copula Modeling but were afraid to ask", *Journal of Hydrologic Engineering*, **12**, (4), 347 - 367

[9] Genest Christian & Louis – Rivest Paul, (1993), "Statistical Inference Procedures for Bivariate Archimedean Copulas", *Journal of the American Statistical Association*, **88**, (423) , 1034 – 1043

[10] Genest C., Ghoudi K. & Rivest L. – P., (1995), "*A semiparametric estimation procedure of dependence parameters in multivariate families of distributions*", Department de mathematiques et de statistique, Universite Laval, Quebec, Canada G1K 7P4

[11] Jaworski Piotr, Durante Fabrizio, Hardle Wolfgang & Rychlik Tomasz, (2010), "*Copula Theory and Its Applications*", Proceedings of the Workshop Held in Warsaw 25 – 26 September 2009, Springer

[12] Kluppelberg Claudia, Stelzer Robert, (2014), "*Dealing with Dependent Risks*", Springer International Publishing

[13] Lehmann E.L., (1966), "Some Concepts of Dependence", *Annals of Mathematical Statistics,* **37**, (5), 1137 – 1153

[14] Mail Jan-Frederik & Scherer Matthias, (2012), "*Simulating Copulas Stochastic Models, Sampling Algorithms, and Applications, with contributions by Claudia Czado, Elke Korn, Ralf Korn & Jakob Stöber*", Series in Quantitative Finance – Vol. 4, Imperial College Press

[15] Mikosch Thomas, (2006), "Copulas: Tales and Facts", *Extremes*, **9**, (1), 3 – 20

[16] Nelsen Roger B., (2006), "*An Introduction to Copulas*", 2nd edition, Springer Series in Statistics, Springer

[17] Weiss Gregor N.F. (2013), "*Identifying mixture copula components using outlier detection methods and goodness-of-fit tests*", Technische Universitat Dortmund