ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΟΜΕΑΣ ΠΙΘΑΝΟΤΗΤΩΝ, ΣΤΑΤΙΣΤΙΚΗΣ
& ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

ΣΥΝΑΡΤΗΣΕΙΣ COPULA: ΜΙΑ ΑΝΑΣΚΟΠΗΣΗ

ΕΥΑΓΓΕΛΙΑ ΤΑΓΚΑΡΕΛΗ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννινα, 2018

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη

ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ

που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 22/06/2018 από την εξεταστική επιτροπή:

Ζωγράφος Κωνσταντίνος, Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων **(Επιβλέπων).**

Λουκάς Σωτήριος, Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Μπατσίδης Απόστολος, Επίκουρος Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

«Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή.»

Ταγκαρέλη Ευαγγελία

Αφιερώνεται
στην οικογένειά μου.

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων και επιβλέποντα κ. Κωνσταντίνο Ζωγράφο για την πολύτιμη βοήθεια του, την καθοδήγηση αλλά και την κατανόηση που έδειξε σε όλο το διάστημα εκπόνησης της παρούσας Μεταπτυχιακής Διατριβής.

Επιπλέον θα ήθελα να ευχαριστήσω και τα άλλα δύο μέλη της τριμελούς επιτροπής τον Καθηγητή κ. Σωτήριο Λουκά και τον Επίκουρο Καθηγητή κ. Απόστολο Μπατσίδη για το χρόνο που αφιέρωσαν για την ανάγνωση και αξιολόγηση της διατριβής.

Οφείλω ακόμα να εκφράσω τις ευχαριστίες μου και στο υπόλοιπο διδακτικό προσωπικό του Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων για τις γνώσεις που μου προσέφεραν.

Τέλος, θα ήθελα να ευχαριστήσω όλους εκείνους που ήταν δίπλα μου και με στήριξαν, μα πάνω από όλους την οικογένειά μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1. Ιστορική αναδρομή

Το πρόβλημα της διδιάστατης (πολυδιάστατης) εξάρτησης έχει απασχολήσει τους ερευνητές για περισσότερο από έναν αιώνα, δεδομένου ότι η ανεξαρτησία των δεδομένων είναι συχνά μία επιθυμητή κατάσταση (Micheas and Zografos, 2006). Για την μελέτη της εξάρτησης μεταξύ δύο, ή περισσότερων τυχαίων μεταβλητών, χρησιμοποιούνται οι λεγόμενες συναρτήσεις copula. Το εδάφιο αυτό, στηρίζεται στη μονογραφία του Nelsen (2006).

Οι συναρτήσεις copula είναι ένα σχετικά πρόσφατο αντικείμενο στην επιστήμη της στατιστικής. Πολλοί επιστήμονες από διαφορετικούς επιστημονικούς κλάδους ασχολούνται με τις συναρτήσεις copula διότι συνήθως τα δεδομένα που μελετώνται σε αυτούς είναι πολυδιάστατα και παρουσιάζουν συσχέτιση.

Ο όρος copula, με μαθηματική έννοια, εισήχθη το 1959 από τον Abe Sklar, ο οποίος χρησιμοποίησε αυτό τον όρο για να περιγράψει και να αποδείξει ένα θεώρημα, που πλέον φέρει το όνομά του, σε μία επιστολή του προς τον Fréchet. Αυτό βέβαια δεν σημαίνει ότι οι συναρτήσεις copula δεν υπήρχαν πριν. Η ιδέα των συναρτήσεων copula είχε εμφανιστεί προηγουμένως στο έργο του Hoeffding (1940, 1941), ο οποίος ασχολήθηκε κυρίως με διδιάστατες τυποποιημένες κατανομές και καθιέρωσε τα καλύτερα δυνατά όρια γι΄αυτές τις συναρτήσεις. Στην δημιουργία αυτών των ορίων συνετέλεσε και ο Fréchet (1951, 1958).

Πολλοί συγγραφείς (Gumbel, 1961, Plackett, 1965, Mardia, 1970, Ali, Mikhail and Haq, 1978, Clayton, 1978, Genest and Mackay, 1986, Nelsen, 1986, 1987) χρησιμοποίησαν συναρτήσεις copula για να παράγουν νέες διδιάστατες κατανομές με προκαθορισμένες τις περιθώριες κατανομές.

Για να εξηγηθεί ο λόγος που ο Sklar χρησιμοποίησε τον όρο copula γι'αυτές τις συναρτήσεις επισημαίνεται αρχικά ότι η λέξη copula στην Λατινική Γλώσσα σημαίνει δεσμός. Στην ουσία είναι ένας όρος της γραμματικής που χρησιμοποιείται σε μία πρόταση για να ενώσει το υποκείμενο με το κατηγόρημα. Η επιλογή της ονομασίας copula από τον Sklar, λοιπόν, δεν ήταν τυχαία καθώς και εδώ υπάρχει ένα είδος δεσμού μεταξύ των περιθώριων κατανομών και της από κοινού συνάρτησης κατανομής. Έτσι ο όρος copula στην ελληνική γλώσσα έχει μεταφραστεί και διαδοθεί ως σύζευξη. Μία copula, λοιπόν, είναι μια συνάρτηση που συνδέει μία πολυδιάστατη συνάρτηση κατανομής με τις περιθώριες μονοδιάστατες συναρτήσεις κατανομών της.

Ένα ερώτημα που θα μπορούσε να δημιουργηθεί είναι γιατί οι συναρτήσεις copula ενδιαφέρουν τους στατιστικούς. Απάντηση στο ερώτημα αυτό δίνεται από τον Fisher (1997) καθώς όπως επισημαίνει ο ίδιος στην Εγκυκλοπαίδεια των Στατιστικών Επιστημών << Οι copulas ενδιαφέρουν τους στατιστικούς για δύο κυρίους λόγους: Πρώτον, ως τρόπος μελέτης των μέτρων κλίμακας χωρίς εξάρτηση και δεύτερον, ως αφετηρία για την οικοδόμηση οικογενειών διδιάστατων κατανομών, … >>.

1.2. Σκοπός της μεταπτυχιακής διατριβής και περίληψη

Σκοπός της μεταπτυχιακής διατριβής είναι η παρουσίαση της έννοιας των συναρτήσεων copula καθώς και η χρησιμότητά τους όταν τα δεδομένα που χρησιμοποιούνται παρουσιάζουν εξάρτηση. Πιο συγκεκριμένα, στο Κεφάλαιο 2, παρουσιάζεται το Θεώρημα του Sklar, το οποίο αποτελεί τη βάση όλης αυτής της θεωρίας, ο ορισμός των συναρτήσεων copula και γίνεται αναφορά στα φράγματα Fréchet-Hoeffding. Επίσης, παρουσιάζονται κάποιες βασικές ιδιότητες των συναρτήσεων copula και γίνεται αναφορά στις copula πυκνότητες. Στη συνέχεια, αναπτύσσεται η περίπτωση των συναρτήσεων copula επιβίωσης (survival copulas), γίνεται αναφορά σε κάποιες γνωστές οικογένειες copula και παρουσιάζονται τα μέτρα εξάρτησης και συσχέτισης των συναρτήσεων copula, συγκεκριμένα ο συντελεστής γραμμικής συσχέτισης του Pearson, ο συντελεστής συσχέτισης τάξης Rho του Spearman και Tau του Kendall καθώς επίσης και ο συντελεστής εξάρτησης ουρών. Στο Κεφάλαιο 3 ασχολούμαστε με τη διδιάστατη συνάρτηση copula ακραίων τιμών (extreme value) δίνοντας αρχικά τον ορισμό αυτών των συναρτήσεων και στη συνέχεια διατυπώνουμε το πώς συνδέεται η έννοια της copula ακραίων τιμών με την έννοια της μεγιστο-ευσταθούς copula. Στη συνέχεια παρουσιάζεται μια διαδικασία κατασκευής copula ακραίων τιμών, η οποία οφείλεται στην εργασία του Pickands (1981). Ακόμη, γίνεται αναφορά σε κάποιες ειδικές περιπτώσεις συναρτήσεων εξάρτησης Pickands αλλά και στην πυκνότητα μιας διδιάστατης συνάρτησης copula ακραίων τιμών. Κλείνουμε το κεφάλαιο αυτό, κάνοντας αναφορά στη μορφή της συνάρτησης εξάρτησης Pickands για κάποιες γνωστές οικογένειες copula αλλά και για κάποιες κατανομές και παρουσιάζοντας τους συντελεστές συσχέτισης των Kendall και Spearman για τις συναρτήσεις αυτές. Στο Κεφάλαιο 4 αναπτύσσουμε μία από τις πιο σημαντικές κλάσεις των συναρτήσεων copula, τις συναρτήσεις Archimedean copulas. Αρχικά, παρουσιάζονται κάποιοι βασικοί ορισμοί αλλά και θεωρήματα γι' αυτές τις συναρτήσεις καθώς και κάποιες αλγεβρικές ιδιότητές τους. Στη συνέχεια, διατυπώνονται τα ισοσταθμικά σύνολα (level sets) μιας Archimedean copula και η πυκνότητα της διδιάστατης συνάρτησης κατανομής. Ακόμη αναφέρουμε τα μέτρα εξάρτησης και συσχέτισης και επικεντρωνόμαστε στις οικογένειες Clayton, Frank και Gumbel. Η διατριβή ολοκληρώνεται με τον επίλογο, την περίληψη στα αγγλικά καθώς και με την βιβλιογραφία.

ΚΕΦΑΛΑΙΟ 2

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΝΝΟΙΑ ΤΩΝ COPULAS

2.1. Εισαγωγή

Σκοπός του κεφαλαίου αυτού είναι μια ανασκόπηση της βιβλιογραφίας, σχετικής με τις συναρτήσεις copula όπως αυτή παρουσιάζεται στις αναφορές στη βιβλιογραφία και ιδιαίτερα στις μονογραφίες των Drouet Mari and Kotz (2001), Nelsen (2006) και Joe (2015). Έτσι στο κεφάλαιο αυτό θα παρουσιασθεί ο ακριβής ορισμός της συνάρτησης copula, το Θεώρημα του Sklar (1959), τα φράγματα Fréchet-Hoeffding για μια συνάρτηση copula καθώς και κάποιες βασικές ιδιότητες των συναρτήσεων copula. Τα αποτελέσματα θα παρουσιασθούν εστιάζοντας στη διδιάστατη περίπτωση, χάριν απλότητας. Το σύνολο σχεδόν από αυτά εξακολουθεί να ισχύει και στην πολυδιάστατη περίπτωση με ανάλογη όμως αύξηση της πολυπλοκότητας του φορμαλισμού.

2.2. Θεώρημα του Sklar (1959)

Πρωταγωνιστικό ρόλο στην ανάπτυξη των συναρτήσεων copula έπαιξε το Θεώρημα του Sklar το οποίο διατυπώθηκε το 1959 και αποτελεί μία σύνδεση, μία σύζευξη μεταξύ της από κοινού κατανομής δύο τυχαίων μεταβλητών και των αντίστοιχων περιθωρίων κατανομών. Το Θεώρημα αυτό διατυπώνεται στη συνέχεια, αρχικά για την περίπτωση της διδιάστατης συνάρτησης κατανομής και η απόδειξη του σκιαγραφείται βασιζόμενη στην πρόσφατη εργασία του Rüschendorf (2009). Πριν όμως από το Θεώρημα του Sklar θα δοθεί μια πρώτη προσέγγιση του ορισμού της συνάρτησης copula, ακολουθώντας τη μονογραφία του Nelsen (2006, σελ. 7). Έτσι, έστω X_1 και X_2 δύο συνεχείς τυχαίες μεταβλητές με αντίστοιχες συναρτήσεις κατανομής $F_1(x_1)$ και $F_2(x_2)$ και από κοινού συνάρτηση κατανομής $F(x_1, x_2)$. Τότε για $(x_1, x_2) \in [-\infty, +\infty] \times [-\infty, +\infty]$, θεωρούμε το σημείο του I^3, $I = [0,1]$, με συντεταγμένες $\big(F_1(x_1), F_2(x_2), F(x_1, x_2)\big)$. Αυτή ακριβώς η απεικόνιση από το $I^2 = [0,1] \times [0,1]$ στο $I = [0,1]$ ονομάζεται copula.

Αυτός ο πρωτόλειος ορισμός της συνάρτησης copula συνδέεται κατευθείαν με το ακόλουθο θεώρημα που οφείλεται στον Sklar (1959) και εκφράζει μια πολυδιάστατη κατανομή σε όρους, αφ' ενός των περιθώριων κατανομών και αφετέρου της συνάρτησης copula, η οποία και σχηματοποιεί την εξάρτηση των τυχαίων μεταβλητών.

Θεώρημα 2.2.1. Αν $F(x, y)$ η από κοινού α.σ.κ. των τυχαίων μεταβλητών X, Y, με $F_1(x), F_2(y)$ τις περιθώριες α.σ.κ., τότε υπάρχει μία διδιάστατη copula C, τέτοια ώστε:

$$F(x, y) = C\big(F_1(x), F_2(y)\big).$$

Απόδειξη. Η απόδειξη που θα δοθεί αναφέρεται στην περίπτωση που οι τυχαίες μεταβλητές X_1 και X_2 είναι συνεχείς και η αντίστοιχη από κοινού κατανομή $F(x_1, x_2)$ είναι συνεχής. Στην περίπτωση γενικών από κοινού κατανομών που περιέχουν και διακριτό μέρος η απόδειξη δίδεται στην εργασία του Rüschendorf (2009) και βασίζεται στον distributional transformation των τυχαίων μεταβλητών X_1 και X_2, που επιτρέπει τον χειρισμό γενικών κατανομών που περιέχουν και διακριτό μέρος.

Ας επανέλθουμε στην περίπτωση συνεχών τυχαίων μεταβλητών X_1 και X_2 και ας θεωρήσουμε τις τυχαίες μεταβλητές U_1 και U_2 που ορίζονται μέσω των μετασχηματισμών ολοκληρώματος (βλέπε, Χαραλαμπίδη, 1990, σελ. 231 σχετικά με την ορολογία)

$$U_i = F_i(X_i), \text{ για } i = 1,2.$$

Τότε, $U_i \sim U(0,1)$ και $X_i = F_i^{-1}(U_i)$, με $F_i^{-1}(u) = \inf\{x \in R: F_i(x) \geq u\}, u \in (0,1)$. Επιστρέφοντας στην απόδειξη του θεωρήματος, και ορίζοντας τη συνάρτηση C να είναι η από κοινού συνάρτηση κατανομής του τυχαίου διανύσματος (U_1, U_2), προκύπτει,

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = P(F_1^{-1}(U_1) \leq x_1, F_2^{-1}(U_2) \leq x_2)$$

$$= P(U_1 \leq F_1(x_1), U_2 \leq F_2(x_2)) = C\big(F_1(x_1), F_2(x_2)\big),$$

και η απόδειξη του θεωρήματος έχει ολοκληρωθεί. ■

Το Θεώρημα του Sklar συνδέεται επίσης και με τον ακόλουθο, τυπικό, ορισμό της copula.

Ορισμός 2.2.1. Ο όρος copula αναφέρεται σε μια συνάρτηση $C: [0,1] \times [0,1] \to [0,1]$ δύο μεταβλητών u, v τέτοια ώστε να ικανοποιούνται οι παρακάτω ιδιότητες:

1. $C(0, u) = C(u, 0) = 0$, για κάθε $u \in [0,1]$ και ομοίως, $C(0, v) = C(v, 0) = 0$, για κάθε $v \in [0,1]$. Επιπλέον,
   $C(1, u) = C(u, 1) = u$, για κάθε $u \in [0,1]$ και ομοίως, $C(1, v) = C(v, 1) = v$, για κάθε $v \in [0,1]$.
2. Η C ικανοποιεί την rectangular inequality
   $$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0,$$
   για κάθε $u_1,\ u_2,\ v_1,\ v_2 \in [0,1]$ τέτοια ώστε $u_1 \leq u_2\ \ και\ v_1 \leq v_2$.
   Αυτή η ανισότητα δηλώνει ότι η C είναι 2-increasing και ειδικότερα σημαίνει ότι η $C(u, v)$ αυξάνεται ως προς τις δύο μεταβλητές.

Απόρροια του ορισμού αυτού και ειδικότερα της 2-increasing ιδιότητας, είναι οι ακόλουθες ιδιότητες της copula, σύμφωνα με τους Drouet Mari and Kotz (2001, σελ. 67).

Ιδιότητα της Συνέχειας – Συνθήκη Lipschitz:

Κάθε συνεχής copula C ικανοποιεί την συνθήκη του Lipschitz, δηλαδή:

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

για κάθε $u_1,\ \ u_2,\ v_1,\ v_2 \in [0,1]$.

Παραγωγισιμότητα:

Εφόσον, η $C(u, v)$ είναι αύξουσα και συνεχής, έχουμε ότι είναι διαφορίσιμη σχεδόν παντού και από την προηγούμενη ιδιότητα προκύπτει ότι

$$0 \leq \frac{\partial}{\partial u} C(u, v) \leq 1 \text{ και } 0 \leq \frac{\partial}{\partial v} C(u, v) \leq 1.$$

Η ύπαρξη των μερικών παραγώγων $\partial C(u, v)/\partial u$ και $\partial C(u, v)/\partial v$ είναι άμεση επειδή οι μονότονες συναρτήσεις είναι διαφορίσιμες σχεδόν παντού. Οι παραπάνω σχέσεις ισχύουν, θέτοντας στην προηγούμενη ιδιότητα, $v_1 = v_2$ και $u_1 = u_2$ αντίστοιχα. Αν $v_1 \leq v_2$ τότε σύμφωνα με τον Nelsen (2006, σελ. 9) η συνάρτηση $u \mapsto C(u, v_2) - C(u, v_1)$ είναι μη φθίνουσα στο $[0, 1]$. Έτσι η ποσότητα $\partial\big(C(u, v_2) - C(u, v_1)\big)/\partial u$ ορίζεται και δεν είναι αρνητική σχεδόν παντού στο $[0, 1]$, από το οποίο προκύπτει ότι $v \mapsto \partial C(u, v)/\partial u$ ορίζεται και είναι μη φθίνουσα σχεδόν παντού στο $[0, 1]$. Για την δεύτερη σχέση τώρα αν $u_1 \leq u_2$ τότε σύμφωνα με τον Nelsen (2006, σελ. 9) η συνάρτηση $v \mapsto C(u_2, v) - C(u_1, v)$ είναι μη φθίνουσα στο $[0, 1]$. Έτσι η ποσότητα $\partial\big(C(u_2, v) - C(u_1, v)\big)/\partial v$ ορίζεται και δεν είναι αρνητική σχεδόν παντού στο $[0, 1]$, από το οποίο προκύπτει ότι $u \mapsto \partial C(u, v)/\partial v$ ορίζεται και είναι μη φθίνουσα σχεδόν παντού στο $[0, 1]$.

Ακολουθούν ορισμένες παρατηρήσεις στο Θεώρημα Sklar, βασιζόμενοι στη βιβλιογραφία και ειδικότερα στους Drouet Mari and Kotz (2001, σελ. 66) και Nelsen (2006).

Παρατηρήσεις 2.2.1. (α) Αν οι περιθώριες κατανομές F_1 και F_2 είναι συνεχείς, τότε η copula C, των X_1 και X_2, είναι μοναδική, είναι η διδιάστατη συνάρτηση κατανομής των $U_1 = F_1(X_1)$ και $U_2 = F_2(X_2)$, με ομοιόμορφες περιθώριες στο $[0, 1]$.

(β) Αν $F_1(x_1)$ και $F_2(x_2)$ είναι συναρτήσεις κατανομής και C μια συνάρτηση copula, τότε η διδιάστατη συνάρτηση $F(x_1, x_2)$ που ορίζεται στο Θεώρημα Sklar είναι μια διδιάστατη συνάρτηση κατανομής με περιθώριες $F_1(x_1)$ και $F_2(x_2)$.

(γ) Η γενίκευση του Θεωρήματος Sklar στις m διαστάσεις είναι άμεση και διατυπώνεται ως ακολούθως:

Έστω οι τυχαίες μεταβλητές $X_1, X_2, \dots, X_m$ με από κοινού α.σ.κ. την $F(x_1, x_2, \dots, x_m)$ και $F_i(x_i)$, $i = 1, \dots, m$ είναι οι περιθώριες α.σ.κ.. Τότε υπάρχει μία m-διάστατη copula C τέτοια ώστε

$$F(x_1, x_2, \dots, x_m) = C\big(F_1(x_1), F_2(x_2), \dots, F_m(x_m)\big).$$

(δ) Τρείς σημαντικές συναρτήσεις copula είναι οι ακόλουθες:

$$C_1(u, v) = u \cdot v,$$
$$C_2(u, v) = min(u, v),$$

και

$$C_3(u, v) = max(u + v - 1, 0),$$

για $u, v \in [0,1]$. Η πρώτη εξ' αυτών είναι, ίσως, η σημαντικότερη γιατί εκπροσωπεί την περίπτωση της ανεξαρτησίας. Πράγματι, αν $C_1(u, v) = u \cdot v$, τότε $F(x_1, x_2) = F_1(x_1)F_2(x_2)$, γεγονός που ισοδύναμα σημαίνει ότι οι τυχαίες μεταβλητές X_1, X_2 είναι ανεξάρτητες.

2.3. Φράγματα Fréchet-Hoeffding για τις συναρτήσεις copula

Όπως έχει ήδη αναφερθεί στο Κεφάλαιο 1, ο Hoeffding (1940, 1941) και έπειτα ο Fréchet (1951) ασχολήθηκαν με τη δημιουργία φραγμάτων για τις συναρτήσεις copula. Έτσι προς τιμήν του έργου και των δύο, δόθηκε στα φράγματα αυτά η ονομασία Fréchet-Hoeffding.

Ως συνέπεια του Θεωρήματος του Sklar, όταν F είναι η από κοινού α.σ.κ. δύο τυχαίων μεταβλητών X, Y με περιθώριες α.σ.κ. F_1 και F_2 αντιστοίχως, τότε για όλα τα $x, y \in [-\infty, +\infty]$ ισχύει ότι:

$$max\{F_1(x) + F_2(y) - 1, 0\} \leq F(x, y) \leq min\{F_1(x),\ F_2(y)\}.$$

Αυτή η ανισότητα είναι γνωστή ως ανισότητα των φραγμάτων Fréchet-Hoeffding. Τα φράγματα αυτά (βλέπε, π.χ., Drouet Mari and Kotz, 2001, σελ. 68) είναι τα ακόλουθα:

- Το κάτω φράγμα ορίζεται να είναι η α.σ.κ.
$$F^-(x, y) = max\{F_1(x) + F_2(y) - 1, 0\}.$$

- Το άνω φράγμα ορίζεται να είναι η α.σ.κ.
$$F^+(x, y) = min\{F_1(x), F_2(y)\}.$$

Ο Fréchet απέδειξε επίσης ότι οι τυχαίες μεταβλητές $\boldsymbol{X}, \boldsymbol{Y}$ **είναι ανεξάρτητες** αν και μόνο αν $F(x, y) = F_1(x)\, F_2(y)$. Ενώ αν, $F(x, y) = F^+(x, y)\ [F^-(x, y)]$, τότε υπάρχει μια αύξουσα [φθίνουσα] συνάρτηση που συνδέει τις τυχαίες μεταβλητές, δηλαδή οι τυχαίες μεταβλητές είναι θετικά [αρνητικά] εξαρτημένες. Με τον όρο θετικά εξαρτημένες εννοούμε ότι η Y είναι αύξουσα συνάρτηση της X, ενώ με τον όρο αρνητικά εξαρτημένες εννοούμε ότι η Y είναι φθίνουσα συνάρτηση της X (βλέπε, π.χ., Schweizer, 1991, Theorem 1.3 και αναφορές εκεί).

2.4. Βασικές ιδιότητες των συναρτήσεων copula

Στο εδάφιο αυτό θα παρουσιαστούν βασικές ιδιότητες των συναρτήσεων copula, υποθέτοντας ότι οι περιθώριες α.σ.κ. F_1, F_2 είναι συνεχείς.

Ιδιότητα 1.

Έστω F είναι η από κοινού α.σ.κ. δύο συνεχών τυχαίων μεταβλητών X, Y με περιθώριες α.σ.κ. F_1, F_2 αντιστοίχως και copula C. Χρησιμοποιώντας τον μετασχηματισμό $u = F_1(x)$ και $v = F_2(y)$ (Whitt, 1976) προκύπτει ότι η μοναδική συνάρτηση copula για $(u, v) \in [0, 1]^2$ έχει την μορφή:

$$C(u, v) = F\left(F_1^{-1}(u), F_2^{-1}(v)\right),$$

όπου F_1^{-1}, F_2^{-1} οι αντίστροφες των συναρτήσεων $F_1(x)$ και $F_2(y)$, αντίστοιχα.

Αυτή η ιδιότητα είναι σημαντική καθώς δίνει την δυνατότητα να κατασκευαστεί η συνάρτηση copula αν είναι γνώστες οι περιθώριες αλλά και η από κοινού κατανομή. Θα παρουσιαστεί ένα παράδειγμα στο οποίο χρησιμοποιείται η ιδιότητα αυτή για την εύρεση της συνάρτησης copula.

Παράδειγμα (Nelsen, 2006, σελ. 28): Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής τη διδιάστατη λογιστική κατανομή του Gumbel (1961) η οποία δίνεται από την σχέση

$$F(x,y) = (1 + e^{-x} + e^{-y})^{-1}, \qquad x, y \in [-\infty, +\infty].$$

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

$$\begin{aligned} F_1(x) &= \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} (1 + e^{-x} + e^{-y})^{-1} \\ &= (1 + e^{-x})^{-1}, x \in [-\infty, +\infty], \end{aligned}$$

και

$$\begin{aligned} F_2(y) &= \lim_{x\to\infty} F(x,y) = \lim_{x\to\infty} (1 + e^{-x} + e^{-y})^{-1} \\ &= (1 + e^{-y})^{-1}, y \in [-\infty, +\infty]. \end{aligned}$$

Οι $F_1(x)$ και $F_2(y)$ είναι γνησίως αύξουσες συναρτήσεις και έτσι για τις αντίστροφές τους έχουμε τα ακόλουθα.

Χρησιμοποιώντας το μετασχηματισμό $u = F_1(x)$ έχουμε

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

και επιλύοντας ως προς x προκύπτει

$$x = \log\left(\frac{u}{1-u}\right)$$

ή

$$x = \log(u) - \log(1-u).$$

Συνεπώς $F_1^{-1}(u) = \log(u) - \log(1-u)$, για κάθε $u \in [0, 1]$.

Ομοίως, θα χρησιμοποιήσουμε τον μετασχηματισμό $v = F_2(y)$ για την εύρεση της αντίστροφης της περιθώριας F_2. Έτσι,

$$v = (1 + e^{-y})^{-1}$$

και επιλύοντας ως προς y προκύπτει

$$y = \log\left(\frac{v}{1-v}\right)$$

ή

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

Συνεπώς $F_2^{-1}(v) = \log(v) - \log(1-v)$, για κάθε $v \in [0,1]$.
Επίσης, τα πεδία τιμών των F_1 και F_2 είναι $R(F_1) = R(F_2) = [0,1]$ και επομένως το πεδίο ορισμού της copula C είναι $DomC = R(F_1) \times R(F_2) = [0,1]^2$.
Έτσι, σύμφωνα με αυτή την ιδιότητα έχουμε ότι η μοναδική συνάρτηση copula, για $(u,v) \in [0,1]^2$, έχει την μορφή:

$$\begin{aligned}
C(u,v) &= F\left(F_1^{-1}(u), F_2^{-1}(v)\right) \\
&= F(\log(u) - \log(1-u), \log(v) - \log(1-v)) \\
&= \left(1 + e^{-\log(u)+\log(1-u)} + e^{-\log(v)+\log(1-v)}\right)^{-1} \\
&= \left(1 + e^{\log\left(\frac{1-u}{u}\right)} + e^{\log\left(\frac{1-v}{v}\right)}\right)^{-1} \\
&= \left(\frac{v+u-uv}{uv}\right)^{-1} \\
&= \frac{uv}{v+u-uv}.
\end{aligned}$$

Κατασκευάστηκε η συνάρτηση copula $C(u,v) = uv/(v+u-uv)$, για $(u,v) \in [0,1]^2$. Η συγκεκριμένη συνάρτηση copula ανήκει στην κατηγορία των Archimedean Copulas, η οποία θα αποτελέσει το αντικείμενο του 4^{ου} κεφαλαίου.

Ιδιότητα 2.

Η συνάρτηση copula παραμένει αμετάβλητη κάτω από γνήσια αύξοντες μετασχηματισμούς (Schweizer and Wolf, 1981 και Nelsen, 2006, Theorem 2.4.3, σελ. 25). Δηλαδή, έστω X και Y τυχαίες μεταβλητές με copula $C(u,v)$. Αν φ, ψ είναι γνήσια αύξουσες συναρτήσεις στο πεδίο τιμών των X, Y τότε:

$$C\left(\varphi(u), \psi(v)\right) = C(u,v).$$

Αυτό σημαίνει ότι αν C είναι η συνάρτηση copula που μοντελοποιεί την από κοινού κατανομή των τυχαίων μεταβλητών Y_1 και Y_2, τότε η ίδια συνάρτηση copula μπορεί να χρησιμοποιηθεί για την από κοινού κατανομή των $(\ln Y_1, \ln Y_2)$.

Ιδιότητα 3.

Για να συνδέσουμε την θεωρία του Fréchet με τις διδιάστατες συναρτήσεις copula αρκεί να θεωρήσουμε ότι $F_1(x) = u$ και $F_2(y) = v$. Έτσι θα έχουμε:

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

και για κάθε copula C θα ισχύει η ακόλουθη ανισότητα:

$$max\{u + v - 1, 0\} \leq C(u, v) \leq min(u, v), \qquad (u, v) \in [0, 1]^2.$$

Τα $max\{u + v - 1, 0\}$ και $min(u, v)$ είναι συναρτήσεις copula και η C^0 εκπροσωπεί την περίπτωση της ανεξαρτησίας.

Ιδιότητα 4.

Οι συναρτήσεις copula αναγνωρίζουν την ανεξαρτησία μεταξύ δύο τυχαίων μεταβλητών. Αυτό συμβαίνει, πράγματι, καθώς αν υποθέσουμε ότι $C(u, v) = uv$, τότε σύμφωνα με την θεωρία του Fréchet οι τυχαίες μεταβλητές X, Y είναι ανεξάρτητες αν και μόνο αν $F(x, y) = F_1(x)F_2(y)$.
Γνωρίζουμε ότι $C(u, v) = F\left(F_1^{-1}(u), F_2^{-1}(v)\right)$, όπου $u = F_1(x)$, $v = F_2(y)$ και F η από κοινού α.σ.κ.. Συνεπώς κάτω από την υπόθεση $C(u, v) = uv$ θα έχουμε:

$$uv = F\left(F_1^{-1}(u), F_2^{-1}(v)\right) \text{ ή } F_1(x)F_2(y) = F(x, y).$$

Ιδιότητα 5.

Αν $C(u, v) = C^+(u, v)$ $[C^-(u, v)]$ τότε οι τυχαίες μεταβλητές X, Y είναι θετικά [αρνητικά] εξαρτημένες.

Στη συνέχεια το ενδιαφέρον θα εστιαστεί στην copula density. Αρχικά θα εισαχθεί η έννοια της copula density. Επιπλέον θα παρουσιαστεί η σχέση της με την από κοινού πυκνότητα αλλά και με τις περιθώριες πυκνότητες.

Πυκνότητα Copula (Copula density)

Ορισμός 2.4.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με από κοινού κατανομή F, περιθώριες F_1 και F_2 και αντίστοιχη συνάρτηση copula C. Αν οι συναρτήσεις αυτές είναι δύο φορές παραγωγίσιμες, τότε η πυκνότητα copula ορίζεται

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}.$$

Με βάση τον ορισμό αυτό είναι εύκολο να προσδιορισθεί η σχέση μεταξύ της από κοινού κατανομής των συνεχών τυχαίων μεταβλητών X και Y, των αντίστοιχων περιθώριων πυκνοτήτων και της πυκνότητας copula. Η σχέση αυτή διατυπώνεται στην πρόταση που ακολουθεί.

Πρόταση 2.4.1. Αν X, Y συνεχείς τυχαίες μεταβλητές με από κοινού πυκνότητα $f(x, y)$ και περιθώριες f_1 και f_2, τότε για την πυκνότητα copula c ισχύει

$$f(x, y) = f_1(x) f_2(y) c\big(F_1(x), F_2(y)\big).$$

Απόδειξη. Από την Ιδιότητα 1 των συναρτήσεων copula C ισχύει

$$C(x, y) = F\big(F_1^{-1}(x), F_2^{-1}(y)\big), \qquad x, y \in [0, 1].$$

Για την απόδειξη χρειάζεται να υπολογιστεί η ποσότητα

$$\frac{\partial^2}{\partial x \partial y} C(x, y) = \frac{\partial^2}{\partial x \partial y} F\big(F_1^{-1}(x), F_2^{-1}(y)\big).$$

Σύμφωνα με τον κανόνα της αλυσίδας και λαμβάνοντας υπόψη ότι

$$\frac{\partial}{\partial x} F_1^{-1}(x) = \frac{1}{F_1'\big(F_1^{-1}(x)\big)} = \frac{1}{f_1\big(F_1^{-1}(x)\big)}$$

και

$$\frac{\partial}{\partial y}F_2^{-1}(y) = \frac{1}{f_2\left(F_2^{-1}(y)\right)},$$

και συμβολίζοντας με F την $F\left(F_1^{-1}(x), F_2^{-1}(y)\right)$ έχουμε ότι

$$\begin{aligned}\frac{\partial}{\partial y}F\left(F_1^{-1}(x), F_2^{-1}(y)\right) &= \frac{\partial F}{\partial F_1^{-1}(x)}\frac{\partial F_1^{-1}(x)}{\partial y} + \frac{\partial F}{\partial F_2^{-1}(y)}\frac{\partial F_2^{-1}(y)}{\partial y} \\ &= \frac{\partial F}{\partial F_2^{-1}(y)}\frac{1}{f_2\left(F_2^{-1}(y)\right)}\end{aligned}$$

και

$$\begin{aligned}\frac{\partial^2}{\partial x \partial y}F\left(F_1^{-1}(x), F_2^{-1}(y)\right) &= \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial F_2^{-1}(y)}\frac{1}{f_2\left(F_2^{-1}(y)\right)}\right) \\ &= \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial F_2^{-1}(y)}\right)\frac{1}{f_2\left(F_2^{-1}(y)\right)}.\end{aligned}$$

Όμως,

$$\begin{aligned}\frac{\partial}{\partial x}F\left(F_1^{-1}(x), F_2^{-1}(y)\right) &= \frac{\partial F}{\partial F_1^{-1}(x)}\frac{\partial F_1^{-1}(x)}{\partial x} + \frac{\partial F}{\partial F_2^{-1}(y)}\frac{\partial F_2^{-1}(y)}{\partial x} \\ &= \frac{\partial F}{\partial F_1^{-1}(x)}\frac{1}{f_1\left(F_1^{-1}(x)\right)}.\end{aligned}$$

Επομένως,

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial F_2^{-1}(y)}\right) = \frac{\partial}{\partial F_1^{-1}(x)}\left(\frac{\partial F}{\partial F_2^{-1}(y)}\right)\frac{1}{f_1\left(F_1^{-1}(x)\right)}.$$

Τελικά έχουμε ότι,

$$\frac{\partial^2}{\partial x \partial y}C(x,y) = \frac{\partial}{\partial F_1^{-1}(x)}\left(\frac{\partial F}{\partial F_2^{-1}(y)}\right)\frac{1}{f_1\left(F_1^{-1}(x)\right)}\frac{1}{f_2\left(F_2^{-1}(y)\right)}$$

$$c(x,y) = \frac{\partial}{\partial F_1^{-1}(x)}\left(\frac{\partial F}{\partial F_2^{-1}(y)}\right)\frac{1}{f_1\left(F_1^{-1}(x)\right)}\frac{1}{f_2\left(F_2^{-1}(y)\right)}$$

και χρησιμοποιώντας τους μετασχηματισμούς $F_1^{-1}(x) = x$ και $F_2^{-1}(y) = y$ προκύπτει ότι

$$c\big(F_1(x), F_2(y)\big) = \frac{\partial}{\partial x}\left(\frac{\partial F(x,y)}{\partial y}\right)\frac{1}{f_1(x)}\frac{1}{f_2(y)} = \frac{f(x,y)}{f_1(x)f_2(y)},$$

δηλαδή αποδείχτηκε το ζητούμενο. ■

2.5. Copulas επιβίωσης (Survival Copulas)

Στο εδάφιο αυτό θα αναφερθούμε στις συναρτήσεις copula οι οποίες ορίζονται μέσω της συνάρτησης επιβίωσης και οι οποίες θα ονομάζονται, στη διατριβή αυτή, copulas επιβίωσης (survival copulas). Στο εδάφιο, αυτό, βασιζόμαστε στο Nelsen (2006, σελ. 32). Οι copulas επιβίωσης, αφού βασίζονται στη συνάρτηση επιβίωσης, ορίζονται στη πράξη για τυχαίες μεταβλητές οι οποίες αντιπροσωπεύουν το χρόνο ζωής ατόμων ή αντικειμένων σε κάποιο πληθυσμό. Στρέφουμε λοιπόν το ενδιαφέρον σε τυχαίες μεταβλητές με τιμές στο διάστημα $[0,+\infty)$. Στο πλαίσιο αυτό, η πιθανότητα να επιβιώσει μια μονάδα πέραν κάποιου χρονικού σημείου x, εκφράζεται από την παρακάτω συνάρτηση επιβίωσης:

$$\bar{F}(x) = P(X > x) = 1 - F(x), \qquad x \geq 0,$$

όπου F είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής X.

Για τις ανάγκες του εδαφίου αυτού θα θεωρηθεί ότι οι χρόνοι ζωής, που μοντελοποιούνται από τις τυχαίες μεταβλητές X και Y, ορίζονται στην πραγματική ευθεία. Έτσι, για ένα ζεύγος τυχαίων μεταβλητών X, Y με από κοινού συνάρτηση κατανομής F, η από κοινού συνάρτηση επιβίωσης ορίζεται ως:

$$\bar{F}(x,y) = P(X > x, Y > y), \text{ για } x, y \in [-\infty, +\infty]$$

και εκφράζει την πιθανότητα και οι δύο μονάδες να έχουν επιβιώσει πέραν κάποιων χρονικών σημείων x, y, αντίστοιχα. Οι περιθώριες της $\bar{F}$ είναι οι εξής:

$$\bar{F}_1(x) = P(X > x) = 1 - F_1(x) = 1 - F(x, +\infty) = \bar{F}(x, +\infty)$$

και

$$\bar{F}_2(y) = P(Y > y) = 1 - F_2(y) = 1 - F(+\infty, y) = \bar{F}(+\infty, y).$$

Ας υποθέσουμε τώρα ότι C είναι η copula των X, Y. Τότε έχουμε:

$$\begin{aligned}\overline{F}(x,y) &= 1 - F_1(x) - F_2(y) + F(x,y)\\ &= \bar{F}_1(x) + \bar{F}_2(y) - 1 + C\big(F_1(x), F_2(y)\big)\\ &= \bar{F}_1(x) + \bar{F}_2(y) - 1 + C\big(1 - \bar{F}_1(x), 1 - \bar{F}_2(y)\big).\end{aligned}$$

Έτσι, αν ορίσουμε την συνάρτηση $\widehat{C}$: $[0,1] \times [0,1] \to [0,1]$ με τύπο:

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

θα έχουμε:

$$\bar{F}(x,y) = \hat{C}\big(\bar{F}_1(x), \bar{F}_2(y)\big).$$

Σημειώνεται ότι η συνάρτηση $\hat{C}$ είναι μία copula και συγκεκριμένα είναι η copula επιβίωσης των X, Y, η οποία συνδέει την από κοινού συνάρτηση επιβίωσης $\bar{F}$ με τις περιθώριες συναρτήσεις επιβίωσης.

Παρατήρηση 2.5.1. Η copula επιβίωσης $\hat{C}$ δεν είναι η από κοινού συνάρτηση επιβίωσης $\bar{C}$ δύο τυχαίων μεταβλητών που κατανέμονται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής τη copula C. Αυτό συμβαίνει καθώς άμεσα αποδεικνύεται ότι ισχύει η παρακάτω σχέση:

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

2.6. Οικογένειες Συναρτήσεων Copula

Έχει προταθεί ένας μεγάλος αριθμός ειδικών μορφών συναρτήσεων copula στη βιβλιογραφία και κάθε μία από αυτές επιβάλλει μια διαφορετική δομή εξάρτησης στα δεδομένα. Οι Hutchinson και Lai (1990), Joe (1997, κεφ. 5), Nelsen (2006, 116-119) και Joe (2015) παρέχουν μία πλήρη κάλυψη των διδιάστατων συναρτήσεων copula και των ιδιοτήτων τους. Στη συνέχεια θα αναφέρουμε κάποιες από αυτές τις συναρτήσεις copula χωρίς να γίνεται λεπτομερής περιγραφή τους.

2.6.1. Product copula

Η Product copula (ή αλλιώς copula ανεξαρτησίας) έχει ήδη αναφερθεί και αποτελεί την απλούστερη συνάρτηση copula. Έχει την ακόλουθη μορφή:

$$C(u,v) = uv, \qquad u,v \in [0,1]$$

δηλαδή, συνδέεται με την ανεξαρτησία δύο τυχαίων μεταβλητών.

Η πυκνότητα της συνάρτησης copula είναι η ποσότητα

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} = \frac{\partial}{\partial u}\left(\frac{\partial C(u,v)}{\partial v}\right),$$

και για $C(u,v) = uv$ άμεσα προκύπτει η πυκνότητα copula

$$c(u,v) = 1, \qquad u,v \in [0,1].$$

2.6.2. Farlie-Gumbel-Morgenstern copula

Μία Farlie–Gumbel–Morgenstern (FGM) copula παίρνει την μορφή:

$$C(u,v,\theta) = uv\big(1 + \theta(1-u)(1-v)\big), \qquad \theta \in [-1,+1], (u,v) \in [0,1]^2.$$

Η FGM copula προτάθηκε, ανεξάρτητα, από τους Farlie (1960), Gumbel (1958) και Morgentern (1956). Αν η παράμετρος εξάρτησης θ ισούται με το μηδέν τότε το FGM copula ισούται με uv, δηλαδή οδηγεί στην ανεξαρτησία.

Η πυκνότητα της συνάρτησης copula είναι η ποσότητα

$$c(u,v,\theta) = \frac{\partial^2 C(u,v,\theta)}{\partial u \partial v} = \frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right).$$

Έχουμε ότι

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

Άρα

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

και

$$\frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right) = 1+\theta(1-v-u+uv)+u\theta(-1+v)+v\theta(-1+u)$$
$$+uv\theta$$
$$= 1+\theta-\theta v-\theta u+\theta uv-u\theta+u\theta v-v\theta+v\theta u+uv\theta$$
$$= 1+\theta-2v\theta-2u\theta+4uv\theta$$
$$= 1+\theta(1-2u)(1-2v).$$

Συνεπώς η FGM πυκνότητα copula είναι

$$c(u,v,\theta) = 1+\theta(1-2u)(1-2v), \qquad \theta \in [-1,+1], (u,v) \in [0,1]^2.$$

2.6.3. Gaussian (Normal) copula

Μία Gaussian Copula ορίζεται ως:

$$C(u,v,\theta) = \Phi_G(\Phi^{-1}(u), \Phi^{-1}(v), \theta)$$

$$= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}}\, e^{\frac{-\left(x^2-2\theta xy+y^2\right)}{2(1-\theta^2)}} dxdy,$$

όπου $\theta \in (-1,1)$, με $(u,v) \in [0,1]^2$.

Στην παραπάνω σχέση η Φ συμβολίζει τη συνάρτηση κατανομής της τυπικής κανονικής κατανομής και Φ_G την αντίστοιχη συνάρτηση της διδιάστατης κανονικής με μέσες τιμές 0, διακυμάνσεις 1 και συντελεστή συσχέτισης $\theta, -1 \leq \theta \leq 1$. Αυτή η συνάρτηση copula προτάθηκε από τον Lee (1983).

Με βάση την Πρόταση 2.4.1. η πυκνότητα copula c δίνεται από τη σχέση

$$c(u,v) = \frac{f\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{f_1\left(\Phi^{-1}(u)\right) f_2\left(\Phi^{-1}(v)\right)},$$

όπου f η από κοινού πυκνότητα της $N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \theta\\ \theta & 1\end{pmatrix}\right)$ κατανομής και f_1, f_2 η πυκνότητα της τυπικής κανονικής $N(0,1)$ κατανομής.

Για $\Phi^{-1}(u) = x$ και $\Phi^{-1}(v) = y$, έχουμε τα ακόλουθα

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in R$$

$$f_2(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}, y \in R$$

και

$$f(x,y) = \frac{1}{2\pi|\boldsymbol{\Theta}|^{1/2}}\ e^{-\frac{1}{2}\boldsymbol{X}^T\boldsymbol{\Theta}^{-1}\boldsymbol{X}},$$

όπου $\boldsymbol{X} = (X,Y)^T$ και $\boldsymbol{\Theta} = \begin{bmatrix}1 & \theta\\ \theta & 1\end{bmatrix}$ ο πίνακας συσχετίσεων.

Έτσι, έχουμε αρχικά ότι

$$\boldsymbol{\Theta}^{-1} = \frac{1}{det\Theta}\begin{pmatrix}1 & -\theta\\ -\theta & 1\end{pmatrix} = \begin{pmatrix}\frac{1}{1-\theta^2} & -\frac{\theta}{1-\theta^2}\\ -\frac{\theta}{1-\theta^2} & \frac{1}{1-\theta^2}\end{pmatrix}.$$

Επομένως, η τετραγωνική μορφή $\boldsymbol{X}^t\boldsymbol{\Theta}^{-1}\boldsymbol{X}$ γίνεται

$$\boldsymbol{X}^t\boldsymbol{\Theta}^{-1}\boldsymbol{X} = (x\ y)\begin{pmatrix}\frac{1}{1-\theta^2} & -\frac{\theta}{1-\theta^2}\\ -\frac{\theta}{1-\theta^2} & \frac{1}{1-\theta^2}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

$$= \frac{x^2}{(1-\theta^2)} + \frac{-\theta xy}{(1-\theta^2)} + \frac{-\theta xy}{(1-\theta^2)} + \frac{y^2}{(1-\theta^2)}$$

$$= \frac{x^2 - 2\theta xy + y^2}{(1-\theta^2)}.$$

Συνεπώς,

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\theta^2}}\ e^{-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}}$$

και έτσι η πυκνότητα copula c γίνεται

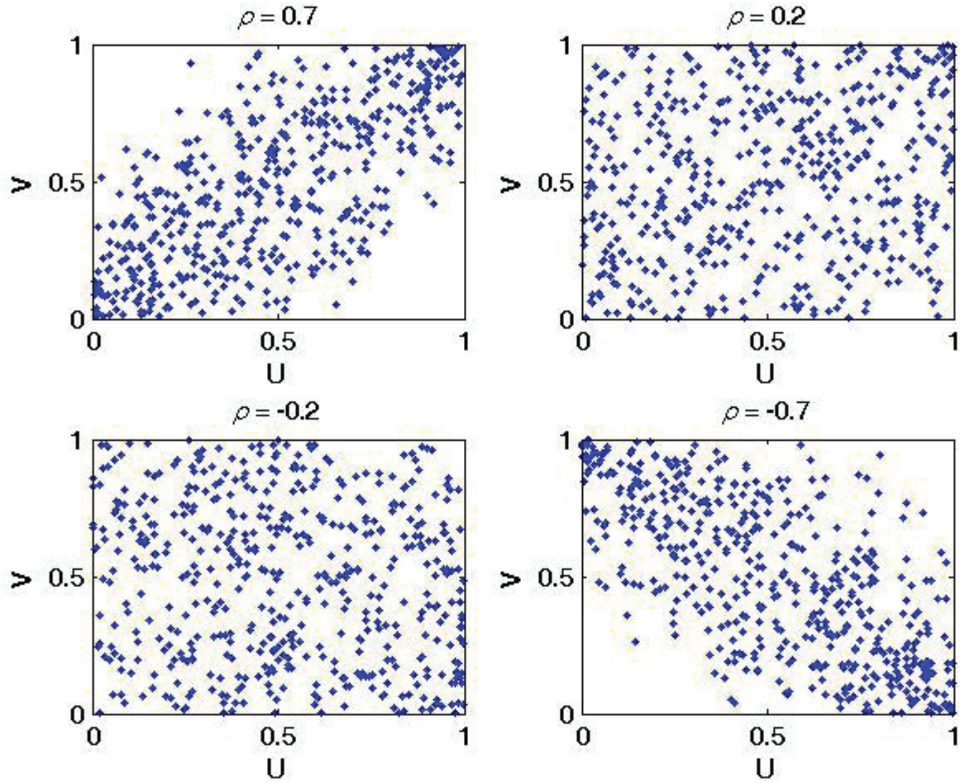
$$c(u,v)=\frac{\dfrac{1}{2\pi\sqrt{1-\theta^2}}\; e^{-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}}}{\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\;\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}}\;.$$

Μετά από στοιχειώδης αλγεβρικές πράξεις προκύπτει ότι η πυκνότητα copula c παίρνει την ακόλουθη μορφή

$$c(u,v)=\frac{1}{\sqrt{1-\theta^2}}\; e^{\frac{-(x^2+y^2)\theta^2+2\theta xy}{2(1-\theta^2)}}\;,\;\; \theta\in(-1,1),$$

για $x=\Phi^{-1}(u)$ και $y=\Phi^{-1}(v)$.

Στη συνέχεια, θα δοθούν τέσσερα διαγράμματα, τα οποία κατασκευάστηκαν χρησιμοποιώντας τις δοθείσες συναρτήσεις της Matlab (βλέπε, https://www.mathworks.com/help/stats/copulas-generate-correlated-samples.html, προσπελάστηκε 15/5/2018). Στο παρακάτω σχήμα έχουν προσομοιωθεί 500 τυχαία σημεία από τη Gaussian copula, με διαφορετικούς συντελεστές συσχέτισης.



Σχήμα 2.1. 500 σημεία από τη Gaussian copula για $\rho = \theta = 0.7$, $\rho = \theta = 0.2$, $\rho = \theta = -0.2$ και $\rho = \theta = -0.7$.

Αυτό που παρατηρούμε είναι πως όταν ο συντελεστής συσχέτισης παίρνει τιμή κοντά στο μηδέν τα σημεία είναι διάσπαρτα σε αυτά τα διαγράμματα (βλέπε, τα αντίστοιχα διαγράμματα στο Σχήμα 2.1. για $\rho = \theta = 0.2$ και $\rho = \theta = -0.2$). Στο πρώτο και στο τέταρτο διάγραμμα, έχουμε θετική και αρνητική γραμμική συσχέτιση, αντίστοιχα (βλέπε, τα αντίστοιχα διαγράμματα στο Σχήμα 2.1. για $\rho = \theta = 0.7$ και $\rho = \theta = -0.7$).

2.6.4. Student's t-copula

Κατά πλήρη αναλογία με την περίπτωση της Gaussian (Normal) Copula, η Student's t copula παίρνει την μορφή:

$$C(u,v,\theta)=\int_{-\infty}^{t_n^{-1}(u)}\int_{-\infty}^{t_n^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}$$

$$\times\left\{1+\frac{(x^2-2\theta xy+y^2)}{n(1-\theta^2)}\right\}^{-\frac{n+2}{2}}\ dxdy\,,$$

όπου $t_n^{-1}(u)$ δηλώνει την αντίστροφη της αθροιστικής συνάρτησης της τυπικής μονοδιάστατης t-κατανομής, με n βαθμούς ελευθερίας και παράμετρο εξάρτησης θ.

Με βάση την Πρόταση 2.4.1. η πυκνότητα copula c δίνεται από τη σχέση

$$c(u,v)=\frac{f\left(t_n^{-1}(u),t_n^{-1}(v)\right)}{f_1\left(t_n^{-1}(u)\right)f_2\left(t_n^{-1}(v)\right)}.$$

Για $t_n^{-1}(u)=x$ και $t_n^{-1}(v)=y$, έχουμε τα ακόλουθα

$$f_1(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

$$f_2(y)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}}$$

και η από κοινού πυκνότητα της διδιάστατης t-κατανομής είναι η

$$f(x,y)=\frac{1}{2\pi\sqrt{1-\theta^2}}\left[1+\frac{1}{n}\boldsymbol{X}^T\boldsymbol{\Theta}^{-1}\boldsymbol{X}\right]^{-\frac{n+2}{2}},$$

με τυχαίο διάνυσμα $\boldsymbol{X} = (X, Y)^T$ και $\boldsymbol{\Theta} = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$ τον πίνακα συσχετίσεων. Οι ποσότητες $\boldsymbol{\Theta}^{-1}$ και $\boldsymbol{X}^t \boldsymbol{\Theta}^{-1} \boldsymbol{X}$ υπολογίστηκαν στην ακριβώς προηγούμενη συνάρτηση copula και είναι ίσες με

$$\boldsymbol{\Theta}^{-1} = \begin{pmatrix} \frac{1}{1-\theta^2} & -\frac{\theta}{1-\theta^2} \\ -\frac{\theta}{1-\theta^2} & \frac{1}{1-\theta^2} \end{pmatrix}$$

και

$$\boldsymbol{X}^t \boldsymbol{\Theta}^{-1} \boldsymbol{X} = \frac{x^2 - 2\theta xy + y^2}{(1-\theta^2)}.$$

Δηλαδή,

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\theta^2}} \left[1 + \frac{1}{n} \frac{x^2 - 2\theta xy + y^2}{(1-\theta^2)}\right]^{-\frac{n+2}{2}}.$$

Συνεπώς η πυκνότητα της συνάρτησης copula είναι η

$$c(u,v) = \frac{\frac{1}{2\pi\sqrt{1-\theta^2}} \left[1 + \frac{1}{n} \frac{x^2 - 2\theta xy + y^2}{(1-\theta^2)}\right]^{-\frac{n+2}{2}}}{\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}}$$

και μετά από στοιχειώδης αλγεβρικές πράξεις, δεδομένης της σχέσης

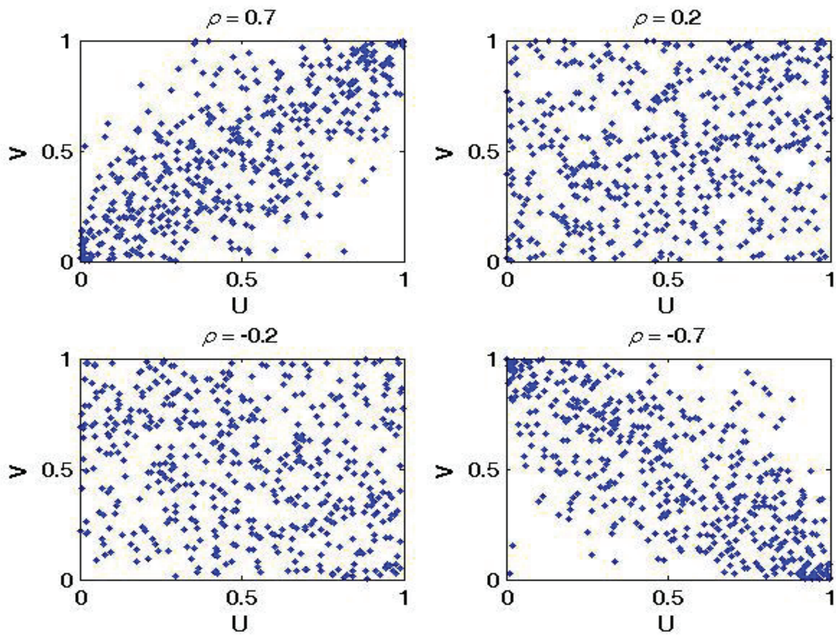
$$\frac{1}{2\pi} = \frac{\Gamma\left(\frac{n+2}{2}\right)}{\pi n \Gamma\left(\frac{n}{2}\right)},$$

προκύπτει ότι η πυκνότητα copula c παίρνει την ακόλουθη μορφή

$$c(u,v) = \frac{\Gamma\left(\frac{n+2}{2}\right) \Gamma\left(\frac{n}{2}\right) \left[1 + \frac{1}{n} \frac{x^2 - 2\theta xy + y^2}{(1-\theta^2)}\right]^{-\frac{n+2}{2}}}{\sqrt{1-\theta^2} \left(\Gamma\left(\frac{n+1}{2}\right)\right)^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}},$$

$\theta \in (-1,1)$, για $x = t_n^{-1}(u)$ και $y = t_n^{-1}(v)$, $(u, v) \in [0, 1]^2$.

Στη συνέχεια, θα δοθούν τέσσερα διαγράμματα, τα οποία κατασκευάστηκαν χρησιμοποιώντας τις δοθείσες συναρτήσεις της Matlab (βλέπε, https://www.mathworks.com/help/stats/copulas-generate-correlated-samples.html, προσπελάστηκε 15/5/2018). Στο παρακάτω σχήμα έχουν προσομοιωθεί 500 τυχαία σημεία από τη t copula, με διαφορετικούς συντελεστές συσχέτισης και n =7 βαθμούς ελευθερίας.



Σχήμα 2.2. 500 σημεία από τη t copula, για $\rho = \theta = 0.7$, $\rho = \theta$ =0.2, $\rho = \theta = -0.2$, $\rho = \theta = -0.7$ και n =7.

Παρατηρείται ότι, όπως και στην περίπτωση της Gaussian copula, έτσι και εδώ όταν ο συντελεστής συσχέτισης παίρνει τιμή κοντά στο μηδέν τα σημεία είναι διάσπαρτα σε αυτά τα διαγράμματα (βλέπε, τα αντίστοιχα διαγράμματα στο Σχήμα 2.2. για $\rho = \theta = 0.2$ και $\rho = \theta = -0.2$). Στο πρώτο και στο τέταρτο διάγραμμα, έχουμε θετική και αρνητική γραμμική συσχέτιση, αντίστοιχα (βλέπε, τα αντίστοιχα διαγράμματα στο Σχήμα 2.2. για $\rho = \theta = 0.7$ και $\rho = \theta = -0.7$).

Σημειώνεται, ότι, τα παραπάνω διαγράμματα που κατασκευάστηκαν για τις Gaussian copula και t copula ήταν ενδεικτικά για να δοθούν και οπτικά κάποια αποτελέσματα για την παράμετρο εξάρτησης θ.

2.6.5. Clayton copula

Η Clayton (1978) copula παίρνει την μορφή:

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

και για $\theta > 0$ είναι της μορφής

$$C(u,v,\theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \qquad \theta \in (0,+\infty).$$

Η Clayton copula, που επίσης αναφέρεται ως copula Cook and Johnson (1981, 1986), αρχικά μελετήθηκε από τους Kimeldorf and Sampson (1975).

Η πυκνότητα της συνάρτησης copula είναι η ποσότητα

$$c(u,v,\theta) = \frac{\partial^2 C(u,v,\theta)}{\partial u \partial v} = \frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right).$$

Έχουμε ότι για $\theta \in (0,+\infty)$

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

Άρα

$$\begin{aligned}\frac{\partial C(u,v,\theta)}{\partial v} &= -\frac{1}{\theta}\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-1}\left(-\theta v^{-\theta-1}\right) \\ &= \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-1}\left(v^{-\theta-1}\right)\end{aligned}$$

και

$$\begin{aligned}\frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right) &= \left(-\frac{1}{\theta} - 1\right)\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-2}\left(-\theta u^{-\theta-1}\right)\left(v^{-\theta-1}\right) \\ &= \left(-\frac{1+\theta}{\theta}\right)\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1+2\theta}{\theta}}\left(-\theta u^{-\theta-1}\right)\left(v^{-\theta-1}\right)\end{aligned}$$

$$= (1+\theta)\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1+2\theta}{\theta}}(uv)^{-(\theta+1)}.$$

Συνεπώς η πυκνότητα copula είναι η

$$c(u,v,\theta) = (1+\theta)\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1+2\theta}{\theta}}(uv)^{-(\theta+1)}, \qquad \theta \in (0,+\infty).$$

2.6.6. Frank copula

Η Frank (1979) copula παίρνει την μορφή:

$$C(u,v,\theta) = -\theta^{-1}\log\left\{1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right\},$$

με $\theta \in (-\infty,+\infty)-\{0\}$.

Οι τιμές $-\infty,+\infty$ της παραμέτρου εξάρτησης θ αντιστοιχούν στο κάτω φράγμα και στο άνω φράγμα του Fréchet αντίστοιχα.

Η πυκνότητα copula είναι η ποσότητα

$$c(u,v,\theta) = \frac{\partial^2 C(u,v,\theta)}{\partial u \partial v} = \frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right).$$

Έχουμε ότι

$$C(u,v,\theta) = -\theta^{-1}\log\left\{1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right\}, \theta \in (-\infty,+\infty)-\{0\}.$$

Έστω $a(u) = e^{-\theta u}-1,\ a(v) = e^{-\theta v}-1,\ a(1) = e^{-\theta}-1$, έτσι έχουμε,

$$C(u,v,\theta) = -\theta^{-1}\log\left\{1+\frac{a(u)a(v)}{a(1)}\right\}, \qquad \theta \in (-\infty,+\infty)-\{0\}.$$

Άρα,

$$\frac{\partial C(u,v,\theta)}{\partial v} = -\theta^{-1}\frac{1}{1+\frac{a(u)a(v)}{a(1)}}\ \frac{1}{a(1)}a(u)\left(-\theta e^{-\theta v}\right)$$

$$= -\theta^{-1}\frac{1}{\frac{a(1)+a(u)a(v)}{a(1)}}\ \frac{1}{a(1)}a(u)\left(-\theta e^{-\theta v}\right)$$

$$= -\theta^{-1} \frac{a(1)}{a(1) + a(u)a(v)} \frac{a(u)\left(-\theta e^{-\theta v}\right)}{a(1)}$$

$$= -\frac{1}{\theta} \frac{a(u)\left(-\theta e^{-\theta v}\right)}{a(1) + a(u)a(v)}$$

$$= \frac{a(u)\left(e^{-\theta v}\right)}{a(1) + a(u)a(v)}$$

και

$$\frac{\partial}{\partial u}\left(\frac{\partial C(u, v, \theta)}{\partial v}\right)$$

$$= \frac{\left(-\theta e^{-\theta u}\right)\left(e^{-\theta v}\right)[a(1) + a(u)a(v)] + a(u)\left(e^{-\theta v}\right)\left(\theta e^{-\theta u}\right)a(v)}{[a(1) + a(u)a(v)]^2}$$

$$= \frac{-\theta e^{-\theta u - \theta v}[a(1) + a(u)a(v)] + \left(\theta e^{-\theta u - \theta v}\right)a(u)a(v)}{[a(1) + a(u)a(v)]^2}$$

$$= \frac{-\theta e^{-\theta u - \theta v}a(1) - \theta e^{-\theta u - \theta v}a(u)a(v) + \left(\theta e^{-\theta u - \theta v}\right)a(u)a(v)}{[a(1) + a(u)a(v)]^2}$$

$$= \frac{-\theta e^{-\theta u - \theta v}a(1)}{[a(1) + a(u)a(v)]^2}.$$

Συνεπώς η πυκνότητα copula είναι η

$$c(u, v, \theta) = \frac{-\theta e^{-\theta u - \theta v}\left(e^{-\theta} - 1\right)}{[e^{-\theta} - 1 + (e^{-\theta u} - 1)(e^{-\theta v} - 1)]^2}, \quad \theta \in (-\infty, +\infty) - \{0\},$$

$$(u, v) \in [0, 1]^2.$$

2.6.7. Gumbel copula

Η Gumbel (1960) copula παίρνει την μορφή:

$$C(u,v,\theta) = e^{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}}}, \qquad \theta \in [1,+\infty).$$

Οι τιμές $1,+\infty$ της παραμέτρου εξάρτησης θ αντιστοιχούν στην ανεξαρτησία και το άνω φράγμα του Fréchet αντίστοιχα.

Η πυκνότητα της συνάρτησης copula είναι η ποσότητα

$$c(u,v,\theta) = \frac{\partial^2 C(u,v,\theta)}{\partial u \partial v} = \frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right).$$

Έχουμε ότι

$$C(u,v,\theta) = e^{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}}}, \qquad \theta \in [1,+\infty).$$

Θέτω $(-\ln u)^{\theta} = b(u)$ και $(-\ln v)^{\theta} = b(v)$, έτσι

$$C(u,v,\theta) = e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}, \qquad \theta \in [1,+\infty).$$

Άρα,

$$\begin{aligned}\frac{\partial C(u,v,\theta)}{\partial v} &= e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\left\{-\frac{1}{\theta}[b(u)+b(v)]^{\frac{1}{\theta}-1}\theta(-\ln v)^{\theta-1}\left(-\frac{1}{v}\right)\right\} \\ &= e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\left\{\frac{1}{v}[b(u)+b(v)]^{\frac{1-\theta}{\theta}}(-\ln v)^{\theta-1}\right\}\end{aligned}$$

και

$$\begin{aligned}\frac{\partial}{\partial u}\left(\frac{\partial C(u,v,\theta)}{\partial v}\right) \\ &= e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\left\{\frac{1}{u}[b(u)+b(v)]^{\frac{1}{\theta}-1}(-\ln u)^{\theta-1}\right\}\left\{\frac{1}{v}[b(u)\right. \\ &\quad \left.+ b(v)]^{\frac{1}{\theta}-1}(-\ln v)^{\theta-1}\right\} \\ &\quad + e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\left\{\frac{1}{v}(-\ln v)^{\theta-1}\frac{1-\theta}{\theta}[b(u)\right. \\ &\quad \left.+ b(v)]^{\frac{1-\theta}{\theta}-1}\theta(-\ln u)^{\theta-1}\left(-\frac{1}{u}\right)\right\}\end{aligned}$$

$$= e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\left\{\frac{1}{u}[b(u)+b(v)]^{\frac{1-\theta}{\theta}}(-\ln u)^{\theta-1}\right\}\left\{\frac{1}{v}[b(u)+b(v)]^{\frac{1-\theta}{\theta}}(-\ln v)^{\theta-1}\right\}$$
$$+ e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\left\{\frac{1}{uv}(-\ln v)^{\theta-1}(\theta-1)[b(u)+b(v)]^{\frac{1-2\theta}{\theta}}(-\ln u)^{\theta-1}\right\}$$

$$= e^{-[b(u)+b(v)]^{\frac{1}{\theta}}}\ \frac{1}{uv}[b(u)+b(v)]^{\frac{2-2\theta}{\theta}}[(-\ln u)(-\ln v)]^{\theta-1}\left\{1+(\theta-1)[b(u)+b(v)]^{-\frac{1}{\theta}}\right\}.$$

Συνεπώς η πυκνότητα copula είναι η

$$c(u,v,\theta) = e^{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}}}\ \frac{1}{uv}\left[(-\ln u)^{\theta}+(-\ln v)^{\theta}\right]^{\frac{2-2\theta}{\theta}}[(-\ln u)(-\ln v)]^{\theta-1}\left\{1+(\theta-1)\left[(-\ln u)^{\theta}+(-\ln v)^{\theta}\right]^{-\frac{1}{\theta}}\right\},\qquad \theta\in[1,+\infty).$$

2.7. Μέτρα εξάρτησης και συσχέτισης και συνάρτηση copula

Λαμβάνοντας υπόψη την ποικιλία των συναρτήσεων copula, ένα ερώτημα είναι το πως σχετίζεται η παράμετρος εξάρτησης που εμπεριέχεται στη συναρτησιακή τους έκφραση με την πιο οικεία έννοια της συσχέτισης. Ένα βασικό μέλημα είναι η ικανότητα ενός μοντέλου copula να συλλάβει την εξάρτηση ανάμεσα στις μεταβλητές με ένα ικανοποιητικό τρόπο.

Στο εδάφιο αυτό θα ασχοληθούμε με τα μέτρα εξάρτησης-συσχέτισης και τη σχέση τους με τις συναρτήσεις copula. Αρχικά, θα παρουσιαστούν κάποιες επιθυμητές ιδιότητες των μέτρων. Έπειτα θα εστιάσουμε το ενδιαφέρον στο συντελεστή γραμμικής συσχέτισης του Pearson, το συντελεστή συσχέτισης Rho του Spearman και Tau του Kendall καθώς επίσης και το συντελεστή εξάρτησης ουράς. Οι τρείς τελευταίοι συντελεστές θα εκφραστούν ως συνάρτηση της copula, ακολουθώντας τον Nelsen (2006).

2.7.1. Επιθυμητές ιδιότητες των συντελεστών εξάρτησης

Η έννοια της εξάρτησης ή της ανεξαρτησίας είναι θεμελιώδης στη θεωρία πιθανοτήτων και στατιστικής. Αντίστοιχη είναι και η επιθυμία ποσοτικοποίησης του βαθμού εξάρτησης δύο τυχαίων μεταβλητών. Λόγω του ενδιαφέροντος που παρουσιάζουν οι έννοιες αυτές υπάρχει μία ευρύτατη βιβλιογραφία μέτρων και εννοιών εξάρτησης. Ενδεικτικά αναφέρουμε τις μονογραφίες των Drouet Mari and Kotz (2001), Nelsen (2006), Balakrishnan and Lai (2009, Chapters 3 and 15), τις εργασίες Rényi (1959), Schweizer and Wolff (1981), Micheas and Zografos (2006) και τις αναφορές που υπάρχουν εκεί.

Έστω X, Y δύο τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και από κοινού συνάρτηση κατανομής F. Οι τυχαίες μεταβλητές X, Y θα λέμε ότι είναι εξαρτημένες ή συσχετισμένες αν δεν είναι ανεξάρτητες, δηλαδή όταν ισχύει ότι :

$$F(x, y) \neq F_1(x)F_2(y) \text{ για κάποια } x, y \in R.$$

Ένα μέτρο ή ένας συντελεστής εξάρτησης εκφράζει το βαθμό απόκλισης από την ανεξαρτησία. Θα συμβολίζεται με $\delta(X, Y)$ και θα πρέπει να ικανοποιεί το ακόλουθο βασικό σύνολο ιδιοτήτων (βλέπε, Embrechts et al., 2002).

i. $\delta(X, Y) = \delta(Y, X)$ (συμμετρία).
ii. $-1 \leq \delta(X, Y) \leq 1$ (κανονικότητα).
iii. $\delta(X, Y) = 0$ αν και μόνο αν οι X, Y είναι ανεξάρτητες.

iv. $\delta(X,Y) = 1 \Leftrightarrow (X,Y)$ comotonic.
Αυτό σημαίνει ότι τα X, Y έχουν τέλεια θετική εξάρτηση όταν ο συντελεστής εξάρτησης παίρνει τη μέγιστη τιμή του. Με την έννοια τέλεια θετική εξάρτηση εννοούμε ότι η Y είναι γνησίως αύξουσα συνάρτηση της X.

v. $\delta(X,Y) = -1 \Leftrightarrow (X,Y)$ countermotonic.
Αυτό σημαίνει ότι τα X, Y έχουν τέλεια αρνητική εξάρτηση όταν ο συντελεστής εξάρτησης παίρνει την ελάχιστη τιμή του. Με την έννοια τέλεια αρνητική εξάρτηση εννοούμε ότι η Y είναι γνησίως φθίνουσα συνάρτηση της X.

vi. Για έναν αυστηρά (γνησίως) μονότονο μετασχηματισμό $T: R \longrightarrow R$ του x ισχύει ότι:
$$\delta(T(X),Y) = \begin{cases} \delta(X,Y) & \text{, αν η } T \text{ είναι αύξουσα} \\ -\delta(X,Y) & \text{, αν η } T \text{ είναι φθίνουσα} \end{cases} .$$

Το σύνολο των παραπάνω ιδιοτήτων ενός μέτρου εξάρτησης έχει εμπλουτιστεί στη βιβλιογραφία (βλέπε, μεταξύ πολλών άλλων, Micheas and Zografos (2006), Balakrishnan and Lai (2009, σελ. 145) και αναφορές που εμφανίζονται εκεί).

2.7.2. Συντελεστής γραμμικής συσχέτισης του Pearson

Ο συντελεστής γραμμικής συσχέτισης Pearson αναπτύχθηκε από τον Karl Pearson, είναι η πιο γνωστή έννοια εξάρτησης και μάλιστα γραμμικής στη βιβλιογραφία και θα παρουσιαστεί στο εδάφιο αυτό ακολουθώντας τους Trivedi and Zimmer (2007).

Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και από κοινού συνάρτηση κατανομής F. Όταν οι διακυμάνσεις των X, Y είναι πεπερασμένες, ο συντελεστής συσχέτισης του Pearson μεταξύ του ζεύγους των μεταβλητών (X,Y) ορίζεται ως εξής:
$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y},$$

όπου $Cov(X,Y) = E(XY) - E(X)E(Y)$ η συνδιασπορά των X, Y και $\sigma_X, \sigma_Y > 0$ οι τυπικές αποκλίσεις των X, Y αντίστοιχα.

Ο συντελεστής συσχέτισης Pearson ικανοποιεί τα ακόλουθα:

i. Αναφέρεται σε γραμμικές και μόνο συσχετίσεις.
ii. $\rho(X,Y) = \rho(Y,X)$ (συμμετρία).
iii. $-1 \leq \rho(X,Y) \leq 1$ (κανονικότητα).
iv. Παραμένει αμετάβλητος υπό γραμμικούς και μόνο μετασχηματισμούς των μεταβλητών.

Παρατηρήσεις 2.7.2.1. (α) Καθώς ο συντελεστής συσχέτισης Pearson εκφράζει την γραμμικότητα, σε περίπτωση που δεν υπάρχει γραμμικότητα, προκύπτει μία απρόβλεπτη συσχέτιση των δύο μεταβλητών, κάτι που είναι σύνηθες όταν υπάρχουν ακραίες τιμές.
(β) Στην περίπτωση που το ζεύγος (X,Y) ακολουθεί μια διδιάστατη κανονική κατανομή τότε η δομή εξάρτησης (copula) καθορίζεται πλήρως από την συσχέτιση και η μηδενική συσχέτιση και η ανεξαρτησία είναι ισοδύναμα. **Όμως,** η γενική μηδενική συσχέτιση δεν συνεπάγεται πάντα την ανεξαρτησία. Για παράδειγμα αν $X \sim N(0,1)$ και $Y = X^2$ τότε $Cov(X,Y) = Cov(X,X^2) = E(X^3) - E(X)E(X^2) = E(X^3) - 0 \cdot E(X^2) = E(X^3)$. Όμως, είναι γνωστό ότι αν η τυχαία μεταβλητή X είναι συνεχής και έχει συμμετρική σ.π.π τότε $E(X^\kappa) = 0$, για $\kappa = 2\nu + 1$. Δηλαδή έχουμε ότι $Cov(X,Y) = 0$ αλλά (X,Y) είναι εξαρτημένες.
(γ) Ο συντελεστής συσχέτισης του Pearson δεν ορίζεται για κάποιες κατανομές με βαριές ουρές (heavy-tailed distributions) για τις οποίες δεν υπάρχουν ροπές δεύτερης τάξης όπως π.χ. η διδιάστατη t-Student κατανομή με βαθμούς ελευθερίας 1 ή 2.
(δ) Ο συντελεστής δεν είναι αμετάβλητος υπό αυστηρά αύξοντες μη-γραμμικούς μετασχηματισμούς. Δηλαδή, $\rho\big(T(X), T(Y)\big) \neq \rho(X,Y)$ για $T: R \to R$ γνησίως αύξουσα και μη γραμμική.

Ο συντελεστής συσχέτισης Pearson έχει αρκετά ακόμη μειονεκτήματα. Πιο συγκεκριμένα, ο συντελεστής συσχέτισης Pearson το μόνο που εκφράζει είναι η γραμμική ή μη σχέση των δύο αυτών μεταβλητών αλλά δεν δίνει καμία πληροφορία για το πώς προέρχεται αυτή η σχέση. Παράλληλα, αν προκύψει γραμμική σχέση, είναι πολύ πιθανό να οφείλεται στον τρόπο που έχει διαμορφωθεί και επιλεχθεί το τυχαίο δείγμα που εξετάζεται. Εξαιτίας, λοιπόν αυτών των μειονεκτημάτων έχουν προταθεί οι συντελεστές βαθμολογικής συσχέτισης (rank correlation), οι οποίοι αναλύονται στη συνέχεια.

2.7.3. Συντελεστής βαθμολογικής συσχέτισης (Rank Correlation)

Σε αυτή την κατηγορία ανήκουν οι συντελεστές Rho του Spearman και Tau του Kendall όπου και οι δύο βασίζονται στην έννοια της συμφωνίας (concordance) και ασυμφωνίας (disconcordance) των τυχαίων μεταβλητών σύμφωνα με τον Nelsen (2006). Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και από κοινού συνάρτηση κατανομής F.

Δύο τυχαίες μεταβλητές X και Y είναι σε συμφωνία, όταν μεγάλες (μικρές) τιμές της μίας τείνουν να συσχετιστούν με μεγάλες (μικρές) τιμές της άλλης. Αντίθετα, λέμε ότι είναι σε ασυμφωνία, όταν μεγάλες (μικρές) τιμές της μίας τείνουν να συσχετιστούν με μικρές (μεγάλες) τιμές της άλλης. Ακολουθεί ένας πιο ακριβής ορισμός για την έννοια της συμφωνίας που διατυπώνεται, στη συνέχεια, σύμφωνα με τον Nelsen (2006, σελ. 157).

Ορισμός 2.7.3.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές και (x_1, y_1) και (x_2, y_2) δύο ζεύγη των παρατηρούμενων τιμών τους. Τότε θα έχουμε συμφωνία των (x_1, y_1) και (x_2, y_2) όταν:

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

ή εναλλακτικά όταν $(x_1 - x_2)(y_1 - y_2) > 0$.

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

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

ή εναλλακτικά όταν $(x_1 - x_2)(y_1 - y_2) < 0$.

Μέτρο συμφωνίας
Αρχικά θα δοθεί ένας ορισμός για την διάταξη δύο συναρτήσεων copula στην διδιάστατη περίπτωση και στην συνέχεια θα διατυπωθεί ο ορισμός του μέτρου συμφωνίας (Scarsini, 1984).

Ορισμός 2.7.3.2. Έστω ότι έχουμε δύο διαφορετικές συναρτήσεις copula C_1, C_2. Τότε θα λέμε ότι η C_1 είναι μικρότερη από την C_2 $(C_1 < C_2)$ όταν ισχύει η παρακάτω ανισότητα:

$$C_1(u,v) \leq C_2(u,v), \text{ για κάθε } (u,v) \in [0,1]^2.$$

Ορισμός 2.7.3.3. (Nelsen, 2006) Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές, M ένα μέτρο συσχέτισής τους και C η copula τους. Τότε θα λέμε πως το M είναι μέτρο συμφωνίας των X, Y και θα το συμβολίζουμε με $M_{X,Y}$ ή M_c εάν ικανοποιεί τις παρακάτω ιδιότητες:

i. Το $M_{X,Y}$ ορίζεται για κάθε ζεύγος X, Y συνεχών τυχαίων μεταβλητών.
ii. $-1 \leq M_{X,Y} \leq 1$ με $M_{X,X} = 1, M_{X,-X} = -1$.
iii. $M_{X,Y} = M_{Y,X}$.
iv. Αν οι X, Y είναι ανεξάρτητες τότε $M_{X,Y} = 0$.
v. $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$.
vi. Αν C_1 και C_2 είναι συναρτήσεις copula τέτοιες ώστε $C_1 < C_2$, τότε $M_{C_1} < M_{C_2}$.
vii. Αν $\{(X_n, Y_n)\}$ είναι μία ακολουθία συνεχών τυχαίων μεταβλητών με copulas C_n, και αν η C_n συγκλίνει σημειακά στη C, τότε $\lim_{n\to\infty} M_{C_n} = M_C$.

2.7.3.1. Συντελεστής συσχέτισης Rho του Spearman

Ο συντελεστής συσχέτισης Rho του Spearman, που συμβολίζεται με ρ_s, σύμφωνα με τον Nelsen (1991), προτάθηκε και πήρε το όνομά του από τον Charles Spearman το 1904 και βασίζεται στην έννοια της συμφωνίας και ασυμφωνίας. Όσα ακολουθούν διατυπώνονται σύμφωνα με τον Nelsen (2006).

Έστω $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ ανεξάρτητα τυχαία διανύσματα με από κοινού συνάρτηση κατανομής F και τα X_1, X_2, X_3 έχουν συνάρτηση κατανομής F_1, ενώ τα Y_1, Y_2, Y_3 έχουν συνάρτηση κατανομής F_2. Τότε ο συντελεστής συσχέτισης ρ_s του Spearman ορίζεται να είναι ανάλογος με την πιθανότητα της συμφωνίας μείον την πιθανότητα της ασυμφωνίας για τα διανύσματα (X_1, Y_1) και (X_2, Y_3) (Kruskal, 1958) και ορίζεται να είναι ο ακόλουθος:

$$\rho_s(X, Y) = 3[P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}].$$

Παρατηρήσεις 2.7.3.1.1. (α) Θα μπορούσε να χρησιμοποιηθεί στον παραπάνω ορισμό εξίσου το διάνυσμα (X_3, Y_2).
(β) Το διάνυσμα (X_1, Y_1) έχει από κοινού συνάρτηση κατανομής F, ενώ το διάνυσμα (X_2, Y_3) έχει από κοινού συνάρτηση κατανομής F_1F_2 αφού τα X_2, Y_3 είναι ανεξάρτητα.

Στη συνέχεια ακολουθεί ένα Θεώρημα (βλέπε, Nelsen, 2006) το οποίο ορίζει τον συντελεστή συσχέτισης ρ_s του Spearman όταν έχουμε την συνάρτηση copula C των μεταβλητών X και Y. Η απόδειξη δίνεται χάριν πληρότητας και η σύνθεσή της, βασίστηκε στον Nelsen (2006).

Θεώρημα 2.7.3.1.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και με copula C. Τότε ο συντελεστής συσχέτισης ρ_s του Spearman ορίζεται ως:

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

$$= 12E(UV) - 3.$$

Απόδειξη. Ο συντελεστής συσχέτισης ρ_s του Spearman ορίζεται ως η συσχέτιση Pearson των U, V σύμφωνα με τους Trivedi and Zimmer (2007, σελ. 22). Έτσι, έχουμε:

$$\rho_s(X,Y) = \rho\big(F_1(X), F_2(Y)\big) = \rho(u,v) = \frac{Cov(U,V)}{\sqrt{Var(U)Var(V)}}$$

$$= \frac{E(UV) - E(U)E(V)}{\sqrt{Var(U)Var(V)}} \; .$$

Σημειώνουμε ότι οι $U = F_1(X)$ και $V = F_2(Y)$ είναι ομοιόμορφες στο $(0,1)$, και έτσι έχουν μέσες τιμές ίσες με $1/2$ και διακυμάνσεις ίσες με $1/12$. Επομένως ο συντελεστής συσχέτισης ρ_s του Spearman γίνεται

$$\rho_s(X,Y) = \frac{E(UV) - \frac{1}{2}\times\frac{1}{2}}{\sqrt{\frac{1}{12}\times\frac{1}{12}}} = \frac{E(UV) - \frac{1}{4}}{\frac{1}{12}} = 12E(UV) - 3.$$

Σκοπός μας τώρα είναι να υπολογίσουμε την ποσότητα $E(UV)$ και η απόδειξή μας θα είναι πλήρης. Δεδομένου ότι, η συνάρτηση κατανομής των (U,V) είναι η copula των (X,Y) έχουμε ότι,

$$E(UV) = \int_0^1 \int_0^1 uv \, dC(u,v)$$

και έτσι η πιο πάνω σχέση γίνεται:

$$\rho_s(X,Y) = 12\int_0^1 \int_0^1 uv \, dC(u,v) - 3 = 12\int_0^1 \int_0^1 C(u,v) du dv - 3$$

$$= \; 12\left(\int_0^1 \int_0^1 C(u,v) du dv - \frac{1}{4}\right).$$

Στο σημείο αυτό αναφέρουμε ότι η τιμή $\frac{1}{4}$ είναι ίση με την τιμή του ολοκληρώματος $\int_0^1 \int_0^1 uv \; du dv$ καθώς,

$$\int_0^1 \int_0^1 uv\, dudv = \int_0^1 v \left(\int_0^1 udu \right) dv = \int_0^1 v \left[\frac{u^2}{2} \right]_0^1 dv$$

$$= \int_0^1 v \frac{1}{2} dv = \frac{1}{2} \left[\frac{v^2}{2} \right]_0^1 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Έτσι, η παραπάνω σχέση για τον συντελεστή συσχέτισης ρ_s του Spearman γίνεται,

$$\rho_s(X,Y) = 12 \left(\int_0^1 \int_0^1 C(u,v) dudv - \int_0^1 \int_0^1 uvdudv \right)$$

$$= 12 \int_0^1 \int_0^1 [C(u,v) - uv] dudv. \qquad \blacksquare$$

Ιδιότητες

Ο συντελεστής συσχέτισης ρ_s του Spearman εκτός από τις ιδιότητες του μέτρου συμφωνίας, σύμφωνα με τον Nelsen (2006, σελ. 169) ικανοποιεί τις επιθυμητές ιδιότητες των συντελεστών εξάρτησης. Δηλαδή,

i. $\rho_s(X,Y) = \rho_s(Y,X)$ (συμμετρία).
ii. $-1 \leq \rho_s(X,Y) \leq 1$ (κανονικότητα).
iii. $\rho_s(X,Y) = 1$ εαν − ν $C = C^+$ εαν − ν υπάρχει τέλεια θετική συσχέτιση.
iv. $\rho_s(X,Y) = -1$ εαν − ν $C = C^-$ εαν − ν υπάρχει τέλεια αρνητική συσχέτιση.
v. Το ρ_s παραμένει αμετάβλητο σε γνησίως μονότονους μετασχηματισμούς των μεταβλητών X, Y.

Επίσης , αν οι τυχαίες μεταβλητές X, Y είναι ανεξάρτητες τότε $\rho_s(X,Y) = 0$.

2.7.3.2. Συντελεστής συσχέτισης Tau του Kendall

Ο συντελεστής, που συμβολίζεται με τ, εισήχθη, σύμφωνα με τον Nelsen (1991), για πρώτη φορά από τον Fechner περίπου το 1900 και διατυπώθηκε εκ νέου από τον Kendall το 1938. Ο συντελεστής αυτός βασίζεται στην έννοια της συμφωνίας και ασυμφωνίας.

Έστω (X_1, Y_1) και (X_2, Y_2) δύο ανεξάρτητα και ισόνομα τυχαία διανύσματα με από κοινού συνάρτηση κατανομής F με τα X_1, X_2 να έχουν συνάρτηση κατανομής F_1, ενώ τα Y_1, Y_2 να έχουν συνάρτηση κατανομής F_2. Τότε σύμφωνα με τον Nelsen (2006) ο συντελεστής συσχέτισης τ του Kendall ισούται με την πιθανότητα της συμφωνίας μείον την πιθανότητα της ασυμφωνίας και ορίζεται να είναι ο ακόλουθος:

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

Στη συνέχεια, βασιζόμενοι στο Nelsen (2006), παρουσιάζουμε ένα Θεώρημα σχετικό με τον ορισμό του συντελεστή συσχέτισης τ του Kendall ως συνάρτηση της copula των υπό θεώρηση τυχαίων μεταβλητών. Η απόδειξη ακολουθεί τα βήματα της απόδειξης του Nelsen (2006, σελ.159) και δίνεται χάριν πληρότητας.

Θεώρημα 2.7.3.2.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και με copula C. Τότε ο συντελεστής συσχέτισης τ του Kendall ορίζεται ως:

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

$$= 4E\big(C(u, v)\big) - 1.$$

Απόδειξη. Έστω (X_1, Y_1) και (X_2, Y_2) δύο ανεξάρτητα τυχαία δείγματα από μία συνεχή διδιάστατη κατανομή, με περιθώριες F_1, F_2 και copula C. Έχουμε ότι:

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

$$= P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - 1 + P\{(X_1 - X_2)(Y_1 - Y_2) > 0\}$$

$$= 2[P\{(X_1 - X_2)(Y_1 - Y_2) > 0\}] - 1$$

$$= 2[P(X_1 > X_2, Y_1 > Y_2) + P(X_1 < X_2, Y_1 < Y_2)] - 1. \qquad (1)$$

Θα υπολογίσουμε τις πιθανότητες $P(X_1 > X_2, Y_1 > Y_2)$ και $P(X_1 < X_2, Y_1 < Y_2)$, χρησιμοποιώντας την copula C. Έστω $c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$ η πυκνότητα που αντιστοιχεί στην copula C και $f(x,y)$ η από κοινού πυκνότητα των X, Y. Για τον υπολογισμό, λοιπόν, της πιθανότητας $P(X_1 > X_2, Y_1 > Y_2)$ έχουμε:

$$P(X_1 > X_2, Y_1 > Y_2) = P(X_2 < X_1, Y_2 < Y_1)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(X_2 \leq x, Y_2 \leq y) f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(x,y) f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C\big(F_1(x), F_2(y)\big) \frac{\partial^2}{\partial x \partial y} F(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C\big(F_1(x), F_2(y)\big) \frac{\partial^2}{\partial x \partial y} C\big(F_1(x), F_2(y)\big)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C(u,v) \frac{\partial^2}{\partial u \partial v} C(u,v) \frac{\partial u}{\partial x} \frac{\partial v}{\partial y}\, dx\, dy$$

$$= \int_0^1 \int_0^1 C(u,v)\, c(u,v)\, du dv$$

$$= \int_0^1 \int_0^1 C(u,v)\, dC(u,v).$$

Ομοίως,

$$P(X_1 < X_2, Y_1 < Y_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(X_1 \leq x, Y_1 \leq y) f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x,y) f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C\big(F_1(x), F_2(y)\big)\, \frac{\partial^2}{\partial x \partial y}\, F(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C\big(F_1(x), F_2(y)\big)\, \frac{\partial^2}{\partial x \partial y}\, C\big(F_1(x), F_2(y)\big)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C(u,v)\, \frac{\partial^2}{\partial u \partial v}\, C(u,v) \frac{\partial u}{\partial x}\frac{\partial v}{\partial y}\, dx\, dy$$

$$= \int_{0}^{1}\int_{0}^{1} C(u,v)\, c(u,v)\, du dv$$

$$= \int_{0}^{1}\int_{0}^{1} C(u,v)\, dC(u,v).$$

Συνεπώς, η σχέση (1) μετά και τον υπολογισμό των πιθανοτήτων γίνεται:

$$\tau(X,Y) = 4\int_{0}^{1}\int_{0}^{1} C(u,v)\, dC(u,v) - 1$$

$$= 4E\big(C(u,v)\big) - 1. \qquad \blacksquare$$

Παρατήρηση 2.7.3.2.1. Ο συντελεστής συσχέτισης τ του Kendall ικανοποιεί και αυτός τις επιθυμητές ιδιότητες των μέτρων εξάρτησης όπως αυτές διατυπώθηκαν νωρίτερα για το συντελεστή συσχέτισης ρ_s του Spearman.

Παρά το γεγονός ότι και οι δύο συντελεστές βαθμολογικής συσχέτισης εκφράζουν την πιθανότητα συμφωνίας μεταξύ δύο τυχαίων μεταβλητών, με copula C, οι τιμές τους συχνά διαφέρουν κατά πολύ. Σύμφωνα λοιπόν με τους Durbin and Stuart (1951) οι συντελεστές αυτοί συνδέονται μεταξύ τους με κάποιες ανισοτικές σχέσεις, οι οποίες διατυπώνονται στο ακόλουθο Θεώρημα, η απόδειξη του οποίου υπάρχει στο σύγγραμμα του Nelsen (2006, σελ. 175) και στους Durbin and Stuart (1951).

Θεώρημα 2.7.3.2.2. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές και τ, ρ_s οι συντελεστές συσχέτισης Tau του Kendall και Rho του Spearman, αντίστοιχα. Τότε ισχύουν οι παρακάτω ανισότητες:

$$-1 \leq 3\tau - 2\rho_s \leq 1,$$

$$\frac{3\tau - 1}{2} \leq \rho_s \leq \frac{1 + 2\tau - \tau^2}{2}, \tau \geq 0 \text{ και}$$

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho_s \leq \frac{1 + 3\tau}{2}, \tau \leq 0.$$

Τέλος, θα παρουσιάσουμε τα πλεονεκτήματα των συντελεστών βαθμολογικής συσχέτισης σε σχέση με τον συντελεστή γραμμικής συσχέτισης του Pearson.

Πλεονεκτήματα

- ❖ Ικανοποιούν όλες τις επιθυμητές ιδιότητες και δεν εξαρτώνται από τις περιθώριες συναρτήσεις κατανομής, αλλά μόνο από την copula C.
- ❖ Παραμένουν αναλλοίωτοι σε γνησίως μονότονους μετασχηματισμούς των τυχαίων μεταβλητών και όχι μόνο σε γραμμικούς.

Παρ' όλα αυτά, λόγω του ότι είναι συντελεστές συσχέτισης και όχι εξάρτησης, δεν μας παρέχουν όλες τις δυνατές πληροφορίες για τη δομή εξάρτησης και υπάρχουν περιπτώσεις που λόγω της πολυπλοκότητας των τύπων τους, δεν είναι εύκολος ο υπολογισμός τους.

2.7.4. Συντελεστής εξάρτησης ουράς (Tail dependence)

Το εδάφιο αυτό επικεντρώνεται στο συντελεστή εξάρτησης ουράς ο οποίος, γενικά μιλώντας, εκφράζει το βαθμό εξάρτησης δύο τυχαίων μεταβλητών στις ουρές της αντίστοιχης κατανομής. Ο συντελεστής εξάρτησης ουράς, όπως και ο συντελεστής βαθμολογικής συσχέτισης, εξαρτάται μόνο από την συνάρτηση copula των τυχαίων μεταβλητών. Είναι ένα μέτρο εξάρτησης που αφορά τις ουρές των κατανομών και μετρά την εξάρτηση των μεταβλητών που παίρνουν τιμές στο επάνω-δεξιά και στο κάτω-αριστερά τεταρτημόριο μιας διδιάστατης κατανομής. Ακολουθεί ο ορισμός του συντελεστή εξάρτησης ουράς σύμφωνα με τον Nelsen (2006).

Ορισμός 2.7.4.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα. Ο συντελεστής εξάρτησης της άνω ουράς λ_U, είναι το όριο (αν αυτό υπάρχει) της δεσμευμένης πιθανότητας, ότι η Y είναι μεγαλύτερη του 100-οστού ποσοστημορίου της F_2, δεδομένου ότι η X είναι μεγαλύτερη του 100-οστού ποσοστημορίου της F_1, καθώς το $t \to 1-$. Δηλαδή:

$$\lambda_U = \lim_{t \to 1-} P[Y > F_2^{-1}(t) | X > F_1^{-1}(t)].$$

Εάν υποθέσουμε τώρα ότι C είναι η συνάρτηση copula των X, Y τότε ο συντελεστής εξάρτησης της άνω ουράς λ_U ορίζεται σύμφωνα με το ακόλουθο θεώρημα (Nelsen, 2006).

Θεώρημα 2.7.4.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και λ_U ο συντελεστής εξάρτησης της άνω ουράς όπως ορίστηκε στον Ορισμό 2.7.4.1. Έστω επίσης η συνάρτηση copula C των X, Y. Αν το όριο στον Ορισμό 2.7.4.1 υπάρχει, τότε

$$\lambda_U = 2 - \lim_{t \to 1-} \frac{1 - C(t,t)}{1-t} \,.$$

Απόδειξη. Για τον συντελεστή εξάρτησης της άνω ουράς λ_U έχουμε

$$\begin{aligned}
\lambda_U &= \lim_{t \to 1-} P[Y > F_2^{-1}(t) | X > F_1^{-1}(t)] \\
&= \lim_{t \to 1-} P[F_2(Y) > t | F_1(X) > t] \\
&= \lim_{t \to 1-} P[V > t | U > t] \\
&= \lim_{t \to 1-} \frac{P[V > t, U > t]}{P[U > t]} \,.
\end{aligned}$$

Στο σημείο αυτό υπενθυμίζουμε ότι σύμφωνα με την Παρατήρηση 2.5.1 η από κοινού συνάρτηση επιβίωσης $\bar{C}$ δύο τυχαίων μεταβλητών που κατανέμονται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής τη copula C ορίζεται ως

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

Άρα έχουμε

$$\begin{aligned}
\lambda_U &= \lim_{t \to 1-} \frac{\bar{C}(t,t)}{P[U > t]} = \lim_{t \to 1-} \frac{1 - 2t + C(t,t)}{1-t} = \lim_{t \to 1-} \frac{2 - 2t + C(t,t) - 1}{1-t} \\
&= \lim_{t \to 1-} \left(\frac{2 - 2t}{1-t} + \frac{C(t,t) - 1}{1-t} \right) \\
&= \lim_{t \to 1-} \frac{2(1-t)}{1-t} + \lim_{t \to 1-} \frac{C(t,t) - 1}{1-t} \\
&= 2 + \lim_{t \to 1-} \frac{C(t,t) - 1}{1-t} \\
&= 2 - \lim_{t \to 1-} \frac{1 - C(t,t)}{1-t}. \qquad \blacksquare
\end{aligned}$$

Εάν, $\lambda_U \in (0, 1]$ θα λέμε ότι η copula C παρουσιάζει άνω εξάρτηση ουράς, ενώ αν $\lambda_U = 0$, τότε θα λέμε ότι η copula C παρουσιάζει άνω ανεξαρτησία ουράς.

Ο ορισμός που ακολουθεί αναφέρεται στο συντελεστή εξάρτησης κάτω ουράς και σύμφωνα με τον Nelsen (2006), διατυπώνεται ως εξής.

Ορισμός 2.7.4.2. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα. Ο συντελεστής εξάρτησης της κάτω ουράς λ_L, είναι το όριο (αν αυτό υπάρχει) της δεσμευμένης πιθανότητας, ότι η Y είναι μικρότερη ή ίση του 100-οστού ποσοστημορίου της F_2, δεδομένου ότι η X είναι μικρότερη ή ίση του 100-οστού ποσοστημορίου της F_1, καθώς το $t \to 0+$. Δηλαδή:

$$\lambda_L = \lim_{t \to 0+} P[Y \leq F_2^{-1}(t) | X \leq F_1^{-1}(t)].$$

Εάν υποθέσουμε τώρα ότι C είναι η copula των X, Y τότε ο συντελεστής εξάρτησης της κάτω ουράς λ_L ορίζεται σύμφωνα με το ακόλουθο θεώρημα (Nelsen, 2006).

Θεώρημα 2.7.4.2. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2 αντίστοιχα και λ_L ο συντελεστής εξάρτησης της κάτω ουράς όπως ορίστηκε στον Ορισμό 2.7.4.2. Έστω επίσης η συνάρτηση copula C των X, Y. Αν το όριο στον Ορισμό 2.7.4.2 υπάρχει, τότε

$$\lambda_L = \lim_{t \to 0+} \frac{C(t,t)}{t}.$$

Απόδειξη. Για τον συντελεστή εξάρτησης της κάτω ουράς λ_L έχουμε

$$\begin{aligned}
\lambda_L &= \lim_{t \to 0+} P[Y \leq F_2^{-1}(t) | X \leq F_1^{-1}(t)] \\
&= \lim_{t \to 0+} P[F_2(Y) \leq t | F_1(X) \leq t] \\
&= \lim_{t \to 0+} P[V \leq t | U \leq t] \\
&= \lim_{t \to 0+} \frac{P[V \leq t, U \leq t]}{P[U \leq t]} \\
&= \lim_{t \to 0+} \frac{C(t,t)}{t}.
\end{aligned}$$

■

Εάν, $\lambda_L \in (0, 1]$ θα λέμε ότι η copula C παρουσιάζει κάτω εξάρτηση ουράς, ενώ αν $\lambda_L = 0$, τότε θα λέμε ότι η copula C παρουσιάζει κάτω ανεξαρτησία ουράς. Τέλος, οι δύο παραπάνω συντελεστές είναι μη παραμετρικοί και εξαρτώνται μόνο από τη copula των X, Y.

ΚΕΦΑΛΑΙΟ 3

COPULAS ΑΚΡΑΙΩΝ ΤΙΜΩΝ

3.1. Εισαγωγή

Σε πολλές επιστημονικές περιοχές εμφανίζεται η ανάγκη στοχαστικής μοντελοποίησης ακραίων συμβάντων (rare events), όπως για παράδειγμα ακραίων μετεωρολογικών φαινομένων (πλημμύρες, τυφώνες κ.λπ.), ακραίων σεισμικών φαινομένων, ακραίων συμπεριφορών σε χρηματοοικονομικού τύπου προβλήματα (αιφνίδιες πολύ μεγάλες ζημιές σε χαρτοφυλάκια). Η πιθανοθεωρητική ή στοχαστική θεωρία των ακραίων τιμών (extreme values) καταπιάνεται με προβλήματα ακριβώς αυτού του τύπου και οικοδομεί πιθανοθεωρητικού τύπου μοντέλα τα οποία βοηθούν στη μοντελοποίηση και την πρόβλεψη της εμφάνισης ακραίων συμβάντων. Υπάρχει μία ευρύτατη βιβλιογραφία σε κατανομές ακραίων τιμών η οποία έχει αναπτυχθεί τα τελευταία χρόνια λόγω του αυξημένου ενδιαφέροντος σε προβλήματα από την οικονομία, την μετεωρολογία, την υδρολογία, μεταξύ πολλών άλλων. Ενδεικτικά αναφέρουμε τα βιβλία των Kotz and Nadarajah (2000) και Reiss and Thomas (2007).

Στη μοντελοποίηση ακραίων συμβάντων σημαντικό ρόλο παίζουν οι κατανομές ακραίων τιμών οι οποίες προκύπτουν ως οι οριακές κατανομές του μεγαλύτερου, $X_{(n)}$, ή του μικρότερου $X_{(1)}$, διατεταγμένου στατιστικού σε ένα τυχαίο δείγμα $X_1, X_2, \dots, X_n$ διατεταγμένου σε αύξουσα τάξη μεγέθους $X_1 \leq X_2 \leq \dots \leq X_n$. Ας θεωρήσουμε το παράδειγμα των Gudendorf and Segers (2010) σύμφωνα με το οποίο το ενδιαφέρον εστιάζεται στην ημερήσια καταγραφή του επιπέδου του νερού σε μία θέση μιας λίμνης. Αν ένα πλημμυρικό φαινόμενο σχετίζεται με την υπέρβαση μιας τιμής ορίου του ύψους του νερού της λίμνης στη συγκεκριμένη θέση, τότε είναι προφανές ότι απαιτείται η γνώση της κατανομής του μέγιστου ύψους του νερού, ώστε να μπορεί να υπολογιστεί η πιθανότητα πλημμυρικού φαινομένου. Συνεχίζοντας το παράδειγμα αυτό, ένα πλημμυρικό φαινόμενο δεν σχετίζεται μόνο με το ύψος του νερού σε μια συγκεκριμένη θέση αλλά με το αντίστοιχο ύψος σε δύο ή περισσότερες θέσεις της λίμνης. Έτσι, ο υπολογισμός της πιθανότητας πλημμυρικού φαινομένου απαιτεί τη γνώση της από κοινού κατανομής των μέγιστων υψών νερού στις συγκεκριμένες θέσεις της λίμνης. Στο πλαίσιο αυτό της μοντελοποίησης ακραίων γεγονότων, οι copulas ακραίων τιμών εφοδιάζουν με τα κατάλληλα μοντέλα για την κατασκευή και τη μοντελοποίηση της εξάρτησης των ακραίων συμβάντων. Στο κεφάλαιο αυτό θα παρουσιαστεί μια ανασκόπηση των συναρτήσεων copula ακραίων τιμών για τη διδιάστατη περίπτωση, ακολουθώντας κυρίως τα βιβλία των Kotz and Nadarajah (2000), Drouet Mari and Kotz (2001), Nelsen (2006), την εργασία των Gudendorf and Segers (2010) και τη βιβλιογραφία που αναφέρεται σ'αυτά.

3.2. Διδιάστατη συνάρτηση copula ακραίων τιμών

Αντικείμενο του εδαφίου αυτού είναι η διατύπωση ενός βασικού θεωρήματος το οποίο αφορά στην από κοινού κατανομή του μεγίστου ανεξάρτητων και ισόνομων τυχαίων μεταβλητών και ειδικότερα της αντίστοιχης συνάρτησης copula. Το θεώρημα αυτό θα οδηγήσει στον ορισμό της συνάρτησης copula ακραίων τιμών, ακολουθώντας τη μονογραφία του Nelsen (2006).

Θεωρούμε το διδιάστατο τυχαίο διάνυσμα (X, Y) και έστω n ανεξάρτητες και ισόνομες παρατηρήσεις $(X_1, Y_1), \ldots, (X_n, Y_n)$ σ'αυτό. Συμβολίζουμε με F την από κοινού κατανομή του (X, Y) και με F_1, F_2 τις αντίστοιχες περιθώριες κατανομές των τυχαίων μεταβλητών X και Y και με C την αντίστοιχη συνάρτηση copula, έτσι ώστε $F(x, y) =$

$C\big(F_1(x), F_2(y)\big)$. Έστω $X_{(n)} = \max_i X_i$ και $Y_{(n)} = \max_i Y_i$ συμβολίζει το n-οστό διατεταγμένο στατιστικό στα τυχαία δείγματα $X_1, \dots, X_n$ και $Y_1, \dots, Y_n$ από τις X και Y αντίστοιχα. Αν συμβολίσουμε με $C_{(n)}$ την συνάρτηση copula του διδιάστατου τυχαίου διανύσματος $\big(X_{(n)}, Y_{(n)}\big)$ το ερώτημα που γεννιέται είναι: ποια η σχέση μεταξύ των συναρτήσεων copulas C και $C_{(n)}$; Απάντηση στο ερώτημα αυτό δίνεται στο ακόλουθο θεώρημα.

Θεώρημα 3.2.1. Αν C η copula του τυχαίου διανύσματος (X, Y) και $C_{(n)}$ η copula του τυχαίου διανύσματος $\big(X_{(n)}, Y_{(n)}\big)$ με $X_{(n)} = \max_i X_i$ και $Y_{(n)} = \max_i Y_i$ για n ανεξάρτητες και ισόνομες παρατηρήσεις $(X_1, Y_1), \dots, (X_n, Y_n)$ του (X, Y), τότε,

$$C_{(n)}(u, v) = C^n\left(u^{1/n}, v^{1/n}\right), \qquad (u, v) \in [0,1]^2 .$$

Απόδειξη. Αν $F_{1(n)}$ και $F_{2(n)}$ συμβολίζουν τις περιθώριες κατανομές των $X_{(n)}$ και $Y_{(n)}$ αντίστοιχα, τότε

$$\begin{aligned} F_{1(n)}(x) &= P\big[X_{(n)} \le x\big] = P[\text{όλα τα } X_\iota \le x, \text{για κάθε } i = 1, \dots, n] \\ &= (P[X_i \le x])^n \\ &= F_1^n(x), \text{για } x \in R \end{aligned}$$

και

$$\begin{aligned} F_{2(n)}(y) &= P\big[Y_{(n)} \le y\big] = P[\text{όλα τα } Y_\iota \le y, \text{για κάθε } i = 1, \dots, n] \\ &= (P[Y_\iota \le y])^n \\ &= F_2^n(y), \text{για } y \in R. \end{aligned}$$

Έτσι λοιπόν για την από κοινού συνάρτηση κατανομής $F_{(n)}$ των $\big(X_{(n)}, Y_{(n)}\big)$ θα έχουμε:

$$\begin{aligned} F_{(n)}(x, y) &= P\big[X_{(n)} \le x, Y_{(n)} \le y\big] = P[\text{όλα τα } X_\iota \le x, \text{όλα τα } Y_\iota \le y] \\ &= [F(x, y)]^n \\ &= \big[C\big(F_1(x), F_2(y)\big)\big]^n \\ &= \left[C\left(\big[F_{1(n)}(x)\big]^{1/n}, \big[F_{2(n)}(y)\big]^{1/n}\right)\right]^n . \end{aligned} \qquad (1)$$

Για την copula $C_{(n)}$, στηριζόμενοι στην σχέση (1), θα έχουμε:

$$C_{(n)}(u,v) = F_{(n)}\left(F_{1(n)}^{-1}(u), F_{2(n)}^{-1}(v)\right)$$
$$= C^n\left(u^{1/n}, v^{1/n}\right),$$

για κάθε $(u,v) \in [0,1]^2$, γεγονός το οποίο αποδεικνύει το θεώρημα. ■

Το προηγούμενο θεώρημα διατυπώνει ότι αν είναι γνωστή μια συνάρτηση copula C, τότε μπορεί να κατασκευαστεί μια συνάρτηση copula $C_{(n)}$ που συνδέεται με τη C όπως στο Θεώρημα 3.2.1. Η $C_{(n)}$ είναι ουσιαστικά η συνάρτηση copula του τυχαίου διανύσματος $\left(X_{(n)}, Y_{(n)}\right)$ των n-οστών στατιστικών $\left(X_{(n)}, Y_{(n)}\right)$ και με τη βοήθειά της θα οριστεί η συνάρτηση copula ακραίων τιμών, στον ορισμό που ακολουθεί, στηριζόμενοι στη μονογραφία του Nelsen (2006).

Ορισμός 3.2.1. Μία συνάρτηση copula C_* ονομάζεται copula ακραίων τιμών (extreme value copula), αν υπάρχει μία συνάρτηση copula C τέτοια ώστε:

$$C_*(u,v) = \lim_{n\to\infty} C^n\left(u^{1/n}, v^{1/n}\right),$$

για $u, v \in [0,1]$. Επιπρόσθετα, η C θα λέγεται ότι ανήκει στην περιοχή έλξης (domain of attraction) της C_*.

Παρατήρηση 3.2.1. Ο ορισμός της copula ακραίων τιμών διατυπώνεται μέσω του ορίου, όταν $n \to \infty$, και αφορά τη συνάρτηση copula του $\left(X_{(n)}, Y_{(n)}\right)$, όταν $n \to \infty$. Λαμβάνοντας υπόψη ότι $X_{(n)} = \max_i X_i$ και $Y_{(n)} = \max_i Y_i$ οι $X_{(n)}$ και $Y_{(n)}$ αναφέρονται στις μέγιστες παρατηρήσεις των X και Y και όταν $n \to \infty$ διαισθητικά οι $X_{(n)}$ και $Y_{(n)}$ περιγράφουν ακραίες τιμές των X και Y. Έτσι, διαισθητικά μιλώντας, η συνάρτηση copula των $\left(X_{(n)}, Y_{(n)}\right)$ αναφέρεται σε ακραίες τιμές και ταιριάζει η ονομασία ως copula ακραίων τιμών.

Σημείωση: Σύμφωνα με τον Nelsen (2006, σελ. 97), αν το κατά συνιστώσα όριο μιας ακολουθίας copulas υπάρχει σε κάθε σημείο του $[0,1] \times [0,1]$, τότε το όριο πρέπει να είναι μία copula. Δηλαδή, η C_* είναι μία συνάρτηση copula.

Ακολουθώντας τους Gudendorf and Segers (2010, σελ. 129) η κατασκευή της copula ακραίων τιμών του Ορισμού 3.2.1 μπορεί να απλοποιηθεί μέσω της ιδέας της μέγιστης ευστάθειας (max-stability) του ακόλουθου ορισμού.

Ορισμός 3.2.2. Μία copula C καλείται μεγιστο-ευσταθής (max-stable) αν για κάθε φυσικό αριθμό $r, r \geq 1,$ και για $(u, v) \in [0,1] \times [0,1]$ ισχύει

$$C(u, v) = C^r\left(u^{1/r}, v^{1/r}\right).$$

Το θεώρημα που ακολουθεί συνδέει την έννοια της copula ακραίων τιμών με την έννοια της μεγιστο-ευσταθούς copula. Η απόδειξή του ακολουθεί τα βήματα του Nelsen (2006, σελ. 97).

Θεώρημα 3.2.2. Μία copula C είναι μία copula ακραίων τιμών αν και μόνο αν είναι μεγιστο-ευσταθής copula.

Απόδειξη. Έστω μία copula ακραίων τιμών C_*. Θα δείξουμε ότι είναι και μεγιστο-ευσταθής copula. Αφού η copula C_* είναι copula ακραίων τιμών τότε σύμφωνα με τον Ορισμό 3.2.1 έχουμε ότι ικανοποιείται η σχέση $C_*(u, v) = \lim_{n\to\infty} C^n\left(u^{1/n}, v^{1/n}\right)$, για μία copula C. Έτσι για κάθε φυσικό αριθμό r, με $r \geq 1$, έχουμε:

$$C_*^r\left(u^{1/r}, v^{1/r}\right) = \lim_{n\to\infty} C^{rn}\left(u^{1/rn}, v^{1/rn}\right) = C_*(u, v),$$

δηλαδή, η C_* είναι μεγιστο-ευσταθής copula.
Αντίστροφα, έστω μία μεγιστο-ευσταθής copula C. Θα δείξουμε ότι είναι copula ακραίων τιμών. Αφού η copula C είναι μεγιστο-ευσταθής τότε σύμφωνα με τον Ορισμό 3.2.2 για κάθε φυσικό αριθμό $r, r \geq 1,$ ισχύει ότι $C(u, v) = C^r\left(u^{1/r}, v^{1/r}\right)$. ∎

Στη συνέχεια, το ενδιαφέρον επικεντρώνεται σε μια διαδικασία κατασκευής copulas ακραίων τιμών, η οποία οφείλεται στην εργασία του Pickands (1981). Ακολουθώντας τους Gudendorf and Segers (2010) η διαδικασία αυτή βασίζεται στο ακόλουθο θεώρημα που διατυπώθηκε από τον Pickands. Η απόδειξή του ξεφεύγει από το σκοπό της διατριβής αυτής και μπορεί να βρεθεί στις εργασίες de Haan and Resnick (1987) και Beirlant et al. (2004).

Θεώρημα 3.2.3. Μία διδιάστατη copula C είναι μία copula ακραίων τιμών αν και μόνο αν υπάρχει ένα Borel μέτρο H ορισμένο στο μοναδιαίο τετράγωνο, $\Delta = \{(w_1, w_2) \in [0, +\infty)^2 : w_1 + w_2 = 1\}$, που ονομάζεται φασματικό μέτρο, τέτοιο ώστε:

$$C(u, v) = e^{-l^*(-\log u, -\log v)}, (u, v) \in (0, 1]^2,$$

όπου η σταθερή συνάρτηση εξάρτησης ουράς $l^*: [0, \infty)^2 \to [0, \infty)$ δίνεται από τον τύπο:

$$l^*(x, y) = \int_0^1 max\{(1 - \mathrm{t})x, ty\}\, dH(1 - t, t), \quad (x, y) \in [0, \infty)^2.$$

Στη συνέχεια, θα αναφέρουμε δύο σημαντικές ιδιότητες της σταθερής συνάρτησης εξάρτησης ουράς l^*, οι οποίες βασίζονται στους Gudendorf and Segers (2010).

Πρόταση 3.2.1. Η σταθερή συνάρτηση εξάρτησης ουράς l^* ικανοποιεί τις παρακάτω ιδιότητες

i. Η l^* είναι ομογενής πρώτης τάξης, δηλαδή

$$l^*\big(a(x, y)\big) = a l^*(x, y), \qquad a > 0, (x, y) \in [0, \infty)^2.$$

ii. $max\{x, y\} \leq l^*(x, y) \leq x + y, \quad (x, y) \in [0, \infty)^2.$

iii. Η l^* είναι κυρτή.

Το προηγούμενο θεώρημα (Θεώρημα 3.2.3) εκφράζει μία copula ακραίων τιμών σε όρους της συνάρτησης l^*. Το θεώρημα αυτό αποτελεί, τη βάση ώστε, στη διδιάστατη περίπτωση, κάθε τέτοια συνάρτηση copula C να μπορεί να εκφραστεί σε όρους μιας συνάρτησης εξάρτησης Pickands

(Pickands dependence function) $A: [0,1] \to [1/2\,,1]$, όπως περιγράφεται στο θεώρημα που ακολουθεί.

Θεώρημα 3.2.4. Μία διδιάστατη συνάρτηση copula C είναι copula ακραίων τιμών αν και μόνο αν

$$C(u,v) = e^{\log(uv)\; A\left(\frac{\log v}{\log(uv)}\right)},$$

με $(u,v) \in (0,1]^2$, $(u,v) \neq (1,1)$ και $A: [0,1] \to [1/2\,,1]$ μία κυρτή πραγματική συνάρτηση τέτοια ώστε $max\{t, 1-t\} \leq A(t) \leq 1$, για κάθε $t \in [0,1]$.

Απόδειξη. Σύμφωνα με το Θεώρημα 3.2.3 έχουμε ότι μία συνάρτηση copula C είναι copula ακραίων τιμών αν και μόνο αν $C(u,v) = e^{-l^*(-\log u, -\log v)}$, $(u,v) \in (0,1]^2$. Η συνάρτηση εξάρτησης Pickands A σύμφωνα με τον Patrick Eschenburg (2013, σελ. 17) ορίζεται ως

$$A(t) = \frac{l^*(-\log u\,, -\log v)}{-\log u - \log v}, t = \frac{\log v}{\log(uv)}\ .$$

Έτσι,

$$A(t) = \frac{l^*(-\log u\,, -\ \log v)}{-\log(uv)}$$

ή, ισοδύναμα,

$$l^*(-\log u\,, -\ \log v) = -A(t)\log(uv).$$

Αντικαθιστώντας την ποσότητα $l^*(-\log u\,, -\log v)$ στην σχέση $C(u,v) = e^{-l^*(-\log u, -\log v)}$, έχουμε:

$$\begin{aligned} C(u,v) &= e^{-(-A(t)\log(uv))} \\ &= e^{\log(uv)\mathrm{A}(\mathrm{t})}, \end{aligned}$$

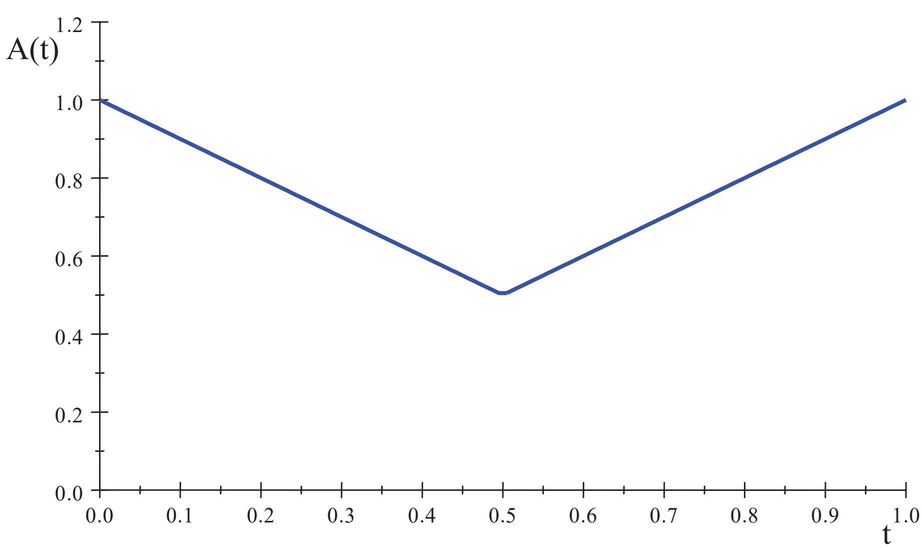
όπου $t = \frac{\log v}{\log(uv)}$. ■

Η συνάρτηση A του θεωρήματος είναι η συνάρτηση εξάρτησης Pickands και μπορεί να δειχθεί ότι ικανοποιεί τις ιδιότητες που σχηματοποιούνται στην ακόλουθη πρόταση (βλέπε, Kotz and Nadarajah, 2000, σελ. 106).

Πρόταση 3.2.2. Η συνάρτηση εξάρτησης Pickands ικανοποιεί τις ιδιότητες

i. Συνδέεται με το μέτρο H του Θεωρήματος 3.2.3 με τη σχέση
$$A(t) = \int_0^1 max\{t(1-q),(1-t)q\}\, H(dq).$$
ii. $A(0) = A(1) = 1$.
iii. $0 \le A^{'}(1) \le 1,\ A^{''}(t) \ge 0,\ max\{t, 1-t\} \le A(t) \le 1,\ 0 \le t \le 1.$
iv. Αν $A(t) = 1$ τότε οι X, Y είναι ανεξάρτητες.
v. Αν $A(t) = max\{t, 1-t\}$ τότε οι X, Y είναι εξαρτημένες.
vi. Η $A(t)$ είναι κυρτή, δηλαδή $A[\lambda y_1 + (1-\lambda)y_2] \le \lambda A(y_1) + (1-\lambda)A(y_2)$.
vii. Αν A_κ είναι συναρτήσεις εξάρτησης τότε η $\sum_{\kappa=1}^{m} \alpha_\kappa A_\kappa$ είναι συνάρτηση εξάρτησης, όπου $\alpha_\kappa \ge 0$ και $\sum_{\kappa=1}^{m} \alpha_\kappa = 1$.

Ακολουθώντας τον Nelsen (2006, σελ. 99) η γραφική παράσταση της συνάρτησης εξάρτησης A είναι της μορφής



Για την συνάρτηση εξάρτησης Pickands A τα πάνω και κάτω όρια έχουν ιδιαίτερη σημασία. Αυτό συμβαίνει καθώς όταν η τιμή της συνάρτησης εξάρτησης Pickands A ισούται με την μονάδα, δηλαδή όταν $A(t) = 1$, οδηγούμαστε στην ανεξαρτησία, δηλαδή $C(u, v) = uv$. Όταν όμως η συνάρτηση εξάρτησης Pickands A ισούται με το κάτω όριο, δηλαδή όταν $A(t) = max\{t, 1 - t\}$, οδηγούμαστε στην τέλεια εξάρτηση, δηλαδή $C(u, v) = max\{u, v\}$. Γενικά, η ανισότητα $A(t) \leq 1$, συνεπάγεται $C(u, v) \geq uv$, δηλαδή οι συναρτήσεις copula ακραίων τιμών είναι positively quadrant dependent, καθώς σύμφωνα με τον Nelsen (2006, σελ. 187), δύο τυχαίες μεταβλητές είναι positively quadrant dependent για κάθε $x, y \in R$ αν

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

ή ισοδύναμα

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

3.3. Ειδικές Περιπτώσεις Συναρτήσεων Εξάρτησης Pickands-Παραδείγματα

Στο εδάφιο αυτό και στο πόρισμα που ακολουθεί θα παρουσιαστεί η σχέση που συνδέει την συνάρτηση εξάρτησης Pickands A και την συνάρτηση εξάρτησης ουράς l^*, ακολουθώντας τον Doyon (2013, σελ. 22).

Πόρισμα 3.3.1. Η συνάρτηση εξάρτησης ουράς l^* δίνεται από τον τύπο

$$l^*(u, v) = (u + v)A\left(\frac{v}{u + v}\right).$$

Απόδειξη. Σύμφωνα με το Θεώρημα 3.2.3 και το Θεώρημα 3.2.4 έχουμε αντίστοιχα ότι

$$C(u, v) = e^{-l^*(-\log u, -\log v)}, \quad (u, v)^T \in (0, 1]^2$$

και

$$C(u, v) = e^{\log (uv)\ A\left(\frac{\log v}{\log (uv)}\right)}, \text{ με } (u, v) \in (0, 1]^2 \text{ και } (u, v) \neq (1, 1).$$

Από αυτές τις δύο σχέσεις έχουμε ότι ισχύει η ακόλουθη ισότητα

$$-l^*(-\log u\,, -\log v) = \log(uv)\, A\left(\frac{\log v}{\log(uv)}\right).$$

Έτσι,

$$l^*(-\log u\,, -\log v) = -(\log u + \log v) A\left(\frac{\log v}{\log u + \log v}\right)$$

ή,

$$l^*(-\log u\,, -\log v) = (-\log u - \log v) A\left(\frac{\log v}{\log u + \log v}\right).$$

Θέτω $-\log u = u$ και $-\log v = v$ και έχω

$$l^*(u, v) = (u + v) A\left(\frac{v}{u + v}\right),$$

γεγονός το οποίο αποδεικνύει το πόρισμα. ■

Στη συνέχεια, θα δοθεί μέσω του επομένου θεωρήματος η πυκνότητα μιας διδιάστατης συνάρτησης copula ακραίων τιμών. Το θεώρημα αυτό διατυπώνεται σύμφωνα με τον Doyon (2013) και για την απόδειξή του παραπέμπουμε στην αναφορά αυτή.

Θεώρημα 3.3.1. Αν υπάρχουν οι παράγωγοι της συνάρτησης εξάρτησης ουράς l^* και είναι καλά ορισμένες τότε η πυκνότητα μιας διδιάστατης συνάρτησης copula ακραίων τιμών C έχει την μορφή

$$c(u, v) = \frac{C(u, v)}{uv}\left[\left(\frac{\partial l^*(x, y)}{\partial x}\,\frac{\partial l^*(x, y)}{\partial y}\right) - \frac{\partial^2 l^*(x, y)}{\partial x \partial y}\right]_{(x,y)=(-\log u, -\log v)},$$

για $(u, v) \in [0, 1]^2$.

Είναι λοιπόν φανερό, ότι για να υπολογιστεί η πυκνότητα μιας διδιάστατης συνάρτησης copula ακραίων τιμών, πρέπει να έχει υπολογιστεί νωρίτερα η συνάρτηση εξάρτησης Pickands A. Εν συνεχεία, για κάποιες γνωστές οικογένειες copula αλλά και για κάποιες κατανομές θα δοθεί η μορφή της συνάρτησης εξάρτησης Pickands A, χωρίς να γίνεται περιγραφή του τρόπου υπολογισμού της.

Marshall – Olkin family

Σύμφωνα με τον Nelsen (2006, σελ. 98) αν έχουμε ότι $A(t) = 1 - min\big(\beta t, \alpha(1-t)\big)$, για $(\alpha, \beta) \in [0,1]^2$ τότε από την σχέση $C(u,v) = e^{\log(uv)\ A\left(\frac{\log v}{\log(uv)}\right)}$, οδηγούμαστε στην οικογένεια των Marshall – Olkin copulas.

Gumbel – Hougaard family

Σύμφωνα με τον Nelsen (2006, σελ. 98) αν έχουμε ότι $A(t) = \left(t^{\theta} + (1-t)^{\theta}\right)^{\frac{1}{\theta}}$, για $\theta \geq 1$, τότε από την σχέση $C(u,v) = e^{\log(uv)\ A\left(\frac{\log v}{\log(uv)}\right)}$, οδηγούμαστε στην οικογένεια των Gumbel – Hougaard copulas.

Bivariate beta distribution

Σύμφωνα με τον Doyon (2013), η παρακάτω διδιάστατη κατανομή Βήτα οφείλεται στους Tawn and Coles (1994) και μπορεί να γραφεί ως

$$F(x,y) = exp\left[-\frac{1}{x}\left\{1 - B\left(q_1 + 1, q_2, \frac{xq_1}{xq_1 + yq_2}\right)\right\}\right.$$
$$\left. - \frac{1}{y}\left\{B\left(q_1, q_2 + 1, \frac{xq_1}{xq_1 + yq_2}\right)\right\}\right],$$

όπου $q_1, q_2 > 0$, $x, y \in [0,1]$,

$$B(\alpha, \beta, \rho) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^{\rho} w^{\alpha - 1}(1-w)^{\beta - 1}\, dw.$$

Η συνάρτηση εξάρτησης Pickands A, στην περίπτωση αυτή, ορίζεται ως

$$A(t) = t\left\{1 - B\left(q_1 + 1, q_2, \frac{(1-t)q_1}{(q_2 - q_1)t + q_1}\right)\right\}$$
$$+(1-t)\left\{B\left(q_1, q_2 + 1, \frac{(1-t)q_1}{(q_2 - q_1)t + q_1}\right)\right\},$$

όπου $t = \frac{y}{x+y}$.

Bivariate asymmetric logistic distribution

Σύμφωνα με τον Doyon (2013), η διδιάστατη asymmetric logistic κατανομή ακραίων τιμών οφείλεται στον Tawn (1988) και βρίσκεται συχνά στην βιβλιογραφία της ανάλυσης επιβίωσης. Η α.σ.κ. της δίνεται από την ακόλουθη σχέση:

$$F_{\psi_1,\psi_2}(x,y) = exp\left\{-\frac{(1-\psi_1)}{x} - \frac{(1-\psi_2)}{y} - \left[\left(\frac{\psi_1}{x}\right)^{\theta} + \left(\frac{\psi_2}{y}\right)^{\theta}\right]^{\frac{1}{\theta}}\right\},$$

όπου $0 \leq \psi_1, \psi_2 \leq 1$ και $\theta > 1$.
Η συνάρτηση εξάρτησης Pickands A ορίζεται ως

$$A(t) = (\psi_2 - \psi_1)t - \psi_2 + 1 + \left[(t\psi_1)^{\theta} + \left(\psi_2(1-t)\right)^{\theta}\right]^{\frac{1}{\theta}},$$

όπου $t = \frac{y}{x+y}$.

Bivariate asymmetric negative logistic distribution

Σύμφωνα με τον Doyon (2013), η διδιάστατη asymmetric negative logistic κατανομή ορίζεται ως

$$F_{\psi_1,\psi_2}(x,y) = exp\left\{-\frac{1}{x} - \frac{1}{y} - \left[\left(\frac{\psi_1}{x}\right)^{\theta} + \left(\frac{\psi_2}{y}\right)^{\theta}\right]^{\frac{1}{\theta}}\right\},$$

όπου $0 \leq \psi_1, \psi_2 \leq 1$ και $\theta < 0$.
Η συνάρτηση εξάρτησης Pickands A δίνεται από τη σχέση

$$A(t) = 1 - \left[(t\psi_1)^{\theta} + \left(\psi_2(1-t)\right)^{\theta}\right]^{\frac{1}{\theta}},$$

όπου $t = \frac{y}{x+y}$.

Hüsler-Reiss distribution

Σύμφωνα με τον Doyon (2013), η κατανομή Hüsler-Reiss (Hüsler and Reiss, 1989) ορίζεται ως

$$F(x,y) = exp\left\{-\frac{1}{y}\Phi\left(s\left(\frac{x}{x+y}\right)\right)-\frac{1}{x}\Phi\left(a-s\left(\frac{x}{x+y}\right)\right)\right\},$$

όπου $s(w)=\frac{\{\alpha^2+2\log(w)-2\log(1-w)\}}{2a}$ και η Φ είναι η συνάρτηση κατανομής της $\mathrm{N}(0,1)$.

Η συνάρτηση εξάρτησης Pickands A δίνεται από

$$A(t) = (1-t)\Phi\left(\frac{a}{2}+\log\left(\frac{1-t}{t}\right)\frac{1}{a}\right)+t\Phi\left(\frac{a}{2}+\log\left(\frac{t}{1-t}\right)\frac{1}{a}\right),$$

όπου $t=\frac{y}{x+y}$.

Asymmetric mixed distribution

Βασιζόμενοι στον Doyon (2013), η διδιάστατη asymmetric mixed κατανομή εισήχθη από τον Tawn (1988) και ορίζεται ως

$$F(x,y) = exp\left\{-\left(\frac{1}{x}+\frac{1}{y}\right)+\frac{(2\beta+\alpha)x+(\beta+\alpha)y}{(x+y)^2}\right\},$$

όπου $\beta\geq 0, \beta+2\alpha\leq 1$ και $\beta+3\alpha\geq 0$.

Η συνάρτηση εξάρτησης Pickands A, στην περίπτωση αυτή, δίνεται από τη σχέση

$$A(t)=\alpha t^3+\beta t^2-(\alpha+\beta)t+1,$$

όπου $t=\frac{y}{x+y}$.

3.4. Copula ακραίων τιμών και συντελεστές συσχέτισης των Kendall και Spearman

Όπως έχει ήδη αναφερθεί στο Κεφάλαιο 2, στα εδάφια 2.7.3.1 και 2.7.3.2, ο συντελεστής συσχέτισης Rho του Spearman καθώς και ο συντελεστής συσχέτισης Tau του Kendall ορίζονται με βάση τη συνάρτηση copula C, αντίστοιχα, ως εξής:

$$\rho_s(x,y) = 12\int_0^1\int_0^1[C(u,v)-uv]dudv,$$

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

Στη συνέχεια, παρουσιάζονται οι αναλυτικές εκφράσεις των προηγούμενων συντελεστών συσχέτισης ως συναρτήσεις της copula ακραίων τιμών C και της αντίστοιχης συνάρτησης Pickands A, σύμφωνα με τον Patrick Eschenburg (2013) καθώς και η απόδειξή τους.

Θεώρημα 3.4.1. Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με copula συνάρτηση C και συνάρτηση εξάρτησης Pickands A. Τότε οι συντελεστές συσχέτισης Rho του Spearman και Tau του Kendall ορίζονται αντίστοιχα ως εξής:

$$\rho_s(x,y) = 12\int_0^1\int_0^1 uv\, dC(u,v) - 3 = 12\int_0^1 \frac{1}{\big(1+A(t)\big)^2}dt - 3,$$

$$\tau(x,y) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = \int_0^1 \frac{t(1-t)}{A(t)}\, dA'(t).$$

Απόδειξη. Θα γίνει η απόδειξη μόνο για τον συντελεστή συσχέτισης Rho του Spearman, καθώς η σχέση για τον συντελεστή συσχέτισης Tau του Kendall αποδεικνύεται παρόμοια**.** Αρχικά, ορίζουμε το ολοκλήρωμα

$$I := \int_0^1\int_0^1 uv\, dC(u,v).$$

Επιπλέον, αν C_1, C_2 είναι συναρτήσεις copulas τότε (Nelsen, 2006, σελ. 164) ισχύει ότι:

$$\int_0^1\int_0^1 C_1(u,v)dC_2(u,v) = \frac{1}{2} - \int_0^1\int_0^1 \frac{\partial}{\partial u}C_1(u,v)\frac{\partial}{\partial v}C_2(u,v)\, dudv. \qquad (2)$$

Εφαρμογή της (2) για $C_1(u,v) = uv$ και $C_2(u,v) = C(u,v)$ δίνει

$$I = \frac{1}{2} - \int_0^1\int_0^1 \frac{\partial}{\partial u}(uv)\frac{\partial}{\partial v}C(u,v)\, dvdu. \qquad (3)$$

Αρχικά, θα υπολογιστεί η τιμή του ολοκληρώματος

$$\int_0^1 \frac{\partial}{\partial u}(uv)\frac{\partial}{\partial v}C(u,v)\,dv,$$

η οποία έχει ως εξής:

$$\begin{aligned}\int_0^1 \frac{\partial}{\partial u}(uv)\frac{\partial}{\partial v}C(u,v)\,dv &= \int_0^1 v\,\frac{\partial C(u,v)}{\partial v}\,dv\\ &= [vC(u,v)]_0^1 - \int_0^1 C(u,v)\;dv\\ &= C(u,1) - \int_0^1 C(u,v)\;dv.\end{aligned}$$

Λόγω της ιδιότητας των συναρτήσεων copulas $C(u,1) = C(1,u) = u$ για κάθε $u \in [0,1]$, έχουμε

$$\int_0^1 \frac{\partial}{\partial u}(uv)\frac{\partial}{\partial v}C(u,v)\,dv = u - \int_0^1 C(u,v)\;dv. \tag{4}$$

Έτσι, κάνοντας αντικατάσταση της σχέσης (4) στη σχέση (3) έχουμε:

$$\begin{aligned}I &= \frac{1}{2} - \int_0^1\left(u - \int_0^1 C(u,v)\;dv\right)du\\ &= \frac{1}{2} - \int_0^1 u\,du + \int_0^1\int_0^1 C(u,v)\;dvdu\\ &= \frac{1}{2} - \left[\frac{u^2}{2}\right]_0^1 + \int_0^1\int_0^1 C(u,v)\;dvdu\\ &= \frac{1}{2} - \frac{1}{2} + \int_0^1\int_0^1 C(u,v)\;dvdu\\ &= \int_0^1\int_0^1 C(u,v)\;dvdu\\ &= \int_0^1\int_0^1 e^{\log(uv)\;A\left(\frac{\log v}{\log(uv)}\right)}\,dudv\\ &:= \int_0^1\int_0^1 f(u,v)\,dudv.\end{aligned}$$

Στη συνέχεια, θέτουμε $u = g(x,t) = x$ και $v = h(x,t) = x^{(1-t)/t}$ και έτσι έχουμε

$$I=\int_0^1\int_0^1 f\big(g(x,t),h(x,t)\big)\left|\frac{\partial(u,v)}{\partial(x,t)}\right|dxdt$$

$$=\int_0^1\int_0^1 e^{\log\left(xx^{(1-t)/t}\right)A\left(\frac{\log x^{(1-t)/t}}{\log\left(xx^{(1-t)/t}\right)}\right)}\left|\frac{\partial(u,v)}{\partial(x,t)}\right|dxdt$$

$$=\int_0^1\int_0^1 e^{\log\left(x^{1+(1-t)/t}\right)A(t)}\left|\frac{\partial(u,v)}{\partial(x,t)}\right|dxdt$$

$$=\int_0^1\int_0^1 e^{\log\left(x^{(t+1-t)/t}\right)A(t)}\left|\frac{\partial(u,v)}{\partial(x,t)}\right|dxdt$$

$$=\int_0^1\int_0^1 e^{\log\left(x^{\frac{1}{t}}\right)A(t)}\left|\frac{\partial(u,v)}{\partial(x,t)}\right|dxdt,$$

όπου $\left|\frac{\partial(u,v)}{\partial(x,t)}\right|=det\begin{pmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial t}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial t}\end{pmatrix}=det\begin{pmatrix}1 & 0\\ \frac{1-t}{t}x^{\frac{1-t}{t}-1} & -\frac{1}{t^2}\log(x)\,x^{(1-t)/t}\end{pmatrix}$

$$=-\frac{1}{t^2}\log(x)\,x^{(1-t)/t}\,.$$

Έτσι, έχουμε

$$I=\int_0^1\int_0^1 e^{\log\left(x^{\frac{1}{t}}\right)A(t)}\left(-\frac{1}{t^2}\log(x)\,x^{(1-t)/t}\right)dxdt$$

$$=\int_0^1\int_0^1 e^{\log\left(x^{\frac{A(t)}{t}}\right)}\left(-\frac{1}{t^2}\log(x)\,x^{(1-t)/t}\right)dxdt$$

$$=\int_0^1\int_0^1 x^{\frac{A(t)}{t}}\left(-\frac{1}{t^2}\log(x)\,x^{(1-t)/t}\right)dxdt$$

$$= \int_0^1 -\frac{1}{t^2}\int_0^1 \log(x)x^{\frac{A(t)+1-t}{t}}\ dxdt$$

$$= \int_0^1 -\frac{1}{t^2}\left[\int_0^1 \log(x)x^{\frac{A(t)+1}{t}-1}\ dx\right]dt$$

$$= \int_0^1 -\frac{1}{t^2}\left\{\left[\log(x)\frac{1}{(1+A(t))/t}x^{\frac{A(t)+1}{t}}\right]_0^1 - \int_0^1 \frac{1}{x}\frac{1}{(1+A(t))/t}x^{\frac{A(t)+1}{t}}dx\right\}dt$$

$$= \int_0^1 -\frac{1}{t^2}\left\{0 - \frac{t}{1+A(t)}\int_0^1 x^{\frac{A(t)+1}{t}-1}\ dx\right\}dt$$

$$= \int_0^1 -\frac{1}{t^2}\left\{-\frac{t}{1+A(t)}\left[\frac{1}{(1+A(t))/t}x^{\frac{A(t)+1}{t}}\right]_0^1\right\}dt$$

$$= \int_0^1 -\frac{1}{t^2}\left\{-\frac{t}{1+A(t)}\ \frac{1}{(1+A(t))/t}\right\}dt$$

$$= \int_0^1 -\frac{1}{t^2}\left\{-\frac{t}{1+A(t)}\ \frac{t}{1+A(t)}\right\}dt$$

$$= \int_0^1 -\frac{1}{t^2}\left\{-\frac{t^2}{(1+A(t))^2}\right\}dt$$

$$= \int_0^1 \frac{1}{(1+A(t))^2}dt\,.$$

Καταλήξαμε, λοιπόν, στην επιθυμητή σχέση,

$$\rho_s(x,y) = 12\int_0^1 \frac{1}{(1+A(t))^2}dt - 3. \qquad \blacksquare$$

ΚΕΦΑΛΑΙΟ 4

ARCHIMEDEAN COPULAS

4.1. Εισαγωγή

Σε αυτό το κεφάλαιο, θα αναφερθούμε σε μία σημαντική κατηγορία συναρτήσεων copula γνωστή ως Archimedean copulas. Οι Archimedean copulas ορίζονται από μια γενική συναρτησιακή έκφραση και δεν προέρχονται από τις πολυδιάστατες συναρτήσεις κατανομών που είναι απόρροια του Θεωρήματος Sklar. Οι συναρτήσεις αυτές, έχουν γενικά μια απλή μορφή και χαρακτηρίζονται από μια σειρά επιθυμητών ιδιοτήτων όπως η προσεταιριστικότητα και μπορούν να εφαρμοστούν για να εκφράσουν τόσο τη θετική, όσο και την αρνητική συσχέτιση (Zhang and Singh, 2006). Οι Archimedean copulas βρίσκουν ένα ευρύ φάσμα εφαρμογών λόγω της ευκολίας με την οποία μπορούν να κατασκευαστούν, εξαιτίας της μεγάλης ποικιλίας των οικογενειών των συναρτήσεων copula που ανήκουν σε αυτή την κλάση καθώς και για τις πολλές και καλές ιδιότητες που έχουν τα μέλη αυτής της οικογένειας, όπως για παράδειγμα η εξάρτηση στις ουρές.

Η κατασκευή των Archimedean copulas βασίζεται σε μία συνάρτηση την οποία θα συμβολίζουμε με φ και θα την ονομάζουμε χαρακτηριστικό γεννήτορα της συνάρτησης copula. Στην ουσία η χρησιμότητά της φαίνεται από το γεγονός ότι, αντί να εξετάζουμε κάθε φορά την Archimedean copula, η οποία είναι συνάρτηση πολλών μεταβλητών, θα εξετάζουμε και μελετούμε τη συνάρτηση φ που είναι συνάρτηση μιας μεταβλητής.

Στο κεφάλαιο αυτό θα οριστεί αρχικά ο χαρακτηριστικός γεννήτορας φ και στη συνέχεια θα παρουσιαστεί η σχέση αυτής της συνάρτησης και των Archimedean copulas σύμφωνα με τους Embrechts, P. Lindskog, F. and McNeil, A. (2001), Nelsen (2006).

4.2. Βασικοί ορισμοί και θεωρήματα

Στο εδάφιο αυτό θα παρουσιάσουμε κάποιους βασικούς ορισμούς ακολουθώντας τη μονογραφία του Nelsen (2006).

Ορισμός 4.2.1. Έστω φ, μία συνεχής, γνησίως φθίνουσα συνάρτηση, όπου $\varphi: [0,1] \to [0,\infty]$, τέτοια ώστε $\varphi(1) = 0$. Η ψευδοαντίστροφη της φ είναι η συνάρτηση $\varphi^{[-1]}: [0,\infty] \to [0,1]$ και ορίζεται ως εξής

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

Η συνάρτηση $\varphi^{[-1]}$ είναι συνεχής και μη αύξουσα στο διάστημα $[0,\infty]$ και γνησίως φθίνουσα στο $[0,\varphi(0)]$. Επιπρόσθετα για την ψευδοαντίστροφη της φ ισχύουν τα εξής:

1) $\varphi^{[-1]}(\varphi(u)) = u$, όπου $u \in [0,1]$.
2) $\varphi\left(\varphi^{[-1]}(t)\right) = \begin{cases} t, & 0 \le t \le \varphi(0) \\ \varphi(0), & \varphi(0) \le t \le \infty \end{cases}$
$= min(t, \varphi(0))$.
3) Αν $\varphi(0) = \infty$, τότε έχουμε ότι $\varphi^{[-1]} = \varphi^{-1}$.

Στη συνέχεια θα διατυπώσουμε ένα θεώρημα σύμφωνα με τον Nelsen (2006), το οποίο θα χρησιμοποιηθεί για τον ορισμό των Archimedean copulas. Η απόδειξη του θεωρήματος δόθηκε από τους Alsina et al. (2005) και βρίσκεται στο βιβλίο του Nelsen (2006, σελ. 111).

Θεώρημα 4.2.1. Έστω φ, μία συνεχής και γνησίως φθίνουσα συνάρτηση, όπου $\varphi: [0,1] \to [0,\infty]$, τέτοια ώστε $\varphi(1) = 0$ και $\varphi^{[-1]}: [0,\infty] \to [0,1]$ η ψευδοαντίστροφη της φ. Τότε, η συνάρτηση $C: [0,1]^2 \to [0,1]$ που δίνεται από τη σχέση

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

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

Το θεώρημα αυτό δίνει τη δυνατότητα διατύπωσης του ορισμού των Archimedean copulas.

Ορισμός 4.2.2. Οι συναρτήσεις copula που έχουν την μορφή

$$C(u,v) = \varphi^{[-1]}\big(\varphi(u) + \varphi(v)\big), \qquad u, v \in [0,1],$$

ονομάζονται Archimedean copulas και η συνάρτηση φ καλείται γεννήτοράς τους. Αν $\varphi(0) = \infty$, τότε έχουμε Archimedean copula με αυστηρό γεννήτορα φ. Στην περίπτωση αυτή, $\varphi^{[-1]} = \varphi^{-1}$, και η συνάρτηση copula $C(u,v) = \varphi^{-1}\big(\varphi(u) + \varphi(v)\big)$ θα ονομάζεται αυστηρή συνάρτηση Archimedean copula.

Από τον ορισμό λοιπόν των Archimedean copulas προκύπτει ότι η αναπαράσταση μιας Archimedean copula μας επιτρέπει να μεταβούμε από τη μελέτη μιας διδιάστατης copula στην εξέταση μιας μονοδιάστατης συνάρτησης φ.

Η ονομασία αυτών των συναρτήσεων copula ως Archimedean copulas εισήχθη από τον Ling (1965), αλλά αναγνωρίστηκε από τους Schweizer and Sklar (1961). Η ονομασία Archimedean δεν ήταν τυχαία, καθώς σύμφωνα με την Αρχιμήδεια ιδιότητα για τους πραγματικούς αριθμούς έχουμε ότι για $a, b \in R$, όπου $a, b > 0$ υπάρχει $n \in Z$ τέτοιος ώστε $na > b$. Κάτι αντίστοιχο ισχύει και για τις συγκεκριμένες συναρτήσεις copula. Για να το αποδείξουμε βέβαια αυτό θα χρειαστούμε τον παρακάτω ορισμό που διατυπώνεται στη μονογραφία του Nelsen (2006, σελ. 122).

Ορισμός 4.2.3. Έστω $u, v \in (0,1)$ και έστω n ένας θετικός ακέραιος αριθμός. Θα ονομάζουμε C −δύναμη του u και θα την συμβολίζουμε με u_C^n τη συνάρτηση του u η οποία ορίζεται από τον αναδρομικό τύπο

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

Πριν διατυπωθεί το παρακάτω θεώρημα που βασίζεται στον Nelsen (2006) θα συμβολίζουμε για το υπόλοιπο της διατριβής αυτής με Φ, το σύνολο των συνεχών, γνησίως φθινουσών (δηλαδή $\varphi^{'}(t) < 0$) και κυρτών ($\varphi^{''}(t) > 0$) συναρτήσεων $\varphi: [0,1] \to [0,\infty]$, με $\varphi(1) = 0$ και $0 < t < 1$.

Θεώρημα 4.2.2. Έστω C μία Archimedean copula, με γεννήτορα φ στο Φ. Για κάθε $u, v \in (0,1)$ υπάρχει θετικός ακέραιος n, τέτοιος ώστε $u_C^n < v$.

Απόδειξη. Βασιζόμενοι στο Nelsen (2006, σελ. 122), έστω $u, v \in (0,1)$. Τότε η $n-$οστή δύναμη της C είναι ίση με $\varphi^{[-1]}\big(n\varphi(u)\big)$. Πράγματι, για $n = 2$, έχουμε ότι

$$u_C^2 = C(u, u_C^1).$$

Όμως από τον ορισμό της $C-$δύναμη του u έχουμε την αρχική συνθήκη $u_C^1 = u$. Έτσι, λοιπόν η παραπάνω σχέση γίνεται

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

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

$$\begin{aligned} u_C^{k+1} &= C\big(u, u_C^k\big) = C\left(u, \varphi^{[-1]}\big(k\varphi(u)\big)\right) \\ &= \varphi^{[-1]}\left(\varphi(u) + \varphi\left(\varphi^{[-1]}\big(k\varphi(u)\big)\right)\right) \\ &= \varphi^{[-1]}\big(\varphi(u) + k\varphi(u)\big). \end{aligned}$$

Αφού τα $\varphi(u), \varphi(v)$ είναι θετικοί πραγματικοί αριθμοί σύμφωνα με την Αρχιμήδεια ιδιότητα υπάρχει ακέραιος n, τέτοιος ώστε $n\varphi(u) > \varphi(v)$. Όμως, επειδή $v > 0$ και αφού η συνάρτηση φ είναι συνεχής και γνησίως φθίνουσα έχουμε ότι $\varphi(v) < \varphi(0)$. Επιπρόσθετα, δεδομένου ότι η συνάρτηση $\varphi^{[-1]}$ είναι συνεχής και γνησίως φθίνουσα στο $[0, \varphi(0)]$ ισχύει ότι

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

δηλαδή αποδείχτηκε το ζητούμενο. ■

Στη συνέχεια, θα παρουσιαστούν ορισμένες αλγεβρικές ιδιότητες των Archimedean copulas στηριζόμενοι στον Nelsen (2006, σελ. 113).

Θεώρημα 4.2.3. Έστω C μία Archimedean copula, με γεννήτορα φ. Τότε ισχύουν τα παρακάτω

i. Η copula C είναι συμμετρική, δηλαδή ισχύει ότι $C(u,v) = C(v,u)$, για κάθε $u, v \in (0,1)$.
ii. Η copula C είναι προσεταιριστική, με την έννοια ότι $C(C(u,v),w) = C\big(u, C(v,w)\big)$, για κάθε $u, v, w \in (0,1)$.
iii. Αν c είναι μία θετική σταθερά, τότε η συνάρτηση $c\varphi$ είναι και αυτή γεννήτορας της C.

Απόδειξη. i) Από τον ορισμό των Archimedean copulas έχουμε ότι

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

για κάθε $u, v \in (0,1)$.

ii)
$$\begin{aligned} C(C(u,v),w) &= \varphi^{[-1]}\Big(\varphi\big(C(u,v)\big) + \varphi(w)\Big) \\ &= \varphi^{[-1]}\Big(\varphi\Big(\varphi^{[-1]}\big(\varphi(u) + \varphi(v)\big)\Big) + \varphi(w)\Big) \\ &= \varphi^{[-1]}\big(\varphi(u) + \varphi(v) + \varphi(w)\big) \\ &= \varphi^{[-1]}\Big(\varphi(u) + \varphi\Big(\varphi^{[-1]}\big(\varphi(v) + \varphi(w)\big)\Big)\Big) \\ &= \varphi^{[-1]}\Big(\varphi(u) + \varphi\big(C(v,w)\big)\Big) \\ &= C\big(u, C(v,w)\big). \end{aligned}$$

iii)
$$\begin{aligned} c\varphi\big(C(u,v)\big) &= c\varphi\Big(\varphi^{[-1]}\big(\varphi(u) + \varphi(v)\big)\Big) = c\big(\varphi(u) + \varphi(v)\big) \\ &= c\varphi(u) + c\varphi(v) \end{aligned}$$ ■

4.3. Ισοσταθμικά σύνολα (level sets) μιας Archimedean copula και πυκνότητα της διδιάστατης συνάρτησης κατανομής

Οι Archimedean copulas δεν είναι όλες απόλυτα συνεχείς καθώς υπάρχουν κάποιες από αυτές οι οποίες έχουν ιδιάζοντα στοιχεία (singular components) με την έννοια ότι η πυκνότητά τους ή μηδενίζεται ή δεν ορίζεται. Θα ορίσουμε, λοιπόν στην συνέχεια την πυκνότητα μιας Archimedean copula αλλά και την πυκνότητα της διδιάστατης συνάρτησης κατανομής που παράγεται από τον γεννήτορα φ, σύμφωνα με τους Genest and MacKay (1986) και Nelsen (2006, σελ. 128).

Κάθε γεννήτορας φ της κλάσης Φ παράγει μια διδιάστατη συνάρτηση κατανομής για το ζεύγος των τυχαίων μεταβλητών X, Y, η οποία ορίζεται ως

$$F(x,y) = \begin{cases} \varphi^{[-1]}\big(\varphi(x) + \varphi(y)\big), & \text{αν } \varphi(x) + \varphi(y) \leq \varphi(0) \\ 0, & \text{αν } \varphi(x) + \varphi(y) > \varphi(0) \end{cases}.$$

Πρόταση 4.3.1. Η πυκνότητα της διδιάστατης συνάρτησης κατανομής δύο συνεχών τυχαίων μεταβλητών X, Y με από κοινού συνάρτηση κατανομής F η οποία γεννάται από την συνάρτηση $\varphi \in \Phi$ είναι της μορφής

$$f(x,y) = -\frac{\varphi''\big(F(x,y)\big)\varphi'(x)\varphi'(y)}{\big[\varphi'\big(F(x,y)\big)\big]^3}$$

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

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

$$F(x,y) = \begin{cases} \varphi^{[-1]}\big(\varphi(x) + \varphi(y)\big), & \text{αν } \varphi(x) + \varphi(y) \leq \varphi(0) \\ 0, & \text{αν } \varphi(x) + \varphi(y) > \varphi(0) \end{cases}.$$

Αν $\varphi(0) = \infty$, τότε έχουμε ότι $\varphi^{[-1]} = \varphi^{-1}$, σύμφωνα με αυτά που ορίστηκαν για την ψευδοαντίστροφη στο δεύτερο εδάφιο αυτού του κεφαλαίου. Τότε, λοιπόν, η $F(x,y)$ είναι γνησίως θετική, εκτός από τα σημεία που έχουν είτε $x = 0$, είτε $y = 0$. Παραγωγίζοντας την $\varphi\big(F(x,y)\big) = \varphi(x) + \varphi(y)$ ως προς x παίρνουμε

$$\varphi'(x) = \varphi'\big(F(x,y)\big)\frac{\partial F(x,y)}{\partial x}. \quad (1)$$

Ξαπαραγωγίζοντας τώρα την σχέση (1) ως προς y παίρνουμε

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

ή

$$\varphi'\big(F(x,y)\big)\frac{\partial^2 F(x,y)}{\partial x \partial y} = -\varphi''\big(F(x,y)\big)\frac{\partial F(x,y)}{\partial y}\frac{\partial F(x,y)}{\partial x}$$

ή

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = -\frac{\varphi''\big(F(x,y)\big)}{\varphi'\big(F(x,y)\big)} \frac{\partial F(x,y)}{\partial y} \frac{\partial F(x,y)}{\partial x}. \qquad (2)$$

Χρησιμοποιώντας την σχέση (1), αντικαθιστούμε την ποσότητα $\frac{\partial F(x,y)}{\partial x}$ με την τιμή $\frac{\varphi'(x)}{\varphi'(F(x,y))}$ και αντίστοιχα την ποσότητα $\frac{\partial F(x,y)}{\partial y}$ με την τιμή $\frac{\varphi'(y)}{\varphi'(F(x,y))}$, το οποίο προκύπτει παραγωγίζοντας την ποσότητα $\varphi\big(F(x,y)\big) = \varphi(x) + \varphi(y)$ ως προς y. Έτσι, η σχέση (2) τελικά γίνεται

$$f(x,y) = -\frac{\varphi''\big(F(x,y)\big)}{\varphi'\big(F(x,y)\big)} \frac{\varphi'(y)}{\varphi'\big(F(x,y)\big)} \frac{\varphi'(x)}{\varphi'\big(F(x,y)\big)}$$

ή

$$f(x,y) = -\frac{\varphi''\big(F(x,y)\big)\varphi'(x)\varphi'(y)}{\left[\varphi'\big(F(x,y)\big)\right]^3},$$

γεγονός που αποδεικνύει το ζητούμενο. ■

Ο ορισμός που ακολουθεί διατυπώνει την έκφραση της πυκνότητας copula που αντιστοιχεί σε μία Archimedean copula.

Ορισμός 4.3.1. Η πυκνότητα της απόλυτα συνεχούς διδιάστατης Archimedean copula C, με γεννήτορα $\varphi \in \Phi$, ορίζεται ως η συνάρτηση

$$c(u,v) = -\frac{\varphi''\big(C(u,v)\big)\varphi'(u)\varphi'(v)}{\left[\varphi'\big(C(u,v)\big)\right]^3},$$

και είναι θετική για όλα τα u, v τέτοια ώστε $\varphi(u) + \varphi(v) \le \varphi(0)$. Όταν $\varphi(u) + \varphi(v) = \varphi(0)$ οι παράγωγοι δεν υπάρχουν.

Αφού ορίσαμε λοιπόν την πυκνότητα, θα δούμε πότε έχουμε ιδιάζον στοιχείο σύμφωνα με τους Genest and MacKay (1986).

Θεώρημα 4.3.1. Η συνάρτηση κατανομής $F(x,y)$ η οποία παράγεται από τη συνάρτηση φ έχει ιδιάζον στοιχείο αν και μόνο αν $\frac{\varphi(0)}{\varphi'(0)} \neq 0$. Αν έχει ιδιάζον στοιχείο τότε αυτό θα βρίσκεται στην καμπύλη $\varphi(x)+\varphi(y) = \varphi(0)$ με πιθανότητα $-\frac{\varphi(0)}{\varphi'(0)}$.

Απόδειξη. Για την απόδειξη αυτού του θεωρήματος χρειάζεται να χρησιμοποιήσουμε το ολοκλήρωμα της πυκνότητας $f(x,y)$, όπου αυτή ορίζεται, δηλαδή για x,y τέτοια ώστε $\varphi(x)+\varphi(y) < \varphi(0)$. Αρχικά θα κάνουμε την αλλαγή μεταβλητών των x,y σε u,v όπου

$$u = F(x,y)$$
$$v = x$$

ή

$$u = \varphi^{[-1]}\big(\varphi(x)+\varphi(y)\big)$$
$$v = x$$

και επιλύοντας ως προς x και y, προκύπτει

$$y = \varphi^{[-1]}\big(\varphi(u)-\varphi(v)\big)$$
$$x = v.$$

Η Ιακωβιανή ορίζουσα είναι ίση με

$$J = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \frac{\varphi'(u)}{\varphi'(y)} & -\frac{\varphi'(v)}{\varphi'(y)} \end{vmatrix} = -\frac{\varphi'(u)}{\varphi'(y)}$$
$$= -\frac{\varphi'(u)}{\varphi'\left(\varphi^{[-1]}\big(\varphi(u)-\varphi(v)\big)\right)}.$$

Η μάζα πιθανότητας που εξηγείται από την πυκνότητα είναι η

$$p = \iint_{\{(x,y):\,\varphi(x)+\varphi(y)<\varphi(0)\}} f(x,y)\,dxdy = \int_0^1 \int_u^1 f\left(v, \varphi^{[-1]}\big(\varphi(u)-\varphi(v)\big)\right) |J|\,dvdu.$$

Για τον υπολογισμό της p, αρχικά θα υπολογίσουμε την ποσότητα $f\left(v, \varphi^{[-1]}(\varphi(u) - \varphi(v))\right)$ που βρίσκεται μέσα στο ολοκλήρωμα και υπολογίζεται σύμφωνα με τον ορισμό της πυκνότητας. Έτσι,

$$f\left(v, \varphi^{[-1]}(\varphi(u) - \varphi(v))\right) =$$

$$-\frac{\varphi''\left(\varphi^{[-1]}\left(\varphi(v) + \varphi\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)\right)\right)\varphi'(v)\varphi'\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)}{\left[\varphi'\left(\varphi^{[-1]}\left(\varphi(v) + \varphi\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)\right)\right)\right]^3}$$

$$= -\frac{\varphi''\left(\varphi^{[-1]}(\varphi(v) + \varphi(u) - \varphi(v))\right)\varphi'(v)\varphi'\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)}{\left[\varphi'\left(\varphi^{[-1]}(\varphi(v) + \varphi(u) - \varphi(v))\right)\right]^3}$$

$$= -\frac{\varphi''\left(\varphi^{[-1]}(\varphi(u))\right)\varphi'(v)\varphi'\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)}{\left[\varphi'\left(\varphi^{[-1]}(\varphi(u))\right)\right]^3}$$

$$= -\frac{\varphi''(u)\varphi'(v)\varphi'\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)}{[\varphi'(u)]^3}.$$

Συνεπώς, η ποσότητα $f\left(v, \varphi^{[-1]}(\varphi(u) - \varphi(v))\right)\ |J|$ γίνεται

$$f\left(v, \varphi^{[-1]}(\varphi(u) - \varphi(v))\right)\ |J| = -\frac{\varphi''(u)\varphi'(v)\varphi'\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)}{[\varphi'(u)]^3}$$

$$\times \quad \frac{\varphi'(u)}{\varphi'\left(\varphi^{[-1]}(\varphi(u) - \varphi(v))\right)}$$

$$= -\frac{\varphi^{''}(u)\varphi^{'}(v)}{[\varphi^{'}(u)]^2},$$

και έτσι αφού έχουμε υπολογίσει τις ποσότητες που βρίσκονται μέσα στο ολοκλήρωμα, είμαστε σε θέση να υπολογίσουμε και το ίδιο το ολοκλήρωμα μετά και την αλλαγή των μεταβλητών. Έτσι, έχουμε

$$p = \int_0^1 \int_u^1 f\left(v, \varphi^{[-1]}\big(\varphi(u) - \varphi(v)\big)\right) |J|\, dvdu$$

$$= \int_0^1 \int_u^1 -\frac{\varphi^{''}(u)\varphi^{'}(v)}{[\varphi^{'}(u)]^2}\, dvdu$$

$$= \int_0^1 -\frac{\varphi^{''}(u)}{[\varphi^{'}(u)]^2}\left[\int_u^1 \varphi^{'}(v)\, dv\right] du$$

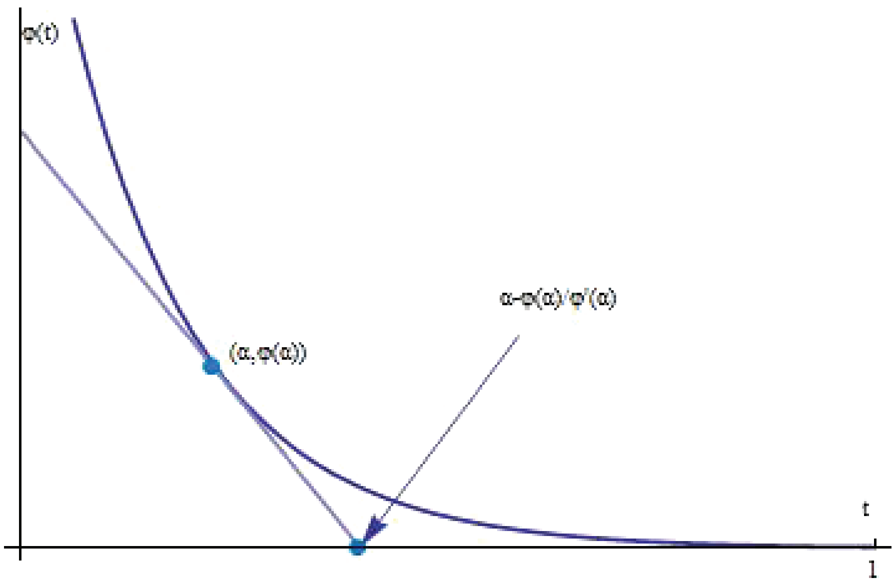
$$= \int_0^1 -\frac{\varphi^{''}(u)}{[\varphi^{'}(u)]^2}[\varphi(v)]_u^1\, du$$

$$= \int_0^1 -\frac{\varphi^{''}(u)}{[\varphi^{'}(u)]^2}\big(\varphi(1) - \varphi(u)\big)\, du.$$

Όμως $\varphi(1) = 0$ και έτσι η μάζα πιθανότητας που εξηγείται από την πυκνότητα γίνεται

$$p = \int_0^1 \frac{\varphi^{''}(u)\varphi(u)}{[\varphi^{'}(u)]^2} du = \int_0^1 \left(-\frac{1}{\varphi^{'}(u)}\right)^{'} \varphi(u)\, du$$
$$= \left[-\frac{1}{\varphi^{'}(u)}\varphi(u)\right]_0^1 - \int_0^1 -\frac{1}{\varphi^{'}(u)}\varphi^{'}(u)\, du$$
$$= \left[-\frac{\varphi(u)}{\varphi^{'}(u)}\right]_0^1 + 1$$
$$= 1 - \frac{\varphi(1)}{\varphi^{'}(1)} + \frac{\varphi(0)}{\varphi^{'}(0)}.$$

Στο σημείο, αυτό, και για την ολοκλήρωση της απόδειξης θεωρούμε την ακόλουθη γραφική παράσταση που προέρχεται από τη μονογραφία του Nelsen (2006, σελ. 129).



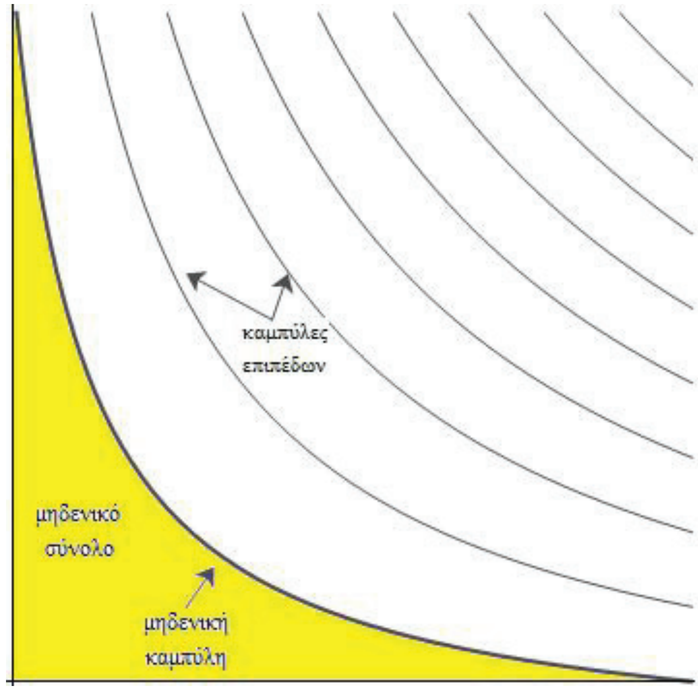
Σχήμα 4.1. Γραφική παράσταση του γεννήτορα φ, μαζί με την εφαπτόμενη στο σημείο $\big(\alpha, \varphi(\alpha)\big)$.

Από την παραπάνω γραφική παράσταση της συνάρτησης φ, παρατηρούμε ότι η ποσότητα $\alpha - \frac{\varphi(\alpha)}{\varphi'(\alpha)}$ είναι το σημείο τομής της εφαπτόμενης της γραφικής παράστασης στο σημείο α με τον οριζόντιο άξονα. Καθώς $\alpha \to 1$, η ποσότητα $-\frac{\varphi(\alpha)}{\varphi'(\alpha)} \to 0$. Επομένως η παραπάνω μάζα πιθανότητας είναι ίση με

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

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

Στο εδάφιο αυτό, επίσης, θα ασχοληθούμε με τα ισοσταθμικά σύνολα (level sets). Τα ισοσταθμικά σύνολα μιας συνάρτησης copula είναι τα $\{(u, v) \in (0,1)^2 : C(u, v) = t,\ 0 < t < 1\}$. Για μία Archimedean copula συνάρτηση, όπου $t > 0$, αυτό το σύνολο αποτελείται από τα σημεία που είναι πάνω στις ισοσταθμικές καμπύλες (level curves) της μορφής $\varphi(u) + \varphi(v) = \varphi(t)$. Για $t = 0$, δηλαδή εκεί που μηδενίζεται η συνάρτηση copula, θεωρούμε ότι έχουμε το μηδενικό σύνολο της C, (zero set), δηλαδή η κίτρινη περιοχή στο επόμενο σχήμα που προέρχεται από το Nelsen (2006, σελ. 124). Η οριακή καμπύλη $\varphi(u) + \varphi(v) = \varphi(0)$ ονομάζεται μηδενική καμπύλη (zero curve) της C. Για τις Archimedean copulas αποδεικνύεται σύμφωνα με το Nelsen (2006) ότι οι καμπύλες επιπέδων είναι κυρτές.



Σχήμα 4.2. Καμπύλες επιπέδων και μηδενική καμπύλη μιας συνάρτησης copula (Nelsen, 2006, σελ. 124).

Ακολουθεί ένα βασικό θεώρημα, το οποίο, θα μας δώσει την μορφή που έχει η συνάρτηση κατανομής των copulas της κλάσης αυτής. Επίσης, διατυπώνεται και άλλο ένα θεώρημα, για το πώς μπορούμε να παράγουμε δύο ανεξάρτητες τυχαίες μεταβλητές βάσει δύο άλλων που είναι ομοιόμορφες στο $(0,1)$ και έχουν από κοινού συνάρτηση κατανομής την Archimedean copula συνάρτηση. Τα παρακάτω θεωρήματα διατυπώνονται σύμφωνα με τον Nelsen (2006). Το πρώτο δίνεται χωρίς απόδειξη, για την οποία παραπέμπουμε στο Nelsen (2006, σελ. 127).

Θεώρημα 4.3.2. Έστω U, V δύο τυχαίες μεταβλητές που κατανέμονται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής μία Archimedean copula C, η οποία παράγεται από την συνάρτηση $\varphi \in \Phi$. Τότε η συνάρτηση κατανομής της τυχαίας μεταβλητής $C(U, V)$ είναι η

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}, \text{για } t \in [0,1].$$

Επιπρόσθετα, η από κοινού συνάρτηση κατανομής των U και $C(U,V)$ είναι η

$$K'_C(s,t) = \begin{cases} s, & s \leq t \\ t - \dfrac{\varphi(t) - \varphi(s)}{\varphi'(t^+)}, & s > t \end{cases},$$

για κάθε $s,t \in [0,1]$.

Θεώρημα 4.3.3. Έστω U,V δύο τυχαίες μεταβλητές που κατανέμονται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής μία Archimedean copula C, η οποία παράγεται από την συνάρτηση $\varphi \in \Phi$. Έστω, επίσης, οι τυχαίες μεταβλητές $S = \frac{\varphi(U)}{\varphi(U)+\varphi(V)}$ και $T = C(U,V)$. Η από κοινού συνάρτηση κατανομής των S και T είναι η

$$H(s,t) = s \cdot K_C(t), \text{για κάθε } (s,t) \in [0,1]^2.$$

Περαιτέρω, οι τυχαίες μεταβλητές S και T είναι ανεξάρτητες και η S κατανέμεται ομοιόμορφα στο $(0,1)$.

Απόδειξη. Η απόδειξη, θα γίνει για την περίπτωση που η συνάρτηση copula C είναι απόλυτα συνεχής και στηρίζεται στον Nelsen (2006, σελ. 130). Η απόδειξη για την γενική περίπτωση υπάρχει στους Genest and Rivest (1993). Πριν προσδιοριστεί η πυκνότητα των τυχαίων μεταβλητών S και T υπενθυμίζουμε ότι η πυκνότητα της απόλυτα συνεχούς διδιάστατης Archimedean copula C με γεννήτορα $\varphi \in \Phi$ ορίζεται ως η ποσότητα

$$c(u,v) = -\frac{\varphi''\big(C(u,v)\big)\varphi'(u)\varphi'(v)}{\left[\varphi'\big(C(u,v)\big)\right]^3}. \tag{3}$$

Γνωρίζοντας την πυκνότητα των (U,V), για την εύρεση της κατανομής των (S,T) θεωρούμε το μετασχηματισμό

$$\begin{matrix} t = C(u,v) \\ s = \dfrac{\varphi(u)}{\varphi(u)+\varphi(v)} \end{matrix} \quad \text{ή} \quad \begin{matrix} t = \varphi^{[-1]}\big(\varphi(u)+\varphi(v)\big) \\ s = \dfrac{\varphi(u)}{\varphi(u)+\varphi(v)} \end{matrix}.$$

Στοιχειώδης άλγεβρα δίνει

$$\begin{cases} \varphi(t) = \varphi(u) + \varphi(v) \\ s = \dfrac{\varphi(u)}{\varphi(u) + \varphi(v)} \end{cases} \text{ ή } \begin{cases} \varphi(t) = \varphi(u) + \varphi(v) \\ s = \dfrac{\varphi(u)}{\varphi(t)} \end{cases} \text{ ή } \begin{cases} \varphi(t) = \varphi(u) + \varphi(v) \\ \varphi(u) = s\varphi(t) \end{cases}$$

$$\text{ή } \begin{cases} \varphi(u) = s\varphi(t) \\ \varphi(v) = \varphi(t) - \varphi(u) \end{cases} \text{ ή } \begin{cases} \varphi(u) = s\varphi(t) \\ \varphi(v) = \varphi(t) - s\varphi(t) \end{cases} \text{ ή } \begin{cases} \varphi(u) = s\varphi(t) \\ \varphi(v) = (1-s)\varphi(t) \end{cases}.$$

Η πυκνότητα, λοιπόν, των τυχαίων μεταβλητών S και T σύμφωνα με την (3), αλλά και χρησιμοποιώντας τους μετασχηματισμούς $u = \varphi^{[-1]}\big(s\varphi(t)\big)$ και $v = \varphi^{[-1]}\big((1-s)\varphi(t)\big)$έχει την ακόλουθη μορφή

$$h(s,t) = -\frac{\varphi''(t)\varphi'\left(\varphi^{[-1]}\big(s\varphi(t)\big)\right)\varphi'\left(\varphi^{[-1]}\big((1-s)\varphi(t)\big)\right)}{[\varphi'(t)]^3}\left|\frac{\partial(u,v)}{\partial(s,t)}\right|,$$

όπου $\left|\dfrac{\partial(u,v)}{\partial(s,t)}\right| = \begin{vmatrix} \dfrac{\partial u}{\partial s} & \dfrac{\partial u}{\partial t} \\ \dfrac{\partial v}{\partial s} & \dfrac{\partial v}{\partial t} \end{vmatrix}$.

Οπότε χρησιμοποιώντας την $\varphi^{[-1]}$ σαν φ^{-1} και γνωρίζοντας ότι $(f^{-1})'(\beta) = \frac{1}{f'(f^{-1}(\beta))}$ έχουμε

$$\frac{\partial u}{\partial s} = \varphi(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big(s\varphi(t)\big)\right)},$$

$$\frac{\partial u}{\partial t} = s\varphi'(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big(s\varphi(t)\big)\right)},$$

$$\frac{\partial v}{\partial s} = -\varphi(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big((1-s)\varphi(t)\big)\right)}$$

και

$$\frac{\partial v}{\partial t} = (1-s)\varphi'(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big((1-s)\varphi(t)\big)\right)}.$$

Επομένως,

$$\left|\frac{\partial(u,v)}{\partial(s,t)}\right| = \varphi(t)(1-s)\varphi'(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big(s\varphi(t)\big)\right)}$$

$$\times\frac{1}{\varphi'\left(\varphi^{[-1]}\big((1-s)\varphi(t)\big)\right)}$$

$$+ s\varphi'(t)\varphi(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big((1-s)\varphi(t)\big)\right)}$$

$$\times\frac{1}{\varphi'\left(\varphi^{[-1]}\big(s\varphi(t)\big)\right)}.$$

Όμως λαμβάνοντας υπόψη ότι $\varphi(u) = s\varphi(t)$ και $\varphi(v) = (1-s)\varphi(t)$ προκύπτει ότι

$$\left|\frac{\partial(u,v)}{\partial(s,t)}\right| = \varphi(t)(1-s)\varphi'(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big(\varphi(u)\big)\right)}\frac{1}{\varphi'\left(\varphi^{[-1]}\big(\varphi(v)\big)\right)}$$

$$+ s\varphi'(t)\varphi(t)\frac{1}{\varphi'\left(\varphi^{[-1]}\big(\varphi(v)\big)\right)}\frac{1}{\varphi'\left(\varphi^{[-1]}\big(\varphi(u)\big)\right)}$$

$$= \frac{\varphi(t)(1-s)\varphi'(t) + s\varphi'(t)\varphi(t)}{\varphi'(u)\varphi'(v)}$$

$$= \frac{\varphi(t)\varphi'(t)}{\varphi'(u)\varphi'(v)}.$$

Συνεπώς, έχουμε ότι

$$h(s,t) = -\frac{\varphi''(t)\varphi(t)\varphi'(t)}{[\varphi'(t)]^3} = -\frac{\varphi''(t)\varphi(t)}{[\varphi'(t)]^2}.$$

Άρα,

$$H(s,t) = \int_0^s\int_0^t h(x,y)\,dydx = \int_0^s\int_0^t -\frac{\varphi''(y)\varphi(y)}{[\varphi'(y)]^2}dydx$$

$$= \int_0^s \int_0^t \left(-\frac{1}{\varphi'(y)}\right)' \varphi(y)dy\, dx.$$

Θα υπολογιστεί αρχικά το ολοκλήρωμα

$$\int_0^t \left(-\frac{1}{\varphi'(y)}\right)' \varphi(y)dy.$$

$$\int_0^t \left(-\frac{1}{\varphi'(y)}\right)' \varphi(y)\, dy = \left[-\frac{1}{\varphi'(y)}\varphi(y)\right]_0^t - \int_0^t -\frac{1}{\varphi'(y)}\varphi'(y)\, dy$$

$$= \left[-\frac{\varphi(y)}{\varphi'(y)}\right]_0^t + t$$

$$= t - \frac{\varphi(t)}{\varphi'(t)} + \frac{\varphi(0)}{\varphi'(0)}.$$

Τελικά,

$$H(s,t) = \int_0^s t - \frac{\varphi(t)}{\varphi'(t)} + \frac{\varphi(0)}{\varphi'(0)} dx = \left[x\left(t - \frac{\varphi(t)}{\varphi'(t)} + \frac{\varphi(0)}{\varphi'(0)}\right)\right]_0^s$$
$$= s\left(t - \frac{\varphi(t)}{\varphi'(t)} + \frac{\varphi(0)}{\varphi'(0)}\right).$$

Η $K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}, t \in [0,1]$, είναι α.σ.κ. Άρα για $t \to -\infty$, $K_C(-\infty) = K_C(0) = 0$, τότε $\frac{\varphi(0)}{\varphi'(0)} = 0$.

Έτσι, αφού όπως ορίσαμε $K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}, t \in [0,1]$ προκύπτει το ζητούμενο

$$H(s,t) = sK_C(t), \text{για κάθε } (s,t) \in [0,1]^2. \qquad \blacksquare$$

4.4. Μέτρα εξάρτησης και συσχέτισης και οικογένειες της κλάσης των συναρτήσεων Archimedean copulas

Στο Κεφάλαιο 2, στο εδάφιο 2.7, ορίστηκε και μελετήθηκε ο συντελεστής βαθμολογικής συσχέτισης Rho του Spearman και Tau του Kendall καθώς επίσης και ο συντελεστής εξάρτησης ουρών για τις συναρτήσεις copula. Αντίστοιχα, στο εδάφιο αυτό, θα ορίσουμε τον συντελεστή βαθμολογικής συσχέτισης Tau του Kendall καθώς επίσης και τον συντελεστή εξάρτησης ουρών για τις Archimedean copulas. Ακολουθούν τα πορίσματα στα οποία διατυπώνονται αυτοί οι συντελεστές σύμφωνα με τον Nelsen (2006, σελ. 163, 215).

Πόρισμα 4.4.1. Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής μία Archimedean copula συνάρτηση C, η οποία παράγεται από τον γεννήτορα $\varphi \in \Phi$. Τότε ο συντελεστής συσχέτισης τ του Kendall για τις X, Y δίνεται από την έκφραση

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

Απόδειξη. Έστω U, V δύο τυχαίες μεταβλητές που κατανέμονται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής C, και έστω K_C η συνάρτηση κατανομής της τυχαίας μεταβλητής $C(U, V)$. Για τον συντελεστή συσχέτισης τ του Kendall γνωρίζουμε ότι ισχύει

$$\begin{aligned}\tau &= 4E\big(C(U,V)\big) - 1 = 4\int_0^1 \mathrm{t}\,\mathrm{d}\,K_C(t) - 1 \\ &= 4\left([tK_C(t)]_0^1 - \int_0^1 K_C(t)\, dt\right) - 1 \\ &= 4\left(K_C(1) - \int_0^1 K_C(t)\, dt\right) - 1.\end{aligned}$$

Όμως όπως έχουμε ορίσει στο Θεώρημα 4.3.2 η συνάρτηση κατανομής της τυχαίας μεταβλητής $C(U, V)$ ορίζεται ως

$$K_C(t) = t - \frac{\varphi(t)}{\varphi^{'}(t^+)}, \text{για } t \in [0,1].$$

Έτσι $K_C(1) = 1 - \frac{\varphi(1)}{\varphi^{'}(1)} = 1$, και επομένως χρησιμοποιώντας αυτές τις δύο σχέσεις ο συντελεστής συσχέτισης τ παίρνει την μορφή

$$\tau = 4\left(1 - \int_0^1 \left(t - \frac{\varphi(t)}{\varphi^{'}(t^+)}\right) dt\right) - 1 = 3 - 4\int_0^1 \left(t - \frac{\varphi(t)}{\varphi^{'}(t^+)}\right) dt$$

$$= 3 - 4\int_0^1 t\, dt + 4\int_0^1 \frac{\varphi(t)}{\varphi^{'}(t^+)}\, dt$$

$$= 3 - 4\left[\frac{t^2}{2}\right]_0^1 + 4\int_0^1 \frac{\varphi(t)}{\varphi^{'}(t)}\, dt$$

$$= 3 - 4\frac{1}{2} + 4\int_0^1 \frac{\varphi(t)}{\varphi^{'}(t)}\, dt$$

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

όπου η ποσότητα $\varphi^{'}(t^+)$ αντικαταστάθηκε από την ποσότητα $\varphi^{'}(t)$ καθώς οι κοίλες συναρτήσεις είναι διαφορίσιμες σχεδόν παντού, δηλαδή αποδείξαμε το ζητούμενο. ■

Πόρισμα 4.4.2. Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής μία Archimedean copula C, η οποία παράγεται από την $\varphi \in \Phi$. Τότε αν το όριο υπάρχει ο συντελεστής εξάρτησης της άνω ουράς δίνεται από τη σχέση:

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

Απόδειξη. Αρχικά υπενθυμίζουμε ότι η από κοινού συνάρτηση επιβίωσης $\bar{C}$ δύο τυχαίων μεταβλητών που κατανέμονται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής τη copula C ορίζεται ως

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

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

Έτσι,

$$\begin{aligned}
\lambda_U &= \lim_{u\to 1-} \frac{\bar{C}(u,u)}{1-u} = \lim_{u\to 1-} \frac{1-2u+C(u,u)}{1-u} \\
&= \lim_{u\to 1-} \frac{1-2u+\varphi^{[-1]}\big(2\varphi(u)\big)}{1-u} \\
&= \lim_{u\to 1-} \frac{2-2u-1+\varphi^{[-1]}\big(2\varphi(u)\big)}{1-u} \\
&= \lim_{u\to 1-} \frac{2(1-u)-1+\varphi^{[-1]}\big(2\varphi(u)\big)}{1-u} \\
&= \lim_{u\to 1-} \frac{2(1-u)}{1-u} + \lim_{u\to 1-} \frac{-1+\varphi^{[-1]}\big(2\varphi(u)\big)}{1-u} \\
&= 2 - \lim_{u\to 1-} \frac{1-\varphi^{[-1]}\big(2\varphi(u)\big)}{1-u}.
\end{aligned}$$

Θέτω στο σημείο αυτό, $s = \varphi(u)$, οπότε $u = \varphi^{[-1]}(s)$ και έτσι έχουμε

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

Πόρισμα 4.4.3. Έστω X,Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής μία Archimedean copula C, η οποία παράγεται από την $\varphi \in \Phi$. Τότε αν το όριο υπάρχει ο συντελεστής εξάρτησης της κάτω ουράς δίνεται από τον τύπο

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

Απόδειξη. Για τον συντελεστή εξάρτησης της κάτω ουράς έχουμε

$$\lambda_L = \lim_{u \to 0+} \frac{C(u,u)}{u} = \lim_{u \to 0+} \frac{\varphi^{[-1]}\big(2\varphi(u)\big)}{u}.$$

Θέτω στο σημείο αυτό, $s = \varphi(u)$, οπότε $u = \varphi^{[-1]}(s)$ και έτσι έχουμε

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

Οικογένειες Archimedean copulas

Τέλος, θα κλείσουμε το εδάφιο αυτό κάνοντας μια αναφορά σε τρείς γνωστές οικογένειες copula οι οποίες ανήκουν στην κλάση των Archimedean copulas και έχουν παρουσιαστεί και στο Κεφάλαιο 2, στο εδάφιο 2.6. Θα παρουσιαστούν επίσης οι γεννήτορες της κάθε οικογένειας χωρίς να γίνει λεπτομερέστατη περιγραφή τους.

Οικογένεια Clayton

Όπως έχει ήδη παρουσιαστεί η οικογένεια Clayton είναι της μορφής

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

και για $\theta > 0$, παίρνει την μορφή

$$C(u,v,\theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \qquad \theta \in (0,+\infty).$$

Ο γεννήτορας της οικογένειας αυτής είναι η συνάρτηση

$$\varphi(t) = \frac{t^{-\theta} - 1}{\theta}, \quad \theta \in (0,+\infty).$$

Οικογένεια Frank

Όπως έχει ήδη παρουσιαστεί η οικογένεια Frank είναι της μορφής

$$C(u,v,\theta) = -\theta^{-1} \log\left\{1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right\},$$

με $\theta \in (-\infty,+\infty) - \{0\}$.

Ο γεννήτορας της οικογένειας αυτής είναι η συνάρτηση

$$\varphi(t) = -\log\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}, \quad \theta \in (-\infty, +\infty) - \{0\},$$

και επιπλέον, ο γεννήτορας αυτός είναι αυστηρός, καθώς $\varphi(0) = \infty$.

Οικογένεια Gumbel

Όπως έχει ήδη παρουσιαστεί η οικογένεια Gumbel είναι της μορφής

$$C(u, v, \theta) = e^{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{\frac{1}{\theta}}}, \qquad \theta \in [1, +\infty).$$

Ο γεννήτορας της οικογένειας αυτής είναι η συνάρτηση

$$\varphi(t) = (-\ln t)^{\theta}, \qquad \theta \in [1, +\infty)\,.$$

και επιπλέον, ο γεννήτορας αυτός είναι αυστηρός, καθώς $\varphi(0) = \infty$.

ΠΕΡΙΛΗΨΗ ΣΤΗΝ ΑΓΓΛΙΚΗ ΓΛΩΣΣΑ

ABSTRACT

The concepts of dependence or independence between two or more random variables is of fundamental importance in statistics and probability theory since independence in the data is always a desirable property (cf. Micheas and Zografos, 2006). In this context, copula theory provides us with tools to formulate dependence or independence in the data. Therefore, the main aim of this thesis is to present and discuss the most basic results of copula theory by following the existing literature, as it is mentioned in the specific places in the text.

Copulas is a relatively recent subject in probability theory and statistics and they have been developed to formulate correlated multivariate data, arising in several disciplines and contexts. Following Fisher (1997), in his article in the Encyclopedia of Statistical Sciences, "Copulas are of interest to statisticians for two main reasons: Firstly, as a way of studying scale-free measures of dependence, and secondly, as a starting point for constructing families of bivariate distributions, ...". The notion of copula was introduced by Abe Sklar in 1959 and the main theorem which inspired the definition of copula function has the name of Sklar. Sklar used the term copula since this function provides a link between the marginal distributions and the joint distribution of two or more random variables. Albeit Sklar formulated the definition of copula, this does not necessarily mean that the copula functions did not exist before. The idea of the copula function has been previously appeared in the papers by Hoeffding (1940, 1941), which were mainly focused on two-dimensional standard distributions. The aforementioned work has moreover established the best possible limits for copula functions which were also elaborated in the papers by Fréchet (1951, 1958). Many authors, such as, Gumbel, Plackett, Mardia, Ali, Mikhail and Haq, Clayton, Joe, Genest, Nelsen, dealt with copulas and some of their contributions in the field appeared in the references.

The aim of this thesis is a presentation of the most important results which are related to copulas. More specifically, Chapter 2 is concentrated to Sklar’s theorem, which is of fundamental importance in copula theory. It is also concentrated to the definition of the copula function and to the Fréchet-Hoeffding bounds. Some basic properties of the copulas are presented and reference is made to the notion of copula density and to the notion of survival copulas. Some known families of copulas are discussed, in the sequel, and the interest is, moreover, focused in the definition and study of some correlation measures, namely the Pearson linear correlation coefficient, the Spearman’s Rho and Kendall’s Tau correlation coefficients. Chapter 3 deals with the two-dimensional extreme value copula. At the beginning, the definition of this specific family of copulas is presented and, then, the connection of this copula with the concept of max-stable copulas is discussed. Then, the Pickand’s (1981) process of constructing copulas of extreme values, is given. In addition, the interest is focused on some special cases of Pickands Dependence and this chapter is finished by a tabulation of the Pickands dependence function of some known copula families and the respective Kendall and Spearman correlation coefficients. The most general class of copula functions, that is the Archimedean copulas, are introduced in Chapter 4. Initially, some basic definitions, properties and results related to Archimedean copulas are presented. Then, the level sets of Archimedean copulas and the respective densities, in the bivariate case, are formulated. Dependency and correlation measures are presented and Clayton, Frank and Gumbel families of copulas are discussed.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΕΛΛΗΝΙΚΗ

Χαραλαμπίδης, Χ. (1990): *Θεωρία πιθανοτήτων και εφαρμογές*. Εκδόσεις Συμμετρία.

ΞΕΝΟΓΛΩΣΣΗ

Ali, M.M., Mikhail N.N and Haq M.S. (1978): A class of bivariate distributions including the bivariate Logistic. *Journal of Multivariate Analysis*, **8**, 405-412.

Alsina, C., Frank, M.J., Schweizer, B. (2005): *Associative Functions on Intervals: A Primer of Triangular Norms.* World Scientific, Hackensack.

Balakrishnan, N. and Lai, C.D. (2009): *Continuous Bivariate Distributions*, Second Edition, Springer.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004): *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd., Chichester.

Clayton, D.G. (1978): A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.

Cook, R.D., Johnson, M.E. (1981): A family of distributions for modelling nonelliptically symmetric multivariate data. *Journal of the Royal Statistical Society, Series B,* **43**, 210–218.

Cook, R.D., Johnson, M.E. (1986): Generalized Burr-Paretologistic distribution with applications to a uranium exploration data set. *Technometrics*, **28**, 123–131.

De Haan, L. and Resnick, S. (1987): On regular variation of probability densities. *Stochastic Process. Appl.*, **25**, 83–95.

Doyon, G. (2013): *On densities of extreme value copulas.* M.Sc. Thesis. Universität Zürich.

Drouet Mari, D. and Kotz, S. (2001): *Correlation and Dependence*. Imperial College Press, London.

Durbin, J. and A. S. Stuart (1951): Inversions and rank correlations. *Journal of Royal Statistical Society, Series B*, **13**, 303–309.

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

Embrechts, P., McNeil, A, and Straumann, D. (2002): Correlation and dependence in risk management: Properties and pitfalls. In: M. A. H. Dempster (ed.): Risk *Management: Value at Risk and Beyond. Cambridge*: Cambridge University Press, 176–223.

Eschenburg, P. (2013): *Properties of extreme-value copulas*. Diplomarbeit. Technische Universität München Fakultät für Mathematik.

Farlie, D.J.G. (1960): The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307–323.

Frank, M.J. (1979): On the simultaneous associativity of F(x,y) and x + y − F(x,y). *Aequationes Mathematicae*, **19**, 194–226.

Fisher N.I. (1997): Copulas. In: *Encyclopedia of Statistical Sciences*, **1**, Wiley, New York, 159-163.

Fréchet, M. (1951): Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon.*, Sect A, **9**, 53-77.

Fréchet, M. (1958): Remarques au sujet de la note précédente. *C.R. Acad. Sci. Paris*, **246**, 2719-2720.

Genest, C. and MacKay, R.J. (1986): The Joy of Copulas: bivariate distributions with uniform marginals. *Amer. Stat.*, **40**, 280–283.

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

Gudendorf, G. and Segers, J. (2010): Extreme-value copulas. *In: Copula Theory and Its Applications.* Proceedings of the Workshop Held in Warsaw, 25-26 September 2009 (Ed. P. Jaworski, F. Durante, W. K. Härdle and T. Rychilk). Springer, 127–145.

Gumbel, E.J. (1958): Distributions a plusiers variables dont les marges sont donness. *Comptes Rendus*, **246**, 2717-2719.

Gumbel, E.J. (1960): Distributions des Valeurs Extremes en Plusieurs Dimensions. *Publications de l'Institute de Statistíque de l'Université de Paris*, **9**, 171–173.

Gumbel, E.J. (1961): Bivariate logistic distributions. *Journal of the American Statistical Association*, **56**, 335-349.

Hoeffding, W. (1940): Masstabinvariante Korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin,* **5,** 179-233 [Reprinted as Scale-invariant correlation theory. In: Fisher, N.I., Sen, P.K. (eds), *The Collected Works of Wassily Hoeffding*. Springer, New York, 57-107].

Hoeffding, W. (1941): Masstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen. *Arkiv für matematischen Wirtschaften und Sozialforschung*, **7**, 49-70 [Reprinted as Scale-invariant correlation measures for discontinuous distributions. In: Fisher, N.I., Sen, P.K. (eds) *The Collected Works of Wassily Hoeffding*. Springer, New York, 109-133].

Hüsler, J. and Reiss, R. (1989): Maxima of Normal Random Vectors: Between Independence and Complete Dependence, *Statistics and Probability Letters*, **7**, 283–286.

Hutchinson, T.P. and Lai, C.D. (1990): *Continuous Bivariate Distributions, Emphasising Applications*. Sydney, Australia: Rumsby.

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

Joe, H. (2015): *Dependence modeling with copulas*. CRC Press.

Kendall, M.G. (1938): A new measure of rank correlation. *Biometrika*, 30, 81–93.

Kimeldorf, G., Sampson, A. (1975): Uniform representations of bivariate distributions. *Communications in Statistics*, **4**, 617–627.

Kotz, S. and Nadarajah, S. (2000): *Extreme value distributions: Theory and Applications.* Imperial College Press, London.

Kruskal, W.H. (1958): Ordinal measures of association. *Journal of the American Statistical Association,* **53**, 814-861.

Lee, L. (1983): Generalized econometric models with selectivity. *Econometrica*, **51**, 507–512.

Ling, C.H. (1965): Representation of associative functions. *Publ. Math. Debrecen*, **12**, 189-212.

Mardia, K.V. (1970): *Families of Bivariate Distributions*. Griffin, London.

Micheas, A. and Zografos, K. (2006): Measuring stochastic dependence using φ-divergence. *Journal of Multivariate Analysis.* **97**, 765-784.

Morgenstern, D. (1956): Einfache beispiele zweidimensionaler verteilungen. *Mitteilingsblatt für Mathematishe Statistik*, **8**, 234–235.

Nelsen, R.B. (1986): Properties of a one-parameter family of bivariate distributions with specified marginals. *Comm. Stat.-Theory Meth.,* **15**, 3277-3285.

Nelsen, R.B. (1987): Discrete bivariate distributions with given marginals and correlation, *Communication in Statistics-Simulation and Computation,* **16**, 199-208.

Nelsen, R.B. (1991): Copulas and association. In: Dall'Aglio G, Kotz S, Salinetti G, (eds): *Advances in Probability Distributions with Given Marginals*. Dordrecht: Kluwer Academic Publishers, 31–74.

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

Pickands, J. (1981): Multivariate extreme value distributions (with discussion). *Bulletin of the International Statistical Institute*, **49**, 859–878.

Plackett, R.L. (1965): A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516-522.

Reiss, R.D. and Thomas, M. (2007): *Statistical Analysis of Extreme Values with applications to insurance, finance, hydrology and other fields*. Birkhäuser Verlag, Basel, Boston, Berlin.

Rényi, A. (1959): On measures of dependence. *Acta Math. Acad. Sci. Hungar*., **10**, 441-451.

Rüschendorf, L. (2009): On the distributional transform, Sklar's theorem and the empirical copula process. *Journal of Statistical Planning and Inference*, **139**, 3921– 3927.

Scarsini, **M.** (1984): On measures of concordance. *Stochastica*, **8**, 201-218.

Schweizer, B. (1991): Thirty years of copulas. In: G. Dall'Aglio, S. Kotz, and G. Salinetti (eds.): *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*. The Netherlands: Kluwer Academic Publishers.

Schweizer, B. and Sklar, A. (1961): Associative functions and statistical triangle inequalities, *Publ. Math. Debrecen*., **8**, 169–186.

Schweizer, B. and Wolff, E.F. (1981): On nonparametric measures of dependence for random variables. *Ann. Stat*., **9**, 879-885.

Sklar, A. (1959): Fonctions de repartition a n dimensions et leurs marges. *Inst. Stat. Univ. Paris Publ.,* **8**, 229-231.

Spearman, C. (1904): The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72–101.

Tawn, J. (1988): Bivariate extreme- value theory, Models and estimation. *Biometrika*, **75**, 397–415.

Tawn, J. and Coles, S. (1994): Statistical methods for multivariate extremes: an application to structural design. *Applied Statistics*, **43**, 1-48.

Trivedi, P.K. and Zimmer D.M. (2007): *Copula modeling: An introduction for practitioners*, Foundations and Trends in Econometrics, **1**, 1-111.

Whitt, W. (1976): Bivariate distributions with given marginals. *Ann. Stat.,* **4**, 1280-1289.

Zhang, L. and Singh, V.P. (2006): Bivariate flood frequency analysis using the copula method. *Journal of Hydrologic Engineering*, **11**, 150- 164.