ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

«Μοντελοποίηση τυχαίων εξαρτημένων μεταβλητών μέσω *copulas* - Οι εφαρμογές τους στη στατιστική»

Γεώργιος Σ. Αρφάνης

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

Πειραιάς
Ιούλιος 2009

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

«Μοντελοποίηση τυχαίων εξαρτημένων μεταβλητών μέσω *copulas* - Οι εφαρμογές τους στη στατιστική»

Γεώργιος Σ. Αρφάνης

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

Πειραιάς
Ιούλιος 2009

UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN APPLIED STATISTICS

"MODELING DEPENDENCE OF RANDOM VARIABLES THROUGH COPULA METHOD – THEIR APPLICATION IN STATISTICS"

By George S. Arfanis

MSc Dissertation

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

Piraeus, Greece

July 2009

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

Τα μέλη της Επιτροπής ήταν:

- ………………………………..…………………..……………………... (Επιβλέπων)
- ………………………………..…………………..……………………...
- ………………………………..……………………………………….....

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

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη

Οι συζεύξεις (*Copulas*) αποτελούν στις μέρες μας ένα από τα πιο πολυσυζητημένα θέματα της στατιστικής επιστήμης με πολλές εφαρμογές, κυρίως στα χρηματοοικονομικά. Η ευρεία χρήση τους έγκειται στο γεγονός ότι συνδέουν πολυδιάστατες συναρτήσεις κατανομών με τις αντίστοιχες μονοδιάστατες περιθώριές τους. Στην εργασία αυτή, δίνεται μια πιο θεωρητική προσέγγιση των συζεύξεων και των ιδιοτήτων τους η οποία αρχίζει με μια ιστορική αναδρομή. Έπειτα δίνονται αναλυτικοί ορισμοί και ιδιότητες των συζεύξεων, όπως το Θεώρημα του *Sklar* και η έννοια της εξάρτησης μεταξύ τυχαίων μεταβλητών. Στη συνέχεια αναλύονται κατηγορίες συζεύξεων όπως η *Marshal-Olkin*, οι Ελλειπτικές συζεύξεις και οι Αρχιμήδειες συζεύξεις όπου δίνεται περισσότερο βάρος, λόγω της ποικιλίας, των πολλών εφαρμογών και της ευκολίας με την οποία προσομοιώνονται οι περισσότερες οικογένειες συζεύξεων που ανήκουν στην κατηγορία αυτή. Επίσης, παρουσιάζονται εκτενώς παραμετρικές και μη-παραμετρικές μέθοδοι εκτίμησης καθώς και τρόποι επιλογής συζεύξεων. Τέλος δίνονται ορισμένοι αλγόριθμοι προσομοίωσης συζεύξεων και εξετάζονται δύο σημαντικές εφαρμογές των συζεύξεων στην στατιστική.

Abstract

Copulas are one of the most frequently discussed issues in statistical science with many applications, particularly in finance. Their wide use is based on the fact that they connect multidimensional distribution functions, with their respective one-dimensional marginals.In this paper, a more theoretical approach of copulas and their properties is given, which starts with a throwback. Then we give analytical definitions and copula properties such as *Sklar's* theorem and the concept of dependence between random variables. We continue by analyzing some copula classes such as the *Marshal*-*Olkin*, the elliptical copulas and the Archimedean copulas which are more thoroughly examined because of their variety, the number of applications they have and the ease of simulating copula families in this class. As a next step we present some parametric and non-parametric methods for estimating copulas and ways of choosing the right copula. Finally we give simulation algorithms and two very important applications of copulas in statistics.

ΚΕΦΑΛΑΙΟ 1

1. Εισαγωγή

Οι συζεύξεις εκφράζουν στην περίπτωση των διδιάστατων κατανομών, τη συναρτησιακή σχέση της αθροιστικής συνάρτησης κατανομής μιας διδιάστατης κατανομής με τις αθροιστικές συναρτήσεις κατανομής των μονοδιάστατων περιθωρίων κατανομών. Η λέξη σύζευξη (*copula*) χρησιμοποιήθηκε για πρώτη φορά το 1959 από τον *Abe Sklar* πριν από εξήντα επτά χρόνια. Αναφορά της θεωρίας σύζευξης συναντάμε και από τον *Hoeffding* κατά τη διάρκεια του 1940. Το πρόβλημα της κατασκευής διδιάστατων κατανομών με δεδομένες περιθώριες ήταν ένα δημοφιλές θέμα στο πρώτο τρίτο του εικοστού αιώνα, όταν οι στατιστικοί αναζητούσαν κατανομές με προσιτό αναλυτικό τύπο, ώστε να περιγράψουν πλήρως μη κανονικά συσχετιζόμενα δεδομένα. Η θεωρία αυτή περιπλανήθηκε για τρεις δεκαετίες στην αφάνεια των θεωρητικών στατιστικών πριν από την επαναπιστοποίηση της και αναδύθηκε ως ένα βασικό αναλυτικό εργαλείο στην παγκόσμια σκηνή του χρηματοπιστωτικού τομέα όταν η ανάπτυξη των υπολογιστών κατέστησε εφικτή την επίλυση προβλημάτων, που στο παρελθόν παρουσίαζαν υπολογιστικές δυσκολίες, με ιδιαίτερη χρησιμότητα στην δομή μοντελοποίησης της εξάρτησης μεταξύ δύο ομάδων τυχαίων μεταβλητών, όπως η εκτίμηση του κινδύνου, τα πιστωτικά μοντέλα, οι τιμές παραγώγων και διαχείρισης χαρτοφυλακίου, για να αναφέρουμε μόνο μερικά. Μπορούν να εφαρμοσθούν στην τιμολόγηση παραγώγων (*Black-Scholes*), στη διαχείριση χαρτοφυλακίου, στην εκτίμηση μέτρων κινδύνου (π.χ. τιμή ρίσκου ή αναμενόμενη αποτυχία), στη συσσώρευση κινδύνου, στη μοντελοποίηση πιστώσεων και σε κάθε είδους πρόβλημα που περιλαμβάνει περισσότερες από μία τυχαίες μεταβλητές. Παρά το γεγονός ότι η πλειοψηφία της βιβλιογραφίας είναι αφιερωμένη στην εφαρμογή των συζεύξειων στο χρηματοπιστωτικό τομέα, οι εφαρμογές της θεωρίας δεν περιορίζονται μόνο στον οικονομικό κόσμο. Κάθε κατάσταση που αφορά περισσότερες από μία τυχαίες μεταβλητές μπορεί να μοντελοποιηθεί και να αναλυθεί με τη βοήθεια της θεωρίας της σύζευξης - παρόλο που όπως συνήθως συμβαίνει με τη στατιστική,όσο περισσότερες μεταβλητές υπάρχουν στο μοντέλο, τόσο πιο πολύπλοκη και χρονοβόρα διαδικασία γίνεται η ανάλυση.

Οι στατιστικολόγοι έδειξαν από πολύ νωρίς ενδιαφέρον για τη σχέση που υπάρχει μεταξύ μιας πολυδιάστατης συνάρτησης κατανομής και των μικρότερης τάξης περιθώριων κατανομών. Αρχικά, κατά τη διάρκεια της δεκαετίας του 1950, οι *M. Frechet* και *C. Dall' Aglio* έκαναν σημαντική δουλειά πάνω σε αυτό το θέμα, μελετώντας τη διδιάστατη και τριδιάστατη συνάρτηση κατανομής, με δεδομένες μονοδιάστατες περιθώριες. Η έννοια των συζεύξεων δόθηκε για πρώτη φορά το 1959 από τον *A. Sklar*, όταν απάντησε στην ερώτηση που του τέθηκε από τον *M. Frechet* για τη σχέση μεταξύ της πολυδιάστατης συνάρτησης πιθανότητας και των μικρότερης διάστασης περιθώριων κατανομών. Εν συντομία, αυτό που απέδειξε ο *A. Sklar* (1959) ήταν ότι αν *H* είναι μια διδιάστατη συνάρτηση κατανομής με περιθώριες $F(x)$ και $G(y)$, τότε υπάρχει μια σύζευξη *C*, την οποία αρχικά μπορούμε να θεωρήσουμε σαν συνάρτηση, έτσι ώστε $H(x,y) = C(F(x),G(y))$. Οι *Kimeldorf* και *Sampson* (1975) αναφέρονται στις συζεύξεις θεωρώντας' τες ομοιόμορφες απεικονίσεις, ενώ οι *Galambos* και *Deheuvels* (1978) αναφέρονται σε αυτές ως εξαρτημένες συναρτήσεις. Μεταξύ του 1959 και 1976, τα περισσότερα αποτελέσματα προέκυψαν στη διάρκεια της ανάπτυξης του μετρικού χώρου πιθανοτήτων, κυρίως κατά τη μελέτη των δυαδικών εφαρμογών σε χώρους συναρτήσεων κατανομών πιθανοτήτων. Αργότερα, οι στατιστικολόγοι ενδιαφέρθηκαν να ορίσουν μη παραμετρικά μέτρα εξάρτησης μεταξύ των τυχαίων μεταβλητών, τα οποία έκτοτε έπαιξαν σημαντικό ρόλο στις πιθανότητες και τη μαθηματική στατιστική. Το 1942 ο *Karl Menger* προτείνει μια πιθανοθεωρητική γενίκευση της θεωρίας των μετρήσιμων χώρων, με την αντικατάσταση της απόστασης $d(p,q)$ σε μια συνάρτηση κατανομής F_{pq}, όπου η τιμή $F_{pq}(x)$ για κάθε πραγματικό x είναι η πιθανότητα η απόσταση μεταξύ του p και q να είναι μικρότερη από x. Η σημαντικότερη δυσκολία στην κατασκευή του μετρικού χώρου πιθανοτήτων εμφανίζεται όταν κάποιος προσπαθεί να βρει την πιθανοθεωρητική αναλογία της τριγωνικής ανισότητας, δηλαδή να βρει ποια είναι η αντίστοιχη σχέση με τις συναρτήσεις κατανομών F_{pq}, F_{qr} και F_{pr} για όλα τα p, q και r που ανήκουν σε ένα σύνολο S. Ο *Menger* πρότεινε ως λύση αυτού του προβλήματος τη σχέση $F_{pr}(x+y) \geq T(F_{pq}(x), F_{qr}(y))$, όπου T είναι μια τριγωνική νόρμα, η *t-norm*.

ΚΕΦΑΛΑΙΟ 2

2 Ορισμοί και Βασικές Ιδιότητες

2.1 Πολυδιάστατες συζεύξεις

2.1.1 Ορισμοί

Ορισμός

Ονομάζουμε *n*-θέσια (*n-place*) μια συνάρτηση της οποίας το πεδίο ορισμού *DomH* είναι υποσύνολο του $\overline{\mathbf{R}}^n$ και σύνολο τιμών *RanH* είναι υποσύνολο του $\mathbf{R}$.

Ορισμός

Έστω $S_1,\ldots,S_n$ μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ όπου $\overline{\mathbf{R}}$ η προεκτεταμένη πραγματική ευθεία $[-\infty,+\infty]$ και έστω *H* μια πραγματική *n*-place συνάρτηση τέτοια ώστε το πεδίο ορισμού της να είναι $DomH = S_1 \times \cdots \times S_n$ και για $a \le b$ ($a_k \le b_k$ για όλα τα *k*) έστω $B=[a,b]$ $(=[a_1,b_1]\times\cdots\times[a_n,b_n])$ ένα n-ορθογώνιο με κορυφές που βρίσκονται στο *DomH*.Τότε ο *H*-όγκος του *B* δίνεται από τον τύπο

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

όπου το άθροισμα Σ υπολογίζεται πάνω στις κορυφές *c* του *B* και το *sgn(c)* ισούται με

$$\operatorname{sgn}(c)=\begin{cases} 1\ , \alpha\nu\ \ c_k = a_k\ \gamma\iota\alpha\ k\ \acute{\alpha}\rho\tau\iota o \\ -1\ , \alpha\nu\ \ c_k = a_k\ \gamma\iota\alpha\ k\ \pi\varepsilon\rho\iota\tau\tau\acute{o} \end{cases}$$

Δηλαδή ο *H*-όγκος του *B* ισούται με την *n*-τάξης διαφορά της *H*

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

όπου οι *n*-τάξης διαφορές της *H* ορίζονται ως

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

Ορισμός

Μία πραγματική n-place συνάρτηση H καλείται n-αύξουσα εάν $V_H\left(B\right)\geq 0$ για όλα τα n-ορθογώνια B των οπίων οι κορυφές βρίσκονται στο $DomH$.

Υποθέτουμε τώρα ότι το πεδίο ορισμού μιας n-place πραγματικής συνάρτησης H είναι $DomH= S_1\times\cdots\times S_n$ και οτι κάθε S_k έχει ένα ελάχιστο στοιχείο a_k. Θα λέμε ότι η H είναι εδραιωμένη (*grounded*) αν $H(t)=0$ για κάθε $t\in DomH$ ώστε $t_k=a_k$ για τουλάχιστον ένα k. Αν S_k είναι μη-κενό και έχει μέγιστο στοιχείο b_k τότε η H έχει περιθώριες. Ειδικότερα οι μονοδιάστατες περιθώριες τις H, H_k έχουν πεδίο ορισμού $Dom\ H_k=S_k$ και ισχύει :

$$H_k\left(x\right)=H\left(b_1,...,b_{k-1},x,b_{k+1},...b_n\right)$$

για κάθε $x\in S_k$.

Λήμμα

Έστω $S_1,...,S_n$ μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ και H μια εδραιωμένη και n-αύξουσα συνάρτηση με πεδίο ορισμού $S_1\times\cdots\times S_n$. Τότε η H είναι αύξουσα δηλαδή αν $\left(t_1,...,t_{k-1},x,t_{k+1},...,t_n\right)$ και $\left(t_1,...,t_{k-1},y,t_{k+1},...,t_n\right)$ βρίσκονται στο $DomH$ με $x\leq y$ τότε $H\left(t_1,...,t_{k-1},x,t_{k+1},...,t_n\right)\leq H\left(t_1,...,t_{k-1},y,t_{k+1},...,t_n\right)$.

Λήμμα

Έστω $S_1,...,S_n$ μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ και H μια εδραιωμένη και n-αύξουσα συνάρτηση με πεδίο ορισμού $S_1\times\cdots\times S_n$.Τότε αν $x=(x_1,...,x_n)$ και $y=(y_1,...,y_n)$ σημεία του $S_1\times\cdots\times S_n$,

$$\left|H(x)-H(y)\right|\leq\sum_{k=1}^{n}\left|H_k(x_k)-H_k(y_k)\right|.$$

Ορισμός

Μία n-διάστατη συνάρτηση κατανομής είναι μια συνάρτηση H με πεδίο ορισμού $S_1\times\cdots\times S_n$ τέτοια ώστε H είναι εδραιωμένη, n-αύξουσα και $H(\infty,...,\infty)=1$.

Ορισμός

Μία *n*-διάστατη σύζευξη (*copula*) είναι μια συνάρτηση *C* με πεδίο ορισμού $[0,1]^n$ τέτοια ώστε

1.*C* εδραιωμένη και *n*-αύξουσα

2.*C* έχει περιθώριες C_k για k=1,2,…,n για τις οποίες $C_k(u)=u$ για κάθε $u \in [0,1]$

Ισοδύναμα η συνάρτηση *C* είναι μία συνάρτηση από το $[0,1]^n$ στο $[0,1]$ με τις παρακάτω ιδιότητες:

1. για κάθε *u* στο $[0,1]^n$, $C(u)=0$ αν τουλάχιστον μια συντεταγμένη του *u* είναι 0 και $C(u)=u_k$ αν όλες οι συντεταγμένες του u είναι 1 εκτός από την u_k.

2. για κάθε *a,b* στο $[0,1]^n$ με $a_i \le b_i$ για όλα τα i ισχύει $V_C([a,b]) \ge 0$.

Θεώρημα

Έστω *C* μία *n*-σύζευξη.Τότε για κάθε *v, u* στο $[0,1]^n$ ισχύει

$$|C(v)-C(u)| \le \sum_{k=1}^{n} |v_k - u_k|$$

και επομένως η *C* ομοιόμορφα συνεχής στο $[0,1]^n$.

Θεώρημα του Sklar

Έστω *H* μία *n*-διάστατη συνάρτηση κατανομής με περιθώριες $F_1,\ldots, F_n$. Τότε υπάρχει μια n-σύζευξη C τέτοια ώστε για κάθε $x \in \overline{\mathbf{R}}^n$,

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

Αν $F_1,\ldots, F_n$ όλες συνεχείς, τότε η *C* είναι μοναδική, διαφορετικά η *C* είναι μοναδικά ορισμένη στο $RanF_1 \times \cdots \times RanF_n$. Αντίστροφα αν *C* είναι μία *n*-σύζευξη και $F_1,\ldots, F_n$ συναρτήσεις κατανομών, τότε η συνάρτηση *H* όπως αυτή ορίζεται παραπάνω είναι μία n-διάστατη συνάρτηση κατανομής με περιθώριες $F_1,\ldots, F_n$.

Από το θεώρημα του Sklar βλέπουμε ότι η δομή εξάρτησης μεταξύ συνεχών πολυδιάστατων συναρτήσεων κατανομών και μονοδιάστατων περιθώριων συναρτήσεων, μπορεί να εκφραστεί μαθηματικά με έναν σύνδεσμο (*copula*).

Πόρισμα

Έστω H μία n-διάστατη συνάρτηση κατανομής με συνεχείς περιθώριες συναρτήσεις $F_1,\dots, F_n$ και σύζευξη C τέτοια ώστε $H(x_1,...,x_n)=C(F_1(x_1),...,F_n(x_n))$. Τότε για κάθε $u\in[0,1]^n$,

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

όπου $F^{-1}(t)$ η γενικευμένη αντίστροφη (*generalized inverse*) της F μονοδιάστατης συνάρτησης κατανομής με $F^{-1}(t)=\inf\{x\in\mathbf{R}\,|\,F(x)\ge t\}$ για κάθε u στο $[0,1]^n$.

2.1.2 Τα Fréchet-Hoeffding όρια για από κοινού συναρτήσεις κατανομών

Έστω οι συναρτήσεις M^n, Π^n και W^n ορισμένες στο $[0,1]^n$ ως εξής :

$$\begin{aligned}&M^n(u)=\min(u_1,...,u_n),\\&\Pi^n(u)=u_1...u_n,\\&W^n(u)=\max(u_1+\cdots+u_n-n+1,0).\end{aligned}$$

Οι συναρτήσεις M^n και Π^n είναι n-σύνδεσμοι για $n\ge 2$, ενώ η συνάρτηση W^n δεν είναι n-σύνδεσμος για $n\ge 3$.

Θεώρημα

Αν C οποιαδήποτε n-σύζευξη, τότε για κάθε $u\in[0,1]^n$,

$$W^n(u)\le C(u)\le M^n(u).$$

Θεώρημα

Για κάθε $n \geq 3$ και κάθε $u \in [0,1]^n$, υπάρχει n-σύζευξη C εξαρτημένη από το u τέτοια ώστε

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

2.1.3 Πολυδιάστατες Συζεύξεις και τυχαίες μεταβλητές

Έστω $X_1,..., X_n$ τυχαίες μεταβλητές με συνεχείς συναρτήσεις κατανομής $F_1,..., F_n$ και από κοινού συνάρτηση κατανομής H. Τότε το τυχαίο διάνυσμα $(X_1,...,X_n)^T$ έχει μοναδικό σύνδεσμο C που δίνεται από τη σχέση $H(x_1,...,x_n) = C(F_1(x_1),...,F_n(x_n))$. Επομένως η έκφραση της σύζευξης της κατανομής του τυχαίου διανύσματος $(X_1,...,X_n)^T$ γίνεται :

$$H(x_1,...,x_n) = P\left\{X_1 \leq x_1,...,X_n \leq x_n\right\} = C(F_1(x_1),...,F_n(x_n))$$

Θεώρημα

Έστω διάνυσμα $(X_1,...,X_n)^T$ τυχαίων μεταβλητών με σύνδεσμο C, τότε $X_1,..., X_n$ είναι ανεξάρτητες αν και μόνο αν $C = \Pi^n$.

Θεώρημα

Έστω διάνυσμα $(X_1,...,X_n)^T$ τυχαίων μεταβλητών με σύνδεσμο C. Αν $a_1,...,a_n$ είναι αυστηρά αύξουσες συναρτήσεις (*strictly increasing*) στα σύνολα τιμών $RanX_1,...,RanX_n$ αντίστοιχα, τότε το διάνυσμα $(a_1(X_1),...,a_n(X_n))^T$ έχει κι αυτό τον σύνδεσμο C.

Απόδειξη

Έστω $F_1,..., F_n$ οι συναρτήσεις κατανομής των $X_1,..., X_n$ και $G_1,..., G_n$ οι συναρτήσεις κατανομής των $a_1(X_1),...,a_n(X_n)$ αντίστοιχα. Έστω ότι το $(X_1,...,X_n)^T$ έχει σύνδεσμο C και το $(a_1(X_1),...,a_n(X_n))^T$ έχει σύνδεσμο C_α. Αφού α_k είναι αυστηρά αύξουσα για κάθε k,

$$G_k(x) = P\left\{a_k(X_k) \leq x\right\} = P\left\{X_k \leq a_k^{-1}(x)\right\} = F_k(a_k^{-1}(x))$$

για κάθε $x \in \overline{\mathbf{R}}$ οπότε,

$$\begin{aligned} C_a(G_1(x_1),...,G_n(x_n)) &= P\left\{a_1(X_1) \leq x_1,...,a_n(X_n) \leq x_n\right\} \\ &= P\left\{X_1 \leq a_1^{-1}(x_1),...,X_n \leq a_n^{-1}(x_n)\right\} \\ &= C(F_1(a_1^{-1}(x_1)),....,F_n(a_n^{-1}(x_n))) \\ &= C(G_1(x_1),...,G_n(x_n)). \end{aligned}$$

Επειδή $X_1,..., X_n$ συνεχείς τυχαίες μεταβλητές και $RanG_1 = \cdots = RanG_n = [0,1]$ έπεται ότι C_a=C στο $\left[0,1\right]^n$.

2.2 Διδιάστατες συζεύξεις

2.2.1 Ορισμοί και Ιδιότητες

Ορισμός

Ονομάζουμε 2-θέσια (*2-place*) μια συνάρτηση της οποίας το πεδίο ορισμού *DomH* είναι υποσύνολο του $\overline{\mathbf{R}}^2$ και σύνολο τιμών *RanH* είναι υποσύνολο του $\mathbf{R}$.

Ορισμός

Έστω S_1 και S_2 μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ και έστω H μια πραγματική 2-place συνάρτηση τέτοια ώστε το πεδίο ορισμού της να είναι $DomH = S_1 \times S_2$ και έστω B=$\left[x_1, y_1\right] \times \left[x_2, y_2\right]$ ένα n-ορθογώνιο με κορυφές που βρίσκονται στο *DomH*.Τότε ο H-όγκος του B δίνεται από τον τύπο

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

και αν υποθέσουμε ότι

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

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

Δηλαδή ο H-όγκος του B θα είναι η δεύτερης τάξης διαφορά της H στο B.

Ορισμός

Μία 2-place συνάρτηση είναι 2-αύξουσα, αν $V_H(B) \geq 0$ για κάθε ορθογώνιο Β του οποίου οι κορυφές βρίσκονται στο πεδίο ορισμού της *H*.

Ο όρος «μία 2-place συνάρτηση είναι 2-αύξουσα» δεν προκύπτει από το γεγονός ότι η συνάρτηση αυτή είναι αύξουσα. Πράγματι αν πάρουμε την συνάρτηση $H(x,y) = \max(x,y)$ ορισμένη στο I^2, διαπιστώνουμε ότι ενώ η *H* είναι αύξουσα ως προς *x* και *y*,

$$V_H(I^2) = \max(1,1) - \max(1,0) - \max(0,1) + \max(0,0) = -1$$

και επομένως η *H* δεν είναι 2-αύξουσα.

Λήμμα

Έστω S_1 και S_2 μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ και έστω *H* μια 2- αύξουσα συνάρτηση τέτοια ώστε το πεδίο ορισμού της να είναι $DomH = S_1 \times S_2$ και έστω $x_1, x_2 \in S_1$ και $y_1, y_2 \in S_2$ με $x_1 \leq x_2$ και $y_1 \leq y_2$. Τότε η συνάρτηση είναι αύξουσα στο S_1 και η συνάρτηση $t \mapsto H(x_2,t) - H(x_1,t)$ είναι αύξουσα στο S_2.

Λήμμα

Έστω S_1 και S_2 μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ και έστω *H* μια 2- αύξουσα συνάρτηση και εδραιωμένη (δηλαδή αν a_1 ελάχιστο στοιχείο του S_1 και a_2 ελάχιστο στοιχείο του S_2 ισχύει $H(x,a_2) = 0 = H(a_1,y)$ για κάθε (x,y) στο $S_1 \times S_2$) με πεδίο ορισμού το $S_1 \times S_2$. Τότε η *H* είναι αύξουσα για κάθε (x,y) στο $S_1 \times S_2$.

Αν τώρα το S_1 έχει μέγιστο στοιχείο b_1 και το S_2 έχει μέγιστο στοιχείο b_2, τότε η συνάρτηση $H: S_1 \times S_2 \to \mathbf{R}$ έχει περιθώριες συναρτήσεις *F*, *G* τέτοιες ώστε,

$$DomF = S_1 \text{ και } F(x) = H(x,b_2) \text{ για κάθε } x \in S_1$$

$$DomG = S_2 \text{ και } G(y) = H(b_1,y) \text{ για κάθε } y \in S_2$$

Λήμμα

Έστω S_1 και S_2 μη-κενά υποσύνολα του $\overline{\mathbf{R}}$ και έστω H μια 2- αύξουσα συνάρτηση με περιθώριες οι οποίες έχουν πεδίο ορισμού το $S_1 \times S_2$. Επιπλέον έστω (x_1, y_1) και (x_2, y_2) ζευγάρια στο $S_1 \times S_2$. Τότε,

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

Ορισμός

Μία συνάρτηση C καλείται 2-σύζευξη όταν έχει πεδίο ορισμού το I^2 και σύνολο τιμών το I και ισχύει

1. για κάθε u,v στο I, $C(u,0) = 0 = C(0,v)$ και $C(u,1) = u, C(1,v) = v$
2. για κάθε u_1, u_2, v_1, v_2 στο I με $u_1 \leq u_2$ και $v_1 \leq v_2$,

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

Επομένως η C είναι φραγμένη και 2-αύξουσα.

Θεώρημα

Έστω C σύζευξη. Τότε για κάθε u_1, u_2, v_1, v_2 στο $DomC$,

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

Επομένως η C ομοιόμορφα συνεχής στο $DomC$.

Από τον παραπάνω τύπο,θέτοντας u_1=u_2 και v_1=v_2 προκύπτει το παρακάτω θεώρημα.

Θεώρημα

Έστω C σύζευξη. Για κάθε v στο I υπάρχει η μερική παράγωγος $\dfrac{\partial C(u,v)}{\partial u}$ για σχεδόν όλα τα u και για τέτοια u και v ισχύει,

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

Ομοίως για κάθε u στο I υπάρχει η μερική παράγωγος $\dfrac{\partial C(u,v)}{\partial v}$ για σχεδόν όλα τα v και για τέτοια u και v ισχύει,

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

Θεώρημα του Sklar

Έστω *H* μία από κοινού συνάρτηση κατανομής με περιθώριες *F* και *G*. Τότε υπάρχει σύζευξη *C* ώστε για κάθε $x, y \in \overline{\mathbf{R}}$,

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

Αν *F, G* συνεχείς, τότε η *C* είναι μοναδική, διαφορετικά η *C* είναι μοναδικά ορισμένη στο $RanF \times RanG$. Αντίστροφα αν *C* είναι μία σύζευξη και *F,G* συναρτήσεις κατανομών, τότε η συνάρτηση *H* όπως αυτή ορίζεται παραπάνω είναι μία από κοινού συνάρτηση κατανομής με περιθώριες *F,G*.

Από τον παραπάνω τύπο και με βάση την γενικευμένη αντίστροφη μιας συνάρτησης, που ορίσαμε σε προηγούμενη ενότητα μπορούμε να βρούμε την σύζευξη C.Ωστόσο σε κάποιες πηγές ορίζεται η έννοια της «κάτι σαν» αντίστροφη (*quasi*-αντίστροφης).

Ορισμός

Έστω *F* συνάρτηση κατανομής. Τότε η *quasi*-αντίστροφη είναι μία συνάρτηση $F^{(-1)}$ με πεδίο ορισμού το *I*, τέτοια ώστε

1. αν το *t* ανήκει στο σύνολο τιμών της *F*, *RanF*, τότε η $F^{(-1)}(t)$ είναι κάθε αριθμός *x* στο $\overline{\mathbf{R}}$ έτσι ώστε $F(x)=t$ για κάθε *t* στο *RanF*
2. αν το *t* δεν ανήκει στο *RanF*, τότε

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

Ειδικότερα, στην περίπτωση που η *F* είναι αυστηρά αύξουσα, ο ορισμός της quasi-αντίστροφης οδηγεί στην συνήθη αντίστροφη συνάρτηση της *F*. Από τον παραπάνω ορισμό και το θεώρημα του *Sklar* προκύπτει ότι

$$C(x, y) = H(F^{(-1)}(x), G^{(-1)}(y))$$

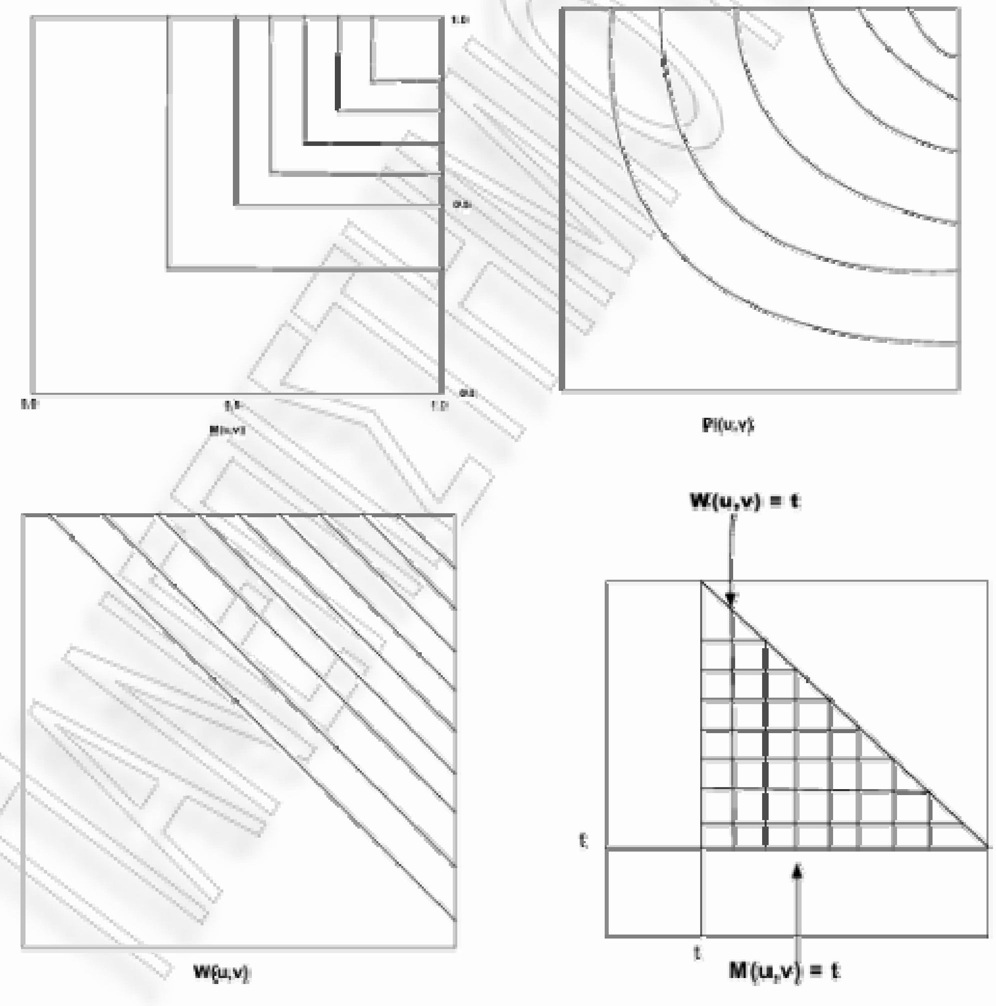
Θεώρημα

Έστω *C* σύζευξη. Τότε για κάθε *(u,v)* στο πεδίο ορισμού της *C* ισχύει

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

Τα όρια της ανισότητας είναι συζεύξεις και θα τα συμβολίζουμε με $W(u,v)$ και $M(u,v)$ αντίστοιχα οπότε προκύπτει η ανισότητα **Fréchet-Hoeffding** όπου με W θα συμβολίζουμε το *Fréchet-Hoeffding* κάτω φράγμα και με M θα συμβολίζουμε το *Fréchet-Hoeffding* άνω φράγμα οπότε,

$$W(u,v) \le C(u,v) \le M(u,v).$$



Ορισμός

Αν C_1 και C_2 είναι συζεύξεις, C_1 μικρότερη της C_2 ($\mathrm{C}_1 \prec \mathrm{C}_2$) αν $\mathrm{C}_1(u) \le \mathrm{C}_2(u)$ για όλα τα $\mathrm{u} \in [0,1]^n$.

Το *Fréchet-Hoeffding* κατώτερο όριο W^2 (για τη διδιάστατη περίπτωση) είναι μικρότερο από κάθε 2-σύζευξη και κάθε *n*-σύζευξη είναι μικρότερη από το *Fréchet-Hoeffding* ανώτερο όριο M^n

2.2.2 Διδιάστατες Συζεύξεις και τυχαίες μεταβλητές

Λήμμα

Έστω X και Y τυχαίες μεταβλητές με κοινή συνάρτηση κατανομής H.Τότε η H είναι το *Fréchet-Hoeffding* άνω φράγμα $M(u,v)$ αν και μόνο αν για κάθε $(x,y) \in \overline{\mathbf{R}}^2$,

$$P[X > x, Y \le y] = 0 \text{ ή } P[X \le x, Y > y] = 0.$$

Απόδειξη

Έστω F και G οι συναρτήσεις κατανομών των μεταβλητών X και Y αντίστοιχα.

$$F(x) = P[X \le x] = P[X \le x, Y \le y] + P[X \le x, Y > y] = H(x,y) + P[X \le x, Y > y]$$

και

$$G(y) = P[Y \le y] = P[X \le x, Y \le y] + P[X > x, Y \le y] = H(x,y) + P[X > x, Y \le y]$$

Οπότε

$$H(x,y) = M(F(x), G(y)) = \min(H(x,y) + P[X \le x, Y > y], H(x,y) + P[X > x, Y \le y])$$

αν και μόνο αν

$$\min(P[X \le x, Y > y], P[X > x, Y \le y]) = 0.$$

Θεώρημα

Έστω X και Y συνεχείς τυχαίες μεταβλητές με σύζευξη $C_{X,Y}$. Έστω τώρα a και b αυστηρά αύξουσες συναρτήσεις στο $RanG \times RanF$. Τότε $C_{a(X),b(Y)} = C_{X,Y}$ και επομένως η $C_{X,Y}$ είναι αναλλοίωτη (*invariant*) κάτω από αυστηρά αύξον μετασχηματισμό των *X* και *Y*.

[Η απόδειξη είναι ίδια με την πολυδιάστατη περίπτωση που αναπτύξαμε σε προηγούμενη ενότητα.]

Βλέπουμε λοιπόν ότι οι 2-συζεύξεις είναι **ομοιόμορφα συνεχείς**, **διαφορίσιμες** και **αναλλοίωτες** υπό συγκεκριμένους μετασχηματισμούς των τ.μ. *X* και *Y*.

Θεώρημα

Έστω X και Y συνεχείς τυχαίες μεταβλητές με σύζευξη $C_{X,Y}$. Έστω τώρα a και b αυστηρά μονότονες συναρτήσεις στο $RanG \times RanF$.Τότε,

Αν a αυστηρά αύξουσα και b αυστηρά φθίνουσα τότε

$$C_{a(X),b(Y)} = u - C_{X,Y}(u, 1-v)$$

1. Αν b αυστηρά αύξουσα και a αυστηρά φθίνουσα τότε

$$C_{a(X),b(Y)} = v - C_{X,Y}(1-u, v)$$

2. Αν a και b αυστηρά φθίνουσες τότε

$$C_{a(X),b(Y)} = u + v - C_{X,Y}(1-u, 1-v)$$

Απόδειξη

Ας υποθέσουμε ότι οι *X* και *Y* ακολουθούν τις κατανομές *F* και *G* ενώ *a(X)* και *b(X)* ακολουθούν τις κατανομές L και *M* με *a* και *b* μονότονες συναρτήσεις έτσι ώστε,

$$C_{a(X),b(Y)}(L(X), M(Y)) = P\{a(X) \leq x, b(Y) \leq y\}$$

Επομένως όταν *b* αυστηρά αύξουσα και *a* αυστηρά φθίνουσα,

$$
\begin{aligned}
C_{a(X),b(Y)} &= P\left\{X \geq a^{-1}(x), b(Y) \leq y\right\} \\
&= P\left\{b(Y) \leq y\right\} - P\left\{X \leq a^{-1}(x), b(Y) \leq y\right\} \\
&= C_{b(Y)}(M(Y)) - C_{X,b(Y)}(F(a^{-1}(x)), M(Y)) \\
&= C_{b(Y)}(M(Y)) - C_{X,b(Y)}(1 - L(x), M(Y)) \\
&= v - C_{X,b(Y)}(1-u, v) \\
&= v - C_{X,Y}(1-u, v)
\end{aligned}
$$

Ομοίως αποδεικνύεται ότι αν *α* αυστηρά αύξουσα και *b* αυστηρά φθίνουσα τότε

$$C_{a(X),b(Y)} = u - C_{X,Y}(u, 1-v)$$

Στην περίπτωση που *α* και *b* αυστηρά φθίνουσες ισχύει

$$
\begin{aligned}
C_{a(X),b(Y)} &= P\left\{X \geq a^{-1}(x), Y \geq b^{-1}(y)\right\} \\
&= 1 - F(a^{-1}(x)) - G(b^{-1}(y)) + C_{X,Y}(F(a^{-1}(x)), G(b^{-1}(y))) \\
&= 1 - (1-u) - (1-v) + C_{X,Y}(1-u, 1-v) \\
&= u + v - 1 + C_{X,Y}(1-u, 1-v)
\end{aligned}
$$

2.3 Συζεύξεις Επιβίωσης και Συμμετρία

2.3.1 Συζεύξεις Επιβίωσης

Σε αυτό το κομμάτι θα αναφερθούμε στις συζεύξεις επιβίωσης οι οποίες έχουν αρκετές ενδιαφέρουσες ιδιότητες και πολλές εφαρμογές στη στατιστική.

Η πιθανότητα ένα άτομο ή μηχάνημα να επιζήσει πέρα του χρόνου x δίνεται από τη συνάρτηση επιβίωσης

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

Όπου *F* η συνάρτηση κατανομής της *X*.

Για τυχαίες μεταβλητές *X* και *Y* με από κοινού συνάρτηση κατανομής *H*, η από κοινού συνάρτηση επιβίωσης δίνεται ως εξής

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

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

Θεωρώντας ότι η *X* και *Y* έχουν σύζευξη *C*, έχουμε

$$\begin{aligned}
\overline{H}(x, y) &= P(X > x, Y > y) \\
&= 1 - P(X \le x \text{ ή } Y \le y) \\
&= 1 - F(x) - G(y) + H(x, y) \\
&= \overline{F}(x) + \overline{G}(y) - 1 + C(F(x), G(y)) \\
&= \overline{F}(x) + \overline{G}(y) - 1 + C(1 - \overline{F}(x), 1 - \overline{G}(y))
\end{aligned}$$

Ορίζουμε τη συνάρτηση $\hat{C} : [0,1] \times [0,1] \to [0,1]$ ως

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

και επομένως

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

Η συνάρτηση $\hat{C}$ αποτελεί την σύζευξη επιβίωσης των *X* και *Y* και συνδέει την από κοινού συνάρτηση επιβίωσης με τις μονοδιάστατες περιθώριές της όπως και μία σύζευξη συνδέει την από κοινού συνάρτηση κατανομής με τις αντίστοιχες περιθώριες.

Προσοχή. Η σύζευξη επιβίωσης $\hat{C}$ **δεν** είναι η συνάρτηση επιβίωσης της σύζευξης C. Η σχέση η οποία συνδέει τις 2 αυτές συναρτήσεις είναι

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

2.3.2 Συμμετρία

Στην ενότητα αυτή θα παραθέσουμε κάποιες ιδιότητες κατανομών και ιδιαίτερα διδιάστατων ώστε να οδηγηθούμε σε κάποιες σημαντικές ιδιότητες των συζεύξεων.

Αν X τυχαία μεταβλητή και α ένας πραγματικός αριθμός, λέμε ότι η X είναι συμμετρική γύρω από το a αν για κάθε x στο $\mathbf{R}$ έχουμε

$$P[X - a \le x] = P[a - X \le x]$$

και επομένως οι $X - a$ και $a - X$ ακολουθούν την ίδια κατανομή.

Όταν ασχολούμαστε με διδιάστατες κατανομές, υπάρχουν διάφορα είδη συμμετρίας που εξαρτώνται από τον τρόπο με τον οποίο συνδέονται μεταξύ τους οι μεταβλητές. Έστω X και Y τυχαίες μεταβλητές και (a,b) σημείο στο $\mathbf{R}^2$. Τότε

1. Το διάνυσμα (X,Y) θα είναι οριακά συμμετρικό (*marginally symmetric*) γύρω από το (a,b), αν X και Y συμμετρικές γύρω από τα a και b αντίστοιχα
2. (X,Y) θα είναι ακτινωτά συμμετρικό (*radially symmetric*) γύρω από το (a,b), αν η από κοινού συνάρτηση κατανομής των $X - a$ και $Y - b$ είναι η ίδια με την από κοινού συνάρτηση κατανομής των $a - X$ και $b - Y$
3. (X,Y) θα είναι από κοινού συμμετρικό (*jointly symmetric*) γύρω από το (a,b), αν τα ζεύγη $(X-a, Y-b), (X-a, b-Y), (a-X, Y-b), (a-X, b-Y)$ έχουν κοινή συνάρτηση κατανομής.

Θεώρημα

Έστω X και Y συνεχείς τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής H και περιθώριες F και G αντίστοιχα. Έστω (a,b) σημείο στο $\mathbf{R}^2$. Τότε το (X,Y) είναι ακτινωτά συμμετρικό γύρω από το (a,b), αν και μόνο αν,

$$H(a+x,b+y)=\overline{H}(a-x,b-y) \text{ για όλα τα } (x,y) \text{ στο } \mathbf{R}^2.$$

Θεώρημα

Έστω X και Y συνεχείς τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής H, περιθώριες F και G αντίστοιχα και σύζευξη C. Επιπλέον έστω ότι X και Y είναι συμμετρικές γύρω από τα a και b αντίστοιχα. Τότε (X,Y) είναι ακτινωτά συμμετρικές γύρω από το (a,b), αν και μόνο αν, $C=\hat{C}$ δηλαδή

$$C(u,v)=u+v-1+C(1-u,1-v)$$

για όλα τα (u,v) στο I^2.

Μία έννοια συνδεδεμένη με τη συμμετρία είναι η εναλλαξιμότητα (*exchangeability*) τυχαίων μεταβλητών. Οι τυχαίες μεταβλητές X και Y θα λέμε ότι είναι εναλλασώμενες (exchangeable) αν τα διανύσματα (X,Y) και (Y,X) κατανέμονται πανομοιότυπα (*identically*).

Θεώρημα

Έστω X και Y συνεχείς τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής H, περιθώριες F και G αντίστοιχα και σύζευξη C. Τότε X και Y είναι εναλλάξιμες αν και μόνο αν $F=G$ και $C(u,v)=C(v,u)$ για όλα τα $(u,v)\in I^2$.

2.4 Εξάρτηση

Οι συζεύξεις παρέχουν ένα φυσικό τρόπο μελέτης και μέτρησης της εξάρτησης μεταξύ τυχαίων μεταβλητών. Ο συντελεστής γραμμικής συσχέτισης χρησιμοποιείται συχνά σαν μέτρο εξάρτησης. Η δημοτικότητά του έγκειται στο γεγονός αφενός ότι υπολογίζεται εύκολα, αφετέρου αποτελεί φυσικό μέτρο εξάρτησης για ελλειπτικές κατανομές όπως η πολυδιάστατη κανονική κατανομή και η πολυδιάστατη κατανομή t. Ωστόσο οι περισσότερες τυχαίες μεταβλητές δεν κατανέμονται «ελλειπτικά» και επομένως η χρήση της γραμμικής συσχέτισης σε τέτοιες περιπτώσεις μπορεί να οδηγήσει σε λανθασμένα αποτελέσματα. Για το λόγο αυτό δεν θα αναφερθούμε περεταίρω στη γραμμική συσχέτιση.

2.4.1 Συμφωνία (*Concordance*)

Έστω $(x,y)^T$ και $(\tilde{x},\tilde{y})^T$ παρατηρήσεις ενός διανύσματος $(X,Y)^T$ από τυχαίες μεταβλητές. Τότε τα $(x,y)^T$ και $(\tilde{x},\tilde{y})^T$ θα λέμε ότι συμφωνούν αν $(x-\tilde{x})(y-\tilde{y})>0$ και ότι δεν συμφωνούν αν $(x-\tilde{x})(y-\tilde{y})<0$.

<u>Θεώρημα</u>

Έστω $(X,Y)^T$ και $(\tilde{X},\tilde{Y})^T$ ανεξάρτητα διανύσματα συνεχών τυχαίων μεταβλητών με από κοινού συναρτήσεις κατανομών H και $\tilde{H}$ αντίστοιχα με περιθώριες F και G. Έστω C και $\tilde{C}$ οι συζεύξεις των $(X,Y)^T$ και $(\tilde{X},\tilde{Y})^T$ ούτως ώστε $H(x,y)=C(F(x),G(y))$ και $\tilde{H}(x,y)=\tilde{C}(F(x),G(y))$. Αν πάρουμε τη διαφορά

$$Q=P\left\{\left(X-\tilde{X}\right)\left(Y-\tilde{Y}\right)>0\right\}-P\left\{\left(X-\tilde{X}\right)\left(Y-\tilde{Y}\right)<0\right\}$$

τότε,

$$Q=Q(C,\tilde{C})=4\iint_{[0,1]^2}\tilde{C}(u,v)dC(u,v)-1$$

<u>Πόρισμα</u>

Έστω $C,\tilde{C}$ και Q όπως αυτά ορίζονται στο παραπάνω θεώρημα. Τότε

1. Q συμμετρική στα ορίσματά της δηλαδή, $Q(C,\tilde{C})=Q(\tilde{C},C)$.
2. Q μη-φθίνουσα σε κάθε όρισμα δηλαδή, αν $C \prec C^{'}$, τότε $Q(C,\tilde{C})\leq Q(C,\tilde{C})$
3. Στην Q οι συζεύξεις μπορούν να αντικατασταθούν από συζεύξεις επιβίωσης δηλαδή $Q(C,\tilde{C})=Q(\hat{C},\hat{\tilde{C}})$

<u>Ορισμός</u>

Ένα αριθμητικό μέτρο εξάρτησης k με πραγματικές τιμές, μεταξύ 2 συνεχών τυχαίων μεταβλητών X και Y με σύζευξη C, θα λέμε ότι είναι μέτρο συμφωνίας αν ισχύουν τα παρακάτω

1. το k ορίζεται για κάθε ζευγάρι X και Y συνεχών τυχαίων μεταβλητών.

2. $-1 \le k_{X,Y} \le 1, k_{X,X} = 1$ και $k_{X,-X} = -1$.

3. $k_{X,Y} = k_{Y,X}$.

4. αν *X* και *Y* ανεξάρτητες, τότε $k_{X,Y} = k_{\Pi} = 0$.

5. $k_{-X,Y} = k_{X,-Y} = -k_{X,Y}$.

6. αν C και $\tilde{C}$ συζεύξεις τέτοιες ώστε $C \prec \tilde{C}$, τότε $k_C \le k_{\tilde{C}}$.

7. αν $\{(X_n, Y_n)\}$ ακολουθία συνεχών τυχαίων μεταβλητών με συζεύξεις C_n και αν η ακολουθία $\{C_n\}$ συγκλίνει σημειακά (pointwise) στο C, τότε $\lim_{n\to\infty} k_{C_n} = k_C$.

Σαν συνέπεια του παραπάνω ορισμού έχουμε ότι αν k μέτρο συμφωνίας των *X* και *Y* και η *Y* είναι αύξουσα συνάρτηση του *X*, τότε $k_{X,Y} = k_M = 1$ ενώ αν η *Y* είναι φθίνουσα συνάρτηση του *X*, τότε $k_{X,Y} = k_W = -1$. Επιπλέον αν a και b αυστηρά αύξουσες συναρτήσεις στο *RanX* και *RanY* αντίστοιχα, τότε $k_{a(X),b(Y)} = k_{X,Y}$.

2.4.2 Μέτρα εξάρτησης

Σε αυτή την ενότητα θα αναφερθούμε μεταξύ των άλλων στα 2 σημαντικότερα μέτρα συσχέτισης με πολλές εφαρμογές κυρίως στη μη-παραμετρική στατιστική, στο τ του *Kendall* και στο ρ του *Spearman*. Αποτελούν ίσως τις καλύτερες εναλλακτικές του συντελεστή γραμμικής συσχέτισης ως μέτρου εξάρτησης για μη-ελλειπτικές κατανομές για τις οποίες ο συντελεστής γραμμικής συσχέτισης δεν παρέχει ασφαλή συμπεράσματα.

<u>Ορισμός</u>

Το τ του *Kendall* για το τυχαίο διάνυσμα $(X,Y)^T$ ορίζεται ως εξής

$$\tau(X,Y) = P\left\{\left(X-\tilde{X}\right)\left(Y-\tilde{Y}\right) > 0\right\} - P\left\{\left(X-\tilde{X}\right)\left(Y-\tilde{Y}\right) < 0\right\}$$

Διαπιστώνουμε λοιπόν ότι το τ του *Kendall* δεν είναι τίποτα άλλο από τη πιθανότητα της συμφωνίας μείον την πιθανότητα της ασυμφωνίας.

Θεώρημα

Έστω $(X,Y)^T$ διάνυσμα συνεχών τυχαίων μεταβλητών με σύζευξη *C*. Τότε το τ του *Kendall* για το $(X,Y)^T$ δίνεται από τον τύπο

$$\tau(X,Y) = Q(C,C) = 4\iint_{[0,1]^2} C(u,v)dC(u,v) - 1$$

Το παραπάνω ολοκλήρωμα είναι η προσδοκώμενη ή μέση τιμή της τυχαίας μεταβλητής $C(U,V)$ όπου $U,V \sim U(0,1)$ με από κοινού συνάρτηση κατανομής *C*.

$$\tau(X,Y) = 4E(C(U,V)) - 1$$

Ορισμός

Το ρ του *Spearman* για το τυχαίο διάνυσμα $(X,Y)^T$ ορίζεται ως εξής

$$\rho_S(X,Y) = 3\left(P\left\{\left(X-\tilde{X}\right)\left(Y-Y'\right) > 0\right\} - P\left\{\left(X-\tilde{X}\right)\left(Y-Y'\right) < 0\right\}\right)$$

Όπου $(X,Y)^T$, $(\tilde{X},\tilde{Y})^T$, $(X',Y')^T$ ανεξάρτητα τυχαία διανύσματα.

Θεώρημα

Έστω $(X,Y)^T$ διάνυσμα συνεχών τυχαίων μεταβλητών με σύζευξη *C*. Τότε το ρ του *Spearman* για το $(X,Y)^T$ δίνεται από τον τύπο

$$\rho_S(X,Y) = 3Q(C,\Pi) = 12\iint_{[0,1]^2} uv\,dC(u,v) - 3 = 12\iint_{[0,1]^2} C(u,v)dudv - 3$$

$$\rho_S(X,Y) = 12\iint_{[0,1]^2} \left(C(u,v) - uv\right) dudv$$

Αν αντικαταστήσουμε την συνάρτηση $\left(C(u,v)-uv\right)$ με την απόλυτή της $\left|C(u,v)-uv\right|$ τότε προκύπτει ένα άλλο μέτρο εξάρτησης, το σ των *Schweizer* και *Wolf*

$$\rho_S(X,Y) = 12\iint_{[0,1]^2} \left|C(u,v)-uv\right| dudv$$

2.4.3 Εξάρτηση Ουράς

Έστω *X* και *Y* συνεχείς τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής *H* και περιθώριες *F* και *G* αντίστοιχα.θα λέμε ότι οι *X* και *Y* είναι εξαρτημένες στη θετική τεταρτοκύκλιο (*positive quadrant dependent*) αν $\mathrm{H}(x,y)-F(x)G(y)\geq 0$ για κάθε $x,y\in\mathbf{R}$. Αν υπάρχει σύζευξη C_{XY} συνδεδεμένη με το ζεύγος *(X,Y)*, τότε θα λέμε ότι οι *X* και *Y* είναι εξαρτημένες στη θετική τεταρτοκύκλιο αν $C_{XY}(u,v)\geq uv$ για όλα τα $u,v\in[0,1]$.

Η εξάρτηση ουράς σχετίζεται με την εξάρτηση στην ουρά του «πάνω δεξιού» τεταρτημόριου (*upper-right-quadrant tail*) ή στην ουρά του «κάτω αριστερού» τεταρτημόριου (*lower-left-quadrant tail*) μιας διδιάστατης κατανομής. Χρησιμοποιείται για τη μελέτη της εξάρτησης μεταξύ ακραίων τιμών (*extreme values*). Η εξάρτηση ουράς μεταξύ 2 τυχαίων μεταβλητών *X* και *Y* εφαρμόζεται στις συζεύξεις και άρα παραμένει αναλλοίωτη υπό αυστηρά αύξοντες μετασχηματισμούς των *X* και *Y*.

<u>Ορισμός</u>

Έστω $(X,Y)^T$ διάνυσμα συνεχών τυχαίων μεταβλητών με περιθώριες συναρτήσεις *F* και *G*. Οι συντελεστές εξάρτησης άνω και κάτω ουράς (*upper and lower tail dependence*) ορίζονται αντίστοιχα ως εξής

$$\lim_{u\to 1^-} P\left\{Y>G^{-1}(u)\middle|X>F^{-1}(u)\right\}=\lambda_U$$
$$\lim_{u\to 0^+} P\left\{Y\leq G^{-1}(u)\middle|X\leq F^{-1}(u)\right\}=\lambda_L$$

Δεδομένου ότι τα παραπάνω όρια υπάρχουν, $\lambda_U,\lambda_L\in[0,1]$.

Για τους συντελεστές λ_U,λ_L ισχύουν τα παρακάτω :

1. Αν $\lambda_U=0$, τότε *X* και *Y* είναι ανεξάρτητες στην άνω ουρά (*upper tail*)
2. Αν $\lambda_U\in(0,1]$, τότε *X* και *Y* είναι εξαρτημένες στην άνω ουρά
3. Αν $\lambda_L=0$, τότε *X* και *Y* είναι ανεξάρτητες στην κάτω ουρά (*lower tail*)
4. Αν $\lambda_L\in(0,1]$, τότε *X* και *Y* είναι εξαρτημένες στην κάτω ουρά

Πόρισμα

Έστω $u \in I$ και C μια 2-σύζευξη, τότε οι συντελεστές εξάρτησης άνω και κάτω ουράς μπορούν να οριστούν ως εξής

$$\lim_{u\to 1^-} \frac{C(u,u)-2u+1}{1-u} = \lambda_U$$
$$\lim_{u\to 0^+} \frac{C(u,u)}{u} = \lambda_L$$

ΚΕΦΑΛΑΙΟ 3

3 Οικογένειες Συζεύξεων

Έχοντας αναπτύξει τους ορισμούς και τις ιδιότητες των συζεύξεων, καθώς και το ρόλο που διαδραματίζουν στην εξάρτηση τυχαίων μεταβλητών, στην ενότητα αυτή θα αναφερθούμε στις σημαντικότερες οικογένειες συζεύξεων παρουσιάζοντας τον τύπο και τις ιδιότητές τους.

3.1 Συζεύξεις Farlie-Gumbel-Morgenstern

Η οικογένεια *Farlie-Gumbel-Morgenstern* είναι αρκετά δημοφιλής και απλή ως προς τη δομή.

Μια διδιάστατη τυχαία μεταβλητή *(X,Y)* με περιθώριες συναρτήσεις κατανομής $F(x)$ και $G(y)$ θα λέμε ότι ανήκει στην οικογένεια **FGM** αν η από κοινού συνάρτηση κατανομής είναι

$$H(x,y) = P(X \leq x, Y \leq y) = F(x)G(y)\{1 + a[1 - F(x)][1 - G(y)]\}$$

για κάθε $x, y \in \mathbf{R}$.

Το α ονομάζεται παράμετρος συνάφειας και πρέπει να ικανοποιεί τη σχέση $-1 \leq a \leq 1$.

Η συσχέτιση των μεταβλητών που ακολουθούν την κατανομή **FGM** εξαρτάται από την παράμετρο *α*. Οι μεταβλητές *X* και *Y* είναι ανεξάρτητες αν *α*=0, θετικά συσχετισμένες αν *α*>0 και αρνητικά συσχετισμένες αν *α*<0.

Η σύζευξη που προκύπτει από την μελέτη της οικογένειας κατανομών **FGM** δίνεται από τον τύπο

$$C(u,v) = uv[1 + a(1-u)(1-v)]$$

με $-1 \leq a \leq 1$ και *u,v* ακολουθούν την ομοιόμορφη κατανομή.

Εξετάζοντας την εξάρτηση μεταξύ των *u,v* βλέπουμε ότι το *τ* του *Kendall* παίρνει την ακόλουθη μορφή.

$$\begin{aligned}\tau &= 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1\\ &= 4E[C(U,V)] - 1\\ &= \frac{2a}{9}\end{aligned}$$

και αφού $-1 \le a \le 1$ έπεται ότι

$$-\frac{2}{9} \le \tau \le \frac{2}{9}.$$

όσον αφορά την εξάρτηση ουράς από τους τύπους

$$\lim_{u\to 1^-}\frac{C(u,u)-2u+1}{1-u} = \lambda_U$$
$$\lim_{u\to 0^+}\frac{C(u,u)}{u} = \lambda_L$$

έπεται ότι $\lambda_U = \lambda_L = 0$ και επομένως οι συζεύξεις της οικογένειας **FGM** δεν έχουν εξάρτηση ουράς.

3.2 Συζεύξεις Marshall-Olkin

Οι συζεύξεις που θα αναλύσουμε παρακάτω ανακαλύφθηκαν από τους *Marshall* και *Olkin* και χρησιμοποιούνται ιδιαίτερα όταν θέλουμε να βρούμε την κοινή κατανομή χρόνων ζωής, όταν οι χρόνοι ζωής αυτοί σχετίζονται ο ένας με τον άλλο. Π.χ όταν μελετάμε χρόνους ζωής ηλεκτρικών λαμπτήρων της ίδιας μάρκας ή ομόλογα εταιρειών στον ίδιο επιχειρηματικό τομέα. Για να μπορέσουμε να μελετήσουμε και να αναλύσουμε αυτές τις συζεύξεις χωρίς πολύπλοκους υπολογισμούς, ξεκινάμε τη διδιάστατη σύζευξη *Marshall-Olkin* και στη συνέχεια κάνουμε μια γενίκευση της πολυδιάστατης περίπτωσης.

3.2.1 Διδιάστατη Σύζευξη Marshall-Olkin

Θεωρούμε X και Y τους χρόνους ζωής 2 στοιχείων, τα οποία υποβάλλονται σε πλήγματα (*shocks*). Επιπλέον ας υποθέσουμε ότι τα πλήγματα ακολουθούν 3 ανεξάρτητες διαδικασίες *Poisson* με παραμέτρους $\lambda_1, \lambda_2, \lambda_{12} \geq 0$ όπου οι δείκτες φανερώνουν το κατά πόσο τα πλήγματα επηρεάζουν μόνο το πρώτο στοιχείο, μόνο το δεύτερο ή και τα 2. Τότε οι χρόνοι Z_1, Z_2 και Z_{12} πραγματοποίησης των πληγμάτων είναι ανεξάρτητες τυχαίες μεταβλητές με εκθετική κατανομή και παραμέτρους λ_1, λ_2 και λ_{12} αντίστοιχα. Άρα,

$$\bar{H}(x,y) = P\{X > x, Y > y\} = P\{Z_1 > x\}P\{Z_2 > y\}P\{Z_{12} > \max(x,y)\}$$

Οι μονοδιάστατες συναρτήσεις επιβίωσης για τις *X* και *Y* είναι

$$\bar{F}(x) = \exp(-(\lambda_1 + \lambda_{12})x) \text{ και } \bar{G}(y) = \exp(-(\lambda_2 + \lambda_{12})y).$$

Επιπλέον αφού $\max(x,y) = x + y - \min(x,y)$,

$$\begin{aligned}\bar{H}(x,y) &= \exp(-(\lambda_1 + \lambda_{12})x - (\lambda_2 + \lambda_{12})y + \lambda_{12}\min(x,y)) \\ &= \bar{F}(x)\bar{G}(y)\min(\exp(\lambda_{12}x), \exp(\lambda_{12}y))\end{aligned}$$

Έστω $a_1 = {\lambda_{12}}/{\lambda_1 + \lambda_{12}}$ και $a_2 = {\lambda_{12}}/{\lambda_2 + \lambda_{12}}$.Τότε $\exp(\lambda_{12}x) = \bar{F}(x)^{-a_1}$ και $\exp(\lambda_{12}y) = \bar{G}(y)^{-a_2}$ και επομένως η σύζευξη επιβίωσης του $(X,Y)^T$ δίνεται από τον τύπο

$$\hat{C}(u,v) = uv\min(u^{-a_1}, v^{-a_2}) = \min(u^{1-a_1}v, uv^{1-a_2})$$

Ο παραπάνω τύπος οδηγεί στη σύζευξη *Marshall-Olkin*

$$C_{a_1,a_2}(u,v) = \min(u^{1-a_1}v, uv^{1-a_2}) = \begin{cases} u^{1-a_1}v, u^{a_1} \geq v^{a_2} \\ uv^{1-a_2}, u^{a_1} \leq v^{a_2} \end{cases}$$

Για να υπολογίσουμε τ του *Kendall* πρέπει να διατυπώσουμε το παρακάτω θεώρημα.

Θεώρημα

Έστω C σύζευξη ώστε το γινόμενο $(\partial C/\partial u)(\partial C/\partial v)$ να είναι ολοκληρώσιμο στο $[0,1]^2$. Τότε

$$\iint_{[0,1]^2} C(u,v)dC(u,v) = \frac{1}{2} - \iint_{[0,1]^2} \frac{\partial}{\partial u}C(u,v)\frac{\partial}{\partial v}C(u,v)dudv$$

Οπότε

$$\begin{aligned}\tau(C_{a_1,a_2}) &= 4\iint_{[0,1]^2} C_{a_1,a_2}(u,v)dC_{a_1,a_2}(u,v) - 1 \\ &= 4\left(\frac{1}{2} - \iint_{[0,1]^2} \frac{\partial}{\partial u}C_{a_1,a_2}(u,v)\frac{\partial}{\partial v}C_{a_1,a_2}(u,v)dudv\right) - 1 \\ &= \frac{a_1 a_2}{a_1 + a_2 - a_1 a_2}\end{aligned}$$

Το ρ του *Spearman* υπολογίζεται ευκολότερα και δίνεται από τον τύπο

$$\begin{aligned}\rho_S(C_{a_1,a_2}) &= 12\iint_{[0,1]^2} C_{a_1,a_2}(u,v)dudv) - 3 \\ &= 12\int_0^1\left(\int_0^{u^{a_1/a_2}} u^{1-a_1} v dv + \int_{u^{a_1/a_2}}^1 u v^{1-a_1} dv\right)du - 3 \\ &= \frac{3a_1 a_2}{2a_1 + 2a_2 - a_1 a_2}\end{aligned}$$

Τέλος οι συζεύξεις *Marshall-Olkin* έχουν εξάρτηση άνω ουράς οπότε

$$\begin{aligned}\lambda_U &= \lim_{u\to 1^-}\frac{\overline{C}(u,u)}{1-u} \\ &= \lim_{u\to 1^-}\frac{1-2u+u^2\min(u^{-a_1},u^{-a_2})}{1-u} \\ &= \min(a_1,a_2)\end{aligned}$$

3.2.2 Η Πολυδιάστατη περίπτωση

Έστω ένα σύστημα n στοιχείων όπου σε κάθε μη-κενό υποσύνολο στοιχείων αντιστοιχεί ένα πλήγμα το οποίο είναι μοιραίο (*fatal*) για όλα τα στοιχεία του υποσυνόλου. Ας υποθέσουμε ότι S το σύνολο των υποσυνόλων του $\{1,\ldots,n\}$. $X_1,\ldots,X_n$ οι χρόνοι ζωής των στοιχείων και υποθέτουμε ότι τα πλήγματα που προσβάλλουν διαφορετικά υποσύνολα s,

$s \in \mathbf{R}$ ακολουθούν ανεξάρτητες διαδικασίες *Poisson* με παραμέτρους λ_S. Έστω Z_S, $s \in \mathbf{R}$ ο χρόνος της πραγματοποίησης του πρώτου πλήγματος για την διαδικασία πληγμάτων του υποσυνόλου *s*. Τότε οι χρόνοι Z_S είναι ανεξάρτητες τυχαίες μεταβλητές με εκθετική κατανομή, παραμέτρους λ_S και $X_j = \min_{s:j\in s} Z_s$ για *j*=1,…,*n*. Υπάρχουν $2^n - 1$ διαδικασίες πληγμάτων κάθε μία 1 προς 1 αντιστοιχισμένη σε ένα μη-κενό υποσύνολο του $\{1,...,n\}$.

Έστω *n*=3.Τότε

$$\begin{aligned} X_1 &= \min(Z_1, Z_{12}, Z_{13}, Z_{123}) \\ X_2 &= \min(Z_2, Z_{12}, Z_{23}, Z_{123}) \\ X_3 &= \min(Z_3, Z_{13}, Z_{23}, Z_{123}) \end{aligned}$$

Επομένως

$$\begin{aligned} \overline{H}(x_1,x_2,x_3) &= P\{X_1 > x_1, X_2 > x_2, X_3 > x_3\} \\ &= P\{Z_1 > x_1\}P\{Z_2 > x_2\}P\{Z_3 > x_3\}\cdot \\ &\quad \cdot P\{Z_{12} > \max(x_1,x_2)\}P\{Z_{13} > \max(x_1,x_3)\}P\{Z_{23} > \max(x_2,x_3)\}\cdot \\ &\quad \cdot P\{Z_{123} > \max(x_1,x_2,x_3)\} \\ &= \exp[-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_3 x_3 - \\ &\quad -\lambda_{12}\max(x_1,x_2) - \lambda_{13}\max(x_1,x_3) - \lambda_{23}\max(x_2,x_3) - \\ &\quad -\lambda_{123}\max(x_1,x_2,x_2)] \end{aligned}$$

Έτσι οι Marshall-Olkin οδηγήθηκαν στην παρακάτω γενίκευση

$$\begin{aligned} \overline{H}(x_1,x_2,x_3) = \exp[&-\sum_1^n \lambda_i x_i - \sum_{i<j}\lambda_{ij}\max(x_i,x_j) - \\ &-\sum_{i<j<k}\lambda_{ijk}\max(x_i,x_j,x_k) - \\ &-\ldots-\lambda_{1,\ldots,n}\max(x_1,\ldots,x_n)] \end{aligned}$$

3.3 Ελλειπτικές Συζεύξεις

Οι ελλειπτικές συζεύξεις είναι οι συζεύξεις των ελλειπτικών κατανομών και αποτελούν πλούσια πηγή πολυδιάστατων κατανομών και επιτρέπουν την μοντελοποίηση αρκετών μορφών εξάρτησης (Π.χ. εξάρτηση ακραίων τιμών). Στην ενότητα αυτή θα παρουσιάσουμε και θα αναλύσουμε κάποιες από αυτές μελετώντας παράλληλα αρκετές από τις ιδιότητές τους.

Για να υπολογίσουμε μέτρο εξάρτησης γι αυτές τις ελλειπτικές συζεύξεις στηριζόμαστε στο παρακάτω θεώρημα.

<u>Θεώρημα</u>

Έστω *X* ένα διάνυσμα τυχαίων μεταβλητών το οποίο ακολουθεί μία ελλειπτική κατανομή και έχει πίνακα συνδιακύμανσης *R*. Τότε

$$\tau(X_i, X_j) = \frac{2}{\pi}\arcsin(R_{ij})$$
$$\rho(X_i, X_j) = \frac{6}{\pi}\arcsin\left(\frac{R_{ij}}{2}\right)$$

Το παραπάνω μέτρο για ελλειπτικές κατανομές βασίζεται στον γραμμικό συντελεστή οπότε δεν μπορεί να μας ωθήσει σε ασφαλή συμπεράσματα για πιθανή μη-γραμμική εξάρτηση μεταξύ των τυχαίων μεταβλητών.

3.3.1 Gaussian Συζεύξεις

<u>Ορισμός</u>

Έστω Φ η μονοδιάστατη τυποποιημένη κανονική συνάρτηση κατανομής και Φ_R η πολυδιάστατη τυποποιημένη κανονική συνάρτηση κατανομής με πίνακα συνδιακύμανσης *R*.Τότε η *Gaussian* σύζευξη ορίζεται ως εξής

$$C_{\Phi_R}(u_1, u_2, \ldots, u_n; R) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n))$$

Στην διδιάστατη περίπτωση, η *Gaussian* σύζευξη παίρνει την ακόλουθη μορφή

$$C_{\Phi_R}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-R_{12}^2)^{1/2}} \exp\left\{-\frac{s^2 - 2R_{12}st + t^2}{2(1-R_{12}^2)}\right\} dsdt$$

Όπου R_{12} ο συντελεστής γραμμικής συσχέτισης της διδιάστατης κανονικής κατανομής.

Ας μελετήσουμε τώρα την εξάρτηση ουράς της διδιάστατης *Gaussian* σύζευξης.

$$\lambda_U = \lim_{u \to 1^-} \frac{C(u,u) - 2u + 1}{1-u}$$

και επειδή

$$P\{V \le v \big| U = u\} = \partial C(u,v)/\partial u \text{ και } P\{V > v \big| U = u\} = 1 - \partial C(u,v)/\partial u$$

εφαρμόζοντας τον κανόνα του *Hopital* έχουμε,

$$\begin{aligned} \lambda_U &= -\lim_{u \to 1^-}(-2 + \frac{\partial}{\partial u} C(u,v)\Big|_{u=v} + \frac{\partial}{\partial v} C(u,v)\big|_{u=v}) \\ &= \lim_{u \to 1^-}(P\{V > u \big| U = u\} + P\{U > u \big| V = u\}) \end{aligned}$$

Όμως οι συζεύξεις της *Gaussian* οικογένειας είναι εναλλασώμενες (*exchangeable*) δηλαδή $C(u,v) = C(v,u)$ για όλα τα $(u,v) \in I^2$ οπότε

$$\lambda_U = 2 \lim_{u \to 1^-} P\{V > u \big| U = u\}$$

Είναι

$$\begin{aligned} \lim_{u \to 1^-} P\{V > u \big| U = u\} &= \lim_{x \to +\infty} P\{\Phi^{-1}(V) > x \big| \Phi^{-1}(U) = x\} \\ &= \lim_{x \to +\infty} P\{X > x \big| Y = x\} \end{aligned}$$

και δεδομένου ότι $Y \big| X = x \sim N(\rho x, 1-\rho^2)$ προκύπτει ότι

$$\begin{aligned} \lambda_U &= 2 \lim_{x \to +\infty} (1 - \Phi((x - \rho x) / \sqrt{1-\rho^2})) \\ &= 2 \lim_{x \to +\infty} (1 - \Phi(x\sqrt{1-\rho} / \sqrt{1+\rho})) \end{aligned}$$

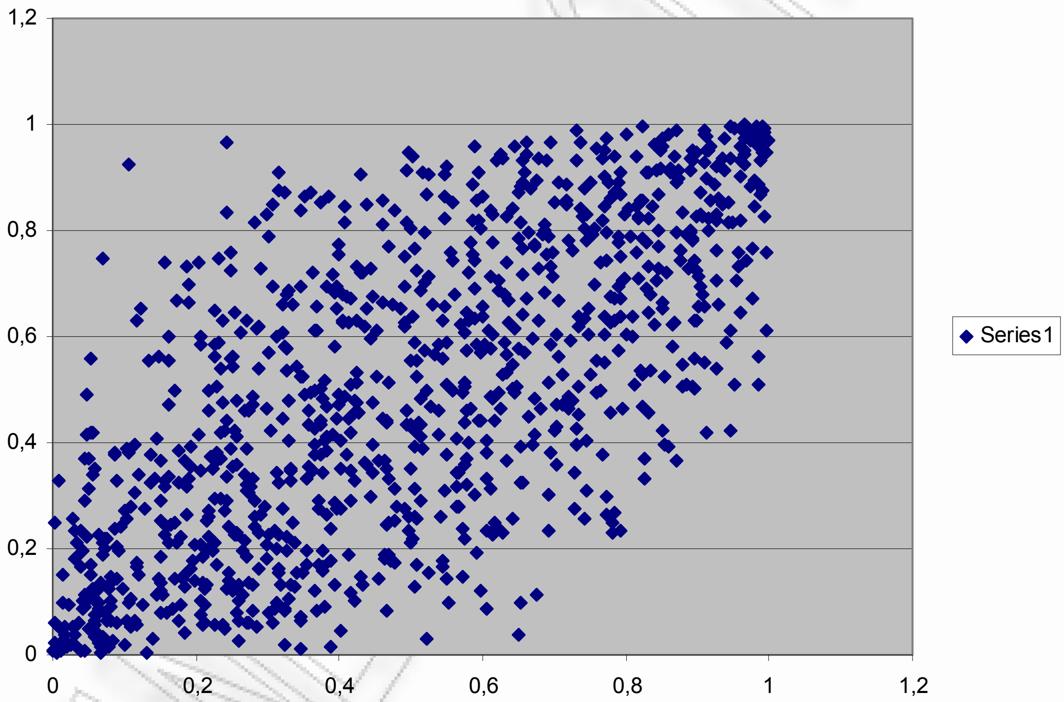
Επομένως $\lambda_U = 0$ για $R_{12} < 1$ και άρα η *Gaussian* σύζευξη *C* με *ρ*<1 δεν έχει εξάρτηση άνω ουράς.Επιπλέον οι ελλειπτικές κατανομές είναι ακτινωτά συμμετρικές (*radially symmetric*) οπότε το όριο

$$\lim_{u \to 1^-} \frac{C(u,u) - 2u + 1}{1-u}$$

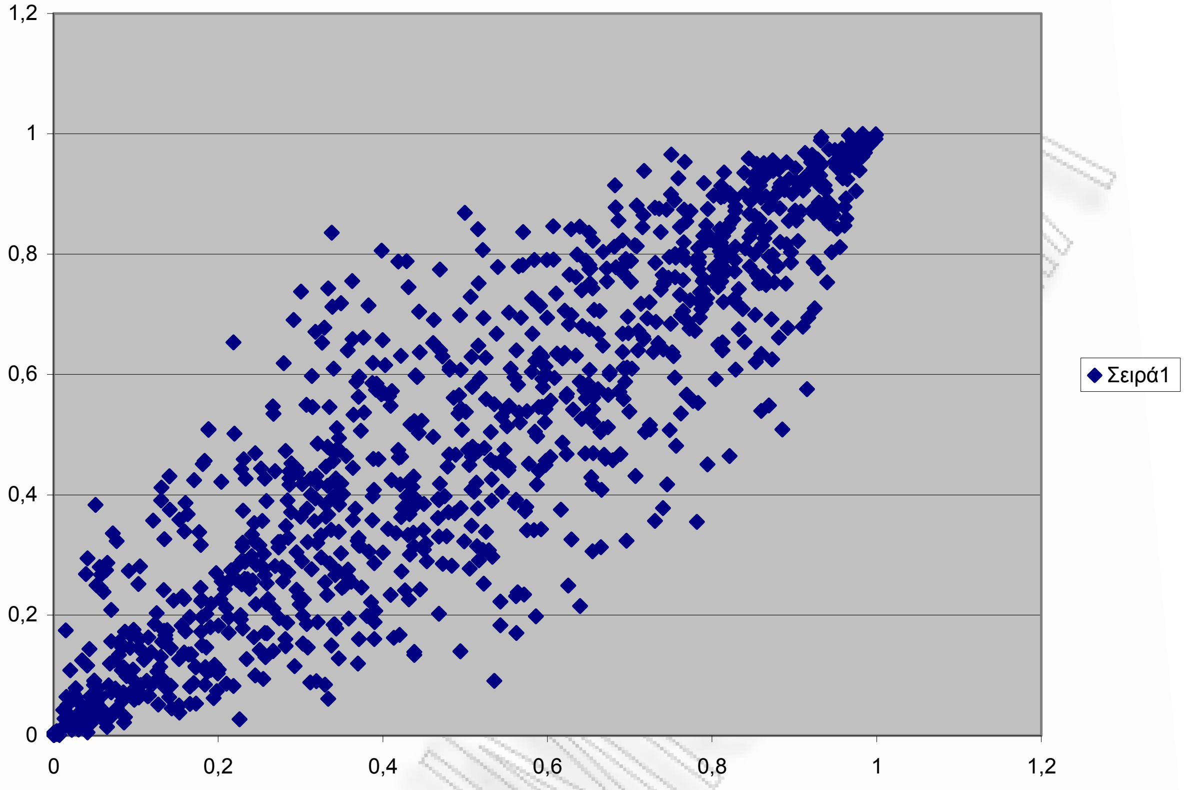
ισοδυναμεί με το

$$\lim_{u \to 0^+} \frac{C(u,u)}{u}$$

και συνεπώς η *Gaussian* σύζευξη *C* με *ρ*<1 δεν έχει εξάρτηση κάτω ουράς.



Gaussian σύζευξη με συντελεστή συσχέτισης 0.7 σε δείγμα μεγέθους 1000.



Gaussian σύζευξη με συντελεστή συσχέτισης 0.9 σε δείγμα μεγέθους 1000.

3.3.2 Student Συζεύξεις

<u>Ορισμός</u>

Έστω t_ν η μονοδιάστατη τυποποιημένη *Student* συνάρτηση κατανομής με ν βαθμούς ελευθερίας και $t_{\nu,R}$ η πολυδιάστατη τυποποιημένη *Student* συνάρτηση κατανομής με πίνακα συνδιακύμανσης R και ν βαθμούς ελευθερίας. Τότε η *Student* σύζευξη ορίζεται ως

$$C_{t_{(\nu,R)}}(u_1,u_2,...,u_n;R)=t_{\nu,R}(t_\nu^{-1}(u_1),t_\nu^{-1}(u_2),...,t_\nu^{-1}(u_n))$$

Στην διδιάστατη περίπτωση, η *Student* σύζευξη παίρνει την ακόλουθη μορφή

$$C_{t_{(\nu,R)}}=\int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{t_\nu^{-1}(v)}\frac{1}{2\pi(1-R_{12}^2)^{1/2}}\exp\left\{-\frac{s^2-2R_{12}st+t^2}{\nu(1-R_{12}^2)}\right\}^{-(\nu+2)/2} dsdt$$

όπου R_{12} ο συντελεστής γραμμικής συσχέτισης της διδιάστατης t_ν κατανομής αν ν>2.

Αν το διάνυσμα $(X_1,X_2)^T$ έχει διδιάστατη τυποποιημένη t-κατανομή με ν βαθμούς ελευθερίας και πίνακα γραμμικής συσχέτισης R, τότε το $X_2 | X_1 = x$ κατανέμεται με κατανομή t με ν+1 βαθμούς ελευθερίας και

$$E(X_2|X_1=x)=R_{12}x,$$

$$Var(X_2|X_1=x)=\left(\frac{\nu+x^2}{\nu+1}\right)(1-R_{12}^2)$$

Θα προσπαθήσουμε να υπολογίσουμε τον συντελεστή εξάρτησης άνω ουράς, ο οποίος λόγω ακτινωτής συμμετρίας θα είναι ισοδύναμος με τον συντελεστή εξάρτησης κάτω ουράς.

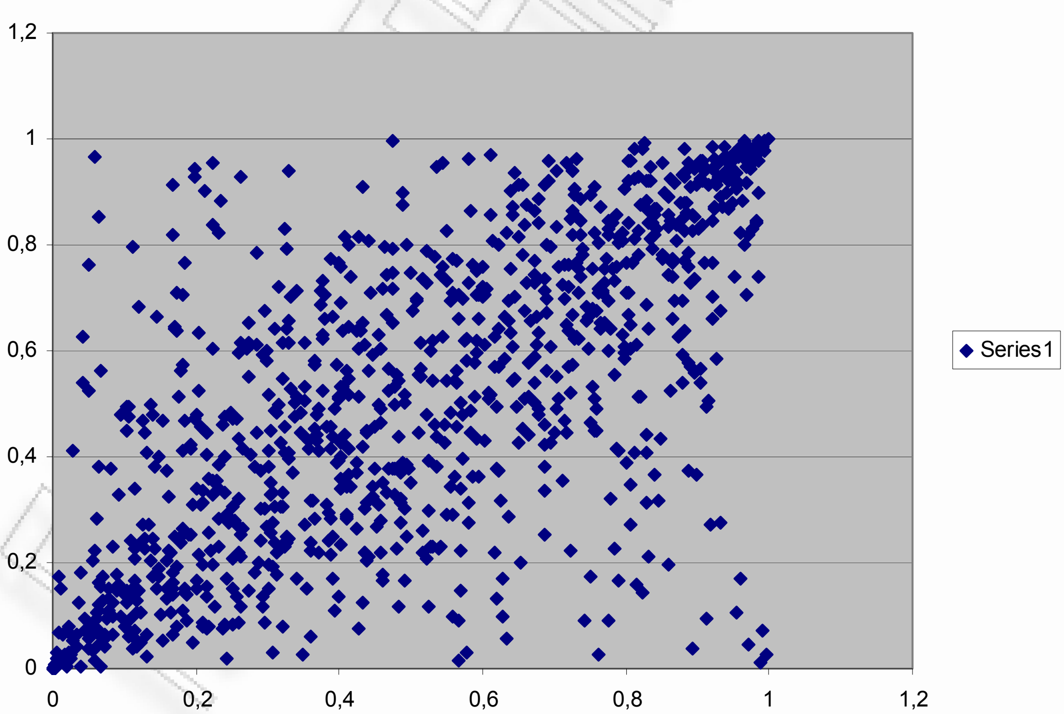
$$\begin{aligned}
\lambda_U &= 2\lim_{x\to+\infty} P\{X_2>x|X_1=x\}\\
&= 2\lim_{x\to+\infty}\left(1-t_{\nu+1}\left(\left(\frac{\nu+1}{\nu+x^2}\right)^{1/2}\frac{x-R_{12}x}{\sqrt{1-\rho^2}}\right)\right)\\
&= 2\lim_{x\to+\infty}\left(1-t_{\nu+1}\left(\left(\frac{\nu+1}{\nu/x^2+1}\right)^{1/2}\frac{\sqrt{1-R_{12}}}{\sqrt{1+R_{12}}}\right)\right)\\
&= 2\left(1-t_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-R_{12}}/\sqrt{1+R_{12}}\right)\right)
\end{aligned}$$

Στον παρακάτω πίνακα βλέπουμε τις τιμές που παίρνει το λ_U για διάφορες τιμές των ν και R_{12}.

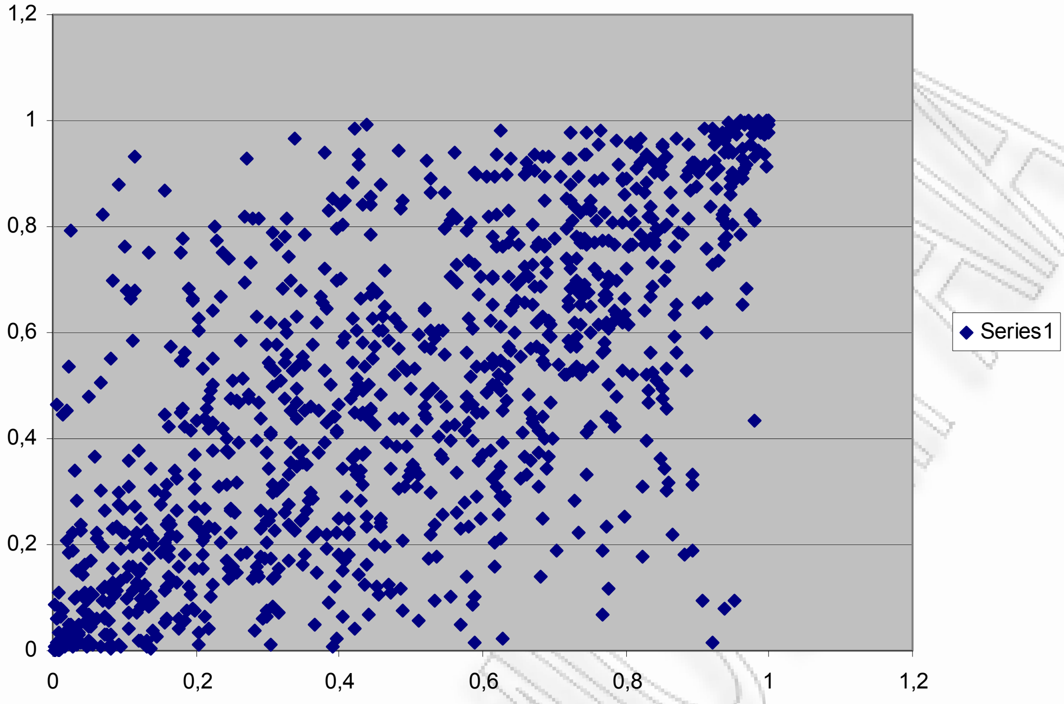
$\nu \setminus R_{12}$	-0.5	0	0.5	0.9	1
2	0.06	0.18	0.39	0.72	1
4	0.01	0.08	0.25	0.63	1
10	0.00	0.01	0.08	0.46	1
∞	0	0	0	0	1

Βλέπουμε ότι όσο αυξάνεται το R_{12}, το λ_U αυξάνεται ενώ όσο αυξάνεται το ν το λ_U μειώνεται. Ειδικότερα το λ_U τείνει στο 0 καθώς οι βαθμοί ελευθερίας τείνουν στο άπειρο, για $R_{12} < 1$.

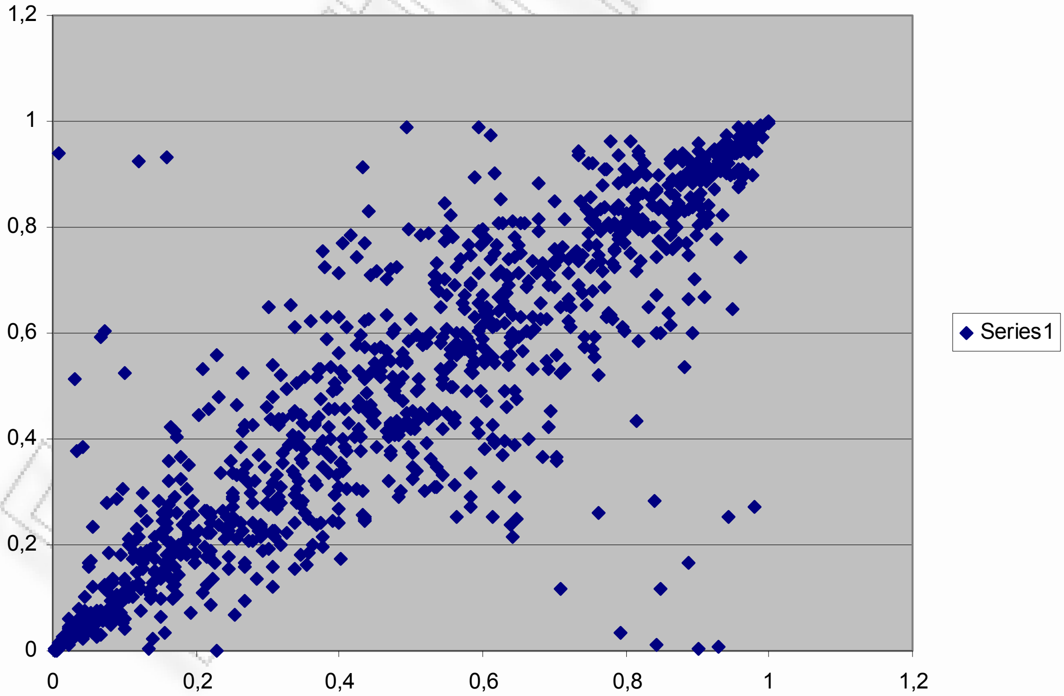
Οδηγούμαστε στο συμπέρασμα ότι η *Student* σύζευξη έχει άνω και κάτω εξάρτηση ουράς ενώ όταν οι βαθμοί ελευθερίας τείνουν στο άπειρο η *Student* σύζευξη συμπεριφέρεται σαν την *Gaussian* δηλαδή δεν έχει εξάρτηση ουρά (άνω και κάτω).



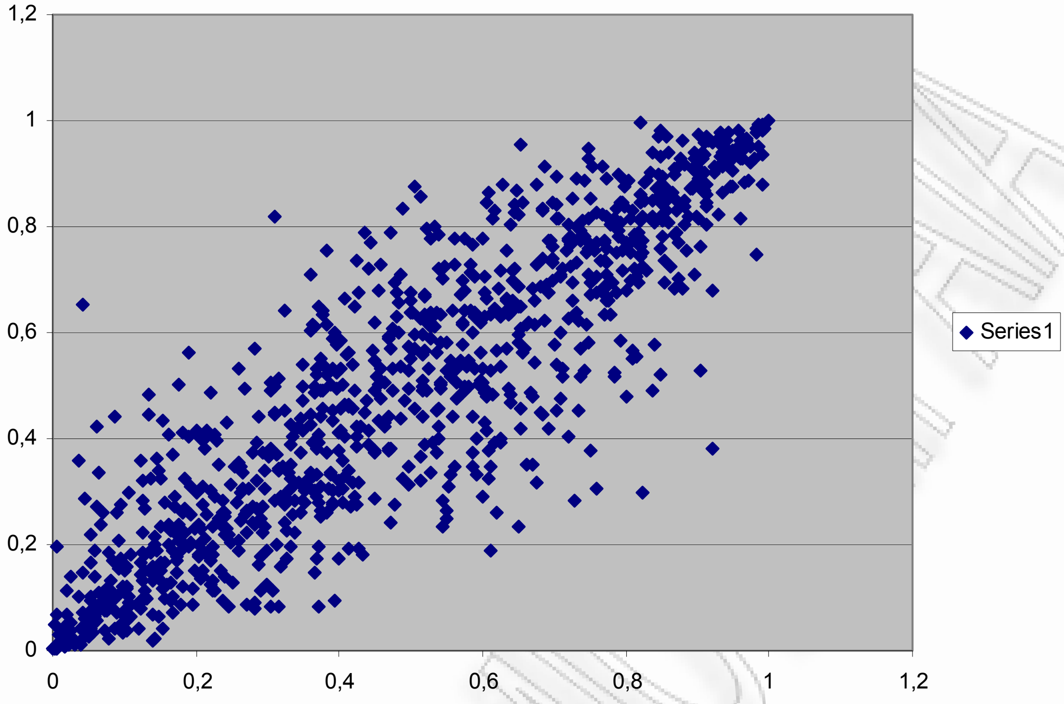
Student σύζευξη με συντελεστή συσχέτισης 0.7 σε δείγμα μεγέθους 1000 και 2 β.ε.



Student σύζευξη με συντελεστή συσχέτισης 0.7 σε δείγμα μεγέθους 1000 και 5 β.ε.



Student σύζευξη με συντελεστή συσχέτισης 0.9 σε δείγμα μεγέθους 1000 και 2 βαθμούς ελευθερίας.



Student σύζευξη με συντελεστή συσχέτισης 0.9 σε δείγμα μεγέθους 1000 και 5 β.ε.

3.4 Αρχιμήδειες Συζεύξεις

Οι ελλειπτικές συζεύξεις ξεχωρίζουν για την ευκολία με την οποία προσημειώνονται. Ωστόσο οι σύνδεσμοι αυτοί έχουν περιορισμένες ιδιότητες ενώ παράλληλα οι υπολογισμοί για τα μέτρα εξάρτησής τους είναι πολύπλοκοι και χρονοβόροι. Υπάρχουν περιπτώσεις, κυρίως σε χρηματοοικονομικές και ασφαλιστικές εφαρμογές, κατά τις οποίες η εξάρτηση μεταξύ μεγάλων ζημιών (πτώση χρηματιστηρίου) είναι ισχυρότερη από αυτήν μεταξύ μεγάλων κερδών. Τέτοιες περιπτώσεις δεν μπορούν να μοντελοποιηθούν με τη χρήση ελλειπτικών συζεύξεων.

Σε αυτή την ενότητα θα αναφερθούμε στις Αρχιμήδειες συζεύξεις που αποτελούν μία από τις σημαντικότερες κατηγορίες συνδέσμων. Σε αντίθεση με τις ελλειπτικές συζεύξεις, οι περισσότερες Αρχιμήδειες συζεύξεις έχουν κλειστούς τύπους έκφρασης. Επιπλέον οι Αρχιμήδειες συζεύξεις, δεν παράγονται από πολυδιάστατες συναρτήσεις κατανομών μέσω του θεωρήματος του *Sklar* όπως συμβαίνει με τις συζεύξεις που εξετάσαμε έως τώρα. Ως εκ τούτου, χρειαζόμαστε κάποιες τεχνικές συνθήκες ώστε να θεωρήσουμε ότι οι πολυδιάστατες προεκτάσεις Αρχιμήδειων 2-συζεύξεων είναι *n*-συζεύξεις. Ένα ακόμα μειονέκτημα των Αρχιμήδειων συζεύξεων είναι ο περιορισμός της επιλογής παραμέτρων με την έννοια ότι κάποια στοιχεία του πίνακα συσχετίσεων πρέπει να είναι ίσα.

Ωστόσο οι Αρχιμήδειες συζεύξεις έχουν πολλές εφαρμογές και η ευρεία χρήση τους οφείλεται στην ευκολία με την οποία κατασκευάζονται, στο μεγάλο πλήθος οικογενειών που ανήκουν στην κατηγορία αυτή και στις ιδιότητες των μελών της.

<u>Ορισμός</u>

Έστω φ μία συνεχής αυστηρά φθίνουσα συνάρτηση από το $\left[0,1\right]$ στο $\left[0,\infty\right]$ τέτοια ώστε $\varphi(1)=0$. Η ψευδο-αντίστροφη (*pseudo-inverse*) της φ είναι η συνάρτηση $\varphi^{[-1]}:\left[0,\infty\right]\to\left[0,1\right]$ με τύπο

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

Η $\varphi^{[-1]}$ είναι συνεχής και φθίνουσα στο $\left[0,\infty\right]$ και αυστηρά φθίνουσα στο $\left[0,\varphi(0)\right]$. Επιπλέον $\varphi^{[-1]}(\varphi(u))=u$ στο $\left[0,1\right]$ και

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

Τέλος αν $\varphi(0)=\infty$, τότε $\varphi^{[-1]}=\varphi^{-1}$.

<u>Θεώρημα</u>

Έστω φ μία συνεχής αυστηρά φθίνουσα συνάρτηση από το $\left[0,1\right]$ στο $\left[0,\infty\right]$ τέτοια ώστε $\varphi(1)=0$ και έστω $\varphi^{[-1]}$ η ψευδο-αντίστροφη της φ. Έστω C μία συνάρτηση από το $\left[0,1\right]^2$ στο $\left[0,1\right]$ με τύπο

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

Τότε C είναι σύζευξη αν και μόνο αν η φ είναι κοίλη.

Συζεύξεις της προηγούμενης μορφής ονομάζονται Αρχιμήδειες και η συνάρτηση φ καλείται γεννήτρια (*generator*) της σύζευξης. Αν $\varphi(0)=\infty$ λέμε ότι η φ είναι αυστηρή (*strict*) γεννήτρια. Στην περίπτωση αυτή $\varphi^{[-1]}=\varphi^{-1}$ και

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

θα λέμε ότι είναι αυστηρή Αρχιμήδεια σύζευξη.

3.4.1 Ιδιότητες Αρχιμήδειων συζεύξεων

<u>Θεώρημα</u>

Έστω C μια Αρχιμήδεια σύζευξη με γεννήτρια φ. Τότε

1. C είναι συμμετρική δηλαδή $C(u,v)=C(v,u)$ για όλα τα u,v στο $[0,1]$.

2. C είναι προσεταιριστική δηλαδή $C(C(u,v),w)=C(u,C(v,w))$ για όλα τα w,u,v στο $[0,1]$.

Πράγματι έχουμε

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

Η προσεταιριστικότητα είναι ιδιότητα των Αρχιμήδειων συζεύξεων και δεν αποτελεί γενική ιδιότητα των συζεύξεων.

Π.χ. Έστω C_θ μία διδιάστατη σύζευξη της *Farlie-Gumbel-Morgenstern* οικογένειας συζεύξεων με τύπο

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

για $\theta\in[-1,1]$. Τότε

$$C_\theta\left(\frac{1}{4},C_\theta\left(\frac{1}{2},\frac{1}{3}\right)\right)\neq C_\theta\left(C_\theta\left(\frac{1}{4},\frac{1}{2}\right),\frac{1}{3}\right)$$

για όλα τα $\theta\in[-1,1]\setminus\{0\}$

και επομένως μόνο η σύζευξη *Π* είναι Αρχιμήδεια από την διδιάστατη *Farlie-Gumbel-Morgenstern* οικογένεια συζεύξεων.

Θεώρημα

Έστω *C* μια Αρχιμήδεια σύζευξη με γεννήτρια *φ* και έστω

$$K_C(t) = V_C(\{(u,v) \in [0,1]^2 \,|\, C(u,v) \le t\})$$

Τότε για κάθε $t \in [0,1]$,

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}$$

Πόρισμα

Αν $(U,V)^T$ έχει συνάρτηση κατανομής *C* όπου *C* είναι Αρχιμήδεια σύζευξη με γεννήτρια *φ*, τότε συνάρτηση K_C που ορίστηκε παραπάνω είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής $C(U,V)$.

Θεώρημα

Κάτω από την υπόθεση του προηγούμενου πορίσματος η από κοινού συνάρτηση κατανομής των μεταβλητών

$$S = \varphi(U)/[\varphi(U)+\varphi(V)] \text{ και } T = C(U,V)$$

δίνεται από τον τύπο

$$H(s,t) = sK_C(t) \text{ για όλα τα } (s,t) \in [0,1]^2$$

Άρα οι *S* και *T* είναι ανεξάρτητες και η *S* κατανέμεται ομοιόμορφα στο $[0,1]$.

3.4.2 Το τ του Kendall για Αρχιμήδειες Συζεύξεις

Θυμόμαστε ότι το τ του *Kendall* για μία σύζευξη *C* μπορεί να εκφραστεί σαν διπλό ολοκλήρωμα της *C* το οποίο τις περισσότερες φορές είναι δύσκολο να υπολογιστεί. Ωστόσο για μία Αρχιμήδεια σύζευξη, το *τ* του *Kendall* μπορεί να εκφραστεί σαν μονό ολοκλήρωμα της γεννήτριας συνάρτησης και της παραγώγου της, όπως φαίνεται στο παρακάτω θεώρημα.

Θεώρημα

Έστω X και Y τυχαίες μεταβλητές με Αρχιμήδεια σύζευξη C και γεννήτρια φ της C. Τότε το τ του Kendall δίνεται από τον τύπο

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

Απόδειξη

Έστω U και V ομοιόμορφα κατανεμημένες τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής C και έστω K_C η συνάρτηση κατανομής της $C(U,V)$. Τότε έχουμε

$$\begin{aligned}
\tau &= 4E\big(C(U,V)\big) - 1 \\
&= 4\int_0^1 t dK_C(t) - 1 \\
&= 4\left(\left[tK_C(t)\right]_0^1 - \int_0^1 K_C(t)dt \right) - 1 \\
&= 3 - 4\int_0^1 K_C(t)dt
\end{aligned}$$

και επειδή

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}$$

έπεται ότι

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

3.4.3 Εξάρτηση Ουράς για Αρχιμήδειες Συζεύξεις

Για τις Αρχιμήδειες Συζεύξεις η εξάρτηση ουράς μπορεί να εκφραστεί μέσα από γεννήτριες.

Θεώρημα

Έστω φ αυστηρή γεννήτρια. Αν το $\varphi^{-1'}(0)$ είναι πεπερασμένο, τότε η

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

δεν έχει εξάρτηση άνω ουράς. Αν $\varphi^{-1'}(0)=-\infty$, τότε η C έχει εξάρτηση άνω ουράς που δίνεται από τον τύπο

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

Απόδειξη

$$\begin{aligned}\lambda_U &= \lim_{u\to 1^-}\left[1-2u+\varphi^{-1}\left(2\varphi(u)\right)\right]/(1-u)\\ &= 2-2\lim_{u\to 0^+}\varphi^{-1'}\left(2\varphi(u)\right)/\varphi^{-1'}\left(2\varphi(u)\right)\\ &= 2-2\lim_{s\to 0^+}\varphi^{-1'}(2s)/\varphi^{-1'}(s)\end{aligned}$$

αν $\varphi^{-1'}(0)\in(-\infty,0)$, τότε το παραπάνω όριο ισούται με 0 και επομένως η C δεν έχει εξάρτηση άνω ουράς.

Θεώρημα

Έστω φ όπως στο προηγούμενο θεώρημα. Τότε συντελεστής εξάρτησης κάτω ουράς για τη σύζευξη $C(u,v)=\varphi^{-1}(\varphi(u)+\varphi(v))$ ισούται με

$$\lambda_L = 2\lim_{s\to\infty}\left[\varphi^{-1'}(2s)/\varphi^{-1'}(s)\right]$$

(Η απόδειξη είναι παρόμοια με του προηγούμενου θεωρήματος)

3.4.4 Οικογένειες Αρχιμήδειων Συζεύξεων

Θα δούμε τώρα της σημαντικότερες οικογένειες συζεύξεων παρουσιάζοντας κάποιες από τις ιδιότητές τους με τη βοήθεια των θεωρημάτων που αναπτύξαμε παραπάνω.

Οικογένεια Gumbel

Έστω $\varphi(t) = (-\ln t)^{\theta}$ με $\theta \geq 1$.

Από τον ορισμό των Αρχιμήδειων Συζεύξεων προκύπτει ότι

$$\begin{aligned} C_{\theta}(u,v) &= \varphi^{-1}(\varphi(u)+\varphi(v)) \\ &= \exp(-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}) \end{aligned}$$

Αυτή η οικογένεια συζεύξεων ονομάζεται οικογένεια *Gumbel* και εξετάζοντας τις ακραίες περιπτώσεις έχουμε

$$C_1 = \Pi \text{ και } \lim_{\theta\to\infty} C_{\theta} = M$$

όπου $\Pi(u,v) = uv$ και $M(u,v) = \min(u,v)$

Για να υπολογίσουμε το *τ* του *Kendall*, χρειαζόμαστε το πηλίκο $\dfrac{\varphi(t)}{\varphi'(t)} = (t\ln t)/\theta$ οπότε

$$\begin{aligned} \tau &= 1 + 4\int_0^1 \frac{t\ln t}{\theta} dt \\ &= 1 + \frac{4}{\theta}\left(\left[\frac{t^2}{2}\ln t\right]_0^1 - \int_0^1 \frac{t}{2}dt\right) \\ &= 1 + \frac{4}{\theta}(0 - 1/4) \\ &= 1 - 1/\theta \end{aligned}$$

Σε ότι έχει να κάνει με την εξάρτηση ουράς, είναι προφανές ότι αυτή η οικογένεια συζεύξεων έχει **μόνο** εξάρτηση άνω ουράς.

Πράγματι,

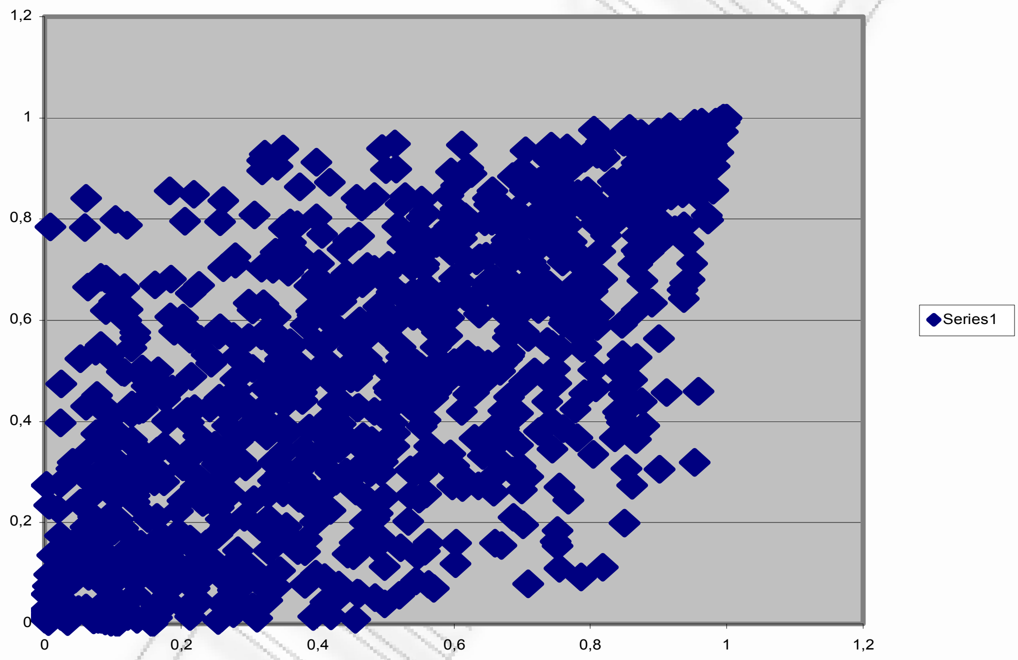
$$\begin{aligned} \lim_{u\to 1^-} \frac{C(u,u) - 2u + 1}{1-u} &= \lim_{u\to 1^-} \frac{1 - 2u + \exp(2^{1/\theta}\ln u)}{1-u} \\ &= \lim_{u\to 1^-} \frac{1 - 2u + u^{2^{1/\theta}}}{1-u} \\ &= 2 - 2^{1/\theta} \end{aligned}$$

Άρα για $\theta > 1$ το λ_U υπάρχει και είναι ίσο με $2 - 2^{1/\theta}$

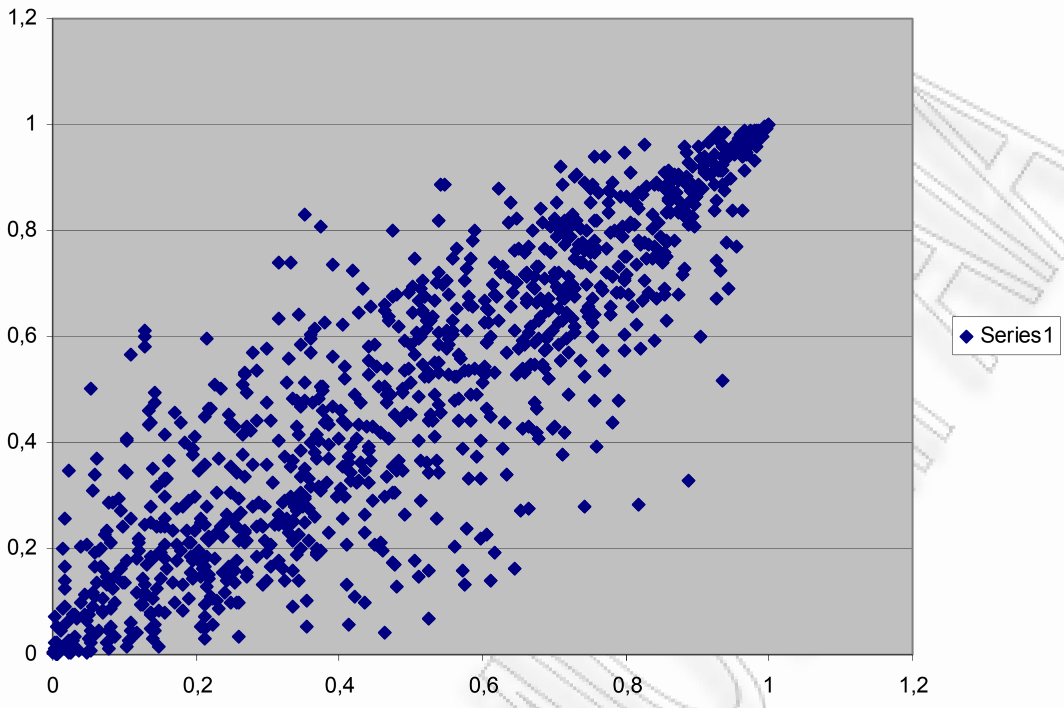
Αντίθετα,

$$\lim_{u\to 0^+} \frac{C(u,u)}{u} = \lambda_L = \lim_{u\to 0^+} \frac{u^{2^{1/\theta}}}{u} = \lim_{u\to 0^+} u^{2^{1/\theta}-1} = 0$$

Οπότε δεν υπάρχει εξάρτηση κάτω ουράς.



Gumbel σύζευξη με συντελεστή συσχέτισης 0.7 σε δείγμα μεγέθους 1000.



Gumbel σύζευξη με συντελεστή συσχέτισης 0.9 σε δείγμα μεγέθους 1000.

Οικογένεια Clayton

Έστω $\varphi(t) = (t^{-\theta} - 1)/\theta$ με $\theta \in [-1, \infty) \setminus \{0\}$.

Οι συζεύξεις που παράγονται από την φ είναι οι

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

για $\theta > 0$ είναι «αυστηρές» και δίνονται από τον τύπο

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

Αυτή η οικογένεια συζεύξεων ονομάζεται οικογένεια *Clayton* και εξετάζοντας τις ακραίες περιπτώσεις έχουμε

$$C_{-1} = W\,,\ \lim_{\theta \to 0} C_\theta = \Pi \ \text{ και } \lim_{\theta \to \infty} C_\theta = \mathrm{M}$$

Για να υπολογίσουμε το τ του *Kendall*, χρειαζόμαστε το πηλίκο

$$\frac{\varphi(t)}{\varphi'(t)} = \left(t^{\theta+1} - t\right)\big/\theta$$

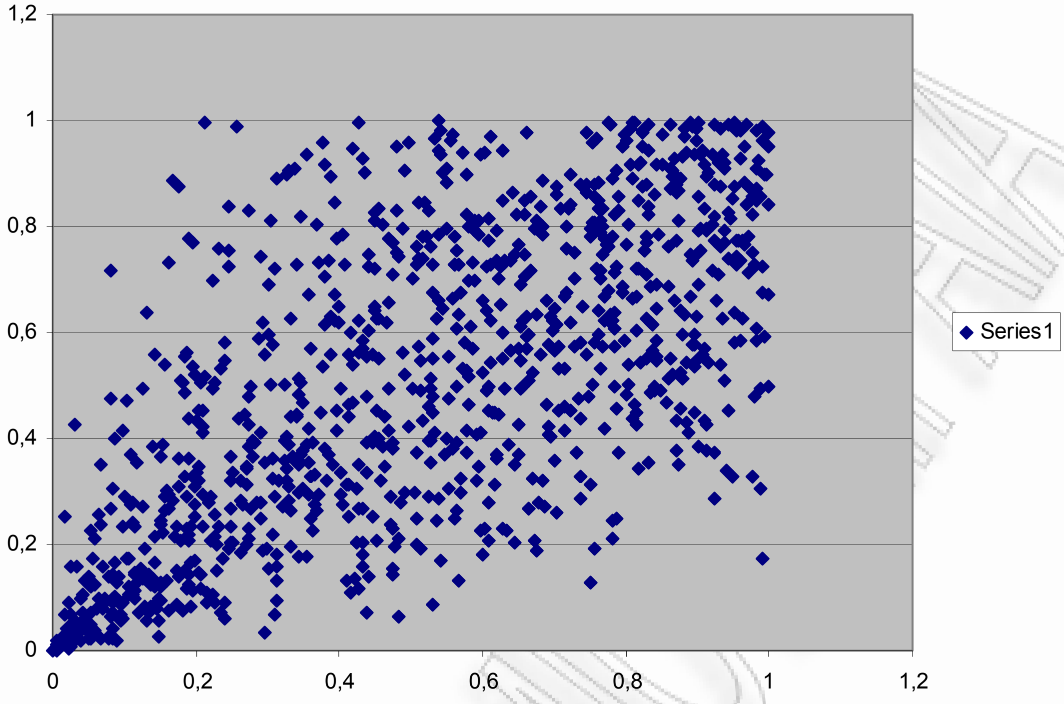
οπότε

$$\begin{aligned}\tau &= 1 + 4\int_0^1 \frac{t^{\theta+1} - t}{\theta} dt \\ &= 1 + \frac{4}{\theta}\left(\frac{1}{\theta+2} - \frac{1}{2}\right) \\ &= \frac{\theta}{\theta+2}\end{aligned}$$

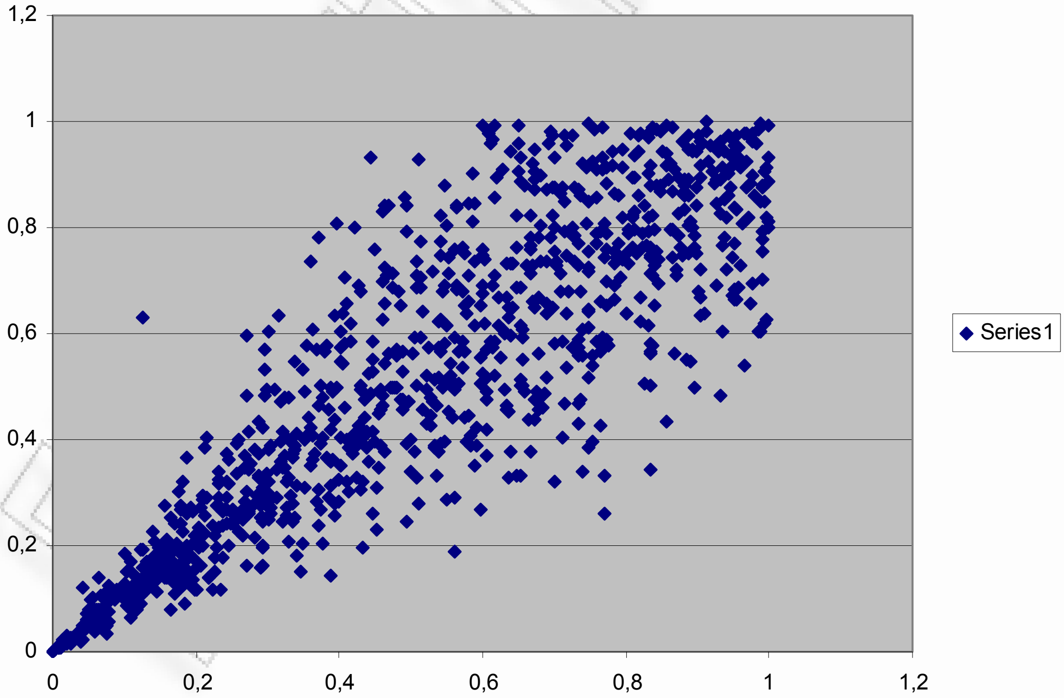
Η οικογένεια συζεύξεων αυτή έχει **μόνο** εξάρτηση κάτω ουράς.

$\varphi^{-1}(s) = (1+\theta s)^{-1/\theta}$ οπότε από προηγούμενα θεωρήματα $\lambda_U = 0$ και

$$\begin{aligned}\lambda_L &= 2\lim_{s\to\infty}\left[\varphi^{-1'}(2s)/\varphi^{-1'}(s)\right] \\ &= 2\lim_{s\to\infty}\left[\frac{(1+2\theta s)^{-1/\theta-1}}{(1+\theta s)^{-1/\theta-1}}\right] \\ &= 2\cdot 2^{-1/\theta-1} \\ &= 2^{-1/\theta}\end{aligned}$$



Clayton σύζευξη με συντελεστή συσχέτισης 0.7 σε δείγμα μεγέθους 1000.



Clayton σύζευξη με συντελεστή συσχέτισης 0.9 σε δείγμα μεγέθους 1000.

Οικογένεια Frank

Έστω $\varphi(t) = -\ln\dfrac{e^{-\theta t}-1}{e^{-\theta}-1}$ με $\theta \in \mathbf{R}\setminus\{0\}$.

Τότε παίρνουμε την οικογένεια *Frank* με τύπο

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

Οι συζεύξεις *Frank* είναι αυστηρές Αρχιμήδειες συζεύξεις και είναι οι μόνες Αρχιμήδειες συζεύξεις που ικανοποιούν τη σχέση

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

Εξετάζοντας τις ακραίες περιπτώσεις οδηγούμαστε στις παρακάτω διαπιστώσεις

$$\lim_{\theta\to-\infty} C_\theta = W\ ,\ \lim_{\theta\to 0} C_\theta = \Pi \ \text{ και } \lim_{\theta\to\infty} C_\theta = \mathrm{M}$$

Αποδεικνύεται ότι για την οικογένεια *Frank*, το *τ* του *Kendall* ορίζεται ως εξής

$$\tau = 1-4\left(1-D_1(\theta)\right)/\theta$$

όπου $D_k(x)$ είναι η συνάρτηση *Debye* που δίνεται από τον τύπο

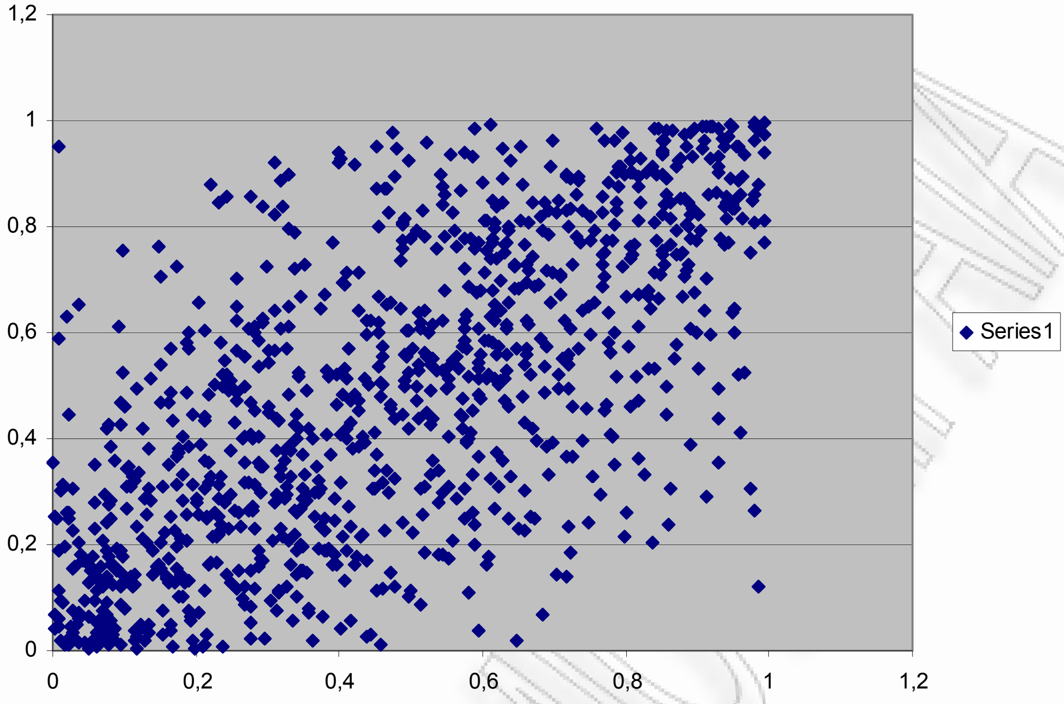
$$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t-1}dt$$

για κάθε θετικό ακέραιο *k*.

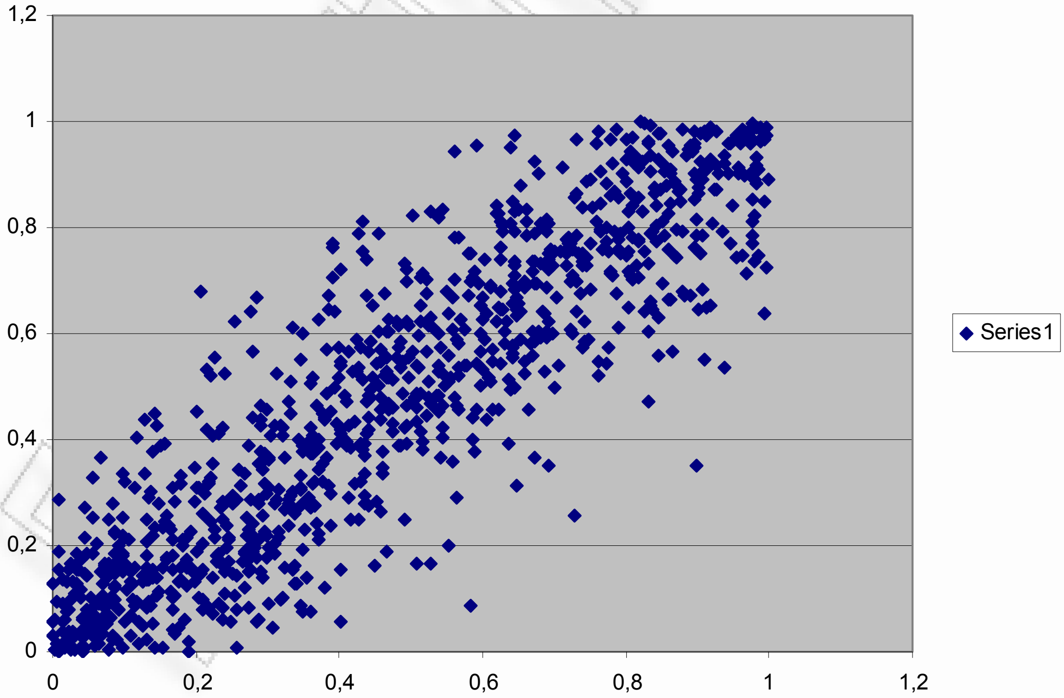
Όσον αφορά την εξάρτηση ουράς, έχουμε

$$\varphi^{-1'}(0) = -\frac{e^\theta-1}{\theta}$$

το οποίο είναι πεπερασμένο και επομένως οι συζεύξεις *Frank* δεν έχουν εξάρτηση άνω ουράς, αλλά και εξάρτηση κάτω ουράς (λόγω συμμετρίας).



Frank σύζευξη με συντελεστή συσχέτισης 0.7 σε δείγμα μεγέθους 1000.



Frank σύζευξη με συντελεστή συσχέτισης 0.9 σε δείγμα μεγέθους 1000.

3.4.5 Πολυδιάστατες Αρχιμήδειες Συζεύξεις

Στην ενότητα αυτή θα εξετάσουμε την κατασκευή μίας πολυδιάστατης προέκτασης Αρχιμήδειων 2-συζεύξεων.

Η έκφραση για τη n-διάστατη σύζευξη γινομένου Π^n με $u=(u_1,...,u_n)^T$, μπορεί να γραφτεί ως εξής

$$\begin{aligned}\Pi^n(u) &= u_1 \cdots u_n \\ &= \exp\left(-\left[(-\ln u_1)+\cdots+(-\ln u_n)\right]\right)\end{aligned}$$

Αυτό οδηγεί στην παρακάτω γενίκευση

$$C^n(u)=\varphi^{[-1]}\left(\varphi(u_1)+\cdots+\varphi(u_n)\right)$$

Στην 3-διάστατη περίπτωση έχουμε,

$$\begin{aligned}C^3(u_1,u_2,u_3) &= \varphi^{[-1]}\left(\varphi\circ\varphi^{[-1]}(\varphi(u_1)+\varphi(u_2)+\varphi(u_3))\right) \\ &= C(C(u_1,u_2),u_3)\end{aligned}$$

και στην 4-διάστατη

$$\begin{aligned}C^4(u_1,u_2,u_3,u_4) &= \varphi^{[-1]}\left(\varphi o \varphi^{[-1]}(\varphi\circ\varphi^{[-1]}(\varphi(u_1)+\varphi(u_2)+\varphi(u_3)+\varphi(u_4)))\right) \\ &= C(C^3(u_1,u_2,u_3),u_4) \\ &= C(C(C(u_1,u_2),u_3),u_4)\end{aligned}$$

Άρα φαίνεται ότι για $n\geq 3$

$$C^n(u_1,...,u_n)=C(C^{n-1}(u_1,...,u_{n-1}),u_n)$$

Αν και η τεχνική αυτή κατασκευής πολυδιάστατων συζεύξεων συνήθως αποτυγχάνει, μπορούμε να υποθέσουμε ότι δεδομένων κάποιων ιδιοτήτων των φ και $\varphi^{[-1]}$, η C^n είναι σύζευξη για $n\geq 3$. Στην παραδοχή αυτή μας οδηγούν ιδιότητες των Αρχιμήδειων συζεύξεων όπως η συμμετρία και η προσεταιριστικότητα.

Ορισμός

Μία συνάρτηση $g(t)$ είναι πλήρως μονότονη (*completely monotone*) στο διάστημα *I*, αν έχει παραγώγους όλων των τάξεων οι οποίες εναλλάσσονται σε πρόσημο, δηλαδή

$$(-1)^k \frac{d^k}{dt^k} g(t) \geq 0$$

για κάθε *t* στο *I* και *k*=0,1,2,…

Αν η $g:[0,\infty) \to [0,\infty)$ είναι πλήρως μονότονη στο $[0,\infty)$ και υπάρχει $t \in [0,\infty)$ τέτοιο ώστε $g(t)=0$, τότε $g(t)=0$ για κάθε $t \in [0,\infty)$. Επομένως αν η ψευδο-αντίστροφη $\varphi^{[-1]}$ μιας Αρχιμήδειας γεννήτριας φ είναι πλήρως μονότονη, τότε $\varphi^{[-1]} > 0$ για κάθε $t \in [0,\infty)$ και άρα $\varphi^{[-1]} = \varphi^{-1}$.

Θεώρημα

Έστω φ μια συνεχής, αυστηρά φθίνουσα συνάρτηση από το $[0,1]$ στο $[0,\infty]$ τέτοια ώστε $\varphi(0)=\infty$ και $\varphi(1)=0$ και έστω φ^{-1} η αντίστροφη της φ. Αν η συνάρτηση C^n από το $[0,1]^n$ στο $[0,1]$ με

$$C^n(u) = \varphi^{-1}\left(\varphi(u_1) + \cdots + \varphi(u_n)\right),$$

τότε C^n είναι μία n-σύζευξη για κάθε $n \geq 2$ αν και μόνο αν η φ^{-1} είναι πλήρως μονότονη στο $[0,\infty)$.

Πόρισμα

Αν η αντίστροφη φ^{-1} μιας αυστηρής γεννήτριας φ μιας Αρχιμήδειας σύζευξης *C* είναι πλήρως μονότονη, τότε $C \succ \Pi$ δηλαδή $C(u,v) \geq uv$ για όλα τα *u,v* στο $[0,1]$.

Προσπαθώντας να εκφράσουμε την πολυδιάστατη προέκταση του γενικού τύπου των Αρχιμήδειων συζεύξεων, θα πρέπει να τονίσουμε ότι μία τέτοια γενίκευση θα ήταν αρκετά πολύπλοκη. Για το λόγο αυτό θα συζητήσουμε ικανές συνθήκες ώστε οι 3-διάστατες και 4-διάστατες προεκτάσεις να είναι 3- και 4-συζεύξεις αντίστοιχα. Οι συνθήκες αυτές μπορούν να μας οδηγήσουν σε γενικεύσεις μεγαλύτερων διαστάσεων.

Η 3-διάστατη γενίκευση της σχέσης

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

είναι

$$\varphi_1^{-1}\left(\varphi_1\circ\varphi_2^{-1}\left(\varphi_2\left(u_1\right)+\varphi_2\left(u_2\right)\right)+\varphi_1\left(u_3\right)\right)=C_1\left(C_2\left(u_1,u_2\right),u_3\right)$$

ενώ η 4-διάστατη γενίκευση είναι

$$\varphi_1^{-1}\left(\varphi_1\circ\varphi_2^{-1}\left(\varphi_2\circ\varphi_3^{-1}\left(\varphi_3\left(u_1\right)+\varphi_3\left(u_2\right)\right)+\varphi_2\left(u_3\right)\right)+\varphi_1\left(u_4\right)\right)=C_1\left(C_2\left(C_3\left(u_1,u_2\right),u_3\right),u_4\right)$$

όπου φ_1,φ_2 και φ_3 γεννήτριες αυστηρών Αρχιμήδειων συζεύξεων και C_i Αρχιμήδειες σύζευξης με γεννήτριες φ_i.

Έστω $L_n=\left\{\varphi:[0,\infty)\to[0,1]\middle|\varphi(0)=1,\varphi(\infty)=0,(-1)^j\varphi^{(j)}\geq 0, j=1,\ldots,n\right\}$

με $n=1,2,\ldots,\infty$ και

$$L_n^*=\left\{\omega:[0,\infty)\to[0,1]\middle|\omega(0)=0,\omega(\infty)=\infty,(-1)^{j-1}\omega^{(j)}\geq 0, j=1,\ldots,n\right\}$$

με $n=1,2,\ldots,\infty$.

Σημειώνουμε ότι $\varphi^{-1}\in L_1$ αν φ είναι η γεννήτρια μιας Αρχιμήδειας σύζευξης. Οι συναρτήσεις που ανήκουν στην κλάση L_n^* είναι συνήθως συνθέσεις της μορφής $\psi^{-1}\circ\varphi$ για $\psi,\varphi\in L_1$.

Σημειώνουμε επίσης ότι για να είναι η $C^n(u)=\varphi^{-1}\left(\varphi(u_1)+\cdots+\varphi(u_n)\right)$, σύζευξη, θα πρέπει επιπλέον

1. $\varphi^{-1}\in L_n$ και
2. Αν $C^n(u)=\varphi^{[-1]}\left(\varphi(u_1)+\cdots+\varphi(u_n)\right)$ είναι σύζευξη για κάθε *n*, τότε η φ^{-1} πρέπει να είναι πλήρως μονότονη.

Για την 3-διάστατη περίπτωση αποδεικνύεται ότι αν φ_1^{-1} και φ_2^{-1} είναι πλήρως μονότονες και $\varphi_1 \circ \varphi_2^{-1} \in L_\infty^*$, τότε η

$$\varphi_1^{-1}\left(\varphi_1 \circ \varphi_2^{-1}\left(\varphi_2\left(u_1\right)+\varphi_2\left(u_2\right)\right)+\varphi_1\left(u_3\right)\right)$$

είναι 3-σύζευξη με περιθώριες της μορφής

$$C^n(u)=\varphi^{[-1]}\left(\varphi(u_1)+\cdots+\varphi(u_n)\right).$$

Αντίστοιχα για την 4-διάστατη περίπτωση, αν $\varphi_1^{-1}, \varphi_2^{-1}$ και φ_3^{-1} είναι πλήρως μονότονες και $\varphi_1 \circ \varphi_2^{-1}, \varphi_2 \circ \varphi_3^{-1} \in L_\infty^*$, τότε η

$$\varphi_1^{-1}\left(\varphi_1 \circ \varphi_2^{-1}\left(\varphi_2 \circ \varphi_3^{-1}\left(\varphi_3\left(u_1\right)+\varphi_3\left(u_2\right)\right)+\varphi_2\left(u_3\right)\right)+\varphi_1\left(u_4\right)\right)$$

είναι 3-σύζευξη με περιθώριες της μορφής

$$C^n(u)=\varphi^{[-1]}\left(\varphi(u_1)+\cdots+\varphi(u_n)\right).$$

Με όμοιο τρόπο γενικεύουμε και σε συζεύξεις μεγαλύτερων διαστάσεων. Στο κομμάτι αυτό ασχοληθήκαμε με αυστηρές Αρχιμήδειες συζεύξεις. Ωστόσο με κάποιους περιορισμούς μπορούμε να γενικεύσουμε τα αποτελέσματα μας για μη-αυστηρές Αρχιμήδειες συζεύξεις.

ΚΕΦΑΛΑΙΟ 4

4 Εκτίμηση, Επιλογή και Έλεγχοι Συζεύξεων

4.1 Μέθοδοι Εκτίμησης Συζεύξεων

Στην ενότητα αυτή θα παρουσιάσουμε κάποιους τρόπους εκτίμησης συζεύξεων των οποίων οι παράμετροι μπορούν να συσχετιστούν με κάποια περιθώρια κατανομή. Στη συνέχεια θα μιλήσουμε για διαστήματα εμπιστοσύνης των εκτιμήσεων και για ασυμπτωτική θεωρία.

Ξεκινώντας ας κάνουμε τις ακόλουθες υποθέσεις. Υποθέτουμε ότι η σύζευξη που θα εκτιμήσουμε ανήκει σε μία οικογένεια $\{C_\theta, \theta \in \Theta\}$, όπου Θ ο χώρος των παραμέτρων. Έστω το τυχαίο διάνυσμα $Y = (Y_1, Y_2, ..., Y_d)$ με κοινή συνάρτηση κατανομής

$$F(y; a_1, a_2, ..., a_d; \theta) = C(F_1(y_1; a_1), F_2(y_2; a_2), ..., F_d(y_d; a_d); \theta)$$

όπου $F_1, F_2, ..., F_d$ μονοδιάστατες συναρτήσεις κατανομής με παραμέτρους $a_1, a_2, ..., a_d$.

Υποθέτουμε ότι η C έχει πυκνότητα c (μεικτές παραγώγους τάξης d) και συμβολίζουμε με f_j την περιθώρια συνάρτηση πυκνότητας πιθανότητας της Y_j για $j \in \{1, 2, ..., d\}$. Τότε το Y έχει συνάρτηση πυκνότητας πιθανότητας

$$f(y; a_1, a_2, ..., a_d; \theta) = c(F_1(y_1; a_1), F_2(y_2; a_2), ..., F_d(y_d; a_d); \theta) \prod_{j=1}^{d} f_j(y_j; a_j)$$

Για ένα δείγμα μεγέθους n με παρατηρήσεις τυχαία διανύσματα $Y_1, Y_2, ..., Y_n$ θεωρούμε τις συναρτήσεις πιθανοφάνειας για τις μονοδιάστατες περιθώριες συναρτήσεις

$$L_j(a_j) = \sum_{i=1}^{n} \log f_j(y_{ij}; a_j), \; j = 1, 2, ..., d\,,$$

και την συνάρτηση πιθανοφάνειας για την από κοινού συνάρτηση κατανομής,

$$L(a_1, a_2, ..., a_d; \theta) = \sum_{i=1}^{n} \log f(y_i; a_1, a_2, ..., a_d; \theta)$$

Όταν κάποιος εκτιμήσει το θ, έχει και εκτίμηση για την σύζευξη.

4.1.1 Η Μέθοδος IFM για περιθώριες

Η μέθοδος *IFM* (*Inference Function for Marginals Method*) αποτελείται από d ξεχωριστές βελτιστοποιήσεις των μονοδιάστατων πιθανοφανειών μαζί με μία βελτιστοποίηση της πολυδιάστατης πιθανοφάνειας ως συνάρτηση του διανύσματος της παραμέτρου εξάρτησης θ.

Η μέθοδος αυτή εκτελείται σε δύο βήματα.

1. οι πιθανοφάνειες $L_1(a_1), L_2(a_2), ..., L_d(a_d)$ των d μονοδιάστατων περιθώριων μεγιστοποιούνται ώστε να προκύψουν εκτιμήσεις $\hat{a}_1, \hat{a}_2, ..., \hat{a}_d$ των $a_1, a_2, ..., a_d$ αντίστοιχα.
2. η συνάρτηση $L(\hat{a}_1, \hat{a}_2, ..., \hat{a}_d; \theta)$ μεγιστοποιείται για θ ώστε να προκύψει εκτίμηση $\hat{\theta}$ της θ.

Κάτω από συνθήκες κανονικότητας δεχόμαστε ότι $(\hat{a}_1, \hat{a}_2, ..., \hat{a}_d; \hat{\theta})$ είναι η λύση της ισότητας

$$\left(\frac{\partial L_1}{\partial a_1}, \frac{\partial L_2}{\partial a_2}, ..., \frac{\partial L_d}{\partial a_d}, \frac{\partial L}{\partial \theta} \right) = 0$$

4.1.2 Η Μέθοδος μέγιστης πιθανοφάνειας

Η μέθοδος αυτή μας βοηθά να πάρουμε τις εκτιμήσεις $\hat{a}_1, \hat{a}_2, ..., \hat{a}_d, \hat{\theta}$ λύνοντας την ισότητα

$$\left(\frac{\partial L}{\partial a_1}, \frac{\partial L}{\partial a_2}, ..., \frac{\partial L}{\partial a_d}, \frac{\partial L}{\partial \theta} \right) = 0$$

4.1.3 Η Μέθοδος CML (*Canonical Maximum Likelihood*)

Η μέθοδος αυτή διαφέρει από την *IFM* μέθοδο καθώς δεν γίνονται υποθέσεις γύρω από την παραμετρική μορφή των περιθώριων κατανομών. Η διαδικασία εκτίμησης πραγματοποιείται σε δύο βήματα, τα οποία είναι

1. Μετασχηματίζουμε τις μεταβλητές $(y_1, y_2, ..., y_d)^i$ δείγματος μεγέθους *n*, σε ομοιόμορφες μεταβλητές $(\hat{u}_1, \hat{u}_2, ..., \hat{u}_d)^i$ χρησιμοποιώντας τις εμπειρικές κατανομές.
2. Εκτιμούμε τις παραμέτρους της σύζευξης μεγιστοποιώντας τη συνάρτηση

$$\sum_{i=1}^{n} \log c((\hat{u}_1, \hat{u}_2, ..., \hat{u}_d)^i ; \theta)$$

Για παράδειγμα μπορούμε να εκτιμήσουμε την παράμετρο *R* της *Gaussian* σύζευξης με την *CML* ή την *IFM* μέθοδο ως εξής

$$\hat{R}_{IFM/CML} = \frac{1}{n} \sum_{i=1}^{n} \varsigma_i^{\mathrm{T}} \varsigma_i$$

όπου $\varsigma_i = (\Phi^{-1}(u_1^i), ..., \Phi^{-1}(u_d^i))$.

Στην περίπτωση αυτή υποθέτουμε ότι $u_j^i = \hat{u}_j^i$ όταν χρησιμοποιούμε την μέθοδο *CML* και $u_j^i = F_j(y_j^i ; \hat{a}_j)$ όταν χρησιμοποιούμε την μέθοδο *IFM* με $j = 1, 2, ..., d$.

Η παρακάτω διαδικασία χρησιμοποιείται για την εκτίμηση της παραμέτρου *R* της t_v-*Student* σύζευξης.

1. Έστω $\hat{R}_1$ η *IFM*/*CML* εκτιμήτρια της παραμέτρου *R* για την *Gaussian* σύζευξη
2. $$\hat{R}_{m+1} = \frac{1}{n}\left(\frac{v+d}{v}\right) \sum_{i=1}^{n} \frac{\varsigma_i^{\mathrm{T}} \varsigma_i}{1 + \frac{1}{v} \varsigma_i \hat{R}_m^{-1} \varsigma_i^{\mathrm{T}}}, \qquad m = 1, 2, ...,$$

   όπου $\varsigma_i = (t_v^{-1}(u_1^i), ..., t_v^{-1}(u_d^i))$
3. Το βήμα δύο επαναλαμβάνεται εωσότου $\hat{R}_{m+1} = \hat{R}_m$.Οπότε η εκτίμηση της παραμέτρου *R* της t_v-*Student* σύζευξη είναι $\hat{R}_{IFM/CML} = \hat{R}_\infty$.

Οι *Marshal* και *Zeevi* (2002) προτείνουν τον παρακάτω αλγόριθμο για την εκτίμηση των παραμέτρων v και R της t_v-*Student* σύζευξης.

1. Μετασχηματίζουμε τις μεταβλητές $(y_1, y_2, ..., y_d)^i$ δείγματος μεγέθους *n*, σε ομοιόμορφες μεταβλητές $(\hat{u}_1, \hat{u}_2, ..., \hat{u}_d)^i$ χρησιμοποιώντας τις εμπειρικές κατανομές.

2. Κάνουμε εκτίμηση του $\hat{R}$ χρησιμοποιώντας το τ του *Kendall* :

$\hat{R}_{ij} = \sin\left(\frac{\pi}{2}\hat{\tau}_{ij}\right)$, με $i, j = 1, 2, ..., n$

3. Βρίσκουμε την εκτίμηση $\hat{v}$ μεγιστοποιώντας την παρακάτω συνάρτηση για $v \in (2, \infty]$

$$\sum_{i=1}^{n} \log c\left(u_1^i, ..., u_n^i; v, \hat{R}\right)$$

όπου

$$c\left(u_1, ..., u_d; v, \hat{R}\right) = \frac{\Gamma((v+d)/2[\Gamma(v+2)]^{d-1}(1+y' R^{-1} y)^{-(v+d)/2}}{|R|^{1/2}[\Gamma((v+1)/2]^d \prod_{i=1}^{d}(1+y_i^2/v)^{-(v+1)/2}}$$

και

$$\begin{aligned} y &= (y_1, ..., y_d) \\ &= (t_v^{-1}(u_1), ..., t_v^{-1}(u_d)) \end{aligned}$$

4.1.4 Μέθοδος Εκτίμησης Αρχιμήδειων Συζεύξεων

Έστω ένα δείγμα $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ τυχαίων διανυσμάτων όμοιων με το τυχαίο διάνυσμα *(X,Y)* και ας υποθέσουμε ότι η σύζευξη *C* συσχετισμένη με το *(X,Y)* είναι Αρχιμήδεια με παράμετρο *α*. Για να βρούμε μία εκτίμηση της παραμέτρου *α*, θα χρησιμοποιήσουμε το τ του *Kendall*. Θυμίζουμε ότι για Αρχιμήδειες συζεύξεις το τ του *Kendall* υπολογίζεται από τον τύπο

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

Η κλασική εκτιμήτρια του τ προκύπτει από τον τύπο

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} sign((X_i - X_j)(Y_i - Y_j))$$

Από τη στιγμή που το τ εκφράζεται ως προς τη φ και φ είναι συνάρτηση της παραμέτρου α, έπεται ότι η εκτίμηση $\hat{a}$ της α, προκύπτει λύνοντας την ισότητα

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

ως προς α.

4.1.5 Παραμετρική Εκτίμηση και μέτρα εξάρτησης

Η μέθοδος αυτή εφαρμόζεται μόνο σε μονοπαραμετρικά διδιάστατα *copulas*. Τα κυριότερα μέτρα εξάρτησης μπορούν να γραφούν σαν μία συνάρτηση της σύζευξης. Σε κάποιες περιπτώσεις υπάρχουν λύσεις ώστε η παράμετρος της σύζευξης να μπορεί εύκολα να γραφτεί σαν συνάρτηση ενός μέτρου εξάρτησης. Στις υπόλοιπες περιπτώσεις είναι απαραίτητο να ακολουθηθούν αριθμητικές διαδικασίες.

Π.χ. για την *Gaussian* σύζευξη έχουμε

$$R_{12} = 2\sin\left(\frac{\pi}{6}\rho_s\right)$$

και

$$R_{12} = \sin\left(\frac{\pi}{2}\tau\right)$$

Για την σύζευξη *Clayton*

$$\alpha = \frac{2\tau}{1-\tau}$$

Για την σύζευξη *Gumbel*

$$\alpha = (1-\tau)^{-1}$$

Για την σύζευξη *(FGM)*

$$\alpha = 3\rho_s$$

και

$$\alpha = \frac{9}{2}\tau$$

4.1.6 Η συνάρτηση εμπειρικής σύζευξης (μη-παραμετρική εκτίμηση)

Τώρα θα δώσουμε μία μη-παραμετρική μέθοδο εκτίμησης μιας διδιάστατης σύζευξης. Έστω ένα δείγμα $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$ τυχαίων διανυσμάτων όμοιων (*iid copies*) με το τυχαίο διάνυσμα *(X,Y)*. Τότε η διδιάστατη εμπειρική συνάρτηση κατανομής που σχετίζεται με το *(X,Y)* είναι

$$H_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x, Y_i \le y\}}$$

με περιθώριες

$$F_n(x) = H_n(x,-\infty) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}$$

και

$$G_n(x) = H_n(-\infty,y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i \le y\}}$$

όπου I_A είναι η δείκτρια συνάρτηση του συνόλου *A*.

Τότε η συνάρτηση εμπειρικής σύζευξης δίνεται από τον τύπο

$$\begin{aligned} C_n(u,v) &= H_n(F_n^{-1}(u), G_n^{-1}(v)) \\ &= \frac{1}{n}\sum_{k=1}^{n} I_{\left\{X_k \le F_n^{-1}(u), Y_k \le G_n^{-1}(v)\right\}} \end{aligned}$$

ή αλλιώς

$$C_n\left(\frac{i}{n},\frac{j}{n}\right) = \frac{\alpha\rho\iota\theta\mu\acute{o}\varsigma\ \zeta\varepsilon\upsilon\gamma\alpha\rho\iota\acute{\omega}\nu\ (x,y)\ \mu\varepsilon\ x \le x_{(i)}, y \le y_{(j)}}{n}$$

με $1 \le i, j \le n$

Σημειώνουμε η συνάρτηση εμπειρικής σύζευξης των $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$ είναι ίδια με αυτήν των ομοιόμορφων τυχαίων μεταβλητών $(U_1,V_1),(U_2,V_2),...,(U_n,V_n)$, όπου $U_i = F(X_i)$ και $V_i = G(Y_i)$ για κάθε $i \in \{1,2,...,n\}$.

Στην πολυδιάστατη περίπτωση τώρα, έστω $\{x_1^{(i)},...,x_d^{(i)}\}$ τα *order statistics* και $\{r_1^i,...,r_d^i\}$ τα *rank statistics* των δεδομένων, με $i=1,...,n$. Έχουμε $x_j^{(r_j^i)} = x_j^i$, με $j=1,...,d$.

Κάθε συνάρτηση

$$\hat{C}\left(\frac{i_1}{n},...,\frac{i_d}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} I_{\left[r_j^i \le i_j\right]}$$

ορισμένη στο

$$l = \left\{\left(\frac{i_1}{n},...,\frac{i_d}{n}\right) : 1 \le j \le d; i_j = 0,...,n\right\}$$

είναι μία εμπειρική σύζευξη.

4.2 Διαστήματα Εμπιστοσύνης

Στην προηγούμενη ενότητα περιγράψαμε τρόπους εκτίμησης παραμέτρων των συζεύξεων. Ωστόσο για να καλύψουμε ολοκληρωτικά την έννοια της εκτίμησης, θα πρέπει να ορίσουμε και διαστήματα εμπιστοσύνης για τις παραμέτρους αυτές.

Υποθέτουμε ότι είναι εύκολο να ορίσουμε δύο στατιστικούς δείκτες (*statistics*) t_1 και t_2 που χαρακτηρίζουν το δείγμα μας, τέτοιους ώστε αν θ είναι η παράμετρος η οποία θέλουμε να μελετήσουμε, να ισχύει

$$P[t_1 \le \theta \le t_2] = 1-a$$

όπου a είναι μία σταθερή πιθανότητα η οποία ονομάζεται συντελεστής εμπιστοσύνης. Τότε το διάστημα $[t_1,t_2]$ καλείται $1-a$ διάστημα εμπιστοσύνης για το θ. Τα t_1 και t_2 ονομάζονται άνω και κάτω όρια του διαστήματος εμπιστοσύνης.

4.3 Ασυμπτωτική Θεωρία

Εδώ θα παρουσιάσουμε τα ασυμπτωτικά αποτελέσματα των μεθόδων εκτίμησης των παραμέτρων των συζεύξεων που περιγράψαμε παραπάνω. Θα μιλήσουμε για την *iid* περίπτωση (*Independent and Identically Distributed Case*) καθώς και για μία προσέγγιση αντιμετώπισης συμμεταβλητών (*covariates*).

4.3.1 Η iid περίπτωση

Στην περίπτωση αυτή υποθέτουμε ότι οι συνθήκες κανονικότητας της ασυμπτωτικής θεωρίας μέγιστης πιθανοφάνειας ισχύουν για το πολυδιάστατο μοντέλο και τις περιθώριές του.

Έστω $\eta = (\alpha_1, \alpha_2, ..., \alpha_d; \theta)$ το διάνυσμα των παραμέτρων και έστω Ψ το διάνυσμα των συναρτήσεων εκτίμησης της ίδιας διάστασης με το η. Έστω ότι $Y, Y_1, Y_2, ..., Y_n$ είναι ανεξάρτητες και ισόνομα κατανεμημένες (*iid*) με συνάρτηση πυκνότητας πιθανότητας *f*.

Ας υποθέσουμε ότι η εκτιμήτρια της η, $\hat{\eta} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_d; \hat{\theta})$ δίνεται από τον παρακάτω τύπο

$$\sum_{i=1}^{n} \Psi(Y_i, \hat{\eta}) = 0$$

και $\dfrac{\partial \Psi^{\mathrm{T}}}{\partial \eta}$ ένας πίνακας $j \times k$ με στοιχεία $\dfrac{\partial \Psi_j(y, \eta)}{\partial \eta_k}$.

Έχει αποδειχθεί (*Joe* και *Xu*) ότι ο πίνακας ασυμπτωτικής συνδιασποράς $n^{\frac{1}{2}}(\hat{\eta} - \eta)^{\mathrm{T}}$, ο οποίος ονομάζεται και πίνακας πληροφορίας του *Godambe*, ισούται με

$$V = D_{\Psi}^{-1} M_{\Psi} (D_{\Psi}^{-1})^{T}$$

όπου,

$$D_{\Psi} = \frac{\partial \Psi^{\mathrm{T}}(Y, \eta)}{\partial \eta}$$

και

$$M_{\Psi} = E[\Psi^{\mathrm{T}}(Y,\eta)\Psi(Y,\eta)]$$

4.3.2 Η περίπτωση συμμεταβλητών (*covariates*)

Εδώ υποθέτουμε ότι έχουμε ανεξάρτητα αλλά όχι ισόνομα (non-identically) κατανεμημένα τυχαία διανύσματα Y_i με $i = 1,2,...,n$ και πυκνότητες $f_i(Y_i;a)$, όπου $a = (a_1, a_2, ..., a_d, \theta)$. Για να συμπεριλάβουμε συμμεταβλητές υποθέτουμε ότι $a_j = a_j(x,\gamma_j), j = 1,2,...,d$ και $\theta = t(x,\gamma_{d+1})$, όπου $a_1, a_2, ..., a_d, t$, είναι συνδετικές συναρτήσεις (*link functions*). Σε αντίθεση με την περίπτωση χωρίς συμμεταβλητές όπου είχαμε $f(y;a_1,a_2,...,a_d;\theta)$, εδώ παίρνουμε την συνάρτηση πυκνότητας πιθανότητας ως εξής

$$\begin{aligned} f_{Y|x}(y|x;\gamma) &= f(y;a_1(x,\gamma_1),a_2(x,\gamma_2),...,a_d(x,\gamma_d),t(x,\gamma_{d+1})) \\ &= c(F_1(y_1;a),F_2(y_2;a),...,F_n(y_n;a))\prod_{i=1}^{n} f_i(y_i;a) \end{aligned}$$

όπου F_i είναι η περιθώρια συνάρτηση κατανομής της Y_i με $i = 1,2,...,n$ και

$$a = (a_1(x,\gamma_1),a_2(x,\gamma_2),...,a_d(x,\gamma_d),t(x,\gamma_{d+1}))$$

Η εκτιμήτρια $\hat{\gamma} = (\hat{\gamma}_1,\hat{\gamma}_2,...,\hat{\gamma}_d,\hat{\gamma}_{d+1})$ της $\gamma = (\gamma_1,\gamma_2,...,\gamma_d,\gamma_{d+1})$ προκύπτει με τη μέθοδο μέγιστης πιθανοφάνειας κάτω από τις ορισμένες συνθήκες και έτσι το αποτέλεσμα ασυμπτωτικής κανονικότητας (*asymptotic normality result*) παίρνει την μορφή

$$n^{-\frac{1}{2}}V_n^{-\frac{1}{2}}(\hat{\gamma}-\gamma)^T \xrightarrow{d} N(0,I),$$

όπου

$$V_n = D_n^{-1}M_n(D_n^{-1})^T$$

με

$$D_n = n^{-1}\sum_{i=1}^{n} E\left[\frac{\partial \Psi^{\mathrm{T}}(Y_i,\gamma)}{\partial \gamma}\right]$$

και

$$M_n = n^{-1}\sum_{i=1}^{n} E\left[\Psi^{\mathrm{T}}(Y_i,\gamma)\Psi(Y_i,\gamma)\right]$$

Η προσέγγιση αυτή μας επιτρέπει να επεκτείνουμε την ασυμπτωτική θεωρία για τυχαία διανυσμάτα με συμμεταβλητές ή για τυχαίες συμμεταβλητές.

4.4 Επιλέγοντας τη «σωστή σύζευξη»

Αφού παρουσιάσαμε διάφορους τρόπους εκτίμησης συζεύξεων, στην ενότητα αυτή θα αναφερθούμε σε τρόπους επιλογής της κατάλληλης σύζευξης για τα εμπειρικά δεδομένα μας.

4.4.1 Επιλογή κατάλληλης σύζευξης με τη χρήση εμπειρικής σύζευξης

Έστω $\{C_k\}_{1\le k\le K}$ ένα σύνολο διαθέσιμων συζεύξεων. Επιλέγουμε τη σύζευξη C_k η οποία ελαχιστοποιεί την παρακάτω «απόσταση» βασισμένη στην διακριτή νόρμα L_n, μεταξύ της σύζευξης C_k και της εμπειρικής σύζευξης.

$$\hat{d}(\hat{C},C_k)=\left(\sum_{i=1}^{n}\cdots\sum_{j=1}^{n}\left[\hat{C}\left(\frac{i}{n},\dots,\frac{j}{n}\right)-C_k\left(\frac{i}{n},\dots,\frac{j}{n}\right)\right]^2\right)^{1/2}$$

Η απόσταση αυτή μπορεί επίσης να χρησιμοποιηθεί για την εκτίμηση του διανύσματος μίας δοθείσας σύζευξης $C(u;\theta)$. Η εκτίμηση της παραμέτρου θ, $\hat{\theta}$ προκύπτει αν ελαχιστοποιήσουμε την σχέση

$$\left(\sum\left[\hat{C}(u)-C(u;\theta)\right]^2\right)^{1/2}.$$

4.4.2 Επιλέγοντας μία Αρχιμήδεια σύζευξη

Η μέθοδος που θα περιγράψουμε και έχει αναπτυχθεί από τους *Genest* και *Rivest* (1993), έχει ως στόχο την επιλογή Αρχιμήδειας σύζευξης με την καλύτερη δυνατή προσαρμογή σε πραγματικά δεδομένα. Μία Αρχιμήδεια σύζευξη έχει αναλυτικό τύπο

$$C(u_1,...,u_n)=\psi^{-1}[\psi(u_1)+...+\psi(u_n)].$$

Άρα είναι προφανές ότι για να επιλέξουμε μία σύζευξη, αρκεί να προσδιορίσουμε την γεννήτριά του, ψ.

Για την διδιάστατη περίπτωση, οι *Genest* και *Rivest* όρισαν μία μονοδιάστατη συνάρτηση που συνδέεται με τη γεννήτρια της Αρχιμήδειας σύζευξης μέσα από τον παρακάτω τύπο

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

Μία μη-παραμετρική εκτίμηση της παραπάνω σχέσης είναι η

$$\hat{K}(z)=\frac{1}{n}\sum_{i=1}^{n}I_{[\theta_i\le z]}$$

όπου,

$$\theta_j=\frac{1}{n-1}\sum_{i=1}^{n}I_{\left[x_1^i<x_1^j,\,x_2^i<x_2^j\right]},\ j=1,...,n.$$

Επιλέγουμε μία παραμετρική αναπαράσταση για τη γεννήτρια ψ. Τότε η παράμετρος α της Αρχιμήδειας σύζευξης που επιλέξαμε μπορεί να εκτιμηθεί από τον τύπο

$$\tau=\binom{T}{2}^{-1}\sum_{i<j}sign((x_1^i-x_1^j)(x_2^i-x_2^j)),$$

όπως είδαμε και παραπάνω.

Η παράμετρος α μπορεί επίσης να εκτιμηθεί με τις μεθόδους *IFM* και *CML*.

Το βέλτιστο *copula* μπορεί να επιλεχθεί ελαχιστοποιώντας την απόσταση στην νόρμα L_2 μεταξύ των $K(z)$ και $\hat{K}(z)$ η οποία είναι

$$d(\hat{K},K)=\int_0^1\left[K(z)-\hat{K}(z)\right]^2dz$$

4.5 Έλεγχοι καλής προσαρμογής (*Goodness-of-Fit Tests*)

Οι έλεγχοι καλής προσαρμογής ή *Goodness-of-Fit Tests* αποτελούν πολύ χρήσιμο εργαλείο στην στατιστική διότι δίνουν συμπεράσματα για το αν ένα μοντέλο είναι κατάλληλο ή όχι για να περιγράψει τα δεδομένα μας. Στις συζεύξεις ειδικότερα και σε περίπτωση που το μοντέλο μας προσεγγίζεται από περισσότερες της μίας συζεύξεις, οι έλεγχοι αυτοί μπορούν με ασφάλεια να μας υποδείξουν ποια από αυτές περιγράφει καλύτερα τι μοντέλο μας.

4.5.1 Το X^2 του Pearson

Ο έλεγχος X^2 βασίζεται σε αυθαίρετη κατηγοριοποίηση δεδομένων ώστε να υπολογιστεί ο στατιστικός συντελεστής του ελέγχου. Εδώ υποθέτουμε ότι τα δεδομένα μας κατηγοριοποιούνται με ίδια πιθανότητα σε όλες τις διαδικασίες ώστε να ελαχιστοποιήσουμε το σφάλμα. Επιπλέον υποθέτουμε ότι τα δεδομένα μας έχουν μετασχηματιστεί ώστε να κατηγοριοποιηθούν σε ομοιόμορφες μεταβλητές οι οποίες είναι ασυσχέτιστες κάτω από την μηδενική υπόθεση, δηλαδή $C_\theta(u,v)=uv$.

Ο κλασικός υπολογισμός για το χ^2 του *Pearson*, είναι ο ακόλουθος

$$\chi^2=\sum_{i=1}^{n}\frac{\left(N_i^{obs}-N_i^{\exp}\right)^2}{N_i^{\exp}}$$

όπου N_i^{obs}, $N_i^{\exp}$ ο αριθμός των «πραγματικών» και «προσδοκώμενων» παρατηρήσεων στην *i*-οστή κατηγορία αντίστοιχα. Για να υπολογίσουμε το *p*-value, απλά σημειώνουμε ότι οι βαθμοί ελευθερίας είναι ίσοι με τις κατηγορίες που δεν είναι κενές και έτσι το *p*-value προκύπτει από στατιστικούς πίνακες.

Αν και ο έλεγχος αυτός είναι ο πιο διαδεδομένος τρόπος ελέγχου καλής προσαρμογής, θα πρέπει να τονιστεί το γεγονός ότι η κατηγοριοποίηση των δεδομένων είναι εντελώς αυθαίρετη και μπορεί να οδηγήσει σε διαφορετικά *p*-values, ανάλογα με την μέθοδο κατηγοριοποίησης που έχει χρησιμοποιηθεί. Για το λόγο αυτό θα περιγράψουμε παρακάτω πέντε διαφορετικές μεθόδους κατηγοριοποίησης για τον έλεγχο χ^2 του *Pearson*.

1. Τομείς (*Sectors*)

Κάθε ζευγάρι παρατηρήσεων έχει διανεμηθεί σε έναν από οχτώ τομείς ανάλογα το πρόσημο της ποσότητας $u_i - 0.5$ και της αναλογίας $r_i = \frac{v_i - 0.5}{u_i - 0.5}$. Κατά αυτόν τον τρόπο μετράμε τη συσπείρωση (*clustering*) γύρω από συγκεκριμένες τιμές του r.

2. Ομόκεντρα Τετράγωνα (*Concentric Squares*)

Κάθε ζευγάρι παρατηρήσεων τοποθετείται σε μία από δέκα κατηγορίες (τετράγωνα) με κέντρο το $(0.5, 0.5)$. Κατά αυτόν τον τρόπο μετράμε τη συσπείρωση γύρω από το κέντρο του συνόλου των δεδομένων.

3. Τεταρτοκύκλια (*Quadrants*)

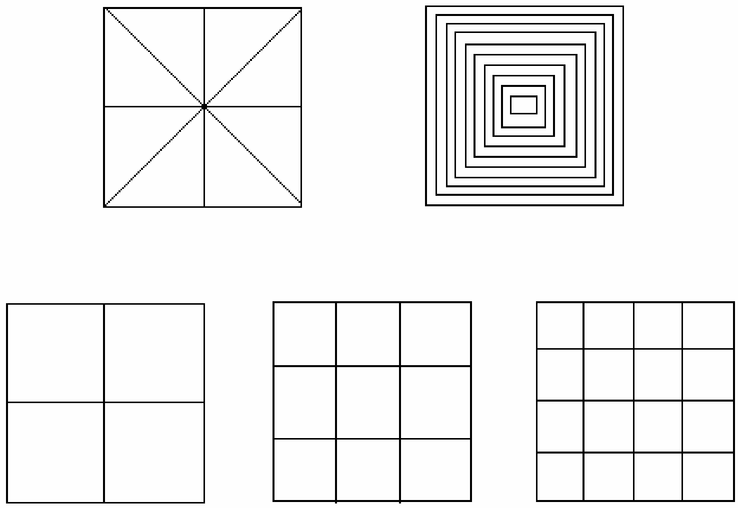
Θεωρώντας σαν προσδιοριστικό σημείο ή αρχή το σημείο $(0.5, 0.5)$, κάθε ζευγάρι παρατηρήσεων είναι τοποθετημένο σε μία από τέσσερις κατηγορίες ανάλογα με το τεταρτοκύκλιο στο οποίο βρίσκεται το ζευγάρι αυτό.

4. 3Χ3 Τετράγωνα

Κάθε ζευγάρι παρατηρήσεων βρίσκεται σε μία από εννιά κατηγορίες.

5. 4Χ4 Τετράγωνα

Κάθε ζευγάρι παρατηρήσεων βρίσκεται σε μία από δεκαέξι κατηγορίες.



Ωστόσο, για την εξισορρόπηση των περιορισμών που επιβλήθηκαν από την μελέτη της αυθαίρετης κατηγοριοποίησης, πρέπει επίσης να εκτελέσουμε δύο ελέγχους για την ισότητα συνεχών κατανομών, κατάλληλα τροποποιημένους για την από κοινού εμπειρική αθροιστική συνάρτηση κατανομής (*cdf*).

4.5.2 Ο Έλεγχος Kolmogorov-Smirnov

Ο έλεγχος των *Kolmogorov-Smirnov* για την ισότητα αθροιστικών συναρτήσεων κατανομών είναι ένας από τους πιο χρησιμοποιημένους ελέγχους για την ισότητα κατανομών. Παρόλο που ο έλεγχος των *Kolmogorov-Smirnov* είναι λιγότερο ισχυρός από άλλους (Π.χ. ο έλεγχος *Anderson-Darling*), παραμένει εύκολος στην εφαρμογή και τον υπολογισμό του *p-value* που σχετίζεται με τον στατιστικό συντελεστή, αφού η κατανομή είναι γνωστή, σε αντίθεση με τον συντελεστή του ελέγχου *Anderson-Darling*.

Έστω $Y_{(i)}$ τα διατεταγμένα δεδομένα μας, τότε ο συντελεστής των *Kolmogorov-Smirnov* ορίζεται ως

$$D^* = \max_{1 \le i \le n} \left\{ \left| F(Y_{(i)} - \frac{i-1}{n}) \right|, \left| F(Y_{(i)} - \frac{i}{n}) \right| \right\}$$

Το p-value που προκύπτει από το στατιστικό αυτό μπορεί εύκολα να υπολογιστεί από το παρακάτω άθροισμα

$$1-Q_{KS}=1-2\sum_{i=1}^{\infty}(-1)^{n}e^{-2i^2\lambda^2}$$

όπου

$$\lambda=\left(\sqrt{n}+0.11+\frac{0.12}{\sqrt{n}}\right)D^{*}$$

μέχρι το άθροισμα να συγκλίνει σε ένα ικανοποιητικό βαθμό ακριβείας.

Εφαρμόζουμε τον έλεγχο *Kolmogorov-Smirnov* στην προσαρμοσμένη αθροιστική συνάρτηση κατανομής αλλά και στην ομοιόμορφη μεταβλητή που προέκυψε από τον μετασχηματισμό των δεδομένων. Μόνο η μία περιθώρια αθροιστική συνάρτηση κατανομής πρέπει να ελεγχθεί καθώς η άλλη περιθώρια είναι απλά η εμπειρική αθροιστική συνάρτηση κατανομής της αρχικής περιθώριας και άρα είναι ομοιόμορφα κατανεμημένη.

4.5.3 Ο Έλεγχος Anderson-Darling

Ο έλεγχος των *Anderson-Darling* χρησιμοποιείται ευρέως στην βιβλιογραφία για τον έλεγχο της κανονικότητας παρόλο που έχουν αναπτυχθεί πίνακες κρίσιμων τιμών για την εκθετική, την *Weibull*, την λογαριθμοκανονική και άλλες κατανομές. Ωστόσο δεν έχει προσδιοριστεί ακόμα σχέση όπως αυτή του ελέγχου *Kolmogorov-Smirnov* και έτσι τα *p-values* που σχετίζονται με το στατιστικό παράγονται συνήθως μέσα από την τεχνική προσομοίωσης *Monte Carlo*.

Ο συντελεστής του ελέγχου *Anderson-Darling* είναι στην ουσία ένα ολοκλήρωμα. Ωστόσο επειδή η εμπειρική αθροιστική συνάρτηση κατανομής εκφράζεται σαν άθροισμα, μπορούμε να εκφράσουμε το ολοκλήρωμα αυτό σαν ένα πεπερασμένο διακριτό άθροισμα. Η διακριτή μορφή του στατιστικού του ελέγχου *Anderson-Darling* είναι

$$A^2=-N-\sum_{i=1}^{N}\frac{2i-1}{N}\left\{\ln F(Y_{(i)})+\ln\left[1-F(Y_{(n+1-i)})\right]\right\}$$

Ο συντελεστής του ελέγχου *Anderson-Darling* χρησιμοποιείται για «πρόχειρη εκτίμηση» και δεν μπορεί να καθορίσει το κατά πόσο μία σύζευξη είναι σημαντικά διαφορετική από την εμπειρική σύζευξη.

Σε πολλές πηγές πάντως δίνονται αρκετές παραλλαγές των παραπάνω ελέγχων και τα στατιστικά των ελέγχων διαμορφώνονται ως εξής

Απόσταση *Kolmogorov – Smirnov*

$$KS = \max_X \left| \hat{F}(\hat{\theta}) - F_T(\theta) \right|$$

Απόσταση *Anderson - Darling*

$$AD = \max_X \frac{\left| \hat{F}(\hat{\theta}) - F_T(\theta) \right|}{\sqrt{F_T(\theta) * [1 - F_T(\theta)]}}$$

όπου $\{X_1, ..., X_T\}$ σετ δεδομένων με παραμέτρους $\theta = \{\theta_1, ..., \theta_T\}$, συνάρτηση κατανομής $\hat{F}(\hat{\theta})$ και εμπειρική συνάρτηση κατανομής $F_T(\theta)$.

ΚΕΦΑΛΑΙΟ 5

5 Προσομοιώσεις και Στατιστικές Εφαρμογές Συζεύξεων

5.1 Προσομοιώσεις Συζεύξεων

Στην ενότητα αυτή θα παρουσιάσουμε κάποιους αλγόριθμους προσομοίωσης τυχαίων μεταβλητών $(u_1,...,u_n)$ για συγκεκριμένες συζεύξεις. Οι μεταβλητές που χρησιμοποιούμε είναι συσχετισμένες και ομοιόμορφα κατανεμημένες. Για να προσομοιώσουμε τις τυχαίες μεταβλητές $x_1,...,x_n$ που ακολουθούν μια πολυδιάστατη κατανομή F με περιθώριες F_i, $i=1,2,...,n$ και γνωστή σύζευξη C, κάθε μεταβλητή u_i θα πρέπει να αντιστραφεί με βάση τις περιθώριες συναρτήσεις κατανομών έτσι ώστε

$$x_i = F_i^{-1}(u_i), i=1,2,...,n.$$

5.1.1 Προσομοίωση της FMG Οικογένειας Συζεύξεων

Ο ακόλουθος αλγόριθμος παράγει διδιάστατες τυχαίες μεταβλητές από τις συζεύξεις της οικογένειας *Farlie – Gumbel – Morgenstern*.

1. Παράγουμε τις ανεξάρτητες και ομοιόμορφα κατανεμημένες στο $(0,1)$ τυχαίες μεταβλητές v_1 και v_2.
2. Θέτουμε $u_1 = v_1$.
3. Υπολογίζουμε τις παρακάτω σχέσεις
$A = \alpha(2u_1 - 1) - 1$ και $B = [1 - \alpha(2u_1 - 1)]^2 + 4\alpha v_2(2u_1 - 1)$
4. Θέτουμε $u_2 = \dfrac{2v_2}{\sqrt{B} - A}$
5. Το διάνυσμα (u_1, u_2) παράγεται από τη σύζευξη *Farlie–Gumbel– Morgenstern*

5.1.2 Προσομοίωση των Συζεύξεων Marshal – Olkin

Έστω τα $l := |S| = 2^n - 1$ μη-κενά υποσύνολα του $\{1,...,n\}$, $s_1,...,s_l$ και ας θέσουμε $\lambda_k := \lambda_{s_k}$ (η παράμετρος του χρόνου Z_{s_k}) για $k = 1,...,l$. Ο ακόλουθος αλγόριθμος παράγει τυχαίες μεταβλητές από την πολυδιάστατη σύζευξη *Marshal* – Olkin.

1. Παράγουμε τις l ανεξάρτητες τυχαίες μεταβλητές $v_1,...,v_l$ που ακολουθούν την $U(0,1)$.

2. Θέτουμε $x_i = \min_{1\le k\le l, i\in s_k, \lambda_k \ne 0} \left(-\ln v_k / \lambda_k\right), i = 1,...,n$.

3. Θέτουμε $\Lambda_i = \sum_{k=1}^{l} 1\{i \in s_k\}\lambda_k, i = 1,...,n$

   όπου Λ_i, η «ένταση» του πλήγματος που «αισθάνθηκε» το στοιχείο *i*.

4. Θέτουμε $u_i = \exp\left(-\Lambda_i x_i\right), i = 1,...,n$.

5. Το διάνυσμα $(x_1,...,x_n)^T$ παράγεται από την n-διάστατη κατανομή *Marshal – Olkin* και το διάνυσμα $(u_1,...,u_n)^T$ παράγεται από την αντίστοιχη *n*-διάστατη *Marshal – Olkin* σύζευξη.

5.1.3 Προσομοίωση της Gaussian Σύζευξης

Για την παραγωγή τυχαίων μεταβλητών από την *Gaussian* σύζευξη μπορούμε να ακολουθήσουμε την εξής διαδικασία. Αν ο πίνακας *R* είναι θετικά ορισμένος, τότε υπάρχουν $n \times n$ πίνακες *A* τέτοιοι ώστε $R = AA^T$. Επιπλέον υποθέτουμε ότι οι τυχαίες μεταβλητές $Z_1,...,Z_n$ είναι ανεξάρτητες και ακολουθούν την τυποποιημένη κανονική κατανομή. Τότε το τυχαίο διάνυσμα $\mu + AZ$ με $Z = (Z_1,...,Z_n)^T$ και $\mu \in \mathbf{R}^n$, ακολουθεί την πολυδιάστατη κανονική κατανομή με διάνυσμα μέσων τιμών μ και πίνακα συνδιασπορών R.

Ο πίνακας *A* μπορεί να προκύψει εύκολα με τη μέθοδο *Cholesky* (*Cholesky decomposition of R*). Η «ανάλυση» αυτή είναι ένας μοναδικός κάτω-τριγωνικός πίνακας *L*, έτσι ώστε $LL^T = R$. Η παραγωγή τυχαίων μεταβλητών από *n*-διάστατες *Gaussian* συζεύξεις γίνεται με τον ακόλουθο αλγόριθμο

1. Βρίσκουμε την «ανάλυση» *Cholesky A* του πίνακα *R*.

2. Παράγουμε n ανεξάρτητες μεταβλητές $z = (z_1, ..., z_n)^T$ που ακολουθούν την τυποποιημένη κανονική κατανομή.

3. Θέτουμε $x = Az$

4. Ορίζουμε τις μεταβλητές u_i θέτοντας $u_i = \Phi(x_i), i = 1, ..., n$, όπου *Φ* είναι η μονοδιάστατη συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής.

5. Το διάνυσμα $(u_1, ..., u_n)^T$ παράγεται από την *n*-διάστατη *Gaussian* σύζευξη, C_R^{Ga} .

5.1.4 Προσομοίωση της Σύζευξης t_v-Student

Για να προσομοιώσουμε μεταβλητές από τη σύζευξη t_v-*Student*, $C_{v,R}^t$,χρησιμοποιούμε τον παρακάτω αλγόριθμο που βασίζεται στην σχέση

$$X = \mu + \frac{\sqrt{v}}{\sqrt{S}} Z$$

όπου $\mu \in R^n$, $S \sim \chi_v^2$ και το τυχαίο διάνυσμα $Z \sim N(0, \Sigma)$ είναι ανεξάρτητα.

1. Βρίσκουμε τον πίνακα *A* όπως προηγουμένως.

2. Παράγουμε *n* ανεξάρτητες μεταβλητές $z = (z_1, ..., z_n)^T$ που ακολουθούν την τυποποιημένη κανονική κατανομή.

3. Παράγουμε μία τυχαία μεταβλητή s που ακολουθεί την κατανομή χ_v^2 και είναι ανεξάρτητη από το *z*.

4. Ορίζουμε το διάνυσμα $y = Az$.

5. Θέτουμε $x = \frac{\sqrt{v}}{\sqrt{s}} y$

6. Ορίζουμε τις μεταβλητές $u_i = t_v(x_i), i = 1, ..., n$

7. Το διάνυσμα $(u_1, ..., u_n)^T$ παράγεται από την *n*-διάστατη *Gaussian* σύζευξη, $C_{v,R}^t$.

5.1.5 Προσομοίωση της Σύζευξης Cook – Johnson

Η σύζευξη *Cook – Johnson* αποτελεί μία πολυδιάστατη επέκταση των Αρχιμήδειων συζεύξεων με από κοινού συνάρτηση κατανομής σύζευξης

$$C(u_1,...,u_n) = \left(\sum_{j=1}^{n} u_j^{-a} - n + 1 \right)^{-1/a}, \ a > 0 .$$

Ο παρακάτω αλγόριθμος αποτελεί ειδική περίπτωση εκείνου που πρότειναν οι Marshal και *Olkin* (1988) για την παραγωγή πολυδιάστατων αποτελεσμάτων από μία σύνθετη σύζευξη και συντελεί στην παραγωγή τυχαίων μεταβλητών από τη σύζευξη *Cook – Johnson* με παράμετρο *α*, ακολουθώντας τα παρακάτω βήματα.

1. Παράγουμε *n* ανεξάρτητες τυχαίες μεταβλητές $y_1,...,y_n$ από την εκθετική κατανομή ($F(x;\lambda) = 1 - e^{-\lambda x}, x > 0$) με παράμετρο $\lambda = 1$.
2. Παράγουμε μία τυχαία μεταβλητή z από την κατανομή $\text{Gamma}(1/a,1)$, ανεξάρτητη από τις $y_1,...,y_n$.
3. Θέτουμε $u_j = (1 + y_j / z)^{-1/a}, j = 1,...,n$.
4. Το διάνυσμα $u = (u_1,...,u_n)$ παράγεται από τη σύζευξη *Cook – Johnson*.

Η σύζευξη *Cook – Johnson* παράγει μεταβλητές με θετική εξάρτηση. Για να πάρουμε μεταβλητές με αρνητική εξάρτηση, πρέπει να ορίσουμε κάποιες μεταβλητές ως $u_i^* = 1 - u_i$.

5.1.6 Προσομοίωση της Εμπειρικής Σύζευξης

Ο παρακάτω αλγόριθμος επιτρέπει την παραγωγή διανυσμάτων τυχαίων μεταβλητών από την εμπειρική σύζευξη.

1. Θεωρούμε ένα διάνυσμα $(x_1^t,...,x_n^t)$ από το δείγμα μας.
2. Χρησιμοποιώντας τις εμπειρικές συναρτήσεις κατανομών $\hat{F}_i$, τροποποιούμε τα στοιχεία του διανύσματος ώστε να προκύψουν οι ομοιόμορφες μεταβλητές $u_i = \hat{F}_i(x_i^t), i = 1,...,n$.

3. Το $(u_1,...,u_n)$ είναι ένα διάνυσμα από ομοιόμορφες $(0,1)$ μεταβλητές οι οποίες είναι εξαρτημένες μέσα στην εμπειρική σύζευξη.

5.1.7 Γενικός τρόπος Προσομοίωσης μιας Σύζευξης

Η μέθοδος αυτή βασίζεται στις δεσμευμένες κατανομές ενός τυχαίου διανύσματος $U=(U_1,...,U_n)$. Στην διδιάστατη περίπτωση έχουμε

$$P[U_2 \le u_2 \,|U_1 = u_1] = C_{2|1}(u_1,u_2)$$

όπου

$$\begin{aligned} C_{2|1}(u_1,u_2) &= \lim_{\Delta u \to 0^+} \frac{C(u_1+\Delta u,u_2)-C(u_1,u_2)}{\Delta u} \\ &= \frac{C(u_1,u_2)}{\partial u_1} \end{aligned}$$

Ο αλγόριθμος είναι ο ακόλουθος

1. Παράγουμε δύο ανεξάρτητες ομοιόμορφες τυχαίες μεταβλητές v_1 και v_2
2. Θέτουμε $u_1 = v_1$
3. Έστω $C(u_2;u_1) = C_{2|1}(u_1,u_2)$.Θέτουμε $u_2 = C^{-1}(v_2;u_1)$.
4. Το διάνυσμα (u_1,u_2), παράγεται από τη σύζευξη C.

Π.χ. Για τη διδιάστατη σύζευξη *Frank* έχουμε

$$C_{2|1}(u_1,u_2) = \frac{\left(e^{-au_2}-1\right)e^{-au_1}}{\left(e^{-a}-1\right)+\left(e^{-au_1}-1\right)\left(e^{-au_2}-1\right)}$$

και

$$C^{-1}(u;u_1) = -\frac{1}{a}\ln\left(1+\frac{u\left(e^{-a}-1\right)}{u+\left(1-u\right)e^{-au_1}}\right).$$

Ο παραπάνω αλγόριθμος μπορεί να γενικευτεί και για την πολυδιάστατη περίπτωση.

1. Παράγουμε δύο ανεξάρτητες ομοιόμορφες τυχαίες μεταβλητές $(v_1,...,v_n)$.

2. Θέτουμε $u_1 = v_1$

3. Έστω $C(u_m; u_1,...,u_{m-1}) = C_{m|1}(u_1,...,u_m)$, $m = 2,...,n$ όπου

$$C_{m|1}(u_1,...,u_m) = P[U_m \le u_m \left|(U_1,...,U_{m-1}) = (u_1,...,u_{m-1})\right.]$$
$$= \frac{\partial^{m-1}_{(u_1,...,u_{m-1})} C(u_1,...,u_m,1,...,1)}{\partial^{m-1}_{(u_1,...,u_{m-1})} C(u_1,...,u_{m-1},1,...,1)}$$

4. Θέτουμε $u_m = C^{-1}(v_m; u_1,...,u_{m-1})$, $m = 2,...,n$

5. Το διάνυσμα $(u_1,...,u_n)$ παράγεται από τη σύζευξη C.

Ο αλγόριθμος αυτός είναι αρκετά πολύπλοκος μιας και η παραπάνω πολυδιάστατη δεσμευμένη συνάρτηση σύζευξης είναι δύσκολο να υπολογιστεί για μεγάλες τιμές του *n*.

5.2 Μελέτη Προσομοιώσεων

Στην ενότητα αυτή θα προσομοιώσουμε τις συζεύξεις *Frank* και *Clayton* μέσω του *Mathematica* και θα μελετήσουμε τα αποτελέσματα των προσομοιώσεων αυτών διερευνώντας κάποιες ακραίες περιπτώσεις.

5.2.1 Προσομοίωση και μελέτη αποτελεσμάτων της σύζευξης Frank

Με βάση τον γενικό τρόπο προσομοίωσης συζεύξεων που μελετήσαμε παραπάνω, ο αλγόριθμος προσομοίωσης της σύζευξης *Frank* στο *Mathematica*, είναι ο εξής

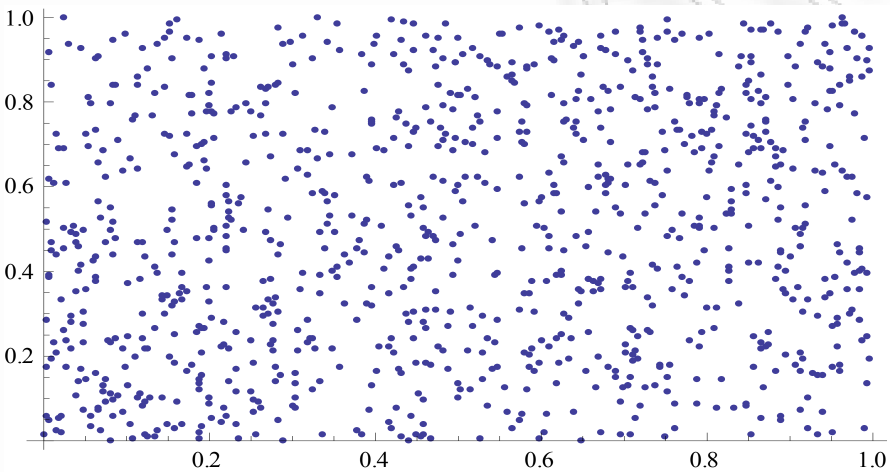
```
θ=1;n=1000;
list={};
Do[
  U1=Random[];
  U2=Random[];
  X1=U1;
```

```
X2=-(1/θ)*Log[1+(U2*(1-Exp[-θ])/(U2*(Exp[-θ*X1]-1)-Exp[-θ*X1]))];
AppendTo[list,{X1,X2}];
,{n}];
```

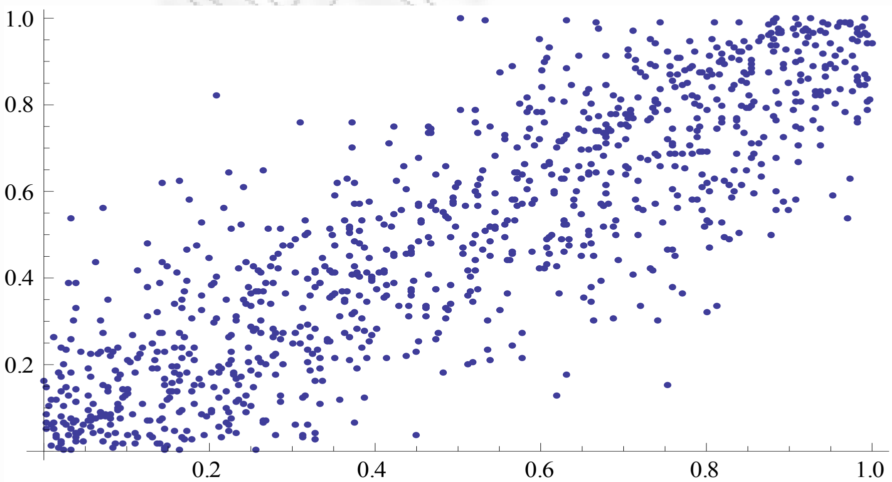
```
ListPlot[list]
```

Τα γραφικά αποτελέσματα που πήραμε για τρεις διαφορετικές τιμές του θ δίνονται παρακάτω.

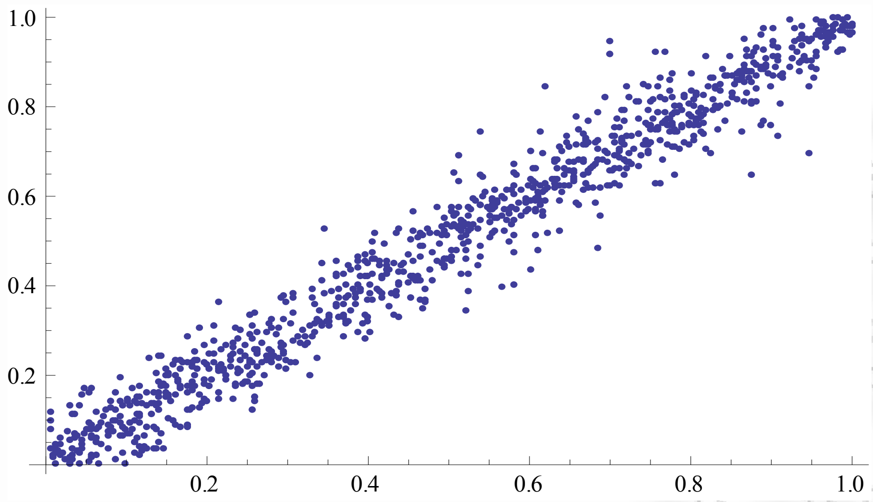
θ=1



θ=10



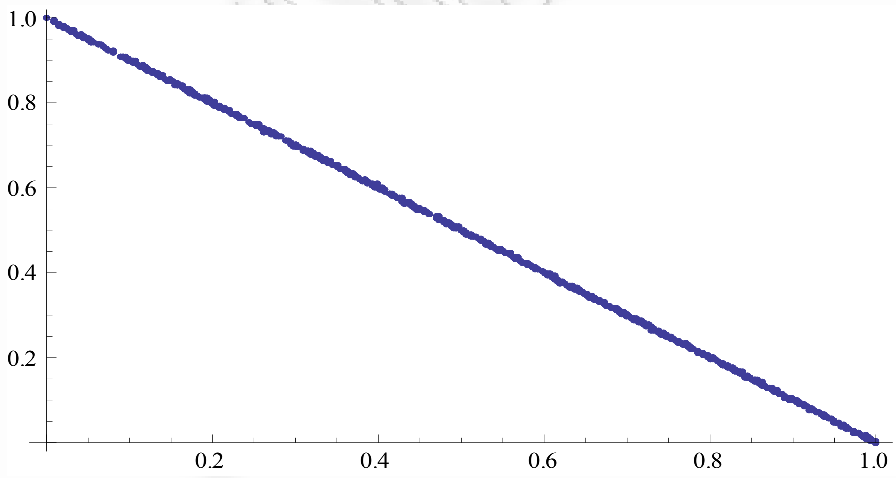
θ=30



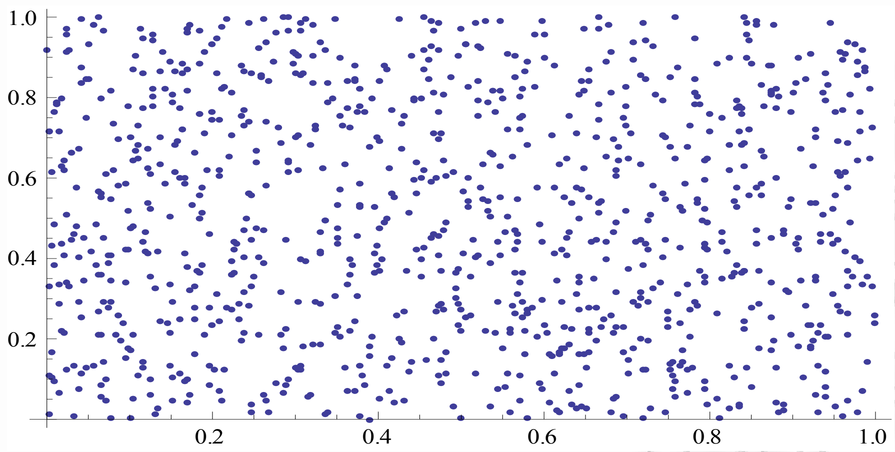
Ακραίες περιπτώσεις για την σύζευξη Frank

Από το τρίτο κεφάλαιο είδαμε ότι οι ακραίες περιπτώσεις της σύζευξης *Frank* είναι $C_{-\infty} = W$, $C_0 = \Pi$, $C_{\infty} = M$. Οι περιπτώσεις αυτές δεν είναι εύκολο να προσομοιωθούν καθώς για τις τιμές αυτές του θ δεν ορίζεται η σύζευξη *Frank*. Ωστόσο, για πολύ οριακές τιμές του *θ* έχουμε

$\boldsymbol{\theta \to -\infty}$



$\theta \to 0$



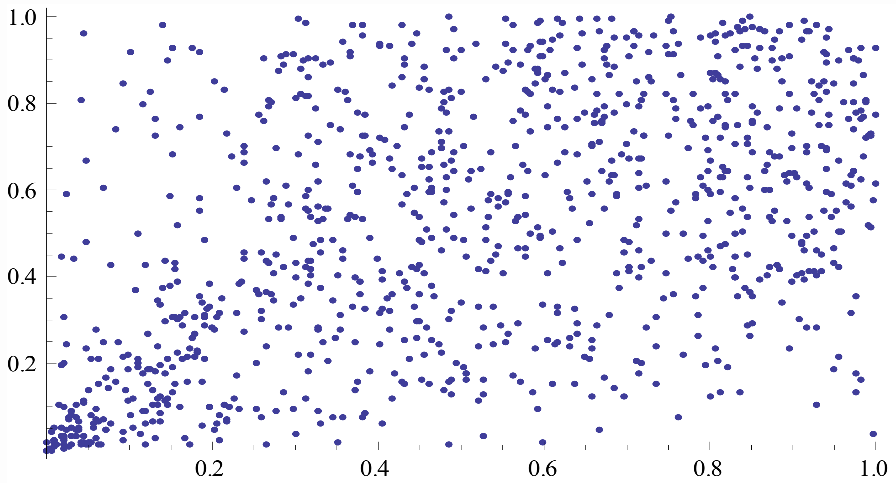
5.2.2 Προσομοίωση και μελέτη αποτελεσμάτων της σύζευξης Clayton

Ο αλγόριθμος προσομοίωσης της σύζευξης *Clayton* στο *Mathematica*, είναι ο εξής

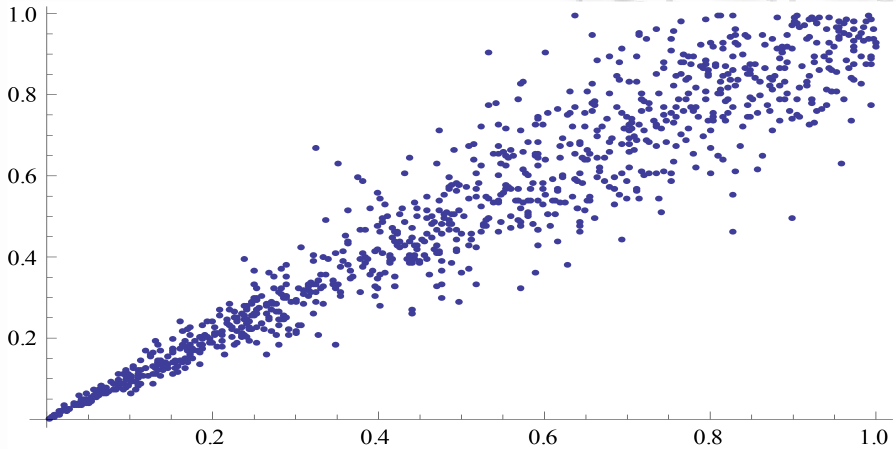
```
θ=1;n=1000;
list={};
Do[
U1=Random[];
 U2=Random[];
 X1=U1;
 X2=((U1^(-θ))*((U2^(-θ/(θ+1)))-1)+1)^(-1/θ);
 AppendTo[list,{X1,X2}];
,{n}];

ListPlot[list]
```

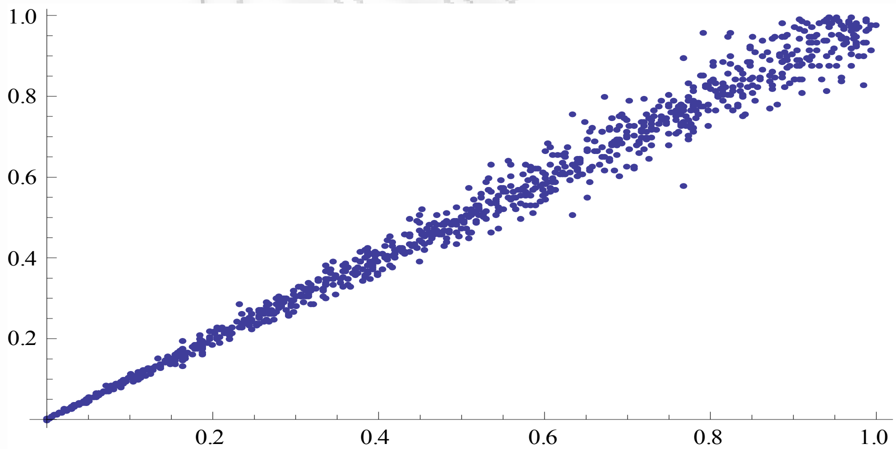
θ=1


1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0

θ=10


1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0

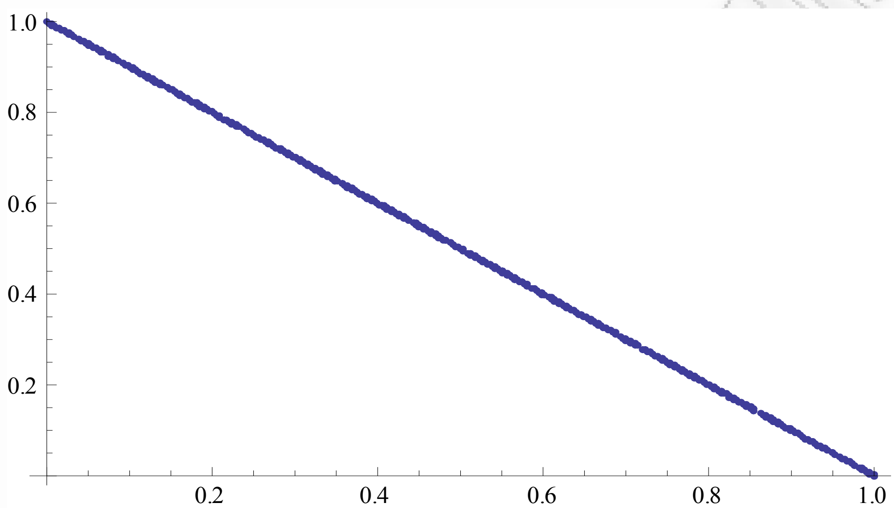
θ=30


1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0

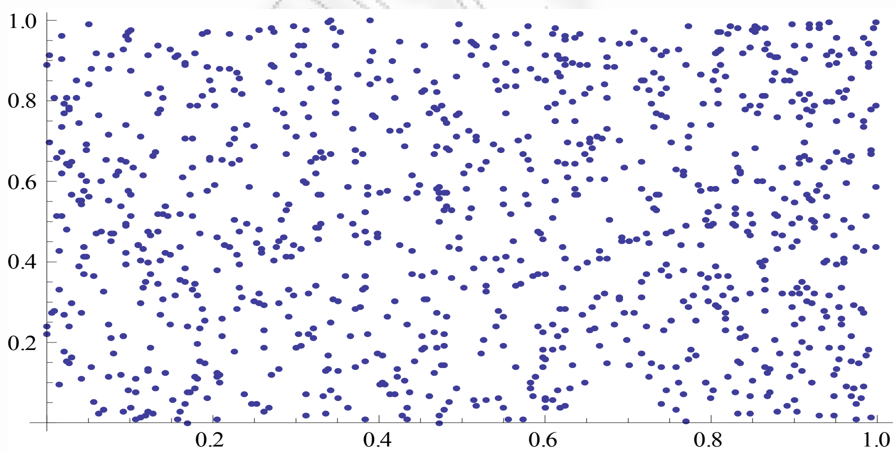
Ακραίες περιπτώσεις για την σύζευξη Clayton

Οι ακραίες περιπτώσεις της σύζευξης *Clayton* είναι $C_{-1} = W$, $C_0 = \Pi$, $C_1 = \dfrac{\Pi}{\Sigma - \Pi}$, $C_\infty = M$ οπότε έχουμε

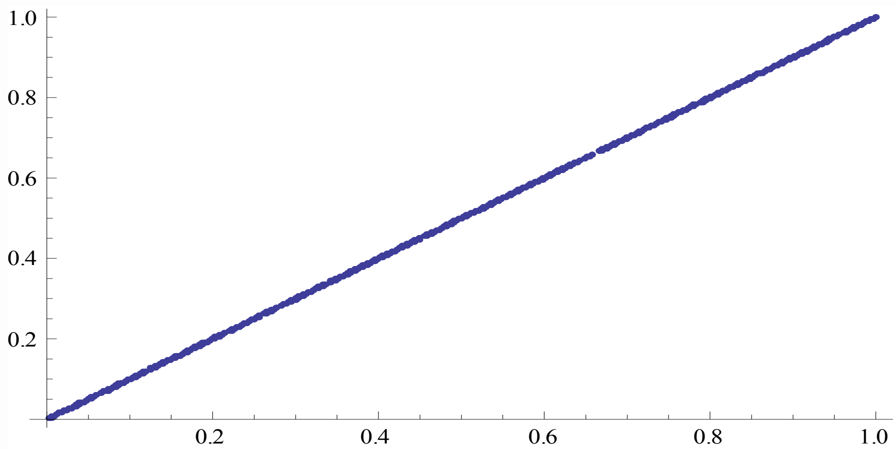
θ→-1



θ→0



θ → ∞



5.2.3 Συμπερασματολογία

Από τα παραπάνω διαγράμματα μπορούμε να διαπιστώσουμε ότι όσο μεγαλώνει το θ, τόσο μεγαλύτερη φαίνεται να είναι η εξάρτηση μεταξύ των μεταβλητών. Αυτό φαίνεται κι από τους εμπειρικούς τύπους για τον υπολογισμό μέτρων εξάρτησης όπως το τ του *Kendall*.

Π.χ. Για θ=6 έχουμε για τη σύζευξη *Frank*

$$\tau = 1 + \frac{4}{\theta}(D_1(\theta) - 1), \qquad \text{με } D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t) - 1}dt$$

και άρα από το *Mathematica* έχω ότι $\tau \approx 0.514$.

Προσομοιώνοντας τη σύζευξη *Frank* για θ=6 και μεταφέροντας το δείγμα στο *Excel*, παίρνουμε με την βοήθεια του εργαλείου *Analyse-it* (*Analyse>Correlation>Kendall*) ότι *tau statistic* = 0.516 το οποίο είναι πολύ κοντά στο 0.514 του εμπειρικού τύπου.

Ακολουθώντας την ίδια διαδικασία για τη σύζευξη *Clayton* βρίσκουμε ότι *tau statistic* = 0.754 το οποίο είναι πολύ κοντά στο εμπειρικό $\tau = \frac{\theta}{\theta + 2} = 0.75$ για θ=6.

5.3 Στατιστικές εφαρμογές των συζεύξεων

Οι συζεύξεις έχουν πολλές εφαρμογές, ειδικά στην στατιστική επιστήμη. Στην ενότητα αυτή θα δώσουμε δύο παραδείγματα τέτοιων εφαρμογών.

5.3.1 Επιβίωση πολλαπλών ζωών (*Survival of multiple lives*)

Στην επιδημιολογία και στις αναλογιστικές μελέτες συχνά εξετάζεται η κοινή θνησιμότητα συνόλων μεγαλύτερων του ενός ατόμου. Π.χ.Αντρόγυνο ή γονείς και παιδιά.

Αρχικά ας δώσουμε τον ορισμό της συνάρτησης κινδύνου ενός τυχαίου χρόνου επιβίωσης.

$$\begin{aligned} h(t) &= -\frac{\partial \log S(t)}{\partial t} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

όπου F είναι η συνάρτηση κατανομής του χρόνου T και S, η οποία ορίζεται ως

$$\begin{aligned} S(t) &= P\left[T > 1\right] \\ &= 1 - F(t) \end{aligned}$$

η συνάρτηση επιβίωσης του T.

Χρησιμοποιώντας το *proportional hazards model* (αναλογικό μοντέλο κινδύνου) του *Cox* (1972), η συνάρτηση κινδύνου εκφράζεται ως

$$h(t, z) = e^{\beta z} b(t)$$

όπου $b(t)$ θα λέμε ότι είναι η αρχική συνάρτηση κινδύνου (“baseline” hazard function) και β είναι ένα διάνυσμα παραμέτρων παλινδρόμησης.

Στην παραπάνω σχέση αν αντικαταστήσουμε το $e^{\beta z}$ με γ έχουμε

$$\begin{aligned} h(t,z) &= \gamma b(t) \\ &= -\frac{\partial \log S(t|\gamma)}{\partial t} \end{aligned}$$

και ολοκληρώνοντας προκύπτει,

$$\partial \log S(t|\gamma) = -\gamma \int_0^t b(s)ds$$

και επομένως

$$S(t|\gamma) = \exp(-\gamma \int_0^t b(s)ds)$$

Άρα $S(t|\gamma) = B(t)^{\gamma}$ όπου $B(t) = \exp(-\int_0^t b(s)ds)$, είναι η συνάρτηση επιβίωσης εκφρασμένη ως προς την αρχική συνάρτηση κινδύνου. Το γ θα ονομάζεται ασθενικότητα αφού όσο μεγαλύτερο είναι, τόσο μικρότερη είναι η $S(t|\gamma)$.

Ας υποθέσουμε τώρα ότι οι p χρόνοι ζωής $T_1, T_2, ..., T_p$ είναι ανεξάρτητοι με δεδομένη ασθενικότητα γ. Τότε η από κοινού πολυδιάστατη συνάρτηση επιβίωσης ορίζεται ως εξής

$$P[T_1 > t_1, T_2 > t_2, ..., T_p > t_p] = E_{\gamma}(P[T_1 > t_1, T_2 > t_2, ..., T_p > t_p|\gamma])$$

Όμως

$$\begin{aligned} P[T_1 > t_1, T_2 > t_2, ..., T_p > t_p|\gamma] &= P[T_1 > t_1|\gamma] \cdot P[T_2 > t_2|\gamma] \cdot ... \cdot P[T_p > t_p|\gamma] \\ &= B_1(t_1)^{\gamma} \cdot B_2(t_2)^{\gamma} \cdot ... \cdot B_p(t_p)^{\gamma} \\ &= [B_1(t_1) \cdot B_2(t_2) \cdot ... \cdot B_p(t_p)]^{\gamma} \end{aligned}$$

οπότε έχουμε

$$\begin{aligned} P[T_1 > t_1, T_2 > t_2, ..., T_p > t_p] &= E_{\gamma}([B_1(t_1) \cdot B_2(t_2) \cdot ... \cdot B_p(t_p)]^{\gamma}) \\ &= E_{\gamma}\{\exp[\gamma \log(B_1(t_1) \cdot B_2(t_2) \cdot ... \cdot B_p(t_p))]\} \end{aligned}$$

Για τον μετασχησμό *Laplace* που ορίζεται ως $E_{\gamma}(e^{-S\gamma})$ βρίσκουμε ότι

$E_{\gamma}(e^{-s\gamma}) = \exp(-s^a)$ αν η γ έχει θετική σταθερή κατανομή τάξης (*order*) a . Έτσι προκύπτει ο ακόλουθος τύπος

$$S_i(t_i) = \exp\{-[-\log B_i(t_i)]^a\}$$

και επομένως,

$$P[T_1 > t_1, T_2 > t_2, ..., T_p > t_p] = \exp\{-\gamma[-\log B_1(t_1) - \log B_2(t_2) - ... - \log B_p(t_p)]\}$$
$$= \exp\{[(-\log S_1(t_1))^{\frac{1}{a}} + (-\log S_2(t_2))^{\frac{1}{a}} + ... + (-\log S_p(t_p))^{\frac{1}{a}}]^a\}$$

Η παραπάνω σχέση αποτελεί σύζευξη.Από τη στιγμή που γνωρίζουμε τις ιδιότητες της σύζευξης αυτής, η σχέση μεταξύ των χρόνων ζωής $T_1, T_2, ..., T_p$ μπορεί εύκολα να κατανοηθεί.

5.3.2 Οι Συζεύξεις στην θεωρία Ακραίων Τιμών

Σκοπός της θεωρίας ακραίων τιμών είναι να συμβάλλει στην μοντελοποίηση ακραίων κινδύνων, δηλαδή συμβάντα τα οποία πραγματοποιούνται με πολύ μικρή πιθανότητα, αλλά μπορεί να έχουν καταστροφικές συνέπειες. Η θεωρία αυτή έχει εφαρμογές στα χρηματοοικονομικά, στα ασφαλιστικά, στην μηχανική και σε άλλους τομείς. Μια σύγχρονη μέθοδος ώστε να κατανοήσουμε την πολυδιάστατη θεωρία ακραίων τιμών είναι μέσα από τις συζεύξεις. Για να είμαστε πιο ακριβείς, τα αποτελέσματα από την μελέτη της θεωρίας αυτής μπορούν να εκφραστούν σε όρους συζεύξεων.

Έστω $(X_{j,1}, X_{j,2})$, με $j = 1, 2, ...$, μια ακολουθία τυχαίων μεταβλητών στο $\mathbf{R} \times \mathbf{R}$ με από κοινού συνάρτηση κατανομής H και περιθώριες συναρτήσεις κατανομών F_{X_i}, με $i = 1, 2$.Έστω (M_1, M_2) σημείο τέτοιο ώστε $M_i = \max\{X_{1,i}, X_{2,i}, ..., X_{N,i}\}$, με $i = 1, 2$.Η διδιάστατη εκδοχή του θεωρήματος των *Fisher – Tippet* λέει ότι αν υπάρχουν κανονικοποιημένες σταθερές $a_{iN} > 0$ και $b_{iN} \in \mathbf{R}, i = 1, 2$ τέτοιες ώστε η από κοινού κατανομή

$$P\left[\frac{M_1 - b_{1N}}{a_{1N}} \le x_1, \frac{M_2 - b_{2N}}{a_{2N}} \le x_2\right] = H^N\left(a_{1N}x_1 + b_{1N}, a_{2N}x_2 + b_{2N}\right)$$

να συγκλίνει καθώς το N τείνει στο ∞, σε μια κατανομή $W(x_1, x_2)$ με μη-εκφυλισμένες περιθώριες, τότε η $W(x_1, x_2)$ είναι μία διδιάστατη κατανομή ακραίων τιμών. Αυτό σημαίνει πρακτικά, ότι η H ανήκει στο MDA της W και άρα $F_{X_i} \in MDA(W_i)$, όπου W_i με $i = 1, 2$

είναι κατανομές ακραίων τιμών. Επιπλέον η $W(x_1, x_2)$ πρέπει να ικανοποιεί την ακόλουθη συνθήκη.

Για κάθε $N \geq 1$ υπάρχουν σταθερές $a_{iN} > 0$ και $b_{iN} \in \mathbf{R}, i = 1,2$ τέτοιες ώστε

$$W^N(x_1, x_2) = W(a_{1N}x_1 + b_{1N}, a_{2N}x_2 + b_{2N})$$

<u>Ορισμός</u>

Ο μέγιστος χώρος έλξης (*Maximum Domain of Attraction*) ή *MDA* που αναφέραμε παραπάνω, για οποιαδήποτε κατανομή *G*, είναι το σύνολο όλων των κατανομών οι οποίες καθώς το *n* τείνει στο ∞ για το κανονικοποιημένο μέγιστο $\frac{M_n - b_n}{a_n}$, συγκλίνουν στην *G* για κάποιες σταθερές a_n και b_n.

<u>Θεώρημα</u>

Έστω $(X_{11}, X_{12}), (X_{21}, X_{22}), ..., (X_{N1}, X_{N2}), ...,$ μια ακολουθία ανεξάρτητων και ισόνομα κατανεμημένων τυχαίων μεταβλητών με από κοινού συνάρτηση κατανομής *H*. Υποθέτουμε ότι υπάρχουν κανονικοποιημένες ακολουθίες $a_{iN} > 0$, $a'_{iN} > 0$, $b_{iN}, b'_{iN} \in \mathbf{R}, i = 1,2$ τέτοιες ώστε

$$H^N\left(a_{1N}x_1 + b_{1N}, a_{2N}x_2 + b_{2N}\right) \to W(x_1, x_2)$$

$$\text{όταν } N \to \infty$$

και

$$H^N\left(a'_{1N}x_1 + b'_{1N}, a'_{2N}x_2 + b'_{2N}\right) \to W'(x_1, x_2)$$

$$\text{όταν } N \to \infty$$

για τις μη-εκφυλισμένες συναρτήσεις $W(x_1, x_2)$ και $W'(x_1, x_2)$.

Τότε υπάρχουν $a_{X_1}, a_{X_2}, \beta_{X_1}, \beta_{X_2}$, τέτοια ώστε ,

$$W_1(x_1) = W'_1(a_{X_1}x_1 + \beta_{X_1})$$

και

$$W_2(x_2) = W'_2(a_{X_2}x_2 + \beta_{X_2})$$

Επιπλέον οι συζεύξεις C_W και $C_{W'}$, ταυτίζονται.

Με αυτό το αποτέλεσμα, οι συζεύξεις μπορούν να μας βοηθήσουν να κατανοήσουμε κάποια ακραία συμβάντα.

Έστω μια σύζευξη *C* με από κοινού συνάρτηση κατανομής *H*. Τότε από την συνέχεια των κατανομών ακραίων τιμών, προκύπτει ότι υπάρχει μοναδική σύζευξη C_* τέτοια ώστε

$$W(x_1, x_2) = C_*(W_1(x_1), W_2(x_2))$$

Ο *Deheuvels* έδειξε ότι

$$C_*(u_1, u_2) = \lim_{N\to\infty} C^N(u_1^{\frac{1}{N}}, u_2^{\frac{1}{N}})$$

και τότε η C_* ικανοποιεί τη σχέση

$$C_*(u_1^t, u_2^t) = C_*^t(u_1, u_2)$$

για κάθε $t > 0$.

Η παραπάνω σχέση δίνει τον ορισμό της σύζευξη ακραίων τιμών. Οι συζεύξεις αυτές έχουν εφαρμογές στις ασφαλιστικές επιστήμες, στον περιβάλλον και στα χρηματοοικονομικά. Για παράδειγμα, η διοίκηση κινδύνου περιλαμβάνει τεχνικές ακραίων τιμών. Πολύ διαδεδομένη στην θεωρία και τις συζεύξεις ακραίων τιμών είναι η εξάρτηση ουράς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Arnold H. (2006): “Dependence Modelling via the Copula Method”, Quantitative Risk Management Group CSIRO Mathematical and Information Sciences
2. 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.
3. Cook R.D and Johnson M.E.(1981) : “A Family of Distributions for Modeling Non-Elliptically Symmetric Multivariate Data”. Journal of the Royal Statistical Society.
4. Damarta S. (2002): “Extreme Value Theory and Copulas”. Master’s thesis, Department of Mathematics. ETH Zurich, Switzerland.
5. Denneberg, D. (1994): “Non-Additive Measure And Integral”. Kluwer Academic Publishers, Boston.
6. Deheuvels P. (1984): “Probabilistic Aspects of Multivariate Extremes”. In Statistical Extremes and Applications. Reidel Publishing Compan
7. Durrleman, V., A. Nikeghbali and T. Roncalli (2000): “Which copula is the right one? ”, Groupe de Recherche Opérationnelle, Crédit Lyonnais, Working Paper.
8. Embrechts, P., F. Lindskog and A. J. Mcneil (2001): “Modelling dependence with copulas and applications to risk management”, ETH Zurich, preprint.
9. Embrechts, P., T. Mikosch, and C. Kluppelberg (1997): “Modelling Extremal Events for Insurance and Finance”. Springer, Berlin.
10. Fang, K.-T., S. Kotz, and K.-W. Ng (1987): “Symmetric Multivariate and Related Distributions”. Chapman & Hall, London.
11. Frees, E. W. and E. A. Valdez (1998): “Understanding relationships using copulas”, North American Actuarial Journal.
12. Genest C. and Rivest L.(1993) : “Statistical Inference Procedures for Archimedean Copulas”. Journal of the American Statistical Association.
13. Gumbel E.J. (1960):“Bivariate Exponential Distributions ”,Journal of American Statistical Association.
14. Joe, H. (1997) : “Multivariate Models and Dependence Concepts”, Chapman & Hall, London, chs 2&5.

15. JoeE, H. and J. J. Xu (1996): “The estimation method of inference functions for margins for multivariate models”, Department of Statistics, University of British Columbia, Technical Report 166.
16. Johnson, M. E. (1987): “Multivariate Statistical Simulation”, Wiley, New York.
17. Johnson, N., and S. Kotz (1972): “Distributions in Statistics: Continuous Multivariate Distributions”. Wiley, New York.
18. Kendall, M., and A. Stuart (1979): “Handbook of Statistics”. Griffin & Company, London.
19. Kolev N., dos Anjos U., and Beatriz Vaz de M. Mendes (2006):“Copulas: A Review and Recent Developments”. Stochastic Models, Taylor & Francis Group.
20. Tchilabalo Abozou Kpanzou (2006): “Copulas in Statistics”, African Institute for Mathematical Sciences (AIMS)
21. Lehmann, E. (1975): “Nonparametrics: Statistical Methods Based on Ranks”. Holden-Day, Inc., San Francisco.
22. Li, D. (1999): “On Default Correlation: A Copula Function Approach”, working paper, RiskMetrics Group, New York.
23. Lindskog, F. (2000): “Modelling Dependence with Copulas”, ETH Zurich.
24. Marshall A.W and Olkin (1967):“A Generalized Bivariate Exponential Distribution. Applied Probability”.
25. Mikosch T. Copulas(2006): “Tales and facts. Springer Science,Business Media, LLC.
26. Nelsen, R. (1999): “An Introduction to Copulas. Springer, New York.
27. Quasada-Molina J.J. (2003): “ What are Copulas? ” Monografias Del Semi. Matem, Garcia de Galdeono.
28. Claudio Romano(2002): “Calibrating and Simulating Copula Functions: An Application to The Italian Stock Market”, working paper.
29. Schweizer, B., and A. Sklar (1983): “Probabilistic Metric Spaces”. North-Holland, New York.
30. Sklar, A. (1996): “Random variables, distribution functions, and copulas,Institute of Mathematical Statistics”, Hayward, CA.