

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ**

**ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ
ΣΕ ΜΟΝΤΕΛΑ ΕΞΑΡΤΗΜΕΝΩΝ
ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ**

Λάμπρος Δ. Κατσάπας

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος
Ειδίκευσης στην Αναλογιστική επιστήμη και Διοικητική
Κινδύνου

**Πειραιάς,
Ιούλιος 2010**

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ**

**ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ
ΣΕ ΜΟΝΤΕΛΑ ΕΞΑΡΤΗΜΕΝΩΝ
ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ**

Λάμπρος Δ. Κατσάπας

Διπλωματική Εργασία

*που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος
Ειδίκευσης στην Αναλογιστική επιστήμη και Διοικητική
Κινδύνου*

**Πειραιάς,
Ιούλιος 2010**

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

Τα μέλη της Επιτροπής ήταν:

- (Επιβλέπων)
-
-

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

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

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

**EXTREME VALUE THEORY FOR DEPENDENT
RANDOM VARIABLES**

BY

Labros D. Katsapas

MSc Dissertation

Submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfillment of the
requirements for the degree of Master of Science in Actuarial
Science and Risk Management

Piraeus, Greece

2010

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΔΙΑ

*Στους γονείς μου
Δημήτρη και Χριστίνα*

РАНЕЕЗНАМО ПЕРПАА

Ευχαριστίες

Ευχαριστώ τον κύριο Μ. Μπούτσικα για την πολύτιμη καθοδήγησή του καθ' όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

Περίληψη

Η Θεωρία Ακραίων Τιμών αποτελεί ξεχωριστό κλάδο της στατιστικής με στόχο την ανάπτυξη θεωρητικών και στατιστικών μοντέλων για την μελέτη προβλημάτων που σχετίζονται με την εμφάνιση ακραίων παρατηρήσεων. Στην παρούσα διατριβή αρχικά παρουσιάζουμε τις βασικές έννοιες και αποτελέσματα της θεωρίας για ακολουθίες ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Επίσης, εισάγονται κάποιες έννοιες από την θεωρία σημειακών διαδικασιών (*point processes*) η οποία αποτελεί ένα σημαντικό πεδίο για την ανάπτυξη θεωρητικών αλλά και στατιστικών αποτελεσμάτων της συγκεκριμένης θεωρίας. Στη συνέχεια υποθέτουμε ότι η ακολουθία των τυχαίων μεταβλητών σχηματίζει μία στάσιμη στοχαστική ανέλιξη και επεκτείνουμε τα αποτελέσματα της θεωρίας για αυτή την περίπτωση. Τέλος, επικεντρώνομαστε στην εκτίμηση μίας σημαντικής παραμέτρου που εκφράζει την ακρότατη εξάρτηση μεταξύ των τ.μ. της στοχαστικής διαδικασίας και ονομάζεται δείκτης ακρότατης εξάρτησης (*extremal index*).

Abstract

Extreme Value Theory (EVT) is a branch of Statistics that concerns theoretical and statistical model developments for the study of problems relevant to the occurrence of extreme events. There are plenty of applications of EVT in many research areas, for example in Hydrology, Meteorology, Geology, Insurance and Finance. The largest part of the associated bibliography deals with results concerning sequences of i.i.d. (independent and identically distributed) random variables. Although, the assumption of independence is very convenient for the development of the relevant theory, there exist many cases where this assumption is not so plausible, raising doubts about the reliability of the results. In the present dissertation we present extreme value results assuming that the observations form a stationary random process. Initially, we present the basic notions and results of the theory for i.i.d sequences of random variables and we also review the concept of point processes. Next we develop the associated theory for stationary sequences of random variables along with alternative approaches through the theory of point processes. Finally, we concentrate on the estimation of an important parameter that reflects the extremal dependence of stationary sequences, known as *extremal index*.

Περιεχόμενα

Εισαγωγή στη θεωρία ακραίων τιμών	1
1 Θεωρία Ακραίων Τιμών για ακολουθίες ανεξάρτητων και ισόνομων τυχαίων μεταβλητών	4
1.1 Ασυμπτωτικές κατανομές μεγίστου	4
1.2 Max stable κατανομές – GEV κατανομή	9
1.3 Ασυμπτωτικές κατανομές ελαχίστου	11
1.4 Περιοχή έλξης κατανομής μεγίστου (Maximum domain of attraction)	12
1.5 Γενικευμένη κατανομή Pareto (Generalized Pareto distribution)	18
1.6 Οριακή κατανομή της r μεγαλύτερης παρατήρησης	21
1.7 Προσέγγιση μέσω Σημειακών στοχαστικών διαδικασιών	22
2 Θεωρία Ακραίων Τιμών για ισχυρά στάσιμες ακολουθίες τυχαίων μεταβλητών	29
2.1 Εισαγωγή	29
2.2 Ασυμπτωτικές κατανομές μεγίστου	32
2.3 Δείκτης ακρότατης εξάρτησης	45
2.4 Προσέγγιση μέσω Σημειακών στοχαστικών διαδικασιών	51
2.5 Στατιστικές συναρτήσεις συστάδων (cluster statistics)	57
3 Εκτίμηση του δείκτη ακρότατης εξάρτησης	65
3.1 Εισαγωγή	65
3.2 Μέθοδος των ομάδων	66
3.3 Μέθοδος των ροών παρατηρήσεων κάτω από ένα συγκεκριμένο κατώφλι u	71
3.4 Μέθοδος των ενδιάμεσων χρόνων	75
3.5 Μελέτη μέσω Monte Carlo προσομοίωσης	78
Βιβλιογραφία	89

Εισαγωγή στη θεωρία ακραίων τιμών

Η θεωρία ακραίων τιμών αποτελεί ξεχωριστό κλάδο της στατιστικής με στόχο την ανάπτυξη θεωρητικών και στατιστικών μοντέλων για την μελέτη προβλημάτων που σχετίζονται με την εμφάνιση ακραίων παρατηρήσεων («πολύ μεγάλων» ή «πολύ μικρών»). Οι απαρχές του κλάδου βρίσκονται στην από κοινού εργασία των Ronald Fisher και Leonard Tippett «*Limiting forms of the frequency distribution of the largest or smallest member of a sample*» (1928). Σε αυτή την εργασία διατυπώνονται οι τρεις οριακές κατανομές της μέγιστης παρατήρησης ενός δείγματος ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Μέχρι και την όγδοη δεκαετία του περασμένου αιώνα το ενδιαφέρον της επιστημονικής κοινότητας για την ανάπτυξη της θεωρίας ακραίων τιμών ήταν σχετικά μικρό κάτι που είχε ως αποτέλεσμα την δημοσίευση ελάχιστων εργασιών πάνω στο συγκεκριμένο αντικείμενο. Ουσιαστικά η ανάπτυξη του κλάδου ξεκινά το 1970 έπειτα από την διδακτορική διατριβή του Laurens de Haan με τίτλο «*On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*».

Μέχρι σήμερα υπάρχουν δύο κυρίως προσεγγίσεις στη θεωρία ακραίων τιμών οι οποίες και εξαρτώνται από την διαδικασία μέσα από την οποία καταγράφονται τα δεδομένα. Η πρώτη προσέγγιση βασίζεται στο θεώρημα των Fisher and Tippett (Θεώρημα 1.1.1) και στην μελέτη της μέγιστης παρατήρησης του δείγματος. Από άποψη στατιστικής ανάλυσης, η διαδικασία που ακολουθείται σύμφωνα με αυτή την προσέγγιση είναι να χωρίσουμε τα δεδομένα μας σε ομάδες και να βασιστούμε στις μέγιστες παρατηρήσεις ανά ομάδα (**blocks method**). Βασική προϋπόθεση αυτής της μεθόδου είναι να έχουμε στην διάθεσή μας τα πλήρη δεδομένα του υπό μελέτη φαινομένου ή τουλάχιστον τις μέγιστες παρατηρήσεις ανά χρονικές περιόδους (π.χ μέγιστη παρατήρηση ανά εβδομάδα ή μήνα). Η δεύτερη μέθοδος έχει ως στόχο τη μελέτη της ουράς της κατανομής των παρατηρήσεων με βάση τις υπερβάσεις πάνω από μία προκαθορισμένη τιμή (κατώφλι). Η μέθοδος αυτή βασίζεται στο αποτέλεσμα του Θεωρήματος 1.5.2 (**Peaks over threshold method**) και συνήθως εφαρμόζεται όταν έχουμε στη διάθεσή μας μόνο τις παρατηρήσεις πάνω από κάποια τιμή.

Το πεδίο εφαρμογών της θεωρίας ακραίων τιμών είναι αρκετά μεγάλο. Ενδεικτικά μπορούμε να αναφέρουμε την επιστήμη της υδρολογίας, της μετεωρολογίας, της γεωλογίας, της ασφάλισης καθώς επίσης και των χρηματοοικονομικών. Για παράδειγμα στην υδρολογία, με σκοπό την κατασκευή φραγμάτων, μείζονος σημασίας πρόβλημα είναι η εκτίμηση της στάθμης των υδάτων ενός ποταμού η οποία δεν θα ξεπεραστεί για τα επόμενα εκατό χρόνια με κάποια συγκεκριμένη πιθανότητα (π.χ $p = 0.01$), καθώς μία πλημύρα μπορεί να επιφέρει σημαντικές καταστροφές στις καλλιέργειες και στην οικονομική δραστηριότητα ενός τόπου. Στην ασφάλιση, μία εντελώς απρόσμενη μεγάλη ζημιά σε ένα χαρτοφυλάκιο κινδύνων, όπως για παράδειγμα οι καταστροφές που ακολουθούν έπειτα από έναν σεισμό, τυφώνα, αεροπορικό δυστύχημα κ.α, μπορεί να έχει ως αποτέλεσμα την αδυναμία της ασφαλιστικής εταιρίας να ανταπεξέλθει στις υποχρεώσεις της. Σκόπιμη λοιπόν είναι η εκτίμηση εκείνου του ποσού το οποίο δεν θα υπερβεί καμία απαίτηση ζημιάς (για το επόμενο έτος) με κάποια συγκεκριμένη πιθανότητα. Σε τέτοιου είδους προβλήματα αλλά και σε πολλά άλλα τα οποία έχουν σχέση με την εκτίμηση παραμέτρων που αφορούν ακραία φαινόμενα, η θεωρία ακραίων τιμών αποτελεί ένα σύγχρονο επιστημονικό «εργαλείο» για την επίλυσή τους. Μία πιο αναλυτική περιγραφή των εφαρμογών της θεωρίας ακραίων τιμών μπορεί να βρεθεί στο βιβλίο *Statistics of Extremes* (Beirlant et al.(2004)).

Αν και στο μεγαλύτερο μέρος της βιβλιογραφίας παρουσιάζονται αποτελέσματα της θεωρίας ακραίων τιμών για ακολουθίες ανεξάρτητων και ισόνομων τυχαίων μεταβλητών, η υπόθεση της ανεξαρτησίας αποτελεί ένα σημαντικό πρόβλημα όσον αφορά την αξιοπιστία των αποτελεσμάτων που μπορούν να προκύψουν. Πιο ρεαλιστικό είναι να υποθέσουμε ότι η ακολουθία των τυχαίων μεταβλητών σχηματίζει μία στάσιμη στοχαστική ανέλιξη και να προσπαθήσουμε να επεκτείνουμε (εάν είναι δυνατόν) τα αποτελέσματα της θεωρίας για αυτή την περίπτωση. Αυτό είναι και το κύριο αντικείμενο της παρούσας εργασίας. Το πλήθος των δημοσιευμένων εργασιών πάνω στο συγκεκριμένο αντικείμενο είναι αρκετά μεγάλο και ενδεικτικές αναφορές παρατίθεται στο τέλος της εργασίας.

Στο πρώτο κεφάλαιο παρουσιάζουμε τις βασικές έννοιες και αποτελέσματα της θεωρίας για ακολουθίες ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Επίσης, στο τέλος του πρώτου κεφαλαίου εισάγονται κάποιες έννοιες από την θεωρία των σημειακών διαδικασιών (*point processes*) η οποία αποτελεί ένα σημαντικό πεδίο για την ανάπτυξη θεωρητικών αλλά και στατιστικών αποτελεσμάτων της θεωρίας ακραίων τιμών. Στο δεύτερο κεφάλαιο

αναπτύσσεται η αντίστοιχη θεωρία που αφορά στάσιμες ακολουθίες τυχαίων μεταβλητών καθώς επίσης παρουσιάζονται και εναλλακτικές προσεγγίσεις μέσω της θεωρίας σημειακών διαδικασιών. Τέλος, στο τρίτο κεφάλαιο που αποτελεί το στατιστικό κομμάτι της θεωρίας, επικεντρωνόμαστε στην εκτίμηση μίας σημαντικής παραμέτρου που προκύπτει από την θεωρία του δευτέρου κεφαλαίου και ονομάζεται δείκτης ακρότατης εξάρτησης (*extremal index*).

РАНЕЕЗНАМО ПЕРПАА

ΚΕΦΑΛΑΙΟ 1

Θεωρία Ακραίων Τιμών για ακολουθίες ανεξάρτητων και ισόνομων τυχαίων μεταβλητών

1.1 Ασυμπτωτικές κατανομές μεγίστου

Σε αυτή την παράγραφο θα αναπτύξουμε το μοντέλο που αποτελεί τον ακρογωνιαίο λίθο της θεωρίας ακραίων τιμών. Η μελέτη μας θα επικεντρωθεί στην πιθανοθεωρητική συμπεριφορά του $\max(X_1, \dots, X_n)$ όπου οι X_1, \dots, X_n είναι ανεξάρτητες και ισόνομες (i.i.d) τυχαίες μεταβλητές με συνάρτηση κατανομής F . Στο εξής θα χρησιμοποιούμε τον συμβολισμό

$$M_n = \max(X_1, \dots, X_n).$$

Για την κατανομή του M_n προκύπτει άμεσα ότι

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = P(X_1 \leq x) \dots P(X_n \leq x) = (F(x))^n.$$

Ωστόσο, σε πρακτικό επίπεδο η γνώση της κατανομής F είναι σπάνια δεδομένη. Οποιαδήποτε προσπάθεια προσέγγισης της με στατιστικές μεθόδους μπορεί να οδηγήσει σε σφάλματα, τα οποία, σε συνδυασμό με τον εκθέτη n μπορούν να οδηγήσουν σε αρκετά λανθασμένες εκτιμήσεις. Μία διαφορετική αντιμετώπιση είναι να δεχτούμε ότι η F είναι άγνωστη και να αναζητήσουμε οριακές οικογένειες κατανομών για το δειγματικό μέγιστο, όπως ανάλογα εργαζόμαστε κατά την προσέγγιση δειγματικών μέσων από την κανονική κατανομή (Κεντρικό Οριακό Θεώρημα). Ορίζουμε

$$z_F = \sup\{x \in R : F(x) < 1\}$$

το δεξιό άκρο των τιμών που μπορούν να πάρουν τα X_i με θετική πιθανότητα («*δεξί άκρο στηρίγματος*»). Το z_F μπορεί να είναι πεπερασμένο, αλλά μπορεί να ισούται και με άπειρο.

Παρατηρούμε ότι

$$P(M_n \leq x) = F^n(x) \rightarrow_{n \rightarrow \infty} \begin{cases} 0, & x < z_F \\ 1, & x \geq z_F \end{cases} \quad (x \in R)$$

Επειδή η ακολουθία $M_n, n = 1, 2, \dots$ είναι μη φθίνουσα θα έχουμε ότι

$$M_n \xrightarrow{n \rightarrow \infty} z_F \quad \text{σχεδόν βέβαια (σ.β)}$$

Η πληροφορία αυτή από μόνη της δεν μπορεί να μας βοηθήσει στον στόχο μας που είναι η εύρεση μιας οριακής κατανομής που να σχετίζεται με το M_n . Για τον λόγο αυτό, όπως και στην περίπτωση του Κεντρικού Οριακού Θεωρήματος, θα αναζητήσουμε οριακές κατανομές για την ποσότητα της μορφής $(M_n - b_n)/a_n$ (**κανονικοποιημένο μέγιστο**) όπου a_n, b_n ακολουθίες με $a_n > 0, b_n \in R, n \in N$. Σκοπός μας λοιπόν είναι να βρούμε αποτελέσματα της μορφής (με \rightarrow_d θα συμβολίζουμε την ασθενή σύγκλιση)

$$\frac{M_n - b_n}{a_n} \rightarrow_d G$$

όπου G μη εκφυλισμένη συνάρτηση κατανομής. Η κατάλληλη επιλογή των ακολουθιών a_n, b_n μπορεί να σταθεροποιήσει τη θέση και την διασπορά του $(M_n - b_n)/a_n$ παρακάμπτοντας έτσι τις δυσκολίες που προκύπτουν από την αρχική ακολουθία $M_n, n = 1, 2, \dots$.

Η απάντηση στο ερώτημα που μας απασχολεί δόθηκε από τους *Fisher&Tippett* (1928). Σε αυτή την εργασία αποδεικνύεται το παρακάτω θεώρημα.

Θεώρημα 1.1.1 (Fisher-Tippett). Έστω X_1, X_2, \dots ακολουθία ανεξάρτητων και ισόνομων (i.i.d) τυχαίων μεταβλητών. Αν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in R, n \in N$ τέτοιες ώστε

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \xrightarrow{n \rightarrow \infty} G(x)$$

για κάθε x , όπου G μη εκφυλισμένη συνάρτηση κατανομής (σ.κ), τότε η G ανήκει σε μία εκ των τριών οικογενειών:

- I. $G(x) = \exp(-e^{-(x-b)/a}), \quad -\infty < x < \infty$
- II. $G(x) = \begin{cases} 0 & , \quad x \leq b \\ \exp(-((x-b)/a)^{-\gamma}) & , \quad x > b \end{cases}$
- III. $G(x) = \begin{cases} \exp(-((b-x)/a)^{\gamma}) & , \quad x < b \\ 1 & , \quad x \geq b \end{cases}$

όπου $a > 0, b \in R$ και $\gamma > 0$.

Παρατήρηση 1.1.2

- Η κατανομή I ονομάζεται **Gumbel** ή **Τύπου I** κατανομή ακραίων τιμών. Η κατανομή II ονομάζεται **Fréchet** ή **Τύπου II** και τέλος, η III ονομάζεται **Weibull** ή **Τύπου III** κατανομή ακραίων τιμών.
- Το b είναι παράμετρος θέσης (**location**), το a παράμετρος κλίμακας (**scale**) και το γ παράμετρος σχήματος (**shape**).
- Οι ακολουθίες κανονικοποίησης δεν είναι μοναδικές. Μπορούμε να βρούμε εναλλακτικές ακολουθίες για τις οποίες το κανονικοποιημένο μέγιστο θα συγκλίνει σε μία κατανομή G' όπου G' είναι ιδίου τύπου με την G , δηλαδή $G' =_d cG + d$ για κάποιες σταθερές a, b . Για παράδειγμα, αν θεωρήσουμε $\tilde{a}_n = a_n / c$, $\tilde{b}_n = b_n - da_n / c$, $d \in R$, $c \neq 0$, εύκολα επαληθεύουμε ότι το $(M_n - \tilde{b}_n) / \tilde{a}_n$ συγκλίνει στην G' .

Το ερώτημα που τίθεται είναι εάν για οποιαδήποτε κατανομή F των αρχικών τυχαίων μεταβλητών X_1, X_2, \dots , το κανονικοποιημένο μέγιστο συγκλίνει σε κάποια από τις τρεις οριακές κατανομές του θεωρήματος *Fisher-Tippett*. Γενικά, η απάντηση είναι αρνητική. Εάν ορίσουμε $u_n(x) = b_n + a_n x$ έχουμε ότι

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P(M_n \leq b_n + a_n x) = P(M_n \leq u_n(x)) = (1 - \bar{F}(u_n(x)))^n = \left(1 - \frac{n\bar{F}(u_n(x))}{n}\right)^n$$

και συνεπώς, κάνοντας τη χρήση της σχέσης

$$\lim_{n \rightarrow \infty} \left(1 + \frac{z_n}{n}\right)^n = e^{\lim_n z_n}$$

αποδεικνύεται το παρακάτω αποτέλεσμα το οποίο αποτελεί αναγκαία και ικανή συνθήκη σύγκλισης του κανονικοποιημένου μεγίστου σε κάποια εκ των τριών κατανομών του Θεωρήματος 1.1.1.

Πρόταση 1.1.3 Για $x \in R$ ισχύει η ισοδυναμία

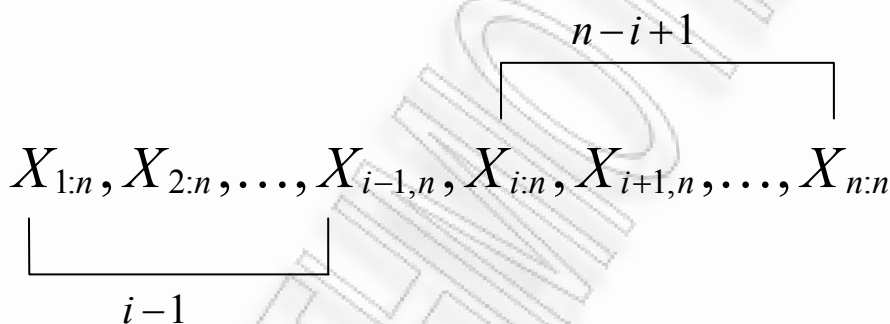
$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} e^{-g(x)} \Leftrightarrow n\bar{F}(b_n + a_n x) \rightarrow_{n \rightarrow \infty} g(x),$$

όπου $g(x) \in [0, \infty]$, $a_n > 0$, $b_n \in R$.

Το παραπάνω αποτέλεσμα βασίζεται στην προσέγγιση της διωνυμικής (*binomial*) κατανομής από την *Poisson*. Έστω λοιπόν, N_n η τ.μ που εκφράζει το πλήθος των X_i που ξεπερνούν το $u_n(x)$. Τότε ισχύει ότι $N_n \sim \text{Binomial}(n, \bar{F}(u_n(x)))$, ενώ για $n \rightarrow \infty$ η N_n συγκλίνει στην κατανομή *Poisson* με παράμετρο $g(x) = \lim_{n \rightarrow \infty} n \bar{F}(u_n(x))$. Εάν τώρα θεωρήσουμε τις διατεταγμένες τ.μ $X_{i,n}$, $i = 1, \dots, n$ όπου

$$X_{1,n} \geq X_{2,n} \geq \dots \geq X_{n,n}$$

παρατηρούμε ότι το ενδεχόμενο $\{η i \text{ μεγαλύτερη παρατήρηση είναι μικρότερη ή ίση της } u_n(x)\}$ είναι ισοπίθανο με το ενδεχόμενο $\{\text{το πλήθος των παρατηρήσεων που είναι μεγαλύτερες της } u_n(x) \text{ είναι το πολύ } i-1\}$, σχηματικά,



Συνεπώς, έχουμε ότι

$$P(X_{i:n} \leq u_n(x)) = P(N_n \leq i-1) \rightarrow_{n \rightarrow \infty} e^{-g(x)} \sum_{j=0}^{i-1} \frac{g(x)^j}{j!}$$

και ειδικότερα, για $i = 1$,

$$P(M_n \leq u_n(x)) = P(X_{1:n} \leq u_n(x)) \rightarrow_{n \rightarrow \infty} e^{-g(x)} \sum_{j=0}^0 \frac{g(x)^j}{j!} = e^{-g(x)}.$$

Παράδειγμα 1.1.4 Έστω τ.μ X_1, X_2, \dots, X_n (i.i.d) που ακολουθούν την εκθετική κατανομή με παράμετρο λ . Επίσης, έστω ακολουθίες $a_n > 0$ και $b_n \in R$ με $a_n = 1/\lambda$, $b_n = \ln n / \lambda$.

Έχουμε ότι

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P\left(\frac{M_n - \ln n / \lambda}{1/\lambda} \leq x\right) = P\left(M_n \leq \frac{\ln n + x}{\lambda}\right) = \left(1 - \bar{F}\left(\frac{\ln n + x}{\lambda}\right)\right)^n$$

$$= \left(1 - e^{-\lambda \frac{\ln n + x}{\lambda}} \right)^n = \left(1 - \frac{e^{-x}}{n} \right)^n \xrightarrow{n \rightarrow \infty} e^{-e^{-x}}, x \in R.$$

Επίσης,

$$n\bar{F}(b_n + a_n x) = ne^{-\lambda \frac{\ln n + x}{\lambda}} = n \frac{e^{-x}}{n} = e^{-x}.$$

Το κανονικοποιημένο μέγιστο λοιπόν, συγκλίνει στην κατανομή *Gumbel* με $b = 0, a = 1$.

Όπως αναφέραμε και παραπάνω, υπάρχουν κατανομές για τις οποίες το $(M_n - b_n)/a_n$ δε συγκλίνει σε κάποια από τις τρεις κατανομές ακραίων τιμών. Το πρόβλημα εμφανίζεται σε κατανομές οι οποίες παρουσιάζουν άλματα στο δεξί άκρο. Γενικά, ισχύουν τα ακόλουθα αποτελέσματα.

Πρόταση 1.1.5. Έστω $z_F < \infty$ και $\bar{F}(z_F^-) = F(z_F) - F(z_F^-) > 0$. Τότε για οποιαδήποτε ακολουθία $u_n, n \in N$ έχουμε ότι

$$P(M_n \leq u_n(x)) \rightarrow_{n \rightarrow \infty} \rho(x)$$

όπου $\rho(x) = 0$ ή $\rho(x) = 1, x \in R$.

Από την παραπάνω πρόταση λοιπόν, συμπεραίνουμε ότι αν η F παρουσιάζει άλμα στο z_F τότε δεν υπάρχει μη εκφυλισμένη κατανομή G τέτοια ώστε

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} G(x).$$

Παρόμοιο αποτέλεσμα ισχύει και για κατανομές με $z_F = \infty$.

Πρόταση 1.1.6 Έστω F σ.κ. με $z_F \leq \infty$ και $g(x) \in (0, \infty)$. Τότε υπάρχει ακολουθία $u_n, n \in N$ με $n\bar{F}(u_n(x)) \rightarrow_{n \rightarrow \infty} g(x)$ αν και μόνο αν

$$\lim_{x \uparrow z_F} \frac{\bar{F}(x)}{F(x^-)} = 1$$

Συνεπώς, για να ισχύει $(M_n - b_n)/a_n \rightarrow_d G$ θα πρέπει τα άλματα της F κοντά στο z_F να «φθίνουν γρήγορα».

Παράδειγμα 1.1.7 (Poisson) Έστω $X \sim Pois(\lambda)$, $\lambda > 0$. Η συνάρτηση πιθανότητας της κατανομής *Poisson* γνωρίζουμε ότι είναι

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

Επίσης, έχουμε ότι

$$(1.1) \quad \frac{\bar{F}(k)}{\bar{F}(k-1)} = \frac{1 - F(k)}{\bar{F}(k-1)} = \frac{\bar{F}(k-1) - F(k) + F(k-1)}{\bar{F}(k-1)} = 1 - \frac{F(k) - F(k-1)}{\bar{F}(k-1)} = 1 - \frac{e^{-\lambda} \lambda^k / k!}{\sum_{j=k}^{\infty} e^{-\lambda} \lambda^j / j!}$$

$$= 1 - \frac{\lambda^k / k!}{\sum_{j=k}^{\infty} \lambda^j / j!} = 1 - \frac{1}{1 + \frac{\lambda}{(k+1)} + \frac{\lambda^2}{(k+1)(k+2)} + \dots} = 1 - \frac{1}{1 + \sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1)\dots(k+s)}}.$$

Όμως ισχύει

$$(1.2) \quad \sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1)\dots(k+s)} \leq \sum_{s=1}^{\infty} \left(\frac{\lambda}{k}\right)^s = \frac{\frac{\lambda}{k}}{1 - \frac{\lambda}{k}}, \quad k > \lambda$$

Συνεπώς για $k \rightarrow \infty$ έχουμε από την (1.1) και (1.2) ότι

$$\frac{\bar{F}(k)}{\bar{F}(k-1)} \rightarrow 0.$$

Από την Πρόταση 1.1.6 επομένως δεν υπάρχει G τέτοια ώστε $P(M_n \leq u_n) \rightarrow_{n \rightarrow \infty} G$.

Παράδειγμα 1.1.8 (Γεωμετρική κατανομή) Έστω $X \sim Geom(p)$. Έχουμε ότι

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots, \quad 0 < p < 1.$$

Επίσης

$$\frac{\bar{F}(k)}{\bar{F}(k-1)} = 1 - (1-p)^{k-1} \left(\sum_{j=k}^{\infty} (1-p)^{j-1} \right)^{-1} = 1 - p \in (0, 1)$$

Συνεπώς, από την Πρόταση 1.1.6 δεν υπάρχει G τέτοια ώστε $P(M_n \leq u_n) \rightarrow_{n \rightarrow \infty} G$.

1.2 Max stable κατανομές – GEV κατανομή

Ορισμός 1.2.1 Μία κατανομή G καλείται *max-stable* αν για κάθε $n = 2, 3, \dots$ υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in R$ τέτοιες ώστε

$$G^n(a_n x + b_n) = G(x).$$

Παράδειγμα 1.2.2 Έστω οι i.i.d τ.μ X, X_1, X_2, \dots με σ.κ G όπου G *max-stable* κατανομή. Τότε έχουμε ότι

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P(M_n \leq a_n x + b_n) = G^n(a_n x + b_n) = G(x).$$

Συνεπώς ,

$$\frac{M_n - b_n}{a_n} \underset{d}{=} X.$$

Παρατήρηση 1.2.3. Αποδεικνύεται ότι οι μοναδικές κατανομές που έχουν την *max-stable* ιδιότητα είναι οι τρεις οριακές κατανομές του θεωρήματος *Fisher-Tippett*. Συνεπώς η κλάση των *max-stable* κατανομών συμπίπτει με την κλάση των μη εκφυλισμένων κατανομών που μπορούν να προκύψουν ως όριο του κανονικοποιημένου μεγίστου. Σύμφωνα με το Θεώρημα *Fisher-Tippett* αυτή η κλάση περιέχει τρεις οικογένειες κατανομών. Για λόγους όμως ενιαίας αντιμετώπισης και εκτίμησης των παραμέτρων είναι προτιμότερο να ενοποιήσουμε αυτές τις τρεις οικογένειες σε μία, την οποία καλούμε *GEV* όπως φαίνεται στο ακόλουθο θεώρημα.

Ορισμός 1.2.4 Η κατανομή **GEV** (*Generalized Extreme Value Distribution*) έχει συνάρτηση κατανομής,

$$G(x) = \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right\}$$

και ορίζεται στο σύνολο $\{x : 1 + \xi \frac{x - \mu}{\sigma} > 0\}$, με παραμέτρους $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$.

Πρόταση 1.2.5. Ισχύει ότι

- Αν $\xi > 0$, $\xi = \frac{1}{\gamma}$, $\sigma = \frac{a}{\gamma}$, $\mu = a + b \Rightarrow$ η *GEV* συμπίπτει με τη κατανομή *Fréchet*
- Αν $\xi < 0$, $\xi = -\frac{1}{\gamma}$, $\sigma = \frac{a}{\gamma}$, $\mu = -a + b \Rightarrow$ η *GEV* συμπίπτει με τη κατανομή *Weibull*
- Αν $\xi \rightarrow 0$, $\sigma = a$, $\mu = b \Rightarrow$ η *GEV* συμπίπτει με τη κατανομή *Gumbel*

Έχοντας ορίσει την κατανομή *GEV* μπορούμε να αναδιατυπώσουμε το Θεώρημα 1.1.1.

Θεώρημα 1.2.6 Αν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in R$ τέτοιες ώστε

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} G(x),$$

όπου G μη εκφυλισμένη κατανομή, τότε η G ανήκει στην οικογένεια GEV με

$$G(x) = \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right\},$$

για x στο σύνολο $\left\{x : 1 + \xi \frac{x - \mu}{\sigma} > 0\right\}$, $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$.

Η κατανομή GEV ως ευρύτερη κατανομή των τριών τύπων του θεωρήματος *Fisher-Tippett* αποτελεί χρήσιμο εργαλείο στην στατιστική ανάλυση των ακραίων παρατηρήσεων, καθώς η παραδοχή ότι το κανονικοποιημένο μέγιστο των δεδομένων ακολουθεί την GEV και η εκτίμηση των παραμέτρων ξ, μ, σ από τα ίδια τα δεδομένα μπορεί να οδηγήσει σε ασφαλέστερα συμπεράσματα από το να υποθέσουμε εκ των προτέρων κάποιον από τους τρεις τύπους κατανομών.

1.3 Ασυμπτωτικές κατανομές ελαχίστου

Σε αρκετές περιπτώσεις μας ενδιαφέρει να γνωρίζουμε την πιθανότητα εμφάνισης ακραίων (π.χ. πολύ μικρών) τιμών κάποιας τυχαίας μεταβλητής X . Για παράδειγμα, αν κατέχουμε μία μετοχή, θέλουμε να ξέρουμε την πιθανότητα η αξία της να πέσει σε αρκετά χαμηλά επίπεδα, ή σε περίπτωση έντονης ξηρασίας μας ενδιαφέρει να γνωρίζουμε την πιθανότητα καταγραφής χαμηλής στάθμης στη λίμνη η οποία υδροδοτεί κάποια μεγάλη πόλη. Είναι σκόπιμο λοιπόν να ορίσουμε την τυχαία μεταβλητή

$$\tilde{M}_n = \min(X_1, \dots, X_n)$$

και να αναζητήσουμε οριακές κατανομές, όπως και στην περίπτωση της M_n , στις οποίες το κανονικοποιημένο ελάχιστο θα συγκλίνει ασθενώς. Από την μελέτη που έχει προηγηθεί για την M_n μπορούν εύκολα να προκύψουν συμπεράσματα για την \tilde{M}_n διότι,

$$\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n).$$

Θέτουμε λοιπόν $Y_i = -X_i, i = 1, \dots, n$ και ορίζουμε

$$L_n = \max(Y_1, \dots, Y_n).$$

Κάνοντας χρήση του Θεωρήματος 1.2.6 έχουμε ότι

$$P(\tilde{M}_n \leq x) = P(-L_n \leq x) = P(L_n \geq -x) = 1 - P(L_n < -x).$$

Για αρκετά μεγάλο n θα έχουμε λοιπόν ότι

$$P(\tilde{M}_n \leq x) \approx 1 - \exp\left\{-\left(1 + \xi \frac{-x - \mu}{\sigma}\right)^{-1/\xi}\right\} = 1 - \exp\left\{-\left(1 - \xi \frac{x + \mu}{\sigma}\right)^{-1/\xi}\right\}.$$

Ισχύει λοιπόν το παρακάτω θεώρημα το οποίο είναι ανάλογο του Θεωρήματος 1.2.6.

Θεώρημα 1.3.1 Αν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in \mathbb{R}$ τέτοιες ώστε

$$P\left(\frac{\tilde{M}_n - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} \tilde{G}(x)$$

όπου \tilde{G} μη εκφυλισμένη κατανομή, τότε η \tilde{G} ανήκει στην οικογένεια GEV (ελαχίστου) με

$$\tilde{G}(x) = \exp\left\{-\left(1 - \xi \frac{x + \mu}{\sigma}\right)^{-1/\xi}\right\}$$

για x στο σύνολο $\left\{x : 1 - \xi \frac{x + \mu}{\sigma} > 0\right\}$, $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$.

1.4 Περιοχή έλξης κατανομής μεγίστου (Maximum domain of attraction)

Ορισμός 1.4.1 Έστω τ.μ. X με σ.κ. F . Λέμε ότι η X ανήκει στην **περιοχή έλξης** μίας κατανομής ακροτάτων G αν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in \mathbb{R}$, $n \in \mathbb{N}$ για τις οποίες ισχύει

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x)$$

και συμβολίζουμε $X \in MDA(G)$ ή $F \in MDA(G)$.

Παρατήρηση 1.4.2. Ανάλογα με τον τύπο της G , όπως ορίζει το Θεώρημα *Fisher-Tippett*, έχουμε τρεις περιοχές έλξης ακροτάτων. Κάθε μία από αυτές περιέχει όλες τις σ.κ F για τις οποίες ισχύει ότι το κανονικοποιημένο μέγιστο δείγματος από αυτές συγκλίνει στην αντίστοιχη κατανομή ακροτάτων.

παράδειγμα 1.4.3. Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες (i.i.d) τ.μ. που ακολουθούν την κατανομή *Cauchy* με σ.π.

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Θεωρούμε τις ακολουθίες $a_n = n/\pi$ και $b_n = 0$. Επιπλέον έχουμε ότι

$$\lim_{z \rightarrow \infty} \frac{\bar{F}(z)}{(\pi z)^{-1}} = \lim_{z \rightarrow \infty} \frac{f(z)}{\pi^{-1}z^{-2}} = \lim_{z \rightarrow \infty} \frac{\pi z^2}{\pi(1+z^2)} = 1,$$

και επομένως $\bar{F}(z) \sim (\pi z)^{-1}$. Συνεπώς,

$$\begin{aligned} P\left(\frac{\pi M_n}{n} \leq x\right) &= \left(1 - \bar{F}(nx/\pi)\right)^n = \left(1 - \frac{n\bar{F}(nx/\pi)}{n\left(\pi \frac{nx}{\pi}\right)^{-1}}\right)^n \rightarrow_{n \rightarrow \infty} \exp\left(-\lim_{n \rightarrow \infty} \frac{n\bar{F}(nx/\pi)}{\left(\pi \frac{nx}{\pi}\right)^{-1}}\right) \\ &= \exp\left(-\frac{1}{x} \lim_{n \rightarrow \infty} \frac{\bar{F}(nx/\pi)}{\left(\pi \frac{nx}{\pi}\right)^{-1}}\right) = e^{-\frac{1}{x}}, \quad x \in \mathbb{R}. \end{aligned}$$

Η κατανομή *Cauchy* λοιπόν ανήκει στην περιοχή έλξης της *Fréchet*.

Παράδειγμα 1.4.4 Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες (i.i.d) τ.μ. που ακολουθούν την ομοιόμορφη κατανομή στο $[0,1]$. Θεωρούμε τις ακολουθίες $a_n = 1/n$ και $b_n = 1$. Επιπλέον, έχουμε ότι

$$P(nM_n - n \leq x) = P(M_n \leq 1 + x/n) = (F(1 + x/n))^n = (1 + x/n)^n \rightarrow_{n \rightarrow \infty} e^x, \quad x < 0.$$

Συνεπώς, η ομοιόμορφη κατανομή στο $[0,1]$ ανήκει στην περιοχή έλξης της *Weibull*.

Παράδειγμα 1.4.5 Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες (i.i.d) τ.μ που ακολουθούν την εκθετική κατανομή με σ.π.

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

Θεωρούμε τις ακολουθίες $a_n = 1/\lambda$ και $b_n = \ln n/\lambda$. Στο Παράδειγμα 1.1.4 είδαμε ότι

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} e^{-e^{-x}}, \quad x \in \mathbb{R}$$

Συνεπώς, η εκθετική κατανομή ανήκει στην περιοχή έλξης της *Gumbel*.

Στο σημείο αυτό γεννάται το ερώτημα εάν υπάρχει κάποιος «εύκολος» τρόπος για να ελέγξουμε αν κάποια σ.κ F ανήκει στην *MDA* κάποιας κατανομής ακροτάτων. Θα διατυπώσουμε λοιπόν κάποιες ικανές και αναγκαίες συνθήκες οι οποίες βοηθούν προς αυτή την κατεύθυνση. Αρχικά θα χρειαστούμε τους παρακάτω ορισμούς.

Ορισμός 1.4.6 Μία θετική, *Lebesgue* μετρήσιμη συνάρτηση h στο $(0, \infty)$, καλείται **ομαλής κύμανσης** (*regularly varying*) στο ∞ με δείκτη $a \in R$ (συμβ. $h \in \mathfrak{R}_a$) αν

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^a, \quad t > 0$$

Ορισμός 1.4.7 Μία θετική, *Lebesgue* μετρήσιμη συνάρτηση L στο $(0, \infty)$, καλείται **βραδείας κύμανσης** (*slowly varying*) στο ∞ (συμβ. $h \in \mathfrak{R}_0$) αν

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0$$

Ορισμός 1.4.8 Μία θετική, *Lebesgue* μετρήσιμη συνάρτηση h στο $(0, \infty)$, καλείται **ταχείας κύμανσης** (*rapidly varying*) στο ∞ με δείκτη $-\infty$ (συμβ. $h \in \mathfrak{R}_{-\infty}$) αν

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = \begin{cases} 0 & , t > 1 \\ \infty & , 0 < t < 1 \end{cases}$$

Παρατήρηση 1.4.9. Ουσιαστικά, με τους τρεις παραπάνω ορισμούς μπορούμε να εξετάσουμε αν κάποια συνάρτηση μεταβάλλεται μέτρια, αργά ή γρήγορα καθώς το x πλησιάζει το άπειρο. Χαρακτηριστικό παράδειγμα συνάρτησης βραδείας κύμανσης είναι η $\ln x$ ή η σταθερή συνάρτηση, ομαλής κύμανσης οι $x^a, x^a \ln(1+x), a \in R$ και ταχείας κύμανσης η e^{-x} . Επίσης, εύκολα διαπιστώνουμε ότι κάθε συνάρτηση ομαλής κύμανσης στο ∞ με δείκτη $a \in R$ μπορεί να γραφεί στην μορφή $x^a L(x)$ όπου L είναι μία κατάλληλη συνάρτηση βραδείας κύμανσης, καθώς και πάλι θα ισχύει

$$\frac{h(tx)}{h(x)} = \frac{(tx)^a L(tx)}{x^a L(x)} = \frac{t^a L(tx)}{L(x)} \xrightarrow{x \rightarrow \infty} t^a, \quad a \in R$$

Θεώρημα 1.4.10. (αναπαράσταση συναρτήσεων ομαλής κύμανσης). Αν $h \in \mathfrak{R}_a, a \in R \cup \{-\infty\}$, τότε η h μπορεί να αναπαρασταθεί ως εξής

$$h(x) = c(x) \exp \left\{ \int_z^x \frac{\delta(u)}{u} du \right\}, x \geq z$$

όπου $z > 0$ και c, δ μετρήσιμες συναρτήσεις με $c(x) \rightarrow c_0 \in (0, \infty)$ και $\delta(x) \rightarrow a$ για $x \rightarrow \infty$.

Ισχύει και το αντίστροφο.

Παρατήρηση 1.4.11. Παρατηρούμε ότι για $a > 0$ ισχύει $\lim_{x \rightarrow \infty} h(x) = \infty$, ενώ για $a < 0$ έχουμε

$$\lim_{x \rightarrow \infty} h(x) = 0.$$

Ορισμός 1.4.12 Δύο σ.κ. F και G έχουν **ισοδύναμη ουρά**, εάν $z_F = z_G$ και

$$\lim_{x \rightarrow z_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c$$

όπου $0 < c < \infty$.

Πρόταση 1.4.13 Έστω F, G σ.κ. ισοδύναμης ουράς. Τότε η F ανήκει στην MDA κάποιας κατανομής ακροτάτων αν και μόνο αν η G ανήκει στην ίδια MDA .

Ορισμός 1.4.14 Έστω σ.κ. F . Ορίζουμε την **γενικευμένη αντίστροφη** της F ως εξής

$$F^{\leftarrow}(t) = \inf \{x \in R : F(x) \geq t\}, \quad 0 < t < 1.$$

Σε αυτό το σημείο είμαστε πλέον σε θέση να διατυπώσουμε τις συνθήκες που εξασφαλίζουν ότι μία κατανομή F ανήκει στην MDA μίας εκ των τριών κατανομών ακροτάτων.

Θεώρημα 1.4.15 (*MDA κατανομής Fréchet*). Μία σ.κ. F ανήκει στην MDA της κατανομής *Fréchet* με σ.κ.

$$G(x) = \begin{cases} 0 & , x \leq 0 \\ \exp(-x^{-\gamma}) & , x > 0 \end{cases} \quad (\gamma > 0)$$

($\beta = 0, a = 1$), αν και μόνο αν $\bar{F}(x) = x^{-\gamma} L(x)$ και $z_F = \infty$, όπου L συνάρτηση βραδείας κύμανσης. Σε αυτή την περίπτωση μπορούμε να επιλέξουμε $b_n = 0$ και $a_n = F^{\leftarrow}(1 - n^{-1})$.

Απόδειξη Έστω $\bar{F}(x) = x^{-\gamma} L(x)$ ή ισοδύναμα $\bar{F} \in \mathfrak{R}_{-\gamma}$, $\gamma > 0$. Αρχικά θα δείξουμε ότι για $n \rightarrow \infty$ ισχύει ότι $a_n \rightarrow \infty$. Έχουμε

$$\lim_{n \rightarrow \infty} \frac{\bar{F}(a_n)}{n^{-1}} = \lim_{n \rightarrow \infty} \frac{n^{-1}}{n^{-1}} = 1 \Rightarrow \bar{F}(a_n) \sim n^{-1}.$$

Συνεπώς, για $n \rightarrow \infty$ έχουμε ότι $\bar{F}(a_n) \rightarrow 0$. Σε συνδυασμό με το γεγονός ότι $z_F = \infty$, έχουμε ότι $a_n \rightarrow \infty$ όταν $n \rightarrow \infty$.

Για $x > 0$ παίρνουμε

$$n\bar{F}(a_n x + b_n) = n\bar{F}(a_n x) = n \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} \bar{F}(a_n) = n \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} n^{-1} \rightarrow_{n \rightarrow \infty} x^{-\gamma}.$$

Για $x < 0$ και $n \rightarrow \infty$ έχουμε

$$F^n(a_n x) \leq F^n(0) \rightarrow 0.$$

Συνεπώς, από την Πρόταση 1.1.3 η F ανήκει στην MDA της $Fréchet$ με $\beta = 0, a = 1$. Η απόδειξη του αντιστρόφου είναι αρκετά τεχνική και παραλείπεται (Για περισσότερες λεπτομέρειες βλ. *Resnick* (1987)).

■

Θεώρημα 1.4.16 (*MDA κατανομής Weibull*). Μία σ.κ. F ανήκει στην MDA της κατανομής *Weibull* με σ.κ.

$$G(x) = \begin{cases} \exp(-(-x)^\gamma) & , x < 0 \\ 0 & , x \geq 0 \end{cases} \quad (\gamma > 0)$$

($\beta = 0, a = 1$), αν και μόνο αν $\bar{F}(z_F - x^{-1}) = x^{-\gamma} L(x)$ και $z_F < \infty$, όπου L συνάρτηση βραδείας κύμανσης. Σε αυτή την περίπτωση μπορούμε να επιλέξουμε $a_n = z_F - F^{\leftarrow}(1 - n^{-1})$ και $b_n = z_F$.

Απόδειξη. Έστω $z_F < \infty$ και $\bar{F}(z_F - x^{-1}) = x^{-\gamma} L(x)$ δηλαδή $\bar{F}(z_F - x^{-1}) \in \mathfrak{R}_{-\gamma}$. Ορίζουμε τη συνάρτηση $F^*(x) = F(z_F - x^{-1})$, $x > 0$. Συνεπώς από το Θεώρημα 1.4.15 έχουμε ότι η F^* ανήκει στην MDA της $Fréchet$ με $\beta = 0, a = 1$. Σε αυτή την περίπτωση μπορούμε να επιλέξουμε

$$a_n^* = F_*^{\leftarrow}(1 - n^{-1}) \quad \text{και} \quad b_n^* = 0.$$

Έχουμε λοιπόν, για $x > 0$

$$F_*^{\leftarrow}(a_n^* x) \rightarrow_{n \rightarrow \infty} \exp(-x^{-\gamma}),$$

δηλαδή

$$F^n(z_F - (a_n^* x)^{-1}) \rightarrow_{n \rightarrow \infty} \exp(-x^{-\gamma}).$$

Θέτουμε $x = -y^{-1}$ και παίρνουμε

$$F^n(z_F + \frac{y}{a_n^*}) \rightarrow_{n \rightarrow \infty} \exp\{-(-y)^\gamma\}, \quad y < 0.$$

Τέλος,

$$\begin{aligned} a_n^* &= F_*^{\leftarrow}(1 - n^{-1}) = \inf\{x \in R : F(z_F - x^{-1}) \geq 1 - n^{-1}\} = \inf\left\{\frac{1}{z_F - u} : F(u) \geq 1 - n^{-1}\right\} \\ &= \frac{1}{z_F - \inf\{u : F(u) \geq 1 - n^{-1}\}} = \frac{1}{z_F - F^{\leftarrow}(1 - n^{-1})} = \frac{1}{z_F - \gamma_n}, \end{aligned}$$

όπου $\gamma_n = F^{\leftarrow}(1 - n^{-1})$.

Συνεπώς,

$$F^n(z_F + (z_F - \gamma_n)y) \rightarrow_{n \rightarrow \infty} \exp\{-(-y)^\gamma\}, \quad y < 0, \text{ δηλαδή,}$$

$$a_n = z_F - \gamma_n = z_F - F^{\leftarrow}(1 - n^{-1}) \quad \text{και} \quad b_n = z_F$$

Για το αντίστροφο παραπέμπουμε στον *Resnick* (1987). ■

Θεώρημα 1.4.17 (*MDA κατανομής Gumbel*). Μία σ.κ. F ανήκει στο *MDA* της κατανομής *Gumbel* με σ.κ. $G(x) = \exp\{-\exp(-x)\}$, $-\infty < x < \infty$, $(\beta = 0, \alpha = 1)$ αν και μόνο αν η F μπορεί να αναπαρασταθεί ως εξής

$$\bar{F}(x) = c(x) \exp\left\{-\int_{\omega}^x \frac{g(t)}{a(t)} dt\right\}, \quad \omega < x < z_F,$$

όπου c και g μετρήσιμες (*Lebesgue*) συναρτήσεις για τις οποίες ισχύει $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$ για $x \uparrow z_F$ και $a(x)$ είναι μία θετική απόλυτα συνεχής (ως προς το μέτρο *Lebesgue*) συνάρτηση με πυκνότητα $a'(x)$ και $\lim_{x \uparrow z_F} a'(x) = 0$.

Για μία τέτοια F μπορούμε να επιλέξουμε $b_n = F^{\leftarrow}(1 - n^{-1})$ και $a_n = a(d_n)$. Όσον αφορά την συνάρτηση a μπορούμε να επιλέξουμε $a(x) = E(X - x | X > x)$, $x < z_F$ όπου $X \sim F$.

Απόδειξη Παραπέμπουμε στον *Resnick* (1987). ■

Παρατήρηση 1.4.18 Αποδεικνύεται ότι οι περιοχές έλξης της *Gumbel*, *Weibull* και *Fréchet* αποτελούνται από όλες τις κατανομές που ικανοποιούν τις αντίστοιχες συνθήκες των θεωρημάτων 1.4.15-1.4.17 και τις κατανομές ισοδύναμης ουράς με αυτές. Για περισσότερες λεπτομέρειες σχετικά με τον χαρακτηρισμό των περιοχών έλξης, παραπέμπουμε στους *Embrechts et al.* (1997) και *de Haan and Ferreira* (2006).

1.5 Γενικευμένη κατανομή Pareto (Generalized Pareto distribution)

Σε ορισμένες περιπτώσεις μελετούμε παρατηρήσεις που υπερβαίνουν ένα υψηλό κατώφλι u . Είναι αρκετά χρήσιμο το γεγονός ότι οι παρατηρήσεις αυτές, ασυμπτωτικά και κάτω από σχετικά ασθενείς συνθήκες, ακολουθούν μια γενικευμένη κατανομή Pareto.

Ορισμός 1.5.1 Η κατανομή *GPD* (Generalized Pareto distribution) έχει σ.κ

$$H_{\xi, \tilde{\sigma}}(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}, \quad y > 0 \text{ και } 1 + \frac{\xi y}{\tilde{\sigma}} > 0$$

με παραμέτρους $\tilde{\sigma} > 0$, $\xi \in \mathbb{R}$.

Θεώρημα 1.5.2. Έστω οι *i.i.d.* τ.μ X_1, X_2, \dots με σ.κ F . Υποθέτουμε ότι για $n \rightarrow \infty$,

$$P(M_n \leq u_n(x)) \rightarrow_d G(x),$$

όπου

$$G(x) = \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right\}, \quad \mu, \sigma > 0, \xi \in \mathbb{R}. \quad (GEV)$$

Τότε ισχύει

$$P(X - u \leq y | X > u) \rightarrow_{u \rightarrow z_F} 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

για κάθε $y > 0$ και $1 + \xi y / \tilde{\sigma} > 0$, όπου $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

Σχέδιο απόδειξης. Από την υπόθεση του Θεωρήματος 1.1.1 έχουμε ότι

$$F^n(x) \approx \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right\}, \quad \mu, \sigma > 0, \xi \in \mathbb{R}$$

Συνεπώς,

$$(1.3) \quad n \ln F(x) \approx - \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi}$$

Από το ανάπτυγμα *Taylor* για την συνάρτηση $\ln x$ έχουμε ότι (όταν $x \approx 1$),

$$(1.4) \quad \ln F(x) \approx -(1 - F(x))$$

Από την (1.3) και (1.4) έχουμε λοιπόν, για $u \rightarrow z_F$

$$1 - F(u) \approx \frac{1}{n} \left(1 + \xi \frac{u - \mu}{\sigma} \right)^{-1/\xi}$$

Όμοια,

$$1 - F(u + y) \approx \frac{1}{n} \left(1 + \xi \frac{u + y - \mu}{\sigma} \right)^{-1/\xi}, y > 0$$

Έτσι λοιπόν παίρνουμε,

$$\begin{aligned} P(X - u > y | X > u) &= \frac{1 - F(u + y)}{1 - F(u)} \approx \frac{1/n (1 + \xi(u + y - \mu)/\sigma)^{-1/\xi}}{1/n (1 + \xi(u - \mu)/\sigma)^{-1/\xi}} = \\ &= \frac{(1 + \xi(u - \mu)/\sigma + \xi y/\sigma)^{-1/\xi}}{(1 + \xi(u - \mu)/\sigma)^{-1/\xi}} = \left(1 + \frac{\xi y}{\sigma + \xi(u - \mu)} \right)^{-1/\xi} \end{aligned}$$

Θέτοντας $\tilde{\sigma} = \sigma + \xi(u - \mu)$ έχουμε,

$$P(X - u \leq y | X > u) \approx 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}} \right)^{-1/\xi}$$

■

Παρατήρηση 1.5.3. Ισχύουν τα ακόλουθα:

- Αν $\xi < 0$ έχουμε ότι, $1 + \xi y/\tilde{\sigma} > 0 \Rightarrow y < -\tilde{\sigma}/\xi$, δηλαδή η ποσότητα $-\tilde{\sigma}/\xi$ αποτελεί άνω-φράγμα για τις τιμές της μεταβλητής y ($z_F = -\tilde{\sigma}/\xi$).
- Αν $\xi > 0$ η y δεν είναι άνω-φραγμένη
- Αν $\xi = 0$ έχουμε ότι

$$\lim_{\xi \rightarrow 0} \left(1 - \left(1 + \frac{\xi y}{\tilde{\sigma}} \right)^{-1/\xi} \right) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), y > 0$$

δηλαδή οριακά θα ισχύει $X - u | X > u \sim \text{Exp}(-1/\tilde{\sigma})$.

Παράδειγμα 1.5.4 (Fréchet). Έστω $F(x) = \exp(-1/x)$. Τότε για $u \rightarrow \infty, y > 0$ έχουμε

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - \exp\{-(u+y)^{-1}\}}{1 - \exp(-u^{-1})} \approx \frac{1 - (1 - (u+y)^{-1})}{1 - (1 - u^{-1})} = \frac{(u+y)^{-1}}{u^{-1}} = \left(1 + \frac{y}{u}\right)^{-1},$$

(GPD με $\xi=1, \tilde{\sigma}=u$)

Θεώρημα 1.5.5 (Ιδιότητες της GPD) 1. Έστω τ.μ $X \sim GPD$ με παραμέτρους $\xi, \tilde{\sigma}$. Τότε $E(X) = \infty$ αν και μόνο αν $\xi \geq 1$.

2. Εάν $\xi < 1$ ισχύουν τα παρακάτω

- $E\left(\left(1 + \frac{\xi}{\tilde{\sigma}} X\right)^{-\kappa}\right) = \frac{1}{1 + \xi\kappa}, \kappa > -\frac{1}{\xi}.$
- $E\left(\left(\ln\left(1 + \frac{\xi}{\tilde{\sigma}} X\right)\right)^\kappa\right) = \xi^\kappa \kappa!, \kappa \in N.$

3. Εάν $\xi < 1/\kappa, \kappa \in N$ τότε $E(X^\kappa) = \frac{\tilde{\sigma}^\kappa}{\xi^{\kappa+1}} \cdot \frac{\Gamma(\xi^{-1} - \kappa)}{\Gamma(1 + \xi^{-1})} \cdot \kappa!.$

4. Ισχύει ότι $\frac{\bar{H}_{\xi, \tilde{\sigma}}(y_1 + y_2)}{\bar{H}_{\xi, \tilde{\sigma}}(y_1)} = \bar{H}_{\xi, \tilde{\sigma} + \tilde{\sigma}_1}(y_2).$

5 Έστω τ.μ $N \sim Pois(\lambda)$, $X_n, n = 1, 2, \dots$ ακολουθία i.i.d τυχαίων μεταβλητών με σ.κ. την GPD. Επίσης, ορίζουμε $M_N = \max(X_1, \dots, X_N)$. Τότε ισχύει

$$P(M_N \leq y) = \exp\left\{-\lambda \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}\right\}$$

6. Έστω τ.μ $X \sim GPD$ με παραμέτρους $\xi < 1, \tilde{\sigma}$. Τότε για $u < z_F$ ισχύει

$$E(X - u \mid X > u) = \frac{\sigma + \xi(u - \mu)}{1 - \xi}, \sigma + \xi(u - \mu) > 0.$$

Παρατήρηση 1.5.6. (α) Ισχύει

$$\frac{P(X - u > y_1 + y_2 \mid X > u)}{P(X - u > y_1 \mid X > u)} = P(X > u + y_1 + y_2 \mid X > u + y_1)$$

Θέτοντας $z_1 = u + y_1$ και $z_2 = y_2$ έχουμε

$$P(X > u + y_1 + y_2 | X > u + y_1) = P(X > z_1 + z_2 | X > z_1).$$

Συνεπώς, η ιδιότητα (4) μας δίνει την πιθανότητα του ενδεχομένου η τ.μ X να ξεπεράσει το κατώφλι $z_1 + z_2$ δεδομένου ότι έχει υπερβεί το κατώφλι z_1 . Η πιθανότητα σε αυτή την περίπτωση δίνεται από την κατανομή GPD με μόνη διαφορά την παράμετρο $\tilde{\sigma}$ όπου τώρα $\tilde{\sigma}' = \tilde{\sigma} + \xi y_1$.

(β) Από την ιδιότητα (5) πληροφορούμαστε ότι αν έχουμε ένα μοντέλο στο οποίο ο αριθμός των υπερβάσεων κατωφλίου ακολουθεί την κατανομή $Poisson$ και οι υπερβάσεις ακολουθούν την GPD τότε η μέγιστη υπέρβαση ακολουθεί την GEV με παραμέτρους $\mu^* = \xi^{-1} \tilde{\sigma} (\lambda^\xi - 1)$ και $\sigma^* = \tilde{\sigma} \lambda^\xi$.

1.6. Οριακή κατανομή της r μεγαλύτερης παρατήρησης

Μέχρι τώρα έχουμε μελετήσει την συμπεριφορά της μέγιστης παρατήρησης σε μία ακολουθία τυχαίων μεταβλητών. Σε αυτό το σημείο θα παρουσιάσουμε την γενίκευση του μοντέλου αυτού και συγκεκριμένα θα ασχοληθούμε με την κατανομή της r μεγαλύτερης παρατήρησης. Έστω και πάλι ανεξάρτητες και ισόνομες τ.μ X_1, X_2, \dots, X_n με συνάρτηση κατανομής F . Συμβολίζουμε με $M_n^{(r)}$ την r -οστή μεγαλύτερη παρατήρηση από τις $X_i, i = 1, \dots, n$.

Θεώρημα 1.6.1 *Εάν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in R$ τέτοιες ώστε*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\},$$

τότε για σταθερό r ισχύει

$$P\left(\frac{M_n^{(r)} - b_n}{a_n} \leq x\right) \rightarrow_{n \rightarrow \infty} G_r(x)$$

στο σύνολο $\left\{x : 1 + \xi \frac{x - \mu}{\sigma} > 0\right\}$, όπου

$$G_r(x) = \exp(-g(x)) \sum_{s=0}^{r-1} \frac{g(x)^s}{s!} \quad \text{και} \quad g(x) = \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}.$$

Για $r = 1$ η G_1 συμπίπτει με την κατανομή GEV .

Θεώρημα 1.6.2. Έστω

$$\mathbf{M}_n = \left(\frac{M_n^{(1)} - b_n}{a_n}, \dots, \frac{M_n^{(r)} - b_n}{a_n} \right)$$

και $\mathbf{x} = (x_1, \dots, x_r)$. Η οριακή από κοινού συνάρτηση κατανομής των r μεγαλύτερων διατεταγμένων παρατηρήσεων δίνεται από την σχέση

$$P(\mathbf{M}_n \leq \mathbf{x}) \rightarrow_{n \rightarrow \infty} \sum_{s_1=0}^1 \sum_{s_2=0}^{2-s_1} \dots \sum_{s_{r-1}=0}^{r-1-s_1-\dots-s_{r-2}} \frac{(\gamma_2 - \gamma_1)^{s_1}}{s_1!} \dots \frac{(\gamma_r - \gamma_{r-1})^{s_{r-1}}}{s_{r-1}!} \exp(-\gamma_r),$$

όπου $\gamma_i = \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{-1/\xi}$, $i = 1, \dots, n$ και $x_1 \geq x_2 \geq \dots \geq x_r$. Αντίστοιχα, για την οριακή από κοινού συνάρτηση πυκνότητας έχουμε

$$f(x_1, x_2, \dots, x_r) = \exp \left\{ - \left(1 + \xi \frac{x_r - \mu}{\sigma}\right)^{-1/\xi} \right\} \prod_{k=1}^r \sigma^{-1} \left(1 + \xi \frac{x_k - \mu}{\sigma}\right)^{\frac{1}{\xi} - 1},$$

όπου μ, σ οι παράμετροι της *GEV* κατανομής και $x_1 \geq x_2 \geq \dots \geq x_r$, $x_k : 1 + \xi \frac{x_k - \mu}{\sigma} > 0$ για $k = 1, 2, \dots, r$. Για $r = 1$ η σ.κ και η σ.π. συμπίπτουν με τις αντίστοιχες της *GEV*.

Παρατήρηση 1.6.3. Οι προσεγγίσεις του θεωρήματος μπορούν να θεωρηθούν έγκυρες μόνο στην περίπτωση όπου το r είναι αρκετά μικρό σε σχέση με το n . Όσο το r αυξάνει σε σχέση με το n , η ταχύτητα σύγκλισης φθίνει με γρήγορο ρυθμό και οι προσεγγίσεις δεν μπορούν να θεωρηθούν αξιόπιστες.

1.7 Προσέγγιση μέσω Σημειακών στοχαστικών διαδικασιών

Ένας ενδιαφέρον τρόπος για να μελετήσουμε την ακραία συμπεριφορά μιας τυχαίας διαδικασίας, είναι μέσω των σημειακών διαδικασιών (*point processes*). Μια σημειακή διαδικασία μπορούμε να την φανταστούμε σαν μια τυχαία κατανομή σημείων στον χώρο π.χ. στο $[a, b] \subset R$ ή στο $[a_1, b_1] \times [a_2, b_2] \subset R^2$. Για τον ακριβή ορισμό της σημειακής διαδικασίας θα χρειαστούμε την έννοια του σημειακού μέτρου (*point measure*) η οποία δίνεται παρακάτω.

Ορισμός 1.7.1 Έστω $E \subseteq \mathbb{R}^k$ και \mathfrak{I}_E η ελάχιστη σ -άλγεβρα η οποία παράγεται από τα ανοιχτά υποσύνολα του E . Για οποιαδήποτε ακολουθία σημείων $(x_i)_{i \geq 1}$ στο E , όχι απαραίτητα διαφορετικά μεταξύ τους, ορίζουμε ως **σημειακό μέτρο** το μέτρο m για το οποίο ισχύει

$$m(A) = \sum_{i=1}^{\infty} \varepsilon_{x_i}(A) = \sum_{i: x_i \in A} 1 = \#\{i : x_i \in A\}, \quad (A \in \mathfrak{I}_E)$$

και $m(K) < \infty$ για κάθε συμπαγές υποσύνολο $K \subset E$ (Το μέτρο ε είναι μέτρο Dirac και με το $\#$ συμβολίζουμε το πλήθος των στοιχείων ενός συνόλου).

Το σημειακό μέτρο m λοιπόν, απαριθμεί τα σημεία του χώρου E τα οποία «περιέχονται» μέσα στο σύνολο A . Συμβολίζουμε με $M_p(E)$ το χώρο όλων των σημειακών μέτρων στο E και με $\mathcal{M}_p(E)$ την ελάχιστη σ -άλγεβρα η οποία περιέχει τα σύνολα της μορφής $\{m \in M_p(E) : m(A) \in B\}$ όπου $A \in \mathfrak{I}_E$ και $B \in \mathcal{B}([0, \infty])$ (B Borel υποσύνολο του $[0, \infty]$), δηλαδή $\mathcal{M}_p(E)$ είναι η ελάχιστη σ -άλγεβρα η οποία κάνει μετρήσιμες τις απεικονίσεις $m \rightarrow m(A)$ για κάθε $A \in \mathfrak{I}_E$.

Ορισμός 1.7.2 Μία μετρήσιμη συνάρτηση N από τον χώρο πιθανότητας (Ω, \mathcal{F}, P) στο χώρο $[M_p(E), \mathcal{M}_p(E)]$ καλείται **σημειακή διαδικασία** στο σύνολο E .

$(N: (\Omega, \mathcal{F}, P) \rightarrow (M_p(E), \mathcal{M}_p(E)))$.

Συνεπώς, η σημειακή διαδικασία N είναι ένα τυχαίο στοιχείο του χώρου $M_p(E)$. Αν διαλέξουμε ένα $\omega \in \Omega$, τότε η απεικόνιση $N(\omega, \cdot)$ είναι σημειακό μέτρο ενώ το $N(\omega, A)$ εκφράζει το πλήθος των σημείων που βρίσκονται στο A ($A \in \mathfrak{I}_E$) μετά την πραγματοποίηση του ω .

Αποδεικνύεται ότι η συνάρτηση $N: \Omega \rightarrow M_p(E)$ είναι σημειακή διαδικασία αν και μόνο αν η απεικόνιση $N: (\Omega, \mathcal{F}) \rightarrow ([0, \infty], \mathcal{B}([0, \infty]))$, $\omega \rightarrow N(\omega, A)$ είναι μετρήσιμη για κάθε $A \in \mathfrak{I}_E$. Έστω τώρα, $m = \sum_{i=1}^{\infty} \varepsilon_{x_i}$ ένα σημειακό μέτρο στο E . Υποθέτουμε ότι $(x_i)_{i \geq 1}$ είναι μια ακολουθία σημείων στο E , όχι απαραίτητα διαφορετικά μεταξύ τους, και θεωρούμε την υπακολουθία $(y_i)_{i \geq 1}$ της που αποτελείται από όλα τα σημεία $\{x_i\}$ που είναι διαφορετικά

μεταξύ τους. Για παράδειγμα, αν $\{x_i, i=1, \dots, 5\} = \{1, 2, 2, 3, 4\}$ τότε $\{y_i\} = \{1, 2, 3, 4\}$. Ορίζουμε ως **πολλαπλότητα** του στοιχείου y_i την ακολουθία $n_i = \#\{j, j \geq 1, y_i = x_j\}$, δηλαδή το n_i εκφράζει τον αριθμό εμφανίσεων του y_i στην αρχική ακολουθία $(x_i)_{i \geq 1}$. Αν γράψουμε $m = \sum_{i=1}^{\infty} n_i \varepsilon_{y_i}$, τότε το m είναι μέτρο το οποίο ονομάζουμε **απλό σημειακό μέτρο (simple point measure)** αν $n_i = 1$ για όλα τα i . Σε διαφορετική περίπτωση, το m ονομάζεται **πολλαπλό σημειακό μέτρο (multiple point measure)**.

Ορισμός 1.7.3 Η σημειακή διαδικασία N ονομάζεται **απλή (simple point process)** αν οι πραγματοποιήσεις της είναι απλά σημειακά μέτρα. Σε αντίθετη περίπτωση ονομάζεται **πολλαπλή (multiple point process)**.

Ορισμός 1.7.4 Έστω σημειακή διαδικασία $N: (\Omega, \mathcal{F}, P) \rightarrow (M_p(E), \mathcal{M}_p(E))$. Η κατανομή της N είναι το μέτρο πιθανότητας $P \circ N^{-1} = P(N \in A)$ με $A \in \mathcal{M}_p(E)$.

Παρατήρηση 1.7.5 Αποδεικνύεται ότι η κατανομή $P \circ N^{-1}$ καθορίζεται μονοσήμαντα από την κατανομή του τυχαίου διανύσματος $(N(A_1), \dots, N(A_m))$ για οποιαδήποτε επιλογή των $A_i \in \mathfrak{S}_E, i=1, \dots, m$ ($m \geq 1$).

Πολύ σημαντική στις εφαρμογές είναι η στοχαστική διαδικασία *Poisson (Poisson process ή Poisson random measure)*, η οποία είναι ειδική περίπτωση σημειακής διαδικασίας. Αποτελεί κατάλληλο μοντέλο για διαδικασίες των οποίων τα γεγονότα συμβαίνουν συνεχώς στο χρόνο και είναι ανεξάρτητα μεταξύ τους. Στην περίπτωση μας, θα δούμε παρακάτω ότι η διαδικασία *Poisson* και συγκεκριμένα η ομογενής διαδικασία *Poisson (Homogeneous Poisson process)* μπορεί να προκύψει ως όριο της διαδικασίας των υπερβάσεων από την ακολουθία $X_i, i=1, 2, \dots$.

Ορισμός 1.7.6 Μία σημειακή διαδικασία N ονομάζεται *Poisson διαδικασία ή μέτρο Poisson (Poisson random measure)*, με μέσο μέτρο (**mean measure**) μ , αν ικανοποιούνται οι παρακάτω συνθήκες,

1. Για $A \in \mathfrak{S}_E$ και $k \geq 0$

$$P(N(A) = k) = \begin{cases} \exp(-\mu(A)) \frac{(\mu(A))^k}{k!} & , \text{αν } \mu(A) < \infty \\ 0 & , \text{αν } \mu(A) = \infty \end{cases}$$

2. Για κάθε $m \geq 1$ και $A_1, \dots, A_m \in \mathfrak{S}_E$ με $A_i \cap A_j = \emptyset$, $i, j = 1, \dots, m$ οι τυχαίες μεταβλητές $N(A_1), \dots, N(A_m)$ είναι ανεξάρτητες (εν συντομία θα συμβολίζουμε τη διαδικασία με $PRM(\mu)$).

Παρατήρηση 1.7.7 Στην ειδική περίπτωση όπου $\mu(A) = \lambda |A|$, $\lambda > 0$ και $|\cdot|$ το μέτρο *Lebesgue*, η διαδικασία ονομάζεται ομογενής διαδικασία *Poisson* με ένταση (*intensity*) λ .

Ορισμός 1.7.8 (ασθενής σύγκλιση σημειακών διαδικασιών) Έστω N, N_1, N_2, \dots σημειακές διαδικασίες με χώρο καταστάσεων τον $E \subseteq \bar{R}^d$ ($\bar{R} = R \cup \{\pm\infty\}$) εφοδιασμένο με τη σ -άλγεβρα \mathfrak{S}_E . Θα λέμε ότι η N_n συγκλίνει ασθενώς στην $N \in M_p(E)$ ($N_n \rightarrow_d N$) αν

$$P(N_n(A_1) = k_1, \dots, N_n(A_m) = k_m) \rightarrow P(N(A_1) = k_1, \dots, N(A_m) = k_m), \quad n \rightarrow \infty,$$

για οποιαδήποτε επιλογή των $A_i \in \mathfrak{S}_E$, $i = 1, \dots, m$ ($m \geq 1$), $k_1, k_2, \dots, k_m \in N$ με $P(N(\partial A_i) = 0) = 1$ όπου ∂A το σύνορο του A .

Θα διατυπώσουμε τώρα το θεώρημα το οποίο μας δίνει τις συνθήκες που πρέπει να ικανοποιούνται ώστε μία σημειακή διαδικασία σε ένα υποσύνολο του R να συγκλίνει ασθενώς σε μία απλή σημειακή διαδικασία.

Θεώρημα 1.7.9 (*Kallenberg*) Έστω N_n και N σημειακές διαδικασίες στο $E = (\alpha, b] \subset R$ με N απλή σημειακή διαδικασία. Υποθέτουμε ότι ισχύουν οι παρακάτω συνθήκες,

$$E(N_n(A)) \rightarrow E(N(A)), \quad n \rightarrow \infty$$

για κάθε $A = (c, d]$ και $\alpha < c < d \leq b$ και

$$P(N_n(B) = 0) \rightarrow P(N(B) = 0), \quad n \rightarrow \infty$$

για κάθε $B = \bigcup_{i=1}^k (c_i, d_i]$ και $\alpha < c_1 < d_1 < \dots < c_k < d_k \leq b$, $k \geq 1$. Τότε $N_n \rightarrow_d N$ στο χώρο $M_p(E)$.

Η σημειακή διαδικασία η οποία παρουσιάζει ενδιαφέρον στη θεωρία ακραίων τιμών είναι η διαδικασία υπερβάσεων κατωφλίου (*point process of exceedances*). Έστω λοιπόν X_1, X_2, \dots ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Θεωρούμε τη σημειακή διαδικασία N_n στο $E = (0,1]$ με

$$(1.5) \quad N_n(\cdot) = \sum_{i=1}^n \varepsilon_{i/n}(\cdot) I\{X_i > u_n\}, \quad (n \geq 1)$$

όπου $I(A)$ η δείκτρια συνάρτηση του A . Μια πραγματοποίηση λοιπόν της N_n , μας δίνει το πλήθος των σημείων $i/n \in E$ ή ισοδύναμα $i \in nE$, στα οποία έχουμε πραγματοποίηση του ενδεχομένου $\{X_i > u_n\}$. Έχοντας κατά νου το Θεώρημα *Kallenberg* θα δείξουμε ότι η σημειακή διαδικασία N_n συγκλίνει ασθενώς σε μία ομογενή διαδικασία *Poisson*. Στην Πρόταση 1.1.3 είδαμε ότι $P(M_n \leq u_n) \rightarrow_{n \rightarrow \infty} e^{-g}$ αν και μόνο αν $n\bar{F}(u_n) \rightarrow_{n \rightarrow \infty} g$. Η ποσότητα $n\bar{F}(u_n)$ εκφράζει την μέση τιμή των υπερβάσεων του κατωφλίου u_n , οι οποίες όπως είδαμε ακολουθούν διωνυμική κατανομή. Θα δείξουμε και πάλι ότι ισχύει η Πρόταση 1.1.3 κάνοντας χρήση της θεωρίας που αναπτύξαμε για τις σημειακές κατανομές.

Θεώρημα 1.7.10 Έστω $(X_n)_{n \geq 1}$ ανεξάρτητες και ισόνομες (i.i.d) τυχαίες μεταβλητές με συνάρτηση κατανομής F . Επίσης, υποθέτουμε ότι για κάθε $g \in (0, \infty)$ υπάρχει ακολουθία $(u_n)_{n \geq 1}$ τέτοια ώστε να ισχύει $n\bar{F}(u_n) \rightarrow_{n \rightarrow \infty} g$. Τότε η σημειακή διαδικασία (1.5) συγκλίνει ασθενώς στην ομογενή διαδικασία *Poisson* N στο $(0,1]$ με ένταση g (η N είναι $PRM(g|\cdot|)$).

Απόδειξη Παρατηρούμε ότι αν N είναι ομογενής *PRM* τότε μπορεί να γραφεί ως $N(\cdot) = \sum_{i=1}^{\infty} \varepsilon_{T_i}(\cdot)$, όπου $T_i = Y_1 + \dots + Y_i$ με Y_i ανεξάρτητες τ.μ που ακολουθούν την εκθετική κατανομή, έστω με παράμετρο g . Συνεπώς, μια ομογενής *PRM* είναι απλή σημειακή διαδικασία. Υποθέτουμε ότι η οριακή διαδικασία N είναι εμφυτευμένη σε μία ομογενή διαδικασία *Poisson* στο $(0, \infty)$. Υπό αυτή την έννοια θα είναι απλή. Μπορούμε λοιπόν να κάνουμε χρήση του Θεωρήματος *Kallenberg*. Έστω λοιπόν $A = (\alpha, b] \subset (0,1]$. Τότε, για κάθε n τέτοιο ώστε $[nb] - [na] \geq 1$, έχουμε ότι

$$N_n(A) = \sum_{i=1}^n \varepsilon_{i/n}(A) I\{X_i > u_n\} = \sum_{\alpha < i/n \leq b} I\{X_i > u_n\} = \sum_{i=[n\alpha]+1}^{[nb]} I\{X_i > u_n\}.$$

Συνεπώς, $N_n(A) \sim \text{Bin}([nb] - [n\alpha], \bar{F}(u_n))$ ($[\cdot]$ η συνάρτηση ακέραιο μέρος). Είναι γνωστό επίσης ότι $[nb] - [n\alpha] \sim n(b - \alpha)$ και από την υπόθεση του θεωρήματος παίρνουμε ότι $\bar{F}(u_n) \sim g/n$. Έχουμε λοιπόν ότι

$$EN_n(A) = ([nb] - [n\alpha])\bar{F}(u_n) \sim (n(b - \alpha))(g/n) = g(b - \alpha) = E(N(A)).$$

Επίσης,

$$\begin{aligned} P(N_n(A) = 0) &= \binom{[nb] - [n\alpha]}{0} (1 - F(u_n))^0 F^{[nb] - [n\alpha]}(u_n) \\ &= F^{[nb] - [n\alpha]}(u_n) \\ &= \exp\{([nb] - [n\alpha]) \ln(1 - \bar{F}(u_n))\} \\ &\rightarrow \exp\{-g(b - \alpha)\}, \quad n \rightarrow \infty \end{aligned}$$

(λαμβάνοντας υπόψη ότι $\ln(1 - \bar{F}(u_n)) \sim -\bar{F}(u_n)$ για $\bar{F}(u_n) \rightarrow 0$). Έστω τώρα $B = \bigcup_{i=1}^k (c_i, d_i]$

όπου $0 < c_1 < d_1 < \dots < c_k < d_k \leq 1$. Έχουμε ότι

$$\begin{aligned} P(N_n(B) = 0) &= P(N_n(c_i, d_i] = 0, i = 1, \dots, k) \\ &= P(\text{δεν έχουμε υπέρβαση του } u_n \text{ στο σύνολο } B) \\ &= P(\max_{c_i < j/n \leq d_i} X_j \leq u_n, i = 1, \dots, k) = P(\max_{[nc_i] < j \leq [nd_i]} X_j \leq u_n, i = 1, \dots, k) \\ &= \prod_{i=1}^k P(\max_{[nc_i] < j \leq [nd_i]} X_j \leq u_n) = \prod_{i=1}^k P(N_n(c_i, d_i] = 0) \\ &\rightarrow \prod_{i=1}^k \exp\{-g(d_i - c_i)\}, \quad n \rightarrow \infty. \end{aligned}$$

Για την ομογενή διαδικασία *Poisson* γνωρίζουμε ότι

$$P(N(B) = 0) = \exp(-g | B|) = \exp\{-g \sum_{i=1}^k (d_i - c_i)\} = \prod_{i=1}^k \exp\{-g(d_i - c_i)\}.$$

Από το Θεώρημα *Kallenberg* λοιπόν έχουμε το ζητούμενο. ■

Γενικότερα ισχύει το παρακάτω αποτέλεσμα,

$$P(X_{i,n} \leq u_n) = P(N_n(0,1] \leq i-1) \rightarrow e^{-g} \sum_{j=0}^{i-1} \frac{g^j}{j!}, \quad i=1, \dots, n, \quad n \rightarrow \infty.$$

Δείξαμε λοιπόν, ότι αν οι υπερβάσεις γίνονται εντελώς τυχαία κατά την εξέλιξη της διαδικασίας και δεν υπάρχει εξάρτηση μεταξύ τους, τότε η ομογενής διαδικασία *Poisson* είναι κατάλληλο μοντέλο για την περιγραφή τους. Υποθέτουμε λοιπόν ότι η τ.μ K εκφράζει τον αριθμό των υπερβάσεων ενός κατωφλίου u και $K \sim Pois(\lambda)$. Επίσης θεωρούμε ότι το μέγεθος των υπερβάσεων Y_1, Y_2, \dots, Y_K είναι ανεξάρτητες και ισόνομες τ.μ με κατανομή την *GPD* με παραμέτρους $\xi, \tilde{\sigma}$ δεδομένου ότι $K \geq 1$. Εάν υποθέσουμε ότι οι τ.μ Y_1, Y_2, \dots, Y_K είναι ανεξάρτητες της τ.μ K θα έχουμε ότι

$$\begin{aligned} P(\max_{1 \leq i \leq K} Y_i \leq x) &= \sum_{n=0}^{\infty} P(\max_{1 \leq i \leq n} Y_i \leq x, K = n) \\ &= P(N = 0) + \sum_{n=1}^{\infty} P(Y_1 \leq x, \dots, Y_n \leq x, K = n) \\ &= e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \left\{ 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}} \right)^{-1/\xi} \right\}^n = \exp \left\{ -\lambda \left(1 + \frac{\xi x}{\tilde{\sigma}} \right)^{-1/\xi} \right\}, \end{aligned}$$

όπου $1 + \xi x / \tilde{\sigma} > 0$.

Το μοντέλο αυτό αναφέρεται στη βιβλιογραφία ως ***Poisson-GPD model***. Το παραπάνω αποτέλεσμα αποτελεί ουσιαστικά απόδειξη του (4) του Θεωρήματος 1.1.5.

Για περισσότερες πληροφορίες σχετικά με τις σημειακές διαδικασίες παραπέμπουμε στους *Daley and Vere-Jones (1988)*, *Resnick (1987)*.

ΚΕΦΑΛΑΙΟ 2

Θεωρία Ακραίων Τιμών για ισχυρά στάσιμες ακολουθίες τυχαίων μεταβλητών

2.1 Εισαγωγή

Στο πρώτο κεφάλαιο της παρούσας εργασίας αναπτύξαμε συνοπτικά τα βασικά αποτελέσματα της θεωρίας ακραίων τιμών για ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Σε αυτό το κεφάλαιο θα επιχειρήσουμε την γενίκευση των αποτελεσμάτων για εξαρτημένες τυχαίες μεταβλητές και δη για ισχυρά στάσιμες ακολουθίες τυχαίων μεταβλητών (*strictly stationary sequences*). Ο ακριβής ορισμός μίας ισχυρά στάσιμης ακολουθίας (ι.σ.α.) δίνεται παρακάτω.

Ορισμός 2.1.1 Μία ακολουθία τυχαίων μεταβλητών $\{X_t, t \in Z\}$ καλείται *ισχυρά στάσιμη* (*strictly stationary*) αν η κατανομή των τυχαίων διανυσμάτων

$$(X_{t_1}, \dots, X_{t_n}) \text{ και } (X_{t_1+l}, \dots, X_{t_n+l})$$

είναι ίδια για κάθε επιλογή των $t_1, \dots, t_n \in Z$, $n \geq 1$ και $l \in Z$.

Από τον ορισμό λοιπόν είναι φανερό ότι μία ι.σ.α. επιδεικνύει την ίδια στοχαστική συμπεριφορά με το πέρασμα του χρόνου. Άμεση συνέπεια αυτού, είναι ότι η μέση τιμή και η διασπορά της είναι σταθερές και ανεξάρτητες του χρόνου. Ωστόσο, η εξάρτηση που υπάρχει ανάμεσα στις τυχαίες μεταβλητές $X_i, i=1, \dots, n$ μπορεί να επηρεάσει σημαντικά τις ποσότητες που μελετήσαμε στο πρώτο κεφάλαιο, όπως για παράδειγμα την συμπεριφορά και το μέγεθος των ακραίων τιμών.

Ένα σύννηθες φαινόμενο που παρατηρείται σε στοχαστικές διαδικασίες οι οποίες παρουσιάζουν θετική εξάρτηση, είναι η «τοπική» συγκέντρωση ακραίων τιμών, δηλαδή η πιθανότητα να παρατηρήσουμε διαδοχικές ακραίες τιμές είναι μεγαλύτερη σε σχέση με μία

διαδικασία που δεν παρουσιάζει εξάρτηση. Συνεπώς, τα μοντέλα που έχουμε αναπτύξει μέχρι τώρα δεν μπορούν να αποτελέσουν αξιόπιστα εργαλεία. Ωστόσο, όπως θα δούμε και παρακάτω, μπορούμε να κάνουμε χρήση των αποτελεσμάτων του πρώτου κεφαλαίου αν η ι.σ.α. ικανοποιεί κάποιες συγκεκριμένες συνθήκες. Οι συνθήκες θα είναι τέτοιες ώστε να αποτρέπεται το φαινόμενο της μακροχρόνιας εξάρτησης μεταξύ των τιμών της, όχι όμως και το φαινόμενο της τοπικής εξάρτησης που αναφέραμε παραπάνω. Σκοπός μας δηλαδή είναι να επιβάλουμε συνθήκες οι οποίες θα καταστούν ασυμπτωτικά ανεξάρτητες τις ακραίες παρατηρήσεις που απέχουν μεταξύ τους κατά επαρκώς μεγάλα χρονικά διαστήματα. Διαισθητικά, είναι σαφές ότι τα χρονικά διαστήματα θα έχουν άμεση σχέση με τον «βαθμό εξάρτησης» των παρατηρήσεων που υπερβαίνουν κάποιο συγκεκριμένο κατώφλι. Όσο πιο μεγάλος είναι ο βαθμός εξάρτησης τόσο μεγαλύτερα θα είναι και τα χρονικά διαστήματα. Παρακάτω θα δούμε ότι ο βαθμός εξάρτησης έχει άμεση σχέση με την παράμετρο που ονομάζουμε δείκτη ακρότατης εξάρτησης (*extremal index*).

Πριν ξεκινήσουμε την γενικότερη ανάλυση της θεωρίας μπορούμε να δούμε τι συμβαίνει σε μία συγκεκριμένη περίπτωση με το παράδειγμα που ακολουθεί.

Παράδειγμα 2.1.2 Ας υποθέσουμε ότι ρίχνουμε n φορές ένα αμερόληπτο ζάρι. Το αποτέλεσμα κάθε ρίψης είναι τυχαία μεταβλητή την οποία την συμβολίζουμε με $Y_i, i = 1, 2, \dots, n$. Θεωρούμε ότι οι τ.μ Y_i είναι ανεξάρτητες και ισόνομες με σ.κ F . Επιπλέον, ορίζουμε τις τυχαίες μεταβλητές $X_i = \max(Y_i, Y_{i+1}), i = 1, 2, \dots, n-1$. Είναι σαφές ότι οι τ.μ X_i δεν είναι ανεξάρτητες καθώς

$$P(X_i \leq x, X_{i+1} \leq y) = F(x) F(\min(x,y)) F(y)$$

και

$$P(X_i \leq x)P(X_{i+1} \leq y) = F^2(x) F^2(y)$$

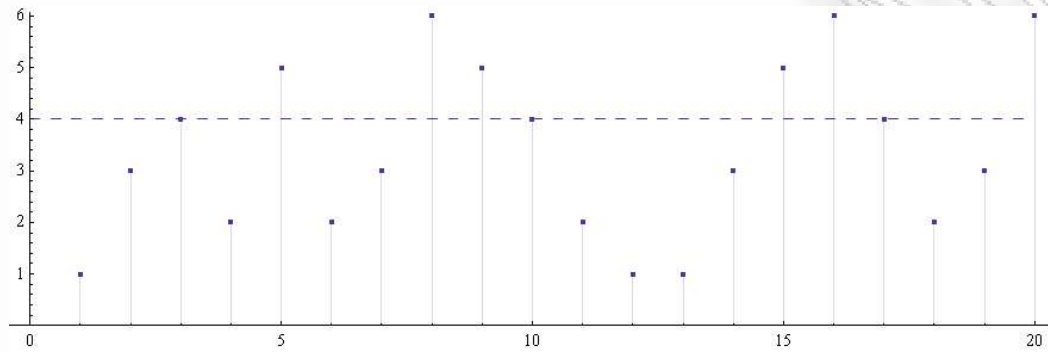
για $x, y = 1, \dots, 6$ και $i = 1, \dots, n-1$. Αν υποθέσουμε ότι $n \rightarrow \infty$ εύκολα διαπιστώνουμε ότι η ακολουθία $\{X_i, i = 1, 2, \dots\}$ είναι ισχυρά στάσιμη. Έστω τώρα ότι έχουμε ρίξει το ζάρι είκοσι φορές και έχουμε πάρει το παρακάτω αποτέλεσμα για τις τιμές της ακολουθίας $\{Y_i, i = 1, 2, \dots, 20\}$,

1, 3, 4, 2, 5, 2, 3, 6, 5, 4, 2, 1, 1, 3, 5, 6, 4, 2, 3, 6.

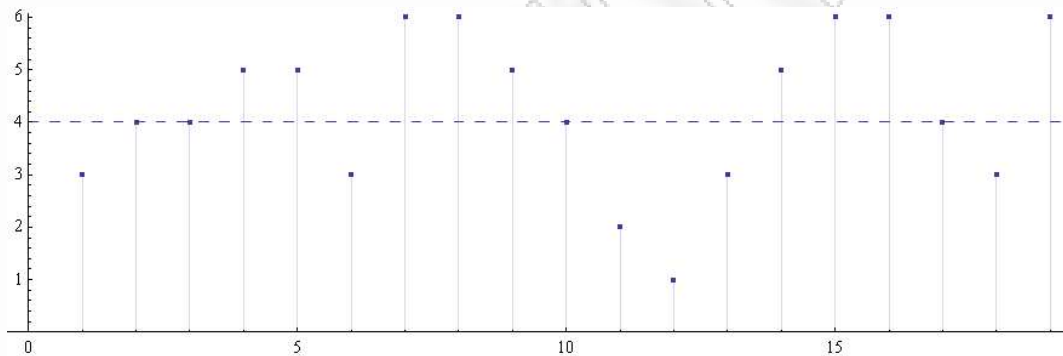
Αντίστοιχα, οι τιμές της ακολουθίας $\{X_i, i=1,2,\dots,19\}$ θα είναι

3, 4, 4, 5, 5, 3, 6, 6, 5, 4, 2, 1, 3, 5, 6, 6, 4, 3, 6.

Στα παρακάτω διαγράμματα απεικονίζονται οι αντίστοιχες τιμές.



$\{Y_i, i=1,\dots,20\}$



$\{X_i, i=1,\dots,19\}$

Αν λοιπόν τώρα θεωρήσουμε ως ακραία αποτελέσματα τις τιμές των διαδικασιών $\{X_i, i = 1, 2, \dots\}$ και $\{Y_i, i = 1, 2, \dots\}$ που είναι μεγαλύτερες ή ίσες του πέντε ($u = 4$), τότε με μία πρώτη ματιά παρατηρούμε ότι τα ακραία γεγονότα (τιμές) της $\{X_i, i=1,2,\dots\}$ στο συγκεκριμένο παράδειγμα εμφανίζονται σε ζεύγη. Αυτό το γεγονός αποτελεί άμεση συνέπεια της εξάρτησης που έχουν οι τυχαίες μεταβλητές X_i . Εάν έχουμε n διαδοχικές υπερβάσεις του κατωφλίου $u=4$ από την ακολουθία $\{Y_i, i=1,2,\dots\}$ στα σημεία $j, j+1,\dots, j+n-1$, οι αντίστοιχες υπερβάσεις της ακολουθίας $\{X_i, i=1,2,\dots\}$ θα συμβαίνουν στα σημεία $j-1, j,\dots, j+n-1$, δίνοντας έτσι μία ακραία παρατήρηση παραπάνω από την $\{Y_i, i=1,2,\dots\}$. Στο συγκεκριμένο παράδειγμα οι χρονικές στιγμές των υπερβάσεων και το πλήθος των διαδοχικών υπερβάσεων είναι οι $j = 5$ με $n = 1, j = 8$ με $n = 2, j = 15$ με $n = 2, j = 20$ με $n = 1, j = 22$ με $n = 2, j = 25$ με $n = 1$ και $j = 28$ με $n = 1$.

2.2 Ασυμπτωτικές κατανομές μεγίστου

Όπως αναφέραμε και παραπάνω στόχος μας είναι να επιβάλλουμε σε μία ισχυρά στάσιμη ακολουθία να ικανοποιεί κάποια συνθήκη η οποία θα αποτρέπει την μακροχρόνια εξάρτηση στα ακραία επίπεδα, έτσι ώστε τα ενδεχόμενα $\{X_i > u\}$ και $\{X_j > u\}$ να είναι ασυμπτωτικώς ανεξάρτητα, για επαρκώς μεγάλες τιμές του κατωφλίου u και χρονική απόσταση μεταξύ των στιγμών i, j που θα καθορίζεται από τον βαθμό εξάρτησης των τυχαίων μεταβλητών $X_i, i = 1, 2, \dots$

Έστω λοιπόν η ι.σ.α. $\{X_i, i = 1, 2, \dots\}$. Για ένα σύνολο θετικών ακεραίων, έστω J , θα χρησιμοποιούμε τον συμβολισμό $M(J) = \max_{i \in J} X_i$ όπου $M(\emptyset) = -\infty$. Θεωρούμε επιπλέον τις ακολουθίες θετικών ακεραίων $(r_n) \in \mathbb{N}, n \in \mathbb{N}$ με $r_n = o(n)$ όταν $n \rightarrow \infty$ και $k_n = [n/r_n]$. Αρχικά, χωρίζουμε το σύνολο $\{1, \dots, n\}$ σε k_n ομάδες (**blocks**) όπου κάθε ομάδα έχει μέγεθος r_n . Το σύνολο των στοιχείων που περιέχει κάθε ομάδα θα το συμβολίζουμε με J_j όπου

$$(2.1) \quad J_j = J_{j,n} = \{(j-1)r_n + 1, \dots, jr_n\}, \quad j=1, \dots, k_n.$$

Στην περίπτωση όπου $k_n r_n < n$ λαμβάνουμε υπόψη και την ομάδα $J_{k_n+1} = \{k_n r_n + 1, \dots, n\}$.

Προφανώς $M_n = \max_j M(J_j)$. Έχουμε λοιπόν ότι

$$(2.2) \quad \begin{aligned} P(M_n \leq u_n) &= P(\max_j M(J_j) \leq u_n) = P\left(\bigcap_{j=1}^{k_n+1} \{M(J_j) \leq u_n\}\right) \\ &= 1 - P\left(\bigcup_{j=1}^{k_n+1} \{M(J_j) > u_n\}\right) \\ &= 1 - P\left(\bigcup_{j=1}^{k_n} \{M(J_j) > u_n\}\right) - P(M(J_{k_n+1}) > u_n) \\ &\quad + P\left(\left(\bigcup_{j=1}^{k_n} \{M(J_j) > u_n\}\right) \cap \{M(J_{k_n+1}) > u_n\}\right), \end{aligned}$$

όπου (u_n) ακολουθία κατωφλίων για την οποία ισχύει ότι $\limsup_{n \rightarrow \infty} n \bar{F}(u_n) < \infty$, δηλαδή ο

αναμενόμενος αριθμός των υπερβάσεων του κατωφλίου u_n είναι πεπερασμένος. Επίσης,

$$(2.3) \quad \begin{aligned} P(M(J_j) > u_n) &= P\left(\bigcup_{i \in J_j} \{X_i > u_n\}\right) \leq \sum_{i=(j-1)r_n+1}^{jr_n} P(X_i > u_n) \\ &= r_n \bar{F}(u_n) = n \frac{r_n}{n} \bar{F}(u_n) \rightarrow_{n \rightarrow \infty} 0. \end{aligned}$$

Επιπλέον, ισχύει ότι

$$\left(\bigcup_{j=1}^{k_n} \{M(J_j) > u_n\}\right) \cap \{M(J_{k_n+1}) > u_n\} \subseteq \{M(J_{k_n+1}) > u_n\}.$$

Συνεπώς,

$$(2.4) \quad P\left(\left(\bigcup_{j=1}^{k_n} \{M(J_j) > u_n\}\right) \cap \{M(J_{k_n+1}) > u_n\}\right) \leq P(\{M(J_{k_n+1}) > u_n\}).$$

Από τις σχέσεις (2.3) και (2.4) προκύπτει το παρακάτω αποτέλεσμα

$$(2.5) \quad P\left(\left(\bigcup_{j=1}^{k_n} \{M(J_j) > u_n\}\right) \cap \{M(J_{k_n+1}) > u_n\}\right) \rightarrow_{n \rightarrow \infty} 0.$$

Από την σχέση (2.3) λοιπόν, βλέπουμε ότι μπορούμε να παραλείψουμε την k_n+1 ομάδα, εφόσον

$$(2.6) \quad P(M(J_{k_n+1}) > u_n) \rightarrow_{n \rightarrow \infty} 0.$$

Πιο αυστηρά, εφαρμόζοντας τις σχέσεις (2.5) και (2.6) στην σχέση (2.2) παίρνουμε την παρακάτω ισότητα

$$(2.7) \quad P(M_n \leq u_n) = P\left(\bigcap_{j=1}^{k_n} \{M(J_j) \leq u_n\}\right) + o(1).$$

Θεωρούμε τώρα τα σύνολα

$$J_j^l = J_{j,n}^l = \{(j-1)r_n + 1, \dots, jr_n - s_n\} \quad \text{και} \quad J_j^k = J_{j,n}^k = \{jr_n - s_n + 1, \dots, jr_n\},$$

όπου (s_n) ακολουθία θετικών ακεραίων για την οποία ισχύει ότι $s_n = o(r_n)$ όταν $n \rightarrow \infty$. Προφανώς $J_j = J_j^l \cup J_j^k$. Ακολουθώντας τον ίδιο συλλογισμό από τον οποίο προέκυψε η (2.3) προκύπτει ότι

$$(2.8) \quad P\left(\bigcup_{j=1}^{k_n} \{M(J_j^k) > u_n\}\right) \leq k_n s_n \bar{F}(u_n) = n \frac{k_n r_n}{n} \frac{s_n}{r_n} \bar{F}(u_n) \rightarrow_{n \rightarrow \infty} 0.$$

Αν θέσουμε $A_j = \{M(J_j^l) \leq u_n\}$ και $B_j = \{M(J_j^k) \leq u_n\}$ τότε από την σχέση (2.7) έχουμε ότι

$$\begin{aligned}
P(M_n \leq u_n) &= P\left(\bigcap_{j=1}^{k_n} \{M(J_j) \leq u_n\}\right) + o(1) = P\left(\bigcap_{j=1}^{k_n} \{M(J_j^l \cup J_j^k) \leq u_n\}\right) + o(1) \\
&= P\left(\bigcap_{j=1}^{k_n} \{\{M(J_j^l) \leq u_n\} \cap \{M(J_j^k) \leq u_n\}\}\right) + o(1) = P\left(\bigcap_{j=1}^{k_n} \{A_j \cap B_j\}\right) + o(1) \\
&= 1 - P\left(\bigcup_{j=1}^{k_n} \{A_j^c \cup B_j^c\}\right) + o(1) = 1 - P\left(\left\{\bigcup_{j=1}^{k_n} A_j^c\right\} \cup \left\{\bigcup_{j=1}^{k_n} B_j^c\right\}\right) + o(1) = 1 - P\left(\bigcup_{j=1}^{k_n} A_j^c\right) \\
(2.9) \quad &- P\left(\bigcup_{j=1}^{k_n} B_j^c\right) + P\left(\left\{\bigcup_{j=1}^{k_n} A_j^c\right\} \cap \left\{\bigcup_{j=1}^{k_n} B_j^c\right\}\right) + o(1).
\end{aligned}$$

Χρησιμοποιώντας την σχέση (2.8) η (2.9) γίνεται

$$(2.10) \quad P(M_n \leq u_n) = 1 - P\left(\bigcup_{j=1}^{k_n} A_j^c\right) + o(1) = P\left(\bigcap_{j=1}^{k_n} A_j\right) + o(1) = P\left(\bigcap_{j=1}^{k_n} \{M(J_j^l) \leq u_n\}\right) + o(1).$$

Επίσης,

$$\begin{aligned}
P(M(J_j) \leq u_n) &= P(M(J_j^l \cup J_j^k) \leq u_n) + o(1) \\
&= P(\{M(J_j^l) \leq u_n\} \cap \{M(J_j^k) \leq u_n\}) + o(1) \\
&= 1 - P(M(J_j^l) > u_n) - P(M(J_j^k) > u_n) \\
(2.11) \quad &+ P(\{M(J_j^l) > u_n\} \cap \{M(J_j^k) > u_n\}) + o(1).
\end{aligned}$$

Όμως από την σχέση (2.8) προκύπτει ότι

$$(2.12) \quad P(M(J_j^k) > u_n) \leq P\left(\bigcup_{j=1}^{k_n} \{M(J_j^k) > u_n\}\right) \leq k_n s_n \bar{F}(u_n) \rightarrow_{n \rightarrow \infty} 0.$$

Χρησιμοποιώντας λοιπόν την σχέση (2.12), η (2.11) δίνει την παρακάτω ισότητα

$$(2.13) \quad P(M(J_j) \leq u_n) = P(M(J_j^l) \leq u_n) + o(1).$$

Από τις σχέσεις (2.9) και (2.10) βλέπουμε λοιπόν ότι

$$P\left(\bigcap_{j=1}^{k_n} \{M(J_j) \leq u_n\}\right) = P\left(\bigcap_{j=1}^{k_n} \{M(J_j^l) \leq u_n\}\right) + o(1),$$

δηλαδή τα ενδεχόμενα $\bigcap_{j=1}^{k_n} \{M(J_j) \leq u_n\}$ και $\bigcap_{j=1}^{k_n} \{M(J_j^l) \leq u_n\}$ είναι ασυμπτωτικά ισοπίθανα και επιπλέον,

$$(2.14) \quad P(M_n \leq u_n) = P\left(\bigcap_{j=1}^{k_n} \{M(J_j) \leq u_n\}\right) + o(1) = P\left(\bigcap_{j=1}^{k_n} \{M(J_j^l) \leq u_n\}\right) + o(1).$$

Επίσης, από την σχέση (2.12) η πιθανότητα των ενδεχομένων $\{M(J_j^k) > u_n\}$, $j=1, \dots, k_n$, μπορεί να θεωρηθεί αμελητέα όταν $n \rightarrow \infty$.

Συνοψίζοντας λοιπόν τα βήματα που ακολουθήσαμε, έχουμε τα εξής:

1. Χωρίσαμε τις τ.μ X_i , $i=1, 2, \dots, n$ σε k_n+1 το πλήθος ομάδες J_j , όπου οι πρώτες k_n περιέχουν r_n το πλήθος τ.μ. (βλ. σχέση (2.1))
2. Δείξαμε ότι η k_n+1 ομάδα μπορεί ασυμπτωτικά να θεωρηθεί αμελητέα (βλ. σχέση (2.6)).
3. Χωρίσαμε κάθε ομάδα J_j σε δύο υποομάδες J_j^l και J_j^k μεγέθους $r_n - s_n$ και s_n αντίστοιχα.
4. Δείξαμε ότι οι ομάδες J_j^k μπορούν να παραληφθούν όταν $n \rightarrow \infty$, εφόσον έχουμε υποθέσει ότι $s_n = o(r_n)$, δηλαδή το μέγεθος των J_j^k συγκριτικά με το μέγεθος των J_j είναι ασυμπτωτικά αμελητέο (βλ. σχέση (2.8)).
5. Καταλήξαμε στο ότι η κατανομή της μέγιστης παρατήρησης μπορεί να προσεγγιστεί από την από κοινού κατανομή των μεγίστων των ομάδων J_j^l , ή ισοδύναμα των J_j . (βλ. σχέση (2.14))

Η σχέση (2.14) μας δίνει μία πρώτη εικόνα για την μορφή της συνθήκης που αναζητούμε. Εάν τα ενδεχόμενα $\{M(J_j^l) \leq u_n\}$, $j=1, \dots, k_n$ μπορούν να θεωρηθούν ανεξάρτητα όταν $n \rightarrow \infty$ τότε από η σχέση (2.14) έχουμε ότι

$$(2.15) \quad P(M_n \leq u_n) = \prod_{j=1}^{k_n} P(M(J_j) \leq u_n) + o(1) = \prod_{j=1}^{k_n} P(M(J_j^l) \leq u_n) + o(1).$$

Η συνθήκη που εξασφαλίζει ότι η σχέση (2.15) ικανοποιείται οριακά είναι η παρακάτω.

Ορισμός (Συνθήκη $D(u_n)$). Μία ακολουθία τ.μ. X_1, X_2, \dots θα λέμε ότι ικανοποιεί την συνθήκη $D(u_n)$ αν ισχύει η ακόλουθη απαίτηση. Για κάθε θετικούς ακέραιους p, q και

$$1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$$

όπου $j_1 - i_p \geq s := s_n = o(n)$ απαιτούμε να ισχύει η παρακάτω σχέση

$$|P(\max_{i \in A_1 \cup A_2} X_i \leq u_n) - P(\max_{i \in A_1} X_i \leq u_n)P(\max_{i \in A_2} X_i \leq u_n)| \leq a_{n,s} \rightarrow 0, \quad n \rightarrow \infty$$

όπου $A_1 = \{i_1, \dots, i_p\}$, $A_2 = \{j_1, \dots, j_q\}$.

Επιβάλλοντας σε μία ι.σ.α να ικανοποιεί την παραπάνω συνθήκη είναι φανερό ότι τα ενδεχόμενα $\{\max_{i \in A_1} X_i \leq u_n\}$ και $\{\max_{i \in A_2} X_i \leq u_n\}$ καθίστανται ασυμπτωτικώς ανεξάρτητα. Η παραπάνω συνθήκη, γενικεύεται (επαγωγικά) για $k = k_n$ το πλήθος σύνολα A_i τα οποία απέχουν μεταξύ τους απόσταση s , δίνοντας την παρακάτω σχέση

$$|P(\max_{i \in \bigcup_{i=1}^{k_n} A_i} X_i \leq u_n) - P(\max_{i \in A_1} X_i \leq u_n) \dots P(\max_{i \in A_{k_n}} X_i \leq u_n)| \leq (k_n - 1)a_{n,s_n}.$$

Αν λοιπόν $s_n = o(r_n)$ και $(k_n - 1)a_{n,s_n} \rightarrow 0$ όταν $n \rightarrow \infty$, τότε

$$(2.16) \quad |P(\bigcap_{j=1}^{k_n} \{M(J_j^l) \leq u_n\}) - \prod_{j=1}^{k_n} P(M(J_j^l) \leq u_n)| \leq (k_n - 1)a_{n,s_n} \rightarrow_{n \rightarrow \infty} 0.$$

Η σχέση (2.16) οδηγεί στην (2.15), την οποία θέλαμε να δείξουμε. Επιπλέον, αν υπάρχει ισχυρή εξάρτηση μεταξύ των τ.μ X_i , $i = 1, 2, \dots, n$, η ακολουθία s_n θα πρέπει να αυξάνει με γρήγορο ρυθμό και σε συνδυασμό με το γεγονός ότι $s_n = o(r_n)$ και $k_n = [n/r_n]$, ο ρυθμός αύξησης της ακολουθίας k_n θα πρέπει να είναι μικρός, εξασφαλίζοντας έτσι ότι $(k_n - 1)a_{n,s_n} \rightarrow 0$ όταν $n \rightarrow \infty$. Στην ειδική περίπτωση όπου έχουμε ανεξαρτησία των τ.μ X_i , $i = 1, 2, \dots, n$, μπορούμε να επιλέξουμε $k_n = n$ και $s_n = 1$.

Άμεση συνέπεια της σχέσης (2.15) είναι το παρακάτω θεώρημα (*Leadbetter (1974)*) το οποίο αποτελεί γενίκευση του Θεωρήματος 1.1.1 για μία ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών X_i , $i = 1, 2, \dots$

Θεώρημα 2.2.1 Έστω $X_i, i = 1, 2, \dots$ μία ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών. Υποθέτουμε ότι υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in R, n \in N$ τέτοιες ώστε

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow_{n \rightarrow \infty} G(x), \quad x \in R$$

όπου G μη εκφυλισμένη συνάρτηση κατανομής. Εάν επιπλέον ισχύει η συνθήκη $D(u_n)$, όπου $u_n = a_n x + b_n$ για κάθε x τέτοιο ώστε $G(x) > 0$, τότε η G ανήκει σε μία εκ των τριών οικογενειών του Θεωρήματος 1.1.1.

Για την απόδειξη του Θεωρήματος θα χρειαστούμε την παρακάτω πρόταση.

Πρόταση 2.2.2 Έστω A, B, A_1, A_2, \dots τυχαίες μεταβλητές και έστω οι ακολουθίες $b_n > 0, d_n > 0, a_n, c_n \in R$. Υποθέτουμε ότι ισχύει η παρακάτω σχέση

$$b_n^{-1}(A_n - a_n) \rightarrow_d A.$$

Τότε

$$(1) \quad d_n^{-1}(A_n - c_n) \rightarrow_d B,$$

αν και μόνο αν, για $n \rightarrow \infty$

$$(2) \quad b_n / d_n \rightarrow a \in [0, \infty) \text{ και } (a_n - c_n) / d_n \rightarrow b \in R.$$

Αν ισχύει η (1) τότε

$$(3) \quad B =_d aA + b$$

και a, b είναι οι μοναδικές σταθερές για τις οποίες ισχύει η σχέση (3).

Επίσης, αν ισχύει η (1) τότε η κατανομή A είναι μη εκφυλισμένη αν και μόνο αν $a > 0$.

Απόδειξη (Θεωρήματος 2.2.1) Από την πρώτη ισότητα της σχέσης (2.15) έχουμε ότι

$$(2.17) \quad P(M_n \leq u_n) = P^k(M_{[n/k]} \leq u_n) + o(1), \quad (k \geq 1).$$

Εφαρμόζοντας λοιπόν την υπόθεση του θεωρήματος στην παραπάνω σχέση, παίρνουμε το παρακάτω αποτέλεσμα

$$P(M_{nk} \leq u_n(x)) = P^k(M_n \leq u_n(x)) + o(1) \rightarrow_{n \rightarrow \infty} G^k(x).$$

Επίσης,

$$P(M_{nk} \leq u_{nk}(x)) \rightarrow_{n \rightarrow \infty} G(x).$$

Έτσι λοιπόν, από την Πρόταση 2.2.2 προκύπτει ότι,

$$a_n/a_{nk} \rightarrow a > 0 \quad \text{και} \quad (b_n - b_{nk})/a_{nk} \rightarrow b \in \mathbb{R}.$$

Επιπλέον,

$$G^k(x) = G(ax+b).$$

Η σ.κ G λοιπόν, είναι *max-stable*. Γνωρίζοντας ότι η κλάση των *max-stable* κατανομών και η κλάση των κατανομών που μπορούν να προκύψουν ως όριο κανονικοποιημένου μεγίστου συμπίπτουν, παίρνουμε το ζητούμενο αποτέλεσμα. ■

Το ερώτημα που τίθεται σε αυτό το σημείο, είναι εάν ισχύει η Πρόταση 1.1.3 στην περίπτωση όπου έχουμε μία ισχυρά στάσιμη ακολουθία τ.μ X_n , $n = 1, 2, \dots$. Με το παρακάτω παράδειγμα βλέπουμε ότι γενικά η απάντηση είναι αρνητική.

Παράδειγμα 2.2.3 Έστω Y_n , $n = 1, 2, \dots$ ακολουθία ανεξάρτητων και ισόνομων τ.μ με σ.κ \sqrt{F} η οποία ικανοποιεί την Πρόταση 1.1.3. Επίσης, έστω ακολουθία ισόνομων τ.μ X_n , $n = 1, 2, \dots$ με σ.κ F τέτοια ώστε $n(1-F(u_n)) \rightarrow g > 0$ όταν $n \rightarrow \infty$ και

$$X_n = \max(Y_n, Y_{n+1}).$$

Είναι φανερό ότι η ακολουθία X_n , $n=1, 2, \dots$ ικανοποιεί την συνθήκη $D(u_n)$ (αν επιλέξουμε $s=2$ τότε $a_{n,s}=0$). Επίσης, για $u_n \rightarrow z_F$ όταν $n \rightarrow \infty$ έχουμε ότι

$$n(1 - \sqrt{F(u_n)}) = \frac{n\bar{F}(u_n)}{1 + \sqrt{F(u_n)}} \rightarrow \frac{g}{2}$$

και

$$P(\tilde{M}_n \leq u_n) \rightarrow e^{-g/2},$$

όπου

$$\tilde{M}_n = \max(Y_1, \dots, Y_n).$$

Συνεπώς,

$$\begin{aligned} P(M_n \leq u_n) &= P(\max(X_1, \dots, X_n) \leq u_n) = P(\max(Y_1, \dots, Y_{n+1}) \leq u_n) \\ &= P(\max(Y_1, \dots, Y_n) \leq u_n) \sqrt{F(u_n)} \rightarrow e^{-g/2}, \quad n \rightarrow \infty. \end{aligned}$$

Έχουμε λοιπόν ότι για μία ισχυρά στάσιμη ακολουθία τ.μ X_n , $n=1,2,\dots$, η απαίτηση $n(1-F(u_n)) \rightarrow g$ δεν εξασφαλίζει την ισχύ της σχέσης $P(M_n \leq u_n) \rightarrow e^{-g}$, όταν $n \rightarrow \infty$. Γενικότερα, αποδεικνύεται ότι για $0 \leq \theta' \leq \theta'' \leq 1$ ισχύει το παρακάτω αποτέλεσμα (*Leadbetter* (1983)),

$$\liminf_{n \rightarrow \infty} P(M_n \leq u_n) = e^{-\theta'g} \quad \text{και} \quad \limsup_{n \rightarrow \infty} P(M_n \leq u_n) = e^{-\theta''g}.$$

Για να υπάρχει λοιπόν το $\lim_{n \rightarrow \infty} P(M_n \leq u_n)$, πρέπει $\theta' = \theta'' = \theta \in [0, 1]$. Σε αυτή την περίπτωση θα ισχύει ότι

$$(2.18) \quad P(M_n \leq u_n) \rightarrow e^{-\theta g}, \quad n \rightarrow \infty.$$

Η παράμετρος θ καλείται δείκτης ακρότατης εξάρτησης (**extremal index**). Ο ακριβής ορισμός του δείκτη ακρότατης εξάρτησης δίνεται παρακάτω.

Ορισμός 2.2.4 Έστω X_n , $n=1,2,\dots$ μία ισχυρά στάσιμη ακολουθία τ.μ με σ.κ F και θ ένας μη αρνητικός πραγματικός αριθμός. Υποθέτουμε ότι για κάθε $g > 0$ υπάρχει ακολουθία (u_n) τέτοια ώστε

$$n(1-F(u_n)) \rightarrow g \quad \text{και} \quad P(M_n \leq u_n) \rightarrow e^{-\theta g}, \quad n \rightarrow \infty.$$

Τότε ο θ ονομάζεται **δείκτης ακρότατης εξάρτησης** της ακολουθίας (X_n) .

Από τον παραπάνω ορισμό, είναι φανερό ότι η ακολουθία (u_n) θα εξαρτάται από το g . Επίσης, αν για τουλάχιστον ένα $g > 0$ υπάρχει ακολουθία (u_n) για την οποία ισχύει ότι $n(1-F(u_n)) \rightarrow g$ όταν $n \rightarrow \infty$, μπορούμε να βρούμε κάποια ακολουθία (y_n) τέτοια ώστε $n(1-F(y_n)) \rightarrow a > 0$ και $a \neq g$. Για παράδειγμα μπορούμε να επιλέξουμε $y_n = u_{[ng/a]}$. Μία πιο αναλυτική περιγραφή του δείκτη ακρότατης εξάρτησης θα δώσουμε στην επόμενη παράγραφο. Στο παραπάνω παράδειγμα βλέπουμε ότι η ακολουθία X_n , $n=1,2,\dots$ έχει δείκτη

ακρότατης εξάρτησης $\theta = 1/2$. Άμεση συνέπεια του παραπάνω ορισμού, είναι η πρόταση που ακολουθεί (Leadbetter (1983)).

Πρόταση 2.2.5 Έστω $X_n, n=1,2,\dots$ μία ισχυρά στάσιμη ακολουθία τ.μ η οποία ικανοποιεί την συνθήκη $D(u_n)$ και επιπλέον $n(1-F(u_n)) \rightarrow g > 0$ όταν $n \rightarrow \infty$. Τότε η ακολουθία (X_n) έχει δείκτη ακρότατης εξάρτησης $\theta \in [0,1]$, αν και μόνο αν

$$(2.19) \quad k \limsup_{n \rightarrow \infty} |P(M_{[n/k]} > u_n) - \theta g/k| \rightarrow 0, \quad k \rightarrow \infty,$$

ή ισοδύναμα

$$(2.20) \quad P(M_{[n/k]} > u_n) \rightarrow \theta g/k + \lambda_k, \quad n \rightarrow \infty,$$

όπου $k\lambda_k \rightarrow 0$ όταν $k \rightarrow \infty$.

Απόδειξη Έστω ότι η ακολουθία (X_n) έχει δείκτη ακρότατης εξάρτησης θ . Τότε από την υπόθεση της πρότασης και την σχέση (2.17) παίρνουμε την σχέση

$$P(M_{[n/k]} \leq u_n) \rightarrow e^{-\theta g/k} = 1 - \theta g/k + o(1/k), \quad n \rightarrow \infty$$

από την οποία προκύπτει η (2.20) ή ισοδύναμα η (2.19). Αντίστροφα, αν ισχύει η σχέση (2.19) έχουμε ότι

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(M_{[n/k]} \leq u_n) &= \limsup_{n \rightarrow \infty} (P(M_{[n/k]} \leq u_n) - 1 + \theta g/k) + 1 - \theta g/k \\ &\leq 1 - \theta g/k + \limsup_{n \rightarrow \infty} |P(M_{[n/k]} > u_n) - \theta g/k| \\ &= 1 - \theta g/k + o(1/k). \end{aligned}$$

Χρησιμοποιώντας την σχέση (2.17), προκύπτει η παρακάτω ανισότητα,

$$\limsup_{n \rightarrow \infty} P(M_n \leq u_n) \leq (1 - \theta g/k + o(1/k))^k.$$

Συνεπώς, αν $k \rightarrow \infty$ παίρνουμε την παρακάτω ανισότητα

$$\limsup_{n \rightarrow \infty} P(M_n \leq u_n) \leq e^{-\theta g}.$$

Όμοια μπορούμε να δείξουμε ότι

$$\liminf_{n \rightarrow \infty} P(M_n \leq u_n) \geq e^{-\theta g}.$$

Από τις παραπάνω ανισότητες προκύπτει τελικά το ζητούμενο, δηλαδή

$$P(M_n \leq u_n) \rightarrow e^{-\theta g}, \quad n \rightarrow \infty.$$

■

Στην περίπτωση όπου υπάρχει ο δείκτης ακρότατης εξάρτησης για μία ισχυρά στάσιμη ακολουθία τ.μ, αποδεικνύεται το παρακάτω θεώρημα (Leadbetter (1983)).

Θεώρημα 2.2.6 Έστω X_n , $n=1,2,\dots$ μία ισχυρά στάσιμη ακολουθία τ.μ. με σ.κ. F η οποία ικανοποιεί την συνθήκη $D(u_n)$ ($u_n = a_n x + b_n$ για κάθε x τέτοιο ώστε $G(x) > 0$) και επιπλέον, έχει δείκτη ακρότατης εξάρτησης $\theta \in [0,1]$. Ορίζουμε την τ.μ $\tilde{M}_n = \max(Y_1, \dots, Y_n)$ όπου Y_i , $i=1,2,\dots$ ακολουθία ανεξάρτητων και ισόνομων τ.μ. με σ.κ. F . Εάν ισχύει ότι

$$P\left\{\frac{\tilde{M}_n - b_n}{a_n} \leq x\right\} \rightarrow G(x), \quad n \rightarrow \infty$$

τότε

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow G^\theta(x), \quad n \rightarrow \infty.$$

Στην περίπτωση όπου $\theta \in (0,1]$ ισχύει και το αντίστροφο.

Από το παραπάνω θεώρημα βλέπουμε ότι εάν $0 < \theta < 1$ οι οριακές κατανομές των κανονικοποιημένων μεγίστων (κάτω από τις ίδιες σταθερές κανονικοποίησης) των ακολουθιών (X_n) και (Y_n) είναι διαφορετικές. Η μεταξύ τους διαφορά δεν έγκειται στον τύπο της οριακής κατανομής, όπως είναι φανερό και από το Θεώρημα 2.2.1, αλλά στις παραμέτρους των δύο κατανομών G και G^θ . Η κατανομή G είναι η GEV με κάποιες παραμέτρους μ, σ, ξ . Έτσι λοιπόν,

$$\begin{aligned} G^\theta(x) &= \exp\left(-\theta\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}\right) \\ &= \exp\left(-\left(\theta^{-\xi} \frac{\sigma + \xi(x-\mu)}{\sigma}\right)^{-1/\xi}\right) \\ &= \exp\left(-\left(1 + \xi \frac{x - (\mu - \frac{\sigma}{\xi}(1-\theta^\xi))}{\sigma\theta^\xi}\right)^{-1/\xi}\right). \end{aligned}$$

Συνεπώς, η κατανομή $G^\theta(x)$ είναι η GEV με παραμέτρους

$$\mu^* = \mu - \frac{\sigma}{\xi}(1 - \theta^\xi) \quad \text{και} \quad \sigma^* = \sigma\theta^\xi.$$

Η περίπτωση $\theta = 0$ δεν συναντάται συχνά στις εφαρμογές και έχει ως συνέπεια το κανονικοποιημένο μέγιστο μίας ί.σ.α. και της αντίστοιχης ακολουθίας ανεξάρτητων και ισόνομων τ.μ. με την ίδια σ.κ. να μην συγκλίνουν «ταυτόχρονα» σε μη εκφυλισμένες κατανομές χρησιμοποιώντας τις ίδιες ακολουθίες κανονικοποίησης. Τέλος, ιδιαίτερο ενδιαφέρον παρουσιάζει η περίπτωση όπου $\theta = 1$, όπου οι κατανομές G και G^θ ταυτίζονται. Σε αυτή την περίπτωση η Πρόταση 1.1.3 ισχύει και για μία ί.σ.α. η οποία ικανοποιεί την συνθήκη $D(u_n)$. Άμεση συνέπεια αυτού είναι ότι μπορούν να εφαρμοστούν τα κριτήρια που αναφέραμε στα Θεωρήματα 1.4.15 - 1.4.17.

Από την ανάλυση που ακολουθεί θα βρούμε μία αναγκαία συνθήκη που πρέπει να ικανοποιεί μία ισχυρά στάσιμη ακολουθία τ.μ έτσι ώστε να έχει δείκτη ακρότατης εξάρτησης $\theta = 1$. Από τις ανισότητες *Bonferroni* έχουμε ότι

$$(2.21) \quad P(M_{[n/k]} \leq u_n) \geq 1 - \sum_{i=1}^{[n/k]} P(X_i > u_n) = 1 - [n/k] \bar{F}(u_n)$$

και ότι

$$(2.22) \quad \begin{aligned} P(M_{[n/k]} \leq u_n) &\leq 1 - \sum_{i=1}^{[n/k]} P(X_i > u_n) + \sum_{1 \leq i < j \leq [n/k]} P(X_i > u_n, X_j > u_n) \\ &= 1 - [n/k] \bar{F}(u_n) + \sum_{1 \leq i < j \leq [n/k]} P(X_i > u_n, X_j > u_n). \end{aligned}$$

Για μία ί.σ.α. είναι εύκολο να επαληθεύσουμε ότι ισχύει η παρακάτω σχέση

$$(2.23) \quad \sum_{1 \leq i < j \leq [n/k]} P(X_i > u_n, X_j > u_n) \leq [n/k] \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n).$$

Συνδυάζοντας τις σχέσεις (2.21), (2.22), (2.23) προκύπτει ότι

$$(2.24) \quad \begin{aligned} 1 - [n/k] \bar{F}(u_n) &\leq P(M_{[n/k]} \leq u_n) \\ &\leq 1 - [n/k] \bar{F}(u_n) + [n/k] \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n). \end{aligned}$$

Αν υποθέσουμε ότι $n(1-F(u_n)) \rightarrow g \geq 0$ όταν $n \rightarrow \infty$ τότε θα ισχύει

$$(2.25) \quad 1 - [n/k] \bar{F}(u_n) \rightarrow_{n \rightarrow \infty} 1 - g/k.$$

Επιπλέον, αν απαιτήσουμε να ισχύει

$$(2.26) \quad \limsup_{n \rightarrow \infty} [n/k] \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) = o(1/k), \quad k \rightarrow \infty$$

από τις σχέσεις (2.17), (2.24)-(2.26) προκύπτει ότι

$$\begin{aligned} (1 - g/k)^k &\leq \liminf_{n \rightarrow \infty} P(M_n \leq u_n) \\ &\leq \limsup_{n \rightarrow \infty} P(M_n \leq u_n) \leq (1 - g/k + o(1/k))^k. \end{aligned}$$

Εάν λοιπόν, $k \rightarrow \infty$ τότε από την παραπάνω σχέση έχουμε ότι

$$P(M_n \leq u_n) \rightarrow e^{-g}, \quad n \rightarrow \infty.$$

Όμοια μπορεί ναδειχθεί ότι αν ικανοποιείται η σχέση (2.26) και $P(M_n \leq u_n) \rightarrow_{n \rightarrow \infty} e^{-g}$, $g \geq 0$ τότε $n(1-F(u_n)) \rightarrow_{n \rightarrow \infty} g$. Η σχέση (2.26) καλείται συνθήκη $D'(u_n)$. Συγκεκριμένα έχουμε τον ακόλουθο ορισμό.

Ορισμός (συνθήκη $D'(u_n)$). Μία ακολουθία τ.μ. X_1, X_2, \dots θα λέμε ότι ικανοποιεί την συνθήκη $D'(u_n)$ αν ισχύει η ακόλουθη απαίτηση

$$\limsup_{n \rightarrow \infty} [n/k] \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) = o(1/k), \quad k \rightarrow \infty$$

ή ισοδύναμα

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} [n/k] \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) = 0.$$

Δείξαμε λοιπόν την παρακάτω πρόταση,

Πρόταση 2.2.7 Έστω X_n , $n=1,2,\dots$ μία ισχυρά στάσιμη ακολουθία τ.μ. η οποία ικανοποιεί τις συνθήκες $D(u_n)$, $D'(u_n)$ και $g \in [0, \infty)$. Τότε θα ισχύει $n(1-F(u_n)) \rightarrow_{n \rightarrow \infty} g$ αν και μόνο αν $P(M_n \leq u_n) \rightarrow_{n \rightarrow \infty} e^{-g}$.

Έχοντας κατά νου την παραπάνω πρόταση, μπορούμε να αναδιατυπώσουμε το Θεώρημα 2.2.6.

Θεώρημα 2.2.8 Έστω X_n , $n=1,2,\dots$ μία ισχυρά στάσιμη ακολουθία τ.μ. με σ.κ. F η οποία ικανοποιεί τις συνθήκες $D(u_n)$, $D'(u_n)$ ($u_n=a_n x+b_n$ για κάθε x τέτοιο ώστε $G(x) > 0$). Ορίζουμε την τ.μ $\tilde{M}_n = \max(Y_1, \dots, Y_n)$ όπου Y_i , $i=1,2,\dots$ ακολουθία ανεξάρτητων και ισόνομων τ.μ. με σ.κ. F . Εάν ισχύει ότι

$$P\left\{\frac{\tilde{M}_n - b_n}{a_n} \leq x\right\} \rightarrow G(x), \quad n \rightarrow \infty$$

τότε και

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \rightarrow G(x), \quad n \rightarrow \infty.$$

Από την παραπάνω πρόταση βλέπουμε ότι μία ι.σ.α. που ικανοποιεί τις συνθήκες $D(u_n)$ και $D'(u_n)$ συμπεριφέρεται (όσον αφορά τα ασυμπτωτικά ακρότατα) σαν μία ακολουθία ανεξάρτητων και ισόνομων τ.μ. Όπως έχουμε ήδη αναφέρει η συνθήκη $D(u_n)$ καθιστά ασυμπτωτικώς ανεξάρτητες τις ακραίες παρατηρήσεις που βρίσκονται επαρκώς μακριά ως προς τον χρόνο, περιορίζοντας έτσι το φαινόμενο τις μακροχρόνιας εξάρτησης. Ωστόσο δεν εμποδίζει την τοπική συγκέντρωση ακραίων παρατηρήσεων, φαινόμενο το οποίο δεν παρατηρείται σε ακολουθία ανεξάρτητων και ισόνομων τ.μ. Αναμένουμε λοιπόν, η συνθήκη $D'(u_n)$ να αποτρέπει την ύπαρξη ενός τέτοιου φαινομένου. Πράγματι, από την σχέση (2.23) και την συνθήκη $D'(u_n)$ προκύπτει άμεσα ότι

$$E\left(\sum_{1 \leq i < j \leq [n/k]} I\{X_i > u_n, X_j > u_n\}\right) \leq [n/k] \sum_{j=2}^{[n/k]} E(I\{X_1 > u_n, X_j > u_n\}) \rightarrow 0, \quad n, k \rightarrow \infty.$$

Συνεπώς, κατά μέσο όρο, εντός της ομάδας μήκους $r_n \sim [n/k_n]$ οι από κοινού υπερβάσεις του κατωφλίου u_n από τις τ.μ X_i , X_j καθίστανται ασυμπτωτικά απίθανες ή, με άλλα λόγια, η πιθανότητα να έχουμε περισσότερες από μία υπερβάσεις εντός της ομάδας τείνει στο μηδέν. Προφανώς η ακολουθία (u_n) θα πρέπει να συγκλίνει «γρήγορα» στο z_F για να εξασφαλίζει την ισχύ της συνθήκης $D'(u_n)$.

Παράδειγμα 2.2.9 Στο Παράδειγμα 2.2.3. είδαμε ότι αν $n(1-F(u_n)) \rightarrow g$ τότε $P(M_n \leq u_n) \rightarrow e^{-g/2}$ όταν $n \rightarrow \infty$ και $g > 0$. Προφανώς η συνθήκη $D'(u_n)$ δεν θα ικανοποιείται στη συγκεκριμένη περίπτωση. Πράγματι, για $j = 2$ έχουμε ότι

$$\begin{aligned} P(X_1 > u_n, X_2 > u_n) &= P(\max(Y_1, Y_2) > u_n, \max(Y_2, Y_3) > u_n) \\ &= P(Y_2 > u_n \text{ ή } (Y_1 > u_n, Y_3 > u_n)), \end{aligned}$$

ενώ για $j \geq 3$ έχουμε

$$P(X_1 > u_n, X_j > u_n) = P^2(X_1 > u_n).$$

Συνεπώς,

$$\begin{aligned} n \sum_{j=2}^{\lfloor n/k \rfloor} P(X_1 > u_n, X_j > u_n) &= nP(Y_2 > u_n \text{ ή } Y_1 > u_n, Y_3 > u_n) + n(\lfloor n/k \rfloor - 2) P^2(X_1 > u_n) \\ &\geq nP(Y_2 > u_n) + n(\lfloor n/k \rfloor - 2) P^2(X_1 > u_n). \end{aligned}$$

Προφανώς, το δεξί μέλος της παραπάνω ανισότητας για $n, k \rightarrow \infty$ συγκλίνει στην ποσότητα $g/2 > 0$ και συνεπώς η συνθήκη $D'(u_n)$ δεν ικανοποιείται.

Παρατήρηση 2.2.10 Κάτω από την υπόθεση ότι μία ισχυρά στάσιμη ακολουθία τ.μ ικανοποιεί τις συνθήκες $D(u_n)$ και $D'(u_n)$ μπορεί να αποδειχθεί ότι οι οριακές κατανομές της r -μεγαλύτερης παρατήρησης και της μικρότερης παρατήρησης ταυτίζονται με τις κατανομές που αναφέραμε στα Θεωρήματα 1.3.1 και 1.6.1 (*Leadbetter* (1974)). Αν παραλείψουμε την συνθήκη $D'(u_n)$ από την υπόθεσή μας, τότε όπως θα δούμε και στην παράγραφο η οριακή κατανομή της r -μεγαλύτερης παρατήρησης ($r \geq 2$) δεν ταυτίζεται με αυτή του Θεωρήματος 1.6.1, όπως είναι φανερό και από το Παράδειγμα 2.2.3 όπου η πρώτη και η δεύτερη μεγαλύτερη παρατήρηση έχουν τις ίδιες οριακές κατανομές.

2.3 Δείκτης ακρότατης εξάρτησης

Στην προηγούμενη παράγραφο αναφερθήκαμε στο δείκτη ακρότατης εξάρτησης, χωρίς να δώσουμε μία ερμηνεία για το τι μπορεί να εκφράζει (στην περίπτωση που υπάρχει) για μία ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών. Στην παρούσα παράγραφο θα δώσουμε

τρεις εναλλακτικές ερμηνείες οι οποίες θα φανούν χρήσιμες στο επόμενο κεφάλαιο όπου θα μελετήσουμε μεθόδους εκτίμησης του δείκτη ακρότατης εξάρτησης από τα δεδομένα που έχουμε στην διάθεσή μας.

Ας υποθέσουμε ότι έχουμε μία ισχυρά στάσιμη ακολουθία (X_n) και για κάθε $g > 0$ υπάρχει ακολουθία (u_n) τέτοια ώστε

$$n\bar{F}(u_n) \rightarrow g, \quad P(M_n \leq u_n) \rightarrow e^{-\theta g}, \quad n \rightarrow \infty.$$

Αν η ακολουθία (X_n) ικανοποιεί την συνθήκη $D(u_n)$ τότε από την σχέση (2.17) έχουμε ότι

$$k_n \ln P(M_{[n/k_n]} \leq u_n) \rightarrow -\theta g, \quad n \rightarrow \infty.$$

Ισχύει ότι $\ln P(M_{[n/k_n]} \leq u_n) \sim -P(M_{[n/k_n]} > u_n)$, όταν $u_n \rightarrow_{n \rightarrow \infty} z_F$. Έτσι λοιπόν, από το παραπάνω όριο προκύπτει ότι

$$k_n P(M_{[n/k_n]} > u_n) \rightarrow \theta g, \quad n \rightarrow \infty,$$

δηλαδή

$$(2.27) \quad \theta = \lim_{n \rightarrow \infty} \frac{k_n P(M_{[n/k_n]} > u_n)}{n\bar{F}(u_n)} = \lim_{n \rightarrow \infty} \frac{P(M_{r_n} > u_n)}{r_n \bar{F}(u_n)}.$$

Όμως,

$$\frac{r_n \bar{F}(u_n)}{P(M_{r_n} > u_n)} = E \left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \mid M_{r_n} > u_n \right).$$

Συνεπώς, για $\theta > 0$ παίρνουμε την σχέση

$$(2.28) \quad \lim_{n \rightarrow \infty} E \left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \mid M_{r_n} > u_n \right) = \frac{1}{\theta},$$

ενώ για $\theta = 0$, θα πρέπει να ισχύει ότι

$$(2.29) \quad \lim_{n \rightarrow \infty} E \left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \mid M_{r_n} > u_n \right) = \infty.$$

Οι σχέσεις (2.28) και (2.29), οριακά εκφράζουν τον αναμενόμενο αριθμό των υπερβάσεων του κατωφλίου u_n εντός των ομάδων μήκους r_n που περιέχουν τουλάχιστον μία υπέρβαση. Οι υπερβάσεις εντός μίας ομάδας θα λέμε ότι σχηματίζουν μία *συστάδα* (*cluster*). Από την σχέση (2.28) βλέπουμε ότι όσο φθίνουν οι τιμές του δείκτη ακρότατης εξάρτησης, ο αναμενόμενος αριθμός των υπερβάσεων αυξάνει. Στην οριακή περίπτωση όπου $\theta = 0$, από την σχέση (2.29) βλέπουμε ότι η ακολουθία (X_n) επιδεικνύει ιδιαίτερα ακραία συμπεριφορά. Έτσι λοιπόν, ο δείκτης ακρότατης εξάρτησης είναι μία παράμετρος η οποία χαρακτηρίζει τον βαθμό εξάρτησης των ενδεχομένων $\{X_i > u_n\}$ μίας ισχυρά στάσιμης ακολουθίας. Η περίπτωση $\theta = 1$ συνεπάγεται ασυμπτωτική ανεξαρτησία των υπερβάσεων ενώ αν $\theta < 1$ οι ακραίες παρατηρήσεις σχηματίζουν συστάδες.

Παράδειγμα 2.3.1 Από το Παράδειγμα 2.2.3 έχουμε ότι

$$\begin{aligned} E\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \mid M_{r_n} > u_n\right) &= \frac{r_n \bar{F}(u_n)}{P(M_{r_n} > u_n)} = \frac{r_n(1-F(u_n))}{(1-F^{r_n/2}(u_n))} \\ &= \frac{-r_n \ln F(u_n)}{-(r_n/2) \ln F(u_n)} + o(1) = 2 + o(1), \end{aligned}$$

δηλαδή $\theta = 1/2$, αποτέλεσμα το οποίο περιμέναμε αφού όπως είδαμε οι ακραίες παρατηρήσεις τείνουν να εμφανίζονται κατά ζεύγη.

Ενδιαφέρον παρουσιάζει μία δεύτερη ερμηνεία του δείκτη ακρότατης εξάρτησης, στην περίπτωση όπου είναι διάφορος του μηδενός, δηλαδή η μέση τιμή του πλήθους των υπερβάσεων εντός μίας συστάδας είναι πεπερασμένος αριθμός, και η οποία δόθηκε από τον *O'Brien* (1987). Για να εξασφαλίσουμε ότι $\theta \in (0,1]$ αρκεί να ισχύει ότι

$$\liminf_{n \rightarrow \infty} \frac{P(M_{r_n} > u_n)}{r_n \bar{F}(u_n)} > 0.$$

Παρατηρούμε ότι για $m \geq 1$ και $M_{j,k} = \max(X_{j+1}, \dots, X_k)$, ($\max(\emptyset) = -\infty$)

$$\begin{aligned} (2.30) \quad P(M_{r_n} > u_n) &\geq \sum_{i=1}^{\lfloor r_n/m \rfloor} P(X_{(i-1)m+1} > u_n, M_{im, r_n} \leq u_n) \\ &\geq \lfloor r_n/m \rfloor \bar{F}(u_n) P(M_{m, r_n} \leq u_n \mid X_1 > u_n). \end{aligned}$$

Από την σχέση (2.30) έχουμε λοιπόν ότι

$$(2.31) \quad \liminf_{n \rightarrow \infty} \frac{P(M_{r_n} > u_n)}{r_n \bar{F}(u_n)} \geq \liminf_{n \rightarrow \infty} \frac{1}{m} P(M_{m, r_n} \leq u_n | X_1 > u_n).$$

Αν απαιτήσουμε λοιπόν, να ισχύει ότι

$$(2.32) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(M_{m, r_n} > u_n | X_1 > u_n) = 0,$$

τότε εξασφαλίζουμε από την σχέση (2.30) ότι

$$\liminf_{n \rightarrow \infty} \frac{P(M_{r_n} > u_n)}{r_n \bar{F}(u_n)} > 0.$$

Παίρνοντας ως δεδομένο ότι ισχύει η σχέση (2.32) έχουμε ότι

$$(2.33) \quad \begin{aligned} P(M_{r_n} > u_n) &= \sum_{i=1}^{r_n} P(X_i > u_n, M_{i, r_n} \leq u_n) \\ &\geq r_n \bar{F}(u_n) P(M_{1, r_n} \leq u_n | X_1 > u_n). \end{aligned}$$

Επιπλέον,

$$(2.34) \quad \begin{aligned} P(M_{r_n} > u_n) &= \sum_{i=1}^{s_n} P(X_i > u_n, M_{0, i-1} \leq u_n) + \sum_{i=s_n+1}^{r_n} P(X_i > u_n, M_{0, i-1} \leq u_n) \\ &\leq s_n \bar{F}(u_n) + (r_n - s_n) \bar{F}(u_n) P(M_{1, s_n} \leq u_n | X_1 > u_n). \end{aligned}$$

και

$$(2.35) \quad \begin{aligned} 0 &\leq P(M_{1, s_n} \leq u_n | X_1 > u_n) - P(M_{1, r_n} \leq u_n | X_1 > u_n) \\ &\leq P(M_{s_n, r_n} > u_n | X_1 > u_n) \rightarrow 0, \quad n \rightarrow \infty, \end{aligned}$$

όπου (s_n) ακολουθία θετικών ακεραίων για την οποία ισχύει ότι $s_n = o(r_n)$ όταν $n \rightarrow \infty$. Από τις σχέσεις (2.33), (2.34) και (2.35), παίρνουμε την σχέση

$$\frac{P(M_{r_n} > u_n)}{r_n \bar{F}(u_n)} = P(M_{1, r_n} \leq u_n | X_1 > u_n) + o(1),$$

δηλαδή,

$$(2.36) \quad \lim_{n \rightarrow \infty} P(M_{1, r_n} \leq u_n | X_1 > u_n) = \theta.$$

Από την σχέση (2.36) βλέπουμε ότι ο δείκτης ακρότατης εξάρτησης εκφράζει την οριακή πιθανότητα έπειτα από μία υπέρβαση κατωφλίου να έχουμε μία σειρά παρατηρήσεων κάτω από το κατώφλι. Μικρές τιμές για αυτή την πιθανότητα, δηλαδή μικρές τιμές του θ , αντικατοπτρίζουν την ακραία συμπεριφορά της διαδικασίας.

Παράδειγμα 2.3.2 Για το μοντέλο του παραδείγματος 2.2.3 έχουμε ότι

$$\begin{aligned}
 P(M_{1,r_n} \leq u_n | X_1 > u_n) &= \frac{P(X_1 > u_n, X_2 \leq u_n, \dots, X_{r_n} \leq u_n)}{P(X_1 > u_n)} \\
 &= \frac{P(Y_2 \leq u_n, Y_3 \leq u_n, \dots, Y_{r_n+1} \leq u_n) - P(Y_1 \leq u_n, Y_3 \leq u_n, \dots, Y_{r_n+1} \leq u_n)}{P(X_1 > u_n)} \\
 &= \frac{F^{r_n/2}(u_n) - F^{(r_n+1)/2}(u_n)}{1 - F(u_n)} = \frac{F^{r_n/2}(u_n)(1 - F^{1/2}(u_n))}{1 - F(u_n)} \\
 &= \frac{F^{r_n/2}(u_n)}{1 + F^{1/2}(u_n)} = \frac{\exp(\frac{1}{2}r_n \ln(1 - \bar{F}(u_n)))}{1 + F^{1/2}(u_n)} \\
 &= \frac{\exp(-\frac{1}{2}r_n \bar{F}(u_n))}{1 + F^{1/2}(u_n)} + o(1) \rightarrow \frac{1}{2}, \quad n \rightarrow \infty.
 \end{aligned}$$

Τέλος, θα δώσουμε μία τρίτη ερμηνεία η οποία βασίζεται στην κατανομή του χρόνου μεταξύ δύο διαδοχικών υπερβάσεων (*Ferro and Segers (2003)*). Ορίζουμε την τ.μ $T(u_n) = \min\{n \geq 1: X_{n+1} > u_n\}$ δεδομένου ότι $X_1 > u_n$ (πιο αυστηρά, αν $T_1 = \min\{n \geq 1: X_n > u_n\}$, ορίζουμε $T(u_n) = \min\{n \geq T_1: X_{n+1} > u_n\} - T_1 + 1$). Η τ.μ. T προφανώς εκφράζει τον χρόνο μεταξύ δύο διαδοχικών υπερβάσεων. Η σ.π. και η σ.κ. της τ.μ T είναι αντίστοιχα

$$P(T(u_n) = n) = P(M_{1,n} \leq u_n, X_{n+1} > u_n | X_1 > u_n)$$

και

$$P(T(u_n) > n) = P(M_{1,n+1} \leq u_n | X_1 > u_n).$$

Στην περίπτωση ανεξάρτητων και ισόνομων τ.μ $X_i, i = 1, 2, \dots$ προκύπτει άμεσα ότι

$$P(T(u_n) > n) = F^n(u_n) = \exp\{n \ln F(u_n)\}.$$

Εάν κανονικοποιήσουμε τον χρόνο T πολλαπλασιάζοντάς τον με $\bar{F}(u_n)$ τότε για $t > 0$ παίρνουμε

$$P(\bar{F}(u_n) T(u_n) > t) = P(T(u_n) > [t / \bar{F}(u_n)]) = \exp\{[t / \bar{F}(u_n)] \ln F(u_n)\} \rightarrow \exp(-t),$$

όταν $n \rightarrow \infty$. Συνεπώς, ο ενδιάμεσος χρόνος, κατάλληλα κανονικοποιημένος, ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda = 1$. Θα δείξουμε ότι παρόμοιο αποτέλεσμα ισχύει και στην περίπτωση όπου έχουμε μία ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών όπου στο αποτέλεσμα θα υπεισέρχεται και ο δείκτης ακρότατης εξάρτησης θ . Υποθέτουμε ότι η ι.σ.α. ικανοποιεί την παρακάτω συνθήκη

$$a_{[cn],s_n} = \max_{1 \leq m \leq [cn]-s_n} \sup |P(B|A) - P(A)| \rightarrow_{n \rightarrow \infty} 0, \quad (P(A) > 0)$$

για κάθε $c > 0$, $s := s_n = o(n)$ ακολουθία θετικών ακεραίων και για όλα τα σύνολα $A \in \mathcal{F}_{1,m}(u_n)$, $B \in \mathcal{F}_{m+s,[cn]}(u_n)$ όπου για $1 \leq j \leq l$ το σύνολο $\mathcal{F}_{j,l}(u_n)$ είναι η ελάχιστη σ-άλγεβρα των ενδεχομένων $\{X_i > u_n\}$, $j \leq i \leq l$. Με την παραπάνω συνθήκη ουσιαστικά απαιτούμε η επιρροή του ενδεχομένου A το οποίο έχει πραγματοποιηθεί, να είναι χρονικά τοπική (δηλ. να «χάνεται» όσο απομακρυνόμαστε χρονικά από αυτό). Έστω λοιπόν $a := a_n = [t / \bar{F}(u_n)]$, τότε για $t > 0$ έχουμε

$$P(\bar{F}(u_n) T(u_n) > t) = P(M_{1,a+1} \leq u_n | X_1 > u_n).$$

Αν θεωρήσουμε την ακολουθία θετικών ακεραίων $r := r_n = o(n)$ τέτοια ώστε $s_n = o(r_n)$, τότε για επαρκώς μεγάλο n θα ισχύει ότι $s_n < r_n$ και επιπλέον

$$P(M_{r,r+s} > u_n | X_1 > u_n) \leq P(M_{r,r+s} > u_n) + a_{n,s} \leq s \bar{F}(u_n) + a_{n,s} \rightarrow 0, \quad n \rightarrow \infty.$$

Συνεπώς,

$$P(M_{1,a+1} \leq u_n | X_1 > u_n) = P(M_{1,r} \leq u_n, M_{r+s,a} \leq u_n | X_1 > u_n) + o(1).$$

Ισχύει ότι $\limsup_{n \rightarrow \infty} (a_n / n) = t / g : c > 0$. Επομένως,

$$\begin{aligned} P(M_{1,r} \leq u_n, M_{r+s,a} \leq u_n | X_1 > u_n) &= P(M_{r+s,a} \leq u_n | M_{1,r} \leq u_n, X_1 > u_n) P(M_{1,r} \leq u_n | X_1 > u_n) \\ &= \{P(M_{r+s,a} \leq u_n) + o(1)\} \{\theta + o(1)\}. \end{aligned}$$

Επίσης, έχουμε ότι $\limsup_{n \rightarrow \infty} (\frac{a_n - (r_n + s_n)}{n}) = t / g := l > 0$. Έτσι λοιπόν παίρνουμε

$$P(\bar{F}(u_n) T(u_n) > t) = \theta P(M_{[nt/g]} \leq u_n) + o(1) = \theta P^{k_n t/g}(M_{r_n} \leq u_n) + o(1) \rightarrow \theta \exp(-\theta t), \quad n \rightarrow \infty.$$

Βλέπουμε λοιπόν ότι ο κανονικοποιημένος ενδιάμεσος χρόνος συγκλίνει κατά κατανομή σε μία τ.μ T_θ η οποία έχει μάζα πιθανότητας $1-\theta$ στο $t = 0$ ενώ για $t > 0$ ακολουθεί την εκθετική κατανομή με παράμετρο θ .

Για να κατανοήσουμε σε ποια περίπτωση μπορεί να έχουμε ενδιάμεσο χρόνο μηδενικό, αρκεί να παρατηρήσουμε ότι κανονικοποιώντας με $\bar{F}(u_n) \sim g/n$, η διαδικασία πραγματοποιείται σε ένα πεπερασμένο διάστημα, συνέπεια του οποίου είναι κάθε ομάδα η οποία περιλαμβάνει μία συστάδα να εκφυλίζεται σε ένα σημείο το οποίο θα έχει πολλαπλότητα ίση με τον αναμενόμενο αριθμό υπερβάσεων ανά συστάδα. Συνεπώς, οι υπερβάσεις οι οποίες βρίσκονται στην ίδια συστάδα θα έχουν μηδενικό ενδιάμεσο χρόνο με αντίστοιχη πιθανότητα $1-\theta$, ενώ οι υπερβάσεις που θα ανήκουν σε διαφορετικές συστάδες θα έχουν ενδιάμεσο χρόνο εκθετικό με πιθανότητα θ . Άμεση συνέπεια αυτού είναι ότι η διαδικασία των σημείων στα οποία εμφανίζονται οι συστάδες μπορεί να προσεγγιστεί από μία διαδικασία *Poisson*. Επειδή κάθε συστάδα θα αποτελείται από έναν αριθμό υπερβάσεων, η διαδικασία των υπερβάσεων ουσιαστικά θα προσεγγίζεται από μια διαδικασία *compound* (σύνθετη) *Poisson*. Θα αναφερθούμε πιο λεπτομερώς σε αυτό στην παράγραφο που ακολουθεί όπου θα αναπτύξουμε την θεωρία ακραίων τιμών για ισχυρά στάσιμες ακολουθίες κάτω από την σκοπιά των σημειακών διαδικασιών.

2.4 Προσέγγιση μέσω Σημειακών στοχαστικών διαδικασιών

Όπως και στην περίπτωση μίας ανεξάρτητης και ισόνομης ακολουθίας τυχαίων μεταβλητών, έτσι και εδώ οι σημειακές διαδικασίες αποτελούν ένα χρήσιμο εργαλείο στην κατανόηση της ακραίας συμπεριφοράς μίας ισχυρά στάσιμης ακολουθίας τυχαίων μεταβλητών. Κύριος σκοπός αυτής της παραγράφου είναι να κατανοήσουμε τη συμπεριφορά των συστάδων που προκύπτουν κατά την εξέλιξη της διαδικασίας X_n , $n=1,2,\dots$. Θα προσπαθήσουμε να απαντήσουμε σε ερωτήματα όπως το πότε είναι δυνατόν να προκύψουν οι συστάδες καθώς επίσης και το ποιο είναι το μέγεθός τους, δηλαδή το πλήθος των υπερβάσεων τις οποίες περιέχουν.

Τα πράγματα είναι εξαιρετικά απλά εάν θεωρήσουμε ότι η ι.σ.α. με συνάρτηση κατανομής F ικανοποιεί τις συνθήκες $D(u_n)$ και $D'(u_n)$. Όπως έχουμε ήδη αναφέρει, και οι δύο αυτές συνθήκες προσδίδουν στην διαδικασία παρόμοια συμπεριφορά με αυτήν της ανεξάρτητης και ισόνομης ακολουθίας τυχαίων μεταβλητών με την ίδια συνάρτηση κατανομής. Εάν θεωρήσουμε λοιπόν, την σημειακή διαδικασία (1.5) του πρώτου κεφαλαίου η οποία καταμετρά το πλήθος των σημείων όπου έχουμε υπέρβαση, αναμένουμε να ισχύει ένα παρόμοιο αποτέλεσμα με αυτό του Θεωρήματος 1.7.10. Πράγματι, από το παρακάτω θεώρημα βλέπουμε ότι για μία ι.σ.α που ικανοποιεί όπως συνθήκες $D(u_n)$ και $D'(u_n)$, η διαδικασία των σημείων στα οποία έχουμε υπερβάσεις προσεγγίζεται από μία διαδικασία *Poisson* (Leadbetter(1974)).

Θεώρημα 2.4.1 Έστω X_n , $n=1,2,\dots$ ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών με συνάρτηση κατανομής F και η οποία ικανοποιεί όπως συνθήκες $D(u_n)$ και $D'(u_n)$ για κάποια ακολουθία (u_n) . Επίσης, υποθέτουμε ότι ισχύει $n\bar{F}(u_n) \rightarrow_{n \rightarrow \infty} g > 0$. Τότε η σημειακή διαδικασία (1.5) συγκλίνει ασθενώς στην ομογενή διαδικασία *Poisson* N στο $E=(0,1]$ με ένταση g .

Απόδειξη Όπως και στην i.i.d περίπτωση θα κάνουμε χρήση του Θεωρήματος 1.7.9. Για να δείξουμε ότι ικανοποιείται η συνθήκη

$$E(N_n(A)) \rightarrow E(N(A)), \quad n \rightarrow \infty$$

ακολουθούμε ακριβώς τα ίδια βήματα όπως και στο Θεώρημα 1.7.10. Μένει λοιπόν να αποδείξουμε ότι ισχύει και η δεύτερη συνθήκη. Έστω λοιπόν, $A = (a,b] \subset (0,1]$. Τότε από την υπόθεση ότι η ακολουθία X_n , $n=1,2,\dots$ είναι ισχυρά στάσιμη, και την Πρόταση 2.2.7, προκύπτει ότι

$$\begin{aligned} P(N_n(A) = 0) &= P(\max_{[na] < i \leq [nb]} X_i \leq u_n) \\ &= P(\max_{1 \leq i \leq [nb] - [na]} X_i \leq u_n) \\ &\rightarrow \exp(-g(b-a)) = P(N(A) = 0), \quad n \rightarrow \infty. \end{aligned}$$

Έστω τώρα $B = \bigcup_{i=1}^k (c_i, d_i]$ όπου $0 < c_1 < d_1 < \dots < c_k < d_k \leq 1$. Εάν θεωρήσουμε τα σύνολα $A_i = \{[nc_i] + 1, \dots, [nd_i]\}$, $i=1, 2, \dots, k$ παρατηρούμε ότι για $l = j+1$, $j=1, \dots, k-1$ η απόσταση

των συνόλων A_i και A_j για επαρκώς μεγάλο n θα είναι $(c_i - d_j)n > s_n = o(n)$, δηλαδή $(k-1)a_{n,s} \rightarrow 0$ όταν $k, n, s \rightarrow \infty$. Έτσι λοιπόν, έχουμε ότι

$$P(N_n(B) = 0) = P(N_n(c_i, d_i] = 0, i = 1, \dots, k) = P\left(\max_{[nc_i] < i \leq [nd_i]} X_i \leq u_n, i = 1, \dots, k\right) \\ = \prod_{i=1}^k P\left(\max_{[nc_i] < i \leq [nd_i]} X_i \leq u_n\right) + o(1) \rightarrow \exp\left(-g \sum_{i=1}^k (d_i - c_i)\right) = P(N(B) = 0), n \rightarrow \infty.$$

■

Η κατάσταση είναι διαφορετική εάν υποθέσουμε ότι ικανοποιείται μόνο η συνθήκη $D(u_n)$. Όπως έχουμε δει, σε αυτή την περίπτωση η διαδικασία (X_n) εμφανίζει συστάδες υπερβάσεων. Η ύπαρξη αυτού του φαινομένου αποτελεί μία ισχυρή ένδειξη ότι το Θεώρημα 2.4.1 δεν μπορεί να ισχύει, διότι η τοπική συγκέντρωση ακραίων τιμών δεν παρουσιάζεται σε μία διαδικασία *Poisson*. Ωστόσο, όταν «συμπιέσουμε» την διαδικασία μέσα στο διάστημα $(0,1]$ το μήκος κάθε ομάδας θα είναι $r_n/n \rightarrow 0$ ($n \rightarrow \infty$) και συνεπώς κάθε ομάδα η οποία αποτελεί συστάδα θα εκφυλίζεται σε ένα σημείο το οποίο θα ενσωματώνει όλες τις υπερβάσεις της συστάδας. Οριακά λοιπόν, το σημείο αυτό θα έχει πολλαπλότητα ίση με τον αριθμό των υπερβάσεων τις οποίες περικλείει. Ορίζουμε λοιπόν την απλή σημειακή διαδικασία

$$(2.37) \quad N_n^*(\cdot) = \sum_{j=1}^{k_n} \varepsilon_{j r_n/n}(\cdot) I\{M(J_j) > u_n\},$$

η οποία «τρέχει» στο σύνολο των ομάδων και αυξάνει κατά μία μονάδα κάθε φορά που κάποια ομάδα περιέχει τουλάχιστον μία υπέρβαση (*point process of cluster positions*). Για την σημειακή διαδικασία (2.37) ισχύει το παρακάτω θεώρημα (*Leadbetter* (1983)).

Θεώρημα 2.4.2 Έστω $X_n, n=1,2,\dots$ ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών η οποία ικανοποιεί την συνθήκη $D(u_n)$ για κάποια ακολουθία (u_n) και έχει δείκτη ακρότατης εξάρτησης $\theta \in (0,1]$. Επίσης, υποθέτουμε ότι ισχύει $n\bar{F}(u_n) \rightarrow_{n \rightarrow \infty} g > 0$. Τότε η σημειακή διαδικασία (2.37) συγκλίνει ασθενώς στην ομογενή διαδικασία *Poisson* N στο $E = (0,1]$ με ένταση θg .

Απόδειξη Προκύπτει εύκολα χρησιμοποιώντας το Θεώρημα 1.7.9. ■

Από το παραπάνω θεώρημα συμπεραίνουμε ότι η ομογενής διαδικασία *Poisson* αποτελεί οριακά ένα κατάλληλο στοχαστικό μοντέλο για την περιγραφή του πλήθους των συστάδων.

Έτσι λοιπόν, κατά μέσο όρο θα εμφανίζονται θg συστάδες στο διάστημα $(0,1]$ ενώ ο χρόνος που μεσολαβεί από την στιγμή εμφάνισης μίας συστάδας μέχρι την επόμενη θα ακολουθεί εκθετική κατανομή με μέση τιμή $1/\theta g$. Σε αυτό το σημείο αξίζει να επισημάνουμε ότι στην τρίτη ερμηνεία που είχαμε δώσει για τον δείκτη ακρότατης κατάστασης είχαμε καταλήξει στο συμπέρασμα ότι οι ενδιάμεσοι χρόνοι ακολουθούν εκθετική κατανομή με μέση τιμή $1/\theta$ και όχι $1/\theta g$. Η διαφορά έγκειται στο γεγονός ότι η κανονικοποίηση είναι διαφορετική. Στην πρώτη περίπτωση είχαμε κανονικοποιήσει με $\bar{F}(u_n) \sim g/n$, ενώ τώρα κανονικοποιούμε με $1/n$.

Όπως έχουμε επισημάνει, κάθε σημείο όπου εμφανίζεται συστάδα έχει πολλαπλότητα ίση με το πλήθος των υπερβάσεων εντός της. Εάν θεωρήσουμε και πάλι την σημειακή διαδικασία (1.5) τότε αναμένουμε να συγκλίνει στη διαδικασία

$$(2.38) \quad \tilde{N}(\cdot) = \sum_{i=1}^{\infty} \xi_i \varepsilon_{\Gamma_i}(\cdot),$$

όπου Γ_i τα σημεία της ομογενούς διαδικασίας *Poisson* N του Θεωρήματος 2.4.2 και ξ_i ακολουθία ανεξάρτητων τυχαίων μεταβλητών, η οποία παίρνει τιμές στο σύνολο των θετικών ακεραίων και έχει συνάρτηση πιθανότητας, έστω π . Βασική προϋπόθεση είναι η συνάρτηση πιθανότητας της τυχαίας μεταβλητής που εκφράζει το πλήθος των υπερβάσεων εντός της συστάδας

$$\pi_n(j) = P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} = j \mid M_{r_n} > u_n\right), \quad j = 1, 2, \dots,$$

να συγκλίνει στην π . Επίσης, παρατηρούμε ότι η σημειακή διαδικασία (2.38) μπορεί να γραφεί και στην μορφή

$$\tilde{N}((a, b]) = \sum_{i=1}^{N((a, b])} \xi_i,$$

για κάποιο διάστημα $(a, b] \subset (0, 1]$. Αν επιπλέον οι τυχαίες μεταβλητές ξ_i είναι ανεξάρτητες της N , τότε η διαδικασία (2.38) δεν είναι τίποτα άλλο παρά μία σύνθετη διαδικασία *Poisson* (*Compound Poisson process*). Η διαισθητική αυτή παρατήρηση επαληθεύεται από το

επόμενο θεώρημα (Hsing et al.(1988)), για την απόδειξη του οποίου απαιτείται μία πιο ισχυρή συνθήκη από την $D(u_n)$. Η συνθήκη δίνεται παρακάτω.

Ορισμός (Συνθήκη $\Delta(u_n)$) Θα θεωρούμε ότι η ακολουθία X_1, X_2, \dots ικανοποιεί την συνθήκη $\Delta(u_n)$ αν ικανοποιεί την ακόλουθη απαίτηση: Έστω ακολουθία θετικών ακεραίων $s := s_n = o(n)$. Για κάθε $A \in \mathcal{F}_{1,l}(u_n)$ και $B \in \mathcal{F}_{l+s,n}(u_n)$, $1 \leq l \leq n-s$, απαιτούμε να ισχύει

$$\max_{1 \leq l \leq n-s} \sup |P(A \cap B) - P(A)P(B)| = a_{n,s_n} \rightarrow 0, \quad n \rightarrow \infty$$

όπου για $1 \leq j \leq l$ το σύνολο $\mathcal{F}_{j,l}(u_n)$, είναι η ελάχιστη σ -άλγεβρα των ενδεχομένων $\{X_i \leq u_n\}$, $j \leq i \leq l$.

Θεώρημα 2.4.3 Έστω $X_n, n=1,2,\dots$ ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών με δείκτη ακρότατης εξάρτησης $\theta \in (0,1]$. Υποθέτουμε ότι υπάρχει ακολουθία (u_n) τέτοια ώστε να ικανοποιείται η συνθήκη $\Delta(u_n)$ και επιπλέον $n\bar{F}(u_n) \rightarrow_{n \rightarrow \infty} g > 0$. Επίσης, υποθέτουμε ότι υπάρχουν ακολουθίες θετικών ακεραίων $(r_n), (s_n)$ και συνάρτηση πιθανότητας π , τέτοιες ώστε $r_n = o(n)$, $s_n = o(r_n)$, $na_{n,s} = o(r_n)$ και $\pi_n(j) \rightarrow \pi(j)$ όταν $n \rightarrow \infty$ για κάθε $j \geq 1$. Τότε η σημειακή διαδικασία (1.5) συγκλίνει ασθενώς στην N στο $E = (0,1]$ όπου $N \sim CP(\theta g, \pi)$.

Σχέδιο απόδειξης Αποδεικνύεται (Daley and Vere-Jones (1988)) ότι η σύγκλιση (ασθενής) της διαδικασίας (1.5) στην $CP(\theta g, \pi)$ ισοδυναμεί με την σύγκλιση του μετασχηματισμού Laplace του τυχαίου διανύσματος $(N_n(A_1), \dots, N_n(A_m))$ στον μετασχηματισμό Laplace του διανύσματος $(N(A_1), \dots, N(A_m))$, $m \geq 1$, $A_i \cap A_j = \emptyset$, για $i \neq j$, $i, j = 1, \dots, m$, και $A_1, \dots, A_m \subset (0,1]$. Ο μετασχηματισμός Laplace των δύο τυχαίων διανυσμάτων είναι αντίστοιχα

$$L_n(t_1, \dots, t_m) = E(\exp\{-\sum_{i=1}^m t_i N_n(A_i)\}) \quad \text{και} \quad L(t_1, \dots, t_m) = \prod_{i=1}^m \exp(-\theta g |A_i| \{1 - \sum_{j \geq 1} \pi(j) e^{-jt_i}\}),$$

όπου $0 \leq t_i < \infty$ ($i=1, \dots, m$). Αρκεί λοιπόν να δείξουμε ότι

$$L_n(t_1, \dots, t_m) \rightarrow L(t_1, \dots, t_m), \quad n \rightarrow \infty,$$

για κάθε $0 \leq t_i < \infty$ ($i = 1, 2, \dots, m$). Θεωρούμε τη διαμέριση $A = \bigcup_{i=1}^{m_n+1} J_i$ ενός συνόλου $A \subseteq (0,1]$, όπου $J_i, i = 1, \dots, m_n$ διαδοχικά σύνολα μήκους r_n/n και $|J_{m_n+1}| < r_n/n$. Προφανώς, θέλουμε να ισχύει $m_n r_n/n \sim |A|$. Από την συνθήκη $\Delta(u_n)$ ακολουθώντας παρόμοιους συλλογισμούς όπως στην Παράγραφο 2.2 μπορεί να δειχθεί ότι ισχύει

$$E(\exp\{-tN_n(A)\}) = (E(\exp\{-tN_n(J_1)\}))^{(n/r_n)|A|} + o(1), \quad n \rightarrow \infty.$$

Παρόμοια προκύπτει ότι

$$L_n(t_1, \dots, t_m) = \prod_{i=1}^m (E(\exp\{-t_i N_n(J_1)\}))^{(n/r_n)|A_i|} + o(1), \quad n \rightarrow \infty.$$

Όμως

$$\begin{aligned} E(\exp\{-tN_n(J_1)\}) &= 1 - P(M_{r_n} > u_n) + \sum_{j \geq 1} \pi_n(j) P(M_{r_n} > u_n) e^{-jt} \\ &= 1 - P(M_{r_n} > u_n) \left(1 - \sum_{j \geq 1} \pi_n(j) e^{-jt} \right). \end{aligned}$$

Γνωρίζουμε ότι $P(M_{r_n} > u_n) \sim (r_n/n)\theta g$ και επιπλέον από την υπόθεση έχουμε ότι $\pi_n(j) \rightarrow \pi(j)$ όταν $n \rightarrow \infty$ για κάθε $j \geq 1$. Συνεπώς,

$$\begin{aligned} L_n(t_1, \dots, t_m) &= \prod_{i=1}^m (E(\exp\{-t_i N_n(J_1)\}))^{(n/r_n)|A_i|} + o(1) \\ &= \prod_{i=1}^m \left(1 - \frac{r_n}{n} \theta g (1 - \sum_{j \geq 1} \pi(j) e^{-jt_i}) \right)^{\frac{n}{r_n}|A_i|} + o(1) \\ &\rightarrow \prod_{i=1}^m \exp(-\theta g |A_i| \{1 - \sum_{j \geq 1} \pi(j) e^{-jt_i}\}), \quad n \rightarrow \infty. \end{aligned}$$

■

Από το παραπάνω θεώρημα διαπιστώνουμε ότι τα μεγέθη των συστάδων είναι οριακά ανεξάρτητες τυχαίες μεταβλητές, οι οποίες επιπλέον είναι ανεξάρτητες και από τις στιγμές εμφανίσεων των συστάδων. Από την υπόθεση ότι $n\bar{F}(u_n) \rightarrow_{n \rightarrow \infty} g > 0$, συμπεραίνουμε ότι κατά μέσο όρο (οριακά) θα έχουμε g υπερβάσεις ενώ από το αποτέλεσμα του θεωρήματος προκύπτει ότι η μέση τιμή του πλήθους των συστάδων θα είναι θg . Συνεπώς, ο

αναμενόμενος αριθμός των υπερβάσεων που θα περιέχει κάθε συστάδα θα είναι $g/\theta g=1/\theta$. Το αποτέλεσμα αυτό συμπίπτει με την σχέση (2.28). Έχουμε λοιπόν ότι

$$(2.39) \quad \lim_{n \rightarrow \infty} \sum_{j \geq 1} j \pi_n(j) = 1/\theta.$$

Παρατήρηση 2.4.4 Η σχέση (2.39) δεν συνεπάγεται ότι $\sum_{j \geq 1} j \pi(j) = 1/\theta$, καθώς από το λήμμα *Fatou* έχουμε ότι

$$\sum_{j \geq 1} j \lim_{n \rightarrow \infty} \pi_n(j) \leq \lim_{n \rightarrow \infty} \sum_{j \geq 1} j \pi_n(j) = 1/\theta,$$

δηλαδή γενικά, $\sum_{j \geq 1} j \pi(j) \leq 1/\theta$.

Κάνοντας χρήση του θεωρήματος 2.4.3 μπορούμε να υπολογίσουμε την οριακή κατανομή της r -μεγαλύτερης παρατήρησης. Εάν θεωρήσουμε την ακολουθία ανεξάρτητων και ισόνομων τ.μ. $K_i, i=1,2,\dots$ με συνάρτηση πιθανότητας π , έχουμε για $r=1, \dots, n$,

$$\begin{aligned} P(X_{r,n} \leq u_n) &= P(N_n(0,1] \leq r-1) \\ &= P(N_n(0,1] = 0) + \sum_{j=1}^{r-1} P(N_n(0,1] = j) \\ &\rightarrow P(N^*(0,1] = 0) + \sum_{j=1}^{r-1} P\left(\sum_{l=1}^{N^*(0,1]} K_l = j\right) \\ &= P(N^*(0,1] = 0) + \sum_{j=1}^{r-1} \sum_{k=1}^j P\left(\sum_{l=1}^{N^*(0,1]} K_l = j \mid N^*(0,1] = k\right) P(N^*(0,1] = k) \\ &= P(N^*(0,1] = 0) + \sum_{j=1}^{r-1} \sum_{k=1}^j P\left(\sum_{l=1}^k K_l = j\right) P(N^*(0,1] = k) \\ &= P(N^*(0,1] = 0) + \sum_{k=1}^{r-1} \sum_{j=k}^{r-1} P(N^*(0,1] = k) P\left(\sum_{l=1}^k K_l = j\right) \\ &= e^{-\theta g} + \sum_{k=1}^{r-1} \sum_{j=k}^{r-1} e^{-\theta g} \frac{(\theta g)^k}{k!} P\left(\sum_{l=1}^k K_l = j\right) \\ &= e^{-\theta g} \left(1 + \sum_{k=1}^{r-1} \sum_{j=k}^{r-1} \frac{(\theta g)^k}{k!} P\left(\sum_{l=1}^k K_l = j\right) \right), \quad n \rightarrow \infty. \end{aligned}$$

2.5 Στατιστικές συναρτήσεις συστάδων (cluster statistics)

Πολλές ποσότητες που παρουσιάζουν ενδιαφέρον στην θεωρία ακραίων τιμών μπορούν να μελετηθούν μέσω ειδικών συναρτήσεων τις οποίες ονομάζουμε «συναρτησοειδή συστάδων» (*cluster functionals*). Μιλώντας σε γενικές γραμμές, τα συναρτησοειδή συστάδων είναι συναρτήσεις οι οποίες έχουν ορίσματα συναρτήσεις (συναρτησοειδή). Στην περίπτωση που τα ορίσματα είναι συναρτήσεις κάποιας τυχαίας μεταβλητής, ονομάζονται «Στατιστικές συναρτήσεις συστάδων» (*cluster statistics*) και αποτελούν γενίκευση των «στατιστικών συναρτήσεων σάρωσης» (*scan statistics*, π.χ. βλ. *Koutras and Balakrishnan* (2002) ή *Glaz* (2001)). Για την ανάπτυξη της θεωρίας μας θα ασχοληθούμε με μία συγκεκριμένη κατηγορία των στατιστικών συναρτήσεων συστάδων την οποία θα αναφέρουμε παρακάτω. Μερικά παραδείγματα ποσοτήτων που μπορούν να μελετηθούν είναι τα εξής:

- Η ποσότητα $\zeta_n = \sum_{i=1}^n I(X_i > u_n)$ η οποία εκφράζει το πλήθος των υπερβάσεων σε μία περίοδο n παρατηρήσεων.
- Η ποσότητα $\zeta_n = \sum_{i=1}^n \max(X_i - u_n, 0) = \sum_{i=1}^n (X_i - u_n)_+$ η οποία εκφράζει την συνολική υπέρβαση του u_n από τις τ.μ X_1, \dots, X_n .
- Η ποσότητα $\zeta_n = \sum_{i=1}^{n-1} I((X_i - u_n)_+, (X_{i+1} - u_n)_+)$, για δείκτρια I τέτοια ώστε $I(x_i, x_{i+1}) = 1$ αν $x_i = 0, x_{i+1} > 0$ και μηδέν διαφορετικά. Η ποσότητα αυτή εκφράζει το πλήθος των «ανοδικών περασμάτων» (*upcrossings*), δηλαδή το πλήθος των σημείων στα οποία η διαδικασία ξεπερνά το κατώφλι u_n μετά από μια παρατήρηση κάτω από το u_n .
- Η ποσότητα $\zeta_n = \sum_{i=1}^{n-m+1} I((X_i - u_n)_+, \dots, (X_{i+m-1} - u_n)_+)$, για δείκτρια I τέτοια ώστε $I(x_i, \dots, x_{i+m-1}) = 1$ αν $x_j > 0, j = i, \dots, i + m - 1$ ($m \leq n$) και μηδέν διαφορετικά. Η ποσότητα αυτή εκφράζει το πλήθος των ροών υπερβάσεων μήκους m (*run statistic*), δηλαδή των περιπτώσεων όπου η διαδικασία παρουσιάζει m διαδοχικές υπερβάσεις στο σύνολο των n παρατηρήσεων.
- Η ποσότητα $\zeta_n = \sum_{i=1}^{n-2} I((X_i - u_n)_+, (X_{i+1} - u_n)_+, (X_{i+2} - u_n)_+)$, για δείκτρια I τέτοια ώστε $I(x_i, x_{i+1}, x_{i+2}) = 1$ αν $x_{i+1} > 0$ και $x_{i+1} > \max(x_i, x_{i+2})$, η οποία εκφράζει το πλήθος των υπερβάσεων που είναι τοπικά μέγιστα, στις n παρατηρήσεις.

Σημαντικά αποτελέσματα που αφορούν τις πιθανοθεωρητικές ιδιότητες των παραπάνω συναρτήσεων συστάδων, αλλά και πολλών άλλων, μπορούν να προκύψουν από την μελέτη

γενικότερων συναρτησοειδών ζ_n . Συγκεκριμένα, θεωρούμε την οικογένεια των συναρτήσεων $\{\zeta_d: R^d \rightarrow R, d = 1, 2, \dots\}$ τέτοια ώστε για κάθε $1 \leq i \leq j \leq k \leq l$,

$$\zeta_{l-i+1}(x_i, \dots, x_l) = \zeta_{k-j+1}(x_j, \dots, x_k)$$

αν $x_i, \dots, x_{j-1} \leq 0$ και $x_{k+1}, \dots, x_l \leq 0$. Εάν $\mathbf{x} = (x_1, x_2, \dots) \in R^\infty$ τότε ορίζουμε

$$\zeta(i, j, \mathbf{x}) = \zeta_{j-i+1}(x_i, \dots, x_j), 1 \leq i \leq j.$$

Παράδειγμα 2.5.1 Για $n \geq 1$ αν $\zeta_n = \sum_{i=1}^n I(X_i > u_n)$, $X_1, \dots, X_{j-1} \leq u_n$ και $X_{k+1}, \dots, X_n \leq u_n$, $1 \leq j \leq k \leq n$, τότε

$$\zeta_n = \sum_{i=1}^n I(X_i > u_n) = \sum_{i=j}^k I(X_i > u_n) = \zeta_{k-j+1}.$$

Παράδειγμα 2.5.2 Για $n \geq 1$ αν $\zeta_n = \sum_{i=1}^{n-1} I((X_i - u_n)_+, (X_{i+1} - u_n)_+)$, $X_1, \dots, X_{j-1} \leq u_n$ και $X_{k+1}, \dots, X_n \leq u_n$, $1 \leq j \leq k \leq n$, όπου I τέτοια ώστε $I(x_i, x_{i+1}) = 1$ αν $x_i < 0, x_{i+1} > 0$ και μηδέν διαφορετικά, τότε

$$\zeta_n = \sum_{i=1}^{n-1} I((X_i - u_n)_+, (X_{i+1} - u_n)_+) = \sum_{i=j-1}^{k-1} I((X_i - u_n)_+, (X_{i+1} - u_n)_+) = \zeta_{k-j+1}.$$

Εύκολα μπορούμε να επαληθεύσουμε ότι και τα υπόλοιπα παραδείγματα που αναφέραμε στην αρχή ανήκουν στην παραπάνω οικογένεια συναρτήσεων.

Στην παρούσα παράγραφο θα μελετήσουμε συναρτήσεις συστάδων της μορφής

$$\zeta_{r_n}(X_1 - u_n, \dots, X_{r_n} - u_n)$$

τις οποίες για λόγους απλότητας θα συμβολίζουμε με $\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\}$. Όπως είναι φανερό και από την ονομασία τους, οι συναρτήσεις συστάδων μας παρέχουν χρήσιμες πληροφορίες για την συμπεριφορά των συστάδων, για το σχηματισμό των οποίων απαιτείται να υπάρχουν υπερβάσεις των κατωφλίων u_n από την διαδικασία. Θα ερευνήσουμε λοιπόν, την ασυμπτωτική συμπεριφορά του $\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\}$ δεδομένου του ενδεχομένου $\{M_{r_n} > u_n\}$. Το παρακάτω θεώρημα μας δίνει μία οριακή προσέγγιση της κατανομής της τ.μ $\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\}$ δεδομένου ότι $M_{r_n} > u_n$.

Θεώρημα 2.5.3 Έστω $X_n, n=1,2,\dots$ ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών. Επίσης, έστω ότι οι ακολουθίες (u_n) και (r_n) ικανοποιούν την σχέση (2.32). Ορίζουμε την ακολουθία

$$a_n(A_n) = P(\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\} \in A_n | X_1 > u_n) - P(\zeta_n \{(X_i - u_n)_{i=2}^{r_n}\} \in A_n, M_{1,r_n} > u_n | X_1 > u_n),$$

όπου A_n Borel υποσύνολο του R . Τότε ισχύει

$$\limsup_{n \rightarrow \infty} |P(\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\} \in A_n | M_{r_n} > u_n) - \theta_n^{-1} a_n(A_n)| = 0,$$

όπου $\theta_n = P(M_{r_n} > u_n) / r_n \bar{F}(u_n)$ ή $\theta_n = P(M_{1,r_n} \leq u_n | X_1 > u_n)$.

Απόδειξη (βλ. Segers (2003)). ■

Το παραπάνω θεώρημα μας διευκολύνει σε μεγάλο βαθμό, καθώς η δέσμευση ως προς το ενδεχόμενο $\{M_{r_n} > u_n\}$ καθιστά δύσκολη, στη γενική περίπτωση, την εύρεση της δεσμευμένης κατανομής της $\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\} | M_{r_n} > u_n$.

Παράδειγμα 2.5.4 Στην προηγούμενη παράγραφο είχαμε δει ότι η σ.π. του μεγέθους μίας συστάδας είναι η

$$\pi_n(j) = P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} = j | M_{r_n} > u_n\right), \quad j = 1, 2, \dots$$

Αν θεωρήσουμε την στατιστική συνάρτηση συστάδας $\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\} = \sum_{i=1}^{r_n} I(X_i > u_n)$ και $A_n = [j, \infty)$, ($j \geq 1$) τότε από το Θεώρημα 2.5.3 έχουμε ότι

$$\begin{aligned} & P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq j | M_{r_n} > u_n\right) \\ &= \theta_n^{-1} \left(P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq j | X_1 > u_n\right) - P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} \geq j, M_{1,r_n} > u_n | X_1 > u_n\right) \right) + o(1) \\ &= \theta_n^{-1} \left(P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} \geq j-1 | X_1 > u_n\right) - P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} \geq j | X_1 > u_n\right) \right) + o(1) \\ &= \theta_n^{-1} P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} = j-1 | X_1 > u_n\right) + o(1). \end{aligned}$$

Συνεπώς,

$$\pi_n(j) = \theta_n^{-1} \left(P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} = j-1 \mid X_1 > u_n\right) - P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} = j \mid X_1 > u_n\right) \right) + o(1).$$

Παράδειγμα 2.5.5 Το αποτέλεσμα του παραδείγματος 2.5.4 μπορεί να χρησιμοποιηθεί για τον υπολογισμό της οριακής σ.π. των συστάδων που εμφανίζονται στο μοντέλο του παραδείγματος 2.2.3. Για $j = 1$ έχουμε

$$P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 1 \mid M_{r_n} > u_n\right) = P(M_{r_n} > u_n \mid M_{r_n} > u_n) = 1.$$

Αν θεωρήσουμε τα ξένα ενδεχόμενα $A_i = \{X_1 > u_n, X_2 \leq u_n, \dots, X_{i-1} \leq u_n, X_i > u_n, X_{i+1} \leq u_n, \dots, X_r \leq u_n\}$, $i = 2, 3, \dots, r$, τότε για $i = 2$ και $i = r$, έχουμε αντίστοιχα ότι $P(A_2) = F^{\frac{r_n-1}{2}} - F^{\frac{r_n}{2}}$ και $P(A_r) = F^{\frac{r_n-1}{2}} - 2F^{\frac{r_n}{2}} + F^{\frac{r_n+1}{2}}$, ενώ για $i = 3, 4, \dots, r-1$, ισχύει ότι $A_i = \emptyset$. Συνεπώς, για $j = 2$ έχουμε ότι

$$\begin{aligned} P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} = 1 \mid X_1 > u_n\right) &= \frac{P\left(\bigcup_{i=2}^{r_n} A_i\right)}{P(X_1 > u_n)} = \frac{P(A_2 \cup A_r)}{P(X_1 > u_n)} \\ &= \frac{2F^{\frac{r_n-1}{2}}(u_n) - 3F^{\frac{r_n}{2}}(u_n) + F^{\frac{r_n+1}{2}}(u_n)}{1 - F(u_n)} \\ &= \frac{F^{\frac{r_n-1}{2}}(u_n)(2 - 3F^{\frac{1}{2}}(u_n) + F(u_n))}{1 - F(u_n)} \\ &= \frac{F^{\frac{r_n-1}{2}}(u_n)(2 - F^{\frac{1}{2}}(u_n))}{1 + F^{\frac{1}{2}}(u_n)} \rightarrow \frac{1}{2}, \quad n \rightarrow \infty. \end{aligned}$$

Επίσης, $\theta_n \rightarrow 1/2$, όταν $n \rightarrow \infty$. Συνεπώς,

$$P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 2 \mid M_{r_n} > u_n\right) \rightarrow 1, \quad n \rightarrow \infty$$

και

$$\begin{aligned} \pi_n(1) &= P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} = 1 \mid M_{r_n} > u_n\right) \\ &= P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 1 \mid M_{r_n} > u_n\right) - P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 2 \mid M_{r_n} > u_n\right) \\ &\rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Όμοια, αν θεωρήσουμε τα ξένα ενδεχόμενα $B_i = \{X_1 > u_n, X_2 \leq u_n, \dots, X_i > u_n, X_{i+1} > u_n, \dots, X_r \leq u_n\}$, $i = 2, 3, \dots, r-1$, τότε για $i = 2$ και $r-1$, έχουμε αντίστοιχα ότι $P(A_2) = F^{\frac{r_n-2}{2}} - F^{\frac{r_n-1}{2}} - F^{\frac{r_n}{2}} + F^{\frac{r_n+1}{2}}$ και $P(A_{r-1}) = F^{\frac{r_n-2}{2}} - 2F^{\frac{r_n-1}{2}} + F^{\frac{r_n}{2}}$, ενώ για $i = 3, \dots, r-2$, έχουμε ότι $P(A_i) = F^{\frac{r_n-1}{2}} - 2F^{\frac{r_n}{2}} + F^{\frac{r_n+1}{2}}$. Συνεπώς για $j=3$ έχουμε ότι

$$\begin{aligned} P\left(\sum_{i=2}^{r_n} I\{X_i > u_n\} = 2 \mid X_1 > u_n\right) &= \frac{P\left(\bigcup_{i=2}^{r_n-1} B_i\right)}{P(X_1 > u_n)} \\ &= \frac{2F^{\frac{r_n-2}{2}}(u_n) + (r_n-7)F^{\frac{r_n-1}{2}}(u_n) - (2r_n-8)F^{\frac{r_n}{2}}(u_n) + (r_n-3)F^{\frac{r_n+1}{2}}(u_n)}{1-F(u_n)} \\ &= \frac{F^{\frac{r_n-2}{2}}(u_n)(F^{\frac{1}{2}}(u_n)-1)(-2+(5-r_n)F^{\frac{1}{2}}(u_n)) + (r_n-3)F(u_n)}{1-F(u_n)} \\ &= \frac{F^{\frac{r_n-2}{2}}(u_n)(3F(u_n) + r_n F^{\frac{1}{2}}(u_n)(1-F^{\frac{1}{2}}(u_n)) - 5F^{\frac{1}{2}}(u_n) + 2)}{1+F^{\frac{1}{2}}(u_n)} \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Συνεπώς,

$$P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 3 \mid M_{r_n} > u_n\right) \rightarrow 0, \quad n \rightarrow \infty$$

και

$$\begin{aligned} \pi_n(2) &= P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} = 2 \mid M_{r_n} > u_n\right) \\ &= P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 2 \mid M_{r_n} > u_n\right) - P\left(\sum_{i=1}^{r_n} I\{X_i > u_n\} \geq 3 \mid M_{r_n} > u_n\right) \\ &\rightarrow 1, \quad n \rightarrow \infty. \end{aligned}$$

Παράδειγμα 2.5.6 Ας θεωρήσουμε την στατιστική συνάρτηση συστάδας $\zeta_n \{(X_i - u_n)_{i=1}^{r_n}\} = \sum_{i=1}^{r_n} I\left(\frac{X_i - u_n}{a_n} > x\right)$ όπου (a_n) είναι ακολουθία για την οποία ισχύει ότι $a_n = g(u_n)$ για κάποια μετρήσιμη συνάρτηση $g > 0$. Επίσης, έστω $A_n = [1, \infty)$. Τότε για $a_n x > 0$ από το Θεώρημα 2.5.3 έχουμε

$$\begin{aligned}
& P\left(\sum_{i=1}^{r_n} I\{X_i - u_n > a_n x\} \geq 1 \mid M_{r_n} > u_n\right) \\
&= P(M_{r_n} - u_n > a_n x \mid M_{r_n} > u_n) \\
&= \theta_n^{-1} \left(P(M_{r_n} - u_n > a_n x \mid X_1 > u_n) - P(M_{1,r_n} - u_n > a_n x, M_{1,r_n} > u_n \mid X_1 > u_n) \right) + o(1) \\
&= \theta_n^{-1} \left(P(\{X_1 - u_n > a_n x\} \cup \{M_{1,r_n} - u_n > a_n x\} \mid X_1 > u_n) - P(M_{1,r_n} - u_n > a_n x \mid X_1 > u_n) \right) + o(1) \\
&= \theta_n^{-1} P(X_1 - u_n > a_n x, M_{1,r_n} - u_n \leq a_n x \mid X_1 > u_n) + o(1) \\
&= \theta_n^{-1} P(M_{1,r_n} - u_n \leq a_n x \mid X_1 - u_n > a_n x) P(X_1 - u_n > a_n x \mid X_1 > u_n) + o(1).
\end{aligned}$$

Εάν

$$\lim_{n \rightarrow \infty} P(M_{1,r_n} - u_n \leq a_n x \mid X_1 > u_n) = \lim_{n \rightarrow \infty} P(M_{1,r_n} \leq u_n \mid X_1 > u_n) = \lim_{n \rightarrow \infty} \theta_n = \theta > 0,$$

τότε

$$\begin{aligned}
& P\left(\sum_{i=1}^{r_n} I\{X_i - u_n > a_n x\} \geq 1 \mid M_{r_n} > u_n\right) \\
&= P(X_1 - u_n > a_n x \mid X_1 > u_n) + o(1) \\
&= \frac{1 - F(u_n + a_n x)}{1 - F(u_n)} + o(1) \rightarrow \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi}, \quad n \rightarrow \infty,
\end{aligned}$$

δηλαδή

$$\begin{aligned}
& P\left(\sum_{i=1}^{r_n} I\{X_i - u_n > a_n x\} = 0 \mid M_{r_n} > u_n\right) = P(M_{r_n} \leq a_n x + u_n \mid M_{r_n} > u_n) \\
&\rightarrow 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi}, \quad n \rightarrow \infty.
\end{aligned}$$

Συνεπώς, η οριακή κατανομή του μεγέθους της μέγιστης υπέρβασης (*cluster maximum-peak excess*) είναι η *Generalized Pareto* κατανομή. Η ακολουθία (a_n) σταθεροποιείται στην παράμετρο $\tilde{\sigma}$ όταν $n \rightarrow \infty$ (*Pickands (1975)*). Για μία πιο άμεση απόδειξη του παραπάνω αποτελέσματος παραπέμπουμε στον *Leadbetter (1991)*. Έχοντας ως βάση το παραπάνω αποτέλεσμα, ορίζουμε τη σημειακή διαδικασία

$$(2.40) \quad N'_n(\cdot) = \sum_{j=1}^{k_n} \frac{M(J_j) - u_n}{a_n} \varepsilon_{j r_n / n}(\cdot) I\{M(J_j) > u_n\},$$

όπου (a_n) είναι ακολουθία για την οποία ισχύει ότι $a_n = g(u_n)$ για κάποια μετρήσιμη συνάρτηση $g > 0$. Η παραπάνω διαδικασία προφανώς εντοπίζει κάθε συστάδα και αθροίζει σταδιακά τα κανονικοποιημένα μεγέθη των υπερβάσεων των μεγίστων παρατηρήσεων. Εάν επιπλέον για $1 \leq j \leq l$ το σύνολο $\mathcal{F}_{j,l}(u_n)$, είναι η ελάχιστη σ -άλγεβρα των ενδεχομένων $\{(X_i - u_n)_+\}$, $j \leq i \leq l$, τότε για την σημειακή διαδικασία (2.40) ισχύει το παρακάτω (διαισθητικά προφανές) αποτέλεσμα (*Leadbetter* (1991)).

Θεώρημα 2.5.7 Έστω X_n , $n=1,2,\dots$ ισχυρά στάσιμη ακολουθία τυχαιών μεταβλητών με δείκτη ακρότατης εξάρτησης $\theta \in (0,1]$. Υποθέτουμε ότι υπάρχει ακολουθία (u_n) τέτοια ώστε

$$\max_{1 \leq m \leq n-s_n} \sup |P(A \cap B) - P(A)P(B)| = a_{n,s_n} \xrightarrow{n \rightarrow \infty} 0,$$

$s_n := s = o(n)$ ακολουθία θετικών ακεραίων και το παραπάνω supremum λαμβάνεται ως προς όλα τα $A \in \mathcal{F}_{1,m}(u_n)$, $B \in \mathcal{F}_{m+s,n}(u_n)$. Επίσης, υποθέτουμε ότι $n\bar{F}(u_n) \xrightarrow{n \rightarrow \infty} g > 0$. Αν υπάρχει ακολουθία θετικών ακεραίων (r_n) τέτοια ώστε $r_n = o(n)$, $s_n = o(r_n)$ και $na_{n,s} = o(r_n)$ όταν $n \rightarrow \infty$, τότε η σημειακή διαδικασία (2.40) συγκλίνει ασθενώς στην N στο $E = (0,1]$ όπου $N \sim CP(\theta g, \pi)$ και π η *Generalized Pareto* κατανομή.

Για μία πιο αναλυτική περιγραφή των ασυμπτωτικών ιδιοτήτων των στατιστικών συναρτήσεων συστάδων (*cluster statistics*) παραπέμπουμε στις εργασίες των *Leadbetter* (1991), *Leadbetter* (1995), *Rootzén et al.* (1998), *Smith et al.* (1997), *Segers* (2003).

ΚΕΦΑΛΑΙΟ 3

Εκτίμηση του δείκτη ακρότατης εξάρτησης

3.1 Εισαγωγή

Αντικείμενο αυτού του κεφαλαίου είναι η εκτίμηση του δείκτη ακρότατης εξάρτησης θ από ένα τυχαίο δείγμα X_1, \dots, X_n το οποίο προέρχεται από μία ισχυρά στάσιμη στοχαστική διαδικασία. Στην Παράγραφο 2.3 είδαμε ότι κάτω από συγκεκριμένες προϋποθέσεις ο δείκτης ακρότατης εξάρτησης μπορεί να προσεγγιστεί με τρεις διαφορετικούς τρόπους. Η πρώτη προσέγγιση προκύπτει άμεσα από τον ορισμό του θ και έχει ως αποτέλεσμα ότι ο θ οριακά εκφράζει το αναμενόμενο πλήθος των υπερβάσεων εντός μιας συστάδας. Στην περίπτωση όπου $\theta > 0$, ο δείκτης ακρότατης εξάρτησης μπορεί να ερμηνευθεί και ως η οριακή πιθανότητα του ενδεχομένου μία υπέρβαση κατώφλιου να ακολουθηθεί από μία σειρά παρατηρήσεων κάτω από το κατώφλι. Τέλος, ο θ μπορεί να ερμηνευθεί ως το (οριακό) ποσοστό των μη μηδενικών ενδιάμεσων χρόνων μεταξύ των διαδοχικών υπερβάσεων καθώς επίσης και ως ο αντίστροφος της μέσης τιμής του ενδιάμεσου χρόνου.

Οι παραπάνω ερμηνείες αποτελούν τα εργαλεία πάνω στα οποία θα βασιστούμε για να κατασκευάσουμε εκτιμήτριες του δείκτη ακρότατης εξάρτησης. Συγκεκριμένα, θα αναπτύξουμε τρεις διαφορετικές μεθόδους οι οποίες θα βασίζονται στις εναλλακτικές ερμηνείες που διατυπώσαμε παραπάνω. Οι μέθοδοι είναι οι εξής:

1. Μέθοδος των ομάδων (*blocks method*).
2. Μέθοδος των ροών παρατηρήσεων κάτω από ένα συγκεκριμένο κατώφλι u (*runs below the threshold method* ή απλά *runs method*).
3. Μέθοδος των ενδιάμεσων χρόνων μεταξύ υπερβάσεων (*interexceedance times method*).

3.2 Μέθοδος των ομάδων

Η μέθοδος των ομάδων βασίζεται στην σχέση (2.17). Από αυτή βλέπουμε ότι για αρκετά μεγάλο n μπορούμε να χωρίσουμε το δείγμα X_1, \dots, X_n σε k ομάδες μεγέθους r ($n \sim kr$) των οποίων οι μέγιστες παρατηρήσεις μπορούν να θεωρηθούν ανεξάρτητες. Επίσης, από τον ορισμό του δείκτη ακρότατης εξάρτησης (Ορισμός 2.2.4) προκύπτει άμεσα ότι

$$\theta = \lim_{n \rightarrow \infty} \left(-\frac{\ln P(M_n \leq u_n)}{n(1 - F(u_n))} \right).$$

Για ακολουθία θετικών ακεραίων $r_n = o(n)$ όταν $n \rightarrow \infty$ προφανώς θα ισχύει

$$\theta_{r_n}(u_{r_n}) := -\frac{\ln P(M_{r_n} \leq u_{r_n})}{r_n(1 - F(u_{r_n}))} \rightarrow \theta, \quad n \rightarrow \infty.$$

Για λόγους απλότητας θα συμβολίζουμε με $F_{r_n}(u_n) := P(M_{r_n} \leq u_n)$. Επίσης, θα υποθέτουμε από εδώ και στο εξής ότι $n = kr$. Για σταθερό κατώφλι u εύλογο είναι να θεωρήσουμε ως εκτιμήτρια της σ.κ. F την δειγματική σ.κ. $\hat{F}(u) := n^{-1} \sum_{i=1}^n I(X_i \leq u)$. Επίσης, ως εκτιμήτρια της F_r μπορούμε να θεωρήσουμε την

$$\hat{F}_r(u) := k^{-1} \sum_{i=1}^k I(M_{(i-1)r, ir} \leq u)$$

όπου $M_{(i-1)r, ir} := \max(X_{(i-1)r+1}, \dots, X_{ir})$. Η \hat{F} προφανώς εκφράζει το ποσοστό των παρατηρήσεων στο δείγμα μεγέθους n οι οποίες δεν υπερβαίνουν το κατώφλι u ενώ η εκτιμήτρια \hat{F}_r εκφράζει το ποσοστό των ομάδων οι οποίες δεν περιέχουν καμία υπέρβαση. Έτσι λοιπόν, από τα παραπάνω έχουμε ότι η

$$(3.1) \quad \hat{\theta}_{bl} = \hat{\theta}_{bl}(n, r, u) := -\frac{\ln \hat{F}_r(u)}{r(1 - \hat{F}(u))}$$

είναι μία εκτιμήτρια του θ . Αν επιπλέον συμβολίζουμε με $Z := Z(n, r, u)$ το πλήθος των ομάδων που περιέχουν τουλάχιστον μία παρατήρηση η οποία υπερβαίνει το u και με $N := N(n, u)$ το πλήθος των υπερβάσεων τότε η εκτιμήτρια (3.1) μπορεί να γραφεί ως εξής:

$$\hat{\theta}_{bl} = -\frac{\ln(1-Z/k)}{r(1-(1-N/n))} = -\frac{\ln(1-Z/k)}{rN/n} \sim \frac{nZ}{krN} = \frac{Z}{N},$$

αν $Z \ll k$ (το Z είναι «μικρό» συγκρινόμενο με το k). Αν συμβολίσουμε με $\hat{\theta}_{bl}^{(1)} = \hat{\theta}_{bl}^{(1)}(n, r, u) := Z/N$, τότε από την παραπάνω σχέση βλέπουμε ότι οι εκτιμήτριες $\hat{\theta}_{bl}$ και $\hat{\theta}_{bl}^{(1)}$ για αρκετά μεγάλο n θα είναι περίπου ίσες. Μπορούμε να πούμε λοιπόν ότι ασυμπτωτικά η $\hat{\theta}_{bl}$ θα ισούται με το πλήθος των συστάδων διαιρεμένο με το πλήθος των υπερβάσεων.

Σε αυτό το σημείο σημαντικό είναι να εξετάσουμε τις ιδιότητες της εκτιμήτριας $\hat{\theta}_{bl}$, όπως η συνέπεια, η μεροληψία καθώς και να βρούμε τη διασπορά της. Τα αποτελέσματα που θα παρουσιάσουμε παρακάτω βασίζονται στην πρόσφατη εργασία των *Robert, Segers and Ferro* (2009). Παρόμοια αποτελέσματα μπορούν να βρεθούν και στις παλαιότερες εργασίες των *Hsing* (1991) και *Weissman and Novak* (1998) όμως από άποψη στατιστικής ανάλυσης πιο εύχρηστα αποτελέσματα παρέχονται στην εργασία των *Robert, Segers and Ferro* (2009). Έστω λοιπόν οι ακολουθίες θετικών ακεραίων $r_n = o(n)$, $s_n = o(r_n)$ και $a_n = o(s_n)$ όταν $n \rightarrow \infty$. Με $L_2(\mathcal{F})$ θα συμβολίζουμε τον χώρο των \mathcal{F} -μετρήσιμων και τετραγωνικά ολοκληρώσιμων (κατά *Lebesgue*) τυχαίων μεταβλητών. Επίσης, ορίζουμε

$$\rho_{n,s}(u) := \max_{k=1, \dots, n-s} \sup |corr(X, Y)|$$

όπου το παραπάνω supremum λαμβάνεται ως προς όλες τις τ.μ. $X \in L_2(\mathcal{F}_{1,k}(u))$, $Y \in L_2(\mathcal{F}_{k+s,n}(u))$, $1 \leq k \leq n-s$, το σύνολο $\mathcal{F}_{j,l}(u)$ είναι η ελάχιστη σ -άλγεβρα των ενδεχομένων $\{X_i \leq u\}$, $1 \leq j \leq l$, $j \leq i \leq l$ και $corr(X, Y)$ ο συντελεστής συσχέτισης των τ.μ. X, Y . Τέλος, με $\pi(j)$ θα συμβολίζουμε, όπως και στο δεύτερο κεφάλαιο, την οριακή σ.π. του πλήθους των υπερβάσεων εντός μιας συστάδας. Έχοντας κατά νου τα παραπάνω, αποδεικνύεται το επόμενο θεώρημα.

Θεώρημα 3.2.1 Έστω X_n , $n = 1, 2, \dots$ ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών με δείκτη ακρότατης εξάρτησης $\theta \in (0, 1]$. Επίσης, έστω (u_r) ακολουθία κατωφλίων για την οποία ισχύει ότι $r(1-F(u_r)) \rightarrow g \in (0, \infty)$ όταν $r \rightarrow \infty$. Εάν

$$\rho_{n,a_n}(u_{r_n}) = o(s_n/n), \quad \sum_{s=s_n}^n \rho_{n,s}(u_{r_n}) = o(r_n)$$

και επιπλέον

$$P\left(\sum_{i=1}^{s_n} I(X_i > u_{r_n}) = j \mid M_{s_n} > u_{r_n}\right) \rightarrow \pi(j), \quad n \rightarrow \infty$$

για κάθε $j \in N$, όπως επίσης και

$$E\left(\left(\sum_{i=1}^{s_n} I(X_i > u_{r_n})\right)^2 \mid M_{s_n} > u_{r_n}\right) \rightarrow \sum_{j=1}^{\infty} j^2 \pi(j) < \infty, \quad n \rightarrow \infty,$$

τότε $\hat{\theta}_{bl}(n, r_n, u_{r_n}) \rightarrow_p \theta$ (κατά πιθανότητα).

Συνεπώς, κάτω από τις προϋποθέσεις του Θεωρήματος 3.2.1 η $\hat{\theta}_{bl}$ είναι συνεπής εκτιμήτρια του θ (*weak consistent*). Όσον αφορά τη μεροληψία $E(\hat{\theta}_{bl}) - \theta$ της εκτιμήτριας $\hat{\theta}_{bl}$, έχουμε ότι αυτή μπορεί να γραφεί ως $(E(\hat{\theta}_{bl}) - \theta_r) + (\theta_r - \theta)$. Εάν $\theta_r(u_r) - \theta = o(1/k_n)$ αποδεικνύεται ότι

$$k_n(E(\hat{\theta}_{bl}(n, r_n, u_{r_n})) - \theta) \rightarrow \mu := \frac{\theta}{2\alpha}(e^\alpha - 1) + \frac{\theta}{\alpha}c^2, \quad n \rightarrow \infty,$$

όπου $\alpha = \theta g$, $c = \sqrt{\text{Var}(X)} / E(X)$ (συντελεστής απόκλισης) και X τ.μ με σ.π.π $\pi(j)$, $j=1,2,\dots$.

Εάν συμβολίσουμε με $m_p = \sum_{j=1}^{\infty} j^p \pi(j)$, $p=1,2,\dots$ τότε η ποσότητα c^2 μπορεί να γραφεί στη μορφή $c^2 = (m_2 - m_1^2) / m_1^2 = m_2 / m_1^2 - 1$. Δεδομένου ότι $\rho_{n,s} \geq a_{n,s}$ (με $a_{n,s}$ της συνθήκης $\Delta(u_n)$) από το Θεώρημα 2.4.3 έχουμε ότι

$$N_{r_n}(u_{r_n}) := \sum_{i=1}^{r_n} I(X_i > u_{r_n}) \rightarrow_d N, \quad n \rightarrow \infty,$$

όπου $N \sim CP(\theta g, \pi)$. Συνεπώς, $\sigma_r^2(u_r) := \text{Var}(N_r(u_r)) \rightarrow \text{Var}(N) = \theta g m_2$, όταν $n \rightarrow \infty$. Άρα λοιπόν,

$$c^2 = \frac{\text{Var}(N)}{\theta g \theta^{-2}} - 1 = \frac{\theta \text{Var}(N)}{g} - 1.$$

Αν συμβολίσουμε

$$N_{(i-1)r, ir}(u) := \sum_{j=(i-1)r+1}^{ir} I(X_j > u), \quad i=1,\dots,k \quad \text{και} \quad \bar{N}(u) := k^{-1} \sum_{i=1}^k N_{(i-1)r, ir}(u)$$

τότε η διασπορά της τ.μ N μπορεί να εκτιμηθεί από την

$$\hat{\sigma}_r^2(u) := k^{-1} \sum_{i=1}^k (N_{(i-1)r, ir}(u) - \bar{N}(u))^2.$$

Επίσης, η παράμετρος g για δεδομένο r μπορεί να εκτιμηθεί από την $\hat{g} := r(1 - \hat{F}(u))$. Έτσι λοιπόν, από τα παραπάνω έχουμε ότι η

$$\hat{c}^2(u) = \frac{\hat{\theta}_{bl}}{\hat{g}} \hat{\sigma}_r^2(u) - 1$$

είναι μία εκτιμήτρια της ποσότητας c^2 . Πρακτικά λοιπόν, εάν \hat{b} είναι η εκτιμήτρια της μεροληψίας της $\hat{\theta}_{bl}$ τότε για μεγάλο n θα έχουμε ότι

$$\hat{b} \approx \frac{\hat{\mu}}{k} = \frac{(2\hat{g})^{-1}(e^{\hat{\theta}_{bl}\hat{g}} - 1) + \hat{g}^{-1}\hat{c}^2}{k}.$$

Τέλος, αν ισχύουν οι προϋποθέσεις του Θεωρήματος 3.2.1 και επιπλέον

$$\sum_{s=1}^n \rho_{n,s}(u_{r_n}) = o(\sqrt{r_n}), \quad \theta_{r_n}(u_{r_n}) - \theta = o(1/\sqrt{k_n}), \quad \limsup_{n \rightarrow \infty} E(N_{r_n}^{2p}) < \infty$$

για κάποιο $p > 1$, αποδεικνύεται ότι

$$\sqrt{k_n}(\hat{\theta}_{bl}(n, r_n, u_{r_n}) - \theta) \rightarrow_d N(0, \sigma^2),$$

όπου

$$\sigma^2 = \frac{\theta^2}{\alpha^2}(e^\alpha - 1 - \alpha) + \frac{\theta^2}{\alpha} c^2.$$

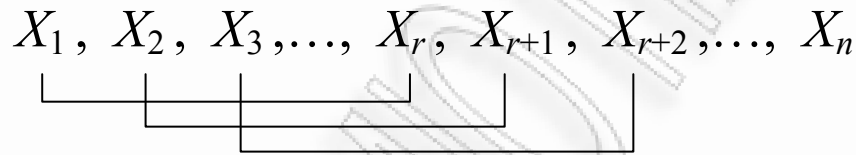
Το διάστημα εμπιστοσύνης συντελεστού εμπιστοσύνης $1-a$ για το θ θα είναι σε αυτή τη περίπτωση το

$$\left(\hat{\theta}_{bl} - \frac{\hat{\mu}}{k} - \frac{\hat{\sigma} z_{a/2}}{\sqrt{k}}, \hat{\theta}_{bl} - \frac{\hat{\mu}}{k} + \frac{\hat{\sigma} z_{a/2}}{\sqrt{k}} \right).$$

Μέχρι στιγμής δεν έχουμε αναφερθεί στην επιλογή της παραμέτρου r . Προφανώς, μπορούμε να εκτιμήσουμε το θ για κάποιο εύρος τιμών του r , να συγκρίνουμε τις διαφορετικές εκτιμήσεις και να πάρουμε μία πρώτη ιδέα για το διάστημα των τιμών που

κυμαίνεται το θ . Γενικά, η επιλογή του r υπόκεινται στα εκάστοτε δεδομένα που έχουμε στην διάθεσή μας. Ωστόσο, παρατηρούμε ότι η διασπορά σ^2 είναι κυρτή συνάρτηση της παραμέτρου α και συνεπώς έχει ολικό ελάχιστο. Αν λοιπόν, για δεδομένο u συμβολίσουμε με α_{min} το α για το οποίο $\hat{\sigma}^2(\alpha_{min}) = \min \hat{\sigma}^2(\alpha)$, τότε το βέλτιστο r (που ελαχιστοποιεί την διασπορά της $\hat{\theta}_{bl}$) μπορεί να βρεθεί από την σχέση $r_{opt} = g_{opt}/(1 - \hat{F}(u))$ όπου $g_{opt} = \alpha_{min}/\hat{\theta}_{bl}$. Όσο αφορά την επιλογή του κατωφλίου u συνήθως επιλέγουμε κάποιο ποσοστιαίο σημείο του δείγματος.

Μία εναλλακτική επιλογή για την εκτίμηση του θ είναι να θεωρήσουμε επικαλυπτόμενες ομάδες (*sliding blocks*) μεγέθους r . Όπως φαίνεται και από το παρακάτω σχήμα



οι ομάδες σε αυτή τη περίπτωση θα είναι «ισχυρά» εξαρτημένες. Ως συνέπεια αυτού, η υπόθεση της ασυμπτωτικής ανεξαρτησίας των μέγιστων παρατηρήσεων των ομάδων από την οποία προέκυψαν τα παραπάνω αποτελέσματα για την εκτιμήτρια $\hat{\theta}_{bl}$, δεν μπορεί να εφαρμοστεί σε αυτή την περίπτωση. Ωστόσο η επιλογή επικαλυπτόμενων ομάδων αντί ανεξάρτητων, όπως θα δούμε και παρακάτω, μας οδηγεί σε μία εκτιμήτρια του θ , την οποία θα συμβολίζουμε με $\hat{\theta}_{sl}$, η οποία έχει καλύτερες ασυμπτωτικές ιδιότητες από την $\hat{\theta}_{bl}$, όπως μικρότερη μεροληψία και διασπορά. Σε αυτή την περίπτωση μπορούμε να εκτιμήσουμε την $F_{r_n}(u_n) := P(M_{r_n} \leq u_n)$ από την

$$\hat{F}_r^{sl}(u) := (n - r + 1)^{-1} \sum_{i=0}^{n-r} I(M_{i,i+r} \leq u).$$

Η εκτιμήτρια του θ που θα χρησιμοποιήσουμε είναι η

$$\hat{\theta}_{sl} = \hat{\theta}_{sl}(n, r, u) := -\frac{\ln \hat{F}_r^{sl}(u)}{r(1 - \hat{F}(u))}.$$

Οι προϋποθέσεις κάτω από τις οποίες προέκυψαν τα παραπάνω ασυμπτωτικά αποτελέσματα για την εκτιμήτρια $\hat{\theta}_{bl}$, παραμένουν και σε αυτή τη περίπτωση ίδιες. Έτσι λοιπόν, αποδεικνύ

εται (Robert, Segers and Ferro (2009)) ότι η $\hat{\theta}_{sl}$ είναι συνεπής εκτιμήτρια του θ ($\hat{\theta}_{sl} \rightarrow_p \theta$) και

$$k_n(E(\hat{\theta}_{sl}(n, r_n, u_{r_n})) - \theta) \rightarrow \mu_{sl} := \frac{\theta}{\alpha^2} (e^\alpha - 1 - \alpha) + \frac{\theta}{\alpha} c^2, \quad n \rightarrow \infty.$$

Προφανώς, $0 \leq \mu_{sl} \leq \mu$, δηλαδή η εκτιμήτρια $\hat{\theta}_{sl}$ έχει ασυμπτωτικά μικρότερη μεροληψία από την $\hat{\theta}_{bl}$. Επίσης, αποδεικνύεται ότι

$$\sqrt{k_n}(\hat{\theta}_{sl}(n, r_n, u_{r_n}) - \theta) \rightarrow_d N(0, \sigma_{sl}^2),$$

όπου

$$\sigma_{sl}^2 = 2 \frac{\theta^2}{\alpha^3} (e^\alpha - 1 - \alpha - \frac{\alpha^2}{2}) + \frac{\theta^2}{\alpha} c^2.$$

Και σε αυτή την περίπτωση έχουμε ότι $\sigma_{sl}^2 \leq \sigma^2$. Τέλος, αν χρησιμοποιήσουμε τους συμβολισμούς

$$N_{i,i+r}(u) := \sum_{j=i+1}^{i+r} I(X_j > u), \quad i = 0, \dots, n-r, \quad \bar{N}_{sl}(u) := (n-r+1)^{-1} \sum_{i=0}^{n-r} N_{i,i+r}(u)$$

και

$$\hat{\sigma}_{sl,r}^2(u) := (n-r+1)^{-1} \sum_{i=0}^{n-r} (N_{i,i+r}(u) - \bar{N}_{sl}(u))^2$$

μία συνεπής εκτιμήτρια της ποσότητας της ποσότητας c^2 είναι η

$$\hat{c}_{sl}^2(u) = \frac{\hat{\theta}_{sl}}{\hat{g}} \hat{\sigma}_{sl,r}^2(u) - 1.$$

Η διαδικασία υπολογισμού του βέλτιστου r είναι ακριβώς η ίδια με αυτή που ακολουθήσαμε στην περίπτωση των ανεξάρτητων ομάδων.

3.3 Μέθοδος των ροών παρατηρήσεων κάτω από ένα συγκεκριμένο κατώφλι u

Η μέθοδος των διαδοχικών παρατηρήσεων βασίζεται στην σχέση (2.36). Όπως και στη περίπτωση των ομάδων, χωρίς βλάβη της γενικότητας και για λόγους απλότητας των συμβολισμών, θα υποθέσουμε ότι $n = kr$. Από την σχέση (2.36) μπορούμε άμεσα να πάρουμε την παρακάτω εκτιμήτρια για το θ ,

$$\hat{\theta}_R = \hat{\theta}_R(n, r, u) := \frac{(n-r+1)^{-1} \sum_{i=1}^{n-r+1} I(X_i > u, M_{i,i+r-1} \leq u)}{n^{-1} \sum_{i=1}^n I(X_i > u)},$$

όπου $M_{i,i+r-1} = \max\{X_{i+1}, \dots, X_{i+r-1}\}$. Η παράμετρος r , η οποία εκφράζει το μέγεθος των $n-r+1$ επικαλυπτόμενων ομάδων που πρέπει να χρησιμοποιήσουμε σε αυτή τη μέθοδο, όπως και στην περίπτωση της μεθόδου των ομάδων, θα εξαρτάται από τα δεδομένα που έχουμε στη διάθεσή μας. Οι υπερβάσεις οι οποίες διαχωρίζονται από λιγότερες από $r-1$ «μη-υπερβάσεις» θα θεωρούμε ότι ανήκουν στην ίδια συστάδα. Προφανώς, αν τα δεδομένα μας παρουσιάζουν «ισχυρή» εξάρτηση στα ακραία επίπεδα (πάνω από ένα αρκετά μεγάλο κατώφλι), το r θα πρέπει να είναι μεγάλο εν αντιθέσει με τη περίπτωση «ασθενούς» εξάρτησης όπου μικρές τιμές του r μπορεί να είναι ικανοποιητικές. Σε κάθε περίπτωση πάντως, είναι λογικό να υποθέσουμε ότι $n \gg r$. Έτσι λοιπόν, $n-r+1 \approx n$ και

$$\hat{\theta}_R = \hat{\theta}_R(n, r, u) := \frac{\sum_{i=1}^n I(X_i > u, M_{i,i+r-1} \leq u)}{\sum_{i=1}^n I(X_i > u)}.$$

Παρατήρηση 3.3.1 Στη βιβλιογραφία όπως και σε αρκετές εργασίες η εκτιμήτρια $\hat{\theta}_R$ δίνεται στη μορφή

$$\hat{\theta}_R := \frac{\sum_{i=1}^{n-r} I(X_i > u, M_{i,i+r} \leq u)}{\sum_{i=1}^n I(X_i > u)}.$$

Σε αυτή την περίπτωση, θα θεωρούμε ότι οι υπερβάσεις οι οποίες διαχωρίζονται από λιγότερες από r «μη-υπερβάσεις» ανήκουν στην ίδια συστάδα. Για μεγάλο δείγμα δεν θα υπάρχει ουσιαστική διαφορά ανάμεσα στις δύο εκτιμήτριες.

Για την εκτιμήτρια $\hat{\theta}_R$ αποδεικνύεται το παρακάτω θεώρημα (*Weissman and Novak* (1998)).

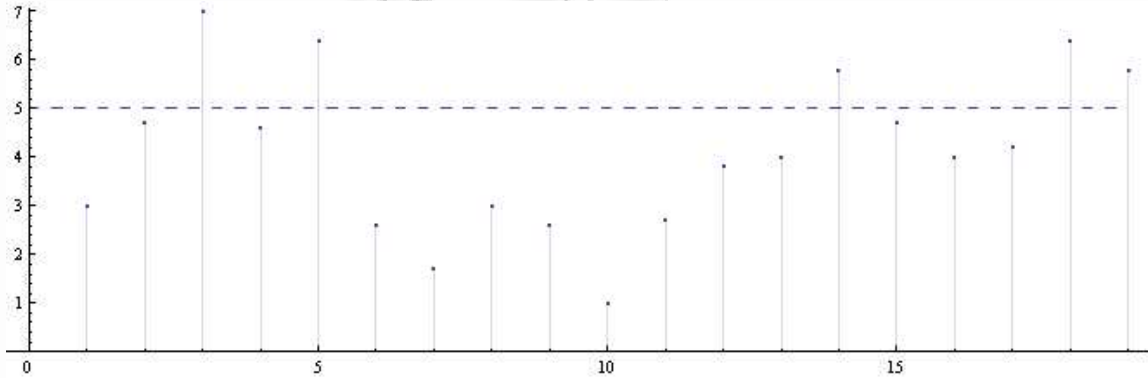
Θεώρημα 3.3.2 Εάν η σχέση (2.36) ισχύει και επιπλέον

$$j_n := \sum_{i=1}^n (1 - \frac{i}{n}) (P(X_{i+1} > u_n, M_{i+1, i+r_n} \leq u_n | X_1 > u_n, M_{1, r_n} \leq u_n) - P(X_1 > u_n, M_{1, r_n} \leq u_n)) \\ = o(n\bar{F}(u_n))$$

τότε για κάθε $r \geq 1$ έχουμε ότι η $\hat{\theta}_R$ είναι συνεπής εκτιμήτρια του θ δηλαδή $\hat{\theta}_R \rightarrow_p \theta$.

Πρέπει σε αυτό το σημείο να αναφέρουμε ότι το παραπάνω αποτέλεσμα ισχύει εάν η ακολουθία κατωφλίων (u_n) είναι τέτοια ώστε $n(1-F(u_n)) \rightarrow \infty$, δηλαδή πηγαίνει «αργά» στο άπειρο και η ακολουθία (r_n) ικανοποιεί τη σχέση $r_n(1-F(u_n)) \rightarrow 0$.

Είναι φανερό ότι ο αριθμητής της εκτιμήτριας $\hat{\theta}_R$ εκφράζει το πλήθος των συστάδων στο δείγμα ενώ ο παρονομαστής το πλήθος των υπερβάσεων. Στην πραγματικότητα, όπως φαίνεται και από το παρακάτω σχήμα όπου $n=19$ και $u=5$, για $r=3$ παίρνουμε τρεις συστάδες ενώ $\sum_{i=1}^{17} I(X_i > u, M_{i, i+r-1} \leq u) = 2$. Αν δηλαδή τουλάχιστον μία από τις παρατηρήσεις $X_{n+1}, \dots, X_{n+r-1}$ είναι μεγαλύτερη από το κατώφλι u , το πλήθος των συστάδων θα είναι $\sum_{i=1}^n I(X_i > u, M_{i, i+r-1} \leq u) + 1$. Για μεγάλο πλήθος υπερβάσεων όμως δεν θα υπάρχει ουσιαστική διαφορά.



Αν τώρα συμβολίσουμε με $Z := \sum_{i=1}^n I(X_i > u, M_{i, i+r-1} \leq u)$ και $N := \sum_{i=1}^n I(X_i > u)$ τότε θα έχουμε ότι

$$\hat{\theta}_R = \frac{Z}{N}.$$

Παρόμοιο αποτέλεσμα είχαμε δει και στη μέθοδο των ομάδων με τη διαφορά ότι ο αριθμητής Z δεν είναι αναγκαστικά ίδιος και στις δύο περιπτώσεις. Για να βρούμε τη μέση τιμή της $\hat{\theta}_R$ θα χρησιμοποιήσουμε το ανάπτυγμα *Taylor* δεύτερης τάξης για τη συνάρτηση $f(Z, N) = Z/N$ στο σημείο $(E(Z), E(N))$. Έπειτα από πράξεις καταλήγουμε στο παρακάτω αποτέλεσμα,

$$(3.2) \quad E(\hat{\theta}_R) = E(Z/N) \approx \frac{E(Z)}{E(N)} \left(1 + \frac{Var(N)}{E^2(N)} - \frac{Cov(Z, N)}{E(N)E(Z)} \right).$$

Προφανώς, για $n \rightarrow \infty$ έχουμε ότι $E(N) \rightarrow g$, $E(Z) \rightarrow \theta g$, $Var(Z) \rightarrow \theta g$, $E(N|Z) \rightarrow Z/\theta$ και $Var(N|Z) \rightarrow Z\sigma^2$, όπου σ^2 είναι η διασπορά του μεγέθους των συστάδων. Συνεπώς, $Var(N) \rightarrow (g/\theta) + \theta g\sigma^2$ και $Cov(Z, N) \rightarrow g$ όταν $n \rightarrow \infty$. Αν αντικαταστήσουμε τα παραπάνω αποτελέσματα στη σχέση (3.2) προκύπτει ότι

$$E(\hat{\theta}_R) = E(Z/N) \approx \theta(1 + \theta\sigma^2/g).$$

Συνεπώς, για αρκετά μεγάλο n θα έχουμε ότι

$$E(\hat{\theta}_R) - \theta \approx \theta^2\sigma^2/g.$$

Τέλος, για την ισχυρά στάσιμη ακολουθία τυχαίων μεταβλητών X_n , $n=1, 2, \dots$, θα χρησιμοποιήσουμε τον παρακάτω συμβολισμό,

$$\phi(k, u) := \sup |P(B|A) - P(B)|, \quad (P(A) > 0)$$

όπου το supremum λαμβάνεται ως προς όλα τα $A \in \mathcal{F}_{1,m}(u)$, $B \in \mathcal{F}_{m+k, \infty}(u)$ και $\mathcal{F}_{j,l}(u)$ η ελάχιστη σ -άλγεβρα των ενδεχομένων $\{X_i > u\}$, $1 \leq j \leq l$, $j \leq i \leq l$. Επιπλέον, συμβολίζουμε

$$\theta_{r_n}^R := P(M_{1,r_n} \leq u_n | X_1 > u_n)$$

και

$$Y_i = I(X_i > u_n, M_{i,i+r_n-1} \leq u_n) - \theta_{r_n}^R I(X_i > u_n).$$

Έχοντας κατά νου τους παραπάνω συμβολισμούς, αποδεικνύεται το επόμενο θεώρημα (*Weissman and Novak (1998)*).

Θεώρημα 3.3.3 Υποθέτουμε ότι $\theta < 1$. Εάν ισχύουν οι προϋποθέσεις του Θεωρήματος 3.3.2 και επιπλέον $\sup_n \phi(kr_n, u_n) \rightarrow 0$, όταν $k \rightarrow \infty$,

$$\frac{\text{Var}(\sum_{i=1}^n Y_i)}{n\text{Var}(Y_1)} \rightarrow \sigma^2 \text{ (σταθερά)}, \quad r_n^2 = o(n\bar{F}(u_n)), \quad \theta_{r_n}^R - \theta = o(1/n\bar{F}(u_n)), \quad (n \rightarrow \infty)$$

τότε

$$\sqrt{n\bar{F}(u_n)}(\hat{\theta}_R - \theta) \rightarrow_d N(0, \sigma^2 \theta(1 - \theta)).$$

3.4 Μέθοδος των ενδιάμεσων χρόνων

Η τρίτη και τελευταία μέθοδος με την οποία θα ασχοληθούμε είναι η μέθοδος των ενδιάμεσων χρόνων. Το πλεονέκτημα αυτής της μεθόδου είναι ότι δεν χρειάζεται να κάνουμε καμία πρότερη επιλογή για κάποια παράμετρο, όπως στις δύο προηγούμενες περιπτώσεις όπου έπρεπε να επιλέξουμε το r . Τα αποτελέσματα που θα παρουσιάσουμε βασίζονται στην εργασία των *Ferro and Segers (2003)*.

Την ιδέα για την εκτιμήτρια του θ που πρέπει να χρησιμοποιήσουμε μας την δίνει το παρακάτω αποτέλεσμα το οποίο και αποδείξαμε στην Παράγραφο 2.3,

$$\bar{F}(u_n) T(u_n) \rightarrow_d T_\theta$$

όπου T_θ τυχαία μεταβλητή η οποία παίρνει την τιμή μηδέν με πιθανότητα $1 - \theta$, ενώ με πιθανότητα θ ακολουθεί εκθετική κατανομή με παράμετρο θ . Ας συμβολίσουμε με

$$N = N(u) := \sum_{i=1}^n I(X_i > u_n)$$

το πλήθος των υπερβάσεων του κατωφλίου u . Επίσης, έστω $1 \leq S_1 < \dots < S_N \leq n$ οι χρονικές στιγμές των υπερβάσεων. Προφανώς, οι ενδιάμεσοι χρόνοι θα είναι $T_i = S_{i+1} - S_i$, $i = 1, \dots, N-1$. Ισχύει ότι

$$\frac{E(T_\theta^2)}{E^2(T_\theta)} = \frac{\theta \frac{2}{\theta^2}}{(\theta \frac{1}{\theta})^2} = \frac{2}{\theta},$$

δηλαδή

$$\theta = \frac{2E^2(T_\theta)}{E(T_\theta^2)}.$$

Αντικαθιστώντας την πρώτη και δεύτερη ροπή της τ.μ T_θ με τις αντίστοιχες δειγματικές, παίρνουμε την παρακάτω εκτιμήτρια του θ ,

$$\hat{\theta}_{\text{int}}^{(1)} = \hat{\theta}_{\text{int}}^{(1)}(n, u) := \frac{2 \left(\sum_{i=1}^{N-1} T_i \right)^2}{(N-1) \sum_{i=1}^{N-1} T_i^2}.$$

Αποδεικνύεται ότι η $\hat{\theta}_{\text{int}}^{(1)}$ είναι συνεπής εκτιμήτρια του θ στη περίπτωση όπου η ακολουθία $X_n, n = 1, 2, \dots$ είναι m -εξαρτημένη (m -dependent).

Είναι φανερό ότι κάθε $T_i, i = 1, \dots, N-1$, συνεισφέρει στην εκτίμηση του θ , αφού $T_i \geq 1$. Ωστόσο, οι «μικροί» ενδιάμεσοι χρόνοι, σύμφωνα με όσα έχουμε αναφέρει, θα πρέπει να μην λαμβάνονται υπόψη (να μοντελοποιούνται ως μηδενικοί). Μια βελτιωμένη εκτιμήτρια (με μικρότερη μεροληψία), στην περίπτωση όπου $\max\{T_i: 1 \leq i \leq N-1\} > 2$, αποδεικνύεται ότι είναι η

$$\hat{\theta}_{\text{int}}^{(2)} = \hat{\theta}_{\text{int}}^{(2)}(n, u) := \frac{2 \left(\sum_{i=1}^{N-1} (T_i - 1) \right)^2}{(N-1) \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2)}.$$

Οι εκτιμήτριες $\hat{\theta}_{\text{int}}^{(1)}$ και $\hat{\theta}_{\text{int}}^{(2)}$ δεν είναι απαραίτητα μικρότερες ή ίσες της μονάδας. Για να αποφύγουμε αυτή τη περίπτωση αρκεί να λάβουμε ως εκτίμηση της θ την $\min(1, \hat{\theta}_{\text{int}}^{(1)})$, αν $\max\{T_i: 1 \leq i \leq N-1\} \leq 2$ ή την $\min(1, \hat{\theta}_{\text{int}}^{(2)})$, αν $\max\{T_i: 1 \leq i \leq N-1\} > 2$.

Με στόχο να πάρουμε ένα διάστημα εμπιστοσύνης για το θ μπορούμε να εφαρμόσουμε τη μέθοδο *Bootstrap*. Πιο συγκεκριμένα, συμβολίζουμε με

$$\tilde{\theta}_{\text{int}} = \begin{cases} \min(1, \hat{\theta}_{\text{int}}^{(1)}) & , \text{ αν } \max\{T_i: 1 \leq i \leq N-1\} \leq 2 \\ \min(1, \hat{\theta}_{\text{int}}^{(2)}) & , \text{ αν } \max\{T_i: 1 \leq i \leq N-1\} > 2 \end{cases}$$

και με n_c το πλήθος των συστάδων στο δείγμα. Όπως έχουμε ήδη αναφέρει στην Παράγραφο 2.3, το θ εκφράζει το ποσοστό των ενδιάμεσων χρόνων που δεν είναι (σχεδόν) «μηδενικοί». Δεδομένου ότι έχουμε $N-1$ ενδιάμεσους χρόνους στο δείγμα μας, είναι λογικό να υποθέσουμε ότι οι $[\tilde{\theta}_{\text{int}}(N-1)]$ μεγαλύτεροι να είναι οι ενδιάμεσοι χρόνοι μεταξύ των συστάδων (*intercluster times*). Προφανώς, οι υπόλοιποι $(N-1)-[\tilde{\theta}_{\text{int}}(N-1)]$ θα είναι οι ενδιάμεσοι χρόνοι εντός των συστάδων (*intracluster times*). Έτσι λοιπόν, θα έχουμε ότι $n_c-1 = [\tilde{\theta}_{\text{int}}(N-1)]$. Αν διατάξουμε τους ενδιάμεσους χρόνους σε φθίνουσα σειρά

$$T_{i_1} > T_{i_2} > \dots > T_{i_{n_c-1}} > T_{i_{n_c}} > \dots > T_{i_{N-1}}$$

τότε, σύμφωνα με τον παραπάνω συλλογισμό, το σύνολο $\{T_{i_j}\}_{j=1}^{n_c-1}$ θα αποτελείται από τους ενδιάμεσους χρόνους μεταξύ των συστάδων. Όμοια, το σύνολο $\{T_{i_j}\}_{j=n_c}^{N-1}$ θα αποτελείται από τους ενδιάμεσους χρόνους εντός των συστάδων. Στην περίπτωση όπου $T_{i_{n_c-1}} = T_{i_{n_c}}$ τότε αρκεί να ελαττώσουμε το n_c έως ότου $T_{i_{n_c-1}} > T_{i_{n_c}}$.

Παράδειγμα 3.4.1 Έστω $N = 6$ και $S_1 = 8, S_2 = 21, S_3 = 22, S_4 = 23, S_5 = 25, S_6 = 28$. Άρα $\tilde{\theta}_{\text{int}} = 0.67, n_c = [(0.67) \cdot 5] + 1 = 4$ και $T_1 = 13, T_2 = 1, T_3 = 1, T_4 = 2, T_5 = 3$. Προφανώς, $T_1 > T_5 > T_4 > T_3 = T_2$ και το σύνολο $\{T_1, T_5, T_4\}$ είναι οι ενδιάμεσοι χρόνοι μεταξύ των συστάδων και το σύνολο $\{T_3, T_2\}$ είναι οι χρόνοι εντός των συστάδων.

Παράδειγμα 3.4.2 Έστω $N = 6$ και $S_1 = 8, S_2 = 21, S_3 = 26, S_4 = 27, S_5 = 28, S_6 = 29$. Άρα $\tilde{\theta}_{\text{int}} = 0.7, n_c = [(0.7) \cdot 5] + 1 = 4$ και $T_1 = 13, T_2 = 5, T_3 = 1, T_4 = 1, T_5 = 1$. Σε αυτή την περίπτωση έχουμε ότι $T_1 > T_2 > T_3 = T_4 = T_5$ και το σύνολο $\{T_1, T_2\}$ είναι οι ενδιάμεσοι χρόνοι μεταξύ των συστάδων και το σύνολο $\{T_3, T_4, T_5\}$ είναι οι χρόνοι εντός των συστάδων.

Ας συμβολίσουμε για λόγους απλότητας $A = \{T_{i_j}\}_{j=1}^{n_c-1}$ και $B = \{T_{i_j}\}_{j=n_c}^{N-1}$. Προφανώς, το σύνολο B ταυτίζεται με το σύνολο

$$\{T_1, \dots, T_{i_1-1}\} \cup \{T_{i_1+1}, \dots, T_{i_2-1}\} \cup \dots \cup \{T_{i_{n_c-1}+1}, \dots, T_{N-1}\} = \bigcup_{j=1}^{n_c} \{T_{i_{j-1}+1}, \dots, T_{i_j-1}\},$$

όπου $i_0 = 1, i_{n_c} = N$ και $\{T_{i_{j-1}+1}, \dots, T_{i_j-1}\} = \emptyset$ αν $i_{j-1} + 1 = i_j, j = 1, \dots, n_c$. Ας συμβολίσουμε με $T^{(j)} = \{T_{i_{j-1}+1}, \dots, T_{i_j-1}\}$ και $B' = \{T^{(1)}, T^{(2)}, \dots, T^{(n_c)}\}$. Αν ισχύουν οι προϋποθέσεις του

Θεωρήματος 2.4.3 είναι λογικό να υποθέσουμε ότι για αρκετά μεγάλο αριθμό παρατηρήσεων η διαδικασία θα προσεγγίζει μία σύνθετη διαδικασία *Poisson*. Συνεπώς, το σύνολο A καθώς και το σύνολο B' θα περιέχουν χρόνους οι οποίοι είναι οριακά ανεξάρτητοι. Επίσης, οι χρόνοι του συνόλου A θα είναι ανεξάρτητοι από τους χρόνους του συνόλου B' . Έπειτα από αυτή την παρατήρηση είμαστε σε θέση να εφαρμόσουμε τη μέθοδο *Bootstrap* για να βρούμε το διάστημα εμπιστοσύνης για το θ . Η διαδικασία που πρέπει να ακολουθήσουμε είναι η εξής:

1. Παράγουμε με τη μέθοδο της προσομοίωσης k τυχαία δείγματα $A^{(i)}$, $i = 1, 2, \dots, k$, όπου τα σύνολα $A^{(i)}$ προκύπτουν έπειτα από δειγματοληψία με επανάθεση από το αρχικό δείγμα A . Την ίδια διαδικασία ακολουθούμε και για το σύνολο B' παίρνοντας k τυχαία δείγματα $B'^{(i)}$, $i = 1, 2, \dots, k$.
2. Θεωρούμε τα σύνολα $G^{(i)} = A^{(i)} \cup B'^{(i)}$ για κάθε $i = 1, 2, \dots, k$.
3. Για κάθε $G^{(i)}$ υπολογίζουμε τον δείκτη ακρότατης κατάστασης $\tilde{\theta}_{\text{int}}^{(i)}$, $i=1, \dots, k$.

Ακολουθώντας τη παραπάνω διαδικασία παίρνουμε k τιμές για την $\tilde{\theta}_{\text{int}}$, $Q = \{\tilde{\theta}_{\text{int}}^{(1)}, \dots, \tilde{\theta}_{\text{int}}^{(k)}\}$. Εάν λοιπόν διατάξουμε τα στοιχεία του συνόλου Q σε αύξουσα σειρά τότε το *Bootstrap* διάστημα εμπιστοσύνης συντελεστού εμπιστοσύνης $1-a$ για το θ είναι το $(2\tilde{\theta}_{\text{int}} - \tilde{\theta}_{\text{int},[(k+1)(1-a/2)]}, 2\tilde{\theta}_{\text{int}} - \tilde{\theta}_{\text{int},[(k+1)a/2]})$ όπου $\tilde{\theta}_{\text{int},[(k+1)(1-a/2)]}$ και $\tilde{\theta}_{\text{int},[(k+1)a/2]}$ είναι το $[(k+1)(1-a/2)]$ -οστό και $[(k+1)a/2]$ -οστό μικρότερο στοιχείο του συνόλου Q αντίστοιχα.

Για περισσότερες πληροφορίες σχετικά με την εκτίμηση του δείκτη ακρότατης εξάρτησης παραπέμπουμε στις εργασίες των *Hsing (1991)*, *Hsing (1993)*, *Weissman and Novak (1998)*, *Ferro and Segers (2003)*, *Robert, Segers and Ferro (2009)*, *Weissman and Cohen (1995)*, *Chernick, Hsing and McCormick (1991)*, *Smith and Weissman (1994)*.

3.5 Μελέτη μέσω Monte Carlo προσομοίωσης

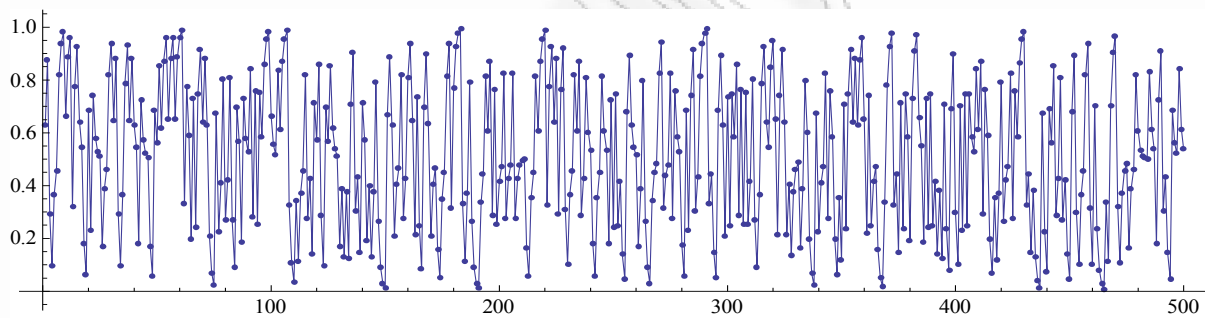
Σε αυτή την παράγραφο θα ασχοληθούμε με την εκτίμηση του δείκτη ακρότατης εξάρτησης βασιζόμενοι σε προσομοιωμένα δεδομένα. Συγκεκριμένα, θα εκτιμήσουμε τον δείκτη ακρότατης εξάρτησης χρησιμοποιώντας τις εκτιμήτριες που παρουσιάσαμε στις

παραπάνω παραγράφους. Αυτό θα γίνει για τρία σύνολα δεδομένων τα οποία θα προκύψουν μέσω προσομοίωσης τριών διαφορετικών διαδικασιών.

(α) Το πρώτο μοντέλο που θα μελετήσουμε είναι η στάσιμη διαδικασία AR(1) η οποία περιγράφεται από την συνθήκη

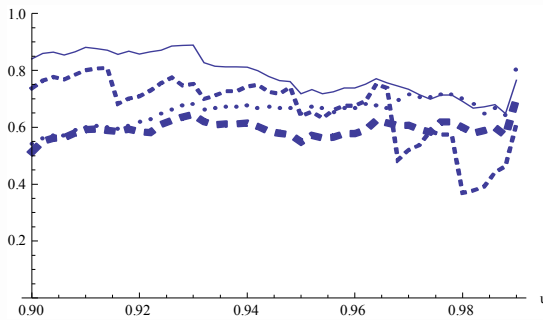
$$X_t = \frac{1}{r_0} X_{t-1} + \varepsilon_t$$

για $t \geq 1$, $r_0 \geq 2$, $X_0 \sim Uniform(0,1)$ και (ε_t) i.i.d ακολουθία τ.μ, ανεξάρτητη της X_0 , με $P(\varepsilon_1 = j/r_0) = 1/r_0$, $j = 0, 1, \dots, r_0 - 1$. Η παραπάνω διαδικασία έχει δείκτη ακρότατης εξάρτησης $\theta = 1 - 1/r_0$ (Chernick et al. (1991)). Προσομοιώνοντας τις X_1, \dots, X_{400} , με $r_0 = 3$ παίρνουμε το παρακάτω διάγραμμα.

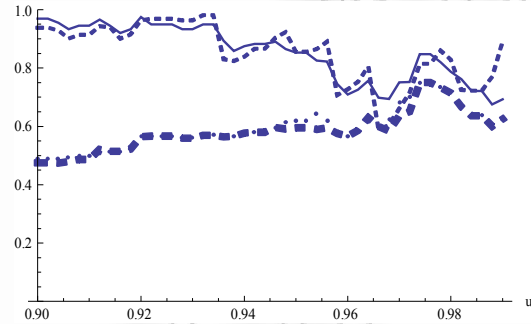


Αν θεωρήσουμε ότι το κατώφλι u παίρνει τιμές στο διάστημα $(0.90, 0.99)$ και επιπλέον συμβολίσουμε με r, z τις παραμέτρους επιλογής στη μέθοδο των ομάδων και στη μέθοδο των ροών αντίστοιχα, τότε προκύπτουν τα παρακάτω διαγράμματα για τις εκτιμήτριες $\hat{\theta}_{bl}^{(1)}$, $\hat{\theta}_{sl}$, $\hat{\theta}_R$ και $\hat{\theta}_{int}^{(2)}$ όπου έχουμε θεωρήσει τρεις διαφορετικές επιλογές των παραμέτρων r, z .

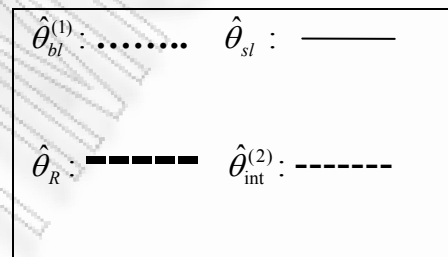
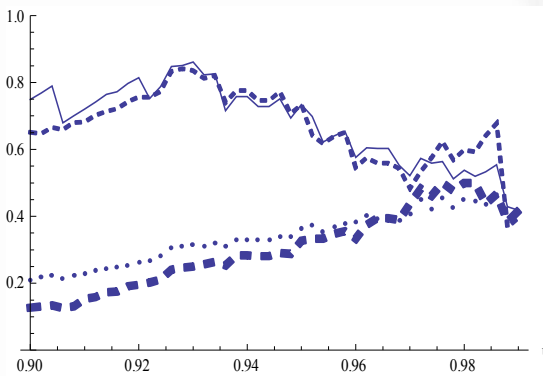
$n = 1000, r = 10, z = 5$



$n = 1000, r = 20, z = 10$



$n = 1000, r = 40, z = 20$



Από τα παραπάνω διαγράμματα φαίνεται ότι οι τέσσερις εκτιμήτριες προσεγγίζουν την θεωρητική τιμή του δείκτη ακρότατης εξάρτησης ($\theta \approx 0.66$) για τιμές του κατωφλίου $u > 0.98$. Στη συνέχεια παράγουμε χίλιες διαφορετικές πραγματοποιήσεις της διαδικασίας X_1, X_2, \dots, X_n ($n = 1000$) και για κάθε μια από αυτές υπολογίσουμε τις τιμές των εκτιμητριών $\hat{\theta}_{bl}^{(1)}$, $\hat{\theta}_{sl}$, $\hat{\theta}_R$, $\hat{\theta}_{int}^{(2)}$. Οι τέσσερις μέσοι όροι αυτών των χιλίων τιμών εκτιμούν τις αναμενόμενες τιμές των $\theta_{bl}^{(1)}$, θ_{sl} , θ_R , $\theta_{int}^{(2)}$. Συγκεκριμένα, κατασκευάζουμε τους παρακάτω πίνακες για διάφορες τιμές των παραμέτρων u, r, z .

$\hat{\theta}_{bl}^{(1)}$	$u = 0.94$	$u = 0.96$	$u = 0.98$	$u = 0.985$
$r = 10$	0.627899	0.645453	0.693684	0.706741
$r = 20$	0.502852	0.561902	0.645984	0.663185
$r = 40$	0.355296	0.437024	0.561918	0.589316

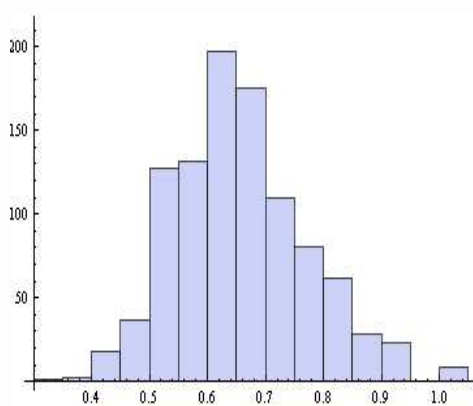
$\hat{\theta}_{sl}$	$u = 0.94$	$u = 0.96$	$u = 0.98$	$u = 0.985$
$r = 10$	0.787785	0.746741	0.746995	0.747332
$r = 20$	0.773846	0.742985	0.742291	0.738161
$r = 40$	0.799509	0.742921	0.737557	0.727753

$\hat{\theta}_R$	$u = 0.94$	$u = 0.96$	$u = 0.98$	$u = 0.985$
$r = 5$	0.611202	0.618529	0.66799	0.677538
$r = 10$	0.478681	0.544398	0.627295	0.648423
$r = 20$	0.307393	0.403809	0.539306	0.572013

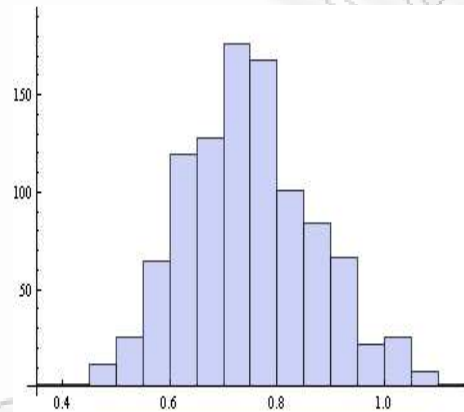
$\hat{\theta}_{int}^{(2)}$	$u = 0.94$	$u = 0.96$	$u = 0.98$	$u = 0.985$
	0.762465	0.750560	0.808535	0.828462

Παρατηρώντας τους παραπάνω πίνακες, οι εκτιμήτριες $\hat{\theta}_{sl}$ και $\hat{\theta}_{int}^{(2)}$ φαίνεται να υπερεκτιμούν τον θ . Συγκεκριμένα, η $\hat{\theta}_{sl}$ παράγει υψηλές εκτιμήσεις του θ (0.72 μέχρι 0.80) για οποιαδήποτε επιλογή των παραμέτρων u, r ενώ η εκτιμήτρια $\hat{\theta}_{int}^{(2)}$ φαίνεται να μην επηρεάζεται σημαντικά από την επιλογή του κατωφλίου u και εκτιμά τον $\theta \approx 0.8$. Αντιθέτως, η εκτιμήτρια $\hat{\theta}_{bl}^{(1)}$ δίνει αξιόπιστες τιμές για τον θ όταν $r = 20$ και $u > 0.98$ ενώ η $\hat{\theta}_R$ προσεγγίζει ικανοποιητικά την θεωρητική τιμή του θ όταν $z = 5$ και $u > 0.98$. Τέλος, μπορούμε να πάρουμε μια εικόνα για την κατανομή των τεσσάρων αυτών εκτιμητριών μέσω

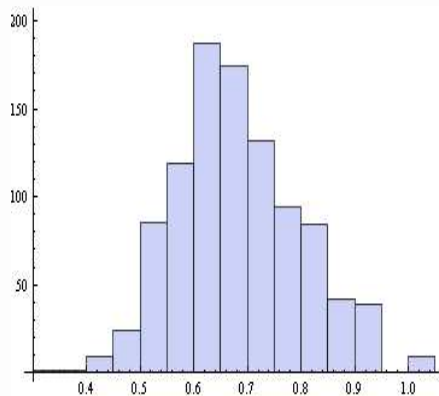
των αντίστοιχων ιστογραμμάτων συχνοτήτων. Για κατώφλι $u = 0.98$, $r = 20$ και $z = 5$ από τα παρακάτω ιστογράμματα φαίνεται ότι οι τέσσερις εκτιμήτριες ακολουθούν περίπου κανονική κατανομή.



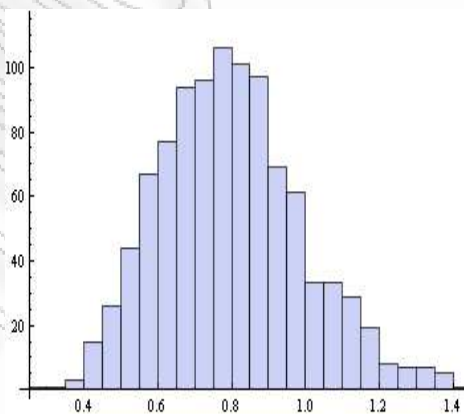
$\hat{\theta}_{bl}^{(1)}$



$\hat{\theta}_{sl}$



$\hat{\theta}_R$



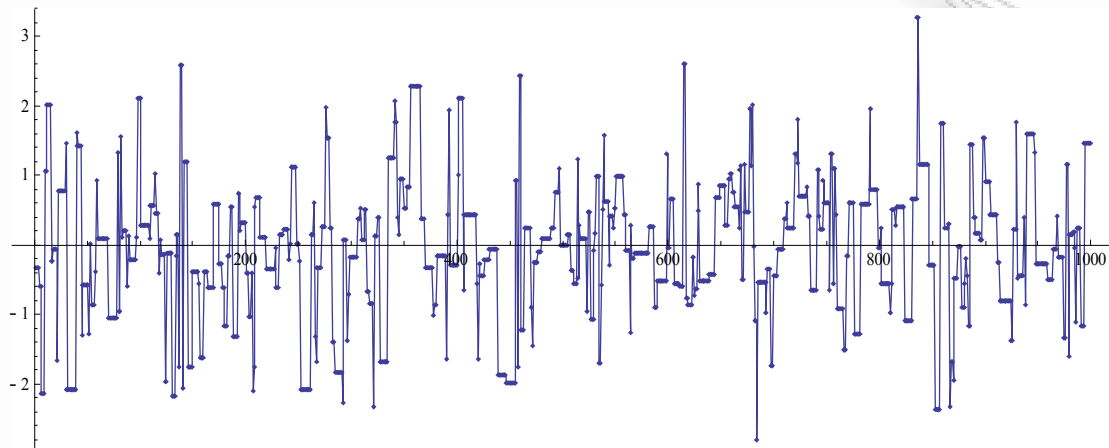
$\hat{\theta}_{int}^{(2)}$

(β) Η δεύτερη διαδικασία με την οποία θα ασχοληθούμε περιγράφεται από τη σχέση

$$X_t = (1 - a_t)X_{t-1} + a_t \varepsilon_t$$

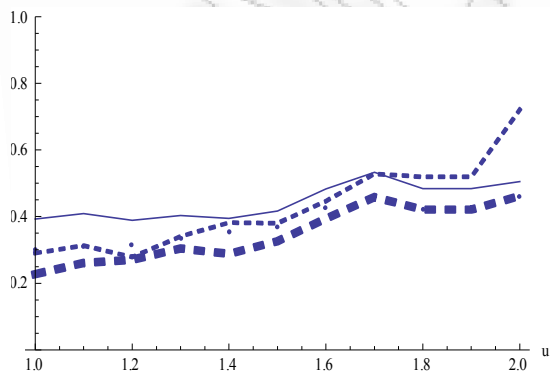
για $t \geq 2$, όπου (ε_t) , (a_t) ανεξάρτητες ακολουθίες i.i.d τ.μ, $X_1 = \varepsilon_1$, $\varepsilon_t \sim N(0,1)$ και $a_t \sim Bernoulli(p)$. Η διαδικασία που περιγράφεται από την παραπάνω συνθήκη έχει δείκτη ακρότατης εξάρτησης $\theta = p$ (Weissman and Novak (1998)). Όπως και στην προηγούμενη

εφαρμογή, αν προσομοιάσουμε $n = 1000$ τιμές για την διαδικασία (X_t) με $p = 0.3$ παίρνουμε το παρακάτω διάγραμμα.

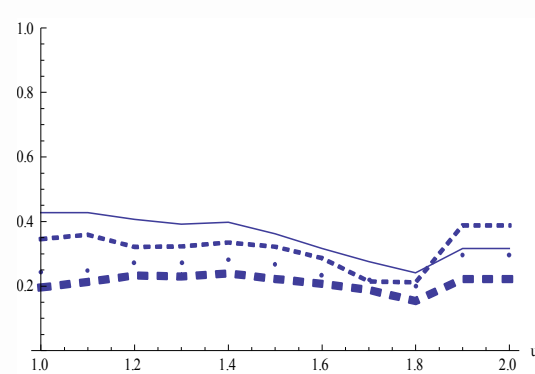


Από το παραπάνω διάγραμμα παρατηρούμε ότι υπάρχει θετική τοπική εξάρτηση. Για παράδειγμα από την στιγμή που η διαδικασία θα πάρει κάποια τιμή στο σύνολο $(0, \infty)$ (αντίστοιχα στο σύνολο $(-\infty, 0)$), έχει την τάση να παραμείνει σε αυτό το σύνολο τιμών. Αν τώρα θεωρήσουμε ότι το κατώφλι u παίρνει τιμές στο διάστημα $(1, 2)$ τα διαγράμματα που προκύπτουν για τις εκτιμήτριες $\hat{\theta}_{bl}^{(1)}$, $\hat{\theta}_{sl}$, $\hat{\theta}_R$ και $\hat{\theta}_{int}^{(2)}$ για τρεις διαφορετικές επιλογές των παραμέτρων r, z , είναι τα παρακάτω.

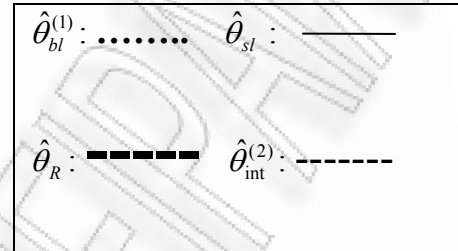
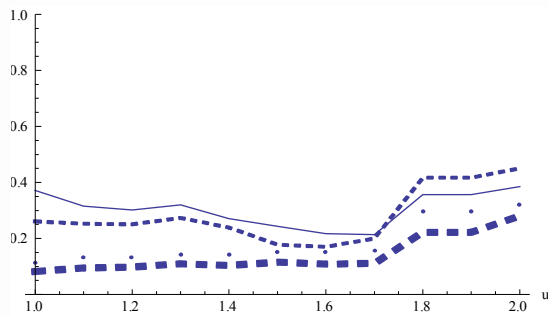
$n = 1000, r = 10, z = 5$



$n = 1000, r = 20, z = 10$



$n = 1000$, $r = 40$, $z = 20$



Παρατηρώντας τα παραπάνω διαγράμματα είναι φανερό ότι οι τέσσερις εκτιμήτριες προσεγγίζουν την θεωρητική τιμή του δείκτη ακρότατης εξάρτησης ($\theta = 0.3$) για τιμές του κατωφλίου $u \in (1.6, 2)$. Αν παράγουμε χίλιες διαφορετικές πραγματοποιήσεις της διαδικασίας X_1, X_2, \dots, X_n ($n = 1000$) και έπειτα υπολογίσουμε τον μέσο όρο των εκτιμήσεων των $\hat{\theta}_{bl}^{(1)}$, $\hat{\theta}_{sl}$, $\hat{\theta}_R$, $\hat{\theta}_{int}^{(2)}$, τότε προκύπτουν οι παρακάτω πίνακες.

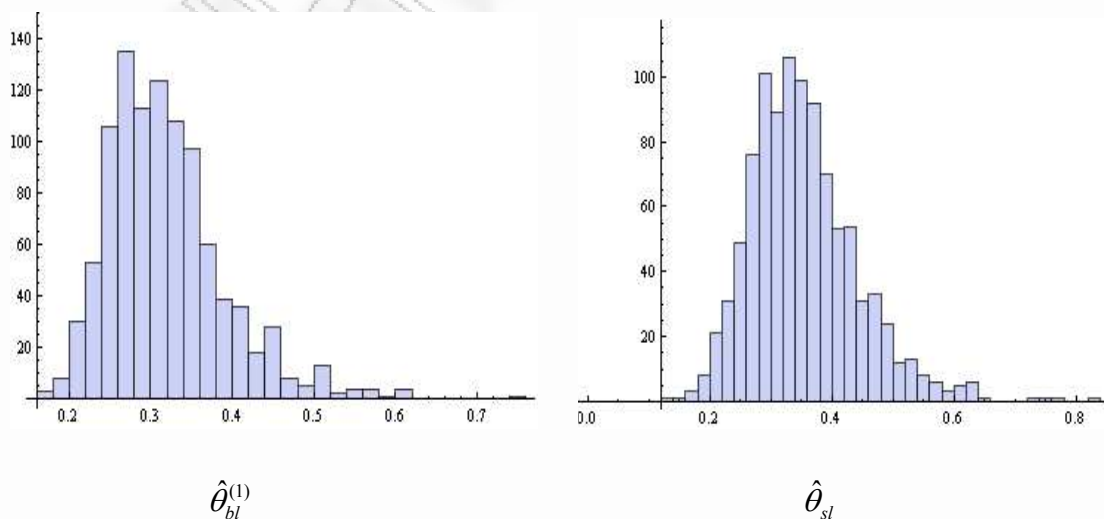
$\hat{\theta}_{bl}^{(1)}$	$u = 1.6$	$u = 1.7$	$u = 1.8$	$u = 1.9$
$r = 10$	0.341789	0.371597	0.365536	0.400054
$r = 20$	0.303940	0.294661	0.316456	0.343824
$r = 40$	0.236936	0.257001	0.277779	0.306898

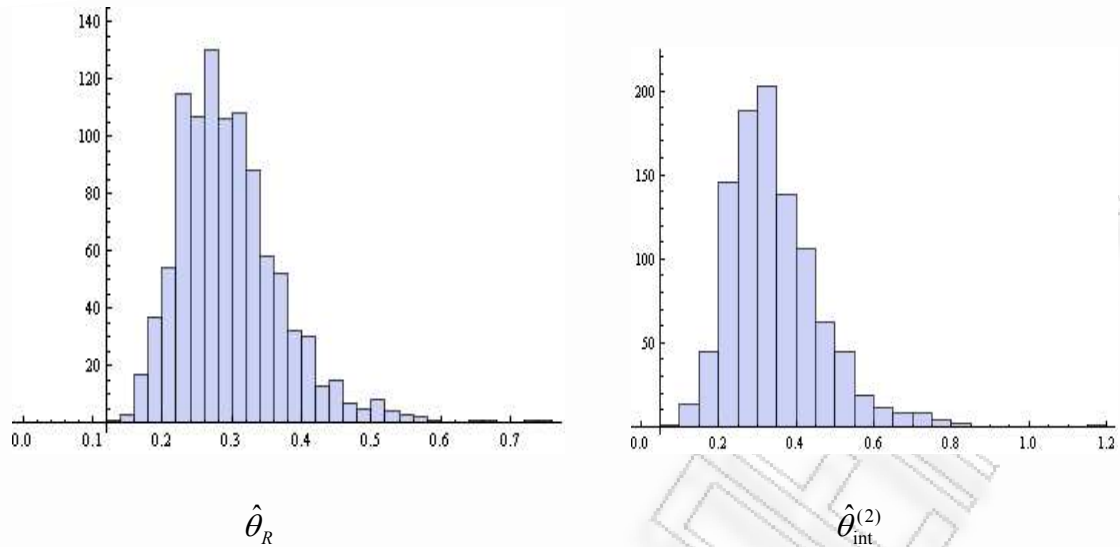
$\hat{\theta}_{sl}$	$u = 1.6$	$u = 1.7$	$u = 1.8$	$u = 1.9$
$r = 10$	0.382698	0.403329	0.385313	0.419266
$r = 20$	0.360509	0.342454	0.351791	0.380409
$r = 40$	0.328965	0.334606	0.350227	0.370179

$\hat{\theta}_R$	$u = 1.6$	$u = 1.7$	$u = 1.8$	$u = 1.9$
$z = 5$	0.288179	0.303383	0.290561	0.331308
$z = 10$	0.253899	0.266754	0.264543	0.310195
$z = 20$	0.218296	0.225647	0.250874	0.284597

$\hat{\theta}_{int}^{(2)}$	$u = 1.6$	$u = 1.7$	$u = 1.8$	$u = 1.9$
	0.337508	0.344945	0.368449	0.392838

Από τους παραπάνω πίνακες βλέπουμε ότι η εκτιμήτρια $\hat{\theta}_{int}^{(2)}$ υπερεκτιμά τον θ για οποιαδήποτε επιλογή του κατωφλίου u . Η $\hat{\theta}_{sl}$ παράγει ικανοποιητικές εκτιμήσεις του θ για $r = 40$ και χαμηλές τιμές του u (1.6 μέχρι 1.7). Η εκτιμήτρια $\hat{\theta}_{bl}^{(1)}$ δίνει αξιόπιστες τιμές όταν $r = 20$ και $1.6 < u < 1.8$ ενώ η $\hat{\theta}_R$ προσεγγίζει ικανοποιητικά την θεωρητική τιμή του θ όταν $z = 5$ για οποιαδήποτε επιλογή του κατωφλίου u μεγαλύτερη από 1.6. Τέλος, αν συμβολίσουμε με r_s την παράμετρο επιλογής στη μέθοδο των επικαλυπτόμενων ομάδων για κατώφλι $u = 1.7$, $r = 20$, $r_s = 40$ και $z = 5$ τα ιστογράμματα συχνοτήτων που προκύπτουν είναι τα παρακάτω.

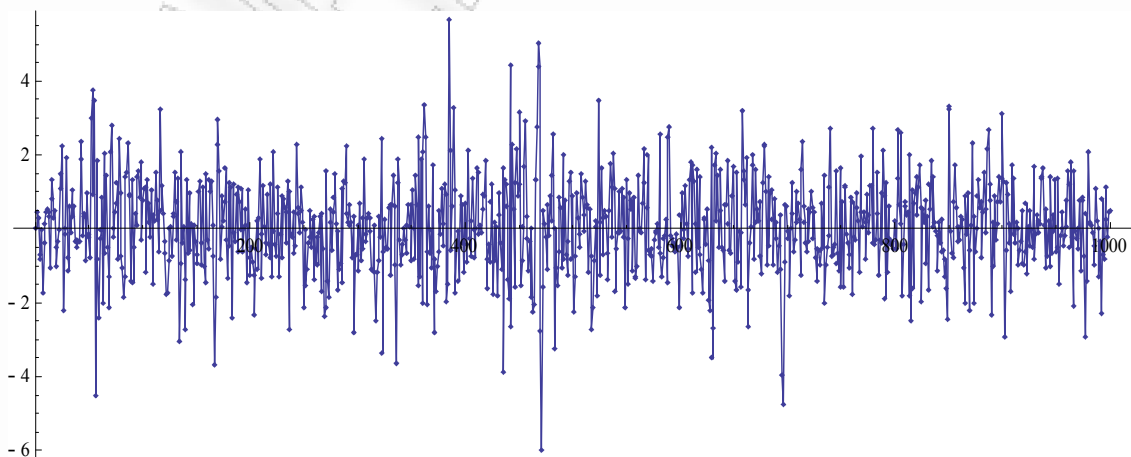




(γ) Η τρίτη και τελευταία διαδικασία με την οποία θα ασχοληθούμε είναι η διαδικασία ARCH(1) η οποία περιγράφεται από τη σχέση

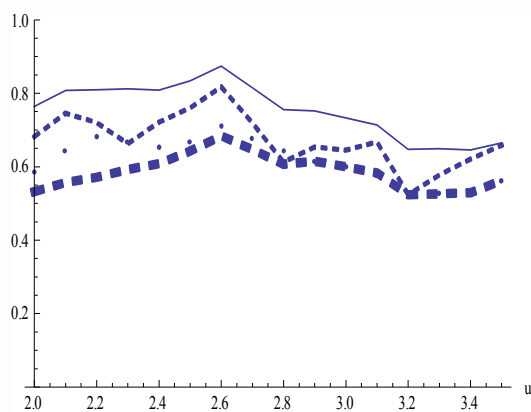
$$X_t = \sqrt{\beta + \lambda X_{t-1}^2} \varepsilon_t$$

για $t \geq 1$, $\beta > 0$, $\lambda > 0$, $X_0 = 0$ και (ε_t) ακολουθία i.i.d τ.μ με $\varepsilon_t \sim N(0,1)$, οι οποίες επιπλέον είναι ανεξάρτητες της X_0 . Θα περιοριστούμε στη μελέτη της συμπεριφοράς της παραπάνω διαδικασίας για σταθερές τιμές των παραμέτρων β , λ και συγκεκριμένα θα θεωρήσουμε ότι $\beta = 1$ και $\lambda = 0.5$. Για αυτές τις τιμές, η διαδικασία ARCH(1) έχει δείκτη ακρότατης εξάρτησης $\theta = 0.835$ (Embrechts et al. (1997)). Προσομοιώνοντας τις X_1, \dots, X_{1000} παίρνουμε το παρακάτω διάγραμμα.

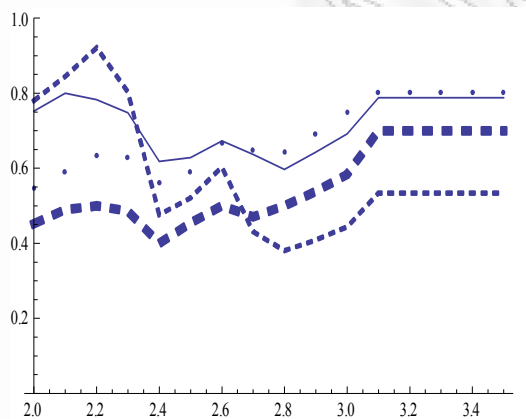


Αν θεωρήσουμε ότι το κατώφλι u παίρνει τιμές στο διάστημα $(2, 3.5)$ τότε παίρνουμε τα παρακάτω διαγράμματα για τις εκτιμήτριες $\hat{\theta}_{bl}^{(1)}$, $\hat{\theta}_{sl}$, $\hat{\theta}_R$ και $\hat{\theta}_{int}^{(2)}$.

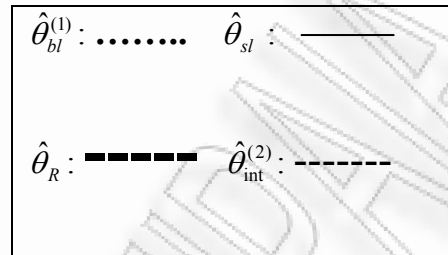
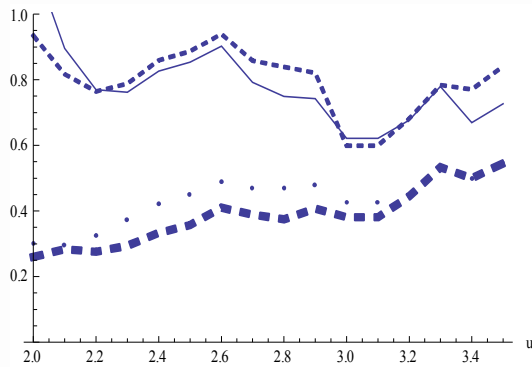
$n = 1000, r = 10, z = 5$



$n = 1000, r = 20, z = 10$



$n = 1000, r = 40, z = 10$



Αν τώρα παράγουμε χίλιες διαφορετικές πραγματοποιήσεις της διαδικασίας X_1, X_2, \dots, X_n ($n = 1000$) και έπειτα υπολογίσουμε τον μέσο όρο των εκτιμήσεων των $\hat{\theta}_{bl}^{(1)}, \hat{\theta}_{sl}, \hat{\theta}_R, \hat{\theta}_{int}^{(2)}$, τότε προκύπτουν οι παρακάτω πίνακες.

$\hat{\theta}_{bl}^{(1)}$	$u = 2.4$	$u = 2.6$	$u = 2.8$	$u = 3$
$r = 10$	0.742697	0.763043	0.782160	0.791951
$r = 20$	0.628589	0.669313	0.706379	0.731009
$r = 40$	0.476917	0.534396	0.589052	0.637873

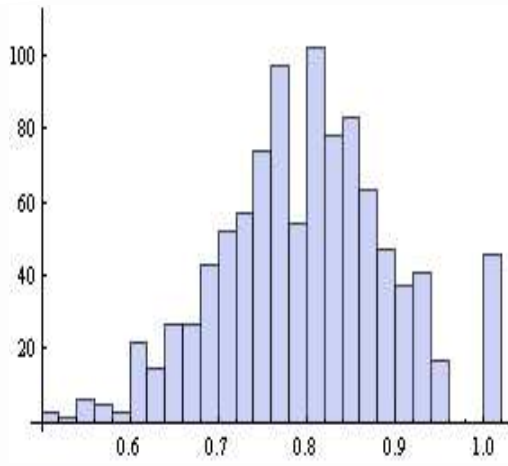
$\hat{\theta}_{sl}$	$u = 2.4$	$u = 2.6$	$u = 2.8$	$u = 3$
$r = 10$	0.876747	0.869165	0.863996	0.860438
$r = 20$	0.861660	0.862876	0.861873	0.854309
$r = 40$	0.875385	0.863219	0.861209	0.865437

$\hat{\theta}_R$	$u = 2.4$	$u = 2.6$	$u = 2.8$	$u = 3$
$z = 5$	0.725472	0.746304	0.760706	0.772428
$z = 10$	0.609782	0.656203	0.691112	0.715390
$z = 20$	0.438395	0.503221	0.563203	0.613351

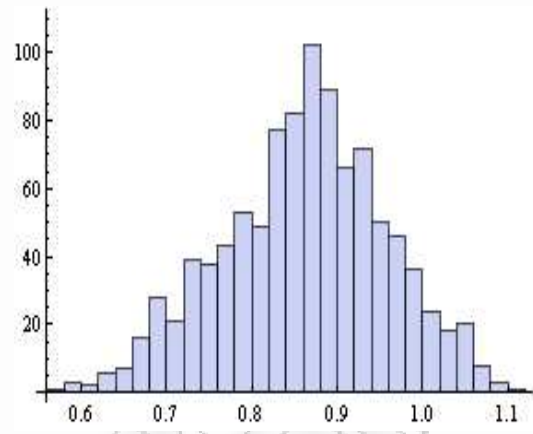
$\hat{\theta}_{int}^{(2)}$	$u = 2.4$	$u = 2.6$	$u = 2.8$	$u = 3$
	0.883904	0.886004	0.899610	0.915228

Παρατηρώντας τους παραπάνω πίνακες, βλέπουμε ότι τις πιο «σταθερές» εκτιμήσεις, οι οποίες φαίνεται να μη επηρεάζονται από την επιλογή του κατωφλίου u , μας δίνουν οι εκτιμήτριες $\hat{\theta}_{sl}$ και $\hat{\theta}_{int}^{(2)}$. Επιπλέον, οι εκτιμήσεις που μας δίνει η $\hat{\theta}_{sl}$ δεν διαφέρουν σημαντικά για τις παραπάνω επιλογές της παραμέτρου r και προσεγγίζουν ικανοποιητικά την θεωρητική τιμή του θ εν αντιθέσει με την $\hat{\theta}_{int}^{(2)}$ η οποία υπερεκτιμά τον θ . Τέλος, η $\hat{\theta}_{bl}^{(1)}$ και η $\hat{\theta}_R$ δίνουν εκτιμήσεις για τον θ χαμηλότερες της θεωρητική τιμής του για οποιαδήποτε επιλογή των παραμέτρων r, z .

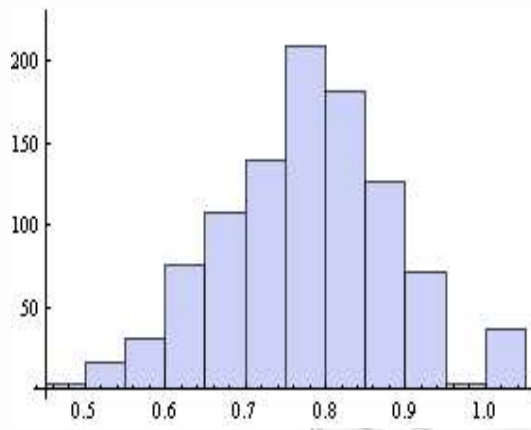
Για κατώφλι $u = 3$, $r = 10$ και $z = 5$ τα ιστογράμματα συχνότητων των εκτιμητριών $\hat{\theta}_{bl}^{(1)}$, $\hat{\theta}_{sl}$, $\hat{\theta}_R$ και $\hat{\theta}_{int}^{(2)}$ είναι τα παρακάτω.



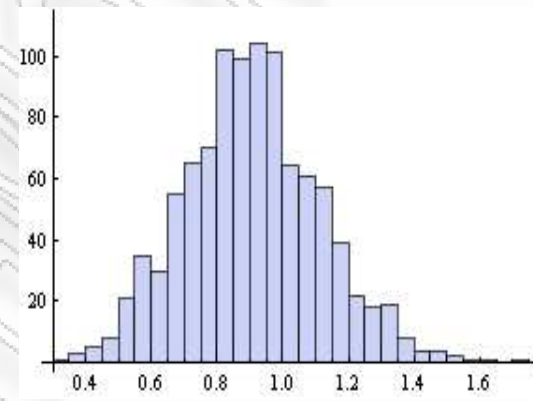
$\hat{\theta}_{bl}^{(1)}$



$\hat{\theta}_{sl}$



$\hat{\theta}_R$



$\hat{\theta}_{int}^{(2)}$

BIBΛΙΟΓΡΑΦΙΑ

1. N. Balakrishnan and M.V. Koutras (2001) *Runs and scans with applications*, John Wiley & Sons.
2. J. Beirlant, Y. Goegebeur, J. Teugels, J. Segers, D. D. Waal and C. Ferro (2004) *Statistics of Extremes*. John Wiley & Sons,Ltd.
3. M. R. Chernick,T. Hsing and W. P. McCormick (1991) *Calculating the extremal index for a class of stationary sequences*. Adv. App. Prob. **23**, 835-850.
4. S.Coles (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag.
5. D.J.Daley and D.Vere-Jones (1988) *An Introduction to the Theory of Point Processes*. Springer, Berlin.
6. P. Embrechts, C. Klüppelberg and T. Mikosch (1997) *Modelling Extremal Events*. Springer-Verlag.
7. C. A. T. Ferro and J. Segers (2003) *Inference for clusters of extreme values*. Journal of the Royal Statistical Society B **65**, 545-556.
8. R. A. Fisher and L. H. C. Tippett (1928) *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Proc. Cambridge Philos. Soc. **24**, 180-190.
9. J. Glaz, J. I. Naus, S. Wallenstein (2001) *Scan Statistics*, Springer-Verlag
10. L. de Haan (1970) *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*. CWI Tract **32**, Amsterdam.
11. L. de Haan and A. Ferreira (2006) *Extreme Value Theory*. Springer-New York.
12. T. Hsing (1991) *On tail index estimation using dependent data*. Annals of Statistics **19**, 1547-1569.
13. T. Hsing (1993) *Extremal index estimation for a weakly dependent stationary sequence*. Annals of Statistics **21**, 2043-2071.
14. T. Hsing, J. Hüsler and M. R. Leadbetter (1988) *On the exceedance point process for a stationary sequence*. Probability Theory and Related Fields **78**, 97-112.
15. S. Kotz and S. Nadarajah (2000) *Extreme Value Distributions*. Imperial College Press.

16. M. R. Leadbetter (1974) *On extreme values in stationary sequences*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **28** , 289-303.
17. M. R. Leadbetter (1983) *Extremes and local dependence in stationary sequences*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **65** , 291-306.
18. M. R. Leadbetter (1991) *On a basis for 'Peaks over Threshold' modeling*. Statistics & Probability Letters **12**, 357-362.
19. M. R. Leadbetter (1995) *On high level exceedance modeling and tail inference*. Journal of Statistical Planning and Inference **45**, 247-260.
20. M. R. Leadbetter, G. Lindgren and H. Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag.
21. M. R. Leadbetter and H. Rootzén (1988) *Extremal theory for stochastic processes*. The Annals of Probability **16**, 431-478.
22. G. L O'Brien (1974) *Limit theorems for the maximum term of a stationary process*. The Annals of Probability **2**, 540-545.
23. G. L O'Brien (1987) *Extreme values for stationary and Markov sequences*. The Annals of Probability **15**, 281-291.
24. J. Pickands (1975) *Statistical inference using extreme order statistics*. Annals of Statistics **3**, 119-131.
25. R.-D. Reiss and M. Thomas (2007) *Statistical Analysis of Extreme Values (third edition)*. Birkhäuser Verlag AG.
26. S. I. Resnick (1987), *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag.
27. C. Y. Robert, J. Segers and C. A. T. Ferro (2009) *A sliding blocks estimator for the extremal index*. Electronic Journal of Statistics **3**, 993-1020.
28. H. Rootzén, M. R. Leadbetter and L. de Haan (1998) *On the distribution of tail array sums for strongly mixing stationary sequences*. Annals of Applied Probability **8**, 868-885.
29. G. Salvadori, C. de Michele, N. T. Kottegoda and R. Rosso (2007) *Extremes in Nature*. Springer-New York.
30. J. Segers (2002) *Extreme events: dealing with dependence*. EURANDOM report 2002-036.

31. J. Segers (2003) *Approximate distributions of cluster of extremes*. Center Discussion Paper 2003-91, Tilburg University, The Netherlands.
32. R. L. Smith and I. Weissman (1994) *Estimating the extremal index*. Journal of the Royal Statistical Society *B* **56**, 515-528.
33. R. L. Smith, J. A. Tawn and S. Coles (1997) *Markov chain models for threshold exceedances*. Biometrika **84**, 249-268.
34. I. Weissman and U. Cohen (1995) *The extremal index and clustering of high values for derived stationary sequences*. Journal of Applied Probability **32**, 972-981.
35. I. Weissman and S. Y. Novak (1998) *On blocks and runs estimators of the extremal index*. Journal of Statistical Planning and Inference **66**, 281-288.