

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ
ΓΙΑ ΑΚΡΑΙΑ ΣΥΜΒΑΝΤΑ**

Πολυξένη Κ. Νομικού

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

Πειραιάς,
Φεβρουάριος 2008

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ
ΓΙΑ ΑΚΡΑΙΑ ΣΥΜΒΑΝΤΑ**

Πολυξένη Κ. Νομικού

Διπλωματική Εργασία

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

Πειραιάς,
Φεβρουάριος 2008

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

Τα μέλη της Επιτροπής ήταν:

- (Επιβλέπων)

-

-

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

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**STATISTICAL METHODS FOR
EXTREMAL EVENTS**

By

Polyxeni K. Nomikou

Msc Dissertation

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

Piraeus,

February 2008

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΔΙΑ

*Σε αυτούς που με υποστήριξαν
και με βοήθησαν
συγγενείς, φίλους, συνεργάτες
& στον σύντροφο μου.*

Περίληψη

Η θεωρία ακραίων τιμών αποτελεί ένα σημαντικό εργαλείο για την εκτίμηση των πιθανοτήτων πραγματοποίησης ακραίων και κατά συνέπεια σπανίων γεγονότων. Πρακτικά εφαρμόζεται σε πολλούς τομείς, όπως η υδρολογία, η μετεωρολογία, η ασφάλιση κ.α. και χρησιμοποιείται για την ανάλυση της συμπεριφοράς των μεγίστων και των ελαχίστων παρατηρήσεων ενός συνόλου παρατηρήσεων, δηλαδή των παρατηρήσεων που βρίσκονται στην ουρά μιας κατανομής. Στη μελέτη των ακραίων παρατηρήσεων, ερχόμαστε αντιμέτωποι με πολλές δυσκολίες, όπως τον εξ' ορισμού μικρό αριθμό των ακραίων παρατηρήσεων, τον προσδιορισμό της κατανομής που ακολουθούν και την επιλογή των κατάλληλων μεθόδων για τη εκτίμηση των παραμέτρων.

Στο Κεφάλαιο 1, θα αιτιολογήσουμε τη χρησιμοποίηση της θεωρίας ακραίων τιμών καθώς και τις βασικές θεωρητικές αρχές που τη διέπουν. Μέσω του Θεωρήματος Fisher – Tippett θα ορίσουμε τους τρεις τύπους κατανομών ακραίων τιμών. Στη συνέχεια θα περιγραφούν οι ιδιότητες των κατανομών αυτών και θα ορισθεί η γενικευμένη τους μορφή καθώς η γενικευμένη κατανομή Pareto. Τέλος θα αναφερθούμε στις ιδιότητες των διατεταγμένων παρατηρήσεων, που αποτελούν τη βάση για ένα μεγάλο αριθμό εκτιμητών.

Στο Κεφάλαιο 2, θα αναφερθούμε στα βασικά διαγραμματικά εργαλείων που χρησιμοποιούνται στην ανάλυση ακραίων τιμών. Στη συνέχεια θα παρουσιάσουμε την εκτίμηση των παραμέτρων της γενικευμένης κατανομής ακραίων παρατηρήσεων μέσω της μεθόδου μέγιστης πιθανοφάνειας. Θα δοθούν οι ορισμοί, μερικά θεωρητικά αποτελέσματα και πρακτικά παραδείγματα για τους εκτιμητές Pickands, Hill και Dekkers, Einmahl and de Haan για την εκτίμηση του ξ . Τέλος θα κάνουμε μια εισαγωγική αναφορά στη μέθοδο των υπερβάσεων πάνω από ένα υψηλό κατώφλι προσαρμόζοντας τη γενικευμένη κατανομή Pareto μέσω της μεθόδου μέγιστης πιθανοφάνειας και στο τρόπο που μπορούμε να προχωρήσουμε σε αποτίμηση κινδύνου χαρτοφυλακίου μέσω των εξαγόμενων αποτελεσμάτων.

Στο Κεφάλαιο 3, θα προχωρήσουμε σε μία πρακτική εφαρμογή της θεωρίας και των μεθόδων που παρουσιάστηκαν στα προηγούμενα κεφάλαια, αναλύοντας τη πορεία των λογαριθμικών αποδόσεων του δείκτη S&P 500. Η ανάλυση θα γίνει εξετάζοντας πρώτα τη συμπεριφορά των μηνιαίων αποδόσεων του δείκτη και στη συνέχεια τη συμπεριφορά των τιμών της ουράς της αρχικής κατανομής των παρατηρήσεων του δείκτη.

Abstract

Extreme Value Theory (EVT) is one of the most important tools when considering probabilities of extreme and thus rare events. It has application in many fields such as hydrology, meteorology, insurance and is used in modeling the behaviour of maxima and minima in a series of observations, i.e. of the tail of the distribution. When studying extremal values, we have to overcome a number of challenges such as the small number of extreme observations in hand, the type of distribution that fits the data and the estimation of both the distribution and the normalization parameters.

As it will be shown in Section 1, the inadequacy of standard statistical models – mainly the Gaussian assumption - when studying extreme values, led to the formulation of a more specialized theory. A vast theoretical background supports EVT, whose main conclusion is the existence of a unique limit law for sample maxima or minima, as expressed in the Fisher – Tippett Theorem. This theorem implies three possible types of the limit law, each with its own characteristics, thus making difficult the choice of the proper limiting distribution. Two generalizations will be presented, considering the way we collect the extreme values from a sample, revealing the Generalized Extreme value Distribution and the Generalized Pareto Distribution. Lastly we will talk about the k^{th} -order statistic, which constitutes the basis for a number of estimators presented subsequently.

In Section 2 we will handle more practical issues, such as graphical exploratory analysis of extreme data and the estimation of the distribution parameters. We will focus mainly on the *Block Maxima* method, the origins of which are traced back in hydrology. Then we will pass on to the Pickands, Hill and Dekkers-Einmahl-de Haan estimators, helpful when estimating the shape parameter ζ of the limiting distribution. Lastly, a brief passage will be made through the *Peaks Over Threshold* method, a useful tool in the analysis of economic and insurance data.

Having presented the theoretical background of EVT and the tools that one can use, in Section 3, we will study a practical application that will serve as an illustration for the results presented in the previous sections. Our main objective will be to assess whether the distribution of the S&P 500 index log returns have heavy tails, and estimate their shape parameters. Analysis will be made using both the *Block Maxima* – examining the behaviour of the monthly maxima and minima – and the *Peaks Over Threshold* – examining the behaviour of the tails of the distribution.

Περιεχόμενα

Κατάλογος Πινάκων	xv
Κατάλογος Σχημάτων	xvi
Κατάλογος Συντομογραφιών	xix
Κεφάλαιο 1, Θεωρία Κατανομών Ακραίων Τιμών	1
1.1 Εισαγωγή	1
1.1.1 Δυσκολίες στην ανάλυση ακραίων τιμών	1
1.1.2 Γραμμικός μετασχηματισμός της M_n	3
1.2 Βασική Θεωρία	5
1.2.1 Ορισμός (Σύγκλιση σε τύπο)	5
1.2.1.2 Πρόταση	5
1.2.1.3 Ορισμός (Γενικευμένη αντίστροφος συνάρτησης κατανομής)	6
1.2.1.4 Πρόταση (Σύγκλιση της γενικευμένης αντιστροφού)	6
1.2.3 Ορισμός (Ομαλά μεταβαλλόμενες συναρτήσεις)	6
1.2.4 Θεώρημα (Ομοιόμορφη σύγκλιση)	7
1.2.5 Θεώρημα (Αναπαράσταση των ομαλά μεταβαλλόμενων συναρτήσεων)	7
1.2.6 Θεώρημα του Karamata	8
1.2.7 Ορισμός (Ταχέως μεταβαλλόμενες συναρτήσεις)	8
1.2.8 Θεώρημα (Ιδιότητες των ταχέως μεταβαλλόμενων συναρτήσεων)	8
1.2.9 Ορισμός (Ομαλά μεταβαλλόμενες ακολουθίες)	8
1.3 Οριακές ιδιότητες κατανομών ακραίων τιμών	9
1.3.1 Πρόταση (Προσέγγιση Poisson)	9
1.3.2 Κατανομές αναλλοίωτου μεγίστου	11
1.3.3 Ορισμός (Κατανομές αναλλοίωτου μεγίστου)	11
1.3.4 Θεώρημα (Οριακή ιδιότητα των νόμων αναλλοίωτου μεγίστου)	12
1.3.5 Θεώρημα Fisher – Tippett	13
1.3.6 Παράδειγμα	14
1.3.7 Παράδειγμα	14
1.3.8 Παράδειγμα	14
1.4 Περιοχή έλξης μεγίστων	15
1.4.1 Ορισμός	15
1.4.2 Πρόταση (Χαρακτηρισμός του $MDA(H)$)	15
1.4.3 Ορισμός (Ισοδυναμία ουρών)	15
1.4.4 Εξέταση της κατανομής Fréchet	16

1.4.5	Θεώρημα	16
1.4.6	Πόρισμα	17
1.4.7	Πρόταση (Ιδιότητα κλειστότητας του $MDA(\Phi_a)$)	17
1.4.8	Εξέταση της κατανομής Weibull	17
1.4.9	Θεώρημα	18
1.4.10	Πόρισμα	18
1.4.11	Πρόταση (Ιδιότητα κλειστότητας του $MDA(\Psi_a)$)	18
1.4.12	Εξέταση της κατανομής Gumbel	19
1.4.13	Ορισμός (Συνάρτησης Von Mises)	20
1.4.14	Παραδείγματα	20
1.4.15	Πρόταση (Ιδιότητες συνάρτησης Von Mises)	21
1.4.16	Πρόταση (Συνάρτηση Von Mises και $MDA(I)$)	21
1.4.17	Θεώρημα (Χαρακτηρισμός I του $MDA(I)$)	22
1.4.18	Θεώρημα (Χαρακτηρισμός II του $MDA(I)$)	22
1.4.19	Πρόταση (Ιδιότητα κλειστότητας του $MDA(I)$)	22
1.4.20	Παρατήρηση (Εμφύτευση του $MDA(\Phi_a)$ στο $MDA(I)$)	23
1.5	Γενικευμένη Οικογένεια Κατανομών Ακραίων Τιμών	23
1.5.1	Ορισμός Jenkinson – Von Mises	24
1.5.2	Θεώρημα (Χαρακτηρισμός του $MDA(H_\xi)$)	25
1.5.3	Ορισμός (Υπερβάλλουσα συνάρτηση κατανομής και μέση υπερβάλλουσα συνάρτηση)	25
1.5.4	Πρόταση (Υπολογισμός της μέσης υπερβάλλουσας συνάρτησης)	26
1.5.5	Ορισμός (Γενικευμένη Κατανομή Pareto)	26
1.6	Διατεταγμένες παρατηρήσεις	27
1.6.1	Πρόταση (Συνάρτηση κατανομής της k -οστής άνω διατεταγμένης τυχαίας μεταβλητής ή παρατήρησης)	28
1.6.2	Θεώρημα (Η από κοινού πυκνότητα των k άνω διατεταγμένων παρατηρήσεων)	29
1.6.3	Ορισμός (Διάκενο διατεταγμένων παρατηρήσεων – spacing)	29
1.6.4	Λήμμα (Μετασχηματισμός ποσοστιαίων σημείων)	30
1.6.5	Πρόταση (Σχεδόν βέβαιη σύγκλιση των διατεταγμένων παρατηρήσεων)	30
1.6.6	Θεώρημα (Οριακός νόμος για το πλήθος των υπερβαλλουσών παρατηρήσεων)	31
1.6.7	Θεώρημα (Οριακές πιθανότητες άνω διατεταγμένων παρατηρήσεων)	32
1.6.8	Πόρισμα (Οριακή κατανομή μιας διατεταγμένης παρατήρησης)	32
	Κεφάλαιο 2, Στατιστικές μέθοδοι ανάλυσης ακραίων παρατηρήσεων	33
2.1	Εισαγωγή	33
2.2	Διαγραμματικά εργαλεία	33
2.2.1	Παράδειγμα (QQ plot έναντι της κατανομής Gumbel)	36

2.2.2	Παράδειγμα (<i>QQ plot</i> έναντι της GEV)	36
2.3	Υπερβάλλουσα συνάρτηση κατανομής και μέση υπερβάλλουσα συνάρτηση	37
2.3.1	Παράδειγμα	38
2.4	Εκτίμηση παραμέτρων για τη GEV	39
2.4.1	Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας	39
2.4.1.1	Παράδειγμα	42
2.4.2	Στάθμη απόδοσης	42
2.4.2.1	Παράδειγμα	44
2.4.3	Εκτίμηση των υψηλών ποσοστιαίων σημείων μέσω της κατατομής πιθανοφάνειας (<i>Profile Likelihood</i>)	45
2.4.3.1	Παράδειγμα	46
2.4.4	Εκτίμηση των παραμέτρων κάτω από την υπόθεση του MDA	47
2.5	Εκτίμηση της παραμέτρου θέσης ξ	49
2.5.1	Εκτίμηση του ξ μέσω του εκτιμητή Pickands	49
2.5.1.1	Θεώρημα (Ιδιότητες του εκτιμητή Pickands)	50
2.5.1.2	Παράδειγμα	51
2.5.2	Εκτίμηση της παραμέτρου σχήματος $\xi = \alpha^{-1} > 0$ μέσω του εκτιμητή Hill	52
2.5.2.1	Παράδειγμα	53
2.5.2.2	Κατασκευή του εκτιμητή Hill μέσω της μεθόδου MLE	54
2.5.2.3	Κατασκευή του εκτιμητή Hill μέσω της ομαλής μεταβλητότητας	55
2.5.2.4	Κατασκευή του εκτιμητή Hill μέσω της μέσης υπερβάλλουσας συνάρτησης	56
2.5.2.5	Θεώρημα (Βασικές ιδιότητες του εκτιμητή Hill)	57
2.5.3	Εκτιμητής του $\xi \in \mathbf{R}$ των Dekkers – Einmahl – de Haan	57
2.5.3.1	Παράδειγμα	58
2.6	Εκτίμηση των σταθερών κανονικοποίησης	59
2.6.1	Λήμμα (Εμφύτευση του $\text{MDA}(H_\xi)$, $\xi \geq 0$ στο $\text{MDA}(A)$)	60
2.7	Εκτίμηση της ουράς και των ποσοστιαίων σημείων	61
2.7.1	Θεώρημα. (Εκτίμηση υψηλών ποσοστιαίων σημείων I)	62
2.7.2	Θεώρημα. (Εκτίμηση υψηλών ποσοστιαίων σημείων II)	63
2.8	Η μέθοδος των υπερβάσεων πάνω από ένα υψηλό κατώφλι	63
2.8.1	Προσαρμογή της GPD	63
2.8.1.1	Παράδειγμα	66
2.8.2	Εκτίμηση Μέγιστης Πιθανοφάνειας	68
2.8.2.1	Παράδειγμα	68
2.8.3	Αποτίμηση Κινδύνου Χαρτοφυλακίου και αναμενόμενο μέγεθος ζημίας	69
2.8.3.1	Παράδειγμα	70

Κεφάλαιο 3, Ανάλυση Δεδομένων	73
3.1 Εισαγωγή	73
3.2 Δείκτης Standard & Poor's 500	73
3.2.1 Υπολογισμός των αποδόσεων ενός δείκτη	73
3.2.2 Διαγραμματική εξέταση των δεδομένων	74
3.2.4 Εύρεση των Block Maxima και Minima	75
3.2.5 Προσαρμογή των αποδόσεων του S&P 500 στην GEV και εκτίμηση των παραμέτρων	78
3.2.6 Εκτίμηση της στάθμης απόδοσης	80
3.3 Εκτίμηση των παραμέτρων μέσω της υπόθεσης MDA	82
3.4 Ανάλυση Peaks Over Threshold	85
3.4.1 Αποτίμηση Κινδύνου Χαρτοφυλακίου και αναμενόμενο μέγεθος ζημίας	89
3.5 Συμπεράσματα	89
Παραρτήματα	91
Π1 Εντολές R	93
Βιβλιογραφία	101

Κατάλογος Πινάκων

2-1	Εκτιμώμενες παράμετροι με τη μέθοδο MLE	42
2-2	Εκτιμώμενες παράμετροι της GPD με τη μέθοδο MLE	69
2-3	Εκτιμήσεις VaR q και ES q των προσαρμοσμένων στην GPD Exp(1) και Cauchy(0,1)	71

ТАНЕЦЫ И ТЕАТР

Κατάλογος Σχημάτων

1-1	Διάγραμμα κανονικής κατανομής	1
2-1	<i>QQ plot</i> της GEV με παραμέτρους $\xi = 0, -0.4, 0.5, 0.7$ έναντι της Gumbel	36
2-2	<i>QQ plot</i> της GEV με παραμέτρους $\xi = 0, 0.1, -0.1, -0.3$ έναντι της GEV με παράμετρο $\xi = -0.4$	37
2-3	Διαγράμματα μέσης υπολοιπόμενης ζωής των $\text{Exp}(1)$, Pareto (1,7,2) και $U(0,1)$	38
2-4	Διαγράμματα στάθμης απόδοσης των προσαρμοσμένων στην GEV $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$	44
2-5	Διάγραμμα στάθμης απόδοσης της προσαρμοσμένης $\text{Exp}(1)$ και της προσαρμοσμένης $\text{Cauchy}(0,1)$	46
2-6	Διάγραμμα του εκτιμητή Pickands για τις διατεταγμένες παρατηρήσεις των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$	52
2-7	Διάγραμμα του εκτιμητή Hill για τις διατεταγμένες παρατηρήσεις των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$	54
2-8	Διάγραμμα του εκτιμητή DEdH για τις διατεταγμένες παρατηρήσεις των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$	59
2-9	Διαγράμματα μέσης υπολοιπόμενης ζωής των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$	67
2-10	<i>QQ plot</i> των υπερβάσεων κατωφλίου των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$ έναντι της εκθετικής κατανομής	67
3-1	Διάγραμμα των τιμών κλεισίματος και των λογαριθμικών αποδόσεων του S&P 500	74
3-2	<i>QQ plot</i> των λογαριθμικών αποδόσεων του δείκτη S&P 500 έναντι της κανονικής κατανομής	75
3-3	Διάγραμμα των Block maxima και Minima των λογαριθμικών αποδόσεων του δείκτη S&P 500	76
3-4	Ιστογράμματα συχνοτήτων των μεγίστων και των ελαχίστων τιμών των λογαριθμικών αποδόσεων του δείκτη S&P 500	77

3-5	<i>QQ plot</i> των μηνιαίων μεγίστων και ελαχίστων λογαριθμικών αποδόσεων του δείκτη S&P 500 έναντι της Gumbel	77
3-6	Εμπειρική σ.κ., <i>PP plot</i> , <i>QQ plot</i> και διάγραμμα πυκνότητας των μηνιαίων μεγίστων έναντι της εκτιμηθέντος μοντέλου	79
3-7	Εμπειρική σ.κ., <i>PP plot</i> , <i>QQ plot</i> και διάγραμμα πυκνότητας των μηνιαίων ελαχίστων έναντι της εκτιμηθέντος μοντέλου	80
3-8	Διαγράμματα στάθμης απόδοσης των μεγίστων για 12 και 24 μήνες	81
3-9	Διαγράμματα στάθμης απόδοσης των ελαχίστων για 12 και 24 μήνες	81
3-10	Διάγραμμα του εκτιμητή Hill για τις θετικές και αρνητικές αποδόσεις	82
3-11	Διαγράμματα των εκτιμητών Pickands, Hill, DEdH για τις θετικές αποδόσεις	83
3-12	Διαγράμματα των εκτιμητών Pickands, Hill, DEdH για τις αρνητικές αποδόσεις	84
3-13	Διάγραμμα μέση υπολοιπούμενης ζωής για τη πάνω και τη κάτω ουρά της κατανομής των λογαριθμικών αποδόσεων	85
3-14	Εκθετικά <i>QQ plot</i> πάνω και τη κάτω ουρά της κατανομής των λογαριθμικών αποδόσεων	86
3-15	Εμπειρική σ.κ., <i>PP plot</i> , <i>QQ plot</i> και διάγραμμα πυκνότητας της πάνω ουρά των λογαριθμικών αποδόσεων έναντι της εκτιμηθέντος μοντέλου	88
3-16	Εμπειρική σ.κ., <i>PP plot</i> , <i>QQ plot</i> και διάγραμμα πυκνότητας της κάτω ουρά των λογαριθμικών αποδόσεων έναντι της εκτιμηθέντος μοντέλου	88

Κατάλογος Συντομογραφιών

τ.μ.	τυχαία μεταβλητή	1
iid	identically and independently distributed	1
σ.κ.	συνάρτηση κατανομής	1
Κ.Ο.Θ.	Κεντρικό Οριακό Θεώρημα	3
MDA	Maximum Domain of Attraction	15
GEV	Generalized Extreme Value Distribution	23
GPD	Generalized Pareto Distribution	26
σ.π.	συνάρτηση πιθανότητας	28
τ.δ.	τυχαίο δείγμα	34
PP plot	Probability-Probability plot	34
QQ plot	Quantile-Quantile Plot	34
ME plot	Mean Excess plot	37
MRL plot	Mean Residual Life plot	37
MLE	Maximum Likelihood Estimation	40
δ.ε.	διάστημα εμπιστοσύνης	41
σ.π.π.	συνάρτηση πυκνότητας πιθανότητας	55
DEdH	Dekkers – Einmahl – de Haan	57
POT	Peaks Over Threshold	65
VaR	Value at Risk	69
ES	Expected Shortfall	70
S&P 500	Standard & Poor's 500	73

ТАНЕЦЫ И МУЗЫКА

ΚΕΦΑΛΑΙΟ 1

ΘΕΩΡΙΑ ΚΑΤΑΝΟΜΩΝ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

1.1 Εισαγωγή

1.1.1 Δυσκολίες στην ανάλυση ακραίων τιμών

Θεωρούμε ένα σύνολο παρατηρήσεων από ανεξάρτητες και ισόνομα κατανεμημένες (*iid*) τυχαίες μεταβλητές (τ.μ.) X_1, X_2, \dots, X_n που προέρχονται από μια άγνωστη συνάρτηση κατανομής (σ.κ.) F . Εκφράζουμε ως

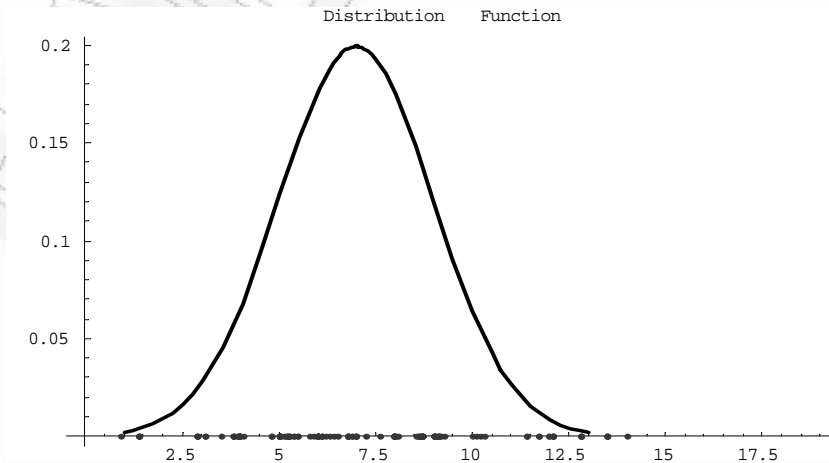
$$M_n = \max(X_1, X_2, \dots, X_n) \quad \text{και} \quad m_n = \min(X_1, X_2, \dots, X_n)$$

τη μέγιστη και την ελάχιστη τιμή των X_i τ.μ. αντίστοιχα. Για αυτές τις τιμές ισχύει ότι $m_n = -\max(-X_1, -X_2, \dots, -X_n)$, επομένως όλα τα αποτελέσματα που μπορούν να εξαχθούν για τα δειγματικά μέγιστα ανάγονται με βάση αυτόν τον μετασχηματισμό και για τα δειγματικά ελάχιστα. Η σ.κ. της τ.μ. M_n προκύπτει από τη σχέση

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

Το συγκεκριμένο μοντέλο δεν έχει ιδιαίτερη πρακτική αξία, δεδομένου ότι στη πλειονότητα των περιπτώσεων που εξετάζουμε, η κατανομή F είναι άγνωστη. Μια μέθοδος που μπορεί να εφαρμοσθεί, είναι να εκτιμηθεί η F από τις παρατηρήσεις που έχουμε και μετά να αντικατασταθεί η εκτίμηση στη παραπάνω σχέση. Η μέθοδος αυτή παρουσιάζει κάποια μειονεκτήματα.

ΣΧΗΜΑ 1-1



Η κυριότερη δυσκολία για την ακριβή εκτίμηση των ουρών της F , είναι ότι η πλειονότητα των παρατηρήσεων συγκεντρώνεται στο κέντρο της κατανομής ενώ (εξ' ορισμού) οι παρατηρήσεις στις ουρές – δηλαδή οι μέγιστες και οι ελάχιστες τιμές – είναι σπάνιες (ΣΧΗΜΑ 1-1). Υπάρχει δε το ενδεχόμενο να χρειάζεται εκτίμηση της $F(x)$ και για τιμές x που υπερβαίνουν τα δειγματικά μέγιστα ή ελάχιστα. Δεδομένης της μορφής της σ.κ. F^n , οποιοσδήποτε διαφορές ή αποκλίσεις υπάρχουν στην εκτίμηση της αρχικής σ.κ. F , μεγεθύνονται στην σ.κ. F^n και σε κάθε περίπτωση οι εκτιμήσεις που θα προκύψουν θα επηρεάζονται σε μεγάλο βαθμό από τις αρχικές μας υποθέσεις

Ας δούμε πρακτικά τι συμβαίνει αν προσαρμόσουμε σε όλες τις παρατηρήσεις ένα μοντέλο που προέρχεται από μια Κανονική $N(\mu, \sigma^2)$, και χρησιμοποιήσουμε για την εκτίμηση των ουρών τη συνάρτηση επιβίωσης της Κανονικής

$$\hat{P}(X > x) = 1 - \hat{F}(x) = 1 - \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right).$$

Αυτή η ενέργεια θα μας έφερνε αντιμέτωπους με τα ακόλουθα προβλήματα. Οι εκτιμήσεις που παίρνουμε για τις παραμέτρους μ και σ της κατανομής, επηρεάζονται κατά κύριο λόγο από τον όγκο των παρατηρήσεων που βρίσκονται στο κέντρο της. Οποιαδήποτε εκτίμηση της καλής προσαρμογής του μοντέλου θα επηρεάζεται και από τις κεντρικές τιμές. Είναι δυνατόν να προσδιοριστούν διαφορετικά μοντέλα – ίδια κατανομή με διαφορετικές παραμέτρους θέσεις και κλίμακας – που ενώ δίνουν καλή προσαρμογή στο κύριο σώμα των παρατηρήσεων, οδηγούν σε διαφορετικές εκτιμήσεις. Ενώ λοιπόν το ενδιαφέρον μας επικεντρώνεται στις ουρές της κατανομής, συμβιβάζομαστε με ένα μοντέλο που ταυτόχρονα με αυτές, δίνει εκτιμήσεις και για το κυρίως σώμα των παρατηρήσεων.

Απομένει να αποδεχθούμε ότι η F είναι άγνωστη και να αναζητήσουμε προσεγγιστικές οικογένειες μοντέλων για την F^n , την οποία θα μπορέσουμε να εκτιμήσουμε χρησιμοποιώντας μόνο ακραίες τιμές. Αυτή η διαδικασία είναι παρόμοια με την εκτίμηση της κατανομής των δειγματικών μέσων, μέσω της κανονικής κατανομής από το Κεντρικό Οριακό Θεώρημα (*Central Limit Theorem*).

Θεωρούμε ένα δείγμα iid τ.μ. X_1, X_2, \dots, X_n με κοινή σ.κ. F , και

$$E(X_i) = \mu, \quad Var(X_i) = \sigma^2 < \infty,$$

ο (πληθυσμιακός) μέσος και η (πληθυσμιακή) διακύμανση αντίστοιχα. Ως

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

συμβολίζουμε τον δειγματικό μέσο. Χρησιμοποιώντας τον νόμο των μεγάλων αριθμών, έχουμε ότι, καθώς $n \rightarrow \infty$,

$$\bar{X}_n \rightarrow \mu \text{ σχεδόν βέβαια.}$$

Επομένως η ασυμπτωτική κατανομή του \bar{X}_n είναι εκφυλισμένη (*degenerate*), δεδομένου ότι συγκλίνει σε ένα σημείο. Χρησιμοποιώντας το Κ.Ο.Θ., ξεπερνάμε αυτό το πρόβλημα που προκύπτει από τη συμπεριφορά της οριακής κατανομής του \bar{X}_n , ενσωματώνοντας έναν γραμμικό μετασχηματισμό, τέτοιο ώστε, για κάθε δεδομένο x , καθώς $n \rightarrow \infty$ να ισχύει

$$P\left(\frac{\bar{X}_n - \mu_n}{\sigma_n} \leq x\right) \rightarrow \Phi(x),$$

όπου

$$\mu_n = E(\bar{X}_n) = \mu, \quad \sigma_n = \sqrt{\text{Var}(\bar{X}_n)} = \sigma / \sqrt{n},$$

και Φ είναι η σ.κ. μιας τυπικής κανονικής τ.μ.. Παρατηρούμε ότι η οριακή και μη εκφυλισμένη κατανομή είναι ίδια ανεξαρτήτως της αρχικής κατανομής F (υπό τη προϋπόθεση ότι $\sigma^2 < \infty$) αν και ο τρόπος με τον οποίο θα κανονικοποιήσουμε εξαρτάται από τις παραμέτρους της αρχικής μας κατανομής. Μέσω της κανονικοποίησης παίρνουμε τη μεταβλητή $n^{1/2} (\bar{X}_n - \mu) / \sigma$ η οποία μεγεθύνει τις διαφορές μεταξύ πληθυσμιακού και δειγματικού μέσου. Από το Κ.Ο.Θ., καταλήγουμε στο μοντέλο $\bar{X}_n \sim N(\mu, \sigma^2 / n)$ για πεπερασμένο n . Το παραπάνω αποτελεί ένα ασυμπτωτικό μοντέλο για μεταβλητές που προέρχονται από αθροίσματα παρατηρήσεων. Επομένως από τη κανονική κατανομή προκύπτει μια ασυμπτωτική προσέγγιση της κατανομής τ.μ. της μορφής $\sum_{i=1}^n X_i$. Με παρόμοιο τρόπο θα εργαστούμε για να πάρουμε ασυμπτωτικές εκτιμήσεις για την κατανομή των δειγματικών μεγίστων ή ελαχίστων, εφαρμόζοντας τη θεωρία ακραίων τιμών.

1.1.2 Γραμμικός μετασχηματισμός της M_n

Θα προχωρήσουμε λοιπόν μελετώντας τη συμπεριφορά της F^n καθώς $n \rightarrow \infty$. Επειδή το ενδιαφέρον μας επικεντρώνεται στη μέγιστη τιμή ενός μεγάλου αριθμού μεταβλητών, αυτό υποδεικνύει ότι για να προκύψουν μοντέλα για τη συμπεριφορά της M_n θα πρέπει να κάνουμε ασυμπτωτικές εκτιμήσεις. Όπως αναφέραμε και παραπάνω, ελπίζουμε ότι θα προκύψει ένας μετασχηματισμός για τη κατανομή της M_n καθώς $n \rightarrow \infty$ ο οποίος θα έχει μικρό βαθμό εξάρ-

τησης από τη μορφή της αρχικής σ.κ. F . Επειδή οι ακραίες τιμές παρατηρούνται κοντά στο άνω φράγμα της F , διαισθητικά μπορούμε να υποστηρίξουμε ότι η ασυμπτωτική συμπεριφορά της τ.μ. M_n συνδέεται με τη δεξιά ουρά της F . Ορίζουμε ως

$$x_F = \sup \{x \in \mathbf{R} : F(x) < 1\} ,$$

το δεξιό οριακό σημείο της F . Προκύπτει ότι για κάθε $x < x_F$ ισχύει

$$P(M_n \leq x) = F^n(x) \rightarrow_{n \rightarrow \infty} 0$$

και όταν $x_F < \infty$ για κάθε $x \geq x_F$ ισχύει

$$P(M_n \leq x) = F^n(x) \rightarrow 1.$$

Επομένως ισχύει ότι, καθώς $n \rightarrow \infty$, $M_n \xrightarrow{P} x_F$, όπου $x_F \leq \infty$. Αφού η ακολουθία (M_n) είναι μη φθίνουσα ως προς το n , συγκλίνει σχεδόν βέβαια, το οποίο συνεπάγεται ότι

$$M_n \rightarrow_{n \rightarrow \infty} x_F, \text{ σχεδόν βέβαια.} \quad (1.1)$$

Επομένως ισχύει ότι

$$P(M_n \leq x) \rightarrow \begin{cases} 0 & \text{αν } x < x_F \\ 1 & \text{αν } x > x_F \end{cases}$$

Επομένως η ακολουθία (M_n) συγκλίνει σε μία εκφυλισμένη (*degenerate*) κατανομή, επειδή συγκλίνει σε ένα σημείο ή στο άπειρο σχεδόν βέβαια. Μπορούμε να παρακάμψουμε αυτή τη συμπεριφορά της M_n χρησιμοποιώντας ένα γραμμικό μετασχηματισμό της μορφής

$$M_n^* = \frac{M_n - d_n}{c_n}.$$

Χρησιμοποιούμε κατάλληλα επιλεγμένες πραγματικές σταθερές $c_n > 0$ και $d_n \in \mathbf{R}$, οι οποίες θα σταθεροποιήσουν τη θέση και τη κλίμακα της M_n^* καθώς το n αυξάνει ξεπερνώντας τις δυσκολίες που προκύπτουν από τη μελέτη της M_n .

Θεωρώντας ότι υπάρχουν κατάλληλα $c_n > 0$ και $d_n \in \mathbf{R}$, θα καταλήξουμε στη σχέση

$$c_n^{-1}(M_n - d_n) \xrightarrow{d}_{n \rightarrow \infty} H$$

για κάποια μη εκφυλισμένη κατανομή H . Επομένως θα μπορέσουμε να εξετάσουμε πιθανότητες της μορφής

$$P(c_n^{-1}(M_n - d_n) \leq x)$$

ή ισοδύναμα

$$P(M_n \leq u_n)$$

όπου $u_n = u_n(x) = c_n x + d_n$.

Ένα βασικό πρόβλημα είναι η εύρεση των οριακών νόμων (κατανομών) των κανονικοποιημένων δειγματικών μεγίστων M_n . Αρχικά μας ενδιαφέρει να προσδιορίσουμε ένα γενικό πλαίσιο για το είδος των μετασχηματισμών που θα χρησιμοποιήσουμε και στη συνέχεια να εκτιμήσουμε τις τιμές των σταθερών $c_n > 0$ και $d_n \in \mathbf{R}$.

Όπως προκύπτει και από το Κ.Ο.Θ. στην §1.1.1, για τυποποιημένες μεταβλητές που προέρχονται από αθροίσματα, υπάρχει μόνο ένας οριακός νόμος που δίνεται υπό τη μορφή της $\bar{X}_n \sim N(\mu, \sigma^2/n)$. Κάτι παρόμοιο υπάρχει και για τα δειγματικά μέγιστα.

1.2 Βασική θεωρία

1.2.1 Ορισμός (Σύγκλιση σε τύπο - *Convergence to Type*) (βλ. Resnick (1987)) Για δυο τ.μ. X και Y γράφουμε $X \stackrel{d}{=} Y$ αν X, Y έχουν την ίδια κατανομή. Θα λέμε ότι οι X και Y (καθώς και οι αντίστοιχες κατανομές τους) ανήκουν στον ίδιο τύπο ή είναι του ίδιου τύπου αν υπάρχουν σταθερές $a \in \mathbf{R}$ και $b > 0$ τέτοιες ώστε

$$X \stackrel{d}{=} bY + a.$$

Αν υποθέσουμε ότι έχουμε δυο σ.κ. $F(x)$ και $G(x)$, τότε αυτές θα καλούνται του ίδιου τύπου αν για κάποια $A > 0, B \in \mathbf{R}$ ισχύει

$$G(x) = F(Ax + B), \text{ για κάθε } x.$$

Το ακόλουθο αποτέλεσμα είναι πολύ χρήσιμο στην ασθενή σύγκλιση, γιατί μας λει ότι ο οριακός νόμος μιας ακολουθίας τ.μ. ορίζεται με μοναδικό τρόπο όσον αφορά μεταβολές στη θέση και τη κλίμακα. Το γεγονός αυτό στη γενίκευσή του αποτελεί ένα πολύ χρήσιμο εργαλείο για όσα θα αναπτύξουμε στη συνέχεια.

1.2.1.2 Πρόταση. Ας υποθέσουμε ότι έχουμε F και G δυο μη εκφυλισμένες κατανομές. Υποθέτουμε ότι έχουμε μια ακολουθία κατανομών $U_n, n \geq 1$. Αν επιλέξουμε ακολουθίες $a_n \geq 0, b_n \in \mathbf{R}, \alpha_n > 0, \beta_n \in \mathbf{R}$ και προχωρήσουμε σε μετασχηματισμό ώστε η U_n να συγκλίνει ασθενώς στις παραπάνω κατανομές,

$$U_n(a_n x + b_n) \rightarrow F(x), \quad U_n(\alpha_n x + \beta_n) \rightarrow G(x), \quad (1.2)$$

τότε ισχύει ότι

$$\alpha_n / a_n \rightarrow A > 0, \quad (\beta_n - b_n) / a_n \rightarrow B \in \mathbf{R} \quad (1.3)$$

και

$$G(x) = F(Ax + B), \quad x \in \mathbf{R}. \quad (1.4)$$

Αν η (1.3) ισχύει, τότε οποιαδήποτε από τις δυο σχέσεις της (1.2) συνεπάγεται την άλλη και ταυτόχρονα ισχύει η (1.4).

1.2.1.3 Ορισμός (Γενικευμένη αντίστροφος μιας συνάρτησης κατανομής) Η γενικευμένη αντίστροφος μιας σ.κ. F ορίζεται

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

και καλείται γενικευμένη αντίστροφος ή συνάρτηση ποσοστιαίων σημείων (*quantile function*) της σ.κ. F . Η ποσότητα $x_t = F^{\leftarrow}(t)$ ορίζει το t -ποσοστιαίο σημείο (t -quantile) της F .

1.2.1.4 Πρόταση. (Σύγκλιση της γενικευμένης αντιστρόφου) Έστω h, h_1, h_2, \dots μη φθίνουσες συναρτήσεις τέτοιες ώστε $\lim_{n \rightarrow \infty} h_n(x) = h(x)$ για κάθε σημείο συνέχειας της h . Τότε

$$\lim_{n \rightarrow \infty} h_n^{\leftarrow}(y) = h^{\leftarrow}(y)$$

για κάθε σημείο συνέχειας της h^{\leftarrow} .

Ομαλά μεταβαλλόμενες (*regularly varying*) συναρτήσεις καλούνται εκείνες που στο δεξί άκρο τους (ασυμπτωτικά) συμπεριφέρονται περίπου σαν συναρτήσεις της μορφής $c \cdot x^\rho$. Πιο συγκεκριμένα δίνεται ο ακόλουθος ορισμός (βλ. de Haan (1970), Feller (1971), Bingham et al. (1987)).

1.2.3 Ορισμός (Ομαλά μεταβαλλόμενες συναρτήσεις) Μια μετρήσιμη συνάρτηση $F: \mathbf{R}_+ \rightarrow \mathbf{R}_+$ θα λέγεται ομαλά μεταβαλλόμενη στο ∞ με δείκτη ρ και θα γράφουμε $F \in \mathfrak{R}_\rho$, αν για $x > 0$ ισχύει

$$\lim_{x \rightarrow \infty} F(tx)/F(x) = t^\rho.$$

Το ρ καλείται εκθέτης ή δείκτης μεταβολής. Επίσης θα λέμε ότι η $F(x^{-1})$ είναι ομαλά μεταβαλλόμενη στο 0 αν και μόνο αν η $F(x)$ είναι ομαλά μεταβαλλόμενη στο ∞ .

Αν $\rho = 0$, τότε η F καλείται βραδέως μεταβαλλόμενη (*slowly varying*) και γενικά συμβολίζεται με $L(x)$. Συναρτήσεις βραδείας κύμανσης είναι οι σταθερές συναρτήσεις, οι συναρτήσεις που συγκλίνουν σε μια σταθερά (στο ∞), και γενικότερα συναρτήσεις που μεταβάλλονται «πολύ αργά» στο ∞ όπως η λογαριθμική συνάρτηση.

Αν η $F \in \mathfrak{R}_\rho$, τότε $F(x) / x^\rho \in \mathfrak{R}_0$ και θέτοντας $L(x) = F(x) / x^\rho$, προκύπτει ότι είναι πάντα δυνατό να συμβολίσουμε μια ρ -μεταβαλλόμενη συνάρτηση (ρ -varying function) ως $x^\rho L(x)$.

1.2.4 Θεώρημα. (Ομοιόμορφη σύγκλιση - Uniform convergence) Αν $h \in \mathfrak{R}_\rho$ τότε για $0 < \rho \leq b < \infty$, το ακόλουθο όριο

$$\lim_{x \rightarrow \infty} h(tx)/h(x) = t^\rho$$

συγκλίνει ομοιόμορφα ως προς το x . Αν $\rho = 0$, τότε θα συγκλίνει σε κάθε διάστημα $[\rho, b]$, αν $\rho < 0$ τότε θα συγκλίνει σε κάθε διάστημα $[\rho, \infty)$ ενώ αν $\rho > 0$ θα συγκλίνει σε κάθε διάστημα $(0, b]$ κάτω από την υπόθεση ότι η h είναι φραγμένη σε κάθε διάστημα της μορφής $(0, x]$, $x > 0$.

1.2.5 Θεώρημα. (Αναπαράσταση των ομαλά μεταβαλλόμενων συναρτήσεων). Αν $h \in \mathfrak{R}_\rho$ για κάποιο $\rho \in \mathbf{R}$, τότε

$$h(x) = c(x) \exp \left\{ \int_z^x \frac{\delta(u)}{u} du \right\}, \quad x \geq z \quad (1.5)$$

για κάποιο $z > 0$, όπου c και δ είναι μετρήσιμες συναρτήσεις, τέτοιες ώστε $c(x) \rightarrow c_0 \in (0, \infty)$, $\delta(x) \rightarrow \rho$ καθώς $x \rightarrow \infty$.

Ισχύει και το αντίστροφο του παραπάνω αποτελέσματος. Ο παραπάνω τρόπος αναπαράστασης μιας ομαλά μεταβαλλόμενη συνάρτησης, καλείται αναπαράσταση Karamata (Karamata (1933)) (*Karamata representation*). Ως άμεσο αποτέλεσμα προκύπτει ότι αν $h \in \mathfrak{R}_\rho$ για κάποιο $\rho \neq 0$, τότε καθώς $x \rightarrow \infty$

$$h(x) \rightarrow \begin{cases} \infty, & \rho > 0, \\ 0, & \rho < 0. \end{cases}$$

Χάριν των ιδιοτήτων της ολοκλήρωσης, μια ρ -μεταβαλλόμενη συνάρτηση συμπεριφέρεται σαν μια συνάρτηση της μορφής x^ρ . Από το επόμενο θεώρημα προκύπτει ότι τα ολοκληρώματα ομαλά μεταβαλλόμενων συναρτήσεων, είναι ομαλά μεταβαλλόμενες συναρτήσεις. Συμβολίζουμε με C^∞ το χώρο των απείρως διαφορίσιμων συναρτήσεων.

1.2.6 Θεώρημα (Το Θεώρημα του Karamata). Έστω ότι υπάρχει $L \in \mathfrak{R}_0$, τοπικά φραγμένη στο $[x_0, \infty)$ για κάποιο $x_0 \geq 0$. Τότε

$$(α) \quad \int_{x_0}^x t^\rho L(t) dt \sim (\rho+1)^{-1} x^{\rho+1} L(x), \text{ για } \rho > -1, x \rightarrow \infty,$$

$$(β) \quad \int_x^\infty t^\rho L(t) dt \sim -(\rho+1)^{-1} x^{\rho+1} L(x), \text{ για } \rho < -1, x \rightarrow \infty.$$

1.2.7 Ορισμός. (Ταχέως μεταβαλλόμενες συναρτήσεις - rapid variation) Μια θετική, μετρήσιμη συνάρτηση h στο $(0, \infty)$ καλείται ταχέως μεταβαλλόμενη με δείκτη $-\infty$ (συμβολίζεται $h \in \mathfrak{R}_{-\infty}$), αν

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

Μια τέτοια συνάρτηση είναι η $h(x) = e^{-x}$.

1.2.8 Θεώρημα. (Ιδιότητες των ταχέως μεταβαλλόμενων συναρτήσεων) (α) Αν η $h \in \mathfrak{R}_{-\infty}$ είναι μη-αύξουσα συνάρτηση, τότε για κάποιο $z > 0$ και για όλα τα $\rho \in \mathbf{R}$,

$$\int_z^\infty t^\rho h(t) dt < \infty \quad \text{και} \quad \lim_{x \rightarrow \infty} \frac{x^{\rho+1} h(x)}{\int_z^\infty t^\rho h(t) dt} = \infty. \quad (1.6)$$

Αντίστροφα, αν για κάποιο $\rho \in \mathbf{R}$, ισχύει η (1.6) και $\int_1^\infty t^\rho h(t) dt < \infty$, τότε $h \in \mathfrak{R}_{-\infty}$.

(β) Αν $h \in \mathfrak{R}_{-\infty}$, και υπάρχουν συναρτήσεις c και δ , τέτοιες ώστε $c(x) \rightarrow c_0 \in (0, \infty)$, $\delta(x) \rightarrow \rho$ καθώς $x \rightarrow \infty$ τότε για κάποιο $z > 0$, καταλήγουμε στη σχέση (1.5). Ισχύει και το αντίστροφο.

1.2.9. Ορισμός. (Ομαλά μεταβαλλόμενες ακολουθίες) (βλ. Bingham et al. (1987)) Μια ακολουθία (c_n) θετικών αριθμών καλείται ομαλά μεταβαλλόμενη με δείκτη $\rho \in \mathbf{R}$ αν

$$\lim_{n \rightarrow \infty} \frac{c_{[tn]}}{c_n} = t^\rho, t > 0.$$

Αν η (c_n) είναι ομαλά μεταβαλλόμενη με δείκτη ρ , τότε $c(x) = c_{[x]}$ ανήκει στο \mathfrak{R}_ρ . Οι ιδιότητες που χαρακτηρίζουν τις ομαλά μεταβαλλόμενες συναρτήσεις ισχύουν και για τις ακολουθίες.

1.3 Οριακές ιδιότητες κατανομών ακραίων τιμών

Μας ενδιαφέρει για κατάλληλα επιλεγμένες σταθερές $c_n > 0$ και $d_n \in \mathbf{R}$ να δούμε ποιες συνθήκες πρέπει να ικανοποιεί η F για να θεωρήσουμε βέβαιη την ύπαρξη του ορίου

$$\lim_{n \rightarrow \infty} P(M_n \leq c_n x + d_n) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n).$$

1.3.1 Πρόταση. (Προσέγγιση Poisson, Poisson approximation) Αν θεωρήσουμε δεδομένο $\tau \in [0, \infty]$ και μια ακολουθία (u_n) πραγματικών αριθμών, τα ακόλουθα είναι ισοδύναμα

$$n\bar{F}(u_n) \rightarrow \tau \quad (1.7)$$

και

$$P(M_n \leq u_n) \rightarrow e^{-\tau}. \quad (1.8)$$

Απόδειξη. Θα αποδείξουμε την ισοδυναμία των σχέσεων εξετάζοντας δυο περιπτώσεις για το τ . Θα επιλέξουμε $0 \leq \tau < \infty$. Αν ισχύει η (1.7) τότε

$$P(M_n \leq u_n) = F^n(u_n) = \left(1 - \bar{F}(u_n)\right)^n = \left(1 - \frac{\tau}{n} - o\left(\frac{1}{n}\right)\right)^n$$

από όπου συνεπάγεται η (1.8). Αντίστροφα, αν ισχύει η (1.8) τότε προκύπτει

$$-\ln P(M_n \leq u_n) = -n \ln F(u_n) = -n \ln(1 - \bar{F}(u_n)) \rightarrow \tau.$$

Ισχύει ότι

$$-\ln(1-x) \sim x, \quad x \rightarrow 0$$

βάσει του οποίου καταλήγουμε ότι

$$n\bar{F}(u_n) = \tau + o(1),$$

όπου $o(1)$ συγκλίνει στο μηδέν, και επομένως οδηγούμαστε στη σχέση (1.7).

Αν επιλέξουμε $\tau = \infty$ και θεωρήσουμε ότι ισχύει η σχέση (1.7), αλλά όχι η (1.8), τότε θα πρέπει να υπάρχει μια υπό-ακολουθία (n_k) τέτοια ώστε

$$P(M_{n_k} \leq u_{n_k}) \rightarrow e^{-\tau'}$$

καθώς $k \rightarrow \infty$ και για κάποιο $\tau' < \infty$. Αλλά σε αυτή τη περίπτωση η σχέση (1.8) προϋποθέτει την ισχύ της (1.7) και επομένως καταλήγουμε ότι

$$n_k \bar{F}(u_{n_k}) \rightarrow \tau' < \infty.$$

Το παραπάνω αναιρεί την αρχική μας υπόθεση ότι $\tau = \infty$. Ομοίως αν ισχύει η (1.8) οδηγούμαστε στην (1.7) για $\tau = \infty$. \square

Προηγουμένως είδαμε στη (1.1) ότι η M_n συγκλίνει σχεδόν βέβαια στο δεξιό οριακό σημείο x_F της σ.κ. F , και είχαμε οδηγηθεί στη δίτιμη οριακή συμπεριφορά της $P(M_n \leq x)$. Γενικά, μέσω της προσέγγισης Poisson μπορεί να αποδειχθεί ότι αν για $x_F < \infty$ ισχύει ότι

$$1 - F(x_{F^-}) = \bar{F}(x_{F^-}) = F(x_F) - F(x_{F^-}) > 0$$

τότε αν υπάρχει κάποια ακολουθία (u_n) τέτοια ώστε να ισχύει $P(M_n \leq u_n) \rightarrow \rho$, το ρ θα είναι υποχρεωτικά 0 ή 1. Το αποτέλεσμα αυτό δείχνει ότι για μια σ.κ. που παρουσιάζει άλμα στο δεξιό πεπερασμένο σημείο της, δε μπορεί να υπάρξει κάποια μη εκφυλισμένη οριακή κατανομή για τη M_n , ανεξάρτητα από τον τρόπο που θα γίνει η κανονικοποίηση. Το ίδιο αποτέλεσμα προκύπτει και για κάποιες κατανομές με άπειρο δεξιό οριακό σημείο. Άμεσο συμπέρασμα των παραπάνω είναι ότι η F πρέπει να είναι συνεχής.

Βεβαίως αυτό δεν αποκλείει το γεγονός ότι και για κάποιες διακριτές κατανομές μπορεί να υπάρχει σύγκλιση. Αν υποθέσουμε ότι έχουμε μια σ.κ. F με δεξιό οριακό σημείο $x_F \leq \infty$ και θεωρήσουμε ότι υπάρχει ένα $\tau \in (0, \infty)$, τότε υπάρχει μια ακολουθία (u_n) που ικανοποιεί τη σχέση $n\bar{F}(u_n) \rightarrow \tau$ αν και μόνο αν

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{F(x^-)} = 1.$$

Το συγκεκριμένο αποτέλεσμα εφαρμόζεται κυρίως σε διακριτές κατανομές με μη πεπερασμένα δεξιά οριακά σημεία. Αν το μέγεθος των αλμάτων της συνάρτησης κατανομής δε φθίνει ικανοποιητικά γρήγορα, τότε δε μπορεί να υπάρξει μια μη φθίνουσα οριακή κατανομή για τα μέγιστα. Φαίνεται ότι υπάρχει μια σύνθετη ασυμπτωτική συμπεριφορά για τα M_n , και ότι η διακριτότητα μιας κατανομής μπορεί να μην οδηγήσει σε σύγκλιση των μεγίστων. Έστω ότι εξετάζουμε τη κατανομή Poisson, δηλαδή

$$P(X = k) = e^{-\lambda} \lambda^k / k!, k \in \mathbb{N}_0, \lambda > 0.$$

Τότε:

$$\frac{\bar{F}(k)}{\bar{F}(k-1)} = 1 - \frac{F(k) - F(k-1)}{\bar{F}(k-1)} = 1 - \frac{\lambda^k}{k! \left(\sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \right)^{-1}} = 1 - \left(1 + \sum_{r=k+1}^{\infty} \frac{k!}{r!} \lambda^{r-k} \right)^{-1}.$$

Αν στο τελευταίο άθροισμα θέσουμε όπου $r - k = s$ τότε προκύπτει ότι

$$\sum_{s=1}^{\infty} \frac{k!}{(s-k)!} \lambda^s = \sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1)(k+2)\dots(k+s)} \leq \sum_{s=1}^{\infty} \left(\frac{\lambda}{k}\right)^s = \frac{\lambda/k}{1-\lambda/k}, k > \lambda.$$

Η τελευταία σχέση τείνει στο μηδέν καθώς $k \rightarrow \infty$ και επομένως καταλήγουμε ότι

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

Από τα παραπάνω προκύπτει ότι δεν υπάρχει μη εκφυλισμένη οριακή κατανομή για τα μέγιστα της κατανομής Poisson.

1.3.2 Κατανομές αναλλοίωτου μεγίστου (*max-stable distributions*)

Το αντικείμενο της μελέτης μας είναι να εκτιμήσουμε τις οριακές κατανομές των μεγίστων M_n αφού έχουν κανονικοποιηθεί μέσω κάποιου κατάλληλου γραμμικού μετασχηματισμού. Κατά την παρουσίαση του Κ.Ο.Θ. στην § 1.1.1 είδαμε ότι η οριακή κατανομή για μεταβλητές που προέρχονται από αθροίσματα παρατηρήσεων είναι μοναδική και είναι η Κανονική κατανομή η οποία είναι αναλλοίωτη ως προς αθροίσματα (*stable ή mean-stable distribution*). Η αντίστοιχη κατανομή για τα μέγιστα M_n^* έχει αποδειχθεί ότι ανήκει σε μια γενικότερη οικογένεια κατανομών, που αποτελείται από τις κατανομές αναλλοίωτου μεγίστου.

1.3.3 Ορισμός. (*Κατανομές Αναλλοίωτου Μεγίστου, max-stable distributions*) Αναλλοίωτου μεγίστου καλούμε μια μη εκφυλισμένη τ.μ. X (ή την αντίστοιχη σ.κ. F) αν ικανοποιεί τη σχέση

$$\max(X_1, X_2, \dots, X_n) \stackrel{d}{=} c_n X + d_n$$

για *i.i.d* τ.μ. $X_1, X_2, \dots, X_n \sim F$, με κατάλληλες σταθερές $c_n > 0$ και $d_n \in \mathbf{R}$ για κάθε $n \geq 2$.

Η παραπάνω ιδιότητα ικανοποιείται από κατανομές από τις οποίες, αν αντλήσουμε δειγματικά μέγιστα, θα καταλήξουμε στην ίδια κατανομή μετά από κατάλληλη μεταβολή στις παραμέτρους θέσης και κλίμακας. Ας υποθέσουμε ότι (X_n) είναι μια ακολουθία από iid τ.μ. από μία κατανομή αναλλοίωτου μεγίστου, τότε μπορούμε να γράψουμε τη παραπάνω σχέση υπό τη μορφή

$$c_n^{-1}(M_n - d_n) \stackrel{d}{=} X. \quad (1.9)$$

Επομένως από το παραπάνω προκύπτει ότι κάθε κατανομή αναλλοίωτου μεγίστου είναι μια οριακή κατανομή για τα μέγιστα από iid τ.μ.. Οι κατανομές αναλλοίωτου μεγίστου αποτελούν το μοναδικό οριακό νόμο για τα μέγιστα M_n^* .

1.3.4 Θεώρημα. (Οριακή ιδιότητα των νόμων αναλλοίωτου μεγίστου) Η οικογένεια των κατανομών αναλλοίωτου μεγίστου συμπίπτει με την οικογένεια όλων των δυνατών μη εκφυλισμένων οριακών νόμων για κατάλληλα κανονικοποιημένα μέγιστα από iid τ.μ..

Απόδειξη. (βλ. Gnedenko (1943)) Για να αποδειχθεί η συγκεκριμένη πρόταση, αρκεί να δειχθεί ότι η οριακή κατανομή των κατάλληλα κανονικοποιημένων μεγίστων είναι μια κατανομή αναλλοίωτου μεγίστου. Υποθέτουμε ότι για κατάλληλες μεταβλητές κανονικοποίησης $c_n > 0$ και $d_n \in \mathbf{R}$ ισχύει

$$F_n(c_n x + d_n) = F^n(c_n x + d_n) \rightarrow H(x), \quad n \rightarrow \infty, \quad x \in \mathbf{R}$$

για κάποια μη εκφυλισμένη κατανομή H . Επομένως για κάθε $k > 0$, θα ισχύει η ισότητα

$$\lim_{n \rightarrow \infty} F^{nk}(c_n x + d_n) = (\lim_{n \rightarrow \infty} F^n(c_n x + d_n))^k = H^k(x), \quad x \in \mathbf{R}.$$

Επιπλέον θα ισχύει ότι

$$\lim_{n \rightarrow \infty} F^{nk}(c_{nk} x + d_{nk}) = H(x), \quad x \in \mathbf{R}.$$

Από την ισότητα αυτή συμπεραίνουμε ότι αν το k είναι σταθερό και $n \rightarrow \infty$, τότε η ακολουθία των συναρτήσεων $F^n(c_{nk} x + d_{nk})$ συγκλίνει σε μια οριακή συνάρτηση. Από τον Ορισμό 1.2.1 και τη Πρόταση 1.2.1.2 είδαμε ότι υπάρχουν σταθερές $\tilde{c}_k > 0$ και $\tilde{d}_k \in \mathbf{R}$ τέτοιες ώστε να ισχύει

$$\lim_{n \rightarrow \infty} \frac{c_{nk}}{c_n} = \tilde{c}_k \quad \text{και} \quad \lim_{n \rightarrow \infty} \frac{d_{nk} - d_n}{c_n} = \tilde{d}_k,$$

τότε για iid τ.μ. Y_1, \dots, Y_k με σ.κ. H θα ισχύει

$$\max(Y_1, \dots, Y_k) \stackrel{d}{=} \tilde{c}_k Y_1 + \tilde{d}_k.$$

□

Πάνω σε αυτή την οριακή ιδιότητα στηρίχθηκαν οι Fisher and Tippett (1928) για να δώσουν το κεντρικό θεώρημα για τους οριακούς νόμους των μεγίστων.

1.3.5 Θεώρημα. (Θεώρημα Fisher – Tjppett, Οριακός Νόμος των Μεγίστων) Έστω (X_n) μια ακολουθία iid τ.μ. Αν υπάρχουν κάποιες σταθερές κανονικοποίησης $c_n > 0$ και $d_n \in \mathbf{R}$ και μια κατανομή, H , μη εκφυλισμένη, τέτοια ώστε να ισχύει

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H, \quad (1.10)$$

τότε η κατανομή H θα είναι του ίδιου τύπου με μια από τρεις ακόλουθες σ.κ.:

Τύπος I: Fréchet

$$\Phi_a(x) = \begin{cases} 0, & \text{αν } x \leq 0 \\ \exp(-x^{-a}), & \text{αν } x > 0 \end{cases} \quad a > 0.$$

Τύπος II: Weibull

$$\Psi_a(x) = \begin{cases} \exp(-(-x)^a), & \text{αν } x \leq 0 \\ 1, & \text{αν } x > 0 \end{cases} \quad a > 0$$

Τύπος III: Gumbel

$$\Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbf{R}.$$

Οι κατανομές αυτές, καλούνται τυπικές κατανομές ακραίων τιμών. Ενώ ο τρόπος με τον οποίο θα εργαστούμε πάνω σε κάθε μια από τις παραπάνω κατανομές είναι διαφορετικός, αποδεικνύεται ότι συνδέονται με στενές μαθηματικές σχέσεις. Αν υποθέσουμε ότι $X > 0$, τότε

$$X \text{ έχει σ.κ. } \Phi_a \Leftrightarrow \ln X^a \text{ έχει σ.κ. } \Lambda \Leftrightarrow -X^{-1} \text{ έχει σ.κ. } \Psi_a.$$

Το καταπληκτικό αποτέλεσμα που προκύπτει είναι ότι αυτές οι κατανομές αποτελούν και τα μοναδικά όρια σύγκλισης των εμπειρικών κατανομών των δειγματικών μεγίστων. Γενικά ισχύει ότι η οικογένεια των οριακών συναρτήσεων για τη $F_n(c_n x + d_n)$ όπου $c_n > 0$ και $d_n \in \mathbf{R}$ είναι κατάλληλα επιλεγμένες σταθερές, δε μπορεί παρά να περιέχεται σε μια από τις οικογένειες που προσδιορίζουν οι τύποι I, II και III, επομένως οποιαδήποτε ακραία τυχαία μεταβλητή X θα ικανοποιεί πάντα τη (1.9).

Αν συνδυάσουμε τη οριακή ιδιότητα των νόμων αναλλοίωτου μεγίστου με το αποτέλεσμα των Fisher και Tjppett, τότε στη πράξη ερχόμαστε αντιμέτωποι με μια από τις τρεις ακόλουθες περιπτώσεις:

$$\text{Fréchet: } M_n \stackrel{d}{=} n^{1/a} X$$

$$\text{Weibull: } M_n \stackrel{d}{=} n^{-1/a} X$$

$$\text{Gumbel: } M_n \stackrel{d}{=} X + \log n$$

Ας δούμε στη πράξη τι μπορεί να συμβεί με τη χρήση ενός γραμμικού μετασχηματισμού της τ.μ. M_n .

1.3.6. Παράδειγμα. Υποθέτουμε ότι έχουμε μια ακολουθία iid τ.μ. X_1, X_2, \dots, X_n που ακολουθούν την εκθετική κατανομή $X_i \sim \text{Exp}(1)$, $i = 1, 2, \dots, n$ με σ.κ. $F(x) = 1 - e^{-x}$, $x > 0$. Για να βρούμε την οριακή κατανομή της M_n θα θεωρήσουμε $c_n = 1$ και $d_n = \log n$ καταλήγοντας σε έναν μετασχηματισμό της μορφής $M_n - \log n$, τότε

$$\begin{aligned} P(M_n - \log n \leq z) &= P(M_n \leq z + \log n) = (F(z + \log n))^n \\ &= (1 - \exp(-z - \log n))^n, \text{ για } z > -\log n \\ &= (1 - \exp(-z)\exp(-\log n))^n = (1 - \exp(-z)/n)^n \\ &\rightarrow \exp[\exp(-z)] \end{aligned}$$

καθώς $n \rightarrow \infty$, για $-\infty < z < \infty$. Η παραπάνω κατανομή ανήκει στην οικογένεια κατανομών Gumbel.

1.3.7. Παράδειγμα. Ας υποθέσουμε τώρα ότι έχουμε μια ακολουθία iid τ.μ. X_1, X_2, \dots, X_n που ακολουθούν την τυπική κατανομή Fréchet με σ.κ. $F(x) = \exp(-1/x)$, $x > 0$, για να βρούμε την οριακή κατανομή της M_n θα θεωρήσουμε $c_n = n$ και $d_n = 0$, τότε

$$P((M_n / n) \leq z) = P(M_n \leq nz) = (F(nz))^n = (\exp(-1/nz))^n = \exp(-1/z)$$

καθώς $n \rightarrow \infty$, για κάθε $z > 0$. Η οριακή κατανομή σε αυτή τη περίπτωση είναι η κατανομή Fréchet.

1.3.8. Παράδειγμα. Ας υποθέσουμε ότι έχουμε μια ακολουθία iid τ.μ. X_1, X_2, \dots, X_n που ακολουθούν την ομοιόμορφη κατανομή $X_i \sim U(0,1)$ με σ.κ. $F(x) = x$, $0 \leq x \leq 1$. Για να βρούμε την οριακή κατανομή της M_n , θα υποθέσουμε ότι για κάποιο σταθερό $z < 0$, ισχύει $n > -z$ και θα επιλέξουμε $c_n = 1/n$ και $d_n = 1$, τότε

$$P((M_n - 1)n \leq z) = P(M_n \leq n^{-1}z + 1) = (F(n^{-1}z + 1))^n \rightarrow e^z \text{ καθώς } n \rightarrow \infty$$

Επομένως η οριακή κατανομή σε αυτή τη περίπτωση είναι η Weibull.

1.4 Περιοχή Έλξης των Μεγίστων (Maximum Domain of Attraction - MDA)

Εξετάζοντας λοιπόν μια κατανομή ακραίων τιμών, H , θέλουμε να βρούμε τις συνθήκες που θα πρέπει να ικανοποιεί η σ.κ. F , ώστε τα κανονικοποιημένα μέγιστα της να συγκλίνουν ασθενώς στο H . Το ερώτημα αυτό μας οδηγεί στη συνέχεια στο να αναζητήσουμε τρόπο επίλογής των μεταβλητών κανονικοποίησης.

1.4.1 Ορισμός. Μια κατανομή F (ή μια τ.μ. X) θα λέμε ότι ανήκει στη περιοχή έλξης των μεγίστων της κατανομής ακραίων τιμών H και θα συμβολίζεται με $F \in \text{MDA}(H)$, αν υπάρχουν σταθερές $c_n > 0$, $d_n \in \mathbf{R}$ τέτοιες ώστε να ισχύει η σχέση (1.10).

Ισοδύναμα και βάσει της υπόθεσης ότι οι κατανομές ακραίων τιμών είναι συνεχείς στο \mathbf{R} , το παραπάνω μπορεί να αποδοθεί και από την ακόλουθη σχέση

$$\lim_{n \rightarrow \infty} P(M_n \leq c_n x + d_n) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x), x \in \mathbf{R}$$

1.4.2 Πρόταση. (Χαρακτηρισμός του $\text{MDA}(H)$) H σ.κ. F ανήκει στη περιοχή έλξης των μεγίστων της κατανομής ακραίων τιμών H με σταθερές κανονικοποίησης $c_n > 0$, $d_n \in \mathbf{R}$ αν και μόνο αν ισχύει ότι

$$\lim_{n \rightarrow \infty} n \overline{F}(c_n x + d_n) = -\ln H(x), x \in \mathbf{R}.$$

Στην περίπτωση που ισχύει $H(x) = 0$, το παραπάνω όριο θα τείνει στο ∞ .

Για κάθε μια από τις κατανομές ακραίων τιμών που περιγράφονται στο Θεώρημα 1.3.5 θα πρέπει να δοθεί ένας χαρακτηρισμός του MDA , δηλαδή να προσδιοριστούν οι σταθερές $c_n > 0$, $d_n \in \mathbf{R}$.

Επειδή η διαδικασία προσδιορισμού των παραμέτρων κανονικοποίησης είναι αρκετά επίπονη, χρησιμοποιούνται κάποιες τεχνικές που την απλοποιούν. Ένα πολύ χρήσιμο εργαλείο που χρησιμοποιείται σε μεγάλο βαθμό στην ανάλυση κατανομών ακραίων τιμών είναι η ισοδυναμία των ουρών των κατανομών (βλ. Von Mises (1936)).

1.4.3 Ορισμός. (Ισοδυναμία ουρών – Tail equivalence) Θα λέμε ότι δυο κατανομές F και G είναι ισοδύναμων ουρών αν έχουν το ίδιο δεξιό οριακό σημείο x_F και για κάποιο $A > 0$ ισχύει

$$\lim_{x \rightarrow x_F} (1 - F(x)) / (1 - G(x)) = A.$$

Είναι εύκολο να δειχθεί ότι αν δύο σ.κ. έχουν ισοδύναμη ουρά τότε ανήκουν στην ίδια περιοχή έλξης μιας κατανομής ακροτάτων χρησιμοποιώντας μάλιστα τις ίδιες σταθερές c_n, d_n . Επομένως μπορούμε να αλλάξουμε τη σ.κ. που μελετάμε με μια πιο απλή σ.κ. ισοδύναμης ουράς και να υπολογίσουμε τις σταθερές κανονικοποίησης για αυτή. Αποδεικνύεται ότι κάθε MDA είναι κλειστό ως προς την ισοδυναμία ουρών.

1.4.4 Εξέταση της κατανομής Fréchet

Ο προσδιορισμός του $\text{MDA}(\Phi_\alpha)$ συνδέεται με τις ομαλά μεταβαλλόμενες κατανομές. Για τη κατανομή $\Phi_\alpha(x) = \exp(-x^{-\alpha})$ και για $\alpha > 0$ χρησιμοποιώντας το ανάπτυγμα Taylor, έχουμε ότι

$$1 - \Phi_\alpha(x) = 1 - \exp(-x^{-\alpha}) \sim x^{-\alpha}, \quad \text{όταν } x \rightarrow \infty,$$

επομένως η ουρά της Φ_α φθίνει με εκθέτη μεταβολής $-\alpha$. Θεωρώντας ότι η ομάδα των κατανομών που ανήκει στην οικογένεια της κατανομής Fréchet, αποτελείται από συναρτήσεις που φθίνουν με εκθέτη μεταβολής $-\alpha$ λόγω της ισοδυναμίας ουρών, πρέπει να επιλέξουμε κατάλληλες σταθερές κανονικοποίησης. Αποδεικνύεται ότι για να έχουμε $F \in \text{MDA}(\Phi_\alpha)$, η σταθερά d_n μπορεί να επιλεγεί ίση με 0 και η c_n μπορεί να προσδιορισθεί μέσα από τη γενικευμένη αντίστροφη ως ακολούθως:

$$c_n = F^{\leftarrow}(1 - n^{-1}) = \inf \{x \in \mathbb{R} : F(x) \geq 1 - n^{-1}\} = \inf \{x \in \mathbb{R} : (1/\bar{F})(x) \geq n\} = (1/\bar{F})^{\leftarrow}(n).$$

1.4.5 Θεώρημα. Μια σ.κ. F ανήκει στο $\text{MDA}(\Phi_\alpha)$, $\alpha > 0$, αν και μόνο αν

$$\bar{F}(x) = x^{-\alpha} L(x)$$

για κάποια βραδέως μεταβαλλόμενη συνάρτηση L . Αν $F \in \text{MDA}(\Phi_\alpha)$, τότε

$$c_n^{-1} M_n \xrightarrow{d} \Phi_\alpha$$

όπου $c_n = F^{\leftarrow}(1 - n^{-1})$ και $d_n = 0$.

Το αποτέλεσμα αυτό υποδηλώνει ότι κάθε $F \in \text{MDA}(\Phi_\alpha)$ έχει ένα μη πεπερασμένο x_F . Επιπλέον οι σταθερές κανονικοποίησης c_n αποτελούν μια ομαλά μεταβαλλόμενη ακολουθία.

Αποδεικνύεται ότι $F \in \text{MDA}(\Phi_\alpha)$ αν και μόνο αν $\bar{F} \in \mathfrak{R}_{-\alpha}$. Πρέπει να σημειώσουμε ότι αυτή η κλάση κατανομών περιέχει κατανομές με πολύ βαριές δεξιές ουρές (*heavy-tailed distributions*). Επομένως αποτελούν τις κατάλληλες κατανομές για τη μοντελοποίηση μεγάλων ασφαλιστικών κινδύνων, κυμάνσεων τιμών κ.α.

Ο Von Mises βρήκε μία απλή συνθήκη για τη πυκνότητα μιας κατανομής, ώστε αυτή να ανήκει σε κάποιο MDA.

1.4.6 Πρόρισμα. Αν η F είναι μια απολύτως συνεχής σ.κ. με πυκνότητα f και ικανοποιεί τη σχέση

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{F(x)} = \alpha > 0$$

τότε $F \in \text{MDA}(\Phi_\alpha)$.

Η κλάση των σ.κ. F με ομαλά μεταβαλλόμενη συνάρτηση επιβίωσης \bar{F} είναι προφανώς κλειστή ως προς την ισοδυναμία ουρών.

1.4.7 Πρόταση. (Ιδιότητα κλειστότητας του $\text{MDA}(\Phi_\alpha)$ - closure property) Έστω F και G συναρτήσεις κατανομών και ισχύει $F \in \text{MDA}(\Phi_\alpha)$, με σταθερές κανονικοποίησης $c_n > 0$, δηλαδή

$$\lim_{n \rightarrow \infty} F^n(c_n x) = \Phi_\alpha(x), \text{ για κάθε } x > 0.$$

Τότε

$$\lim_{n \rightarrow \infty} G^n(c_n x) = \Phi_\alpha(cx), \text{ για κάθε } x > 0$$

για κάποιο $c > 0$ αν και μόνο αν F και G είναι κατανομές ισοδύναμων ουρών με

$$\lim_{n \rightarrow \infty} \frac{\bar{F}(x)}{\bar{G}(x)} = c^\alpha.$$

Επομένως μπορούμε να πούμε ότι το MDA της Fréchet περιλαμβάνει τις σ.κ. που ικανοποιούν τη συνθήκη Von Mises, και τις κατανομές που είναι ισοδύναμων ουρών. Είναι εύκολο να επαληθευτεί ότι στην $\text{MDA}(\Phi_\alpha)$ ανήκουν οι κατανομές Pareto, Cauchy, Loggamma, κ.α..

1.4.8 Εξέταση της κατανομής Weibull

Για να μελετήσουμε το MDA της κατανομής Weibull θα επιλέξουμε $\alpha > 0$. Πρέπει να είμαστε ιδιαίτερα προσεκτικοί, γιατί όλες οι σ.κ. F που ανήκουν στο MDA της Weibull έχουν δεξιό πεπερασμένο σημείο x_F . Η συγκεκριμένη κατανομή μπορεί να αποδοθεί από τη σχέση

$$\Psi_\alpha(-x^{-1}) = \Phi_\alpha = \exp\{-x^{-\alpha}\}, x > 0$$

επομένως περιμένουμε ότι το MDA της θα συνδέεται στενά με αυτό της Fréchet. Αν για $\alpha > 0$ πάρουμε ανάπτυγμα Taylor, μπορούμε να δούμε ότι και αυτή φθίνει με τον ίδιο τρόπο.

1.4.9 Θεώρημα. Μια σ.κ. F ανήκει στο MDA της Ψ_α , $\alpha > 0$ αν και μόνο αν για κάποια βραδέως μεταβαλλόμενη συνάρτηση L ισχύει ότι $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ και $x_F < \infty$. Τότε

$$c_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha$$

όπου η σταθερά κανονικοποίησης d_n επιλέγεται $d_n = x_F$ και $c_n = x_F - F^{\leftarrow}(1 - n^{-1})$.

Αποδεικνύεται ότι $F \in \text{MDA}(\Psi_\alpha)$ αν και μόνο αν για $x_F < \infty$ ισχύει ότι $\bar{F}(x_F - x^{-1}) \in \mathfrak{R}_{-\alpha}$, δηλαδή ότι το MDA της Weibull αποτελείται από σ.κ. F που είναι φραγμένες από τα δεξιά, για τον λόγο αυτό καλείται και οικογένεια κατανομών με μικρή ουρά (*short-tailed distributions*). Οι συναρτήσεις αυτές μπορεί να μην είναι κατάλληλες για τη μελέτη ακραίων ασφαλιστικών κινδύνων ή χρηματοοικονομικών μεταβολών επειδή $x_F < \infty$. Αν και προφανώς στις πρακτικές εφαρμογές πάντα υπάρχει ένα άνω όριο, είναι προτιμότερο να μη συμπεριλαμβανουμε την παράμετρο x_F και να προτιμήσουμε άλλες κατανομές που μας επιτρέπουν να έχουμε αυθαίρετα μεγάλες τιμές σε ένα δείγμα. Τέτοιες κατανομές ανήκουν στα $\text{MDA}(\Phi_\alpha)$ και $\text{MDA}(A)$.

Όπως και στη περίπτωση του MDA της Fréchet, ο Von Mises έθεσε και εδώ μια συνθήκη η οποία και αυτή αποτελεί απόρροια της ομαλής μεταβολής.

1.4.10 Πρόρισμα. Αν για μια απολύτως συνεχή σ.κ. F με πυκνότητα f η οποία είναι θετική σε κάποιο πεπερασμένο διάστημα (z, x_F) , ικανοποιείται το ακόλουθο όριο

$$\lim_{x \rightarrow x_F} \frac{(x_F - x)f(x)}{\bar{F}(x)} = a > 0$$

τότε $F \in \text{MDA}(\Psi_\alpha)$.

1.4.11 Πρόταση. (Ιδιότητα κλειστότητας του $\text{MDA}(\Psi_\alpha)$) Έστω ότι εξετάζουμε δυο σ.κ. F και G με πεπερασμένα δεξιά σημεία $x_F = x_G < \infty$ και ότι ισχύει $F \in \text{MDA}(\Psi_\alpha)$ με σταθερά κανονικοποίησης $c_n > 0$. Αυτό συνεπάγεται ότι

$$\lim_{n \rightarrow \infty} F^n(c_n x + x_F) = \Psi_\alpha(x), \text{ για κάθε } x > 0,$$

τότε θα ισχύει ότι

$$\lim_{n \rightarrow \infty} G^n(c_n x + x_F) = \Psi_\alpha(cx), \text{ για κάθε } x > 0$$

για κάποιο $c > 0$ αν και μόνο αν F και G είναι κατανομές ισοδύναμων ουρών με

$$\lim_{x \rightarrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c^{-\alpha}.$$

Επομένως μπορούμε να πούμε ότι το MDA της Weibull περιλαμβάνει της σ.κ. που ικανοποιούν τη συνθήκη Von Mises και τις κατανομές που είναι ισοδύναμων ουρών.

1.4.12 Εξέταση της κατανομής Gumbel

Το MDA της κατανομής Gumbel καλύπτει ένα μεγάλο κομμάτι κατανομών ακραίων τιμών F . Παρόλο που δεν υπάρχει σύνδεση με τη θεωρία των ομαλά μεταβαλλόμενων συναρτήσεων όπως για τα MDA της Fréchet και της Weibull, μπορούμε να βρούμε κάποιες προεκτάσεις της θεωρίας που να επιτρέπουν ένα πλήρη χαρακτηρισμό του MDA της Gumbel.

Όπως και για τις δυο προηγούμενες σ.κ., αν πάρουμε ανάπτυγμα Taylor για τη Gumbel προκύπτει ότι

$$1 - A(x) \sim e^{-x}, x \rightarrow \infty,$$

δηλαδή η συνάρτηση επιβίωσης της κατανομής Gumbel φθίνει στο μηδέν με εκθετικό ρυθμό. Τα αποτελέσματα της θεωρητικής έρευνας που έχει πραγματοποιηθεί δείχνουν ότι η οικογένεια κατανομών Gumbel περιλαμβάνει σ.κ. με πολύ διαφορετικές ουρές – από κατανομές με ελαφρώς βαριές ουρές (*moderately heavy tailed distributions*), όπως η Λογαριθμοκανονική, μέχρι συναρτήσεις με ελαφριές ουρές (*light tailed distributions*), όπως η Κανονική. Γενικά καλείται οικογένεια των κατανομών με ουρά μεσαίου μεγέθους (*medium-tailed distributions*). Επίσης τόσο η περίπτωση του $x_F < \infty$ όσο και του $x_F = \infty$ είναι δυνατές. Αυτά τα χαρακτηριστικά του MDA της κατανομής τη καθιστούν ως τη πιο ενδιαφέρουσα ανάμεσα σε όλες τις κατανομές ακραίων τιμών. Φυσικό επακόλουθο της ποικιλομορφίας που παρουσιάζουν οι ουρές της κατανομής, είναι και η ποικιλομορφία των σταθερών κανονικοποίησης. Ενώ στις περιπτώσεις των MDA των κατανομών Fréchet και Weibull ο προσδιορισμός των σταθερών κανονικοποίησης γινόταν με άμεση εφαρμογή της θεωρίας των ομαλά μεταβαλλόμενων συναρτήσεων, για τον προσδιορισμό των ίδιων παραμέτρων για την Gumbel χρειαζόμαστε πιο εξειδικευμένα αποτελέσματα. Για να διευκολυνθούμε στην ανάλυση μας θα περιοριστούμε σε μια απολύτως συνεχή $F \in \text{MDA}(A)$ η οποία έχει μια απλή αναπαράσταση, όπως αυτή δόθηκε από τον Von Mises (βλ. Von Mises (1936)). Αυτές οι κατανομές παρέχουν ένα σημαντικό δομικό κομμάτι αυτού του MDA, και κατά συνέπεια αποτελούν και αντικείμενο εκτενούς μελέτης.

1.4.13 Ορισμός. (Συνάρτηση Von Mises) Έστω F μια σ.κ. με δεξιό οριακό σημείο (endpoint) $x_F \leq \infty$. Υποθέτουμε ότι υπάρχει $z < x_F$ τέτοιο ώστε η F να αναπαρίσταται ως ακολούθως:

$$\bar{F}(x) = c \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\}, \quad z < x < x_F \quad (1.11)$$

όπου c είναι μια θετική σταθερά, $a(\cdot)$ είναι μια θετική και απολύτως συνεχής συνάρτηση με πυκνότητα a' για την οποία ισχύει ότι

$$\lim_{x \rightarrow x_F} a'(x) = 0.$$

Σε αυτή τη περίπτωση καλούμε την F συνάρτηση Von Mises, και η $a(\cdot)$ καλείται βοηθητική συνάρτηση της F .

Η (1.11) αρχικά μπορεί να μην θυμίζει κάτι γνωστό, όμως μπορούμε να δούμε πως κάποιες ήδη γνωστές κατανομές αποτελούν συνάρτηση Von Mises.

1.4.14 Παραδείγματα

Ας εξετάσουμε την Εκθετική κατανομή με παράμετρο λ της οποίας η συνάρτηση επιβίωσης ορίζεται ως

$$\bar{F}(x) = \exp(-\lambda x), \quad x \geq 0, \lambda > 0.$$

Επομένως η Εκθετική εκφράζεται ως μια συνάρτηση Von Mises με βοηθητική συνάρτηση $a(x) = \lambda^{-1}$.

Ας εξετάσουμε την Weibull της οποίας η συνάρτηση επιβίωσης ορίζεται ως

$$\bar{F}(x) = \exp(-cx^\tau), \quad x \geq 0, c, \tau > 0.$$

Προκύπτει ότι η Weibull είναι μια συνάρτηση Von Mises με βοηθητική συνάρτηση

$$a(x) = c^{-1} \tau^{-1} x^{-1-\tau}, \quad x > 0.$$

Ως μια πιο γενική περίπτωση θα εξετάσουμε μια σ.κ. F με δεξιό σημείο $x_F \leq \infty$. Υποθέτουμε ότι υπάρχει κάποιο $z < x_F$ τέτοιο ώστε η F να είναι διπλά διαφορίσιμη στο (z, x_F) με θετική πυκνότητα $f = F'$ και $F''(x) < 0$ για $z < x < x_F$. Προκύπτει ότι η F είναι συνάρτηση Von Mises με βοηθητική συνάρτηση $a = \bar{F} / f$, όπου η αντίστροφη της βοηθητικής είναι η συνάρτηση κινδύνου της F , αν και μόνο αν ισχύει

$$\lim_{x \rightarrow x_F} \bar{F}(x) F''(x) / f^2(x) = -1. \quad (1.12)$$

Πράγματι έστω ότι $z < x < x_F$, αν θέσουμε

$$Q(x) = -\ln \bar{F}(x)$$

και

$$a(x) = 1 / Q'(x) = \bar{F}(x) / f(x) > 0,$$

τότε η F αναπαρίσταται όπως στην (1.11). Επιπλέον ισχύει ότι η

$$a'(x) = -\frac{\bar{F}(x)F''(x)}{f^2(x)} - 1$$

και η (1.12) είναι ισοδύναμες με $\lim_{x \rightarrow x_F} a'(x) = 0$. Η συνθήκη (1.12) εφαρμόζεται σε πολλές κατανομές που αποτελούν αντικείμενο ενδιαφέροντος για τη μελέτη ακραίων τιμών, όπως είναι και η κανονική.

1.4.15 Πρόταση. (Ιδιότητες της συνάρτησης Von Mises) Κάθε συνάρτηση F Von Mises είναι απολύτως συνεχής στο (z, x_F) , με θετική πυκνότητα f . Η βοηθητική συνάρτηση μπορεί να επιλεγεί από τη σχέση

$$a(x) = \bar{F}(x) / f(x).$$

Επιπλέον ισχύουν τα ακόλουθα:

Αν $x_F = \infty$, τότε $\bar{F} \in \mathfrak{R}_{-\infty}$ και

$$\lim_{x \rightarrow \infty} \frac{xf'(x)}{F(x)} = \infty$$

Αν $x_F < \infty$, τότε $\bar{F}(x_F - x^{-1}) \in \mathfrak{R}_{-\infty}$ και

$$\lim_{x \rightarrow x_F} \frac{(x_F - x)f'(x)}{F(x)} = \infty.$$

Βασιζόμενοι στα παραπάνω, θα δείξουμε ότι οι συναρτήσεις Von Mises ανήκουν στο MDA της Gumbel και επιπλέον ότι η συγκεκριμένη μορφή συνάρτησης επιβίωσης επιτρέπει τον υπολογισμό της σταθεράς κανονικοποίησης c_n από τη βοηθητική συνάρτηση.

1.4.16. Πρόταση. (Συνάρτηση Von Mises και MDA(λ)) Αν η σ.κ. F είναι συνάρτηση Von Mises, τότε $F \in \text{MDA}(\lambda)$. Μια πιθανή επιλογή σταθερών κανονικοποίησης είναι

$$d_n = F^{\leftarrow}(1 - n^{-1}) \text{ και } c_n = a(d_n)$$

όπου a είναι η βοηθητική συνάρτηση της F .

Η συνάρτηση Von Mises δε χαρακτηρίζει πλήρως το $\text{MDA}(A)$, παρόλα αυτά χρησιμοποιώντας κάποια παραλλαγή του ορισμού, μπορούμε μέσω αυτής να δώσουμε πλήρη χαρακτηρισμό του.

1.4.17 Θεώρημα. (Χαρακτηρισμός I του $\text{MDA}(A)$) Η σ.κ. F με δεξιό οριακό σημείο $x_F \leq \infty$ ανήκει στο $\text{MDA}(A)$ αν και μόνο αν υπάρχει κάποιο $z < x_F$ τέτοιο ώστε η F να έχει την αναπαράσταση ως

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

όπου c και g είναι μετρήσιμες συναρτήσεις που ικανοποιούν τις συνθήκες $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$ καθώς $x \rightarrow x_F$ και $a(x)$ είναι μια θετική απολύτως συνεχής συνάρτηση με πυκνότητα $a'(x)$, η οποία ικανοποιεί το όριο $\lim_{x \rightarrow x_F} a'(x) = 0$. Για να έχει η F την παραπάνω αναπαράσταση μπορούμε να επιλέξουμε τις σταθερές κανονικοποίησης $d_n = F^{\leftarrow}(1 - n^{-1})$ και $c_n = a(d_n)$. Μια πιθανή επιλογή για τη συνάρτηση a είναι

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt, \quad x < x_F. \quad (1.13)$$

Υπάρχει και ένας άλλος τρόπος να δοθεί ο χαρακτηρισμός του $\text{MDA}(A)$.

1.4.18 Θεώρημα. (Χαρακτηρισμός II του $\text{MDA}(A)$) Μια σ.κ. F ανήκει στο $\text{MDA}(A)$ αν και μόνο αν υπάρχει κάποια θετική συνάρτηση \tilde{a} τέτοια ώστε να ισχύει

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x + t \tilde{a}(x))}{\bar{F}(x)} = e^{-t}.$$

Μια πιθανή μορφή της συνάρτησης $a = \tilde{a}$ είναι και πάλι η (1.13).

Στην § 1.4.3 αναφέραμε ένα πολύ σημαντικό εργαλείο για την ανάλυση κατανομών ακραίων τιμών, την ισοδυναμία ουρών. Η συγκεκριμένη ιδιότητα εφαρμόζεται και στη κατανομή Gumbel.

1.4.19 Πρόταση. (Ιδιότητα κλειστότητας του $\text{MDA}(A)$) Έστω F και G σ.κ. με τα ίδια οριακά δεξιά σημεία $x_F = x_G$ με την F να ανήκει στο $\text{MDA}(A)$ με σταθερές κανονικοποίησης $c_n > 0$ και $d_n \in \mathbf{R}$ τέτοιες ώστε να ισχύει

$$\lim_{n \rightarrow \infty} F^n(c_n x + x_F) = A(x), \quad x \in \mathbf{R}.$$

Τότε θα ισχύει

$$\lim_{n \rightarrow \infty} G^n(c_n x + x_F) = A(x+b), \quad x \in \mathbf{R}$$

αν και μόνο αν οι συναρτήσεις F και G είναι ισοδύναμων ουρών με

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = e^b.$$

Επομένως το $MDA(A)$ αποτελείται από συναρτήσεις Von Mises και από τις συναρτήσεις ισοδύναμων ουρών αυτών.

1.4.20. Παρατήρηση. (Εμφύτευση του $MDA(\Phi_\alpha)$ στο $MDA(A)$). Έστω X με σ.κ. F που ανήκει στο $MDA(\Phi_\alpha)$ με σταθερά c_n . Ορίζουμε $X^* = \ln(1 \vee X)$ με σ.κ. F^* . Από τις § 1.4.1 και 1.4.5 γνωρίζουμε ότι η $F \in MDA(\Phi_\alpha)$, αν και μόνο αν ισχύει

$$\lim_{n \rightarrow \infty} n \bar{F}(c_n x) = \lim_{n \rightarrow \infty} \frac{\bar{F}(c_n x)}{\bar{F}(c_n)} = x^{-a}, \quad x > 0$$

το οποίο υποδηλώνει ότι

$$\lim_{n \rightarrow \infty} \frac{\bar{F}^*(a^{-1} c_n + \ln c_n)}{\bar{F}^*(\ln c_n)} = \lim_{n \rightarrow \infty} \frac{\bar{F}(c_n \exp(a^{-1} x))}{\bar{F}(c_n)} = e^{-x}, \quad x \in \mathbf{R}.$$

Επομένως η F^* ανήκει στο $MDA(A)$ με σταθερές $c_n^* = a^{-1}$ και $d_n = \ln c_n$. Ως βοηθητική συνάρτηση της Von Mises μπορούμε να επιλέξουμε τη

$$a^*(x) = \int_x^\infty \frac{\bar{F}^*(y)}{\bar{F}^*(x)} dy.$$

1.5. Γενικευμένη Οικογένεια Κατανομών Ακραίων Τιμών (Generalized Extreme Value Distribution, GEV)

Για να προχωρήσουμε στην παρουσίαση της γενικευμένης οικογένειας κατανομών ακραίων τιμών μπορούμε να ορίσουμε μια παραμετρική οικογένεια $(H_\xi)_{\xi \in \mathbf{R}}$, η οποία περιλαμβάνει τις τυπικές κατανομή ακραίων τιμών ως εξής:

$$H_\xi = \begin{cases} \Phi_{1/\xi}, & \text{αν } \xi > 0 \\ A, & \text{αν } \xi = 0 \\ \Psi_{-1/\xi}, & \text{αν } \xi < 0 \end{cases}$$

Η συνθήκη $F \in \text{MDA}(H_\xi)$ αποδίδει την απόλυτη προσέγγιση (*ultimate approximation*)

$$F^n(c_n x + d_n) \approx H_\xi(x)$$

για κατάλληλα επιλεγμένες σταθερές $c_n > 0$ και $d_n \in \mathbf{R}$.

Επομένως, μια μονοπαραμετρική αναπαράσταση των τριών τυπικών κατανομών ακραίων τιμών δίνεται εισάγοντας τη παράμετρο ξ , έτσι ώστε

$\xi = \alpha^{-1} > 0$, να αντιστοιχεί στη κατανομή Fréchet Φ_α

$\xi = 0$, να αντιστοιχεί στη κατανομή Gumbel A

$\xi = -\alpha^{-1} < 0$, να αντιστοιχεί στη κατανομή Weibull Ψ_α

Η ενοποίηση των τριών αρχικών οικογενειών κατανομών σε μια οικογένεια, απλοποιεί τη στατιστική ανάλυση (βλ. Von Mises (1936), Jenkinson (1955)). Τα δειγματικά δεδομένα είναι αυτά που καθορίζουν τη συμπεριφορά της ουράς της κατανομής και μας οδηγούν στην επιλογή του κατάλληλου ξ , γεγονός που δε μας δεσμεύει να προβούμε σε μια υποκειμενική επιλογή της κατάλληλης κατανομής πριν ακόμα προχωρήσουμε στην ανάλυση.

1.5.1 Ορισμός (Ορισμός Jenkinson - Von Mises - Αναπαράσταση της GEV)_Ορίζουμε τη σ.κ. H_ξ

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}) & \text{αν } \xi \neq 0 \\ \exp(-\exp(-x)) & \text{αν } \xi = 0 \end{cases}$$

όπου $1 + \xi x > 0$. Επομένως το στήριγμα της H_ξ αντιστοιχεί σε $x > -\xi^{-1}$ αν $\xi > 0$, σε $x < -\xi^{-1}$ αν $\xi < 0$ και σε $x \in \mathbf{R}$ αν $\xi = 0$.

Η H_ξ καλείται τυπική γενικευμένη κατανομή ακραίων τιμών. Μπορούμε να θεωρήσουμε την οικογένεια γενικευμένων κατανομών θέσης και κλίμακας $H_{\xi;\mu;\psi}$ αντικαθιστώντας το x παραπάνω με $(x - \mu) / \psi$ για $\mu \in \mathbf{R}$, $\psi > 0$. Το στήριγμα της κατανομής θα πρέπει να προσαρμοστεί κατάλληλα. Και αυτή η σ.κ. καλείται GEV.

Θεωρούμε τη σ.κ. H_0 ως το όριο H_ξ καθώς $\xi \rightarrow 0$. Μια γενική αναπαράσταση της GEV για όλα τα $\xi \in \mathbf{R}$, δίνεται ως ακολούθως

$$H_\xi(x) = \exp(-(1 + \xi x)^{-1/\xi}), \quad 1 + \xi x > 0.$$

Η GEV μας δίνει μια ενοποιημένη αναπαράσταση των τριών κατανομών ακραίων τιμών Fréchet, Gumbel και Weibull και όπως και σε αυτές τις κατανομές ακραίων τιμών έτσι και στη GEV υπάρχει MDA. Ορίζουμε $U(t) = F^{\leftarrow}(1 - t^{-1})$, $t > 0$.

1.5.2. Θεώρημα. (Χαρακτηρισμός του $\text{MDA}(H_\xi)$) Για $\xi \in \mathbf{R}$ οι ακόλουθες προτάσεις είναι ισοδύναμες:

(α) $F \in \text{MDA}(H_\xi)$

(β) Υπάρχει κάποια θετική, μετρήσιμη συνάρτηση $a(\cdot)$ τέτοια ώστε για $1 + \xi x > 0$

$$\lim_{u \rightarrow x_F} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{αν } \xi \neq 0 \\ e^{-x}, & \text{αν } \xi = 0 \end{cases} \quad (1.14)$$

(γ) Για $x, y > 0, y \neq 1,$

$$\lim_{s \rightarrow x_F} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1}, & \text{αν } \xi \neq 0 \\ \frac{\ln x}{\ln y}, & \text{αν } \xi = 0 \end{cases}.$$

Το όριο (1.14) έχει μια πολύ ενδιαφέρουσα ερμηνεία. Αν X είναι μια τ.μ. μια σ.κ. $F \in \text{MDA}(H_\xi)$, τότε το όριο (1.14) μπορεί να αποδοθεί με τον ακόλουθο τύπο

$$\lim_{u \rightarrow x_F} P\left(\frac{X - u}{a(u)} > x \mid X > u\right) = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{αν } \xi \neq 0 \\ e^{-x}, & \text{αν } \xi = 0 \end{cases}. \quad (1.15)$$

Το όριο (1.15) δίνει μια συναρτησιακή προσέγγιση για τις τ.μ. που ξεπερνάνε κάποιο υψηλό κατώφλι u . Η υπερβάλλουσα συνάρτηση κατανομής και η μέση υπερβάλλουσα συνάρτησή μας δίνουν μια πιο ακριβή ερμηνεία της έννοιας.

1.5.3. Ορισμός. (Υπερβάλλουσα συνάρτηση κατανομής και μέση υπερβάλλουσα συνάρτηση)

Έστω X μια τ.μ. με σ.κ. F και δεξιό οριακό σημείο x_F . Για ένα σταθερό $u < x_F$, η

$$F_u(x) = P(X - u \leq x \mid X > u), \quad x \geq 0$$

καλείται υπερβάλλουσα σ.κ. (Excess Distribution Function) της τ.μ. X (ή της σ.κ. F) πάνω από το κατώφλι (threshold) u . Η συνάρτηση

$$e(u) = E(X - u \mid X > u)$$

καλείται μέση υπερβάλλουσα συνάρτηση (mean excess function) ή συνάρτηση μέσης υπολοίπου ζωής (mean residual life function) της τ.μ. X .

Τα υπερβάλλοντα ποσά (ή σημεία ή παρατηρήσεις) πάνω από κάποιο ορισμένο κατώφλι παίζουν σημαντικό ρόλο σε πολλά επιστημονικά πεδία. Π.χ. η F_u είναι γνωστή στη θεωρία

αξιοπιστίας και στη βιοστατιστική ως σ.κ. υπερβάλλουσας ή υπολειπόμενης ζωής. Στις ασφάλειες F_u είναι συνήθως η σ.κ. *excess-of-loss* δηλαδή υπερβάλλουσας ζημίας.

1.5.4. Πρόταση. (Υπολογισμός της μέσης υπερβάλλουσας συνάρτησης) Υποθέτουμε ότι X είναι μια θετική τ.μ. με σ.κ. F και πεπερασμένη αναμενόμενη τιμή. Τότε

$$e(u) = \int_u^{x_F} (x-u) dF(x) / \bar{F}(u) = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) dx, \quad 0 < u < x_F. \quad (1.16)$$

Όταν η F είναι συνεχής ισχύει

$$\bar{F}(x) = \frac{e(0)}{e(x)} \exp \left\{ - \int_0^x \frac{1}{e(u)} du \right\}, \quad x > 0.$$

Από την παραπάνω πρόταση προκύπτει ότι μια συνεχής σ.κ. ορίζεται με μοναδικό τρόπο από τη μέση υπερβάλλουσα συνάρτησή της.

Είδαμε ότι η GEV μας δίνει μια ενοποιημένη συναρτησιακή μορφή για τις τρεις τυπικές κατανομές των μεγίστων τιμών μιας ακολουθίας iid τ.μ.. Η γενικευμένη κατανομή Pareto και οι αντίστοιχοι κλάδοι της εμφανίζονται ως οριακή κατανομή για την (κανονικοποιημένη) υπερβάλλουσα τιμή μιας τ.μ.. Αντί λοιπόν να εξετάζουμε τις μέγιστες τιμές, εξετάζουμε τις ουρές τους, δηλαδή εξ'ορισμού τις τιμές εκείνες που υπερβαίνουν κάποιο πολύ μεγάλο κατώφλι u , το οποίο ικανοποιεί τη συνθήκη $u \rightarrow \infty$.

Είδαμε ότι η σχέση (1.16) δίνει μια συναρτησιακή προσέγγιση για τις τ.μ. που ξεπερνάνε κάποιο υψηλό κατώφλι u και μας οδηγεί στον ακόλουθο ορισμό.

1.5.5. Ορισμός. (Γενικευμένη Κατανομή Pareto –Generalized Pareto Distribution (GPD))

Ορίζουμε τη σ.κ. G_ξ ως

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, & \text{αν } \xi \neq 0 \\ 1 - e^{-x}, & \text{αν } \xi = 0 \end{cases}$$

όπου $x \geq 0$ αν $\xi \geq 0$ και $0 \leq x \leq -1/\xi$ αν $\xi < 0$.

Η G_ξ καλείται γενικευμένη κατανομή Pareto. Μπορούμε να θεωρήσουμε την οικογένεια γενικευμένων κατανομών θέσης και κλίμακας $G_{\xi, \nu, \beta}$ αντικαθιστώντας το x παραπάνω με $(x - \nu) / \beta$ για $\nu \in \mathbb{R}$, $\beta > 0$. Το στήριγμα της κατανομής θα πρέπει να προσαρμοστεί καταλλήλως. Και αυτή η σ.κ. καλείται GPD.

Μπορούμε να θεωρήσουμε τη σ.κ. G_0 ως το όριο G_ξ καθώς $\xi \rightarrow 0$. Μια γενική αναπαράσταση της GPD για όλα τα $\xi \in \mathbf{R}$, δίνεται ως ακολούθως

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi, \beta),$$

όπου για $\xi \geq 0$, $x \in D(\xi, \beta)$ είναι ίσο με $[0, \infty)$ ενώ για $\xi < 0$ είναι ίσο με $[0, -\beta/\xi]$. Όταν θα λέμε ότι η X έχει GPD με παραμέτρους ξ και β , θα εννοούμε ότι η X έχει σ.κ. $G_{\xi,\beta}$.

1.6. Διατεταγμένες παρατηρήσεις (Order Statistics)

Κατά την προηγούμενη ανάλυση μελετήσαμε τη συμπεριφορά των μεγαλύτερων τιμών μέσα σε ένα δείγμα παρατηρήσεων. Σε αυτό το κομμάτι θα προσπαθήσουμε να δούμε την από κοινού συμπεριφορά των διατεταγμένων πλέον τιμών, που ξεπερνούν κάποιο ορισμένο κατώφλι. Η συμπεριφορά αυτών των τιμών μας δίνει σημαντικές πληροφορίες για τη δεξιά ουρά της εκάστοτε σ.κ. που θα εξετάζουμε.

Ας θεωρήσουμε ότι έχουμε ένα δείγμα iid τ.μ. X_1, \dots, X_n μιας μη εκφυλισμένης σ.κ. F . Αν διατάξουμε τις τιμές της ακολουθίας κατά αύξουσα σειρά, προκύπτει το διατεταγμένο δείγμα $X_{n,n} \leq \dots \leq X_{1,n}$, όπου $X_{n,n} = \min(X_1, \dots, X_n)$ και $X_{1,n} = \max(X_1, \dots, X_n)$ είναι η ελάχιστη και η μέγιστη τιμή αντίστοιχα. Ορίζουμε λοιπόν τη τ.μ. $X_{k,n}$ ως τη k -οστή άνω διατεταγμένη τιμή (k^{th} upper order statistic). Η σχέση των διατεταγμένων τιμών με την αρχική σ.κ., μπορεί να μην είναι προφανής, όμως αν ορίσουμε την εμπειρική συνάρτηση κατανομής (empirical df) ως ακολούθως

$$F_x(n) = \frac{1}{n} \text{card} \{i : 1 \leq i \leq n, X_i \leq x\} = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}, \quad x \in \mathbf{R},$$

όπου I_A είναι μια δείκτρια συνάρτηση στο σύνολο A και παρατηρήσουμε ότι ισχύει

$$X_{k,n} \leq x \text{ αν και μόνο αν } \sum_{i=1}^n I_{\{X_i > x\}} < k \quad (1.17)$$

τότε καταλήγουμε ότι η πιθανότητα, η k -οστή άνω διατεταγμένη τιμή να ξεπερνάει τη τιμή x , είναι ίση με τη πιθανότητα η εμπειρική συνάρτηση κατανομής στο x να παίρνει τιμή μεγαλύτερη του $(1 - k/n)$, δηλαδή

$$P(X_{k,n} \leq x) = P\left(F_n(x) > 1 - \frac{k}{n}\right).$$

Οι άνω διατεταγμένες παρατηρήσεις ενός δείγματος από μία κατανομή είναι χρήσιμες για την εκτίμηση τόσο της ουράς της κατανομής, των ποσοστιαίων σημείων της καθώς και για πιθανότητες της μορφής $P(X > x)$.

Πέραν από την εμπειρική σ.κ., μπορούμε να δούμε και τη σχέση των διατεταγμένων τιμών με την εμπειρική συνάρτηση ποσοστιαίων σημείων F_n^{\leftarrow} . Αν η σ.κ. F είναι συνεχής (κάτι που συμβαίνει στις περισσότερες περιπτώσεις), τότε μπορούμε να υποθέσουμε ότι ισχύει $X_{n,n} < \dots < X_{1,n}$, οπότε σε αυτή τη περίπτωση η F_n^{\leftarrow} εκφράζεται ως μια απλή συνάρτηση των διατεταγμένων τιμών και συγκεκριμένα

$$F_n^{\leftarrow}(t) = X_{k,n} \text{ για } 1 - (k/n) < t \leq 1 - (k-1)/n$$

για $k = 1, \dots, n$. Στην παρακάτω πρόταση δίνεται η σ.κ. των διατεταγμένων τυχαίων μεταβλητών.

1.6.1 Πρόταση. (Συνάρτηση κατανομής της k -οστής άνω διατεταγμένης τυχαίας μεταβλητής ή παρατήρησης) Αν για $k = 1, 2, \dots, n$, συμβολίζουμε με $F_{k,n}$ τη σ.κ. της $X_{k,n}$, τότε θα ισχύουν τα ακόλουθα:

(α)

$$F_{k,n}(x) = \sum_{r=0}^{k-1} \binom{n}{r} \bar{F}^r(x) F^{n-r}(x).$$

(β) Αν η F είναι συνεχής, τότε

$$F_{k,n}(x) = \int_{-\infty}^x f_{k,n}(z) dF(z)$$

όπου

$$f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!} F^{n-k}(x) \bar{F}^{k-1}(x).$$

Η $f_{k,n}$ μπορεί να θεωρηθεί ως η πυκνότητα της $F_{n,k}$, όχι ως προς το μέτρο Lebesgue, αλλά ως προς το μέτρο πιθανότητας με σ.κ. F .

Με παρόμοιο τρόπο καταλήγουμε στο τύπο της από κοινού κατανομής όλων των διατεταγμένων παρατηρήσεων. Αν η σ.κ. F είναι απολύτως συνεχής με σ.π. f , τότε η από κοινού πυκνότητα των (X_1, \dots, X_n) ως γνωστό είναι

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i), \quad (x_1, \dots, x_n) \in \mathbf{R}^n.$$

Δεδομένου τώρα ότι οι τιμές (X_1, \dots, X_n) προέρχονται από συνεχή σ.κ., μπορούν να τοποθετηθούν με $n!$ τρόπους και επομένως κάθε συλλογή διατεταγμένων τιμών $(X_{k,n})_{k=1, \dots, n}$ μπορεί να προέρχεται από $n!$ διαφορετικά δείγματα. Επομένως η από κοινού σ.π. του διατεταγμένου δείγματος γράφεται

$$f_{X_{1,n}, \dots, X_{n,n}}(x_1, \dots, x_n) = n! \prod_{i=1}^n f(x_i), \quad x_n < \dots < x_1.$$

1.6.2. Θεώρημα. (Η από κοινού πυκνότητα των k άνω διατεταγμένων παρατηρήσεων) Αν η σ.κ. F είναι απολύτως συνεχής με πυκνότητα f , τότε η από κοινού πυκνότητα των k άνω διατεταγμένων παρατηρήσεων δίνεται από τον τύπο

$$f_{X_{1,n}, \dots, X_{k,n}}(x_1, \dots, x_k) = \frac{n!}{(n-k)!} F^{n-k}(x_k) \prod_{i=1}^k f(x_i), \quad x_k < \dots < x_1.$$

Μια ποσότητα που προκύπτει φυσικά από τον ορισμό των διατεταγμένων παρατηρήσεων και αποτελεί τη θεμέλια λίθο για τον εκτιμητή *Hill* που θα δούμε παρακάτω, είναι το διάκενο ή απόσταση (*spacing*) μεταξύ των διατεταγμένων παρατηρήσεων και ορίζεται ως η διαφορά μεταξύ δυο διαδοχικών διατεταγμένων παρατηρήσεων.

1.6.3 Ορισμός. (Διάκενο διατεταγμένων παρατηρήσεων – *spacing*) Διάκενα μεταξύ των παρατηρήσεων X_1, \dots, X_n ορίζονται οι διαφορές

$$X_{k,n} - X_{k+1,n}, \quad k = 1, \dots, n-1.$$

Για τ.μ. με πεπερασμένο αριστερό (δεξιό) οριακό σημείο $\tilde{x}_F(x_F)$ ορίζουμε τη n -οστό (0-οστό) διάκενο ως

$$X_{n,n} - X_{n+1,n} = X_{n,n} - \tilde{x}_F \quad (X_{0,n} - X_{1,n} = x_F - X_{1,n}).$$

Οι διατεταγμένες παρατηρήσεις που προκύπτουν από ένα δείγμα είναι πολύ χρήσιμες για τη μελέτη υπερβάλλουσων πιθανοτήτων πάνω από κάποιο κατώφλι u . Αποδεικνύεται ότι μπορούμε να εξάγουμε συμπεράσματα για τις k -άνω διατεταγμένες παρατηρήσεις, μέσω των αντίστοιχων διατεταγμένων παρατηρήσεων της ομοιόμορφης κατανομής.

1.6.4 Λήμμα. (Μετασχηματισμός ποσοστιαίων σημείων) Έστω X_1, \dots, X_n iid τ.μ. με σ.κ. F και U_1, \dots, U_n iid τ.μ. που προέρχονται από την $U(0,1)$. Αν ορίσουμε ως $U_{n,n} < \dots < U_{1,n}$ τις διατεταγμένες U_1, \dots, U_n , τότε ισχύουν τα ακόλουθα:

(α) $F_n^{\leftarrow}(U_1) \stackrel{d}{=} X_1$

(β) Για κάθε $n \in \mathbb{N}$, ισχύει ότι $(X_{1,n}, \dots, X_{n,n}) \stackrel{d}{=} (F^{\leftarrow}(U_{1,n}), \dots, F^{\leftarrow}(U_{n,n}))$

(γ) Η τ.μ. $F(X_1)$ ακολουθεί την $U(0,1)$ αν και μόνο αν η F είναι συνεχής.

Ο παραπάνω μετασχηματισμός συνδέει την Ομοιόμορφη κατανομή γενικά με κάποια κατανομή F . Μέσω αυτής της σχέσης μπορούμε να παράγουμε (*generate*) τυχαίους αριθμούς (μέθοδος αντιστροφής).

Αν θελήσουμε να μελετήσουμε τη συμπεριφορά των ποσοστιαίων σημείων και να βρούμε τις εκτιμήσεις για τις ουρές των κατανομών, τότε θα πρέπει να γνωρίζουμε τις ασυμπτωτικές ιδιότητες των $X_{k,n}$.

1.6.5 Πρόταση. (Σχεδόν βέβαιη σύγκλιση των διατεταγμένων παρατηρήσεων) Έστω F μια σ.κ. με δεξιό (αριστερό) οριακό σημείο $x_F \leq \infty$ ($\tilde{x}_F \geq -\infty$) και $(k(n))$ μια μη φθίνουσα ακολουθία ακεραίων αριθμών τέτοια ώστε να ισχύει $\lim_{n \rightarrow \infty} n^{-1}k(n) = c \in [0,1]$.

(α) Τότε $X_{k(n),n} \xrightarrow{a.s.} x_F$ (\tilde{x}_F) αν ισχύει ότι $c = 0$ ($c = 1$).

(β) Αν υποθέσουμε ότι $c \in (0, 1)$ έτσι ώστε να υπάρχει μια μοναδική λύση $x(c)$ για την εξίσωση $\bar{F}(x) = c$, τότε θα ισχύει $X_{k(n),n} \rightarrow x(c)$, σχεδόν βέβαια.

Έστω X_1, \dots, X_n iid τ.μ. με σ.κ. F . Όπως έχει αναφερθεί και στη § 1.3.1, αν έχουμε μια ακολουθία (u_n) πραγματικών αριθμών και ισχύει ότι $0 \leq \tau \leq \infty$, τότε τα ακόλουθα είναι ισοδύναμα

$$\lim_{n \rightarrow \infty} P(X_{1,n} \leq u_n) = e^{-\tau} \quad \text{αν και μόνο αν} \quad \lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau \quad (1.18)$$

Το ερώτημα είναι αν μπορούμε να επεκτείνουμε τη σχέση (1.18) ώστε να ισχύει για κάθε $X_{k,n}$ για δεδομένο $k \in \mathbb{N}$ ή ακόμα αν μπορούμε να πάρουμε οριακές πιθανότητες για τις

$X_{k,n}, \dots, X_{1,n}$ για κάποιο δεδομένο k . Αν θεωρήσουμε ότι για $n \in \mathbf{N}$ ο αριθμός των X_1, \dots, X_n που ξεπερνάνε το κατώφλι u_n εκφράζεται από τη σχέση

$$B_n = \sum_{i=1}^n I_{\{X_i > u_n\}},$$

τότε η B_n είναι μια διωνυμική τ.μ. με παραμέτρους n και $\bar{F}(u_n)$. Πρέπει να σημειώσουμε ότι οι υπερβάσεις $\{X_i > u_n\}$ τείνουν να είναι λιγότερες καθώς αυξάνεται η τιμή του κατωφλίου u_n , ενώ ταυτόχρονα τείνουν να αυξάνονται καθώς αυξάνεται το μέγεθος του δείγματος που εξετάζουμε. Θα πρέπει λοιπόν με κάποιο τρόπο να εξισορροπήσουμε αυτά τα δυο αποτελέσματα έτσι ώστε να ισχύει $E(B_n) = n\bar{F}(u_n) \rightarrow \tau$ καθώς $n \rightarrow \infty$, για να μπορέσουμε μέσω της προσέγγισης Poisson να καταλήξουμε ότι

$$B_n \xrightarrow{d} P(\tau).$$

Για να επιτευχθεί αυτό θα πρέπει να επιλέξουμε το κατώφλι u_n , έτσι ώστε ο αριθμός των υπερβάλλουσων τιμών να συγκλίνει.

1.6.6 Θεώρημα. (Οριακός νόμος για το πλήθος των υπερβαλλουσών παρατηρήσεων) Υποθέτουμε ότι (u_n) είναι μια ακολουθία στο \mathbf{R} , τέτοια ώστε να ισχύει $n\bar{F}(u_n) \rightarrow \tau$ για κάποιο $\tau \in [0, \infty]$, καθώς $n \rightarrow \infty$. Τότε θα ισχύει

$$\lim_{n \rightarrow \infty} P(B_n \leq k) = e^{-\tau} \sum_{r=0}^k \frac{\tau^r}{r!}, \quad k \in \mathbf{N}_0. \quad (1.19)$$

Για $\tau = 0$ ερμηνεύουμε τη δεξιά σχέση ως 1, για $\tau = \infty$ ως 0. Αν η (1.19) ισχύει για κάποιο $k \in \mathbf{N}_0$, τότε θα ισχύει και ότι $n\bar{F}(u_n) \rightarrow \tau$ καθώς $n \rightarrow \infty$ και επομένως η (1.19) θα ισχύει για κάθε $k \in \mathbf{N}_0$.

Η προσέγγιση Poisson όπως εκφράζεται από τη (1.18) μας επιτρέπει να παράγουμε ασυμπτωτικά αποτελέσματα για τη k -οστή διατεταγμένη τιμή. Από τον ορισμό του B_n και από την (1.17) οδηγούμαστε στη σχέση

$$P(B_n < k) = P(X_{k,n} \leq u_n), \quad 1 \leq k \leq n,$$

η οποία μέσω της (1.19) μας οδηγεί στο ακόλουθο αποτέλεσμα.

1.6.7 Θεώρημα. (Οριακές πιθανότητες άνω διατεταγμένων παρατηρήσεων). Υποθέτουμε ότι (u_n) είναι μια ακολουθία στο \mathbf{R} , τέτοια ώστε να ισχύει $n\bar{F}(u_n) \rightarrow \tau$ για κάποιο $\tau \in [0, \infty]$ καθώς $n \rightarrow \infty$. Τότε θα ισχύει

$$\lim_{n \rightarrow \infty} P(X_{k,n} \leq u_n) = e^{-\tau} \sum_{r=0}^{k-1} \frac{\tau^r}{r!}, \quad k \in \mathbf{N}. \quad (1.20)$$

Για $\tau = 0$ ερμηνεύουμε τη δεξιά σχέση ως 1, για $\tau = \infty$ ως 0. Αν η (1.20) ισχύει για κάποιο $k \in \mathbf{N}$, τότε θα ισχύει και ότι $n\bar{F}(u_n) \rightarrow \tau$ καθώς $n \rightarrow \infty$ και επομένως η (1.20) θα ισχύει για κάθε $k \in \mathbf{N}$.

1.6.8 Πρόρισμα. (Οριακή κατανομή μιας διατεταγμένης παρατήρησης) Θέτουμε $u_n = c_n + d_n$ και $\tau = \tau(x) = -\ln H(x)$ και υποθέτουμε ότι $F \in \text{MDA}(H)$ με παραμέτρους κανονικοποίησης $c_n > 0$ και $d_n \in \mathbf{R}$. Τότε ορίζουμε

$$H^{(k)}(x) = H(x) \sum_{r=0}^{k-1} \frac{(-\ln H(x))^r}{r!}, \quad x \in \mathbf{R}.$$

Αν υπάρχει x τέτοιο ώστε $H(x) = 0$, τότε και το $H^{(k)}(x) = 0$. Τότε για κάθε $k \in \mathbf{N}$,

$$\lim_{n \rightarrow \infty} P(c_n^{-1}(X_{k,n} - d_n) \leq x) = H^{(k)}(x). \quad (1.21)$$

Αν τώρα για κάποιο $k \in \mathbf{N}$ ισχύει

$$\lim_{n \rightarrow \infty} P(c_n^{-1}(X_{k,n} - d_n) \leq x) = G(x)$$

για κάποια μη εκφυλισμένη σ.κ. G , τότε $G = H^{(k)}$ για κάποια κατανομή ακραίων τιμών H και η σχέση (1.21) θα ισχύει για κάθε $k \in \mathbf{N}$.

ΚΕΦΑΛΑΙΟ 2

ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΑΚΡΑΙΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

2.1 Εισαγωγή

Στο Κεφάλαιο 1 παρουσιάσαμε τη βασική θεωρία ακραίων τιμών και τα κυριότερα θεωρητικά μοντέλα που χρησιμοποιούνται. Στη πράξη η μελέτη ακραίων συμβάντων ξεκινάει από την στατιστική ανάλυση δεδομένων όπως π.χ. μεγάλες ζημιές που μπορεί να επηρεάσουν το χαρτοφυλάκιο μιας ασφαλιστικής εταιρείας, επίπεδα βροχόπτωσης, μεγάλες κυμάνσεις στις τιμές των χρηματοοικονομικών προϊόντων κ.α.. Είναι εσφαλμένη η αντίληψη ότι μπορούμε να πάρουμε οποιαδήποτε από τα παραπάνω δεδομένα και να τα εισάγουμε σε κάποιο από τα μοντέλα που αναφέραμε στο Κεφάλαιο 1. Τα μοντέλα αυτά δεν είναι παρά τα εργαλεία που θα χρησιμοποιήσει ο εκάστοτε ερευνητής, για να εξάγει ασφαλή επιστημονικά συμπεράσματα από την ανάλυση των δεδομένων. Σημαντικό κομμάτι αυτής της ανάλυσης είναι η σωστή και αντικειμενική παρουσίαση των δεδομένων, η εξέταση κατάλληλων ερωτημάτων και η απόδοση σαφών απαντήσεων που δε θα παραβιάζουν τις υποθέσεις των μοντέλων. Το ερώτημα λοιπόν είναι ποια εργαλεία θα χρησιμοποιήσουμε για να έχουμε μια πλήρη και μαθηματικά ορθή ανάλυση.

2.2 Διαγραμματικά Εργαλεία

Αν κάποιος εξετάσει τη πλειονότητα της βιβλιογραφίας πάνω στη στατιστική ανάλυση θα παρατηρήσει ότι γίνεται εκτενής χρήση διαγραμμάτων. Οι ερευνητές πάντα τονίζουν την ανάγκη μελέτης των ίδιων των δεδομένων πριν από την ανάλυση τους. Τα συνήθη διαγράμματα που χρησιμοποιούνται είναι τα ιστογράμματα (*histograms*), τα θηκογράμματα (*box-plot*), τα διαγράμματα των εμπειρικών σ.κ. και τα διαγράμματα διασποράς (*scatter plots*) στις πολυδιάστατες περιπτώσεις. Αντικείμενο της αρχικής αναζήτησης είναι η εύρεση εκείνης της σ.κ. F που αποτελεί ένα καλό μοντέλο για iid τ.μ. X_1, \dots, X_n . Αυτές οι τ.μ. ορίζουνε το διατε-

ταγμένο δείγμα $X_{n,n} \leq \dots \leq X_{1,n}$ και βάσει αυτού θα σχεδιάσουμε ένα γράφημα πιθανοτήτων (*PP plot*).

Το βασικό θεωρητικό κομμάτι για τον σχεδιασμό του *PP-plot*, δίνεται από το Λήμμα 1.6.4. στο οποίο δηλώνεται ότι αν η F είναι συνεχής, οι τ.μ. $F(X_i) = U_i$ για $i = 1, \dots, n$, θα ακολουθούν την $U(0,1)$. Επιπλέον θα ισχύει

$$(F(X_{k,n}))_{k=1,\dots,n} \stackrel{d}{=} (U_{k,n})_{k=1,\dots,n}$$

από το οποίο συνεπάγεται ότι

$$EF(X_{k,n}) = \frac{n-k+1}{n+1}, \quad k = 1, \dots, n$$

και στη περίπτωση που εξετάζουμε την εμπειρική συνάρτηση κατανομής F_n της F έχουμε

$$F_n(X_{k,n}) = \frac{n-k+1}{n}, \quad k = 1, \dots, n.$$

Το γράφημα το οποίο αποτελείται από τα σημεία,

$$\left\{ \left(F(X_{k,n}), \frac{n-k+1}{n+1} \right) : k = 1, \dots, n \right\}$$

καλείται *PP plot*, ενώ το γράφημα το οποίο αποτελείται από τα σημεία

$$\left\{ \left(X_{k,n}, F^{\leftarrow} \left(\frac{n-k+1}{n+1} \right) \right) : k = 1, \dots, n \right\} \quad (2.1)$$

καλείται γράφημα ποσοστιαίων σημείων (*QQ plot*). Το *PP plot* ή το *QQ plot* χρησιμοποιείται για να εξετάσουμε αν πράγματι οι παρατηρήσεις προέρχονται από τη δεδομένη F . Πράγματι, σύμφωνα με τα παραπάνω, αν το τ.δ. προέρχεται από την F τότε τα n σημεία στα γραφήματα αυτά θα πρέπει να βρίσκονται κοντά στη διαγώνιο.

Στη βιβλιογραφία έχουν δοθεί κάποιες παραλλαγές του τύπου (2.1) της μορφής

$$\{(X_{k,n}, F^{\leftarrow}(p_{k,n})) : k = 1, \dots, n\}$$

όπου $p_{k,n}$ είναι κάποια συγκεκριμένη θέση μέσα στο διάγραμμα με συνηθέστερη επιλογή την

$$p_{k,n} = \frac{n-k+\delta_k}{n+\gamma_k},$$

όπου (δ_k, γ_k) είναι κατάλληλα επιλεγμένα για να επιτρέπουν τη διόρθωση συνέχειας των δεδομένων. Στη πλειονότητα των περιπτώσεων χρησιμοποιείται ο τύπος

$$p_{k,n} = \frac{n-k+0,5}{n}.$$

Συνήθως τα δεδομένα ελέγχονται με βάση μια οικογένεια κατανομών της μορφής $F((\cdot - \mu) / \psi)$, όπου σε κάποιες περιπτώσεις, όπως στη περίπτωση της Τυποποιημένης Κανονικής, μ και ψ είναι ο μέσος και η τυπική απόκλιση του X αντίστοιχα.

Η βασική ιδέα για τη χρησιμοποίηση των QQ plot προέκυψε από τη παρατήρηση ότι για κάποιες οικογένειες κατανομών τα ποσοστιαία σημεία συνδέονται γραμμικά με τα αντίστοιχα ποσοστιαία σημεία μιας τυπικής κατανομής από αυτή τη συγκεκριμένη οικογένεια. Επειδή η γραμμικότητα ενός γραφήματος μπορεί να ελεγχθεί εύκολα, με τη χρήση των QQ plot μπορούμε να προσδιορίσουμε αν ένα συγκεκριμένο μοντέλο έχει καλή προσαρμογή πάνω στη κατανομή του τ.δ. που έχουμε στη διάθεση μας.

Κατά την ανάλυση ακραίων τιμών, η συνήθης πρακτική είναι τα QQ plot να έχουν ως βάση τα ποσοστιαία σημεία της κατανομής Gumbel. Για παράδειγμα, αν έχουμε n μέγιστες τιμές (π.χ. από n σύνολα παρατηρήσεων), και επιθυμούμε να ελέγξουμε αν αυτές προέρχονται από την Gumbel, μπορούμε να ελέγξουμε αν τα n σημεία του QQ plot

$$\{(X_{k,n}, \Lambda^{\leftarrow}(p_{k,n})) : k = 1, \dots, n\} = \{(X_{k,n}, -\ln(-\ln p_{k,n})) : k = 1, \dots, n\}$$

βρίσκονται κοντά σε μια ευθεία γραμμή με θετική κλίση. Το παραπάνω QQ plot καλείται διπλό λογαριθμικό (*double logarithmic plots*). Αν το γράφημα παρουσιάζει κοίλη (*concave*) απόκλιση από την ευθεία γραμμή, τότε υπάρχει η ένδειξη ότι οι παρατηρήσεις προέρχονται από κάποια κατανομή με βαριά ουρά. Αντίθετα στη περίπτωση που υπάρχει κυρτή (*convex*) απόκλιση από την ευθεία, τότε υπάρχει ένδειξη ότι οι παρατηρήσεις προέρχονται από κάποια κατανομή με ελαφριά ουρά.

Στη περίπτωση της GEV $H_{\xi;\mu;\psi}$ που ορίζεται ως

$$H_{\xi;\mu;\psi}(x) = \exp\left(-\left(1 + \xi \frac{x - \mu}{\psi}\right)^{-1/\xi}\right), \quad 1 + \xi(x - \mu) / \psi > 0 \quad (2.2)$$

εκτός από τη παράμετρο θέσης μ , κλίμακας ψ εισάγεται και η παράμετρος σχήματος ξ , γεγονός που κάνει την ερμηνεία του QQ -plot αρκετά δύσκολη. Στη συγκεκριμένη περίπτωση χρησιμοποιούμε ένα QQ -plot το οποίο αποτελείται από τα σημεία

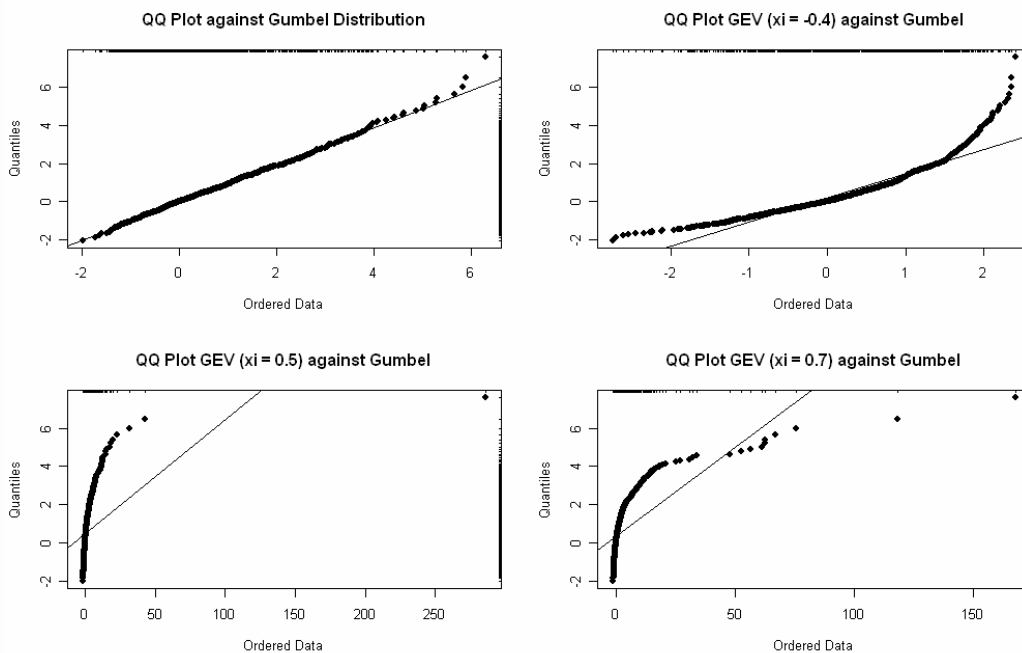
$$\left\{ \left(X_{k,n}, H_{\xi;0,1}^{\leftarrow}(p_{k,n}) = -\frac{1}{\xi} (1 - (-\log(p_{k,n}))^{-\xi}) \right) : k = 1, \dots, n \right\},$$

όπου το ξ εκτιμάται με κάποιο τρόπο από τα δεδομένα. Στη συνέχεια τα μ , ψ εκτιμώνται εμπειρικά από το QQ plot. Το ψ εκτιμάται από την κλίση (*slope*) της ευθείας που σχηματίζουν τα δεδομένα και το μ από το ύψος της τομής του κάθετου άξονα και της ευθείας που σχηματίζουν τα δεδομένα (*intercept*).

2.2.1 Παράδειγμα (*QQ plot* έναντι της κατανομής Gumbel)

Κατασκευάσαμε τα *QQ plot* προσομοιωμένων παρατηρήσεων ($n = 1000$) της GEV με παράμετρο $\xi = -0.4$, $\xi = 0.5$ και $\xi = 0.7$ αντίστοιχα, έναντι της Gumbel ($n = 1000$) (ΣΧΗΜΑ 2-1). Στο πρώτο γράφημα παρατηρούμε τη πολύ καλή προσαρμογή της Gumbel πάνω στην ευθεία γραμμή. Στο δεύτερο γράφημα όπου $\xi < 0$, δηλαδή οι μέγιστες τιμές ακολουθούν την Weibull, βλέπουμε την ύπαρξη κυρτής απόκλισης από την ευθεία γραμμή. Στα δυο τελευταία γραφήματα όπου $\xi > 0$, δηλαδή οι μέγιστες τιμές ακολουθούν την Fréchet, βλέπουμε την ύπαρξη κοίλης απόκλισης από την ευθεία γραμμή.

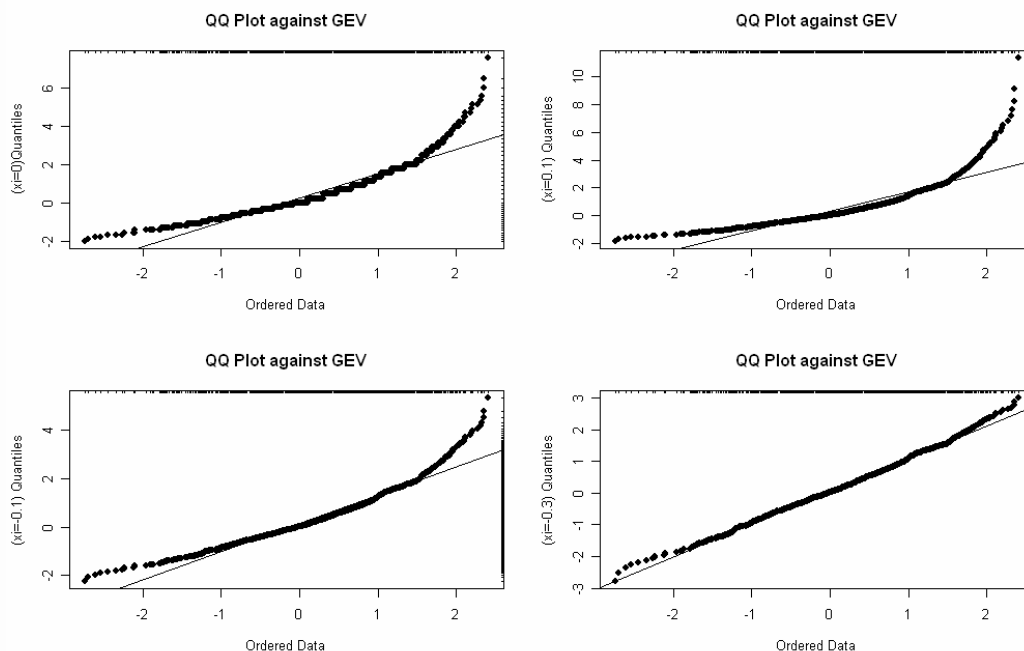
ΣΧΗΜΑ 2-1



2.2.2 Παράδειγμα (*QQ plot* έναντι της GEV)

Εξετάζοντας τα *QQ plot* των προσομοιωμένων παρατηρήσεων ($n = 1000$) της GEV με παράμετρο $\xi = -0.4$ έναντι της GEV με παράμετρο $\xi = 0$, $\xi = 0.1$, $\xi = -0.1$ και $\xi = -0.3$ αντίστοιχα (ΣΧΗΜΑ 2-2), παρατηρούμε ότι όσο το ξ πλησιάζει προς το -0.4 το *QQ plot* γίνεται πιο γραμμικό, γεγονός που μας οδηγεί στο συμπέρασμα ότι υπάρχει καλή προσαρμογή του μοντέλου πάνω στις παρατηρήσεις.

ΣΧΗΜΑ 2-2



2.3 Υπερβάλλουσα συνάρτηση κατανομής και μέση υπερβάλλουσα συνάρτηση

Υπενθυμίζεται ότι, για ένα σταθερό $u < x_F$, η $e(u) = E(X - u | X > u)$ καλείται μέση υπερβάλλουσα συνάρτηση (*mean excess function*) της τ.μ. X . Μπορούμε να εφαρμόσουμε ένα διαγραμματικό έλεγχο για τη συμπεριφορά της ουράς μιας κατανομής, βασιζόμενοι στην εμπειρική μέση υπερβάλλουσα συνάρτηση $e_n(u)$. Αν X_1, \dots, X_n iid τ.μ. με σ.κ. F και εμπειρική σ.κ. F_n και ορίσουμε $\Delta_n(u) = \{i : i = 1, \dots, n, X_i > u\}$ τότε

$$e_n(u) = \frac{1}{F_n(u)} \int_u^\infty \bar{F}_n(y) dy = \frac{1}{\text{card} \Delta_n(u)} \sum_{i \in \Delta_n(u)} (X_i - u), \quad u \geq 0$$

με τη παραδοχή ότι $0/0 = 0$. Ένα διάγραμμα μέσης υπερβάλλουσας συνάρτησης (*mean excess plot, ME plot*) ή μέσης υπολειπόμενης ζωής (*mean residual life plot, MRL plot*) αποτελείται από τα σημεία

$$\{(X_{k,n}, e_n(X_{k,n})) : k = 1, \dots, n\}.$$

Η μέθοδος του *ME plot* χρησιμοποιείται μόνο διαγνωστικά, για να διαχωρίσουμε τα μοντέλα ανάλογα με το αν οι υπό εξέταση κατανομές έχουν ελαφριά ή βαριά ουρά. Κατά την ερμηνεία αυτών των διαγραμμάτων χρειάζεται να είμαστε ιδιαίτερα προσεκτικοί. Για τον υπολογισμό της $e_n(u)$ για μεγάλες τιμές του ορίου u , οι παρατηρήσεις που έχουμε στη διάθεση μας είναι

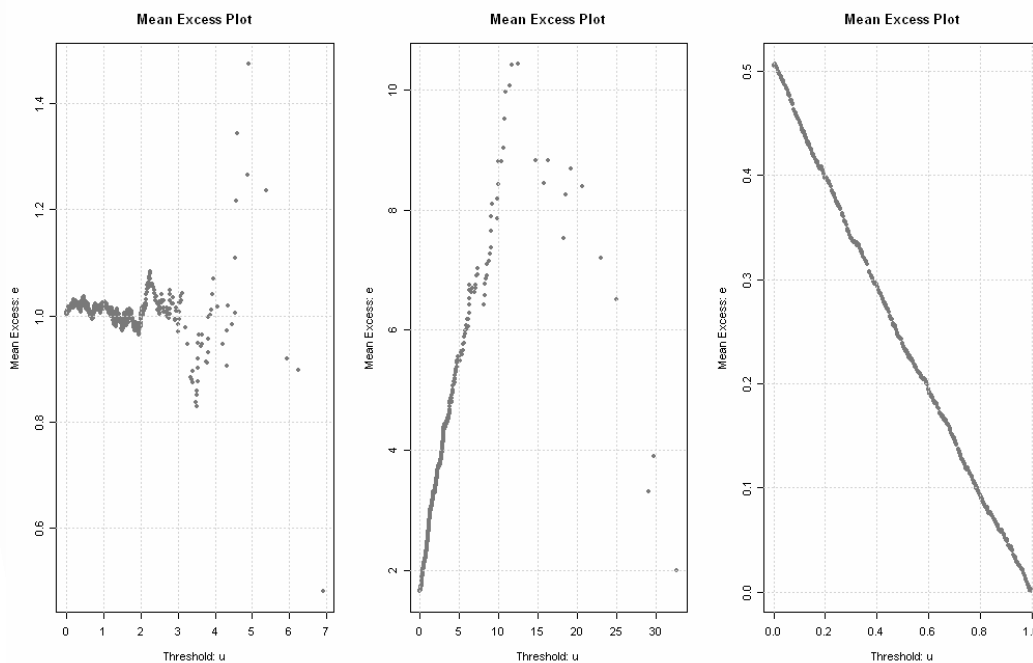
λίγες και αυτό έχει ως αποτέλεσμα τα διαγράμματα να παρουσιάζουν μεγάλη ευαισθησία όσο πλησιάζουμε στο τέλος του εύρους των δεδομένων.

2.3.1 Παράδειγμα

Σχεδιάσαμε τα *ME plot* (διαγράμματα σε όλο το σύνολο των παρατηρήσεων) των προσομοιωμένων παρατηρήσεων ($n = 1000$) των $\text{Exp}(1)$, $\text{Pareto}(1,7, 2)$ και $U(0,1)$ αντίστοιχα (ΣΧΗΜΑ 2-3). Παρατηρούμε ότι το διάγραμμα της $\text{Exp}(1)$ παρουσιάζει σταθερή συμπεριφορά γύρω από μια γραμμή με σχεδόν μηδενική κλίση, το οποίο αποτελεί τυπική συμπεριφορά κατανομής με εκθετική ουρά. Το διάγραμμα της $\text{Pareto}(1,7, 2)$ έχει ανοδική πορεία, γεγονός που υποδεικνύει ότι οι παρατηρήσεις ακολουθούν κατανομή με βαριά ουρά. Τέλος το διάγραμμα της $U(0,1)$ παρουσιάζει καθοδική πορεία, γεγονός που υποδεικνύει ότι οι παρατηρήσεις ακολουθούν κατανομή με ελαφριά ουρά. Πρέπει να επισημανθεί ότι στα δυο πρώτα διαγράμματα, υπάρχει ασταθής συμπεριφορά καθώς το κατώφλι u παίρνει μεγαλύτερες τιμές, γεγονός που δυσχεραίνει στην ερμηνεία των διαγραμμάτων.

ΣΧΗΜΑ 2-3

(Από δεξιά προς τα αριστερά) *ME plot* των προσομοιωμένων παρατηρήσεων των $\text{Exp}(1)$, $\text{Pareto}(1,7, 2)$ και $U(0,1)$



2.4 Εκτίμηση παραμέτρων για την GEV

2.4.1 Εκτίμηση με τη μέθοδο μέγιστης πιθανοφάνειας (*Maximum Likelihood Estimation*)

Ορίσαμε τη GEV στη (2.2) ως

$$H_{\xi;\mu,\psi}(x) = \exp\left(-\left(1 + \xi \frac{x - \mu}{\psi}\right)^{-1/\xi}\right), \quad 1 + \xi(x - \mu) / \psi > 0$$

Η παράμετρος $\theta = (\xi, \mu, \psi) \in \mathbf{R} \times \mathbf{R} \times \mathbf{R}_+$, αποτελείται από τη παράμετρο θέσης μ , τη παράμετρο κλίμακας ψ και τη παράμετρο σχήματος ξ . Όπως είδαμε και στο πρώτο κεφάλαιο, η H_ξ είναι η οριακή κατανομή των κανονικοποιημένων μεγίστων iid τ.μ.. Αν υποθέσουμε ότι τα δεδομένα ικανοποιούν τη συνθήκη

$$X_1, \dots, X_n \text{ iid από } H_\theta \quad (2.3)$$

τότε μπορούμε να προχωρήσουμε σε παραμετρική εκτίμηση μέσω της κλασσικής στατιστικής μεθοδολογίας. Η συγκεκριμένη υπόθεση δεν είναι ορθή, αλλά ισχύει προσεγγιστικά (όταν κάθε X_i είναι η μέγιστη τιμή από ένα μεγάλο πλήθος παρατηρήσεων). Επειδή μας επιτρέπει να εκτιμήσουμε εύκολα τις παραμέτρους, στη συνέχεια θα θεωρούμε ότι (σχεδόν) ισχύει.

Στην υδρολογία είναι δυνατόν να πάρουμε δεδομένα της μορφής (2.3) αν μπορούμε να εκφράσουμε τα X_i ως μέγιστες τιμές ενός δείγματος μιας ανεξάρτητης χρονικής περιόδου μεγέθους s , η οποία επιλέγεται συνήθως ίση με 365 μέρες. Το διάστημα του ενός έτους επιλέγεται για να αποφευχθεί η εποχικότητα των περιόδων του έτους. Επομένως τα αρχικά δεδομένα έχουν τη μορφή

$$\begin{aligned} X^{(1)} &= (X_1^{(1)}, \dots, X_s^{(1)}) \\ X^{(2)} &= (X_1^{(2)}, \dots, X_s^{(2)}) \\ &\vdots \\ X^{(n)} &= (X_1^{(n)}, \dots, X_s^{(n)}) \end{aligned}$$

όπου τα διανύσματα $(X^{(i)})$ υποθέτουμε ότι είναι iid, αλλά μέσα σε κάθε διάνυσμα τα στοιχεία $X^{(i)}$ δεν είναι απαραίτητο να είναι ανεξάρτητα. Η χρονική περίοδος s επιλέγεται έτσι ώστε οι παραπάνω συνθήκες να ικανοποιούνται. Το βασικό iid δείγμα από τη H_θ πάνω στο οποίο θα βασιστούμε για να κάνουμε στατιστικές εκτιμήσεις θα αποτελείται από

$$X_i = \max(X_1^{(i)}, \dots, X_s^{(i)}), \quad i = 1, \dots, n. \quad (2.4)$$

Η διαδικασία εκτίμησης σε δεδομένα της μορφής (2.4) καλείται προσαρμογή ετησίων μεγίστων (*fitting of annual maxima*) ή πιο γενικά των μεγίστων των περιόδων (*block maxima analysis*). Η μορφή (2.3) αντιστοιχεί στη τυπική παραμετρική περίπτωση της στατιστικής εκτίμησης και επομένως στη πλειονότητα των περιπτώσεων θα επιλύεται με τη χρήση της με-

θόδου της μέγιστης πιθανοφάνειας. Υποθέτουμε λοιπόν ότι η H_θ έχει πυκνότητα h_θ . Η συνάρτηση πιθανοφάνειας που προκύπτει από τα δεδομένα $X = (X_1, \dots, X_n)$ δίνεται από τον τύπο

$$L(\theta; X) = \prod_{i=1}^n h_\theta(X_i) I_{\{1+\xi(X_i-\mu)/\psi > 0\}}.$$

Ορίζουμε με $l(\theta; X) = \ln L(\theta; X)$ τη συνάρτηση λογάριθμο-πιθανοφάνειας. Ο εκτιμητής μέγιστης πιθανοφάνειας (*maximum likelihood estimator* – MLE) του θ θα είναι ίσος με

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} l(\theta; X),$$

δηλαδή η $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ θα μεγιστοποιεί την $l(\theta; X)$ σε ένα κατάλληλο παραμετρικό χώρο Θ . Η συνάρτηση λογάριθμο-πιθανοφάνειας της $H_{\xi, \mu, \psi}$ όταν $\xi \neq 0$, είναι

$$l((\xi, \mu, \psi); X) = -n \ln \psi - (1 + 1/\xi) \sum_{i=1}^n \ln \left\{ 1 + \xi \left(\frac{X_i - \mu}{\psi} \right) \right\} - \sum_{i=1}^n \left\{ 1 + \xi \left(\frac{X_i - \mu}{\psi} \right) \right\}^{-1/\xi}, \quad (2.5)$$

όπου $1 + \xi (X_i - \mu) / \psi > 0$, $i = 1, \dots, n$. Στη περίπτωση που εξετάζουμε την $H_{0, \mu, \psi}$ (*Gumbel*), τότε θα πάρουμε ότι

$$l((0, \mu, \psi); X) = -n \ln \psi - \sum_{i=1}^n \exp \left\{ -\frac{X_i - \mu}{\psi} \right\} - \sum_{i=1}^n \frac{X_i - \mu}{\psi}.$$

Παίρνοντας μερικές παραγώγους ως προς μ και ψ στη παραπάνω συνάρτηση, προκύπτουν οι ακόλουθες εξισώσεις

$$0 = n - \sum_{i=1}^n \exp \left\{ -\frac{X_i - \mu}{\psi} \right\},$$

$$0 = n + \sum_{i=1}^n \frac{X_i - \mu}{\psi} \left(\exp \left\{ -\frac{X_i - \mu}{\psi} \right\} - 1 \right).$$

Είναι προφανές ότι δεν υπάρχουν αναλυτικές εκφράσεις για τα μ , ψ που ικανοποιούν αυτές τις εξισώσεις. Αν εξετάσουμε τη περίπτωση H_ξ όπου $\xi \neq 0$, τότε θα βρεθούμε αντιμέτωποι με ακόμα πιο περίπλοκες εξισώσεις. Όπως αναφέραμε και παραπάνω, η χρήση MLE ισχύει στη πλειονότητα των περιπτώσεων και μας δίνει αξιόπιστους και συνεπείς εκτιμητές, οι οποίοι ακολουθούν την κανονική κατανομή ασυμπτωτικά. Στις περιπτώσεις που το στήριγμα της αρχικής σ.κ. αποτελείται από άγνωστες παραμέτρους, εφαρμόζονται κάποιες αξιόπιστες αριθμητικές μέθοδοι για την εκτίμηση του MLE του $\hat{\theta}_n$, αλλά αν το υπό εξέταση δείγμα είναι αρκετά μικρό, δε μπορούμε να είμαστε σίγουροι για τις ιδιότητες αυτού του εκτιμητή. Έχει αποδειχθεί ότι οι “καλές” ιδιότητες του MLE ισχύουν όταν $\xi > -1/2$ (βλ. Smith (1985)). Όταν το ξ βρίσκεται στο διάστημα $(-1, -1/2]$, οι εκτιμητές μέγιστης πιθανοφάνειας δεν ακολου-

θούν την κανονική κατανομή ασυμπτωτικά. Αν $\xi < -1$ τότε θεωρείται απίθανο να προκύψουν εκτιμήσεις από τη μέθοδο της μέγιστης πιθανοφάνειας. Αν ισχύει ο περιορισμός ότι $\xi > -1/2$, τότε οι εκτιμώμενες παράμετροι $(\hat{\mu}, \hat{\psi}, \hat{\xi})$ αποδεικνύεται ότι ακολουθούν κατά προσέγγιση την πολυμεταβλητή κανονική κατανομή με μέσο (μ, ψ, ξ) και πίνακα διακυμάνσεων – συνδιακυμάνσεων ίσο με τον αντίστροφο του πίνακα πληροφορίας Fisher. Όμοια, στη περίπτωση που εξετάζουμε τη κατανομή Gumbel, αν το μέγεθος του δείγματος n είναι αρκετά μεγάλο, τότε αποδεικνύεται ότι οι εκτιμώμενες παράμετροι $(\hat{\mu}, \hat{\psi})$ ακολουθούν πολυμεταβλητή κανονική με μέσο (μ, ψ) και πίνακα διακυμάνσεων - συνδιακυμάνσεων ίσο με τον αντίστροφο του πίνακα πληροφορίας Fisher. Τα στοιχεία της κυρίας διαγωνίου του πίνακα διακυμάνσεων συνδιακυμάνσεων, αντιστοιχούν στις διακυμάνσεις των αντίστοιχων παραμέτρων, επιτρέποντάς μας να πάρουμε δ.ε. για τις παραμέτρους. Η τεχνική του MLE είναι χρήσιμη στις ασφαλιστικές και χρηματοοικονομικές εφαρμογές καθώς σε αυτά τα πεδία θεωρούμε ότι οι υπό εξέταση συναρτήσεις δεν έχουν δεξιό φράγμα, δηλαδή $\xi \geq 0$.

Στη περίπτωση που δεν επιθυμούμε να εξετάσουμε τη μεγαλύτερη παρατήρηση, όπως προκύπτει από τα δειγματικά δεδομένα, αλλά τις k μεγαλύτερες παρατηρήσεις μέσα σε κάθε περίοδο, τότε αυτή η ποσότητα δεδομένων θα αντιστοιχούσε με

$$X_{k,s}^{(i)} \leq \dots \leq X_{1,s}^{(i)} = X_i, i = 1, \dots, n.$$

Βάσει της θεωρίας της μέγιστης πιθανοφάνειας, για αυτές τις $k \times n$ παρατηρήσεις θα πρέπει να χρησιμοποιήσουμε την από κοινού πυκνότητα των ανεξάρτητων διανυσμάτων

$$(X_{k,s}^{(i)}, \dots, X_{1,s}^{(i)}), i = 1, \dots, n.$$

Στη πλειονότητα των περιπτώσεων δεν είναι απαραίτητο να υποθέσουμε ότι για κάθε i τα παραπάνω διανύσματα προέρχονται από iid δεδομένα. Αν ίσχυε αυτό, τότε η εκτίμηση μέγιστης πιθανοφάνειας θα στηριζόταν στην από κοινού πυκνότητα των k άνω διατεταγμένων τιμών μιας GEV κατανομής, δηλαδή

$$\frac{s!}{(s-k)!} H_{\theta}^{s-k}(x_k) \prod_{\ell=i}^k h_{\theta}(x_{\ell}), x_k < \dots < x_1.$$

Ανάλογα με τη τιμή του θ , οι τιμές του x θα ικανοποιούν τους διάφορους περιορισμούς. Σαν συμπέρασμα των παραπάνω μπορούμε να πούμε ότι το πλεονέκτημα της διαδικασίας MLE είναι ότι μπορεί να εφαρμοσθεί σε ένα μεγάλο πεδίο κατανομών τροποποιώντας ελάχιστα τις αρχικές μας υποθέσεις και τη μεθοδολογία.

2.4.1.1 Παράδειγμα

Προσομοιώσαμε $n = 20000$ παρατηρήσεις από τις $\text{Exp}(1)$, $\text{Cauchy}(a = 0, b = 1)$ και $\text{U}(0,1)$. Σε κάθε ένα από αυτά τα σύνολα, χωρίσαμε τις παρατηρήσεις σε εκατό (100) ισομεγέθη υποσύνολα ($n_i = 200$) και θεωρήσαμε τα αντίστοιχα *block maxima*. Οι εκτιμηθείσες παράμετροι παραθέτονται στον ακόλουθο πίνακα.

ΠΙΝΑΚΑΣ 2-1

Εκτιμώμενες παράμετροι με τη μέθοδο MLE

Κατανομή	Παράμετρος		
	$\hat{\xi}$	$\hat{\mu}$	$\hat{\psi}$
$\text{Exp}(1)$	-0.01738893	4.65371365	1.00127219
$\text{Cauchy}(0, 1)$	1.101327	30.645080	32.490178
$\text{U}(0, 1)$	-0.960917113	0.990106384	0.009506422

Εκ πρώτης όψεως μπορούμε να υποστηρίξουμε ότι η μέθοδος MLE δίνει καλές εκτιμήσεις για όλες τις παραπάνω περιπτώσεις. Στην περίπτωση της εκθετικής κατανομής μπορούμε να χρησιμοποιήσουμε τη συνάρτηση λογάριθμο-πιθανοφάνειας $l((0, \mu, \psi); X)$ αντί της $l((\xi, \mu, \psi); X)$ λαμβάνοντας τις εκτιμήσεις $(\hat{\mu}, \hat{\psi}) = (4.6441831, 0.9955024)$. Αναφορικά με την $\text{U}(0,1)$, πρέπει να επισημανθεί ότι τα συγκεκριμένα αποτελέσματα προέκυψαν μετά από πολλές διαδοχικές ανεπιτυχείς επαναλήψεις της διαδικασίας εκτίμησης (ορισμένες φορές η επαναληπτική προσεγγιστική μέθοδος δεν συνέκλινε). Όπως αναφέρθηκε στο Παράδειγμα 1.3.8, η οριακή κατανομή της $\text{U}(0,1)$ είναι η Weibull με $\xi = -1$. Όπως έχουμε επισημάνει παραπάνω, στην περίπτωση αυτή δεν είναι πάντοτε δυνατό να προκύψουν εκτιμήτριες μέγιστης πιθανοφάνειας, και αν προκύψουν δε θα είναι αξιόπιστες.

2.4.2 Στάθμη απόδοσης (Return level)

Ας ορίσουμε ως z_p το άνω p ποσοστιαίο σημείο της GEV, ο εκτιμητής του οποίου είναι

$$z_p = \begin{cases} \mu - \frac{\psi}{\xi} [1 - \{-\log(1-p)\}^{-\xi}], & \xi \neq 0 \\ \mu - \psi \log\{-\log(1-p)\}, & \xi = 0 \end{cases}, \quad H_{(\xi; \mu, \psi)}(z_p) = 1 - p \quad (2.6)$$

και προκύπτει από την αντιστροφή της (2.2). Το ποσοστιαίο σημείο z_p είναι η στάθμη απόδοσης, δηλαδή το επίπεδο εκείνο που αναμένεται να ξεπεραστεί μια φορά κάθε $1/p$ περιόδους. Το διάστημα των $1/p$ περιόδων καλείται περίοδος απόδοσης (*return period*). Ορίζοντας ως y_p το άνω p ποσοστιαίο σημείο της τυπικής εκθετικής κατανομής, δηλαδή $y_p = -\log(1-p)$, και αντικαθιστώντας στη (2.6), καταλήγουμε στη

$$z_p = \begin{cases} \mu - \frac{\psi}{\xi} [1 - y_p^{-\xi}], & \xi \neq 0 \\ \mu - \psi \log y_p, & \xi = 0 \end{cases} \quad (2.7)$$

Η (2.7) μας επιτρέπει να ελέγξουμε διαγραμματικά τα ποσοστιαία σημεία της GEV έναντι αυτών της εκθετικής, μέσω ενός λογαριθμικού γραφήματος *QQ Plot*, τα σημεία του οποίου ορίζονται ως

$$\{(-\log y_p, z_p) : p \in (0,1)\}.$$

Το γράφημα αυτό καλείται διάγραμμα στάθμης απόδοσης (*return level plot*). Αποδεικνύεται ότι αν $\xi = 0$, το γράφημα είναι γραμμικό, αν $\xi < 0$ το γράφημα είναι κοίλο με ασυμπτωτικό όριο $\mu - \psi / \xi$, καθώς $p \rightarrow 0$, ενώ αν $\xi > 0$ το γράφημα είναι κυρτό και δεν έχει πεπερασμένο όριο. Τα διαγράμματα αυτά είναι κατάλληλα τόσο για την επαλήθευση όσο και την απεικόνιση των αποτελεσμάτων.

Αντικαθιστώντας τις εκτιμήσεις που προέκυψαν από τη μέθοδο της μέγιστης πιθανοφάνειας στη (2.7), τότε η εκτίμηση μέγιστης πιθανοφάνειας του z_p , για $0 < p < 1$, θα είναι

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\psi}}{\hat{\xi}} [1 - y_p^{-\hat{\xi}}], & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\psi} \log y_p, & \hat{\xi} = 0 \end{cases}.$$

Επιπλέον, εφαρμόζοντας τη γνωστή μέθοδο δέλτα (*delta method*) μπορούμε να εκτιμήσουμε την διακύμανση των ποσοστιαίων σημείων

$$\text{Var}(\hat{z}_p) = \nabla_{z_p}^T \cdot V \cdot \nabla_{z_p},$$

όπου V είναι ο πίνακας διακυμάνσεων-συνδιακυμάνσεων των $(\hat{\mu}, \hat{\psi}, \hat{\xi})$ και

$$\nabla_{z_p}^T = \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \psi}, \frac{\partial z_p}{\partial \xi} \right] = \left[1, -\xi^{-1} (1 - y_p^{-\xi}), \psi \xi^{-2} (1 - y_p^{-\xi}) - \psi \xi^{-1} y_p^{-\xi} \log y_p \right],$$

αντικαθιστώντας τα μ, ψ, ξ στον παραπάνω τύπο από τα $(\hat{\mu}, \hat{\psi}, \hat{\xi})$. Στη συνέχεια μπορούμε να κατασκευάσουμε δ.ε. για τα ποσοστιαία σημεία z_p .

Αν σχεδιάσουμε ένα διάγραμμα στάθμης απόδοσης τα σημεία του οποίου ορίζονται ως,

$$\{(-\log y_p, \hat{z}_p) : p \in (0,1)\}$$

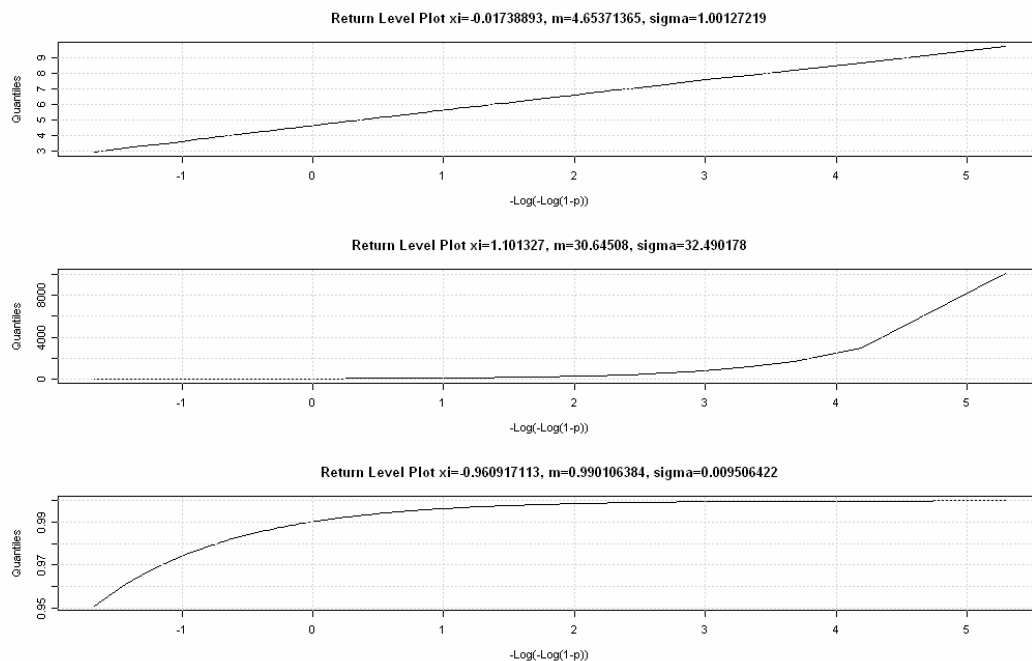
μπορούμε να ελέγξουμε και οπτικά σε ποια οικογένεια κατανομών ακραίων τιμών ανήκουν οι παρατηρήσεις που προσαρμόσαμε.

2.4.2.1 Παράδειγμα

Χρησιμοποιώντας τις εκτιμήσεις που προέκυψαν από την προσαρμογή των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ στη GEV (§ 2.4.1) αναπαραστήσαμε γραφικά τις αντίστοιχες στάθμες απόδοσης (ΣΧΗΜΑ 2-4):

ΣΧΗΜΑ 2-4

Από πάνω προς τα κάτω: Διαγράμματα στάθμης απόδοσης για τα μέγιστα περιόδων της $\text{Exp}(1)$, της $\text{Cauchy}(0,1)$ και της $\text{U}(0,1)$.



Τα διαγράμματα που προέκυψαν ουσιαστικά επιβεβαιώνουν το γεγονός ότι οι οριακές κατανομές των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ ακολουθούν κατανομές με ελαφριά, βαριά και μεσαία ουρά αντίστοιχα. Βεβαίως στη περίπτωση της $\text{U}(0,1)$ πρέπει να είμαστε επιφυλακτικοί ως προς τα συμπεράσματα μας, δεδομένου ότι οι εκτιμώμενες παράμετροι, από τις οποίες προέκυψαν και τα ποσοστιαία σημεία δεν είναι αξιόπιστες.

2.4.3 Εκτίμηση των υψηλών ποσοστιαίων σημείων μέσω της κατατομής πιθανοφάνειας (Profile Likelihood)

Μια μέθοδος που παρέχει πιο αξιόπιστα διαστήματα εμπιστοσύνης είναι αυτή της κατατομής πιθανοφάνειας. Η συνάρτηση λογάριθμο-πιθανοφάνειας ορίζεται ισοδύναμα και ως $l(\theta_i; \theta_{-i})$, όπου θ_{-i} περιλαμβάνει όλα τα στοιχεία του θ εκτός του θ_i . Η συνάρτηση κατατομής λογάριθμο-πιθανοφάνειας για το θ_i ορίζεται ως

$$\ell_p(\theta_i) = \max_{\theta_{-i}} \ell(\theta_i, \theta_{-i}).$$

Δηλαδή για κάθε τιμή του θ_i , η συνάρτηση κατατομής λογάριθμο-πιθανοφάνειας είναι η μεγιστοποιημένη λογάριθμο-πιθανοφάνεια ως προς τα υπόλοιπα στοιχεία του θ . Στη περίπτωση της GEV, η κατατομή πιθανοφάνειας για οποιαδήποτε από τις παραμέτρους μ , ψ και ξ εκτιμάται ως ακολούθως. Έστω ότι ενδιαφερόμαστε για την εκτίμηση του ξ , ορίζουμε ως $\theta^{(1)} = (\xi)$ και $\theta^{(2)} = (\mu, \psi)$, θέτουμε $\xi = \xi_0$ και μεγιστοποιούμε ως προς τις παραμέτρους μ και ψ ,

$$\ell_p(\theta^{(1)}) = \max_{\theta^{(2)}} \ell(\theta^{(1)}, \theta^{(2)})$$

Η συγκεκριμένη διαδικασία επαναλαμβάνεται για ένα εύρος τιμών του ξ_0 (π.χ. 0.1, 0.11, ..., 0.2). Οι μεγιστοποιημένες τιμές που προκύπτουν δίνουν μια προσέγγιση της κατατομής πιθανοφάνειας του ξ στο διάστημα που κινείται το ξ_0 . Αποδεικνύεται ότι αν $\hat{\theta}_0$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας του μοντέλου $\theta_0 = (\theta^{(1)}, \theta^{(2)})$, τότε κάτω από ορισμένες συνθήκες και για μεγάλο μέγεθος του αρχικού δείγματος n , ισχύει

$$D_p(\theta^{(1)}) = 2\{\ell(\hat{\theta}_0) - \ell_p(\theta^{(1)})\} \sim \chi_1^2$$

από όπου μπορούμε να κατασκευάσουμε δ.ε..

Μπορούμε να εφαρμόσουμε την παραπάνω διαδικασία και στην περίπτωση που θέλουμε να πάρουμε εκτιμήσεις για μία συνάρτηση των παραμέτρων, όπως το ποσοστιαίο σημείο z_p . Αρκεί να εκφράσουμε το z_p ως μια από τις παραμέτρους της GEV. Από τη (2.6) μπορεί να εύκολα να προκύψει ότι

$$\mu = z_p + \frac{\psi}{\xi} [1 - \{-\log(1-p)\}^{-\xi}]. \quad (2.8)$$

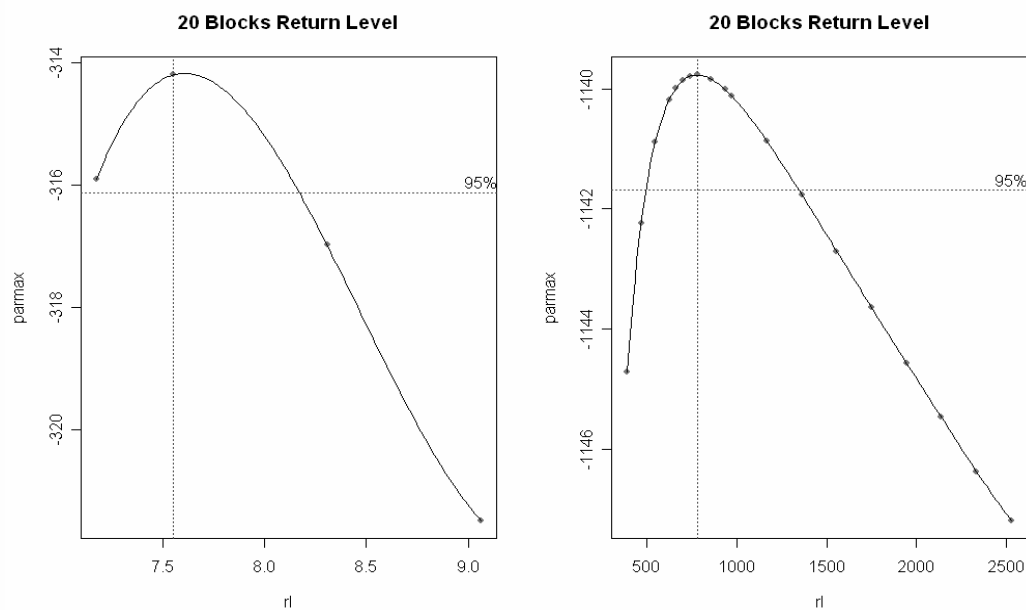
Αντικαθιστώντας τη (2.8) στη (2.5) μπορούμε να εκφράσουμε τη GEV ως συνάρτηση των (z_p, μ, ψ) και να εκτιμήσουμε τα υψηλά ποσοστιαία σημεία καθώς και να κατασκευάσουμε διαστήματα εμπιστοσύνης. Η εκτίμηση του z_p , μας επιτρέπει να προσδιορίσουμε το επίπεδο απόδοσης που αναμένεται να ξεπεραστεί μια φορά στις $1/p$ επόμενες περιόδους.

2.4.3.1 Παράδειγμα

Χρησιμοποιώντας τις εκτιμήσεις μέγιστης πιθανοφάνειας των παραμέτρων που προέκυψαν από τις προσομοιωμένες παρατηρήσεις ($n = 20000$) των $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $U(0,1)$ στη GEV (§ 2.4.1) με τη μέθοδο *block maxima*, προχωρήσαμε σε εκτίμηση της στάθμης απόδοσης z_p , όπως περιγράψαμε πιο πάνω στα επόμενα $1/p = 20$ *block*. Στο παρακάτω διάγραμμα σχεδιάζεται η καμπύλη της κατατομής log-πιθανοφάνειας για $1/p = 20$, επιτρέποντάς μας να προσδιορίσουμε ένα δ.ε. για τη στάθμη απόδοσης.

ΣΧΗΜΑ 2-5

Διάγραμμα κατατομής log-πιθανοφάνειας της στάθμης απόδοσης για τα δεδομένα που προέρχονται από $\text{Exp}(1)$ και $\text{Cauchy}(0,1)$



Απεικονίζοντας τη συνάρτηση κατατομής λογάριθμο-πιθανοφάνειας των αποδόσεων (ΣΧΗΜΑ 2-5), προκύπτει ότι η στάθμη απόδοσης στα επόμενα 20 *blocks*, δηλαδή η τιμή αυτή που αναμένουμε να ξεπεραστεί στα επόμενα 20 *blocks*, είναι 7.552192 για τις παρατηρήσεις από την $\text{Exp}(1)$ και 778.2506 για τις παρατηρήσεις από την $\text{Cauchy}(0,1)$. Στη περίπτωση της $U(0,1)$ δε μπορέσαμε να εξάγουμε εκτιμήσεις, κάτι που αναμέναμε από το γεγονός ότι $\xi = -1$. Στη περίπτωση της $\text{Exp}(1)$ το εκτιμώμενο (από την συνάρτηση κατατομής λογάριθμο-πιθανοφάνειας) δ.ε. συντελεστού 95% είναι (7.174583, 8.161301) ενώ για την $\text{Cauchy}(0,1)$ είναι (496.6725, 1346.295).

2.4.4 Εκτίμηση των παραμέτρων κάτω από την υπόθεση του MDA

Ας χαλαρώσουμε κάπως τον αυστηρό χαρακτήρα της υπόθεσης (2.3) και ας υποθέσουμε ότι για κάποιο $\xi \in \mathbf{R}$ ισχύει ότι,

$$X_1, \dots, X_n \text{ iid από } F \in \text{MDA}(H_\xi) \quad (2.9)$$

Από τη θεωρία του πρώτου κεφαλαίου, γνωρίζουμε ότι αν $F \in \text{MDA}(H_\xi)$ τότε ισοδύναμα

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln H_\xi(x) \quad (2.10)$$

για κατάλληλα επιλεγμένες παραμέτρους (c_n) και (d_n) και το x ανήκει σε ένα κατάλληλο χώρο ανάλογα με το πρόσημο του ξ . Για να καταλάβουμε ποια είναι η ουσιώδης διαφορά μεταξύ των σχέσεων (2.3) και (2.9), Ας υποθέσουμε ότι εξετάζουμε τη κατανομή Fréchet, δηλαδή ότι $\xi = 1/\alpha > 0$. Σε αυτή τη περίπτωση η (2.3) σημαίνει ότι το δείγμα X_1, \dots, X_n ακολουθεί ακριβώς την κατανομή Fréchet, δηλαδή

$$\bar{F}(x) = 1 - \exp\{-x^{-\alpha}\}, x > 0.$$

Από την άλλη η (2.9) μπορεί να γραφτεί ως

$$\bar{F}(x) = x^{-\alpha} L(x), x > 0,$$

για κάποια βραδέως μεταβαλλόμενη συνάρτηση L . Σε αυτή τη περίπτωση η εκτίμηση της ουράς της κατανομής $\bar{F}(x)$ γίνεται ακόμα πιο περίπλοκη, λόγω του μη παραμετρικού χαρακτήρα της συνάρτησης L . Επομένως ενώ η (2.3) αντιστοιχεί σε παραμετρικές υποθέσεις, η (2.9) αντιστοιχεί στο παραμετρικό κομμάτι α και στο μη παραμετρικό κομμάτι L . Λόγω αυτής της διαφοράς η (2.9) χρησιμοποιείται κυρίως για την εκτίμηση κατανομών με βαριά ουρά.

Μια άμεση συνέπεια της (2.10) είναι ότι για μεγάλο $u = c_n x + d_n$ ισχύει ότι

$$n\bar{F}(u) \approx \left(1 + \xi \frac{u - d_n}{c_n}\right)^{-1/\xi}$$

και επομένως ο εκτιμητής της ουράς θα μπορούσε να πάρει τη μορφή

$$\hat{\bar{F}}(u) = \frac{1}{n} \left(1 + \hat{\xi} \frac{u - \hat{d}_n}{\hat{c}_n}\right)^{-1/\hat{\xi}} \quad (2.11)$$

για κατάλληλους εκτιμητές $\hat{\xi}, \hat{c}_n, \hat{d}_n$. Καθώς η σχέση (2.9) εκφράζει μια ιδιότητα της ουράς της κατανομής, η εκτίμηση του ξ μπορεί να στηριχθεί στις k -άνω διατεταγμένες τιμές $X_{k,n} \leq \dots \leq X_{1,n}$.

Σε αυτή τη περίπτωση τίθενται οι ακόλουθες υποθέσεις,

$$(α) k(n) \rightarrow \infty \quad \text{και} \quad (β) n/k(n) \rightarrow \infty. \quad (2.12)$$

Δηλαδή χρησιμοποιούμε ένα αρκετά μεγάλο αριθμό διατεταγμένων τιμών. Δεδομένου όμως ότι ενδιαφερόμαστε για τις ιδιότητες της ουράς της κατανομής, θα πρέπει να βεβαιωθούμε ότι η μελέτη μας θα επικεντρωθεί μόνο στις άνω διατεταγμένες τιμές.

Από την (2.11) μπορούμε σε γενικές γραμμές να εκτιμήσουμε το ποσοστιαίο σημείο $x_p = F^{\leftarrow}(p)$ για δεδομένο $p \in (0,1)$ ως ακολούθως

$$\hat{x}_p = \hat{d}_n + \frac{\hat{c}_n}{\hat{\xi}} \left((n(1-p))^{-\hat{\xi}} - 1 \right) \quad (2.13)$$

Το ενδιαφέρον μας θα επικεντρωθεί στην εκτίμηση των p -ποσοστιαίων σημείων έξω από το δείγμα X_1, \dots, X_n . Αυτό σημαίνει ότι επιλέγουμε το $p = p_n$ έτσι ώστε $p > 1 - 1/n$. Με αυτό τον τρόπο η εμπειρική σ.κ. ικανοποιεί τη σχέση $\bar{F}_n(p) = 0$ και δε δίνει πληροφορίες για τα ποσοστιαία σημεία. Για να εκτιμήσουμε καλά τα $\hat{\xi}, \hat{c}_n, \hat{d}_n$ στην (2.13) χρειάζεται να χρησιμοποιήσουμε κάποιο τέχνασμα. Θα υποθέσουμε χάριν ευκολίας συμβολισμού ότι $n/k \in \mathbb{N}$ και θα χρησιμοποιήσουμε μια υπό-ακολουθία (n/k) με $k = k(n)$ που να ικανοποιεί τις υποθέσεις (2.12). Το ποσοστιαίο σημείο x_n θα εκτιμάται από τη σχέση

$$\hat{x}_p = \hat{d}_{n/k} + \frac{\hat{c}_{n/k}}{\hat{\xi}} \left(\left(\frac{n}{k} (1-p_n) \right)^{-\hat{\xi}} - 1 \right). \quad (2.14)$$

Χρησιμοποιούμε αυτή την τεχνική γιατί θέλουμε να πάρουμε εκτιμήσεις σε δυο επίπεδα. Πρώτον, πρέπει να βρούμε μια αξιόπιστη εκτιμήτρια για το ξ , όπου οι συνθήκες (2.12) θα αποτελούν ένα φυσικό επακόλουθο. Δεύτερον, θέλουμε να εκτιμήσουμε τις παραμέτρους κανονικοποίησης c_n και d_n οι οποίες ορίζονται από τα ποσοστιαία σημεία της F . Για παράδειγμα, αν εξετάζουμε τη Fréchet, γνωρίζουμε από το Κεφάλαιο 1 ότι $c_n = F^{\leftarrow}(1 - n^{-1})$. Επομένως η εκτίμηση του c_n είναι ισοδύναμη της εκτίμησης του x_p στα ακρότατα του συνόλου των δεδομένων μας. Με το να χρησιμοποιήσουμε μια υπό-ακολουθία (n/k) , αποφεύγουμε το όριο $1 - n^{-1}$ παίρνοντας εκτιμήσεις στο πιο ασφαλές όριο $1 - (n/k)^{-1}$. Η εκτίμηση του $c_{n/k}$ περιορίζεται στην εκτίμηση ποσοστιαίων σημείων μέσα στο εύρος του συνόλου των δεδομένων μας. Παρόμοια είναι και η προσέγγιση της εκτίμησης του $d_{n/k}$. Αυτή η διαδικασία είναι παρόμοια και για τις Gumbel και Weibull.

Αναφορικά με τα παραπάνω, μια ποσότητα που είναι χρήσιμο να εκτιμηθεί και που συνδέεται στενά με τα ποσοστιαία σημεία x_p είναι η ακόλουθη

$$x_{p,r} = F^{\leftarrow}(p^{1/r}), r \in \mathbb{N},$$

όπου $x_p = x_{p,1}$. Το $x_{p,1}$ ερμηνεύεται ως ακολούθως

$$p = F^r(x_{p,r}) = P(\max(X_{n+1}, \dots, X_{n+r}) \leq x_{p,r})$$

Δηλαδή το $x_{p,r}$ είναι η στάθμη την οποία, δεν υπερβαίνουν οι επόμενες r παρατηρήσεις X_{n+1}, \dots, X_{n+r} με πιθανότητα p . Η εκτίμηση του $x_{p,r}$ προκύπτει από την (2.14) ως

$$\hat{x}_{p,r} = \hat{d}_{n/k} + \frac{\hat{c}_{n/k}}{\hat{\xi}} \left(\left(\frac{n}{k} (1 - p^{1/r}) \right)^{-\hat{\xi}} - 1 \right).$$

2.5 Εκτίμηση της παραμέτρου θέσης ξ

Οι τεχνικές για την εξαγωγή εκτιμήσεων γύρω από τη συμπεριφορά της ουράς μιας κατανομής έχουν ως θεωρητικό υπόβαθρο την οριακή συμπεριφορά των κατανομών ακραίων τιμών. Μια βασική ομάδα εκτιμητών αναπτύχθηκε από τους Hill (1975), Pickands (1975) κ.α. χρησιμοποιώντας διατεταγμένες τιμές. Οι διαδικασίες που προτάθηκαν μας επιτρέπουν να εκτιμήσουμε τη παράμετρο θέσης ξ , η οποία είναι ο δείκτης ουράς (*tail-index*) μιας κατανομής και συχνά στη βιβλιογραφία συμβολίζεται με γ . Το γ καλείται και παράμετρος σχήματος (*shape parameter*) και αποτελεί ένα από τα κύρια αντικείμενα μελέτης, αφού από τη τιμή του καθορίζεται και η συμπεριφορά των ακραίων τιμών.

2.5.1 Εκτίμηση του ξ μέσω του εκτιμητή Pickands (*Pickands estimator*)

Η βασική αρχή για την εφαρμογή του εκτιμητή, είναι μπορέσουμε να θέσουμε μια υπόθεση ισοδύναμη του $F \in \text{MDA}(H_\xi)$. Το κλειδί αυτής της διαδικασίας είναι το Θεώρημα 1.4.23 όπου αποδεικνύεται ότι $F \in \text{MDA}(H_\xi)$ αν η $U(t) = F^{\leftarrow}(1 - t^{-1})$ ικανοποιεί τη σχέση

$$\lim_{t \rightarrow \infty} \frac{U(2t) - U(t)}{U(t) - U(t/2)} = 2^\xi.$$

Αποδεικνύεται επίσης ότι αν για κάποια θετική συνάρτηση c ισχύει $\lim_{t \rightarrow \infty} c(t) = 2$, τότε

$$\lim_{t \rightarrow \infty} \frac{U(c(t)t) - U(t)}{U(t) - U(t/c(t))} = 2^\xi. \quad (2.15)$$

Χρησιμοποιώντας τη (2.15) θα προσπαθήσουμε να κατασκευάσουμε ένα εμπειρικό εκτιμητή. Υποθέτουμε λοιπόν ότι $V_{n,n} \leq \dots \leq V_{1,n}$ είναι το διατεταγμένο δείγμα των iid τ.μ. V_1, \dots, V_n με κοινή σ.κ. Pareto της μορφής $F_V(x) = 1 - x^{-1}$, $x \geq 1$. Είναι εύκολο να αποδειχθεί ότι

$$(X_{k,n})_{k=1, \dots, n} \stackrel{d}{=} (U(V_{k,n}))_{k=1, \dots, n}$$

όπου X_1, \dots, X_n είναι iid με σ.κ. F . Η διατεταγμένη τιμή $V_{k,n}$ αποτελεί το εμπειρικό $(1 - k/n)$ ποσοστιαίο σημείο της F_V . Χρησιμοποιώντας το Λήμμα 1.6.4 προκύπτει ότι

$$\frac{k}{n} V_{k,n} \xrightarrow{P} 1$$

όταν $k = k(n) \rightarrow \infty$ και $k/n \rightarrow 0$. Ειδικότερα

$$V_{k,n} \xrightarrow{P} \infty \quad \text{και} \quad V_{2k,n}/V_{k,n} \xrightarrow{P} 1/2$$

Από αυτές τις σχέσεις και την (2.15) προκύπτει ότι

$$\frac{U(V_{k,n}) - U(V_{2k,n})}{U(V_{2k,n}) - U(V_{4k,n})} \xrightarrow{P} 2^\xi$$

Από τα παραπάνω και την (2.15) ορίζουμε τον εκτιμητή Pickands ως ακολούθως

$$\hat{\xi}_{k,n}^{(P)} = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}. \quad (2.16)$$

2.5.1.1 Θεώρημα. (Ιδιότητες του εκτιμητή Pickands) Υποθέτουμε ότι (X_n) είναι μια iid ακολουθία με σ.κ. $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbf{R}$. Έστω $\hat{\xi}^{(P)} = \hat{\xi}_{k,n}^{(P)}$ ο εκτιμητής Pickands (2.16).

(α) (Ασθενής συνέπεια) Αν ισχύει ότι $k \rightarrow \infty$, $k/n \rightarrow 0$, καθώς $n \rightarrow \infty$, τότε

$$\hat{\xi}_{k,n}^{(P)} \xrightarrow{P} \xi.$$

(β) (Ισχυρή συνέπεια) Αν ισχύει ότι $k/n \rightarrow 0$, $k/\ln \ln n \rightarrow \infty$, καθώς $n \rightarrow \infty$, τότε

$$\hat{\xi}^{(P)} \rightarrow_{n \rightarrow \infty} \xi \text{ σχεδόν βέβαια.}$$

(γ) (Ασυμπτωτική κανονικότητα) Κάτω από κάποιες επιπλέον υποθέσεις για το k και την F (βλ. Dekkers and de Haan (1989)) ισχύει ότι

$$\sqrt{k}(\hat{\xi}^{(P)} - \xi) \xrightarrow{d} N(0, u(\xi))$$

όπου

$$u(\xi) = \frac{\xi^2 (2^{2\xi+1} + 1)}{(2(2^\xi - 1) \ln 2)^2}.$$

Ένα ιδιαίτερο χαρακτηριστικό του συγκεκριμένου εκτιμητή είναι ότι η μεγαλύτερη παρατήρηση δεν χρησιμοποιείται για την εκτίμηση. Αυτό μπορεί να δικαιολογηθεί από την άποψη ότι η μεγαλύτερη παρατήρηση θα είναι αυτή που θα προσδώσει και τη μεγαλύτερη αβεβαιότητα στις εκτιμήσεις. Το μειονέκτημα του, είναι ότι παρουσιάζει ευαισθησία στην επιλογή των ενδιάμεσων διατεταγμένων τιμών που χρησιμοποιούνται για την εκτίμηση. Ακόμα και μια μικρή μεταβολή στο k μπορεί να δώσει τελείως διαφορετική τιμή στον εκτιμητή.

Είδαμε λοιπόν ότι ο υπολογισμός του εκτιμητή Pickands περιλαμβάνει και μια ακολουθία διατεταγμένων τιμών, οι οποίες αυξάνουν με ρυθμό n . Δεν είναι σπάνιο στην ανάλυση του εκτιμητή να χρησιμοποιηθεί το διάγραμμα Pickands (*Pickands Plot*), δηλαδή ένα διάγραμμα του οποίου τα σημεία ορίζονται ως

$$\{(k, \hat{\xi}_{k,n}^{(P)}) : k = 1, \dots, n\}$$

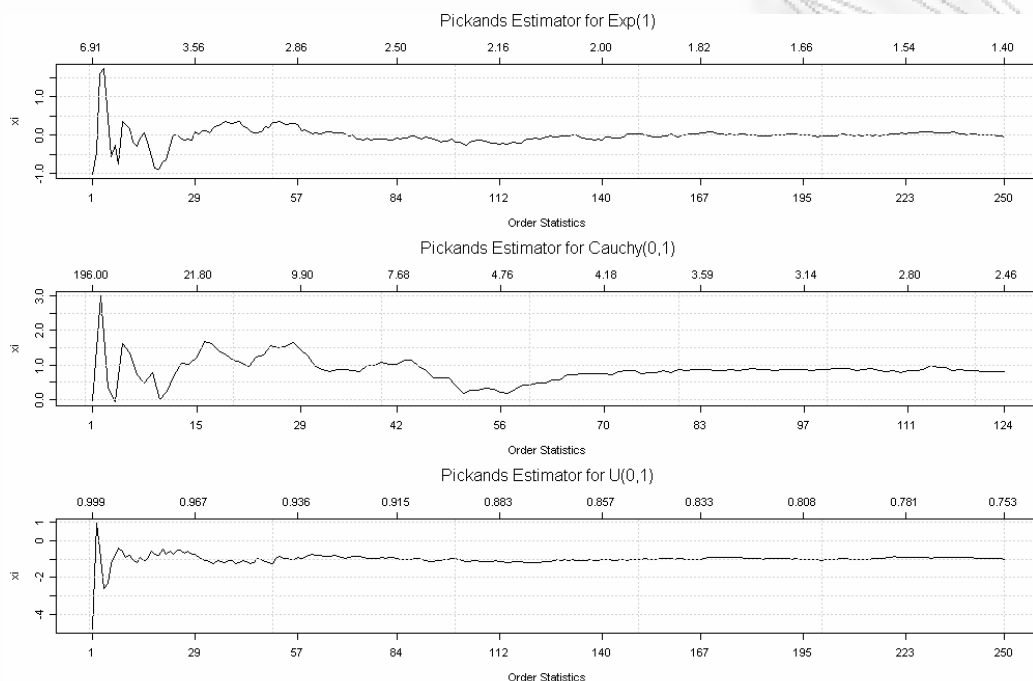
ώστε να επιλέξουμε το κατάλληλο k καθώς και το κατώφλι u που αντιστοιχεί σε αυτό. Ως κατώφλι u θεωρούμε εκείνη τη παρατήρηση πάνω από την οποία «ξεκινάει» η ουρά της κατανομής των παρατηρήσεων του υπό εξέταση τ.δ. (εκτενέστερη αναφορά θα γίνει στη § 2.8). Η ερμηνεία ενός τέτοιου διαγράμματος δεν είναι πάντα εύκολη και δεν υπάρχει κάποια γενική οδηγία που μπορεί να δοθεί. Διαισθητικά μπορούμε να υποστηρίξουμε ότι το $\hat{\xi}_{k,n}^{(P)}$ θα πρέπει να επιλεγεί από εκείνη τη περιοχή των k , που δίνουν σχεδόν οριζόντιο διάγραμμα.

2.5.1.2 Παράδειγμα

Αποδώσαμε γραφικά τη πορεία του εκτιμητή Pickands για προσομοιωμένα δεδομένα ($n = 1000$) από $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ αντίστοιχα (ΣΧΗΜΑ 2-6). Στον οριζόντιο άξονα απεικονίζονται οι διατεταγμένες τιμές (συγκεκριμένα απεικονίζεται η θέση μιας παρατήρησης μέσα στο διατεταγμένο δείγμα), στον κάθετο άξονα οι τιμές του ξ , ενώ στον πάνω άξονα απεικονίζονται οι τιμές του κατωφλίου u . Εξετάζοντας το διάγραμμα της $\text{Exp}(1)$, παρατηρούμε την εκτίμηση του ξ καθώς προχωράμε από τη μεγαλύτερη διατεταγμένη παρατήρηση προς τη μικρότερη, δηλαδή καθώς απομακρυνόμαστε από την ουρά της υπό εξέταση κατανομής προς τις πιο κεντρικές παρατηρήσεις. Ο πάνω άξονας των κατωφλίων u , παρουσιάζει εκείνη τη τιμή του τ.δ. (σε απόλυτους όρους) πάνω από την οποία παίρνουμε παρατηρήσεις για να εκτιμήσουμε το ξ . Παρατηρούμε ότι για τις $\text{Exp}(1)$ και $\text{U}(0,1)$ οι τιμές του εκτιμητή κυμαίνονται γύρω από το $\xi = 0$ και $\xi = -1$. Στη περίπτωση της $\text{Cauchy}(0,1)$ βλέπουμε ότι ο εκτιμητής

δίνει τιμές γύρω από το $\zeta = 1$, αλλά είναι δύσκολο να εντοπίσουμε κάποια σημεία που το διάγραμμα να παρουσιάζει γραμμικότητα, έτσι ώστε να επιλεγεί το κατάλληλο k .

ΣΧΗΜΑ 2-6



2.5.2 Εκτίμηση της παραμέτρου σχήματος $\zeta = \alpha^{-1} > 0$ μέσω του εκτιμητή Hill

Από το 1975 που ο Hill εξέδωσε την εργασία του πάνω στην εκτίμηση του δείκτη κανονικής μεταβολής της ουράς μιας κατανομής, έδωσε το έναυσμα για μια μεγάλη ερευνητική προσπάθεια. Πολλοί ήταν οι ερευνητές που προσπάθησαν να αποδείξουν την ισχύ του, ενώ άλλοι εστίασαν στις αδυναμίες του παρουσιάζοντας κάποιους άλλους εκτιμητές. Παρόλα αυτά η δημοτικότητα του συγκεκριμένου εκτιμητή στους επιστημονικούς κύκλους παραμένει αμείωτη. Ένας από τους βασικούς λόγους είναι τόσο η απλότητα του ορισμού του όσο και το γεγονός ότι στη πλειονότητα των περιπτώσεων που είναι απαραίτητη η ανάλυση ακραίων τιμών έχουμε κατανομές με βαριά ουρά ($\zeta > 0$).

Υποθέτουμε ότι X_1, \dots, X_n είναι iid με σ.κ. $F \in \text{MDA}(\Phi_\alpha)$, $\alpha > 0$, επομένως θα ισχύει

$$\bar{F}(x) = x^{-\alpha} L(x), \quad x > 0,$$

για κάποια βραδέως μεταβαλλόμενη συνάρτηση L . Κατανομές με τέτοιες ουρές αποτελούν τη βάση για τη μοντελοποίηση φαινομένων με βαριά ουρά. Για πολλές εφαρμογές η γνώση του δείκτη μεταβολής α είναι πολύ σημαντική. Ο εκτιμητής Hill του $\alpha = \zeta^{-1}$ ορίζεται ως

$$\hat{a}^{(H)} = \hat{a}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1} \quad (2.17)$$

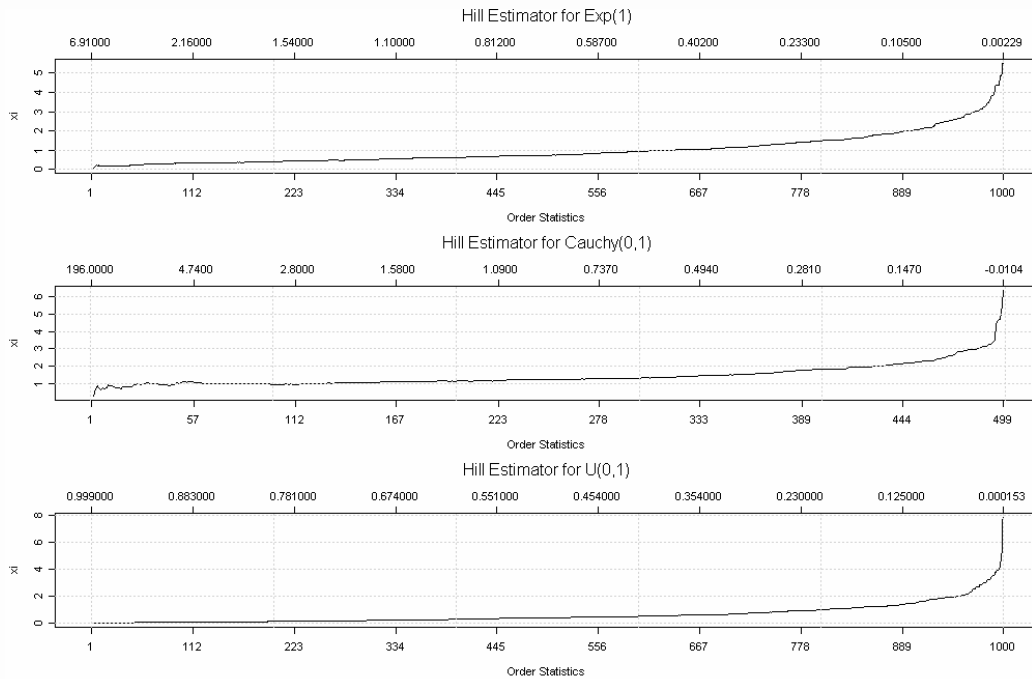
όπου το k επιλέγεται με κατάλληλο τρόπο έτσι ώστε $k = k(n) \rightarrow \infty$. Όπως και στη περίπτωση του εκτιμητή Pickands συνήθως χρησιμοποιούμε μια αύξουσα ακολουθία άνω διατεταγμένων τιμών. Είναι αξιοσημείωτο ότι, πέρα από τη παραπάνω μορφή του εκτιμητή, μπορούμε να πάρουμε και άλλες ασυμπτωτικά ισοδύναμες μορφές με αρκετά διαφορετικές μεθόδους (π.χ. μέθοδο μέγιστης πιθανοφάνειας).

Όπως ισχύει και για τον εκτιμητή Pickands, μπορούμε να δώσουμε μια συνοπτική παρουσίαση του εκτιμητή Hill γραφικά για διάφορες τιμές του k . Το διάγραμμα Hill, το οποίο αποτελείται από τα σημεία $\{(k, \hat{\xi}_{k,n}^{(H)}) : k = 2, \dots, n\}$ είναι αρκετά χρήσιμο εργαλείο στην εύρεση του βέλτιστου k καθώς και του κατωφλίου u που αντιστοιχεί σε αυτό. Οι ομαλές γραμμές ενός διαγράμματος Hill πάνω σε μια ορισμένη περιοχή του εύρους των k , υποδεικνύουν που πρέπει να αναζητήσουμε το βέλτιστο k .

2.5.2.1 Παράδειγμα

Αναπαραστήσαμε γραφικά τη πορεία του εκτιμητή Hill για προσομοιωμένα δεδομένα ($n = 1000$) από $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ αντίστοιχα (ΣΧΗΜΑ 2-7). Στον οριζόντιο άξονα απεικονίζονται οι διατεταγμένες τιμές, στον κάθετο άξονα οι τιμές του ξ , ενώ στον πάνω άξονα απεικονίζονται οι τιμές του κατωφλίου u . Σε αντίθεση με το διάγραμμα του εκτιμητή Pickands που απεικονίζει τις εκτιμήσεις του ξ για το 25% του συνόλου των διατεταγμένων παρατηρήσεων, στο διάγραμμα του εκτιμητή Hill απεικονίζονται οι εκτιμήσεις για όλο το σύνολο των διατεταγμένων παρατηρήσεων. Όπως και στη περίπτωση του διαγράμματος Pickands παρατηρούμε την εκτίμηση του ξ , καθώς προχωράμε από τη μεγαλύτερη διατεταγμένη παρατήρηση προς τη μικρότερη. Το ανοδικό τμήμα του διαγράμματος προς το τέλος του εύρους του συνόλου των διατεταγμένων παρατηρήσεων, μας δείχνει πως επηρεάζεται η εκτίμηση του ξ από παρατηρήσεις που βρίσκονται στο κέντρο της κατανομής. Παρατηρούμε ότι για τη περίπτωση της $\text{Cauchy}(0,1)$ όπου ισχύει ότι $\xi = 1$, το διάγραμμα κινείται γύρω από τη τιμή $\xi = 1$ για τιμές του k στο διάστημα $(60, 200)$. Αντίθετα στις περιπτώσεις των $\text{Exp}(1)$ και $\text{U}(0,1)$, είναι ορατή η αδυναμία του εκτιμητή να προσδιορίσει το βέλτιστο k . Το προφανές πρόβλημα που αντιμετωπίζουμε με τον εκτιμητή Hill είναι ότι περιορίζεται στη περίπτωση $\xi > 0$, δηλαδή στην κατανομή Fréchet.

ΣΧΗΜΑ 2-7



2.5.2.2 Κατασκευή του εκτιμητή Hill μέσω της μεθόδου MLE (βλ. Hill (1975))

Υποθέτουμε ότι X είναι μια τ.μ. με σ.κ. F τέτοια ώστε για $a > 0$,

$$P(X > x) = \bar{F}(x) = x^{-a}, x \geq 1.$$

Αν θεωρήσουμε ότι $Y = \ln X$, τότε η σ.κ. της Y θα είναι

$$P(Y > y) = e^{-ay}, y \geq 0$$

δηλαδή η Y θα ακολουθεί την εκθετική κατανομή και επομένως ο MLE του a θα είναι

$$\hat{a}_n = \bar{Y}_n^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_j \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_{j,n} \right)^{-1}.$$

Ως γενίκευση του παραπάνω μπορούμε να θεωρήσουμε ότι

$$\bar{F}(x) = Cx^{-a}, x \geq u > 0 \quad (2.18)$$

με το u να είναι άγνωστο. Αν θεωρήσουμε ότι $C = u^a$, τότε προκύπτει άμεσα ο MLE του a

$$\hat{a}_n = \left(\frac{1}{n} \sum_{j=1}^n \ln \left(\frac{X_{j,n}}{u} \right) \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_{j,n} - \ln u \right)^{-1}.$$

Συνήθως στη πράξη δε γνωρίζουμε την ακριβή παραμετρική μορφή του παραδείγματος που δώσαμε, αλλά υπό τη συνθήκη $F \in \text{MDA}(\Phi_a)$ υποθέτουμε ότι η \bar{F} συμπεριφέρεται σαν μια κατανομή Pareto πάνω από ένα ορισμένο κατώφλι u . Ορίζουμε λοιπόν

$$K = \text{card} \{i : X_{i,n} > u, i = 1, \dots, n\}. \quad (2.19)$$

Δεδομένου ότι ισχύει το ενδεχόμενο $\{K = k\}$, ο εκτιμητής MLE του a και του C της (2.18) προσδιορίζεται αν μεγιστοποιήσουμε τη πυκνότητα των $\{X_{k,n}, \dots, X_{1,n}\}$. Από το Θεώρημα 1.6.2 παίρνουμε ότι

$$f_{X_{1,n}, \dots, X_{k,n}}(x_1, \dots, x_k) = \frac{n!}{(n-k)!} (1 - Cx_k^{-a})^{n-k} C^k a^k \prod_{i=1}^k x_i^{-(a+1)}, \quad u < x_k < \dots < x_1.$$

Από τη παραπάνω σ.π.π. μπορούμε να υπολογίσουμε τον δεσμευμένο MLE

$$\hat{a}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k \ln \left(\frac{X_{j,n}}{X_{k,n}} \right) \right)^{-1} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \quad \hat{C}_{k,n} = \frac{k}{n} X_{k,n}^{\hat{a}_{k,n}^{(H)}}.$$

Επομένως ο εκτιμητής Hill έχει την ίδια μορφή με τον MLE του μοντέλου (2.18), μόνο που αντί για την τιμή u έχουμε το τυχαίο όριο $X_{k,n}$, όπου το k ορίζεται από την (2.19). Επίσης από την παραπάνω διαδικασία προκύπτει άμεσα και μια εκτίμηση για την ουρά $\bar{F}(x)$

$$\widehat{\bar{F}}(x) = \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\hat{a}_{k,n}^{(H)}} \quad (2.20)$$

και για το p -ποσοστιαίο σημείο.

$$\hat{x}_p = \left(\frac{n}{k} (1-p) \right)^{-1/\hat{a}_{k,n}^{(H)}} X_{k,n}.$$

Από την (2.20) προκύπτει ένας εκτιμητής για την υπερβάλλουσα σ.κ. $F_u(x - u)$, $x \geq u$, αν χρησιμοποιήσουμε την ισότητα $F_u(x - u) = 1 - \bar{F}(x) / \bar{F}(u)$.

2.5..2.3 Κατασκευή του εκτιμητή Hill μέσω της ομαλής μεταβλητότητας

Αυτή η προσέγγιση ακολουθεί τα βήματα για τη κατασκευή του εκτιμητή Pickands, δηλαδή η εκτίμηση στηρίζεται σε μια κατάλληλη μορφή του $F \in \text{MDA}(\Phi_a)$ (βλ. de Haan (1994)).

Πράγματι $F \in \text{MDA}(\Phi_a)$ αν και μόνο αν ισχύει

$$\lim_{t \rightarrow \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-a}, x > 0.$$

Χρησιμοποιώντας ολοκλήρωση κατά μέρη, παίρνουμε ότι

$$\int_t^{\infty} (\ln x - \ln t) dF(x) = \int_t^{\infty} \frac{\overline{F}(x)}{x} dx,$$

έτσι ώστε από το θεώρημα Karamata (Θεώρημα 1.2.7) να προκύπτει ότι

$$\frac{1}{\overline{F}(t)} \int_t^{\infty} (\ln x - \ln t) dF(x) \rightarrow 1/a, t \rightarrow \infty. \quad (2.21)$$

Για να πάρουμε εκτιμητή από το παραπάνω αποτέλεσμα μπορούμε να χρησιμοποιήσουμε δυο μεθόδους. Μπορούμε να αντικαταστήσουμε το F με έναν εκτιμητή, όπου η προφανής επιλογή είναι η εμπειρική σ.κ.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}} = \frac{1}{n} \sum_{i=1}^n I_{\{X_{i,n} \leq x\}}.$$

Εναλλακτικά μπορούμε να αντικαταστήσουμε το t με ένα αρκετά υψηλό και εξαρτώμενο από το δείγμα επίπεδο, δηλαδή θεωρούμε $t = X_{k,n}$ για κάποιο $k = k(n)$. Η επιλογή του t εξαρτάται από το ότι $X_{k,n} \rightarrow \infty$, σχεδόν βέβαια αν ισχύει ότι $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$. Από την (2.21) προκύπτει ο ακόλουθος εκτιμητής

$$\frac{1}{\overline{F}(X_{k,n})} \int_{X_{k,n}}^{\infty} (\ln x - \ln X_{k,n}) dF_n(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \ln X_{j,n} - \ln X_{k,n}$$

ο οποίος εκτός από τον παράγοντα $k-1$, είναι της μορφής $(\hat{a}^{(H)})^{-1}$ όπως δίνεται από την (2.17). Πρέπει να επισημάνουμε ότι η αλλαγή από το k στο $k-1$ είναι ασυμπτωτικά ασήμαντη.

2.5.2.4 Κατασκευή του εκτιμητή Hill μέσω της μέσης υπερβάλλουσας συνάρτησης

Αυτή η μέθοδος ουσιαστικά αποτελεί μια επαναδιατύπωση της διαδικασίας που περιγράψαμε στην § 2.5.2.3. Ο λόγος που διαχωρίζεται από τη κατασκευή μέσω της ομαλής μεταβλητότητας είναι η ενδιαφέρουσα διατύπωση που χρησιμοποιείται (βλ. Embrechts et al. (1997)). Υποθέτουμε ότι X είναι μια τ.μ. με σ.κ. $F \in \text{MDA}(\Phi_a)$, $a > 0$. Χάριν ευκολίας θα υποθέσουμε ότι $X > 1$ σχεδόν βέβαια. Μπορούμε να ξαναγράψουμε την (2.21) ως ακολούθως

$$E(\ln X - \ln t | \ln X > \ln t) \rightarrow 1/a, t \rightarrow \infty.$$

Αν ορίσουμε ως $u = \ln t$ και ως $e^*(u)$ την μέση υπερβάλλουσα συνάρτηση του $\ln X$ τότε προκύπτει

$$e^*(u) \rightarrow 1/a, u \rightarrow \infty.$$

Σε αυτή τη περίπτωση ο εκτιμητής Hill μπορεί να ερμηνευθεί ως η εμπειρική μέση υπερβάλλουσα συνάρτηση του $\ln X$ υπολογιζόμενη στο κατώφλι $u = \ln X_{k,n}$ δηλαδή $e_n^*(\ln X_{k,n})$.

2.5.2.5 Θεώρημα (Βασικές ιδιότητες του εκτιμητή Hill) Υποθέτουμε ότι (X_n) είναι μια αυστηρά στάσιμη ακολουθία με οριακή κατανομή F που ικανοποιεί για κάποιο $a > 0$ και $L \in \mathfrak{R}_0$,

$$\bar{F}(x) = P(X > x) = x^{-a}L(x), x > 0.$$

Έστω $\hat{a}^{(H)} = \hat{a}_{k,n}^{(H)}$ ο εκτιμητής Hill (2.17)

(α) (Ασθενής συνέπεια) Αν υποθέσουμε ότι η (X_n) ικανοποιεί μία από τις ακόλουθες συνθήκες:

- είναι μια ακολουθία iid τ.μ. (βλ. Mason (1982)),
- είναι ασθενώς εξαρτημένη (βλ. Rootzén et al. (1992)),
- αποτελεί μια γραμμική στοχαστική ανέλιξη (βλ. Resnick and Stărică (1992, 1996)),

τότε αν $k \rightarrow \infty, k/n \rightarrow 0$ για $n \rightarrow \infty$, θα ισχύει ότι $\hat{a}^{(H)} \xrightarrow{P} a$.

(β) (Ισχυρή συνέπεια) Αν $k/n \rightarrow 0, k/\ln \ln n \rightarrow \infty$ για $n \rightarrow \infty$ και (X_n) είναι μια iid ακολουθία, τότε $\hat{a}^{(H)} \rightarrow a$ σχεδόν βέβαια (βλ. Deheuvels, Häusler and Mason (1988)).

(γ) (Ασυμπτωτική Κανονικότητα) Κάτω από κάποιες επιπλέον υποθέσεις για το k και για την F ισχύει

$$\sqrt{k}(\hat{a}^{(H)} - a) \xrightarrow{d} N(0, a^2).$$

Το παραπάνω θεώρημα δε πρέπει να ληφθεί ως επιβεβαίωση του ότι ο εκτιμητής Hill είναι πάντα τέλειος, απλώς αποτελεί επιβεβαίωση ότι γενικά ισχύουν οι βασικές στατιστικές ιδιότητες.

2.5.3 Εκτιμητής του $\xi \in \mathbf{R}$ των Dekkers – Einmahl – de Haan (DEdH)

Ένα μειονέκτημα του εκτιμητή Hill είναι ότι έχει σχεδιαστεί για κατανομές που πληρούν την προϋπόθεση $F \in \text{MDA}(H_\xi)$, $\xi > 0$. Οι Dekkers, Einmahl and de Haan (1989) επέκτειναν τον εκτιμητή Hill ώστε να καλύπτει όλη τη κλάση των H_ξ , $\xi \in \mathbf{R}$. Πρότειναν τον εκτιμητή

$$\hat{\xi} = 1 + H_n^{(1)} + \frac{1}{2} \left(\frac{(H_n^{(1)})^2}{H_n^{(2)}} - 1 \right)^{-1} \quad (2.22)$$

όπου

$$H_n^{(1)} = \frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k+1,n})$$

είναι (περίπου) ο αντίστροφος εκτιμητής Hill (*reciprocal Hill's estimator*) και

$$H_n^{(2)} = \frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k+1,n})^2 .$$

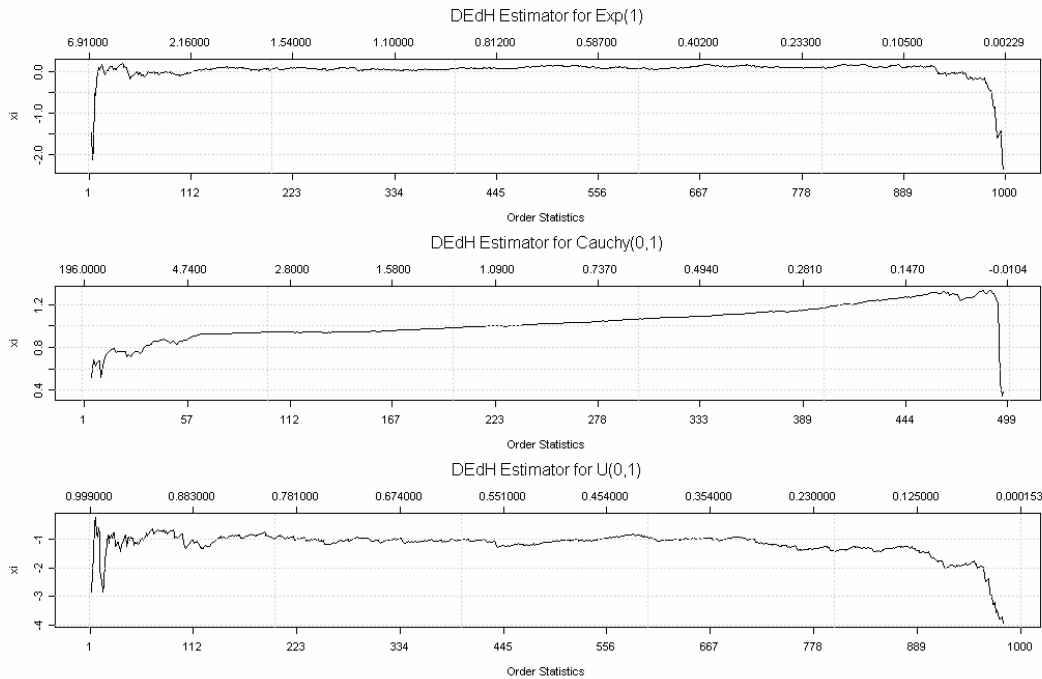
Οι ποσότητες $H_n^{(1)}$ και $H_n^{(2)}$ μπορούν να ερμηνευθούν σαν εμπειρικές ροπές. Αναλόγως το $\hat{\xi}$ ερμηνεύεται σαν εκτιμητής ροπής του ξ (*moment estimator*). Οι Dekkers, Einmahl and de Haan (1989) απέδειξαν την ισχυρή συνέπεια και την ασυμπτωτική κανονικότητα του παραπάνω εκτιμητή.

Όπως και στη περίπτωση των εκτιμητών Pickands και Hill, μπορούμε να δώσουμε μια συνοπτική παρουσίαση του εκτιμητή DEdH γραφικά για διάφορες τιμές του k . Το διάγραμμα αποτελείται από τα σημεία $\{(k, \hat{\xi}) : k = 1, \dots, n\}$.

2.5.3.1 Παράδειγμα

Εξετάζοντας τη γραφική αναπαράσταση του εκτιμητή DEdH για προσομοιωμένα δεδομένα ($n = 1000$) από $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ αντίστοιχα (ΣΧΗΜΑ 2-8), παρατηρούμε ότι είναι εμφανής η υπεροχή του συγκεκριμένου εκτιμητή σε σύγκριση με τους εκτιμητές Pickands και Hill, δεδομένου ότι και στις τρεις περιπτώσεις το διάγραμμα παρουσιάζει γραμμικότητα γύρω από τη τιμή του ξ που αναμένουμε. Πρέπει να σημειώσουμε ότι για τη σχεδίαση των συγκεκριμένων διαγραμμάτων, περιορίσαμε το εύρος των τιμών του $\hat{\xi}$, δεδομένου ότι η πρώτη εκτίμηση που προέκυψε και για τις τρεις περιπτώσεις ήταν τόσο μεγάλη, που αλλοίωσε τη κλίμακα του διαγράμματος. Σε αντίθεση με τον εκτιμητή Hill, το καθοδικό τμήμα του διαγράμματος προς το τέλος του εύρους του συνόλου των διατεταγμένων παρατηρήσεων, μας δείχνει πως επηρεάζεται η εκτίμηση του ξ από παρατηρήσεις που βρίσκονται στο κέντρο της κατανομής.

ΣΧΗΜΑ 2-8



Παρουσιάσαμε τρεις διαφορετικές μεθόδους για την εκτίμηση της παραμέτρου ξ . Ποια ακριβώς από τις μεθόδους αυτές θα χρησιμοποιήσει ο ερευνητής στην πράξη εξαρτάται από τις πιθανές τιμές που μπορεί να πάρει το ξ και από τις ακριβείς ιδιότητες της αρχικής σ.κ. F . Σε γενικές γραμμές οι εκτιμητές Pickands και DEdH λειτουργούν για $\xi \in \mathbf{R}$, αν και έχει αποδειχθεί ότι για $\xi > -2$ ο εκτιμητής DEdH έχει μικρότερη διακύμανση από τον Pickands.

2.6 Εκτίμηση των σταθερών κανονικοποίησης

Ας επανέλθουμε λοιπόν στην αρχική υπόθεση ότι έχουμε X_1, \dots, X_n iid δεδομένα με σ.κ. $F \in \text{MDA}(H_\xi)$. Όπως έχει αναφερθεί και σε προηγούμενες παραγράφους, αυτή η σχέση για κατάλληλα επιλεγμένες σταθερές $c_n > 0$ και $d_n \in \mathbf{R}$ είναι ισοδύναμη με

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H_\xi.$$

Αυτή η σχέση θα ισχύει αν και μόνο αν

$$\overline{nF}(c_n x + d_n) \rightarrow -\ln H_\xi(x), \quad n \rightarrow \infty, x \in \mathbf{R}.$$

Κατά τη παρουσίαση των MDA στο Κεφάλαιο 1, είδαμε τη σχέση που συνδέει τις ακολουθίες (c_n) και (d_n) με την \overline{F} . Ενδεικτικά, στη περίπτωση της κατανομής Gumbel όπου $\xi = 0$ και το δεξιό οριακό σημείο $x_F = \infty$, οι σταθερές κανονικοποίησης προκύπτουν από τις σχέσεις

$$c_n = a(d_n), \quad d_n = F^{\leftarrow}(1 - n^{-1}), \quad (2.23)$$

όπου $a(\cdot)$ είναι μια βοηθητική συνάρτηση, η οποία ορίζεται ως

$$a(x) = \int_x^{\infty} \frac{\overline{F}(y)}{\overline{F}(x)} dy.$$

Από τις παραπάνω σχέσεις, διαπιστώνουμε ότι ενώ χρειαζόμαστε τις σταθερές c_n και d_n για να πάρουμε εκτιμήσεις για την ουρά της κατανομής και τα ποσοστιαία σημεία, από την (2.23) οι σταθερές ορίζονται σαν συνάρτηση της ουράς της κατανομής.

Ας προσπαθήσουμε να προσεγγίσουμε το πρόβλημα λίγο διαφορετικά. Ας θεωρήσουμε τη πιο γενική περίπτωση $F \in \text{MDA}(H_\xi)$, $\xi \geq 0$, η οποία περιλαμβάνει τις κατανομές Gumbel και Fréchet. Στη §1.4.20 είδαμε πως χρησιμοποιώντας τον λογαριθμικό μετασχηματισμό $x^* = \ln(1 \vee x)$, $x \in \mathbf{R}$ μπορούμε να ενσωματώσουμε το $\text{MDA}(\Phi_\alpha)$ στο $\text{MDA}(A)$. Αν συνδυάσουμε το παραπάνω με τη Πρόταση 1.4.16, τότε προκύπτει το ακόλουθο αποτέλεσμα.

2.6.1 Λήμμα. (Εμφύτευση του $\text{MDA}(H_\xi)$, $\xi \geq 0$ στο $\text{MDA}(A)$) Υποθέτουμε ότι X_1, \dots, X_n είναι iid τ.μ. με σ.κ. $F \in \text{MDA}(H_\xi)$, $\xi \geq 0$, δεξιά οριακό σημείο $x_F = \infty$ και σταθερές κανονικοποίησης $c_n > 0$ και $d_n \in \mathbf{R}$. Τότε οι X_1^*, \dots, X_n^* είναι iid τ.μ. με σ.κ. $F^* \in \text{MDA}(A)$ και βοηθητική συνάρτηση

$$a^*(t) = \int_t^{\infty} \frac{\overline{F^*}(y)}{\overline{F^*}(t)} dy.$$

Οι σταθερές κανονικοποίησης μπορούν να προσδιοριστούν από τις σχέσεις

$$d_n^* = (F^*)^{\leftarrow}(1 - n^{-1}), \quad c_n^* = a^*(d_n^*) = \int_{d_n^*}^{\infty} \frac{\overline{F^*}(y)}{\overline{F^*}(d_n^*)} dy \sim n \int_{d_n^*}^{\infty} \overline{F^*}(y) dy.$$

Αν θεωρήσουμε ότι οι εκτιμητές μας θα πρέπει να προέρχονται από τις k μεγαλύτερες διατεταγμένες τιμές $X_{k,n}, \dots, X_{1,n}$, όπου $k = k(n) \rightarrow \infty$, τότε από το παραπάνω λήμμα μπορούμε να πάρουμε εκτιμητές αν αντικαταστήσουμε την σ.κ. F^* με την εμπειρική σ.κ. F_n^* :

$$\begin{aligned} \hat{d}_{n/k}^* &= X_{k+1,n}^* = \ln(1 \vee X_{k+1,n}), \\ \hat{c}_{n/k}^* &= \frac{n}{k} \int_{\hat{d}_{n/k}^*}^{\infty} \overline{F_n^*}(y) dy = \frac{n}{k} \int_{\ln(X_{k+1,n})}^{\ln(X_{1,n})} \overline{F_n^*}(y) dy = \frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k+1,n}. \end{aligned} \quad (2.24)$$

Η (2.24) αποτελεί μια μορφή του εκτιμητή Hill. Πρέπει να σημειώσουμε ότι η μεταβολή από το k στο $k + 1$ στην (2.24) είναι ασήμαντη ασυμπτωτικά.

Για να προχωρήσουμε τώρα στον προσδιορισμού του F από την F^* , θα χρησιμοποιήσουμε τον ακόλουθο μετασχηματισμό.

$$\frac{n}{k} P(X^* > c_{n/k}^* x + d_{n/k}^*) = \frac{n}{k} P(X > \exp\{c_{n/k}^* x + d_{n/k}^*\}), \quad x > 0.$$

Αφού έχουμε θεωρήσει ότι $F^* \in \text{MDA}(A)$, τότε όπως δείξαμε και στο Κεφάλαιο 1, το αριστερό μέρος της παραπάνω ισότητας θα συγκλίνει στο e^{-x} καθώς $n \rightarrow \infty$ και υπό την προϋπόθεση ότι $n/k \rightarrow \infty$. Επομένως προκύπτει ο ακόλουθος εκτιμητής

$$(\bar{F}(x))^\wedge = \frac{k}{n} (\exp\{-\hat{d}_{n/k}^* + \ln x\})^{-1/\hat{c}_{n/k}^*} = \frac{k}{n} \left(\frac{x}{X_{k+1,n}} \right)^{-1/\hat{c}_{n/k}^*}.$$

Αυτός ο εκτιμητής δόθηκε από τον Hill για το μοντέλο (2.18). Σαν εκτιμητής ποσοστιαίων σημείων προκύπτει ο

$$\hat{x}_p = \left(\frac{n}{k} (1-p) \right)^{-\hat{c}_{n,k}^*} X_{k+1,n}.$$

2.7 Εκτίμηση της ουράς και των ποσοστιαίων σημείων

Όπως και προηγουμένως θα υποθέσουμε ότι έχουμε X_1, \dots, X_n iid δεδομένα με σ.κ. $F \in \text{MDA}(H_\xi)$, $\xi \geq 0$. Θεωρούμε ότι $0 < p < 1$ και x_p εκφράζει το αντίστοιχο p -ποσοστιαίο σημείο. Ο λόγος για τον οποίο θέσαμε τις προϋποθέσεις των MDA είναι για αν μπορέσουμε να εκτιμήσουμε ποσοστιαία σημεία έξω από το εύρος των δεδομένων που έχουμε στη διάθεσή μας, δηλαδή $p > 1 - 1/n$, το οποίο ισοδυναμεί με την εύρεση εκτιμητή για την ουρά $\bar{F}(x)$ για μεγάλο x . Αν έχουμε προσδιορίσει τους εκτιμητές της παραμέτρου ξ καθώς και των σταθερών κανονικοποίησης c_n και d_n , οι εκτιμητές των x_p και $\bar{F}(x)$ μπορούν να προκύψουν άμεσα από τον ορισμό του $F \in \text{MDA}(H_\xi)$. Αν $U(t) = F^\leftarrow(1 - t^{-1})$ έτσι ώστε $x_p = U(1/(1 - p))$, τότε αν συμβολίσουμε με $U_n(t) = F_n^\leftarrow(1 - t^{-1})$ την εμπειρική συνάρτηση ποσοστιαίων σημείων, έχουμε ότι

$$U_n\left(\frac{n}{k-1}\right) = F_n^\leftarrow\left(1 - \frac{k-1}{n}\right) = X_{k,n}, \quad k = 1, \dots, n.$$

Επομένως προκύπτει ότι το $X_{k,n}$ είναι ο φυσικός εκτιμητής του $(1 - (k - 1)/n)$ - ποσοστιαίου σημείου. Το εύρος $[X_{n,n}, X_{1,n}]$ των δεδομένων μας επιτρέπει να πάρουμε εκτιμήσεις μέσα από τα δειγματικά δεδομένα μέχρι και το $(1 - n^{-1})$ -ποσοστιαίο σημείο.

Σύμφωνα με το επόμενο θεώρημα, μπορούμε να χρησιμοποιήσουμε το $X_{k,n}$ για την εκτίμηση του x_p .

2.7.1 Θεώρημα. (Εκτίμηση υψηλών ποσοστιαίων σημείων I) Υποθέτουμε ότι X_1, \dots, X_n είναι iid δεδομένα με σ.κ. $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbf{R}$ και ότι η σ.κ. F έχει σ.π. f . Υποθέτουμε ότι η πυκνότητα U' ανήκει στο $\mathfrak{R}_{\xi-1}$. Θέτουμε $p = p_n$ και $k = k(n) = [n(1 - p_n)]$, όπου $[x]$ είναι το ακέραιο μέρος του x . Αν ισχύουν οι συνθήκες $p_n \rightarrow 1$ και $n(1 - p_n) \rightarrow \infty$ τότε

$$\sqrt{2k} \frac{X_{k,n} - x_p}{X_{k,n} - X_{2k,n}} \xrightarrow{d} N(0, 2^{2\xi+1} \xi^2 / (2^\xi - 1)^2). \quad (2.25)$$

Στο Θεώρημα 1.4.23 χαρακτηρίσαμε την $F \in \text{MDA}(H_\xi)$ μέσω της ασυμπτωτικής συμπεριφοράς της U :

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\xi - 1}{y^\xi - 1}, \quad x, y > 0, y \neq 1.$$

Αν $\xi = 0$ το παραπάνω όριο ερμηνεύεται ως $\ln x / \ln y$ και επομένως μπορούμε να το ξαναγράψουμε ως

$$U(tx) = \frac{x^\xi - 1}{1 - y^\xi} (U(t) - U(ty))(1 + o(1)) + U(t). \quad (2.26)$$

Χρησιμοποιώντας αυτή τη σχέση, είναι δυνατόν να προσδιορίσουμε έναν εκτιμητή για τα ποσοστιαία σημεία x_p έξω από εύρος των δεδομένων μας. Αντικαθιστούμε το U με U_n στην (2.26) και θέτουμε $y = 1/2$, $x = (k - 1)/(n(1 - p))$ και $t = n/(k - 1)$. Στη συνέχεια αντικαθιστούμε το ξ με έναν κατάλληλο εκτιμητή $\hat{\xi}$ και παραβλέποντας τη ποσότητα $o(1)$ η οποία είναι κοντά στο 0, προκύπτει ο ακόλουθος εκτιμητής

$$\hat{x}_p = \frac{(k/(n(1 - p)))^{\hat{\xi}} - 1}{1 - 2^{-\hat{\xi}}} (X_{k,n} - X_{2k,n}) + X_{k,n} \quad (2.27)$$

Το ακόλουθο αποτέλεσμα αποδείχθηκε από τους Dekkers and de Hann (1989).

2.7.2 Θεώρημα. (Εκτίμηση υψηλών ποσοστιαίων σημείων II) Υποθέτουμε ότι X_1, \dots, X_n είναι iid τ.μ. με σ.κ. $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbf{R}$ και ότι $\lim_{n \rightarrow \infty} n(1-p) = c$ για κάποιο $c > 0$. Θεωρούμε ότι το \hat{x}_p ορίζεται από την (2.27) με $\hat{\xi}$ να είναι ο εκτιμητής Pickands όπως ορίζεται από την (2.25). Τότε για κάθε σταθερό $k > c$,

$$\frac{\hat{x}_p - x_p}{X_{k,n} - X_{2k,n}} \xrightarrow{d} Y$$

όπου

$$Y = \frac{(k/c)^\xi - 2^{-\xi}}{1 - 2^{-\xi}} + \frac{1 - (Q_k/c)^\xi}{\exp\{\xi H_k\} - 1}.$$

Οι τ.μ. H_k και Q_k είναι ανεξάρτητες και η Q_k ακολουθεί τη κατανομή Gamma με παράμετρο $2k + 1$ και

$$H_k = \sum_{j=k+1}^{2k} \frac{E_j}{j}$$

για iid εκθετικές τ.μ. E_1, E_2, \dots με μέση τιμή 1.

Για το Θεώρημα 2.7.1 δεν υπάρχουν αποτελέσματα αναφορικά με την επιλογή του βέλτιστου k . Η εκτίμηση του \hat{x}_p επηρεάζεται σε μεγάλο βαθμό από την επιλογή του k και παρουσιάζει πολλές και απότομες αυξομειώσεις. Έχει αποδειχθεί ότι για δείγματα μικρού και μεσαίου μεγέθους, ο συγκεκριμένος εκτιμητής δε δίνει συνεπείς εκτιμήσεις. Για τη συνέπεια του εκτιμητή $\hat{\xi}$ που αποτελεί μέρος του \hat{x}_p , θα πρέπει να θέσουμε τον περιορισμό $k = k(n) \rightarrow \infty$.

2.8. Η μέθοδος των υπερβάσεων πάνω από ένα υψηλό κατώφλι

2.8.1 Προσαρμογή της GPD

Η προηγούμενη ανάλυση βασίστηκε στην υπόθεση ότι τα δεδομένα εξετάζουμε προέρχονται από μία GEV ή ότι ανήκουν στο MDA αυτής. Μια διαφορετική προσέγγιση είναι να στηριχθούμε στον αριθμό των υπερβάλλουσων τιμών πάνω από κάποιο υψηλό κατώφλι. Υποθέτουμε ότι X_1, \dots, X_n αποτελούν iid δεδομένα με σ.κ. $F \in \text{MDA}(H_\xi)$ για κάποιο $\xi \in \mathbf{R}$. Θα επιλέξουμε αρχικά ένα υψηλό κατώφλι u και θα εκφράσουμε με

$$N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\}$$

το πλήθος των X_1, \dots, X_n που ξεπερνούν το κατώφλι u , ενώ οι αντίστοιχες υπερβάσεις θα εκφράζονται με Y_1, \dots, Y_{N_u} . Είδαμε ότι η υπερβάλλουσα σ.κ. ορίζεται ως

$$F_u(y) = P(X - u \leq y | X > u) = P(Y \leq y | X > u), y \geq 0$$

η οποία μπορεί να αποδοθεί και ως

$$\bar{F}(u+y) = \bar{F}(u)\bar{F}_u(y). \quad (2.28)$$

Η GPD με παραμέτρους $\xi \in \mathbf{R}$ και $\beta > 0$ έχει συνάρτηση κατανομής

$$G_{\xi, \beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, & \text{αν } \xi \neq 0 \\ 1 - e^{-x/\beta}, & \text{αν } \xi = 0 \end{cases}, x \in D(\xi, \beta),$$

όπου $D(\xi, \beta)$ είναι $[0, \infty)$ αν $\xi \geq 0$ και $[0, -\beta/\xi]$ αν $\xi < 0$ (Pickands (1975), Balkema and de Haan (1974)). Αποδεικνύεται ότι για κάθε $\xi \in \mathbf{R}$ θα ισχύει ότι $F \in \text{MDA}(H_\xi)$ αν και μόνο αν για κάποια θετική συνάρτηση β ισχύει

$$\lim_{u \rightarrow x_F} \sup_{0 < x < x_F - u} |\bar{F}_u(x) - \bar{G}_{\xi, \beta(u)}(x)| = 0.$$

Αν στηριχθούμε σε αυτό το αποτέλεσμα, τότε, για μεγάλο u , διαισθητικά οδηγούμαστε στο ακόλουθο αποτέλεσμα

$$\bar{F}_u(y) \approx \bar{G}_{\xi, \beta(u)}(y). \quad (2.29)$$

Παρατηρούμε ότι το β είναι συνάρτηση του κατώφλιου u , το οποίο επιλέγεται αρκετά μεγάλο, έτσι ώστε τα ξ και $\beta = \beta(u)$ να εκτιμώνται από τις υπερβάλλουσες παρατηρήσεις, και κατά συνέπεια οι εκτιμήσεις που θα προκύπτουν να εξαρτώνται από το u .

Από τη σχέση (2.28) υποδεικνύεται μια μέθοδος για την εκτίμηση της ουράς της F εκτιμώντας τις $\bar{F}_u(y)$ και $\bar{F}(u)$ χωριστά. Ο εκτιμητής της $\bar{F}(u)$ προκύπτει από την εμπειρική σ.κ.

$$\left(\widehat{\bar{F}(u)}\right) = \bar{F}_n(u) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i > u\}} = \frac{N_u}{n},$$

ενώ ο εκτιμητής της $\bar{F}_u(y)$ προκύπτει από την (2.29)

$$\left(\widehat{\bar{F}_u(y)}\right) = \bar{G}_{\hat{\xi}, \hat{\beta}}(y)$$

για κατάλληλα επιλεγμένα $\hat{\xi} = \hat{\xi}_{N_u}$ και $\hat{\beta} = \hat{\beta}_{N_u}$.

Ο εκτιμητής που προκύπτει για την ουρά $\bar{F}(u+y)$ για $y > 0$ θα είναι της μορφής

$$(\widehat{F}(u+y)) = (\widehat{F}(u))(\widehat{F}_u(y)) = \frac{N_u}{n} \left(1 + \frac{\widehat{\xi}}{\widehat{\beta}} y\right)^{-1/\widehat{\xi}}. \quad (2.30)$$

Στη περίπτωση που εξετάζουμε τις κατανομές Fréchet και Gumbel ($\xi \geq 0$), θα πρέπει να θέσουμε τον περιορισμό ότι $y \geq 0$ στην (2.30) για να εκτιμήσουμε την \widehat{F} στην πάνω ουρά. Από τα παραπάνω προκύπτει ότι ένας εκτιμητής ποσοστιαίων σημείων x_p θα είναι

$$\hat{x}_p = u + \frac{\hat{\beta}}{\widehat{\xi}} \left(\left(\frac{n}{N_u} (1-p) \right)^{-\widehat{\xi}} - 1 \right). \quad (2.31)$$

Επιπλέον αν ισχύει ότι $\widehat{\xi} < 0$ μπορούμε να βρούμε έναν εκτιμητή για το δεξιό οριακό σημείο x_F της F από την ακόλουθη σχέση

$$\hat{x}_F = u - \hat{\beta} / \widehat{\xi},$$

η οποία προκύπτει αν αντικαταστήσουμε το $\hat{x}_F = \hat{x}_1$ στην (2.31).

Η μέθοδος αυτή είναι μια καθαρά διαισθητική προσέγγιση του προβλήματος και έχει τις ρίζες της στην υδρολογία. Είναι γνωστή με την ονομασία Υπερβάσεις Κατωφλίου (*Peaks Over Threshold – POT*). Για την επιλογή του βέλτιστου κατωφλίου u , αντιμετωπίζουμε το ίδιο πρόβλημα όπως και με την επιλογή του βέλτιστου k για τον εκτιμητή Hill. Για μια πολύ υψηλή τιμή του u , ο αριθμός των τιμών που το υπερβαίνουν είναι πολύ μικρός και κατά συνέπεια οι εκτιμητές μας έχουν πολύ μεγάλη διακύμανση. Για μια μικρή τιμή του κατωφλίου u , οι εκτιμητές που προκύπτουν παρουσιάζουν μεγάλη μεροληψία.

Μια μέθοδος για την επιλογή του u είναι μέσω της γραμμικότητας της μέσης υπερβάλλουσας συνάρτησης $e(u)$ για την GPD. Αν έχουμε μια τ.μ. X με σ.κ. $G_{\xi, \beta}$, τότε η μέση υπερβάλλουσα συνάρτηση θα είναι της μορφής

$$e(u) = E(X - u | X > u) = \frac{\beta + \xi u}{1 - \xi}, \quad u \in D(\xi, \beta), \quad \xi < 1, \quad (2.32)$$

επομένως η $e(u)$ θα είναι γραμμική. Από την §2.1.2 είδαμε ότι η εμπειρική μέση υπερβάλλουσα συνάρτηση για ένα δείγμα X_1, \dots, X_n iid είναι

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \mathcal{A}_n(u)} (X_i - u), \quad u > 0,$$

όπου $N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\} = \text{card}\mathcal{A}_n(u)$. Η εμπειρική σ.κ. θα χρησιμοποιηθεί για τη γραφική προσέγγιση του βέλτιστου u ακολουθώντας την εξής λογική: θα επιλέξουμε εκείνο το $u > 0$ τέτοιο ώστε η $e_n(u)$ να είναι σχεδόν γραμμική για $x \geq u$. Το πρόβλημα που αντιμετωπίζεται σε αυτή τη διαδικασία είναι η ερμηνεία του σχεδόν γραμμική.

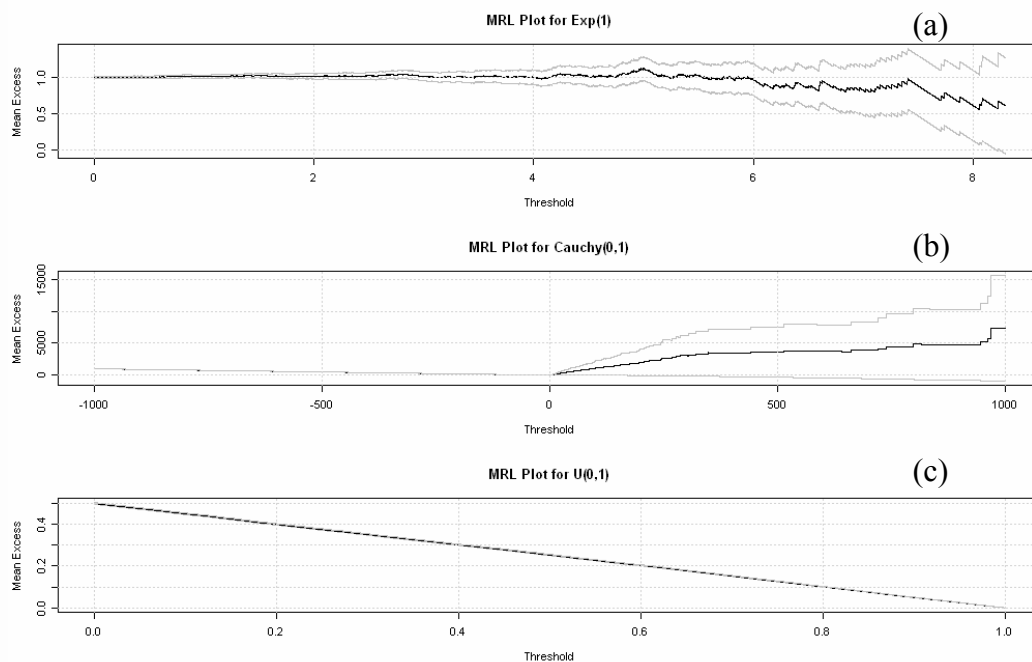
Έχοντας προσδιορίσει το κατώφλι u , μπορούμε να ελέγξουμε διαγραμματικά τη συμπεριφορά της ουράς της κατανομής, μέσω ενός *QQ plot*, έναντι της GPD. Το διάγραμμα αυτό μας επιτρέπει πρώτα να ελέγξουμε αν η κατανομή των υπερβάσεων του κατωφλίου u έχει βαριά ή ελαφριά ουρά, επιλέγοντας το $\xi = 0$. Σε αυτή τη περίπτωση το *QQ plot* των παρατηρήσεων της ουράς γίνεται έναντι της εκθετικής κατανομής. Αν οι υπερβάσεις του κατωφλίου u , ακολουθούν κατανομή με ελαφριά ουρά, τότε το *QQ plot* θα είναι σχεδόν γραμμικό. Αν οι υπερβάσεις του κατωφλίου u , ακολουθούν κατανομή με βαριά ουρά, τότε το *QQ plot* θα παρουσιάζει κοίλη απόκλιση από την ευθεία προσαρμογής. Στη περίπτωση που οι υπερβάσεις κατωφλίου ακολουθούν κατανομή με μεσαία ουρά, τότε το *QQ plot* θα παρουσιάζει κυρτή απόκλιση από την ευθεία προσαρμογής. Στην συνέχεια, μπορούμε να μεταβάλλουμε τη τιμή του ξ , έτσι ώστε να προσδιορίσουμε διαγραμματικά τη μορφή της GPD στην οποία προσαρμόζονται καλύτερα οι παρατηρήσεις της ουράς.

2.8.1.1 Παράδειγμα

Σχεδιάσαμε τα *MRL plot* προσομοιωμένων ($n = 20000$) παρατηρήσεων από $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ (ΣΧΗΜΑ 2-9). Στο *MRL plot* (a) της $\text{Exp}(1)$, παρατηρούμε ότι, όπως αναμενόταν, το διάγραμμα είναι σχεδόν γραμμικό για όλα τα u . Στο *MRL plot* (b) της $\text{Cauchy}(0,1)$ παρατηρούμε ότι μετά το 0 περίπου το διάγραμμα είναι σχεδόν γραμμικό. Αναφέρουμε ότι περιορίσαμε το διάγραμμα αυτό σε εύρος κατωφλίων $-1000 < u < 1000$, γιατί το αρχικό σύνολο παρατηρήσεων είχε πολύ μεγάλο εύρος και επιπλέον θέλαμε να περιορίσουμε το διάγραμμα στο κομμάτι εκείνο που θεωρητικά «ξεκινάει» η ουρά της κατανομής. Επίσης στο *MRL plot* (c) της $\text{U}(0,1)$ το διάγραμμα είναι γραμμικό για όλα τα u .

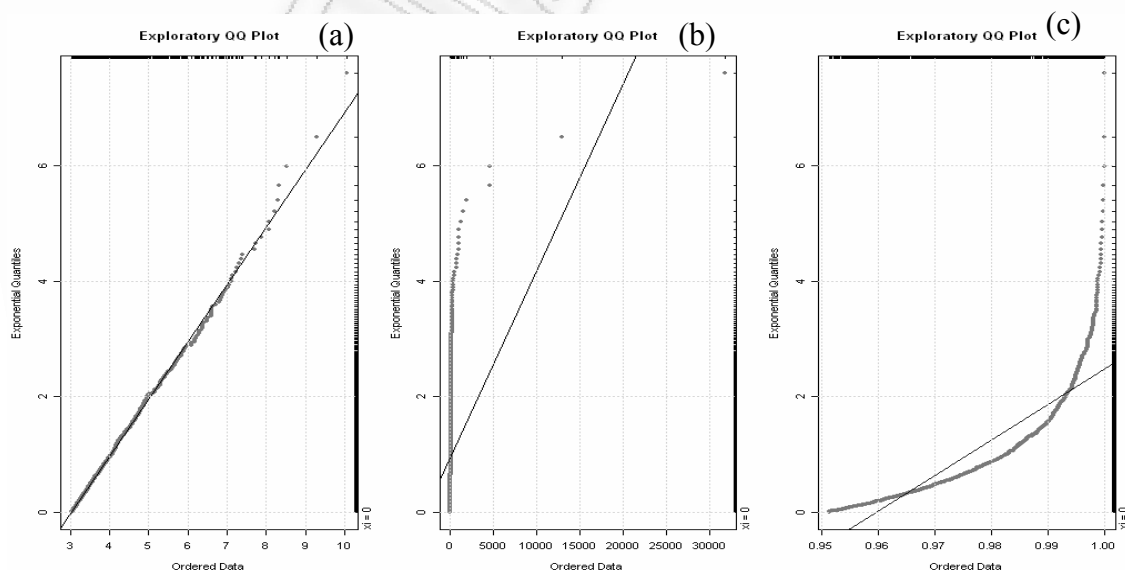
Για κάθε μια από τις παραπάνω κατανομές επιλέξαμε ένα κατώφλι u πάνω από το οποίο βρίσκεται το 5% των παρατηρήσεων, δηλαδή $u_e = 3.03899$, $u_c = 6.037972$ και $u_u = 0.9512253$ για τις $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$ αντίστοιχα. Σε αυτό το σημείο πρέπει να επισημάνουμε ότι οι McNeil and Saladin (1997) μέσω προσομοιώσεων έδειξαν ότι για κατανομές που ανήκουν στο MDA (H_1), όπως και η Cauchy, το επιλεγόμενο κατώφλι δεν είναι ποτέ ικανοποιητικά υψηλό για να δώσει καλές εκτιμήσεις.

ΣΧΗΜΑ 2-9



Σχεδιάζοντας τα *QQ plot* έναντι της GPD για $\xi = 0$ (ΣΧΗΜΑ 2-10), παρατηρούμε ότι για την $\text{Exp}(1)$ (a) το διάγραμμα παρουσιάζει πολύ καλή προσαρμογή. Στη περίπτωση των $\text{Cauchy}(0,1)$ (b) και $\text{U}(0,1)$ (c), παρατηρούμε κοίλη και κυρτή απόκλιση αντίστοιχα από την ευθεία προσαρμογής, δηλαδή την ύπαρξη βαριάς και μεσαίας ουράς.

ΣΧΗΜΑ 2-10



2.8.2 Εκτίμηση Μέγιστης Πιθανοφάνειας

Αν υποθέσουμε ότι τα δεδομένα $\mathbf{Y} = (Y_1, \dots, Y_n)$ που εκφράζουν τις υπερβάσεις κατοφλίου είναι iid με κοινή σ.κ. μια GPD με παραμέτρους ξ και β (γεγονός που, όπως είδαμε, ισχύει προσεγγιστικά για μεγάλο u), τότε θα έχουν πυκνότητα

$$f(x) = \frac{\xi}{\beta} \left(1 + \xi \frac{x}{\beta}\right)^{\frac{1}{\xi}-1}, \quad x \in D(\xi, \beta).$$

Η συνάρτηση λογάριθμο-πιθανοφάνειας θα είναι ίση με

$$\ell((\xi, \beta); \mathbf{Y}) = -n \ln \beta - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} Y_i\right).$$

Οι μεταβλητές της παραπάνω συνάρτησης θα πρέπει να ικανοποιούν τον περιορισμό $Y_i \in D(\xi, \beta)$ διαφορετικά θεωρούμε ότι $\ell((\xi, \beta); \mathbf{Y}) = -\infty$. Στη περίπτωση $\xi = 0$, η συνάρτηση λογάριθμο-πιθανοφάνειας είναι ίση με

$$\ell((0, \beta); \mathbf{Y}) = -n \ln \beta - \frac{1}{\beta} \sum_{i=1}^n Y_i.$$

και επομένως $\hat{\beta} = \bar{Y}$. Στην περίπτωση που $\xi \neq 0$ οι εκτιμήτριες μέγιστης πιθανοφάνειας $\hat{\xi}_n, \hat{\beta}_n$ υπολογίζονται αριθμητικά μέσω κατάλληλης επαναληπτικής προσεγγιστικής μεθόδου.

Η συγκεκριμένη μέθοδος λειτουργεί καλά αν $\xi > -1/2$. Σε αυτή τη περίπτωση θα ισχύει ότι

$$n^{1/2} \left(\hat{\xi}_n - \xi, \frac{\hat{\beta}_n}{\beta} - 1 \right) \xrightarrow{d} N(0, M^{-1}),$$

όπου

$$M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & 1 \\ 1 & 2 \end{pmatrix}$$

και $N(\mu, \Sigma)$ είναι η διμεταβλητή κανονική κατανομή με διάνυσμα μέσου μ και πίνακα συνδιακυμάνσεων Σ .

2.8.2.1 Παράδειγμα

Προσομοιώσαμε $n = 20000$ παρατηρήσεις από τις κατανομές $\text{Exp}(1)$, $\text{Cauchy}(0,1)$ και $\text{U}(0,1)$. Για κάθε ένα από αυτά τα σύνολα παρατηρήσεων, εκτιμήσαμε το κατώφλι u πάνω από το οποίο βρίσκονται 1000 παρατηρήσεις (5% των παρατηρήσεων) και προχωρήσαμε στη

προσαρμογή των δεδομένων πάνω στη GPD για τις παρατηρήσεις που ξεπερνούν το συγκεκριμένο κατώφλι. Τα αποτελέσματα παραθέτονται στον ακόλουθο πίνακα.

ΠΙΝΑΚΑΣ 2-2

Εκτιμώμενες παράμετροι της GPD με τη μέθοδο MLE

Κατανομή	Κατώφλι (u)	Παράμετρος	
		$\hat{\xi}$	$\hat{\beta}$
Exp(1)	3.03899	-0.002734266	1.007298028
Cauchy(0,1)	6.037972	1.001978	6.155700
U (0,1)	0.9512253	NaN	NaN

Παρατηρούμε ότι η εκτίμηση του ξ από τη προσαρμογή των ουρών των κατανομών στη GPD για την Cauchy(0,1) και την Exp(1), δεν απέχει πολύ από τα αποτελέσματα της μεθόδου *block maxima*. Στη περίπτωση της U(0,1) δε μπορέσαμε να εκτιμήσουμε τις παραμέτρους, κάτι που κατά κύριο λόγο οφείλεται στο γεγονός ότι το $\xi < -1/2$.

2.8.3 Αποτίμηση Κινδύνου Χαρτοφυλακίου (*Value at risk*) και αναμενόμενο μέγεθος ζημίας (*Expected Shortfall*)

Ως αποτίμηση κινδύνου χαρτοφυλακίου (VaR) ορίζεται μια πιθανή (με μικρή πιθανότητα πραγματοποίησης $p = 5\%$ ή 1%) απώλεια που μπορεί να πραγματοποιηθεί σε ένα συγκεκριμένο χρονικό ορίζοντα. Χρησιμοποιείται από χρηματοπιστωτικά και επενδυτικά ιδρύματα ως μέτρο του κινδύνου αγοράς (*market risk*) των χαρτοφυλακίων τους. Η VaR ενός χαρτοφυλακίου που αποφέρει τυχαίο κέρδος $-X$ (δηλ. ζημία X), σε επίπεδο σημαντικότητας $p \in (0,1)$, δίνεται από τον τύπο

$$\text{VaR}_p = \inf \{x \in \mathbb{R} : P(X > x) \leq 1 - p\} = \inf \{x \in \mathbb{R} : F(x) \geq p\}.$$

Στη θεωρία πιθανοτήτων η VaR ορίζεται ως το ποσοστιαίο σημείο της κατανομής ζημιάς F , δηλαδή

$$\text{VaR}_p = F^{\leftarrow}(p).$$

Για ορισμένη πιθανότητα $p > F(u)$ και μέσω της (2.31), αποδεικνύεται ότι ένας εκτιμητής του VaR θα είναι

$$\hat{\text{VaR}}_p = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1-p) \right)^{-\hat{\xi}} - 1 \right).$$

Ως αναμενόμενο μέγεθος ζημίας (ES) στο p επίπεδο ορίζεται η αναμενόμενη απόδοση του χαρτοφυλακίου, δεδομένου ότι κάποια απώλεια πραγματοποιείται το πολύ μέχρι το p ποσοστιαίο σημείο του εύρους των παρατηρήσεων. Εναλλακτικά μπορούμε να πούμε ότι το ES είναι το αναμενόμενο μέγεθος απώλειας που ξεπερνάει τη VaR, δηλαδή

$$\begin{aligned} ES_p &= E[X | X > VaR_p] \\ ES_p &= VaR_p + E[X - VaR_p | X > VaR_p] \end{aligned} \quad (2.33)$$

όπου το δεύτερο μέρος της ισότητας είναι ο μέσος της υπερβάλλουσας σ.κ. $F_{VaR_p}(y)$ πάνω από το κατώφλι VaR_p . Αν στη (2.33) εισάγουμε το κατώφλι u (βλ. McNeil (1999)), τότε προκύπτει ότι

$$ES_p = VaR_p + E[(X - u) - (VaR_p - u) | (X - u) > (VaR_p - u)]. \quad (2.34)$$

Από την (2.34) και μέσω της (2.32) ο εκτιμητής του ES, υπό τη προϋπόθεση ότι $\xi < 1$, γράφεται ως

$$\hat{ES}_p = \frac{VaR_p}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}}.$$

Αποδεικνύεται ότι το ES αυξάνεται καθώς αυξάνεται η τιμή του p και ότι η τιμή του είναι πάντα μεγαλύτερη από την αντίστοιχη τιμή του VaR.

Τόσο το VaR όσο και η ES αποτελούν πολύ χρήσιμα εργαλεία, στις περιπτώσεις κατανομών με ελαφριά και μεσαία ουρά. Για την εξέταση των κατανομών με βαριά ουρά, προφανώς λόγω της μεγάλης αβεβαιότητας που χαρακτηρίζει τέτοιες κατανομές η VaR και η ES θα πρέπει να αντιμετωπίζονται με επιφυλακτικότητα, όπως και κάθε εργαλείο που χρησιμοποιούμε για την εξέταση τέτοιων κατανομών.

2.8.3.1 Παράδειγμα

Χρησιμοποιώντας τα αποτελέσματα που προέκυψαν κατά τη προσαρμογή των προσομοιωμένων παρατηρήσεων των $Exp(1)$, $Cauchy(0,1)$ στη GPD (§ 2.8.2) εκτιμήσαμε τις τιμές των VaR και ES για $p = 0.99, 0.995, 0.999, 0.9995$ και 0.9999 . Να σημειωθεί ότι για την $Cauchy(0,1)$ η εκτίμηση του ES ήταν αδύνατη δεδομένου ότι η παράμετρος ξ εκτιμήθηκε μεγαλύτερη του 1 (και επομένως η αναμενόμενη τιμή είναι άπειρη). Ερμηνεύοντας τα αποτελέσματα για την $Exp(1)$, υπάρχει πιθανότητα 1% μία παρατήρηση X_i να ξεπεράσει την τιμή 4.657614 και δεδομένου ότι την ξεπέρασε τότε η τιμή της αναμένεται να είναι 5.657752.

ΠΙΝΑΚΑΣ 2-3

Εκτιμήσεις VaR_p και ES_p και τα θεωρητικά ποσοστιαία σημεία προσομοιωμένων παρατηρήσεων από $Exp(1)$ και $Cauchy(0,1)$.

p	Exp(1)			Cauchy (0,1)		
	<i>Theoretical Quantiles</i>	\hat{VaR}_p	\hat{ES}_q	<i>Theoretical Quantiles</i>	\hat{VaR}_p	\hat{ES}_q
0.9900	4.609696	4.657614	5.657752	31.8205	30.74098	NaN
0.9950	5.298512	5.352094	6.350338	63.6567	61.67215	NaN
0.9990	6.892868	6.959559	7.953420	318.309	309.76773	NaN
0.9995	7.577362	7.649680	8.641659	636.619	620.49112	NaN
0.9999	9.161711	9.247056	10.234680	3183.1	3112.7694	NaN

ТАНЕЦЫ И МУЗЫКА

Κεφάλαιο 3, Ανάλυση Δεδομένων

3.1 Εισαγωγή

Έχοντας αναφερθεί στη Θεωρία Κατανομών ακραίων τιμών στο Κεφάλαιο 1 και σε μερικές από τις στατιστικές μεθόδους που εφαρμόζονται για την ανάλυση τους στο Κεφάλαιο 2, σε αυτό το κεφάλαιο θα δούμε δυο πρακτικές εφαρμογές. Η μελέτη των δεδομένων έγινε με το στατιστικό πακέτο R (R-2.6.0) και χρησιμοποιήθηκε το πακέτο fExtremes (240.10068 & 260.72) (βλ. Παράρτημα 1).

3.2 Δείκτης Standard & Poor' s 500

Τα δεδομένα που θα εξετάσουμε προέρχονται από την επεξεργασία των ιστορικών τιμών (*historical data*) του δείκτη Standard & Poor' s 500 (S&P 500) για τη περίοδο 01/01/1977 – 30/06/2007*. Ο S&P 500 περιλαμβάνει τις μετοχές 500 εταιρειών μεγάλης κεφαλαιοποίησης (*Large-Cap* ή *blue chip*), δηλαδή εταιρειών με κεφαλαιοποίηση μεγαλύτερη των 10 εκ. δολαρίων Αμερικής (USD). Η πλειονότητα αυτών των εταιρειών λειτουργεί στην Αμερική και οι μετοχές τους αποτελούν αντικείμενο συναλλαγής στις δυο μεγαλύτερες χρηματιστηριακές αγορές της χώρας, το χρηματιστήριο της Νέας Υόρκης (*New York Stock Exchange*) και το Nasdaq. Η πορεία του, αντικατοπτρίζει τη δυναμικότητα αυτών των 500 εταιρειών υψηλής απόδοσης, γεγονός που έχει οδηγήσει τους αναλυτές να θεωρούν ότι ο δείκτης αποτελεί καθρέπτη της Αμερικανικής αγοράς. Πριν από τη κρίση που επικράτησε στα Αμερικανικά χρηματιστήρια με τις εταιρείες dot-com, ο δείκτης έφτασε στο όριο των 1552.87 (24/3/2000) και στη συνέχεια λόγω της επιφυλακτικότητας της αγορά έχασε περίπου 50% ης αξίας του, πέφτοντας κάτω από τις 800 μονάδες. Τον Μάιο του 2007 ο δείκτης έκλεισε στις 1530.23 μονάδες και μέχρι τον Ιούνιο του 2007 συνέχιζε να έχει ανοδική πορεία.

3.2.1 Υπολογισμός των αποδόσεων ενός δείκτη

Αυτό που μας ενδιαφέρει δεν είναι η πορεία του δείκτη αλλά η απόδοση (*return*) που έχει ο δείκτης σε ημερήσια βάση. Για τη μετατροπή των αρχικών δεδομένων σε μια σειρά αποδόσεων μπορούμε εφαρμόσουμε τη μέθοδο των απλών αποδόσεων (*simple returns*) ή των λο-

γαριθμικών αποδόσεων (*log returns*). Στη συγκεκριμένη εργασία, η ανάλυση των δεδομένων θα γίνει χρησιμοποιώντας τις λογαριθμικές αποδόσεις.

Συμβολίζουμε με R_t την απόδοση, και με p_t την τιμή του δείκτη τη χρονική στιγμή t . Οι λογαριθμικές αποδόσεις ορίζονται ως

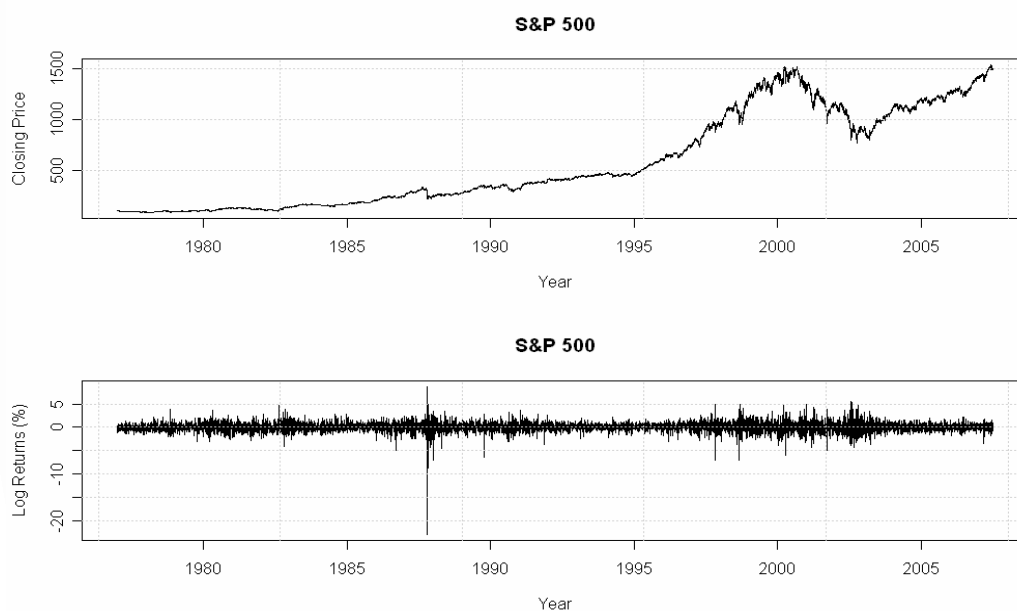
$$R_t = \ln(p_t / p_{t-1}) \times 100.$$

Επίσης αγνοούμε την καταβολή μερισμάτων ή υποθέτουμε ότι έχει ήδη γίνει προσαρμογή στα αρχικά δεδομένα. Υπογραμμίζεται ότι, σύμφωνα με το κλασικό μοντέλο Black-Scholes οι λογαριθμικές αποδόσεις μπορούν να θεωρηθούν iid τ.μ. που ακολουθούν την Κανονική κατανομή.

3.2.2. Διαγραμματική εξέταση των δεδομένων

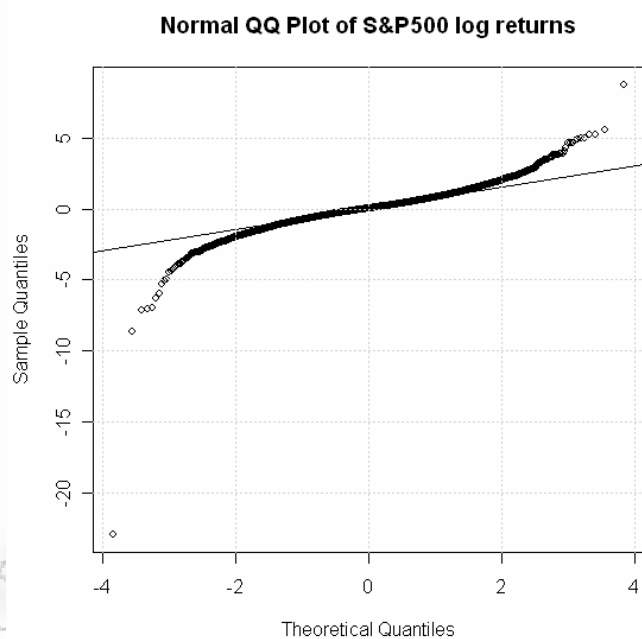
Αν εξετάσουμε τη πορεία των τιμών κλεισίματος του δείκτη S&P 500 και των λογαριθμικών αποδόσεων του για τη περίοδο 1/1/1997 – 30/06/2007 (ΣΧΗΜΑ 3-1), παρατηρούμε αρχικά την ανοδική τάση που έχει ο δείκτης. Επιπλέον βλέπουμε ότι στις περιόδους που ο δείκτης είχε ανοδική πορεία, οι αποδόσεις ήταν σχετικά σταθερές γύρω από το μηδέν. Αυτό που μας προκαλεί εντύπωση είναι η απότομη μεταβολή που παρουσιάζεται στις αποδόσεις μεταξύ 1985 και 1990.

ΣΧΗΜΑ 3-1



Για να εξετάσουμε την υπόθεση ότι τα δεδομένα είναι κανονικά και ισόνομα, ελέγχουμε το *QQ plot* έναντι της κανονικής κατανομής όπου προκύπτει ότι το διάγραμμα έχει μορφή S, δηλαδή υπάρχει σχετικά καλή προσαρμογή στη κανονική κατανομή του συνόλου των αποδόσεων που συγκεντρώνονται γύρω από το μέσο της κατανομής. (ΣΧΗΜΑ 3-2) ενώ στις ουρές παρατηρούμε ότι υπάρχει απόκλιση των παρατηρήσεων από την κανονική κατανομή. Ως συνέπεια αυτού οποιαδήποτε εκτίμηση κάναμε στις ουρές κάτω από την υπόθεση της κανονικότητας θα περιείχε ουσιαστικά σφάλματα. Δεδομένου ότι οι πρώτες παρατηρήσεις βρίσκονται κάτω από τη ευθεία προσαρμογής και οι τελευταίες παρατηρήσεις βρίσκονται πάνω από αυτήν, μπορούμε να υποστηρίξουμε ότι οι λογαριθμικές αποδόσεις ακολουθούν κατανομή με βαριά ουρά.

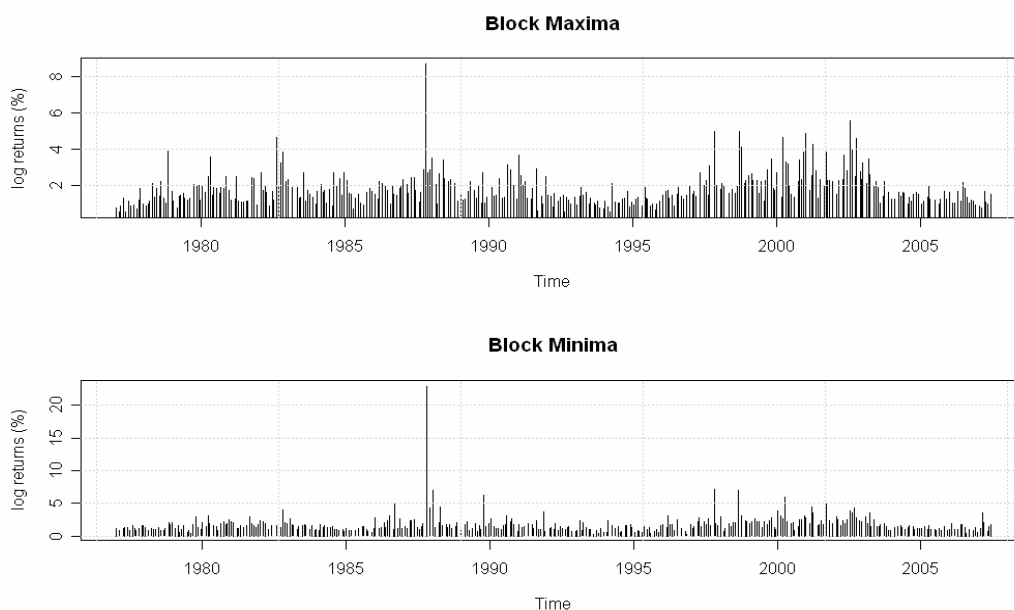
ΣΧΗΜΑ 3-2



3.2.4 Εύρεση των Block Maxima και Minima

Αναφερθήκαμε στο Κεφάλαιο 2 στην προσαρμογή των ετησίων μεγίστων, όπου X_i εκφράζονται ως τα μέγιστα ανεξάρτητων χρονικών περιόδων μεγέθους $s = 365$. Δεδομένου ότι έχουμε να αναλύσουμε ένα δείκτη, ο οποίος αντικατοπτρίζει την τρέχουσα οικονομική πραγματικότητα, θα μελετήσουμε τη πορεία του σε πιο μικρά χρονικά διαστήματα, αυτά των μηνών.

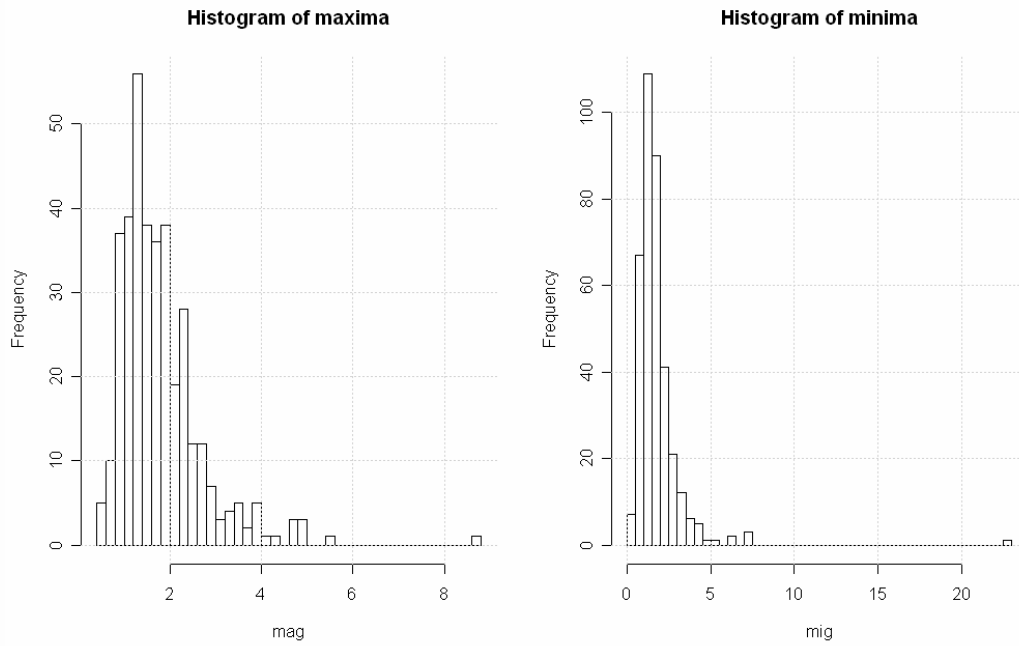
ΣΧΗΜΑ 3-3



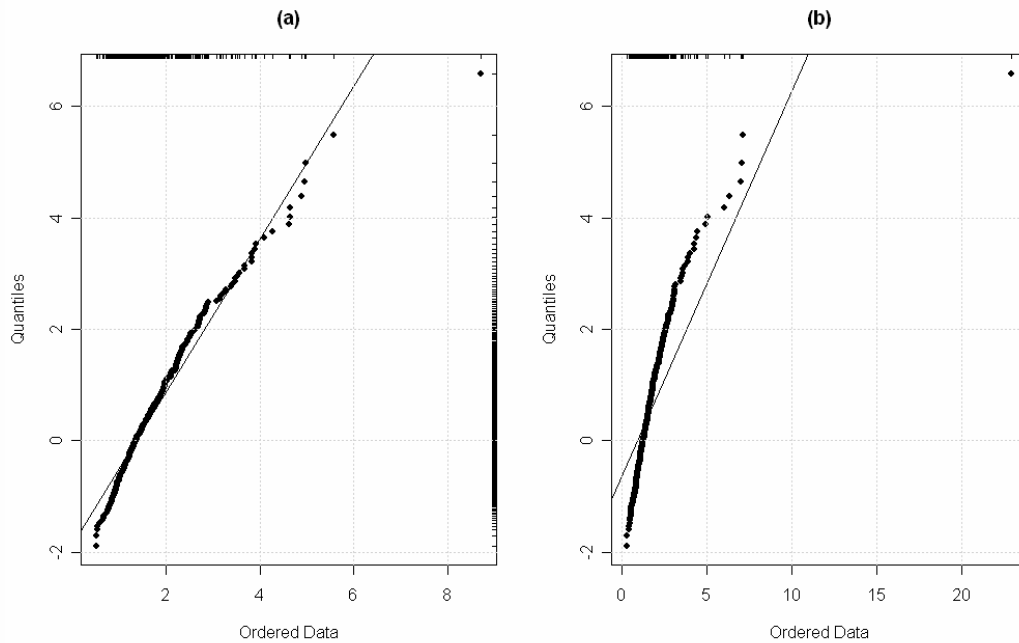
Στα διαγράμματα των μηνιαίων μεγίστων και ελαχίστων του δείκτη (ΣΧΗΜΑ 3-3), βλέπουμε για τις ίδιες ακριβώς περιόδους τη συμπεριφορά των *log returns* (σε απόλυτη κλίμακα). Η πρώτη διαφορά που διαπιστώνουμε μεταξύ των δυο διαγραμμάτων είναι στη κλίμακα τους. Στις τιμές των μηνιαίων ελαχίστων (*Block Minima*) συμπεριλαμβάνεται η τιμή της «Black Monday», που ξεπερνάει κατά πολύ τις υπόλοιπες αρνητικές αποδόσεις. Μπορούμε να υποστηρίξουμε ότι, με μόνη εξαίρεση αυτή τη τιμή, η συμπεριφορά των μηνιαίων μεγίστων (*Block Maxima*) δε διαφέρει σημαντικά από αυτή των μηνιαίων ελαχίστων.

Εξετάζοντας τα ιστογράμματα των μηνιαίων μεγίστων και ελαχίστων αντίστοιχα (ΣΧΗΜΑ 3-4), παρατηρούμε ότι τα μηνιαία μέγιστα παρουσιάζουν κυρτότητα προς τα δεξιά. Μια συντηρητική εκτίμηση που μπορούμε να κάνουμε είναι ότι τα μέγιστα ακολουθούν κατανομή με «μεσαία» ουρά. Τα μηνιαία ελάχιστα παρουσιάζουν και αυτά μια κυρτότητα προς τα δεξιά, αλλά δε μπορούμε να υποστηρίξουμε με βεβαιότητα ότι ακολουθούν κάποια κατανομή με «βαριά» ουρά. Και σε αυτό το διάγραμμα, παρατηρούμε ότι υπάρχει μια μεμονωμένη παρατήρηση που απομακρύνεται αισθητά από το σύνολο των υπολοίπων παρατηρήσεων. Η συγκεκριμένη παρατήρηση θα μπορούσε να θεωρηθεί ως έκτροπη (*outlier*) και να απαλειφθεί από το σύνολο των παρατηρήσεων. Δεδομένου όμως ότι επιθυμούμε να εξετάσουμε τη συμπεριφορά των ακραίων τιμών και η συγκεκριμένη παρατήρηση προφανώς ανήκει σε αυτή τη κατηγορία, θα προχωρήσουμε στην ανάλυση συμπεριλαμβάνοντας και αυτή τη τιμή στο σύνολο των μηνιαίων ελαχίστων.

ΣΧΗΜΑ 3-4



ΣΧΗΜΑ 3-5



Έχοντας εξάγει τα μηνιαία μέγιστα και ελάχιστα από το αρχικό σύνολο παρατηρήσεων, θα προσπαθήσουμε να δούμε μέσω *QQ plot* (ΣΧΗΜΑ 3-5) σε ποια οικογένεια κατανομών μπορεί να ανήκουν. Ελέγχοντας το *QQ plot* των μηνιαίων μεγίστων έναντι της Gumbel (a), παρατηρούμε ότι υπάρχει μια ελαφρά κοίλη απόκλιση από την ευθεία προσαρμογής, γεγονός που

θα μπορούσε να μεταφραστεί σαν «βαριά» ουρά. Στο ίδιο ακριβώς συμπέρασμα οδηγούμαστε και από την εξέταση του QQ plot των μηνιαίων ελαχίστων έναντι της Gumbel (b).

3.2.5 Προσαρμογή των αποδόσεων του S&P 500 στην GEV και εκτίμηση των παραμέτρων

Βάση για τη προσαρμογή των δεδομένων θα αποτελέσουν τα μηνιαία μέγιστα και ελάχιστα που έχουμε πάρει. Η προσαρμογή των συνόλων αυτών στη GEV, θα διευκολύνει στη πρόβλεψη μελλοντικών κερδών ή απωλειών ενός επενδυτή ή στην εκτίμηση της πορείας της αγοράς σε μηνιαία βάση. Για τη χρονική περίοδο που εξετάζουμε έχουμε 366 μήνες, τους οποίους θα θέσουμε σαν περιόδους (*blocks*).

Στο Κεφάλαιο 2 (§ 2.4.1) αναπτύχθηκε θεωρητικά η μέθοδος εκτίμησης των παραμέτρων της GEV μέσω της μεθόδου MLE. Πρακτικά ο τρόπος εκτίμησης είναι αρκετά απλός, δεδομένου ότι χρησιμοποιείται κάποια στατιστικό πακέτο. Στο πακέτο fExtremes του R υπάρχει έτοιμο σύνολο εντολών με τα οποία γίνεται η εκτίμηση των παραμέτρων, βάση του οποίου είναι ένας επαναληπτικός αλγόριθμος. Για την εύρεση των παραμέτρων, ο αλγόριθμος αυτός ξεκινάει από κάποιες αρχικές τιμές (οι μ και ψ ορίζονται ως κάποια συνάρτηση των δεδομένων και το ξ ως σταθερά). Κάθε φορά που πραγματοποιείται κάποια εκτίμηση, τα αποτελέσματα αντικαθίστανται στον αρχικό τύπο και ξεκινάει μια καινούργια διαδικασία εκτίμησης, μέχρι να εντοπισθούν από τον αλγόριθμο οι μέγιστες τιμές. Το συγκεκριμένο πακέτο παρέχει τη δυνατότητα στον ερευνητή να εργαστεί με το αρχικό σύνολο δεδομένων, αφού μπορεί να ορίσει τα *blocks* στα οποία θα χωριστούν οι παρατηρήσεις απ' ευθείας στον τύπο.

Από τη προσαρμογή των δεδομένων στην GEV και μέσω της MLE, προκύπτει ότι τα μέγιστα ακολουθούν την GEV με παραμέτρους $\hat{\xi} = 0.1542760$, $\hat{\mu} = 1.3682504$ και $\hat{\psi} = 0.5519405$, δηλαδή

$$H_{\hat{\xi}, \hat{\mu}, \hat{\psi}}(x) = \exp \left\{ - \left(1 + 0.1542760 \left(\frac{x - 1.3682504}{0.5519405} \right) \right)^{-1/0.1542760} \right\}, \quad 1 + \hat{\xi} x > 0.$$

Τα αντίστοιχα τυπικά σφάλματα είναι 0.04076116, 0.03257193 και 0.02521920. Κάτω από την υπόθεση της ασυμπτωτικής κανονικότητας των MLE's, τα αντίστοιχα 95% ασυμπτωτικά δ.ε. είναι (0.0743856, 0.2341664), (1.304411, 1.432090) και (0.5025118, 0.6013692). Από την εκτίμηση του ξ και από το ασυμπτωτικό δ.ε., μπορούμε να υποστηρίξουμε ότι τα μηνιαία μέγιστα ακολουθούν κατανομή Fréchet, δηλαδή κατανομή με «βαριά» ουρά. Το αποτέλεσμα αυτό έρχεται να επιβεβαιώσει τα αρχικά συμπεράσματα από την εξέταση των διαγραμμάτων.

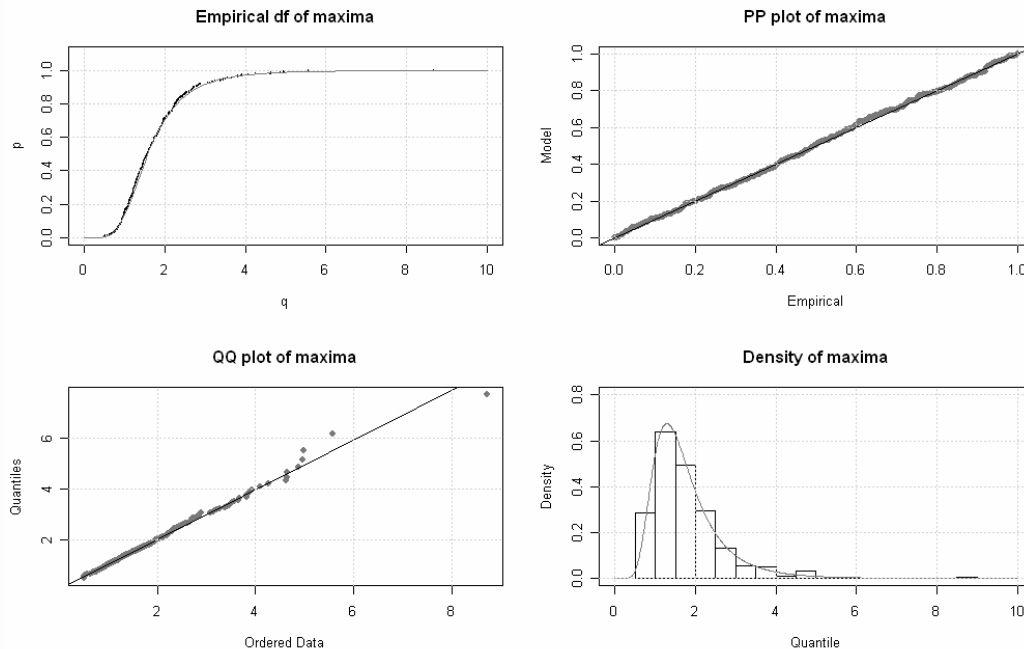
Όμοια, τα ελάχιστα ακολουθούν τη GEV με παραμέτρους $\hat{\xi} = 0.2035556$, $\hat{\mu} = 1.2464518$ και $\hat{\psi} = 0.5880980$, δηλαδή έχουμε ότι

$$H_{\hat{\xi}, \hat{\mu}, \hat{\psi}}(x) = \exp \left\{ - \left(1 + 0.2035556 \left(\frac{x - 1.2464518}{0.588098} \right) \right)^{-1/0.2035556} \right\}, \quad 1 + \hat{\xi} x > 0.$$

Σε αυτή τη περίπτωση τα αντίστοιχα τυπικά σφάλματα είναι 0.03733067, 0.03426558 και 0.02702005, ενώ τα αντίστοιχα δ.ε. είναι (0.1303888, 0.2767224), (1.179292, 1.313611) και (0.5351397, 0.6410563). Όπως και στη περίπτωση των μηνιαίων μεγίστων, και εδώ το συμπέρασμα είναι ότι τα μηνιαία ελάχιστα ακολουθούν τη κατανομή Fréchet.

Για να ελέγξουμε τα αποτελέσματα της προσαρμογής των μηνιαίων μεγίστων και ελαχίστων πάνω στη GEV, σχεδιάσαμε κάποια γραφήματα για να συγκρίνουμε τις εμπειρικές παρατηρήσεις με τα εκτιμηθέντα μοντέλα. (ΣΧΗΜΑ 3-6). Αναφορικά με τα μηνιαία μέγιστα, διαπιστώνουμε από το γράφημα της εμπειρικής σ.κ., το *PP plot*, όσο και από το *QQ plot* πολύ καλή προσαρμογή παρά την ύπαρξη κάποιων εκτροπών παρατηρήσεων. Το ίδιο συμπέρασμα προκύπτει και από το διάγραμμα της πυκνότητας.

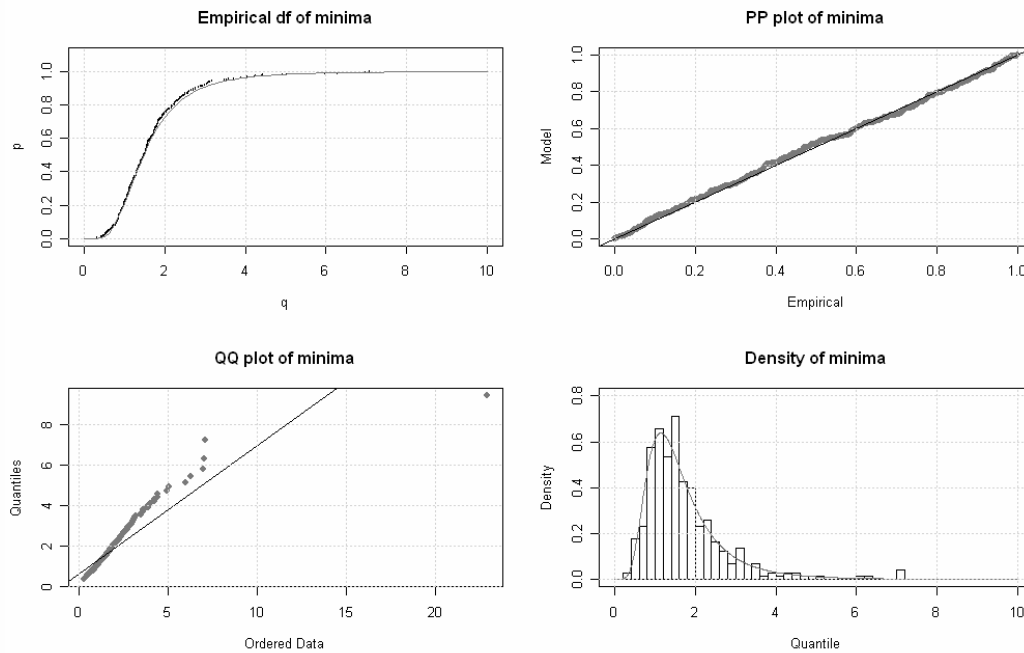
ΣΧΗΜΑ 3-6



Εξετάζοντας την περίπτωση των μηνιαίων ελαχίστων (ΣΧΗΜΑ 3-7), παρατηρούμε ότι ενώ το *PP plot* δίνει πολύ καλή προσαρμογή, το *QQ plot* παρουσιάζει κοίλη απόκλιση από

την ευθεία προσαρμογής. Αυτό ίσως να οφείλεται στην ύπαρξη της έκτροπης παρατήρησης που «έλκει» τη γραμμή προσαρμογής. Το γενικό συμπέρασμα στο οποίο μπορούμε να καταλήξουμε είναι μια ικανοποιητική προσαρμογή του μοντέλου στην εμπειρική κατανομή.

ΣΧΗΜΑ 3-7

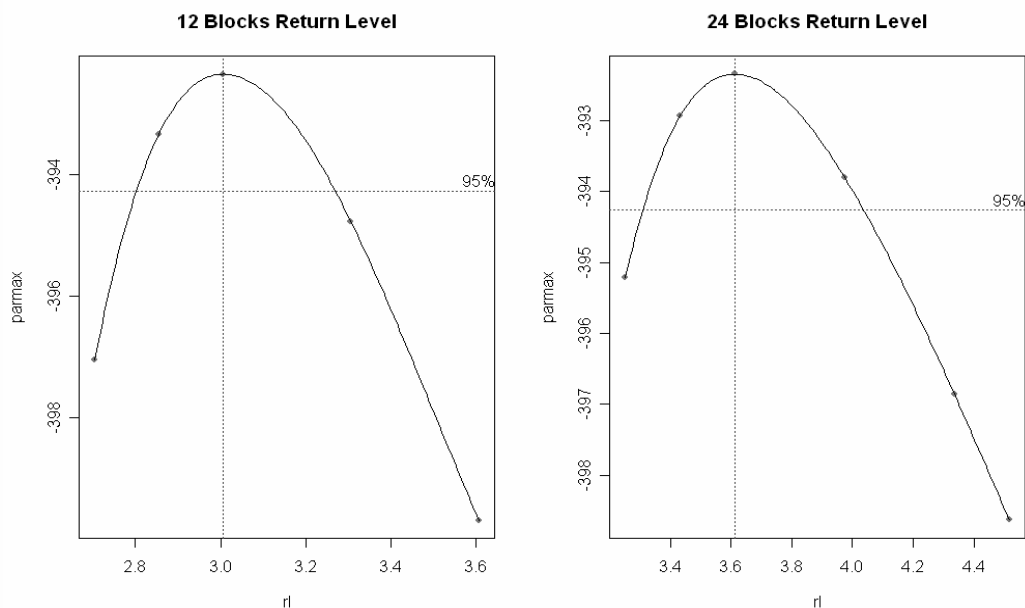


3.2.6 Εκτίμηση της στάθμης απόδοσης

Έχοντας εκτιμήσει τη κατανομή GEV που ακολουθούν τα μηνιαία μέγιστα και ελάχιστα, μπορούμε να υπολογίσουμε τις περιόδους απόδοσης, δηλαδή για κάθε ένα από τα υπό εξέταση σύνολα παρατηρήσεων να πάρουμε εκτιμήσεις για εκείνα τις στάθμες αποδόσεων που θα ξεπεραστούν μια φορά μέσα σε k μήνες.

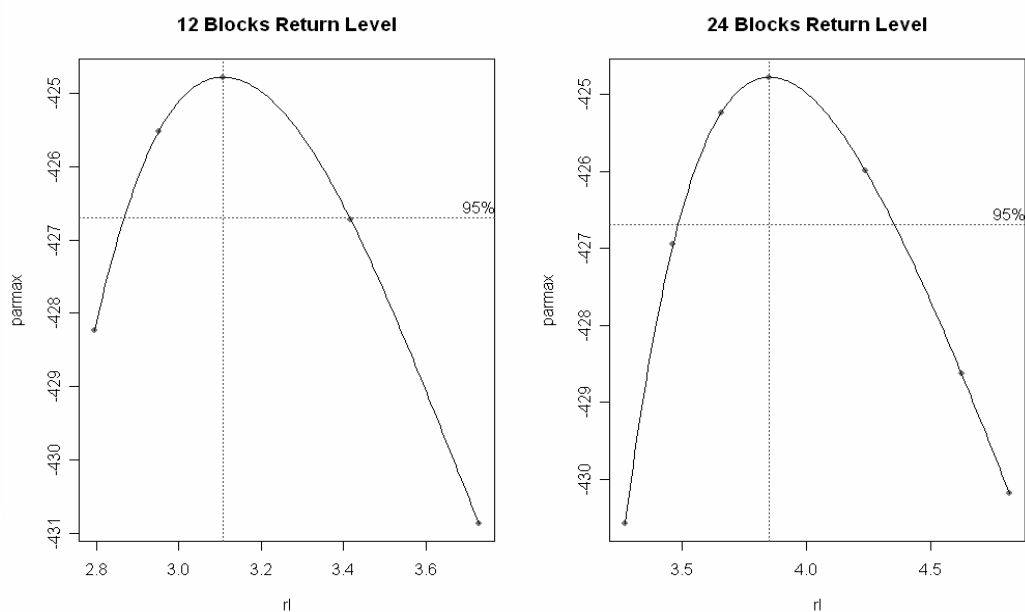
Εφαρμόζοντας τη διαδικασία όπως περιγράφηκε στη § 2.6.3, εκτιμήσαμε ότι η στάθμη των θετικών λογαριθμικών αποδόσεων που θα ξεπεραστεί μια φορά στους 12 μήνες είναι 3.00483%, ενώ η στάθμη απόδοσης που θα ξεπεραστεί μια φορά τους 24 μήνες είναι 3.613095%. Η καμπύλη της συνάρτησης κατατομής λογάριθμο-πιθανοφάνειας (ΣΧΗΜΑ 3-8), μας δίνει τα κατά προσέγγιση 95% δ.ε., τα οποία είναι (2.804054%, 3.266112%) και (3.315332%, 4.033412%) για τους 12 και 24 μήνες αντίστοιχα.

ΣΧΗΜΑ 3-8



Για τα μηνιαία ελάχιστα, προκύπτει ότι η απόδοση εκείνη που θα ξεπεραστεί μια φορά στους 12 μήνες είναι -3.10654% , ενώ η απόδοση -3.850732% θα ξεπεραστεί μια φορά στους 24 μήνες. Τα αντίστοιχα δ.ε. όπως προκύπτουν από την συνάρτηση κατατομής λογάριθμο-πιθανοφάνειας (ΣΧΗΜΑ 3-9), είναι $(-3.414072\%, -2.870818\%)$ και $(-4.349006, -3.489847\%)$.

ΣΧΗΜΑ 3-9



3.3 Εκτίμηση των παραμέτρων μέσω της υπόθεσης MDA

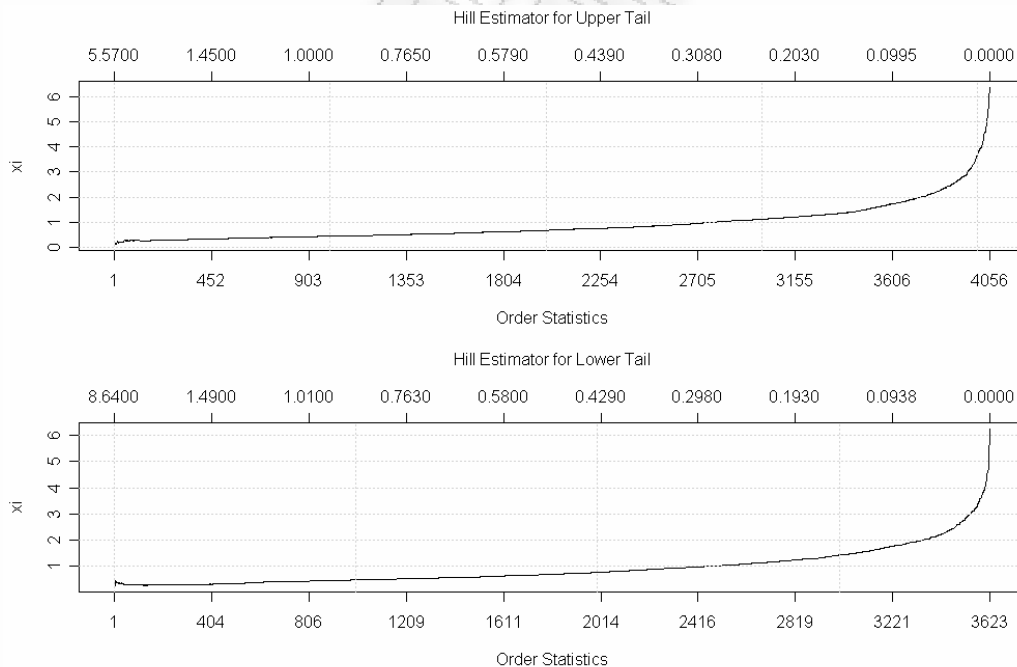
Γνωρίζουμε ότι ο εκτιμητής Hill ορίζεται ως

$$\hat{a}^{(H)} = \hat{a}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1}$$

υπό την προϋπόθεση ότι τα δεδομένα του αρχικού μας δείγματος X_1, \dots, X_n είναι iid με σ.κ. F που ανήκει στο $MDA(\Phi_\alpha)$, $\alpha > 0$, και για την οποία θα ισχύει ότι η συνάρτηση επιβίωσης της θα είναι ίση με $x^{-\alpha} L(x)$ για $x > 0$ και για κάποια βραδέως μεταβαλλόμενη συνάρτηση L . Εφαρμόζοντας τον συγκεκριμένο τύπο, προκύπτουν τα διαγράμματα Hill plot. Στα διαγράμματα αυτά το $\xi = 1/\alpha$ απεικονίζεται ως συνάρτηση του k , δηλαδή των διατεταγμένων τιμών (κάτω οριζόντιος άξονας) και των κατωφλίων u (πάνω οριζόντιος άξονας) σε απόλυτους όρους. Βλέπουμε ότι όσο αυξάνεται η τιμή του k , αυξάνεται και η τιμή του ξ ενώ ταυτόχρονα φθίνει ο αριθμός των παρατηρήσεων που ξεπερνάνε το κατώφλι u .

ΣΧΗΜΑ 3-10

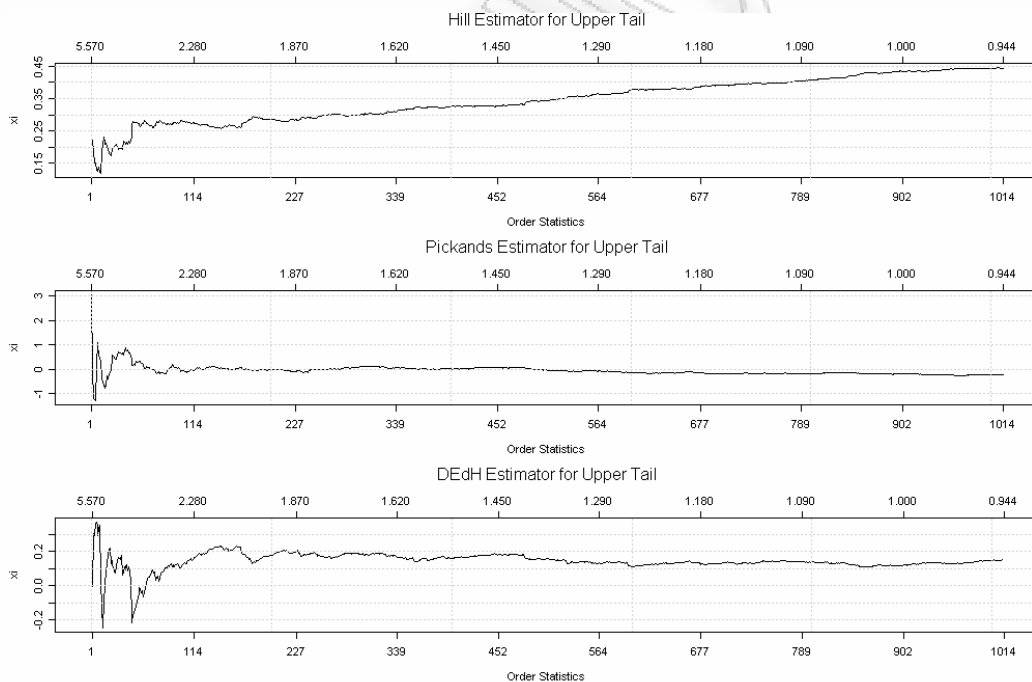
Hill plot των θετικών (πάνω) και των αρνητικών αποδόσεων (κάτω).



Όπως αναφέρθηκε στο Κεφάλαιο 2, για να προσδιορίσουμε τη βέλτιστη τιμή του k διαγραμματικά, πρέπει να εντοπίσουμε εκείνη τη περιοχή που το διάγραμμα είναι σχεδόν οριζόντιο. Στα διαγράμματα (ΣΧΗΜΑ 3-10) που παριστούν γραφικά τη πορεία του εκτιμητή Hill

για τις άνω διατεταγμένες και κάτω διατεταγμένες τιμές αντίστοιχα σε όλο το σύνολο των παρατηρήσεων, είναι δύσκολο να προσδιοριστεί οπτικά η τιμή του ξ . Τα δυο διαγράμματα, τόσο των θετικών όσο και των αρνητικών αποδόσεων, φαίνεται να είναι οριζόντια για $k < 400$, δηλαδή για τιμές του ξ κοντά στο μηδέν. Για $k > 400$, τα δυο διαγράμματα ακολουθούν ανοδική πορεία, δηλαδή σταδιακά χρησιμοποιούνται παρατηρήσεις από το κέντρο της κατανομής για τη εκτίμηση του ξ . Δε πρέπει να παρερμηνεύσουμε την αδυναμία του διαγράμματος Hill ως ένδειξη ότι οι άνω και οι κάτω διατεταγμένες τιμές δεν ακολουθούν κατανομή με βαριά ουρά. Οι Drees, de Haan and Resnick (2005) αναφέρουν ότι ένα Hill plot δίνει πολύ καλές εκτιμήσεις του ξ , μόνο αν η αρχική κατανομή των παρατηρήσεων είναι Pareto ή πολύ κοντά στη Pareto. Επίσης επισήμαναν το γεγονός ότι πολλές φορές το διάγραμμα μένει για μικρό αριθμό παρατηρήσεων στην περιοχή του ξ .

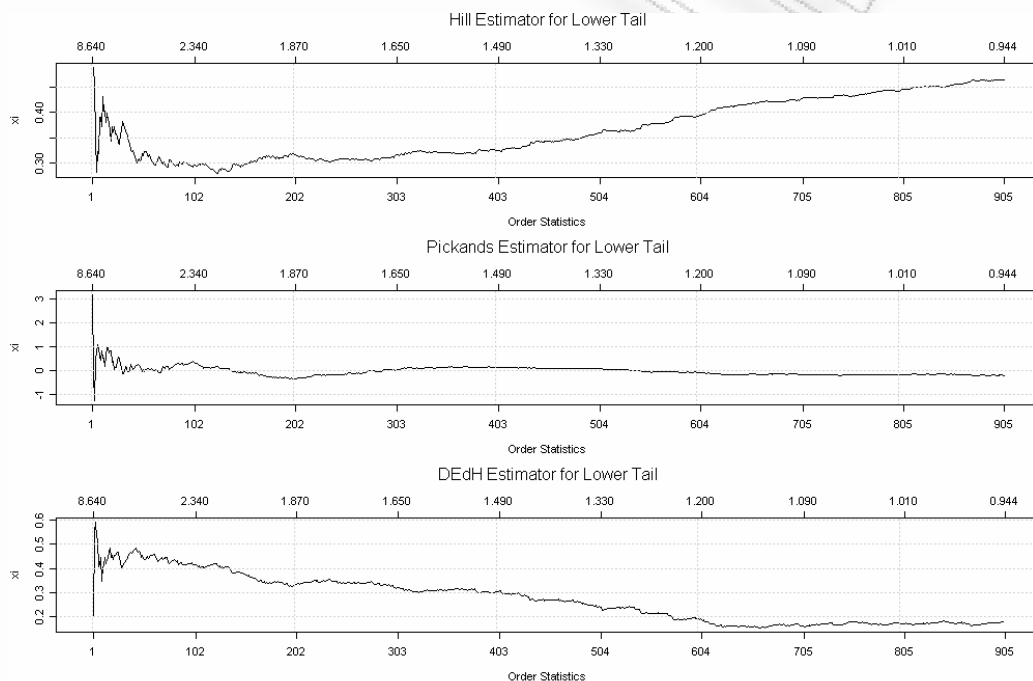
ΣΧΗΜΑ 3-11



Θα επανεξετάσουμε τον εκτιμητή Hill συγκρίνοντας τον με τους εκτιμητές Pickands και DEdH. Για να διευκολύνουμε στη σύγκριση μεταξύ των αποτελεσμάτων των εκτιμητών και για να πάρουμε την ακριβή εικόνα του τι γίνεται στις πρώτες διατεταγμένες τιμές, θα περιορίσουμε τις εκτιμήσεις στο $\frac{1}{4}$ του συνόλου των παρατηρήσεων (εύρος εκτίμησης του εκτιμητή Pickands). Όπως προκύπτει από τα διαγράμματα (ΣΧΗΜΑ 3-11) ο εκτιμητής Hill φαίνεται να βρίσκεται στη περιοχή $0.15 < \xi < 0.25$. Πιο σταθερή συμπεριφορά κατά μήκος των

δεδομένων που εξετάζουμε έχει ο εκτιμητής Pickands. Συγκεκριμένα το γράφημα φαίνεται να γίνεται γραμμικό για $k \in (114, 500)$, δηλαδή το ξ εκτιμάται κατά προσέγγιση κοντά στο μηδέν. Το γράφημα του εκτιμητή DEdH εμφανίζει και αυτό σταθερή συμπεριφορά και γίνεται γραμμικό για $k \geq 600$, με το ξ να εκτιμάται πάλι κοντά στο μηδέν. Σε αντίθεση με τους Pickands και DEdH, ο εκτιμητής Hill αδυνατεί να δώσει μια σταθερή εκτίμηση. Παρατηρούμε ότι κάθε εκτιμητής προσεγγίζει το ξ , σε διαφορετικό k , και δεδομένης της ευαισθησίας του Pickands στη τιμή του επιλεγόμενου k , θεωρείται πιο ασφαλές να επιλέξουμε τον εκτιμητή DEdH.

ΣΧΗΜΑ 3-12



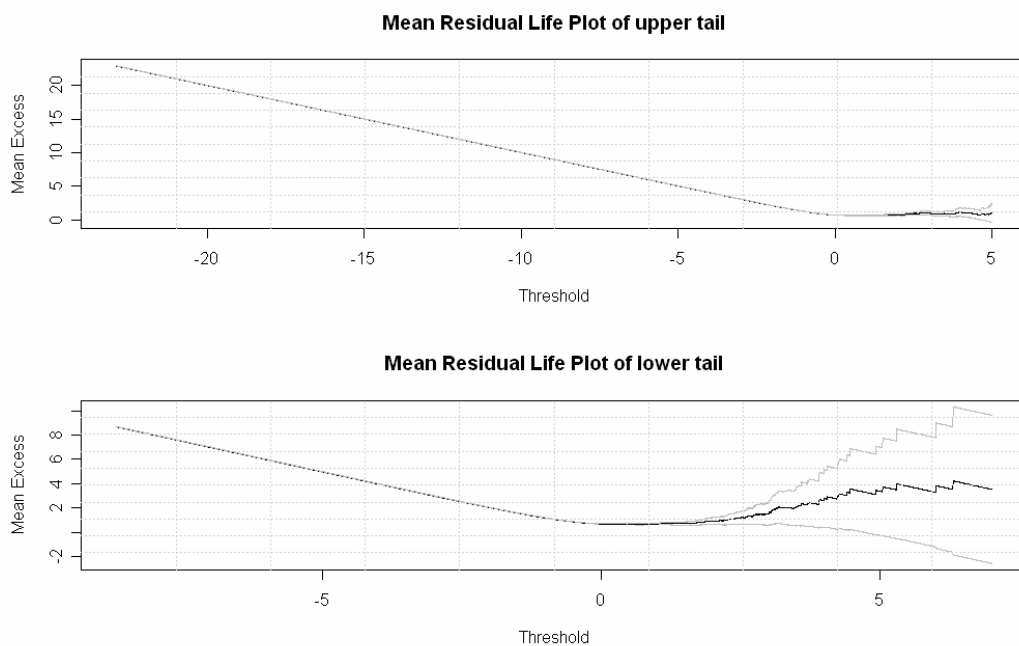
Αναφορικά με τη κάτω ουρά (ΣΧΗΜΑ 3-12) ο εκτιμητής Hill παρουσιάζει την ίδια συμπεριφορά όπως και με τις άνω διατεταγμένες παρατηρήσεις, δηλαδή κινείται γύρω από τιμή 0.3 στη περιοχή $80 < k < 150$. Ο Pickands φαίνεται και πάλι να έχει τη πιο σταθερή συμπεριφορά. Το διάγραμμα Pickands γίνεται γραμμικό στη περιοχή $400 < k < 500$, εκτιμώντας το ξ κοντά στο μηδέν. Το γράφημα του εκτιμητή DEdH παρουσιάζει αρκετά τμήματα στα οποία είναι γραμμικό, δυσχεραίνοντας στην επιλογή του κατάλληλου k .

Από τα παραπάνω καθίσταται προφανές ότι από τα διαγράμματα είναι πολλές φορές αδύνατο να επιλεγεί το βέλτιστο k για την εκτίμηση του ξ .

3.4 Ανάλυση Peaks Over Threshold

Από την προηγούμενη ανάλυση διαπιστώσαμε ότι οι άνω και οι κάτω τιμές του αρχικού δείγματος των λογαριθμικών αποδόσεων του δείκτη S&P 500 ακολουθούν τη κατανομή GEV. Σε αυτό το σημείο της ανάλυσης θα επικεντρωθούμε στην ανάλυση των υπερβαλλόντων τιμών πάνω από κάποιο συγκεκριμένο κατώφλι u . Για να επιλέξουμε το κατώφλι θα θέσουμε να κινηθούμε αρχικά μέσω διαγραμμάτων.

ΣΧΗΜΑ 3-13



Θα χρησιμοποιήσουμε το διάγραμμα της εμπειρικής μέσης υπερβάλλουσας συνάρτησης, το οποίο προκύπτει από την απεικόνιση των σημείων

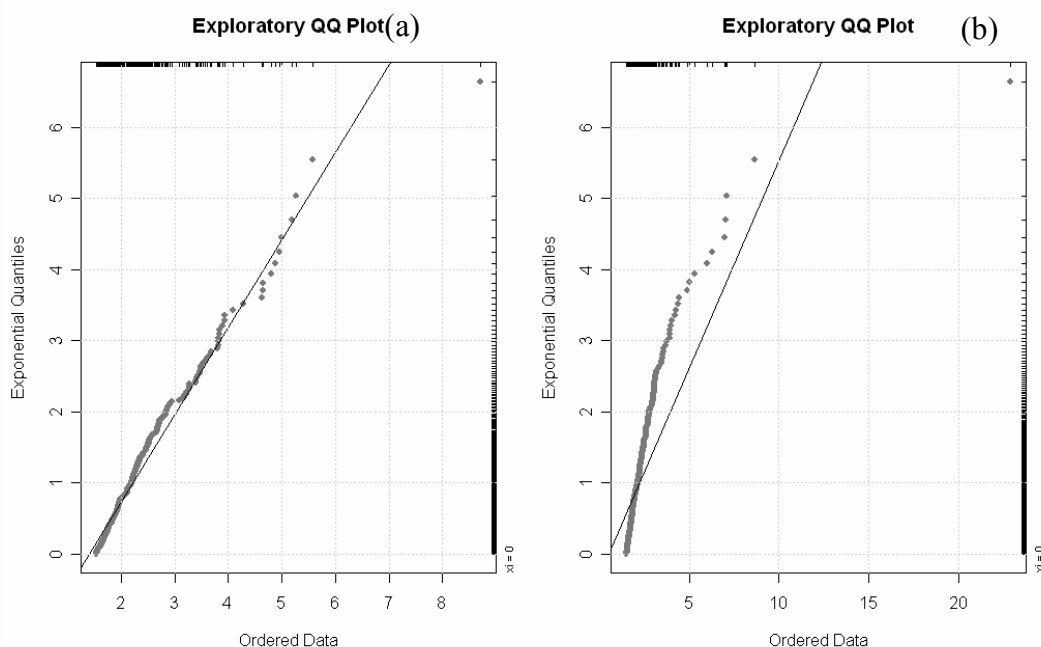
$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u), u > 0.$$

Η εμπειρική σ.κ. θα χρησιμοποιηθεί για τη γραφική προσέγγιση του βέλτιστου u ακολουθώντας την εξής λογική: θα επιλέξουμε εκείνο το $u > 0$ τέτοιο ώστε η $e_n(u)$ να είναι σχεδόν γραμμική για $x \geq u$. Στο πάνω διάγραμμα (ΣΧΗΜΑ 3-13), που αποδίδεται η πορεία της μέσης υπερβάλλουσας συνάρτησης για τη πάνω ουρά, βλέπουμε ότι το κατώφλι u βρίσκεται στη περιοχή $0.2 < u < 2$. Στο κάτω διάγραμμα βλέπουμε ότι το κατώφλι u βρίσκεται στη περιοχή $0 < u < 2$. Για να επιβεβαιώσουμε τα παραπάνω αποτελέσματα, ζητήσαμε από το πακέτο να υπολογίσει εκείνα τα u τόσο για τη πάνω όσο και για τη κάτω ουρά, πάνω από τα οποία βρί-

σκεται το 5% των παρατηρήσεων. Για τη πάνω ουρά μας έδωσε τη τιμή $u = 1.534901$ ενώ για τη κάτω ουρά μας έδωσε το όριο $u = 1.513367$. Και στις δυο περιπτώσεις το πλήθος των παρατηρήσεων που υπερβαίνουν το όριο να είναι $n = 384$, ήτοι το 5% των συνολικών μας παρατηρήσεων. Παρατηρούμε ότι τα παραπάνω αποτελέσματα, εμπίπτουν στη περιοχή όπου τα διαγράμματα της μέσης υπερβάλλουσας συνάρτησης είναι γραμμικά. Για τη προσαρμογή των δεδομένων των ουρών στη κατανομή GPD μπορούμε να χρησιμοποιήσουμε τα παραπάνω κατώφλια, θεωρώντας ότι αποτελούν μια ασφαλή επιλογή για την ανάλυση μας.

ΣΧΗΜΑ 3-14

Εκθετικά *QQ plot* για την πάνω (a) και την κάτω (b) ουρά



Σχεδιάσαμε τα εκθετικά *QQ plot* για τις δυο ουρές των συνόλων των παρατηρήσεων των λογαριθμικών αποδόσεων του δείκτη S&P 500 (ΣΧΗΜΑ 3-14). Παρατηρούμε ότι στη περίπτωση της κάτω ουράς εμφανίζεται μια ελαφριά κοίλη απόκλιση από την ευθεία προσαρμογής, και αποτελεί ένδειξη ότι η πάνω ουρά ακολουθεί κατανομή που ανήκει στο $MDA(H_\xi)$, $\xi > 0$. Στο ίδιο συμπέρασμα καταλήγουμε και από την εξέταση του *QQ plot* για τη κάτω ουρά.

Για να προχωρήσουμε στην εκτίμηση των παραμέτρων της GPD, θα χρησιμοποιήσουμε το πακέτο fExtremes. Η λογική της εκτίμησης είναι ίδια όπως και στη προσαρμογή των δεδομένων στη GEV. Χρησιμοποιείται ένας επαναληπτικός αλγόριθμος πάνω στις εξισώσεις που προκύπτουν από τη συνάρτηση λογάριθμο-πιθανοφάνειας

$$\ell((\xi, \beta); \mathbf{X}) = -n \ln \beta - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} X_i \right).$$

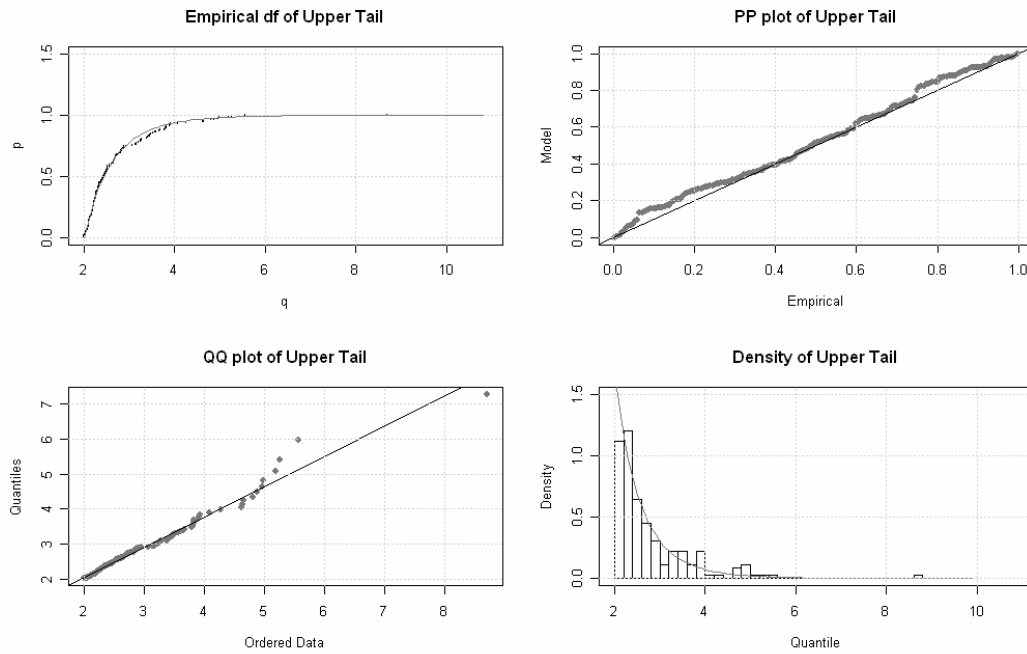
Τα αποτελέσματα που μας δίνει το πρόγραμμα, χρησιμοποιώντας την εκτίμηση της μέγιστης πιθανοφάνειας είναι τα ακόλουθα. Οι εκτιμώμενες παράμετροι για τη πάνω ουρά είναι $\hat{\xi} = 0.1291$ με τυπικό σφάλμα 0.05723 και $\hat{\beta} = 0.5984$ με τυπικό σφάλμα 0.04571. Για τις εκτιμώμενες παραμέτρους και χρησιμοποιώντας την υπόθεση της ασυμπτωτικής κανονικότητας, μπορούμε να πάρουμε 95% δ.ε.. Για το $\hat{\xi}$ το 95% δ.ε. είναι (0.0169819, 0.2413120) ενώ για το $\hat{\beta}$ είναι (0.5087756, 0.6879621). Οι εκτιμώμενες παράμετροι για τη κάτω ουρά είναι $\hat{\xi} = 0.2747$ με τυπικό σφάλμα 0.05964 και $\hat{\beta} = 0.5187$ με τυπικό σφάλμα 0.04003 με αντίστοιχα 95% δ.ε. (0.1577943, 0.3915663) και (0.4401875, 0.5971161). Παρατηρούμε ότι οι εκτιμηθείσες τιμές του ξ που προέκυψαν από τη προσαρμογή των παρατηρήσεων στη GPD διαφέρουν (αν και ελάχιστα) από αυτές που προέκυψαν κατά τη διαδικασία προσαρμογής στη GEV. Αυτό συμβαίνει γιατί είναι δύσκολο να οριστεί ένα κατώφλι μέσα στο σύνολο των παρατηρήσεων, που να ορίζει από πιο σημείο «ξεκινάει» η ουρά της κατανομής. Με το σχόλιο αυτό θέλουμε να τονίσουμε πόσο δύσκολο είναι να εξάγουμε εκτιμήσεις από πραγματικά δεδομένα. Δεδομένου ότι και στις δυο περιπτώσεις ισχύει ότι το $\hat{\xi} > -1/2$, μπορούμε να υποστηρίξουμε ότι η μέθοδος της μέγιστης πιθανοφάνειας μας έδωσε πολύ καλές εκτιμήσεις.

Για να ελέγξουμε τα αποτελέσματα της προσαρμογής της πάνω και της κάτω ουράς στη GPD, σχεδιάσαμε τα γραφήματα της εμπειρικής συνάρτησης προσαρμογής και το διάγραμμα πυκνότητας των εμπειρικών παρατηρήσεων, καθώς και τα *PP plot* και *QQ plot* των εμπειρικών τιμών έναντι αυτών που προκύπτουν από το εκτιμηθέν μοντέλο. Για το σχεδιασμό των διαγραμμάτων, θεωρήσαμε ως παράμετρο θέσης (τιμή εκκίνησης) για τη προσαρμοσμένη GPD τη πρώτη τιμή του αρχικού συνόλου παρατηρήσεων που ξεπερνάει το κατώφλι u .

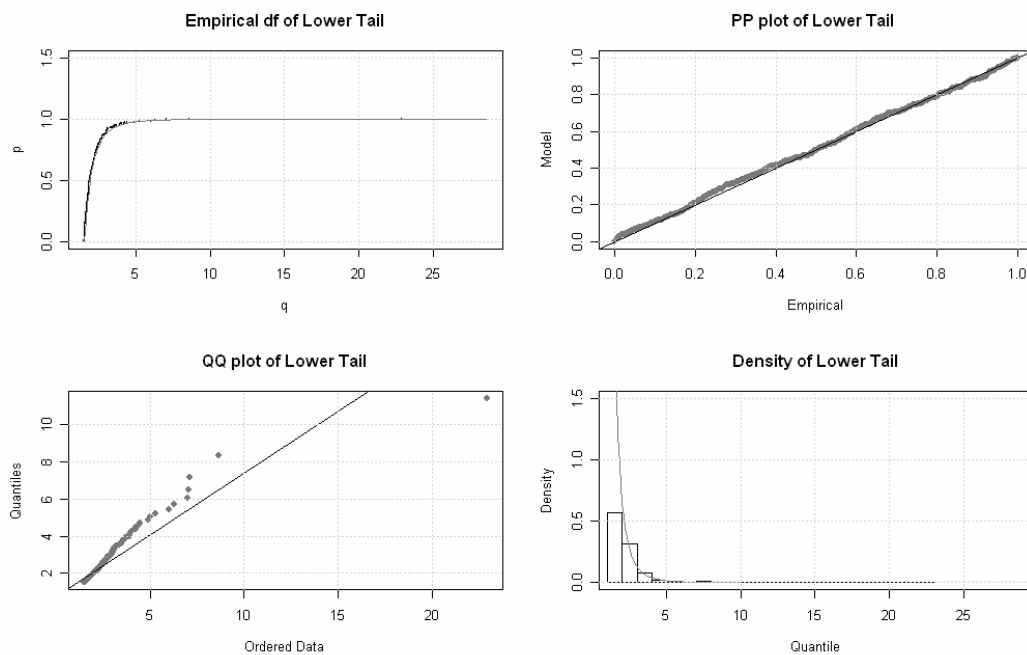
Εξετάζοντας τα γραφήματα για τη πάνω ουρά του αρχικού συνόλου παρατηρήσεων (ΣΧΗΜΑ 3-15), διαπιστώνουμε ικανοποιητική προσαρμογή του μοντέλου της GPD. Για τη κάτω ουρά (ΣΧΗΜΑ 3-16), βλέπουμε και πάλι την ίδια εικόνα που προέκυψε από την ανάλυση των μηνιαίων ελαχίστων (ΣΧΗΜΑ 3-7). Ενώ όλα τα διαγράμματα υποδεικνύουν καλή προσαρμογή, το *QQ plot*, παρουσιάζει κοίλη απόκλιση από την ευθεία προσαρμογής. Και πάλι βλέπουμε την ύπαρξη μίας έκτροπης παρατήρησης («Black Monday»), η οποία επηρεά-

ζει στην προσαρμογή του μοντέλου. Ανεξαρτήτως της εικόνας που προκύπτει από το QQ $plot$, το γενικό συμπέρασμα είναι ότι τα εκτιμηθέντα μοντέλα ανήκουν στο $MDA(H_\xi)$, $\xi > 0$.

ΣΧΗΜΑ 3-15



ΣΧΗΜΑ 3-16



3.4.1 Αποτίμηση Κινδύνου Χαρτοφυλακίου αναμενόμενο μέγεθος ζημίας

Στο Κεφάλαιο 2, κάναμε αναφορά στην αποτίμηση κινδύνου χαρτοφυλακίου (VaR) και στο αναμενόμενο μέγεθος ζημίας (ES). Βάσει των αποτελεσμάτων που προέκυψαν από την προσαρμογή των παρατηρήσεων στη GPD, μπορούμε να εκτιμήσουμε τις ποσότητες VaR και ES. Για τη πάνω ουρά προκύπτει ότι για επίπεδο σημαντικότητας 5%

$$\hat{VaR}_p = 1.533656 \text{ και } \hat{ES}_q = 2.220475,$$

ενώ για επίπεδο σημαντικότητας 1%

$$\hat{VaR}_p = 2.603655 \text{ και } \hat{ES}_q = 3.449095.$$

Για τη κάτω ουρά, για επίπεδο σημαντικότητας 5%, έχουμε

$$\hat{VaR}_p = -1.513636 \text{ και } \hat{ES}_q = -2.228857$$

και για επίπεδο σημαντικότητας 1%

$$\hat{VaR}_p = -2.563530 \text{ και } \hat{ES}_q = -3.676575.$$

Ερμηνεύοντας αυτά τα αποτελέσματα που προέκυψαν από τα ιστορικά δεδομένα των αποδόσεων του S&P 500, μπορούμε να πούμε ότι ένα *log-return* θα υπερβεί την τιμή 1.533 με πιθανότητα 5% ενώ, δεδομένου ότι υπερέβη αυτή την τιμή, αναμένεται γύρω στο 2.22. Αντίστοιχα συμπεράσματα μπορούμε να εξάγουμε και για τις υπόλοιπες παραπάνω περιπτώσεις. Πρέπει να αναφέρουμε ότι η εκτιμήσεις που προκύπτουν από το VaR αντιστοιχούν σε ημερήσιες αποδόσεις.

3.5 Συμπεράσματα

Από την ανάλυση που προηγήθηκε μπορούμε να υποστηρίξουμε ότι οι λογαριθμικές αποδόσεις του δείκτη S&P 500 ακολουθούν κατανομή με «βαριά» ουρά. Σε αυτό το συμπέρασμα συνηγορούν τόσο τα αποτελέσματα της μεθόδου *block maxima*, όσο και τα αποτελέσματα της μεθόδου *peaks over threshold*. Το ενδεχόμενο μιας πολύ μεγάλης μεταβολής στις αποδόσεις, είτε θετικής είτε αρνητικής, που να ξεπερνάει τις προηγούμενες καταγεγραμμένες τιμές, είναι αρκετά πιθανό. Βεβαίως, όπως ισχύει και στη περίπτωση των περισσότερων χρηματοοικονομικών δεικτών, μπορεί μελλοντικά η αγορά να εισέλθει σε μια νέα περίοδο και τα μοντέλα που εκτιμήθηκαν με τα μέχρι τώρα ιστορικά στοιχεία να χρειάζονται επανεξέταση.

ΓΑΛΕΠΣΤΗΜΟ ΠΕΡΑΙΑ

* Οι ιστορικές τιμές του δείκτη αντλήθηκαν από την ηλεκτρονική διεύθυνση <http://finance.yahoo.com/q/hp?s=%5EGSPC>

ΠΑΡΑΡΤΗΜΑ

Π1 Εντολές R

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

ТАНЕЦЫ И МУЗЫКА

Π1 Εντολές R

Οι υπολογισμοί και τα διαγράμματα που παρουσιάστηκαν στη παρούσα εργασία έγιναν με τη χρήση του στατιστικού πακέτου R και του επιπρόσθετου πακέτου (add-on) fExtremes (Version 260.72). Θα παραθέσουμε τις βασικές εντολές που χρησιμοποιήθηκαν από το πακέτο fExtremes, καθώς και κάποιες εντολές που σχεδιάστηκαν για αυτή την εργασία. Περισσότερες πληροφορίες αναφορικά με τις εντολές του πακέτου fExtremes μπορούν να βρεθούν στις διευθύνσεις

<http://www.rmetrics.org>

<http://cran.r-project.org/src/contrib/Descriptions/fExtremes.html>

Εντολές πακέτου fExtremes

blockMaxima. Εξάγει τα μηνιαία ή ετήσια μέγιστα μιας σειράς παρατηρήσεων. Απαραίτητη προϋπόθεση για τον προσδιορισμό των μεγίστων είναι το σύνολο των παρατηρήσεων να έχει ορισθεί ως **time.series**. Για την εξαγωγή των μηνιαίων ελαχίστων πρέπει να αντιστραφεί το πρόσημο της σειράς.

findThreshold. Υπολογίζει το κατώφλι u σε μια σειρά παρατηρήσεων, έτσι ώστε ένας ορισμένος αριθμός παρατηρήσεων να βρίσκεται πάνω από αυτό.

qqparetoPlot. Σχεδιάζει ένα *QQ plot* για παρατηρήσεις πάνω από κάποιο κατώφλι u έναντι της GPD για διάφορες τιμές του ξ . Αν το ξ της GPD οριστεί ίσο με 0, τότε το *QQ Plot* σχεδιάζεται έναντι της Εκθετικής κατανομής.

mePlot. Σχεδιάζει το διάγραμμα της μέσης υπολοιπόμενης ζωής μιας σειράς παρατηρήσεων πάνω από συνεχώς αυξανόμενα κατώφλια u .

mrlPlot. Σχεδιάζει το διάγραμμα της μέσης υπολοιπόμενης ζωής για την πάνω ουρά μιας κατανομής. Για την εξέταση της κάτω ουράς, αντιστρέφουμε το πρόσημο της σειράς των παρατηρήσεων.

dgev. Υπολογίζει τη πυκνότητα της GEV για διάφορες τιμές των (ξ, μ, ψ) .

pgenv. Υπολογίζει τη σ.π. της GEV για διάφορες τιμές των (ξ, μ, ψ) .

qgev. Υπολογίζει τα ποσοστιαία σημεία της GEV για διάφορες τιμές των (ξ, μ, ψ) .

rgenv. Προσομοιώνει τ.μ. από την GEV για διάφορες τιμές των (ξ, μ, ψ) .

gevFit. Προσαρμόζει τις μέγιστες (ελάχιστες) τιμές που προκύπτουν από τη μέθοδο *block maxima* στη GEV και δίνει εκτιμήσεις για τις παραμέτρους (ξ, μ, ψ) . Τα *blocks* ορίζονται είτε αριθμητικά (n αριθμός *blocks* σε όλη τη σειρά παρατηρήσεων) είτε σε μηνιαία ή ετήσια βάση αν οι παρατηρήσεις έχουν ορισθεί ως **time.series**.

gumbelFit. Προσαρμόζει τις μέγιστες (ελάχιστες) τιμές που προκύπτουν από τη μέθοδο *block maxima* στη Gumbel και δίνει εκτιμήσεις για τις παραμέτρους (μ, ψ) . Τα *blocks* ορίζονται είτε αριθμητικά (n αριθμός *blocks* σε όλη τη σειρά παρατηρήσεων) είτε σε μηνιαία ή ετήσια βάση αν οι παρατηρήσεις έχουν ορισθεί ως **time.series**.

summary. Παράγει διαγνωστικά διαγράμματα για τα μοντέλα που εκτιμήθηκαν από τη προσαρμογή στη GEV ή στη Gumbel.

gevrlevelPlot. Υπολογίζει και σχεδιάζει τη στάθμη απόδοσης (z_p) του k -block (το k ορίζεται από το χρήστη) καθώς και το 95% δ.ε. βάσει το εκτιμηθέντος μοντέλου που προέκυψε από τη προσαρμογή των παρατηρήσεων στη GEV, μέσω της μεθόδου της κατατομής πιθανοφάνειας.

dgpd. Υπολογίζει τη πυκνότητα της GPD για διάφορες τιμές των (ξ, β) .

pgpd. Υπολογίζει τη σ.π. της GPD για διάφορες τιμές των (ξ, β) .

qgpd. Υπολογίζει τα ποσοστιαία σημεία της GPD για διάφορες τιμές των (ξ, β) .

rgpd. Προσομοιώνει τ.μ. από την GPD για διάφορες τιμές των (ξ, β) .

gpdFit. Προσαρμόζει τη πάνω (κάτω) ουρά της κατανομής μιας σειράς παρατηρήσεων στη GPD και δίνει εκτιμήσεις για τις παραμέτρους (ξ, β) . Το κατώφλι u ορίζεται από το χρήστη.

gpdRiskMeasures. Εκτιμά τη VaR και το ES βάσει του εκτιμηθέντος μοντέλου.

Εντολές εργασίας

qqgumbelPlot.

Σχεδιάζει το *QQ plot* μιας σειράς παρατηρήσεων έναντι της Gumbel και δημιουργεί έναν πίνακα αποτελεσμάτων.

Ορισμός συνάρτησης qqgumbelPlot(x)

Παράμετροι

x διάνυσμα παρατηρήσεων

Κώδικας συνάρτησης

```
function (x, doplot = TRUE, labels = TRUE, ...)
{
  x = as.vector(x)
  n=length(x)
  y = rexp(n,1)
  y1=exp(-y)
  y2= -log(-log(ppoints(y1)))
  if (doplot) {
    if (labels) {
      xlab = "Ordered Data"
      ylab = "Quantiles"
      main = "QQ Plot against Gumbel Distribution"
    }
    else {
      xlab = ""
      ylab = ""
      main = ""
    }
  }
  z = sort(x)
  plot(z, y2, pch = 19, xlab = xlab, ylab = ylab, main = main, ...)
  rug(z, ticksize = 0.01, side = 3)
  rug(y2, ticksize = 0.01, side = 4)
  abline(lsfite(z, y2))
}
result = data.frame(x = sort(x), y2)
if (doplot)
  return(invisible(result))
else return(result)
}
```

qqgenPlot.

Σχεδιάζει το *QQ plot* μιας σειράς παρατηρήσεων έναντι της GEV για διάφορες τιμές της παραμέτρου ξ και δημιουργεί έναν πίνακα αποτελεσμάτων.

Ορισμός συνάρτησης qqgenPlot(x, xi)

Παράμετροι

x διάνυσμα παρατηρήσεων

xi τιμή της παραμέτρου ξ . Προεπιλεγμένη τιμή (default) είναι το 0.

Κώδικας συνάρτησης

```
function (x, xi = 0, doplot = TRUE,
  labels = TRUE, ...)
{
  x = as.vector(x)
  if (is.data.frame(x))
    x = x[, 1]
  xi=xi
  if (xi == 0) {y = -log(-log(ppoints(x)))}
  if (xi != 0) {y = (-1/xi)*(1-(-log(ppoints(x)))^(-xi))}
  if (doplot) {
    if (labels) {
      xlab = "Ordered Data"
      ylab = "Quantiles"
      ylab = paste("(xi=", xi, ") ", ylab, sep = "")
      main = "QQ Plot against GEV"
    }
    else {
      xlab = ""
      ylab = ""
      main = ""
    }
  }
  z = sort(x)
  plot(z, y, pch = 19, col = "black", xlab = xlab,
    ylab = ylab, main = main, ...)
  rug(z, ticksize = 0.01, side = 3)
  rug(y, ticksize = 0.01, side = 4)
  abline(lsfilt(z, y))
}
result = data.frame(x = sort(x), y)
if (doplot)
  return(invisible(result))
else return(result)
}
```

returnLevelPlot. Σχεδιάζει το διάγραμμα στάθμης απόδοσης της GEV για διάφορες τιμές των παραμέτρων (ξ, μ, ψ).

Ορισμός συνάρτησης returnLevelPlot(xi, mu, beta,n)

Παράμετροι

xi τιμή της παραμέτρου ξ .
mu τιμή της παραμέτρου μ .
beta τιμή της παραμέτρου ψ .
n αριθμός παρατηρήσεων.

Κώδικας συνάρτησης

```
function (xi, mu, beta, n, doplot = TRUE, labels = TRUE, ...)  
{  
  xi=xi  
  mu=mu  
  beta=beta  
  n=n  
  rate=1  
  y=rexp(n,rate)  
  y1= -log(1-ppoints(y))  
  if (xi == 0) {zu=mu-beta*log(y1)}  
  if (xi!= 0) {zu=mu-((beta/xi)*(1-(y1^(-xi))))}  
  if (doplot) {  
    if (labels) {  
      xlab = "-Log(-Log(1-p))"  
      ylab = "Quantiles"  
      main = paste("Return Level Plot xi = ", xi, ", m = ", mu, ", sigma = ",beta, sep = "")  
    }  
    else {  
      xlab = ""  
      ylab = ""  
      main = ""  
    }  
    z=sort(zu)  
    y2=-log(y1)  
    plot(y2, zu, type="l", col = "black", log="x", xlab = xlab,  
         ylab = ylab, main = main, ...)  
  }  
  result = data.frame(y1,sort(zu))  
  if (doplot)  
    return(invisible(result))  
  else return(result)  
}
```

pickandsPlot.

Σχεδιάζει το διάγραμμα του εκτιμητή Pickands για το ξ , συναρτήσει του k (κάτω οριζόντιος άξονας) και του κατωφλίου u (πάνω οριζόντιος άξονας) για ένα σύνολο παρατηρήσεων και δίνει σε μορφή πίνακα την εκτίμηση της παραμέτρου ξ για διάφορες τιμές του k .

Ορισμός συνάρτησης pickandsPlot (x)

Παράμετροι

x σύνολο παρατηρήσεων (ορισμένο ως διάνυσμα).

Κώδικας συνάρτησης

```
function (x, start = 1, doplot = TRUE, labels=TRUE,...)
{
  data = x
  ordered1 = rev(sort(abs(x[x > 0])))
  n = length(ordered1)
  k=1:n
  n1 = n%%4
  k1 = 1:(n%%4)
  pickands = log((ordered1[k1] - ordered1[2 * k1])/(ordered1[2 *
    k1] - ordered1[4 * k1]))/log(2)
  x = trunc(seq(from = min(n1, length(data)), to = start))
  if (doplot) {
    plot(k1, pickands, type = "l", xlab = xlab,
      ylab = ylab, axes = FALSE, ...)
    grid()
    pos = floor(seq(1, length(x), length = 10))
    axis(1, at = x[pos], lab = paste(x[pos]), tick = TRUE)
    axis(2)
    threshold = signif(findThreshold(data, x), 3)
    axis(3, at = x[pos], lab = paste(format(threshold[pos])),
      tick = TRUE)
    box()
  }
  if (labels) {
    title(xlab = "Order Statistics", ylab = "xi")
    mtext("Pickands Estimator", side = 3, line = 3)
  }
}
else {
  main = xlab = ylab = ""
}
res <- cbind(k1, pickandsest = pickands)
return(res)
}
```

Σημείωση: για τη σύνταξη της εντολής βασιστήκαμε στην εντολή **shaparmPickands** του πακέτου **fExtremes**.

hillPlot.

Σχεδιάζει το διάγραμμα του εκτιμητή Hill για το ξ , συναρτήσει του k (κάτω οριζόντιος άξονας) και του κατωφλίου u (πάνω οριζόντιος άξονας) για ένα σύνολο παρατηρήσεων. Επιπλέον δίνει σε μορφή πίνακα την εκτίμηση της παραμέτρου ξ για διάφορες τιμές του k .

Ορισμός συνάρτησης hillPlot(x, ni)

Παράμετροι

x

σύνολο παρατηρήσεων (ορισμένο ως διάνυσμα).

ni

διαιρεί το πλήθος των διατεταγμένων παρατηρήσεων, περιορίζοντας το διάγραμμα του εκτιμητή στις πρώτες n / ni παρατηρήσεις. Προεπιλεγμένη τιμή είναι το 1.

Κώδικας συνάρτησης

```
function(x, ni = 1, start = 1, doplot=TRUE, labels=TRUE,...)
{
  data = x
  ordered = rev(sort(data))
  ordered = ordered[ordered > 0]
  n = length(ordered)
  k = 1:n
  loggs = log(ordered)
  avesumlog = cumsum(loggs)/(1:n)
  xihat = c(NA, (avesumlog - loggs)[2:n])
  n1 = n%//%ni
  k1 = 1:n1
  loggs1 = log(ordered)[k1]
  avesumlog1 = cumsum(loggs1)/(1:n1)
  xihat1 = c(NA, (avesumlog1 - loggs1)[2:n1])
  x = trunc(seq(from = min(n1, length(data)), to = start))
  if (doplot) {
    plot(k1, xihat1, type = "l", xlab = "",
         ylab = "", axes = FALSE, ...)
    pos = floor(seq(1, length(k1), length = 10))
    axis(1, at = x[pos], lab = paste(x[pos]), tick = TRUE)
    axis(2)
    threshold = signif(findThreshold(data, x), 3)
    axis(3, at = x[pos], lab = paste(format(threshold[pos])),
         tick = TRUE)
    box()
    grid()
    if (labels) {
      title(xlab = "Order Statistics", ylab = "xi")
      mtext("Hill Estimator", side = 3, line = 3)
    }
  }
  else {
    main = xlab = ylab = ""
  }
  res <- cbind(k = 1:n, hill = xihat)
  return(res)
}
```

Σημείωση: για τη σύνταξη της συνάρτησης βασιστήκαμε στην συνάρτηση **shaparmHill** του πακέτου **fExtremes**.

DEdHPlot.

Σχεδιάζει το διάγραμμα του εκτιμητή DEdH για το ζ , συναρτήσσει του k (κάτω οριζόντιος άξονας) και του κατωφλίου u (πάνω οριζόντιος άξονας) για ένα σύνολο παρατηρήσεων. Επιπλέον δίνει σε μορφή πίνακα την εκτίμηση της παραμέτρου ζ για διάφορες τιμές του k .

Ορισμός συνάρτησης DEdHPlot(x, ni)

Παράμετροι

x σύνολο παρατηρήσεων (ορισμένο ως διάνυσμα).
ni διαιρεί το πλήθος των διατεταγμένων παρατηρήσεων, περιορίζοντας το διάγραμμα του εκτιμητή στις πρώτες n / ni παρατηρήσεις. Προεπιλεγμένη τιμή είναι το 1.

Κώδικας συνάρτησης

```
function(x, ni = 1, start = 1, doplot = TRUE, labels = TRUE, ...)  
{  
  data = x  
  ordered = rev(sort(abs(x[x > 0])))  
  n = length(ordered)  
  n1 = n%%ni  
  ns0 = 2  
  nm = n1 - 1  
  ns = ns0  
  nsp = ns + 1  
  bod = c(cumsum(log(ordered))[ns:nm]/(ns:nm) - log(ordered)[nsp:n1])  
  bid = c(cumsum((log(ordered))^2)[ns:nm]/(ns:nm) -  
    2 * cumsum(log(ordered))[ns:nm] * log(ordered)[nsp:n1]/(ns:nm) +  
    ((log(ordered))^2)[nsp:n1])  
  dehaan = c(NA, (1 + bod + (0.5/(bod^2/bid - 1)))[1:n1-1])  
  x = trunc(seq(from = min(n1, length(data)), to = start))  
  if(doplot) {  
    plot(1:n1-1, dehaan, type = "l", xlab = "", ylab = "", axes = FALSE, ...)  
    pos = floor(seq(1, length(x), length = 10))  
    axis(1, at = x[pos], lab = paste(x[pos]), tick = TRUE)  
    axis(2)  
    threshold = signif(findThreshold(data, x), 3)  
    axis(3, at = x[pos], lab = paste(format(threshold[pos])),  
      tick = TRUE)  
    box()  
    if(labels) {  
      title(xlab = "Order Statistics", ylab = option)  
      mtext("Threshold", side = 3, line = 3)  
    }  
  }  
  res <- cbind(1:n1, DEdH = dehaan)  
  return(res)  
}
```

Σημείωση: για τη σύνταξη της εντολής βασιστήκαμε στην εντολή **shaparmDEHaan** του πακέτου **fExtremes**.

BIBΛΙΟΓΡΑΦΙΑ

- Balkema, A.A. and Haan, L. de (1974). Residual life time at great age. *Annals of Probability*, **2**, pp. 792 – 804.
- Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.
- Coles, S., (2001). *An Introduction to Statistical Modelling of Extreme Value*, Springer – Verlag, London.
- Deheuvels, P., Haeusler, E., and Mason, D, M. (1988). Almost sure convergence of the Hill estimator, *Math. Proc. Cambridge Philos. Soc.*, **104**, pp. 371 – 381.
- Dekkers, A.L.M. and Haan, L. de (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics*, **17**, pp. 1795 – 1832.
- Dekkers, A.L.M., Einmahl, J.H.J. and Haan, L. de (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, **17**, pp. 1833 – 1855.
- Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot, *The Annals of Statistics*, **28**, pp. 254 – 274.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: for insurance and finance*, Springer – Verlag, Berlin.
- Feller, W. (1971), *An introduction to Probability Theory and Its Applications*, volume II, Wiley, New York
- Fisher, R.A. and Tippett, L.H.C. (1928). Limiting Forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, pp. 180 – 190
- Gnedenko, B., (1943) Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, *The Annals of Mathematics*, **44**, pp. 423 – 453
- Haan, L. de (1970), On regular variation and its application to the weak convergence of sample extremes, Thesis, University of Amsterdam / Mathematical Centre tract, **32**
- Haan, L. de (1994). Extreme Value Statistics. In: Galambos et al. (Eds.) *Extreme Value Theory and Applications*, Kluwer, Dordrecht. pp 93 – 112.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**, pp. 1163 – 1174
- Jenkinson, A.F. (1955). The Frequency distribution of the annual maximum (or minimum) values of meteorological events, *Quarterly Journal of the Royal Meteorological Society*, **81**, pp. 158 – 172
- Karamata, J. (1933). *Sur un mode de croissance régulière. Théorèmes fondamentaux*, *Bull. Soc. Math. France* **61**, pp. 55 – 62
- Mason, D.M. (1982). Laws of large numbers for sums of extreme values, *Annals of Probability*, **10**, pp. 756 – 764.

- McNeil, A.J. and Saladin, T., (1997). The peaks over thresholds method for estimating high quantiles of loss distributions, *Proceedings of 28th International ASTIN Colloquium*.
- McNeil, A.J. (1998). Calculating quantile risk measures for financial return series using extreme value theory, *ETH E-Collection*.
<http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=85>
- McNeil, A.J. (1999). Extreme Value Theory for Risk Managers, *Internal Modelling and CAD II* published by *RISK Books*, pp. 93 – 113
- Mises, R. Von (1936). La distribution de la plus grande de n valeurs, *Selected Papers II*, *American Mathematical Society*, pp. 271 – 294
- Pickands, J. III (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, pp. 119 – 131
- Resnick, S.I., (1987). *Extreme Values, Regular Variation and Point Processes*, Springer – Verlag, New York. pp. 7, 13
- Resnick, S.I., Stărică, C. (1995). Consistency of Hill's estimator for dependent data, *Journal of Applied Probability*, **32**, pp. 139 – 167
- Resnick, S.I., Stărică, C. (1996). Asymptotic behaviour of Hill's estimator for autoregressive data, *Communications in Statistics - Stochastic Models*, **13**, pp. 703 – 721.
- Rootzén, H., Leadbetter, M.R. and de Haan, L. (1992). *Tail and quantile estimators for strongly mixing stationary processes*, Report, Departments of Statistics, University of North Carolina.
- Smith, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases, *Biometrika*, **72**, pp. 67 – 90