

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ
ΚΙΝΔΥΝΟΥ

Η ΓΕΝΙΚΕΥΜΕΝΗ ΚΑΤΑΝΟΜΗ
ΑΚΡΑΙΩΝ ΤΙΜΩΝ ΚΑΙ ΕΦΑΡΜΟΓΕΣ
ΤΗΣ ΣΤΟΝ ΑΝΑΛΟΓΙΣΜΟ

Μαριάννα Δ. Τατσή

Διπλωματική Εργασία

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

Πειραιάς
Ιούνιος 2015

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

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής Μ.Κούτρας (Επιβλέπων)
- Αναπλ. Καθηγητής Χ.Χατζηκωνσταντινίδης
- Επικ. Καθηγητής Μ.Μπούτσινας

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

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND INSURANCE
SCIENCE

POSTGRADUATE PROGRAM IN ACTUARIAL
SCIENCE AND RISK MANAGEMENT

**THE GENERALIZED EXTREME VALUE
DISTRIBUTION AND ITS APPLICATIONS
IN ACTUARIAL SCIENCE**

By

Marianna D. Tatsi

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 in Science in Actuarial Science and Risk Management

Piraeus

Greece

November 2014

Ευχαριστίες

Θα ήθελα να ευχαριστήσω καταρχήν τον κο. Κούτρα για την υπομονή και την καθοδήγηση του καθόλο το χρονικό διάστημα της εκπόνησης της πτυχιακής εργασίας καθώς και την τριμελή επιτροπή για τον χρόνο που αφιέρωσε στην ανάγνωση της και τα πολύτιμα σχόλια της. Επίσης, ήθελα να ευχαριστήσω ιδιαίτερώς τον κο. Χατζηκωνσταντινίδη για τον ρεαλισμό του και την προσήλωσή του στη λεπτομέρεια καθώς και τον κο. Μπούτσινα για την επιμονή του στις αριθμητικές εφαρμογές. Επιπρόσθετα θα ήθελα να ευχαριστήσω τον κο. Τολίκα για τις πολύ ενδιαφέρουσες εργασίες που μας παρουσίασε, οι οποίες μου έδωσαν κάποιες ιδέες σχετικά με τις αριθμητικές εφαρμογές της πτυχιακής αλλά τον κο. Σεβρόγλου που πίστεψε σε εμένα και το έδειχνε με κάθε ευκαιρία. Όλα τα παραπάνω με βοήθησαν καθόλη τη χρονική περίοδο που παρέμεινα στο Πανεπιστήμιο Πειραιώς και είμαι ευγνώμων προς όλους. Τέλος θα ήθελα να ευχαριστήσω τον Θεωρή Καραμήτρου από το γραφείο του Νέου Ψυχικού και την κα. Τάνια Μονογιού, αναλογίστρια της Ευρωπαϊκής Πίστης για την παροχή των δεδομένων ζημιών.

Περίληψη

Στην παρούσα εργασία γίνεται μία σύντομη παρουσίαση της Θεωρίας Ακραίων Τιμών και της χρησιμότητάς της στις ασφαλίσσεις έναντι ακραίων γεγονότων. Παρουσιάζουμε μια εκτενή αναφορά στις μεθόδους εκτίμησης παραμέτρων των κατανομών ακροτάτων Gumbel, Frechet και Weibull συγκρίνοντας τις παραγόμενες εκτιμήτριες, μέσω αριθμητικών υπολογισμών. Παρουσιάζεται η Γενικευμένη Κατανομή Ακραίων Τιμών (GEV) και δείχνεται πως αυτή μπορεί να χρησιμοποιηθεί για την περιγραφή πραγματικών δεδομένων απωλειών. Επιπρόσθετα, υπολογίζουμε τα δειγματικά μέτρα κινδύνου, αξία σε κίνδυνο και αναμενόμενο έλλειμα, συγκρίνοντάς τα με τα αντίστοιχα της εκτιμημένης κατανομής. Τέλος, υπολογίζουμε τη σύνθετη συνάρτηση απωλειών (συνολικές απώλειες χαρτοφυλακίου) χρησιμοποιώντας τη μέθοδο Monte Carlo με στόχο την πρόβλεψη των μέτρων κινδύνου και την ανάλυση ευαισθησίας αυτών για διαφορετικά επίπεδα εμπιστοσύνης.

Abstract

In this thesis we offer a brief presentation of extreme value theory and discuss its usefulness in Insurance and extreme events. An extensive description to parameter estimation methods for extreme value distributions i.e. Gumbel, Frechet, Weibull is given along with numerical comparison of the efficiency of the estimators by the aid of a simulation study. Generalized Extreme Value distribution (GEV) is covered as well and an illustration is provided on how it can be adapted to real loss data. In addition we calculate sample risk measures Value-at-Risk and Expected Shortfall and compare them to the corresponding measures of the estimated distribution. Finally, we conducted an estimation of the aggregate loss function using a Monte Carlo simulation to forecast risk measures and analyze their sensitivity against different confidence levels.

Περιεχόμενα

Κατάλογος Πινάκων	5
Κατάλογος Σχημάτων	7
1 Εισαγωγή	9
2 Βασικά στοιχεία Θεωρίας Ακραίων Τιμών	13
2.1 Παραδείγματα	13
2.2 Βασική Θεωρία	14
2.3 Περιοχή έλξης κατανομών ακροτάτων	22
3 Εκτίμηση παραμέτρων για τις κατανομές ακραίων τιμών	28
3.1 Μέθοδος μέγιστης πιθανοφάνειας	28
3.2 Μέθοδος ροπών	29
3.3 Μέθοδος των σταθμισμένων ροπών πιθανότητας	30
3.4 Κριτήρια αποδοτικότητας μεθόδων	31
3.5 Εκτίμηση των παραμέτρων της κατανομής <i>Gumbel</i>	32
3.5.1 Εκτίμηση των παραμέτρων μ και β	35
3.5.2 Σύγκριση μεθόδων εκτίμησης	37
3.6 Εκτίμηση των παραμέτρων της κατανομής <i>Frechet</i>	41
3.6.1 Εκτίμηση των παραμέτρων s και α	43
3.6.2 Σύγκριση μεθόδων εκτίμησης	45
3.7 Εκτίμηση των παραμέτρων της κατανομής <i>Weibull</i>	49
3.7.1 Εκτίμηση των παραμέτρων σ και α	51
3.7.2 Σύγκριση μεθόδων εκτίμησης	55
4 Γενικευμένη κατανομή ακραίων τιμών	60
4.1 Ορισμός γενικευμένης κατανομής	60
4.2 Εκτίμηση Παραμέτρων	63
4.2.1 Μέθοδος μέγιστης πιθανοφάνειας	63
4.2.2 Μέθοδος των σταθμισμένων ροπών πιθανότητας	65

4.2.3	Μικτές Μέθοδοι	67
4.3	Δείκτης Ουράς	69
4.3.1	Μέθοδος εκτιμήτριας <i>Pickands</i>	71
4.3.2	Μέθοδος εκτιμήτριας <i>Hill</i>	73
4.4	Μέτρα Κινδύνου	76
4.4.1	Αξία σε κίνδυνο	77
4.4.2	Αναμενόμενο Έλλειμα	78
4.5	Στάθμη απόδοσης και Χρονική Περίοδος απόδοσης	79
4.6	Μέθοδος μέγιστων υποσυνόλων	80
4.7	Μέθοδος τιμών πάνω από κάποιο κατώφλι	82
5	Αριθμητική Εφαρμογή	85
5.1	Συχνότητα και Ύψος Απαιτήσεων	93
5.2	Κατανομή Συνολικών Απωλειών και Ανάλυση Ευαισθησίας	100
6	Σύνοψη	107
7	Παράρτημα	109
7.1	Σταθερά <i>Euler</i>	109
7.2	Συνάρτηση Γάμμα	109
7.3	Συνάρτηση Ζήτα	109
7.4	Συνάρτηση δι-Γάμμα	110
7.5	Συνεκτικό μέτρο κινδύνου	110
	Αναφορές	112

Κατάλογος Πινάκων

1	Ποσοστά ζημίας σεισμών στην Καλιφόρνια για τα ετη 1971-1994 (Πηγή: A.M. Best, Τμήμα ασφαλίσεων της Καλιφόρνια, Ινστιτούτο Ασφαλιστικών Πληροφοριών)	11
2	Συνοπτικός πίνακας κατανομών ακροτάτων.	24

3	Μέση τιμή, διακύμανση και ασυμμετρία της κατανομής Gumbel για διάφορες τιμές των παραμέτρων.	34
4	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου μέγιστης πιθανοφάνειας για την κατανομή Gumbel.	38
5	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου ροπών πιθανότητας για την κατανομή Gumbel.	38
6	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου σταθμισμένων ροπών πιθανότητας για την κατανομή Gumbel.	39
7	Μέση τιμή, διακύμανση και ασυμμετρία της κατανομής Fréchet για διάφορες τιμές των παραμέτρων.	43
8	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου μέγιστης πιθανοφάνειας για την κατανομή Fréchet.	45
9	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου σταθμισμένων ροπών πιθανότητας για την κατανομή Fréchet.	46
10	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου ροπών πιθανότητας για την κατανομή Fréchet.	46
11	Μέση τιμή, διακύμανση και ασυμμετρία της κατανομής Weibull για διάφορες τιμές των παραμέτρων.	51
12	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου μέγιστης πιθανοφάνειας για την κατανομή Weibull.	55
13	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου σταθμισμένων ροπών πιθανότητας για την κατανομή Weibull.	56
14	Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου ροπών για την κατανομή Weibull.	56
15	Πίνακας περιγραφικών στοιχείων δεδομένων συχνοτήτων	88
16	Πίνακας περιγραφικών στοιχείων δεδομένων απαιτήσεων	92
17	Ύψη ζημιών που αποκλίνουν από το διάγραμμα Q-Q plot.	92
18	Var και ES δεδομένων συχνοτήτων.	95
19	Var και ES δεδομένων απαιτήσεων.	96
20	Var και ES προσωμοιωμένων δεδομένων απαιτήσεων.	98
21	Πίνακας περιγραφικών στοιχείων αθροιστικής κατανομής απωλειών	103

22	Var και ES της αθροιστικής κατανομής απωλειών.	103
----	--	-----

Κατάλογος Σχημάτων

1	Συναρτήσεις πυκνότητας πιθανότητας Gumbel, Fréchet και Weibull για $\alpha = 1$	19
2	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Fréchet για $\alpha = (0.5, 1, 1.7, 2, 2.5)$	19
3	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull για $\alpha = [0.8, 1, 2, 3.5, 5]$	20
4	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Gumbel για σταθερή τιμή της παραμέτρου μ και ενδεικτικές τιμες της παραμέτρου β	33
5	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Gumbel για σταθερή τιμή της παραμέτρου β και ενδεικτικές τιμες της παραμέτρου μ	33
6	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Fréchet για σταθερή τιμή της παραμέτρου s και ενδεικτικές τιμές της παραμέτρου α	42
7	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Fréchet για σταθερή τιμή της παραμέτρου α και ενδεικτικές τιμές της παραμέτρου s	42
8	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull για σταθερή τιμή της παραμέτρου σ και για ενδεικτικές τιμές της παραμέτρου α	50
9	Συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull για σταθερή τιμή της παραμέτρου α και για ενδεικτικές τιμές της παραμέτρου σ	50
10	Γενικευμένη κατανομή ακραίων τιμών για διάφορες τιμές του ξ	61
11	Μέθοδος μέγιστων μπλοκ.	81
12	Αναμενόμενες και μη-αναμενόμενες ζημιές	86

13	Αθροιστική κατανομή απωλειών	87
14	Boxplot των δεδομένων απαιτήσεων	88
15	Γραφική απεικόνιση των δεδομένων συχνοτήτων.	89
16	Ιστόγραμμα των δεδομένων συχνοτήτων.	89
17	QQ-plot των δεδομένων συχνοτήτων	90
18	Γραφική απεικόνιση των δεδομένων απαιτήσεων.	90
19	Ιστόγραμμα των δεδομένων απαιτήσεων.	91
20	QQ-plot των δεδομένων απαιτήσεων.	91
21	QQ-plot των δεδομένων απαιτήσεων προσαρμόζοντας την κανονική κατανομή.	97
22	QQ-plot των δεδομένων απαιτήσεων προσαρμόζοντας τη γενικευμένη Pareto.	97
23	Σύγκριση εμπειρικής και θεωρητικής κατανομής.	98
24	Ιστόγραμμα πιθανότητας της αθροιστικής κατανομής απωλειών. .	104
25	Ιστόγραμμα συχνοτήτων της αθροιστικής κατανομής απωλειών. .	104
26	Σχηματική απεικόνιση των μέτρων κινδύνου μεταβάλλοντας το επίπεδο εμπιστοσύνης.	105

1 Εισαγωγή

Ο κίνδυνος ακραίου γεγονότος είναι παρών σε όλους τους τομείς της διαχείρισης κινδύνων, στον κίνδυνο αγοράς, τον πιστωτικό, τον ασφαλιστικό αλλά και στις καθημερινές διαδικασίες με τη μορφή του λειτουργικού κινδύνου. Μια από τις μεγαλύτερες προκλήσεις ενός διαχειριστή κινδύνου είναι η εφαρμογή κατάλληλων εργαλείων διαχείρισης που λαμβάνουν υπόψη τη μοντελοποίηση σπάνιων αλλά με μεγάλες επιπτώσεις γεγονότων, επιτρέποντας έτσι τη μέτρηση των συνέπειών τους. Η θεωρία ακραίων τιμών κατέχει πολύ σημαντικό ρόλο στην μελέτη και την ανάπτυξη των εργαλείων αυτών.

Η κλασική μαθηματική προσέγγιση στη μοντελοποίηση κινδύνων γίνεται μέσω της θεωρίας πιθανοτήτων. Οι κίνδυνοι είναι τυχαίες μεταβλητές που αναπαριστούν μελλοντικές καταστάσεις οι οποίες αντιστοιχούν σε κέρδη ή ζημιές. Αυτοί οι κίνδυνοι μπορούν να μελετηθούν ατομικά ή σαν μέρος στοχαστικών διαδικασιών όπου οι παρόντες εξαρτώνται από τους παρελθοντικούς. Οι πιθανές τιμές που λαμβάνουν οι τυχαίες μεταβλητές, που αναπαριστούν τους κινδύνους, έχουν μια κατανομή πιθανότητας η οποία είναι εν μέρη γνωστή, παρόλο που οι παρελθοντικοί κίνδυνοι έχουν καταγραφεί με ακρίβεια. Ακραίο γεγονός έχουμε όταν η τυχαία μεταβλητή λαμβάνει τιμές από την ουρά της κατανομής πιθανότητας.

Η αρχή της θεωρίας ακραίων τιμών τοποθετείται γύρω στο 1925, όταν ο Άγγλος στατιστικός Leonard H.C. Tippett προσελήφθη στο ινστιτούτο Shirley του Manchester. Εκεί, με στόχο να κάνει το βαμβακερό νήμα ισχυρότερο, εφάρμοσε στατιστικές μεθόδους σε διάφορα προβλήματα υφαντικής που χρειάστηκε να μελετήσει όπως, στην εύρεση της σχέσης μεταξύ του μήκους ενός δοκιμαστικού νήματος και της αντοχής του, στην εκτίμηση της διακύμανσης του πάχους κατά μήκος ενός νήματος καθώς και του ρυθμού θραύσης του κατά την ύφανση (βλ [52]).

Στις μελέτες του, συνειδητοποίησε ότι η δύναμη του βαμβακερού νήματος ελέγχεται από την αντοχή των πιο αδύναμων ινών από τις οποίες αποτελείται. Πιο συγκεκριμένα, αν υποθέσουμε ότι μια ακολουθία τυχαίων μεταβλητών X_1, X_2, \dots, X_n περιγράφει το πάχος των ινών ενός νήματος πλήθους n , τότε

η συνολική αντοχή του εξαρτάται από τη ποσότητα $\min(X_1, X_2, \dots, X_n)$.

Το 1928, με τη βοήθεια του Ronald A. Fisher διατύπωσε το θεώρημα Fisher-Tippett αποδεικνύοντας ότι μετά από κατάλληλη κανονικοποίηση το μέγιστο ενός δείγματος ανεξάρτητων και ισόνομων τυχαίων μεταβλητών μπορεί να συγκλίνει μόνο σε μία από τρεις πιθανές κατανομές (βλ [19]). Το θεώρημα δόθηκε με πιο λεπτομερή απόδειξη το 1943 από τον Boris V. Gnedenko (βλ [21]) ενώ αργότερα, το 1958, ο Emil Gumbel κωδικοποίησε τη θεωρία ακραίων τιμών στο βιβλίο του *Statistics of Extremes* (βλ [27]).

Η θεωρία ακραίων τιμών είναι ένας πολύ σημαντικός κλάδος της στατιστικής και έχει πληθώρα εφαρμογών σε πολλούς τομείς. Ένα από τα πιο σημαντικά ζητήματα συζήτησης και έρευνας στον τομέα των ασφαλειών για παράδειγμα, είναι οι απώλειες λόγω του κινδύνου που προέρχεται από ασφαλίσεις περιουσίας-απώλειας (property-casualty). Συνήθως ο κίνδυνος αυτός αντισταθμίζεται από την αντασφάλιση χρησιμοποιώντας τα catastrophe futures, ειδικά διαμορφωμένα μελλοντικά συμβόλαια υποχρεωτικής εκπλήρωσης της αναγραφόμενης υποχρεώσεως. Πιο συγκεκριμένα, αναφερόμαστε σε υπερβάλλουσα ασφάλιση (excess insurance) όταν οι απώλειες που προκύπτουν από ένα συγκεκριμένο συμβάν υπερβαίνουν ένα προκαθορισμένο ποσό είτε σε επίπεδο ενός μεμονωμένου κινδύνου (risk excess) είτε σε συνολικό επίπεδο απωλειών (catastrophe).

Για να δικαιολογήσουμε τον παραπάνω ισχυρισμό αρκεί να δούμε τον Πίνακα 1. Ο πίνακας, δείχνει τα ποσοστά ζημίας (loss ratio) μέσα σε ένα έτος για ασφάλιση κατά σεισμού στην Καλιφόρνια από το 1971 έως το 1994 (βλ [33]). Να σημειώσουμε ότι ο όρος loss ratio αναφέρεται στο λόγο των συνολικών απαιτήσεων που πληρώνονται από την ασφαλιστική εταιρεία, συμπεριλαμβανομένων των εξόδων προσαρμογής, προς τα ασφάλιστρα που εισέπραξε κατά τη διάρκεια ενός έτους χρήσης για τα ασφαλιστήρια συμβόλαια που διαχειρίζεται και εκφράζεται σε ποσοστό. Ας υποθέσουμε τώρα πως θέλουμε να κάνουμε μια εκτίμηση για το ύψος της τιμής του ποσοστού ζημίας (loss ratio) για το έτος 1994. Εφαρμόζοντας γραμμική παλινδρόμηση στα δεδομένα η πρόβλεψή μας για την τιμή του loss ratio προκύπτει ίση με 31.7, ενώ σε επίπεδο εμπιστοσύνης 95% το αντίστοιχο διάστημα εμπιστοσύνης που προκύπτει είναι το $[0.0, 83.0]$.

Έτος	% LossRatio	Έτος	% LossRatio
1971	17.4	1983	2.9
1972	0.0	1984	5.0
1973	0.6	1985	1.3
1974	3.4	1986	9.3
1975	0.0	1987	22.8
1976	0.0	1988	11.5
1977	0.7	1989	129.8
1978	1.5	1990	47.0
1979	2.2	1991	17.2
1980	9.2	1992	12.8
1981	0.9	1993	3.2
1982	0.0	1994	;

Πίνακας 1. Ποσοστά ζημίας σεισμών στην Καλιφόρνια για τα έτη 1971-1994 (Πηγή: A.M. Best, Τμήμα ασφαλίσεων της Καλιφόρνια, Ινστιτούτο Ασφαλιστικών Πληροφοριών)

Ωστόσο η πραγματική τιμή που κατέγραψε το έτος αυτό ήταν 2272.7. Οι απώλειες που προέκυψαν από περιουσία που είχε ασφαλισθεί, εκτιμήθηκαν γύρω στα 12.5 δισεκατομμύρια δολάρια ενώ οι συνολικές απώλειες έφτασαν τα 30 δισεκατομμύρια. Από τα παραπάνω προκύπτει ότι η εφαρμογή κλασικών μεθόδων στατιστικής στη μοντελοποίηση ακραίων τιμών, μπορεί να οδηγήσει σε εσφαλμένα αποτελέσματα ακόμη και στον χαρακτηρισμό τους ως outliers με αποτέλεσμα, την αφαίρεσή τους από τη στατιστική ανάλυση. Η ασφαλιστική βιομηχανία πρέπει να επαναξιολογήσει τον κίνδυνο της εμφάνισης ακραίων τιμών και να χρησιμοποιήσει τη θεωρία ακραίων τιμών για τη μοντελοποίηση τέτοιων κινδύνων.

Είναι λοιπόν σημαντικό να τεθεί το ακόλουθο ερώτημα: πόσο πιθανό είναι η μεγαλύτερη παρατήρηση ενός δείγματος υπό μελέτη να είναι μικρότερη ή ίση από μία συγκεκριμένη τιμή;

Στη μελέτη χρηματοοικονομικών και ασφαλιστικών δεδομένων, όπως για παράδειγμα χρηματοοικονομικών αποδόσεων ή ασφαλιστικών απαιτήσεων, κρίνεται αναγκαία η χρήση κατανομών οι οποίες προσεγγίζουν αποτελεσματικά δεδομένα που εμφανίζουν ακραίες τιμές. Η πιο γνωστή από τις κατανομές αυτές είναι η γενικευμένη κατανομή ακραίων τιμών.

Στα κεφάλαια που ακολουθούν δίνεται μια σύντομη περιγραφή βασικών θεωρημάτων και ορισμών της θεωρίας ακραίων τιμών. Έπειτα παρουσιάζονται εκτενώς τα χαρακτηριστικά και οι μέθοδοι εκτίμησης παραμέτρων της γενικευμένης κατανομής ακραίων τιμών και των ειδικών περιπτώσεων της Gumbel, Fréchet και Weibull, για τις οποίες γίνεται σύγκριση της αποτελεσματικότητας των μεθόδων εκτίμησης. Στα δύο τελευταία κεφάλαια γίνεται εφαρμογή της θεωρίας σε πραγματικά δεδομένα ζημιών και μια ολοκληρωμένη μελέτη σχετικά με τα μέτρα κινδύνου ακολουθεί. Τέλος, μέσω υπολογιστικής προσομοίωσης δημιουργούμε την αθροιστική κατανομή απωλειών και εφαρμόζουμε ανάλυση ευαισθησίας έναντι του επιπέδου εμπιστοσύνης δίνοντας έμφαση στη σύγκριση των μέτρων κινδύνου.

2 Βασικά στοιχεία Θεωρίας Ακραίων Τιμών

2.1 Παραδείγματα

Στη μοντελοποίηση ακραίων τιμών έχουν προταθεί διάφορες προσεγγίσεις για κάθε περίπτωση. Παρακάτω παραθέτουμε δύο παραδείγματα ώστε να γίνουν περισσότερο κατανοητοί οι ορισμοί που θα εισαγάγουμε στη συνέχεια.

Το 1953 ο συνδυασμός της υψηλής ανοιξιάτικης παλίρροιας με μια σφοδρή καταιγίδα προκάλεσε τεράστιες απώλειες κυρίως στις ακτές της Σκωτίας και της Ολλανδίας (βλ [29]). Για να αποφευχθούν παρόμοια γεγονότα στο μέλλον η Ολλανδική κυβέρνηση σχημάτισε μια επιτροπή, υπεύθυνη για τον υπολογισμό του ύψους των φραγμάτων που θα έπρεπε να χρησιμοποιηθούν ώστε το ύψος της στάθμης του νερού να μη τα ξεπερνά.

Ουσιαστικά το πρόβλημα ήταν ο προσδιορισμός του ύψους των φραγμάτων έτσι ώστε η πιθανότητα ύπαρξης παλίρροιας σε ένα δεδομένο έτος να ισούται με p όπου $p \ll 1$, της τάξης $[10^{-4}, 10^{-3}]$. Πιο συγκεκριμένα, έστω τυχαίες ανεξάρτητες και ισόνομες παρατηρήσεις X_1, X_2, \dots, X_n οι οποίες περιγράφουν το ύψος της στάθμης του νερού με άγνωστη συνάρτηση πυκνότητας πιθανότητας και αθροιστική συνάρτηση πιθανότητας F . Ας υποθέσουμε ότι $p < \frac{1}{n} \in \mathbb{R}$. Συνεπώς, η εκτίμηση που πρέπει να γίνει είναι ο προσδιορισμός μιας θετικής πραγματικής τιμής, έστω x_p , έτσι ώστε $F(x_p) = 1 - p$. Ουσιαστικά πρέπει να υπολογιστεί το $(1 - p)$ -ποσοστημόριο της F (βλ [28]). Βάσει ιστορικών τιμών, το έτος 1570 είχε καταγραφεί η τιμή (NAP+4) μέτρα όπου η συντομογραφία NAP προέρχεται από τα γερμανικά αρχικά του normal Amsterdam water level και είναι η αναφορική τιμή του ύψους του νερού για την Ολλανδία. Το αναφερόμενο έτος 1953 καταγράφηκε μια τιμή ύψους (NAP+3.85) μέτρα. Άρα θα ήταν λογικό να πούμε ότι μια γρήγορη προσέγγιση για τη συχνότητα τέτοιων ακραίων τιμών για το ύψος του νερού θα είναι 383 έτη. Η αναφορά της επιτροπής που μελέτησε τα δεδομένα κατέληξε σε μια προσέγγιση (NAP+5.14) μέτρα κάθε 10.000 έτη βασιζόμενη σε 166 ιστορικές παρατηρήσεις (βλ [28]).

Κάνοντας μια λογική αντιστοίχιση με το παραπάνω παράδειγμα οι μεθοδολογίες μπορούν να μεταφερθούν για τη μελέτη ασφαλιστικών απαιτήσεων. Όταν εξετάζουμε μια σειρά από ασφαλιστικές απαιτήσεις είναι σύνηθες να τις λαμβάνουμε υπόψιν ως μια κατανομή. Αυτό διότι η φιλοσοφία της ασφάλισης έχει δομηθεί γύρω από κάποιες βασικές αρχές. Συγκεκριμένα την αρχή της ανεξαρτησίας των απαιτήσεων ή πιο συγκεκριμένα τη χαμηλή συσχέτιση ανάμεσα στις απαιτήσεις. Πρακτικά, μια κατανομή απαιτήσεων αποτελείται από δύο μέρη, το κεντρικό κομμάτι της κατανομής όπου βρίσκονται οι συνήθεις (κοντά στη μέση τιμή) παρατηρήσεις και το άκρο της κατανομής (ουρά).

Ένας ασφαλιστής γνωρίζει καλά ότι ένα μεμονωμένο γεγονός μπορεί να έχει ως αποτέλεσμα πολλαπλούς θανάτους, τραυματισμούς και απώλεια περιουσίας όπως για παράδειγμα ένα μαζικό τροχαίο ή φωτιά που μεταδόθηκε σε ένα μεγάλο αριθμό κτιρίων. Παρακάτω δίνουμε ένα κομμάτι της βασικής θεωρίας που χρησιμοποιείται στη θεωρία ακραίων τιμών ώστε να γίνουν περισσότερο κατανοητά τα κεφάλαια που θα ακολουθήσουν.

2.2 Βασική Θεωρία

Ας υποθέσουμε λοιπόν ότι X_1, \dots, X_n είναι ημερήσιες ή μηνιαίες απαιτήσεις που εγείρονται σε μια ασφαλιστική εταιρία σε ένα έτος. Συμβολίζουμε με Z_n τη μέγιστη ετήσια απαίτηση, δηλαδή

$$Z_n = \max(X_1, \dots, X_n). \quad (1)$$

Παρακάτω δίνουμε τους ορισμούς σύγκλιση κατά πιθανότητα και σύγκλιση κατά κατανομή που θα χρησιμοποιήσουμε στη συνέχεια.

Ορισμός 2.2.1 (Σύγκλιση κατά κατανομή). *Μια ακολουθία τυχαίων μεταβλητών X_1, X_2, \dots, X_n λέγεται ότι συγκλίνει κατά κατανομή ή συγκλίνει ασθενώς σε μία τυχαία μεταβλητή X αν*

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

για όλα τα $x \in \mathbb{R}$ για τα οποία η F είναι συνεχής. Εδώ οι F_n και F είναι οι αθροιστικές κατανομές των τυχαίων μεταβλητών X_n και X αντίστοιχα.

Η σύγκλιση κατά κατανομή θα συμβολίζεται με $X_n \xrightarrow{L} X$.

Ορισμός 2.2.2 (Σύγκλιση κατά πιθανότητα). Μια ακολουθία τυχαίων μεταβλητών X_1, X_2, \dots, X_n λέγεται ότι συγκλίνει κατά πιθανότητα σε μία τυχαία μεταβλητή X και συμβολίζεται με $X_n \xrightarrow{P} X$ αν για κάθε $\epsilon > 0$ ισχύει

$$\lim_{n \rightarrow \infty} P\{|X_n - X| \geq \epsilon\} = 0. \quad (2)$$

Να σημειώσουμε εδώ ότι η σύγκλιση κατά πιθανότητα συνεπάγει τη σύγκλιση κατά κατανομή ενώ το αντίθετο δεν ισχύει γενικά.

Έστω λοιπόν $Z_n = \max(X_1, \dots, X_n)$ όπως ορίσαμε νωρίτερα. Οι X_1, X_2, \dots, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με αθροιστική συνάρτηση κατανομής $F(x)$ και έστω x^* το δεξιό άκρο των τιμών που μπορούν να λάβουν οι τυχαίες μεταβλητές X_i με θετική πιθανότητα, δηλαδή

$$x^* := \sup\{x > 0 : F(x) < 1\}. \quad (3)$$

Τότε θα έχουμε

$$Z_n = \max(X_1, X_2, \dots, X_n) \xrightarrow{P} x^*, \quad \text{καθώς } n \rightarrow \infty \quad (4)$$

Πράγματι, η πιθανότητα μια απαίτηση να μην υπερβαίνει ένα συγκεκριμένο όριο, έστω x , θα είναι

$$P(Z_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^n P(X_i \leq x) = F^n(x), \quad (5)$$

από το οποίο προκύπτει άμεσα ότι

$$\begin{aligned} F^n(x) &\rightarrow 0, & \text{για } x < x^*, \\ F^n(x) &\rightarrow 1, & \text{για } x \geq x^*. \end{aligned}$$

Στην πράξη το πρόβλημα έγκειται στον προσδιορισμό της αθροιστικής κατανομής F , η οποία είναι συνήθως άγνωστη. Πρέπει λοιπόν να αναρωτηθούμε ποιες είναι οι πιθανές κατανομές που θα μπορούσαν να περιγράφουν τα Z_n καθώς το n πλησιάζει στο άπειρο.

Σε αυτό το σημείο είναι χρήσιμο να δώσουμε μια αντιστοίχιση με το κεντρικό οριακό θεώρημα ώστε να ακολουθήσουμε παρόμοια λογική για τα Z_n απαντώντας έτσι στα παραπάνω ερωτήματα.

Ας θεωρήσουμε λοιπόν το άθροισμα $S_n = X_1 + X_2 + \dots + X_n$. Τότε η αναμενόμενη μέση τιμή των S_n θα είναι $n\mu$ και η τυπική απόκλιση θα είναι $\sigma n^{1/2}$. Σύμφωνα με το Κεντρικό Οριακό Θεώρημα, χωρίς να χρησιμοποιούμε αυστηρούς ορισμούς μπορούμε να πούμε ότι τα S_n θα ακολουθούν την κανονική κατανομή, $N(n\mu, \sigma^2 n)$ καθώς το n πλησιάζει το άπειρο. Κανονικοποιούμε τα S_n ως εξής

$$M_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \quad (6)$$

Το κεντρικό οριακό θεώρημα αναφέρει ότι τα M_n θα συγκλίνουν στην $N(0, 1)$. Παρόμοια μπορούμε να υπολογίσουμε την κατανομή των Z_n χωρίς να γνωρίζουμε την αθροιστική κατανομή $F(x)$ των X_1, X_2, \dots, X_n . Για να γίνουμε πιο αυστηροί εισαγάγουμε τους παρακάτω ορισμούς

Ορισμός 2.2.3 (Μεγιστο-ευστάθεια). Έστω X_1, X_2, \dots μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών με συνάρτηση κατανομής F . Η κατανομή που αντιστοιχεί στην F θα λέγεται *μεγιστο-ευσταθής* (*max-stable*) αν υπάρχουν ακολουθίες κανονικοποίησης $a_n > 0, b_n \in \mathbb{R}$ τέτοιες ώστε

$$\max(X_1, \dots, X_n) \stackrel{L}{\cong} a_n X + b_n. \quad (7)$$

Με απλά λόγια, αν μια συνάρτηση κατανομής F είναι μεγιστο-ευσταθής, σημαίνει ότι το κανονικοποιημένο μέγιστο, $\frac{Z_n - b_n}{a_n}$ θα ακολουθεί πάλι την ίδια συνάρτηση κατανομής F .

Ορισμός 2.2.4 (Περιοχή έλξης). Έστω X_1, X_2, \dots μια ακολουθία τυχαίων μεταβλητών με συνάρτηση κατανομής F . Η κατανομή που αντιστοιχεί στην F θα ανήκει στη μέγιστη περιοχή έλξης (*maximum domain of attraction*) μιας μη-εκφυλισμένης συνάρτησης G αν για κατάλληλες ακολουθίες $a_n > 0, b_n \in \mathbb{R}$ ισχύει ότι

$$\lim_{n \rightarrow \infty} F^n(a_n + b_n x) = G(x). \quad (8)$$

Αντίστοιχα, θα ανήκει στην ελάχιστη περιοχή έλξης (*minimum domain of attraction*) μιας μη-εκφυλισμένης συνάρτησης L αν για κατάλληλες ακολουθίες $c_n > 0, d_n \in \mathbb{R}$ ισχύει ότι

$$\lim_{n \rightarrow \infty} 1 - [1 - F(c_n + d_n x)]^n = L(x) \quad (9)$$

δηλαδή,

$$1 - [1 - F(c_n + d_n x)]^n \rightarrow L(x) \quad \forall x \quad L \in C^0. \quad (10)$$

Αποδεικνύεται ότι μόνο μερικές κατανομές G μπορούν να προκύψουν ως ασυμπτωτικό όριο κατανομής των κανονικοποιημένων μεγίστων Z_n . Έτσι προκύπτει το πιο σημαντικό θεώρημα της θεωρίας ακραίων τιμών.

Θεώρημα 2.2.1 (Fisher-Tippet (1928), Gnedenko (1943)). *Αν υπάρχουν ακολουθίες κανονικοποίησης $a_n > 0$ και $b_n \in \mathbb{R}$ έτσι ώστε*

$$\lim_{n \rightarrow \infty} P\left\{\frac{Z_n - b_n}{a_n} \leq x\right\} = G(x) \quad (11)$$

για κάποια μη-εκφυλισμένη συνάρτηση κατανομής G τότε η G θα είναι του ίδιου τύπου με μία από τις επόμενες οικογένειες κατανομών,

$$G_0(x) \equiv \Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R} \quad (12)$$

$$G_{1,\alpha}(x) \equiv \Phi_\alpha(x) = \begin{cases} 0 & \text{αν } x \leq 0 \\ e^{-x^{-\alpha}} & \text{αν } x > 0 \end{cases} \quad (13)$$

$$G_{2,\alpha}(x) \equiv \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha} & \text{αν } x \leq 0 \\ 1 & \text{αν } x > 0 \end{cases} \quad (14)$$

όπου α είναι μια θετική σταθερά. Η $\Lambda(x)$ ονομάζεται κατανομή *Gumbel*, η $\Phi_\alpha(x)$ κατανομή *Fréchet* και η $\Psi_\alpha(x)$ κατανομή *αντεστραμμένη Weibull*.

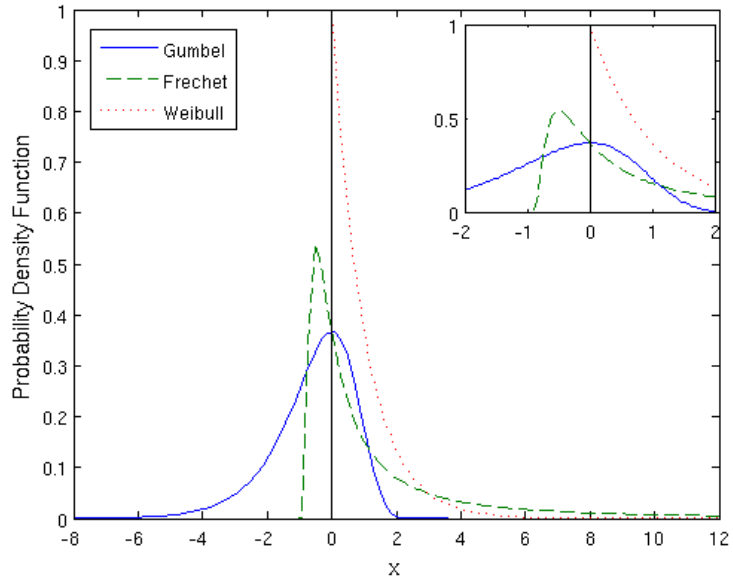
Οι κατανομές $\Lambda, \Phi_\alpha, \Psi_\alpha$ ονομάζονται τυπικές κατανομές ακραίων τιμών (*extreme value distributions*) ή κατανομές ακροτάτων, ενώ οι κατανομές που είναι ίδιου τύπου με αυτές καλούνται απλώς κατανομές ακροτάτων [34].

Ορισμός 2.2.5 (Κατανομές ιδίου τύπου). Έστω X_1, X_2 δυο τυχαίες μεταβλητές με αθροιστικές συναρτήσεις κατανομής F_1 και F_2 αντίστοιχα. Οι F_1 και F_2 είναι του ιδίου τύπου αν υπάρχουν κατάλληλες σταθερές $c > 0, d \in \mathbb{R}$ με $X_2 \stackrel{L}{=} cX_1 + d$ έτσι ώστε να ισχύει

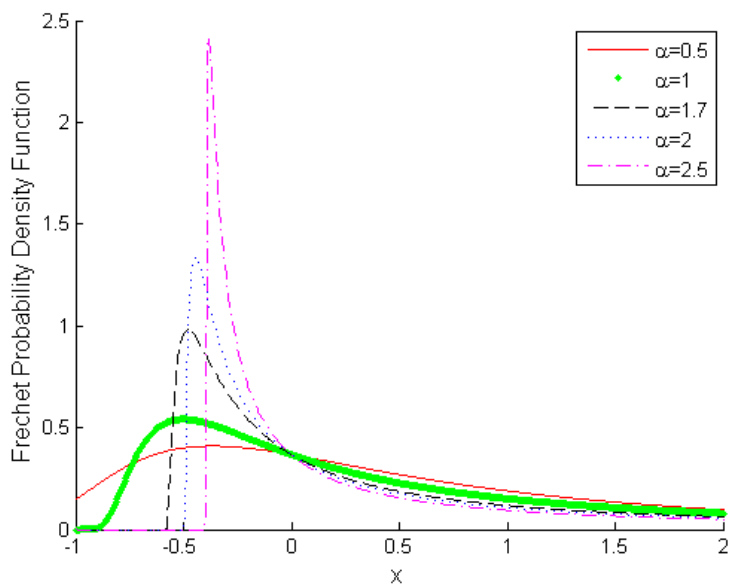
$$F_2 = F_1\left(\frac{x-d}{c}\right). \quad (15)$$

Αποδεικνύεται ότι οι μόνες μη-εκφυλισμένες λύσεις της συναρτησιακής εξίσωσης που προκύπτει από τον ορισμό των μεγιστο-ευσταθών κατανομών, είναι οι κατανομές Gumbel, Fréchet και Weibull.

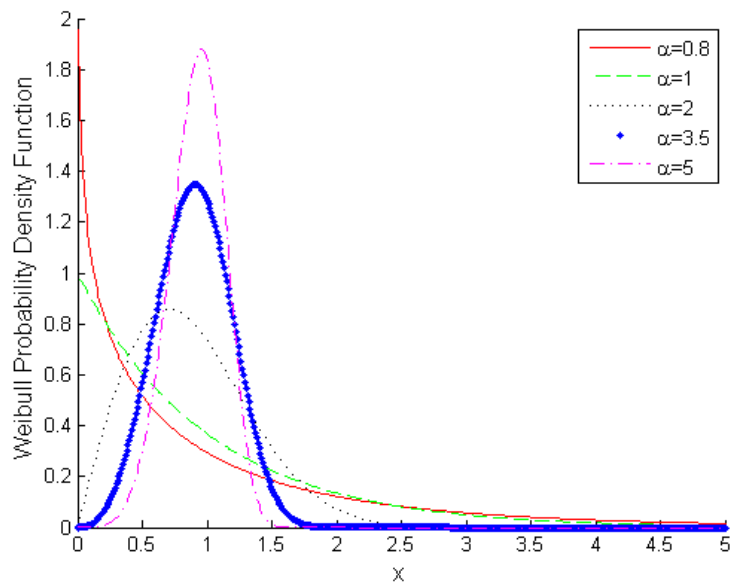
Παρακάτω παραθέτουμε διάγραμμα των τριών κατανομών για $\alpha = 1$ και διαγράμματα των κατανομών Weibull και Fréchet για ενδεικτικές τιμές του α .



Σχήμα 1. Συναρτήσεις πυκνότητας πιθανότητας Gumbel, Fréchet και Weibull για $\alpha = 1$.



Σχήμα 2. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Fréchet για $\alpha = (0.5, 1, 1.7, 2, 2.5)$.



Σχήμα 3. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull για $\alpha = [0.8, 1, 2, 3.5, 5]$.

Παρατηρούμε πως όταν $\alpha = 1$ η κατανομή Weibull ταυτίζεται με την εκθετική κατανομή ενώ όταν $\alpha = 2$ αποκαλείται κατανομή Rayleigh. Θυμίζουμε ότι η αθροιστική συνάρτηση κατανομής της Rayleigh δίνεται από τον παρακάτω τύπο

$$F(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}, \quad x \in [0, \infty]. \quad (16)$$

Αυξάνοντας την τιμή της παραμέτρου, πιο συγκεκριμένα από $\alpha = 3$ και άνω, η κατανομή Weibull θυμίζει την κανονική κατανομή.

Είναι ενδιαφέρον να αναφέρουμε στο σημείο αυτό κάποιες πληροφορίες για τις οριακές κατανομες.

Η κατανομή Gumbel έλαβε το όνομά της από τον Γερμανό μαθηματικό Emil Julius Gumbel και περιγράφεται αναλυτικά σε σχετικές έρευνες από τον ίδιο (βλ [47], [26]). Μια τέτοια κατανομή θα μπορούσε να χρησιμοποιηθεί λόγου χάρη, για να περιγράψει την κατανομή του μέγιστου ύψους της στάθμης του νερού ενός ποταμού σε ένα συγκεκριμένο έτος, αν υπήρχε καταγραφή ιστορικών δεδομένων για το μέγιστο ύψος τα προηγούμενα έτη. Ο ίδιος ο Gumbel τη χρησιμοποιούσε συχνά στις μελέτες του καθώς όπως εξήγησε χαρακτηριστικά το 1958, ‘it seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.’

Η κατανομή Fréchet έλαβε το όνομά της από τον Γάλλο μαθηματικό Maurice Fréchet σε σχετική δημοσίευσή του το 1927 στην οποία αναφέρεται σε μια πιθανή οριακή κατανομή των μεγίστων υπό την προϋπόθεση κατάλληλης κανονικοποίησής τους. Συγκεκριμένα, υπέθεσε όρια της μορφής

$$\lim_{n \rightarrow \infty} Pr\left(\frac{X_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x) = G(x) \quad (17)$$

καταλήγωντας στην Φ ως πιθανό όριο (βλ [20]). Την ίδια περίοδο ο Tippett, σε έρευνα του, προσπαθούσε να ‘προσδιορίσει την οριακή κατανομή για ένα μεγάλο αριθμό δειγμάτων’ (βλ [54]). Και οι δύο κατέληξαν στις οριακές κατανομές αλλά μόνο ο Fréchet έδινε έμφαση στη σημαντικότητα της ευστάθειας. Έτσι ήταν ο πρώτος που εισήγαγε τον όρο αλλά και τη σημασία της μεγιστο-ευστάθειας (max-stability). Η κατανομή έχει σημαντικές εφαρμογές στις χρηματοοικονο-

μικές αγορές διότι λόγω της βαριάς ουράς περιγράφει επαρκώς χρηματοοικονομικές αποδόσεις.

Η κατανομή Weibull έλαβε το όνομά της από τον Σουηδό μηχανικό και επιστήμονα Waloddi Weibull γνωστό για την έρευνα του στην αντοχή υλικών. Παρόλο που η κατανομή Weibull χρησιμοποιήθηκε αρχικά για να περιγράψει ελάχιστες ακραίες τιμές που παρουσιάζονταν σε μελέτες για την αντοχή των υλικών, χρησιμοποιείται ευρέως σε πολλούς τομείς. Χρησιμοποιείται με μεγάλη επιτυχία στη θεωρία αξιοπιστίας, για παράδειγμα για να μελετήσει τη θραύση τάσης των ηλεκτρικών κυκλωμάτων (βλ [7], [30]). Επίσης, έχει εφαρμογή και σε άλλους κλάδους, όπως στη φυσική σε μελέτες κρυστάλλωσης, στην κλιματολογία σε μελέτες των παλιρροιών ακόμα και στη γνωστική ψυχολογία όπου περιγράφει το χρόνο που χρειάζεται κάποιος για να ολοκληρώσει μια εργασία (βλ [28], [17], [9]). Είναι ενδιαφέρον ότι ο ίδιος ο Weibull ισχυριζόταν ότι η κατανομή μπορούσε να βρει εφαρμογή σε ένα ευρύ φάσμα προβλημάτων (βλ [55]). Μάλιστα, εξέθετε επτά παραδείγματα, που αποδείκνυαν τον ισχυρισμό του και αναφέρονταν στην αντοχή του ατσαλιού αλλά ακόμα και στο ύψος των ενήλικων ανδρών στη Βρετανία. Αργότερα ο E.J. Gumbel απέδειξε ότι ο αριθμός αποτυχιών ενός εξαρτήματος μιας μηχανής (failure rate) είναι ανάλογος του χρόνου λειτουργίας της μηχανής υψωμένου στην $\alpha + 1$, όπου α η παράμετρος σχήματος όπως θα δούμε αργότερα.

2.3 Περιοχή έλξης κατανομών ακροτάτων

Παρακάτω παρουσιάζουμε συνοπτικά τις ικανές και αναγκαίες συνθήκες ώστε μια συνάρτηση κατανομής να ανήκει στην περιοχή έλξης μιας από τις κατανομές ακροτάτων καθώς και μερικά παραδείγματα για την καλύτερη κατανόηση των θεωρημάτων.

Ορισμός 2.3.1 (Μέγιστη περιοχή έλξης βάσει ουράς). *Μια συνάρτηση κατανομής F ανήκει στην μέγιστη περιοχή έλξης (maximum domain of attraction) μιας κατανομής ακροτάτων $G_i, i = 0, 1, 2$ με ακολουθίες κανονικοποίησης*

$a_n > 0, b_n \in \mathfrak{R}$ αν και μόνο αν

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = -\ln G_i(x). \quad (18)$$

Ορισμός 2.3.2 (Συνάρτηση ομαλής κύμανσης). Μια θετική μετρήσιμη συνάρτηση $h \in (0, \infty)$ ονομάζεται συνάρτηση ομαλής κύμανσης (*regularly varying*) στο άπειρο με δείκτη $u \in \mathfrak{R}$ και συμβολίζεται $h \in \mathcal{R}_u$ αν

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^u, \quad t \geq 0. \quad (19)$$

Αν $u = 0$ τότε καλείται βραδείας κύμανσης. Αποδεικνύεται εύκολα ότι κάθε συνάρτηση $h \in \mathcal{R}$ μπορεί να γραφεί $h(x) = x^u L(x)$.

Θεώρημα 2.3.1. (Περιοχή έλξης της Gumbel) Έστω F μια συνάρτηση κατανομής. Τότε η F ανήκει στη μέγιστη περιοχή έλξης της Gumbel αν και μόνο αν μπορεί να αναπαρασταθεί ως εξής

$$\bar{F}(x) = c(x)e^{-\int_z^x \frac{g(t)}{a(t)} dt}, \quad z < x < z^*, \quad (20)$$

για κάποιο $z < x^*$, όπου c, g είναι συναρτήσεις για τις οποίες ισχύει

$$c(x) \rightarrow c_0 > 0, \quad (21)$$

$$g(x) \rightarrow 1 \quad (22)$$

όταν $x \uparrow x^*$ και $\lim_{x \uparrow x^*} a'(x) = 0$. Επίσης θα ισχύει,

$$\frac{Z_n - b_n}{\alpha_n} \xrightarrow{\mathcal{L}} \Lambda. \quad (23)$$

Θεώρημα 2.3.2. (Περιοχή έλξης της Frechét) Έστω F μια συνάρτηση κατανομής. Τότε η F ανήκει στη μέγιστη περιοχή έλξης της Frechét αν και μόνο αν $\bar{F} \in \mathcal{R}_{-u}, u > 0$. Επίσης, θα ισχύει

$$\frac{Z_n - b_n}{\alpha_n} \xrightarrow{\mathcal{L}} \Phi_\alpha. \quad (24)$$

Κατανομή Ακροτάτων	MDA	Κατανομή
$\Lambda(x)$	$\bar{F}(x) = c(x)e^{-\int_z^x \frac{g(t)}{a(t)} dt}, \quad z < x < z^*$	Εκθετική, Κανονική, Gamma, Lognormal, Weibull
$\Phi(x)$	$\bar{F}(x) = x^{-u}L(x) \in \mathcal{R}_{-u}$	Pareto, Loggamma, Cauchy, Burr
$\Psi(x)$	$\bar{F}(x^* - x^{-1}) = x^{-u}L(x) \in \mathcal{R}_{-u}$	Ομοιόμορφη, Beta

Πίνακας 2. Συνοπτικός πίνακας κατανομών ακροτάτων.

Θεώρημα 2.3.3. (Περιοχή έλξης της Weibull) Έστω F μια συνάρτηση κατανομής. Τότε η F ανήκει στη μέγιστη περιοχή έλξης της Weibull αν και μόνο αν $x^* < \infty$ και $\bar{F}(x^* - x^{-1}) \in \mathcal{R}_{-u}, u > 0$. Επίσης, θα ισχύει

$$\frac{Z_n - x^*}{\alpha_n} \xrightarrow{\mathcal{L}} \Psi_\alpha. \quad (25)$$

Είναι σημαντικό να τονίσουμε πως αν οι παρατηρήσεις ενός δείγματος υπό μελέτη ακολουθούν την κανονική, εκθετική, λογαριθμοκανονική, γάμμα ή Weibull κατανομή, τότε όπως θα δούμε αργότερα, τα μέγιστα θα ακολουθούν την κατανομή Gumbel. Παρακάτω παραθέτουμε ένα συγκεντρωτικό πίνακα παραδειγμάτων κατανομών μεταξύ αυτών και συνήθων κατανομών όπως αναφέρθηκε παραπάνω και την περιοχή έλξης των κατανομών ακροτάτων στην οποία ανήκουν. Να σημειώσουμε ότι βάσει του ορισμού 2.3.2, η συνάρτηση ομαλής κύμανσης με δείκτη $-u$ συμβολίζεται \mathcal{R}_{-u} .

Παρακάτω παρουσιάζουμε παραδείγματα για την περιοχή έλξης της κάθε κατανομής ακροτάτων. Έστω ότι έχουμε ανεξάρτητες και ισόνομες τυχαίες μεταβλητές για τις οποίες ισχύει ότι $X_1, X_2, \dots, X_n \sim \text{Exp}(\lambda)$, όπου λ είναι η παράμετρος της εκθετικής κατανομής. Αν επιλέξουμε $b_n = \frac{1}{\lambda} \ln n$ και $a_n = \frac{1}{\lambda}$ προκύπτει άμεσα ότι,

$$\begin{aligned} Pr\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= Pr\left(\frac{Z_n - \frac{\ln n}{\lambda}}{\frac{1}{\lambda}} \leq x\right) = Pr\left(Z_n \leq \frac{\ln n + x}{\lambda}\right) \\ &= 1 - \bar{F}\left(\frac{\ln n + x}{\lambda}\right)^n = (1 - e^{-\lambda \frac{\ln n + x}{\lambda}})^n \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n \rightarrow e^{-e^{-x}}, x \in \mathbb{R} \end{aligned} \quad (26)$$

όπου \bar{F} είναι η ουρά της εκθετικής κατανομής, δηλαδή $\bar{F} = 1 - F$. Γενικά αποδεικνύεται ότι για μια ακολουθία k_n , ισχύει ότι $(1 + \frac{k_n}{n})^n \rightarrow e^{\lim_{n \rightarrow \infty} k_n}$ (η απόδειξη δίνεται στο Παράρτημα). Η $e^{-e^{-x}}$ είναι η αθροιστική συνάρτηση κατανομής της Gumbel. Αποδείξαμε ότι η κατανομή Gumbel ανήκει στο σύνολο των μη-εκφυλισμένων κατανομών που μπορούν να προκύψουν ως όριο κάποιου κανονικοποιημένου μεγίστου.

Να σημειώσουμε ότι οι ακολουθίες a_n, b_n που εξασφαλίζουν ότι το κανονικοποιημένο μέγιστο $\frac{Z_n - b_n}{a_n}$ συγλίνει σε μια οριακή κατανομή δεν είναι μοναδικές. Επιλέγοντας διαφορετικές ακολουθίες, έστω a'_n, b'_n , μπορούμε να καταλήξουμε σε μια άλλη οριακή κατανομή η οποία όμως αποδεικνύεται ότι θα είναι πάντα ιδίου τύπου με την προηγούμενη.

Επιπρόσθετα αποδεικνύεται εύκολα ότι η κατανομή Gumbel είναι μεγιστοευσταθής. Αν υποθέσουμε ότι $X_1, X_2, \dots, X_n \sim \Lambda$, και ότι $b_n = \ln n$ και $a_n = 1$, τότε θα έχουμε,

$$\begin{aligned} Pr\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= Pr(Z_n \leq \ln n + x) = \Lambda(\ln n + x)^n \\ &= e^{-ne^{-(\ln n + x)}} = e^{-e^{-x}} \end{aligned} \quad (27)$$

Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν τη κατανομή Cauchy με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\pi(1+x)}, \quad x \in \mathbb{R} \quad (28)$$

Για κατάλληλες ακολουθίες $b_n = 0$ και $a_n = \frac{n}{\pi}$ θα έχουμε

$$\begin{aligned} Pr\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= Pr\left(\frac{\pi Z_n}{n} \leq x\right) = Pr\left(Z_n \leq \frac{nx}{\pi}\right) \\ &= [1 - \bar{F}\left(\frac{nx}{\pi}\right)]^n, \quad x \in \mathbb{R} \end{aligned} \quad (29)$$

όμως επειδή ισχύει

$$\lim_{x \rightarrow \infty} \frac{\bar{F}}{(\pi x)^{-1}} = \lim_{x \rightarrow \infty} \frac{f(x)}{\pi^{-1}x^{-2}} = \lim_{x \rightarrow \infty} \frac{x^2}{1+x^2} = 1 \quad (30)$$

προκύπτει τελικά ότι

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \bar{F}\left(\frac{nx}{\pi}\right)\right)^n &= \lim_{n \rightarrow \infty} \left(1 - \frac{\bar{F}\left(\frac{nx}{\pi}\right)}{nx\left(\pi\frac{nx}{\pi}\right)^{-1}}\right)^n \\ &= \exp\left(-\lim_{n \rightarrow \infty} \frac{\bar{F}\left(\frac{nx}{\pi}\right)}{nx\left(\pi\frac{nx}{\pi}\right)^{-1}}\right) = e^{-x^{-1}}, \quad x \in \mathbb{R} \end{aligned} \quad (31)$$

και επομένως εδώ η οριακή κατανομή είναι η κατανομή Fréchet. Επιπλέον μπορούμε εύκολα να αποδείξουμε ότι η κατανομή Fréchet είναι μεγιστο-ευσταθής. Έστω ότι $X_1, X_2, \dots, X_n \sim \Phi$ και $b_n = 0, a_n = n^{\frac{1}{\alpha}}$. Τότε θα έχουμε,

$$\begin{aligned} Pr\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= Pr\left(Z_n \leq xn^{\frac{1}{\alpha}}\right) = \Phi^n(xn^{\frac{1}{\alpha}}) \\ &= [e^{-(xn^{\frac{1}{\alpha}})^{-\alpha}}]^n = e^{-x^{-\alpha}}, \quad x > 0. \end{aligned} \quad (32)$$

Ακολουθώντας παρόμοια λογική με τις προηγούμενες δύο κατανομές είναι χρήσιμο να αναφέρουμε ότι αν οι παρατηρήσεις του δείγματος υπο μελέτη ακολουθούν ομοιόμορφη ή κατανομή Beta τότε όπως θα δείξουμε παρακάτω, τα

μέγιστα θα ακολουθούν την κατανομή Weibull .

Για παράδειγμα, έστω ανεξάρτητες και ισόνομες τυχαίες μεταβλητές για τις οποίες ισχύει $(X_1, X_2, \dots, X_n) \sim Uniform[0, 1]$. Προκύπτει άμεσα ότι αν υποθέσουμε κατάλληλες ακολουθίες $b_n = 1$ και $a_n = \frac{1}{n}$ θα έχουμε,

$$\begin{aligned} Pr\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= Pr(nZ_n - n \leq x) = Pr\left(Z_n \leq 1 + \frac{x}{n}\right) \\ &= \left(F\left(1 + \frac{x}{n}\right)\right)^n = \left(1 + \frac{x}{n}\right)^n \rightarrow e^x, \quad x < 0 \end{aligned} \quad (33)$$

η οποία είναι η αθροιστική συνάρτηση κατανομής της αντεστραμμένης Weibull για $\alpha = 1$.

Έστω τώρα ότι έχουμε ανεξάρτητες και ισόνομες τυχαίες μεταβλητές για τις οποίες ισχύει $X_1, X_2, \dots, X_n \sim \Psi$. Για να δείξουμε ότι η Weibull είναι μεγιστοευσταθής κατανομή αρκεί να υποθέσουμε τις ακολουθίες $b_n = 0$ και $a_n = n^{-\frac{1}{\alpha}}$. Έτσι, για το κανονικοποιημένο μέγιστο θα έχουμε

$$\begin{aligned} Pr\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= Pr\left(Z_n \leq xn^{-\frac{1}{\alpha}}\right) = \Psi^n(xn^{-\frac{1}{\alpha}}) \\ &= \left[e^{-(-xn^{-\frac{1}{\alpha}})^\alpha}\right]^n = e^{-(-x)^\alpha}, \quad x \leq 0. \end{aligned} \quad (34)$$

Οι τρεις τύποι κατανομών που ορίσαμε σε παραπάνω κεφάλαιο μπορούν να συνδυαστούν σε μία κατανομή, τη γενικευμένη κατανομή ακραίων τιμών, για την οποία θα μιλήσουμε εκτενώς σε επόμενο κεφάλαιο. Στις επόμενες παραγράφους θα δώσουμε τη θεωρία των μεθόδων εκτίμησης παραμέτρων και έπειτα θα αναλύσουμε την κάθε οριακή κατανομή ξεχωριστά.

3 Εκτίμηση παραμέτρων για τις κατανομές ακραίων τιμών

3.1 Μέθοδος μέγιστης πιθανοφάνειας

Έστω x_1, x_2, \dots, x_n οι παρατηρηθείσες τιμές για ένα δείγμα ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από μια κατανομή με άγνωστη συνάρτηση πυκνότητας πιθανότητας η οποία ανήκει σε μια οικογένεια κατανομών $\{f(\cdot|\theta), \theta \in \Theta\}$ όπου το $\theta = (\theta_1, \theta_2, \dots)$ είναι ένα διάνυσμα παραμέτρων. Η τιμή του θ είναι άγνωστη. Ο στόχος είναι να βρεθεί μια εκτιμήτρια $\hat{\theta}$ η οποία να έχει 'καλές' ιδιότητες. Η απο κοινού κατανομή των παρατηρήσεων θα είναι

$$f(x_1, x_2, \dots, x_n) = f(x_1|\theta)f(x_2|\theta) \cdots f(x_n|\theta).$$

Αν υποθέσουμε ότι x_1, x_2, \dots, x_n είναι σταθεροί αριθμοί και θ είναι η μεταβλητή της συνάρτησης, τότε η συνάρτηση πιθανοφάνειας θα είναι

$$\mathcal{L}(\theta|x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta). \quad (35)$$

Συνήθως, για ευκολία στους υπολογισμούς, χρησιμοποιούμε το λογάριθμο της συνάρτησης πιθανοφάνειας, δηλαδή,

$$l(\theta|x_1, x_2, \dots, x_n) = \ln \mathcal{L}(\theta|x_1, x_2, \dots, x_n) = \sum_{i=1}^n \ln f(x_i|\theta). \quad (36)$$

Ας υποθέσουμε τώρα για ευκολία ότι έχουμε μόνο δύο παραμέτρους δηλαδή, $\theta = (\theta_1, \theta_2)$. Παραγωγίζοντας την παραπάνω εξίσωση ως προς κάθε παράμετρο ξεχωριστά και θέτοντάς τη ίση με το 0 καταλήγουμε στο παρακάτω σύστημα εξισώσεων,

$$\frac{\partial l(\theta)}{\partial \theta_1} = 0, \quad (37)$$

$$\frac{\partial l(\theta)}{\partial \theta_2} = 0. \quad (38)$$

Απο το παραπάνω σύστημα μπορούν να υπολογισθούν οι εκτιμήτριες $(\hat{\theta}_1, \hat{\theta}_2)$ των παραμέτρων θ_1, θ_2 είτε με αλγεβρικές μεθόδους είτε χρησιμοποιώντας κάποια επαναληπτική αριθμητική μέθοδο.

3.2 Μέθοδος ροπών

Ας υποθέσουμε ότι έχουμε ένα τυχαίο πείραμα το οποίο σχετίζεται με μια παρατηρούμενη μεταβλητή X που λαμβάνει θετικές τιμές. Η κατανομή των X έχει ένα αριθμό άγνωστων παραμέτρων έστω $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ που λαμβάνουν τιμές στο χώρο παραμέτρων Θ . Επαναλαμβάνουμε το πείραμα n φορές έτσι ώστε να παράγουμε ένα τυχαίο δείγμα μεγέθους n από τη κατανομή του X . Θα έχουμε λοιπόν μία σειρά ανεξάρτητων και ισόνομων τυχαίων μεταβλητών, X_1, X_2, \dots, X_n που ακολουθεί τη κατανομή της X .

Η μέθοδος των ροπών είναι μια τεχνική εκτίμησης παραμέτρων που βασίζεται στην εξίσωση των θεωρητικών ροπών και των εμπειρικών ροπών. Αρχικά ορίζουμε τη j -οστή ροπή της X ως εξής

$$\mu_j(\theta) = E_\theta(X^j), \quad j = 1, 2, \dots \quad (39)$$

Εδώ να σημειώσουμε ότι από το συμβολισμό δίνεται έμφαση στην εξάρτηση των ροπών από το διάνυσμα των παραμέτρων. Επίσης η ροπή για $j = 1$ θα μας δίνει φυσικά τη μέση τιμή της X η οποία συμβολίζεται αλλιώς με μ . Έπειτα ορίζουμε τη j -οστή εμπειρική ροπή του δείγματος ως εξής:

$$M_j(X) = \frac{1}{n} \sum_{i=1}^n X_i^j, \quad j = 1, 2, \dots \quad (40)$$

Είναι σαφές ότι η ποσότητα $M_j(X)$ είναι η δειγματική μέση τιμή του j δείγματος $(X_1^j, X_2^j, \dots, X_n^j)$ από την κατανομή της X . Αντίστοιχα και εδώ η $M_1(X)$, θα είναι η συνήθης δειγματική μέση τιμή που συμβολίζεται απλά με M ενώ η $M_2(X)$ είναι η διακύμανση και συμβολίζεται με S . Γνωρίζουμε ότι η $M_j(X)$ είναι μια αμερόληπτη εκτιμήτρια της $\mu_j(\theta)$ για κάθε j . Συνεπώς, για να κατασκευάσουμε τις εκτιμήτριες αρκεί να λύσουμε το κάτωθι σύστημα k

εξισώσεων με k αγνώστους [56],

$$\mu_j(\theta_1, \theta_2, \dots, \theta_k) = M_j(X_1, X_2, \dots, X_n), \quad j = 1, 2, \dots, k. \quad (41)$$

3.3 Μέθοδος των σταθμισμένων ροπών πιθανότητας

Η τεχνική της μεθόδου των ροπών αναφέρεται στην εξίσωση των θεωρητικών (ροπές του μοντέλου) και των εμπειρικών ροπών όπως ήδη αναφέραμε παραπάνω. Παρόλο που οι γενικές ιδιότητες των εκτιμητών που εξάγονται από τη μέθοδο μπορεί να είναι αναξιόπιστες, μας δίνει ένα αρχικό προσεγγιστικό διάλυμα των παραμέτρων, μια αρχική δηλαδή πρώτη εκτίμηση, η οποία χρησιμοποιείται συνήθως ως είσοδος στην εφαρμογή αριθμητικών μεθόδων για την επίτευξη ακριβέστερης εκτίμησης.

Ως σταθμισμένες βάσει πιθανοτήτων ροπές (probability weighted moments) μιας τυχαίας μεταβλητής X με συνάρτηση κατανομής

$$F(x) = P(X \leq x)$$

ορίζονται οι ποσότητες [23]

$$M_{p,r,s} = E[X^p \{F(X)\}^r \{1 - F(X)\}^s], \quad p, r, s \in \mathbb{R}. \quad (42)$$

Οι ροπές αυτές θα ήταν πιο χρήσιμες στους υπολογισμούς μας αν η αντίστροφη της συνάρτησης κατανομής μπορούσε να γραφεί σε κλειστή μορφή. Τότε εκτελώντας την αλλαγή μεταβλητής $F(x) = y$ ο τύπος 42 δίνει,

$$M_{p,r,s} = \int_0^1 \{F^{-1}(y)\}^p y^r (1 - y)^s dy \quad (43)$$

διευκολύνοντας με αυτόν τον τρόπο τον υπολογισμό των ροπών. Οι ποσότητες $M_{p,0,0}$ είναι οι συνηθισμένες μη-κεντρικές ροπές της X , στις οποίες αναφερθήκαμε στην κλασική μέθοδο των ροπών. Οι ροπές $M_{1,r,s}$ μπορεί ωστόσο να είναι προτιμότερες για τον υπολογισμό των παραμέτρων της κατανομής της X

από τη στιγμή που η παρουσία μόνο της πρώτης τάξης της X στην έκφραση $M_{1,r,s}$ σημαίνει ότι η σχέση ανάμεσα στις παραμέτρους και τις ροπές λαμβάνει μια απλούστερη μορφή απο ότι στη συμβατική έκφραση. Όταν τα r, s είναι ακέραιοι αριθμοί τότε η έκφραση $F^r(1 - F)^s$ μπορεί να γραφεί σαν γραμμικός συνδυασμός δυνάμεων είτε της F είτε της $1 - F$, οπότε είναι φυσικό να χρησιμοποιήσουμε μια από τις ακόλουθες ροπές

$$\alpha_s \equiv M_{1,r,0} = E[X\{F(X)\}^r], \quad r = 0, 1, 2, \dots \quad (44)$$

$$\beta_r \equiv M_{1,0,s} = E[X\{1 - F(X)\}^s], \quad s = 0, 1, 2, \dots \quad (45)$$

Ας υποθέσουμε λοιπόν ότι επιλέγουμε την Σχέση 45 για τον υπολογισμό των εκτιμητριών $\hat{\theta}_1, \hat{\theta}_1$ των παραμέτρων θ_1, θ_2 . Μια αμερόληπτη εκτιμήτρια της M_s δίνεται από (βλ [36]),

$$\hat{M}_s = \frac{1}{n} \sum_{i=1}^n x_i \frac{\binom{n-i}{s}}{\binom{n-1}{s}}, \quad s = 0, 1, 2, \dots \quad (46)$$

Χρησιμοποιώντας τις άμεσες εκφράσεις των M_0, M_1 και εξισώνοντας με τις δειγματικές εκτιμήτριες \hat{M}_0, \hat{M}_1 λαμβάνουμε τις εκτιμήτριες των παραμέτρων

$$\hat{\theta}_1 = \frac{\hat{M}_0 - 2\hat{M}_1}{\ln 2}, \quad (47)$$

$$\hat{\theta}_2 = \hat{M}_0 - \gamma \hat{\theta}_1. \quad (48)$$

όπου γ είναι η σταθερά Euler.

3.4 Κριτήρια αποδοτικότητας μεθόδων

Στις επόμενες υπο-παραγράφους παραθέτουμε τις τρεις κατανομές Gumbel, Fréchet και Weibull καθώς και τις μεθόδους εκτίμησης των παραμέτρων τους. Τα κριτήρια τα οποία χρησιμοποιήσαμε για να αξιολογήσουμε τις εκτιμήσεις αυτών είναι τα παρακάτω:

Η αμεροληψία (Bias) είναι η απόσταση μεταξύ της μέσης τιμής του συνόλου των εκτιμήσεων και της πραγματικής τιμής της παραμέτρου που εκτιμάται. Έστω θ η παράμετρος προς εκτίμηση και $\hat{\theta}$ η εκτίμηση της παραμέτρου τότε η

αμεροληψία ορίζεται ως:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta \quad (49)$$

$$= E(\hat{\theta} - \theta). \quad (50)$$

Όπως βλέπουμε και από την παραπάνω σχέση η αμεροληψία ισούται και με την αναμενόμενη τιμή του σφάλματος ($\hat{\theta} - \theta$). Το μέσο τετραγωνικό σφάλμα εκφράζει την απόσταση καταμέσο όρο, του συνόλου των εκτιμήσεων σε σχέση με τη παράμετρο που εκτιμάται δηλαδή,

$$MSE(\hat{\theta}) = E[(\hat{\theta}(x) - \theta)^2]. \quad (51)$$

3.5 Εκτίμηση των παραμέτρων της κατανομής *Gumbel*

Η αθροιστική συνάρτηση κατανομής της κατανομής Gumbel με δύο παραμέτρους δίνεται από τον τύπο,

$$\Lambda(x; \mu, \beta) = e^{-e^{-\frac{x-\mu}{\beta}}}, \quad x \in \mathbb{R} \quad (52)$$

όπου μ είναι η παράμετρος θέσης και β η παράμετρος κλίμακας.

Παραγωγίζοντας τη Σχέση 52 θα λάβουμε τη συνάρτηση πυκνότητας πιθανότητας,

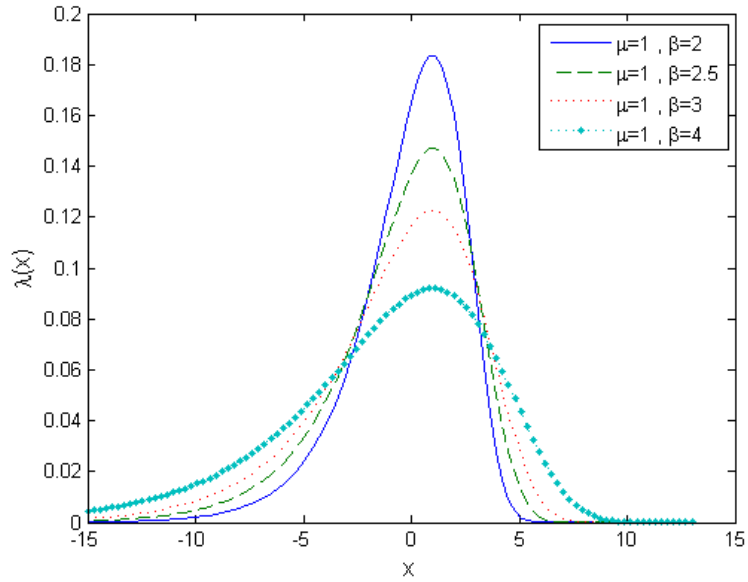
$$\lambda(x) = \frac{1}{\beta} e^{-(z+e^{-z})}, \quad \text{όπου } z = \frac{x-\mu}{\beta}. \quad (53)$$

Η αντίστροφη συνάρτηση της αθροιστικής συνάρτησης κατανομής θα δίνεται από τον παρακάτω τύπο,

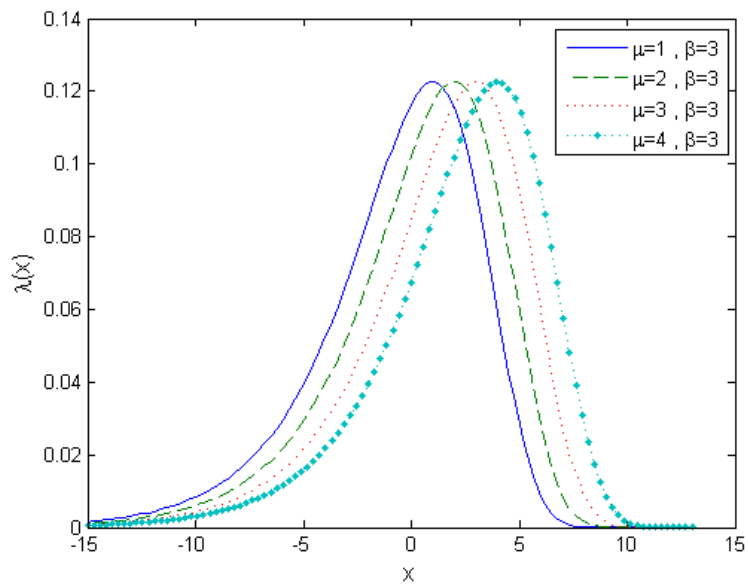
$$Q_{\Lambda(x)}(p) = \mu - \beta \ln(-\ln(p)). \quad (54)$$

Διότι,

$$\begin{aligned} \Lambda(x) = y &\Leftrightarrow e^{-e^{-\frac{x-\mu}{\beta}}} = y \Leftrightarrow -e^{-\frac{x-\mu}{\beta}} = \ln y \\ &\Leftrightarrow -\frac{x-\mu}{\beta} = \ln(-\ln y) \\ &\Leftrightarrow x = \mu - \beta \ln(-\ln y). \end{aligned}$$



Σχήμα 4. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Gumbel για σταθερή τιμή της παραμέτρου μ και ενδεικτικές τιμές της παραμέτρου β .



Σχήμα 5. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Gumbel για σταθερή τιμή της παραμέτρου β και ενδεικτικές τιμές της παραμέτρου μ .

Η μέση τιμή και η διακύμανση της κατανομής με αθροιστική συνάρτηση κατανομής την 52 θα δίνονται από τους παρακάτω τύπους,

$$E(X) = \mu + \beta\gamma, \quad (55)$$

$$Var(X) = \frac{\pi^2}{6}\beta^2. \quad (56)$$

Με γ συμβολίζεται η σταθερά Euler(Παράρτημα 7.1). Ο συντελεστής ασυμμετρίας της κατανομής Gumbel είναι ίσος με

$$skew_{\Lambda(x)} = \frac{12\sqrt{6}\zeta(3)}{\pi^3} \simeq 1.14 \quad (57)$$

όπου $\zeta(\cdot)$ είναι η συνάρτηση ζήτα του Riemann.

μ	β	$E(X)$	$Var(X)$	$skew_{\Lambda(x)}$
1	2	0.0378	0.0035	1.3614
1	2.5	0.0363	0.0026	1.0502
1	3	0.0348	0.0020	0.8194
1	4	0.0319	0.0013	0.4986
1	3	0.0348	0.0020	0.8194
2	3	0.0361	0.0020	0.7814
3.5	3	0.0374	0.0019	0.7633
4	3	0.0378	0.0019	0.7631

Πίνακας 3. Μέση τιμή, διακύμανση και ασυμμετρία της κατανομής Gumbel για διάφορες τιμές των παραμέτρων.

Η ροπογεννήτρια δίνεται συναρτήσει της συνάρτησης Γάμμα από τον τύπο,

$$MGF_{\Lambda(x)} = \Gamma(1 - \beta t)e^{\mu t}. \quad (58)$$

Οι συναρτήσεις $\zeta(\cdot)$ και $\Gamma(\cdot)$ αναφέρονται αναλυτικά στο Παράρτημα 7.2 7.3.

3.5.1 Εκτίμηση των παραμέτρων μ και β

Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν τη κατανομή Gumbel. Σύμφωνα με τη μέθοδο μέγιστης πιθανοφάνειας που αναφέραμε παραπάνω ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι,

$$l(\mu, \beta) = -n \ln(\beta) - \sum_{i=1}^n \left(\frac{x_i - \mu}{\beta} \right) - \sum_{i=1}^n e^{-\left(\frac{x_i - \mu}{\beta} \right)}. \quad (59)$$

Για να καταλήξουμε σε εκφράσεις που θα μας δίνουν τις παραμέτρους αρκεί να μεγιστοποιήσουμε την παραπάνω εξίσωση ως προς το διάνυσμα (μ, β) . Για να το επιτύχουμε αυτό υπολογίζουμε τις μερικές παραγώγους της Σχέσης 59 ως προς τις παραμέτρους (μ, β) ,

$$\frac{\partial l(\mu, \beta)}{\partial \mu} = \frac{1}{\beta} \left[n - \sum_{i=1}^n e^{-\left(\frac{x_i - \mu}{\beta} \right)} \right] \quad (60)$$

$$\frac{\partial l(\mu, \beta)}{\partial \beta} = \sum_{i=1}^n \left(\frac{x_i - \mu}{\beta^2} \right) - \frac{n}{\beta} - \sum_{i=1}^n \left(\frac{x_i - \mu}{\beta^2} \right) e^{-\left(\frac{x_i - \mu}{\beta} \right)} \quad (61)$$

και τις θέτουμε ίσες με το μηδέν. Οι εκτιμήτριες των παραμέτρων προκύπτουν επιλύοντας το παρακάτω ζεύγος εξισώσεων (βλ [39]),

$$\hat{\mu} = \hat{\beta} (\ln(n) - \ln \sum_{i=1}^n e^{-\frac{x_i}{\hat{\beta}}}), \quad (62)$$

$$\hat{\beta} = \bar{x} - \frac{\sum_{i=1}^n x_i e^{-\frac{x_i}{\hat{\beta}}}}{\sum_{i=1}^n e^{-\frac{x_i}{\hat{\beta}}}}. \quad (63)$$

Να σημειώσουμε ότι η εκτιμήτρια $\hat{\beta}$ υπολογίζεται απευθείας από τη Σχέση 63 και η εκτιμήτρια $\hat{\mu}$ προκύπτει αντικαθιστώντας την εκτιμήτρια $\hat{\beta}$ στη Σχέση 62 η οποία λύνεται με τη χρήση κάποιας αριθμητικής μεθόδου.

Χρησιμοποιώντας τη μέθοδο των σταθμισμένων ροπών πιθανότητας για την εκτίμηση των παραμέτρων θα έχουμε [35],

$$\beta_r = \int_0^1 Q_{\Lambda(x)}(p) \Lambda^r d\Lambda = \frac{1}{r+1} [\mu + \beta(\gamma + \ln(r+1))] \quad (64)$$

όπου γ είναι η σταθερά Euler, Λ είναι η αθροιστική συνάρτηση κατανομής της Gumbel και $Q_{\Lambda(x)}(p)$ είναι η αντίστροφη συνάρτηση κατανομής της. Θέτουμε $u = -\ln f$, $m = (r+1)u$ για να κάνουμε αλλαγή μεταβλητών και έπειτα από αριθμητικούς υπολογισμούς καταλήγουμε στις παρακάτω εκτιμήτριες,

$$\hat{\mu} = (r+1)\hat{\beta}_r - \hat{\beta}[\ln(r+1) + \gamma], \quad (65)$$

$$\hat{\beta} = \frac{(r+1)\hat{\beta}_r - (s+1)\hat{\beta}_s}{\ln(r+1)\ln(s+1)}. \quad (66)$$

όπου

$$\hat{\beta}_r = \frac{1}{n} \sum_{i=1}^n \frac{\binom{i-1}{r}}{\binom{n-1}{r}} x_{(i)}, \quad (67)$$

$$\hat{\beta}_s = \frac{1}{n} \sum_{i=1}^n \frac{\binom{i-1}{s}}{\binom{n-1}{s}} x_{(i)}. \quad (68)$$

για $r = 1, \dots, n$, $s = 1, \dots, n$ και $i = 1, \dots, n$.

Τέλος χρησιμοποιούμε τη μέθοδο ροπών για την εκτίμηση των παραμέτρων. Η συνήθης ροπή τάξης j δίνεται από τον τύπο,

$$\mu_j = E(X^j) = \sum_{k=0}^j \binom{k}{j} \beta^{j-k} \mu^k \int_{\mathbb{R}} x^{j-k} e^{-(x+e^{-x})} dx, \quad j = 1, 2, 3, \dots \quad (69)$$

Για $j = 1$ και $j = 2$ αντίστοιχα θα έχουμε,

$$\mu_1 = \mu + \beta\gamma \quad (70)$$

$$\mu_2 = \beta^2 J + 2\beta\mu\gamma + \mu^2 \quad (71)$$

όπου $\gamma = 0,577215$ και $J = \frac{\pi^2}{6} \simeq 1.978$. Οι εκτιμήτριες των παραμέτρων προκύπτουν από το παρακάτω σύστημα (βλ [39]),

$$\hat{\beta} = \left(\frac{\bar{\chi}^2 - \bar{\chi}^2}{J - \gamma^2} \right)^{\frac{1}{2}} \quad (72)$$

$$\hat{\mu} = \bar{\chi} - \gamma \hat{\beta} \quad (73)$$

όπου $\bar{\chi}$ και $\bar{\chi}^2$ είναι οι δειγματικές ροπές πρώτης και δεύτερης τάξης αντίστοιχα.

3.5.2 Σύγκριση μεθόδων εκτίμησης

Παρακάτω παρουσιάζουμε τα αποτελέσματα της σύγκρισης των μεθόδων μέγιστης πιθανοφάνειας, ροπών και σταθμισμένων ροπών πιθανότητας. Με τη βοήθεια του προγράμματος R δημιουργήσαμε 5000 σετ τυχαίων δειγμάτων χωρισμένα σε μικρά, μεσαία και μεγάλα δείγματα και εφαρμόζοντας τις παραπάνω μεθόδους για δύο ζευγάρια παραμέτρων καταλήξαμε στα ακόλουθα αποτελέσματα. Για το σετ παραμέτρων ($\mu = 0.5, \beta = 2$) και ως προς την παράμετρο θέσης, η αποδοτικότερη μέθοδος είναι η μέθοδος των ροπών για μικρά ($n = 10$) και μεσαία δείγματα $n = 50$, ενώ στα μεγαλύτερα μεγέθη δειγμάτων $n = 100, 1000$ η μέθοδος που επικρατεί είναι η μέθοδος μέγιστης πιθανοφάνειας. Ωσον αφορά στη μέθοδο σταθμισμένων ροπών πιθανότητας παρατηρούμε ότι είναι εξίσου αποδοτική και 'κοντά' στην μέθοδο μέγιστης πιθανοφάνειας αλλά εμφανίζει μέσο τετραγωνικό σφάλμα μια τάξη μεγέθους μεγαλύτερο από την μέθοδο μέγιστης πιθανοφάνειας. Ως προς την παράμετρο κλίμακας β , η μέθοδος που φαίνεται να είναι εξαιρετικά αποδοτική είναι η μέθοδος σταθμισμένων ροπών πιθανότητας για μικρά και μεσαία μεγέθη δειγμάτων. Για ακόμη μία φορά η μέθοδος μέγιστης πιθανοφάνειας επικρατεί στα μεγάλα σετ δεδομένων. Είναι ενδιαφέρον να πούμε ότι για τις τιμές παραμέτρων ($\mu = 3, \beta = 4$) λαμβάνουμε πολύ ανακριβή εκτίμηση για το μικρό μέγεθος δείγματος $n = 10$ για όλες τις μεθόδους. Από το τελευταίο μπορούμε να συμπεράνουμε ότι η φύση των παραμέτρων έχει ως αποτέλεσμα το φαινόμενο αυτό. Για τα υπόλοιπα μεγέθη δειγμάτων η εκτίμηση των παραμέτρων είναι σχετικά ακριβής ιδιαίτερα στη περίπτωση της μεθόδου ροπών πιθανότητας.

Για την πληρότητα της βιβλιογραφίας αναφέρουμε την παρακάτω έρευνα που συγκρίνει τις ίδιες μεθόδους για τιμές των παραμέτρων $\mu = 0$ και $\beta = 1$. Οι

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\mu = 0.5$	0.0799	0.443	0.0176	0.0899	0.005	0.043	0.002	0.0046
$\beta = 2$	-0.147	0.242	-0.032	0.048	-0.017	0.023	-0.0004	0.0023
$\mu = 3$	0.126	1.77	-0.024	0.347	0.017	0.179	0.0025	0.0173
$\beta = 4$	-0.298	0.938	-0.063	0.2015	-0.036	-0.0955	-0.0043	0.0095

Πίνακας 4. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου μέγιστης πιθανοφάνειας για την κατανομή Gumbel.

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\mu = 0.5$	0.0025	0.440	0.003	0.089	-0.001	0.043	0.001	0.0046
$\beta = 2$	0.0049	0.348	-0.004	0.064	-0.003	0.031	0.0006	0.003
$\mu = 3$	-0.028	0.834	-0.0046	0.348	-0.0035	0.180	-0.004	0.0179
$\beta = 4$	-0.0037	1.360	0.0027	0.2751	-0.008	0.1257	0.00016	0.0127

Πίνακας 5. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου ροπών πιθανότητας για την κατανομή Gumbel.

προσωμοιώσεις έγιναν σε δείγματα μεγέθους 5 έως 100 παρατηρήσεων. Τα συμπεράσματα που προέκυψαν από τη σύγκριση των μεθόδων μέγιστης πιθανοφάνειας, ροπών και σταθμισμένων ροπών πιθανότητας ήταν τα ακόλουθα (βλ [39]). Αρχικά, ως προς τη παράμετρο θέσης μ , η μέθοδος των σταθμισμένων ροπών πιθανότητας είναι αποτελεσματικότερη από τις μεθόδους μέγιστης πιθανοφάνειας και ροπών ενώ η μέθοδος μέγιστης πιθανοφάνειας είναι καλύτερη από τη μέθοδο ροπών για όλα τα μεγέθη δειγμάτων. Ως προς τη παράμετρο κλίμακας β , οι μέθοδοι μέγιστης πιθανοφάνειας και σταθμισμένων ροπών είναι εξίσου αποτελεσματικές και καλύτερες από τη μέθοδο ροπών για όλα τα δείγματα που μελετήθηκαν. Είναι σημαντικό να αναφέρουμε ότι στα δείγματα μικρού και μεσαίου μεγέθους η μέθοδος των σταθμισμένων ροπών πιθανότη-

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\mu = 0.5$	0.008	1.184	-0.002	0.240	-0.005	0.117	0.002	0.012
$\beta = 2$	0.004	0.240	-0.003	0.058	-0.004	0.045	0.0008	0.005
$\mu = 3$	-0.033	4.749	-0.0081	0.967	-0.007	0.477	0.004	0.046
$\beta = 4$	-0.0037	1.3635	-0.0027	0.2756	-0.0085	0.122	0.00017	0.0127

Πίνακας 6. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου σταθμισμένων ροπών πιθανότητας για την κατανομή *Gumbel*.

τας είναι εμφανώς αποτελεσματικότερη από τις άλλες δύο μεθόδους. Γενικά η μέθοδος των σταθμισμένων ροπών πιθανότητας είναι περισσότερο αξιόπιστη και δίνει πιο ακριβή αποτελέσματα. Αμέσως επόμενη στην κατάταξη έρχεται η μέθοδος μέγιστης πιθανοφάνειας. Να σημειώσουμε εδώ ότι για τη μέθοδο των σταθμισμένων ροπών τα παραπάνω αποτελέσματα ισχύουν για ακέραιες τιμές των r, s .

```

1
2 for (j in 1:iter)
3 {
4   data_gum = rgumbel(n, loc, scale)
5   # max lik
6
7   parameters_mle = fitdist(data_gum, "gumbel", "mle", start=list(
8     loc=1, scale=1))
9
10  estim_mle_gum_loc[j] = parameters_mle$estimate[1]
11  estim_mle_gum_sca[j] = parameters_mle$estimate[2]
12
13  ## method of moments
14
15  lmom_gum = lmoms(data_gum, 2)

```

```

16  par_gum_mom = pargum(lmom_gum, checklmom=TRUE)
17  loc_gum_lm[j] = par_gum_mom$para[1]
18  sca_gum_lm[j] = par_gum_mom$para[2]
19
20  ## probability weighted moment method
21  pwm_gum_gener = pwm(data_gum, 4)
22  sca_gum_pwm[j] = (2* pwm_gum_gener$betas[2] - pwm_gum_gener$
    betas[1])/log(2)
23  loc_gum_pwm[j] = (2* pwm_gum_gener$betas[2]) - (scale *euler) -
    (scale * log(2))
24
25  }
26
27  loc_mean = sum(estim_mle_gum_loc)/iter
28  sca_mean = sum(estim_mle_gum_sca)/iter
29  bias_loc = loc_mean - loc
30  bias_sca = sca_mean - scale
31  mse_loc = sum((estim_mle_gum_loc - loc_mean)^2)/iter
32  mse_sca = sum((estim_mle_gum_sca - sca_mean)^2)/iter

```

3.6 Εκτίμηση των παραμέτρων της κατανομής Fréchet

Η αθροιστική συνάρτηση κατανομής της κατανομής Fréchet με δύο παραμέτρους, θα είναι

$$\Phi(x; s, \alpha) = e^{-\left(\frac{x}{s}\right)^{-\alpha}}, \quad x > 0 \quad (74)$$

όπου s είναι η παράμετρος κλίμακας και α η παράμετρος σχήματος.

Παραγωγίζοντας τη Σχέση 74 θα λάβουμε τη συνάρτηση πυκνότητας πιθανότητας,

$$\phi(x) = \frac{\alpha}{s} \left(\frac{x}{s}\right)^{-1-\alpha} e^{-\left(\frac{x}{s}\right)^{-\alpha}}. \quad (75)$$

Παρακάτω παραθέτουμε ένα ενδεικτικό διάγραμμα που δίνει έμφαση στη παράμετρο σχήματος α , και πώς αυτή επηρεάζει την ουρά της κατανομής.

Η αντίστροφη συνάρτηση της αθροιστικής συνάρτησης κατανομής θα δίνεται από τον παρακάτω τύπο,

$$Q_{\Phi(x)}(p) = s(-\ln p)^{-\frac{1}{\alpha}}. \quad (76)$$

Διότι,

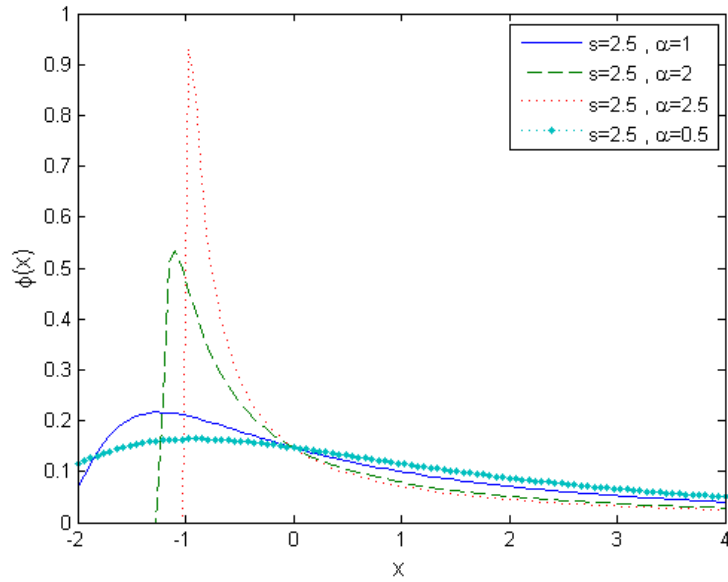
$$\begin{aligned} \Phi(x) = y &\Leftrightarrow e^{-\left(\frac{x}{s}\right)^{-\alpha}} = y \Leftrightarrow -\left(\frac{x}{s}\right)^{-\alpha} = \ln y \\ &\Leftrightarrow x = s\left(-\frac{1}{\ln y}\right)^{\frac{1}{\alpha}} \\ &\Leftrightarrow x = s(-\ln y)^{-\frac{1}{\alpha}}. \end{aligned}$$

Η μέση τιμή και η διακύμανση της κατανομής με αθροιστική συνάρτηση κατανομής την 74 θα δίνονται από τους παρακάτω τύπους

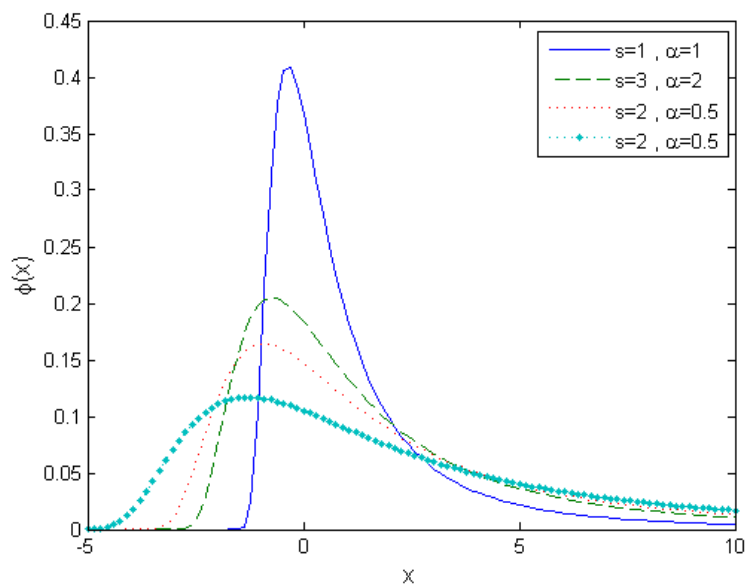
$$E(X) = s\Gamma\left(1 - \frac{1}{\alpha}\right), \quad \alpha > 1, \quad (77)$$

$$Var(X) = s^2\left(\Gamma\left(1 - \frac{2}{\alpha}\right) - \left(\Gamma\left(1 - \frac{1}{\alpha}\right)\right)^2\right), \quad \alpha > 2. \quad (78)$$

Για τιμές του $\alpha \leq 1$ και $\alpha \leq 2$, η μέση τιμή και η διακύμανση αντίστοιχα ισούνται με άπειρο. Ο συντελεστής ασυμμετρίας της κατανομής είναι ανεξάρτητος από τις παραμέτρους της κατανομής και είναι ίσος με



Σχήμα 6. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Fréchet για σταθερή τιμή της παραμέτρου s και ενδεικτικές τιμές της παραμέτρου α .



Σχήμα 7. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Fréchet για σταθερή τιμή της παραμέτρου α και ενδεικτικές τιμές της παραμέτρου s .

$$skew_{\Phi(x)} = \frac{\Gamma(1 - \frac{3}{\alpha}) - 3\Gamma(1 - \frac{2}{\alpha})\Gamma(1 - \frac{1}{\alpha}) + 2\Gamma^3(1 - \frac{1}{\alpha})}{\sqrt{(\Gamma(1 - \frac{2}{\alpha}) - \Gamma^2(1 - \frac{1}{\alpha}))^3}}. \quad (79)$$

α	s	$E(X)$	$Var(X)$	$skew_{\Phi(x)}$
1	2.5	0.1117	0.0033	0.4908
2	2.5	0.1017	0.0141	2.0265
2.5	2.5	0.0993	0.0258	3.3031
0.5	2.5	0.1117	0.0014	-0.1123
0.5	1	0.1478	0.0169	0.7336
0.5	2	0.1261	0.0027	0.1059
0.5	2.5	0.1117	0.0014	-0.1123
0.5	3.5	0.0876	$5.2920 \cdot 10^{-4}$	-0.1672

Πίνακας 7. Μέση τιμή, διακύμανση και ασυμμετρία της κατανομής Fréchet για διάφορες τιμές των παραμέτρων.

3.6.1 Εκτίμηση των παραμέτρων s και α

Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την κατανομή Fréchet. Σύμφωνα με τη μέθοδο μέγιστης πιθανοφάνειας, ο λογάριθμος της συνάρτησης πιθανοφάνειας θα δίνεται από τον τύπο

$$l(s, \alpha) = n \ln \alpha + n\alpha \ln s - (1 + \alpha) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left(\frac{s}{x_i}\right)^\alpha. \quad (80)$$

Για να βρούμε τις εκτιμήτριες για τις παραμέτρους s και α αρκεί να παραγωγίσουμε τη Σχέση 80 ως προς τη κάθε παράμετρο ξεχωριστά και θέσουμε τις εξισώσεις που προκύπτουν ίσες με το μηδέν. Συνεπώς, καταλήγουμε στο παρακάτω σύστημα εξισώσεων,

$$\frac{\partial l(s, \alpha)}{\partial \alpha} = \frac{n}{\alpha} + n \ln s - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left(\frac{s}{x_i}\right)^\alpha \ln \frac{s}{x_i} = 0 \quad (81)$$

$$\frac{\partial l(s, \alpha)}{\partial s} = \frac{n\alpha}{s} - \alpha \sum_{i=1}^n s^{\alpha-1} \left(\frac{1}{x_i}\right)^\alpha = 0 \quad (82)$$

Οι παραπάνω εξισώσεις λύνονται με χρήση κάποιας αριθμητικής μεθόδου για παράδειγμα του αλγορίθμου Newton-Raphson.

Για τη μέθοδο ροπών πιθανότητας αρκεί να εξισώσουμε την θεωρητική ροπή τρίτης τάξης με τη τιμή του δειγματικής ροπής και να λύσουμε ως προς τη παράμετρο σχήματος χρησιμοποιώντας μια επαναληπτική αριθμητική μέθοδο δηλαδή,

$$\frac{\sqrt{\Gamma(1 - \frac{2}{\alpha}) - \Gamma^2(1 - \frac{1}{\alpha})}}{\Gamma(1 - \frac{1}{\alpha})} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}, \quad (83)$$

όπου $\bar{y} = \sum_{i=1}^n \frac{1}{n} x_i$. Χρησιμοποιώντας τη μέθοδο σταθμισμένων ροπών πιθανότητας θα έχουμε

$$\beta_r = E(y\Phi^r(y)) = \int_{-\infty}^{\infty} y\Phi^r(y)\phi(y)dy = \int_0^1 y(\Phi)\Phi^r(y)d\Phi, \quad (84)$$

όπου Φ είναι η συνάρτηση κατανομής της κατανομής Frechét και

$$b_r = \frac{1}{n \binom{n-1}{r}} \sum_{i=1}^n \binom{i-1}{r}, \quad (85)$$

όπου $y_{(1)} \leq y_{(2)} \dots \leq y_{(n)}$ διατεταγμένη σειρά του δείγματος y_1, y_2, \dots, y_n .

Μετά από μερικούς αριθμητικούς υπολογισμούς η παραπάνω σχέση γίνεται:

$$\beta_r = \frac{1}{r+1} \left(\frac{1}{s}(r+1)\right)^{\frac{1}{\alpha}} \Gamma\left(1 - \frac{1}{\alpha}\right). \quad (86)$$

Για $r = 0, 1$ λαμβάνουμε το παρακάτω σύστημα εξισώσεων

$$\beta_0 = \left(\frac{1}{s}\right)^{1/\alpha} \Gamma\left(1 - \frac{1}{\alpha}\right), \quad (87)$$

$$\beta_1 = \left(\frac{2}{s}\right)^{1/\alpha} \frac{1}{2} \Gamma\left(1 - \frac{1}{\alpha}\right). \quad (88)$$

Όμως ισχύει ότι,

$$b_0 = \sum_{i=1}^n \frac{y_i}{n}, \quad (89)$$

$$b_1 = \sum_{i=1}^n \frac{(i-1)y_{(i)}}{n(n+1)}. \quad (90)$$

Εξισώνοντας τις ποσότητες β_0 με b_0 και β_1 με b_1 λαμβάνουμε τις εκτιμήτριες των παραμέτρων s και α ως εξής:

$$\hat{\alpha} = \left(\frac{\ln b_1 - \ln b_0}{\ln 2} + 1 \right)^{-1}, \quad (91)$$

$$\hat{s} = \left(\frac{b_0}{\Gamma(1 - \frac{1}{\hat{\alpha}})} \right)^{-\hat{\alpha}}. \quad (92)$$

3.6.2 Σύγκριση μεθόδων εκτίμησης

Επειδή δεν υπάρχει κάποια μελέτη σύγκρισης των παραπάνω μεθόδων για την κατανομή Frechét συγκρίνουμε τις τρεις μεθόδους μέσω κώδικα που αναπτύξαμε στην R. Για την εκτίμηση της παραμέτρου σχήματος, δημιουργήσαμε 5000 σετ τυχαίων δεδομένων για μεγέθη δείγματος $n = 10, 50, 100, 1000$ και τιμές της παραμέτρου $\alpha = 1, 3, 5, 7$. Για να συγκρίνουμε τα αποτελέσματά μας χρησιμοποιήσαμε το μέσο τετραγωνικό σφάλμα και την αμεροληψία. Παρακάτω παραθέτουμε αναλυτικούς πίνακες για τις τρεις μεθόδους εκτίμησης και σχόλια για την αποδοτικότητα των μεθόδων.

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\alpha = 3$	0.513	1.099	0.0807	0.121	0.0399	0.058	0.004	0.005
$\alpha = 5$	0.836	2.973	0.144	0.340	0.073	0.157	0.006	0.014
$\alpha = 7$	1.210	6.308	0.193	0.682	0.089	0.323	0.007	0.030

Πίνακας 8. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου μέγιστης πιθανοφάνειας για την κατανομή Frechét.

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\alpha = 3$	0.517	1.257	0.120	0.227	0.067	0.118	0.008	0.014
$\alpha = 5$	0.643	3.381	0.141	0.541	0.075	0.267	0.007	0.026
$\alpha = 7$	0.828	6.539	0.157	0.962	0.071	0.495	0.005	0.049

Πίνακας 9. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου σταθμισμένων ροπών πιθανότητας για την κατανομή Frechét.

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\alpha = 3$	0.627	1.075	0.024	0.953	0.012	0.937	0.001	0.917
$\alpha = 5$	-0.312	0.254	-0.045	0.433	-0.023	0.449	-0.002	0.467
$\alpha = 7$	-0.373	1.733	-0.063	1.872	-0.033	1.890	-0.003	1.905

Πίνακας 10. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου ροπών πιθανότητας για την κατανομή Frechét.

Για τιμή της παραμέτρου $\alpha = 3$ και το δείγμα μεγέθους $n = 10$ η μέθοδος μέγιστης πιθανοφάνειας είναι η πιο αποδοτική με μικρή διαφορά από τη μέθοδο σταθμισμένων ροπών πιθανότητας. Πρέπει να τονίσουμε ωστόσο ότι και οι δύο μέθοδοι εμφανίζουν σχετικά υψηλό μέσο τετραγωνικό σφάλμα σε σχέση με τη μέθοδο ροπών. Για το δείγμα μεσαίου μεγέθους η μέθοδος μέγιστης πιθανοφάνειας κυριαρχεί εμφανίζοντας μεγάλη απόκλιση στην αμεροληψία σε σχέση με τις άλλες δύο μεθόδους ενώ παράλληλα έχει και το χαμηλότερο μέσο τετραγωνικό σφάλμα. Για τα μεγέθη δειγμάτων $n = 100, 1000$ η μέθοδος μέγιστης πιθανοφάνειας είναι με διαφορά η πιο αποδοτική σε όρους αμεροληψίας και μέσου σφάλματος. Καθώς η τιμή της παραμέτρου σχήματος αυξάνει ($\alpha = 5, 7$) η μέθοδος σταθμισμένων ροπών πιθανότητας είναι σαφώς καλύτερη στο μικρό δείγμα, $n = 10$. Αποτελεί ενδιαφέρον ωστόσο ότι σχεδόν διπλασιάζεται όχι μόνο μέθοδο σταθμισμένων ροπών αλλά και στη μέθοδο πιθανοφάνειας.

Η εικόνα αλλάζει καθώς το μέγεθος του δείγματος αυξάνεται με τη μέθοδο μέγιστης πιθανοφάνειας να φαίνεται ως η πιο αποδοτική για τα υπόλοιπα μεγέθη δειγμάτων για τιμή της παραμέτρου σχήματος ίση με 5 ενώ για $\alpha = 7$ η μέθοδος σταθμισμένων ροπών πιθανότητας είναι η πιο κατάλληλη σε όρους αμεροληψίας. Η μέθοδος ροπών είναι η πιο αδύναμη για όλες τις παραπάνω τιμές της παραμέτρου και όλα τα δείγματα. Όπως είναι γνωστό, οι ροπές της κατανομής Frechét υπάρχουν και μπορούν να υπολογιστούν μόνο όταν η τιμή της της παραμέτρου σχήματος είναι αυστηρά μεγαλύτερη από τη τάξη ροπής που θέλουμε να υπολογίσουμε (βλ [45]). Η μέθοδος που και πάλι επικρατεί για τα δείγματα $n = 10, 100$ είναι η μέθοδος σταθμισμένων ροπών πιθανότητας ενώ η μέθοδος μέγιστης πιθανοφάνειας είναι αποδοτικότερη στα μεγαλύτερα μεγέθη δειγμάτων.

```

1 iter=5000
2 shape=7
3 for (i in 1:iter)
4 {
5   data_fre = rfrechet(n, location=0, shape, scale=1)
6   #### maximum likelihood
7   x= fitdist(data_fre, "invweibull", "mle", start=list(shape=3,
8     scale=1)) # fitdist is an object
9   estim_mle_fre[i] = x$estimate[1]
10
11  #### moments
12  mom_fre = lmoms(data_fre, 2)
13  cv = sqrt(mom_fre$lambda[2])/mom_fre$lambda[1]
14  func = function(z) ((sqrt(gamma(1-2/z)) - (gamma(1-1/z)^2)) /
15    gamma(1-1/z)) - cv
16  alpha_estimations = uniroot(func, c(-100000, 100000))
17  alpha_fre_lm[i]= alpha_estimations$root[1]
18
19  ### weighted moments
20  pwm_fre = pwm(data_fre, 2)
21  alpha_fre_pwm[i] = (((log(pwm_fre$betas[2]) - log(pwm_fre$
22    betas[1])) / log(2)) + 1) ^ -1

```

```
20 }
21
22 alpha_mean = sum(estim_mle_fre)/iter
23 bias_alpha = alpha_mean - shape
24 mse_alpha = sum((estim_mle_fre - alpha_mean)^2)/iter
25 alpha_fre_lm_mean = sum(alpha_fre_lm)/iter
26 bias_alpha_lm = alpha_fre_lm_mean - shape
27 mse_alpha_lm = sum((alpha_fre_lm - alpha_fre_lm_mean)^2)/iter
28 alpha_fre_pwm_mean = sum(alpha_fre_pwm)/iter
29 bias_alpha_pwm = alpha_fre_pwm_mean - shape
30 mse_alpha_pwm = sum((alpha_fre_pwm - alpha_fre_pwm_mean)^2)/iter
```

3.7 Εκτίμηση των παραμέτρων της κατανομής Weibull

Η αθροιστική συνάρτηση κατανομής της Weibull θα δίνεται από τον τύπο

$$\Psi(x; \sigma, \alpha) = 1 - e^{-\left(\frac{x}{\sigma}\right)^\alpha}, \quad x \geq 0 \quad (93)$$

όπου σ είναι η παράμετρος κλίμακας και α είναι η παράμετρος σχήματος.

Παραγωγίζοντας τη Σχέση 93 θα λάβουμε τη συνάρτηση πυκνότητας πιθανότητας, $\psi(x)$,

$$\psi(x) = \frac{\alpha}{\sigma} \left(\frac{x}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{x}{\sigma}\right)^\alpha}, \quad x > 0 < \infty. \quad (94)$$

Η αντίστροφη συνάρτηση κατανομής της αθροιστικής κατανομής θα δίνεται από τον τύπο

$$Q_{\Psi(x)}(p) = -\sigma(\ln(1-p))^{\frac{1}{\alpha}}, \quad 0 < p < 1, \quad (95)$$

διότι,

$$\begin{aligned} \Psi(x) = y &\Leftrightarrow 1 - e^{-\left(\frac{x}{\sigma}\right)^\alpha} = y \Leftrightarrow e^{-\left(\frac{x}{\sigma}\right)^\alpha} = 1 - y \\ &\Leftrightarrow -\left(\frac{x}{\sigma}\right)^\alpha = (\ln(1-y))^{\frac{1}{\alpha}} \\ &\Leftrightarrow x = -\sigma(\ln(1-y))^{\frac{1}{\alpha}}. \end{aligned}$$

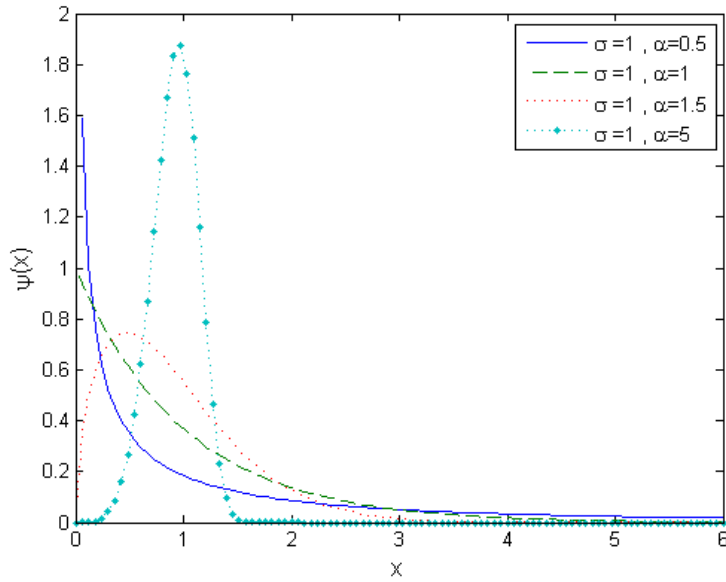
Η μέση τιμή και η διακύμανση της κατανομής με αθροιστική συνάρτηση κατανομής την 93 καθώς και ο συντελεστής ασυμμετρίας της κατανομής Weibull θα δίνονται από τους παρακάτω τύπους

$$E(X) = \sigma\Gamma\left(1 + \frac{1}{\alpha}\right), \quad (96)$$

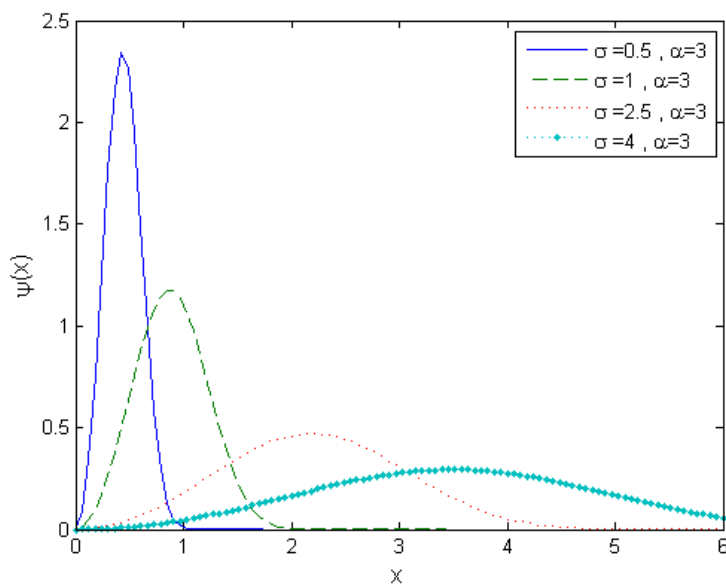
$$Var(X) = \sigma^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right], \quad (97)$$

$$skew_{\Psi(X)} = \frac{\Gamma\left(1 + \frac{3}{\alpha}\right)Var^3 - 3E(X)Var^2 - E(X)^3}{Var^3}. \quad (98)$$

Είναι ενδιαφέρον να αναφέρουμε ότι όταν η παράμετρος σχήματος α είναι μεγάλη τότε η μέση τιμή και η διακύμανση της κατανομής δίνονται προσεγγιστικά από



Σχήμα 8. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull για σταθερή τιμή της παραμέτρου σ και για ενδεικτικές τιμές της παραμέτρου α .



Σχήμα 9. Συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull για σταθερή τιμή της παραμέτρου α και για ενδεικτικές τιμές της παραμέτρου σ .

τους παρακάτω τύπους (βλ [41])

$$E_{\alpha \gg 1}(X) \approx 1 - \frac{\gamma}{\alpha} + \frac{\pi^2/6 + \gamma^2}{2\alpha^2}, \quad (99)$$

$$Var_{\alpha \gg 1}(x) \approx \frac{\pi^2}{6\alpha^2}. \quad (100)$$

Για $\alpha > 3.6$ η συμπεριφορά της κατανομής προσεγγίζει την κανονική κατανομή και μάλιστα ο συντελεστής ασυμμετρίας ισούται με 0 (βλ [44]).

α	σ	$E(X)$	$Var(X)$	$skew_{\Psi(x)}$
0.5	1	0.0431	0.0354	8.1547
1	1	0.0426	0.0212	4.4157
1.5	1	0.0402	0.0196	3.8370
5	1	0.0396	0.0521	6.3812
3	0.5	0.0394	0.0661	6.7839
3	1	0.0396	0.0325	4.9814
3	2.5	0.0396	0.0121	2.8725
3	4	0.0396	0.0069	2.0413

Πίνακας 11. Μέση τιμή, διακύμανση και ασυμμετρία της κατανομής Weibull για διάφορες τιμές των παραμέτρων.

Η ροπογεννήτρια δίνεται συναρτήσει της συνάρτησης Γάμμα από τον τύπο

$$MGF_{\Psi(x)} = \sum_{n=0}^{\infty} \frac{(t\sigma)^n}{n!} \Gamma\left(1 + \frac{n}{\alpha}\right). \quad (101)$$

3.7.1 Εκτίμηση των παραμέτρων σ και α

Σύμφωνα με τη μέθοδο μεγιστής πιθανοφάνειας ο λογάριθμος της συνάρτησης πιθανοφάνειας θα είναι

$$l(\sigma, \alpha) = n \ln \alpha + (\alpha - 1) \sum_{i=1}^n \ln x_i - \alpha n \ln \sigma - \sum_{i=1}^n \left(\frac{x_i}{\sigma}\right)^\alpha. \quad (102)$$

Για να καταλήξουμε στις εκτιμήτριες των παραμέτρων αρκεί να παραγωγίσουμε τη Σχέση 102 ως προς κάθε μια από τις παραμέτρους και να θέσουμε ίσες με το μηδέν τις δύο εξισώσεις που θα προκύψουν. Έτσι καταλήγουμε στο ακόλουθο σύστημα εξισώσεων,

$$\frac{\partial l(\sigma, \alpha)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \ln x_i - \frac{1}{\sigma} \sum_{i=1}^n x_i^\alpha \ln x_i = 0, \quad (103)$$

$$\frac{\partial l(\sigma, \alpha)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n x_i^\alpha = 0. \quad (104)$$

Η λύση του παρακάτω συστήματος μας δίνει τις εκτιμήτριες $\hat{\alpha}$, $\hat{\sigma}$ για τις παραμέτρους α , σ .

$$\frac{1}{\hat{\alpha}} = \frac{\sum_{i=1}^n x_i^{\hat{\alpha}} \ln x_i}{\sum_{i=1}^n x_i^{\hat{\alpha}}} - \frac{1}{n} \sum_{i=1}^n \ln x_i, \quad (105)$$

$$\hat{\sigma} = \frac{1}{n} \left(\sum_{i=1}^n x_i^{\hat{\alpha}} \right)^{\hat{\alpha}}. \quad (106)$$

Παρατηρούμε ότι η εύρεση των εκτιμητριών σε κλειστή μορφή δεν είναι δυνατή. Ωστόσο, μπορούμε να εκτιμήσουμε τη παράμετρο α με τη χρήση κάποιας αριθμητικής μεθόδου και έπειτα μέσω της Σχέσης 106 να υπολογίσουμε τη παράμετρο σ . Η μέθοδος μέγιστης πιθανοφάνειας χρησιμοποιείται ευρέως για την εκτίμηση των παραμέτρων της κατανομής Weibull αλλά έχουν παρατηρηθεί προβλήματα στη σύγκλιση των αριθμητικών μεθόδων (βλ [56]).

Οι παράμετροι της κατανομής Weibull, δε μπορούν να εκφραστούν ρητά από τις συνηθισμένες ροπές. Ωστόσο, μπορούν να υπολογισθούν χρησιμοποιώντας κεντρικές ροπές. Η ροπή τάξης j δίνεται από τον τύπο [10]

$$\mu_j = \sigma^j \sum_{k=0}^j (-1)^k \binom{j}{k} \Gamma\left(1 + \frac{j-k}{\alpha}\right) \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^k, \quad j = 1, 2, 3... \quad (107)$$

Η μέση τιμή και οι ροπές τάξης δύο και τρία, δίνονται συναρτήσει των παραμέ-

των της κατανομής Weibull ως εξής:

$$\mu = \sigma\Gamma_1, \quad (108)$$

$$\mu_2 = \sigma^2(\Gamma_2 - \Gamma_1^2), \quad (109)$$

$$\mu_3 = \sigma^3(\Gamma_3 - 3\Gamma_1\Gamma_2 + 2\Gamma_1^3), \quad (110)$$

όπου $\Gamma_k = \Gamma(1 + \frac{k}{\alpha})$. Παρουσιάζει ενδιαφέρον ότι ο συντελεστής ασυμμετρίας της της κατανομής είναι αναλλοίωτος στο μετασχηματισμό των δεδομένων ως προς τη θέση και τη κλίμακα. Όπως φαίνεται παρακάτω, δεν εξαρτάται από τις παραμέτρους θέσης και κλίμακας δηλαδή,

$$sk = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \sqrt{\frac{(\Gamma_3 - 3\Gamma_1\Gamma_2 + 2\Gamma_1^3)^2}{(\Gamma_2 - \Gamma_1^2)^3}}. \quad (111)$$

Η παράμετρος σχήματος, α , μπορεί να υπολογισθεί βρίσκοντας τις λύσεις της εξίσωσης 111 αντικαθιστώντας το αριστερό σκέλος με το δειγματικό της ανάλυσης. Έπειτα μπορούμε να εκτιμήσουμε τις υπόλοιπες παραμέτρους από τις παραπάνω σχέσεις.

Είναι ενδιαφέρον να σημειώσουμε ότι αντιθέτως με τη μέθοδο των ροπών, στη μέθοδο σταθμισμένων ροπών πιθανότητας, οι παράμετροι της κατανομής Weibull μπορούν να εκφραστούν ρητά συναρτήσει των συνήθων ροπών. Ακολουθώντας τη μέθοδο και θέτοντας για ευκολία $M_r = M_{1,0,r}$ έχουμε [24]:

$$M_r = \sigma \frac{\Gamma(1 + \frac{1}{\alpha})}{(1+r)^{1+\frac{1}{\alpha}}}. \quad (112)$$

Έτσι θα έχουμε τις παρακάτω εκτιμήτριες,

$$\hat{\alpha} = \frac{\ln 2}{\ln(\frac{M_0}{2M_1})}, \quad (113)$$

$$\hat{\sigma} = \frac{M_0}{\Gamma(\frac{\ln(\frac{M_0}{M_1})}{\ln 2})}. \quad (114)$$

Επειδή δεν υπάρχει μια μελέτη που να συγκρίνει τις μεθόδους μέγιστης πιθανοφάνειας, ροπών και σταθμισμένων ροπών πιθανότητας, θα εισαγάγουμε

μια μέθοδο που χρησιμοποιείται συνήθως για την εκτίμηση των παραμέτρων της κατανομής Weibull.

Η μέθοδος ονομάζεται μέγιστο γινόμενο διαστημάτων (maximum product of spacing (mps)) και βασίζεται στη μεγιστοποίηση της γεωμετρικής μέσης τιμής ισόνομων και ανεξάρτητα κατανεμημένων διαστημάτων ενός δείγματος y_i όπου

$$y_i = \Psi(x_i) = 1 - e^{-\left(\frac{x_i}{\sigma}\right)^\alpha} \quad (115)$$

για διατεταγμένο δείγμα $x_1 < x_2 < \dots < x_n$ το οποίο προέρχεται από την κατανομή Weibull (βλ [8]). Το διάστημα ορίζεται ως εξής

$$D_i = y_i - y_{i-1} = \int_{x_{i-1}}^{x_i} \psi(x; \sigma, \alpha) dx, \quad i = 1, 2, \dots, n+1, \quad (116)$$

όπου $x_0 = \Psi^{-1}(0)$, $x_{n+1} = \Psi^{-1}(1)$ και προφανώς $\sum_{i=1}^{n+1} D_i = 1$. Τα σημεία x_0, x_{n+1} είναι τα άκρα του διαστήματος στο οποίο λαμβάνει τιμές το παραπάνω ολοκλήρωμα. Στη πραγματικότητα μπορούν να λάβουν οποιεσδήποτε τιμές έτσι ώστε $F(x_0) = 0$ και $F(x_{n+1}) = 1$, άρα θα μπορούσαν να οριστούν ως $(-\infty, \infty)$. Σύμφωνα με τα παραπάνω στη μέθοδο αυτή αρκεί να μεγιστοποιήσουμε τη γεωμετρική μέση τιμή δηλαδή,

$$G = \left\{ \prod_{i=1}^{n+1} D_i \right\}^{\frac{1}{n+1}} \quad (117)$$

ή ισοδύναμα να μεγιστοποιήσουμε το λογάριθμο της Σχέσης 117 δηλαδή,

$$h = \ln G. \quad (118)$$

Μετά απο αρκετούς αριθμητικούς υπολογισμούς οι εκτιμήτριες των παραμέτρων της κατανομής Weibull δίνονται από τις παρακάτω εξισώσεις,

$$\frac{\partial h}{\partial \sigma} = \frac{\sigma}{\alpha} \frac{1}{n+1} \sum_{i=1}^n (S_i + Y_i) = 0, \quad (119)$$

$$\frac{\partial h}{\partial \alpha} = \frac{1}{n+1} \sum_{i=1}^n (T_i - \alpha^{-1} Y_i \ln Y_i) = 0, \quad (120)$$

όπου

$$Y_i = \left(\frac{x_i}{\sigma}\right)^\alpha, \quad (121)$$

$$Z_i = 1 - e^{Y_i - Y_{i-1}}, \quad (122)$$

$$S_i = \frac{Y_i - Y_{i-1}}{Z_i}, \quad (123)$$

$$T_i = \frac{Y_i \ln Y_i - Y_{i-1} \ln Y_{i-1}}{\alpha Z_i}. \quad (124)$$

3.7.2 Σύγκριση μεθόδων εκτίμησης

Σε αριθμητική εφαρμογή που αναπτύξαμε στο Matlab συγκρίνουμε τις μεθόδους μέγιστης πιθανοφάνειας, ροπών και σταθμισμένων ροπών πιθανότητας. Για την εκτίμηση της παραμέτρου σχήματος, δημιουργήσαμε 5000 σετ τυχαίων δεδομένων με απλό αλγόριθμο Monte Carlo για μεγέθη δείγματος $n = 10, 50, 100, 1000$ και τιμές της παραμέτρου $\alpha = 0.5, 1, 1.5, 5$. Για να συγκρίνουμε τα αποτελέσματά μας χρησιμοποιήσαμε το μέσο τετραγωνικό σφάλμα και την αμεροληψία. Παρακάτω παραθέτουμε αναλυτικούς πίνακες για τις τρεις μεθόδους εκτίμησης.

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\alpha = 0.5$	0.085	0.039	0.014	0.003	0.005	0.001	$2.3 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$
$\alpha = 1$	0.162	0.141	0.029	0.014	0.006	0.014	0.001	$6.1 \cdot 10^{-4}$
$\alpha = 1.5$	0.252	0.352	0.040	0.031	0.021	0.015	0.002	0.001
$\alpha = 5$	0.846	3.645	0.148	0.358	0.069	0.168	0.006	0.015

Πίνακας 12. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου μέγιστης πιθανοφάνειας για την κατανομή Weibull.

Παρατηρούμε ότι η μέθοδος σταθμισμένων ροπών πιθανότητας είναι αποδοτικότερη για όλα τα μεγέθη δειγμάτων και για όλες τις τιμές της παραμέτρου σχήματος α . Για δείγματα πλήθους 1000 η μέθοδος μέγιστης πιθανοφάνειας

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\alpha = 0.5$	0.054	0.036	0.013	0.005	0.007	0.002	$2.7 \cdot 10^{-4}$	$2.9 \cdot 10^{-4}$
$\alpha = 1$	0.062	0.112	0.004	0.015	0.002	0.007	$7.1 \cdot 10^{-4}$	$6.9 \cdot 10^{-4}$
$\alpha = 1.5$	0.068	0.252	0.008	0.030	0.004	0.014	$8.9 \cdot 10^{-4}$	0.001
$\alpha = 5$	0.362	2.882	0.065	0.383	0.037	0.183	0.002	0.017

Πίνακας 13. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου σταθμισμένων ροπών πιθανότητας για την κατανομή Weibull.

	$n = 10$		$n = 50$		$n = 100$		$n = 1000$	
	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>	<i>Bias</i>	<i>MSE</i>
$\alpha = 0.5$	0.050	0.035	0.013	0.005	0.007	0.002	$9.8 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
$\alpha = 1$	0.060	0.108	0.007	0.014	0.005	0.007	0.001	$6.7 \cdot 10^{-4}$
$\alpha = 1.5$	0.079	0.230	0.008	0.031	0.005	0.014	0.001	0.001
$\alpha = 5$	0.402	3.212	0.056	0.373	0.022	0.173	0.005	0.017

Πίνακας 14. Αμεροληψία και μέσο τετραγωνικό σφάλμα της μεθόδου ροπών για την κατανομή Weibull.

είναι περισσότερο ακριβής ιδιαίτερα για τιμές της παραμέτρου μικρότερες της μονάδας. Πιο συγκεκριμένα για $\alpha = 0.5$ η μέθοδος ροπών είναι η βέλτιστη εκτιμήτρια όταν πρόκειται για μικρά δείγματα ενώ η εκτιμήτρια σταθμισμένων ροπών φαίνεται να είναι πιο ακριβής για δείγματα πλήθους 100 και 1000. Όταν η τιμή της παραμέτρου ισούται με τη μονάδα η μέθοδος ροπών εξακολουθεί να επικρατεί στα μικρά μεγέθη δειγμάτων ενώ και πάλι η εκτιμήτρια των σταθμισμένων ροπών πιθανότητας είναι πιο αποδοτική στα μεγαλύτερα δείγματα.

Σε πρόσφατη έρευνα [2] γίνεται σύγκριση των έξι συνηθέστερων μεθόδων που χρησιμοποιούνται για την εκτίμηση των παραμέτρων της κατανομής Weibull μεταξύ των οποίων είναι οι μέθοδοι μέγιστης πιθανοφάνειας, ροπών και μέγιστου γινόμενου διαστημάτων. Η σύγκριση των μεθόδων έγινε σε διάφορα με-

γένη δειγμάτων μεγέθους n , ($n = 10, 20, 50, 100$) τα οποία προέρχονται από αναλογιστικά και χρηματοοικονομικά δεδομένα αλλά και από δεδομένα ποιοτικού ελέγχου και μηχανικής στα οποία εμφανίζεται συνήθως η κατανομή Weibull. Έπειτα χρησιμοποιήθηκε απλός αλγόριθμος Monte Carlo για τη δημιουργία 5000 σετ δεδομένων για κάθε μέγεθος δείγματος. Πιο συγκεκριμένα η παράμετρος σχήματος, α , έλαβε τις τιμές $\alpha = 0.5, 0.75, 1, 1.5, 3, 4, 6, 9$ με τις τιμές 1.5, 3, 4, 6, 9 να αντιστοιχούν σε αναλογιστικά και χρηματοοικονομικά δεδομένα. Η αποδοτικότητα των μεθόδων εκτίμησης εκτιμήθηκε χρησιμοποιώντας τη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) δηλαδή

$$RMSE = \sqrt{Var + Bias^2}. \quad (125)$$

Είναι σαφές ότι για μια αμερόληπτη εκτιμήτρια η ρίζα του μέσου τετραγωνικού σφάλματος ισούται απλά με το μέσο τετραγωνικό σφάλμα. Συγκρίνοντας τις μεθόδους μέγιστης πιθανοφάνειας, ροπών και μέγιστου γινομένου διαστημάτων κατέληξαν στα παρακάτω αποτελέσματα (βλ [2]).

Γενικά η μέθοδος μέγιστης πιθανοφάνειας είναι εξαιρετικά αποδοτική όταν η παράμετρος σχήματος είναι στο διάστημα $[0.5, 4]$ ακόμη και σε μικρά δείγματα πλήθους $n = 10$. Ωστόσο η μέθοδος μέγιστου γινομένου διαστημάτων είναι καλύτερη όταν η παράμετρος σχήματος παίρνει τιμές στο διάστημα $[0.5, 1.5]$. Για μεγάλες τιμές της παραμέτρου δηλαδή για $\alpha \geq 6$ η μέθοδος ροπών είναι η αποδοτικότερη μέθοδος όσον αφορά δείγματα πλήθους $n \leq 20$. Να σημειώσουμε εδώ ότι για $\alpha = 3$ η μέθοδος ροπών παρουσιάζει μικρότερη τιμή στη ρίζα του μέσου τετραγωνικού σφάλματος, το οποίο ήταν μη-αναμενόμενο, παρόλα αυτά όμως διπλασιάζει τη μεροληψία (Bias). Ως πληρότητα βιβλιογραφίας παραπέμπουμε τον αναγνώστη σε έρευνα σύγκρισης 5 διαφορετικών μεθόδων για την κατανομή Weibull συμπεριλαμβανομένων των μεθόδων μέγιστης πιθανοφάνειας και ροπών [53].

```

1
2 clc; clear;
3 iter = 5000;
4 sigma_wei = 1;

```

```

5 alpha_wei = 5;
6 %--Likelihood estimation
7 for n=[10 50 100 1000]
8
9 sumestimation=0;
10 estim_mle = zeros(2,iter);
11 %need to uncomment matlabpool
12 %matlabpool 4
13     for i=1:iter
14
15         randomwei = wblrnd(sigma_wei,alpha_wei,n,1);
16         % wblrnd(scale,shape,M,N) MxN array
17         estim_mle(:,i) = mle(randomwei, 'Distribution', 'wbl');
18
19 %----method of moments
20
21         m2=moment(randomwei,2);
22         m3=moment(randomwei,3);
23         sk_sample = sqrt(m3^2/m2^3);
24         estim_mom(i) = solve( sqrt(((gamma(1+3/a)-3*gamma(1+1/a)
25             ) * gamma(1+2/a) + ...
26                 2*(gamma(1+1/a))^2)^2) / ((gamma(1+2/a)-(gamma
27                 (1+1/a))^2)^3) ) == sk_sample);
28
29 %--method of probability weighted moments
30         pwm = lmom(randomwei.pwm,4);
31         ratio = pwm(1)/(2*pwm(2));
32         alpha(i) = log(2)/log(ratio);
33
34     end
35
36 alpha_mean = sum(estim_mle(2,:))/iter;
37 bias = alpha_mean-alpha_wei
38 tmp_varian = power(estim_mle(2,i)-alpha_mean,2);
39 varian = sum(tmp_varian)/iter;
40 mse = sum(power(estim_mle(2,:)-alpha_wei,2))/iter

```

```
41
42 shape_mom_mean = sum(estim_mom)/iter;
43 bias_mom = shape_mom_mean - alpha_wei
44 tmp_varian_mom = power(a-shape_mom_mean,2);
45 varian_mom = sum(tmp_varian_mom)/iter;
46 mse_mom = sum(power(shape_mom_mean-alpha_wei,2))/iter
47
48
49 shape_pwm_mean = sum(alpha(1,:))/iter;
50 bias_pwm = shape_pwm_mean-alpha_wei
51 tmp_varian_pwm = power(alpha-shape_pwm_mean,2);
52 varian_pwm = sum(tmp_varian_pwm)/iter;
53 mse_pwm = sum(power(alpha-alpha_wei,2))/iter
54
55 end
56 %matlabpool close
```

4 Γενικευμένη κατανομή ακραίων τιμών

4.1 Ορισμός γενικευμένης κατανομής

Οι τρεις τύποι κατανομών που παρουσιάστηκαν και αναλύθηκαν σε προηγούμενα κεφάλαια μπορούν να συνδυαστούν σε μια γενικευμένη κατανομή. Είναι ενδιαφέρον να αναφέρουμε ότι ο Richard von Mises, μαθηματικός Αυστριακής καταγωγής, ήταν ο πρώτος που πρότεινε το συνδυασμό των τριών τύπων κατανομών σε ένα τύπο που αργότερα ονομάστηκε γενικευμένη κατανομή ακραίων τιμών. Ο ίδιος έδωσε συνθήκες για να ανήκει μια κατανομή στο μέγιστο πεδίο έλξης μιας κατανομής ακραίων τιμών, οι οποίες συχνά καλούνται συνθήκες von Mises(βλ [43]).

Ορισμός 4.1.1 (Γενικευμένη κατανομή ακραίων τιμών). *Μια τυχαία μεταβλητή Z θα λέμε ότι έχει γενικευμένη κατανομή ακραίων τιμών (Generalized Extreme Value distribution) με παράμετρο θέσης μ , παράμετρο κλίμακας $\sigma > 0$ και παράμετρο σχήματος ξ αν η αθροιστική συνάρτηση κατανομής της δίνεται από τον παρακάτω τύπο*

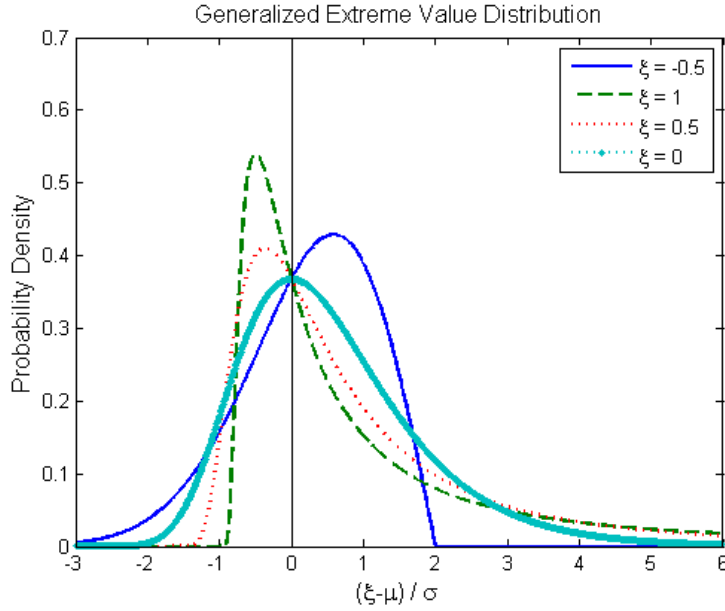
$$G_{\xi}(x) = \begin{cases} e^{-[1+\xi\frac{x-\mu}{\sigma}]^{-\frac{1}{\xi}}} & \text{αν } \xi \neq 0 \\ e^{-e^{[-\frac{x-\mu}{\sigma}]} } & \text{αν } \xi = 0. \end{cases} \quad (126)$$

Η κατανομή αυτή είναι γνωστή ως γενικευμένη κατανομή ακραίων τιμών και θα συμβολίζεται $GEV(x; \mu, \sigma, \xi)$, $G_{\xi}(x)$ ή απλά $G(x)$. Η παράμετρος ξ είναι αυτή που καθορίζει τον τύπο των οριακών κατανομών. Για παράδειγμα η $GEV(1, 1, \frac{1}{\alpha})$ για $\alpha > 0$ θα είναι ίδιου τύπου με την $\Phi_{\alpha}(x)$ ενώ η $GEV(-1, -1, -\frac{1}{\alpha})$ θα είναι ίδιου τύπου με την $\Psi_{\alpha}(x)$, αφού σε αυτές τις περιπτώσεις προκύπτουν οι τύποι που αντιστοιχούν στην Fréchet και τη Weibull. Παρατηρούμε ότι λαμβάνοντας το όριο στο μηδέν του πρώτου κλάδου της Σχέσης 126 δηλαδή,

$$\lim_{\xi \rightarrow 0} e^{-[1+\xi\frac{x-\mu}{\sigma}]^{-\frac{1}{\xi}}} \quad (127)$$

καταλήγουμε στον δεύτερο κλάδο της Σχέσης 126. Η παράμετρος σχήματος ξ αναφέρεται συνήθως και ως δείκτης ουράς (tail index) της κατανομής. Αυτό

διότι κατέχει σημαντικό ρόλο στον προσδιορισμό της συμπεριφοράς της ουράς της κατανομής. Παραθέτουμε ένα διάγραμμα της γενικευμένης κατανομής ακραίων τιμών για τιμές του ξ που αντιστοιχούν στις τρεις οριακές κατανομές απο όπου φαίνεται ότι για μεγάλες τιμές του ξ έχουμε βαριά ουρά. Γενικά ισχύει



Σχήμα 10. Γενικευμένη κατανομή ακραίων τιμών για διάφορες τιμές του ξ

ότι για τιμές του $\xi < 0, \xi = 0, \xi > 0$ η G_ξ είναι ίδιου τύπου με την Weibull, Gumbel και Frechét αντίστοιχα. Συγκεκριμένα ισχύει,

$$\Lambda = G_0(x) \quad (128)$$

$$\Phi_\alpha(x) = G_{1/\alpha}(\alpha(x - 1)) \quad (129)$$

$$\Psi_\alpha(x) = G_{-1/\alpha}(\alpha(1 + x)) \quad (130)$$

Η συνάρτηση πυκνότητας πιθανότητας της γενικευμένης κατανομής ακραίων τιμών θα βρισκείται παραγωγίζοντας τη Σχέση 126.

$$g(x) = \begin{cases} \frac{1}{\sigma} [1 + \xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}-1} e^{[1+\xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}}} & \text{αν } \xi \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{(-e[-\frac{x-\mu}{\sigma}])} & \text{αν } \xi = 0. \end{cases} \quad (131)$$

Η αντίστροφη συνάρτηση κατανομή της αθροιστικής κατανομής της Σχέσης 126 δίνεται από τον τύπο

$$Q_{G_\xi(x)}(p) = \begin{cases} \mu + \frac{\sigma}{\xi}[(-\ln p)^{-\xi} - 1] & \text{αν } \xi \neq 0 \\ \mu + \sigma(-\ln(-\ln y)) & \text{αν } \xi = 0 \end{cases} \quad (132)$$

διότι για $\xi \neq 0$ έχουμε,

$$\begin{aligned} G_{\xi \neq 0}(x) = y &\Leftrightarrow -(1 + \xi \frac{x - \mu}{\sigma})^{-\frac{1}{\xi}} = \ln y \\ &\Leftrightarrow x = \mu + \frac{\sigma}{\xi}[(\ln y)^{-\xi} - 1] \end{aligned}$$

και για $\xi = 0$ έχουμε,

$$\begin{aligned} G_{\xi=0}(x) = y &\Leftrightarrow e^{-e^{-\frac{x-\mu}{\sigma}}} = y \\ &\Leftrightarrow \frac{x - \mu}{\sigma} = -\ln(-\ln y) \\ &\Leftrightarrow x = \mu + \sigma(-\ln(-\ln y)). \end{aligned}$$

Η μέση τιμή και η διακύμανση της κατανομής δίνεται από τους παρακάτω τύπους

$$E(X) = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - \Gamma(1 - \xi)) & \text{αν } \xi < 1, \xi \neq 0 \\ \mu + \sigma\gamma & \text{αν } \xi = 0 \end{cases} \quad (133)$$

$$Var(X) = \begin{cases} \frac{\sigma^2}{\xi^2}(\Gamma(1 - 2\xi) - \Gamma^2(1 - \xi)) & \text{αν } \xi < \frac{1}{2}, \xi \neq 0 \\ \frac{\pi}{6}\sigma^2 & \text{αν } \xi = 0 \end{cases} \quad (134)$$

όπου $\gamma = 0.5772$ είναι η σταθερά Euler. Τέλος, ο συντελεστής ασυμμετρίας της γενικευμένης κατανομής ορίζεται ως

$$skew_{G_\xi} := \frac{E[\{z - E(X)\}^3]}{Var(X)^{\frac{3}{2}}}$$

και είναι ίσος με 0 στο $\xi_0 \simeq -2.8$. Επιπλέον $skew_{G_\xi} > 0$ για $\xi > \xi_0$ και $skew_{G_\xi} < 0$ για $\xi < \xi_0$. Για την Gumbel έχουμε $skew_{G_0} \simeq 1.14$.

4.2 Εκτίμηση Παραμέτρων

Για την εκτίμηση των παραμέτρων της γενικευμένης κατανομής ακραίων τιμών έχουν προταθεί διάφορες μέθοδοι. Οι δύο επικρατέστερες και αποτελεσματικότερες είναι η μέθοδος μέγιστης πιθανοφάνειας (βλ [47]) και η μέθοδος των σταθμισμένων ροπών πιθανότητας (βλ [32]). Σε σχετική έρευνα από τον Hosking, (1990) γίνεται αναφορά στη μέθοδο L-ροπών (L-moments method) η οποία είναι βασισμένη σε ποσότητες που εκφράζονται ως γραμμικοί συνδιασμοί διατεταγμένων όρων. Πρόκειται για τη γενικευμένη μέθοδο της μεθόδου των ροπών. Σύγκριση της αποτελεσματικότητας των μεθόδων έχει γίνει σε αρκετές μελέτες, Madsen et al (1997), Dupuis & Field (1998) και Dupuis (1999b). Παρόλο που η μέθοδος των σταθμισμένων ροπών πιθανότητας παράγει μεροληπτικές εκτιμήτριες φαίνεται να είναι προτιμότερη από τις εκτιμήτριες της μεθόδου μέγιστης πιθανοφάνειας διότι οδηγούν σε ακραία ποσοστοιμύρια με χαμηλή διακύμανση (βλ [32]). Σχετικά με την εκτίμηση της παραμέτρου ξ , έχουν αντιμετωπιστεί κάποιες δυσκολίες. Αναφορικά, για $\xi < -1$ οι εκτιμήτριες μέγιστης πιθανοφάνειας δεν ορίζονται και για $-1 < \xi < -\frac{1}{2}$ εμφανίζονται σοβαρά προβλήματα εκτίμησης (βλ [51]) ενώ όταν $\xi > \frac{1}{2}$ οι ροπές δεύτερης τάξης και άνω δεν υπάρχουν. Σε σχετικά πρόσφατες έρευνες εκτίμησης των παραμέτρων της γενικευμένης κατανομής προτείνεται ο συνδυασμός μεθόδων (βλ [1]) γνωστές και ως μικτές μέθοδοι (mixed methods) οι οποίες φαίνεται να είναι περισσότερο αποτελεσματικές στην εκτίμηση παραμέτρων για την κατανομή ακραίων τιμών.

4.2.1 Μέθοδος μέγιστης πιθανοφάνειας

Ο λογάριθμος της συνάρτησης πιθανοφάνειας για τη γενικευμένη κατανομή ακραίων τιμών θα είναι

$$l(x_1, x_2, \dots, x_n | \mu, \sigma, \xi) = \tag{135}$$
$$\begin{cases} -n \ln \sigma - (1 + \frac{1}{\xi}) \sum_{i=1}^n \ln[1 + \xi(\frac{x_i - \mu}{\sigma})] - \sum_{i=1}^n [1 + \xi(\frac{x_i - \mu}{\sigma})]^{\frac{1}{\xi}} & \text{αν } \xi \neq 0 \\ -n \ln \sigma - \sum_{i=1}^n (\frac{x_i - \mu}{\sigma}) - \sum_{i=1}^n e^{(-\frac{x_i - \mu}{\sigma})} & \text{αν } \xi = 0 \end{cases}$$

για $1 + \xi\left(\frac{x_i - \mu}{\sigma}\right) > 0$, $i = 1, \dots, n$ αλλιώς $l(x_1, x_2, \dots, x_n | \mu, \sigma, \xi) = -\infty$. Για την εκτίμηση των παραμέτρων μεγιστοποιούμε την παραπάνω εξίσωση ως προς το διάνυσμα (μ, σ, ξ) . Η συνηθέστερη αριθμητική μέθοδος που εφαρμόζεται για την επίλυση του παραπάνω συστήματος είναι η μέθοδος Newton-Raphson. Σαφώς υπάρχει πάντα η περίπτωση εμφάνισης προβλημάτων στη μέθοδο όπως για παράδειγμα μη-σύγκλιση του αλγορίθμου.

Είναι σημαντικό να αναφέρουμε ότι οι εκτιμήτριες μέγιστης πιθανοφάνειας $\hat{\mu}$, $\hat{\sigma}$, $\hat{\xi}$ ακολουθούν ασυμπτωτικά μια πολυδιάστατη κανονική κατανομή με μέση τιμή $(\mu, \sigma, \xi)^T$ και πίνακα διασποράς τον I^{-1} όπου I είναι ο εσιανός πίνακας με

$$I = \begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \mu \partial \xi} \\ \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \sigma^2} & \frac{\partial^2 l}{\partial \sigma \partial \xi} \\ \frac{\partial^2 l}{\partial \mu \partial \xi} & \frac{\partial^2 l}{\partial \sigma \partial \xi} & \frac{\partial^2 l}{\partial \xi^2} \end{pmatrix}.$$

Αν s_{ii} είναι το i -διαγώνιο στοιχείο του I^{-1} ισχύει προσεγγιστικά ότι

$$\hat{\mu} \sim N(\mu, s_{11}), \quad (136)$$

$$\hat{\sigma} \sim N(\sigma, s_{22}), \quad (137)$$

$$\hat{\xi} \sim N(\xi, s_{33}). \quad (138)$$

Έτσι τα προσεγγιστικά διαστήματα εμπιστοσύνης με συντελεστή εμπιστοσύνης $1 - \alpha$ για τις παραμέτρους μ, σ, ξ θα είναι τα παρακάτω

$$(\hat{\mu} - \sqrt{s_{11}}z_{\alpha/2}, \hat{\mu} + \sqrt{s_{11}}z_{\alpha/2}), \quad (139)$$

$$(\hat{\sigma} - \sqrt{s_{22}}z_{\alpha/2}, \hat{\sigma} + \sqrt{s_{22}}z_{\alpha/2}), \quad (140)$$

$$(\hat{\xi} - \sqrt{s_{33}}z_{\alpha/2}, \hat{\xi} + \sqrt{s_{33}}z_{\alpha/2}). \quad (141)$$

όπου z_{α} είναι το άνω α -σημείο της κανονικής κατανομής με μέση τιμή μηδέν και διασπορά τη μονάδα.

4.2.2 Μέθοδος των σταθμισμένων ροπών πιθανότητας

Οι σταθμισμένες ροπές πιθανότητας της γενικευμένης κατανομής ακραίων τιμών για $\xi \neq 0$ δίνονται από,

$$\begin{aligned}
 \beta_r = M_{1,r,0} &= \int_0^1 [\mu + \sigma \frac{\{1 - (-\ln F)^\xi\}}{\xi}] F^r dF & (142) \\
 &= \int_0^\infty \{ \mu + \sigma \frac{(1 - u^\xi)}{\xi} e^{-(r+1)u} du \} \\
 &= (\mu + \frac{\sigma}{\xi}) \int_0^\infty e^{-(r+1)u} du - \frac{\sigma}{\xi} \int_0^\infty u^\xi e^{-(r+1)u} du \\
 &= (\mu + \frac{\sigma}{\xi})(r+1)^{-1} - \frac{\sigma}{\xi}(r+1)^{-1-\xi} \Gamma(1+\xi) \\
 &= (r+1)^{-1} [\mu + \sigma \frac{\{1 - (r+1)^{-\xi} \Gamma(1+\xi)\}}{\xi}], \quad \xi > -1.
 \end{aligned}$$

Δηλαδή,

$$\beta_r = (r+1)^{-1} [\mu + \sigma \frac{\{1 - (r+1)^{-\xi} \Gamma(1+\xi)\}}{\xi}], \quad \xi > -1. \quad (143)$$

Όταν $\xi \leq -1$ τότε η ποσότητα β_0 και τα υπόλοιπα β_r δεν υπάρχουν.

Από τη Σχέση (143) έχουμε,

$$\beta_0 = \mu + \frac{\sigma \{1 - \Gamma(1+\xi)\}}{\xi}, \quad (144)$$

$$2\beta_1 - \beta_0 = \frac{\sigma \Gamma(1+\xi)(1 - 2^{-\xi})}{\xi}, \quad (145)$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1 - 3^{-\xi}}{1 - 2^{-\xi}}. \quad (146)$$

Οι εκτιμήτριες των παραμέτρων είναι η λύση του συστήματος 144-146 για τις παραμέτρους μ, σ, ξ όταν αντικαταστήσουμε τα β_r από τις εκτιμήτριές τους, b_r . Για να εκτιμήσουμε τη παράμετρο ξ αρκεί να λύσουμε την παρακάτω εξίσωση,

$$\frac{3b_2 - b_0}{2b_1 - b_0} = \frac{1 - 3^{-\xi}}{1 - 2^{-\xi}}. \quad (147)$$

Η εξίσωση 147 απαιτεί τη χρήση επαναληπτικής μεθόδου. Γνωρίζουμε ότι συνήθως η παράμετρος ξ είναι μέσα στο διάστημα $[-\frac{1}{2}, \frac{1}{2}]$ με αποτέλεσμα η ποσότητα $\frac{1-3^{-\xi}}{1-2^{-\xi}}$ να είναι σχεδόν γραμμική σε αυτό το διάστημα. Έτσι, μπορούμε να τη προσεγγίσουμε από πολυώνυμο μικρής τάξης. Η προσεγγιστική εκτιμήτρια της παραμέτρου ξ θα είναι

$$\hat{\xi} = 7.8590c + 2.9554c^2, \quad \text{όπου } c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\ln 2}{\ln 3}. \quad (148)$$

Το σφάλμα που προκύπτει από την παραπάνω προσέγγιση είναι μικρό, της τάξης του $9 \cdot 10^{-4}$. Έχοντας την εκτιμήτρια για την παράμετρο ξ μπορούμε να υπολογίσουμε τις παραμέτρους μ, σ από τις εξισώσεις 144 και 145. Έτσι θα έχουμε,

$$\hat{\sigma} = \frac{(2b_1 - b_0)\hat{\xi}}{\Gamma(1 + \hat{\xi})(1 - 2^{-\hat{\xi}})}, \quad (149)$$

$$\hat{\mu} = b_0 + \frac{\hat{\sigma}}{\hat{\xi}} \{\Gamma(1 + \hat{\xi}) - 1\}. \quad (150)$$

Οι εξισώσεις 148,149,150 είναι οι εκτιμήτριες της GEV.

Όσον αφορά τις ιδιότητες των εκτιμητριών μπορεί να αποδειχθεί ότι το διάνυσμα $(b_0, b_1, b_2)^T$ ακολουθεί προσεγγιστικά τη πολυδιάστατη κανονική κατανομή με μέση τιμή $M = (b_0, b_1, b_2)^T$ και πίνακα συνδιακύμανσης $n^{-1}GVG^T$ όπου G είναι ο εσιανός 3×3 πίνακας ως προς το διάνυσμα των $(b_0, b_1, b_2)^T$ δηλαδή,

$$n^{-1}GVG^T = n^{-1} \begin{pmatrix} \sigma^2 w_{11} & \sigma^2 w_{12} & \sigma w_{13} \\ \sigma^2 w_{12} & \sigma^2 w_{22} & \sigma w_{23} \\ \sigma w_{13} & \sigma w_{23} & w_{33} \end{pmatrix}$$

Τα w_{ij} είναι συναρτήσεις της παραμέτρου ξ και έχουν πολύπλοκη αλγεβρική μορφή. Για περισσότερες λεπτομέρειες παραπέμπουμε τον αναγνώστη στη σχετική βιβλιογραφία (βλ [32], [48]).

Είναι σημαντικό να σημειώσουμε ότι το κύριο μειονέκτημα της μεθόδου είναι στον περιορισμό της όταν πρόκειται για κατανομές πυκνότητας πιθανότητας με πολύ βαριά ουρά, πιο συγκεκριμένα όταν πρόκειται για τιμές της παραμέτρου

ξ μεγαλύτερες από 0.5 . Στις περιπτώσεις όπου $\xi > 0.5$, οι ασυμπτωτικές ιδιότητες της εκτιμήτριας των σταθμισμένων ροπών πιθανότητας $\hat{\xi}$ δεν ισχύουν με αποτέλεσμα να μη μπορούμε να υπολογίσουμε ασυμπτωτικά διαστήματα εμπιστοσύνης. Έτσι χρησιμοποιούνται οι γενικευμένες σταθμισμένες ροπές πιθανότητας (generalized probability weighted moments), επεκτείνοντας το εύρος τιμών που μπορεί να λάβει η παράμετρος ξ (βλ [14]).

4.2.3 Μικτές Μέθοδοι

Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε τις παραμέτρους της γενικευμένης κατανομής ακραίων τιμών χρησιμοποιώντας το συνδυασμό δύο μεθόδων, της μεθόδου μέγιστης πιθανοφάνειας και της μεθόδου L-ροπών από την οποία θα λάβουμε τις πρώτες δύο ροπές. Έστω μια ακολουθία τυχαίων μεταβλητών X_1, X_2, \dots, X_n με συνάρτηση κατανομής την κατανομή ακραίων τιμών όπως ορίζεται στη Σχέση 126. Επιπλέον υποθέτουμε ότι η παράμετρος σχήματος ξ βρίσκεται στο διάστημα $-\frac{1}{2} < \xi < \frac{1}{2}$. Αυτή η συνθήκη εξασφαλίζει ότι οι ροπές δεύτερης τάξης υπάρχουν και ότι οι εκτιμήτριες της μεθόδου μέγιστης πιθανοφάνειας είναι συνεπείς ως προς τις συνθήκες κανονικότητας (βλ [47]). Θυμόμαστε ότι ο λογάριθμος της μεθόδου μέγιστης πιθανοφάνειας για $\xi \neq 0$ είναι

$$\ln L(\theta) = -n \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{\frac{1}{\xi}}. \quad (151)$$

Επίσης, όπως είδαμε και στη μέθοδο των σταθμισμένων ροπών πιθανότητας, οι δύο πρώτες ροπές δίνονται από τις παρακάτω σχέσεις

$$\lambda_1 = E(X) = \mu + \frac{\sigma}{\xi} (1 - \Gamma(1 + \xi)), \quad (152)$$

$$\lambda_2 = \frac{1}{2} E(|X_1 - X_2|) = \frac{\sigma}{\xi} (1 - 2^{-\xi}) \Gamma(1 + \xi). \quad (153)$$

Στόχος μας είναι να εκτιμήσουμε το διάνυσμα παραμέτρων $\theta = (\mu, \sigma, \xi)$ της κατανομής ακραίων τιμών. Έστω λοιπόν

$$\lambda = \lambda(\theta) \quad (154)$$

μια παραμετροποίηση του θ όπου $\lambda_3 = \xi$ και $\lambda(\theta)$ είναι μια συνεχώς διαφορίσιμη και ένα προς ένα απεικόνιση ανάμεσα στα λ, θ έτσι ώστε ο λογάριθμος της πιθανοφάνειας $\ln L(\lambda)$ να είναι ισοδύναμος με τον λογάριθμο $\ln L(\theta)$ της Σχέσης 151. Επίσης υποθέτουμε ότι το διάνυσμα παραμέτρων λ εκτιμάται από το διάνυσμα $\hat{\lambda}$, τις δειγματικές ροπές δηλαδή

$$\hat{\lambda}_1 = \frac{1}{n} \sum_{i=1}^n X_i, \quad (155)$$

$$\hat{\lambda}_2 = \frac{1}{n(n-1)} \sum_{i < j} |X_i - X_j|. \quad (156)$$

Έτσι, αν υποθέσουμε ότι $\lambda_3 = \xi$ και αντικαταστήσουμε το διάνυσμα παραμέτρων $\theta = (\mu, \sigma, \xi)$ με το διάνυσμα $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ στις Σχέσεις 152, 153 καταλήγουμε στις ακόλουθες σχέσεις για τα μ, σ :

$$\mu = \lambda_1 - \frac{\lambda_2(1 - \Gamma(1 + \lambda_3))}{(1 - 2^{-\lambda_3})\Gamma(1 + \lambda_3)}, \quad (157)$$

$$\sigma = \frac{\lambda_2 \lambda_3}{(1 - 2^{-\lambda_3})\Gamma(1 + \lambda_3)}, \quad (158)$$

$$\xi = \lambda_3. \quad (159)$$

Μετά από αριθμητικούς υπολογισμούς οι εκτιμήτριες δίνονται από τις παρακάτω σχέσεις:

$$\hat{\mu} = \hat{\lambda}_1 - \frac{\hat{\lambda}_2(1 - \Gamma(1 + \hat{\lambda}_3))}{(1 - 2^{-\hat{\lambda}_3})\Gamma(1 + \hat{\lambda}_3)}, \quad (160)$$

$$\hat{\sigma} = \frac{\hat{\lambda}_2 \hat{\lambda}_3}{(1 - 2^{-\hat{\lambda}_3})\Gamma(1 + \hat{\lambda}_3)}, \quad (161)$$

$$\hat{\lambda}_3 = \arg \max_{\lambda_3} \ln L(\lambda). \quad (162)$$

Αποδεικνύεται (βλ [1]) ότι η από κοινού κατανομή $\sqrt{n}(\hat{\theta} - \theta)$ ακολουθεί προσεγγιστικά κανονική κατανομή με μέση τιμή 0 και πίνακα συνδιακύμανσης $J^T V J$

με

$$V = BCB^T$$

$$C = \begin{pmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & 0 \\ 0 & 0 & I_{33}^{(\lambda)} \end{pmatrix}$$

$$B^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ I_{31}^{(\lambda)} & I_{32}^{(\lambda)} & I_{33}^{(\lambda)} \end{pmatrix}$$

όπου,

$$C_{ij} = \lim_{n \rightarrow \infty} \text{cov}(\hat{\lambda}_i, \hat{\lambda}_j), \text{ για } i, j = 1, 2 \quad (163)$$

και

$$J = \frac{\partial \theta^T}{\partial \lambda} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\mu - \lambda_1}{\lambda_2} & \frac{\sigma}{\lambda_2} & 0 \\ \frac{\sigma}{\xi} D(1 + \xi) + (\mu - \lambda_1) \frac{\ln 2}{1 - 2\xi} & \frac{\sigma}{\xi} (1 - \xi D(1 + \xi) + \xi \frac{\ln 2}{1 - 2\xi}) & 1 \end{pmatrix} \quad (164)$$

όπου $D(x) = \frac{d}{dx} \ln \Gamma(x)$ είναι η συνάρτηση δι-γάμμα (Παράρτημα 7.4).

4.3 Δείκτης Ουράς

Οι μέθοδοι που παρουσιάσαμε παραπάνω έχουν σαν στόχο την εύρεση των εκτιμητριών και των τριών παραμέτρων ταυτόχρονα, προσδιορίζοντας έτσι την κατανομή που ακολουθούν τα δεδομένα μας. Ωστόσο, υπάρχει ένας διαφορετικός τρόπος προσδιορισμού της εν λόγω κατανομής, μέσω της εκτίμησης του δείκτη ουράς, ξ . Εκτιμώντας τον δείκτη ουράς, μπορούμε να αποφανθούμε ποια από τις τρεις κατανομές ακραίων τιμών Gumbel, Frechét και Weibull είναι η κατανομή που ακολουθούν τα δεδομένα μας και έπειτα να προσδιορίσουμε τις υπόλοιπες παραμέτρους.

Τα τελευταία χρόνια έχουν γίνει αρκετές μελέτες προς αυτή την κατεύθυνση. Τυπικά μπορούμε να χωρίσουμε τις μελέτες αυτές σε δύο κατηγορίες:

έρευνες βασισμένες σε στατιστική διατεταγμένων δειγμάτων (extreme order statistics) και μελέτες βασισμένες σε παρατηρήσεις πάνω από ένα κατώφλι. Στην πρώτη κατηγορία ανήκουν οι μελέτες των Hill(1975), Pickands(1975) και Weissman(1985) ενώ στη δεύτερη κατηγορία μελέτες των Smith and Weissman(1985), Smith(1987) και Davidson and Smith(1990).

Έστω X_1, \dots, X_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση κατανομής F η οποία ανήκει στη μέγιστη περιοχή έλξης της γενικευμένης κατανομής ακραίων τιμών για κάποιο $\xi \in \mathfrak{R}$ δηλαδή,

$$X_1, \dots, X_n \text{ iid από } F \in MDA(G_\xi). \quad (165)$$

Προκύπτει άμεσα από τον Ορισμό 2.3.1 αντικαθιστώντας την συνάρτηση κατανομής της γενικευμένης κατανομής ακραίων τιμών για $\xi \neq 0$ (Σχέση 126) ότι για μεγάλες τιμές της ποσότητας $v = a_n x + b_n$ θα ισχύει

$$n\bar{F} \approx (1 + \xi \frac{v - b_n}{a_n})^{-\frac{1}{\xi}}, \quad (166)$$

έτσι, η εκτιμήτρια της ουράς λαμβάνει τη μορφή

$$\hat{F}(u) = n^{-1} (1 + \xi \frac{\bar{v} - \bar{b}_n}{\hat{a}_n})^{-\frac{1}{\xi}}, \quad (167)$$

για κατάλληλες εκτιμήτριες $\hat{\xi}, \hat{b}_n, \hat{a}_n$. Σύμφωνα με την Σχέση 165, η εκτίμηση της παραμέτρου ξ μπορεί να βασιστεί σε στατιστική k -άνω διατεταγμένων δειγμάτων δηλαδή,

$$X_{k,n} \leq X_{k+1,n} \leq \dots \leq X_{n,n}. \quad (168)$$

Εδώ να αναφέρουμε ότι σε αυτή την ιδέα έχουν βασιστεί οι μελέτες για την εύρεση εκτιμητριών του δείκτη ουράς οι οποίες συνήθως επιβάλλουν τους παρακάτω περιορισμούς για την επιλογή του k :

$$k(n) \rightarrow \infty, \quad (169)$$

$$\frac{n}{k(n)} \rightarrow \infty. \quad (170)$$

Οι παραπάνω συνθήκες εξασφαλίζουν ότι το πλήθος των διατεταγμένων όρων θα είναι επαρκώς μεγάλο αλλά παράλληλα θα δοθεί περισσότερη εστίαση σε

άνω διατεταγμένους όρους καθώς αυτό που μας ενδιαφέρει περισσότερο είναι οι ιδιότητες της ουράς της κατανομής. Η τιμή του k είναι πολύ σημαντική και όπως θα δούμε παρακάτω οι εκτιμήτριες είναι πολύ ευαίσθητες στην επιλογή του. Με απλά λόγια μια μικρή αλλαγή στην τιμή του k μπορεί να επιφέρει μεγάλη απόκλιση στον υπολογισμό τους επηρεάζοντας μάλιστα τις απαραίτητες ιδιότητες που θα θέλαμε να έχουν οι εκτιμήτριές μας όπως συνέπεια και ασυμπτωτική κανονικότητα.

Από την Σχέση 167 είμαστε σε θέση να εκτιμήσουμε το ποσοστημόριο $x_p = F^{-1}(p)$ για $p \in (0, 1)$ ως εξής:

$$\hat{x}_p = \hat{b}_n + \frac{\hat{a}_n}{\hat{\xi}} ((n(1-p))^{\hat{\xi}} - 1). \quad (171)$$

Πρέπει να αναφέρουμε ότι το $p = p_n$ επιλέγεται με τέτοιο τρόπο ώστε να ισχύει $p > 1 - \frac{1}{n}$. Για να λάβουμε όμως κάποιες εκτιμήσεις για τα ξ, b_n, a_n χρειάζεται να χρησιμοποιήσουμε μια υποσειρά του δείγματος X_1, \dots, X_n . Συγκεκριμένα θα πρέπει να χρησιμοποιήσουμε την υποσειρά $\frac{k}{n}$ με το $k = k(n)$ να επαληθεύει τους περιορισμούς που συζητήσαμε νωρίτερα. Έτσι, η Σχέση 171 γίνεται

$$\hat{x}_p = \hat{b}_{n/k} + \frac{\hat{a}_{n/k}}{\hat{\xi}} \left(\left(\frac{n}{k} (1 - p_n) \right)^{\hat{\xi}} - 1 \right). \quad (172)$$

Με αυτό τον τρόπο απομακρυνόμαστε από το κρίσιμο σύνορο $(1 - \frac{1}{n})$ και μεταφερόμαστε στο πιο ασφαλές σύνορο $(1 - \frac{k}{n})$.

4.3.1 Μέθοδος εκτιμήτριας Pickands

Η βασική ιδέα πίσω από την εκτιμήτρια Pickands είναι η εύρεση μιας ισοδύναμης συνάρτησης κατανομής που να ανήκει στη μέγιστη περιοχή έλξης της γενικευμένης κατανομής ακραίων τιμών η οποία θα περιέχει μια πιο βολική έκφραση της παραμέτρου ξ . Από το Θεώρημα 3.4.5 (βλ [15]) προκύπτει ότι

$$\lim_{t \rightarrow \infty} \frac{U(2t) - U(t)}{U(t) - U(t/2)} = 2^{\xi}, \quad (173)$$

όπου $U(t) = F^{-1}(1 - \frac{1}{t})$. Επίσης θα ισχύει η ακόλουθη ομοιόμορφη ιδιότητα:

$$\lim_{t \rightarrow \infty} \frac{U(c(t)t) - U(t)}{U(t) - U(t/c(t))} = 2^{\xi}, \quad (174)$$

για κάθε θετική συνάρτηση c για την οποία ισχύει $\lim_{t \rightarrow \infty} c(t) = 2$. Το επόμενο βήμα είναι να κατασκευάσουμε μια εμπειρική εκτιμήτρια χρησιμοποιώντας την Σχέση 174. Για να το επιτύχουμε θεωρούμε το διατεταγμένο δείγμα

$$V_{n,n} \leq \dots \leq V_{1,n} \quad (175)$$

ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με απο κοινού συνάρτηση κατανομής Pareto

$$F_V(x) = 1 - \frac{1}{x}, \quad x \geq 1. \quad (176)$$

Συνεπώς θα ισχύει ότι

$$(X_{k,n})_{k=1,\dots,n} \xrightarrow{\mathcal{L}} (U(V_{k,n}))_{k=1,\dots,n}, \quad (177)$$

όπου X_1, \dots, X_n είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση κατανομής F . Εδώ να προσθέσουμε ότι η ποσότητα $V_{k,n}$ είναι το εμπειρικό $(1 - \frac{k}{n})$ ποσοστημόριο της F_V . Είναι σχετικά εύκολο να δούμε ότι ισχύει η παρακάτω σχέση

$$\frac{k}{n} V_{k,n} \xrightarrow{\mathcal{P}} 1, \quad n \rightarrow \infty, \quad (178)$$

όταν $k = k(n) \rightarrow \infty$ και $k/n \rightarrow 0$. Πιο συγκεκριμένα,

$$V_{k,n} \xrightarrow{\mathcal{P}} \infty \quad (179)$$

$$\frac{V_{2k,n}}{V_{k,n}} \xrightarrow{\mathcal{P}} 0, \quad n \rightarrow \infty. \quad (180)$$

Συνδυάζοντας το παραπάνω με την Σχέση 174 προκύπτει ότι

$$\frac{U(V_{k,n}) - U(V_{2k,n})}{U(V_{2k,n}) - U(V_{4k,n})} \xrightarrow{\mathcal{P}} \frac{1}{2}, \quad n \rightarrow \infty. \quad (181)$$

Έτσι η εκτιμήτρια Pickand ορίζεται ως εξής

$$\hat{\xi}_{k,n}^P = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}. \quad (182)$$

Η εκτιμήτρια Pickand είναι ασθενώς συνεπής δηλαδή,

$$\hat{\xi}_{k,n}^P \xrightarrow{\mathcal{P}} \xi, \quad n \rightarrow \infty \quad (183)$$

υπό την προϋπόθεση ότι ισχύουν οι παρακάτω συνθήκες (βλ [31]):

$$k \rightarrow \infty \quad (184)$$

$$\frac{k}{n} \rightarrow 0. \quad (185)$$

Αποδεικνύεται επίσης (βλ [46]) ότι η εκτιμήτρια Pickand ακολουθεί ασυμπτωτικά την κανονική κατανομή δηλαδή,

$$\sqrt{k}(\hat{\xi} - \xi) \xrightarrow{P} N(0, v(\xi)), \quad n \rightarrow \infty, \quad (186)$$

όπου

$$v(\xi) = \frac{\xi^2(2^{2\xi+1} + 1)}{(2(2^\xi - 1) \ln 2)^2}. \quad (187)$$

Στο σημείο αυτό θα πρέπει να τονίσουμε ότι η εκτιμήτρια Pickands της Σχέσης 182 εμπεριέχει μια σειρά από όρους άνω-διατεταγμένων παρατηρήσεων. Έτσι όπως θα δούμε και στην αριθμητική εφαρμογή μπορεί κάποιος να ορίσει το σχήμα Pickands (Pickands-plot) ως εξής

$$\{(k, \hat{\xi}_{k,n}^P) : k = 1, \dots, n\}, \quad (188)$$

ώστε να γίνει πιο εύκολη η επιλογή του k . Συνήθως, η βέλτιστη επιλογή για το k είναι στο σημείο όπου το σχήμα Pickands είναι σχετικά οριζόντιο. Είναι σημαντικό να τονίσουμε πως η επιλογή του k είναι βασική για την εκτιμήτρια που θα λάβουμε. Με άλλα λόγια η εκτιμήτρια της Σχέσης 182 είναι πολύ ευαίσθητη στην επιλογή του k . Έτσι, μια μικρή αλλαγή στην τιμή του μπορεί να επιφέρει μεγάλη απόκλιση στην εκτιμήτρια. Επιπλέον, η επιλογή του k είναι αυτή που καθορίζει τις απαραίτητες στατιστικές ιδιότητες που πρέπει να έχει η εκτιμήτρια μας όπως συνέπεια και ασυμπτωτική κανονικότητα όπως άλλωστε είδαμε και παραπάνω.

4.3.2 Μέθοδος εκτιμήτριας Hill

Ένας άλλος τρόπος για να εκτιμήσουμε το Δείκτη Ουράς ο οποίος δε χρησιμοποιεί κάποια παραμετρική μορφή της συνάρτησης κατανομής αλλά εστιάζει

στη συμπεριφορά της ουράς της κατανομής προτάθηκε από τον Hill(1975). Η εκτιμήτρια αυτή χρησιμοποιείται για την εκτίμηση του Δείκτη Ουράς ξ , όταν η άνω (ή κάτω) ουρά της κατανομής είναι της μορφής

$$1 - F(x) = Cx^{-\xi}. \quad (189)$$

Πιο αναλυτικά, ας υποθέσουμε ότι οι τυχαίες μεταβλητές X_1, \dots, X_n είναι ανεξάρτητα και ισόνομα κατανεμημένες με αθροιστική συνάρτηση κατανομής $F \in MDA(\Phi_\xi)$, $\xi > 0$, έτσι ώστε

$$\bar{F}(x) = x^{-\xi}L(x), \quad x > 0 \quad (190)$$

όπου L μια συνάρτηση βραδείας κύμανσης. Για να μπορέσουμε να μοντελοποιήσουμε δεδομένα με βαριά ουρά είναι σημαντικό να γνωρίζουμε την τιμή του δείκτη ξ . Για παράδειγμα, αν $\xi < 2$ τότε $EX_1^2 = \infty$. Η περίπτωση αυτή απαντάται περισσότερο σε ασφαλιστικά δεδομένα (βλ [31]).

Εμπειρικές μελέτες σε ουρές λογαριθμικών αποδόσεων στα χρηματοοικονομικά έχουν δείξει ότι η τιμή του δείκτη ξ είναι στο διάστημα $(3, 4)$ (βλ [25]). Έτσι η κατασκευή ασυμπτωτικών διαστημάτων εμπιστοσύνης είναι αμφιλεγόμενη καθώς θα πρέπει να επαληθεύεται το κριτήριο της τέταρτης ροπής. Ορίζουμε την εκτιμήτρια Hill έτσι ώστε

$$\alpha = \xi^{-1} > 0, \quad (191)$$

έτσι η εκτιμήτρια Hill δίνεται από τον ακόλουθο τύπο

$$\hat{\alpha}^H = \alpha_{\hat{k},n}^H = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \quad (192)$$

όπου $k = k(n) \rightarrow \infty$ με τέτοιο τρόπο ώστε όπως και στην εκτιμήτρια Pickands να δημιουργείται μια αύξουσα ακολουθία άνω διατεταγμένων όρων. Ένα από τα πλεονεκτήματα της εκτιμήτριας Hill είναι ότι μέσω διάφορων μεθόδων εκτίμησης μπορούν να προκύψουν διάφορες ασυμπτωτικές εκδοχές της $\hat{\alpha}^H$.

Έστω για παράδειγμα ότι μια τυχαία μεταβλητή X έχει αθροιστική συνάρτηση κατανομής F έτσι ώστε για $\alpha > 0$ να ισχύει

$$P(X > x) = \bar{F}(x) = x^{-\alpha}, \quad x \geq 1. \quad (193)$$

Προκύπτει άμεσα ότι η συνάρτηση κατανομής της $Y = \ln X$ είναι

$$P(Y > y) = e^{-\alpha y}, \quad y \geq 0, \quad (194)$$

δηλαδή ότι η Y έχει εκθετική κατανομή με παράμετρο α και η εκτιμήτρια της παραμέτρου δίνεται ως εξής:

$$\hat{\alpha}_n = \bar{Y}_n^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_j \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_{j,n} \right)^{-1}. \quad (195)$$

Αν γενικεύσουμε τα παραπάνω τότε μπορούμε να δούμε εύκολα ότι

$$\bar{F}(x) = Cx^{-\alpha}, \quad x \geq u \geq 0, \quad (196)$$

με u να είναι γνωστό. Αν επιπλέον υποθέσουμε ότι η παραπάνω σχέση είναι καλώς ορισμένη και $C = u^\alpha$ τότε προκύπτει η εκτιμήτρια της παραμέτρου ως εξής:

$$\hat{\alpha}_n = \left(\frac{1}{n} \sum_{j=1}^n \ln \left(\frac{X_{j,n}}{u} \right) \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_{j,n} - \ln u \right)^{-1}. \quad (197)$$

Ανάλογα με την υπόθεση που κάνει κάποιος για την κατανομή υπό μελέτη η παραπάνω εκτιμήτρια εφαρμόζεται υπό όρους. Η εκτιμήτρια Hill είναι ασθενώς συνεπής αρκεί να ισχύουν οι παρακάτω προϋποθέσεις (βλ [40]):

$$k = k(n) \rightarrow \infty \quad (198)$$

$$\frac{k(n)}{n} \rightarrow 0, \quad \text{καθώς } n \rightarrow \infty. \quad (199)$$

Υποθέτοντας πως η ουρά της κατανομής συμπεριφέρεται σαν την κατανομή Pareto πάνω από κάποιο συγκεκριμένο κατώφλι έστω u , τότε μετά από κάποιους αριθμητικούς υπολογισμούς έχουμε

$$\hat{\alpha}_{k,n}^H = \left(\frac{1}{n} \sum_{j=1}^k \ln \left(\frac{X_{j,n}}{X_{k,n}} \right) \right)^{-1} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \quad (200)$$

$$\hat{C}_{k,n} = \frac{k}{n} X_{k,n}^{\hat{\alpha}_{k,n}^H}. \quad (201)$$

Οπότε η εκτιμήτρια Hill έχει την ίδια μορφή με την εκτιμήτρια μέγιστης πιθανοφάνειας με μόνη διαφορά την αντικατάσταση του ντετερμινιστικού όρου u από το τυχαίο κατώφλι $X_{k,n}$ με το k να ορίζεται ως εξής:

$$K = \text{card}\{i : X_{i,n} > u, i = 1, \dots, n\}. \quad (202)$$

Πρέπει να αναφέρουμε ότι αυτή η μέθοδος τείνει να υπερ-εκτιμά την συμπεριφορά της ουράς όταν ο Δείκτης Ουράς είναι κοντά στο 2 και το δείγμα δεν είναι αρκετά μεγάλο. Ένα δείγμα της τάξης 10^4 είναι κατάλληλο για την σωστή εκτίμηση του.

4.4 Μέτρα Κινδύνου

Όπως αναφέραμε νωρίτερα, τα μεγάλα κέρδη και οι ζημίες αντιστοιχούν συνήθως στην ουρά μιας κατανομής που αναπαριστά χρηματοοικονομικές και αναλογιστικές αποδόσεις. Έρευνες στην εφαρμογή της θεωρία ακραίων τιμών στη μέτρηση χρηματοοικονομικού κινδύνου έδειξαν ότι οι ουρές των αποδόσεων των χρηματιστηρίων ακολουθούν την κατανομή Fréchet (βλ [38]). Το 1997 δόθηκε μια συνοπτική και περιεκτική μελέτη στη βασική θεωρία της θεωρίας ακραίων τιμών Fréchet (βλ [15]). Οι Danielson και deVries χρησιμοποίησαν μια ημι-παραμετρική προσέγγιση βασισμένη στην εκτιμήτρια Hill (Hill-estimator) για τη μέτρηση του δείκτη ουράς (tail index) (βλ [11]), ενώ οι Embrechts et.al. χρησιμοποίησαν μια παραμετρική τεχνική εκτίμησης η οποία ήταν βασισμένη σε ένα ασυμπτωτικό αποτέλεσμα της υπερβάλλουσας κατανομής (excess-distribution) που υπερβαίνει κάποιο υψηλό κατώφλι η οποία ονομάζεται μέθοδος τιμών πάνω από κάποιο κατώφλι (peaks over threshold) (βλ [16]).

Στις επόμενες παραγράφους θα αναφερθούμε στα δύο συνηθέστερα μέτρα κινδύνου, την αξία σε κίνδυνο (VaR) και το αναμενόμενο έλλειμμα (cVaR) που χρησιμοποιούνται ευρέως για την περιγραφή της ουράς κατανομών απωλειών ή κερδών αλλά και στις δύο επικρατέστερες μεθόδους στη θεωρία ακραίων τιμών, τη μέθοδο των μέγιστων μπλοκ (block maxima method) και τη μέθοδο τιμών πάνω από κάποιο κατώφλι. Δίνουμε τον ορισμό της γενικευμένης αντίστροφης

συνάρτησης και της συνάρτησης ποσοστημορίων που θα χρησιμοποιήσουμε στη συνέχεια.

Ορισμός 4.4.1 (Γενικευμένη αντίστροφος συνάρτηση). Έστω μια αύξουσα συνάρτηση $T : \mathfrak{R} \rightarrow \mathfrak{R}$. Η γενικευμένη αντίστροφος συνάρτηση της T ορίζεται να είναι η

$$T^{-1}(y) := \inf\{x \in \mathfrak{R} : T(x) \geq y\}, \text{ όπου } \inf\{\emptyset\} = \infty. \quad (203)$$

Ορισμός 4.4.2 (Συνάρτηση ποσοστημορίων). Έστω μια αθροιστική συνάρτηση κατανομής F . Η γενικευμένη αντίστροφος της F ονομάζεται συνάρτηση ποσοστημορίων. Πιο συγκεκριμένα για $\alpha \in (0, 1)$ το α -ποσοστημόριο της F δίνεται από την σχέση

$$q_\alpha(F) := F^{-1}(\alpha) = \inf\{x \in \mathfrak{R} : F(x) \geq \alpha\}. \quad (204)$$

Για μια τυχαία μεταβλητή X χρησιμοποιούμε το συμβολισμό $q_\alpha(X) := q_\alpha(F)$. Προφανώς αν η F είναι συνεχής και αυστηρώς αύξουσα τότε θα ισχύει $q_\alpha(F) = F^{-1}(\alpha)$ όπου F^{-1} είναι η συνήθης αντίστροφος συνάρτηση της F .

4.4.1 Αξία σε κίνδυνο

Η αξία σε κίνδυνο (Value-at-Risk) ή απλώς VaR είναι ένα μέτρο κινδύνου που προσδιορίζει τη μεγαλύτερη αναμενόμενη ζημία σε ένα ορισμένο χρονικό διάστημα με βάση κάποιο επίπεδο εμπιστοσύνης. Χρησιμοποιήθηκε για πρώτη φορά από την JP Morgan το 1990. Ορίζεται ως το α -ποσοστημόριο της μελλοντικής κατανομής απωλειών ή κερδών σε ένα προκαθορισμένο χρονικό ορίζοντα. Πιο συγκεκριμένα, έστω μια τυχαία μεταβλητή X με συνεχή συνάρτηση αθροιστικής κατανομής F . Το VaR τότε ορίζεται ως εξής:

$$VaR_\alpha = \inf\{x | F_X(x) \geq 1 - \alpha\}. \quad (205)$$

Μετά από μετασχηματισμό έχουμε,

$$VaR_\alpha = q_\alpha(1 - \alpha), \quad (206)$$

όπου q είναι συνάρτηση ποσοστημορίων (quantile function) σε επίπεδο εμπιστοσύνης a , της αθροιστικής συνάρτησης κατανομής F .

Το επίπεδο εμπιστοσύνης είναι συνήθως πολύ μικρό ιδιαίτερα στις χρηματοοικονομικές αποδόσεις όπου έχουμε υψηλές συχνότητες ακραίων τιμών. Ωστόσο, αυτό το μέτρο κινδύνου έχει δυο σημαντικά μειονεκτήματα (βλ [4], [5]), δεν είναι συνεκτικό μέτρο κινδύνου διότι δεν επαληθεύει την ιδιότητα της υποαθροιστικότητας (Παράρτημα 7.5) ενώ παράλληλα δε μας δείχνει το πιθανό μέγεθος μιας απώλειας που το υπερβαίνει. Ωστόσο, έχει αποδειχθεί ότι για τις περισσότερες πρακτικές εφαρμογές η ιδιότητα της υποαθροιστικότητας ισχύει, με αποτέλεσμα να μη χρειάζεται να χρησιμοποιήσουμε ένα πιο περίπλοκο μέτρο ρίσκου μόνο για λόγους συνεκτικότητας (βλ [12]).

4.4.2 Αναμενόμενο Έλλειμα

Οι Artzner et.al.(1997,1999) πρότειναν ένα συνεκτικό μέτρο ως αντικαταστάτη του VaR. Το μέτρο αυτό ονομάζεται αναμενόμενο έλλειμα (Expected Shortfall ή conditional Value-at-Risk), ή απλά ES και παρέχει μια εκτίμηση του πιθανού μεγέθους της ζημίας που υπερβαίνει το VaR.

Υποθέτουμε μια τυχαία μεταβλητή X , τότε το αναμενόμενο έλλειμα κινδύνου ορίζεται ως εξής

$$ES_a = E[X|X > VaR_a]. \quad (207)$$

Πρόταση 4.4.2.1 (Αναμενόμενο Έλλειμμα). Έστω X τυχαία μεταβλητή που περιγράφει απώλειες, με $E(|X|) < \infty$ και συνάρτηση κατανομής F . Το αναμενόμενο έλλειμα σε επίπεδο εμπιστοσύνης $a \in (0, 1)$ ορίζεται ως:

$$ES_a = \frac{1}{1-a} \int_a^1 q_a(F) dx = \frac{1}{1-a} \int_a^1 VaR_x dx, \quad (208)$$

όπου $q_a(F)$ είναι η συνάρτηση ποσοστημορίων (quantile function) της F .

Μπορεί να αποδειχθεί ότι το αναμενόμενο έλλειμα είναι υποαθροιστικό μέτρο κινδύνου και άρα αποτελεσματικότερο από το VaR αφού ορίζεται ως η μέση τιμή όλων των ζημιών που το ξεπερνούν. Έτσι, αντί να ορίσουμε κάποιο κατάλληλο επίπεδο εμπιστοσύνης, a και στηρίζουμε τις προβλέψεις μας στο VaR,

παίρνουμε τον μέσο όρο των VaR σε όλα τα επίπεδα εξετάζοντας έτσι περαιτέρω την ουρά της κατανομής. Προφανώς το αναμενόμενο έλλειμα εξαρτάται μόνο από τη κατανομή F . Με απλά λόγια, το αναμενόμενο έλλειμα απαντά στο ερώτημα: Ποιο είναι το μέγεθος του χρηματοοικονομικού κινδύνου που πιθανώς να χρειαστεί να αντιμετωπίσουμε; Παρακάτω δίνουμε ένα ενδεικτικό παράδειγμα που δείχνει τη σχέση του VaR και του ES.

4.5 Στάθμη απόδοσης και Χρονική Περίοδος απόδοσης

Παρακάτω δίνουμε βασικούς ορισμούς που χρειάζονται για την αναφορά δυο πολύ σημαντικών μεθόδων εκτίμησης όπως θα δούμε στην επόμενη παράγραφο της μεθόδου μέγιστων μπλοκ και της μεθόδου τιμών πάνω από κάποιο κατώφλι.

Ορισμός 4.5.1 (Χρονική περίοδος απόδοσης). Έστω $\{X_n\}$ μια ακολουθία ισόνομων και ανεξάρτητων τυχαίων μεταβλητών με συνεχή συνάρτηση κατανομής F και u ένα δεδομένο κατώφλι. Θεωρούμε την ακολουθία $I_{X_n > u}$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών Bernoulli με πιθανότητα επιτυχίας $p = 1 - F(u)$ όπου,

$$I_{X_n > u} = \begin{cases} 1 & \text{αν } X_n < u, \\ 0 & \text{αλλιώς.} \end{cases} \quad (209)$$

Ο χρόνος πρώτης υπέρβασης του u θα δίνεται ως εξής:

$$L(u) = \min\{n \geq 1 \text{ έτσι ώστε } X_n > u\}. \quad (210)$$

Τότε η μέση τιμή του χρόνου υπέρβασης θα είναι ίση με

$$E[L(u)] = \frac{1}{p},$$

και θα ονομάζεται χρονική περίοδος απόδοσης του γεγονότος $X_n > u$.

Είναι προφανές ότι η $L(u)$ ακολουθεί γεωμετρική κατανομή με παράμετρο p . Εξ' ορισμού η χρονική περίοδος απόδοσης είναι η αναμενόμενη περίοδος για να ξανασυμβεί το γεγονός που αντιστοιχεί στο κατώφλι u . Έτσι μπορούμε να δώσουμε τον ορισμό της στάθμης απόδοσης ως εξής:

Ορισμός 4.5.2 (Στάθμη απόδοσης). Έστω X_n μια ακολουθία ισόνομων και ανεξάρτητων τυχαίων μεταβλητών με συνεχή συνάρτηση κατανομής F . Αν

$$F(z_p) = 1 - p \quad (211)$$

τότε το z_p ονομάζεται *στάθμη απόδοσης* που αντιστοιχεί στη χρονική περίοδο απόδοσης $\frac{1}{p}$.

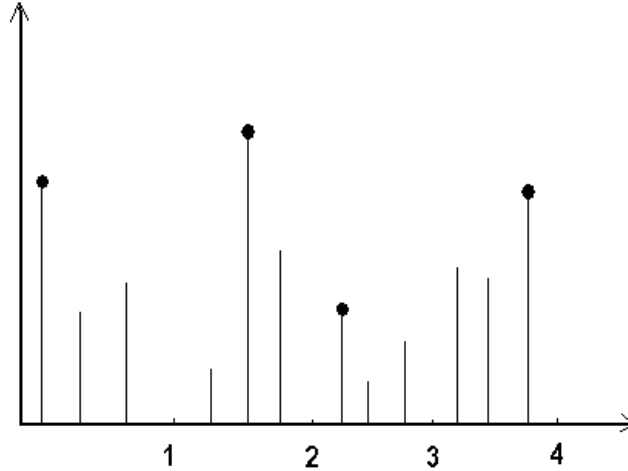
Με άλλα λόγια, η στάθμη απόδοσης είναι ένα κατώφλι z_p το οποίο δεν υπερβαίνει καμία παρατήρηση X_i με πιθανότητα $1 - p$.

4.6 Μέθοδος μέγιστων υποσυνόλων

Η μέθοδος μέγιστων υποσυνόλων (Block Maxima method) αναπτύχθηκε από τον Gumbel (βλ [27]). Είναι μια από τις δύο κυριότερες μεθόδους που χρησιμοποιείται για την εύρεση της κατανομής που ακολουθούν οι μέγιστες, ακραίες τιμές ενός δείγματος. Η μέθοδος βασίζεται στις μέγιστες ή ελάχιστες παρατηρήσεις σε συγκεκριμένες περιόδους. Η μεθοδολογία που ακολουθείται έγκειται στην επιλογή κατάλληλου μήκους, n , μιας χρονικής περιόδου block πλήθους m και την εισαγωγή των παρατηρήσεων σε κάθε χρονική περίοδο. Οι χρονικές περίοδοι μπορεί να είναι επικαλυπτόμενες ή μη-επικαλυπτόμενες. Στη περίπτωση όπου οι περίοδοι δεν επικαλύπτονται, οι επιλεγμένες παρατηρήσεις αποτελούν τις ακραίες παρατηρήσεις και ονομάζονται block maxima. Η λογική πίσω από τη μέθοδο είναι ότι το δείγμα των μέγιστων τιμών ανα περίοδο που δημιουργείται μετά το χωρισμό του αρχικού δείγματος μπορεί να προσεγγιστεί, για επαρκώς μεγάλο αριθμό δείγματος, από την κατανομή ακραίων τιμών με παραμέτρους μ , σ και ξ . Παρακάτω παραθέτουμε ένα πρόχειρο διάγραμμα για την ευκολότερη κατανόηση της μεθόδου για μη-επικαλυπτόμενες περιόδους.

Οι παρατηρήσεις με τις κουκίδες αναπαριστούν τα block maxima για τεσσέρις περιόδους με τρεις παρατηρήσεις να αντιστοιχούν σε κάθε περίοδο. Η μέγιστη τοπική παρατήρηση συμβολίζεται με κουκίδα.

Η μέθοδος χρησιμοποιείται συνήθως όταν η διαθέσιμη πληροφορία που έχουμε είναι μόνο τα block maxima διότι για την εφαρμογή της δεν απαιτούνται



Σχήμα 11. Μέθοδος μέγιστων μπλοκ.

οι υπόλοιπες παρατηρήσεις που ανήκουν στο υποσύνολο αυτό. Είναι σχετικά εύκολη στη προσαρμογή στις περιπτώσεις όπου, λόγω της φύσης του προβλήματος, οι χρονικές περίοδοι εμφανίζονται με ένα φυσικό τρόπο και αντιστοιχούν σε ένα έτος, μήνα ή ακόμη και εβδομάδα. Στη μελέτη μας θα χρησιμοποιήσουμε μόνο ισόνομες και ανεξάρτητα κατανομημένες παρατηρήσεις. Μπορεί για παράδειγμα να εμφανίζουν περιοδική εποχικότητα ή να υφίσταται κάποιου είδους εξάρτηση μικρού εύρους εσωτερικά κάποιων block όχι όμως και μεταξύ των block. Όταν το μέγεθος του block τείνει στο άπειρο θα ισχύει,

$$Pr(Z_n \leq z_p) = G(z_p) = 1 - p. \quad (212)$$

Συνδυάζοντας τη Σχέση 211 και τη Σχέση 126 και λαμβάνοντας κάθε κλάδο ξεχωριστά προκύπτει το ακραίο ποσοστημόριο δηλαδή,

$$z_p = G^{-1}(1 - p) = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - [-\ln(1 - p)]^{-\xi}] & \text{αν } \xi \neq 0, \\ \mu - \sigma \ln[-\ln(1 - p)] & \text{αν } \xi = 0. \end{cases} \quad (213)$$

Οι παράμετροι μ, σ, ξ θα αντικατασταθούν από τις εκτιμήτριες $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ κάποιας μεθόδου εκτίμησης, συνήθως της μεθόδου μέγιστης πιθανοφάνειας. Έστω M_{nj} η μέγιστη παρατήρηση του j -οστού Block. Τότε το καινούριο δείγμα που δημιουργείται είναι

$$M_{n1}, \dots, M_{nm}. \quad (214)$$

Αν υποθέσουμε ότι θα αντικαταστήσουμε τις παραμέτρους από τις εκτιμήτριες μέγιστης πιθανοφάνειας, πρέπει να χρησιμοποιήσουμε σχετικά μεγάλο μήκος block ώστε ανεξαρτήτα με το αν οι παρατηρήσεις είναι εξαρτημένες να μπορούμε να υποθέσουμε ότι είναι ανεξάρτητες. Σε αυτή τη περίπτωση μπορούμε εύκολα να γράψουμε τη συνάρτηση μέγιστης πιθανοφάνειας ως εξής

$$l(M_{n1}, \dots, M_{nm}; \mu, \sigma, \xi) = \quad (215)$$

$$- m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \ln\left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right) - \sum_{i=1}^m \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right)^{-1/\xi},$$

η οποία μεγιστοποιείται έτσι ώστε $\sigma > 0$ και $1 + \xi(M_{ni} - \mu)/\sigma > 0$ για κάθε i .

4.7 Μέθοδος τιμών πάνω από κάποιο κατώφλι

Μια εναλλακτική μέθοδος είναι η μέθοδος τιμών πάνω από κάποιο κατώφλι (Peaks over threshold method) η οποία περιγράφει τη συμπεριφορά τιμών που υπερβαίνουν ένα προκαθορισμένο υψηλό κατώφλι (βλ [50], [13], [37]). Σε αυτή τη μέθοδο χρησιμοποιούνται δύο είδη ανάλυσης και εκτίμησης, η ημι-παραμετρική μέθοδος γνωστή και ως Εκτιμήτρια Hill (βλ [6]) (Hill estimator approach) και η παραμετρική μέθοδος που βασίζεται στη γενικευμένη κατανομή Pareto (βλ [16]). Για λόγους απλότητας θα αναλύσουμε τη δεύτερη προσέγγιση.

Δεδομένου ενός κατωφλιού u θέλουμε να μελετήσουμε όλες τις τιμές που υπερβαίνουν το κατώφλι. Έστω X μια τυχαία μεταβλητή με συνάρτηση κατανομής F . Η κατανομή των τιμών που υπερβαίνουν το κατώφλι u ορίζεται ως εξής:

$$F_u(y) = Pr\{X - u \leq y | X > u\}, \quad (216)$$

για $0 \leq y < x_0 - u$ όπου $x_0 \leq \infty$ είναι το δεξιό άκρο της κατανομής F . Η συνάρτηση κατανομής που δίνεται από τη Σχέση 216 ονομάζεται υπερβάλλουσα

κατανομή (excess distribution function) και αναπαριστά την πιθανότητα μια ζημία να υπερβεί το κατώφλι u το πολύ κατα y , δεδομένου ότι υπερβαίνει όντως το u . Είναι πολύ χρήσιμο να γίνει η παρατήρηση ότι η υπερβάλλουσα κατανομή, F_u , μπορεί να γραφεί συναρτήσει της κατανομής F ως εξής:

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}. \quad (217)$$

Μπορεί να αποδειχθεί ότι, για επαρκώς υψηλό κατώφλι u , η παραπάνω κατανομή μπορεί να προσεγγιστεί από τη γενικευμένη κατανομή Pareto (Generalized Pareto Distribution) διότι όσο το κατώφλι αυξάνεται, η υπερβάλλουσα κατανομή τιμών, $F_u(y)$, συγκλίνει στην γενικευμένη κατανομή Pareto δηλαδή,

$$F_u(y) \rightarrow G_{\xi,\sigma}(y), \quad (218)$$

όπου

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\sigma})_+^{-\frac{1}{\xi}} & \text{αν } \xi \neq 0, \\ 1 - e^{-\frac{y}{\sigma}} & \text{αν } \xi = 0, \end{cases} \quad (219)$$

για $\sigma > 0$ και $-\infty < \xi < \infty$. Η κατανομή $G_{\xi,\sigma}(y)$ είναι η γενικευμένη κατανομή Pareto. Θέτοντας $x = u + y$ και χρησιμοποιώντας τις Σχέσεις 218 και 219, μπορούμε να γράψουμε την F ως εξής,

$$F(x) = (1 - F(x))G_{\xi,\sigma}(x - u) + F(u), \quad x > u. \quad (220)$$

Χρησιμοποιώντας τις εκτιμήτριες μέγιστης πιθανοφάνειας των παραμέτρων της κατανομής και την εμπειρική υπερβάλλουσα κατανομή καταλήγουμε στην εξίσωση εκτίμησης της ουράς της κατανομής δηλαδή,

$$F(\hat{x}) = 1 - \frac{N_u}{N} (1 + \hat{\xi} \frac{x - u}{\hat{\sigma}})^{-\frac{1}{\hat{\xi}}}. \quad (221)$$

Έτσι θα έχουμε

$$VaR_a = \hat{z}_a = u + \frac{\hat{\sigma}}{\hat{\xi}} \left(\frac{N_u}{N} (1 - p)^{-\frac{1}{\hat{\xi}}} - 1 \right), \quad (222)$$

όπου N_u είναι ο αριθμός των παρατηρήσεων που υπερβαίνουν το κατώφλι u . Το αναμενόμενο έλλειμμα που αντιστοιχεί στην παραπάνω αξία σε κίνδυνο μπορεί να υπολογισθεί εύκολα. Έτσι, από τη Σχέση 208 και τη Σχέση 222 έχουμε:

$$ES_a = \frac{1}{1-a} \int_a^1 VaR_x dx = \frac{VaR_a}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}. \quad (223)$$

Η μέθοδος αυτή διαφοροποιείται από τη μέθοδο μέγιστων υποσυνόλων διότι εδώ αναλόγως με την επιλογή του κατωφλιού μπορούμε να προσδιορίσουμε τον αριθμό του δείγματος που το ξεπερνά. Ωστόσο, υπάρχει ένα σοβαρό μειονέκτημα. Οι τιμές που τελικά θα κρατήσουμε ως ακραίες τιμές μετά απο τον προσδιορισμό του κατωφλιού μπορεί να αντιστοιχούν σε χρονικά δυσανάλογες και άρα μη άμεσα συγκρίσιμες χρονικές στιγμές με αποτέλεσμα να μην υπάρχει μια καθολική αιτία που να εξηγεί την συχνότητά τους. Να σημειώσουμε εδώ ότι είναι εξαιρετικά χρήσιμο να δούμε πώς η παράμετρος ξ επηρεάζει τον λόγο των δύο μέτρων κινδύνου για μεγάλες τιμές του επιπέδου εμπιστοσύνης a . Συνεπώς κάποιος μπορεί να υπολογίσει το όριο του λόγου των δύο μέτρων κινδύνου δηλαδή,

$$\lim_{a \rightarrow 1} \frac{ES_a}{VaR_a} = \begin{cases} (1-\xi)^{-1} & \text{αν } \xi \geq 0 \\ 1 & \text{αν } \xi < 0 \end{cases}. \quad (224)$$

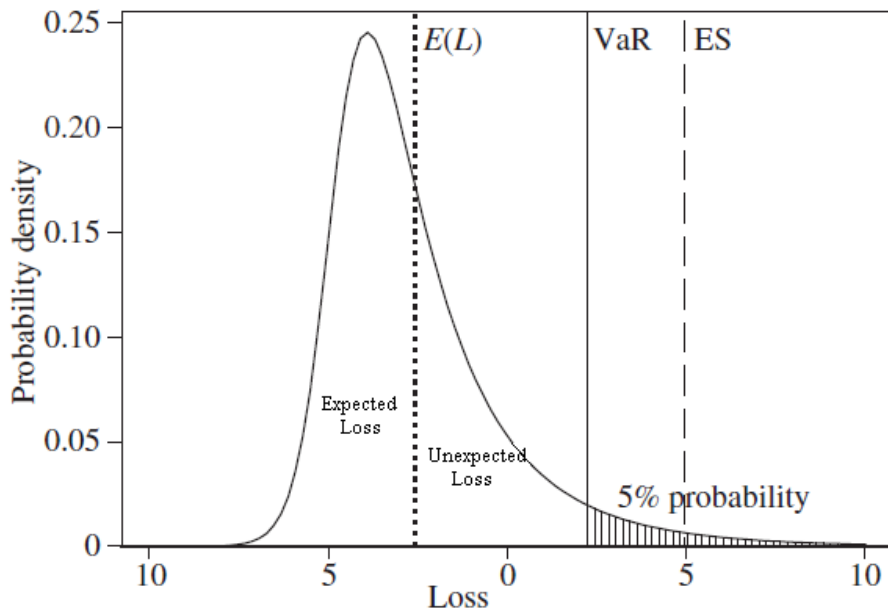
5 Αριθμητική Εφαρμογή

Οι αναμενόμενες ζημιές, οι οποίες αναφέρονται και ως μέσες απώλειες ή συνήθεις ζημιές, είναι γενικά διαχειρίσιμες. Ωστόσο οι μη αναμενόμενες ζημιές, αυτές δηλαδή που αποκλίνουν από τις συνήθεις, μπορούν να θέσουν σε σοβαρό κίνδυνο τη βιωσιμότητα μιας ασφαλιστικής εταιρείας. Ένα από τα πιο δύσκολα καθήκοντα στη διαχείριση κινδύνου είναι η κατάλληλη ρύθμιση του Κεφαλαίου σε Κίνδυνο (Capital at Risk, CaR) που αποσκοπεί στην κάλυψη μη-αναμενόμενων ζημιών. Με απλά λόγια, ο υπολογισμός της κεφαλαιακής απαίτησης για κάθε κίνδυνο που αντιμετωπίζει ένας ασφαλιστικός οργανισμός είναι μια διαδικασία που απαιτεί προσοχή και συνδέεται άμεσα με το επίπεδο του κινδύνου που αναλαμβάνει αλλά και με το κόστος που ενδέχεται να έχει κάποιος 'αυστηρές' προβλέψεις. Ως αυστηρές προβλέψεις εννοούμε κατα κύριο λόγο, τον προσδιορισμό του κατάλληλου επιπέδου εμπιστοσύνης.

Όπως συζητήσαμε παραπάνω, το Value at Risk είναι ένα μέτρο κινδύνου που προσδιορίζει την μεγαλύτερη αναμενόμενη ζημία σε ένα ορισμένο χρονικό διάστημα με βάση κάποιο επίπεδο εμπιστοσύνης. Το αναμενόμενο έλλειμα (Expected Shortfall, ES) από την άλλη είναι η μέση τιμή των απωλειών που είναι μεγαλύτερες από το VaR. Πρέπει να σημειώσουμε ότι το αναμενόμενο έλλειμα δίνει μέγιστη πληροφορία για την κεφαλαιακή επάρκεια διότι συμπεριλαμβάνει ακραίες ζημιές μεγαλύτερες του VaR.

Οι μη-αναμενόμενες ζημιές εκφράζονται από τη διαφορά μεταξύ του VaR και των αναμενόμενων ζημιών όπως φαίνεται στο Σχήμα 12.

Στον υπολογισμό αυτών των μέτρων κινδύνου λαμβάνουμε υπόψιν όχι μόνο το ύψος των ζημιών αλλά και τη συχνότητα αυτών. Οι δύο αυτοί παράγοντες, συχνότητα και ύψος ζημιών, είναι σχετικά δύσκολο να συνδυαστούν διότι υπάρχουν γεγονότα με χαμηλή συχνότητα αλλά μεγάλο κόστος και αρκετά γεγονότα με υψηλή συχνότητα αλλά χαμηλό κόστος ζημιών. Επιπλέον το γεγονός ότι η συχνότητα, περιγράφεται συνήθως από μια διακριτή κατανομή ενώ το ύψος των απωλειών από κάποια συνεχή κατανομή καθιστά τον συνδυασμό τους ακόμη πιο δύσκολο. Έχοντας αναγνωρίσει τις κατανομές που περιγράφουν τη συχνότητα και το ύψος απωλειών το επόμενο βήμα είναι ο συνδυασμός αυτών



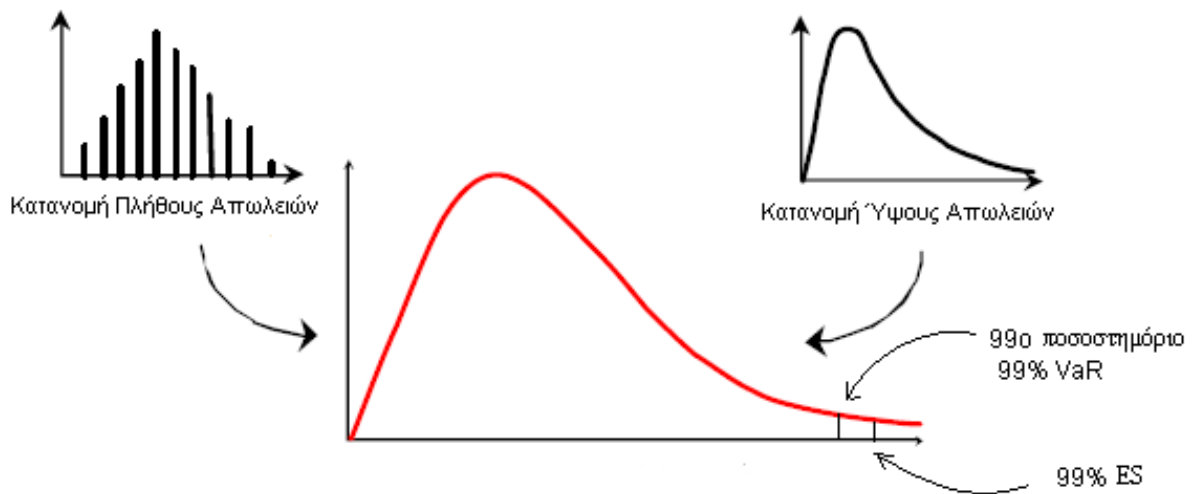
Σχήμα 12. Αναμενόμενες και μη-αναμενόμενες ζημιές

για τη δημιουργία της αθροιστικής κατανομής απωλειών (βλ. Σχήμα 13).

Τα βήματα λοιπόν που ακολουθούνται είναι τα παρακάτω:

1. Αναγνώριση της κατανομής που ακολουθούν οι συχνότητα και το ύψος απωλειών.
2. Συνδυασμός των δύο κατανομών για την δημιουργία της κατανομής συνολικών απωλειών (aggregate loss distribution).
3. Υπολογισμός των μέτρων κινδύνου στο επιθυμητό επίπεδο εμπιστοσύνης.

Υπάρχουν όμως και ανοιχτού τύπου υπολογισμοί όπως για παράδειγμα η χρήση ενός αλγορίθμου προσομοίωσης. Ο πιο εύκολος τρόπος εφαρμογής μιας τέτοιας μεθόδου είναι ο αλγόριθμος Monte Carlo (Plain MC). Σύμφωνα με αυτόν μπορούμε να παράγουμε διάφορα σενάρια για το πλήθος και το ύψος των συνολικών απωλειών δημιουργώντας τυχαίους αριθμούς από τις κατανομές που

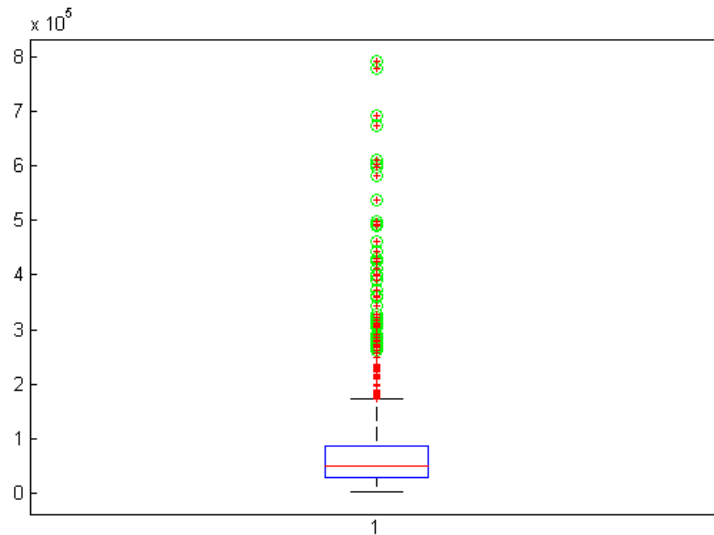


Σχήμα 13. Αθροιστική κατανομή απωλειών

τα περιγράφουν. Έπειτα η διαδικασία έγκειται στην εύρεση των ποσοστημορίων για τον υπολογισμό των μέτρων κινδύνου. Είναι σαφές ότι για μία τέτοια μέθοδο όσο πιο μεγάλος είναι ο αριθμός των δειγμάτων που θα παράξουμε τόσο μεγαλύτερη ακρίβεια θα έχουμε στις προβλέψεις μας. Στη παράγραφο 5.2 εφαρμόζουμε την παραπάνω μεθοδολογία για δεδομένα της ασφαλιστικής εταιρείας Ευρωπαϊκής Πίστης που αναφέρονται σε ζημιές περιουσίας σε ημερήσιο επίπεδο χρονικής διάρκειας τριων ετών. Πρόκειται για το χρονικό διάστημα 5/01/2006 έως 31/12/2008. Το χρονικό διάστημα αυτό είναι διάρκειας 1092 ημερολογιακών ημερών με 734 εργάσιμες ημέρες. Στο δείγμα των εργάσιμων ημερών παρατηρούμε 53 ακραίες παρατηρήσεις. Ο προσδιορισμός των ακραίων παρατηρήσεων υπολογίστηκε σύμφωνα με την παρακάτω σχέση:

$$Extremevalues = quartile(75\%) + 3iqr = Q_3 + 3iqr \quad (225)$$

όπου $iqr = Q(3) - Q(1)$ είναι το ενδοτεταρτημοριακό εύρος. Στο Σχήμα 14 απεικονίζονται με πράσινο χρώμα οι ακραίες τιμές. Στους παρακάτω πίνακες



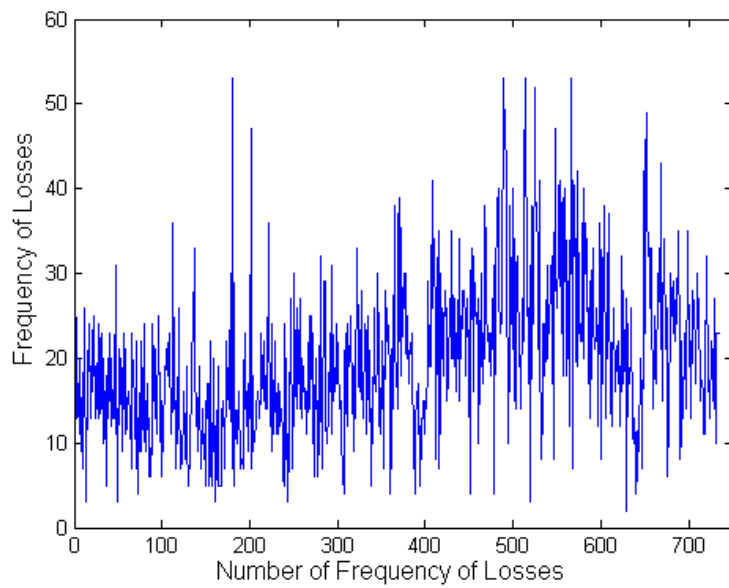
Σχήμα 14. Boxplot των δεδομένων απαιτήσεων

Περιγραφικά	Τιμή
Μέση Τιμή	20
Διακύμανση	84.05
Τυπική Απόκλιση	9.16
Ελάχιστη Τιμή	2
Μέγιστη Τιμή	53
Συντελεστής Ασυμμετρίας	0.75
Κυρτότητα	3.72

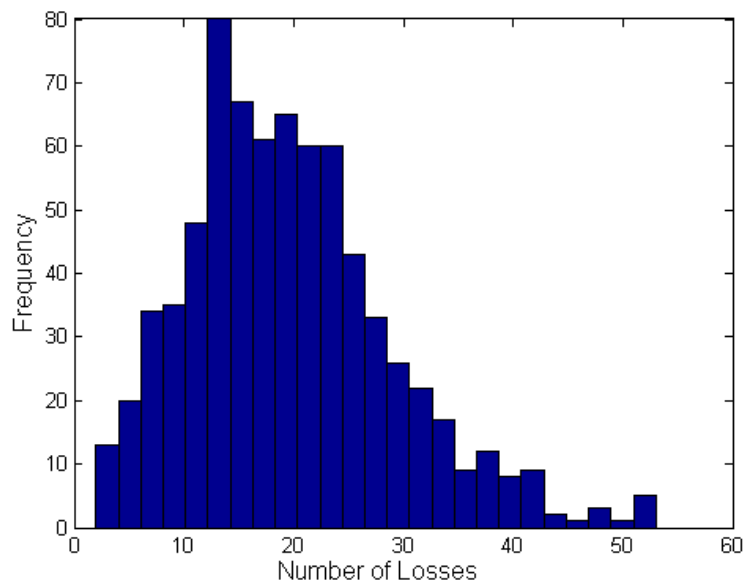
Πίνακας 15. Πίνακας περιγραφικών στοιχείων δεδομένων συχνοτήτων

αναφέρουμε συνοπτικά μερικά περιγραφικά στοιχεία των δεδομένων μας ενώ στην παράγραφο 5.1 γίνεται μια ολοκληρωμένη μελέτη για τον προσδιορισμό των κατανομών συχνοτήτων και απωλειών που περιγράφουν τα δεδομένα μας.

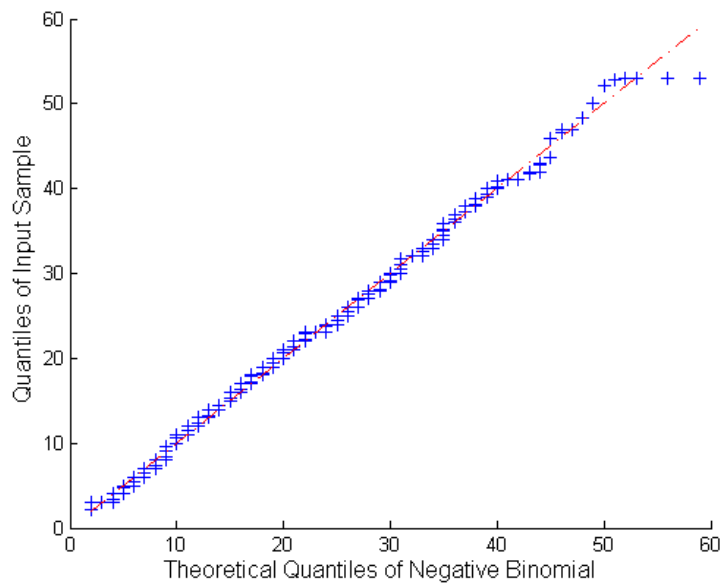
Από τα παρακάτω διαγράμματα παρατηρούμε ότι είναι πολύ πιθανό η κατανομή που περιγράφει τα δεδομένα συχνοτήτων να είναι η αρνητική διωνυμική.



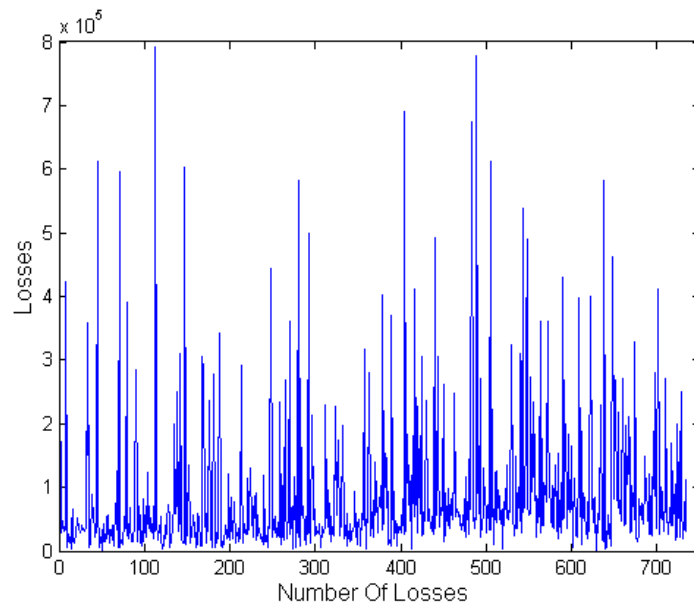
Σχήμα 15. Γραφική απεικόνιση των δεδομένων συχνοτήτων.



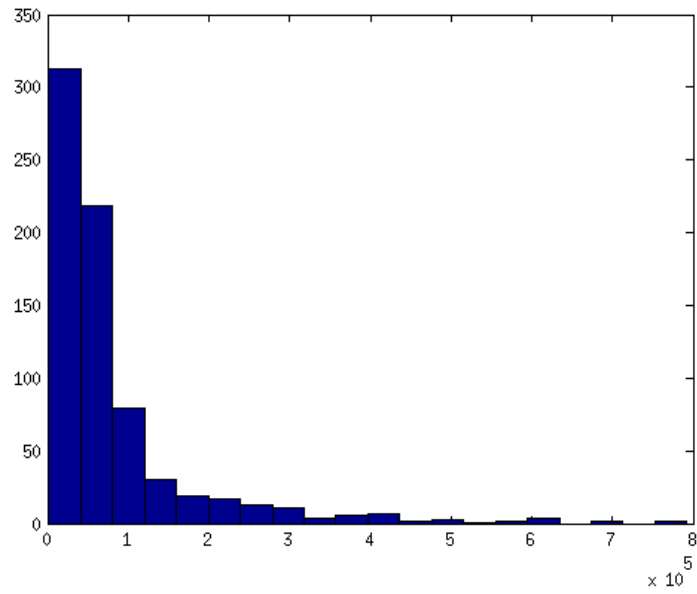
Σχήμα 16. Ιστόγραμμα των δεδομένων συχνοτήτων.



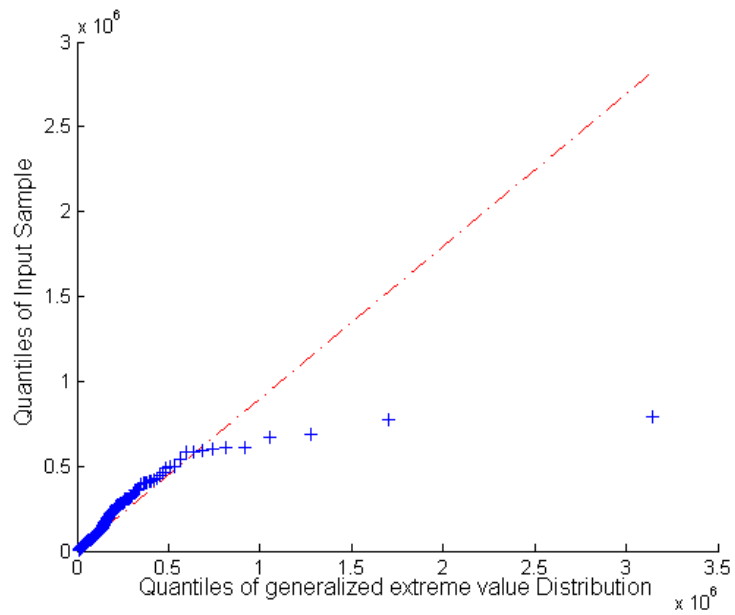
Σχήμα 17. QQ-plot των δεδομένων συχνότητας



Σχήμα 18. Γραφική απεικόνιση των δεδομένων απαιτήσεων.



Σχήμα 19. Ιστόγραμμα των δεδομένων απαιτήσεων.



Σχήμα 20. QQ-plot των δεδομένων απαιτήσεων.

Περιγραφικά	Τιμή
Μέση Τιμή	$8.4 \cdot 10^4$
Διακύμανση	$1.17 \cdot 10^{10}$
Τυπική Απόκλιση	$1.08 \cdot 10^5$
Ελάχιστη Τιμή	$1.77 \cdot 10^3$
Μέγιστη Τιμή	$7.92 \cdot 10^5$
Συντελεστής Ασυμμετρίας	3.16
Κυρτότητα	14.92

Πίνακας 16. Πίνακας περιγραφικών στοιχείων δεδομένων απαιτήσεων

Ημερομηνία	Ύψος Ζημίας
18/12/2007	$6.74 \cdot 10^5$
30/9/2007	$6.91 \cdot 10^5$
28/12/2007	$7.77 \cdot 10^5$
29/12/2006	$7.92 \cdot 10^5$

Πίνακας 17. Ύψη ζημιών που αποκλίνουν από το διάγραμμα Q-Q plot.

Κάνοντας ένα διάγραμμα ποσοστημορίων (QQ-plot) των δεδομένων μας σε αντιπαραβολή με τα θεωρητικά ποσοστημόρια της αρνητικής διωνυμικής κατανομής αυτή η υποψία ισχυροποιείται (Σχήμα 17). Όσον αφορά τα δεδομένα απωλειών βλέπουμε ότι ακολουθούν κατανομή με βαριά ουρά. Από το διάγραμμα ποσοστημορίων έχουμε ενδείξεις ότι η Γενικευμένη κατανομή ακραίων τιμών είναι αυτή που προσαρμόζεται καλύτερα στα δεδομένα μας.

Στο Σχήμα 20, υπάρχουν τιμές που αποκλίνουν από την διχοτόμο. Τα σημεία αυτά αντιστοιχούν σε τιμές του δείγματος απωλειών που φαίνονται στον Πίνακα 17. Θα μπορούσε κανείς να ισχυριστεί ότι ιδιαίτερα για τις δύο τελευταίες τιμές πρόκειται για συσσώρευση ζημιών λόγω της λήξης του εκάστοτε διαχειριστικού έτους χρήσης.

5.1 Συχνότητα και Ύψος Απαιτήσεων

Δυο από τις πιο συνήθεις κατανομές που ακολουθούν οι συχνότητες ζημιών είναι η διωνυμική ή η κατανομή Poisson [42]. Το γεγονός ότι η διακύμανση του αριθμού των συχνοτήτων είναι αρκετά μεγαλύτερη από την μέση τιμή είναι ένα επιπλέον δείγμα που μας οδηγεί να πιστέψουμε ότι δεδομένα συχνοτήτων μας φαίνεται να ακολουθούν την αρνητική διωνυμική και όχι την κατανομή Poisson όπως γίνεται συνήθως.

Ας υποθέσουμε ότι έχουμε μια σειρά ανεξάρτητων δοκιμών Bernoulli με κάθε μία δοκιμή να έχει δύο πιθανά ενδεχόμενα τα οποία ονομάζουμε επιτυχία και αποτυχία. Σε κάθε δοκιμή η πιθανότητα της επιτυχίας είναι p και η πιθανότητα της αποτυχίας είναι $(1 - p)$. Παρατηρούμε αυτή τη σειρά δοκιμών μέχρι να έχει ολοκληρωθεί ένας προκαθορισμένος αριθμός αποτυχιών, έστω r . Τότε ο τυχαίος αριθμός επιτυχιών, X , τις οποίες έχουμε παρατηρήσει θα ακολουθεί την αρνητική διωνυμική κατανομή δηλαδή $X \sim NB(r, p)$.

Η συνάρτηση πιθανότητας της αρνητικής διωνυμικής δίνεται από τη σχέση,

$$f(x; r, p) \equiv Pr(X = x) = \binom{x+r-1}{r-1} (1-p)^x p^r, \quad x = 0, 1, 2, \dots \quad (226)$$

Η μέση τιμή και η διακύμανση της αρνητικής διωνυμικής θα είναι

$$E(X) = \frac{r(1-p)}{p}, \quad (227)$$

$$Var(X) = \frac{r(1-p)}{p^2}. \quad (228)$$

Η ροπογεννήτρια της κατανομής δίνεται από την παρακάτω σχέση:

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \binom{x+r-1}{r-1} (1-p)^x p^r. \quad (229)$$

Για να υπολογίσουμε τις εκτιμήτριες της παραπάνω κατανομής χρησιμοποιούμε την μέθοδο μέγιστης πιθανοφάνειας. Ο λογάριθμος της συνάρτησης πιθανοφάνειας θα είναι,

$$l(r, p) = \sum_{i=1}^n \ln(\Gamma(x_i + r)) - \sum_{i=1}^n \ln x_i! - N \ln \Gamma(r) + \sum_{i=1}^n x_i \ln(p) + Nr \ln(1 - p). \quad (230)$$

Για να βρούμε το μέγιστο λαμβάνουμε τις μερικές παραγώγους ως προς τις παραμέτρους r, p και θέτουμε τις εξισώσεις που προκύπτουν ίσες με μηδέν δηλαδή,

$$\frac{\partial l(r, p)}{\partial p} = \sum_{i=1}^n x_i \frac{1}{p} - nr \frac{1}{1-p} = 0, \quad (231)$$

$$\frac{\partial l(r, p)}{\partial r} = \sum_{i=1}^n \psi(x_i + r) - n\psi(r) + n \ln(1 - p) = 0, \quad (232)$$

όπου $\psi(k)$ είναι η συνάρτηση δι-γάμμα. Λύνοντας την Σχέση 231 και αντικαθιστώντας στην Σχέση 232 λαμβάνουμε το παρακάτω σύστημα εξισώσεων,

$$p = \frac{\sum_{i=1}^n x_i}{nr + \sum_{i=1}^n x_i}, \quad (233)$$

$$\sum_{i=1}^n \psi(x_i + r) - n\psi(r) + n \ln\left(\frac{\sum_{i=1}^n x_i}{nr + \sum_{i=1}^n x_i}\right) = 0. \quad (234)$$

Λύνοντας το παραπάνω σύστημα χρησιμοποιώντας κάποια αριθμητική μέθοδο λαμβάνουμε τις εκτιμήτριες \hat{r}, \hat{p} των παραμέτρων της αρνητικής διωνυμικής. Χρησιμοποιώντας τα δεδομένα συχνοτήτων οι εκτιμήτριες των παραμέτρων και τα αντίστοιχα διαστήματα εμπιστοσύνης με συντελεστή 95% είναι

$$\hat{r} = 5.97 \quad , \quad [5.173, 6.773], \quad (235)$$

$$\hat{p} = 0.23 \quad , \quad [0.207, 0.256]. \quad (236)$$

Ο στατιστικός έλεγχος χ^2 (Chi-squared) που χρησιμοποιείται συνήθως για τον έλεγχο καλής προσαρμογής δεδομένων επιβεβαιώνει τον ισχυρισμό μας επαληθεύοντας την μηδενική υπόθεση. Σύμφωνα με τον έλεγχο αυτό υπολογίζεται η παρακάτω ποσότητα

$$Q = \sum_{k=0}^{\infty} \frac{(n_k - E_k)^2}{E_k}, \quad (237)$$

όπου E_k είναι ο αναμενόμενος αριθμός παρατηρήσεων. Ο αριθμός βαθμών ελευθερίας υπολογίζεται ως ο αριθμός των κελιών μείον τον αριθμό των εκτιμημένων παραμέτρων αφού έχουμε αφαιρέσει τη μονάδα. Άρα τελικά η μηδενική υπόθεση επαληθεύεται όταν ισχύει

$$Q < \chi_{d,a}^2, \quad (238)$$

όπου a είναι το επίπεδο σημαντικότητας και d οι βαθμοί ελευθερίας.

Η αξία σε κίνδυνο για την αρνητική διωνυμική δίνεται από τον τύπο [22]

$$VaR_\alpha(p) = \inf\{c \in N_0 | H_{r,c}(p) \geq \alpha\} \quad \text{για } \alpha \in (0, 1). \quad (239)$$

Ενώ το αναμενόμενο έλλειμα δίνεται από την παρακάτω σχέση [22]

$$cVaR_\alpha(p) = \frac{\frac{r(1-p)}{p}[1 - H_{r,z_p}(p) + \binom{r+z(p)}{z(p)}p^r(1-p)^{z(p)} + z(p)[H_{r,p}(p) - \alpha]]}{1 - \alpha}, \quad (240)$$

όπου $H_{r,p}$ είναι η αντίστροφη συνάρτηση κατανομής της αρνητικής διωνυμικής. Στον Πίνακα 18 παραθέτουμε τις τιμές των μέτρων κινδύνου για επίπεδα εμπιστοσύνης 95% και 99%.

Επίπεδο εμπιστοσύνης	VaR	ES
95%	38	44
99%	47	52

Πίνακας 18. *VaR και ES δεδομένων συχνότητας.*

Παρατηρούμε ότι παρόλο που το αναμενόμενο έλλειμα εξόρισμού, είναι πιο αυστηρό μέτρο κινδύνου, χρησιμοποιώντας μεγαλύτερο επίπεδο εμπιστοσύνης (99%) η αξία σε κίνδυνό μας οδηγεί σε υψηλότερη συχνότητα ημερήσιων απωλειών.

Ας υποθέσουμε ότι θέλουμε να προσαρμόσουμε στα δεδομένα ύψους ζημιών την κανονική κατανομή ή την κατανομή Pareto. Με έναν γρήγορο γραφικό έλεγχο παρατηρούμε ότι οι δύο αυτές κατανομές δεν αποτελούν καλή επιλογή (βλ. αντίστοιχα Q-Q Plots στα σχήματα 21, 22).

Συγκρίνοντας την εμπειρική κατανομή των δεδομένων μας και την γενικευμένη κατανομή ακραίων τιμών παρατηρούμε ότι βρίσκονται πολύ κοντά (βλ Σχήμα 23). Εφαρμόζοντας τον στατιστικό έλεγχο Kolmogorov-Smirnov, ο οποίος ποσοτικοποιεί αυτή τη διαφορά, επιβεβαιώνεται ο ισχυρισμός μας. Ο υπολογισμός της εν λόγω διαφοράς βασίζεται στην μεγαλύτερη κάθετη απόσταση μεταξύ της εμπειρικής και της θεωρητικής κατανομής, έτσι δεν γίνεται οποιαδήποτε υπόθεση για την κατανομή που μπορεί να ακολουθούν τα προς μελέτη δεδομένα.

Προσαρμόζοντας λοιπόν τη γενικευμένη κατανομή ακραίων τιμών στα δεδομένα μας λαμβάνουμε τις παρακάτω εκτιμήτριες και τα αντίστοιχα διαστήματα εμπιστοσύνης με συντελεστή 95%.

$$\hat{\xi} = 0.55, \quad [0.4795, 0.6283], \quad (241)$$

$$\hat{\mu} = 30874.3, \quad [28438, 33519] \quad (242)$$

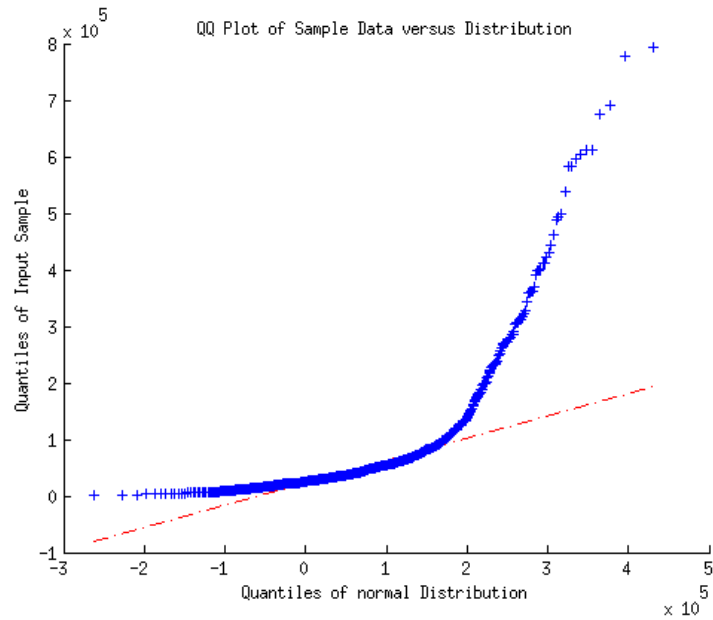
$$\hat{\sigma} = 36065.9, \quad [33513.8, 38618] \quad (243)$$

Η εκτιμήτρια της παραμέτρου ξ που καθορίζει ποια κατανομή ακραίων τιμών ακολουθούν τα δεδομένα μας είναι θετική άρα τα δεδομένα ζημιών ακολουθούν την κατανομή Frechét.

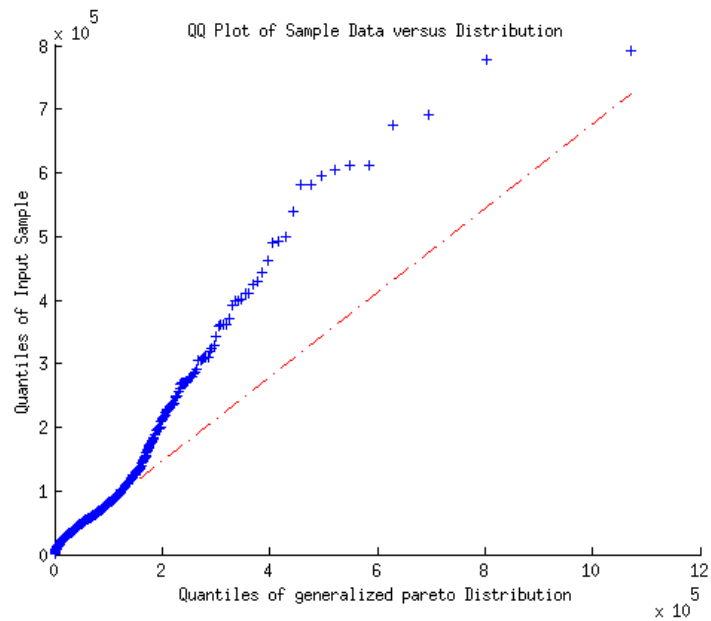
Στον Πίνακα 19 παραθέτουμε τις τιμές των μέτρων κινδύνου για επίπεδα εμπιστοσύνης 95% και 99% που υπολογίσαμε από τα δεδομένα μας. Οι τύποι για τον υπολογισμό των μέτρων κινδύνου για τη γενικευμένη κατανομή ακραίων τιμών δίνονται από τις σχέσεις 206 και 208.

Επίπεδο εμπιστοσύνης	VaR	ES
95%	$3.06 \cdot 10^5$	$4.61 \cdot 10^5$
99%	$5.97 \cdot 10^5$	$6.80 \cdot 10^5$

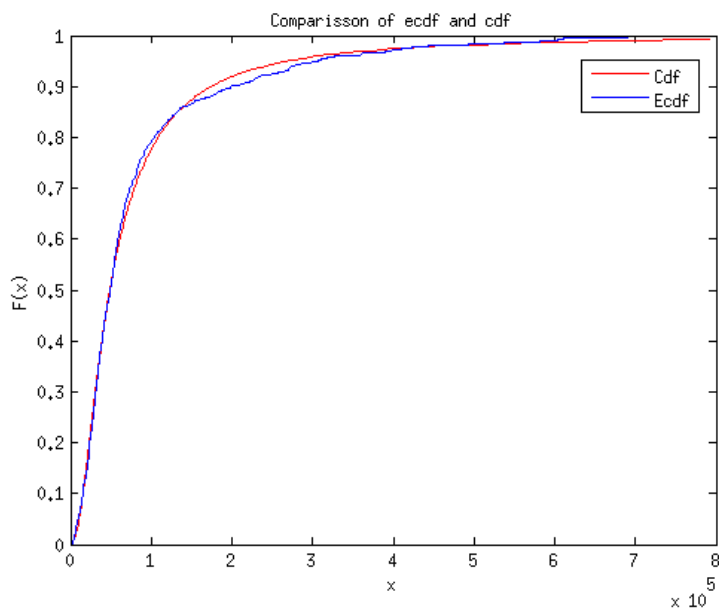
Πίνακας 19. Var και ES δεδομένων απαιτήσεων.



Σχήμα 21. QQ-plot των δεδομένων απαιτήσεων προσαρμόζοντας την κανονική κατανομή.



Σχήμα 22. QQ-plot των δεδομένων απαιτήσεων προσαρμόζοντας τη γενικευμένη Pareto.



Σχήμα 23. Σύγκριση εμπειρικής και θεωρητικής κατανομής.

Έχοντας εκτιμήσει τις παραμέτρους της κατανομής που ακολουθούν τα ύψη ζημιών μπορούμε να παράξουμε επαρκή αριθμό δειγμάτων υπολογίζοντας τα μέτρα κινδύνου και να τα συγκρίνουμε με τις πραγματικές τιμές που έχουμε υπολογίσει από τα δεδομένα μας. Όπως φαίνεται από την σύγκριση του Πίνακα 19 και του Πίνακα 20 οι τιμές του VaR και ES είναι αρκετά κοντά το οποίο επαληθεύει ότι τα δεδομένα μας ακολουθούν τη γενικευμένη κατανομή ακραίων τιμών. Να σημειώσουμε εδώ ότι χρησιμοποιήσαμε 10^5 δείγματα με πλήθος τιμών 10^5 .

Επίπεδο εμπιστοσύνης	VaR	ES
95%	$2.92 \cdot 10^5$	$4.59 \cdot 10^5$
99%	$5.89 \cdot 10^5$	$6.78 \cdot 10^5$

Πίνακας 20. Var και ES προσωμοιωμένων δεδομένων απαιτήσεων.

```
1 % check var and cvar in scenarios with these parameters
2 ksi= par_est_mle(1);
3 sig = par_est_mle(2);
4 m = par_est_mle(3);
5 losses_scenario = gevrnd(ksi, sig, m,10^5,10^5);
6 prc_scenario_loss = prctile(losses_scenario, p);
7 prc_final = sum(prc_scenario_loss(1,:)) / (10^3)
```

5.2 Κατανομή Συνολικών Απωλειών και Ανάλυση Ευαισθησίας

Όπως συζητήσαμε στην εισαγωγή του κεφαλαίου μπορούμε μέσω της προσωμοίωσης Monte Carlo, να δημιουργήσουμε σενάρια για το ύψος και τη συχνότητα ζημιών. Συγκεκριμένα, μέσω αυτής και σε επίπεδο χαρτοφυλακίου, μπορούμε να υπολογίσουμε την κατανομή συνολικών απωλειών και τα μέτρα κινδύνου για δεδομένο επίπεδο εμπιστοσύνης κάνοντας έτσι προβλέψεις για το ακραίο ύψος ζημιών σε προκαθορισμένο χρονικό ορίζοντα (για παράδειγμα, ενός μήνα ή ενός έτους.)

Έστω N η συχνότητα των απαιτήσεων από συμβάντα ζημιών περιουσίας σε ένα προκαθορισμένο χρονικό ορίζοντα και έστω X_1, X_2, \dots η πρώτη, η δεύτερη κ.ο.κ απαίτηση ζημίας. Στο αθροιστικό μοντέλο κινδύνου, οι αθροιστικές ζημιές που πηγάζουν από το χαρτοφυλάκιο στην υπο μελέτη χρονική περίοδο, περιγράφονται από το τυχαίο άθροισμα

$$S_N = \sum_{j=1}^N X_j. \quad (244)$$

Ο αριθμός των ζημιών N , είναι τυχαία μεταβλητή και σχετίζεται με τη συχνότητά τους. Κάθε X_j είναι επίσης τυχαία μεταβλητή η οποία περιγράφει το ύψος των απωλειών. Έτσι, η κατανομή συνολικών απωλειών θα περιγράφεται από το αθροιστικό μοντέλο κινδύνου της μορφής

$$S = X_1 + X_2 + \dots + X_N. \quad (245)$$

Να σημειώσουμε εδώ ότι δύο βασικές υποθέσεις που γίνονται στο συγκεκριμένο μοντέλο είναι ότι οι τυχαίες μεταβλητές X_1, X_2, \dots είναι ισόνομα κατανομημένες και ότι οι τυχαίες μεταβλητές N, X_1, X_2, \dots είναι αμοιβαίως ανεξάρτητες. Η αναμενόμενη μέση τιμή των συνολικών απωλειών είναι το γινόμενο της αναμενόμενης τιμής του αριθμού των απωλειών και του ατομικού ύψους ζημίας δηλαδή,

$$E(S) = E(N)E(X), \quad (246)$$

ενώ η διακύμανση των συνολικών απωλειών είναι το άθροισμα δύο παραγόντων από τους οποίους ο πρώτος αναφέρεται στη μεταβλητότητα του ατομικού ύψους ζημίας και ο δεύτερος στη μεταβλητότητα του αριθμού των απωλειών. Πιο συγκεκριμένα ισχύει ο τύπος

$$\text{Var}(S) = E(N)\text{Var}(X) + E^2(X)\text{Var}(N). \quad (247)$$

Αν υποθέσουμε ότι οι ροπογεννήτριες των κατανομών συχνοτήτων και ύψους ζημιών υπάρχουν τότε

$$M_S(t) = M_N(\ln M_X(t)). \quad (248)$$

Όπως αναφέραμε στην εισαγωγή του κεφαλαίου, ο προσδιορισμός της κατανομής της S είναι αρκετά περίπλοκος έτσι χρησιμοποιούμε προσωμοίωση. Ο λόγος για τον οποίο μας ενδιαφέρει να βρούμε την κατανομή συνολικών απωλειών είναι για να προσδιορίσουμε για παράδειγμα το ασφάλιστρο ενός συμβολαίου μέσω της πρόβλεψης μελλοντικών ζημιών. Η ποσότητα της οποίας ο υπολογισμός μας ενδιαφέρει συνήθως είναι τα ποσοστημόρια της κατανομής αυτής σε ένα δεδομένο επίπεδο εμπιστοσύνης δηλαδή τα

$$q_\alpha = F_S^{-1}(1 - \alpha), \quad \alpha \in (0, 1). \quad (249)$$

Η προσωμοίωση Monte Carlo που χρησιμοποιήθηκε στην παρούσα εργασία μπορεί να συνοψιστεί στα ακόλουθα βήματα [49]:

- Δημιουργία $i = 1, \dots, 10^5$ τυχαίου αριθμού σεναρίων για τη συχνότητα ζημιών απο την αρνητική διωνυμική με παραμέτρους (\hat{r}, \hat{p}) .
- Δημιουργία αντίστοιχων $m(i)$ τυχαίου αριθμού σεναρίων για το ύψος ζημιών απο την γενικευμένη κατανομή ακραίων τιμών με παραμέτρους $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.
- Υπολογισμός της συνολικής ζημίας για κάθε σενάριο από το προηγούμενο βήμα αθροίζοντας τις τιμές $m(i)$.
- Εύρεση των μέτρων κινδύνου.

Να σημειωθεί ότι οι παραπάνω εκτιμήτριες έχουν υπολογισθεί στα πραγματικά δεδομένα ζημιών που συζητήσαμε στη προηγούμενη παράγραφο και είναι οι ακόλουθες

$$(\hat{r}, \hat{p}) = (5.97, 0.23), \quad (250)$$

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (30874.3, 36065.9, 0.55). \quad (251)$$

Ο παραπάνω αλγόριθμός υλοποιήθηκε με την βοήθεια της συνάρτησης `cellfun` του Matlab η οποία μας επιτρέπει να χρησιμοποιούμε πολύ λιγότερες επαναληπτικές ρουτίνες. Η ιδέα γύρω από την υλοποίηση αυτή μπορεί να φανεί στον παρακάτω κώδικα στο οποίο παραθέτουμε μόνο το κύριο κομμάτι του υπολογισμού παραλείποντας τις λεπτομέρειες.

```

1 % create N numbers of freq from negative binomial and
   corresponding losses from gev
2 ksi= par_est_mle(1);
3 sig = par_est_mle(2);
4 m = par_est_mle(3);
5 N = 1e5;
6 Freq_scenario = num2cell(nbinrnd(fitnbin(1), fitnbin(2), N, 1));
7 Losses_scenario = cellfun(@(x) sum(gevrnd(ksi, sig, m, x, 1)), ...
8                           Freq_scenario, '
                               UniformOutput', false
                               );

```

Στον παρακάτω πίνακα παραθέτουμε κάποια περιγραφικά στοιχεία της κατανομής συνολικών απωλειών.

Περιγραφικά	Τιμή
Μέση Τιμή	$1.80 \cdot 10^7$
Διακύμανση	$2.39 \cdot 10^{12}$
Τυπική Απόκλιση	$1.54 \cdot 10^6$
Ελάχιστη Τιμή	$6.30 \cdot 10^3$
Μέγιστη Τιμή	$7.94 \cdot 10^7$
Συντελεστής Ασυμμετρίας	11.69
Κυρτότητα	370.80

Πίνακας 21. Πίνακας περιγραφικών στοιχείων αθροιστικής κατανομής απωλειών

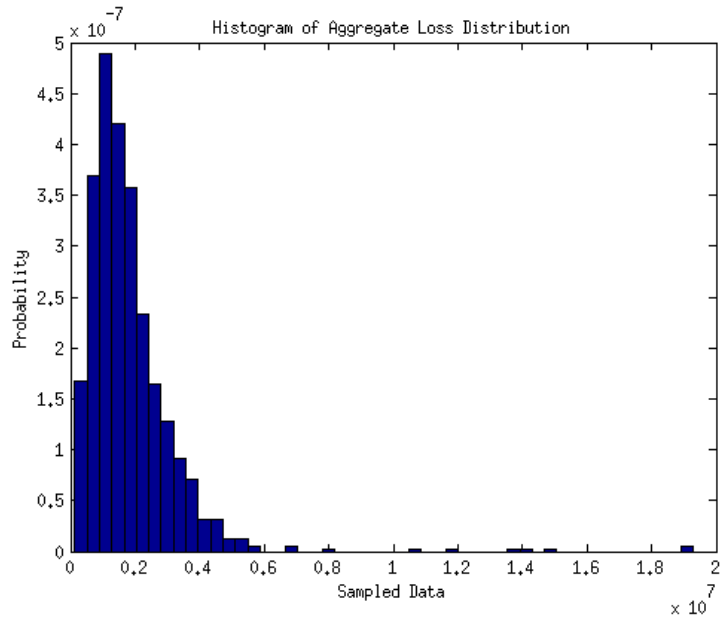
Μετά από τους παραπάνω υπολογισμούς παραθέτουμε στο Σχήμα 24 τη συνάρτηση πυκνότητας πιθανότητας της κατανομής συνολικών ζημιών μέσω ενός κανονικοποιημένου ιστογράμματος ενώ στο Σχήμα 25 προσαρμόζουμε με βάση το ιστόγραμμα συχνοτήτων την ομαλοποιημένη κατανομή πυκνότητας (density fit).

Σύμφωνα με την παραπάνω μελέτη η κατανομή που προσαρμόζεται στα δεδομένα συχνοτήτων είναι η αρνητική διωνυμική ενώ τα ύψη ζημιών ακολουθούν την γενικευμένη κατανομή ακραίων ζημιών. Έχοντας λοιπόν προσομοιώσει την κατανομή συνολικών απωλειών μπορούμε να υπολογίσουμε τις τιμές των μέτρων κινδύνου για διάφορες τιμές του συντελεστή εμπιστοσύνης στον Πίνακα 22.

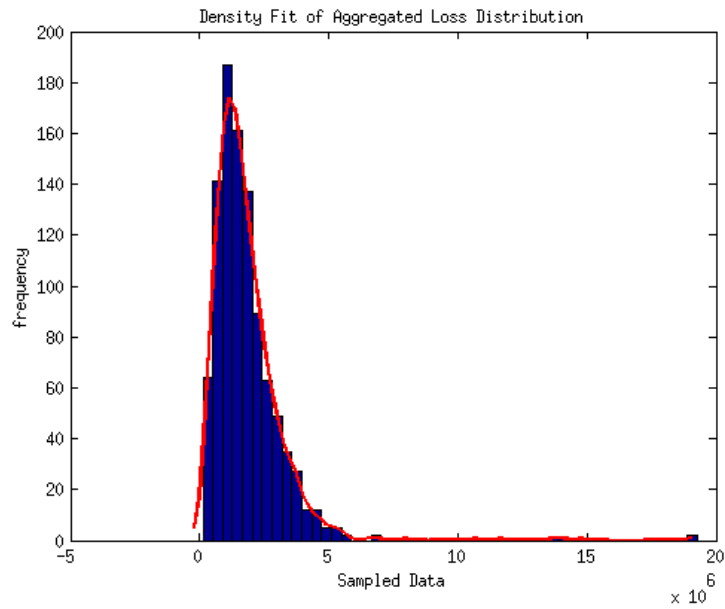
Επίπεδο εμπιστοσύνης	95%	99%	99.9%
VaR	$0.39 \cdot 10^7$	$0.62 \cdot 10^7$	$1.51 \cdot 10^7$
ES	$0.58 \cdot 10^7$	$1.05 \cdot 10^7$	$3.16 \cdot 10^7$

Πίνακας 22. VaR και ES της αθροιστικής κατανομής απωλειών.

Οι τιμές των VaR και ES που θεωρούνται λογικές, επαφίονται στην κατάλληλη επιλογή της κατανομής συνολικών ζημιών. Αν η κατανομή που έχουμε διαλέξει να περιγράψει τα δεδομένα μας δεν είναι η καταλληλότερη τότε μπορεί να οδηγηθούμε σε εντελώς λανθασμένα συμπεράσματα. Ειδικότερα στην περι-



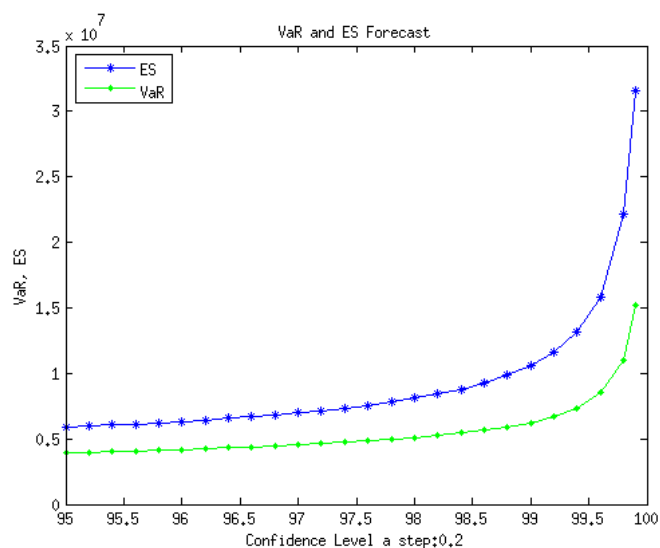
Σχήμα 24. Ιστόγραμμα πιθανότητας της αθροιστικής κατανομής απωλειών.



Σχήμα 25. Ιστόγραμμα συχνοτήτων της αθροιστικής κατανομής απωλειών.

πτωση όπου η καταλληλότερη κατανομή έχει πιο βαριά ουρά, οι τιμές του ES θα αποκλίνουν αισθητά.

Παρακάτω, αναφορικά με το επίπεδο εμπιστοσύνης, προχωράμε στη συνέχεια σε ανάλυση ευαισθησίας του VaR και ES σε σχέση με το επίπεδο εμπιστοσύνης για τον χρονικό ορίζοντα τριών ετών. Το βήμα αύξησης του επιπέδου εμπιστοσύνης έχει τεθεί ίσο με 0.2.



Σχήμα 26. Σχηματική απεικόνιση των μέτρων κινδύνου μεταβάλλοντας το επίπεδο εμπιστοσύνης.

```

1 j=95:.2:99.9;
2 j(length(j)+1) = 99.9;
3 VaR_comp = prctile(total_loss, j);
4 for i=1:length(j)
5 CVaR_comp(i) = mean(total_loss(total_loss>VaR_comp(i)));
6 end

```

Όπως ήταν αναμενόμενο παρατηρούμε ότι καθώς το επίπεδο εμπιστοσύνης αυξάνεται, το VaR και ES αυξάνονται. Για τα επίπεδα εμπιστοσύνης 95% έως 98% παρατηρούμε μια μικρή σχεδόν γραμμική αύξηση του VaR και ES. Για τιμές του επιπέδου εμπιστοσύνης άνω του 98,5% η κλίση των καμπυλών αυξάνεται απότομα. Φαίνεται ότι τα VaR και ES είναι περισσότερο ευαίσθητα σε μεγαλύτερα επίπεδα εμπιστοσύνης ενώ το ES ακολουθεί παρόμοια συμπεριφορά με το VaR η οποία όμως είναι πιο έντονη καθώς το επίπεδο εμπιστοσύνης αυξάνεται.

Η επιλογή του κατάλληλου επιπέδου εμπιστοσύνης εξαρτάται όχι μόνο από το επίπεδο του κινδύνου που μπορεί να αναλάβει μια ασφαλιστική επιχείρηση (risk appetite) αλλά και από το κόστος που αντιμετωπίζει κάνοντας προβλέψεις χρησιμοποιώντας υψηλά επίπεδα εμπιστοσύνης και άρα διακρατώντας υψηλά κεφάλαια για την αντιμετώπιση ζημιών. Σύμφωνα με τα παραπάνω θα προτείναμε ένα επίπεδο εμπιστοσύνης στο 98.8%.

6 Σύνοψη

Στα πρώτα δύο κεφάλαια της εργασίας γίνεται εκτενής αναφορά στα βασικά θεωρήματα της θεωρίας ακραίων τιμών μέσω παραδειγμάτων αλλά και μια σύντομη ιστορική αναδρομή που αναδεικνύει την χρησιμότητα της σε διάφορες επιστήμες εστιάζοντας κυρίως στις ασφαλίσεις. Οι κατανομές Gumbel, Frechet και Weibull αναλύονται στο τρίτο κεφάλαιο όπου αναπτύσσονται τα χαρακτηριστικά και οι μέθοδοι εκτίμησής τους. Οι τρεις μέθοδοι που χρησιμοποιούνται για την εκτίμηση των παραμέτρων των κατανομών ακροτάτων είναι οι μέθοδοι μέγιστης πιθανοφάνειας, ροπών και σταθμισμένων ροπών πιθανότητας. Η αποτελεσματικότητα των μεθόδων αξιολογείται με βάση την αμεροληψία και το μέσο τετραγωνικό σφάλμα χρησιμοποιώντας προσωμοίωση Monte Carlo για μικρά, μεσαία και μεγάλα δείγματα για διάφορες τιμές των παραμέτρων.

Η γενικευμένη κατανομή ακραίων τιμών αναλύεται στο τέταρτο κεφάλαιο όπου αναφέρονται οι μέθοδοι εκτίμησης των παραμέτρων της. Έπειτα γίνεται μια αναφορά μεθόδων προσδιορισμού του δείκτη ουράς (μέθοδοι Hill, Pickands), από τον οποίο γίνεται ο προσδιορισμός της κατάλληλης κατανομής ακροτάτων. Επιπρόσθετα, παρουσιάζουμε τα μέτρα κινδύνου και τις δύο κυριότερες προσεγγίσεις προσδιορισμού και μελέτης της κατανομής ακραίων παρατηρήσεων ενός δείγματος, αυτές των μεθόδων Block maxima και Peaks over threshold.

Στην αριθμητική εφαρμογή προσαρμόζουμε την γενικευμένη κατανομή ακραίων τιμών σε πραγματικά δεδομένα ζημιών από όπου γίνεται σαφές ότι δε θα μπορούσαν να περιγραφούν χρησιμοποιώντας άλλη κατανομή. Υπολογίζουμε τα μέτρα κινδύνου τα οποία είναι απαραίτητα για τον προσδιορισμό σημαντικών ποσοτήτων που σε καθημερινή βάση χρειάζεται μια ασφαλιστική εταιρεία και αφορούν για παράδειγμα τον προσδιορισμό του ασφαλιστρού αλλά και της κεφαλαιακής επάρκειας για κάλυψη ζημιών (εποπτικός έλεγχος). Προσομοιώνοντας σενάρια για τα ύψη ζημιών από τη γενικευμένη κατανομή ακραίων τιμών, παρατηρούμε ότι τα μέτρα κινδύνου είναι πολύ κοντά στις πραγματικές τιμές. Τέλος, εστιάζουμε στη δημιουργία της κατανομής συνολικών απωλειών ενός χαρτοφυλακίου, δημιουργώντας σενάρια για τις κατανομές που ακολουθούν η συχνότητα και τα ύψη ζημιών, με τη βοήθεια προσωμοίωσης Monte Carlo. Έ-

τσι, υπολογίζοντας τα μέτρα κινδύνου προβλέπουμε ουσιαστικά τις μέγιστες απώλειες που μπορούν να προκύψουν στο εν λόγω ασφαλιστικό χαρτοφυλάκιο.

Η θεωρία ακραίων τιμών καθίσταται ολοένα και πιο σημαντική κερδίζοντας έδαφος ιδιαίτερα στον τομέα των ασφαλίσεων. Παρατηρούμε μάλιστα ότι σε κάποιες περιπτώσεις η εφαρμογή της είναι αναγκαία καθώς οι παραδοσιακές μέθοδοι αποτυγχάνουν. Η γενικευμένη κατανομή ακραίων τιμών και οι ειδικές περιπτώσεις αυτής περιγράφουν με επιτυχία δεδομένα ζημιών που εμπεριέχουν ακραίες παρατηρήσεις. Έτσι, είναι πιο εύκολο να απαντήσουμε σε ερωτήματα που αφορούν στη πρόβλεψη της συχνότητας αλλά και της δριμύτητας απωλειών καθώς και την πιθανότητα με την οποία μπορούν να συμβούν με τη βοήθεια προσομοιώσεων που λαμβάνουν υπόψιν ιστορικά δεδομένα. Η μαθηματική θεμελίωση της θεωρίας ακραίων τιμών, που παρουσιάζεται αναλυτικά στη παρούσα εργασία, συνεχίζεται με ταχείς ρυθμούς καθώς το επιστημονικό αυτό πεδίο έχει ακόμη πολλές δυνατότητες, θεωρητικές και πρακτικές, εφαρμόσιμες σε πολλές επιστήμες.

7 Παράρτημα

7.1 Σταθερά Euler

Η σταθερά Euler, είναι το όριο στο άπειρο της διαφοράς της αρμονικής σειράς μείον το φυσικό λογάριθμο, δηλαδή

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \ln(n) \right) \simeq 0,5772.. \quad (252)$$

7.2 Συνάρτηση Γάμμα

Η συνάρτηση Γάμμα ή αλλιώς ολοκλήρωμα δεύτερου είδους του Euler δίνεται από τον παρακάτω τύπο

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx. \quad (253)$$

Αποδεικνύεται εύκολα ότι

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1). \quad (254)$$

Για $\alpha = n + 1$, $n \in \mathbb{Z}_+$ η παραπάνω Σχέση δίνει $\Gamma(n + 1) = n\Gamma(n)$ και με συνεχόμενη εφαρμογή του τύπου αυτού προκύπτει ότι $\Gamma(n + 1) = n!$.

7.3 Συνάρτηση Ζήτα

Η συνάρτηση ζ του Riemann είναι το άθροισμα της άπειρης σειράς

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s = \sigma + it, \quad (255)$$

η οποία συγκλίνει μόνο όταν το πραγματικό μέρος του s είναι θετικό, $\Re(s) > 1$. Συγκεκριμένα για $s = 3$ θα έχουμε

$$\zeta(3) = 1 + \frac{1}{2^3} + \frac{1}{3^3} + \dots \simeq 1,202. \quad (256)$$

Η $\zeta(3)$ ονομάζεται σταθερά του Apéry, ο οποίος απέδειξε ότι είναι άρρητος αριθμός.

7.4 Συνάρτηση δι-Γάμμα

Η συνάρτηση δι-Γάμμα ορίζεται ως η παράγωγος του λογαρίθμου της συνάρτησης Γάμμα δηλαδή:

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}. \quad (257)$$

Συνδέεται με τους αρμονικούς αριθμούς μέσω της παρακάτω σχέσης,

$$\psi(n) = H_{n-1} - \gamma, \quad (258)$$

όπου H_n είναι ο n -οστός αρμονικός αριθμός

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k}$$

και γ η σταθερά Euler. Η συνάρτηση δι-Γάμμα γράφεται σε ολοκληρωτική μορφή ως εξής:

$$\psi(x) = \int_0^{\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1 - e^{-t}} \right) dt. \quad (259)$$

Για τη σταθερά Euler ισχύουν οι τύποι:

$$\gamma = -\Gamma'(1) = -\psi(1). \quad (260)$$

7.5 Συνεκτικό μέτρο κινδύνου

Έστω το σύνολο όλων των κινδύνων. Ένα μέτρο κινδύνου, έστω ρ , είναι συνεκτικό όταν για δύο κινδύνους X, Y ικανοποιούνται οι παρακάτω ιδιότητες:

$$\begin{aligned} (\text{Κανονικότητα}) \quad & \rho(0) = 0, \\ (\text{Υποπροσθετικότητα}) \quad & \rho(X + Y) \leq \rho(X) + \rho(Y), \\ (\text{Μονοτονία}) \quad & \rho(X) \geq \rho(Y), \quad \text{αν } X \leq Y, \\ (\text{Translation Invariance}) \quad & \rho(X + ar) = \rho(X) - a, \\ (\text{Positive Homogeneity}) \quad & \text{Για } \forall \lambda \geq 0, \quad \rho(\lambda X) = \lambda \rho(X). \end{aligned}$$

Στους παραπάνω τύπους το r παριστάνει την απόδοση της εκάστοτε επενδυτικής θέσης και το a είναι ένα ποσό που μπορεί να αφαιρεθεί από την αρχική επένδυση.

Με άλλα λόγια, όταν προστίθεται ένα ποσό σε μια επενδυτική θέση, ο κίνδυνος της θέσης μειώνεται τόσο, όσο αυξήθηκε το συνολικό ποσό της επένδυσης. Πρακτικά η ιδιότητα της υποαθροιστικότητας σημαίνει ότι ο συνολικός κίνδυνος δε μπορεί να είναι μεγαλύτερος από το άθροισμα δύο επιμέρους κινδύνων (βλ. [5], [3]).

Για τα αναφερόμενα στο παράρτημα ο ενδιαφερόμενος αναγνώστης μπορεί να συμβουλευτεί την αναφορά (βλ [18]).

Αναφορές

- [1] Ailliot P., Thompson C. & Thomson P. (2011). Mixed methods for fitting the GEV distribution. *Water Resources Research*, **47** (5).
- [2] Akram M. & A. H. (2014). Comparison of estimators of the Weibull distribution. *Journal of Statistical Theory and Practice*, **8** (2), 238–259.
- [3] Artzner P., Delbaen F., Eber J. & Heath D. (1999). Coherent measures of risk. *Mathematical Finance* *9*, (3), 203–228.
- [4] Artzner P., Delbaen F., Eber J.M. & Heath D. (1997). Thinking coherently. *Risk*, **10**, 68–71.
- [5] Artzner P., Delbaen F., Eber J.M. & Heath D. (1999). Coherent measures of risk. *Mathematical Finance*, **9** (3), 203–228.
- [6] Beirlant J., Vynckier P. & Teugels J. (1996). Excess functions and estimation of the extreme-value index. *Bernoulli*, **2** (4), 293–318.
- [7] Cacciari M., Mazzanti G. & Montanari G. (1996). Comparison of maximum likelihood unbiasing methods for the estimation of the Weibull parameters. *IEEE Transactions*, **3**, 18–27.
- [8] Cheng R. & Amin N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society*, **45** (3), 394–403.
- [9] Cousineau D., Goodman V. & Shiffrin M. (2002). Extending statistics of extremes to distributions varying on position and scale, and implication models. *Journal of Mathematical Psychology*, **46**, 431–454.

- [10] Cran G. (1988). Moments estimators for the 3-parameter Weibull distribution. *IEEE Transactions on Reliability*, **37**, 360–363.
- [11] Danielson J. & deVries C. (1997). Value-at-Risk and extreme events. *Discussion Paper, LSE Financial Markets Group, London School of Economics*, **273**.
- [12] Danielsson J., Jorgesen B., Samorodnitsky G., Sarma M. & de Vries C. (2005). Subadditivity re-examined: the case for Value-at-Risk. *London School of Economics*.
- [13] Davison A. & Smith R. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*, **52 (3)**, 393–442.
- [14] Diebolt J., Guillou A., Naveau P. & Ribereau P. (2008). Improving probability-weighted moment methods for the Generalized Extreme Value distribution. *Journal REVSTAT*, **6 (1)**, 33–50.
- [15] Embrechts P., Kluppelberg C. & Mikoch T. (1999). Modelling extremal events : for insurance and finance. *Applications of Mathematics, Stochastic Modelling and Applied Probability, Springer*.
- [16] Embrechts P., Resnick S. & Samorodnitsky G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, **3 (2)**, 30–41.
- [17] Feng J. (1998). Generalization errors of the simple perceptron. *Journal Physics A:Mathematical and General*, **31**, 4037–4048.
- [18] Finney L., Giordano R. & D W. (2001). Thomas' calculus. *Crete University Press*.
- [19] Fisher R.A. & 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**, 190–190.

- [20] Frechet M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Societe Polonaise de Mathematique*, **6**, 93–116.
- [21] Gnedenko B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics, Translated and reprinted in: Breakthroughs in Statistics, Springer-Verlag, eds. Kotz S. and Johnson N.L.(1922), Vol. I, pg:195-225*, **44**, 423–453.
- [22] Gob R. (2011). Estimating value at risk and conditional value at risk for count variables. *Quality and Reliability Engineering International*, **27 (5)**, 659–672.
- [23] Greenwood A., Landwehr M., Matalas N. & Wallis J. (1979). Probability weighted moments:definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, **15 (5)**, 1049–1054.
- [24] Grender J., Reich R. & Dell T. (1990). Theory and derivation for Weibull parameter probability weighted moment estimators. *Research paper SO-260 USA Agriculture Dept. Southern Forest Experiment Station*.
- [25] Guillaume M., Dacarogna M., Dave R., Muller A., Olsen B. & Pictet V. (1997). From the bird's eye to the microscope. a survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics, Springer*, **1**, 95–127.
- [26] Gumbel E. (1935). Les valeurs extrêmes des distribution statistiques. *Annales de l'institut Henri Poincaré*, **5 (2)**, 115–158.
- [27] Gumbel E. (1958). Statistics of extremes. *Columbia University Press, New York*.
- [28] de Haan L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica*, **44 (2)**, 45–68.

- [29] Hickey K.R. (2001). The storm of 31 January to 1 February 1953 and its impact on Scotland. *Scottish Geographical Journal*, **117** (4), 283–295.
- [30] Hirose H. (1996). Maximum likelihood estimation in the 3-parameter Weibull distribution: A look through the Generalized Extreme-Value distribution. *IEEE Transactions*, **3**, 43–45.
- [31] Hogg V. & Klugman A. (1984). Loss distributions. *Wiley, New York*.
- [32] Hosking J.R.M., Wallis J.R. & Wood E.F. (1985). Estimation of the Generalized Extreme Value Distribution by the method of probability-weighted moments. *Technometrics*, **27** (3), 251–261.
- [33] Jaffe D. & Russell T. (1996). Working paper: catastrophe insurance, capital markets and uninsurable risk. *Financial Institutions Center, The Wharton School, Philadelphia*, 96–112.
- [34] Kotz S. & Nadarajah S. (2000). Extreme value distributions: Theory and applications. *Imperial college press*.
- [35] Landwehr M., Matalas N. & Wallis J. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, **15** (5), 1055–1064.
- [36] Landwehr M.J., Matalas N. & Wallis J. (1979). Estimation of parameters and quantiles of Wakeby distributions. *Water Resources Research*, **15** (6), 1361–1379.
- [37] Leadbetter M. (1991). On a basis for "peaks over threshold" modelling. *Statistics and Probability Letters*, **12** (4), 357–362.
- [38] Longin F. (1996). The asymptotic distribution of extreme stock market returns. *Journal of Business*, **6** (3), 383–408.

- [39] Mahdi S. & Cenac M. (2005). Estimating parameters of Gumbel distribution using the methods of moments, probability weighted moments and maximum likelihood. *Revista de Matematica : Theoria Y Aplicaciones*, **12** (1), 151–156.
- [40] Mason D.M. (1982). Laws of large numbers for sums of extreme values. *The annals of Probability*, **10** (3), 754–764.
- [41] Mc Ewan R. & Parresol B. (1991). Moment expressions and summary statistics for the complete and truncated Weibull distribution. *Communication in Statistics-Theory and Methods*, **20** (4), 1361–1372.
- [42] McNeil A., Frey R. & Embrechts P. (2005). Quantitative risk management: Concepts, techniques and tools. *Princeton University Press*.
- [43] von Mises R. (1936). La distribution de la plus grande de n valeurs. *Rev. Mathematique de l'Union Interbalcanique*, **1**, 141–160.
- [44] Mudholkar G. & Kollia G. (1994). Generalized Weibull family:a structural analysis. *Communication in Statistics-Theory and Methods*, **23** (4), 1149–1171.
- [45] Muraleedharan G. & Soares C. (2014). Characteristic and moment generating functions of Generalised Pareto and Weibull distributions. *Journal of Scientific Research and Reports*, **14** (3), 1861–1874.
- [46] Pickands J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, **3** (1), 119–131.
- [47] Prescott P. & Walden A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, **67**, 723–724.
- [48] Rao R. (1973). Linear statistical inference and its applications. *New York:John Wiley 2nd ed..*

- [49] Shevchenko P. (2010). Calculation of aggregate loss distributions. *The Journal of Operational Risk*, **5** (2), 3–40.
- [50] Smith R. (1990). Extreme value analysis of environmental time series: An application to trend detection in ground level ozone. *Journal of Statistical Science*, **4**, 367–377.
- [51] Smith R.L. (1986). Extreme value theory based on the r largest annual events. *Journal of Hydrology*, **86**, 27–43.
- [52] Stanton R. (1987). The work of L.H.C. Tippett. *Ars Textrina*, **7**, 179–185.
- [53] Teimouri M., S. H. & Nadarajah S. (2011). Comparison of estimation methods for the Weibull distribution. *Journal of Theoretical and Applied Statistics*, **47** (1), 93–109.
- [54] Tippett L. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364–387.
- [55] Weibull W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 293–297.
- [56] Zanakis S. & Kyparisis J. (1986). A review of maximum likelihood estimation methods for the 3-parameter Weibull distribution. *Journal of Statistical Computation and Simulation*, **25**, 53–73.