ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΜΕΛΕΤΗ ΤΗΣ ΕΜΦΑΝΙΣΗΣ ΜΕΓΑΛΩΝ ΖΗΜΙΩΝ ΣΕ ΕΝΑ ΧΑΡΤΟΦΥΛΑΚΙΟ ΑΣΦΑΛΙΣΗΣ ΑΥΤΟΚΙΝΗΤΩΝ

Γεώργιος Γ. Μανθόπουλος

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 2013

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΜΕΛΕΤΗ ΤΗΣ ΕΜΦΑΝΙΣΗΣ ΜΕΓΑΛΩΝ ΖΗΜΙΩΝ ΣΕ ΕΝΑ ΧΑΡΤΟΦΥΛΑΚΙΟ ΑΣΦΑΛΙΣΗΣ ΑΥΤΟΚΙΝΗΤΩΝ

Γεώργιος Γ. Μανθόπουλος

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 2013

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

Τα μέλη της Επιτροπής ήταν:

- Αναπληρωτής Καθηγητής Πολίτης Κωνσταντίνος (Επιβλέπων)
- Επίκουρος Καθηγητής Τζαβελάς Γεώργιος
- Λέκτορας Ψαρράκος Γεώργιος

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

UNIVERISTY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN ACTUARIAL SCIENCE AND RISK MANAGEMENT

STUDY OF THE EMERGENCE OF LARGE LOSSES IN A CAR INSURANCE PORTFOLIO

By

Georgios G. Manthopoulos

MSc Dissertation

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

Piraeus, Greece

June 2013

Στην οικογένεια μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω όσους συντέλεσαν στην ολοκλήρωση της παρούσας Διπλωματικής εργασίας. Αρχικά, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα Επίκουρο Καθηγητή κ. Πολίτη Κωνσταντίνο για τη γνώση, την καθοδήγηση και τις πολύτιμες συμβουλές που μου προσέφερε καθ' όλη τη διάρκεια της συγγραφής της Διπλωματικής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής Επίκουρο Καθηγητή κ. Γιώργο Τζαβελλά και Λέκτορα κ. Γιώργο Ψαρράκο για τη συμμετοχή τους στην εξεταστική επιτροπή της παρούσας Διπλωματικής εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω θερμά τα μέλη της οικογένειας μου που με στηρίζουν σε όλα τα χρόνια της ακαδημαϊκής μου πορείας, τους φίλους και τους συμφοιτητές μου για την ηθική συμπαράστασή τους.

Περίληψη

Η συνεχής και σωστή αξιολόγηση του κινδύνου στον ασφαλιστικό κλάδο έτσι ώστε να υπάρχουν επαρκή αποθέματα στην κάθε εταιρεία έχει τις ρίζες της στη δεκαετία του 1970, αναπροσδιορίστηκε στα μέσα της δεκαετίας του 1990 ενώ σήμερα, με την έλευση του Solvency II, έχει γίνει πιο επιτακτική από ποτέ. Ιδιαίτερα σημαντικό ρόλο για τον προσδιορισμό του επαρκούς αποθέματος είναι τόσο ο αριθμός όσο και το μέγεθος των αποζημιώσεων που καταφθάνουν σε συγκεκριμένο χρονικό διάστημα σε μια ασφαλιστική εταιρεία.

Στην παρούσα διπλωματική εργασία αρχικά παρουσιάζονται ορισμοί και τεχνικές της θεωρίας κινδύνου για την μελέτη της κατανομής του αριθμού και του μεγέθους των αποζημιώσεων. Στη συνέχεια εξετάζεται η Θεωρία των Ακραίων Τιμών η οποία αναφέρεται σε παρατηρήσεις που ξεπερνούν κάποιο προκαθορισμένο μέγιστο όριο. Τέλος όλα τα θεωρητικά αποτελέσματα εφαρμόζονται σε πραγματικά δεδομένα, με χρήση του στατιστικού πακέτου R, ενώ τα συμπεράσματα της ανάλυσης ερμηνεύονται πρακτικά έτσι ώστε να μπορούν να χρησιμοποιηθούν άμεσα από την οποιαδήποτε ασφαλιστική εταιρεία για τη μελλοντική εκτίμηση του κινδύνου της.

Abstract

The continuous and proper risk assessment in the insurance industry, which has been implemented so that there is adequate reserve amount to each company, was originated in the 1970s, was redefined in the mid 1990s and nowadays, with the advent of Solvency II, has become more urgent than ever. Both the number and the amount of the compensations, which arrive during a pre-specified period at an insurance company, are the most important features so as to best estimate the sufficient reserve amount needed to be held by the company.

In this thesis, definitions and techniques of the risk theory have been initially illustrated in order to study the number and the size of claims. Furthermore, extreme value theory, which is typically applied to observations that exceed a predefined maximum, is meticulously presented. Finally, all the theoretical results have been applied to real world data, with the aid of the statistical software R. The conclusions of the analysis have been practically interpreted so that they could be used by any insurance company for future risk assessment.

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κατάλογος Συντομογραφιών

GEV	Generalized Extreme Value Distribution – Γενικευμένη κατανομή Ακραίων τιμών
GPD	Generalized Pareto Distribution – Γενικευμένη κατανομή Pareto
MDA	Maximum Domain of Attraction – Περιοχή έλξης Ακροτάτων
POT	Peak Over Threshold – *Υπερβάσεις πάνω από ένα όριο*
δ.ε.	Διάστημα εμπιστοσύνης
ε.μ.π.	Εκτιμητής μεγίστης πιθανοφάνειας
σ.κ.	Συνάρτηση κατανομής
σ.π.π	Συνάρτηση πυκνότητας πιθανότητας
τ.μ.	Τυχαία μεταβλητή

Κεφάλαιο 1

Εισαγωγή

1.1 Σκοπός της εργασίας

Τα άτομα, οι κοινωνίες και κατ' επέκταση οι επιχειρήσεις για να αναπτυχθούν έχουν ανάγκη να νιώθουν ασφαλείς απέναντι σε επικείμενες δυσμενείς αλλαγές που μπορεί να πραγματοποιηθούν. Για αυτό το λόγο προσπαθούν να προβλέψουν, όσο είναι δυνατόν, αλλά και να ελαχιστοποιήσουν τις απώλειες σε οποιοδήποτε αρνητική μεταβολή μπορεί να προκύψει. Ένας τρόπος για να αντισταθμίσουν τους κινδύνους που διατρέχουν ιδιαίτερα τα άτομα αλλά και οι επιχειρήσεις είναι η αγορά ασφαλιστικής κάλυψης ή αντασφαλιστικής αντίστοιχα αν μιλάμε για ασφαλιστικές εταιρίες. Η κύρια ενασχόληση των ασφαλιστικών και αντασφαλιστικών επιχειρήσεων και ιδιαίτερα των αναλογιστικών τμημάτων τους, είναι η μελέτη των αποζημιώσεων που μπορεί να προκύψουν έτσι ώστε να είναι συνεπείς προς τις υποχρεώσεις που έχουν αναλάβει έναντι των ασφαλισμένων τους. Η διαδικασία αξιολόγησης των κινδύνων από τη μεριά τους δηλαδή αποτελεί ζωτικής σημασίας λειτουργία η οποία προϋποθέτει την σωστή τιμολόγηση των κινδύνων που αναλαμβάνουν καθώς και αποτελεσματική διαχείριση των αποθεμάτων έτσι ώστε να είναι φερέγγυες απέναντι στους ασφαλισμένους και συνεπείς στις εκτιμήσεις που ζητούν οι εποπτικές αρχές. Η ασφάλεια που νιώθουν οι ασφαλισμένοι στηρίζεται κατά κύριο λόγο στην επάρκεια των αποθεμάτων που έχει στην κατοχή της η ασφαλιστική εταιρία. Η πρώτη προσπάθεια για τη σωστή αξιολόγηση των κινδύνων και δημιουργία επαρκούς αποθεμάτων έχει τις βάσεις της στο έτος 1973 όταν γεννήθηκε η ιδέα της υιοθέτησης μιας κοινοτικής οδηγίας για τον έλεγχο των ασφαλιστικών εταιριών υπό την ονομασία Solvency I (Περιθώριο φερεγγυότητας I). Αυτή η οδηγία σκοπό είχε να καθορίσει το ύψος και την ποιότητα των αποθεμάτων που πρέπει να έχει στην κατοχή της η ασφαλιστική επιχείρηση έτσι ώστε να βρίσκεται σε θέση να αποπληρώσει τις υποχρεώσεις της. Η Τρίτη γενιά του Solvency I που εφαρμόζεται από τα μέσα της δεκαετίας του '90 είναι αυτή που χρησιμοποιείται μέχρι και σήμερα. Εδώ και λίγα χρόνια προετοιμάζεται μια αυστηρότερη οδηγία με την ονομασία Solvency II με σκοπό να συμπληρώσει τις ελλείψεις του Solvency I καθώς και να επικαιροποιηθεί στα σημερινά οικονομικά δεδομένα. Από τα παραπάνω καταλαβαίνουμε πως υπάρχει διάθεση από όλους τους εμπλεκόμενους φορείς για μεγαλύτερη ασφάλεια και αποτελεσματικότερη διαχείριση

των κινδύνων που αναλαμβάνουν οι ασφαλιστικές εταιρίες. Η μελέτη που θα ακολουθήσει σκοπό έχει να δώσει στις ασφαλιστικές εταιρίες μια επιπλέον πληροφορία ειδικότερα όσον αφορά το μέγεθος της μέγιστης αποζημίωσης έτσι ώστε να μπορεί να το χρησιμοποιήσει στις εκτιμήσεις της για το ύψος των αποθεμάτων που θα πρέπει να έχει στην κατοχή της.

1.2 Διάρθρωση της εργασίας

Η παρούσα διπλωματική εργασία καλείται να μελετήσει τις ζημιές της ασφαλιστικής επιχείρησης που θα προέλθουν από τις διεκδικήσεις των ασφαλισμένων, λόγω της έλευσης του ασφαλισμένου κινδύνου, με σκοπό να μπορεί να γίνει εκτίμηση για τα αποθέματα τα οποία θα πρέπει να κρατήσει. Η μελέτη που θα πραγματοποιηθεί αφορά το μέγεθος και των αριθμό των αποζημιώσεων έτσι ώστε να παρατηρηθεί αν οι ποσότητες αυτές ακολουθούν κάποια κατανομή. Επίσης θα γίνει προσπάθεια να μοντελοποιηθούν οι πολύ μεγάλες αποζημιώσεις με τέτοιο τρόπο ώστε η ασφαλιστική επιχείρηση να μπορεί να κάνει προβλέψεις για τις μέγιστες αποζημιώσεις που θα προκύψουν. Συγκεκριμένα το δεύτερο κεφάλαιο της συγκεκριμένης εργασίας επικεντρώνεται θεωρητικά στη στατιστική μελέτη του αριθμού και του μεγέθους των αποζημιώσεων αναπτύσσοντας τις κυριότερες κατανομές πάνω στα μοντέλα που προτείνει η θεωρία κινδύνου. Το τρίτο κεφάλαιο βασίζεται σε μια ταχέως αναπτυσσόμενη τις τελευταίες δεκαετίες θεωρία που αφορά τη συμπεριφορά των ακραίων παρατηρήσεων. Η θεωρία ακραίων τιμών, όπως ονομάζεται, μελετά την κατανομή των παρατηρήσεων που ξεπερνούν κάποιο όριο. Στα κεφάλαια 4 και 5 θα γίνει η προσπάθεια να εφαρμοσθούν σε ένα πραγματικό χαρτοφυλάκιο ασφαλιστηρίων συμβολαίων αυτοκινήτου τα όσα θεωρητικά αναπτύχθηκαν στα κεφάλαια 2 και 3. Ιδιαίτερα στο κεφάλαιο 5 θα αναπτυχθεί η θεωρία ακραίων τιμών και με τη βοήθεια του στατιστικού πακέτου R θα γίνουν οι υπολογισμοί που θα προβλέπουν το ποσό που δε θα ξεπερνάει η μέγιστη αποζημίωση για μια σειρά χρονικών περιόδων. Στο κεφάλαιο 6 θα παρουσιαστούν συγκεντρωτικά τα αποτελέσματα της μελέτης με σκοπό να εξαχθούν πολύτιμα συμπεράσματα που θα συνεισφέρουν στην πληροφόρηση της ασφαλιστικής επιχείρησης όσον αφορά το ύψος των αποθεμάτων που θα πρέπει να κρατάει έτσι ώστε να είναι σε θέση να αποπληρώσει της υποχρεώσεις της ακόμα και στην περίπτωση ακραίων αποζημιώσεων.

Κεφάλαιο 2

Μελέτη και τρόποι εκτίμησης των κατανομών για το μέγεθος και τον αριθμό των αποζημιώσεων

Η αναλογιστική επιστήμη είναι άρρηκτα συνδεδεμένη με τις κατανομές απώλειας. Αυτό συμβαίνει διότι το βασικό χαρακτηριστικό της ασφάλισης έγκειται στο γεγονός της αποζημίωσης του ασφαλισμένου από την ασφαλιστική εταιρία σε ένα τυχαίο χρόνο και κατά ένα τυχαίο ποσό, ποσότητες οι οποίες μπορούν να αποδοθούν με τη μορφή πιθανοθεωρητικών κατανομών. Ειδικότερα στην περίπτωση που το μέγεθος της αποζημίωσης εκφράζεται με μορφή πιθανότητας τότε η κατανομή αυτή καλείται κατανομή απώλειας. Από την παραπάνω έκφραση προκύπτει ότι οι πληρωμές που αποδίδει η ασφαλιστική εταιρία στους ασφαλισμένους διέπονται από τρία βασικά χαρακτηριστικά:

- Τον αριθμό των αποζημιώσεων
- Το ποσό της αποζημίωσης
- Το χρόνο πληρωμής της αποζημίωσης

2.1 Κατανομή για το μέγεθος των αποζημιώσεων

Η εύρεση της κατανομής που ακολουθούν οι αποζημιώσεις προϋποθέτει αρχικά την επιλογή μιας μεθόδου προσέγγισης των δεδομένων. Η μέθοδος η οποία θα επιλεγεί θα εξαρτάται από την ποιότητα και το μέγεθος του δείγματος, και θα οδηγήσει στην εξαγωγή ασφαλών συμπερασμάτων για την κατανομή του πληθυσμού από τον οποίο προήλθε το δείγμα. Οι πιο διαδεδομένοι μέθοδοι είναι:

1)Η εμπειρική μέθοδος
2)Η παραμετρική μέθοδος και
3) Η μέθοδος Bayes

2.1.1 Εμπειρική μέθοδος εκτίμησης

Η εμπειρική μέθοδος εκτίμησης βασίζεται στην εμπειρική συνάρτηση κατανομής δηλαδή στην συνάρτηση κατανομής των δεδομένων. Το σημαντικότερο πλεονέκτημα αυτής της μεθόδου είναι το γεγονός ότι είναι ιδιαίτερα απλή στην εφαρμογή της καθώς και ότι αποτελεί ίσως την πιο αξιόπιστη μέθοδο στην περίπτωση που είναι διαθέσιμος μεγάλος αριθμός παρατηρήσεων.

Σκοπός της οποιασδήποτε διαδικασίας εκτίμησης είναι να χρησιμοποιήσουμε τα στοιχεία του δείγματος έτσι ώστε να πάρουμε πληροφορίες για τον πληθυσμό από τον οποίο επιλέχθηκε. (*Klugman et al,1998*).

Έστω $X_1,\ldots,X_n$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση κατανομής $F_X(x)$ και από κοινού συνάρτηση κατανομής $F_X(x_1,\ldots x_n) = F_X(x_1)\ldots F_X(x_n)$.

Ορισμός 2.1.1 : Η εμπειρική κατανομή εξάγεται από ένα δείγμα ορίζοντας πιθανότητα $\frac{1}{n}$ σε κάθε παρατήρηση. Συγκεκριμένα η συνάρτηση κατανομής ορίζεται ως:

$$F_n(x) = \frac{\alpha\rho\iota\theta\mu\acute{o}\varsigma\ \tau\omega\nu\ x_j \le x}{n} = F_{\hat{X}}(x),$$

ενώ η συνάρτηση πιθανότητας ορίζεται ως:

$$f_n(x) = \frac{\alpha\rho\iota\theta\mu\acute{o}\varsigma\ \tau\omega\nu\ x_j = x}{n}.$$

i) Εμπειρική εκτίμηση του μέσου

Ο μέσος της εμπειρικής κατανομής υπολογίζεται εύκολα από την ακόλουθη σχέση:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

ii) Εμπειρική εκτίμηση των υπολοίπων ροπών

Η n-οστη ροπή μιας τ.μ. Χ περί την αρχή ορίζεται ως:

$$\mu_n' = \mathrm{E}(X^n).$$

Ενώ η κεντρική ροπή ως:

$$\mu_n = \mathrm{E}[(\mathrm{X} - \mu)^n].$$

Διακύμανση:

$$\sigma^2 = \mu_2 = \mu_2' - \mu^2.$$

Τυπική απόκλιση:

$$\sigma = \sqrt{\sigma^2}.$$

Συντελεστής μεταβλητότητας:

$$\frac{\sigma}{\mu}.$$

Οι εμπειρικοί εκτιμητές είναι:

$$\hat{\mu}_n' = \mathrm{E}(\hat{\mathrm{X}}^n) = \frac{1}{n}\sum_{j=1}^{n} x_j{}^n \quad \text{και}$$

$$\hat{\mu}_n = \mathrm{E}[(\hat{\mathrm{X}} - \hat{\mu})^n] = \frac{1}{n}\sum_{j=1}^{n} (x_j - \bar{x})^n$$

iii) Ποσοστημόρια

Τα ποσοστημόρια παριστάνουν τιμή των διατεταγμένων δεδομένων, όπου τουλάχιστον το 100*p% αυτών των δεδομένων είναι κάτω από αυτήν την τιμή και το 100*(1-p) % είναι πάνω από αυτή την τιμή. Αν συμβολίσουμε με π_p το p-ποσοστημόριο μιας συνάρτησης κατανομής F, τότε

$$F(\pi_p -) \le p \le F(\pi_p)$$

Όπου $F(\pi_p -) = \lim_{h \to 0} F(\pi_p - h)$.

iv) Εκτίμηση μέσω διαστήματος εμπιστοσύνης

Οι παραπάνω εκτιμητές αναφέρονται σε σημειακή εκτίμηση. Δηλαδή την καλύτερη δυνατή προσπάθεια εκτίμησης μιας συγκεκριμένης τιμής σε ένα τυχαίο δείγμα. Όσο καλή και να είναι αυτή η εκτίμηση δεν αναμένεται να συμπίπτει απόλυτα με την ακριβή τιμή του πληθυσμού. Για αυτό το λόγο υπάρχει η δυνατότητα εκτίμησης της συγκεκριμένης τιμής μέσω ενός διαστήματος εμπιστοσύνης.

v) Διάστημα εμπιστοσύνης : Το διάστημα εμπιστοσύνης δίνει ένα εύρος εκτιμώμενων τιμών το οποίο είναι πιθανό να περικλείει την άγνωστη παράμετρο του πληθυσμού η οποία καλείται να εκτιμηθεί. Στην πραγματικότητα ένα διάστημα εμπιστοσύνης με επίπεδο εμπιστοσύνης 95% σημαίνει ότι για έναν μεγάλο αριθμό δειγμάτων από τον ίδιο πληθυσμό τα διαστήματα εμπιστοσύνης θα περικλείουν την πραγματική ποσότητα του πληθυσμού στο 95% των περιπτώσεων. Το επίπεδο σημαντικότητας είναι η πιθανότητα α η οποία συνδέεται με το διάστημα εμπιστοσύνης και συνήθως εκφράζεται με τη μορφή ποσοστού. Για παράδειγμα α=5% σημαίνει ότι το επίπεδο σημαντικότητας είναι 5% και το επίπεδο εμπιστοσύνης είναι το $(1-a)\% = 95\%$.

vi) Διάστημα εμπιστοσύνης κατά προσέγγιση (*approximate confidence interval*)

Επειδή πολλές φορές είναι αρκετά δύσκολο να κατασκευαστεί ένα διάστημα εμπιστοσύνης, προτιμάται όταν είναι εφικτό να κατασκευαστεί ένα διάστημα εμπιστοσύνης κατά προσέγγιση. Ας υποθέσουμε ότι έχουμε έναν εκτιμητή $\hat{\theta}$ τέτοιον ώστε:

$$\mathrm{E}(\hat{\theta}) = \theta\,,\ Var(\hat{\theta}) = u(\theta).$$

Τότε, θεωρώντας ότι η κατανομή του $\hat{\theta}$ προσεγγίζεται ικανοποιητικά από την κανονική κατανομή, το (1-α)% διάστημα εμπιστοσύνης κατά προσέγγιση καταλήγει να είναι:

$$1\text{-}\alpha = \Pr[\hat{\theta} - z_{\alpha/2}\sqrt{u(\hat{\theta})} \le \theta \le \hat{\theta} + z_{\alpha/2}\sqrt{u(\hat{\theta})}].$$

(*Klungman et al,1998*).

vii) Αξιολόγηση του εκτιμητή

Αφού υπολογιστεί η εκτίμηση της ποσότητας που μας ενδιαφέρει είναι βασικό να αξιολογηθεί η ποιότητα του εκτιμητή έτσι ώστε να αποφασιστεί το αν ο εκτιμητής που έχει επιλεγεί έχει όλα τα στοιχεία τα οποία θα τον χαρακτηρίσουν ως αξιόπιστο. Σκοπός αυτής της διαδικασίας είναι η επιλογή ενός εκτιμητή ο οποίος να προσεγγίζει ικανοποιητικά την ποσότητα που επιθυμούμε να εκτιμήσουμε. Αυτή η επιλογή θα γίνει μέσω κριτηρίων και ιδιοτήτων που πρέπει να πληρούν αυτοί οι εκτιμητές.

Ορισμός 2.1.2: Έστω $\delta = \delta(\underset{\sim}{\mathrm{X}})$ ένας εκτιμητής της συνάρτησης $g(\theta)$. Μέσο τετραγωνικό σφάλμα του δ καλείται η ποσότητα:

$$\mathrm{MT\Sigma}(\delta\,,\theta) = E_\theta[\{\ \delta(\underset{\sim}{\mathrm{X}}) - g(\theta)\ \}\ ]^2$$

Αν θεωρήσουμε το ΜΤΣ ως την απόσταση ενός εκτιμητή από την τιμή που εκτιμά, είναι φυσικό να προτιμούμε εκτιμητές που απέχουν όσο γίνεται λιγότερο από την προς εκτίμηση τιμή. Άρα μεταξύ δύο εκτιμητών του $g(\theta)$ θα προτιμήσουμε αυτόν με το μικρότερο ΜΤΣ. (*Ηλιόπουλος, 2006*)

Ορισμός 2.1.3: Η μεροληψία ενός εκτιμητή $\hat{\theta}$ ισούται με:

$$b_\theta(\hat{\theta}) = E(\hat{\theta}) - \theta\,.$$

Στη περίπτωση πού η μεροληψία ενός εκτιμητή είναι ίση με 0, τότε αυτός ο εκτιμητής ονομάζεται αμερόληπτος.

Ορισμός 2.1.4: Ένας αμερόληπτος εκτιμητής του g(θ) που έχει τη μικρότερη διασπορά από όλους τους αμερόληπτους εκτιμητές του g(θ) για κάθε δυνατή τιμή του θ, λέγεται Αμερόληπτα Ομοιόμορφος Ελαχίστης Διασποράς (ΑΟΕΔ) εκτιμητής του g(θ). *(Ηλιόπουλος, 2006).*

2.1.2 Παραμετρική εκτίμηση

Αν και η εμπειρική εκτίμηση λειτουργεί αρκετά καλά σε κάποιες περιπτώσεις η ύπαρξη κάποιων σοβαρών μειονεκτημάτων οδήγησε στην εύρεση άλλων μεθόδων προσέγγισης όπως η παραμετρική μέθοδος εκτίμησης.

Ορισμός 2.1.5: Παραμετρική οικογένεια κατανομών είναι μια συλλογή συναρτήσεων κατανομών της οποίας κάθε μέλος ξεχωρίζεται μέσω ενός συγκεκριμένου αριθμού μεταβλητών που ονομάζονται παράμετροι. Και συγκεκριμένα η $F(x;\theta):\theta\in\Theta$ (*Klungman et al,1998*).

i) Σημειακές εκτιμήσεις παραμέτρων

Η μέθοδος των ροπών

Ορισμός 2.1.6: Έστω X~ $f(x)$. Για k=1,2,…, εάν $E|X|^k<\infty$,τότε η ποσότητα $\mu_k = E(X^k)$ ονομάζεται k-ροπή της X (ή της f) (*Ηλιόπουλος, 2006).*

Ορισμός 2.1.7: Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$ένα τυχαίο δείγμα. Για $k=1,2,\ldots,n$ η στατιστική συνάρτηση

$$m_k = m_k(\underset{\sim}{X}) := \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

ονομάζεται *κ* - δειγματική ροπή *(Ηλιόπουλος ,2006).*

Η μέθοδος αυτή στηρίζεται στο νόμο των μεγάλων αριθμών και αναφέρει πως για να εκτιμηθεί η άγνωστη ποσότητα θ θα πρέπει να εξισωθούν οι πληθυσμιακές με τις δειγματικές ροπές και συγκεκριμένα όσες διαστάσεις έχει η ποσότητα θ τόσες ροπές θα πρέπει να εξισωθούν. Επομένως αν το μέγεθος του δείγματος είναι αρκετά μεγάλο τότε κατά προσέγγιση ισχύει ότι:

$$m_k \approx \mu_k(\theta).$$

ii) Βελτιστοποιημένοι εκτιμητές

Το μεγαλύτερο πρόβλημα των εκτιμητών με τη μέθοδο των ροπών είναι ότι παρά το γεγονός ότι προσεγγίζουν «καλά» κάποια χαρακτηριστικά δεν έχουν καλή προσαρμογή σε ένα μεγάλο εύρος παρατηρούμενων τιμών. Πρέπει να λαμβάνεται υπόψη πως η εμπειρική κατανομή δεν ταυτίζεται απόλυτα με την παραμετρική κατανομή και για αυτό πρέπει να βρεθούν τρόποι που να δείχνουν το πόσο κοντά βρίσκονται οι δυο τους.

Υπάρχουν διάφοροι μέθοδοι για να το ανακαλύψουμε αυτό όπως οι εκτιμητές ελαχίστων αποστάσεων, οι εκτιμητές διαστημάτων ελαχίστης απόστασης, οι εκτιμητές μεγίστης πιθανοφάνειας κ.α.

Εκτιμητές μεγίστης πιθανοφάνειας (maximum likelihood estimators)

Πρόκειται ίσως για τους εκτιμητές με τις καλύτερες στατιστικές ιδιότητες και την καλύτερη εφαρμογή στα περισσότερα είδη δεδομένων. Συχνά στη βιβλιογραφία η έκφραση εκτιμητής μιας παραμέτρου συνήθως αναφέρεται στον εκτιμητή μεγίστης πιθανοφάνειας. Η φιλοσοφία πίσω από αυτή τη μέθοδο είναι αρκετά απλή. Γίνεται η προσπάθεια εκτίμησης της τιμής της άγνωστης παραμέτρου η οποία μεγιστοποιεί την πιθανότητα να παρατηρηθούν οι τιμές του τυχαίου δείγματος.

Ορισμός 2.1.8: Η από κοινού πυκνότητα πιθανότητας $f(\underset{\sim}{x};\theta)$ στο παρατηρηθέν σημείο $\underset{\sim}{x} = (x_1, x_2, \ldots x_n)$ θεωρούμενη ως συνάρτηση του θ λέγεται συνάρτηση πιθανοφάνειας και συμβολίζεται ως:

$$L(\theta \setminus \underset{\sim}{x}) := L(\theta) := \prod_{j=1}^{n} L_j(\theta) := \prod_{j=1}^{n} f(x_j;\theta) := f(\underset{\sim}{x};\theta), \quad \theta \in \Theta .$$

Εφόσον έχει υπολογιστεί η συνάρτηση πιθανοφάνειας τότε ψάχνουμε να βρούμε την τιμή της παραμέτρου η οποία μεγιστοποιεί αυτή τη συνάρτηση.

Ορισμός 2.1.9: Έστω $\underset{\sim}{x}$ η παρατηρηθείσα τιμή του $\underset{\sim}{X}$ και $\hat{\theta}(\underset{\sim}{x})$ ένα σημείο του Θ τέτοιο ώστε:

$$L(\hat{\theta}(\underset{\sim}{x}) \setminus \underset{\sim}{x}) = \max_{\theta \in \Theta} L(\theta \setminus \underset{\sim}{x}).$$

Η τιμή $\hat{\theta}(\underset{\sim}{x})$ ονομάζεται εκτίμηση μεγίστης πιθανοφάνειας του θ και η στατιστική συνάρτηση $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$ λέγεται εκτιμητής μεγίστης πιθανοφάνειας του θ (*Ηλιόπουλος 2006).*

iv) Συμπεριφορά της δεξιάς ουράς

Η παρούσα εργασία επικεντρώνεται στην ύπαρξη και τη συμπεριφορά της δεξιάς ουράς μιας κατανομής. Η ύπαρξη διαδικασιών οι οποίες μπορούν να επιφέρουν μεγάλες ζημιές με μια συχνότητα οδήγησε τα άτομα και τις επιχειρήσεις στην αγορά ασφαλιστικών καλύψεων. Με αυτό τον τρόπο ο κίνδυνος των μεγάλων ζημιών μεταβιβάζεται στην ασφαλιστική εταιρία. Για τις κατανομές οι οποίες επεκτείνονται μέχρι το άπειρο το ζήτημα έγκειται στο πόσο γρήγορα η συνάρτηση πυκνότητας πιθανότητας πλησιάζει το μηδέν. Όσο πιο αργά συμβαίνει αυτό τόσο πιο "βαριά" δεξιά ουρά λέμε ότι έχει η κατανομή μας.

Υπάρχουν διάφορες συναρτήσεις που συνδέονται με τη συνάρτηση κατανομής και οι οποίες δείχνουν την πιθανότητα εμφάνισης μεγάλων ζημιών. Ένα παράδειγμα είναι η συνάρτηση επιβίωσης η οποία ορίζεται ως:

$$S(x) = 1 - F(x) = \begin{cases} \int_x^{\infty} f(t)dt, & \text{για συνεχείς τ.μ} \\ \sum_{t=x+1}^{\infty} f(t), & \text{για διακριτές τ.μ} \end{cases}$$

α) Ύπαρξη των ροπών

Στην περίπτωση συνεχούς τυχαίας μεταβλητής η k-οστη ροπή υπολογίζεται ως: $\int_0^\infty x^k f(x)dx$. Ανάλογα με την συνάρτηση πυκνότητας και με την τιμή του αριθμού k το ολοκλήρωμα ενδέχεται να μην μπορεί να υπολογιστεί. Για τ.μ που παίρνει μόνο θετικές τιμές αν η αντίστοιχη συνάρτηση πιθανότητας παίρνει πολύ μεγάλες τιμές όσο το x μεγαλώνει τότε πολλαπλασιάζοντας την με το x^k τότε η τιμές θα είναι πολύ μεγάλες και το ολοκλήρωμα δεν θα συγκλίνει. Όσο πιο πολλές ροπές μπορούν να υπολογιστούν τόσο πιο ελαφριά δεξιά ουρά έχει η κατανομή.

β) Οριακοί λόγοι *(Limiting Ratios)*

Ένας άλλος τρόπος που αποτελεί ένδειξη για το αν μια κατανομή έχει πιο ΄΄βαριά΄΄ δεξιά ουρά από μια άλλη είναι το αποτέλεσμα του ορίου καθώς $x \to \infty$ του λόγου των δύο κατανομών. Συγκεκριμένα αν: $\lim_{x\to\infty} \frac{S_1(x)}{S_2(x)}$ τείνει στο άπειρο τότε η κατανομή 1 έχει πιο βαριά δεξιά ουρά ενώ στην περίπτωση που το όριο τείνει στο μηδέν τότε η κατανομή 2 έχει πιο βαριά δεξιά ουρά.

Παράδειγμα 2.1.1 Σύγκριση των ουρών των κατανομών Pareto και Weibull με το κριτήριο των οριακών λόγων:

Έστω η τ.μ X ακολουθεί την κατανομή Pareto(a,b) τότε

$$f(x) = \frac{ab^a}{(x+b)^{a+1}}, \quad a > 0$$

$$S_X(x) = \left(\frac{b}{x+b}\right)^a$$

Για a=6 και b=17 η μέση της τιμή ισούται με:

$$E(X) = \frac{b}{a-1} = \frac{17}{5} = 3,4$$

Έστω ακόμη πως η τ.μ Y ακολουθεί την Weibull (k,λ) τότε

$$f(y) = \frac{k}{\lambda}(\frac{y}{\lambda})^{k-1} \exp(-(\frac{y}{\lambda})^k), \qquad x > 0.$$

$$S_Y(y) = \exp\left(-\left(y/\lambda\right)^k\right)$$

Για k=0.8 και λ=3 η μέση της τιμή ισούται με:

$$E(Y) = \lambda\Gamma\left(1+\frac{1}{k}\right) = 3,4$$

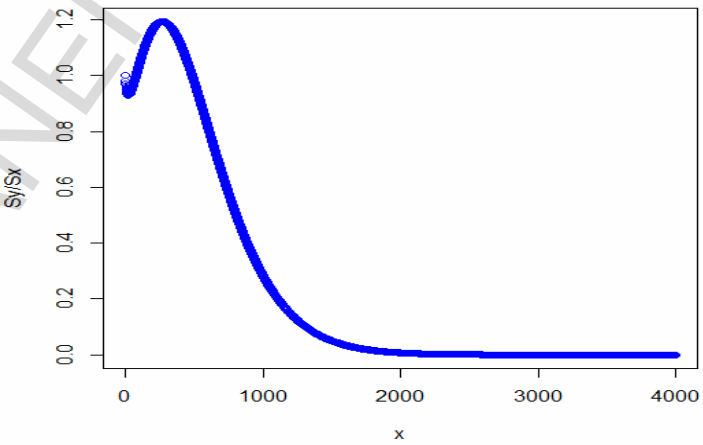
Δηλαδή και οι δύο κατανομές έχουν την ίδια μέση τιμή.

Κριτήριο οριακού λόγου

$$\lim_{x\to\infty}\frac{S_Y(x)}{S_X(x)} = \lim_{x\to\infty}\frac{\exp\left(-\left(x/\lambda\right)^k\right)}{\left(\frac{b}{x+b}\right)^a}$$

Είναι γνωστό ότι οι εκθετικές συναρτήσεις συγκλίνουν γρηγορότερα από τις πολυωνυμικές έτσι στην περίπτωση του παραπάνω ορίου ο αριθμητής συγκλίνει στο μηδέν πιο γρήγορα από τον παρονομαστή που αποκλίνει. Σύμφωνα λοιπόν με το κριτήριο των οριακών λόγων εφόσον το όριο της συναρτήσης επιβίωσης της Weibull προς αυτήν της Pareto τείνει στο μηδέν τότε η κατανομή Pareto έχει πιο βαριά δεξιά ουρά.

Σχήμα 2.1.1: Ο λόγος των συναρτήσεων επιβίωσης των κατανομών Weibull(0.8,3) και Pareto(6,17)



γ) Ένταση θνησιμότητας

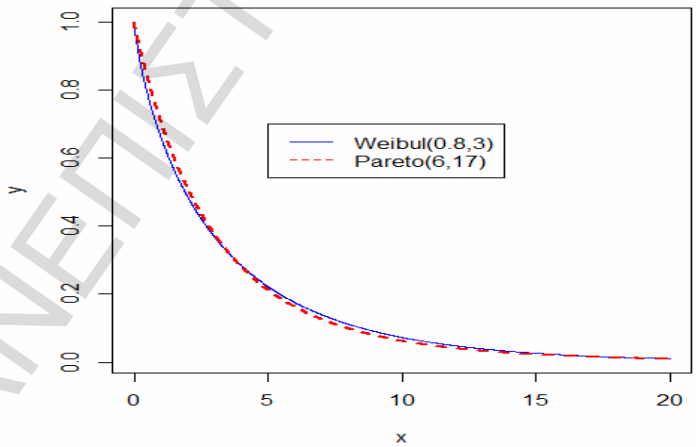
Μια ακόμη ένδειξη για την ύπαρξη βαριάς δεξιάς ουράς περιέχεται στην πληροφορία που προέρχεται από την ένταση θνησιμότητας. Στην περίπτωση που η ένταση θνησιμότητας είναι φθίνουσα συνάρτηση τότε η κατανομή που εξετάζεται θα έχει βαριά δεξιά .

Ορισμός 2.1.10: Ο λόγος $\lambda(x) = \frac{f(x)}{S(x)}, \quad x \geq 0$ ονομάζεται ένταση θνησιμότητας (*force of mortality*), βαθμίδα αποτυχίας (*failure rate*) ή ρυθμός κινδύνου (*hazard rate*). (*Klungman et al,1998*)

Παράδειγμα 2.1.2 Ενδείξεις για τη ύπαρξη βαριών δεξιών ουρών των κατανομών Pareto και Weibull με το κριτήριο των οριακών λόγων:

Από τα δεδομένα του παραδείγματος 2.1.1 για τις κατανομές Weibull(0.8,3) και Pareto(6,17) θα υπολογίσουμε τις εντάσεις θνησιμότητας τους.

Σχήμα 2.1.2: Συναρτήσεις επιβίωσης των κατανομών Pareto(6,17) και Weibull(0.8,3)



Οι συγκεκριμένες κατανομές θεωρείται ότι έχουν βαριές δεξιές ουρές όπως παρατηρείται και στο παραπάνω γράφημα. Μένει να εξακριβωθεί το παραπάνω συμπέρασμα και με το κριτήριο της έντασης θνησιμότητας.

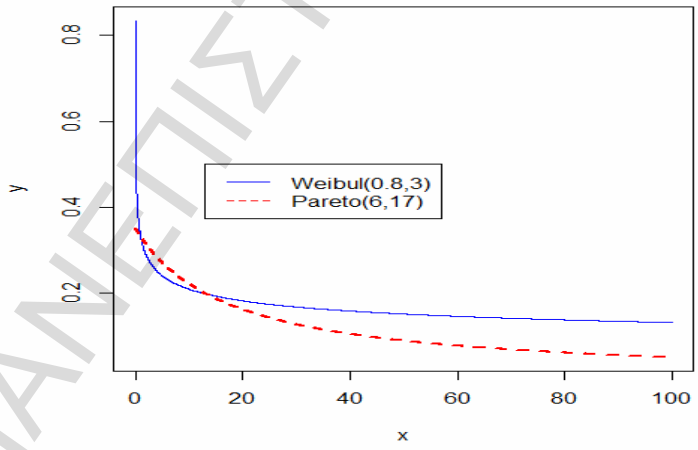
Ένταση θνησιμότητας Pareto :

$$\lambda(x)=\frac{f(x)}{S(x)}=\frac{\frac{ab^a}{(x+b)^{a+1}}}{\frac{b^a}{(x+b)^a}}=\frac{a}{x+b}=\frac{6}{x+17}$$

Ένταση θνησιμότητας Weibull:

$$\lambda(x)=\frac{f(x)}{S(x)}=\left(\frac{k/\lambda\left(y/\lambda\right)^{k-1}\exp\left(-\left(-y/\lambda\right)^k\right)}{\exp\left(-\left(-y/\lambda\right)^k\right)}\right)$$
$$=k/\lambda\left(y/\lambda\right)^{k-1}=0{,}8/3\left(y/3\right)^{0{,}8-1}$$

Σχήμα 2.1.3: Ένταση θνησιμότητας των κατανομών Pareto(6,17) και Weibull(0.8,3)



Από το παραπάνω γράφημα διακρίνεται η φθίνουσα πορεία της έντασης θνησιμότητας και των δύο προαναφερθέντων κατανομών γεγονός που ενισχύει τους αρχικούς ισχυρισμούς μας ως προς την ύπαρξη βαριών δεξιών ουρών. Αξίζει να σημειωθεί πως το κριτήριο της έντασης

θνησιμότητας ενέχει ποιοτική ερμηνεία και όχι ποσοτική. Αυτό σημαίνει πως ενδεχόμενη φθίνουσα πορεία της συγκεκριμένης ποσότητας αποτελεί ένδειξη για την ύπαρξη βαριάς δεξιάς ουράς αλλά δεν ενδείκνυται για συγκρίσεις μεταξύ διαφορετικών κατανομών.

2.1.3 Εκτίμηση κατά Bayes

Η προσέγγιση κατά Bayes, αντίθετα με την παραμετρική προσέγγιση, θεωρεί ότι όλες οι άγνωστες ποσότητες είναι τυχαίες μεταβλητές και χρησιμοποιούνται κατανομές πιθανότητας για να περιγράψουν την κατάσταση της γνώσης μας για αυτές τις ποσότητες. Στην πραγματικότητα η γενική έκφραση υπολογισμού είναι μια εφαρμογή του πολλαπλασιαστικού τύπου και του θεωρήματος ολικής πιθανότητας.

Ποιοτικά η Μπευζιανή προσέγγιση ξεκινά με μια κατανομή πιθανότητας η οποία περιγράφει το επίπεδο της γνώσης μας αναφορικά με τις άγνωστες ποσότητες πριν συλλεγούν δεδομένα και στη συνέχεια χρησιμοποιεί τα παρατηρηθέντα δεδομένα για να επανακαθορίσει την κατανομή αυτή. *(Ι. Πανάρετος & Ε. Ξεκαλάκη (2000))*

Ορισμός 2.1.12: **Θεώρημα Bayes**

Έστω $\{B_1, B_2, ..., B_n\}$ μια διαμέριση ενός δειγματικού χώρου Ω, τέτοια ώστε $P(B_i) > 0$ για όλα τα $i = 1,2,...,n$. Τότε, για κάθε ενδεχόμενο Α του ίδιου δειγματικού χώρου με $P(A) > 0$, ισχύει:

$$P(B_i \setminus A) = \frac{P(A \setminus B_i)P(B_i)}{\sum_{j=1}^{n} P(A \setminus B_j)P(B_j)}, \quad i = 1,2,...,n.$$

(Κούτρας, 2004).

i) Συναρτήσεις κινδύνου και Μπεϋζιανοί εκτιμητές

Έχοντας καταλήξει στην επανακαθορισμένη κατανομή η ανάλυση μπορεί να θεωρείται ολοκληρωμένη. Στην ουσία όμως κάποιος συγκεκριμένος εκτιμητής με ένα περιθώριο λάθους ίσως να είναι προτιμότερος.

Ορισμός 2.1.13: Συνάρτηση απώλειας *(Loss function)* $L(\hat{\theta},\theta)$ είναι μια συνάρτηση η οποία επιβαρύνει την ζημιά στην περίπτωση που η εκτιμώμενη τιμή απομακρύνεται από την πραγματική.

Παραδείγματα τέτοιων συναρτήσεων απώλειας είναι τα παρακάτω:

α) Τετραγωνική συνάρτηση απώλειας (squared-error loss) $L(\hat{\theta},\theta) = (\hat{\theta}-\theta)^2$

β) Απόλυτη απώλεια (absolute loss) $L(\hat{\theta},\theta) = \mid \hat{\theta}-\theta \mid$

γ) Απώλεια μηδέν-ένα (zero-one loss) $L(\hat{\theta},\theta) = 0$ αν $\hat{\theta} = 0$ και 1 αλλιώς.

Ορισμός 2.1.14: Ο Μπευζιανός εκτιμητής για μια δοσμένη συνάρτηση ζημιάς είναι η συνάρτηση που ελαχιστοποιεί την αναμενόμενη ζημιά δοθέντος της εκ των υστέρων κατανομής.

2.1.4 Υποψήφιες κατανομές για το μέγεθος των αποζημιώσεων

Συχνά είναι επιθυμητό να βρεθεί μια αναλυτική έκφραση για μια κατανομή απώλειας ιδιαίτερα στην περίπτωση που τα εμπειρικά δεδομένα δεν είναι επαρκή. Υπάρχουν πολλές κατανομές στην στατιστική αλλά δεν είναι όλες κατάλληλες για να προσαρμόζονται σε δεδομένα μεγέθους αποζημιώσεων. Οι πιο συνηθισμένες κατανομές για τις αποζημιώσεις είναι η Pareto, η εκθετική, η log-normal,η Γάμμα, η Weibull κ.α .

i) Λογαριθμοκανονική κατανομή

Έστω τυχαία μεταβλητή X η οποία ακολουθεί την κανονική κατανομή με συνάρτηση πυκνότητας πιθανότητας

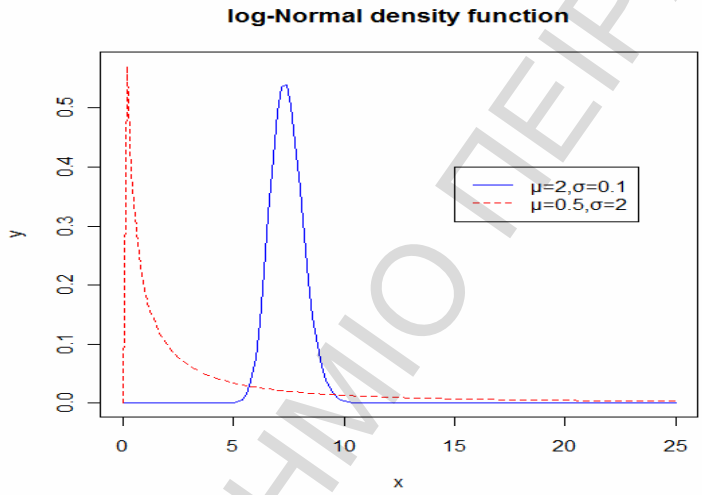
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty.$$

Έστω τώρα $Y = e^X$ έτσι ώστε $X = \log Y$.Η συνάρτηση πυκνότητας πιθανότητας της Y είναι η

$$f_X(y) = \frac{1}{\sigma\sqrt{2\pi}} \left\{-\frac{(\log y-\mu)^2}{2\sigma^2}\right\}, \quad 0 < y < \infty.$$

Η λογαριθμοκανονική κατανομή λόγω της σχετικά «βαριάς» δεξιάς ουράς είναι μια κατανομή που εφαρμόζει καλά σε αρκετές περιπτώσεις δεδομένων ζημιών. Όμως έχει και κάποια μειονεκτήματα όπως το γεγονός ότι η ροπογεννήτρια της δεν υπάρχει. Στο Σχήμα 2.1.4 απεικονίζεται η συνάρτηση πυκνότητας πιθανότητας της λογαριθμοκανονικής κατανομής με παραμέτρους μ=2,σ=0.1 (μπλέ γραμμή) και μ=0.5,σ=2 (κόκκινη διακεκομμένη γραμμή).

Σχήμα 2.1.4: Συνάρτηση πυκνότητας πιθανότητας της logNormal



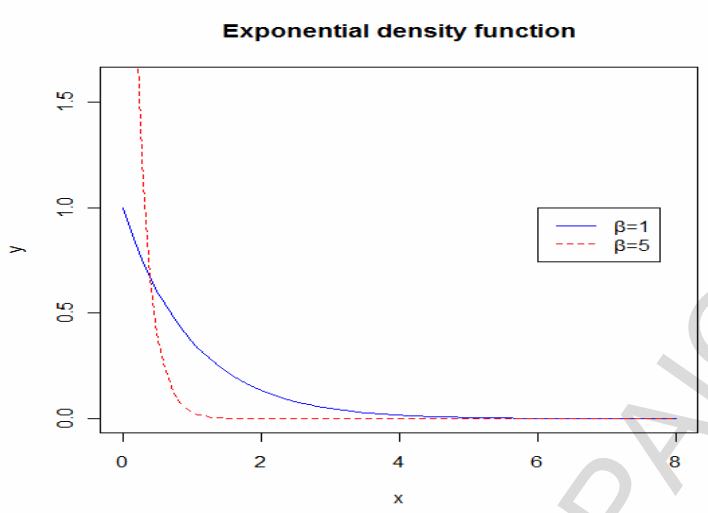
ii) Εκθετική κατανομή

Η συνάρτηση πυκνότητας της εκθετικής κατανομής είναι:

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

Η εκθετική κατανομή χρησιμοποιείται σε πολλούς ασφαλιστικούς κινδύνους λόγω των καλών μαθηματικών ιδιοτήτων που έχει. Ένα από τα μειονεκτήματα της εκθετικής πυκνότητας είναι ότι είναι μονότονη και φθίνουσα σε όλο το πεδίο ορισμού της το οποίο δυσκολεύει την εφαρμογή της σε πρακτικές εφαρμογές. Στο Σχήμα 2.1.5 απεικονίζεται η συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής με παραμέτρους b=1(μπλέ γραμμή) και b=5 (κόκκινη διακεκομμένη γραμμή).

Σχήμα 2.1.5: Συνάρτηση πυκνότητας πιθανότητας εκθετικής κατανομής



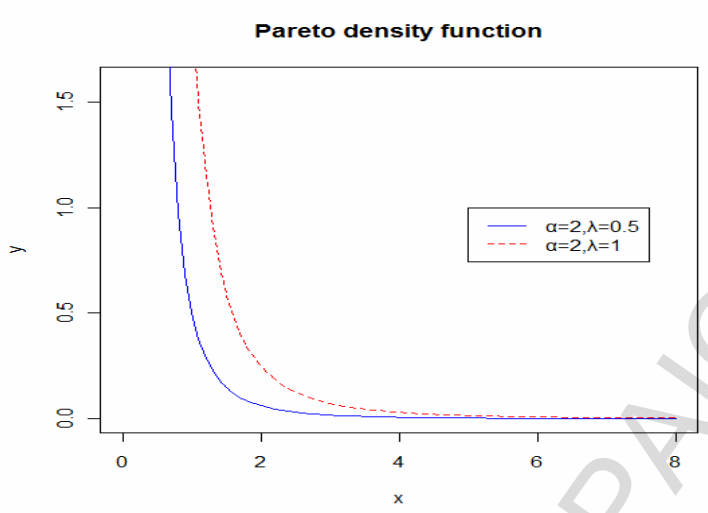
iii) Κατανομή Pareto

Έστω X τυχαία μεταβλητή ακολουθεί την εκθετική κατανομή με παράμετρο λ και η τυχαία μεταβλητή Y ακολουθεί την κατανομή Gamma(α,λ) τότε η X\Y ακολουθεί κατανομή Pareto με συνάρτηση πυκνότητας

$$f(x) = \frac{a\lambda^{a}}{(\lambda + x)^{a+1}}, \quad x > 0$$

Η κατανομή Pareto χρησιμοποιείται πολύ συχνά στην μοντελοποίηση του ύψους των ασφαλιστικών αποζημιώσεων κυρίως λόγω της πολύ βαριάς δεξιάς ουράς της. Τα μειονεκτήματα της οφείλονται όπως και στη log-normal στην έλλειψη της ροπογεννήτριας της και στο γεγονός πως η πυκνότητα είναι μονότονα φθίνουσα με αποτέλεσμα να μην προσαρμόζεται σε πρακτικές εφαρμογές *(Burnecki et al, 2010).* Στο Σχήμα 2.1.6 απεικονίζεται η συνάρτηση πυκνότητας πιθανότητας της κατανομής Pareto με παραμέτρους α=2,λ=0.5 (μπλέ γραμμή) και α=2, λ=1 (κόκκινη διακεκομμένη γραμμή).

Σχήμα 2.1.6: Συνάρτηση πυκνότητας πιθανότητας κατανομής Pareto



iv) Κατανομή Burr

Εμπειρικά έχει αποδειχτεί πως η κατανομή Pareto προσαρμόζεται με αρκετή επιτυχία στη κατανομή μεγέθους των αποζημιώσεων ιδιαίτερα στις περιπτώσεις που συμβαίνουν μεγάλες ζημιές. Σε πολλές περιπτώσεις όμως υπάρχει η ανάγκη εύρεσης μιας κατανομής που έχει βαριά ουρά αλλά παρουσιάζει μεγαλύτερη ευελιξία από την Pareto. Ένα τέτοιο παράδειγμα είναι η κατανομή Burr που έχει συνάρτηση πυκνότητας η οποία δεν είναι μονότονη.

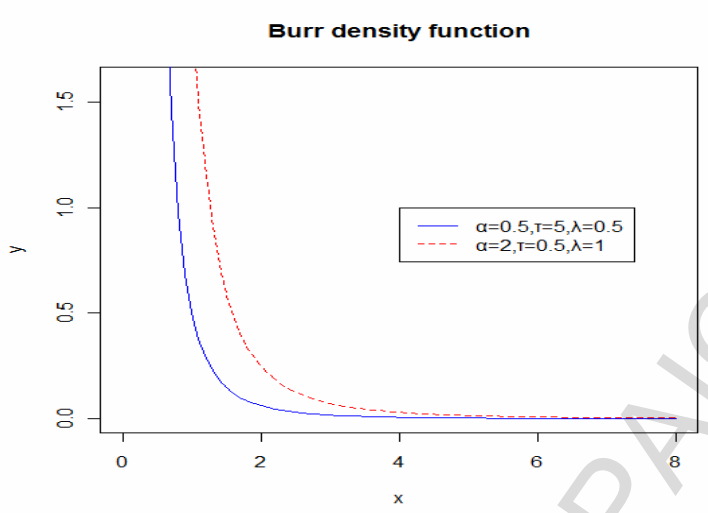
Έστω η τυχαία μεταβλητή Y ακολουθεί την κατανομή Pareto (α,λ) τότε η κατανομή της $X = Y^{1/\tau}$ ονομάζεται κατανομή Burr με συνάρτηση πυκνότητας

$$f(x) = \tau\alpha\lambda^{\alpha}\frac{x^{\tau-1}}{(\lambda+x)^{\alpha+1}}, \quad x>0$$

(Burnecki et al, 2010).

Στο Σχήμα 2.1.7 απεικονίζεται η συνάρτηση πυκνότητας πιθανότητας της κατανομής Burr με παραμέτρους a=0.5,b=5,λ=0.5 (μπλέ γραμμή) και a=2,b=0.5,λ=1 (κόκκινη διακεκομμένη γραμμή).

Σχήμα 2.1.7: Συνάρτηση πυκνότητας πιθανότητας της κατανομής Burr



2.2 Κατανομή για το πλήθος των αποζημιώσεων

Οι κατανομές που απαριθμούν το πλήθος των αποζημιώσεων που καταφθάνουν στην ασφαλιστική εταιρία είναι συνήθως διακριτές και έχουν θετική πιθανότητα μόνο για θετικές τιμές. Ο λόγος που χρησιμοποιούνται ξεχωριστά μοντέλα για το μέγεθος και για τον αριθμό των απαιτήσεων είναι γιατί ιδιαίτερα τα μοντέλα που απαριθμούν τον αριθμό των απαιτήσεων είναι εύκολο να προσδιορισθούν από απλές και συνηθισμένες διακριτές κατανομές. Έστω N η τυχαία μεταβλητή που παριστάνει τον αριθμό των απαιτήσεων. Τότε η συνάρτηση πιθανότητας της είναι: $p_k = \Pr(N = k), \quad k = 0,1,2,\ldots$. Η πιθανογεννήτρια της N είναι η $P(z)$ και υπολογίζεται από τη σχέση

$$P(z) = P_N(z) = E(z^N) = \sum_{k=0}^{\infty} p_k z^k .$$

Η ροπογεννήτρια της N είναι η

$$M_N(k) = E(e^{kN}) = \sum_{k=0}^{\infty} e^{kN} \cdot p_k .$$

2.2.1 Οι κυριότερες διακριτές κατανομές

i) κατανομή Poisson

Η συνάρτηση πιθανότητας της κατανομής Poisson είναι:

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,1,2\ldots$$

Η πιθανογεννήτρια της Poisson είναι:

$$P_N(z) = e^{\lambda(z-1)}, \quad \lambda > 0$$

Η ροπογεννήτρια της είναι:

$$M_N(z) = \exp(\lambda(e^z - 1)).$$

Η κατανομή Poisson είναι μια κατανομή με δύο πολύ χαρακτηριστικές ιδιότητες που συνοψίζονται στα παρακάτω θεωρήματα.

Θεώρημα 2.2.1: Έστω $N_1, N_2, \ldots N_n$ ανεξάρτητες και ισόνομές μεταβλητές από κατανομή Poisson με αντίστοιχες παραμέτρους $\lambda_1, \lambda_2, \ldots \lambda_n$. Τότε η $N = N_1 + N_2 +, \ldots + N_n$ ακολουθεί και αυτή κατανομή Poisson με παραμέτρους $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$ (*Klungman et al,2004*)

Θεώρημα 2.2.2: Έστω η τυχαία μεταβλητή N που απαριθμεί τον αριθμό των ενδεχομένων ακολουθεί κατανομή Poisson με μέση τιμή λ. Ας υποτεθεί πως κάθε ενδεχόμενο χωρίζεται σε m διαφορετικά ανεξάρτητα ενδεχόμενα με αντίστοιχες πιθανότητες $p_1, p_2, \ldots, p_m$. Τότε και τα $N_1, N_2, \ldots N_m$ είναι τυχαίες μεταβλητές που ακολουθούν Poisson με μέση τιμή $\lambda p_1, \lambda p_2, \ldots, \lambda p_m$,αντίστοιχα. (*Klungman et al,2004*)

ii) Διωνυμική κατανομή

Η διωνυμική κατανομή αποτελεί μια γενίκευση της κατανομής Bernoulli. Η διωνυμική κατανομή μετράει τον αριθμό των επιτυχιών σε n ανεξάρτητες δοκιμές του ίδιου πειράματος με την προυπόθεση ότι η πιθανότητα επιτυχίας παραμένει σταθερή και ίση με p. Μια σημαντική ιδιότητα της κατανομής αυτής είναι το γεγονός ότι ο μέσος της κατανομής είναι μεγαλύτερος από τη διακύμανση της. Η συνάρτηση πιθανότητας της διωνυμικής κατανομής είναι:

$$p_k = \Pr(N = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 1, 2, \ldots n$$

και η πιθανογεννήτρια της :

$$P_N(z) = [1 + p(z-1)]^n, 0 < p < 1.$$

iii) Αρνητική διωνυμική κατανομή

Η αρνητική Διωνυμική κατανομή χρησιμοποιείται συχνά σαν εναλλακτική της κατανομής Poisson. Παίρνει τιμές μόνο στον θετικό ημιάξονα αλλά το γεγονός ότι έχει δύο παραμέτρους την κάνει περισσότερο ευέλικτη από την κατανομή Poisson. Η αρνητική διωνυμική κατανομή μετράει τον αριθμό των αποτυχιών μέχρι την εμφάνιση τη r επιτυχίας. Συμβολίζεται με Nb (n,p) και η συνάρτηση πιθανότητας της δίνεται από τον τύπο

$$p_k = \binom{n+k-1}{k} p^n (1-p)^k \ , k = 0,1,2,\ldots$$

ενώ πιθανογεννήτρια της από τον τύπο $P_N(k) = \left(\frac{p}{1-(1-p)k} \right)^n$.

Η αρνητική διωνυμική σε αντίθεση με την διωνυμική έχει το χαρακτηριστικό ότι η μέση τιμή της είναι μικρότερη από την διακύμανση της.

iv) Γεωμετρική κατανομή

Η Γεωμετρική κατανομή αποτελεί μια ειδική περίπτωση της αρνητικής διωνυμικής κατανομής. Συγκεκριμένα είναι η αρνητική διωνυμική για την τιμή της παραμέτρου $n = 1$. Μια χαρακτηριστική ιδιότητα της γεωμετρικής κατανομής ονομάζεται ιδιότητα της έλλειψης μνήμης. Στην περίπτωση που η γεωμετρική κατανομή παρουσιάζει των αριθμό των απαιτήσεων ένα παράδειγμα της ιδιότητας έλλειψης μνήμης μπορεί να είναι το ακόλουθο: Αν με ανεξάρτητες δοκιμές Bernoulli ο αριθμός των αποτυχιών είναι ίσος με k τότε η πιθανότητα να χρειαστούν επιπλέον περισσότερες από r αποτυχίες για να εμφανιστεί η πρώτη απαίτηση ισούται με την πιθανότητα να απαιτηθούν r αποτυχίες μέχρι την πρώτη απαίτηση. Επομένως την στιγμή k το παρελθόν δεν υπάρχει και η διαδικασία είναι σαν να ξεκινάει από την αρχή.

Η συνάρτηση πιθανότητας της γεωμετρικής κατανομής είναι η : $p_k = p(1-p)^k \ , k = 0,1,2,\ldots$

Η πιθανογεννήτρια της γεωμετρικής κατανομής είναι: $P_N(k) = \frac{p}{1-(1-p)k}$

2.2.2 Οικογένεια κατανομών Panjer (a,b,0)

Μια πολύ γνωστή κλάση απαριθμητριών κατανομών στην αναλογιστική επιστήμη είναι η κλάση κατανομών του Panjer. Η κλάση κατανομών μελετήθηκε αρχικά από τον Katz (1965)

αλλά πήρε το όνομα της από τον Panjer επειδή εκείνος ανακάλυψε τον αναδρομικό αλγόριθμο που βοηθάει στον υπολογισμό της κατανομής του αθροίσματος των απαιτήσεων όταν ο αριθμός των απαιτήσεων ακολουθεί μια εκ των κατανομών της κλάσης Panjer. Ο παρακάτω ορισμός χαρακτηρίζει τις κατανομές που ανήκουν στην οικογένεια κατανομών του Panjer.

Ορισμός 2.2.1: Έστω p_k η συνάρτηση πιθανότητας μιας διακριτής τυχαίας μεταβλητής. Η κατανομή αυτή ανήκει στην οικογένεια κατανομών (a,b,0) όταν υπάρχουν σταθερές a ,b τέτοιες ώστε

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 1, 2, 3, \ldots$$

Η διωνυμική, η αρνητική διωνυμική (άρα και γεωμετρική) και η κατανομή Poisson είναι οι μοναδικές κατανομές που ανήκουν στην οικογένεια (a,b,0). Ο λόγος που γίνεται η κατηγοριοποίηση σε αυτήν την οικογένεια κατανομών είναι για να γίνεται ευκολότερα η διάκριση της κατανομής που ακολουθούν τα δεδομένα μεταξύ των προαναφερθέντων διακριτών κατανομών. Στον παρακάτω πίνακα φαίνονται οι τιμές των παραμέτρων a.b της κάθε κατανομής που ανήκουν στην κλάση (a,b,0).

Πίνακας 2.1.1: Οι τιμές των παραμέτρων της οικογένειας Panjer για τις τρείς διακριτές κατανομές

κατανομή	a	b	P_0
P(λ)	0	λ	$e^{-\lambda}$
B(n,p)	$\frac{-p}{(1-p)}$	$\frac{(n+1)p}{(1-p)}$	$(1-p)^n$
NB(r,p)	$(1-p)$	$(r-1)(1-p)$	p^r
G(p)	$(1-p)$	0	p

Μετασχήματίζοντας τον αριθμό των απαιτήσεων στη μορφή

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} = ak + b, \quad k = 1, 2, 3, \ldots$$

Η συνάρτηση του k είναι γραμμική με αποτέλεσμα μια γραφική αναπαράσταση της ευθείας προσδιορίζει την κλίση της και συνεπώς επιλέγεται αναλόγως η κατανομή που πιθανόν να ακολουθούν τα δεδομένα σύμφωνα με την τιμή της παραμέτρου a .

2.2.3 Οικογένεια κατανομών (a,b,1)

Οι κατανομές που ανήκουν στην κλάση (a,b,0) δεν περιγράφουν πάντα σωστά τα χαρακτηριστικά και το σχήμα κάποιων δεδομένων στην πράξη. Αυτό μπορεί να συμβαίνει διότι σε δεδομένα σπάνιων ζημιών μπορεί να υπάρχει αυξημένη πιθανότητα στο σημείο μηδέν ενώ σε κάποια άλλα είδη δεδομένων η πιθανότητα στην τιμή μηδέν να πλησιάζει το μηδέν. Στις παραπάνω περιπτώσεις δεν υπάρχει στο σημείο μηδέν καλή εφαρμογή των δεδομένων με την επιλεγμένη κατανομή. Για αυτό το λόγο υπάρχει η δυνατότητα μετατροπής της πιθανότητας στο μηδέν στις κατανομές που ανήκουν στην κλάση κατανομών (a,b,0).

Ορισμός 2.2.2:Εστω p_k η συνάρτηση πιθανότητας μιας διακριτής τυχαίας μεταβλητής. Η κατανομή αυτή ανήκει στην οικογένεια κατανομών (a,b,1) όταν υπάρχουν σταθερές a , b τέτοιες ώστε

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 2,3,4,\ldots$$

Η μόνη διαφορά της κλάσης κατανομών (a,b,1) είναι ότι ο αναδρομικός τύπος υπολογισμού ξεκινάει από την p_1 και όχι από την p_0 που ξεκινούσε στην κλάση κατανομών (a,b,0).Είναι φανερό πως η κλάση (a,b,0) αποτελεί μια υποκατηγορία της ευρύτερης κλάσης (a,b,1).

Από την κλάση κατανομών (a,b,1) διαχωρίζουμε την υποκατηγορία των κατανομών για τις οποίες $p_0 = 0$ και η οποία ονομάζεται αποκομμένη στο μηδέν (*zero-truncated*) και στην υποκατηγορία όπου $p_0 > 0$ και η οποία ονομάζεται τροποποιημένη στο μηδέν (*zero-modified*). Οι τροποποιημένες στο μηδέν (*zero-modified)* κατανομές μπορούν να εκφραστούν και ως κατανομές της κλάσης (a,b,0) και μιας εκφυλισμένης κατανομής πιθανότητα μόνο στο μηδέν. Οι αποκομμένες στο μηδέν (*zero-truncated*) κατανομές μπορούν να εκφραστούν και ως τροποποιημένες στο μηδέν (*zero-modified)* κατανομές με συγκεκριμένη τροποποίηση στο $p_0 = 0$.

2.2.4 Σύνθετα μοντέλα συχνότητας

Μια μεγαλύτερη κλάση κατανομών μπορεί να δημιουργηθεί συνθέτοντας οποιεσδήποτε δύο διακριτές κατανομές. Ο όρος σύνθεση αντικατοπτρίζει το γεγονός ότι η πιθανογεννήτρια της νέας σύνθετης κατανομής εξαρτάται από τις πιθανογεννήτριες των δύο επιλεγμένων διακριτών κατανομών. Έστω N απαριθμήτρια τυχαία μεταβλητή με πιθανογεννήτρια $P_N(z)$ και $M_1, M_2, \ldots$ απαριθμήτριες τυχαίες μεταβλητές με πιθανογεννήτρια $P_M(z)$. Υποθέτοντας ότι M_j ανεξάρτητο από το N τότε η πιθανογεννήτρια του τυχαίου αθροίσματος $S = M_1 + M_2 + \cdots + M_N$ είναι η :

$$P_S(z) = P_N[P_M(z)],$$

όπου N, M καλούνται πρωταρχική (*primary*) και δευτερεύουσα (*secondary*) κατανομή αντίστοιχα.

2.3 Κατανομές απώλειας για το άθροισμα των απαιτήσεων

Σκοπός της ασφαλιστικής επιχείρησης είναι η αποτελεσματική διαχείριση των κινδύνων. Με την ασφάλιση μεγάλου αριθμού ατόμων ο ατομικός κίνδυνος μετατρέπεται σε ένα μοντέλο συνολικού αθροιστικού κινδύνου το οποίο είναι διαχειρίσιμο. Το συσσωρευμένο ποσό μπορεί να είναι οι απαιτήσεις των ασφαλισμένων από την ασφαλιστική εταιρία ή ακόμα και οι απαιτήσεις της ασφαλιστικής εταιρίας από μια αντασφαλιστική εταιρία. Έστω τα τυχαία ποσά των πληρωμών $X_1, X_2, \ldots, X_N$ όπου N ο τυχαίος αριθμός των απαιτήσεων. Στο **συλλογικό μοντέλο κινδύνου** ***(collective risk model)*** η συνολική αθροιστική απώλεια θεωρείται ένα τυχαίο άθροισμα S

$$S = X_1 + X_2 + \ldots + X_N ,$$

όπου X είναι τυχαία μεταβλητή του μεγέθους των αποζημιώσεων και N τυχαία μεταβλητή του αριθμού των αποζημιώσεων. Στην περίπτωσή που το N=0 τότε και το S=0. Το τυχαίο άθροισμα S είναι μια σύνθετη κατανομή της σφοδρότητας και της συχνότητας των απαιτήσεων. Ο διαχωρισμός του αθροίσματος S σε σφοδρότητα και συχνότητα προσφέρει αρκετά πλεονεκτήματα στη μοντελοποίηση του. Μια μεγάλη αύξηση στον αριθμό των συμβολαίων της ασφαλιστικής επιχείρησης θα επιφέρει την αναπροσαρμογή μόνο της κατανομής που ακολουθεί ο αριθμός των αποζημιώσεων. Αντίστοιχα πληθωριστικές

μεταβολές ή άλλες αλλαγές καλύψεων των συμβολαίων (αφαιρετέα ποσά, ανώτατα όρια κάλυψης κλπ) θα επιφέρουν αλλαγές στην κατανομή του μεγέθους των απαιτήσεων.

Στο **μοντέλο ατομικού κινδύνου** ***(individual risk model)*** το συνολικό άθροισμα των απαιτήσεων περιγράφεται από το άθροισμα

$$S = X_1 + X_2 + \ldots + X_n,$$

όπου τα X_j είναι ανεξάρτητα αλλά όχι κατ' ανάγκη ισόνομα και ο αριθμός των απαιτήσεων είναι γνωστός και ίσος με n.

Σχέση μεταξύ των δύο μοντέλων

Στην περίπτωση που υπάρχει ανάγκη μοντελοποίησης της κατανομής του αθροίσματος των απαιτήσεων όχι ενός αλλά πολλών διαφορετικών χαρτοφυλακίων ή κλάδων ασφαλιστικών συμβολαίων ή πολλών διαφορετικών καλύψεων σε ένα ασφαλιστήριο συμβόλαιο τότε προκύπτει η ανάγκη σύνδεσης του ατομικού μοντέλου κινδύνου με το συλλογικό. Έστω Y_i είναι το άθροισμα των απαιτήσεων του i ασφαλιστικού κλάδου. Τότε για το μοντέλο του ατομικού κινδύνου το συνολικό άθροισμα των απαιτήσεων ισούται με $S = Y_1 + Y_2 + \cdots + Y_n$. Έστω τώρα ότι N_i ο αριθμός των απαιτήσεων στον i ασφαλιστικό κλάδο έτσι ώστε $Y_i = X_1 + X_2 + \cdots + X_{Ni}$ και $N = N_1 + N_2 + \cdots + N_n$. Στην συγκεκριμένη περίπτωση το φαινομενικά απλό μοντέλο ατομικού κινδύνου ανά κλάδο έκρυβε ένα μοντέλο συλλογικού κινδύνου στον κάθε ασφαλιστικό κλάδο ξεχωριστά.

2.3.1 Σύνθεση μοντέλων για τις συνολικές ζημιές

Έστω S το άθροισμα των απαιτήσεων στο συλλογικό μοντέλο κινδύνου. Η επιλογή ενός μοντέλου για την εκτίμηση της κατανομής του S περιλαμβάνει τρία βήματα.

1. Επιλογή της κατανομής για την τυχαία μεταβλητή N βασισμένη στα δεδομένα.
2. Επιλογή της κατανομής για την τυχαία μεταβλητή X_j βασισμένη στα δεδομένα.
3. Υπολογισμός του αθροίσματος των απαιτήσεων S με την χρησιμοποίηση των κατανομών των N, X

(Klungman et al, 2004

Στην περίπτωση που τα πρώτα δύο βήματα έχουν αναπτυχθεί αρκετά υπάρχουν τρόποι ώστε να υπολογισθεί αριθμητικά το συνολικό άθροισμα των απαιτήσεων.

Το τυχαίο άθροισμα $S = X_1 + X_2 + \cdots + X_N$ έχει συνάρτηση κατανομής:

$$\begin{aligned} F_S(x) &= \Pr(S \le x) \\ &= \sum_{n=0}^{\infty} p_n \Pr(S \le x \setminus N = n) \\ &= \sum_{n=0}^{\infty} p_n F_X^{*n}(x), \end{aligned}$$

όπου $F_X(x) = \Pr(X \le x)$ είναι η συνάρτηση κατανομής της X_j και $p_n = \Pr(N = n)$. Η κατανομή της S καλείται σύνθετη κατανομή. Η $F_X^{*n}(x)$ ονομάζεται ΄΄n-οστή συνέλιξη΄΄ της συνάρτησης κατανομής της X και υπολογίζεται ως:

$$F_X^{*0}(x) = \begin{cases} 0, & x < 0, \\ 1, & x \ge 0, \end{cases}$$

και

$$F_X^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*(k-1)}(x-y) dF_X(y) \quad \text{για } k = 1,2,\ldots$$

(Klungman et al, 2004).

Στην περίπτωση που η X είναι συνεχής τυχαία μεταβλητή με θετική πιθανότητα μόνο για θετικές τιμές τότε ο παραπάνω τύπος καταλήγει στην παρακάτω απλουστευμένη μορφή:

$$F_X^{*k}(x) = \int_0^x F_X^{*(k-1)}(x-y) f_X(y) dy \quad \text{για } k = 2,3,\ldots$$

ενώ η αντίστοιχη πυκνότητα δίνεται από την αναδρομική σχέση:

$$f_X^{*k}(x) = \int_0^x f_X^{*(k-1)}(x-y) f_X(y) dy \quad \text{για } k = 2,3,\ldots$$

Με βάση τα παραπάνω, προκύπτει ότι η κατανομή της S έχει συνάρτηση πυκνότητας για $x > 0$:

$$f_S(x) = \sum_{n=1}^{\infty} p_n f_X^{*n}$$

επιπρόσθετα με μια μάζα πιθανότητας στο μηδέν.

Στην περίπτωση που η X είναι διακριτή τυχαία μεταβλητή με πιθανότητα στα σημεία $0,1,2,\ldots$ τότε η συνάρτηση κατανομής της είναι η:

$$F_X^{*k}(x) = \sum_{y=0}^{x} F_X^{*(k-1)}(x-y) f_X(y) \text{ για } x = 0,1,\ldots, k = 2,3,\ldots$$

και η αντίστοιχη συνάρτηση πιθανότητας:

$$f_X^{*k}(x) = \sum_{y=0}^{x} f_X^{*(k-1)}(x-y) f_X(y) \text{ για } x = 0,1,\ldots, k = 2,3,\ldots$$

Η πιθανογεννήτρια του αθροίσματος των απαιτήσεων ισούται με:

$$\begin{aligned} P_S(z) &= E[z^S] \\ &= P_N[P_X(z)]. \end{aligned}$$

Ο μεταΣχήματισμός Laplace και η ροπογεννήτρια της σύνθετης κατανομής του αθροίσματος των απαιτήσεων υπολογίζονται ως:

$$\begin{aligned} L_S(z) &= E(e^{-zS}) \\ &= P_N[L_X(z) \end{aligned}$$

και

$$M_S(z) = P_N[M_X(z)].$$

2.3.2 Υπολογισμός της κατανομής του αθροίσματος των απαιτήσεων

Ο υπολογισμός της κατανομής του αθροίσματος των απαιτήσεων ακόμα και στις πιο απλές περιπτώσεις συνήθως είναι μια δύσκολη διαδικασία. Για αυτό το λόγο έχουν αναπτυχθεί εναλλακτικοί τρόποι αριθμητικής εκτίμησης των αποτελεσμάτων για συγκεκριμένες περιπτώσεις κατανομών.

i) προσεγγιστική κατανομή

Η προσεγγιστική κατανομή χρησιμοποιεί την μέθοδο των ροπών για να εκτιμήσει τις παραμέτρους της κατανομής. Το μεγάλο πλεονέκτημα αυτής της μεθόδου είναι το γεγονός ότι αποτελεί μια ιδιαίτερα απλή στην εφαρμογή της μέθοδο. Υπάρχουν όμως και σοβαρά μειονεκτήματα αυτής της μεθόδου. Ένα από αυτά είναι ότι δεν υπάρχει τρόπος να

αξιολογηθεί πόσο καλή είναι αυτή η προσέγγιση. Κάθε διαφορετική προσέγγιση δίνει διαφορετικά αποτελέσματα ειδικότερα στη δεξιά ουρά της κατανομής και ο μόνος τρόπος να βελτιωθεί είναι η χρησιμοποίηση περισσοτέρων ροπών. Επίσης η προσεγγιστική κατανομή δεν μπορεί να αναπαραστήσει με ακρίβεια κάποια ειδικά χαρακτηριστικά της πραγματικής κατανομής όπως για παράδειγμα στην περίπτωση που υπάρχει ένα ανώτατο όριο ιδίας κράτησης τότε η κατανομή της σφοδρότητας των απαιτήσεων θα παρουσιάζει αυξημένη πιθανότητα εμφάνισης σε αυτό το όριο που θα αποτελεί και το μέγιστο της κατανομής.

ii) Η αναδρομική μέθοδος

Έστω η κατανομή του μεγέθους των αποζημιώσεων $f_X(x)$ παίρνει τιμές $0,1,2,\ldots,m$ που αντιστοιχούν σε πολλαπλάσια κάποιας νομισματικής μονάδας και η τιμή m είναι η μέγιστη πληρωμή. Έστω ακόμα πως η κατανομή του αριθμού των αποζημιώσεων p_k ανήκει στην κλάση κατανομών του Panjer (a,b,1). Τότε η κατανομή του αθροίσματος των αποζημιώσεων περιγράφεται από την σχέση:

$$f_S(x) = \frac{[p_1 - (a+b)p_0]f_X(x) + \sum_{y=1}^{x\wedge m}(a+by/x)f_X(y)f_S(x-y)}{1-af_X(0)},$$

όπου $x \wedge m$ είναι ο συμβολισμός για το $\min(x,m)$.

2.3.3 Κατασκευή διακριτών κατανομών

Η αναδρομική μέθοδος αναπτύχθηκε κυρίως για τη χρήση διακριτών κατανομών. Αν και η υπάρχει η δυνατότητα να κατασκευαστεί αναδρομικός τύπος και για συνεχείς κατανομές προτιμάται λόγω ευκολίας η μετατροπή των συνεχών κατανομών σε διακριτές με μια μέθοδο η οποία ονομάζεται διακριτοποίηση και χρησιμοποίηση του τύπου της αναδρομικής μεθόδου για διακριτές κατανομές. Ο ευκολότερος τρόπος κατασκευής διακριτής κατανομής για το μέγεθος των απαιτήσεων από μια συνεχή κατανομή είναι η απόδοση μάζας πιθανότητας σε πολλαπλάσια μιας ΄΄βολικής΄΄ μονάδα μέτρησης. Μια κατανομή με αυτά τα χαρακτηριστικά ονομάζεται αριθμητική κατανομή αφού ορίζεται στον θετικό ημιάξονα. Η κατασκευή αξιόπιστης αριθμητικής κατανομής προυποθέτει την διατήρηση των χαρακτηριστικών ιδιοτήτων της αρχικής συνεχούς κατανομής. Οι πιο διαδεδομένοι τρόποι διακριτοποίησης μιας συνεχούς κατανομής είναι οι ακόλουθοι:

i. Μέθοδος στρογγυλοποίησης *(method of rounding)*

Έστω f_j η πιθανότητα στο jh σημείο όπου h είναι η μονάδα μέτρησης και $j = 1, 2, \ldots$ Οι τιμές της τυχαίας μεταβλητής X στρογγυλοποιούνται στο πλησιέστερο πολλαπλάσιο του h με εξαίρεση την πιθανότητα στο σημείο 0.

$$f_0 = \Pr\left(X < \frac{h}{2} \right) = F_X\left(\frac{h}{2} - 0 \right),$$

$$f_j = \Pr\left(jh - \frac{h}{2} \le X < jh + \frac{h}{2} \right)$$

$$= F_X\left(jh + \frac{h}{2} - 0 \right) - F_X\left(jh - \frac{h}{2} - 0 \right), \quad j = 1, 2, \ldots$$

Δηλαδή για :

$X \in \left[0, \frac{h}{2} \right] \rightarrow$ Οι τιμές στρογγυλοποιούνται στο 0.

$X \in \left[\frac{h}{2}, \frac{3h}{2} \right] \rightarrow$ Οι τιμές στρογγυλοποιούνται στο h.

$X \in \left[\frac{3h}{2}, \frac{5h}{2} \right] \rightarrow$ Οι τιμές στρογγυλοποιούνται στο 2h.

(Χατζηκωνσταντινίδης,2011)

ii. Μέθοδος διατήρησης των ροπών *(method of local method matching)*

Σε αυτή τη μέθοδο κατασκευάζεται η νέα διακριτή κατανομή με την προυπόθεση ότι οι p πρώτες ροπές της αριθμητικής κατανομής συμπίπτουν με τις αντίστοιχες πρώτες ροπές της αρχικής συνεχούς κατανομής. Έστω ένα διάστημα μήκους ph το οποίο συμβολίζεται $[x_k, x_k + ph)$ όπου p είναι ο αριθμός των ροπών και h το βήμα. Θέτοντας μάζα πιθανότητας $m_0^k, m_1^k, \ldots, m_p^k$ στα σημεία $x_k, x_k + h, \ldots, x_k + ph$ έτσι ώστε οι πρώτες p ροπές να συμπίπτουν. Το σύστημα των εξισώσεων που ικανοποιούν αυτές τις συνθήκες είναι το ακόλουθο:

$$\sum_{j=0}^{p} \left(x_k + jh \right)^r m_j^k = \int_{x_k - 0}^{x_k + ph - 0} x^r \, dF_X(x), \quad r = 0, 1, 2, \ldots . p,$$

Στη συνέχεια θεωρώντας το διάστημα $[x_{k+1}, x_{k+1} + ph)$ έτσι ώστε $x_{k+1} = x_k + ph$. Θέτοντας όπου $x_0 = 0$ η διακριτή κατανομή έχει τις εξής μάζες πιθανότητας.

$$f_0 = m_0^0, \qquad f_1 = m_1^0, \qquad f_2 = m_2^0, \ldots,$$
$$f_p = m_p^0 + m_0^1, \quad f_{p+1} = m_1^1, \quad f_{p+2} = m_2^1, \ldots$$

ΚΕΦΑΛΑΙΟ 3

Εκτίμηση των ακραίων παρατηρήσεων με χρήση της Θεωρίας Ακραίων Τιμών

3.1 Εισαγωγή στη θεωρία ακραίων τιμών

Η θεωρία ακραίων τιμών αποτελεί έναν ξεχωριστό κλάδο της Στατιστικής όπου σκοπό έχει τη μελέτη στοχαστικών μοντέλων που αφορούν την εμφάνιση ακραίων παρατηρήσεων. Το πλήθος των εφαρμογών της θεωρίας των ακραίων τιμών είναι πολυάριθμο, γεγονός που την καθιστά ένα πολύτιμο εργαλείο για πολλούς επιστήμονες και που εξηγεί εν μέρει την γρήγορη ανάπτυξη της συγκεκριμένης θεωρίας τις τελευταίες δεκαετίες. Η αρχή για τη μελέτη της θεωρίας ακραίων τιμών έγινε από τους Fisher και Tippett το 1928 αν και ο κλάδος γνώρισε ιδιαίτερη άνθηση μετά το 1970. Αναφορικά με την θεωρία ακραίων τιμών ενδιαφέρον εξέφρασαν επιστήμονες από πολλούς κλάδους όπως οι:

- Μετεωρολόγοι για την πρόβλεψη ακραίων καιρικών φαινομένων όπως πλημμύρες, τυφώνες, κ.α.
- Μηχανικοί για την κατασκευή φραγμάτων σε ποτάμια, λίμνες όπου τους ενδιαφέρει τόσο η στάθμη των υδάτων όσο και η αξιοπιστία των υλικών και συνολικά της κατασκευής. κ.α.
- Γεωλόγοι για το ενδεχόμενο σεισμών.
- Οικονομολόγοι για το ενδεχόμενο εμφάνισης ακραίων φαινομένων που μπορούν να οδηγήσουν σε ζημιές στα χαρτοφυλάκια επενδύσεων. κ.α.

Η παρούσα εργασία σκοπό έχει να αξιοποιήσει την συνεισφορά της θεωρίας ακραίων τιμών στην ασφάλιση και συγκεκριμένα στην μελέτη «πολύ» μεγάλων αποζημιώσεων που μπορούν να προκύψουν σε ένα χαρτοφυλάκιο ασφάλισης αυτοκινήτων. Θα πρέπει να διευκρινιστεί σε αυτό το σημείο ότι οι κυριότερες αναφορές του συγκεκριμένου κεφαλαίου έχουν προέλθει από τους Coles (2001) και Μπούτσικας (2008).

3.1.1 Μοντέλα της κλασσικής θεωρίας ακραίων τιμών

Το παρόν κεφάλαιο επικεντρώνεται στην ανάπτυξη του μοντέλου που αποτελεί τη βάση της θεωρίας ακραίων τιμών και συγκεκριμένα στη συμπεριφορά του μεγίστου μιας ακολουθίας τυχαίων μεταβλητών.

Έστω

$$M_n = \max(X_1, \ldots, X_n),$$

όπου $X_1, \ldots, X_n$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση κατανομής F.

Η κατανομή της M_n εξάγεται εύκολα για όλες τις τιμές του n:

$$\begin{aligned} P(M_n \le x) &= P(X_1 \le x, \ldots, X_n \le x) \\ &= P(X_1 \le x), \ldots, P(X_n \le x) \\ &= (F(x))^n. \end{aligned}$$

Στην πραγματικότητα επειδή η κατανομή της F είναι άγνωστη δεν ενδείκνυται να χρησιμοποιηθεί ο παραπάνω τύπος και θα πρέπει να βρεθούν άλλοι τρόποι για να εκτιμηθεί η ζητούμενη κατανομή. Μια πρώτη προσέγγιση θα μπορούσε να είναι η εκτίμηση της κατανομής της F από τα δεδομένα των παρατηρήσεων όμως και αυτή η προσέγγιση δεν θα έδινε ασφαλή συμπεράσματα αφού μικρές αποκλίσεις στην κατανομή της F θα επιφέρει μεγάλες αποκλίσεις στην κατανομή της F^n.

Μια συνηθισμένη τακτική προσέγγισης είναι να θεωρηθεί ότι η κατανομή της F είναι άγνωστη και να προσεγγιστεί η οριακή κατανομή της F^n μόνο από τα δεδομένα των ακραίων παρατηρήσεων. Η συγκεκριμένη διαδικασία είναι εφικτή μέσω μιας διαδικασίας κανονικοποίησης, παρόμοιας φιλοσοφίας με την κανονικοποίηση που ισχύει για τον δειγματικό μέσο από το Κ.Ο.Θ.

Έστω x_F ,το δεξί άκρο του στηρίγματος της F ,

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

δηλαδή, το δεξιό άκρο των τιμών που μπορούν να πάρουν τα X_i με θετική πιθανότητα.

$$P(M_n \le x) = F(x)^n \to \begin{cases} 0, & x < x_F \\ 1, & x > x_F \end{cases}$$

(Μπούτσικας 2008)

Η ακολουθία των τ.μ. M_n είναι αύξουσα οπότε για $n \to \infty$ θα ισχύει ότι $M_n \to x_F$ σχεδόν βέβαια. Επομένως το δειγματικό μέγιστο M_n συγκλίνει στο πληθυσμιακό μέγιστο x_F. Η συμβολή της συγκεκριμένης πληροφορίας έγκειται στο γεγονός ότι θυμίζει κατά μια έννοια το κεντρικό οριακό θεώρημα (Κ.Ο.Θ.) και συγκεκριμένα τη σύγκλιση του δειγματικού μέσου στο μέσο του πληθυσμού. Με παρόμοια ίσως κανονικοποίηση με αυτή που συμβαίνει στο Κ.Ο.Θ. θα μπορούσε λοιπόν να βρεθεί μια οριακή κατανομή του μεγίστου ανεξάρτητη από την αρχική κατανομή των δεδομένων.

Έστω M_n^* ένας μετασχηματισμός του M_n έτσι ώστε:

$$M_n^* = \frac{M_n - b_n}{a_n},$$

όπου a_n, b_n ακολουθίες με $a_n > 0, b_n \in R$. Με κατάλληλη επιλογή των a_n, b_n σταθεροποιείται η μέση τιμή και η διακύμανση της M_n^* όσο το n μεγαλώνει αποφεύγοντας τις δυσκολίες που προκύπτουν με την κατανομή της M_n. Η λύση στο πρόβλημα της εύρεσης της οριακής κατανομής της M_n^* προήλθε από τους Fisher και Tippett το 1928 όταν και απέδειξαν ότι το μέγιστο ενός δείγματος τυχαίων μεταβλητών ύστερα από μια κανονικοποίηση συγκλίνει σε μια εκ των τριών κατανομών Gumbel, Frechet ή Weibull.

Θεώρημα 3.1.1 (Fisher-Tippett theorem) **:** Έστω $(X_1, X_2, \ldots, X_n)$ μια ακολουθία από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές , $M_n = \max(X_1, \ldots, X_n)$. Αν υπάρχουν ακολουθίες πραγματικών αριθμών (a_n, b_n) έτσι ώστε $a_n > 0$ και

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

όπου G μια μη-εκφυλισμένη συνάρτηση κατανομής , τότε η G ανήκει σε μια εκ των τριών ακόλουθων κατανομών.

$$Gumbel: G(x) = \exp\left\{-\exp\left[-\frac{x-b}{a}\right]\right\}, x \in R \cdot$$

$$Frechet: G(x) = \begin{cases} 0 & , x \le b \\ \exp\left\{ -\left(\frac{x-b}{a} \right)^{-\gamma} \right\} & , x > b \end{cases}$$

$$Weibull: G(x) = \begin{cases} \exp\left\{ -\left[-\left(\frac{x-b}{a} \right) \right]^{\gamma} \right\} & , x < b \\ 1 & , x \ge b \end{cases}$$

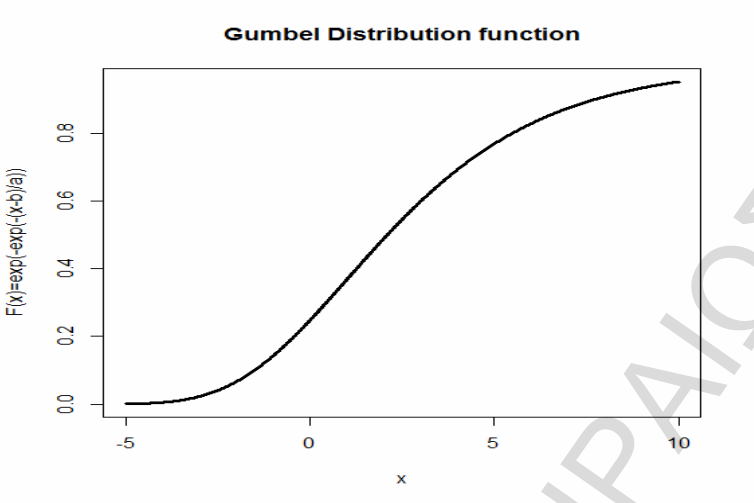
όπου $a > 0, b \in R$ και $\gamma > 0$.

Το παραπάνω θεώρημα δηλώνει πως το δειγματικό μέγιστο $\frac{M_n - b_n}{a_n}$ συγκλίνει σε μια κατανομή εκ των Gumbel, Frechet ή Weibull οι οποίες αποτελούν την οικογένεια κατανομών των ακραίων τιμών. Κάθε μία από τις κατανομές έχει μια παράμετρο θέσης b, μια παράμετρο κλίμακας a και η κατανομές Frechet και Weibull έχουν μια επιπρόσθετη παράμετρο σχήματος γ. Επίσης το παραπάνω θεώρημα δηλώνει πως με κατάλληλες ακολουθίες a_n, b_n η M_m^* θα συγκλίνει σχεδόν κατά αποκλειστικότητα σε μια από τις τρείς προαναφερθείσες κατανομές ανεξάρτητα από την κατανομή της F. Οι μόνες περιπτώσεις που δεν μπορεί να κανονικοποιηθεί το μέγιστο ανεξάρτητα από την επιλογή των ακολουθιών a_n, b_n συμβαίνει όταν :

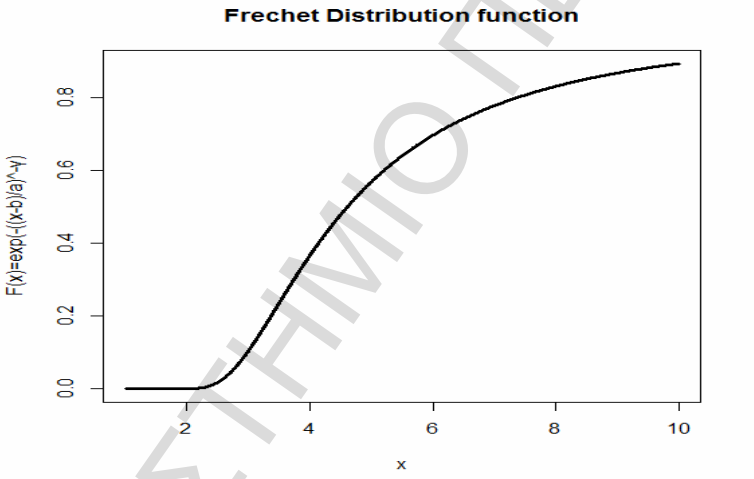
- τα X_i έχουν μια κατανομή η οποία λαμβάνει μια μέγιστη τιμή με θετική πιθανότητα
- ή στην περίπτωση που το x_F δεν είναι πεπερασμένο και τα άλματα από αριστερά προς τα δεξιά δεν φθίνουν.

Τα παρακάτω γραφήματα δείχνουν τη μορφή που έχουν οι συναρτήσεις κατανομής των Gumbel, Frechet και Weibull αντίστοιχα.

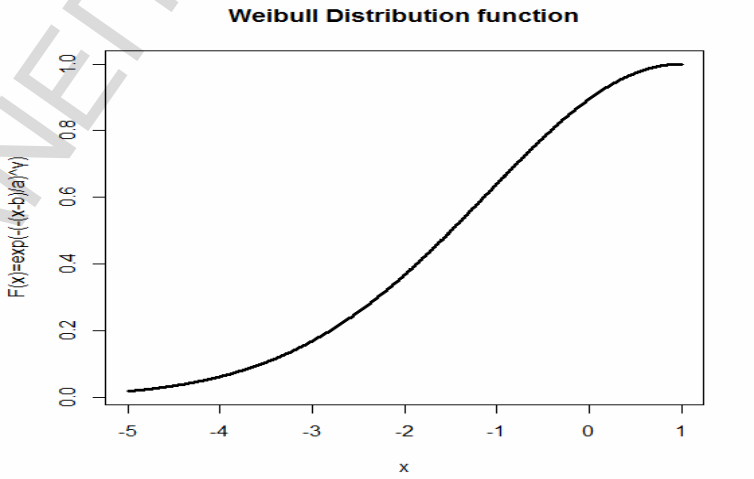
Σχήμα 3.1.1: Συνάρτηση κατανομής Gumbel για a=3,b=1



Σχήμα 3.1.2: Συνάρτηση κατανομής Frechet για a=3, b=1, γ=1



Σχήμα 3.1.3: Συνάρτηση κατανομής Weibull για a=3, b=1, γ=1



3.1.2 Περιοχή έλξης κατανομής μεγίστου (maximum domain of attraction)

Η οριακή κατανομή ενός κανονικοποιημένου μεγίστου M_n^*, όπως έχει αναφερθεί στην προηγούμενη ενότητα, έχει (όταν υπάρχει) τον ίδιο τύπο με μια εκ των Gumbel, Frechet και Weibull. Επομένως σε κάθε κατανομή των X_i θα αντιστοιχεί μια μοναδική οριακή κατανομή. Το ερώτημα λοιπόν που θα πρέπει να απαντηθεί είναι σε ποιες κατανομές των X_i αντιστοιχεί κάθε μια εκ των τριών οριακών κατανομών. Απάντηση στο παραπάνω ερώτημα έρχεται να δώσει η έννοια της περιοχής έλξης ακροτάτων.

Ορισμός 3.1.2: Έστω τ.μ. X με σ.κ. F. Θεωρούμε πως η X ανήκει στην περιοχή έλξης μιας κατανομής ακροτάτων G αν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in R, n \in N$ για τις οποίες ισχύει

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

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

Το χαρακτηριστικό μιας κατανομής το οποίο καθορίζει σε ποια περιοχή έλξης ανήκει είναι η μορφή της δεξιάς ουράς της, δηλαδή πόσο γρήγορα συγκλίνει στο μηδέν η $\bar{F}(x)$ όταν $x \to x_F$. Ένας λοιπόν τρόπος κατηγοριοποίησης των κατανομών μπορεί να είναι σύμφωνα με την ταχύτητα σύγκλισης της δεξιάς ουράς τους στο μηδέν. Ένας τέτοιος καθορισμός γίνεται με τις συναρτήσεις ομαλής κύμανσης.

Ορισμός 3.1.3: Μια θετική, Lebesgue μετρήσιμη συνάρτηση h στο $(0,\infty)$ καλείται ομαλής κύμανσης (*regurarly varying*) στο ∞ με δείκτη $a \in R$ (συμβ. $h \in \Re_a$) αν

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

Ορισμός 3.1.4: Μια θετική, Lebesgue μετρήσιμη συνάρτηση L στο $(0,\infty)$ καλείται βραδείας κύμανσης (*slowly* varying) στο ∞ (συμβ. $h \in \Re_0$) αν

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

Ορισμός 3.1.5: Μια θετική, Lebesgue μετρήσιμη συνάρτηση h στο $(0,\infty)$ καλείται ταχείας κύμανσης (*rapidly varying*) στο ∞ με δείκτη $-\infty$ (συμβ. $h \in \Re_{-\infty}$) αν

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

Οι τρείς παραπάνω ορισμοί δείχνουν την ταχύτητα με την οποία κάποια συνάρτηση μεταβάλλεται καθώς η ανεξάρτητη μεταβλητή της πλησιάζει το άπειρο. Μερικά παραδείγματα συναρτήσεων που έχουν τις παραπάνω ιδιότητες είναι τα ακόλουθα: Συνάρτηση ομαλής κύμανσης είναι σίγουρα η x^a με $a \in R$,συνάρτηση βραδείας κύμανσής είναι η $\ln(x)$ ενώ ταχείας κύμανσης η e^{-x}.

Περιοχές έλξης με βάση τη μορφή της δεξιάς ουράς

Είναι προφανές πως δυο κατανομές μπορεί να έχουν διαφορετική μορφή στο R αλλά η δεξιά ουρά τους να έχει την ίδια συμπεριφορά. Δύο τέτοιες κατανομές από εδώ και στο εξής θα τις ονομάσουμε κατανομές «ισοδύναμης ουράς». Ο παρακάτω ορισμός (*Boutsikas,2008)* καταδεικνύει ακριβώς αυτό το χαρακτηριστικό.

Ορισμός 3.1.6: Δυο συναρτήσεις κατανομής F, G έχουν ισοδύναμη ουρά αν $x_F = x_G$ και, για κάποιο $c>0$,

$$\lim_{x\to x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c.$$

Είναι εύκολο να δειχθεί ότι αν δύο σ.κ. έχουν ισοδύναμη ουρά τότε ανήκουν στην ίδια περιοχή έλξης μιας κατανομής ακροτάτων.

Περιοχές έλξης των τριών οριακών κατανομών

Για να βρούμε ποιες κατανομές ανήκουν στην περιοχή έλξης της κάθε μιας οριακής κατανομής θα πρέπει να βρεθούν οι σ.κ που έχουν ισοδύναμη ουρά με αυτές. Στον παρακάτω πίνακα φαίνονται κάποιες από τις κατανομές που ανήκουν στις περιοχές έλξης της κάθε μιας οριακής κατανομής καθώς και οι σταθερές κανονικοποίησης για τις Gumbel, Frechet και Weibull.

Πίνακας 3.1.1: Κατανομές που ανήκουν στις περιοχές έλξης των οριακών κατανομών

Κατανομή ακροτάτων	Weibull	Gumbel	Frechet
Σταθερές κανονικοποίησης	$a_n = x_F - F^{\leftarrow}(1-n^{-}$ $b_n = x_F$	$a_n = a(b_n)$ $b_n = F^{\leftarrow}(1-n^{-1})$	$a_n = F^{\leftarrow}(1-n$ $b_n = 0$
Παραδείγματα	Uniform, Beta	Exponential,Weibull ,Gamma,Normal, Lognormal	Caushy,Pareto, Loggamma, Burr

Πηγή:*Μπούτσικας,2008*

3.2 Η μέθοδος Block Maxima

Αφού στο πρώτο μέρος του κεφαλαίου εξετάστηκε περιληπτικά το πιθανοθεωρητικό μέρος που διέπει την θεωρία ακραίων τιμών στη συνέχεια θα εξεταστεί η δημιουργία μοντέλων εκτίμησης με βάση τα ιστορικά δεδομένα. Η πιο διαδεδομένη μέθοδος εκτίμησης και πρόβλεψης των ακραίων συμβάντων ιδιαίτερα στην περίπτωση δεν που έχουμε πλήρη δεδομένα είναι η μέθοδος Block Maxima. Μερικά παραδείγματα προβλημάτων που χρησιμοποιείται η συγκεκριμένη μέθοδος για την πρόβλεψη ακραίων παρατηρήσεων μπορεί να είναι τα εξής: Ποιο είναι το ποσό που δεν θα υπερβεί καμιά αποζημίωση το επόμενο έτος με πιθανότητα 99%, ή ποια η πιθανότητα το ύψος βροχής σε μια περιοχή να ξεπεράσει ένα προκαθορισμένο υψηλό κατώφλι στη διάρκεια του επόμενου έτους. Ένα σημαντικό θέμα που χρειάζεται να δώσουμε προσοχή είναι το μέγεθος του διαστήματος που θα χωριστούν τα δεδομένα. Ο λόγος είναι πως διαστήματα με λίγες παρατηρήσεις οδηγούν σε σφάλμα στην εκτίμηση ενώ μεγάλο εύρος παρατηρήσεων οδηγεί σε μεγαλύτερη διακύμανση στην εκτίμηση.

3.2.1 Γενικευμένη κατανομή ακραίων τιμών (GEV Distribution)

Η διαδικασία που ακολουθείται με χρήση της μεθόδου Block Maxima είναι η εξής:

Έστω $X_1, X_1, \ldots, X_m$ είναι ανεξάρτητες τ.μ. που προέρχονται από κάποια άγνωστη κατανομή. Θεωρούμε πως $m = kn$ δηλαδή χωρίζουμε τα δεδομένα σε k υποσύνολα μεγέθους n το καθένα και συμβολίζουμε $Y_1, Y_2, \ldots, Y_k$ τις μέγιστες τιμές σε κάθε υποσύνολο οι οποίες ονομάζονται μέγιστα ομάδων (Block Maxima). Στη συνέχεια προσπαθούμε να ανακαλύψουμε μέσω των περιοχών έλξης των οριακών κατανομών ποια είναι η κατανομή που ακολουθούν τα τοπικά μέγιστα. Εφόσον βρεθεί η οριακή κατανομή των μεγίστων το μόνο που απομένει είναι η εκτίμηση των παραμέτρων της κατανομής. Η παραπάνω διαδικασία ακολουθείται όταν υπάρχει βεβαιότητα ως προς την οριακή κατανομή που ακολουθούν τα μέγιστα. Διαφορετικά χρειάζεται μια μέθοδος για την επιλογή του κατάλληλου τύπου της οριακής κατανομής αφού σε αυτήν την περίπτωση, εσφαλμένη επιλογή της κατανομής οδηγεί σε παραπλανητικά συμπεράσματα. Τα λανθασμένα συμπεράσματα αυτά προέρχονται κυρίως από το γεγονός πως οι τρείς τύποι των οριακών κατανομών που περιγράφηκαν στο θεώρημα 3.1.1 παρουσιάζουν σημαντικές διαφορές, κυρίως λόγω διαφορετικής συμπεριφοράς των δεξιών ουρών τους. Συγκεκριμένα από τη συνάρτηση κατανομής φαίνεται καθαρά πως η ανώτερη τιμή που μπορεί να πάρει η μεταβλητή x στην περίπτωση της κατανομής Weibull είναι συγκεκριμένη και πρέπει να είναι μικρότερη ή ίση με τη τιμή της μεταβλητής b. Στις κατανομές Gumbell και Frechet η τιμή της μεταβλητής x δεν έχει κάποιον περιορισμό και μπορεί να πάρει οποιαδήποτε τιμή μέχρι το $+\infty$. Ένα άλλο παράδειγμα που δείχνει τις διαφορές των προαναφερθέντων οριακών κατανομών είναι ότι ο εκθέτης στην πυκνότητα της Gumbel είναι εκθετικής μορφής ενώ της Frechet πολυωνυμικής γεγονός που οφείλεται στο ρυθμό με τον οποίο φθίνουν οι δύο κατανομές.

Μια πιο σωστή προσέγγιση προσφέρεται από μια αναδιατύπωση των μοντέλων που αναφέρθηκαν στο θεώρημα των Fisher και Tippett. Αποδεικνύεται πως οι οικογένειες κατανομών των Gumbell, Frechet και Weibull μπορούν να συνδυαστούν σε μια οικογένεια κατανομών που η συνάρτηση κατανομής της έχει την ακόλουθη μορφή:

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

όπου $-\infty < \mu < \infty$, $\sigma > 0$ και $-\infty < \xi < \infty$. Αυτή είναι η **γενικευμένη κατανομή ακραίων**

τιμών (GEV). Το παραπάνω μοντέλο περιέχει τρείς παραμέτρους: μια παράμετρο θέσης μ, μια παράμετρο κλίμακας σ και μια παράμετρο σχήματος ξ. Οι κατανομές Frechet και Weibull αντιστοιχούν στις περιπτώσεις ξ>0 και ξ<0 αντίστοιχα και η περίπτωση ξ=0 της γενικευμένης κατανομής ακραίων τιμών αντιστοιχεί στην κατανομή Gumbel (*Coles, 2001*). Τελικά φαίνεται πως η ενοποίηση των τριών οριακών κατανομών σε μια γενικευμένη διευκολύνει την επιλογή της καταλληλότερης κατανομής αφού τα δεδομένα πια καθορίζουν ποια θα είναι αυτή αφήνοντας και ένα περιθώριο αβεβαιότητας το οποίο περιγράφεται από την παράμετρο ξ. Αναλύοντας το θεώρημα 3.1.1 για μεγάλο αριθμό n προυπόθεση αποτελεί η χρήση του τύπου των γενικευμένων κατανομών ακραίων τιμών για τη μοντελοποίηση του μεγίστου μεγάλων ακολουθιών.

$$\Pr\left\{ \left. (M_n - b_n) \right/ a_n \le x \right\} \approx G(x)$$

για μεγάλο n. Ισοδύναμα

$$\Pr\{M_n \le x\} \approx G\left\{ \left. (x - b_n) \right/ a_n \right\}$$
$$= G^*(x),$$

όπου η G^* αποτελεί ένα άλλο μέλος της GEV οικογένειας κατανομών. Δηλαδή αν η κατανομή M_n^* αποτελεί ένα μέλος της GEV για μεγάλο n τότε και η κατανομή της M_n περιγράφεται από ένα άλλο μέλος της GEV. Πρέπει να σημειωθεί όμως πως οι παράμετροι της G όπως είναι φυσικό θα διαφέρουν από τις παραμέτρους της G^*. Η φυσιολογική αυτή παρατήρηση διευκόλυνε την δημιουργία μιας νέας προσέγγισης η οποία ομαδοποιεί τα δεδομένα σε ακολουθίες παρατηρήσεων μεγέθους n (όπου το n για παράδειγμα μπορεί να είναι οι παρατηρήσεις ενός έτους). Στη συνέχεια δημιουργεί μια ακολουθία m μεγίστων στα οποία χρησιμοποιείται η γενικευμένη κατανομή ακραίων τιμών δηλαδή $M_{n,1},\ldots M_{n,m}$ όπου προσαρμόζεται η GEV.

3.2.2 Μεγιστο-ευσταθείς κατανομές (*Max-stable*)

Max-stable κατανομές είναι αυτές οι κατανομές για τις οποίες το κανονικοποιημένο μέγιστο M_n^* των X_i ακολουθεί και αυτό με τη σειρά του την ίδια κατανομή με αυτήν που ακολουθούν τα αρχικά δεδομένα X_i.

Ορισμός 3.2.1: Μια κατανομή G ονομάζεται max-stable κατανομή αν, για κάθε $n = 2,3,\ldots,$ υπάρχουν σταθερές $\alpha_n > 0$ και β_n τέτοιες ώστε

$$G^n\left(\alpha_n x + \beta_n\right) = G(x).$$

Αφού G^n είναι η συνάρτηση κατανομής της $M_n = \max\{X_1,\ldots,X_n\}$ όπου τα X_i είναι ανεξάρτητες μεταβλητές οι οποίες έχουν συνάρτηση κατανομής G, η ιδιότητα της μέγιστο ευστάθειας ικανοποιείται από κατανομές για τις οποίες η διαδικασία εξαγωγής δειγματικών μεγίστων καταλήγει να ακολουθεί την ίδια κατανομή με την G με μόνη διαφορά τις παραμέτρους θέσης και κλίμακας (*Coles, 2001*).

Θεώρημα 3.2.2: Μια κατανομή είναι μέγιστο-ευσταθής αν, και μόνο αν, ανήκει στην κατηγορία των γενικευμένων κατανομών ακραίων τιμών.

3.2.3 Στάθμη απόδοσης

Μια ποσότητα η οποία έχει μεγάλο ενδιαφέρον στη θεωρία ακραίων τιμών είναι το κατώφλι z_p το οποίο δεν θα υπερβεί καμία παρατήρηση για την επόμενη ορισμένη από εμάς χρονική περίοδο (block) με πιθανότητα $1-p$. Αν συμβολίσουμε με Τ το πλήθος των χρονικών περιόδων μέχρι να υπάρξει μια παρατήρηση που να ξεπερνάει αυτό το υψηλό κατώφλι τότε η τ.μ Τ ακολουθεί την γεωμετρική κατανομή με πιθανότητα επιτυχίας p κάτι που σημαίνει ότι $E(\mathrm{T}) = 1/p$. Δηλαδή θα υπάρχουν παρατηρήσεις που θα ξεπερνάνε το z_p κατα μέσο όρο κάθε $1/p$ χρονικές περιόδους. Ο παρακάτω ορισμός *(Boutsikas,2008)* προσδιορίζει ακριβώς αυτό το αποτέλεσμα.

Ορισμός 3.2.3: Το κατώφλι z_p το οποίο υπερβαίνουν τα Block Maxima κατά μέσο όρο κάθε $1/p$ χρονικές περιόδους καλείται άνω όριο απόδοσης ή στάθμη απόδοσής (return level) για $1/p$ χρονικές περιόδους απόδοσης (return period).

Η στάθμη απόδοσης υπολογίζεται ώς:

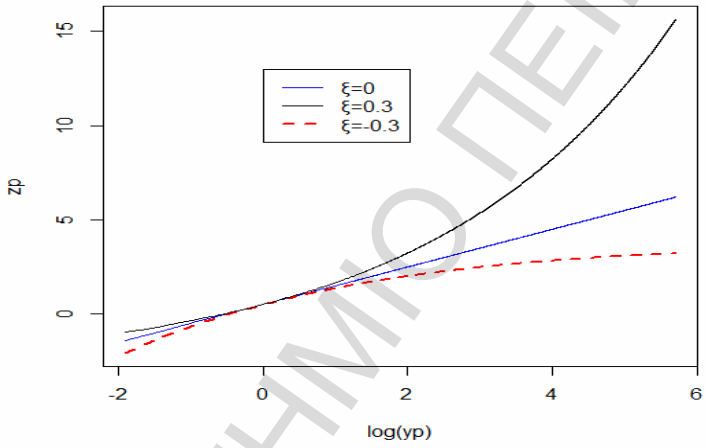
$$z_p = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\}, & \xi = 0 \end{cases}$$

Γραφικό ενδιαφέρον παρουσιάζει η στάθμη απόδοσης στην περίπτωση που $y_p = -\log(1-p)$ έτσι ώστε:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_p{}^{-\xi}\right], \xi \neq 0 \\ \mu - \sigma \log y_p, \quad \xi = 0 \end{cases}$$

Το παρακάτω γράφημα παρουσιάζει την στάθμη απόδοσής z_p ως προς την $\log(y_p)$ με παραμέτρους $\mu = 0.5, \sigma = 1$ για διάφορες τιμές του ξ.

Σχήμα 3.2.1: Γράφημα της συνάρτησης z_p ως προς την $\log(y_p)$ για ξ=0 , ξ=0,3 και ξ=-0,3



3.2.4 Εκτίμηση των παραμέτρων της GEV

Η εφαρμογή των παραπάνω αποτελεσμάτων σε πραγματικά δεδομένα προϋποθέτει την εκτίμηση των τριών παραμέτρων της κατανομής από τα Block Maxima. Οι μέθοδοι που έχουν προταθεί για την εκτίμηση των παραμέτρων είναι πολλοί και η κάθε μια έχει πλεονεκτήματα και μειονεκτήματα. Μερικές από αυτές τις μεθόδους είναι η εκτίμηση μέσω γραφημάτων πιθανότητας , μέσω εξισώσεων των δειγματικών και θεωρητικών ροπών, μέσω συναρτήσεων μεγίστης πιθανοφάνειας κλπ. Οι καλές ιδιότητες και η ευρεία υιοθέτηση τους σε πολύπλοκα μοντέλα κατέστησαν τις μεθόδους που βασίζονται σε εκτιμήσεις μεγίστης πιθανοφάνειας αρκετά διάσημες.

Το πρόβλημα που προκύπτει με τις μεθόδους μεγίστης πιθανοφάνειας για την εκτίμηση της GEV είναι ότι δεν ικανοποιούνται οι συνθήκες ομαλότητας που απαιτούνται έτσι ώστε να ισχύουν τα ασυμπτωτικά αποτελέσματα της μεθόδου. Ο λόγος που δεν ικανοποιούνται είναι

ότι το στήριγμα της GEV εξαρτάται από τις παραμέτρους μ, σ, ξ και επομένως δεν μπορούμε να βασιστούμε αυτόματα στις ασυμπτωτικές ιδιότητες των εκτιμητών μεγίστης πιθανοφάνειας (ε.μ.π.). Αν και δεν μπορούμε να βασιστούμε απόλυτα, γενικά έχουν αποδειχθεί ότι ισχύουν τα ακόλουθα συμπεράσματα :

- Για $\xi > -0.5$ οι εκτιμητές μεγίστης πιθανοφάνειας έχουν τις ασυμπτωτικές τους ιδιότητες.
- Για $-1 < \xi < -0.5$ οι εκτιμητές μεγίστης πιθανοφάνειας υπάρχουν αλλά δεν έχουν τις ασυμπτωτικές τους ιδιότητες.
- Για $\xi < -1$ οι εκτιμητές μεγίστης πιθανοφάνειας δεν μπορούν να υπολογισθούν.

3.2.5 Εκτίμηση των παραμέτρων με τη μέθοδο μεγίστης πιθανοφάνειας

Αφού έχουμε υποθέσει πως οι τιμές των μεγίστων $Y_1, Y_2, \ldots, Y_k$ είναι ανεξάρτητες μεταξύ τους ο λογάριθμος της συνάρτησης πιθανοφάνειας για $\xi \neq 0$ είναι ο ακόλουθος:

$$\ell(\mu,\sigma,\xi) = -k\log\sigma - (1+1/\xi)\sum_{i=1}^{k}\log\left[1+\xi\left(\frac{y_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{k}\left[1+\xi\left(\frac{y_i-\mu}{\sigma}\right)\right]^{-1/\xi}$$

δοθέντος ότι

$$1+\xi\left(\frac{y_i-\mu}{\sigma}\right) > 0\,, \quad \textbf{για } i = 1,\ldots,m\,.$$

Στην περίπτωση που το $\xi = 0$ η συνάρτηση πιθανοφάνειας είναι διαφορετική και ο λογάριθμος της ισούται με:

$$\ell(\mu,\sigma) = -k\log\sigma - \sum_{i=1}^{k}\left(\frac{y_i-\mu}{\sigma}\right) - \sum_{i=1}^{k}\exp\left\{-\left(\frac{y_i-\mu}{\sigma}\right)\right\}$$

Οι εκτιμητές μεγίστης πιθανοφάνειας $(\hat{\mu},\hat{\sigma},\hat{\xi})$ των μεταβλητών μ,σ,ξ δεν μπορούν να υπολογισθούν αναλυτικά γιατί το στήριγμα εξαρτάται από τις παραμέτρους της δείκτριας συνάρτησης . Ο μόνος τρόπος υπολογισμού των ε.μ.π. είναι μέσω αριθμητικών μεθόδων βελτιστοποίησης όπως η μέθοδος Newton-Raphson με ιδιαίτερη προσοχή όμως στο να μην παραβιάζονται οι υποθέσεις της συνάρτησης.

Οι εκτιμήτριες $(\hat{\mu},\hat{\sigma},\hat{\xi})$ ακολουθούν προσεγγιστικά πολυωνυμική κατανομή με μέση τιμή (μ,σ,ξ) και πίνακα διασπορά τον αντίστροφο του πίνακα πληροφορίας.

3.2.6 Εκτίμηση της στάθμης απόδοσης

Αφού έχουν υπολογιστεί οι εκτιμήτριες $(\hat{\mu},\hat{\sigma},\hat{\xi})$, η εκτίμηση της στάθμης απόδοσης προκύπτει από την αντικατάσταση των εκτιμητριών αυτών στον τύπο της στάθμης απόδοσης. Συγκεκριμένα παίρνουμε ότι

$$\hat{z}_p = \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\xi}}\left[1 - y_p^{-\hat{\xi}}\right], & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log y_p, & \hat{\xi} = 0 \end{cases}$$

όπου $y_p = -\log(1-p)$.

Η διακύμανση της μεταβλητής $\hat{z}_p$ δίνεται από την ακόλουθη σχέση με χρήση της μεθόδου δέλτα:

$$Var(\hat{z}_p) \approx \nabla z_p{}^T V \nabla z_p,$$

όπου V είναι ο πίνακας διακύμανσης – συνδιακύμανσης των $(\hat{\mu},\hat{\sigma},\hat{\xi})$ και

$$\begin{aligned} \nabla z_p{}^T &= \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi}\right] \\ &= \left[1, -\xi^{-1}(1 - y_p^{-\xi}), \sigma\xi^{-2}(1 - y_p^{-\xi}) - \sigma\xi^{-1} y_p^{-\xi}\log y_p\right]. \end{aligned}$$

Εναλλακτικά για την εκτίμηση της στάθμης απόδοσης μπορεί να χρησιμοποιηθεί η profile deviance function $D_p(z_p)$ η οποία οδηγεί σε διάστημα εμπιστοσύνης για την z_p. Η διαδικασία που ακολουθείται για την εφαρμογή της μεθόδου προϋποθέτει αρχικά μια νέα παραμετροποίηση αλλά αυτή τη φορά ως προς z_p. Δηλαδή στη τη θέση μιας παραμέτρου παίρνει η z_p π.χ. ως ακολούθως:

$$\mu = z_p + \frac{\sigma}{\xi}\left(1 - \left(-\lg(1-p)\right)^{-\xi}\right)$$

Ο λογάριθμος της πιθανοφάνειας είναι $\ell(z_p,\sigma,\xi)$ και ισχύει ότι

$$D_p(z_p) = 2\left\{\ell(\hat{z}_p, \hat{\sigma}, \hat{\xi}) - \max_{\sigma,\xi}(z_p, \sigma, \xi)\right\} \square \chi_i^2$$

Συνεπώς το σύνολο

$$\left\{z_p : D_p\left(z_p\right) \le \chi_i^2\right\}$$

είναι ένα $(1-a)$ διάστημα εμπιστοσύνης για το z_p.

Έλεγχος καλής προσαρμογής των δεδομένων στην GEV

Στην περίπτωση που το μέγεθός των blocks είναι πολύ μεγάλο $(n \to \infty)$ τότε τα block maxima ανήκουν στην GEV. Σε πραγματικά δεδομένα όμως όπου το n είναι πεπερασμένο θα πρέπει να ελεγθεί το κατά πόσο τα δεδομένα ανήκουν στην GEV έτσι ώστε στη συνέχεια να εκτιμηθούν οι παράμετροί τους. Ένας εύκολος τρόπος να ελεγχθεί η παραπάνω πρόταση είναι τα Q-Q plots. Εάν τα block maxima ακολουθούν την GEV τότε τα διατεταγμένα σημεία που απεικονίζουν τα maxima θα πρέπει να βρίσκονται κοντά στη διαγώνιο του γραφήματος.

3.3 Η μέθοδος των υπερβάσεων πάνω από ένα όριο (POT)

Στην εισαγωγή της μεθόδου block maxima αναφέραμε πως χρησιμοποιείται ιδιαίτερα στην περίπτωση που τα διαθέσιμα δεδομένα δεν είναι πλήρη αλλά είναι γνωστά ίσως μόνο τα μέγιστα κάποιων διαστημάτων. Σε περίπτωση που είναι διαθέσιμα διαφορετικών ειδών δεδομένα υπάρχουν και διαφορετικές μέθοδοι οι οποίες μπορούν να δώσουν καλύτερα αποτελέσματα. Μια μέθοδος είναι αυτή της r-μεγαλύτερης παρατήρησης η οποία μοντελοποιεί την r-μεγαλύτερη παρατήρηση και όχι τη μέγιστη. Αν και αυτή η μέθοδος δίνει καλά αποτελέσματα σπανίως υπάρχουν δεδομένα σε αυτή τη μορφή. Η ανάλυση της ενότητας που ακολουθεί, αφορά τη μέθοδο που μοντελοποιεί της υπερβάσεις πάνω από ένα ορισμένο όριο (***peak over threshold method-POT***). Η μέθοδος αυτή χρησιμοποιείται στην περίπτωση όπου τα διαθέσιμα δεδομένα είναι αναλυτικά (π.χ. ημερήσια στοιχεία).

Έστω $X_1, X_2, \ldots$ μια ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών που ακολουθούν μια άγνωστη συνάρτηση κατανομής F. Θεωρούνται ακραίες οι παρατηρήσεις οι οποίες υπερβαίνουν κάποιο υψηλό κατώφλι u. Μια πρώτη ποσότητα που θα πρέπει να εκτιμηθεί σε αυτό το σημείο είναι η κατανομή της υπέρβασης μιας τυχαίας μεταβλητής X

πάνω από αυτό το κατώφλι u δοθέντος ότι η X το έχει υπερβεί.

$$F_u(y) = \Pr(X - u \le y \setminus X > u) = 1 - \frac{1 - F(u+y)}{1 - F(u)}, \quad y > 0$$

Στον παραπάνω τύπο για να βρεθεί η F_u πρέπει να είναι γνωστή η F. Επειδή σε πρακτικές εφαρμογές αυτό δεν είναι πάντα εφικτό, θα πρέπει να αναζητηθεί αν υπάρχει οριακή κατανομή της F_u και ακόμα αν αυτή είναι ανεξάρτητη από την αρχική κατανομή των δεδομένων.

3.3.1 Η Γενικευμένη κατανομή Pareto

Η γενικευμένη κατανομή Pareto αναπτύχθηκε ως μια κατανομή η οποία μπορεί να μοντελοποιήσει τις ουρές ενός μεγάλου πλήθους κατανομών. Όπως έχει αναφερθεί και στην αρχή του κεφαλαίου (*Coles, 2001*) ότι αν $X_1, X_2, \ldots X_n$ είναι μια ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με συνάρτηση κατανομής F τότε:

$$M_n = \max(X_1, X_2, \ldots, X_n)$$

και για μεγάλο n

$$\Pr(M_n \le x) \approx G(x),$$

όπου

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

Τότε για μεγάλο u η συνάρτηση κατανομής της $(X-u)$ δοθέντος ότι $X > u$ είναι η :

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

όπου ορίζεται ως:

$$\left\{y : y > 0 \; \kappa\alpha\iota \; (1 + \xi y/\tilde{\sigma}) > 0\right\},$$

όπου

$$\tilde{\sigma} = \sigma + \xi(u - \mu).$$

Η κατανομή $H(\cdot, \tilde{\sigma}, \xi)$ ονομάζεται **γενικευμένη κατανομή Pareto (*GPD*)**. Από τον τύπο της GPD βλέπουμε ότι οι παράμετροι της εξαρτώνται από τις παραμέτρους της GEV. Η

παράμετρος ξ είναι η ίδια και στις δύο κατανομές και ανάλογα με το μέγεθος της προσδίδει και διαφορετικά χαρακτηριστικά στην GPD. Συγκεκριμένα αν:

- $\xi < 0$ η GPD έχει άνω όριο το $u - \tilde{\sigma}/\xi$.
- $\xi > 0$ η GPD δεν έχει άνω όριο και επεκτείνεται μέχρι το άπειρο.
- $\xi = 0$ η GPD εκφυλίζεται και καταλήγει στην εκθετική κατανομή με παράμετρο $1/\tilde{\sigma}$.

Γίνεται σαφές ότι για κατανομές ασφαλιστικών αποζημιώσεων, οι οποίες μάλιστα συνήθως έχουν βαριά ουρά, η περίπτωση που μας απασχολεί είναι για $\xi > 0$.

3.3.2 Η επιλογή του ανώτατου ορίου u

Προτού εκτιμηθούν οι παράμετροι της GPD θα πρέπει να επιλεγεί το κατάλληλο ανώτερο όριο u οι υπερβάσεις του οποίου θα αποτελέσουν την ενδεδειγμένη GPD. Η σωστή επιλογή αυτού του ορίου είναι εξέχουσας σημασίας για τη σωστή προσέγγιση της κατανομής που θα ακολουθούν τα δεδομένα. Συγκεκριμένα εάν το u επιλεγεί αρκετά μικρό τότε η F_u ενδέχεται να μην προσεγγίζεται ικανοποιητικά από την GPD. Σε αντίθετη περίπτωση αν το u είναι αρκετά μεγάλο τότε ενδέχεται να μην υπάρξουν αρκετές υπερβάσεις πάνω από αυτό με αποτέλεσμα να μην υπάρχουν δεδομένα για ασφαλή εξαγωγή συμπερασμάτων. Για την επιλογή του βέλτιστου δυνατού κατωφλίου υπάρχουν δύο μέθοδοι οι οποίες θα παρουσιαστούν στη συνέχεια.

- **1η μέθοδος**

Σε αυτή τη μέθοδο χρησιμοποιείται η μέση τιμή της GPD και στηρίζεται στο γεγονός πως αν ένα κατώφλι u_0 είναι αρκετά μεγάλο ώστε $F_{u_0} \approx GPD$ τότε και για οποιοδήποτε $u > u_0$, θα ισχύει ότι $F_u \approx GPD$.

Πιο συγκεκριμένα, έχουμε:

$$e(u) = E(X - u \setminus X > u) \approx \frac{\tilde{\sigma}}{1+\xi} = \frac{\sigma + \xi(u-\mu)}{1+\xi}, \; u > u_0$$

Η δεσμευμένη αυτή μέση τιμή είναι γραμμική συνάρτηση του u και ονομάζεται μέση υπερβάλουσα απώλεια (mean excess loss) ή μέση υπολειπόμενη ζωή(mean residual life). Η

$e(u)$ μπορεί να εκτιμηθεί από τον αριθμό των υπερβάσεων των X_i πάνω από το u η οποία είναι η εμπειρική μέση υπερβάλλουσα απώλεια.

$$\hat{e}(u) = \frac{1}{k(u)} \sum_{i=X_i>u} (X_i - u)$$

όπου $k(u)$ είναι το πλήθος των X_i που υπερβαίνουν το όριο u. Πάνω από το όριο u_0 το γράφημα της μέσης υπολειπόμενης ζωής θα πρέπει να είναι γραμμικό για $u > u_0$. Τελικά δηλαδή η επιλογή του ορίου u θα γίνει με τη βοήθεια του γραφήματος και θα ισούται με το σημείο του γραφήματος πάνω από το οποίο η γραφική παράσταση θα είναι περίπου γραμμική. Θα πρέπει να σημειωθεί πως η συγκεκριμένη μέθοδος επιλογής του ορίου δεν είναι αξιόπιστη στη περίπτωση όπου δεν υπάρχουν αρκετές υπερβάσεις (Boutsikas 2008).

- **2η μέθοδος**

  Και η δεύτερη μέθοδος επιλογής του βέλτιστου κατωφλίου εκτιμάται γραφικά. Συγκεκριμένα, εκτιμούνται αρχικά οι παράμετροι $\mu, \tilde{\sigma}$ της GPD για διάφορες τιμές του u. Αφού η $F_u \approx GPD$ η εκτίμηση του ξ δεν πρέπει να επηρεάζεται από το u και το ενδεδειγμένο u θα είναι το σημείο πάνω από το οποίο η $\tilde{\sigma} = \sigma + \xi(u - \mu)$ μεταβάλλεται γραμμικά ως πρός u.

3.3.3 Εκτίμηση των παραμέτρων της γενικευμένης κατανομής Pareto

Αφού έχει επιλεχθεί το κατάλληλο κατώφλι u, το επόμενο βήμα είναι η εκτίμηση των παραμέτρων της γενικευμένης κατανομής Pareto. Η συνηθέστερη μέθοδος εκτίμησης είναι και σε αυτή τη περίπτωση η μέθοδος μεγίστης πιθανοφάνειας. Ας συμβολίσουμε με $y_1, y_2, \ldots, y_k$ τις k υπερβάσεις των X_i πάνω από το όριο u. Τότε ο λογάριθμος της συνάρτησης πιθανοφάνειας θα ισούται με :

$$\ell(\tilde{\sigma}, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^{k} \log(1 + \xi y_i / \tilde{\sigma}), \ \xi \neq 0$$

όπου $(1 + \tilde{\sigma}^{-1} \xi y_i) > 0$ για $i = 1, \ldots, k$

ενώ για $\xi = 0$ θα ισούται με

$$\ell(\tilde{\sigma}) = -k \log \tilde{\sigma} - \tilde{\sigma}^{-1} \sum_{i=1}^{k} y_i \bullet$$

Θα πρέπει να σημειωθεί ότι, όπως και στην block maxima, έτσι και σε αυτή τη μέθοδο δεν υπάρχει αναλυτική λύση για τη μεγιστοποίηση της παραπάνω συνάρτησης και για αυτό το λόγο θα χρησιμοποιηθούν αριθμητικές μέθοδοι για τον υπολογισμό της.

3.3.4 Εκτίμηση της στάθμης απόδοσης

Συχνά είναι χρήσιμο να εκτιμήσουμε τη στάθμη απόδοσης η οποία θα δώσει την πιθανότητα να ξεπεράσει μια παρατήρηση το κατώφλι x_m κάθε $1/m$ χρονικές περιόδους. Ας υποθέσουμε πως η γενικευμένη κατανομή Pareto με παραμέτρους $\tilde{\sigma}, \xi$ είναι το κατάλληλο μοντέλο για τις υπερβάσεις της μεταβλητής X πάνω από το όριο u. Τότε για $x > u$ θα ισχύει ότι :

$$\Pr(X > x \setminus X > u) \approx H(x-u) = \left(1 + \xi \frac{x-u}{\tilde{\sigma}}\right)^{-1/\xi}$$

και επομένως

$$\Pr(X > x) \approx \Pr(X > u)\left(1 + \xi \frac{x-u}{\tilde{\sigma}}\right)^{-1/\xi}.$$

Το κατώφλι x_m που θα ξεπερνάει μια παρατήρηση κάθε $1/m$ παρατηρήσεις σημαίνει ότι $\Pr(X > x_m) = 1/m$. Θέτοντας όπου x το x_m και λύνοντας ως προς αυτό παίρνουμε ότι

$$x_m \approx u + \frac{\tilde{\sigma}}{\xi}\left(\left(m \Pr(X > u)\right)^{\xi} - 1\right).$$

Η εκτίμηση της στάθμης απόδοσης για m παρατηρήσεις θα είναι

$$\hat{x} \approx u + \frac{\hat{\tilde{\sigma}}}{\hat{\xi}}\left(\left(\frac{mk(u)}{n}\right)^{\hat{\xi}} - 1\right),$$

όπου $\hat{\tilde{\sigma}}, \hat{\xi}$ είναι οι ε.μ.π. των παραμέτρων και η ποσότητα $k(u)/n$ είναι το ποσοστό των υπερβάσεων πάνω από το u που αποτελεί τον ε.μ.π. της $\Pr(X > u)$. Είναι δυνατόν μέσω των

ασυμπτωτικών ιδιοτήτων των κατανομών να κατασκευαστούν διαστήματα εμπιστοσύνης για το x_m με τη μέθοδο Δέλτα ή εναλλακτικά με τη χρησιμοποίηση της profile deviance function.

3.3.5 Έλεγχος καλής προσαρμογής των δεδομένων στην GPD

Όπως και στην προσαρμογή των δεδομένων στην GEV έτσι και για την καλή προσαρμογή των δεδομένων στην GPD θα χρησιμοποιηθούν τα P-P Plots και τα Q-Q Plots. Ο λόγος που πρέπει να γίνει αυτό το βήμα είναι για να βεβαιωθούμε πως τα σημεία που προκύπτουν με τη μέθοδο POT ακολουθούν κάποια GPD έτσι ώστε στη συνέχεια να εκτιμηθούν με ασφάλεια οι παρεάμετροι της GPD. Συγκεκριμένα για να ελεγθεί αν τα δεδομένα προέρχονται από μια GPD κατανομή με τη βοήθεια των Q-Q Plots θα πρέπει τα k διατεταγμένα POT να βρίσκονται κοντά στην κύρια διαγώνιο του γραφήματος.

Κεφάλαιο 4

Εφαρμογή σε χαρτοφυλάκιο ασφάλισης αυτοκινήτων

4.1 Εισαγωγή

Στο παρόν κεφάλαιο θα εφαρμοστούν σε ένα χαρτοφυλάκιο ασφάλισης αυτοκινήτων οι μέθοδοι εκτίμησης των κατανομών οι οποίες παρουσιάστηκαν στο δεύτερο κεφάλαιο της παρούσας εργασίας. Τα δεδομένα που θα χρησιμοποιηθούν είναι στοιχεία πραγματικών ζημιών από γνωστή ασφαλιστική εταιρία. Τα διαθέσιμα στοιχεία αφορούν το μέγεθος της ζημιάς ανά ατύχημα καθώς και την ακριβή ημερομηνία αναγγελίας της ζημιάς για τα έτη 2006-07-08. Τα προγράμματα που πρόκειται να χρησιμοποιηθούν για την στατιστική επεξεργασία των δεδομένων είναι κατά κύριο λόγο η R (γλώσσα προγραμματισμού S) καθώς και το στατιστικό πακέτο SPSS και το excel.

Στην μελέτη που θα ακολουθήσει γίνεται προσπάθεια να διερευνηθούν τα δεδομένα και να προκύψουν χρήσιμα συμπεράσματα για τις κατανομές που ακολουθούν κάποιες σημαντικές ποσότητες όπως αυτή του μεγέθους και του αριθμού των αποζημιώσεων.

Πιο συγκεκριμένα θα γίνει προσπάθεια να παρουσιαστούν τα δεδομένα και να περιγραφεί με απλό τρόπο αρχικά η ποσότητα που περιγράφει τον αριθμό των απαιτήσεων οι οποίες καταφθάνουν στην ασφαλιστική εταιρία και να διερευνηθεί ποια κατανομή προσεγγίζει καλύτερα αυτή την ποσότητα. Στη συνέχεια το επόμενο μέρος θα αφιερωθεί στο μέγεθος των απαιτήσεων και θα εξεταστεί αν ακολουθεί κάποια από τις γνωστές κατανομές.

Τέλος θα αναλυθούν τα αποτελέσματα που προέκυψαν από την ανάλυση των δεδομένων για την εξαγωγή συμπερασμάτων.

4.2 Μελέτη της κατανομής του αριθμού των αποζημιώσεων

Μια ποσότητα με ιδιαίτερο ενδιαφέρον στην αναλογιστική επιστήμη είναι ο αριθμός των αποζημιώσεων που καταφθάνουν στην ασφαλιστική εταιρία. Σκοπός της μελέτης που θα ακολουθήσει είναι το κατά πόσο ο αριθμός αυτός μπορεί να προσεγγιστεί από μια πιθανοθεωρητική κατανομή.

4.2.1 Περιγραφή των δεδομένων

Τα δεδομένα τα οποία είναι διαθέσιμα για την ανάλυση περιλαμβάνουν τις μη μηδενικές αποζημιώσεις για την περίοδο από 1/1/2006 έως τις 31/12/2008. Τα συγκεκριμένα στοιχεία για να μπορέσουν να μελετηθούν απαραίτητη προϋπόθεση είναι να ομαδοποιηθούν με ξεκάθαρο και βολικό τρόπο και εν συνεχεία να μελετηθεί αν το πλήθος της κάθε ομάδας μπορεί να θεωρηθεί ως μια τυχαία μεταβλητή που προέρχεται από κάποια γνωστή κατανομή. Η ομαδοποίηση που χρησιμοποιήθηκε έγινε με βάση κάποιο χρονικό προσδιορισμό και συγκεκριμένα επιλέχθηκε ο ένας ημερολογιακός μήνας.

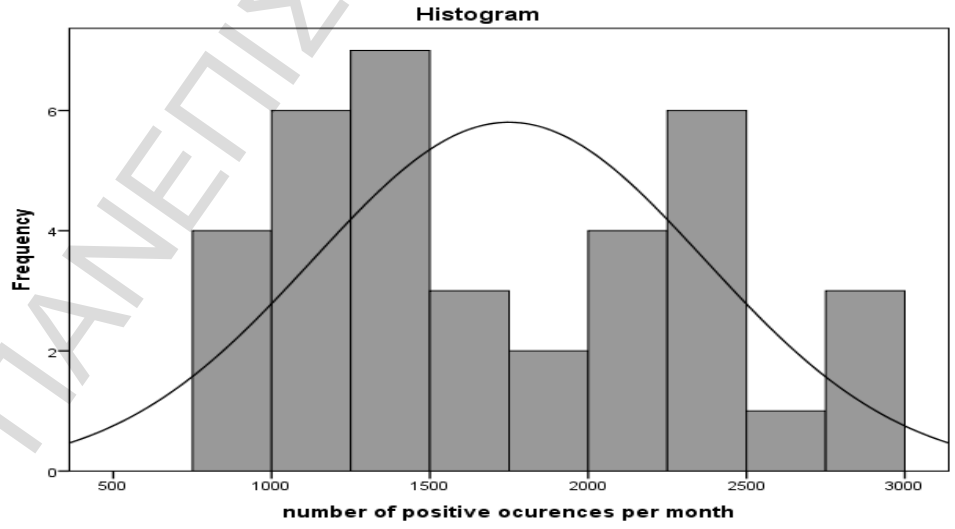
4.2.2 Στατιστική ανάλυση δεδομένων

Για να κατανοηθεί το είδος των δεδομένων καλύτερα, θα πρέπει αρχικά να παρουσιαστούν μερικά περιγραφικά στοιχεία τα οποία και θα υποδείξουν τον τρόπο με τον οποίο θα συνεχιστεί η ανάλυση.

I. Ανάλυση του συνόλου των παρατηρήσεων

Αρχικά θα ήταν χρήσιμο ένα ιστόγραμμα συχνοτήτων για να φανεί γραφικά αν η κατανομή του πλήθους των αποζημιώσεων για τα ομαδοποιημένα ανά μήνα δεδομένα θυμίζει ολόκληρη ή, τουλάχιστον σε κάποιο μέρος της, κάποια γνωστή κατανομή.

Σχήμα 4.2.1: Ιστόγραμμα του πλήθους των θετικών αποζημιώσεων ανά μήνα



Στο παραπάνω ιστόγραμμα παρουσιάζεται η συχνότητα του αριθμού των θετικών αποζημιώσεων ανά μήνα κατά την τριετία για την οποία εξετάζεται το δείγμα. Αν και το

δείγμα των 36 παρατηρήσεων είναι αρκετά μικρό και στην πραγματικότητα δεν μπορούν να εξαχθούν ασφαλή συμπεράσματα, παρ΄ όλα αυτά η κατανομή των δεδομένων δεν φαίνεται να ακολουθεί κάποια γνωστή κατανομή και εκ πρώτης όψεως θα μπορούσαμε να κατευθυνθούμε σε μια περίπτωση δικόρυφης κατανομής.

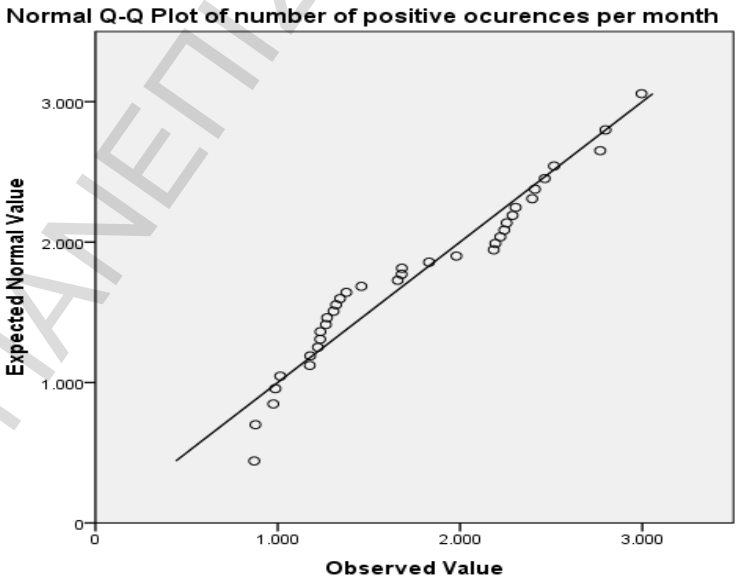
Πίνακας 4.2.1: Πίνακας περιγραφικών στοιχείων των θετικών αποζημιώσεων ανά μήνα

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
# of positive ocurrences per month	36	871	2996	1749,36	618,508	382552,3

Στον παραπάνω πίνακα διακρίνονται μερικά από τα πιο χαρακτηριστικά περιγραφικά στοιχεία της κατανομής των δεδομένων. Παρατηρούμε πως η μέση τιμή των δεδομένων ισούται με 1749,36 ενώ η διακύμανση τους με 382552,3 αντίστοιχα.

Αν και εκ πρώτης όψεως τα δεδομένα δεν φαίνεται να ακολουθούν την κανονική κατανομή υπάρχουν κάποια στατιστικά test τα οποία μπορούν να δείξουν το κατά πόσο η παραπάνω γραφική αίσθηση που δόθηκε από το ιστόγραμμα είναι αληθής ή όχι. Συγκεκριμένα στο ακόλουθο σχήμα παρουσιάζεται το Q-Q Plot για την κανονική κατανομή.

Σχήμα 4.2.2: Q-Q plot του πλήθους των θετικών αποζημιώσεων ανά μήνα για κανονική κατανομή



Το Q-Q Plot δεν επιδεικνύει πολύ καλή προσαρμογή των δεδομένων στη κανονική κατανομή παρ΄όλα αυτά υπάρχουν και στατιστικά test που δεν στηρίζονται σε γραφικές μεθόδους όπως το Kolmogorov-Smirnov test που παρουσιάζεται στον παρακάτω πίνακα.

Πίνακας 4.2.2: Kolmogorov Smirnov test για το πλήθος των θετικών αποζημιώσεων ανά μήνα

One-Sample Kolmogorov-Smirnov Test		
number of positive ocurences per month		
N		36
Normal Parameters	Mean	1749,36
	Std. Deviation	618,508
Most Extreme Differences	Absolute	,171
	Positive	,171
	Negative	-,148
Kolmogorov-Smirnov Z		1,028
Asymp. Sig. (2-tailed)		,241

Τα αποτελέσματα που προκύπτουν από το Kolmogorov-Smirnov test διαψεύδουν τις αρχικές προβλέψεις για την κατανομή που ακολουθούν τα δεδομένα αφού το p-value $\approx 0{,}241 > 0{,}05$ που σημαίνει πώς δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%. Οι εκτιμήσεις των παραμέτρων της κανονικής κατανομής που δόθηκαν από τα δεδομένα με την υιοθέτηση της μηδενικής υπόθεσης είναι $\mu = 1749,36$ και $\sigma^2 = 382.552,14$.

Όπως έχει αναφερθεί και παραπάνω αφού οι παρατηρήσεις είναι μόλις 36 τα αποτελέσματα των διαφόρων ελέγχων δεν μπορούν να οδηγήσουν σε ασφαλή συμπεράσματα και για αυτό το λόγο μια πιο ενδελεχής έρευνα χρειάζεται στη συγκεκριμένη περίπτωση.

II. Ανάλυση των δεδομένων ανά κατηγορίες

Παρατηρώντας το ιστόγραμμα συχνοτήτων του συνόλου των δεδομένων η αίσθηση που παίρνουμε είναι ότι το σχήμα του ιστογράμματος δεν αντικατοπτρίζει κάποια γνωστή κατανομή και ότι ίσως θα πρέπει να αναζητηθεί μια κατανομή η οποία να παρουσιάζει δύο

κορυφές. Ένας τρόπος με τον οποίο μπορεί να κατασκευασθούν τέτοιου είδους κατανομές είναι με μίξη δύο γνωστών κατανομών (βλ. Klungman et al,1998). Έστω τ.μ. X_i

έτσι ώστε $f_i(x) = P(X_i = x)$ για $i = 1,\ldots,n$ και επίσης σταθερά ω_i τέτοια ώστε

$0 < \omega_i < 1$ και $\sum_{i=1}^{n} \omega_i = 1$. Τότε η συνάρτηση πυκνότητας πιθανότητας της μικτής κατανομής θα

είναι

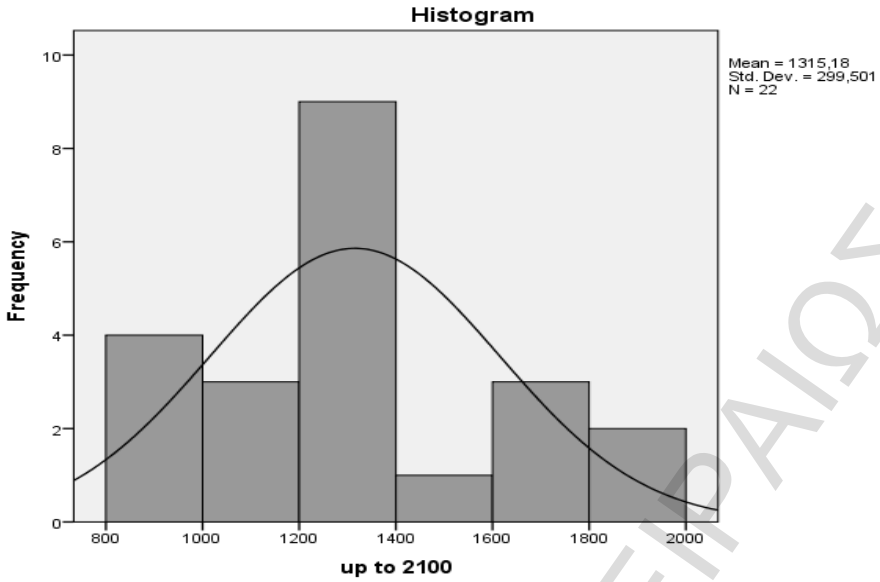
$$f(x) = \sum_{i=1}^{n} \omega_i f_i(x).$$

Η συνέχεια λοιπόν της ανάλυσης προϋποθέτει των διαχωρισμό της αρχικής κατανομής των δεδομένων σε δύο επιμέρους κατανομές. Ύστερα από διαδοχικές δοκιμές το σημείο στο οποίο θα επιλέξουμε να χωριστεί η αρχική κατανομή των δεδομένων είναι το 2.100, σημείο το οποίο φαίνεται να φθίνει η πρώτη κατανομή και να ξεκινάει η δεύτερη.

- Ανάλυση για τις παρατηρήσεις μέχρι το σημείο 2.100

Η παρακάτω ανάλυση περιλαμβάνει όλες τις παρατηρήσεις που δεν ξεπερνάνε την τιμή των 2.100 εμφανίσεων ζημιών ανά μήνα. Σαν πρώτο βήμα θα παραστήσουμε γραφικά τις παρατηρήσεις αυτές με ένα ιστόγραμμα για να ανακαλύψουμε αν Σχήματικά θυμίζει κάποια κατανομή.

Σχήμα 4.2.3: Ιστόγραμμα του πλήθους των αποζημιώσεων ανά μήνα εως την τιμή 2.100



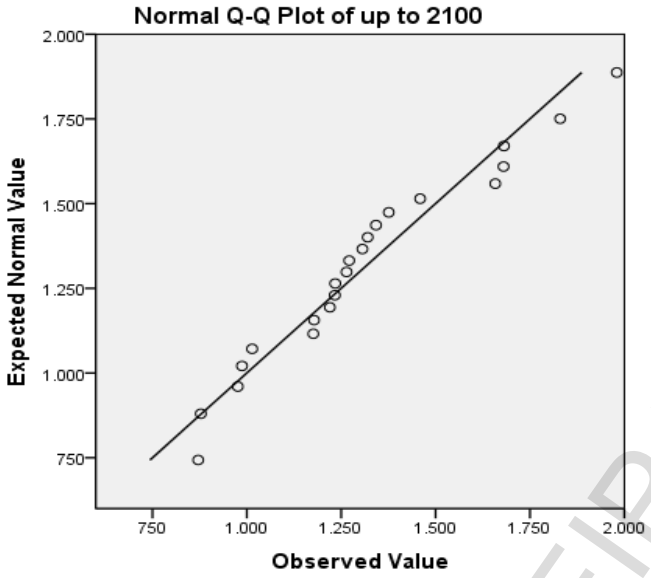
Από το ιστόγραμμα συχνοτήτων παρατηρούμε πως η κατανομή που ακολουθούν τα δεδομένα θα μπορούσε να είναι η κανονική κατανομή. Τα περιγραφικά στοιχεία για τις συγκεκριμένες παρατηρήσεις παρουσιάζονται στον παρακάτω πίνακα

Πίνακας 4.2.3: Πίνακας περιγραφικών στοιχείων του πλήθους των αποζημιώσεων ανά μήνα εως την τιμή 2.100

Statistics	
	up to 2.100
N	22
Mean	1315,18
Median	1267,50
Variance	89700,918
Skewness	,567
Std. Error of Skewness	,491
Minimum	871
Maximum	1980

Ένας άλλος γραφικός τρόπος πέρα από το ιστόγραμμα ο οποίος μπορεί να δώσει ενδείξεις για την καταλληλότητα της επιλογής της κανονικής κατανομής είναι η κατασκευή ενός Q-Q Plot για την κανονική κατανομή.

Σχήμα 4.2.4: Q-Q plot για την κανονική κατανομή του πλήθους των αποζημιώσεων ανά μήνα εως την τιμή 2.100



Από το παραπάνω σχήμα βλέπουμε πως οι παρατηρήσεις ακολουθούν την γραμμή της κανονικής κατανομής αλλά δεν μπορούμε να αποφασίσουμε για την ορθότητα αυτής της επιλογής. Ο πίνακας που ακολουθεί περιλαμβάνει τα αποτελέσματα του Kolmogorov-Smirnov test σύμφωνα με τον οποίο θα επιβεβαιωθεί η όχι την υπόθεση της κανονικότητας των παρατηρήσεων.

Πίνακας 4.2.4: Kolmogorov Smirnov test του πλήθους των αποζημιώσεων ανά μήνα εως την τιμή 2.100

One-Sample Kolmogorov-Smirnov Test		
		up to 2.100
N		22
Normal Parameters	Mean	1315,18
	Std. Deviation	299,501
Most Extreme Differences	Absolute	,147
	Positive	,147
	Negative	-,101
Kolmogorov-Smirnov Z		,689
Asymp. Sig. (2-tailed)		,730

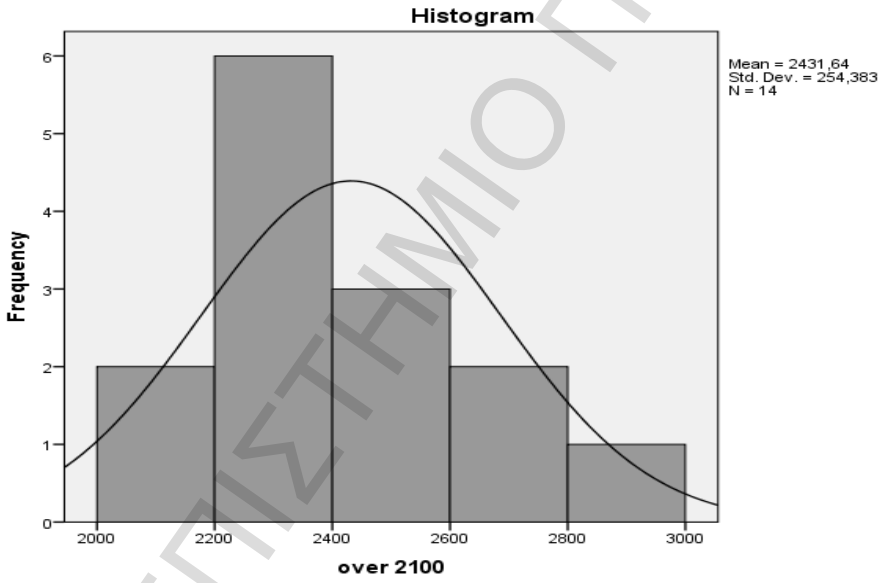
Παρατηρώντας το p-value του ελέγχου το οποίο ανέρχεται στο 73% δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση με αποτέλεσμα να δίνεται η δυνατότητα να θεωρήσουμε

πως η κατανομή που ακολουθούν οι μη μηδενικές αποζημιώσεις ανά μήνα έως του αριθμού των 2.100 ακολουθούν κανονική κατανομή με μέση τιμή 1315,18 και τυπική απόκλιση 299,501.

- Ανάλυση για τις παρατηρήσεις πάνω από την τιμή 2.100

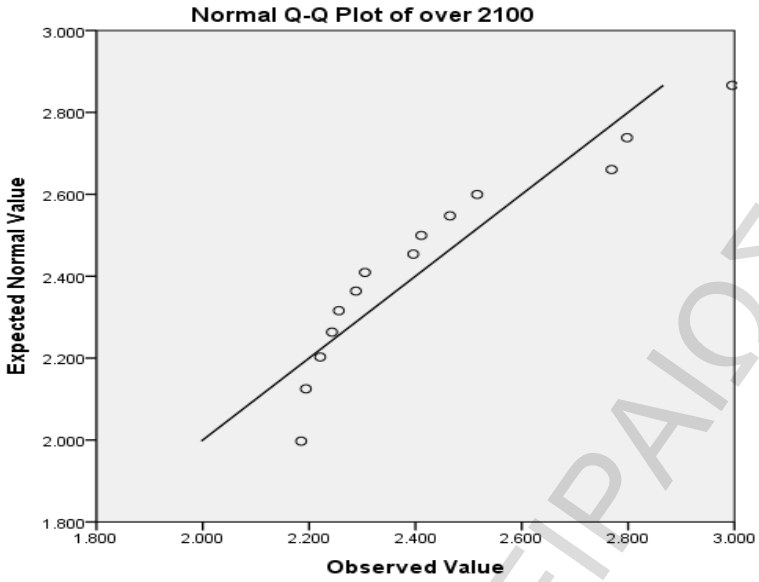
Στη ανάλυση που ακολουθεί και αφορά τις παρατηρήσεις πάνω από την τιμή 2.100 θα πρέπει να τονιστεί ότι ορισμένα αποτελέσματα που πρόκειται να βγουν ενδέχεται να μην είναι απολύτως σωστά αφού το δείγμα είναι εξαιρετικά μικρό (μόλις 14 παρατηρήσεις). Αρχικά θα κατασκευαστεί το ιστόγραμμα συχνοτήτων για να φανεί το σχήμα της κατανομής.
\

Σχήμα 4.2.5: Ιστόγραμμα του πλήθους των αποζημιώσεων ανά μήνα από την τιμή 2.100



Από το ιστόγραμμα φαίνεται πως υπάρχει πιθανότητα να ακολουθούν κανονική κατανομή τα δεδομένα. Ωστόσο ενδέχεται και κάποια άλλη κατανομή με βαριά δεξιά ουρά να προσαρμόζεται πιο ικανοποιητικά από την κανονική. Στη συνέχεια θα ακολουθήσει το Q-Q Plot για την κανονική κατανομή καθώς και το Kolmogorov-Smirnov test για την υπόθεση της κανονικής κατανομής.

Σχήμα 4.2.6: Q-Q plot του πλήθους των αποζημιώσεων ανά μήνα από την τιμή 2.100



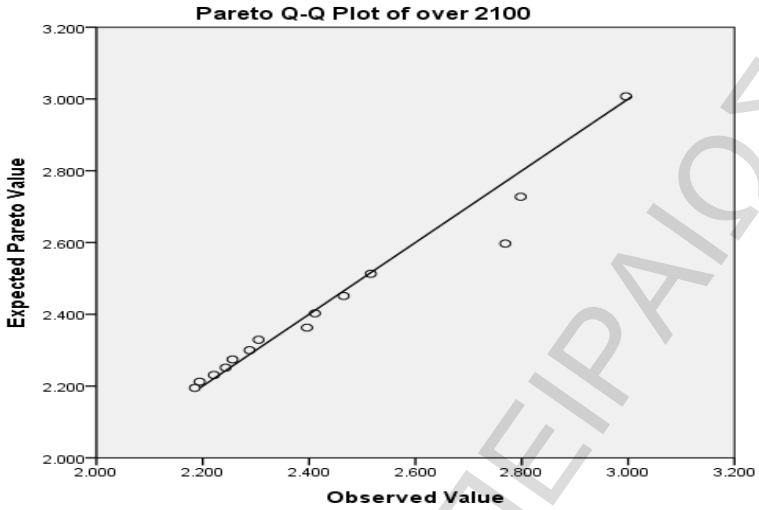
Από το Q-Q Plot παρατηρούμε πως τα δεδομένα πιθανόν να μην ακολουθούν την κανονική κατανομή

Πίνακας 4.2.5: Kolmogorov Smirnov test του πλήθους των αποζημιώσεων ανά μήνα από την τιμή 2.100

One-Sample Kolmogorov-Smirnov Test		
		over 2.100
N		14
Normal Parameters	Mean	2431,64
	Std. Deviation	254,383
Most Extreme Differences	Absolute	,191
	Positive	,191
	Negative	-,166
Kolmogorov-Smirnov Z		,714
Asymp. Sig. (2-tailed)		,689

Η τιμή του p-value στο Kolmogorov-Smirnov test ανέρχεται στο 0,689>0,05 που σημαίνει πως δεν μπορεί να απορριφθεί η υπόθεση πως τα δεδομένα ακολουθούν κανονική κατανομή. Στην περίπτωση όμως αυτή ίσως είναι προτιμότερο να ελεγχθεί και κάποια άλλη κατανομή αφού όπως έχει αναφερθεί και προηγουμένως ο πολύ μικρός αριθμός παρατηρήσεων επηρεάζει την ορθότητα των στατιστικών test. Μια κατανομή που θα πρέπει να εξεταστεί λόγω της μακριάς ουράς της είναι η κατανομή Pareto. Στο παρακάτω σχήμα παρουσιάζεται το Q-Q Plot των δεδομένων ως προς την κατανομή Pareto.

Σχήμα 4.2.7: Q-Q plot για την κατανομή Pareto του πλήθους των αποζημιώσεων ανά μήνα από την τιμή 2.100



Παρατηρώντας το Q-Q Plot διακρίνεται μια πολύ καλή προσαρμογή των δεδομένων στην Pareto πράγμα που δεν συνέβη για την κανονική κατανομή ασχέτως του αποτελέσματος του Kolmogorov-Smirnov test. Εφόσον η κατανομή των δεδομένων φαίνεται να ακολουθεί την κατανομή Pareto θα πρέπει να γίνει και εκτίμηση των παραμέτρων αυτής και στη συνέχεια να κατασκευαστεί ένας έλεγχος Kolmogorov Smirnov για να αποφασιστεί αν η υπόθεση ότι τα δεδομένα πάνω από τις 2.100 παρατηρήσεις ακολουθούν την κατανομή Pareto απορρίπτεται η όχι, καθώς και αν οι εκτιμήσεις αυτές είναι σωστές. Ένας ασφαλής τρόπος για εκτιμηθούν οι παράμετροι της κατανομής Pareto είναι η μέθοδος των ροπών. Στην περίπτωση της κατανομής Pareto με μια παράμετρο είναι γνωστό το κάτω όριο η σ.π.π της δίνεται από τον τύπο :

$$f(x) = \frac{a x_{\mathrm{m}}{}^{a}}{x^{a+1}}, \quad x > x_m$$

Σε αυτή την περίπτωση λοιπόν η άγνωστη παράμετρος είναι μόνο το a αφού το x_m είναι το κάτω όριο που στην περίπτωση αυτή είναι ο αριθμός 2.100. Για να βρεθεί η άγνωστη παράμετρος αρκεί να εξισωθεί η πρώτη δειγματική με την πρώτη πληθυσμιακή ροπή.

$m_k \approx \mu_k(\theta)$.

$$\frac{a x_m}{a-1} = 2431 \Rightarrow a = 7,344 \bullet$$

Από την εκτίμηση με την μέθοδο των ροπών παίρνουμε ότι η κατανομή που θα εξεταστεί είναι η Pareto(7,344,2.100).

Το επόμενο βήμα που θα γίνει είναι ένας έλεγχος Kolmogorov Smirnov για να αποφασιστεί εάν η υπόθεση ότι τα δεδομένα ακολουθούν κατανομή Pareto είναι σωστή.

Πίνακας 4.2.6: Kolmogorov Smirnov test για το πλήθος των αποζημιώσεων ανά μήνα από την τιμή 2.100

Two-sample Kolmogorov-Smirnov test
data: over.2.100 and Pareto (7.344,2.100)
D = 0.2143, p-value = 0.9205
alternative hypothesis: two-sided

Από τον παραπάνω πίνακα παρατηρούμε πως το p-value του ελέγχου είναι 0,92 με αποτέλεσμα να μην μπορούμε να απορρίψουμε την υπόθεση πως τα δεδομένα ακολουθούν την κατανομή Pareto (7.344,2.100).

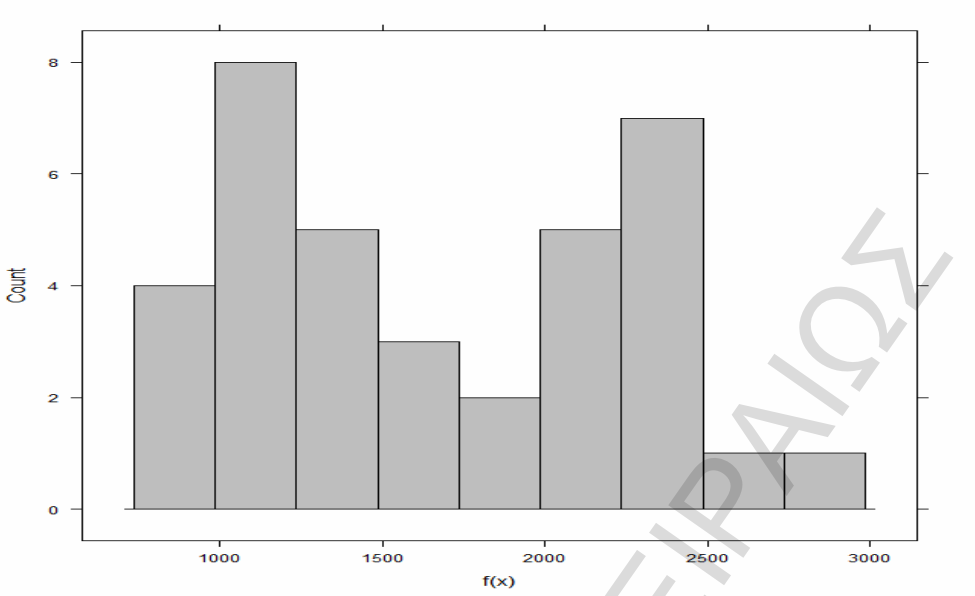
Με βάση τα παραπάνω υποθέτουμε πως η κατανομή της συχνότητας των αποζημιώσεων ακολουθεί μια μεικτή κατανομή που αποτελείται από μια κανονική κατανομή καθώς και μια κατανομή Pareto. Η μορφή της κατανομής του αριθμού των αποζημιώσεων ανά μήνα περιγράφεται από τον παρακάτω τύπο:

$$f(x) = p * f_1(x) + (1-p) * f_2(x)$$

Όπου η $f_1(x)$ είναι η κανονική κατανομή κατανομή N(1315.18,89700) ενώ η $f_2(x)$ είναι η κατανομή Pareto(7.344,2.100) και το p αντιστοιχεί στον αριθμό 0,61.

Για τον έλεγχο της υπόθεσης ότι τα δεδομένα προέρχονται από την συγκεκριμένη δικόρυφη κατανομή θα κατασκευαστεί το ιστόγραμμα συχνοτήτων της ενώ θα πραγματοποιηθεί και ένα έλεγχος Kolmogorov Smirnov για να αποφασιστεί αν τα πραγματικά συνολικά δεδομένα που αφορούν τον αριθμό των μη μηδενικών αποζημιώσεων μπορούν να περιγραφούν αποτελεσματικά από την κατανομή που έχουμε κατασκευάσει.

Σχήμα 4.2.8: Ιστόγραμμα του αναμενόμενου πλήθους των μη μηδενικών αποζημιώσεων



Το παραπάνω ιστόγραμμα του αναμενόμενου αριθμού των μη μηδενικών αποζημιώσεων φαίνεται να μοιάζει με το ιστόγραμμα των παρατηρηθέντων μη μηδενικών αποζημιώσεων (**βλ. Σχήμα 4.2.1**). Αυτό που απομένει είναι και το αποτέλεσμα του στατιστικού ελέγχου το οποίο θα καθορίσει για την ορθότητα η όχι της αρχικής υπόθεσης.

Πίνακας 4.2.7: Kolmogorov Smirnov test για την κατανομή των πλήθους των αποζημιώσεων ανά μήνα με την θεωρητική κατανομή

Two-sample Kolmogorov-Smirnov test
data: number of positive occurrences per month and f(x)
D = 0.1667, p-value = 0.7069
alternative hypothesis: two-sided

Το p-value του Kolmogorov Smirnov test είναι 0,7 που σημαίνει πως δεν μπορούμε να απορρίψουμε την υπόθεση πως ο αριθμός των μη μηδενικών αποζημιώσεων ανά μήνα ακολουθεί την κατανομή με σ.π.π.:

$$f(x) = p * f_1(x) + (1-p) * f_2(x),$$

όπου η $f_1(x)$ είναι η σ.π.π. της κανονικής κατανομής N(1315.18,89700) ενώ η $f_2(x)$ η σ.π.π. είναι της κατανομής Pareto(7.344,2.100) και το p αντιστοιχεί στον αριθμό 0,61.

4.3 Μελέτη της κατανομής του μεγέθους των αποζημιώσεων

Το μέγεθος των αποζημιώσεων οι οποίες καταφθάνουν σε μια ασφαλιστική εταιρία αποτελεί την πιο σημαντική ποσότητα για την συνέχιση της λειτουργίας της γιατί στην πραγματικότητα αυτό το μέγεθος είναι που καλείται να αποδώσει τους ασφαλισμένους ως αποτέλεσμα της ζημιάς που έχει προκύψει. Στην μελέτη που ακολουθεί γίνεται μα προσπάθεια για να περιγραφεί η τυχαία μεταβλητή των αποζημιώσεων και να προσδιοριστεί αν είναι δυνατόν η κατανομή του.

4.3.1 Περιγραφή των δεδομένων

Τα διαθέσιμα δεδομένα για το μέγεθος των αποζημιώσεων αποτελούνται από 89779 αποζημιώσεις για τα τρία εξεταζόμενα έτη συμπεριλαμβανομένων και των μηδενικών αποζημιώσεων.

4.3.2 Στατιστική ανάλυση δεδομένων

Για να κατανοηθεί το είδος των δεδομένων καλύτερα θα πρέπει αρχικά να παρουσιαστούν μερικά περιγραφικά στοιχεία τα οποία και θα υποδείξουν τον τρόπο με τον οποίο θα συνεχιστεί η ανάλυση.

I. Ανάλυση του συνόλου των παρατηρήσεων

Μια πρώτη ιδέα για τα δεδομένα μπορούμε να πάρουμε από τον παρακάτω πίνακα ο οποίος περιλαμβάνει τα περιγραφικά μέτρα για την μεταβλητή του μεγέθους των αποζημιώσεων.

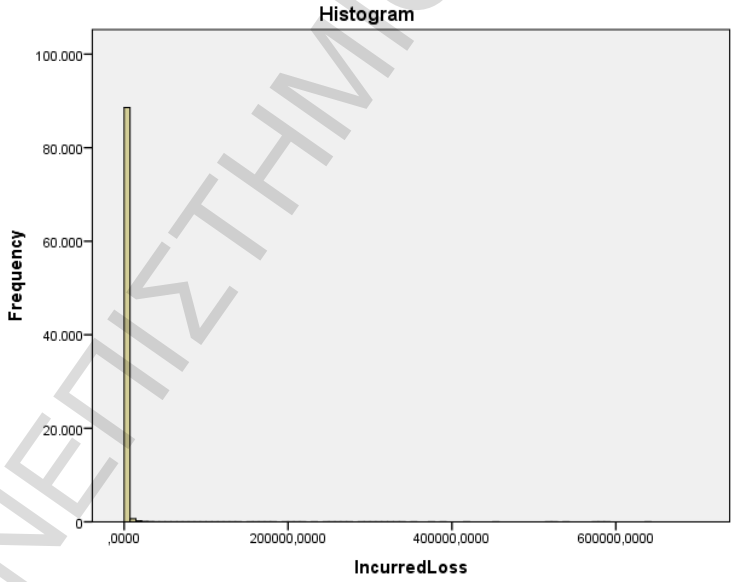
Πίνακας 4.3.1: Πίνακας περιγραφικών στοιχείων της μεταβλητής των αποζημιώσεων

Statistics	
IncurredLoss	
N	89779
Mean	884,683
Median	40,510
Std. Deviation	9389,711
Variance	88166677,236

Skewness	38,423
Std. Error of Skewness	0,008
Kurtosis	1839,207
Std. Error of Kurtosis	0,016

Από τον παραπάνω πίνακα περιγραφικών στοιχείων παρατηρούμε πως η μέση τιμή των δεδομένων ισούται με 884,683 και η διακύμανση τους με 88166677,236. Επίσης από την τιμή της ασυμμετρίας που ισούται 38,423 περιμένουμε η κατανομή των δεδομένων να παρουσιάζει «βαριά» δεξιά ουρά. Στο παρακάτω σχήμα παρουσιάζεται το ιστόγραμμα συχνοτήτων των δεδομένων.

Σχήμα 4.3.1: Ιστόγραμμα της μεταβλητής των αποζημιώσεων



Από το ιστόγραμμα δεν είναι δυνατόν να εξαχθούν συμπεράσματα αφού υπάρχει πολύ μεγάλη συγκέντρωση παρατηρήσεων κοντά στο μηδέν με αποτέλεσμα να μην εμφανίζονται καν στο γράφημα οι μεγάλες παρατηρήσεις οι οποίες είναι αυτές που ενδιαφέρουν περισσότερο.

ΙΙ. Ανάλυση περικομμένων δεδομένων

Όπως αναφέρθηκε προηγουμένως τα δεδομένα παρουσιάζουν υψηλή συγκέντρωση κοντά στο μηδέν με αποτέλεσμα να μην είναι δυνατή η ανάλυση των παρατηρήσεων. Ένας τρόπος για να καταστεί δυνατή η ανάλυση είναι να περικοπούν τα δεδομένα στο μηδέν ή κοντά σε αυτό έτσι ώστε να μεγαλώσει η πιθανότητα εμφάνισης των ακραίων δεδομένων που αποτελούν αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας.

- **Πρώτη περίπτωση : περικομμένες στο μηδέν**

Οι μηδενικές παρατηρήσεις αποτελούν ένα πολύ μεγάλο ποσοστό των συνολικών διαθέσιμων παρατηρήσεων συγκεκριμένα μηδενικές είναι 26.802 παρατηρήσεις. Ο παρακάτω πίνακας περιλαμβάνει τα περιγραφικά στοιχεία για τη συγκεκριμένη περίπτωση

Πίνακας 4.3.2: Πίνακας στοιχείων περιγραφικής στατιστικής της περικομμένης στο 0 μεταβλητής των αποζημιώσεων

Statistics	
IncurredLoss	
N	62977
Mean	1261,190139
Std. Deviation	11189,9392681
Variance	125214740,823
Skewness	32,263
Std. Error of Skewness	,010
Kurtosis	1294,748
Std. Error of Kurtosis	,020
Maximum	639845,7000

Τα δεδομένα του πίνακα είναι διαφοροποιημένα τώρα αφού λείπουν οι μηδενικές παρατηρήσεις με την μέση τιμή να διαμορφώνεται στην τιμή 1261 και την τυπική απόκλιση σε 11189. Το ιστόγραμμα συχνοτήτων για τις μη μηδενικές αποζημιώσεις παρουσιάζει ωστόσο την ίδια μορφή με το ιστόγραμμα των συνολικών αποζημιώσεων που απεικονίζεται παραπάνω. Ο λόγος είναι ότι συνεχίζουν να υπάρχουν πάρα πολλές παρατηρήσεις με τιμές κοντά στο μηδέν και συνεπώς δεν μπορούν να εξαχθούν συμπεράσματα ούτε σε αυτή την περίπτωση.

- **Δεύτερη περίπτωση : περικομμένες σε άλλα σημεία**

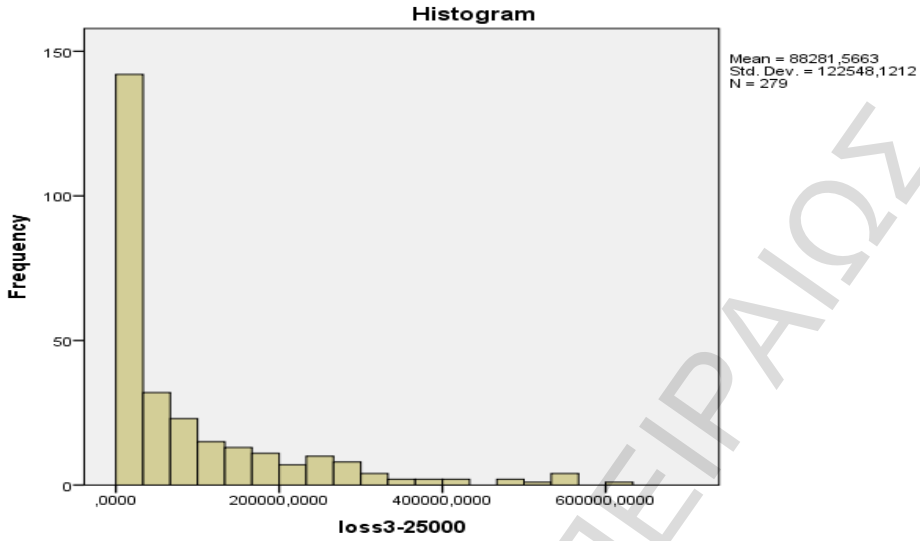
Η υψηλή συγκέντρωση παρατηρήσεων σε μικρές τιμές των αποζημιώσεων δυσκολεύει την προσαρμογή σε κάποια από τις γνωστές κατανομές. Για τον λόγο αυτό έγινε μια περαιτέρω προσπάθεια μελέτης των παρατηρήσεων πάνω από μια προκαθορισμένη τιμή. Ύστερα από δοκιμές για την επιλογή του κατάλληλου σημείου που θα περικοπούν τα δεδομένα επιλέχθηκε η τιμή των 25.000 ευρώ πάνω από την οποία θα γίνει η ανάλυση των δεδομένων. Στην πραγματικότητα δημιουργήθηκε μια νέα μεταβλητή η $loss = incurredloss - 25000$. Στην μελέτη που ακολουθεί εξετάζεται η νέα μεταΣχήματισμένη μεταβλητή loss. Στον παρακάτω πίνακα παρουσιάζονται τα περιγραφικά στοιχεία της μεταβλητής loss.

Πίνακας 4.3.3: Πίνακας περιγραφικών στοιχείων της περικομμένης στο 25.000 μεταβλητής των αποζημιώσεων

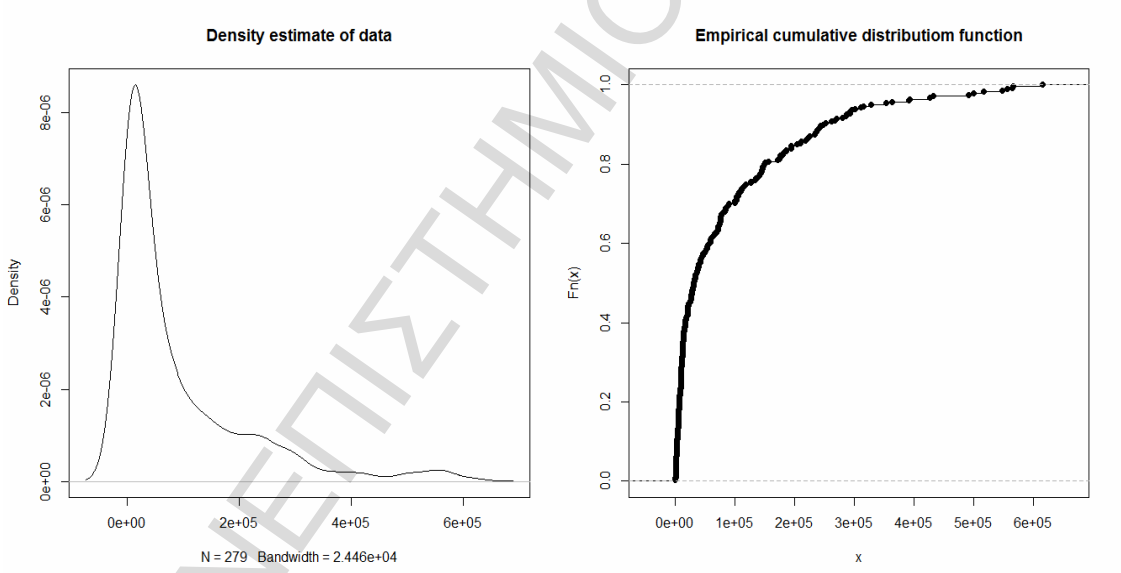
Statistics	
	loss
N	279
Mean	88281,566
Std. Error of Mean	7336,764
Median	31016,64
Std. Deviation	122548,121
Variance	15018042008,307
Skewness	2,066
Std. Error of Skewness	,146
Kurtosis	4,343
Std. Error of Kurtosis	,291

Από τον πίνακα περιγραφικών στοιχείων παρατηρούμε αρχικά μια δραματική μείωση του πλήθους των παρατηρήσεων αφού μόνο 279 υπερβαίνουν τις 25.000 ευρώ, ενώ η μέση τιμή των παρατηρήσεων ανέρχεται στην τιμή 88281 και η διακύμανση σε 15018042008. Ακόμα η ασυμμετρία της παίρνει την τιμή 2,066 που σημαίνει ότι περιμένουμε βαριά δεξιά ουρά. Στο παρακάτω σχήματα παρουσιάζονται το ιστόγραμμα συχνοτήτων των περικομμένων δεδομένων το διάγραμμα της συνάρτησης πυκνότητας πιθανότητας καθώς και το διάγραμμα της εμπειρικής συνάρτησης κατανομής.

Σχήμα 4.3.2: Ιστόγραμμα της περικομμένης στο σημείο 25.000 μεταβλητής των αποζημιώσεων



Σχήμα 4.3.3: Διάγραμμα της σ.π.π. και της εμπειρικής σ.κ. της περικομμένης στο 25.000 μεταβλητής των αποζημιώσεων

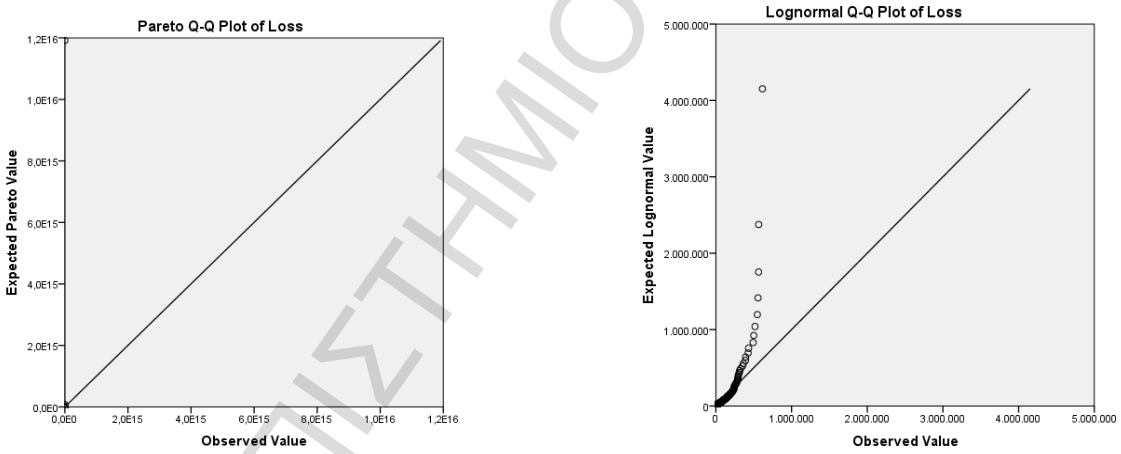


Από τα παραπάνω γραφήματα τα οποία πραγματοποιήθηκαν με τη χρήση του στάτιστικού προγράμματος R παρατηρούμε πως η κατανομή που ακολουθούν τα δεδομένα θα είναι κάποια με μεγάλη θετική ασυμμετρία. Στη συνέχεια θα γίνει προσπάθεια για τον προσδιορισμό της κατανομής που ακολουθούν οι παρατηρήσεις. Αρχικά θα διενεργηθούν κάποια γραφικά test και στη συνέχεια ανάλογα με τα αποτελέσματα που θα προκύψουν θα γίνουν και κάποια άλλα στατιστικά test.

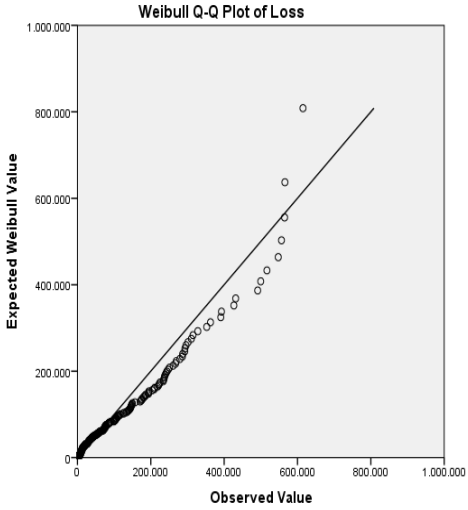
- Q-Q Plots

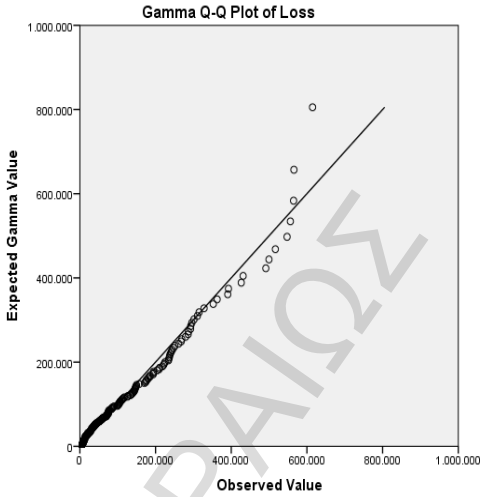
Ένας εύκολος και αρκετά διαδεδομένος τρόπος γραφικού έλεγχου προσαρμογής των δεδομένων σε μια κατανομή είναι το Q-Q Plot. Ο έλεγχος αυτός δείχνει το κατά πόσο οι παρατηρήσεις βρίσκονται κοντά με την θεωρητική κατανομή η οποία έχει επιλεγεί. Στην περίπτωση κατά την οποία η θεωρητική κατανομή που επιλέχθηκε ταυτίζεται με την κατανομή των δεδομένων τότε οι παρατηρήσεις θα βρίσκονται πάνω στην ευθεία γραμμή του γραφήματος. Στα παρακάτω γραφήματα παρουσιάζονται τα Q-Q Plots για κάποιες από τις γνωστές κατανομές με θετική ασυμμετρία με παραμέτρους όπως αυτές εκτιμώνται με την βοήθεια του στατιστικού πακέτου SPSS.

Σχήμα 4.3.4: Q-Q plot για την κατανομή Pareto και lognormal της περικομμένης στο σημείο 25.000 μεταβλητής των αποζημιώσεων



Σχήμα 4.3.5: Q-Q plot για την κατανομή Weibull και Gamma της περικομμένης στο σημείο 25.000 μεταβλητής των αποζημιώσεων





Από τα παραπάνω Q-Q Plots μόνο αυτό της κατανομής Γάμμα είναι αυτό το οποίο φαίνεται να είναι κοντά με τα δεδομένα. Οι εκτιμήσεις των παραμέτρων για την κατανομή Γάμμα δίνονται στον παρακάτω πίνακα

Πίνακας 4.3.4: Πίνακας εκτίμησης των παραμέτρων της κατανομής Gamma της περικομμένης στο σημείο 25.000 μεταβλητής των αποζημιώσεων

Estimated Distribution Parameters		
		Loss
Gamma Distribution	Shape	,519
	Scale	170125

- Kolmogorov Smirnov test

Τα Q-Q Plots όπως έχει προαναφερθεί είναι γραφικά test που ως στόχο έχουν να δείξουν την εφαρμογή των δεδομένων σε μια κατανομή αλλά στην πραγματικότητα λειτουργούν μόνο ως μια ένδειξη καλής προσαρμογής. Στην περίπτωση που τα διαθέσιμα δεδομένα είναι αρκετά υπάρχουν στατιστικά test τα οποία αποτυπώνουν με ακρίβεια την απόσταση των πραγματικών δεδομένων από τις θεωρητικές τιμές και δοθέντος ενός επιπέδου σημαντικότητας δίνουν αποτέλεσμα για την αποδοχή ή όχι της υπόθεσης ότι τα δεδομένα ακολουθούν κάποια συγκεκριμένη κατανομή. Το πιο γνωστό στατιστικό test για την συγκεκριμένη περίπτωση είναι το Kolmogorov Smirnov test. Εφόσον από τα Q-Q plots που προηγήθηκαν φάνηκε ότι τα δεδομένα μας είναι πιθανόν να ακολουθούν την κατανομή Γάμμα τότε θα διενεργήσουμε ένα έλεγχο Kolmogorov Smirnov για να αποφασίσουμε εάν

ευσταθεί αυτή η υπόθεση. Στους παρακάτω πίνακες διενεργήθηκε Kolmogorov Smirnov test για την κατανομή Γάμμα στην μεταβλητή Loss που αντιστοιχεί στις περικομμένες παρατηρήσεις στο σημείο 25.000 .

Πίνακας 4.3.5: Kolmogorov Smirnov test για την κατανομή Gamma της περικομμένης στο σημείο 25.000 μεταβλητής των αποζημιώσεων

Two-sample Kolmogorov-Smirnov test
data: k and Loss
D = 0.0681, p-value = 0.5371
alternative hypothesis: two-sided

Το p-value του ελέγχου είναι 0,5371 που σημαίνει πως δεν μπορούμε να απορρίψουμε την υπόθεση πως τα δεδομένα προέρχονται από την κατανομή Gamma(0,519 , 170125).

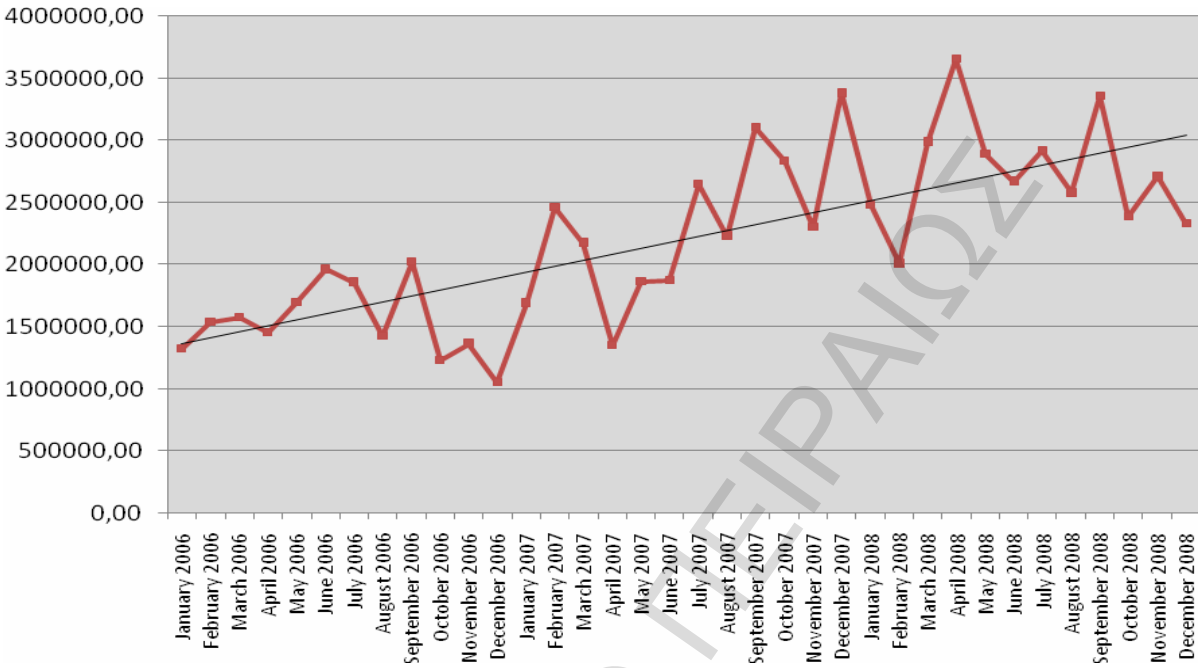
4.4 Μελέτη της χρονικής εξέλιξης των αποζημιώσεων

Ένα σημαντικό κομμάτι της ανάλυσης ενός χαρτοφυλακίου κινδύνων είναι η χρονική εξέλιξη των μεγεθών που το απαρτίζουν για να αποσαφηνιστεί το αν υπάρχει κάποια τάση ή περιοδικότητα και το πού μπορεί να οφείλεται αυτή. Στη συνέχεια αυτής της ενότητας θα παρατεθούν γραφικά τα διαγράμματα των κυριοτέρων ποσοτήτων σε για να παρατηρηθεί η χρονική εξέλιξη τους.

4.4.1 Συνολικές μηνιαίες αποζημιώσεις

Στο παρακάτω γράφημα παρουσιάζονται οι συνολικές αποζημιώσεις του κλάδου αυτοκινήτου σε μηνιαία κλίμακα για την εξεταζόμενη τριετία.

Σχήμα 4.4.1: Μηνιαίο διάγραμμα των συνολικών αποζημιώσεων

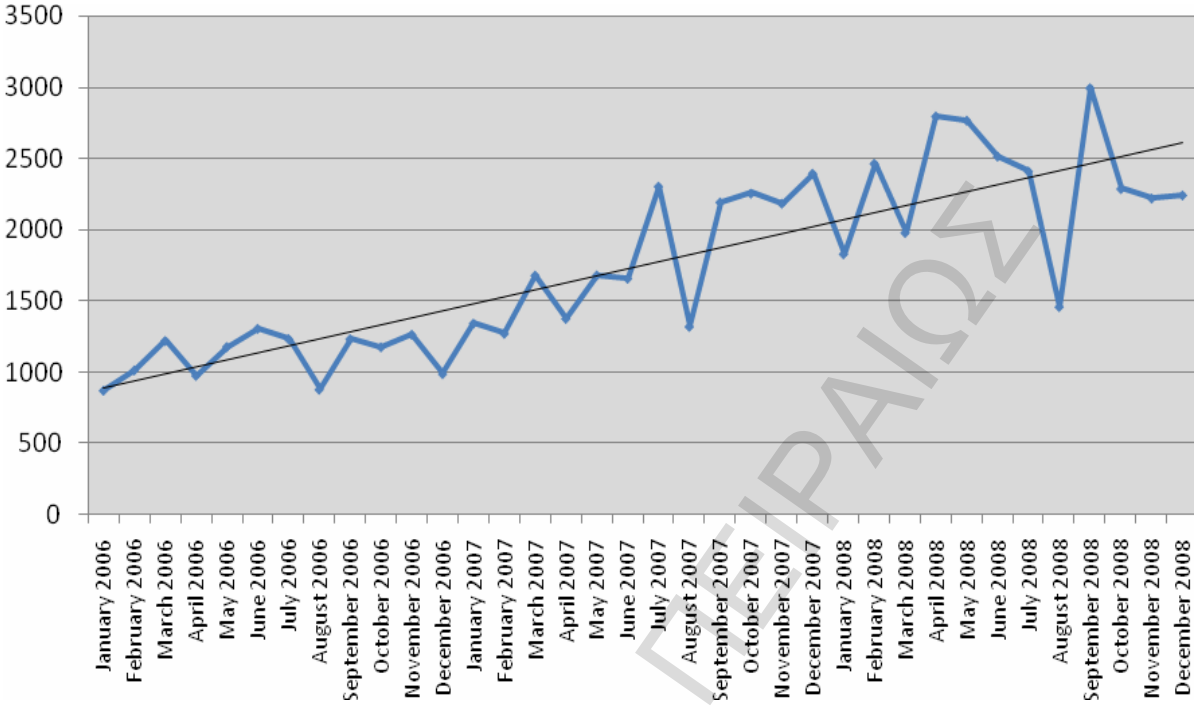


Από το παραπάνω χρονοδιάγραμμα παρατηρούμε πως υπάρχει μια αυξητική τάση στις μηνιαίες αποζημιώσεις. Προτού βγουν όμως συμπεράσματα για την αλλαγή της οδηγικής συμπεριφοράς των ασφαλισμένων του συγκεκριμένου χαρτοφυλακίου θα πρέπει να επισημανθεί πως ενδέχεται η αύξηση αυτή να προέρχεται από αύξηση της έκθεσης στον κίνδυνο. Από το παραπάνω γίνεται ξεκάθαρο πως δεν μπορεί να βγεί ασφαλές συμπέρασμα για το αν η αύξηση των αποζημιώσεων μεταφράζεται σε μεγαλύτερη ζημιά για την ασφαλιστική επιχείρηση αφού αυτή μπορεί να εκμηδενίζεται από την εισροή ασφαλίστρων από νέους ασφαλισμένους. Η μόνη ασφαλής παρατήρηση που μπορεί να γίνει είναι πως φαίνεται ότι ο μήνας Σεπτέμβριος είναι ο μήνας που και τις τρείς διαδοχικές χρονιές παρουσιάζει από τα υψηλότερα επίπεδα αποζημιώσεων μέσα στο έτος.

4.4.2 Αριθμός των αποζημιώσεων ανά μήνα

Στο παρακάτω γράφημα απεικονίζεται ο αριθμός των μη μηδενικών αποζημιώσεων ανά μήνα.

Σχήμα 4.4.2: Μηνιαίο διάγραμμα του αριθμού των απαιτήσεων

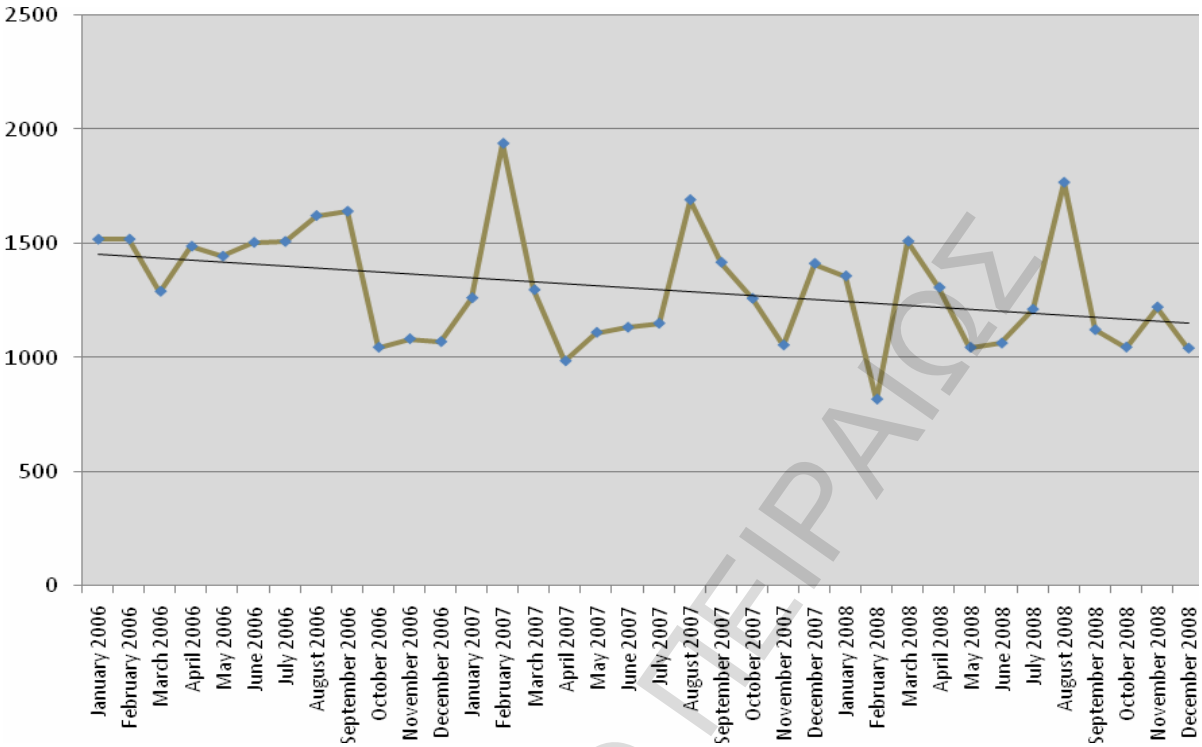


Στο παραπάνω γράφημα του αριθμού των μηνιαίων αποζημιώσεων παρατηρείται μια αυξανόμενη τάση στων αριθμό των ατυχημάτων που καταφθάνουν στην ασφαλιστική εταιρία με την πάροδο του χρόνου. Όπως συνέβη και στο μέγεθος των μηνιαίων αποζημιώσεων έτσι και σε αυτή την περίπτωση δεν είναι ασφαλές να εξαχθεί το συμπέρασμα ότι οι ασφαλισμένοι προκαλούν περισσότερα ατυχήματα από ότι παλιά μιας και δεν έχει ληφθεί υπόψη το ενδεχόμενο μιας ταυτόχρονης αύξησης της έκθεσης στον κίνδυνο. Ένα στοιχείο που γίνεται ξεκάθαρο παρατηρώντας το παραπάνω γράφημα είναι το γεγονός πως σε κάθε μια από τις τρείς χρονιές ο χαμηλότερος αριθμός ατυχημάτων παρατηρείται τον μήνα Αύγουστο. Μια εξήγηση για το γεγονός μπορεί να είναι η μειωμένη κίνηση των αυτοκινήτων το συγκεκριμένο μήνα αφού οι περισσότεροι Έλληνες αποφασίζουν να πραγματοποιήσουν τις καλοκαιρινές τους διακοπές εκείνη την περίοδο.

4.4.3 Μέση αποζημίωση ανά μήνα

Στο παρακάτω γράφημα απεικονίζεται η μέση αποζημίωση η οποία καταφθάνει στην ασφαλιστική επιχείρηση ανά μήνα.

Σχήμα 4.4.3: Μηνιαίο διάγραμμα του μεγέθους της μέσης απαίτησης ανά μήνα



Από το παραπάνω μηνιαίο διάγραμμα παρατηρείται μια τάση μείωσης της μέσης αποζημίωσης με την πάροδο του χρόνου. Το παρόν γράφημα αποτελεί μια σύνδεση των προηγουμένων δύο διαγραμμάτων και τα αποτελέσματα του προέρχονται από τα αντίστοιχα αποτελέσματα τους. Η σταδιακή μείωση της μέσης αποζημίωσης παρά την αύξηση του μεγέθους των μηνιαίων αποζημιώσεων οφείλεται στην μεγάλη αυξηση του αριθμού των μη μηδενικών αποζημιώσεων που παρουσιάστηκε στο Σχήμα 4.4.2.

4.4.4 Συμπεράσματα για την κατανομή των αποζημιώσεων

Το συγκεκριμένο κεφάλαιο ασχολήθηκε τόσο με την ανάλυση των δεδομένων για την περιγραφή της κατανομής που ακολουθεί ο αριθμός των αποζημιώσεων όσο και με την εύρεση της κατανομής που ακολουθεί η μεταβλητή της σφοδρότητας των αποζημιώσεων. Αν και τα αποτελέσματα ήταν θετικά παρ'όλα αυτά λόγω των μετασχηματισμών που χρησιμοποιήθηκαν δεν ήταν δυνατόν να προσεγγιστεί η κατανομή των αθροιστικών αποζημιώσεων. Το συμπέρασμα που βγαίνει είναι ότι στην περίπτωση πραγματικών

δεδομένων, όπως είναι το συγκεκριμένο χαρτοφυλάκιο αποζημιώσεων ασφαλιστηρίων συμβολαίων αυτοκινήτου, δεν είναι πάντα εφικτό να περιγραφεί κατάλληλα από τις γνωστές οικογένειες κατανομών. Το αναμενόμενο δηλαδή ως έναν βαθμό αποτέλεσμα της έως τώρα ανάλυσης έρχεται να επιβεβαιώσει το γεγονός ότι ορισμένες κατηγορίες προβλημάτων μπορούν να επιλυθούν μόνο με ειδικές τεχνικές ανάλυσης προορισμένες για συγκεκριμένα είδη δεδομένων. Το επόμενο κεφάλαιο της συγκεκριμένης διπλωματικής εργασίας επικεντρώνεται σε μια ραγδαία αναπτυσσόμενη τα τελευταία χρόνια θεωρία η οποία στηρίζεται στην μελέτη των ακραίων τιμών.

Κεφάλαιο 5

Εφαρμογή της θεωρίας Ακραίων Τιμών σε πραγματικά δεδομένα

5.1 Εισαγωγή

Η θεωρία των ακραίων τιμών μελετά την εμφάνιση των ακραίων παρατηρήσεων. Τα αποτελέσματα του Κεφαλαίου 4 που προέκυψαν από την ανάλυση των δεδομένων

αποτελούν μια πρώτη ανάλυση για την εξαγωγή αρχικών συμπερασμάτων όσον αφορά την κατανομή αλλά και τη γενική τάση των αποζημιώσεων. Ωστόσο, η παρούσα εργασία επικεντρώνεται στην εξαγωγή συμπερασμάτων όσον αφορά τις πολύ μεγάλες αποζημιώσεις με αποτέλεσμα η έως τώρα ανάλυση να μην βοηθά σε μεγάλο βαθμό στην εκτίμηση αυτών των ακραίων παρατηρήσεων. Το κενό αυτό καλείται να καλύψει η θεωρία των ακραίων τιμών που αναφέρθηκε θεωρητικά στο Κεφάλαιο 3.

Αρχικά στο πρώτο μέρος του παρόντος κεφαλαίου θα εφαρμοσθεί στα διαθέσιμα δεδομένα των αποζημιώσεων η μέθοδος Block Maxima. Συγκεκριμένα θα γίνει η προσπάθεια να μελετηθεί η κατανομή που ακολουθούν οι μέγιστες παρατηρήσεις ανά κάποιο ορισμένο διάστημα έτσι ώστε να μπορεί να αποσαφηνιστεί το ύψος του ποσού που δεν θα ξεπεράσει η μέγιστη αποζημίωση στον επόμενο χρόνο για ένα δεδομένο διάστημα εμπιστοσύνης.

Στο δεύτερο μέρος του κεφαλαίου θα χρησιμοποιηθεί η μέθοδος Peak Over Threshold. Σύμφωνα με αυτή την μέθοδο θα εξεταστούν οι παρατηρήσεις οι οποίες ξεπερνάνε κάποιο προκαθορισμένο όριο. Επίσης θα γίνει προσπάθεια σύμφωνα με τα διαθέσιμα δεδομένα να εκτιμηθεί ένα καινούριο ανώτατο όριο. Στη συνέχεια, όπως και στη μέθοδο Block Maxima θα εκτιμηθεί και πάλι το ποσό που δεν θα ξεπεράσει καμία αποζημίωση σε ένα δεδομένο επίπεδο σημαντικότητας κατά το επόμενο έτος.

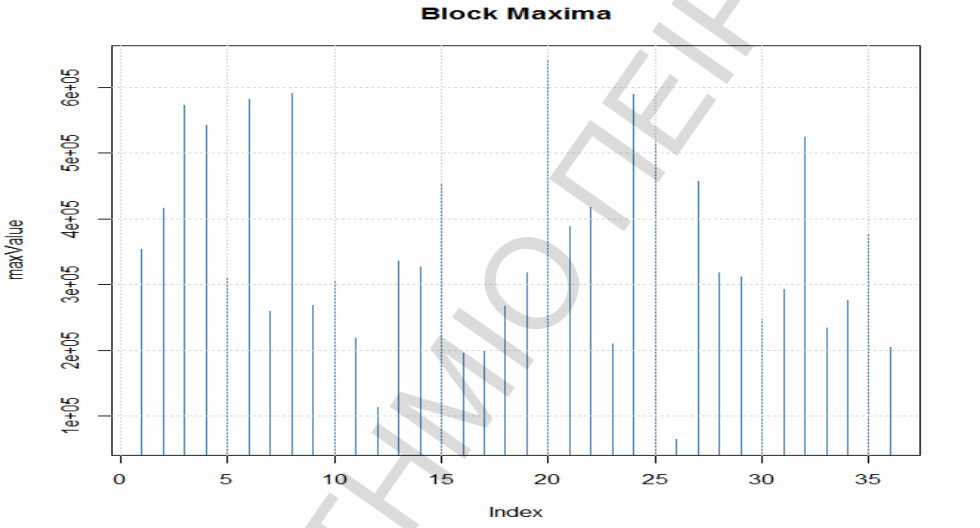
5.2 Εφαρμογή της μεθόδου Block Maxima στα δεδομένα

Η μέθοδος Block Maxima χρησιμοποιείται συνήθως όταν τα διαθέσιμα δεδομένα δεν είναι πλήρη αλλά υπάρχουν μόνο οι μέγιστες παρατηρήσεις ανά διάστημα. Η μέθοδος αυτή χρησιμοποιεί την πληροφορία που της δίνει το ύψος της μέγιστης παρατήρησης ανά κάποιο χρονικό διάστημα έτσι ώστε αυτές οι παρατηρήσεις να ακολουθούν την Γενικευμένη κατανομή Pareto (GEV). Οταν εκτιμηθεί η GEV τότε είναι εύκολο να υπολογιστεί η στάθμη απόδοσης (βλ. Κεφ. 3.2.2). Με αυτό τον τρόπο δίνεται η δυνατότητα στην ασφαλιστική επιχείρηση να προβλέψει το μέγεθος της μέγιστης αποζημίωσης ανά κάποιο χρονικό διάστημα και να κρατάει αντίστοιχο ποσό κεφαλαίων για μια αποτελεσματικότερη διαχείριση των αποζημιώσεων που θα προκύψουν. Για την εφαρμογή της συγκεκριμένης μεθόδου θα χρησιμοποιηθεί το στατιστικό πρόγραμμα R αφού προσφέρει ικανοποιητικό αριθμό ρουτινών για τον υπολογισμό της.

5.2.1 Η Block Maxima για τα μηνιαία μέγιστα

Στο σημείο αυτό επιλέχθηκαν τα μέγιστα για χρονικό διάστημα ενός ημερολογιακού μήνα από το σύνολο των ατομικών αποζημιώσεων για τα τρία εξεταζόμενα έτη. Ο λόγος που επιλέχθηκε το συγκεκριμένο χρονικό διάστημα είναι γιατί ο ένας ημερολογιακός μήνας περιέχει μεγάλο πλήθος παρατηρήσεων με αποτέλεσμα να ελαχιστοποιείται το λάθος εκτίμησης. Στο παρακάτω σχήμα παρουσιάζονται γραφικά οι μέγιστές αποζημιώσεις ανά μήνα ενώ στον αμέσως επόμενο πίνακα παραθέτονται μερικά περιγραφικά στοιχεία για τη συγκεκριμένη μεταβλητή.

Σχήμα 5.2.1: Διάγραμμα μεγέθους των μέγιστων αποζημιώσεων ανά μήνα



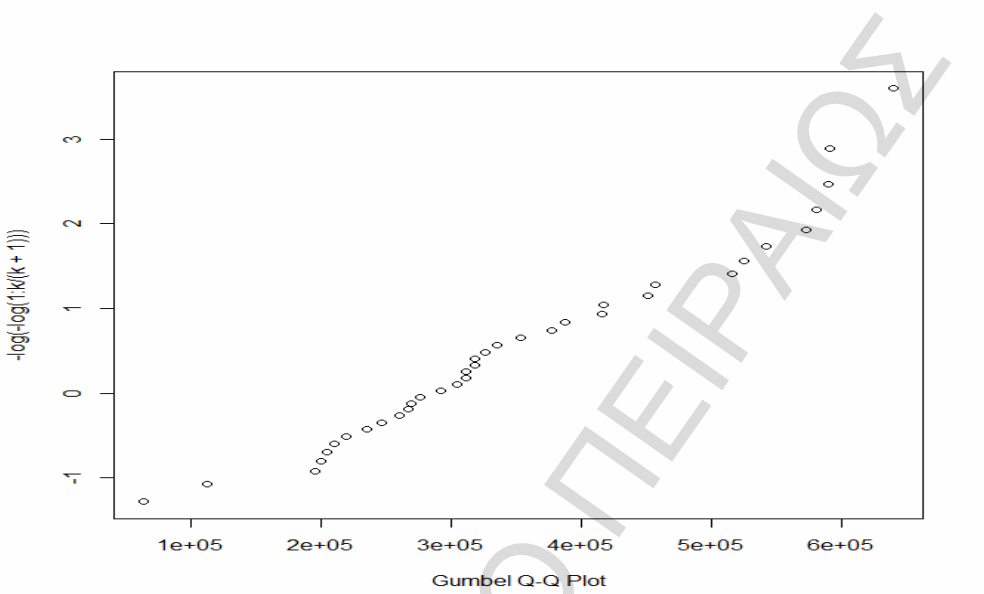
Πίνακας 5.2.1: Πίνακας περιγραφικών στοιχείων των μέγιστων αποζημιώσεων ανά μήνα

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
64020	256500	317800	352600	452700	639800

Σκοπός της ανάλυσης είναι να καθορισθεί η κατανομή που ακολουθούν οι μέγιστες παρατηρήσεις ανά μήνα. Όπως έχει αναφερθεί και στο Κεφάλαιο 3 της παρούσας εργασίας η κατανομή των μεγίστων αναμένεται να ακολουθεί μία κατανομή εκ των Gumbel , Frechet και Weibull. Για μεγαλύτερη ευκολία στον υπολογισμό των κατανομών, αυτές μπορούν να ενοποιηθούν υπό τον γενικό όρο γενικευμένη κατανομή ακραίων τιμών (GEV) έτσι ώστε η συνάρτηση πυκνότητας και των τριών κατανομών να είναι κοινή και να διαχωρίζονται μεταξύ τους από την τιμή μιας παραμέτρου.

Στο παρακάτω σχήμα παρουσιάζεται το Q-Q Plot των μηνιαίων μέγιστων παρατηρήσεων ανά μήνα με την κατανομή Gumbel

Σχήμα 5.2.2: Q-Q Plot των μηνιαίων μέγιστων με την κατανομή Gumbel

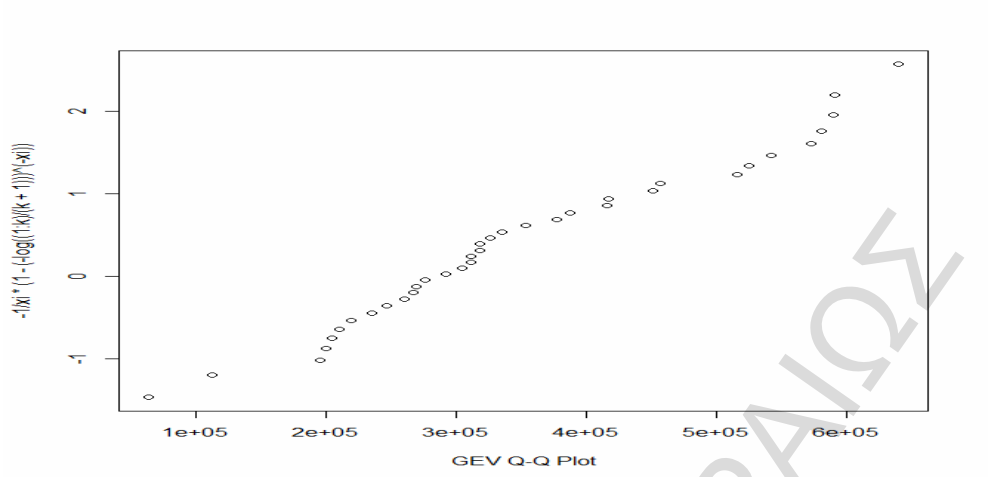


.

Από το Q-Q Plot με την κατανομή Gumbel παρατηρούμε πως τα δεδομένα βρίσκονται κοντά στη διαγώνιο αλλά ίσως μια άλλη κατανομή να δίνει καλύτερα αποτελέσματα. Ύστερα από διαδοχικές δοκιμές για την παράμετρο ξ της κατανομής GEV παρατηρήθηκε πως για ξ=-0,2 τα αποτελέσματα του Q-Q Plot είναι τα καλύτερα που μπορούν να παρατηρηθούν. Για ξ<0, όπως στη συγκεκριμένη περίπτωση, η γενικευμένη κατανομή ακραίων τιμών (GEV) αντιστοιχεί στην κατανομή Weibull όπως αυτή περιγράφηκε στο Κεφάλαιο 3.

Το Q-Q Plot της GEV για ξ=-0,2 δίνεται στο παρακάτω σχήμα.

Σχήμα 5.2.3: Q-Q Plot των μηνιαίων μέγιστων με την κατανομή GEV για ξ=-0,2



Τα παραπάνω Q-Q Plot δίνουν μια πρώτη αίσθηση για το ποια κατανομή φαίνεται να ακολουθούν οι μέγιστες παρατηρήσεις ανά μήνα στην πραγματικότητα όμως δεν αποτελούν αποτελεσματικό κριτήριο ενώ ταυτόχρονα δεν εκτιμώνται και οι ακριβείς παράμετροι της κατανομής.

5.2.2 Εκτίμηση των παραμέτρων της GEV

Ύστερα από τον διαχωρισμό των παρατηρήσεων σε ομάδες που πληρούν κάποια κριτήρια και την γραφική προσπάθεια εκτίμησης κάποιων εκ των παραμέτρων, το επόμενο βήμα στην ανάλυση είναι να εκτιμηθούν οι παράμετροι της κατανομής που ακολουθούν τα δεδομένα. Στον πίνακα που ακολουθεί παρουσιάζονται οι ακριβείς εκτιμήσεις των παραμέτρων της GEV με τη μέθοδο μεγίστης πιθανοφάνειας.

Πίνακας 5.2.2: Πίνακας εκτίμησης των παραμέτρων της GEV για τη μεταβλητή των μέγιστων μηνιαίων παρατηρήσεων

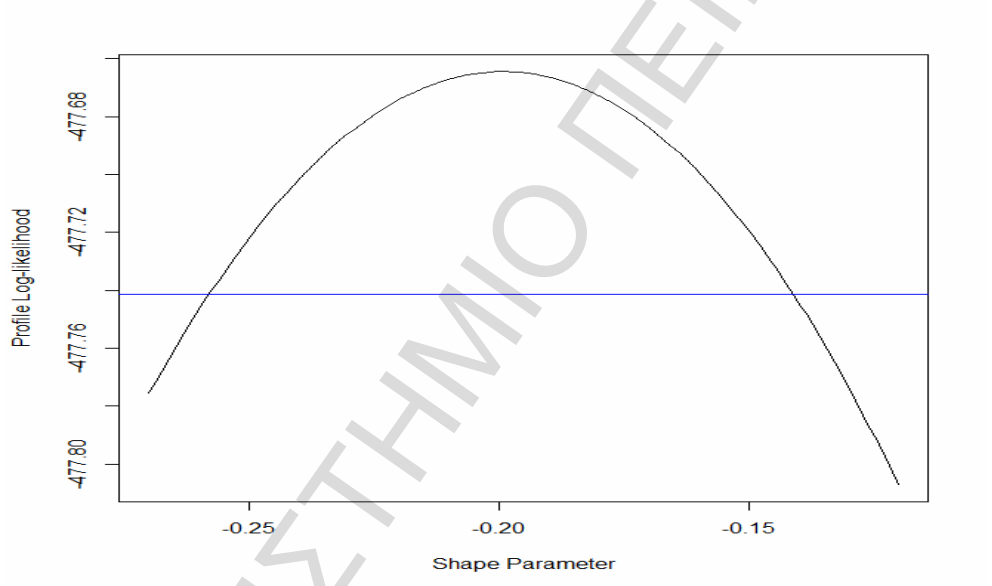
Estimation Type:	**GEV mle**	
Estimated Parameters:		
$\hat{\xi}$	$\hat{\mu}$	$\hat{\sigma}$
-1.801873e-01	**2.910448e+05**	**1.307734e+05**
Estimated standard errors		
0.1161111	**NaN**	**3883.5280541**
Converged	**0 , Neg.logarithm of loglikelihood = 477.7413**	

Ο πίνακας 5.2.5 παρουσιάζει εκτός από τις εκτιμήσεις των παραμέτρων και των τυπικών αποκλίσεων τους και δύο άλλες πολύ σημαντικές ποσότητες. Η πρώτη που παρουσιάζεται

είναι η σύγκλιση στην συγκεκριμένη κατανομή όπου το αποτέλεσμα 0 σημαίνει πως η σύγκλιση είναι επιτυχημένη. Ενώ η δεύτερη είναι η τιμή του αρνητικού λογαρίθμου της συνάρτησης πιθανοφάνειας.

Όπως έχει αναφερθεί προηγουμένως η παράμετρος ξ της GEV είναι εξέχουσας σημασίας μιας και καθορίζει ποια κατανομή ακολουθούν τα δεδομένα για αυτό το λόγο είναι αρκετά ενδιαφέρον να κατασκευαστεί ένα διάστημα εμπιστοσύνης για την τιμή αυτής της παραμέτρου. Στο παρακάτω γράφημα παρουσιάζεται ένα 95% διάστημα εμπιστοσύνης για την τιμή της παραμέτρου ξ.

Σχήμα 5.2.4: Γράφημα για το 95% δ.ε. της παραμέτρου ξ της κατανομής GEV

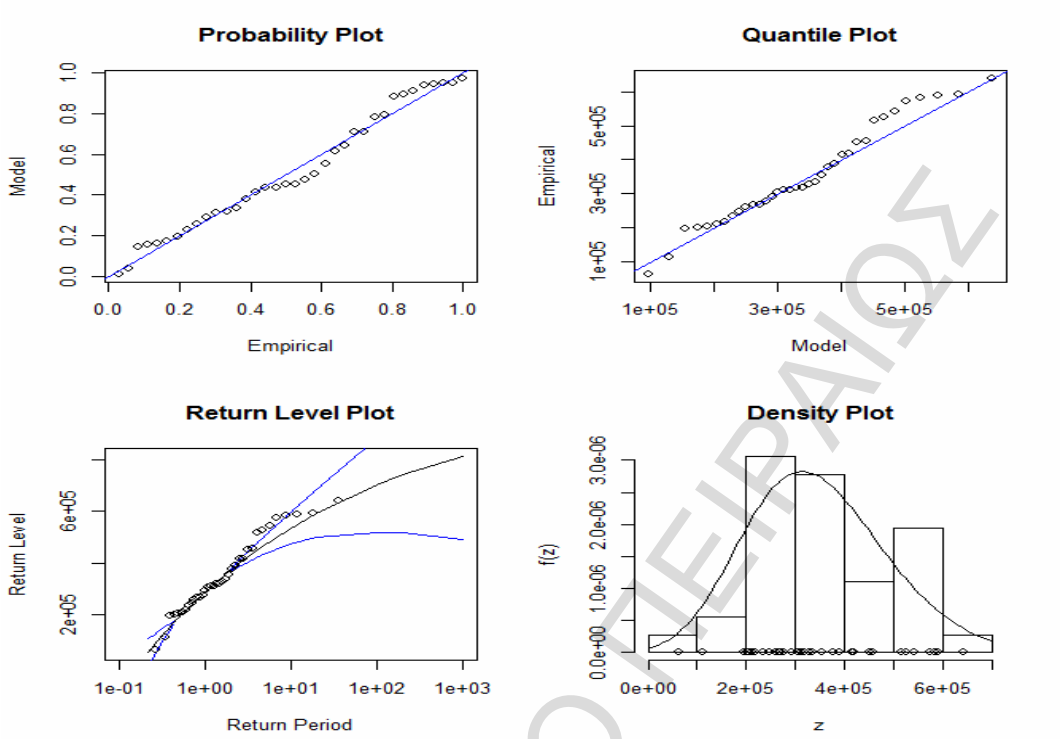


Τα αποτελέσματα του σχήματος 5.2.4 δείχνουν πως ένα 95% διάστημα εμπιστοσύνης για την παράμετρο ξ είναι

$$\Pr(-0,26 < \xi < -0,14) = 0,95$$

Ο πίνακας 5.2.2 παρουσιάζει τις εκτιμήσεις των παραμέτρων για την κατανομή GEV όμως επιθυμούμε να αποφασίσουμε αν αυτές οι εκτιμήσεις είναι σωστές. . Ο έλεγχος της καλής προσαρμογής της Γενικευμένης κατανομής ακραίων τιμών μπορεί να γίνει είτε με γραφικές μεθόδου είτε με κάποιο στατιστικό test. Στο παρακάτω σχήμα παρουσιάζονται γραφικές απεικονίσεις καλής προσαρμογής στην GEV για τη μεταβλητή των μέγιστων αποζημιώσεων ανά μήνα.

Σχήμα 5.2.5: Σύνολο γραφημάτων καλής προσαρμογής στην GPD



Το Σχήμα 5.2.5 παρουσιάζει τέσσερα επιμέρους γραφήματα. Τα γραφήματα 1,2,4 παρουσιάζουν τα P-P plot, Q-Q plot και Density plot αντίστοιχα και δείχνουν μια πολύ καλή προσαρμογή των δεδομένων στην GEV. Το γράφημα 3 απεικονίζει το διάγραμμα της στάθμης απόδοσής ποσότητα η οποία θα εξεταστεί σε επόμενη ενότητα.

Επειδή οι γραφικές απεικονίσεις δεν αποτελούν αξιόπιστο κριτήριο για την καλή προσαρμογή των δεδομένων σε μια κατανομή θα διενεργηθεί επίσης και ένα Kolmogorov - Smirnov test. Συγκεκριμένα θα παραχθούν 1000 τιμές από μια GEV με τις εκτιμήσεις των παραμέτρων από την GEV που εκτιμήθηκαν από τον Πίνακα 5.2.2 και θα συγκριθούν με τις 36 τιμές των μηνιαίων μεγίστων για να αποφασιστεί αν προέρχονται από την ίδια κατανομή.

Πίνακας 5.2.3: Kolmogorov Smirnov test για την προσαρμογή των δεδομένων με την GEV

Two-sample Kolmogorov-Smirnov test

data: Max.loss.per.month and s*
D = 0.1667, p-value = 0.7069
alternative hypothesis: two-sided
*s=1000 simulated values from GEV with the previously estimated parameters

Το Kolmogorov-Smirnov test δείχνει πως δεν μπορούμε να απορρίψουμε την υπόθεση πως οι παρατηρήσεις των μέγιστων μηνιαίων παρατηρήσεων ακολουθούν την GEV κατανομή με παραμέτρους ξ**= -1.801873e-01,** μ**=2.910448e+05 και** σ**=1.307734e+05**

5.2.3 Εκτίμηση της στάθμη απόδοσης για την μέθοδο Block Maxima

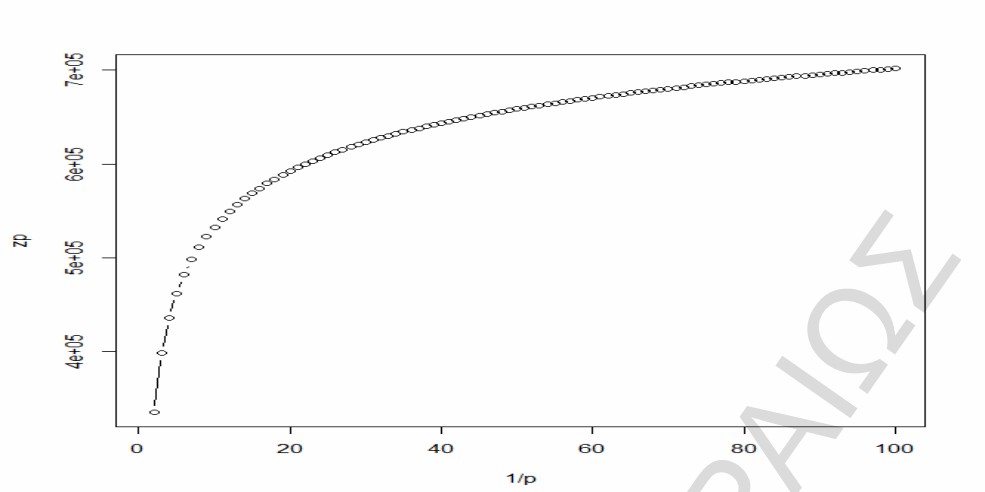
Η στάθμη απόδοσης (return level) όπως έχει οριστεί και στο Κεφάλαιο 3.2.2 είναι το κατώφλι το οποίο δεν θα υπερβεί καμία παρατήρηση με πιθανότητα $1-p$ για το επόμενο χρονικό διάστημα το οποίο θα οριστεί όπου

$$p = \frac{1}{\textit{επιθυμητός αριθμός Blocks}}\bullet$$

Θα συμβολίζουμε αυτό το κατώφλι z_p. Ένας άλλος ορισμός για τη στάθμη απόδοσής είναι ότι εκτιμάται το z_p έτσι ώστε ένα block maximum θα υπερβαίνει κατά μέσο όρο αυτό το κατώφλι κάθε 1/p χρονικές περιόδους. Η εκτίμηση της R (μέσω του πακέτου extRemes) για αυτό το υψηλό κατώφλι είναι $\hat{z}_p = 549.165$. Δηλαδή η μέγιστη αποζημίωση τον επόμενο μήνα δεν θα ξεπεράσει το ποσό των 549.165 με πιθανότητα $p = \frac{1}{12} = 91,66\%$. Εναλλακτικά, η μέγιστη ετήσια αποζημίωση θα ξεπεράσει το ποσό των 549.165 ευρώ κατά μέσο όρο μια φορά το χρόνο.

Στο παρακάτω γράφημα παρουσιάζεται η στάθμη απόδοσής για διάφορες χρονικές περιόδους που κατασκευάστηκε με το πακέτο extremes της R (βλ.Παράρτημα 1).

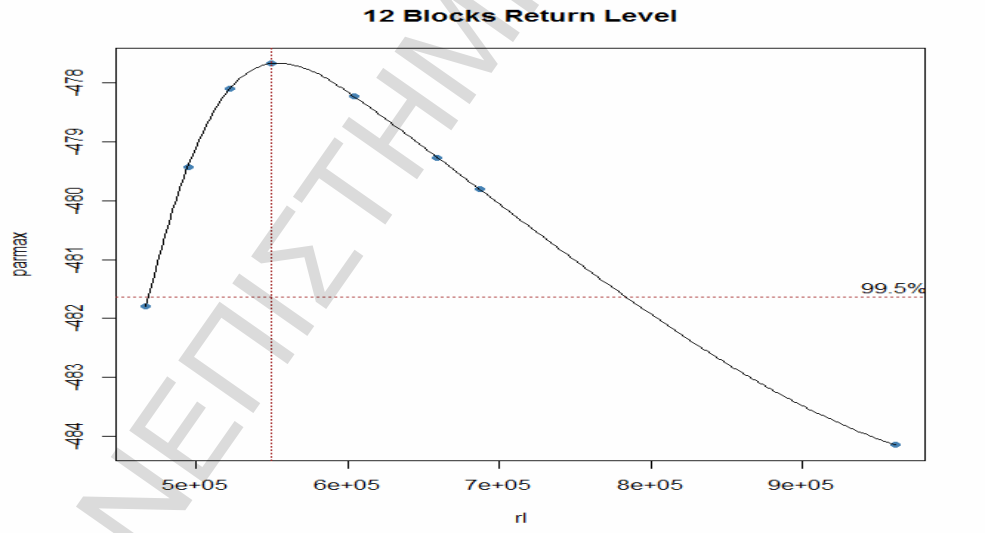
Σχήμα 5.2.6: Γράφημα στης στάθμης απόδοσης για διάφορες χρονικές περιόδους



Στο παραπάνω γράφημα φαίνεται πως η μέγιστη μηνιαία αποζημίωση θα ξεπεράσει το ποσό των 600000 κατά μέσο όρο μια φορά κάθε 24 μήνες.

Στο παρακάτω γράφημα απεικονίζεται ένα 99,5% διάστημα εμπιστοσύνης για την στάθμη απόδοσής σε χρονικό διάστημα ενός έτους.

Σχήμα 5.2.7: Γράφημα του 99.5% δ.ε. της στάθμης απόδοσης για περίοδο ενός έτους



Στο σχήμα 2.5.7 παρουσιάζεται η εκτίμηση της στάθμης απόδοσης για το επόμενο έτος που όπως και προηγουμένως υπολογίστηκε αντιστοιχεί σε μια τιμή γύρω στις 55000 ευρώ ενώ ταυτόχρονα κατασκευάστηκε και ένα 99,5% διάστημα εμπιστοσύνης για την τιμή της στάθμης απόδοσης το οποίο είναι

$$\Pr(470.000 < \hat{z}_p < 780.000) = 0,995 \bullet$$

5.3 Εφαρμογή της μεθόδου Peak over Threshold στα δεδομένα

Η μέθοδος Block Maxima αν και δίνει αρκετά καλά αποτελέσματα για τις μέγιστες παρατηρήσεις ανά κάποιο χρονικό διάστημα εντούτοις δεν λαμβάνει υπόψη της πως ενδέχεται να υπάρχουν περισσότερες από μια αρκετά υψηλές παρατηρήσεις μέσα σε αυτό. Το πρόβλημα αυτό έρχεται να το λύσει μια μέθοδος που ονομάζεται Peak Over Threshold (POT). Η μέθοδος Peak Over Threshold θεωρείται μια πιο ολοκληρωμένη μέθοδος από την Block Maxima γιατί λαμβάνει πληροφορία από όλα τα δεδομένα και επομένως από όλες τις υπερβάσεις που ξεπερνούν ένα ανώτατο όριο. Σύμφωνα με όσα θεωρητικά αναπτύχθηκαν στο Κεφάλαιο 3 αυτές οι υπερβάσεις ακολουθούν προσεγγιστικά μια κατανομή που ονομάζεται Γενικευμένη κατανομή Pareto (Generalized Pareto Distribution ή GEV). Σκοπός λοιπόν της επόμενης ενότητας είναι με τη βοήθεια του στατιστικού προγράμματος R να γίνει η εκτίμηση του ορθότερου ανώτατού ορίου, των παραμέτρων της GPD καθώς και η εκτίμηση της στάθμης απόδοσης.

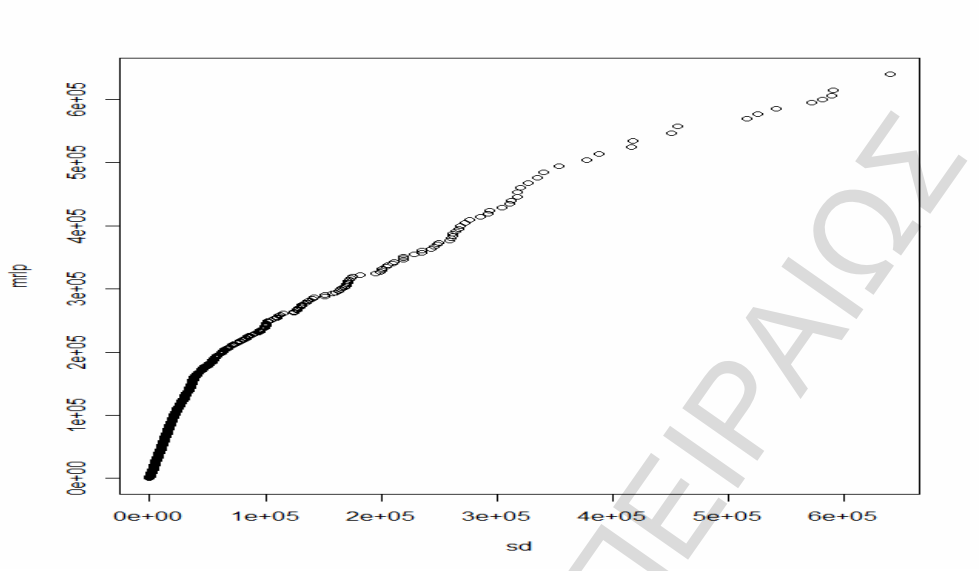
5.3.1 Η επιλογή του ανώτατου ορίου

Σύμφωνα με αυτήν την μέθοδο θα εξεταστούν οι παρατηρήσεις οι οποίες ξεπερνούν κάποιο υψηλό κατώφλι και στη συνέχεια θα εξεταστεί αν αυτές ακολουθούν την γενικευμένη Pareto κατανομή. Η επιλογή του ορίου αυτού είναι εξέχουσας σημασίας για τους ενδιαφερόμενους επιστήμονες και φυσικά και για τις ασφαλιστικές εταιρίες. Η εύρεση ενός τέτοιο ορίου αποτελεί έναν ξεχωριστό κλάδο έρευνας ακόμα και στην εποχή μας όμως παρ'όλα αυτά υπάρχουν κάποιες τεχνικές οι οποίες παρουσιάζουν αρκετές ενδείξεις για την επιλογή ενός αξιόπιστου ανώτατου ορίου. Οι μέθοδοι που θα χρησιμοποιηθούν για την επιλογή του «βέλτιστου» κατωφλίου είναι γραφικές και άρα εκ των πραγμάτων όχι πολύ ακριβείς.

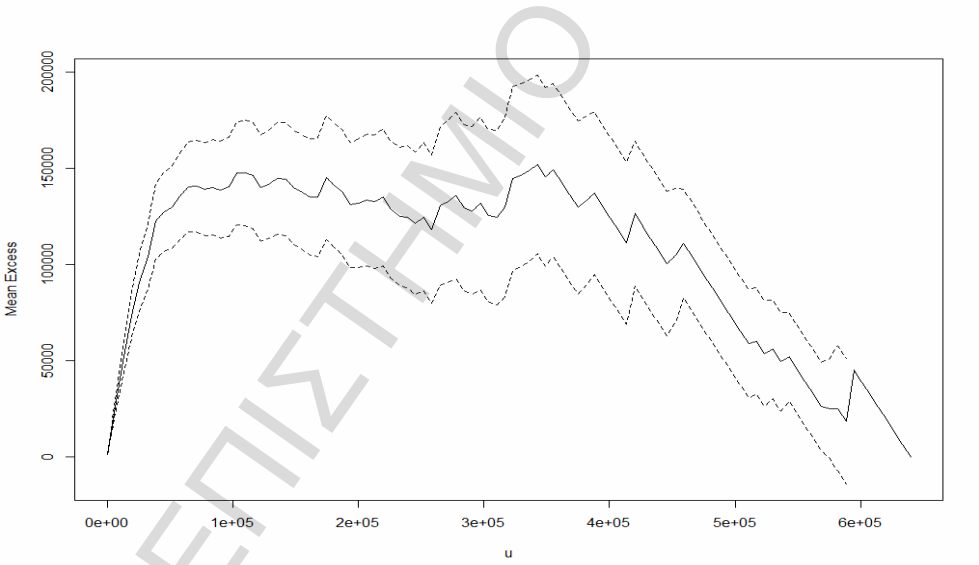
- **$1^{η}$ μέθοδος**

Ένας τρόπος επιλογής του κατάλληλου ανώτατου ορίου είναι μέσω της συνάρτησης μέσης υπερβάλλουσας απώλειας (mean excess loss function) ή όπως αναφέρεται στους ιατρικούς κλάδους ή τους κλάδους της αξιοπιστίας συνάρτηση μέσης υπολειπόμενης ζωής (mean residual life function). Συγκεκριμένα όπως έχει αναφερθεί και στην ενότητα 3.3.2 το κατάλληλο όριο θα βρίσκεται στο σημείο πάνω από το οποίο το γράφημα της συνάρτησης θα παρουσιάζει γραμμικότητα. Τα δύο παρακάτω γραφήματα είναι ισοδύναμα και παρουσιάζουν τη χρήση της συνάρτησης μέσης υπερβάλλουσας απώλειας στα διαθέσιμα δεδομένα.

Σχήμα 5.3.1: Γράφημα της συνάρτησης της μέσης υπερβάλλουσας απώλειας



Σχήμα 5.3.2: Γράφημα της μέσης υπολειπόμενης ζωής

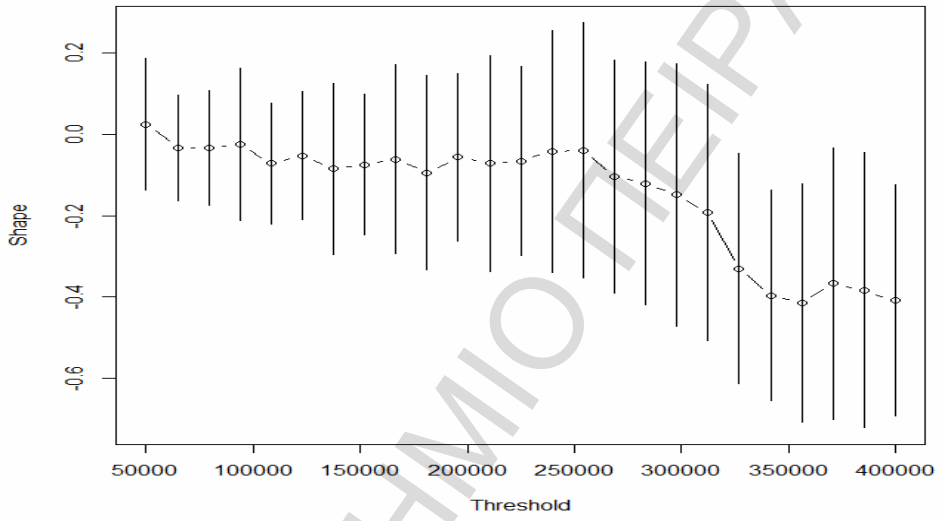


Στο σχήμα 5.3.1 παρουσιάζεται το γράφημα της μέσης υπερβάλλουσας συνάρτησης από το οποίο παρατηρείται ότι η συνάρτηση παρουσιάζει μια γραμμικότητα για όρια από το σημείο των 90.000 ευρώ και πάνω. Από το Σχήμα 5.3.2 παρατηρείται μια γραμμικότητα για τη συνάρτηση της μέσης υπολειπόμενης ζωής για τιμές ορίων από 90.000 έως 250.000 ευρώ. Από τα δύο παραπάνω διαγράμματα φαίνεται ότι ένα όριο $u > 90.000$ ευρώ θα μπορούσε να είναι ένα αξιόπιστο όριο.

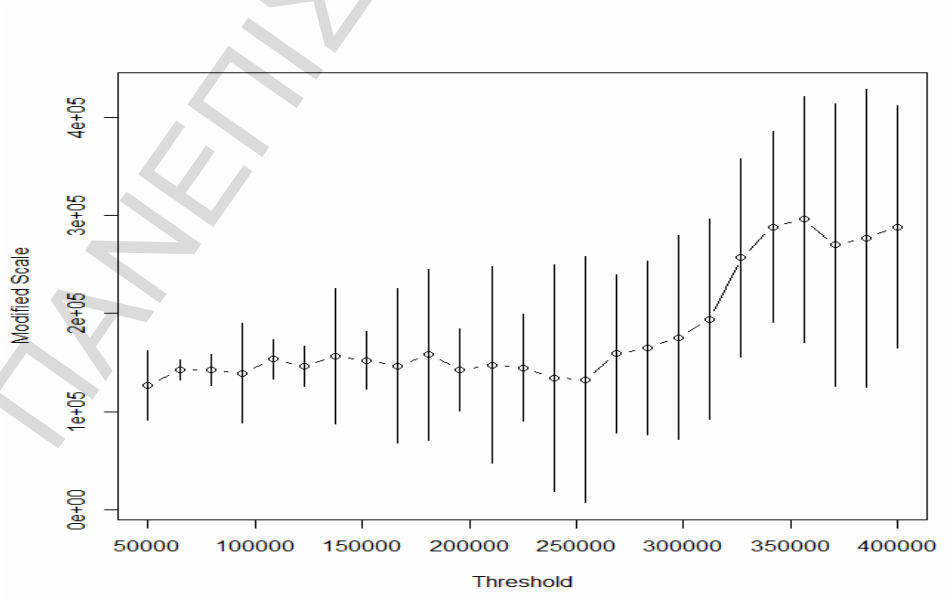
- **2η μέθοδος**

Η συγκεκριμένη μέθοδος για την επιλογή του καταλληλότερου υψηλού κατωφλίου βασίζεται στην εκτίμηση των παραμέτρων $\hat{\sigma}, \xi$ για διάφορες τιμές ορίων. Εφόσον από τη θεωρία είναι γνωστό ότι $F_u \sim GPD$ τότε θα πρέπει η εκτίμηση του $\hat{\sigma}$ να μεταβάλλεται γραμμικά ως προς u ενώ το ξ να μην εξαρτάται από το u. Στα παρακάτω διαγράμματα παρουσιάζονται οι εκτιμήσεις των παραμέτρων της Γενικευμένης κατανομής Pareto για πλήθος κατωφλίων.

Σχήμα 5.3.3: Διάγραμμα της παραμέτρου $\hat{\sigma}$ για πλήθος ανώτατων ορίων



Σχήμα 5.3.4: Διάγραμμα της παραμέτρου ξ για πλήθος ανώτατων ορίων



Στα σχήματα 5.3.3 και 5.3.4 παρατηρείται ότι οι εκτιμήσεις παραμέτρων παραμένουν περίπου σταθερές για όρια από 110.000 έως 250.000 ευρώ. Σύμφωνα και με τις δύο παραπάνω μεθόδους έχουμε αρκετές ενδείξεις για να θεωρήσουμε πως ένα όριο κοντά στις 110.000 ευρώ είναι ένα αρκετά ικανοποιητικό όριο για προσαρμοστεί η GPD στα δεδομένα.

5.3.2 Εκτίμηση των παραμέτρων της GPD

Ύστερα από την επιλογή του καλύτερου κατωφλίου πάνω από το οποίο θα εξεταστούν οι παρατηρήσεις θα πρέπει να εκτιμηθούν και οι παράμετροι της GPD. Η ανάλυση που ακολουθεί αφορά την εκτίμηση των παραμέτρων της GPD πάνω από το όριο που επιλέχθηκε προηγουμένως καθώς και η εκτίμηση των παραμέτρων για την κατανομή GPD για το όριο που χρησιμοποιούν οι περισσότερες ασφαλιστικές για τις «μεγάλες ζημιές» και ανέρχεται στις 150.000 ευρώ. Οι εκτιμήσεις θα γίνουν με τη χρήση του πακέτου POT της R για τη μέθοδο μεγίστης πιθανοφάνειας αν και υπάρχει η δυνατότητα να χρησιμοποιηθεί μεγάλο πλήθος μεθόδων εκτίμησης (βλ. Παράρτημα 2) όπου η επιλογή της βέλτιστης επιλογής μεθόδου ξεφεύγει από τη σκοπιά της παρούσας διπλωματικής εργασίας.

Πινακας 5.3.1: Πίνακας εκτίμησης παραμέτρων της κατανομής GPD με τη μέθοδο μεγίστης πιθανοφάνειας.

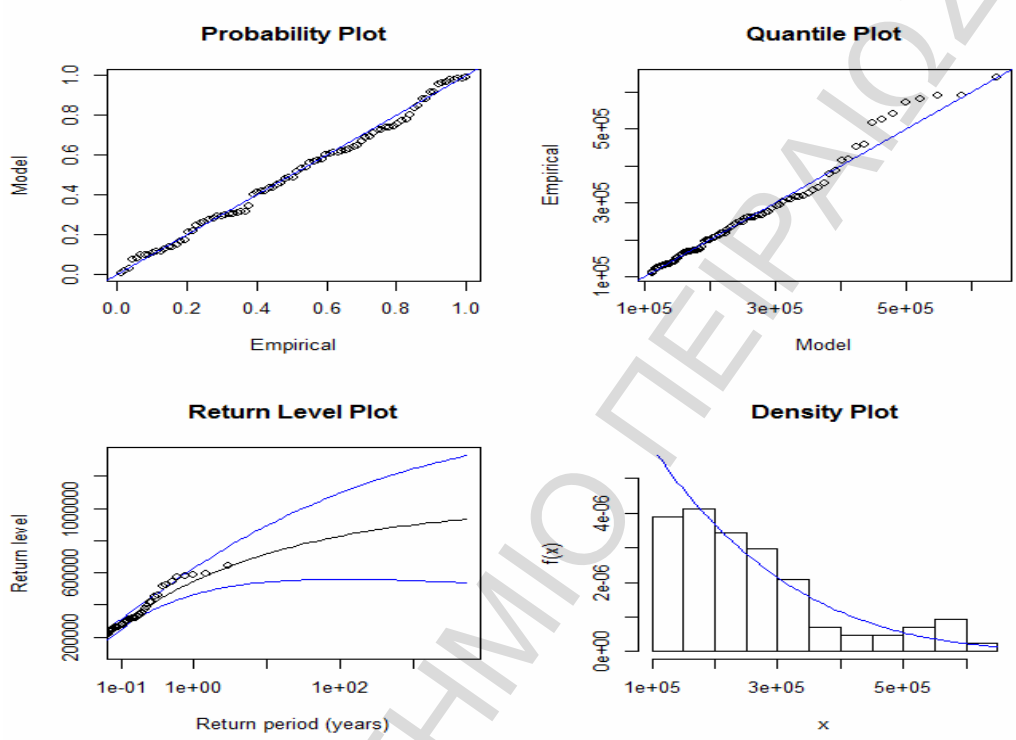
MLE estimation
Threshold Call: 110000
Number Above: 87
Estimates
scale shape
1.474e+05 -7.996e-02
Standard Errors
scale shape
1.887e+03 7.393e-02
Convergence: successful

MLE estimation
Threshold Call: 150000
Number Above: 70
Estimates
scale shape
1.386e+05 -6.411e-02
Standard Errors
scale shape
2.267e+04 1.119e-01
Convergence: successful

Οι παραπάνω πίνακες παρουσιάζουν τις εκτιμήσεις των παραμέτρων για τα δύο διαφορετικά όρια. Ένα σημαντικό στοιχείο το οποίο παρουσιάζεται είναι το γεγονός πως υπάρχει σύγκλιση των δεδομένων στην GPD. Ο έλεγχος της καλής προσαρμογής της

Γενικευμένης Pareto κατανομής μπορεί να γίνει είτε με γραφικές μεθόδου είτε με κάποιο στατιστικό test. Στο παρακάτω σχήμα παρουσιάζονται γραφικές απεικονίσεις καλής προσαρμογής στην GPD για το επιλεγμένο όριο των 110.000 ευρώ.

Σχήμα 5.3.5: Σύνολο γραφημάτων καλής προσαρμογής στην GPD



Το Σχήμα 5.3.5 παρουσιάζει τέσσερα επιμέρους γραφήματα. Τα γραφήματα 1,2,4 παρουσιάζουν τα P-P plot, Q-Q plot και Density plot αντίστοιχα και δείχνουν μια πολύ καλή προσαρμογή των δεδομένων στην GPD. Το γράφημα 3 απεικονίζει το διάγραμμα της στάθμης απόδοσής και φαίνεται πως η μέγιστη αποζημίωση θα ξεπερνάει τις 600.000 ευρώ μια φορά περίπου τα δύο χρόνια. Οι γραφικές απεικονίσεις δεν είναι πάντα αρκετά ακριβείς για αυτό το λόγο θα διενεργηθεί ένα Kolmogorov-Smirnov test και για τα δύο όρια έτσι ώστε να αποφασιστεί αν οι παρατηρήσεις που ξεπερνούν τα δύο όρια ακολουθούν την Γενικευμένη κατανομή Pareto. Στους παρακάτω πίνακες παρουσιάζονται δύο Kolmogorov-Smirnov test για τα δύο όρια ελέγχοντας αν οι παρατηρήσεις οι οποίες ξεπερνούν αυτά τα όρια προέρχονται από την ίδια κατανομή με μια GPD όπου οι παράμετροι της είναι οι ίδιες οι οποίες είχαν εκτιμηθεί προηγουμένως από στον Πίνακα 5.3.1

Πίνακας 5.3.2: Kolmogorov-Smirnov test για τις κατανομές των δεδομένων πάνω από το υψηλό κατώφλι.

Two-sample Kolmogorov-Smirnov test	Two-sample Kolmogorov-Smirnov test
data: simulated_values and over_110000	data: simulated_values and over_150000
D = 0.0638, p-value = 0.8744	D = 0.0715, p-value = 0.8691
alternative hypothesis: two-sided	alternative hypothesis: two-sided

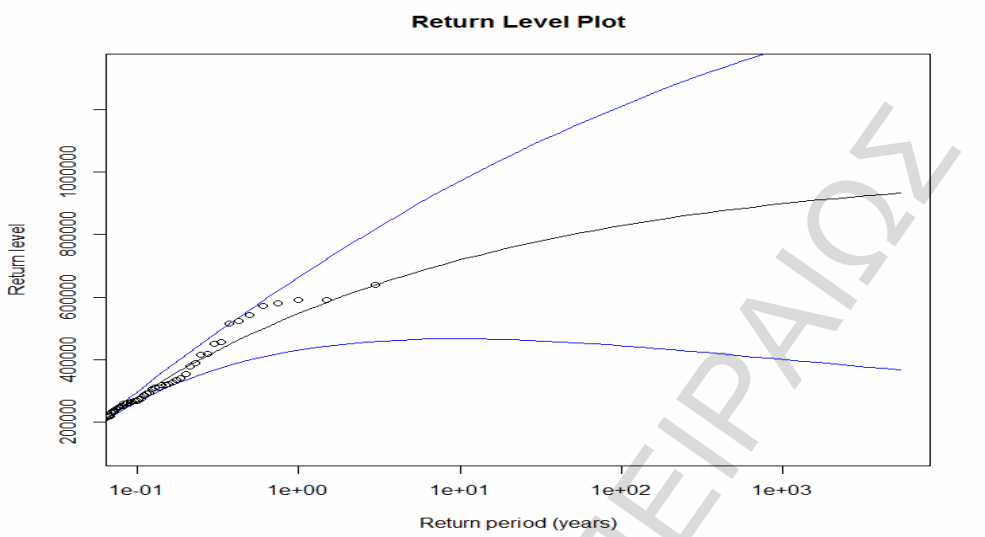
Όπως ήταν αναμενόμενο το Kolmogorov-Smirnov test έδειξε (p-value1 = 0.8744>0.05 , p-value2 = 0.8691>0.05) ότι δεν μπορούμε να απορρίψουμε την υπόθεση ότι οι παρατηρήσεις πάνω από το όριο των 110.000 ευρώ καθώς και από αυτό των 150.000 ακολουθούν την Γενικευμένη κατανομή Pareto με παραμέτρους αυτές που παρουσιάζονται στους πίνακες του σχήματος 5.3.2 σε επίπεδο σημαντικότητας 95%.

5.3.3 Εκτίμηση της στάθμης απόδοσης της GPD

Όπως έχει αναφερθεί και στην προηγούμενη ενότητα ως στάθμη απόδοσης καλείται το κατώφλι το οποίο δεν θα υπερβεί η επόμενη παρατήρηση με πιθανότητα $1-p$. Στην συγκεκριμένη περίπτωση, με την μέθοδο POT, οι παρατηρήσεις δεν χωρίζονται σε μηνιαία μέγιστα όπως στην Block Maxima αλλά περιλαμβάνονται όλες. Το αποτέλεσμα είναι πως η στάθμη απόδοσης αναφέρεται στην πιθανότητα να μην υπερβεί η επόμενη αποζημίωση ένα υψηλό κατώφλι. Η εκτίμηση της R (μέσω του πακέτου extRemes) για αυτό το κατώφλι, θεωρώντας ότι το σύνολο των παρατηρήσεων των τριών ετών είναι 90.000, διαμορφώνεται στις 547.336,1. Δηλαδή η πιθανότητα η επόμενη αποζημίωση να ξεπεράσει τις 547.336,1 είναι ίση με $p = \frac{1}{30.000}$ ή εναλλακτικά, η μέγιστη αποζημίωση θα ξεπεράσει τις 547.336,1 κατά μέσο όρο μια φορά κάθε 30.000 αποζημιώσεις (δηλ. μια φορά το χρόνο).

Στο παρακάτω γράφημα παρουσιάζεται ένα 99,5% διάστημα εμπιστοσύνης για το τη στάθμη απόδοσής

Σχήμα 5.3.6: Γράφημα του 99,5 διαστήματος εμπιστοσύνης για τη στάθμη απόδοσης

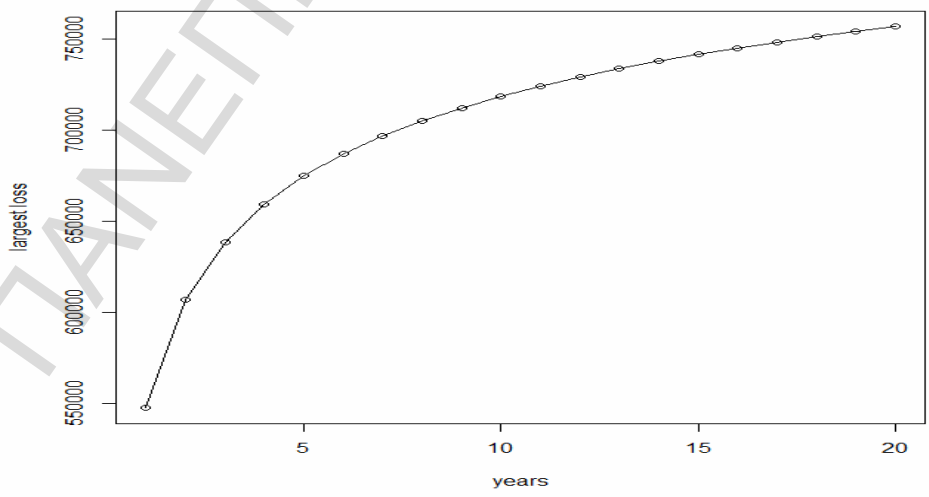


Στο σχήμα 5.3.6 παρουσιάζεται το 99,5 διάστημα εμπιστοσύνης της στάθμης απόδοσης για ένα έτος το οποίο είναι περίπου:

$$\Pr(400000 < \hat{z}_p < 630000) = 0,995$$

Ένα σημαντικό αποτέλεσμα εκτός από την εκτιμώμενη σε κάποιο συγκεκριμένο σημείο στάθμη απόδοσης είναι η διαχρονική εξέλιξη της στάθμης απόδοσης για ένα μεγάλο πλήθος χρονικών περιόδων. Το γράφημα που ακολουθεί εκφράζει ακριβώς αυτή την διαχρονική εξέλιξη.

Σχήμα 5.3.7: Γράφημα της διαχρονικής εξέλιξης της στάθμης απόδοσης



Από το σχήμα 5.3.7 παρατηρείται πως η ανώτερη αποζημίωση θα ξεπεράσει κατά μέσο όρο της 600.000 ευρώ μια φορά κάθε δύο χρόνια, αποτέλεσμα που επιβεβαιώνεται και από τη μέθοδο Block Maxima και συγκεκριμένα από το σχήμα 5.2.6.

5.4 Συμπεράσματα ανάλυσης θεωρίας ακραίων τιμών

Στο παρόν κεφάλαιο που προηγήθηκε έγινε προσπάθεια να αναλυθούν τα δεδομένα με τη χρήση της θεωρίας ακραίων τιμών έτσι ώστε να βγουν αποτελέσματα που αφορούν τις ακραίες παρατηρήσεις. Η ενότητα 5.2 έδωσε βάση στην ανάλυση των δεδομένων με τη χρήση της μεθόδου Block Maxima ενώ η ενότητα 5.3 ασχολήθηκε με την μέθοδο Peak Over Threshold. Τα αποτελέσματα που προέκυψαν από την παραπάνω ανάλυση είναι επαρκώς επιστημονικά τεκμηριωμένα και παρουσιάζουν με απλό και ξεκάθαρο τρόπο την εκτίμηση των ανώτατων αποζημιώσεων που ενδέχεται να καταφθάσουν στην ασφαλιστική εταιρία σε διάφορές μελλοντικές χρονικές περιόδους όπως επίσης αποτελούν χρήσιμα αποτελέσματα για την οριακή κατανομή των αποζημιώσεων πάνω από ένα ορισμένο όριο. Το γεγονός ότι η ανώτερη αποζημίωση που θα προκύψει στην ασφαλιστική επιχείρηση σε μια συγκεκριμένη χρονική περίοδο συμπίπτει και επιβεβαιώνεται και από τις δύο εξεταζόμενες μεθόδους αποτελεί ένα αναμενόμενο πλην όμως ευχάριστο αποτέλεσμα το οποίο συνηγορεί στην αξιοπιστία των μεθόδων που εφαρμόστηκαν.

Κεφάλαιο 6

Συμπεράσματα

6.1 Συμπεράσματα ανάλυσης δεδομένων

Όπως έχει αναφερθεί από την εισαγωγή της παρούσας εργασίας οι ασφαλιστικές εταιρίες έχουν ανάγκη να κρατούν κατάλληλο απόθεμα στα ταμεία τους. Οι λόγοι είναι πολύ απλοί καθώς από τη μια πλευρά πρέπει να είναι προστατευμένες έτσι ώστε να συνεχίζουν να λειτουργούν ακόμα και σε περιπτώσεις που θα προκύψουν απαιτήσεις υψηλότερες από τις συνηθισμένες ενώ συγχρόνως με αυτό τον τρόπο θα νιώθουν πιο ασφαλείς και οι ασφαλισμένοι. Από την άλλη πλευρά πρέπει να πληρούν τα κριτήρια που τους θέτουν οι εποπτικές αρχές αφού, ιδιαίτερα από την 1/1/2014 θα μπει σε εφαρμογή το Περιθώριο Φερεγγυότητας II (Solvency II). Η λύση στο πρόβλημα των εταιριών δεν είναι απλή γιατί το αυξημένο απόθεμα αποτελεί συνήθως ένα περιουσιακό στοιχείο το οποίο δεν μπορεί να επενδυθεί επαρκώς. Ο μόνος σωστός συνεπής και αξιόπιστος τρόπος είναι ο υπολογισμός που στηρίζεται σε επιστημονικά τεκμηριωμένα και εφαρμοσμένα αποτελέσματα.

Στην παρούσα διπλωματική εργασία έγινε προσπάθεια να εφαρμοστούν θεωρητικά υποδείγματα σε ένα πραγματικό χαρτοφυλάκιο ασφάλισης αυτοκινήτων. Η μελέτη που πραγματοποιήθηκε μπορεί να χαρακτηριστεί ότι πέτυχε σε ένα βαθμό το σκοπό της, αφού τα αποτελέσματα που προέκυψαν είναι απλά και εφαρμόσιμα ενώ σε μεγάλο βαθμό προσδίδουν ιδιαιτέρως χρήσιμες πληροφορίες στη λήψη αποφάσεων. Ωστόσο, επισημαίνουμε ότι μία πληρέστερη ανάλυση θα μπορούσε να αξιοποιήσει και άλλα εργαλεία που θα βοηθούσαν σε μια πιο ολοκληρωμένη μελέτη για τον όγκο και τη ροή των αποζημιώσεων (π.χ. χρήση δεδομένων από περισσότερα έτη, ανάλυση με χρήση χρονοσειρών, μελέτη της έκθεσης στον κίνδυνο κλπ).

Στις επόμενες παραγράφους θα περιγραφούν συνοπτικά τα αποτελέσματα που προέκυψαν από την ανάλυση των δεδομένων στα κεφάλαια 4 και 5 με χρήση των πολυάριθμων πακέτων που περιλαμβάνει το πρόγραμμα R.

Το κεφάλαιο 4 αποτελεί σε ένα βαθμό την εφαρμογή μέρους της θεωρίας που παρουσιάστηκε στο κεφάλαιο 2. Αρχικά στην ενότητα 4.2 αναλύθηκε η ποσότητα του αριθμού των αποζημιώσεων που καταφθάνουν σε μια ασφαλιστική εταιρία. Από τα δεδομένα των 3 ετών που υπήρχαν διαθέσιμα έγινε η προσπάθεια να εντοπιστεί αν υπάρχει μια

κατανομή την οποία να ακολουθεί η μεταβλητή του αριθμού των αποζημιώσεων. Τα δεδομένα του αριθμού των μη μηδενικών αποζημιώσεων που υπήρχαν για τα έτη από το 2006 έως και το 2008 ομαδοποιήθηκαν σε μηνιαίες αποζημιώσεις. Αν και οι παρατήσεις ήταν πολύ λίγες για να γίνει αξιόπιστη ανάλυση παρόλα αυτά από το ιστόγραμμα συχνοτήτων φάνηκε πως αντί να γίνει προσπάθεια εύρεσης μιας κατανομής που να προσαρμόζεται καλά στα δεδομένα ίσως θα ήταν προτιμότερο να εξεταστεί η περίπτωση μιας δικόρυφης κατανομής που να προέρχεται από μίξη δύο γνωστών κατανομών. Το αποτέλεσμα ήταν να χωριστούν τα δεδομένα σε δύο διαστήματα και να αναλυθούν ξεχωριστά. Ύστερα από δοκιμές αποφασίστηκε να χωριστούν οι παρατηρήσεις στο σημείο όπου οι μηνιαίες τιμές δεν θα ξεπερνούσαν την τιμή των 2.100 αποζημιώσεων και σε μια άλλη κλάση που θα περιελάμβανε τις υπόλοιπες. Οι γραφικοί και παραμετρικοί έλεγχοι έδειξαν πως οι παρατηρήσεις που δεν ξεπερνούσαν την τιμή 2.100 ακολουθούσαν την κανονική κατανομή με αντίστοιχες παραμέτρους N(1315,89700) και οι παρατηρήσεις που υπερβαίνανε την τιμή 2.100 ακολουθούσαν την μονοπαραμετρική Pareto με αντίστοιχες παραμέτρους Pareto(7.344,2.100). Επόμενο βήμα στη συγκεκριμένη ανάλυση ήταν να γίνει η μίξη των κατανομών και να πραγματοποιηθεί ένα Kolmogorov- Smirnov test για την καλή προσαρμογή των δεδομένων και των δύο κλάσεων με την καινούργια μικτή κατανομή. Τελικά η μικτή κατανομή κατέληξε να είναι η ακόλουθη:

$$f(x) = 0,61 * f_1(x) + 0,39 * f_2(x)$$

όπου $f_1(x)$ είναι η σ.π.π. της κανονικής κατανομής N(1315.18,89700) ενώ η $f_2(x)$ η σ.π.π. είναι της κατανομής Pareto(7.344,2.100). Το p-value του Kolmogorov- Smirnov test υπολογίστηκε σε 0,7 που σημαίνει ότι δεν μπορούμε να απορρίψουμε την υπόθεση πως τα δεδομένα ακολουθούν την παραπάνω μίξη κατανομών.

Η ενότητα 4.3 της παρούσας εργασίας αναφέρεται στην πολύ σημαντική ποσότητα του μεγέθους των αποζημιώσεων. Λόγω της μεγάλης συγκέντρωσης των παρατηρήσεων σε τιμές κοντά στο μηδέν αποφασίστηκε να περικοπούν τα δεδομένα σε κάποια τιμή και να εξεταστούν οι υπερβάλλουσες αυτήν την τιμή παρατηρήσεις. Το σημείο που επιλέχθηκε ύστερα από δοκιμές είναι το ποσό των 25.000 ευρώ το οποίο αφαιρέθηκε από όλες τις παρατηρήσεις και εξετάστηκαν φυσικά μόνο οι θετικές τιμές. Από την ανάλυση που ακολούθησε και ύστερα από γραφικούς αλλά και παραμετρικούς ελέγχους, οι εναπομείνασες

τιμές των 279 παρατηρήσεων ακολουθούν κατανομή Γάμμα με τις αντίστοιχες παραμέτρους Gamma(0.519 , 170125).

Στην ενότητα 4.4 πραγματοποιήθηκε μια χρονολογική ανάλυση των παραπάνω ποσοτήτων από την οποία φάνηκε πως ενώ ο αριθμός των αποζημιώσεων τριπλασιάστηκε το ποσό των αποζημιώσεων διπλασιάστηκε. Το γεγονός αυτό δείχνει πως αν και δεν υπήρχαν διαθέσιμα τα στοιχεία του αριθμού των ασφαλισμένων φαίνεται να αυξήθηκε το μερίδιο αγοράς της εταιρίας ενώ ταυτόχρονα έγινε και καλύτερη επιλογή των ασφαλισμένων ως προς την ικανότητα να φέρνουν μικρότερες ζημιές.

Το κεφάλαιο 5 στηρίχθηκε στα όσα θεωρητικά αναπτύχθηκαν στο κεφάλαιο **3** και αφορά την ανάλυση των δεδομένων με τη χρήση της θεωρίας ακραίων τιμών. Τα συγκεκριμένα κεφάλαια αποτελούν τα σημαντικότερα στοιχεία της συγκεκριμένης εργασίας καθώς επικεντρώνονται μόνο στις πολύ ακραίες αποζημιώσεις ενώ ο τρόπος υπολογισμού των αποτελεσμάτων στηρίζεται σε συγκεκριμένα κριτήρια και όχι σε κάποιους μετασχηματισμούς που επηρεάζονται από την ποιότητα και την ποσότητα των ασφαλισμένων.

Η ενότητα 5.2 περιγράφει την μέθοδο Block Maxima η οποία χρησιμοποιεί τη μέγιστη παρατήρηση κάθε χρονικού διαστήματος έτσι ώστε το σύνολο των μεγίστων να ακολουθεί την κατανομή GEV. Συγκεκριμένα για την εφαρμογή αυτής της μεθόδου χωρίστηκαν τα δεδομένα σε μηνιαίες κλάσεις και μοντελοποιήθηκαν σύμφωνα με την κατανομή GEV. Το αποτέλεσμα αυτής της μεθόδου ήταν ότι η μέγιστη αποζημίωση που θα καταφθάσει στην ασφαλιστική επιχείρηση θα ξεπεράσει το ποσό των 550.000 ευρώ μια φορά το χρόνο ενώ το ποσό των το ποσό των 600.000 ευρώ σε περίοδο δύο. Επίσης υπολογίστηκαν και τα αντίστοιχα διαστήματα εμπιστοσύνης για το μέγεθος της μέγιστης αποζημίωσης.

Στην ενότητα 5.3 η μέθοδος που εξετάστηκε ήταν η Peak Over Threshold. Η συγκεκριμένη μέθοδος αποτελεί μια βελτιωμένη προσέγγιση για την εκτίμηση της μέγιστης αποζημίωσης αφού μοντελοποιεί όλες τις παρατηρήσεις οι οποίες ξεπερνούν ένα συγκεκριμένο υψηλό κατώφλι. Το πρώτο βήμα της ανάλυσης ήταν να καθοριστεί ποιο ήταν το συγκεκριμένο κατώφλι το οποίο αν ξεπερνούσαν οι αποζημιώσεις θα θεωρούνταν ακραίες. Σύμφωνα με δύο γραφικές μεθόδους που χρησιμοποιήθηκαν το όριο πάνω από το οποίο μια παρατήρηση θεωρείται ακραία και επομένως περιλαμβάνεται στην ανάλυση είναι οι 110.000 ευρώ. Επόμενο βήμα ήταν να εκτιμηθούν οι παράμετροι της οριακής κατανομής των μεγίστων (κατανομή GPD) καθώς και το ποσό που δε θα ξεπεράσει η μέγιστη αποζημίωση για μια σειρά χρονικών περιόδων. Το αποτέλεσμα που προέκυψε από την εφαρμογή της μεθόδου

POT ήταν πως η μέγιστη ετήσια αποζημίωση θα ξεπεράσει το ποσό των 547.000 ευρώ μια φορά το χρόνο ενώ για χρονική περίοδο δύο ετών τις 600.000 ευρώ αντίστοιχα.

Τέλος, αξίζει να παρατηρήσουμε πως αν και οι μέθοδοι που χρησιμοποιήθηκαν στο κεφάλαιο 5 για τον υπολογισμό της μέγιστης αποζημίωσης διέφεραν, εντούτοις τα αποτελέσματα που προέκυψαν συμφωνούν μεταξύ τους γεγονός που ενισχύει την αξιοπιστία των εξεταζόμενων μεθόδων.

Παράρτημα

Παράρτημα 1

Παρακάτω παρατίθενται συνοπτικά ο κώδικας στην R που χρησιμοποιήθηκε για την δημιουργία των γραφημάτων, τον υπολογισμό των εκτιμήσεων των παραμέτρων και της στάθμης απόδοσής για την μέθοδο Block Maxima τα οποία παρουσιάστηκαν αναλυτικά στην ενότητα 5.2 της παρούσας εργασίας. Συγκεκριμένα παρατίθενται τρείς πανομοιότυποι τρόποι από τρία διαφορετικά πακέτα της R.

Κώδικας 1 (1^{ος} τρόπος)

```
## method Block Maxima
library(extRemes)

l=na.omit(Max.loss.per.month)
k=36; m=1; BlockMaxima=rep(0,k)
for(i in 1:k){ BlockMaxima [i]=max(Max.loss.per.month[((i-1)*m+1):(i*m)])}
plot(BlockMaxima,type="h")
##gev qqplot
xi<--0.2; k<-length(BM);
plot(sort(BlockMaxima,na.last=TRUE),xlab="GEV Q-Q Plot",-1/xi*(1-(-
log((1:k)/(k+1)))^(-xi)))

##parameter estimation
t=gev.fit(BM)
gev.profxi(t,-0.27,-0.12)
gev.diag(t)

##estimation of level plot for 1 year
mu=t$mle[1]; sigma=t$mle[2]; xi=t$mle[3]
p=1/12;w=-log(-log(1-p));
zp=mu-sigma/xi*(1-exp(xi*w));
print(zp)
```

```
rl=return.level(t, conf = 0.005, rperiods= c(2,3,6,9,12,15,18,21,24), make.plot = TRUE)

p=1/1:60;w=-log(-log(1-p));
zp=mu-sigma/xi*(1-exp(xi*w));
plot(1/p,zp,type="b")
```

Κώδικας 2 (2ος τρόπος)

```
library(fExtremes)
## Block Maxima {fextremes}
l=na.omit(Max.loss.per.month)
blockMaxima(l, block=1,doplot=TRUE)
summary(l)
##parameter estimation
gF1=gevFit (Max.loss.per.month, block=1)
gF1

##estimation of level plot and confidence intervals
gevrlevelPlot(gF1, kBlocks = 12, ci = 0.995)
gevrlevelPlot(gF1, kBlocks = 6, ci = 0.995)
gevrlevelPlot(gF1, kBlocks = 36, ci = 0.995)
```

Κώδικας 3 (3ος τρόπος)

```
Library(evir)
##parameter estimation
c=gev(l,block=1)
plot.gev(c)
##estimation of level plot and confidence intervals
rlevel.gev(c,k.blocks=12,add=FALSE)
```

Κώδικας 4 (Έλεγχος καλής προσαρμογής)

```
s=gevSim(model = list(xi = -0.18, mu = 291044, beta = 130773), n = 36, seed = NULL)
ks.test(Max.loss.per.month,s)
```

Παράρτημα 2

Παρακάτω παρατίθενται συνοπτικά ο κώδικας στην R που χρησιμοποιήθηκε για την δημιουργία των γραφημάτων, τον υπολογισμό των εκτιμήσεων των παραμέτρων και της στάθμης απόδοσής για την μέθοδο Peak Over Threshold τα οποία παρουσιάστηκαν αναλυτικά στην ενότητα 5.3 της παρούσας εργασίας. Συγκεκριμένα παρατίθενται δύο πανομοιότυποι τρόποι από τρία διαφορετικά πακέτα της R.

Κώδικας 1 (1ος τρόπος)

```
Library(extRemes)
sd=sort(IncurredLoss);k=length(sd);mrlp=rep(0,k)
for(i in 1:k){mrlp[i]=sum(sd[i:k])/(k-i+1)}

##threshold selection
##mean residual life plots
plot(sd,mrlp)##1st plot
mrl.plot(IncurredLoss)##2nd plot
gpd.fitrange(IncurredLoss,118000,120000,nint=4,show=TRUE)

library(ismev)
##parameter estimation
##estimation method "Nelder-Mead"
b=gpd.fit(IncurredLoss,110000,npy=30000)
```

```
##estimation of return level
u=110000
n=89779
b=gpd.fit(IncurredLoss,u,npy=30000)
gpd.diag(b)
gpd.profxi(b,-0.22,-0.17,conf=0.995)
sigma2=b$mle[1];xi2=b$mle[2]
k=b$nexc;m=30000*1:50;
xm=u+sigma2/xi2*((m*k/n)^xi2-1)
xm
plot(m/30000,xm,type="o")
```

Κώδικας 2 (2^{ος} τρόπος)

```
## mean residual life plot{POT}
mrlplot(IncurredLoss,u.range=c(0,400000),nt = 100)
##threshold selection
tcplot(IncurredLoss,u.range=c(50000,400000))

##parameter estimation (with various methods)
a=fitgpd(IncurredLoss,110000, est = "mle")
a
fitgpd(IncurredLoss,110000, est = "moments")
fitgpd(IncurredLoss,110000, est = "pickands")
##estimation of level plot and confidence intervals
t=rlplot(b, ci = 0.005, add.ci = TRUE)
t
```

Κώδικας 3 (Έλεγχος καλής προσαρμογής)

```
##kolmogorov smirnov for GPD parameters

##u=110000
over_110000=IncurredLoss[IncurredLoss>110000]
length(over_110000)
simulated_values=rgpd(10000, loc = 110000, scale = 147200, shape =-0.08 )
ks.test(simulated_values,over_110000)


##u=150000
over_150000=IncurredLoss[IncurredLoss>150000]
length(over_150000)
simulated_values=rgpd(10000, loc = 150000, scale = 138600, shape =-0.064 )
ks.test(simulated_values,over_150000)
```

Βιβλιογραφία

Ελληνική βιβλιογραφία

1. Αντζουλάκος Δ. (2009), *Ανάλυση Δεδομένων με τη Χρήση Στατιστικών Πακέτων Εισαγωγή στο R.* Σημειώσεις παραδόσεων. Πανεπιστήμιο Πειραιώς ΠΜΣ στην Εφαρμοσμένη Στατιστική.

2. Ηλιόπουλος Γ. (2006), *Βασικές μέθοδοι εκτίμησης παραμέτρων*. Εκδόσεις Σταμούλης.

3. Κατσάπας Λ. (2010), *Θεωρία ακραίων τιμών σε μοντέλα εξαρτημένων τυχαίων μεταβλητών*, Διπλωματική εργασία για το ΠΜΣ «Αναλογιστική Επιστήμη και Διοικητική Κινδύνου» Πανεπιστήμιο Πειραιώς.

4. Κούτρας Μ. (2004*), Εισαγωγή στις Πιθανότητες* (Μέρος Ι). Εκδόσεις Σταμούλης.

5. Μπούτσικας Μ. (2008), *Σημειώσεις διαλέξεων στη «Θεωρία Ακραίων τιμών»*. ΠΜΣ στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου, (Τμ. Στατιστικής & Ασφ. Επιστ., Πανεπ. Πειραιώς).

6. Πανάρετος Ι. & Ξεκαλάκη Ε.,(2000), *Εισαγωγή στη στατιστική σκέψη, τόμος ΙΙ, Εισαγωγή στις Πιθανότητες και στην στατιστική συμπερασματολογία*. Πανάρετος Ι.

7. Παρτσακουλάκης Β. (2012), *Στατιστική ανάλυση και προβλέψεις σε περιβαλλοντολογικά μοντέλα με τη χρήση της Θεωρίας Ακραίων τιμών*, Διπλωματική εργασία για το ΠΜΣ «Εφαρμοσμένη Στατιστική» Πανεπιστήμιο Πειραιώς.

8. Φωκιανός Κ. & Χαραλάμπους Χ. (2010), *Εισαγωγή στην R Πρόχειρες Σημειώσεις*. Τμ .Μαθηματικών & Στατιστικής, Πανεπιστήμιο Κύπρου.

9. Χατζηκωνσταντινίδης Ε. (2010), Σημειώσεις στο μάθημα Θεωρία Κινδύνου Ι. Π.Μ.Σ. Αναλογιστική Επιστήμη και Διοικητική Κινδύνου, Πανεπιστήμιο Πειραιώς.

Ξενόγλωσση βιβλιογραφία

1. Burnecki, Krzysztof, Misiorek, Adam and Weron, Rafal (2010), Loss Distributions. *Munich Personal RePEc Archive* http://mpra.ub.uni-muenchen.de/id/eprint/22163

2. Coles S. (2001) *An introduction to statistical modeling of extreme values.* Springer-Verlag.

3. Embrechts P, Kluppelberg C. and Mikosh T. (1996*), Modeling Extremal Events for Insurance and Finance.* Springer-Verlag

4. Gilleland E., Katz R., and Greg Young G. (2013), Extreme value toolkit. Package 'extRemes'. http://cran.r-project.org/web/packages/extRemes/extRemes.pdf

5. Katz, L. (1965), Unified treatment of a broad class of discrete probability distributions. In "Classical and Contagious Discrete Distributions". *Pergamon Press, Oxford, 175–182*

6. Klugman S., Panjer H., Wilmot (1998), *Loss Models: From Data to Decis*ions. John Wiley and Sons.

7. Lee Wo-Chiang (2009), *Applying Generalized Pareto Distribution to the Risk Management of Commerce Fire Insurance.*

http://ir.lib.au.edu.tw/bitstream/987654321/2656/1/CF07-conf.2009_li_01.pdf

8. McNeil A. (1996), Estimating the Tails of Loss Severity Distributions using Extreme Value Theory. *ASTINBULLETIN, Vol 27. No I 1997 pp 117-137.*

9. Panjer H. (2006*) Operational Risk: Modeling analytics*. John Wiley and Sons.

10. Reiss R. and Thomas M. (2001), *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and other Fields*. Birkhauser

11. Ribatet M. (2007), *POT:* Modeling Peaks Over a Threshold. *The Newsletter of the R Project volume 7.*

12. Ribatet M. (2011), *A User's Guide to the POT Package (Version 1.4*). http://cran.r-project.org/web/packages/POT/vignettes/POT.pdf.

13. Ribatet M. (2013), Generalized Pareto Distribution and Peaks Over Threshold: Package 'POT'. http://cran.r-project.org/web/packages/POT/POT.pdf

14. Ricci V. (2005), Fitting Distributions with R. http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

15. Scarrott C. and MacDonald A. (2012), A Review of Extreme Value Threshold Estimation and Uncertainty Quantificatio. REVSTAT-*Statistical Journal, volume 10, Number 1, March 2012, 33-60*

16. Wuertz D. (2013), *Rmetrics - Extreme Financial Market Data*: Package 'fExtremes'. http://cran.rproject.org/web/packages/fExtremes/fExtremes.pdf