

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

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

**ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΟΥ ΥΨΟΥΣ  
ΑΤΟΜΙΚΗΣ ΖΗΜΙΑΣ ΜΕΣΩ SPLICED  
ΚΑΤΑΝΟΜΩΝ**

**Μιχαήλ Στράτος**

**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  
Ιούλιος  
2021**

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

Τα μέλη της Επιτροπής ήταν:

- Ευστάθιος Χατζηκωνσταντινίδης (Επιβλέπων)
- Γεώργιος Τζαβελάς
- Χαράλαμπος Ευαγγελάρας

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

**UNIVERSITY OF PIRAEUS**



**DEPARTMENT OF STATISTICS AND INSURANCE  
SCIENCE**

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

**MODELLING OF LOSS PAYMENTS  
WITH SPLICED DISTRIBUTIONS**

**Michail Stratos**

**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 in Actuarial  
Science and Risk Management

**Piraeus  
July  
2021**



## **ΕΥΧΑΡΙΣΤΙΕΣ**

Με αφορμή την ολοκλήρωση της παρούσας διπλωματικής εργασίας θα ήθελα να ευχαριστήσω τους καθηγητές αλλά και όλο το προσωπικό του Πανεπιστημίου Πειραιώς που σε μία τόσο δύσκολη και πρωτόγνωρη κατάσταση κατέβαλλε κάθε δυνατή προσπάθεια με σκοπό την μετάδοση της γνώσης προς τους φοιτητές και την ομαλή ολοκλήρωση του μεταπτυχιακού προγράμματος σπουδών.

## ΕΙΣΑΓΩΓΗ

Σε ορισμένες περιπτώσεις κατά την μοντελοποίηση των ατομικών ζημιών ενός χαρτοφυλακίου παρατηρούμε ότι σε κάποια διαστήματα τιμών οι ζημιές παρουσιάζουν υψηλή συχνότητα - χαμηλή σφοδρότητα ενώ σε άλλα παρουσιάζουν χαμηλή συχνότητα - υψηλή σφοδρότητα. Με στόχο να προταθεί μία αποτελεσματική πρακτική για την μοντελοποίηση της ιδιάζουσας αυτής συμπεριφοράς των δεδομένων παρουσιάζεται η έννοια της συγκολλημένης κατανομής, δηλαδή ενός πιθανοθεωρητικού μοντέλου που προκύπτει από τον συνδυασμό  $n$  – κατανομών βαριάς ουράς,  $n \geq 2$ , και εφαρμόζεται σε περιπτώσεις όπου οι συνήθεις ζημιοκατανομές δεν προσφέρουν ποιοτικά αποτελέσματα. Ο αρχικός ορισμός εισήχθη από τους Cooray – Ananda (2005) αλλά η έρευνα πάνω στις συγκολλημένες κατανομές στηρίζεται στον γενικότερο ορισμό που δόθηκε από τον Scollnik (2007). Θα παρουσιασθεί ο ορισμός των νέων μοντέλων και θα μελετηθεί η περίπτωση των δύο συνιστωσών ( $n = 2$ ). Έπειτα, θα αναλυθούν οι συγκολλημένες κατανομές Lognormal – Pareto (Scollnik, 2007), Lognormal – GPD (Scollnik, 2007), Weibull – GPD (Scollnik – Sun, 2012) και Gamma – GPD (Teodorescu – Vernic, 2013). Θα δοθούν αναλυτικοί τύποι για τον υπολογισμό των ροπών καθώς και για τα βασικά μέτρα κινδύνου. Για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας των παραμέτρων θα χρησιμοποιηθεί το στατιστικό πακέτο R ενώ στην συνέχεια θα πραγματοποιηθεί μελέτη προσομοίωσης με σκοπό να μετρηθεί η απόδοσή τους. Τέλος, θα παρουσιασθούν δύο εφαρμογές των συγκολλημένων κατανομών σε πραγματικά δεδομένα.

## ABSTRACT

In some cases, during the modelling process of the loss payments we observe that in some ranges the loss amounts seem to display high frequency – low severity while in other ranges they show low frequency – high severity. In order to propose a proper way to model the claims special behaviour we present the concept of spliced distributions, which means a probabilistic model that derives from the combination of  $n$  – fat tail distributions,  $n \geq 2$ , and it is applied in situations that the common loss distributions fail to provide the expected outcomes. The initial definition was given by Cooray – Ananda (2005) but the research regarding the spliced distributions is based on the more general definition of Scollnik (2007). The definition of the new models will be presented and the case of two components will be studied ( $n = 2$ ). Thereafter, the spliced distributions Lognormal – Pareto (Scollnik, 2007), Lognormal – GPD (Scollnik, 2007), Weibull – GPD (Scollnik – Sun, 2012) and Gamma – GPD (Teodorescu – Vernic, 2013) will be analyzed. The analytical expressions for the calculation of the moments as well as the basic risk measures will be given. For the derivation of the maximum likelihood estimators the statistical software R will be used. Also, a simulation study will be conducted in order to assess the estimator's quality. Finally, two applications of spliced distributions on real data sets will be presented.

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ</b> .....	1
<b>ΚΕΦΑΛΑΙΟ 2: ΤΟ ΜΟΝΤΕΛΟ ΤΗΣ ΣΥΓΚΟΛΛΗΜΕΝΗΣ ΚΑΤΑΝΟΜΗΣ</b> .....	3
2.1 Συγκολλημένη κατανομή n-συνιστωσών .....	3
2.2 Συγκολλημένη κατανομή 2-συνιστωσών .....	5
<b>ΚΕΦΑΛΑΙΟ 3: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ LOGNORMAL – PARETO (<math>\alpha, \theta, \sigma, \mu, r</math>)</b> .	11
3.1 Εισαγωγή .....	11
3.2 Χαρακτηριστικά μεγέθη της κατανομής $LP(\alpha, \theta, \sigma, \mu, r)$ .....	15
3.3 Αριθμητική εφαρμογή των $VaRp - TVaRp$ .....	25
3.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $LP(\alpha, \theta, \sigma, \mu, r)$ .....	28
<b>ΚΕΦΑΛΑΙΟ 4: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ LOGNORMAL – GPD (<math>\alpha, \theta, \lambda, \sigma, \mu, r</math>)</b> .....	33
4.1 Εισαγωγή .....	33
4.2 Χαρακτηριστικά μεγέθη της κατανομής $LGPD(\alpha, \theta, \lambda, \sigma, \mu, r)$ .....	35
4.3 Αριθμητική εφαρμογή για σύγκριση $VaRp - TVaRp$ .....	44
4.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $LGPD(\alpha, \theta, \lambda, \sigma, \mu, r)$ .....	46
<b>ΚΕΦΑΛΑΙΟ 5: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ GAMMA – GPD (<math>\alpha, \beta, \gamma, \delta, \theta, r</math>)</b> .....	50
5.1 Εισαγωγή .....	50
5.2 Χαρακτηριστικά μεγέθη της κατανομής $GGPD(\alpha, \beta, \gamma, \delta, \theta, r)$ .....	51
5.3 Αριθμητική εφαρμογή για σύγκριση $VaRp - TVaRp$ .....	59
5.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $GGPD(\alpha, \beta, \gamma, \delta, \theta, r)$ .....	61
<b>ΚΕΦΑΛΑΙΟ 6: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ WEIBULL – GPD (<math>\alpha, \theta, \lambda, \tau, \varphi, r</math>)</b> .....	65
6.1 Εισαγωγή .....	65
6.2 Χαρακτηριστικά μεγέθη της κατανομής $WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$ .....	66
6.3 Αριθμητική εφαρμογή για σύγκριση $VaRp - TVaRp$ .....	77
6.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$ .....	79
<b>ΚΕΦΑΛΑΙΟ 7: ΜΕΛΕΤΗ ΠΡΟΣΟΜΟΙΩΣΗΣ ΓΙΑ ΤΙΣ ΣΥΓΚΟΛΛΗΜΕΝΕΣ ΚΑΤΑΝΟΜΕΣ</b> .....	83



7.1 Σχεδιασμός προσομοίωσης.....	83
7.2 Μελέτη προσομοίωσης για την συγκολλημένη $LP(\alpha, \theta, \sigma)$ .....	86
7.3 Μελέτη προσομοίωσης για την συγκολλημένη $LGPD(\alpha, \theta, \lambda, \sigma)$ .....	87
7.4 Μελέτη προσομοίωσης για την συγκολλημένη $GGPD(\beta, \gamma, \delta, \theta)$ .....	89
7.5 Μελέτη προσομοίωσης για την συγκολλημένη $WGPD(\alpha, \theta, \lambda, \tau)$ .....	90
<b>ΚΕΦΑΛΑΙΟ 8: ΕΦΑΡΜΟΓΗ ΣΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ</b> .....	92
8.1 Εισαγωγή.....	92
8.2 Προσαρμογή στο σετ δεδομένων “beaonre” .....	93
8.3 Προσαρμογή στο σετ δεδομένων “lossalae” .....	97
<b>ΚΕΦΑΛΑΙΟ 9: ΣΥΜΠΕΡΑΣΜΑΤΑ</b> .....	102
<b>ΠΑΡΑΡΤΗΜΑ</b> .....	103
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b> .....	126

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

<b>Σχήμα 3.1</b> Ενδεικτική γραφική παράσταση των κατανομών LP, Lognormal, Pareto.....	15
<b>Σχήμα 3.2</b> Συναρτήσεις πυκνότητας της κατανομής LP( $\alpha, \theta, \sigma$ ).....	18
<b>Σχήμα 3.3</b> 3-Δ και 2-Δ απεικόνιση της LP( $\alpha, 2, 0.5$ ) για επιλεγμένες τιμές της παραμέτρου $\alpha$ .....	19
<b>Σχήμα 3.4</b> Μέση τιμή και Διασπορά της κατανομής LP ( $\alpha, 5, \sigma$ ) στις 3 διαστάσεις.....	22
<b>Σχήμα 3.5</b> Λοξότητα και Κύρτωση της κατανομής LP ( $\alpha, 5, \sigma$ ) στις 3 διαστάσεις.....	22
<b>Σχήμα 3.6</b> Διάγραμμα σύγκρισης των VaR <sub>p</sub> – TVaR <sub>p</sub> για τα δύο σείτ παραμέτρων.....	28
<b>Σχήμα 4.1</b> Ενδεικτική γραφική παράσταση των κατανομών LGPD, Lognormal, GPD.....	34
<b>Σχήμα 4.2</b> 3-Δ και 2-Δ απεικόνιση της LGPD( $\alpha, 2, 0.5, 0.5$ ) για επιλεγμένες τιμές της παραμέτρου $\alpha$ .....	37
<b>Σχήμα 4.3</b> Συναρτήσεις πυκνότητας της LGPD ( $\alpha, \theta, \lambda, \sigma$ ).....	38
<b>Σχήμα 4.4</b> Λοξότητα και Κύρτωση της κατανομής LGPD ( $5, \theta, \lambda, 0.5$ ) στις 3 διαστάσεις .....	41
<b>Σχήμα 4.5</b> Μέση τιμή και Διασπορά της κατανομής LGPD ( $5, \theta, \lambda, 0.5$ ) στις 3 διαστάσεις.....	41
<b>Σχήμα 4.6</b> Διαγράμματα σύγκρισης των VaR <sub>p</sub> για τις κατανομές LP – LGPD.....	45
<b>Σχήμα 4.7</b> Διαγράμματα σύγκρισης των TVaR <sub>p</sub> για τις κατανομές LP – LGPD.....	46
<b>Σχήμα 5.1</b> Ενδεικτική γραφική παράσταση των GGPD, Gamma, GPD.....	51
<b>Σχήμα 5.2</b> 3-Δ και 2-Δ απεικόνιση της GGPD( $\beta, 20, 2, 5$ ) για επιλεγμένες τιμές της παραμέτρου $\beta$ .....	53
<b>Σχήμα 5.3</b> Συναρτήσεις πυκνότητας της κατανομής GGPD( $\beta, \gamma, \delta, \theta$ ).....	54
<b>Σχήμα 5.4</b> Λοξότητα και Κύρτωση της κατανομής GGPD( $\beta, 2, \delta, 3$ ) στις 3 διαστάσεις.....	56
<b>Σχήμα 5.5</b> Μέση τιμή και Διασπορά της κατανομής GGPD( $\beta, 2, \delta, 3$ ) στις 3 διαστάσεις.....	57
<b>Σχήμα 5.6</b> Διάγραμμα σύγκρισης των VaR <sub>p</sub> για τις GGPD, Gamma, GPD.....	60
<b>Σχήμα 5.7</b> Διάγραμμα σύγκρισης των TVaR <sub>p</sub> για τις GGPD, Gamma, GPD.....	61
<b>Σχήμα 6.1</b> Ενδεικτική γραφική παράσταση των πυκνοτήτων των WGPD, Weibull, GPD.....	66
<b>Σχήμα 6.2</b> 3-Δ και 2-Δ απεικόνιση της WGPD( $2, \theta, 10, 3$ ) για επιλεγμένες τιμές της παραμέτρου $\theta$ .....	69
<b>Σχήμα 6.3</b> Συναρτήσεις πυκνότητας της WGPD( $\alpha, \theta, \lambda, \tau$ ).....	70
<b>Σχήμα 6.4</b> Λοξότητα και Κύρτωση της κατανομής WGPD( $5, \theta, 10, \tau$ ) στις 3 διαστάσεις.....	73
<b>Σχήμα 6.5</b> Μέση τιμή και Διασπορά της κατανομής WGPD( $5, \theta, 10, \tau$ ) στις 3 διαστάσεις.....	74

<b>Σχήμα 6.6</b> Διαγράμματα σύγκρισης των $\text{VaR}_p$ για τα τέσσερα σεντ παραμέτρων.....	78
<b>Σχήμα 6.7</b> Διαγράμματα σύγκρισης των $\text{TVaR}_p$ για τα τέσσερα σεντ παραμέτρων.....	79
<b>Σχήμα 7.1</b> Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους $\alpha, \theta, \sigma$ .....	86
<b>Σχήμα 7.2</b> Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους $\alpha, \theta, \lambda, \sigma$ .....	88
<b>Σχήμα 7.3</b> Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους $\beta, \gamma, \delta, \theta$ .....	89
<b>Σχήμα 7.4</b> Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους $\alpha, \theta, \lambda, \tau$ .....	91
<b>Σχήμα 8.1</b> Ιστόγραμμα των δεδομένων “beaonre”.....	94
<b>Σχήμα 8.2</b> Q–Q plot των δεδομένων έναντι της κατανομής WGPD.....	97
<b>Σχήμα 8.3</b> Ιστόγραμμα των δεδομένων “lossalae”.....	98
<b>Σχήμα 8.4</b> Q–Q plot των δεδομένων έναντι της κατανομής LGPD.....	101

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

<b>Πίνακας 3.1</b> Αριθμητικά αποτελέσματα των $VaR_p$ ανά επίπεδο σημαντικότητας για τα δύο σετ παραμέτρων των κατανομών LP, Lognormal, Pareto.....	26
<b>Πίνακας 3.2</b> Αριθμητικά αποτελέσματα των $TVaR_p$ ανά επίπεδο σημαντικότητας για τα δύο σετ παραμέτρων των κατανομών LP, Lognormal, Pareto.....	27
<b>Πίνακας 7.1</b> Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων $\alpha, \theta, \sigma$ για επιλεγμένες τιμές του $n$ .....	87
<b>Πίνακας 7.2</b> Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων $\alpha, \theta, \lambda, \sigma$ για επιλεγμένες τιμές του $n$ .....	88
<b>Πίνακας 7.3</b> Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων $\beta, \gamma, \delta, \theta$ για επιλεγμένες τιμές του $n$ .....	90
<b>Πίνακας 7.4</b> Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων $\alpha, \theta, \lambda, \tau$ για επιλεγμένες τιμές του $n$ .....	91
<b>Πίνακας 8.1</b> Περιγραφικά μέτρα των δεδομένων “beaonre”.....	94
<b>Πίνακας 8.2</b> Αποτελέσματα της ανάλυσης των δεδομένων “beaonre”.....	95
<b>Πίνακας 8.3</b> Θεωρητικά – Δειγματικά $VaR_p$ ανά επίπεδο σημαντικότητας.....	96
<b>Πίνακας 8.4</b> Περιγραφικά μέτρα των δεδομένων “lossalae”.....	98
<b>Πίνακας 8.5</b> Αποτελέσματα της ανάλυσης των δεδομένων “lossalae”.....	99
<b>Πίνακας 8.6</b> Θεωρητικά – Δειγματικά $VaR_p$ ανά επίπεδο σημαντικότητας.....	100
<b>Πίνακας 8.7</b> Θεωρητικά – Δειγματικά $TVaR_p$ ανά επίπεδο σημαντικότητας.....	100

## ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Σε ορισμένες περιπτώσεις κατά την μελέτη του ύψους των ατομικών ζημιών ενός χαρτοφυλακίου οι αναλογιστές έρχονται αντιμέτωποι με δεδομένα που παρουσιάζουν θετική ασυμμετρία και έχουν βαριά ουρά. Με σκοπό την μοντελοποίηση του ύψους των ζημιών εφαρμόζονται ζημιοκατανομές όπως πχ, Lognormal, Weibull, Gamma, Pareto, GPD. Συχνά, μέσα στο ίδιο σετ δεδομένων παρατηρούνται ζημιές με μεγάλη συχνότητα και μικρή σφοδρότητα όπως επίσης και ζημιές με μικρή συχνότητα και μεγάλη σφοδρότητα. Παρότι τα προαναφερθέντα μοντέλα διευκολύνουν κατά την εφαρμογή τους λόγω της απλότητας τους, έχει διαπιστωθεί πως σε κάποιες περιπτώσεις αδυνατούν να περιγράψουν ικανοποιητικά την “ιδιάζουσα” συμπεριφορά των ζημιών του χαρτοφυλακίου. Κατανομές όπως οι Lognormal, Weibull και Gamma αποτελούν λογικές επιλογές για την περιγραφή των μικρότερων ζημιών του συνόλου των δεδομένων αλλά δεν εμφανίζουν ικανοποιητικά αποτελέσματα όσον αφορά τις μεγάλες ζημιές, δηλαδή τις ζημιές στην ουρά της κατανομής. Για τον λόγο αυτόν χρησιμοποιούνται μοντέλα με πιο βαριά ουρά όπως οι Pareto, GPD. Το μειονέκτημα αυτών είναι ότι τείνουν να υπερεκτιμούν την ουρά των δεδομένων με αποτέλεσμα ο αναλογιστής να θέτει υψηλά ασφάλιστρα. Συνεπώς, είναι αναγκαία η κατασκευή μίας νέας κλάσης κατανομών, οι οποίες θα παρουσιάζουν μεγαλύτερη ευελιξία και θα συλλαμβάνουν επαρκώς τα χαρακτηριστικά των ατομικών ζημιών. Θα πρέπει δηλαδή τα πλεονεκτήματα των βασικών ζημιοκατανομών να συνδυαστούν σε ένα ενιαίο μοντέλο με μεγαλύτερη αποδοτικότητα. Οι νέες αυτές κατανομές ονομάζονται συγκολλημένες (*Spliced – Composite distributions*). Η μεθοδολογία για την κατασκευή της νέας κλάσης εμφανίζεται για πρώτη φορά στην έρευνα των Cooray – Ananda (2005). Σύμφωνα με την μέθοδο αυτή, το σύνολο των δεδομένων χωρίζεται σε ένα σημείο  $\theta$  και η συγκολλημένη κατανομή παράγεται από τις Lognormal και Pareto (συμβολ. Lognormal – Pareto). Οι Cooray – Ananda ορίζουν την συνάρτηση πυκνότητας του νέου μοντέλου να είναι ανάλογη της συνάρτησης πυκνότητας της Lognormal για τις ζημιές που είναι κάτω από το σημείο  $\theta$  και ανάλογη της συνάρτησης πυκνότητας της κατανομής Pareto για τις ζημιές που ξεπερνούν το σημείο  $\theta$ . Τα αποτελέσματα της ανάλυσης που πραγματοποιούν οι Cooray - Ananda είναι ικανοποιητικά καθώς η πρώτη συγκολλημένη κατανομή επιλέγεται έναντι των Lognormal, Pareto, Gamma, Inverse Gaussian ως η καταλληλότερη για την περιγραφή των δεδομένων (Cooray – Ananda, 2005). Στην συνέχεια, ο Scollnik (2007)

αποδεικνύει πως ο τρόπος ορισμού της συγκολλημένης κατανομής από τους Cooray – Ananda είναι άκρως περιοριστικός καθώς εύκολα το μοντέλο τους ανάγεται σε διακριτή μίξη κατανομών όπου τα βάρη μίξης είναι εκ των προτέρων γνωστά. Ο Scollnik στην έρευνα του γενικεύει τον ορισμό της συγκολλημένης κατανομής διορθώνοντας την ατέλεια του αρχικού μοντέλου. Επιτρέπει δηλαδή στα βάρη μίξης να είναι άγνωστα και να εκτιμώνται από το εκάστοτε σετ δεδομένων χρησιμοποιώντας την έννοια της διακριτής μίξης κατανομών ως το κύριο εργαλείο για την κατασκευή των συγκολλημένων κατανομών. Η έρευνα συνεχίζεται με τις μελέτες των Scollnik (2007), Scollnik – Sun (2012) και Teodorescu – Vernic (2013), όπου παρουσιάζονται περισσότερα μοντέλα συγκολλημένων κατανομών όπως οι Lognormal – GPD, Gamma – Pareto, Gamma – GPD, Weibull – Pareto, Weibull – GPD. Ένα από τα κοινά χαρακτηριστικά των παραπάνω κατανομών είναι ότι για την συνάρτηση πυκνότητας τους υποτίθεται η συνέχεια και η διαφορισμότητα στο σημείο  $\theta$ , εκεί δηλαδή όπου χωρίζεται το σύνολο των δεδομένων. Οι υποθέσεις αυτές έχουν σκοπό να οδηγήσουν σε ένα πιο ομαλό μοντέλο ενώ ταυτόχρονα δημιουργούν συσχετίσεις μεταξύ των αγνώστων παραμέτρων. Για τους εκτιμητές μέγιστης πιθανοφάνειας (ε.μ.π.) των παραμέτρων δεν είναι εφικτή η εξαγωγή αναλυτικών τύπων συνεπώς καθίσταται αναγκαία η χρήση υπολογιστικών πακέτων για την εκτίμηση τους (π.χ, R). Οι κατανομές που θα παρουσιασθούν αναλυτικά είναι οι Lognormal – Pareto, Lognormal – GPD, Gamma – GPD, Weibull – GPD. Σχετικά με τα μέτρα κινδύνου VaR και TVaR, είναι εφικτός ο υπολογισμός τους μέσω κλειστών τύπων οι οποίοι παρουσιάζονται στην συνέχεια. Έπειτα, Θα πραγματοποιηθεί μελέτη προσομοίωσης για κάθε μία από τις παραπάνω κατανομές με σκοπό να ελεγχθεί η απόδοση των ε.μ.π. που προκύπτουν από τις αριθμητικές μεθόδους του πακέτου R. Τέλος, θα πραγματοποιηθεί αριθμητική εφαρμογή των συγκολλημένων κατανομών σε δύο σετ πραγματικών δεδομένων με στόχο να παρουσιασθούν και να αναλυθούν τα πλεονεκτήματα των νέων κατανομών σε σχέση με τις βασικές ζημιοκατανομές.

## ΚΕΦΑΛΑΙΟ 2: ΤΟ ΜΟΝΤΕΛΟ ΤΗΣ ΣΥΓΚΟΛΛΗΜΕΝΗΣ ΚΑΤΑΝΟΜΗΣ

### 2.1 Συγκολλημένη κατανομή n-συνιστωσών

Αν υποθέσουμε ότι οι ζημιές του χαρτοφυλακίου περιγράφονται μέσω της τυχαίας μεταβλητής  $X$ , δίνουμε τον παρακάτω ορισμό:

**Ορισμός.** Η τ.μ  $X$  ακολουθεί την συγκολλημένη κατανομή των n-συνιστωσών (*n-component spliced distribution*)  $Y_1, Y_2, \dots, Y_n$  όταν η συνάρτηση πυκνότητας πιθανότητας (σ.π.π)  $f_X(x)$  μπορεί να εκφραστεί ως εξής:

$$f_X(x) = \begin{cases} c_1 f_{Y_1}(x), & d_0 < x < d_1 \\ c_2 f_{Y_2}(x) & , d_1 < x < d_2 \\ \vdots & \\ c_n f_{Y_n}(x) & , d_{n-1} < x < d_n \end{cases} \quad (2.1)$$

όπου  $c_i > 0, i = 1, \dots, n$  είναι σταθερές κανονικοποίησης,  $0 < d_0 < d_1 < \dots < d_n$  κατάλληλοι θετικοί αριθμοί και  $f_{Y_i}(x), i = 1, \dots, n$  είναι οι σ.π.π των τ.μ  $Y_i$ .

**Σχόλιο.** Θα λέμε ότι η πυκνότητα  $f_X(x)$  είναι ανάλογη της  $f_{Y_1}(x)$  στο  $(d_0, d_1)$ , είναι ανάλογη της  $f_{Y_2}(x)$  στο  $(d_1, d_2)$ , κ.ο.κ.

**Ορισμός.** Έστω  $X, Y_1, Y_2, \dots, Y_n$  τ.μ με σ.π.π  $f_X(x), f_{Y_1}(x), f_{Y_2}(x), \dots, f_{Y_n}(x)$  και  $w_i \in (0,1), i = 1, \dots, n$  τέτοια ώστε  $\sum_{i=1}^n w_i = 1$ . Τότε θα λέμε ότι η τ.μ  $X$  είναι διακριτή μίξη (*discrete mixture model*) των τ.μ  $Y_1, Y_2, \dots, Y_n$  με αντίστοιχα βάρη μίξης  $w_1, w_2, \dots, w_n$ , όταν:

$$f_X(x) = \sum_{i=1}^n w_i f_{Y_i}(x)$$

Με την βοήθεια των παραπάνω ορισμών θα αποδείξουμε την εξής σημαντική πρόταση:

**Πρόταση 2.1.** Έστω η τ.μ  $X$  που ακολουθεί την συγκολλημένη κατανομή των  $Y_1, Y_2, \dots, Y_n$  τότε η σ.π.π  $f_X(x)$  μπορεί να γραφεί εναλλακτικά ως εξής:

$$f_X(x) = \begin{cases} w_1 f_{Y_1}^*(x), & d_0 < x < d_1 \\ w_2 f_{Y_2}^*(x) & , d_1 < x < d_2 \\ \vdots & \\ w_n f_{Y_n}^*(x) & , d_{n-1} < x < d_n \end{cases} \quad (2.2)$$

όπου  $w_i \in (0,1)$ ,  $i = 1, \dots, n$ ,  $\sum_{i=1}^n w_i = 1$  και  $f_{Y_i}^*(x)$  αναπαριστά την συνάρτηση πυκνότητας της τ.μ  $Y_i^*$  στο διάστημα  $(d_{i-1}, d_i)$ ,  $i = 1, \dots, n$ .

**Απόδειξη.** Πράγματι, έστω σ.π.π  $f_X(x)$  που δίνεται από την σχέση 2.1, τότε θα έχουμε:

$$f_X(x) = \begin{cases} c_1 f_{Y_1}(x), & d_0 < x < d_1 \\ c_2 f_{Y_2}(x) & , d_1 < x < d_2 \\ \vdots & \\ c_n f_{Y_n}(x) & , d_{n-1} < x < d_n \end{cases}$$

Θα κατασκευάσουμε κατάλληλες συναρτήσεις πυκνότητας  $f_{Y_1}^*(x), f_{Y_2}^*(x), \dots, f_{Y_n}^*(x)$  και βάρη μίξης  $w_1, w_2, \dots, w_n$ .

Ορίζουμε τις περικομμένες τ.μ (*truncated random variables*):

$$Y_i^* = Y_i | d_{i-1} < Y_i < d_i, \quad i = 1, \dots, n.$$

Η πυκνότητα της τ.μ  $Y_i^*$  ισούται με:

$$f_{Y_i}^*(x) = \frac{f_{Y_i}(x)}{P(d_{i-1} < Y_i < d_i)}, \quad d_{i-1} < x < d_i$$

Επίσης ορίζουμε βάρη μίξης:

$$w_i = c_i P(d_{i-1} < Y_i < d_i), \quad i = 1, \dots, n.$$

Τώρα, αν στην  $f_X(x)$  πολλαπλασιάσουμε και διαιρέσουμε τον  $i$ -οστό κλάδο με την πιθανότητα  $P(d_{i-1} < Y_i < d_i)$ ,  $i = 1, \dots, n$ , τότε θα έχουμε:

$$f_X(x) = \begin{cases} w_1 f_{Y_1}^*(x), & d_0 < x < d_1 \\ w_2 f_{Y_2}^*(x) & , d_1 < x < d_2 \\ \vdots & \\ w_n f_{Y_n}^*(x) & , d_{n-1} < x < d_n \end{cases}$$

Ισοδύναμα, επειδή οι κλάδοι της συνάρτησης πυκνότητας ορίζονται σε ξένα διαστήματα, λόγω του ότι  $d_0 < d_1 < \dots < d_n$ , είναι εφικτό να εκφραστεί σε συμπυκνωμένη μορφή ως εξής:

$$f_X(x) = \sum_{i=1}^n w_i f_{Y_i}^*(x)$$

Δηλαδή η  $X$  είναι διακριτή μίξη των περικομμένων τ.μ  $Y_1^*, Y_2^*, \dots, Y_n^*$ . Για να ολοκληρωθεί η απόδειξη μένει να δείξουμε ότι οι αριθμοί  $w_i$  είναι πράγματι βάρη μίξης, δηλαδή ότι ισχύει:



$$\sum_{i=1}^n w_i = 1$$

Αφού η συνάρτηση  $f_X(x)$  ορίζει πυκνότητα θα έχουμε:

$$\begin{aligned} \int_0^{\infty} f(x)dx &= \int_{d_0}^{d_1} c_1 f_{Y_1}(x)dx + \int_{d_1}^{d_2} c_2 f_{Y_2}(x)dx + \dots + \int_{d_{n-1}}^{d_n} c_n f_{Y_n}(x)dx = \sum_{i=1}^n \int_{d_{i-1}}^{d_i} c_i f_{Y_i}(x)dx \\ &= \sum_{i=1}^n c_i P(d_{i-1} < Y_i < d_i) = 1 \end{aligned}$$

Άρα, από τον ορισμό των  $w_i$  θα ισχύει ότι :

$$\sum_{i=1}^n w_i = \sum_{i=1}^n c_i P(d_{i-1} < Y_i < d_i) = 1$$

**Σχόλιο.** Απο την πρόταση 2.1 συμπεραίνουμε ότι η συγκολλημένη κατανομή τυχαίων μεταβλητών είναι ειδική περίπτωση της διακριτής μίξης τυχαίων μεταβλητών.

## 2.2 Συγκολλημένη κατανομή 2-συνιστωσών

Ιδιαίτερο πρακτικό ενδιαφέρον για την μοντελοποίηση μεγεθών ατομικών ζημιών παρουσιάζει η περίπτωση όπου  $n = 2$ . Έστω  $d_0 = 0, d_1 = d, d_2 = \infty$ .

Έστω η τ.μ  $X$  η οποία ακολουθεί την συγκολλημένη κατανομή των τ.μ  $Y_1, Y_2$  με σ.κ  $F_{Y_1}(x), F_{Y_2}(x)$  αντίστοιχα και συναρτήσεις δεξιάς ουράς  $\bar{F}_{Y_1}(x), \bar{F}_{Y_2}(x)$  αντίστοιχα . Τότε από την σχέση 2.1 καταλήγουμε πως η σ.π.π εκφράζεται ως εξής:

$$f_X(x) = \begin{cases} c_1 f_{Y_1}(x) & , 0 < x \leq d \\ c_2 f_{Y_2}(x) & , d < x < \infty \end{cases}$$

Το παραπάνω μοντέλο συγκολλημένης κατανομής μπορεί να αναχθεί σε ένα ισοδύναμο μοντέλο διακριτής μίξης δύο τυχαίων μεταβλητών όπως είδαμε στην πρόταση 2.1 ως εξής:

$$f_X(x) = \begin{cases} w f_{Y_1^*}(x) & , 0 < x \leq d \\ (1-w) f_{Y_2^*}(x) & , d < x < \infty \end{cases}$$

δηλαδή η τ.μ  $X$  είναι διακριτή μίξη των τ.μ  $Y_1^* = Y_1 | Y_1 < d, Y_2^* = Y_2 | Y_2 > d$  με βάρη μίξης  $w, 1-w$  αντίστοιχα.

Η συνάρτηση κατανομής (σ.κ) της τ.μ  $Y_1^* \in (0, d)$  είναι:

$$F_{Y_1^*}(x) = \begin{cases} 0 & , x < 0 \\ \frac{F_{Y_1}(x)}{F_{Y_1}(d)} & , 0 < x < d \\ 1 & , x \geq d \end{cases}$$

και η σ.π.π είναι:

$$f_{Y_1^*}(x) = \frac{f_{Y_1}(x)}{F_{Y_1}(d)} , \quad 0 < x \leq d$$

Η σ.κ της τ.μ  $Y_2^* \in (d, \infty)$  είναι:

$$F_{Y_2^*}(x) = \begin{cases} 0 & , 0 < x \leq d \\ \frac{F_{Y_2}(x) - F_{Y_2}(d)}{1 - F_{Y_2}(d)} & , x > d \end{cases}$$

και η σ.π.π είναι:

$$f_{Y_2^*}(x) = \frac{f_{Y_2}(x)}{1 - F_{Y_2}(d)} , \quad x > d$$

Στην συνέχεια θα αναφερθούμε στα βασικά χαρακτηριστικά της τ.μ  $X$ .

**Πρόταση 2.2.** Η συνάρτηση κατανομής της τ.μ  $X$  θα δίνεται από τον παρακάτω τύπο:

$$F_X(x) = \begin{cases} w \frac{F_{Y_1}(x)}{F_{Y_1}(d)} & , 0 < x \leq d \\ w + (1 - w) \left[ 1 - \frac{\bar{F}_{Y_2}(x)}{\bar{F}_{Y_2}(d)} \right] & , x > d \end{cases}$$

**Απόδειξη.** Για  $0 < x < d$ , έχουμε:

$$F_X(x) = \int_0^x f_X(t) dt = \int_0^x w f_{Y_1^*}(t) dt = w \int_0^x \frac{f_{Y_1}(t)}{F_{Y_1}(d)} dt = w \frac{F_{Y_1}(x)}{F_{Y_1}(d)}$$

Για  $x > d$  είναι:

$$\begin{aligned} F_X(x) &= \int_0^x f_X(t) dt = \int_0^d f_X(t) dt + \int_d^x f_X(t) dt = w \int_0^d f_{Y_1^*}(t) dt + (1 - w) \int_d^x f_{Y_2^*}(t) dt \\ &= w \int_0^d \frac{f_{Y_1}(t)}{F_{Y_1}(d)} dt + (1 - w) \int_d^x \frac{f_{Y_2}(t)}{1 - F_{Y_2}(d)} dt \\ &= w \frac{F_{Y_1}(d)}{F_{Y_1}(d)} + (1 - w) \frac{F_{Y_2}(x) - F_{Y_2}(d)}{1 - F_{Y_2}(d)} = w + (1 - w) \left[ 1 - \frac{1 - F_{Y_2}(x)}{1 - F_{Y_2}(d)} \right] \\ &= w + (1 - w) \left[ 1 - \frac{\bar{F}_{Y_2}(x)}{\bar{F}_{Y_2}(d)} \right] \end{aligned}$$

**Πρόταση 2.3.** Η συνάρτηση δεξιάς ουράς της τ.μ  $X$  θα υπολογίζεται από τον τύπο:

$$\bar{F}_X(x) = \begin{cases} 1 - w \frac{F_{Y_1}(x)}{F_{Y_1}(d)}, & 0 < x \leq d \\ (1 - w) \left[ \frac{\bar{F}_{Y_2}(x)}{\bar{F}_{Y_2}(d)} \right], & d < x < \infty \end{cases}$$

**Απόδειξη.** Για  $0 < x < d$ , έχουμε:

$$\bar{F}_X(x) = 1 - F_X(x) = 1 - w \frac{F_{Y_1}(x)}{F_{Y_1}(d)}$$

Για  $x > d$  είναι:

$$\bar{F}_X(x) = 1 - F_X(x) = 1 - w - (1 - w) \left[ 1 - \frac{1 - F_{Y_2}(x)}{1 - F_{Y_2}(d)} \right] = (1 - w) \left[ \frac{\bar{F}_{Y_2}(x)}{\bar{F}_{Y_2}(d)} \right]$$

**Πρόταση 2.4.** Η ροπή τάξεως  $-k$  της τ.μ  $X$  δίνεται από τον τύπο:

$$E(X^k) = wE(Y_1^{*k}) + (1 - w)E(Y_2^{*k})$$

όπου  $k \in \{1, 2, \dots\}$  τέτοιο ώστε  $E(Y_1^{*k}) < \infty, E(Y_2^{*k}) < \infty$ .

**Απόδειξη .** Πράγματι, θα έχουμε :

$$\begin{aligned} E(X^k) &= \int_0^\infty x^k f_X(x) dx = \int_0^d x^k f_X(x) dx + \int_d^\infty x^k f_X(x) dx = \\ &= w \int_0^d x^k f_{Y_1^*}(x) dx + (1 - w) \int_d^\infty x^k f_{Y_2^*}(x) dx = wE(Y_1^{*k}) + (1 - w)E(Y_2^{*k}) \end{aligned}$$

Κατά την μελέτη των ζημιοκατανομών συχνά επικεντρωνόμαστε στον υπολογισμό διαφόρων μεγεθών που προκύπτουν μέσα από τις ροπές τάξεως  $-k$ , για  $k = 1, 2, 3, 4$ . Πέρα από την μέση τιμή και την διασπορά της κατανομής εστιάζουμε επίσης στα παρακάτω μεγέθη:

### Λοξότητα

Είναι το μέτρο ασυμμετρίας που χαρακτηρίζει την κατανομή γύρω από την μέση τιμή της. Στις περιπτώσεις των κατανομών που μελετάμε οι τιμές της λοξότητας θα είναι θετικές και θα δηλώνουν την βαριά ουρά των κατανομών καθώς επίσης και την συγκέντρωση της μάζας της κατανομής στα αριστερά. Η λοξότητα θα δίνεται από τον τύπο:

$$E \left[ \left( \frac{X - \mu}{\sigma} \right)^3 \right] = \frac{E(X^3) - 3E(X)E(X^2) + 2E^3(X)}{\sigma^3} \quad (2.3)$$

όπου  $\mu = E(X)$ ,  $\sigma = \sqrt{E(X^2) - E^2(X)}$  συμβολ. την μέση τιμή και την τυπική απόκλιση της τ.μ  $X$  αντίστοιχα.

### **Κύρτωση**

Είναι το μέτρο ασυμμετρίας που προσδιορίζει τον βαθμό συγκέντρωσης των τιμών της κατανομής γύρω από την μέση τιμή και δείχνει την “αιχμηρότητα” της πυκνότητας της κατανομής. Η κύρτωση θα δίνεται από τον τύπο:

$$E \left[ \left( \frac{X - \mu}{\sigma} \right)^4 \right] = \frac{E(X^4) - 4E(X)E(X^3) + 6E^2(X)E(X^2) - 4E^3(X)E(X) + E^4(X)}{\sigma^4} \quad (2.4)$$

όπου  $\mu = E(X)$ ,  $\sigma = \sqrt{E(X^2) - E^2(X)}$  συμβολ. την μέση τιμή και την τυπική απόκλιση της τ.μ  $X$  αντίστοιχα.

Ένα σύννηθες μέτρο κινδύνου για μια τ.μ  $X \sim F$  είναι το Value at Risk σε ένα καθορισμένο επίπεδο σημαντικότητας  $p$  και συμβολίζεται με  $VaR_p(X)$ . Είναι ουσιαστικά το  $p$  – κάτω ποσοστιαίο σημείο της κατανομής και δηλώνει εκείνη την τιμή της κατανομής όπου με πιθανότητα  $1 - p$  θα παρατηρηθεί τιμή μεγαλύτερη από αυτή.

$$VaR_p(X) = F_X^{-1}(p)$$

Το μέτρο αυτό μας παρέχει μία εικόνα για το πόσο βαριά είναι η ουρά της κατανομής που μελετάμε και μπορεί να χρησιμοποιηθεί ως κριτήριο για την προσαρμογή των ζημιοκατανομών σε ένα σετ δεδομένων.

Ανάλογα την φύση των δεδομένων που μελετάμε επικεντρωνόμαστε σε διαφορετικά χαρακτηριστικά της κατανομής π.χ αν μελετάμε ασφαλιστικές αποζημιώσεις είναι λογικό η έρευνα μας να στρέφεται στις καταστροφικές ζημιές, δηλαδή αυτές που παρουσιάζουν υψηλή σφοδρότητα και χαμηλή συχνότητα και εμφανίζονται στην ουρά της ζημιοκατανομής. Συνεπώς, θα πρέπει να εστιάσουμε στην ουρά της κατανομής και να προσδιορίσουμε τα ποσά εκείνα που, βάσει την συμπεριφορά των ζημιών, μια εταιρεία θα πρέπει να διαθέτει με σκοπό να εξασφαλίσει την λειτουργία της ακόμα και στο ακραίο σενάριο που συμβαίνει με πιθανότητα 95%, 99%, 99.5% δηλαδή τα  $VaR_{0.95}$ ,  $VaR_{0.99}$ ,  $VaR_{0.995}$ .

**Πρόταση 2.5.** Το Value at Risk για την τ.μ  $X$  σε ένα επίπεδο σημαντικότητας  $p$  θα υπολογίζεται από τον κάτωθι τύπο:

$$VaR_p(X) = \begin{cases} F_{Y_1^*}^{-1} \left[ \frac{p}{w} F_{Y_1^*}(d) \right] & , 0 \leq p < w \\ F_{Y_2^*}^{-1} \left[ \frac{p - w + (1 - p)F_{Y_2^*}(d)}{1 - w} \right] & , w \leq p < 1 \end{cases}$$

**Απόδειξη.** Πράγματι, έστω ότι:

$$F_X(x) = p$$

Θέλουμε να λύσουμε την παραπάνω εξίσωση ως προς  $x$  δηλαδή να έχουμε:

$$x = F_X^{-1}(p)$$

Οπότε θα διακρίνουμε περιπτώσεις για το  $p$ :

Αν  $0 \leq p < w$  τότε:

$$F_X(x) = p \Rightarrow w \frac{F_{Y_1^*}(x)}{F_{Y_1^*}(d)} = p \Rightarrow F_{Y_1^*}(x) = \frac{pF_{Y_1^*}(d)}{w} \Rightarrow x = F_{Y_1^*}^{-1} \left[ \frac{p}{w} F_{Y_1^*}(d) \right]$$

Αρα θα είναι:

$$VaR_p(X) = F_{Y_1^*}^{-1} \left[ \frac{p}{w} F_{Y_1^*}(d) \right]$$

Αν  $w \leq p < 1$  τότε:

$$\begin{aligned} F_X(x) = p &\Rightarrow w + (1 - w) \left[ 1 - \frac{\bar{F}_{Y_2^*}(x)}{\bar{F}_{Y_2^*}(d)} \right] = p \Rightarrow \bar{F}_{Y_2^*}(x) = 1 - [1 - F_{Y_2^*}(d)] \left[ \frac{1 - p}{1 - w} \right] \\ &\Rightarrow x = F_{Y_2^*}^{-1} \left[ \frac{p - w + (1 - p)F_{Y_2^*}(d)}{1 - w} \right] \end{aligned}$$

Αρα θα είναι:

$$VaR_p(X) = F_{Y_2^*}^{-1} \left[ \frac{p - w + (1 - p)F_{Y_2^*}(d)}{1 - w} \right]$$

Ένα εξίσου σημαντικό μέτρο κινδύνου είναι το Tail Value at Risk σε ένα επίπεδο σημαντικότητας  $p$  και συμβολίζεται με  $TVaR_p(X)$ . Ορίζεται ως η μέση τιμή της τ.μ  $X$  δεδομένου ότι αυτή ξεπερνάει το  $VaR_p(X)$ . Το  $TVaR_p(X)$  δίνει μία πιο ουσιαστική εικόνα για την ουρά της κατανομής σε σχέση με το  $VaR_p(X)$  διότι σε πολλές περιπτώσεις δύο κατανομές τυχαίνει να έχουν ίσα  $p$  – κάτω ποσοστιαία σημεία αλλά να εμφανίζουν τελείως διαφορετική συμπεριφορά στην ουρά. Επομένως για την πιο ολοκληρωμένη μελέτη της ζημιοκατανομής ορίζεται το  $TVaR_p(X)$  ως εξής:

$$TVaR_p(X) = E(X|X > VaR_p(X)) = \frac{1}{1-p} \int_{VaR_p(X)}^{\infty} x f_X(x) dx$$

**Πρόταση 2.6.** Το Tail Value at Risk για την τ.μ  $X$  σε καθορισμένο επίπεδο σημαντικότητας  $p$  ορίζεται ως εξής:

$$TVaR_p(X) = \begin{cases} \int_{VaR_p(X)}^d \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_d^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx, & VaR_p(X) \leq d \\ \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx & , \quad VaR_p(X) > d \end{cases}$$

**Απόδειξη.** Πράγματι, θα έχουμε:

$$\begin{aligned} TVaR_p(X) &= E(X|X > VaR_p(X)) = \frac{1}{1-p} \int_{VaR_p(X)}^{\infty} x f_X(x) dx = \\ &= \frac{1}{1-p} \int_{VaR_p(X)}^{\infty} x [r f_{Y_1}^*(x) + (1-r) f_{Y_2}^*(x)] dx = \\ &= \int_{VaR_p(X)}^{\infty} \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx \end{aligned}$$

Θα διακρίνουμε περιπτώσεις για το  $VaR_p(X)$ .

Αν  $VaR_p(X) \leq d$  τότε είναι:

$$TVaR_p(X) = \int_{VaR_p(X)}^d \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_d^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Αφού οι τ.μ  $Y_1^*, Y_2^*$  ορίζονται στα διαστήματα  $(0, d), (d, \infty)$  αντίστοιχα.

Αν  $VaR_p(X) > d$  τότε είναι:

$$\int_{VaR_p(X)}^{\infty} \frac{xr}{1-p} f_{Y_1}^*(x) dx = 0$$

εφόσον η τ.μ  $Y_1^* \in (0, d)$ .

Συνεπώς θα ισχύει:

$$TVaR_p(X) = \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

## ΚΕΦΑΛΑΙΟ 3: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ LOGNORMAL – PARETO ( $\alpha, \vartheta, \sigma, \mu, r$ )

### 3.1 Εισαγωγή

Οι αναλογιστές συχνά έρχονται αντιμέτωποι με δεδομένα τα οποία παρουσιάζουν υψηλή θετική ασυμμετρία (*skewness*). Ορισμένα από τα μοντέλα που εφαρμόζονται για την περιγραφή των δεδομένων είναι η κατανομή Pareto και η κατανομή Lognormal. Η κατανομή Pareto αποτελεί μία αξιόπιστη επιλογή για την μοντελοποίηση των μεγάλων τιμών των δεδομένων ενώ η Lognormal μοντελοποιεί πιο αποτελεσματικά τις μικρότερες τιμές, αυτές που εμφανίζουν μεγαλύτερη συχνότητα.

Στην έρευνα των Cooray – Ananda (2005) αναφέρεται η πρώτη συγκολλημένη κατανομή για  $n = 2$  που έχει ως αριστερό άκρο την κατανομή Lognormal και ως δεξί άκρο την κατανομή Pareto. Η αρχική ιδέα ήταν να χωριστεί σε δύο μέρη το σύνολο των δεδομένων, έστω σε ένα σημείο  $\theta$ , και η συνάρτηση πυκνότητας να αλλάζει την μορφή της εκατέρωθεν του σημείου αυτού. Για να επιτευχθεί αυτό, οι Cooray – Ananda όρισαν την συνάρτηση πυκνότητας να είναι ανάλογη της κατανομής Lognormal μέχρι το σημείο  $\theta$  και ανάλογη της κατανομής Pareto από αυτό το σημείο και έπειτα. Δηλαδή:

$$f(x) = \begin{cases} cf_{Y_1}(x), & 0 < x \leq \theta \\ cf_{Y_2}(x), & \theta < x < \infty \end{cases} \quad (3.1)$$

όπου

$$f_{Y_1}(x) = \frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}, \quad x > 0, \sigma > 0, \mu \in \mathbb{R}$$

και

$$f_{Y_2}(x) = \frac{\alpha\theta^\alpha}{x^{a+1}}, \quad x > \theta, a > 0, \theta > 0$$

είναι οι συναρτήσεις πυκνότητας πιθανότητας των κατανομών Lognormal( $\mu, \sigma^2$ ), Pareto ( $\alpha, \theta$ ) αντίστοιχα. Η συγκολλημένη κατανομή Lognormal – Pareto (LP) έχει τέσσερις άγνωστες παραμέτρους,  $\alpha > 0, \sigma > 0, \theta > 0, \mu \in \mathbb{R}$ .

Το σημείο  $\theta$  όπου χωρίζεται το σύνολο των δεδομένων δεν είναι εκ των προτέρων γνωστό αλλά εντάσσεται όπως είδαμε στο μοντέλο ως άγνωστη παράμετρος η οποία θα πρέπει να εκτιμηθεί

από τα δεδομένα. Οι Cooray – Ananda με σκοπό να αποφύγουν τα προβλήματα συνέχειας και διαφορισιμότητας της συνάρτησης πυκνότητας στο σημείο  $\theta$  έκαναν δύο υποθέσεις.

1. Υπόθεση συνέχειας στο σημείο  $\theta$ :

$$f(\theta -) = f(\theta +) \Rightarrow f_{Y_1}(\theta) = f_{Y_2}(\theta)$$

2. Υπόθεση διαφορισιμότητας στο σημείο  $\theta$ :

$$f'(\theta -) = f'(\theta +) \Rightarrow f'_{Y_1}(\theta) = f'_{Y_2}(\theta)$$

Αυτές οι υποθέσεις οδηγούν σε ένα πιο ομαλό μοντέλο ενώ ταυτόχρονα μειώνουν το πλήθος των παραμέτρων από τέσσερις σε δύο. Μέσω των παραπάνω υποθέσεων προέκυψαν οι παρακάτω σχέσεις μεταξύ των παραμέτρων:

$$\frac{\ln\theta - \mu}{\sigma} = \alpha\sigma = \kappa \quad (3.2)$$

και

$$e^{-k^2} = 2\pi k^2 \quad (3.3)$$

Στην μελέτη των Cooray – Ananda (2005) έχει αποδειχθεί ότι η σταθερά  $\kappa$  ισούται με:

$$\kappa = 0.372238898$$

όπου έχει αποδειχθεί επίσης για την σταθερά κανονικοποίησης  $c$  της σχέσης 3.1 ότι:

$$c = \frac{1}{1 + \Phi(k)}$$

Συνεπώς, για την σταθερά  $c$  θα είναι:

$$c = 0.6078501$$

Έπειτα από κατάλληλη αναπαραμέτρηση η συνάρτηση πυκνότητας μπορεί να γραφεί ως εξής:

$$f(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}} \exp\left\{-\frac{\alpha^2}{2k^2} \ln^2\left(\frac{x}{\theta}\right)\right\} & , 0 < x \leq \theta \\ \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}} & , \theta < x < \infty \end{cases}$$

όπου οι άγνωστες παράμετροι είναι οι  $\alpha > 0, \theta > 0$ .

Η συγκολλημένη κατανομή LP προσαρμόστηκε για πρώτη φορά στο σετ δεδομένων “Danish fire insurance loss data” (Cooray – Ananda, 2005) και η απόδοσή της συγκρίθηκε με αυτές των Lognormal, Pareto, Inverse Gaussian, Gamma, Weibull. Η σύγκριση έγινε ως προς τις τιμές της συνάρτησης λογαριθμο – πιθανοφάνειας (Loglikelihood - LL), το στατιστικό των Kolmogorov – Smirnov και τις τιμές των p – values για τους ελέγχους καλής προσαρμογής  $\chi^2$ . Η κατανομή LP παρουσίασε αισθητά καλύτερα αποτελέσματα σε σχέση τις προηγούμενες



κατανομές ως προς όλα τα μέτρα σύγκρισης. Βάσει των προηγούμενων αποτελεσμάτων οι Cooray – Ananda κατέληξαν πως τέτοιου είδους κατανομές αποτελούν πιο λογικές επιλογές για την μοντελοποίηση του ύψους ατομικών ζημιών διότι συνδυάζουν τα χαρακτηριστικά των επιμέρους κατανομών προσφέροντας μεγαλύτερη ευελιξία και περιγραφικότητα κατά την μελέτη των δεδομένων.

Παρότι, σε αυτή την ιδέα στηρίχθηκε η δημιουργία και άλλων συγκολλημένων κατανομών όπως η Exponential - Pareto (Teodorescu – Vernic, 2006) και η Weibull – Pareto (Ciumara, 2006) αποδείχθηκε από τον Scollnik (Scollnik, 2007) πως ο τρόπος που οι Cooray – Ananda όρισαν την συγκολλημένη LP ήταν άκρως περιοριστικός διότι το μοντέλο μπορούσε να αναχθεί μέσω απλών αλγεβρικών πράξεων σε ένα μοντέλο διακριτής μίξης μεταξύ της περικομμένης (αριστερά) στο σημείο  $\theta$  Lognormal και της περικομμένης (δεξιά) στο σημείο  $\theta$  Pareto με εκ των προτέρων γνωστά και σταθερά βάρη μίξης (39.2% και 60.8% αντίστοιχα). Δηλαδή ο Scollnik μέσα από την μεθοδολογία της πρότασης 2.1 έδειξε πως η σχέση 3.1 γράφεται ως εξής:

$$f(x) = \begin{cases} \psi f_{Y_1^*}(x) & , \quad 0 < x \leq \theta \\ (1 - \psi) f_{Y_2^*}(x) & , \quad \theta < x < \infty \end{cases} \quad (3.4)$$

όπου

$$f_{Y_1^*}(x) = \frac{f_{Y_1}(x)}{F_{Y_1}(\theta)}, \quad x \leq \theta$$

και

$$f_{Y_2^*}(x) = \frac{f_{Y_2}(x)}{1 - F_{Y_2}(\theta)}, \quad x > \theta$$

Συνεπώς, μέσω της σχέσης 3.1 και 3.4 συμπεραίνουμε ότι για το βάρος μίξης  $\psi$  θα είναι:

$$\psi = c F_{Y_1}(\theta)$$

Αφού  $Y_1 \sim \text{Lognormal}(\mu, \sigma^2)$  τότε από την σχέση 3.2 θα ισχύει:

$$F_{Y_1}(\theta) = \Phi\left(\frac{\ln\theta - \mu}{\sigma}\right) = \Phi(\kappa) = 0.64514$$

Αρα τελικά θα ισχύει:

$$\psi = 0.392$$

Αρα για την πυκνότητα της σχέσης 3.4 ισχύει ότι:

$$f(x) = 0.392 * f_{Y_1^*}(x) + 0.608 * f_{Y_2^*}(x), \quad x \in \mathbb{R}$$

Πράγματι λοιπόν το μοντέλο των Cooray – Ananda δεν έχει κάποια πρακτική σημασία διότι απαιτεί εν τέλει το 39.2% των δεδομένων να προέρχονται από την περικομμένη Lognormal ενώ το υπόλοιπο 60.8% να προέρχεται από την περικομμένη Pareto κάτι το οποίο είναι προφανώς αδύνατο να συμβαίνει σε όλες τις περιπτώσεις. Συνεπώς, ο Scollnik στηρίχθηκε στην ιδέα των Cooray – Ananda και γενίκευσε τον ορισμό των συγκολλημένων κατανομών ξεκινώντας από την σχέση 3.4 . Για τον Scollnik η συγκολλημένη κατανομή είναι η διακριτή μίξη των περικομμένων κατανομών (στο σημείο  $\theta$ ) με βάρη που είναι άγνωστα και εντάσσονται στις παραμέτρους του μοντέλου. Δηλαδή το μοντέλο του Scollnik είναι το παρακάτω:

$$f(x) = r * f_{Y_1}^*(x) + (1 - r) * f_{Y_2}^*(x), \quad x \in \mathbb{R}$$

Στην βιβλιογραφία το μοντέλο των Cooray – Ananda αναφέρεται ως “First composite model” ενώ αντίστοιχα του Scollnik αναφέρεται ως “Second composite model” και οι άγνωστες παράμετροι του μοντέλου είναι οι  $0 \leq r \leq 1, \alpha > 0, \theta > 0, \sigma > 0, \mu \in \mathbb{R}$ .

Στην επόμενη ενότητα θα μελετήσουμε την συγκολλημένη κατανομή Lognormal – Pareto με τον τρόπο που ορίστηκε από τον Scollnik.

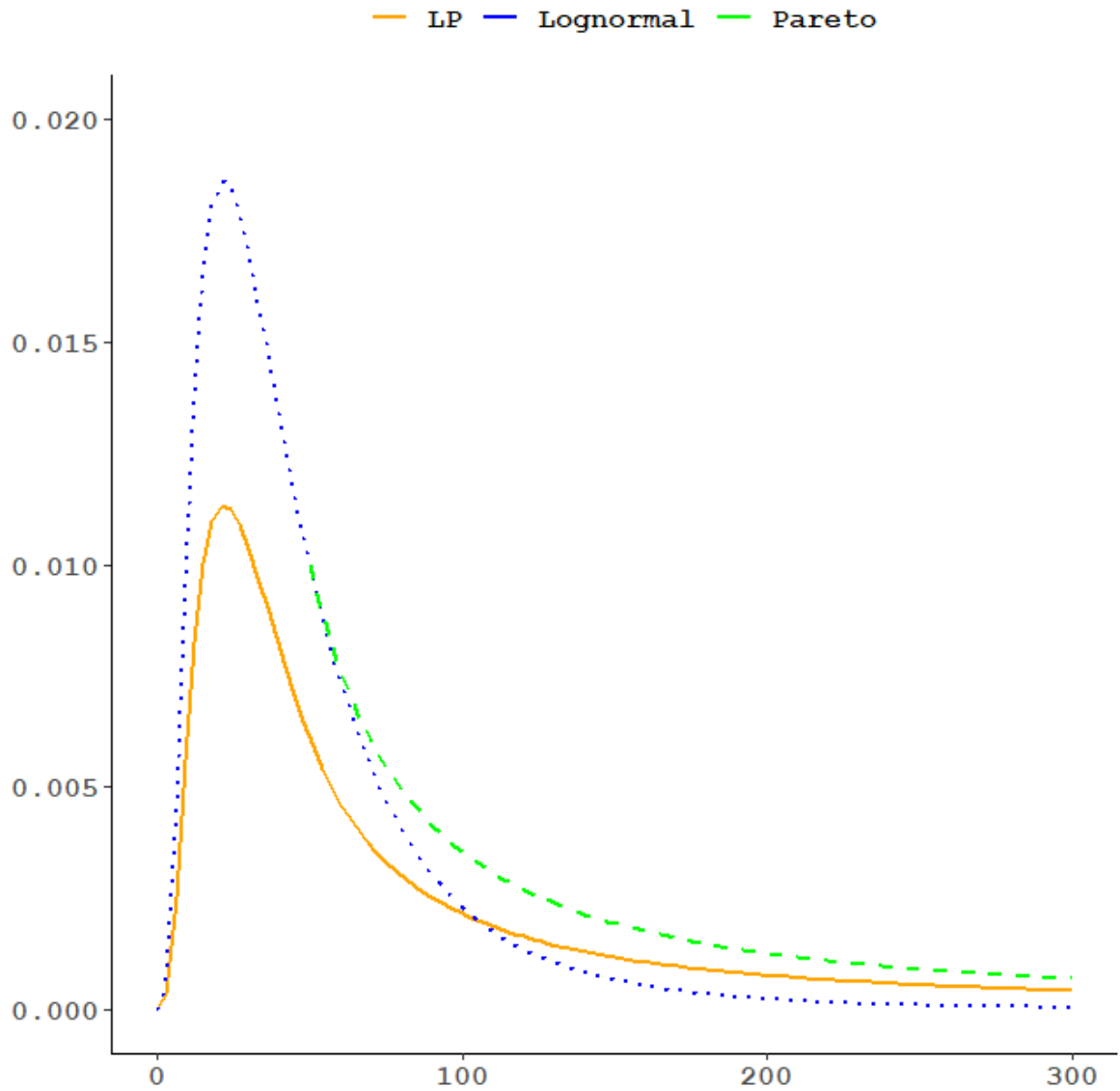
Η συγκολλημένη κατανομή LP θα είναι η διακριτή μίξη των κατανομών Lognormal, Pareto περικομμένων στο σημείο  $\theta$ .

Η σ.π.π της Lognormal( $\mu, \sigma^2$ ) περικομμένης στο σημείο  $\theta$  δίνεται από τον παρακάτω τύπο:

$$f_{Y_1}^*(x) = \frac{\frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, \quad x \leq \theta, \sigma > 0, \mu \in \mathbb{R}$$

ενώ για την σ.π.π της περικομμένης κατανομής Pareto( $\alpha, \theta$ ) στο σημείο  $\theta$  θα ισχύει ότι:

$$f_{Y_2}^*(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0$$



Σχήμα 3.1 Ενδεικτική γραφική παράσταση των κατανομών LP, Lognormal, Pareto

### 3.2 Χαρακτηριστικά μεγέθη της κατανομής LP ( $\alpha, \vartheta, \sigma, \mu, r$ )

**Ορισμός.** Αν υποθέσουμε τ.μ  $X \sim LP(\alpha, \theta, \sigma, \mu, r)$  τότε η συνάρτηση πυκνότητας πιθανότητας θα δίνεται από τον εξής τύπο:

$$f(x; \alpha, \theta, \sigma, \mu, r) = \begin{cases} \frac{(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{r \frac{x\sigma}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}}, & 0 < x \leq \theta \\ (1-r) \frac{a\theta^\alpha}{x^{\alpha+1}}, & \theta < x < \infty \end{cases}$$

όπου  $r$  είναι το βάρος μίξης,  $0 \leq r \leq 1$ ,  $\theta > 0$ ,  $\sigma > 0$ ,  $\alpha > 0$ ,  $\mu \in \mathbb{R}$  οι άγνωστες παράμετροι της κατανομής και  $\Phi(\cdot)$  είναι η σ.κ της τυποποιημένης κανονικής κατανομής.

Στο σχήμα 3.1 παρουσιάζεται η κατανομή LP σε σχέση με τις κατανομές που την απαρτίζουν. Παρατηρούμε ότι η συγκολλημένη κατανομή LP εμφανίζει πιο αργή σύγκλιση στο μηδέν απ ότι η κατανομή Lognormal ενώ ταυτόχρονα η ουρά της είναι πιο “ελαφριά” από την ουρά της κατανομής Pareto. Η συγκολλημένη κατανομή ενσωματώνει την συμπεριφορά της κατανομής Lognormal για τις μικρές – μεσαίες ζημιές οι οποίες εμφανίζουν μεγαλύτερη συχνότητα ενώ ταυτόχρονα στην ουρά παρουσιάζει χαρακτηριστικά παρόμοια με αυτά της κατανομής Pareto.

Ο Scollnik έκανε τις υποθέσεις συνέχειας και διαφορισιμότητας όπως ακριβώς και οι Cooray – Ananda. Οι παράμετροι του μοντέλου μειώθηκαν από πέντε σε τρεις. Στην πρόταση 3.1 αποδεικνύεται η μεθοδολογία που ακολούθησε ο Scollnik καθώς και οι σχέσεις που συνδέουν τις παραμέτρους.

**Πρόταση 3.1.** Αν η  $f(x; \alpha, \theta, \sigma, \mu, r)$  είναι συνεχής και διαφορίσιμη στο σημείο  $\theta$  τότε για τις παραμέτρους  $r, \mu$  θα ισχύουν οι σχέσεις:

$$\mu = \ln\theta - \alpha\sigma^2 \quad (3.5)$$

και

$$r = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)e^{\frac{(\alpha\sigma)^2}{2}}}{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)e^{\frac{(\alpha\sigma)^2}{2}} + 1} \quad (3.6)$$

**Απόδειξη.** Για να καταλήξουμε στις σχέσεις μεταξύ των παραμέτρων εργαζόμαστε ως εξής:

Από την υπόθεση συνέχειας στο σημείο  $\theta$ :

$$\lim_{x \rightarrow \theta^-} f(x) = \lim_{x \rightarrow \theta^+} f(x)$$

προκύπτει ότι:

$$\frac{r e^{-\frac{(\ln\theta - \mu)^2}{2\sigma^2}}}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right) \sqrt{2\pi\sigma^2\theta^2}} = \frac{(1-r)\alpha\theta^\alpha}{\theta^{\alpha+1}}$$

Λύνοντας την παραπάνω εξίσωση ως προς  $r$  θα έχουμε ότι:

$$r = \frac{\sqrt{2\pi\alpha^2\sigma^2}\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)e^{\frac{(\ln\theta - \mu)^2}{2\sigma^2}}}{\sqrt{2\pi\alpha^2\sigma^2}\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)e^{\frac{(\ln\theta - \mu)^2}{2\sigma^2}} + 1} \quad (3.7)$$

Από την υπόθεση διαφορισιμότητας στο σημείο  $\theta$ :

$$\lim_{x \rightarrow \theta^-} f'(x) = \lim_{x \rightarrow \theta^+} f'(x)$$

καταλήγουμε στην εξής σχέση:

$$\frac{r e^{-\frac{1}{2\sigma^2}(\ln\theta - \mu)^2} \left(1 + \frac{\ln\theta - \mu}{\sigma^2}\right)}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)\sqrt{2\pi\sigma^2}} = a(a+1)(1-r)$$

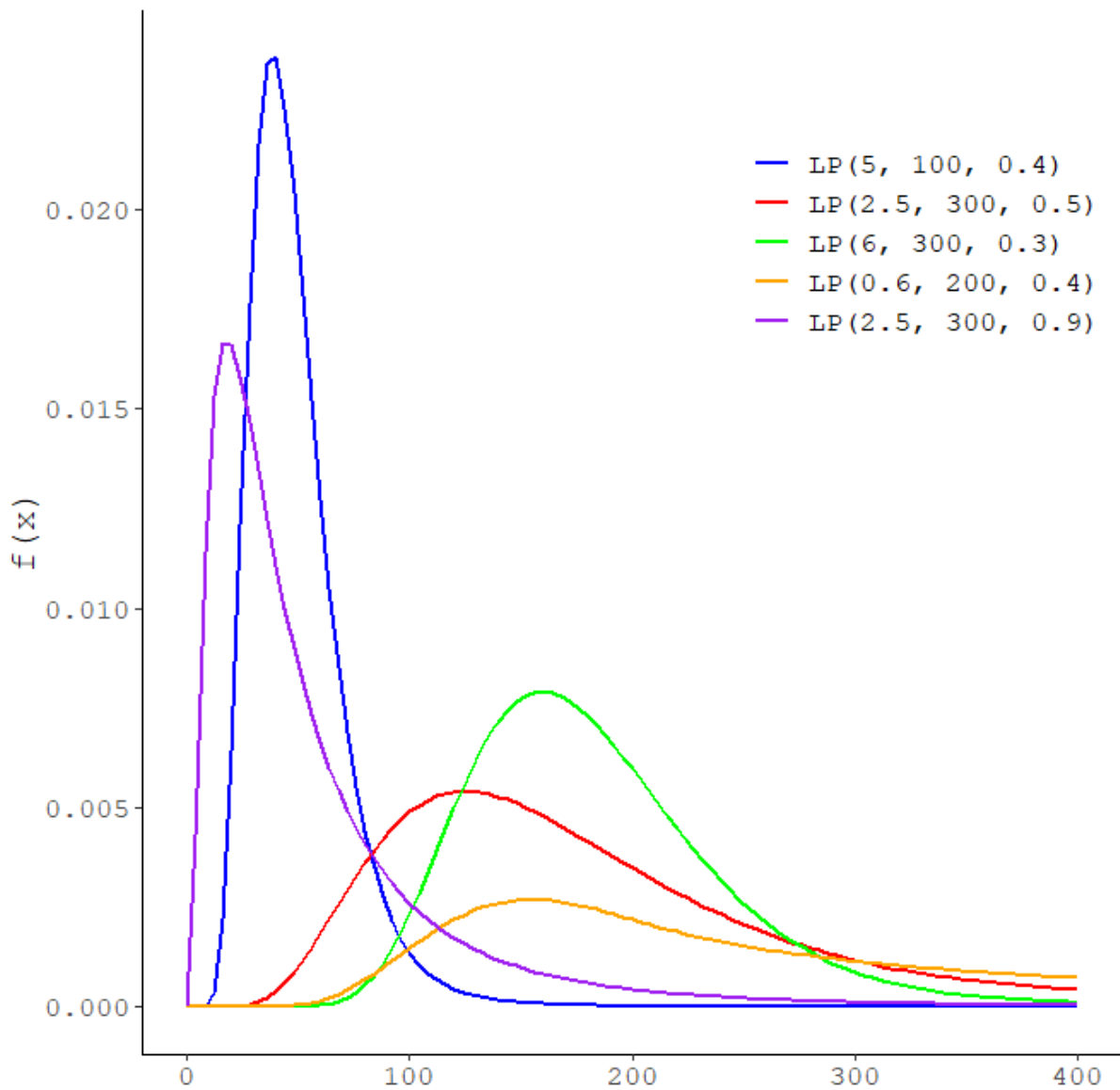
Στην παραπάνω εξίσωση εφαρμόζουμε την σχέση 3.7 οπότε καταλήγουμε στην ζητούμενη έκφραση για την παράμετρο  $\mu$ .

$$\mu = \ln\theta - \alpha\sigma^2$$

Επομένως, ισχύει η εξής απλοποιημένη έκφραση για την παράμετρο  $r$

$$r = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)e^{\frac{(\alpha\sigma)^2}{2}}}{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)e^{\frac{(\alpha\sigma)^2}{2}} + 1}$$

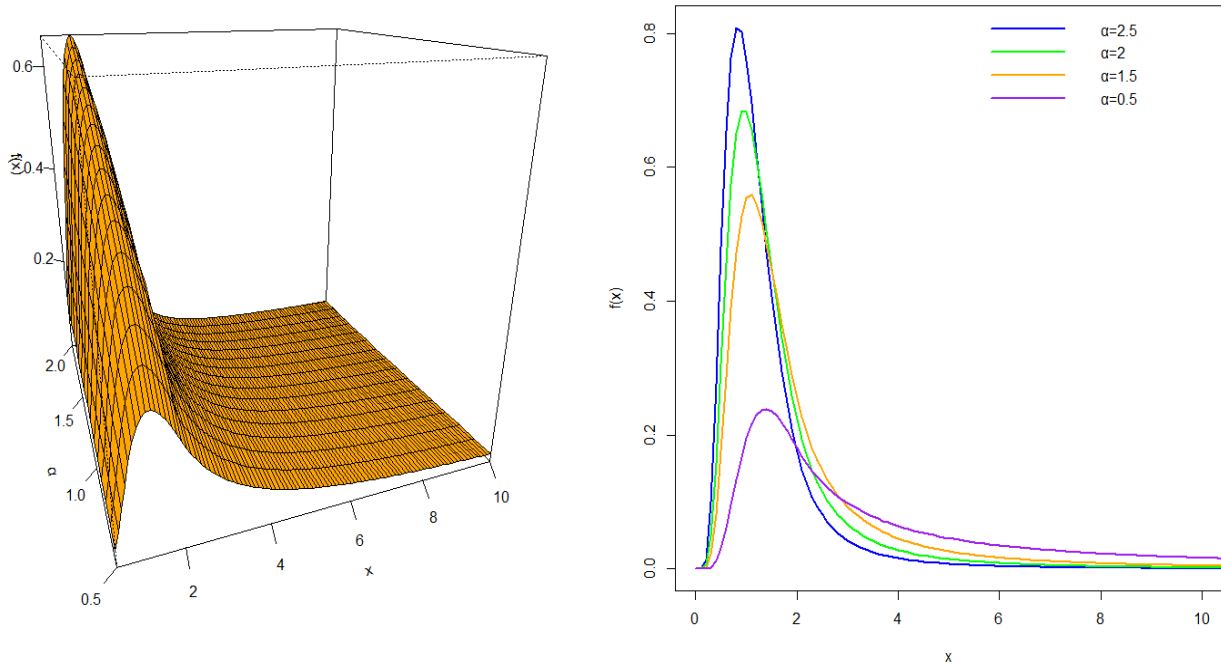
Παρατηρούμε ότι στο μοντέλο του Scollnik, σε αντίθεση με το μοντέλο των Cooray – Ananda, το βάρος μίξης  $r$  δεν είναι σταθερός και εκ των προτέρων γνωστός αριθμός αλλά προσδιορίζεται με βάση τις υπόλοιπες παραμέτρους και ισχύει ότι  $0 \leq r \leq 1$ .



**Σχήμα 3.2** Συναρτήσεις πυκνότητας της κατανομής  $LP(\alpha, \theta, \sigma)$

**Σχόλιο.** Έπειτα από τις υποθέσεις συνέχειας και διαφορισιμότητας για την σ.π.π της  $LP(\alpha, \theta, \sigma, \mu, r)$  οι άγνωστες παράμετροι μειώθηκαν από πέντε σε τρεις. Συνεπώς για τον προσδιορισμό της συγκολλημένης κατανομής  $LP$  απαιτούνται μόνο οι παράμετροι  $\alpha, \theta, \sigma$  καθώς οι παράμετροι  $r, \mu$  εκφράζονται συναρτήσει αυτών. Δηλαδή ισχύει:

$$LP(\alpha, \theta, \sigma, \mu, r) \equiv LP(\alpha, \theta, \sigma)$$



Σχήμα 3.3 3-Δ και 2-Δ απεικόνιση της  $LP(\alpha, 2, 0.5)$  για επιλεγμένες τιμές της παραμέτρου  $\alpha$

Από την πρόταση 2.2 καταλήγουμε στον παρακάτω τύπο για την συνάρτηση κατανομής της  $LP(\alpha, \theta, \sigma)$ :

$$F(x; \alpha, \theta, \sigma) = \begin{cases} r \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, & 0 < x < \theta \\ r + (1-r) \left[1 - \left(\frac{\theta}{x}\right)^\alpha\right], & x \geq \theta \end{cases}$$

Συνεπώς, εύκολα καταλήγουμε στην συνάρτηση δεξιάς ουράς:

$$\bar{F}(x; \alpha, \theta, \sigma) = \begin{cases} 1 - r \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, & 0 < x \leq \theta \\ (1-r) \left(\frac{\theta}{x}\right)^\alpha, & \theta < x < \infty \end{cases}$$

Το βελτιωμένο μοντέλο του Scollnik εφαρμόστηκε στο σει δεδομένων “Danish fire insurance” (Scollnik 2007, Scollnik – Sun 2012) και συγκρίθηκε με το μοντέλο των Cooray – Ananda ως προς τις τιμές των LL, AIC, BIC. Τα αποτελέσματα έδειξαν πως το ανανεωμένο μοντέλο υπερτερεί σημαντικά έναντι του αρχικού συνεπώς η μεθοδολογία του Scollnik οδηγεί σε καλύτερα αποτελέσματα κατά την μοντελοποίηση και την περιγραφή των δεδομένων.

**Πρόταση 3.2.** Οι ροπές  $k$  - τάξης της κατανομής LP ( $\alpha, \theta, \sigma$ ) δίνονται μέσω του παρακάτω τύπου:

$$E(X^k) = r \frac{e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{\ln\theta - (\mu + k\sigma^2)}{\sigma}\right)}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} + (1-r) \frac{\alpha\theta^k}{\alpha - k}, \quad a > k$$

**Απόδειξη.** Έχει ήδη αναφερθεί ότι η τ.μ  $X \sim \text{LP}(\alpha, \theta, \sigma)$  είναι η διακριτή μίξη των τ.μ  $Y_1^*, Y_2^*$  για τις οποίες ισχύει:

$$Y_1^* = X_1 | X_1 < \theta$$

και

$$Y_2^* = X_2 | X_2 > \theta$$

όπου  $X_1 \sim \text{Lognormal}(\mu, \sigma^2)$  και  $X_2 \sim \text{Pareto}(\alpha, \theta)$ . Τότε από την Πρόταση 2.4 ισχύει:

$$E(X^k) = rE(Y_1^{*k}) + (1-r)E(Y_2^{*k})$$

Θα υπολογίσουμε αρχικά την ποσότητα  $E(Y_1^{*k})$ :

$$E(Y_1^{*k}) = \int_0^\infty x^k f_{X_1|X_1 < \theta}(x) dx = \frac{\int_0^\theta x^k f_{X_1}(x) dx}{P(X_1 < \theta)} = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \int_0^\theta x^{k-1} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} dx$$

Αν θέσουμε:

$$A = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)}$$

Και κάνουμε αλλαγή μεταβλητής στο ολοκλήρωμα ως εξής:

$$y = \ln x$$

$$dx = e^y dy$$

Τότε θα έχουμε:

$$\begin{aligned} E(Y_1^{*k}) &= A \int_{-\infty}^{\ln\theta} e^{ky} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = A \int_{-\infty}^{\ln\theta} e^{ky} e^{-\left(\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2}\right)} dy \\ &= A \int_{-\infty}^{\ln\theta} e^{-\left(\frac{y^2 - 2y\mu + \mu^2 - 2ky\sigma^2}{2\sigma^2}\right)} dy \\ &= A \int_{-\infty}^{\ln\theta} e^{-\left(\frac{y^2 - 2y(\mu + k\sigma^2) + \mu^2 + (\mu + k\sigma^2)^2 - (\mu + k\sigma^2)^2}{2\sigma^2}\right)} dy \end{aligned}$$



$$\begin{aligned}
&= A e^{\frac{(\mu + \kappa \sigma^2)^2 - \mu^2}{2\sigma^2}} \int_{-\infty}^{\ln \theta} e^{-\frac{(y - (\mu + \kappa \sigma^2))^2}{2\sigma^2}} dy \\
&= A \sqrt{2\pi\sigma^2} e^{\frac{(\mu + \kappa \sigma^2)^2 - \mu^2}{2\sigma^2}} \int_{-\infty}^{\ln \theta} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - (\mu + \kappa \sigma^2))^2}{2\sigma^2}} dy
\end{aligned}$$

Αναγνωρίζουμε το ολοκλήρωμα ως την τιμή της συνάρτησης κατανομής της  $N(\mu + \kappa \sigma^2, \sigma^2)$  στο σημείο  $\ln \theta$ . Δηλαδή:

$$\int_{-\infty}^{\ln \theta} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - (\mu + \kappa \sigma^2))^2}{2\sigma^2}} dy = \Phi\left(\frac{\ln \theta - (\mu + \kappa \sigma^2)}{\sigma}\right)$$

Άρα για την μέση τιμή θα είναι:

$$E(Y_1^{*k}) = A \sqrt{2\pi\sigma^2} e^{\frac{(\mu + \kappa \sigma^2)^2 - \mu^2}{2\sigma^2}} \Phi\left(\frac{\ln \theta - (\mu + \kappa \sigma^2)}{\sigma}\right) = \frac{e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{\ln \theta - (\mu + \kappa \sigma^2)}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}$$

Στην συνέχεια θα υπολογίσουμε την ποσότητα  $E(Y_2^{*k})$ :

$$E(Y_2^{*k}) = \int_0^{\infty} x^k f_{X_2|X_2>\theta}(x) dx = \frac{\int_{\theta}^{\infty} x^k f_{X_2}(x) dx}{P(X_2 > \theta)} = \int_{\theta}^{\infty} \frac{x^k a \theta^a}{x^{a+1}} dx = a \theta^a \int_{\theta}^{\infty} x^{k-(a+1)} dx$$

Αν ισχύει ότι  $a > k$  τότε:

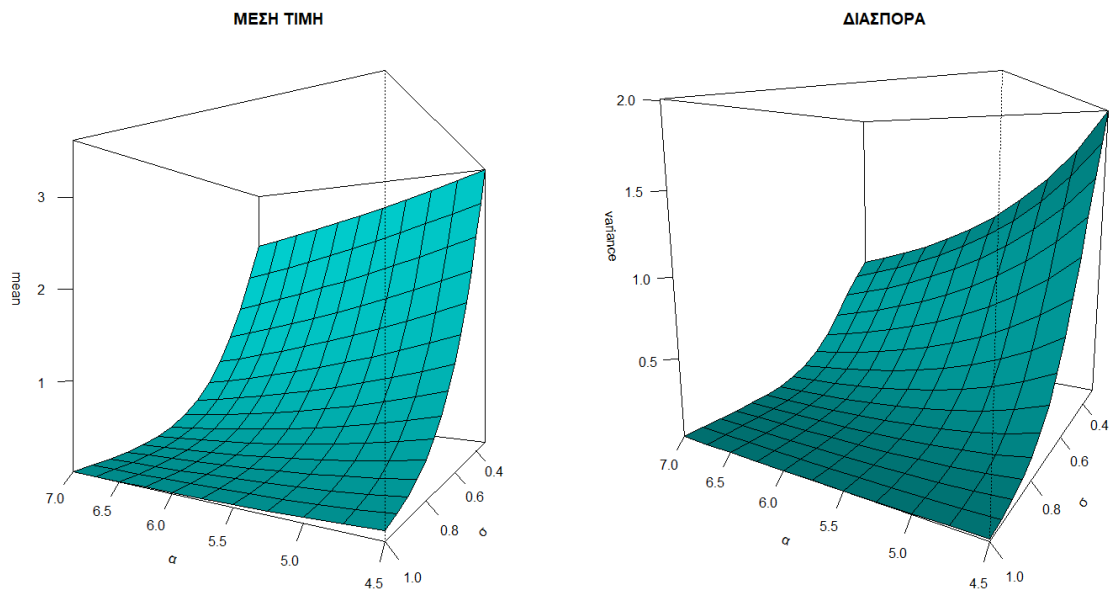
$$E(Y_2^{*k}) = \frac{\alpha \theta^k}{\alpha - k}$$

Συνεπώς, για την ροπή τάξης  $-k$  της τ.μ  $X$  θα ισχύει:

$$E(X^k) = r \frac{e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{\ln \theta - (\mu + \kappa \sigma^2)}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} + (1 - r) \frac{\alpha \theta^k}{\alpha - k}, \quad a > k$$

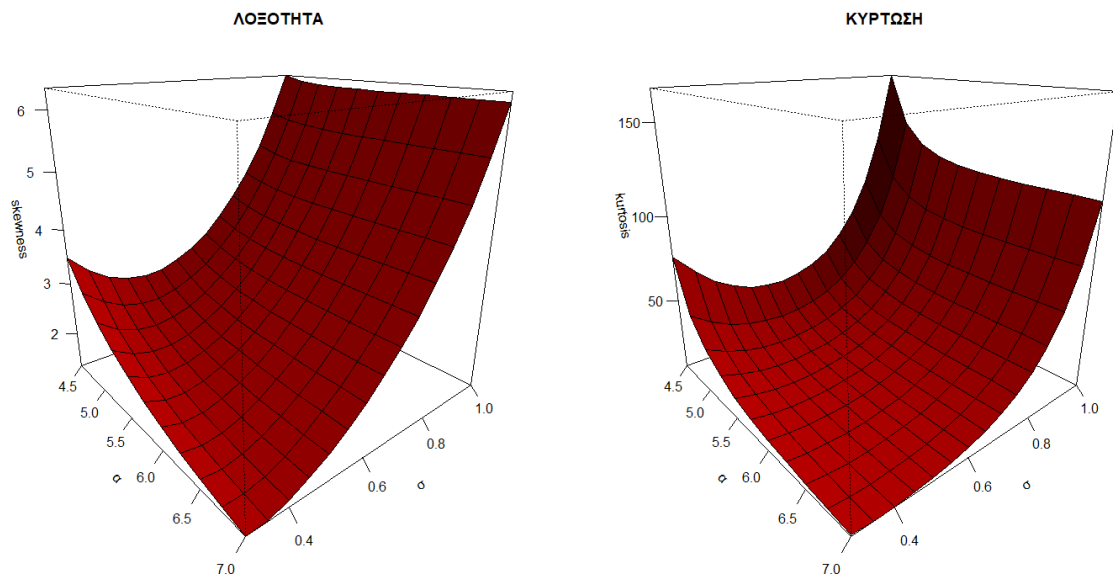
Συνεπώς, από τον παραπάνω κλειστό τύπο χρησιμοποιώντας τις ροπές  $k$  - τάξης για  $k = 1, 2, 3, 4$  μπορούμε να υπολογίσουμε την μέση τιμή, την διασπορά καθώς και τους συντελεστές κύρτωσης και λοξότητας της κατανομής LP  $(\alpha, \theta, \sigma)$  από τις σχέσεις 2.3 και 2.4.

Στο σχήμα 3.4 παρουσιάζεται η μέση τιμή και η διασπορά της κατανομής LP  $(\alpha, 5, \sigma)$  ως προς τις μεταβολές των παραμέτρων  $\alpha, \sigma$ .



Σχήμα 3.4 Μέση τιμή και Διασπορά της κατανομής  $LP(a, 5, \sigma)$  στις 3 διαστάσεις

στο σχήμα 3.5 παρουσιάζεται η λοξότητα και η κύρτωση της κατανομής  $LP(a, 5, \sigma)$  ως προς τις μεταβολές των παραμέτρων  $a, \sigma$ .



Σχήμα 3.5 Λοξότητα και Κύρτωση της κατανομής  $LP(a, 5, \sigma)$  στις 3 διαστάσεις

**Πρόταση 3.3.** Το Value at Risk της LP ( $\alpha, \theta, \sigma$ ) δίνεται από τον τύπο:

$$VaR_p(X) = \begin{cases} e^{\mu + \sigma \Phi^{-1}\left[\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]}, & 0 \leq p < r \\ \theta \left(\frac{1-p}{1-r}\right)^{-\frac{1}{\alpha}}, & r \leq p < 1 \end{cases}$$

**Απόδειξη.** Πράγματι, αν ακολουθήσουμε την μεθοδολογία της πρότασης 2.5 έχουμε ότι:

$$F(x) = p$$

Θέλουμε να λύσουμε την παραπάνω εξίσωση ως προς  $x$  δηλαδή να έχουμε:

$$x = F^{-1}(p)$$

Οπότε θα διακρίνουμε περιπτώσεις για το  $p$ :

Αν  $0 \leq p < r$  τότε:

$$\begin{aligned} F(x) = p &\Rightarrow r \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} = p \Rightarrow \frac{\ln x - \mu}{\sigma} = \Phi^{-1}\left(\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right) \\ &\Rightarrow x = e^{\mu + \sigma \Phi^{-1}\left[\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]} \end{aligned}$$

Άρα θα είναι:

$$VaR_p(X) = e^{\mu + \sigma \Phi^{-1}\left[\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]}$$

Αν  $r \leq p < 1$  τότε:

$$\begin{aligned} F(x) = p &\Rightarrow r + (1-r) \left[1 - \left(\frac{\theta}{x}\right)^\alpha\right] = p \Rightarrow \frac{p-r}{1-r} = 1 - \left(\frac{\theta}{x}\right)^\alpha \Rightarrow \left(\frac{x}{\theta}\right)^{-\alpha} = \frac{1-r}{1-p} \\ &\Rightarrow x = \theta \left(\frac{1-p}{1-r}\right)^{-\frac{1}{\alpha}} \end{aligned}$$

Άρα θα είναι:

$$VaR_p(X) = \theta \left(\frac{1-p}{1-r}\right)^{-\frac{1}{\alpha}}$$

**Πρόταση 3.4.** Το Tail Value at Risk της LP ( $\alpha, \theta, \sigma$ ) δίνεται από τον τύπο:

Αν  $VaR_p(X) \leq \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \frac{r}{1-p} \frac{e^{\mu + \frac{\sigma^2}{2}}}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} \left[ \Phi\left(\frac{\ln \theta - \mu - \sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln(VaR_p(X)) - \mu - \sigma^2}{\sigma}\right) \right] + \frac{1-r}{(1-p)(\alpha-1)} \frac{a\theta}{\alpha-1}$$

Αν  $VaR_p(X) > \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \frac{1-r}{(1-p)} \frac{\alpha\theta^\alpha}{(\alpha-1)} VaR_p(X)^{1-\alpha}$$

**Απόδειξη.** Πράγματι, από την πρόταση 2.6 έχουμε ότι:

Αν  $VaR_p(X) > \theta$  τότε είναι:

$$\begin{aligned} TVaR_p(X) &= \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx = \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} \frac{\alpha\theta^\alpha}{x^{\alpha+1}} dx = \\ &= \alpha\theta^\alpha \frac{1-r}{1-p} \int_{VaR_p(X)}^{\infty} x^{-\alpha} dx \end{aligned}$$

Αν ισχύει ότι  $\alpha > 1$  τότε είναι:

$$TVaR_p(X) = \frac{1-r}{(1-p)} \frac{\alpha\theta^\alpha}{(\alpha-1)} VaR_p(X)^{1-\alpha}$$

Αν  $VaR_p(X) \leq \theta$  τότε ισχύει:

$$TVaR_p(X) = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_{\theta}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Για ευκολία θα θέσουμε:

$$A_1 = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx$$

και

$$B_1 = \int_{\theta}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Άρα θα έχουμε:

$$A_1 = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx = \frac{r}{1-p} \int_{VaR_p(X)}^{\theta} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} dx$$

Κάνουμε αλλαγή μεταβλητής στο ολοκλήρωμα ως εξής:

$$x = e^{\mu+y}$$

$$dx = e^{\mu+y} dy$$

Και έχουμε:

$$\begin{aligned}
A_1 &= \frac{r}{1-p} \frac{1}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \int_{\ln(\text{VaR}_p(X)) - \mu}^{\ln\theta - \mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\mu+y} e^{-\frac{y^2}{2\sigma^2}} dy = \\
&= \frac{r}{1-p} \frac{e^{\mu+\frac{\sigma^2}{2}}}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \int_{\ln(\text{VaR}_p(X)) - \mu}^{\ln\theta - \mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\sigma^2)^2}{2\sigma^2}} dy
\end{aligned}$$

Αναγνωρίζουμε το ολοκλήρωμα ως την πιθανότητα μίας τ.μ που ακολουθεί την κατανομή  $N(\sigma^2, \sigma^2)$  να βρίσκεται μεταξύ των τιμών  $\ln(\text{VaR}_p(X)) - \mu$  και  $\ln\theta - \mu$ . Δηλαδή:

$$\int_{\ln(\text{VaR}_p(X)) - \mu}^{\ln\theta - \mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\sigma^2)^2}{2\sigma^2}} dy = \Phi\left(\frac{\ln\theta - \mu - \sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln(\text{VaR}_p(X)) - \mu - \sigma^2}{\sigma}\right)$$

Άρα, τελικά θα είναι:

$$A_1 = \frac{r}{1-p} \frac{e^{\mu+\frac{\sigma^2}{2}}}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \left[ \Phi\left(\frac{\ln\theta - \mu - \sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln(\text{VaR}_p(X)) - \mu - \sigma^2}{\sigma}\right) \right]$$

Έπειτα, αν ισχύει ότι  $\alpha > 1$  θα έχουμε:

$$B_1 = \frac{1-r}{1-p} \int_{\theta}^{\infty} \frac{\alpha\theta^\alpha}{x^\alpha} dx = \frac{1-r}{(1-p)(\alpha-1)} \alpha\theta$$

Δηλαδή θα είναι:

$$\text{TVaR}_p(X) = \frac{r}{1-p} \frac{e^{\mu+\frac{\sigma^2}{2}}}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \left[ \Phi\left(\frac{\ln\theta - \mu - \sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln(\text{VaR}_p(X)) - \mu - \sigma^2}{\sigma}\right) \right] + \frac{1-r}{(1-p)(\alpha-1)} \alpha\theta$$

### 3.3 Αριθμητική εφαρμογή των $\text{VaR}_p - \text{TVaR}_p$

Τα μεγέθη  $\text{VaR}_p$  και  $\text{TVaR}_p$  είναι σημαντικά διότι μέσα από την μελέτη αυτών μπορούμε να βγάλουμε συμπεράσματα για την ουρά της κατανομής, να συγκρίνουμε τα θεωρητικά ποσοστιαία σημεία με τα αντίστοιχα δειγματικά και να προσδιορίσουμε τα απαραίτητα ποσά που θα πρέπει, βάσει του μοντέλου, να εξασφαλιστούν για τα ακραία σενάρια (με πιθ. 95%, 99%, 99.5%).

Παρακάτω παρουσιάζεται η σύγκριση των μεγεθών αυτών για τις κατανομές  $\text{LP}(\alpha, \theta, \sigma)$ ,  $\text{Lognormal}(\mu, \sigma^2)$ ,  $\text{Pareto}(\alpha, \theta)$  για δύο διαφορετικά σενάρια παραμέτρων. Ορισμένα

από τα αριθμητικά αποτελέσματα συνοψίζονται στους πίνακες 3.1 και 3.2 ενώ παρουσιάζονται γραφικά στο σχήμα 3.6.

Επιλέχθηκε ως 1<sup>ο</sup> σετ παραμέτρων  $\alpha = 4, \theta = 20, \sigma = 0.2$  και ως 2<sup>ο</sup> σετ παραμέτρων  $\alpha = 5, \theta = 1, \sigma = 0.5$ . Από τις σχέσεις 3.5 και 3.6 μπορούμε να υπολογίσουμε και τις παραμέτρους  $\mu, r$  για το 1<sup>ο</sup> σετ είναι  $\mu = 2.835, r = 0.685$  ενώ για το 2<sup>ο</sup> σετ είναι  $\mu = -1.25, r = 0.992$ . Στην πρώτη περίπτωση όπου  $r = 0.685$  φαίνεται πως κατά την μίξη των κατανομών Lognormal, Pareto δίνεται περισσότερο βάρος στην 1<sup>η</sup> επιτρέποντας όμως ένα ποσοστό κοντά στο 31,5% για την 2<sup>η</sup> συνεπώς αναμένουμε την συγκολλημένη κατανομή LP να εμφανίζει αρκετά πιο βαριά ουρά από αυτή της Lognormal. Στην 2<sup>η</sup> περίπτωση όπου το βάρος μίξης είναι  $r = 0.992$  παρατηρούμε πως μόνο ένα ποσοστό κοντά στο 0.8% χρησιμοποιείται για την κατανομή Pareto συνεπώς η κατανομή LP πρόκειται να είναι παρόμοια με την Lognormal.

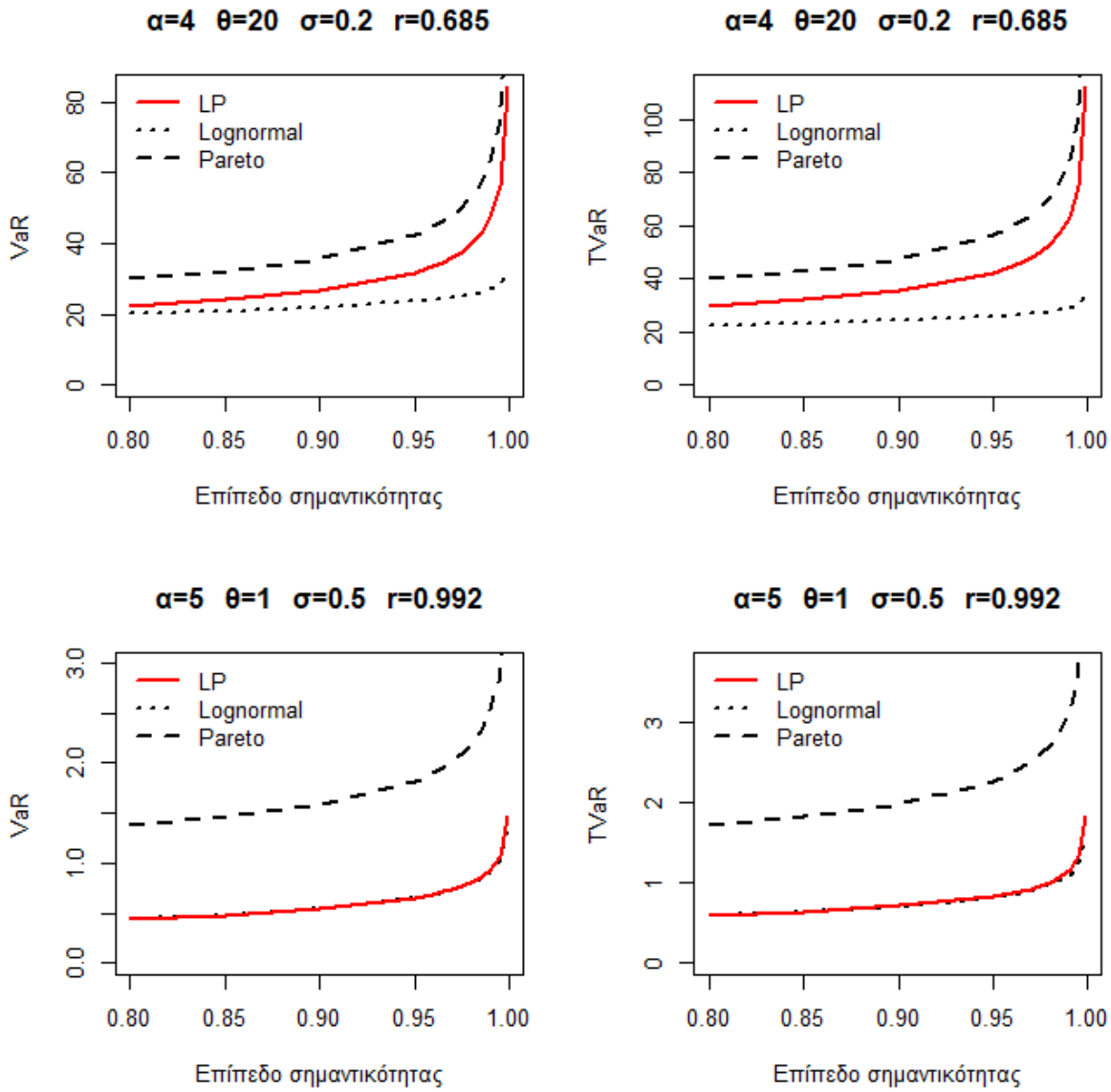
Παράμετροι	Επίπεδο σημαντικότητας	VaR		
		Lognormal	Pareto	LP
$\alpha=4$ $\theta=20$ $\sigma=0,2$	80%	20,167	29,906	22,401
	85%	20,968	32,137	24,072
	90%	22,022	35,565	26,64
	95%	23,681	42,294	31,681
	99%	27,139	63,245	47,374
	99.5%	28,528	75,212	56,337
$\alpha=5$ $\theta=1$ $\sigma=0,5$	80%	0,4364	1,379	0,436
	85%	0,481	1,461	0,481
	90%	0,543	1,584	0,544
	95%	0,652	1,82	0,654
	99%	0,916	2,511	0,931
	99.5%	1,038	2,885	1,069

**Πίνακας 3.1** Αριθμητικά αποτελέσματα των  $VaR_p$  ανά επίπεδο σημαντικότητας για τα δύο σετ παραμέτρων των κατανομών LP, Lognormal, Pareto.

Παράμετροι	Επίπεδο σημαντικότητας	TVaR		
		Lognormal	Pareto	LP
$\alpha=4$	80%	22,651	39,875	29,869
	85%	23,351	42,849	32,096
$\theta=20$	90%	24,294	47,42	35,52
	95%	25,819	56,393	42,241
$\sigma=0,2$	99%	29,101	84,327	63,165
	99.5%	30,444	100,282	75,117
$\alpha=5$	80%	0,594	1,724	0,599
	85%	0,64	1,826	0,646
$\theta=1$	90%	0,705	1,981	0,714
	95%	0,819	2,275	0,835
$\sigma=0,5$	99%	1,1	3,139	1,164
	99.5%	1,23	3,606	1,337

**Πίνακας 3.2** Αριθμητικά αποτελέσματα των  $TVaR_p$  ανά επίπεδο σημαντικότητας για τα δύο σετ παραμέτρων των κατανομών LP, Lognormal, Pareto.

Πράγματι, απο τους πίνακες 3.1 και 3.2 φαίνεται πως για το 1<sup>ο</sup> σετ η κατανομή LP έχει πιο βαριά ουρά σε σχέση με την Lognormal ενώ ταυτόχρονα η ουρά της φαίνεται να συγκλίνει στο μηδέν σημαντικά πιο γρήγορα από αυτή της Pareto. Για το 2<sup>ο</sup> σετ παρατηρούμε πως οι ουρές των Lognormal, LP είναι πανομοιότυπες κάτι το οποίο είναι λογικό εφόσον το περισσότερο βάρος (99.2%) έχει δοθεί στην κατανομή Lognormal.



Σχήμα 3.6 Διάγραμμα σύγκρισης των  $VaR_p - TVaR_p$  για τα δύο σετ παραμέτρων

### 3.4 Εκτιμητές μέγιστης πιθανοφάνειας για την LP ( $\alpha, \theta, \sigma, \mu, r$ )

Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την συγκολλημένη κατανομή LP( $\alpha, \theta, \sigma, \mu, r$ ) με συνάρτηση πυκνότητας πιθανότητας (σ.π.π):

$$f(x; \alpha, \theta, \sigma, \mu, r) = \begin{cases} r f_{Y_1}^*(x; \mu, \sigma) & , 0 < x \leq \theta \\ (1 - r) f_{Y_2}^*(x; \alpha, \theta) & , \theta < x < \infty \end{cases}$$

όπου

$$f_{Y_1}^*(x; \mu, \sigma) = \frac{(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{x\sigma \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, \quad x \leq \theta$$



και

$$f_{Y_2}^*(x; \alpha, \theta) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}, \quad x > \theta$$

Χωρίς βλάβη της γενικότητας υποθέτουμε ότι  $x_1 \leq x_2 \leq \dots \leq x_n$  και ότι υπάρχει θετικός ακέραιος αριθμός  $m$  τέτοιος ώστε:

$$x_m \leq \theta \leq x_{m+1}$$

Έχει αποδειχθεί μέσω της πρότασης 3.1 ότι:

$$LP(\alpha, \theta, \sigma, \mu, r) \equiv LP(\alpha, \theta, \sigma)$$

Συνεπώς, η συνάρτηση πιθανοφάνειας θα εξαρτάται μόνο από τις παραμέτρους  $\alpha, \theta, \sigma$  και θα είναι η εξής:

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \alpha, \theta, \sigma) &= \prod_{i=1}^n f(x_i; \alpha, \theta, \sigma) = \prod_{i=1}^m r f_{Y_1}^*(x_i; \mu, \sigma) \prod_{i=m+1}^n (1-r) f_{Y_2}^*(x_i; \alpha, \theta) \\ &= r^m (1-r)^{n-m} \prod_{i=1}^m f_{Y_1}^*(x_i; \mu, \sigma) \prod_{i=m+1}^n f_{Y_2}^*(x_i; \alpha, \theta) \end{aligned}$$

Υπενθυμίζουμε από την πρόταση 3.1 ότι οι παράμετροι  $\mu, r$  υπολογίζονται συναρτήσει των τριών παραμέτρων  $\alpha, \theta, \sigma$ . Θα χρησιμοποιήσουμε την δείκτρια συνάρτηση  $I_A$  η οποία ορίζεται ως εξής για  $A \subseteq \mathbb{R}$ :

$$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

Λογαριθμίζοντας την συνάρτηση πιθανοφάνειας θα έχουμε:

$$\begin{aligned} \ln(L(\alpha, \theta, \sigma)) &= l(\alpha, \theta, \sigma) \\ &= m \ln(r) + (n-m) \ln(1-r) + \sum_{i=1}^m \ln f_{Y_1}^*(x_i, \mu, \sigma) + \sum_{i=m+1}^n \ln f_{Y_2}^*(x_i, \alpha, \theta) \end{aligned}$$

Άρα θα είναι:

$$\begin{aligned} l(\alpha, \theta, \sigma) &= m \ln(r) + (n-m) \ln(1-r) - \sum_{i=1}^m \ln(x_i) - \frac{m}{2} \ln(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu]^2 - m \ln \left[ \Phi \left( \frac{\ln\theta - \mu}{\sigma} \right) \right] \\ &\quad + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) + (n-m) \ln(\alpha\theta^\alpha) - \end{aligned}$$

$$-(\alpha + 1) \sum_{i=m+1}^n \ln(x_i) + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) \quad (3.8)$$

Για τις παραμέτρους  $\mu, r$  ισχύει ότι υπολογίζονται συναρτήσεις των παραμέτρων  $\alpha, \theta, \sigma$  από τις σχέσεις 3.5 και 3.6. Συνεπώς είναι:

$$\mu = \mu(\alpha, \theta, \sigma)$$

και

$$r = r(\alpha, \theta, \sigma)$$

Η παραπάνω συνάρτηση καλείται συνάρτηση λογαριθμο-πιθανοφάνειας και πρέπει να μεγιστοποιηθεί ως προς τις παραμέτρους  $\alpha, \theta, \sigma$ . Θα πρέπει να λάβουμε υπόψη ότι τα στηρίγματα των πυκνοτήτων  $f_{Y_1}^*(x), f_{Y_2}^*(x)$  συμπεριλαμβάνονται στην σχέση 3.8 μέσω των δεικτριών συναρτήσεων.

Αν υπάρχει  $k \in \{1, \dots, m\}$ :  $x_k > \theta$  τότε θα είναι:

$$\ln(I_{\{x_k \leq \theta\}}) = -\infty$$

Επίσης αν υπάρχει  $k \in \{m+1, \dots, n\}$ :  $x_k \leq \theta$  τότε θα είναι:

$$\ln(I_{\{x_k > \theta\}}) = -\infty$$

Επομένως για την εύρεση των ε.μ.π θα πρέπει να λυθεί το παρακάτω σύστημα:

$$\begin{cases} \frac{\partial l(\alpha, \theta, \sigma)}{\partial \alpha} = 0 \\ \frac{\partial l(\alpha, \theta, \sigma)}{\partial \theta} = 0 \\ \frac{\partial l(\alpha, \theta, \sigma)}{\partial \sigma} = 0 \end{cases}$$

Για την 1<sup>η</sup> εξίσωση είναι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \sigma)}{\partial \alpha} &= \frac{m}{r} \frac{\partial r}{\partial \alpha} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \alpha} + \frac{\partial \mu}{\partial \alpha} \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu] - \\ &- m \frac{\partial \log \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \alpha} + (n-m) \left[ \ln(\theta) + \frac{1}{\alpha} \right] - \sum_{i=m+1}^n \ln(x_i) = 0 \end{aligned}$$

Για την 2<sup>η</sup> εξίσωση είναι:

$$\frac{\partial l(\alpha, \theta, \sigma)}{\partial \theta} = \frac{\partial \mu}{\partial \theta} \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu] + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) -$$

$$-m \frac{\partial \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \theta} + (n - m) \frac{a}{\theta} + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) = 0$$

Για την 3<sup>η</sup> εξίσωση είναι:

$$\frac{\partial l(\alpha, \theta, \sigma)}{\partial \sigma} = \frac{m}{r} \frac{\partial r}{\partial \sigma} - \frac{(n - m)}{1 - r} \frac{\partial r}{\partial \sigma} - \frac{m}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^m [\ln(x_i) - \mu]^2$$

$$-m \frac{\partial \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \sigma} + \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu] = 0$$

Παρατηρούμε πως η τιμή του ακεραίου  $m$  είναι άγνωστη και επηρεάζει την εξίσωση 3.8 συνεπώς θα πρέπει να λάβουμε υπόψη όλες τις δυνατές τιμές του  $m$  κατά την επίλυση του παραπάνω συστήματος. Ο παρακάτω αλγόριθμος για την εύρεση των ε.μ.π έχει προταθεί στην έρευνα των Cooray – Ananda (2005):

### **Βήμα 1**

Για κάθε τιμή του  $m = 1, 2, \dots, n - 1$  βρίσκουμε τις υποψήφιες ε.μ.π  $\hat{\alpha}, \hat{\theta}, \hat{\sigma}$  λύνοντας το παρακάτω σύστημα :

$$\begin{cases} \frac{\partial l(\alpha, \theta, \sigma)}{\partial \alpha} = 0 \\ \frac{\partial l(\alpha, \theta, \sigma)}{\partial \theta} = 0 \\ \frac{\partial l(\alpha, \theta, \sigma)}{\partial \sigma} = 0 \end{cases}$$

Αν για κάποια τιμή του  $m$  ισχύει ότι :

$$x_m \leq \hat{\theta} \leq x_{m+1}$$

Τότε ισχύει ότι:

$$\hat{\alpha} = \alpha^{\text{εμπ}}, \hat{\theta} = \theta^{\text{εμπ}}, \hat{\sigma} = \sigma^{\text{εμπ}}$$

### **Βήμα 2**

Αν δεν προκύψει λύση για το  $\theta$  που να ικανοποιεί την παραπάνω σχέση τότε υπάρχουν δύο περιπτώσεις: είτε  $m = n$  είτε  $m = 0$ , επομένως προτείνεται για την μοντελοποίηση των δεδομένων να χρησιμοποιηθεί μόνο μία κατανομή εκ των  $f_{Y_1}^*(x), f_{Y_2}^*(x)$ .

Η εξαγωγή ενός κλειστού τύπου για τον υπολογισμό των εκτιμητών μέγιστης πιθανοφάνειας είναι ανέφικτη συνεπώς ο υπολογισμός αυτών πραγματοποιείται με αριθμητικές μεθόδους στο στατιστικό πακέτο R με την βοήθεια της συνάρτησης “fitdistr” του πακέτου “MASS”. Στο

κεφάλαιο 7 θα πραγματοποιηθεί μελέτη προσομοίωσης ώστε να ελεγχθεί η ποιότητα των εκτιμητών μέγιστης πιθανοφάνειας που προκύπτουν από τις αριθμητικές βελτιστοποιήσεις του πακέτου R.

## ΚΕΦΑΛΑΙΟ 4: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ LOGNORMAL – GPD ( $\alpha, \vartheta, \lambda, \sigma, \mu, r$ )

### 4.1 Εισαγωγή

Ο Scollnik (2007) χρησιμοποίησε εκ νέου την έννοια της διακριτής μίξης κατανομών ώστε να δημιουργήσει ένα νέο μοντέλο που θα έχει ως βάση τις κατανομές Lognormal και Generalized Pareto distribution (GPD). Η συνάρτηση κατανομής της GPD ορίζεται ως εξής:

$$G(x) = \begin{cases} 1 - (1 + \xi x/\tau)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-x/\tau) & , \quad \xi = 0 \end{cases}$$

όπου  $\tau > 0$ , και έχει ως στήριγμα τα  $x \geq 0$  όταν  $\xi \geq 0$  ενώ έχει ως στήριγμα τα  $0 \leq x \leq -\frac{\tau}{\xi}$  όταν  $\xi < 0$ . Στην περίπτωση όπου  $\xi = 0$  το μοντέλο ανάγεται στην εκθετική κατανομή. Στην περίπτωση μας, μας ενδιαφέρει να μελετήσουμε δεδομένα από ασφαλιστικές ζημιές οπότε εστιάζουμε το ενδιαφέρον στην περίπτωση όπου  $x \geq 0$  άρα και  $\xi \geq 0$ . Το παραπάνω μοντέλο είναι γνωστό και ως Lomax ή Type II Pareto και η σ.κ έπειτα από αναπαραμέτρηση γράφεται ως εξής:

$$G(x) = 1 - \left(1 + \frac{x - \theta}{\alpha\beta}\right)^{-\alpha}, \quad x > \theta$$

όπου  $\theta > 0, \alpha > 0, \beta > 0$  και η σ.π.π μπορεί να γραφεί ως εξής:

$$g(x) = \frac{\alpha(\alpha\beta)^\alpha}{(\alpha\beta - \theta + x)^{\alpha+1}}, \quad x > \theta$$

όπου  $\theta > 0, \alpha > 0, \beta > 0$ . Εναλλακτικά αν θέσουμε  $\lambda = \alpha\beta - \theta$  τότε η σ.π.π είναι:

$$g(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta$$

Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π) της Lognormal( $\mu, \sigma^2$ ) δίνεται από τον τύπο:

$$f(x) = \frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}, \quad x > 0, \sigma > 0, \mu \in \mathbb{R}$$

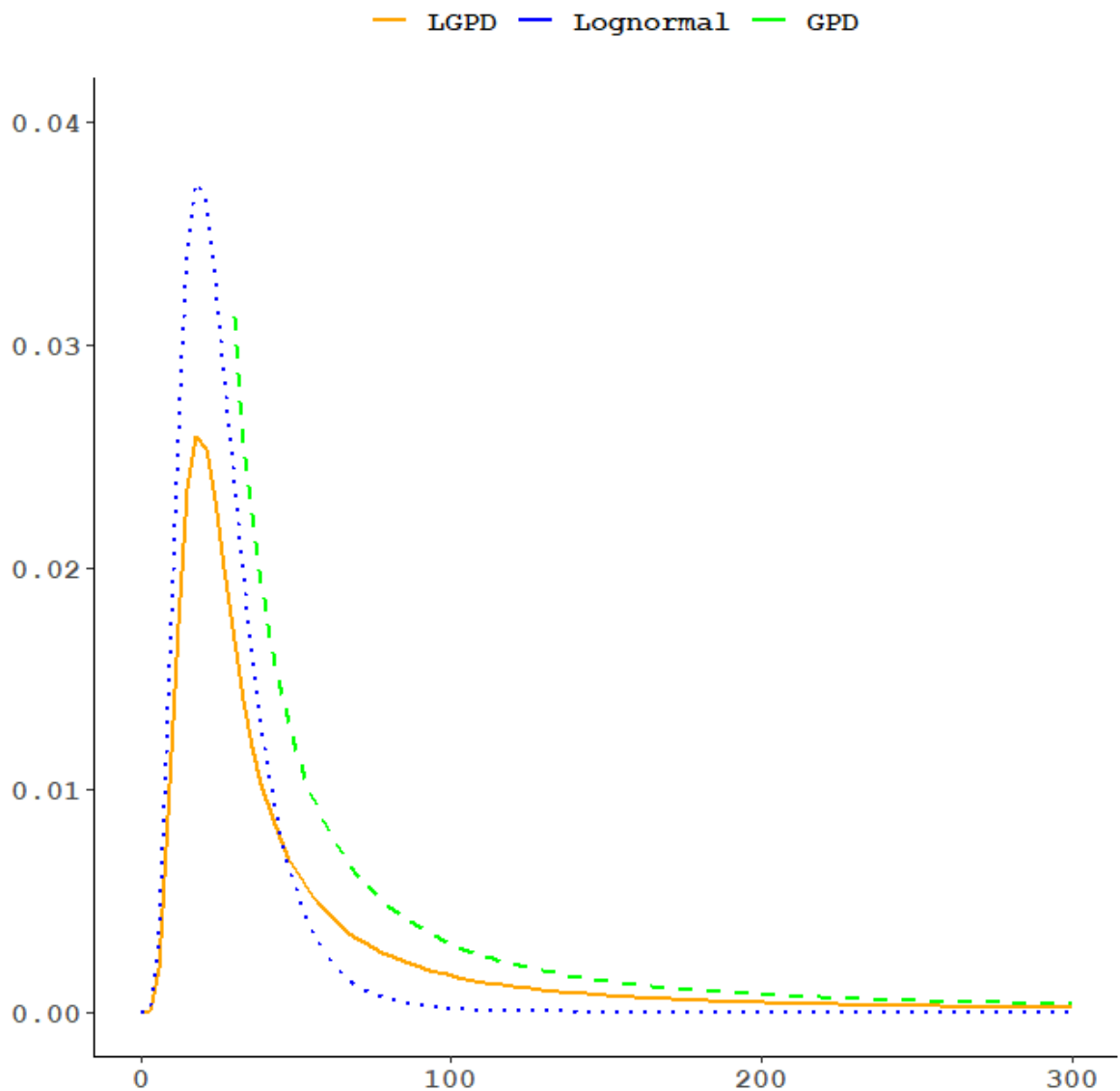
Η συγκολλημένη κατανομή Lognormal – GPD (LGPD) θα είναι η διακριτή μίξη των κατανομών Lognormal, GPD περικομμένων στο κατώφλι  $\theta$ .

Η σ.π.π της Lognormal( $\mu, \sigma^2$ ) κατανομής, περικομμένης στο σημείο  $\theta$ , δίνεται από τον παρακάτω τύπο:

$$f_{Y_1}^*(x) = \frac{\frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, \quad x \leq \theta, \sigma > 0, \mu \in \mathbb{R}$$

ενώ για την σ.π.π της περικομμένης κατανομής  $GPD(\alpha, \theta, \lambda)$  στο σημείο  $\theta$  θα ισχύει ότι:

$$f_{Y_2}^*(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \lambda > -\theta$$



**Σχήμα 4.1** Ενδεικτική γραφική παράσταση των κατανομών *LGPD*, *Lognormal*, *GPD*.

## 4.2 Χαρακτηριστικά μεγέθη της κατανομής LGPD ( $\alpha, \theta, \lambda, \sigma, \mu, r$ )

**Ορισμός.** Αν υποθέσουμε τ.μ  $X \sim \text{LGPD}(\alpha, \theta, \lambda, \sigma, \mu, r)$  τότε η συνάρτηση πυκνότητας πιθανότητας θα δίνεται από τον παρακάτω τύπο:

$$f(x; \alpha, \theta, \lambda, \sigma, \mu, r) = \begin{cases} r \frac{\frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} & , 0 < x \leq \theta \\ (1-r) \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}} & , \theta < x < \infty \end{cases}$$

όπου  $r$  είναι το βάρος μίξης,  $0 \leq r \leq 1$ ,  $\alpha > 0, \theta > 0, \sigma > 0, \mu \in \mathbb{R}, \lambda > -\theta$  οι άγνωστες παράμετροι της κατανομής.

Στο σχήμα 4.1 παρατηρούμε, όπως και στην περίπτωση της κατανομής LP ότι η κατανομή LGPD παρουσιάζει βαριά ουρά που συγκλίνει στο μηδέν πιο γρήγορα απ'ότι η ουρά της GPD αλλά πιο αργά απ'ότι η ουρά της Lognormal.

Ο Scollnik με την ίδια λογική όπως και στην κατανομή LP ( $\alpha, \theta, \sigma$ ) υπέθεσε την συνέχεια και την διαφορισιμότητα της σ.π.π της κατανομής LGPD ( $\alpha, \theta, \lambda, \sigma, \mu, r$ ) στο σημείο  $\theta$  με σκοπό να καταλήξει σε ένα πιο ομαλό μοντέλο. Οι υποθέσεις αυτές όπως και πριν, δημιουργούν συσχετίσεις μεταξύ των παραμέτρων και μειώνουν το πλήθος αυτών κατά δύο.

**Πρόταση 4.1.** Αν η  $f(x; \alpha, \theta, \lambda, \sigma, \mu, r)$  είναι συνεχής και διαφορίσιμη στο σημείο  $\theta$  τότε για τις παραμέτρους  $r, \mu$  θα ισχύουν οι σχέσεις:

$$\mu = \ln \theta - \left(\frac{\alpha\theta - \lambda}{\lambda + \theta}\right) \sigma^2 \quad (4.1)$$

και

$$r = \frac{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)e^{\frac{(\ln\theta - \mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)e^{\frac{(\ln\theta - \mu)^2}{2\sigma^2}} + \lambda + \theta} \quad (4.2)$$

**Απόδειξη.** Πράγματι, από την υπόθεση συνέχειας στο σημείο  $\theta$

$$\lim_{x \rightarrow \theta^-} f(x) = \lim_{x \rightarrow \theta^+} f(x)$$

προκύπτει η παρακάτω εξίσωση

$$\frac{r e^{-\frac{(\ln\theta - \mu)^2}{2\sigma^2}}}{\theta \sqrt{2\pi\sigma^2} \Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} = \frac{(1-r)\alpha(\lambda + \theta)^\alpha}{(\lambda + \theta)^{\alpha+1}}$$

λύνοντας την παραπάνω εξίσωση ως προς την παράμετρο  $r$  έχουμε ότι:

$$r = \frac{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right) e^{-\frac{(\ln\theta - \mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right) e^{-\frac{(\ln\theta - \mu)^2}{2\sigma^2}} + \lambda + \theta}$$

Έπειτα, από την υπόθεση διαφορισιμότητας στο σημείο  $\theta$

$$\lim_{x \rightarrow \theta^-} f'(x) = \lim_{x \rightarrow \theta^+} f'(x)$$

καταλήγουμε στην παρακάτω εξίσωση:

$$\frac{r e^{-\frac{(\ln\theta - \mu)^2}{2\sigma^2}}}{\theta^2 \sqrt{2\pi\sigma^2} \Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \left[1 + \frac{\ln\theta - \mu}{\sigma^2}\right] = (1-r)\alpha(\alpha+1)(\lambda + \theta)^{-2}$$

Αν αντικαταστήσουμε στην παραπάνω εξίσωση την παράμετρο  $r$  τότε για την παράμετρο  $\mu$  θα ισχύει η ζητούμενη σχέση:

$$\mu = \ln\theta - \left(\frac{\alpha\theta - \lambda}{\lambda + \theta}\right)\sigma^2$$

Συνεπώς η κατανομή LGPD προσδιορίζεται πλήρως αν γνωρίζουμε μόνο τις τέσσερις από τις έξι παραμέτρους. Ισχύει δηλαδή:

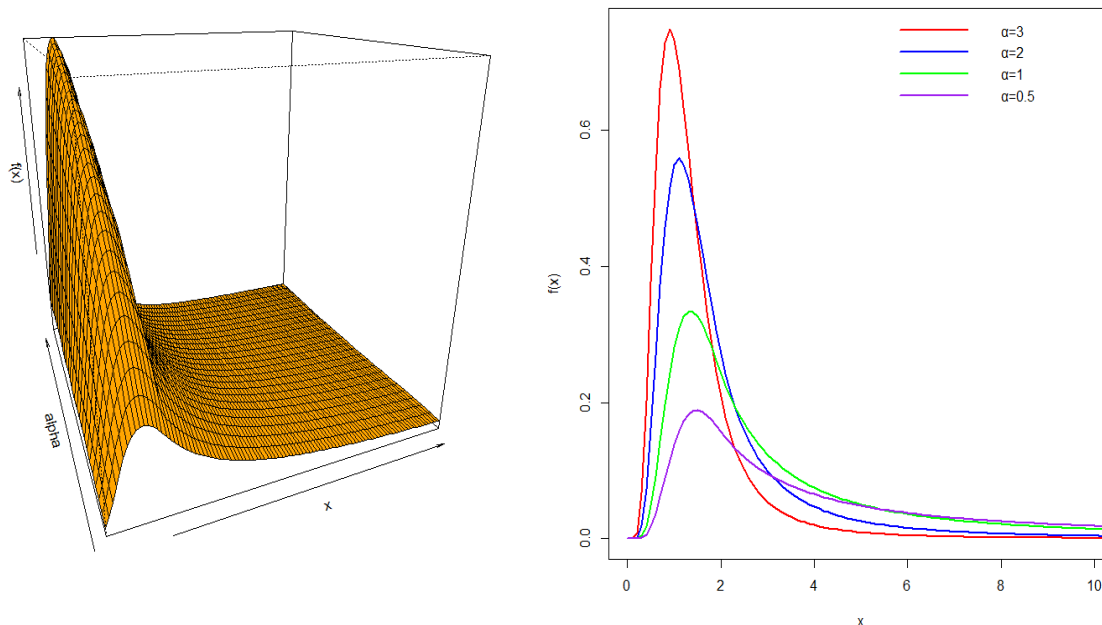
$$\text{LGPD}(\alpha, \theta, \lambda, \sigma, \mu, r) \equiv \text{LGPD}(\alpha, \theta, \lambda, \sigma)$$

Στην περίπτωση όπου  $\lambda = 0$  τότε ισχύει ότι:

$$\text{LGPD}(\alpha, \theta, \lambda, \sigma) \equiv \text{LP}(\alpha, \theta, \sigma)$$

Συνεπώς, η κατανομή LP είναι ειδική περίπτωση της LGPD.

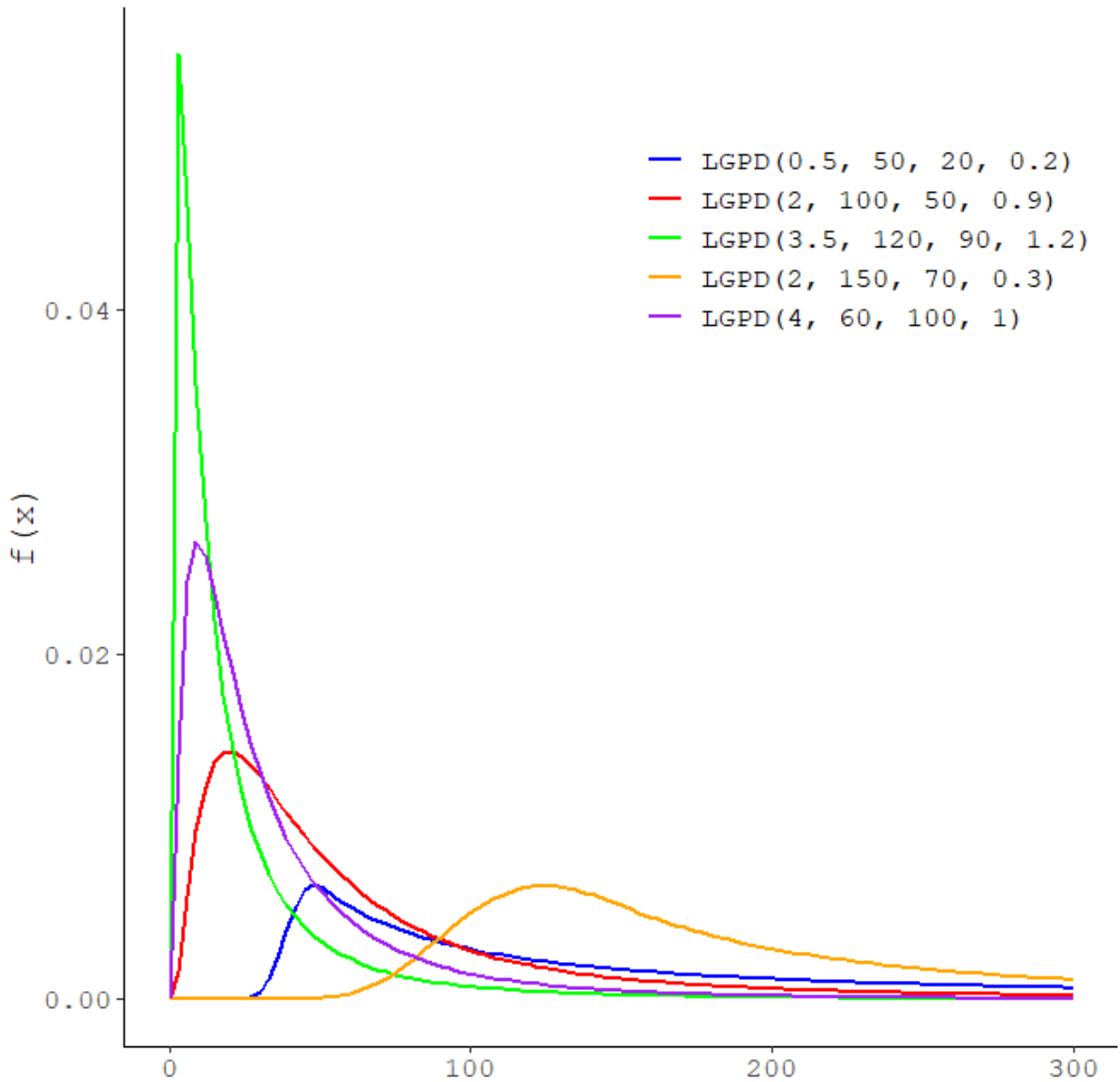




**Σχήμα 4.2** 3-Δ και 2-Δ απεικόνιση της  $LGPD(a, 2, 0.5, 0.5)$  για επιλεγμένες τιμές της παραμέτρου  $a$

Η κατανομή LGPD αναφέρεται στην βιβλιογραφία ως “Third composite model” και η απόδοση της έχει συγκριθεί πάνω στο σετ δεδομένων “Danish fire insurance” τόσο με το αρχικό μοντέλο των Cooray – Ananda όσο και με το προηγούμενο μοντέλο του Scollnik. Η νέα αυτή κατανομή οδηγεί σε καλύτερα αποτελέσματα σε σχέση με τις πρώτες συνεπώς επιλέγεται έναντι αυτών για την περιγραφή του συγκεκριμένου σετ δεδομένων (Scollnik 2007, Scollnik – Sun 2012).

Σχετικά με την επιλογή του σημείου  $\theta$  έχει προταθεί ένας εναλλακτικός τρόπος ο οποίος αναφέρεται στην έρευνα των Pigeon M., Denuit M. (2011). Σύμφωνα με αυτή την μεθοδολογία το σημείο  $\theta$  δεν αποτελεί άγνωστη παράμετρο του μοντέλου αλλά θεωρείται τ.μ η οποία ακολουθεί κάποια κατανομή, π.χ Lognormal, Gamma. Στην παραπάνω έρευνα κατασκευάζεται μία νέα συγκολλημένη κατανομή η οποία θεωρείται επέκταση του Second composite model. Για το κατώφλι  $\theta$  χρησιμοποιείται η κατανομή Lognormal. Η απόδοση του νέου μοντέλου είναι αισθητά καλύτερη σε σχέση με αυτή των First composite model και Second composite model ως προς τις τιμές των NLL, AIC ενώ δεν παρουσιάζει κάποια σημαντική βελτίωση σε σχέση με το Third composite model. Τα αποτελέσματα δείχνουν πως η εναλλακτική μέθοδος για την επιλογή του σημείου  $\theta$  δεν προσφέρει ουσιαστική βελτίωση ενώ ταυτόχρονα αποτελεί μία πιο πολύπλοκη διαδικασία για την κατασκευή των μοντέλων.



**Σχήμα 4.3** Συναρτήσεις πυκνότητας της  $LGPD(\alpha, \theta, \lambda, \sigma)$

Από την Πρόταση 2.2 μπορούμε να εξάγουμε την συνάρτηση κατανομής της  $LGPD(\alpha, \theta, \lambda, \sigma)$  η οποία δίνεται από τον παρακάτω τύπο:

$$F(x; \alpha, \theta, \lambda, \sigma) = \begin{cases} r \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} & , 0 < x < \theta \\ r + (1 - r) \left[ 1 - \left(\frac{\lambda + x}{\lambda + \theta}\right)^{-\alpha} \right] & , x \geq \theta \end{cases}$$

συνεπώς είναι εύκολο να καταλήξουμε στην συνάρτηση δεξιάς ουράς της κατανομής που περιγράφεται από τον ακόλουθο τύπο:

$$\bar{F}(x; \alpha, \theta, \lambda, \sigma) = \begin{cases} 1 - r \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, & 0 < x \leq \theta \\ (1 - r) \left(\frac{\lambda + x}{\lambda + \theta}\right)^{-\alpha}, & \theta < x < \infty \end{cases}$$

**Πρόταση 4.2.** Οι ροπές  $k$  - τάξης της κατανομής LGPD  $(\alpha, \theta, \lambda, \sigma)$  δίνονται μέσω του παρακάτω κλειστού τύπου:

$$E(X^k) = r \frac{e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{\ln \theta - (\mu + k\sigma^2)}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} + (1 - r) \alpha \sum_{m=0}^k \binom{k}{m} \frac{(-\lambda)^{k-m} (\lambda + \theta)^m}{\alpha - m}, \alpha > k$$

**Απόδειξη.** Ισχύει ότι η τ.μ  $X \sim \text{LGPD}(\alpha, \theta, \lambda, \sigma)$  είναι η διακριτή μίξη των τ.μ  $Y_1^*, Y_2^*$  με βάρη  $r, 1 - r$  αντίστοιχα για τις οποίες ισχύει:

$$Y_1^* = X_1 | X_1 < \theta$$

και

$$Y_2^* = X_2 | X_2 > \theta$$

όπου  $X_1 \sim \text{Lognormal}(\mu, \sigma^2)$  και  $X_2 \sim \text{GPD}(\alpha, \theta, \lambda)$  τότε από Πρόταση 2.4. ισχύει:

$$E(X^k) = rE(Y_1^{*k}) + (1 - r)E(Y_2^{*k})$$

Για τον υπολογισμό της μέσης τιμής  $E(Y_1^{*k})$  επαναλαμβάνουμε τα βήματα όπως στην πρόταση 3.2. Συνεπώς, μένει να υπολογισθεί η ποσότητα  $E(Y_2^{*k})$ :

Υπενθυμίζουμε το διωνυμικό ανάπτυγμα που θα εφαρμόσουμε στην συνέχεια της απόδειξης:

$$(x + a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$$

Έχουμε ότι:

$$E(Y_2^{*k}) = \int_{\theta}^{\infty} \frac{x^k a (\theta + \lambda)^a}{(\lambda + x)^{a+1}} dx$$

Συνεπώς, αν κάνουμε αλλαγή μεταβλητής στο ολοκλήρωμα ως εξής:

$$y = x + \lambda$$

$$dy = dx$$

Τότε καταλήγουμε στο εξής:

$$\int_{\theta}^{\infty} \frac{x^k a(\theta + \lambda)^a}{(\lambda + x)^{a+1}} dx = \int_{\lambda+\theta}^{\infty} \frac{(y - \lambda)^k a(\theta + \lambda)^a}{y^{a+1}} dy = a(\lambda + \theta)^\alpha \int_{\lambda+\theta}^{\infty} \frac{\sum_{m=0}^k \binom{k}{m} y^m (-\lambda)^{k-m}}{y^{a+1}} dy$$

$$= \alpha(\lambda + \theta)^\alpha \sum_{m=0}^k \binom{k}{m} (-\lambda)^{k-m} \int_{\lambda+\theta}^{\infty} y^{m-(a+1)} dy$$

Τότε, αν  $\alpha > m$  θα έχουμε:

$$= \frac{\alpha}{\alpha - m} (\lambda + \theta)^\alpha \sum_{m=0}^k \binom{k}{m} (-\lambda)^{k-m} (\lambda + \theta)^{m-\alpha}$$

Δηλαδή, θα είναι:

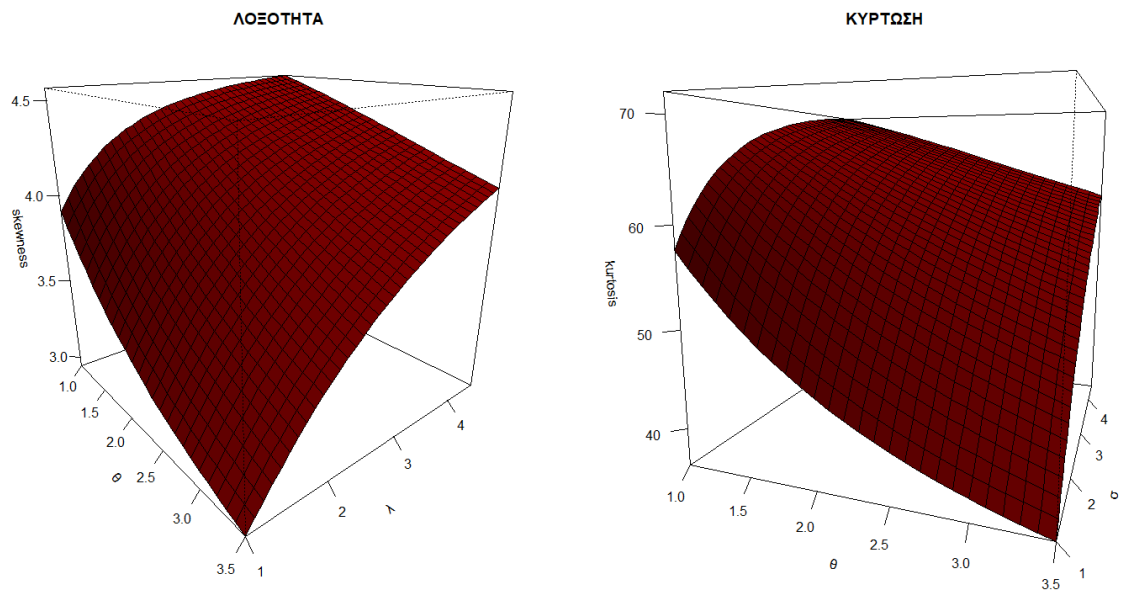
$$E(Y_2^{*k}) = \frac{\alpha}{\alpha - m} \sum_{m=0}^k \binom{k}{m} (-\lambda)^{k-m} (\lambda + \theta)^m$$

Συνεπώς, για την ροπή τάξης  $-k$  της τ.μ  $X$  θα ισχύει:

$$E(X^k) = r \frac{e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{\ln\theta - (\mu + k\sigma^2)}{\sigma}\right)}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} + (1 - r)\alpha \sum_{m=0}^k \binom{k}{m} \frac{(-\lambda)^{k-m} (\lambda + \theta)^m}{\alpha - m}, \alpha > k$$

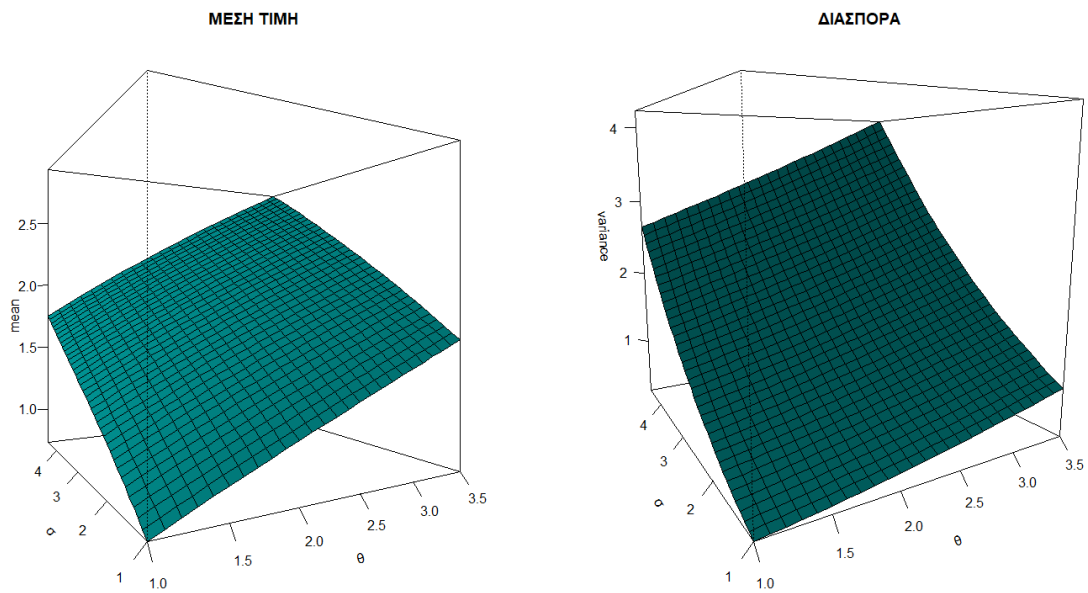
Από τον παραπάνω κλειστό τύπο για τις ροπές της κατανομής LGPD ( $\alpha, \theta, \lambda, \sigma$ ) μπορούμε μέσω των  $k$ -ροπών για  $k = 1, 2, 3, 4$  να υπολογίσουμε την μέση τιμή, την διασπορά καθώς και τους συντελεστές κύρτωσης και λοξότητας από τις σχέσεις 2.4 και 2.5.

Στο σχήμα 4.4 αναπαρίστανται γραφικά η λοξότητα και η κύρτωση της LGPD( $5, \theta, \lambda, 0.5$ ) για επιλεγμένες τιμές των παραμέτρων  $\theta, \lambda$ .



Σχήμα 4.4 Λοξότητα και Κύρτωση της κατανομής  $LGPD(5, \theta, \lambda, 0.5)$  στις 3 διαστάσεις

Στο σχήμα 4.5 αναπαρίστανται γραφικά η μέση τιμή και η διασπορά της  $LGPD(5, \theta, \lambda, 0.5)$  για επιλεγμένες τιμές των παραμέτρων  $\theta, \lambda$ .



Σχήμα 4.5 Μέση τιμή και Διασπορά της κατανομής  $LGPD(5, \theta, \lambda, 0.5)$  στις 3 διαστάσεις

**Πρόταση 4.3.** Το Value at Risk της LGPD ( $\alpha, \theta, \lambda, \sigma, \mu, r$ ) δίνεται από τον κάτωθι τύπο:

$$VaR_p(X) = \begin{cases} e^{\mu + \sigma \Phi^{-1}\left[\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]} & , 0 \leq p < r \\ (\lambda + \theta) \left(\frac{1-p}{1-r}\right)^{\frac{1}{\alpha}} - \lambda & , r \leq p < 1 \end{cases}$$

**Απόδειξη.** Πράγματι, έστω ότι:

$$F(x) = p$$

Θέλουμε να λύσουμε την παραπάνω εξίσωση ως προς  $x$  δηλαδή να έχουμε:

$$x = F^{-1}(p)$$

Οπότε θα διακρίνουμε περιπτώσεις για το  $p$ :

Αν  $0 \leq p < r$  τότε:

$$\begin{aligned} F(x) = p &\Rightarrow r \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)} = p \Rightarrow \frac{\ln x - \mu}{\sigma} = \Phi^{-1}\left(\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right) \\ &\Rightarrow x = e^{\mu + \sigma \Phi^{-1}\left[\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]} \end{aligned}$$

Άρα θα είναι:

$$VaR_p(X) = e^{\mu + \sigma \Phi^{-1}\left[\frac{p}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]}$$

Αν  $r \leq p < 1$  τότε:

$$\begin{aligned} F(x) = p &\Rightarrow r + (1-r) \left[1 - \left(\frac{\lambda + x}{\lambda + \theta}\right)^{-\alpha}\right] = p \Rightarrow \left(\frac{\lambda + x}{\lambda + \theta}\right)^{-\alpha} = \frac{1-p}{1-r} \\ &\Rightarrow \lambda + x = (\lambda + \theta) \left(\frac{1-p}{1-r}\right)^{-\frac{1}{\alpha}} \\ &\Rightarrow x = (\lambda + \theta) \left(\frac{1-p}{1-r}\right)^{\frac{1}{\alpha}} - \lambda \end{aligned}$$

Δηλαδή έχουμε:

$$VaR_p(X) = (\lambda + \theta) \left(\frac{1-p}{1-r}\right)^{\frac{1}{\alpha}} - \lambda$$

**Πρόταση 4.4.** Το Tail Value at Risk της LGPD ( $\alpha, \theta, \lambda, \sigma, \mu, r$ ) για καθορισμένο επίπεδο σημαντικότητας  $p$  δίνεται από τον παρακάτω τύπο:

Αν  $VaR_p(X) > \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \frac{1-r}{1-p} \alpha(\lambda + \theta)^\alpha \left[ \frac{(VaR_p(X) + \lambda)^{1-\alpha}}{\alpha - 1} - \frac{\lambda}{\alpha} (VaR_p(X) + \lambda)^{-\alpha} \right]$$

Αν  $VaR_p(X) \leq \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \frac{r}{1-p} \frac{e^{\mu + \frac{\sigma^2}{2}}}{\Phi\left(\frac{\ln\theta - \mu}{\sigma}\right)} \left[ \Phi\left(\frac{\ln\theta - \mu - \sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln(VaR_p(X)) - \mu - \sigma^2}{\sigma}\right) \right] +$$

$$+ \frac{1-r}{1-p} \alpha(\lambda + \theta)^\alpha \left[ \frac{(\lambda + \theta)^{1-\alpha}}{\alpha - 1} - \frac{\lambda}{\alpha} (\lambda + \theta)^{-\alpha} \right]$$

**Απόδειξη.** Πράγματι, από την πρόταση 2.6 έχουμε ότι:

Αν  $VaR_p(X) > \theta$  τότε:

$$TVaR_p(X) = \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx = \frac{1-r}{1-p} \int_{VaR_p(X)}^{\infty} \frac{x\alpha(\lambda + \theta)^\alpha}{(x + \lambda)^{\alpha+1}} dx$$

Κάνουμε αλλαγή μεταβλητής στο παραπάνω ολοκλήρωμα ως εξής:

$$y = x + \lambda$$

$$dy = dx$$

και έχουμε:

$$TVaR_p(X) = \frac{1-r}{1-p} \int_{VaR_p(X)+\lambda}^{\infty} \alpha(\lambda + \theta)^\alpha \frac{(y - \lambda)}{y^{\alpha+1}} dy =$$

$$= \frac{1-r}{1-p} \alpha(\lambda + \theta)^\alpha \left[ \int_{VaR_p(X)+\lambda}^{\infty} \frac{1}{y^\alpha} dy - \int_{VaR_p(X)+\lambda}^{\infty} \frac{\lambda}{y^{\alpha+1}} dy \right]$$

Αν ισχύει ότι  $\alpha > 1$  τότε είναι:

$$TVaR_p(X) = \frac{1-r}{1-p} \alpha(\lambda + \theta)^\alpha \left[ \frac{(VaR_p(X) + \lambda)^{1-\alpha}}{\alpha - 1} - \frac{\lambda}{\alpha} (VaR_p(X) + \lambda)^{-\alpha} \right]$$

Αν  $VaR_p(X) \leq \theta$  θα έχουμε:

$$TVaR_p(X) = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_{\theta}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Για ευκολία θα θέσουμε:

$$A_1 = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx$$

και

$$B_1 = \int_{\theta}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Το ολοκλήρωμα  $A_1$  έχει ήδη υπολογισθεί στην απόδειξη της Πρότασης 3.4. Οπότε μένει να υπολογίσουμε το ολοκλήρωμα  $B_1$ .

$$B_1 = \frac{1-r}{1-p} \int_{\theta}^{\infty} \frac{x\alpha(\lambda+\theta)^\alpha}{(x+\lambda)^{\alpha+1}} dx$$

Κάνουμε αλλαγή μεταβλητής στο παραπάνω ολοκλήρωμα ως εξής:

$$y = x + \lambda$$

$$dy = dx$$

και έχουμε:

$$B_1 = \frac{1-r}{1-p} \int_{\theta+\lambda}^{\infty} \alpha(\lambda+\theta)^\alpha \frac{(y-\lambda)}{y^{\alpha+1}} dy =$$

$$= \frac{1-r}{1-p} \alpha(\lambda+\theta)^\alpha \left[ \int_{\theta+\lambda}^{\infty} \frac{1}{y^\alpha} dy - \int_{\theta+\lambda}^{\infty} \frac{\lambda}{y^{\alpha+1}} dy \right]$$

Αν ισχύει ότι  $a > 1$  τότε είναι:

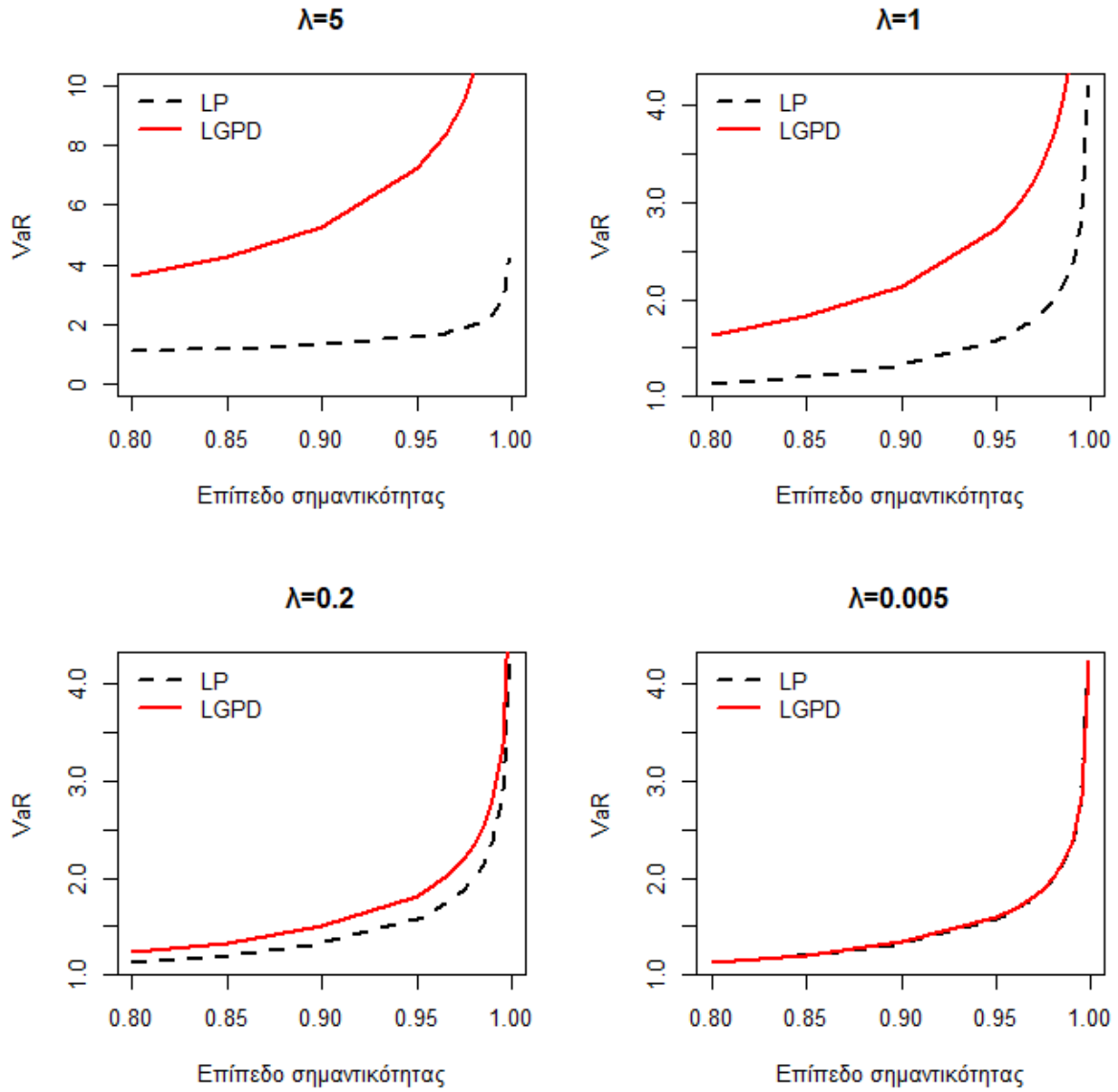
$$B_1 = \frac{1-r}{1-p} \alpha(\lambda+\theta)^\alpha \left[ \frac{(\theta+\lambda)^{1-\alpha}}{\alpha-1} - \frac{\lambda}{\alpha} (\theta+\lambda)^{-\alpha} \right]$$

### 4.3 Αριθμητική εφαρμογή για σύγκριση $VaR_p - TVaR_p$

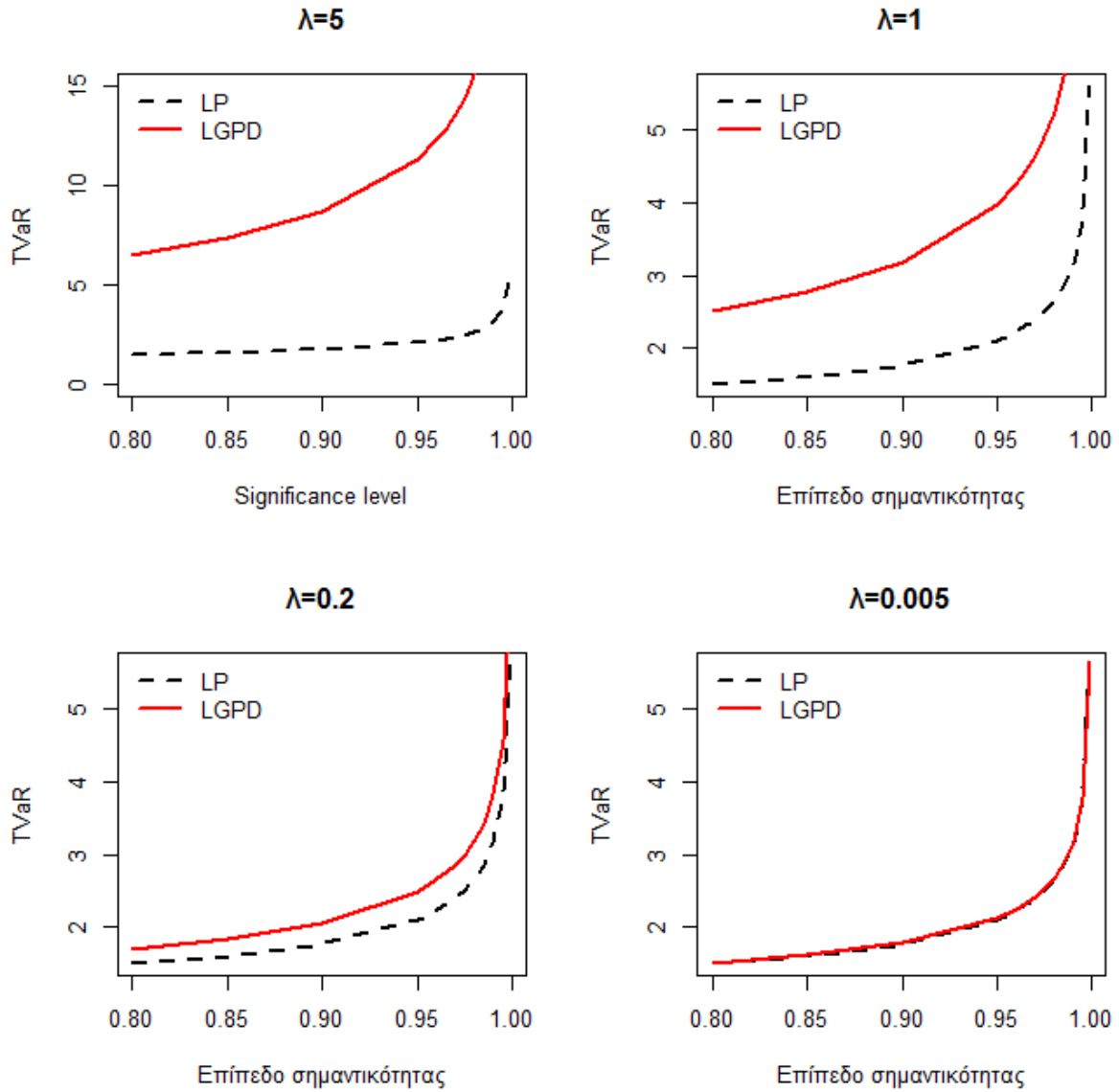
Όπως έχουμε ήδη αναφέρει η κατανομή  $LP(\alpha, \theta, \sigma)$  είναι ειδική περίπτωση της  $LGPD(\alpha, \theta, \lambda, \sigma)$  για  $\lambda = 0$ . Παρακάτω θα συγκρίνουμε γραφικά τα μεγέθη  $VaR_p - TVaR_p$  για παραμέτρους  $\alpha = 4, \theta = 1, \sigma = 0.2$  για διαφορετικές τιμές της παραμέτρου  $\lambda$ . Τα αποτελέσματα των παραπάνω μεγεθών είναι ικανά να μας περιγράψουν το πόσο πιο βαριά είναι η ουρά της μίας κατανομής σε σχέση με την άλλη. Όπως διαπιστώνουμε στα σχήματα 4.6 και 4.7 η κατανομή  $LGPD$  έχει πιο βαριά ουρά από την  $LP$  καθώς οι τιμές της παραμέτρου  $\lambda$  αυξάνονται ενώ για



μικρές τιμές του  $\lambda$  η ουρά της κατανομής LGPD εμφανίζει παρόμοια συμπεριφορά με την ουρά της LP.



**Σχήμα 4.6** Διαγράμματα σύγκρισης των  $VaR_p$  για τις κατανομές LP – LGPD



**Σχήμα 4.7** Διαγράμματα σύγκρισης των  $TVaR_p$  για τις κατανομές  $LP - LGPD$

#### 4.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $LGPD(\alpha, \theta, \lambda, \sigma, \mu, r)$

Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την συγκολλημένη κατανομή  $LGPD(\alpha, \theta, \lambda, \sigma, \mu, r)$  με συνάρτηση πυκνότητας πιθανότητας (σ.π.π):

$$f(x; \alpha, \theta, \sigma, \mu, r) = \begin{cases} r f_{Y_1}^*(x; \mu, \sigma) & , 0 < x \leq \theta \\ (1-r) f_{Y_2}^*(x; \alpha, \theta, \lambda) & , \theta < x < \infty \end{cases}$$

όπου

$$f_{Y_1}^*(x; \mu, \sigma) = \frac{\frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, \quad x \leq \theta$$

και

$$f_{Y_2}^*(x; \alpha, \theta, \lambda) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta$$

Χωρίς βλάβη της γενικότητας υποθέτουμε ότι  $x_1 \leq x_2 \leq \dots \leq x_n$  και ότι υπάρχει θετικός ακέραιος αριθμός  $m$  τέτοιος ώστε:

$$x_m \leq \theta \leq x_{m+1}$$

Έχει αποδειχθεί στην πρόταση 4.1. ότι:

$$LGPD(\alpha, \theta, \lambda, \sigma, \mu, r) \equiv LGPD(\alpha, \theta, \lambda, \sigma)$$

Συνεπώς, η συνάρτηση πιθανοφάνειας θα εξαρτάται μόνο από τις παραμέτρους  $\alpha, \theta, \lambda, \sigma$  και θα είναι η εξής:

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \alpha, \theta, \lambda, \sigma) &= \prod_{i=1}^n f(x_i; \alpha, \theta, \lambda, \sigma) = \prod_{i=1}^m r f_{Y_1}^*(x_i; \mu, \sigma) \prod_{i=m+1}^n (1-r) f_{Y_2}^*(x_i; \alpha, \theta, \lambda) \\ &= r^m (1-r)^{n-m} \prod_{i=1}^m f_{Y_1}^*(x_i; \mu, \sigma) \prod_{i=m+1}^n f_{Y_2}^*(x_i; \alpha, \theta, \lambda) \end{aligned}$$

Λογαριθμίζοντας την συνάρτηση πιθανοφάνειας έχουμε:

$$\begin{aligned} \ln(L(\alpha, \theta, \lambda, \sigma)) &= l(\alpha, \theta, \lambda, \sigma) \\ &= m \ln(r) + (n-m) \ln(1-r) + \sum_{i=1}^m \ln f_{Y_1}^*(x_i, \mu, \sigma) + \sum_{i=m+1}^n \ln f_{Y_2}^*(x_i, \alpha, \theta, \lambda) \end{aligned}$$

Συνεπώς θα είναι:

$$\begin{aligned} l(\alpha, \theta, \lambda, \sigma) &= m \ln(r) + (n-m) \ln(1-r) - \sum_{i=1}^m \ln(x_i) - \frac{m}{2} \ln(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu]^2 + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) - m \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right] \\ &\quad + (n-m) \ln[\alpha(\lambda + \theta)^\alpha] \\ &\quad - (\alpha + 1) \sum_{i=m+1}^n \ln(x_i + \lambda) + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) \end{aligned} \quad (4.3)$$

Για τις παραμέτρους  $\mu, r$  ισχύει ότι υπολογίζονται συναρτήσει των παραμέτρων  $\alpha, \theta, \lambda, \sigma$  από τις σχέσεις 4.1 και 4.2. Συνεπώς είναι:

$$\mu = \mu(\alpha, \theta, \lambda, \sigma)$$

και

$$r = r(\alpha, \theta, \lambda, \sigma)$$

Η παραπάνω συνάρτηση θα πρέπει να μεγιστοποιηθεί ως προς τις μεταβλητές  $\alpha, \theta, \lambda, \sigma$ . Παρατηρούμε πως ο ακέραιος  $m$  επηρεάζει την εξίσωση 4.3 καθώς είναι αδύνατο να γνωρίζουμε την πραγματική του τιμή. Για την αντιμετώπιση του προβλήματος προτείνεται ο αλγόριθμος της ενότητας 3.4. Για την εύρεση των ε.μ.π είναι αναγκαία η επίλυση του παρακάτω συστήματος:

$$\begin{cases} \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \alpha} = 0 \\ \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \theta} = 0 \\ \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \lambda} = 0 \\ \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \sigma} = 0 \end{cases}$$

Για την 1<sup>η</sup> εξίσωση θα είναι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \alpha} &= \frac{m}{r} \frac{\partial r}{\partial \alpha} - \frac{n-m}{1-r} \frac{\partial r}{\partial \alpha} + \frac{\partial \mu}{\partial \alpha} \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu] + (n-m) \left[ \ln(\lambda + \theta) + \frac{1}{\alpha} \right] \\ &\quad - \sum_{i=m+1}^n \ln(x_i + \lambda) - m \frac{\partial \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \alpha} = 0 \end{aligned}$$

Για την 2<sup>η</sup> εξίσωση είναι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \theta} &= \frac{m}{r} \frac{\partial r}{\partial \theta} - \frac{n-m}{1-r} \frac{\partial r}{\partial \theta} + \frac{\partial \mu}{\partial \theta} \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu] + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) + \frac{(n-m)a}{\lambda + \theta} \\ &\quad + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) - m \frac{\partial \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \theta} = 0 \end{aligned}$$

Για την 3<sup>η</sup> εξίσωση είναι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \lambda} &= \frac{m}{r} \frac{\partial r}{\partial \lambda} - \frac{n-m}{1-r} \frac{\partial r}{\partial \lambda} + \frac{\partial \mu}{\partial \lambda} \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu] + \frac{(n-m)a}{\lambda + \theta} \\ &\quad - (\alpha + 1) \sum_{i=m+1}^n \frac{1}{x_i + \lambda} - m \frac{\partial \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \lambda} = 0 \end{aligned}$$

Τέλος, για την 4<sup>η</sup> εξίσωση θα είναι:

$$\frac{\partial l(\alpha, \theta, \lambda, \sigma)}{\partial \sigma} = \frac{m}{r} \frac{\partial r}{\partial \sigma} - \frac{n-m}{1-r} \frac{\partial r}{\partial \sigma} - \frac{m}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^m [\ln(x_i) - \mu]^2 + \frac{\partial \mu}{\partial \sigma} \frac{1}{\sigma^2} \sum_{i=1}^m [\ln(x_i) - \mu]$$

$$- m \frac{\partial \ln \left[ \Phi \left( \frac{\ln \theta - \mu}{\sigma} \right) \right]}{\partial \sigma} = 0$$

Η επίλυση των παραπάνω εξισώσεων δεν μπορεί να πραγματοποιηθεί αναλυτικά συνεπώς η εκτίμηση των ε.μ.π θα επιτυγχάνεται με την βοήθεια αριθμητικών μεθόδων. Η εύρεση των ε.μ.π γίνεται μέσω του στατιστικού πακέτου R και της συνάρτησης “fitdistr”. Στο κεφάλαιο 7 θα πραγματοποιηθεί μελέτη προσομοίωσης με σκοπό να ελεγχθεί η απόδοση (μέση μεροληψία και μέσο τετρ. σφάλμα) των ε.μ.π που προκύπτουν από τις αριθμητικές βελτιστοποιήσεις.

## ΚΕΦΑΛΑΙΟ 5: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ GAMMA – GPD ( $\alpha, \beta, \gamma, \delta, \vartheta, r$ )

### 5.1 Εισαγωγή

Στην έρευνα των Teodorescu – Vernic (2013) παρουσιάζεται μία νέα συγκολλημένη κατανομή η οποία έχει ως συνιστώσες τις κατανομές Gamma και GPD. Ο τρόπος με τον οποίο όρισαν την συνάρτηση πυκνότητας βασίζεται στην μεθοδολογία του Scollnik, δηλαδή αν η τ.μ  $X$  ακολουθεί την συγκολλημένη κατανομή Gamma – GPD (GGPD) και το σύνολο των δεδομένων χωρίζεται σε ένα κατώφλι  $\theta$  τότε αυτή εκφράζεται ως διακριτή μίξη των περικομμένων κατανομών Gamma και GPD στο σημείο  $\theta$ .

Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π) της  $\text{Gamma}(\delta, \beta)$  δίνεται από τον τύπο:

$$f(x) = \frac{\beta^\delta x^{\delta-1} e^{-\beta x}}{\Gamma(\delta)}, \quad x > 0, \delta > 0, \beta > 0$$

ενώ η σ.π.π της κατανομής  $\text{GPD}(\alpha, \theta, \lambda)$  εκφράζεται από τον κάτωθι τύπο:

$$g(x) = \frac{\alpha(\alpha\beta)^\alpha}{(\alpha\beta - \theta + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \beta > 0$$

ο οποίος έπειτα από την αναπαράμετρηση  $\lambda = \alpha\beta - \theta$  γράφεται ως εξής:

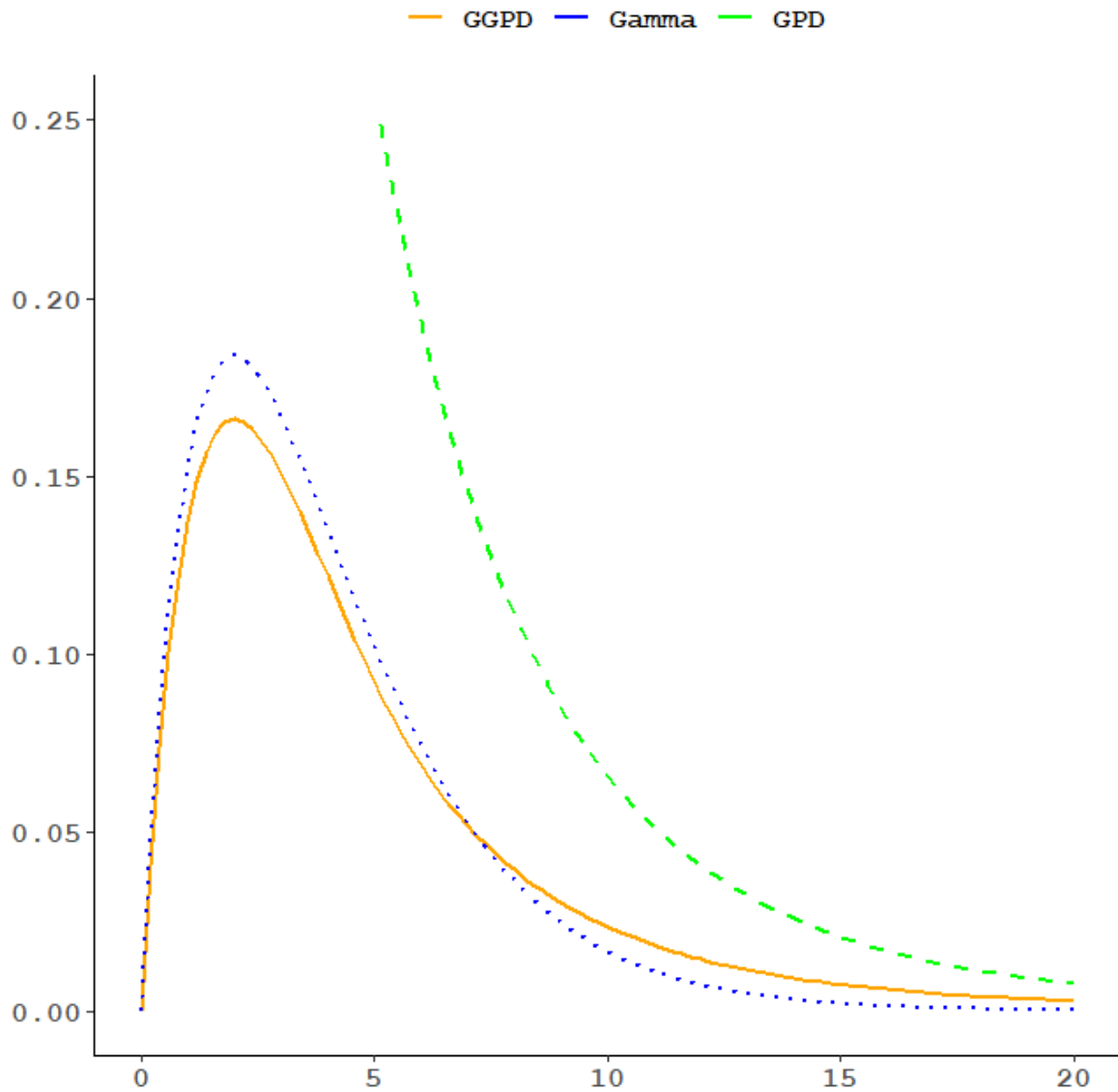
$$g(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \lambda > -\theta$$

Η σ.π.π της  $\text{Gamma}(\delta, \beta)$  κατανομής, περικομμένης στο σημείο  $\theta$ , δίνεται από τον παρακάτω τύπο:

$$f_{Y_1}^*(x) = \frac{\beta^\delta x^{\delta-1} e^{-\beta x}}{\Gamma(\delta; \beta\theta)}, \quad x \leq \theta, \delta > 0, \beta > 0$$

ενώ για την σ.π.π της περικομμένης κατανομής  $\text{GPD}(\alpha, \theta, \lambda)$  στο σημείο  $\theta$  θα ισχύει ότι:

$$f_{Y_2}^*(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \lambda > -\theta$$



Σχήμα 5.1 Ενδεικτική γραφική παράσταση των GGPD, Gamma, GPD

## 5.2 Χαρακτηριστικά μεγέθη της κατανομής GGPD ( $\alpha, \beta, \gamma, \delta, \theta, r$ )

**Ορισμός.** Αν υποθέσουμε τ.μ  $X \sim \text{GGPD}(\alpha, \beta, \gamma, \delta, \theta, r)$  τότε η συνάρτηση πυκνότητας πιθανότητας (σ.π.π) θα δίνεται από τον παρακάτω τύπο:

$$f(x; \alpha, \beta, \gamma, \delta, \theta, r) = \begin{cases} r \frac{\beta^\delta}{\Gamma(\delta; \beta\theta)} x^{\delta-1} e^{-\beta x}, & 0 < x \leq \theta \\ (1-r) \frac{\alpha(\gamma+\theta)^\alpha}{(\gamma+x)^{\alpha+1}}, & \theta < x < \infty \end{cases}$$

όπου  $r$  είναι το βάρος μίξης,  $0 \leq r \leq 1$ ,  $\alpha > 0, \beta > 0, \delta > 0, \theta > 0, \gamma > -\theta$  οι άγνωστες παράμετροι της κατανομής και  $\Gamma(\delta; \beta\theta) = \int_0^{\beta\theta} x^{\delta-1} e^{-x} dx$  είναι η μη-πλήρης συνάρτηση Γάμμα.

Με την ίδια λογική, όπως και ο Scollnik, οι Teodorescu – Vernic (2013) υπέθεσαν την συνέχεια και την διαφορισιμότητα της κατανομής GGPD. Οι υποθέσεις αυτές οδηγούν σε ένα πιο ομαλό μοντέλο και το πλήθος των παραμέτρων μειώνεται από έξι σε τέσσερις.

**Πρόταση 5.1.** Αν η  $f(x; \alpha, \beta, \gamma, \delta, \theta, r)$  είναι συνεχής και διαφορίσιμη στο σημείο  $\theta$  τότε για τις παραμέτρους  $r, \alpha$  θα ισχύουν οι σχέσεις:

$$\alpha = \frac{(\gamma + \theta)(\beta\theta - \delta + 1)}{\theta} - 1 \quad (5.1)$$

και

$$r = \frac{\alpha(\beta\theta - \delta + 1)\Gamma(\delta; \beta\theta)}{\alpha(\beta\theta - \delta + 1)\Gamma(\delta; \beta\theta) + (\alpha + 1)(\beta\theta)^\delta e^{-\beta\theta}} \quad (5.2)$$

**Απόδειξη.** Πράγματι, απο την υπόθεση συνέχειας στο σημείο  $\theta$ :

$$\lim_{x \rightarrow \theta^-} f(x) = \lim_{x \rightarrow \theta^+} f(x)$$

Καταλήγουμε στην παρακάτω σχέση:

$$\frac{r\beta^\delta \theta^{\delta-1} e^{-\beta\theta}}{\Gamma(\delta; \beta\theta)} = \frac{(1-r)a(\gamma + \theta)^\alpha}{(\gamma + \theta)^{\alpha+1}}$$

Λύνοντας την παραπάνω εξίσωση ως προς την παράμετρο  $r$  έχουμε ότι:

$$r = \frac{a\Gamma(\delta; \beta\theta)}{\alpha\Gamma(\delta; \beta\theta) + (\gamma + \theta)\beta^\delta \theta^{\delta-1} e^{-\beta\theta}} \quad (5.3)$$

Επίσης, απο την υπόθεση διαφορισιμότητας της σ.π.π στο σημείο  $\theta$ :

$$\lim_{x \rightarrow \theta^-} f'(x) = \lim_{x \rightarrow \theta^+} f'(x)$$

προκύπτει η ακόλουθη σχέση:

$$\frac{r\beta^\delta \theta^{\delta-1} e^{-\beta\theta}}{\Gamma(\delta; \beta\theta)} \left[ \beta - \frac{\delta - 1}{\theta} \right] = \frac{(1-r)\alpha(\alpha + 1)}{(\gamma + \theta)^2}$$

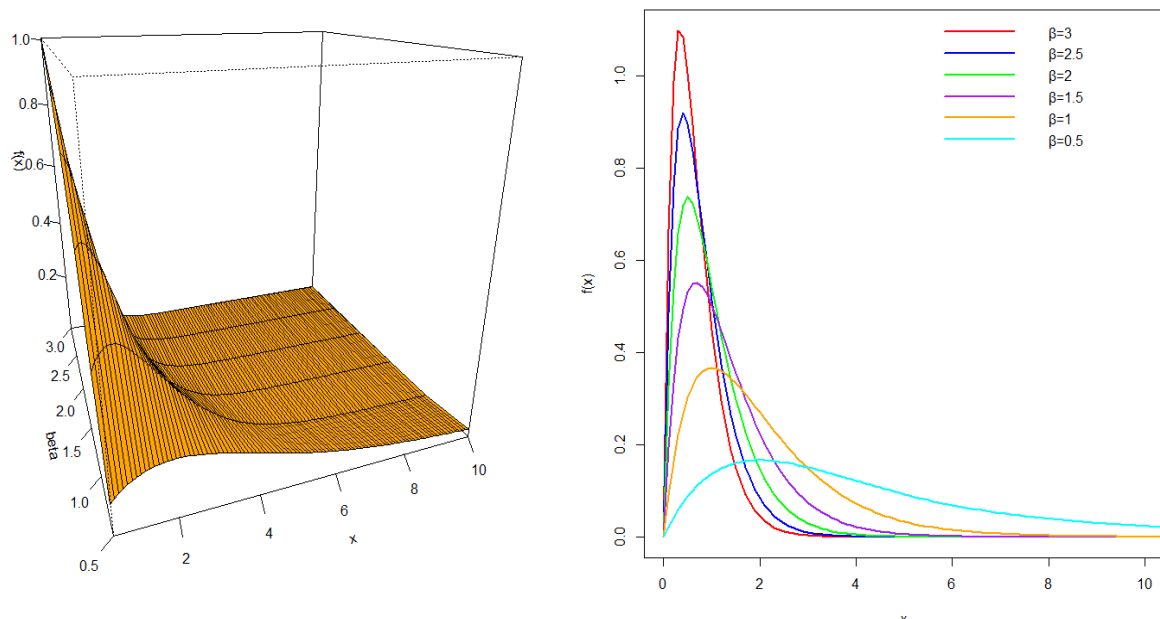
Η παραπάνω εξίσωση με την βοήθεια της σχέσης 5.3 μας δίνει την εξής έκφραση για την παράμετρο  $\alpha$ .

$$\alpha = \frac{(\gamma + \theta)(\beta\theta - \delta + 1)}{\theta} - 1$$

Συνεπώς, καταλήγουμε στην εξής απλοποιημένη έκφραση για την παράμετρο  $r$ :

$$r = \frac{\alpha(\beta\theta - \delta + 1)\Gamma(\delta; \beta\theta)}{\alpha(\beta\theta - \delta + 1)\Gamma(\delta; \beta\theta) + (\alpha + 1)(\beta\theta)^\delta e^{-\beta\theta}}$$





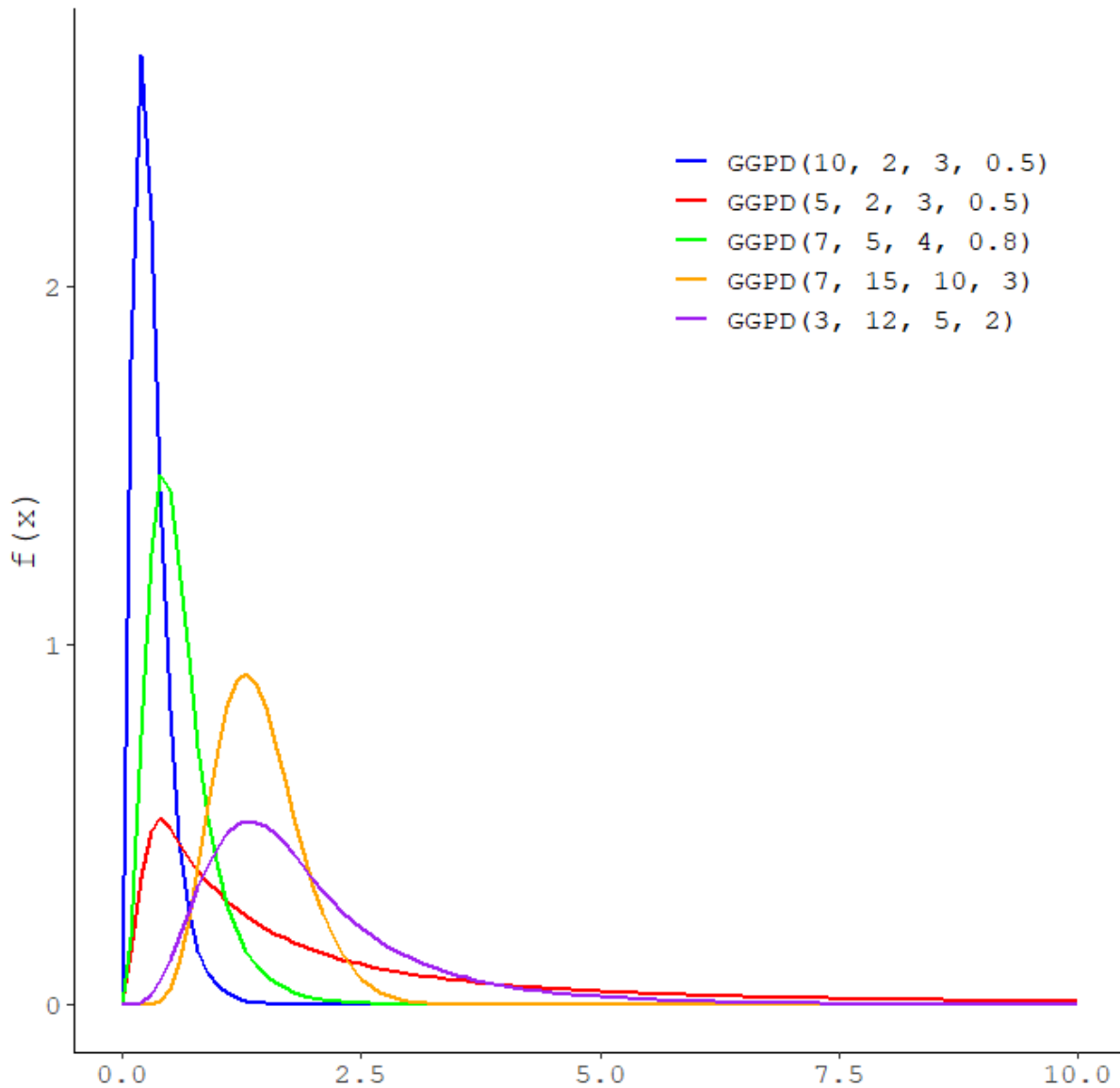
**Σχήμα 5.2** 3-Δ και 2-Δ απεικόνιση της  $GGPD(\beta, 20, 2, 5)$  για επιλεγμένες τιμές της παραμέτρου  $\beta$ .

Επομένως για προσδιορισθεί πλήρως η κατανομή GGPD αρκεί να γνωρίζουμε τις τέσσερις από τις έξι παραμέτρους, έστω τις  $\beta, \gamma, \delta, \theta$  καθώς οι παράμετροι  $\alpha, r$  γράφονται ως συναρτήσεις αυτών.

Δηλαδή ισχύει ότι:

$$GGPD(\alpha, \beta, \gamma, \delta, \theta, r) \equiv GGPD(\beta, \gamma, \delta, \theta)$$

Από την συγκολλημένη κατανομή GGPD μπορούν να προκύψουν και άλλα μοντέλα αυτής της κλάσης κατανομών. Για  $\delta = 1$  και  $\gamma = 0$  καταλήγουμε στο μοντέλο Exp – Pareto ενώ για  $\delta = 1$  καταλήγουμε στο Exp – GPD που περιγράφονται στις έρευνες των Teodorescu - Vernic (2006, 2009). Για  $\gamma = 0$  έχουμε το μοντέλο Gamma – Pareto που αναλύεται και αυτό στην έρευνα των Teodorescu – Vernic (2013).



**Σχήμα 5.3** Συναρτήσεις πυκνότητας της κατανομής  $GGPD(\beta, \gamma, \delta, \theta)$

Από την Πρόταση 2.2 προκύπτει ο παρακάτω τύπος για την συνάρτηση κατανομής της GGPD  $(\beta, \gamma, \delta, \theta)$

$$F(x; \beta, \gamma, \delta, \theta) = \begin{cases} r \frac{\Gamma(\delta; \beta x)}{\Gamma(\delta; \beta \theta)}, & 0 < x \leq \theta, \\ 1 - (1 - r) \left(\frac{\gamma + \theta}{\gamma + x}\right)^\alpha, & \theta < x < \infty \end{cases}$$

Συνεπώς εύκολα καταλήγουμε στον ακόλουθο τύπο για την συνάρτηση δεξιάς ουράς.

$$\bar{F}(x; \beta, \gamma, \delta, \theta) = \begin{cases} 1 - r \frac{\Gamma(\delta; \beta x)}{\Gamma(\delta; \beta \theta)}, & 0 < x \leq \theta \\ (1 - r) \left(\frac{\gamma + \theta}{\gamma + x}\right)^\alpha, & \theta < x < \infty \end{cases}$$

**Πρόταση 5.2.** Οι ροπές  $\kappa$  – τάξης της κατανομής  $GGPD(\beta, \gamma, \delta, \theta)$  δίνονται απο τον παρακάτω τύπο:

$$E(X^k) = r \frac{\Gamma(\kappa + \delta; \beta\theta)}{\beta^\kappa \Gamma(\delta; \beta\theta)} + (1-r)\alpha \sum_{m=0}^k \binom{k}{m} \frac{(-\gamma)^{k-m} (\gamma + \theta)^m}{\alpha - m}, \quad \alpha > \kappa$$

**Απόδειξη.** Πράγματι, ισχύει ότι η τ.μ  $X \sim \text{GGPD}(\beta, \gamma, \delta, \theta)$  είναι η διακριτή μίξη των τ.μ  $Y_1^*, Y_2^*$  με βάρη  $r$  και  $1-r$  αντίστοιχα.

με

$$Y_1^* = X_1 | X_1 < \theta$$

και

$$Y_2^* = X_2 | X_2 > \theta$$

όπου  $X_1 \sim \text{Gamma}(\delta, \beta)$  και  $X_2 \sim \text{GPD}(\alpha, \theta, \gamma)$ . Τότε ισχύει από Πρόταση 2.4 ότι:

$$E(X^k) = rE(Y_1^{*k}) + (1-r)E(Y_2^{*k})$$

Έχουμε ήδη παρουσιάσει τον υπολογισμό της μέσης τιμής  $E(Y_2^{*k})$  στην απόδειξη της πρότασης 4.2 συνεπώς μένει να υπολογίσουμε την μέση τιμή  $E(Y_1^{*k})$ . Για την ολοκλήρωση των παρακάτω υπολογισμών θα χρειαστούμε την μη – πλήρη συνάρτηση Γάμμα η οποία ορίζεται ως εξής:

$$\Gamma(\alpha; \beta) = \int_0^\beta x^{\alpha-1} e^{-x} dx$$

Συνεπώς, έχουμε:

$$\begin{aligned} E(Y_1^{*k}) &= \int_0^\infty x^k f_{X_1|X_1 < \theta}(x) dx = \frac{\int_0^\theta x^k f_X(x) dx}{P(X_1 < \theta)} = \frac{\int_0^\theta x^{k+\delta-1} e^{-\beta x} \beta^\delta dx}{\Gamma(\delta) P(X_1 < \theta)} \\ &= \frac{\int_0^\theta x^{k+\delta-1} e^{-\beta x} dx}{\int_0^\theta x^{\delta-1} e^{-\beta x} dx} \end{aligned}$$

Αν κάνουμε αλλαγή μεταβλητής στα παραπάνω ολοκληρώματα ως εξής:

$$\beta x = y$$

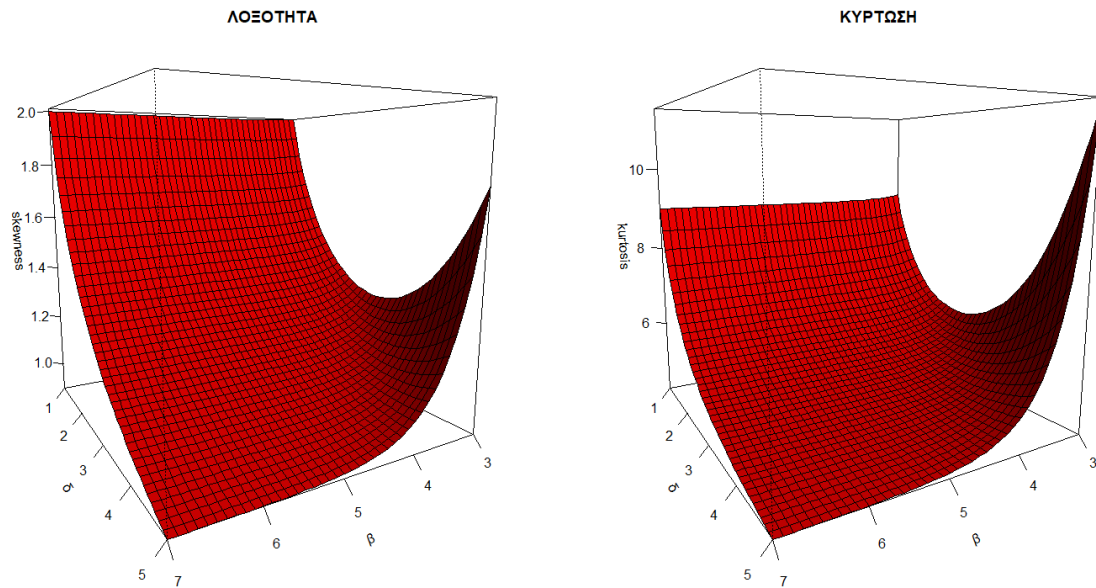
$$\beta dx = dy$$

Τότε θα έχουμε:

$$E(Y_1^{*k}) = \frac{\int_0^{\beta\theta} x^{k+\delta-1} e^{-y} dy}{\beta^k \int_0^{\beta\theta} x^{\delta-1} e^{-y} dy} = \frac{\Gamma(\kappa + \delta; \beta\theta)}{\beta^\kappa \Gamma(\delta; \beta\theta)}$$

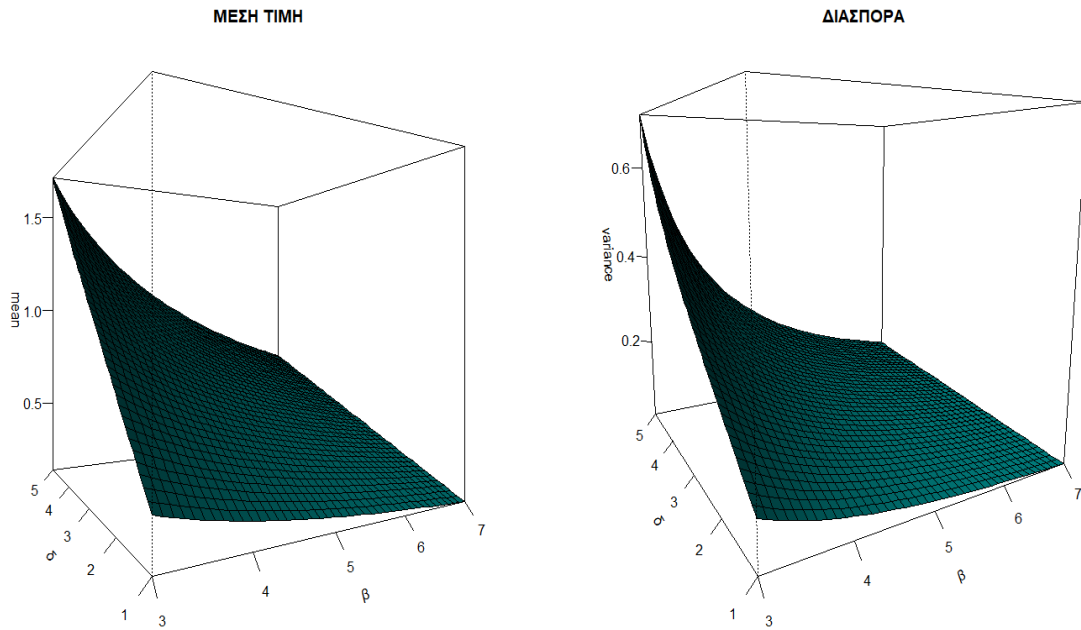
Από τον παραπάνω κλειστό τύπο για τις ροπές της κατανομής  $GGPD(\beta, \gamma, \delta, \theta)$  μπορούμε να υπολογίσουμε την μέση τιμή, την διασπορά και τους συντελεστές κύρτωσης και λοξότητας της κατανομής από τις σχέσεις 2.3 και 2.4.

Στο σχήμα 5.4 απεικονίζεται η λοξότητα και η κύρτωση της κατανομής  $GGPD(4, \gamma, \delta, 3)$  για επιλεγμένες τιμές των παραμέτρων  $\gamma, \delta$ .



**Σχήμα 5.4** Λοξότητα και Κύρτωση της κατανομής  $GGPD(\beta, 2, \delta, 3)$  στις 3 διαστάσεις

Στο σχήμα 5.5 απεικονίζεται η μέση τιμή και η διασπορά της κατανομής  $GGPD(4, \gamma, \delta, 3)$  για επιλεγμένες τιμές των παραμέτρων  $\gamma, \delta$ .



Σχήμα 5.5 Μέση τιμή και Διασπορά της κατανομής  $GGPD(\beta, 2, \delta, 3)$  στις 3 διαστάσεις

Για το Value at Risk της κατανομής  $GGPD(\beta, \gamma, \delta, \theta)$  δεν μπορούμε να καταλήξουμε σε κλειστό τύπο διότι η επίλυση της εξίσωσης

$$F(x) = p$$

για  $p \leq r$  απαιτεί τον υπολογισμό της μη – πλήρους συνάρτησης Γάμμα

$$\Gamma(\delta; \beta x) = \int_0^{\beta x} x^{\delta-1} e^{-x} dx$$

Η μη – πλήρης συνάρτηση Γάμμα είναι εφικτό να υπολογισθεί με την βοήθεια του στατιστικού πακέτου R και της συνάρτησης “gammainc” στο πακέτο “pracma”.

Συνεπώς, για τον υπολογισμό του  $VaR_p$  της κατανομής  $GGPD$  θα εφαρμόσουμε τις αριθμητικές μεθόδους για την επίλυση εξισώσεων που μας παρέχει το πακέτο R. Ο αλγόριθμος βρίσκεται στο παράρτημα.

**Πρόταση 5.3.** Για το μέτρο κινδύνου  $TVaR_p(X)$  της κατανομής  $GGPD(\beta, \gamma, \delta, \theta)$  ισχύει ο παρακάτω τύπος:

Αν  $VaR_p(X) > \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \alpha(\lambda + \theta)^\alpha \frac{1 - r}{1 - p} \left[ \frac{(VaR_p(X) + \lambda)^{1-\alpha}}{\alpha - 1} - \frac{\lambda}{\alpha} (VaR_p(X) + \lambda)^{-\alpha} \right]$$

Αν  $VaR_p(X) \leq \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \alpha(\lambda + \theta)^\alpha \frac{1-r}{1-p} \left[ \frac{(\theta + \lambda)^{1-\alpha}}{\alpha - 1} - \frac{\lambda}{\alpha} (\theta + \lambda)^{-\alpha} \right] + \frac{r}{\beta(1-p)\Gamma(\delta; \beta\theta)} \left[ \Gamma(\delta + 1; \beta\theta) - \Gamma(\delta + 1; \beta VaR_p(X)) \right]$$

**Απόδειξη.** Πράγματι, από την Πρόταση 2.6:

Αν  $VaR_p(X) \leq \theta$  τότε είναι:

$$TVaR_p(X) = \int_{VaR_p(X)}^d \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_d^\infty \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Για ευκολία θέτουμε

$$A_1 = \int_{VaR_p(X)}^d \frac{xr}{1-p} f_{Y_1}^*(x) dx$$

και

$$B_1 = \int_d^\infty \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Για το ολοκλήρωμα  $B_1$  έχουμε ήδη αποδείξει την διαδικασία υπολογισμού στην απόδειξη της πρότασης 4.4 συνεπώς για την περίπτωση όπου  $VaR_p(X) \leq \theta$  μένει να υπολογίσουμε το ολοκλήρωμα  $A_1$ . Θα έχουμε:

$$A_1 = \frac{r}{1-p} \int_{VaR_p(X)}^\theta \frac{x\beta^\delta x^{\delta-1} e^{-\beta x}}{\Gamma(\delta; \beta\theta)} dx$$

Στο παραπάνω ολοκλήρωμα κάνουμε αλλαγή μεταβλητής ως εξής:

$$\beta x = y$$

$$\beta dx = dy$$

Άρα θα έχουμε:

$$\begin{aligned} A_1 &= \frac{r\beta^\delta}{(1-p)\Gamma(\delta; \beta\theta)} \int_{\beta VaR_p(X)}^{\beta\theta} \frac{y^\delta}{\beta^{\delta+1}} e^{-y} dy = \\ &= \frac{r}{\beta(1-p)\Gamma(\delta; \beta\theta)} \left[ \int_0^{\beta\theta} y^\delta e^{-y} dy - \int_0^{\beta VaR_p(X)} y^\delta e^{-y} dy \right] = \end{aligned}$$

$$= \frac{r}{\beta(1-p)\Gamma(\delta; \beta\theta)} \left[ \Gamma(\delta + 1; \beta\theta) - \Gamma(\delta + 1; \beta VaR_p(X)) \right]$$

Αν  $VaR_p(X) > \theta$  τότε από Πρόταση 2.6 θα είναι:

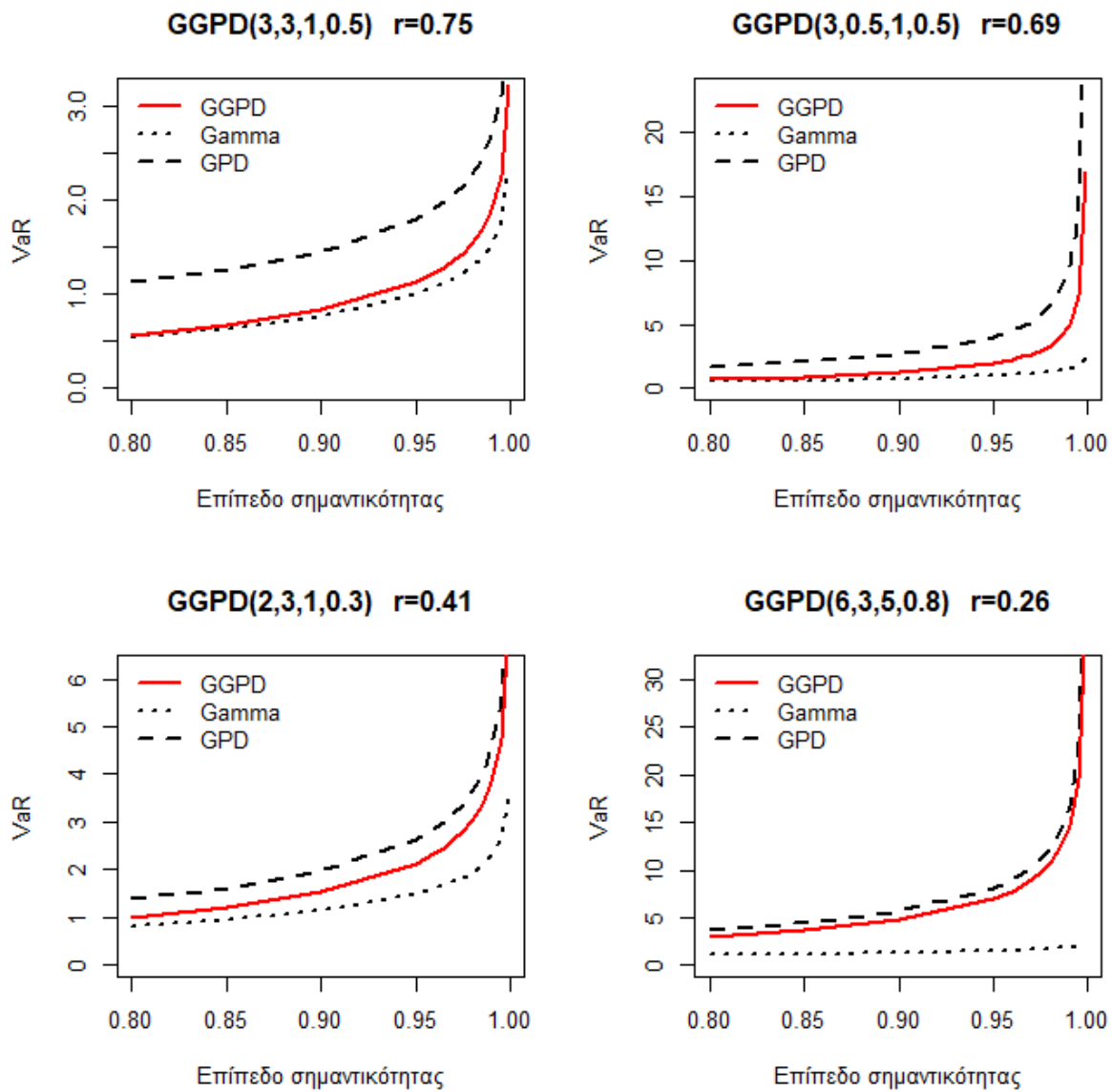
$$TVaR_p(X) = \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Για το παραπάνω ολοκλήρωμα, όπως και στην περίπτωση του  $A_1$  έχουμε ήδη εργασθεί στην απόδειξη της πρότασης 3.4, θα είναι δηλαδή:

$$TVaR_p(X) = \frac{1-r}{1-p} \alpha(\lambda + \theta)^\alpha \left[ \frac{(VaR_p(X) + \lambda)^{1-\alpha}}{\alpha - 1} - \frac{\lambda}{\alpha} (VaR_p(X) + \lambda)^{-\alpha} \right], \quad \alpha > 1$$

### 5.3 Αριθμητική εφαρμογή για σύγκριση $VaR_p - TVaR_p$

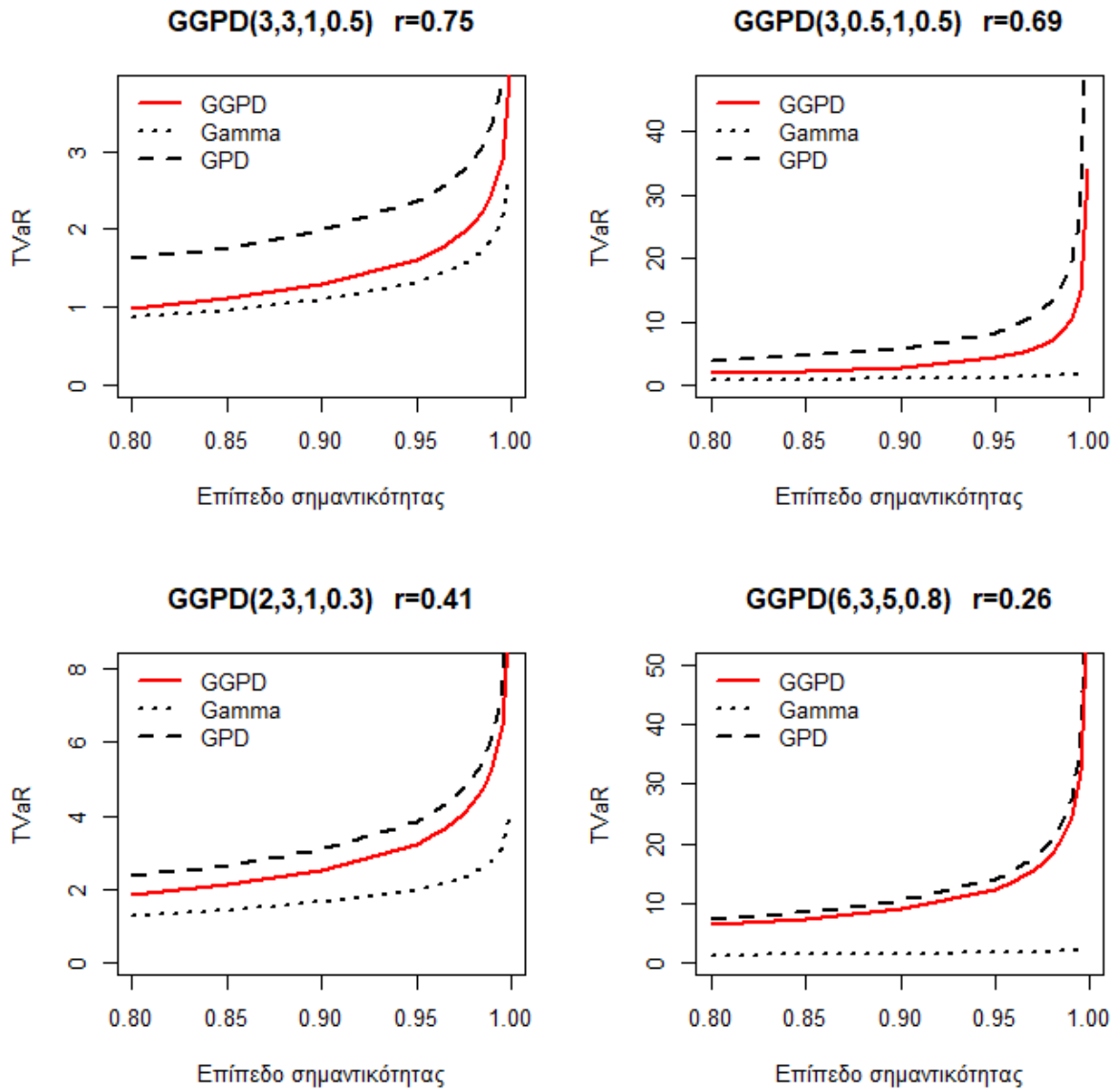
Στην παρακάτω εφαρμογή θα παρουσιάσουμε γραφικά τα αποτελέσματα των  $VaR_p - TVaR_p$  των κατανομών GGPD( $\alpha, \beta, \gamma, \delta, \theta, r$ ), Gamma( $\delta, \beta$ ), GPD( $\alpha, \theta, \gamma$ ) για τέσσερα διαφορετικά σετ παραμέτρων. Για 1<sup>ο</sup> σετ επιλέξαμε  $\beta = 3, \gamma = 3, \delta = 1, \theta = 0.5$ , για το 2<sup>ο</sup> σετ οι παράμετροι είναι  $\beta = 3, \gamma = 0.5, \delta = 1, \theta = 0.5$ , για το 3<sup>ο</sup> σετ είναι  $\beta = 2, \gamma = 3, \delta = 1, \theta = 0.3$  ενώ για το 4<sup>ο</sup> θα είναι  $\beta = 6, \gamma = 3, \delta = 5, \theta = 0.8$ . Οι παράμετροι  $\alpha, r$  προκύπτουν από τις σχέσεις 5.1 και 5.2. Τα μεγέθη  $VaR_p - TVaR_p$  μας υποδεικνύουν το πόσο βαριά είναι η ουρά της GGPD σε σχέση με τις ουρές των Gamma, GPD.



Σχήμα 5.6 Διάγραμμα σύγκρισης των  $VaR_p$  για τις GGPD, Gamma, GPD

Παρατηρούμε στο σχήμα 5.6 πως για διαφορετικά σεί παραμέτρων οι τιμές της παραμέτρου  $r$  μεταβάλλονται σημαντικά με αποτέλεσμα στο 1<sup>ο</sup> σεί ( $r = 0.75$ ) η GGPD να δίνει το περισσότερο βάρος στην Gamma συνεπώς να παρουσιάζουν παρόμοια συμπεριφορά στην ουρά. Αντίθετα όμως, στην περίπτωση του 4<sup>ου</sup> σεί ( $r = 0.26$ ) η GGPD δίνει το περισσότερο βάρος στην GPD κάτι το οποίο επιβεβαιώνουμε από τα σχήματα 5.6 και 5.7 καθώς οι δύο κατανομές εμφανίζουν παρόμοια συμπεριφορά όσον αφορά τις ακραίες τιμές τους.





**Σχήμα 5.7** Διάγραμμα σύγκρισης των  $TVaR_p$  για τις  $GGPD$ ,  $Gamma$ ,  $GPD$

#### 5.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $GGPD$ ( $\alpha, \beta, \gamma, \delta, \theta, r$ )

Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την συγκολλημένη κατανομή  $GGPD(\alpha, \beta, \gamma, \delta, \theta, r)$  με συνάρτηση πυκνότητας πιθανότητας (σ.π.π):

$$f(x; \alpha, \beta, \gamma, \delta, \theta, r) = \begin{cases} r f_{Y_1}^*(x; \beta, \delta) & , 0 < x \leq \theta \\ (1-r) f_{Y_2}^*(x; \alpha, \theta, \gamma) & , \theta < x < \infty \end{cases}$$

Όπου,

$$f_{Y_1}^*(x; \beta, \delta) = \frac{\beta^\delta}{\Gamma(\delta; \beta\theta)} x^{\delta-1} e^{-\beta x}, \quad x \leq \theta$$

και

$$f_{Y_2}^*(x; \alpha, \theta, \gamma) = \frac{\alpha(\gamma + \theta)^\alpha}{(\gamma + x)^{\alpha+1}}, \quad x > \theta$$

Χωρίς βλάβη της γενικότητας υποθέτουμε ότι  $x_1 \leq x_2 \leq \dots \leq x_n$  και ότι υπάρχει θετικός ακέραιος αριθμός  $m$  τέτοιος ώστε:

$$x_m \leq \theta \leq x_{m+1}$$

Έχει αποδειχθεί στην πρόταση 5.1. ότι:

$$GGPD(\alpha, \beta, \gamma, \delta, \theta, r) \equiv GGPD(\beta, \gamma, \delta, \theta)$$

Συνεπώς, η συνάρτηση πιθανοφάνειας θα εξαρτάται μόνο από τις παραμέτρους  $\beta, \gamma, \delta, \theta$  και θα είναι η εξής:

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \beta, \gamma, \delta, \theta) &= \prod_{i=1}^n f(x_i; \beta, \gamma, \delta, \theta) = \prod_{i=1}^m r f_{Y_1}^*(x_i; \beta, \delta) \prod_{i=m+1}^n (1-r) f_{Y_2}^*(x_i; \alpha, \theta, \gamma) \\ &= r^m (1-r)^{n-m} \prod_{i=1}^m f_{Y_1}^*(x_i; \beta, \delta) \prod_{i=m+1}^n f_{Y_2}^*(x_i; \alpha, \theta, \gamma) \end{aligned}$$

Λογαριθμίζοντας την συνάρτηση πιθανοφάνειας έχουμε τα παρακάτω:

$$\log(L(\beta, \gamma, \delta, \theta)) = l(\beta, \gamma, \delta, \theta)$$

$$= m \ln(r) + (n-m) \ln(1-r) + \sum_{i=1}^m \ln f_{Y_1}^*(x_i, \beta, \delta) + \sum_{i=m+1}^n \ln f_{Y_2}^*(x_i, \alpha, \theta, \gamma)$$

Επομένως θα είναι:

$$\begin{aligned} l(\beta, \gamma, \delta, \theta) &= m \ln(r) + (n-m) \ln(1-r) + m \delta \ln(\beta) + (\delta-1) \sum_{i=1}^m \ln(x_i) - \beta \sum_{i=1}^m x_i \\ &\quad - m \ln[\Gamma(\delta; \beta \theta)] + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) + (n-m) \ln[a(\gamma + \theta)^\alpha] \\ &\quad - (\alpha+1) \sum_{i=m+1}^n \ln(\gamma + x_i) + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) \end{aligned} \quad (5.4)$$

Για τις παραμέτρους  $\alpha, r$  ισχύει ότι υπολογίζονται συναρτήσει των παραμέτρων  $\beta, \gamma, \delta, \theta$  από τις σχέσεις 5.1 και 5.2. Συνεπώς είναι:

$$a = a(\beta, \gamma, \delta, \theta)$$

και

$$r = r(\beta, \gamma, \delta, \theta)$$

Η συνάρτηση της σχέσης 5.4 πρέπει να μεγιστοποιηθεί ως προς τις παραμέτρους  $\beta, \gamma, \delta, \theta$  δηλαδή θα πρέπει να λυθεί το παρακάτω σύστημα εξισώσεων:

$$\begin{cases} \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \beta} = 0 \\ \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \gamma} = 0 \\ \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \delta} = 0 \\ \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \theta} = 0 \end{cases}$$

Για την 1<sup>η</sup> εξίσωση έχουμε:

$$\begin{aligned} \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \beta} &= \frac{m}{r} \frac{\partial r}{\partial \beta} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \beta} + \frac{m\delta}{\beta} - \sum_{i=1}^m x_i - m \frac{\partial \ln[\Gamma(\delta; \beta\theta)]}{\partial \beta} \\ &+ \frac{\partial \alpha}{\partial \beta} (n-m) \left[ \frac{1}{a} + \ln(\gamma + \theta) \right] - \frac{\partial \alpha}{\partial \beta} \sum_{i=m+1}^n \ln(\gamma + x_i) = 0 \end{aligned}$$

Για την 2<sup>η</sup> εξίσωση θα είναι:

$$\begin{aligned} \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \gamma} &= \frac{m}{r} \frac{\partial r}{\partial \gamma} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \gamma} - m \frac{\partial \ln[\Gamma(\delta; \beta\theta)]}{\partial \gamma} + \frac{\partial \alpha}{\partial \gamma} (n-m) \left[ \frac{1}{a} + \ln(\gamma + \theta) \right] \\ &+ \frac{\alpha(n-m)}{\gamma + \theta} - \frac{\partial \alpha}{\partial \gamma} \sum_{i=m+1}^n \ln(\gamma + x_i) - (a+1) \sum_{i=m+1}^n \frac{1}{\gamma + x_i} = 0 \end{aligned}$$

Έπειτα για την 3<sup>η</sup> εξίσωση θα έχουμε:

$$\begin{aligned} \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \delta} &= \frac{m}{r} \frac{\partial r}{\partial \delta} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \delta} + m \ln(\beta) + \sum_{i=1}^m \ln(x_i) - m \frac{\partial \ln[\Gamma(\delta; \beta\theta)]}{\partial \delta} \\ &+ \frac{\partial \alpha}{\partial \delta} (n-m) \left[ \frac{1}{a} + \ln(\gamma + \theta) \right] - \frac{\partial \alpha}{\partial \delta} \sum_{i=m+1}^n \ln(\gamma + x_i) = 0 \end{aligned}$$

Τέλος για την 4<sup>η</sup> εξίσωση προκύπτει ότι:

$$\begin{aligned} \frac{\partial l(\beta, \gamma, \delta, \theta)}{\partial \theta} &= \frac{m}{r} \frac{\partial r}{\partial \theta} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \theta} - m \frac{\partial \ln[\Gamma(\delta; \beta\theta)]}{\partial \theta} + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) \\ &+ \frac{\partial \alpha}{\partial \theta} (n-m) \left[ \frac{1}{a} + \ln(\gamma + \theta) \right] + \frac{\alpha(n-m)}{\gamma + \theta} - \frac{\partial \alpha}{\partial \theta} \sum_{i=m+1}^n \ln(\gamma + x_i) \\ &+ \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) = 0 \end{aligned}$$

Η τιμή του ακεραίου αριθμού  $m$  είναι άγνωστη και επηρεάζει την εξίσωση της σχέσης 5.4. Στην έρευνα των Cooray – Ananda (2005) και Teodorescu – Vernic (2013) έχει προταθεί ένας

αλγόριθμος επίλυσης του παραπάνω συστήματος, ο οποίος αναφέρεται και στην ενότητα 3.4. Μέσω του προηγούμενου συστήματος εξισώσεων είναι αδύνατο να καταλήξουμε σε κλειστούς τύπους για τις ε.μ.π συνεπώς η εκτίμηση αυτών θα γίνεται με την βοήθεια αριθμητικών βελτιστοποιήσεων που παρέχει το στατιστικό πακέτο R. Η συνάρτηση “fitdistr” χρησιμοποιείται για την εύρεση των ε.μ.π. Στο κεφάλαιο 7 θα πραγματοποιήσουμε μελέτη προσομοίωσης με σκοπό να ελεγχθεί η απόδοση των ε.μ.π που προκύπτουν από τους υπολογισμούς της συνάρτησης “fitdistr”.

## ΚΕΦΑΛΑΙΟ 6: Η ΣΥΓΚΟΛΛΗΜΕΝΗ ΚΑΤΑΝΟΜΗ WEIBULL – GPD

$(\alpha, \vartheta, \lambda, \tau, \varphi, r)$

### 6.1 Εισαγωγή

Στην έρευνα των Scollnik – Sun (2012) αναπτύσσεται μία νέα συγκολλημένη κατανομή η οποία έχει ως κύριες συνιστώσες την κατανομή Weibull και την κατανομή GPD . Αν υποθέσουμε ότι η συμπεριφορά των δεδομένων αλλάζει εκατέρωθεν ενός σημείου  $\theta$  τότε η συγκολλημένη κατανομή Weibull – GPD (WGPD) εκφράζεται ως διακριτή μίξη της περικομμένης (αριστερά) Weibull και της περικομμένης (δεξιά) GPD στο σημείο  $\theta$ .

Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π) της Weibull( $\tau, \varphi$ ) δίνεται από τον τύπο:

$$f(x) = \left(\frac{\tau}{x}\right) \left(\frac{x}{\varphi}\right)^{\tau} e^{-\left(\frac{x}{\varphi}\right)^{\tau}}, \quad x > 0, \varphi > 0, \tau > 0$$

ενώ η σ.π.π της κατανομής GPD( $\alpha, \beta, \theta$ ) εκφράζεται από τον κάτωθι τύπο:

$$g(x) = \frac{\alpha(\alpha\beta)^{\alpha}}{(\alpha\beta - \theta + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \beta > 0$$

ο οποίος έπειτα από την αναπαράμετρηση  $\lambda = \alpha\beta - \theta$  γράφεται ως εξής:

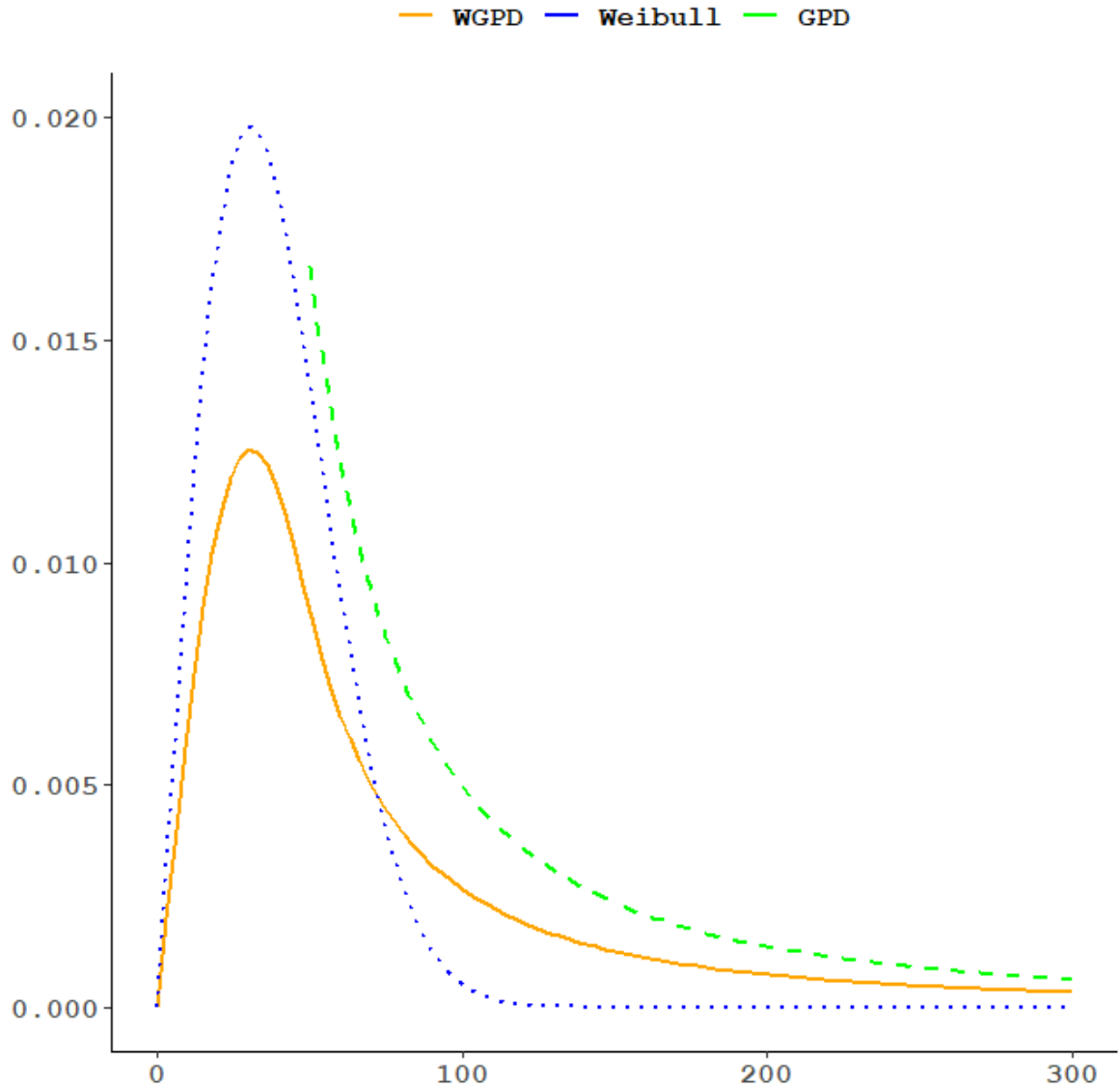
$$g(x) = \frac{\alpha(\lambda + \theta)^{\alpha}}{(\lambda + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \lambda > -\theta$$

Η σ.π.π της κατανομής Weibull( $\tau, \varphi$ ) περικομμένης στο σημείο  $\theta$  δίνεται από τον παρακάτω τύπο:

$$f_{Y_1}^*(x) = \frac{\left(\frac{\tau}{x}\right) \left(\frac{x}{\varphi}\right)^{\tau} e^{-\left(\frac{x}{\varphi}\right)^{\tau}}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^{\tau}}}, \quad x \leq \theta, \varphi > 0, \tau > 0$$

ενώ για την σ.π.π της περικομμένης κατανομής GPD( $\alpha, \theta, \lambda$ ) στο σημείο  $\theta$  θα ισχύει ότι:

$$f_{Y_2}^*(x) = \frac{\alpha(\lambda + \theta)^{\alpha}}{(\lambda + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \lambda > -\theta$$



Σχήμα 6.1 Ενδεικτική γραφική παράσταση των πυκνοτήτων των  $WGPD$ ,  $Weibull$ ,  $GPD$

## 6.2 Χαρακτηριστικά μεγέθη της κατανομής $WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$

**Ορισμός.** Αν υποθέσουμε τ.μ  $X \sim WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$  τότε η συνάρτηση πυκνότητας πιθανότητας (σ.π.π) θα δίνεται από τον παρακάτω τύπο:

$$f(x; \alpha, \theta, \lambda, \tau, \varphi, r) = \begin{cases} r \frac{\left(\frac{\tau}{x}\right) \left(\frac{x}{\varphi}\right)^\tau e^{-\left(\frac{x}{\varphi}\right)^\tau}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}}, & 0 < x \leq \theta \\ (1 - r) \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, & \theta \leq x < \infty \end{cases}$$

όπου  $r$  είναι το βάρος μίξης,  $0 \leq r \leq 1, a > 0, \theta > 0, \tau > 0, \varphi > 0, \lambda > -\theta$  οι άγνωστες παράμετροι της κατανομής.

Παρατηρούμε στο σχήμα 6.1 ότι η κατανομή WGPD έχει ουρά που συγκλίνει πιο γρήγορα στο μηδέν απ' ότι η ουρά της GPD ενώ βλέπουμε πως έχει πιο βαριά ουρά από την κατανομή Weibull.

Οι Scollnik – Sun υπέθεσαν την συνέχεια και την διαφορισιμότητα της συνάρτησης πυκνότητας της WGPD στο κατώφλι  $\theta$ . Αυτές οι υποθέσεις οδήγησαν σε ένα πιο ομαλό μοντέλο ενώ ταυτόχρονα δημιουργήθηκαν συσχετίσεις μεταξύ των παραμέτρων κάτι το οποίο ελάττωσε τον αριθμό τους από έξι σε τέσσερις.

**Πρόταση 6.1.** Αν η  $f(x; \alpha, \theta, \lambda, \tau, \varphi, r)$  είναι συνεχής και διαφορίσιμη στο σημείο  $\theta$  τότε για τις παραμέτρους  $\varphi, r$  θα ισχύουν οι παρακάτω σχέσεις:

$$\varphi = \theta \left[ \frac{\alpha\theta - \lambda}{(\lambda + \theta)\tau} + 1 \right]^{-\frac{1}{\tau}} \quad (6.1)$$

και

$$r = \frac{\frac{\alpha}{\tau}}{\frac{\lambda + \theta}{\theta} \frac{\left(\frac{\theta}{\varphi}\right)^\tau}{e^{\left(\frac{\theta}{\varphi}\right)^\tau} - 1} + \frac{\alpha}{\tau}} \quad (6.2)$$

**Απόδειξη.** Πράγματι, από την υπόθεση συνέχειας στο σημείο  $\theta$

$$\lim_{x \rightarrow \theta^-} f(x) = \lim_{x \rightarrow \theta^+} f(x)$$

Καταλήγουμε στην παρακάτω σχέση:

$$\frac{r \left(\frac{\tau}{\theta}\right) \left(\frac{\theta}{\varphi}\right)^\tau e^{-\left(\frac{\theta}{\varphi}\right)^\tau}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} = (1 - r) \frac{a(\lambda + \theta)^\alpha}{(\lambda + \theta)^{a+1}}$$

Αν λύσουμε την παραπάνω εξίσωση ως προς την παράμετρο  $r$  τότε προκύπτει ότι:

$$r = \frac{\varphi^\tau \alpha \theta \left( e^{\left(\frac{\theta}{\varphi}\right)^\tau} - 1 \right)}{\varphi^\tau \alpha \theta \left( e^{\left(\frac{\theta}{\varphi}\right)^\tau} - 1 \right) + \tau \theta^\tau (\lambda + \theta)}$$

Αν διαιρέσουμε αριθμητή και παρονομαστή με την ποσότητα

$$\tau \varphi^\tau \left[ e^{\left(\frac{\theta}{\varphi}\right)^\tau} - 1 \right]$$

Τότε προκύπτει η κάτωθι σχέση για την παράμετρο  $r$ :

$$r = \frac{\frac{\alpha}{\tau}}{\frac{\lambda + \theta}{\theta} \frac{\left(\frac{\theta}{\varphi}\right)^\tau}{e^{\left(\frac{\theta}{\varphi}\right)^\tau} - 1} + \frac{\alpha}{\tau}}$$

Έπειτα από την υπόθεση διαφορισιμότητας της σ.π.π στο σημείο  $\theta$

$$\lim_{x \rightarrow \theta^-} f'(x) = \lim_{x \rightarrow \theta^+} f'(x)$$

Προκύπτει η ακόλουθη σχέση:

$$\frac{r \tau \left(\frac{\theta}{\varphi}\right)^\tau e^{-\left(\frac{\theta}{\varphi}\right)^\tau} \left[ \tau \left(\frac{\theta}{\varphi}\right)^\tau - \tau + 1 \right]}{\theta^2 \left[ 1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau} \right]} = \frac{\alpha(\alpha + 1)(1 - r)}{(\lambda + \theta)^2}$$

Η παραπάνω εξίσωση αντικαθιστώντας την παράμετρο  $r$  γίνεται:

$$(\lambda + \theta) \left[ \tau \left(\frac{\theta}{\varphi}\right)^\tau - \tau + 1 \right] = (\alpha + 1)\theta$$

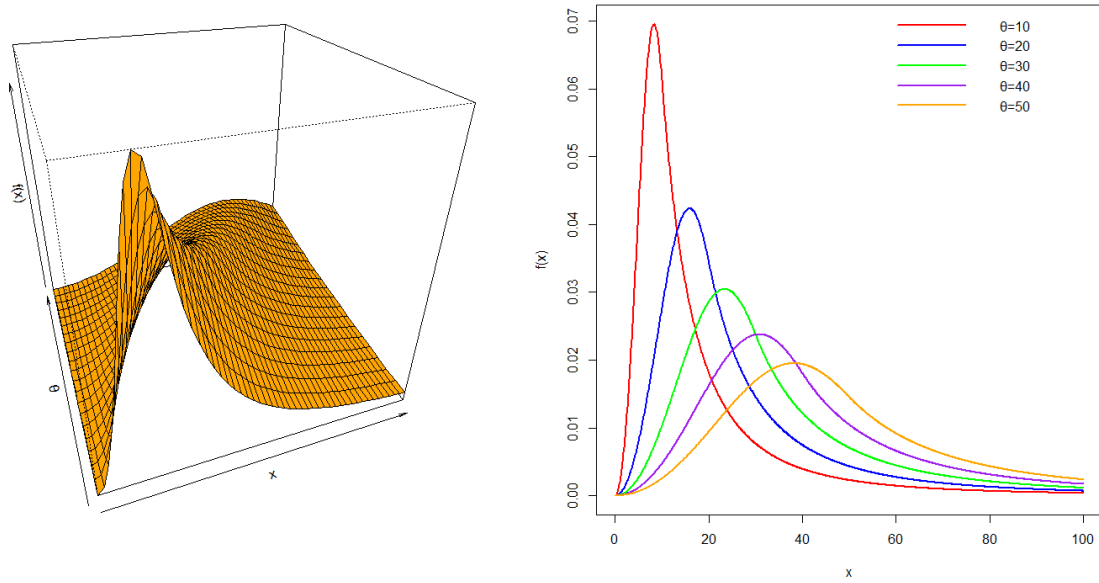
Λύνουμε ως προς  $\left(\frac{\theta}{\varphi}\right)^\tau$  και έχουμε:

$$\left(\frac{\theta}{\varphi}\right)^\tau = \frac{(\alpha + 1)\theta + (\tau - 1)(\lambda + \theta)}{\tau(\lambda + \theta)}$$

Άρα τελικά θα είναι:

$$\varphi = \theta \left[ \frac{\alpha\theta - \lambda}{(\lambda + \theta)\tau} + 1 \right]^{-\frac{1}{\tau}}$$



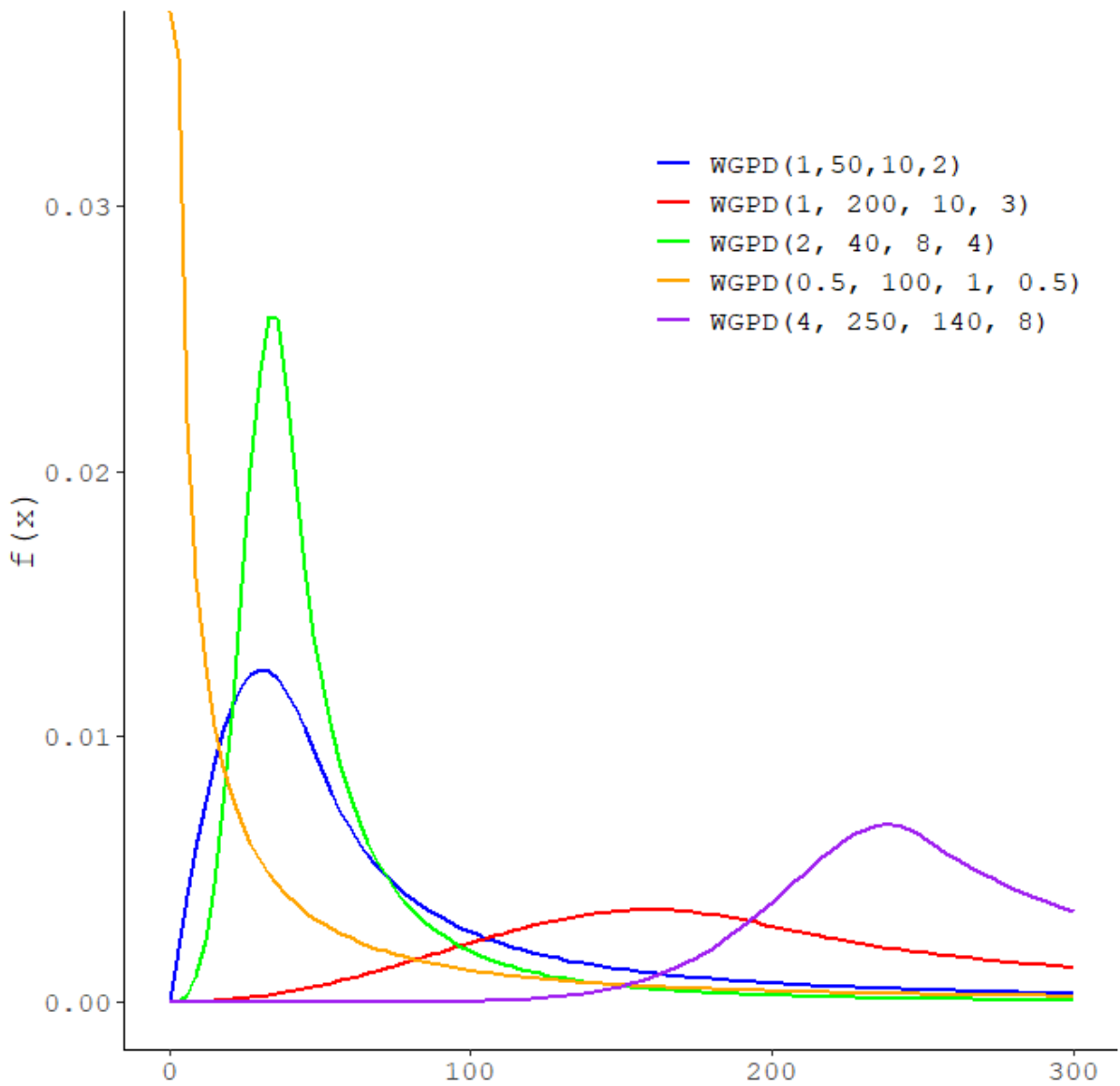


**Σχήμα 6.2** 3-Δ και 2-Δ απεικόνιση της  $WGP(2, \theta, 10, 3)$  για επιλεγμένες τιμές της παραμέτρου  $\theta$ .

Επομένως για προσδιορισθεί πλήρως η κατανομή  $WGP$  αρκεί να γνωρίζουμε τις τέσσερις από τις έξι παραμέτρους, έστω τις  $\alpha, \theta, \lambda, \tau$  καθώς οι παράμετροι  $\varphi, r$  γράφονται ως συναρτήσεις αυτών.

Δηλαδή ισχύει ότι:

$$WGP(\alpha, \theta, \lambda, \tau, \varphi, r) \equiv WGP(\alpha, \theta, \lambda, \tau)$$



**Σχήμα 6.3** Συναρτήσεις πυκνότητας της  $WGPD(\alpha, \theta, \lambda, \tau)$

Από την Πρόταση 2.2 μπορούμε να εξάγουμε τον ακόλουθο τύπο για την συνάρτηση κατανομής της  $WGPD(\alpha, \theta, \lambda, \tau)$

$$F(x; \alpha, \theta, \lambda, \tau) = \begin{cases} r \frac{1 - e^{-\left(\frac{x}{\varphi}\right)^\tau}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} & , 0 < x \leq \theta \\ r + (1 - r) \left[ 1 - \left(\frac{\gamma + \theta}{\gamma + x}\right)^\alpha \right] & , x > \theta \end{cases}$$

Συνεπώς εύκολα μπορούμε να καταλήξουμε στον παρακάτω τύπο ο οποίος περιγράφει την συνάρτηση δεξιάς ουράς

$$\bar{F}(x; \alpha, \theta, \lambda, \tau) = \begin{cases} 1 - r \frac{1 - e^{-\left(\frac{x}{\varphi}\right)^\tau}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} & , 0 < x \leq \theta \\ (1 - r) \left(\frac{\gamma + \theta}{\gamma + x}\right)^\alpha & , \theta < x < \infty \end{cases}$$

Στην βιβλιογραφία το παραπάνω μοντέλο αναφέρεται ως "Third Weibull – Pareto". Στις μελέτες των Ciumara (2006) και Cooray (2009) κατασκευάζεται το "First Weibull - Pareto" με συνιστώσες την Weibull και Pareto χρησιμοποιώντας τον αρχικό ορισμό των συγκολλημένων κατανομών. Το "First Weibull – Pareto" παρουσίαζε το ίδιο πρόβλημα με αυτό της LP, των Cooray – Ananda, δηλαδή οδηγούσε στην διακριτή μίξη των περικομμένων κατανομών εφαρμόζοντας εκ των προτέρων γνωστά βάρη μίξης. Ως "Second Weibull – Pareto" αναφέρεται το μοντέλο που για δεξί άκρο έχει την κατανομή Pareto αντί της GPD (συμβολ. με WP) και βασίζεται στον γενικότερο ορισμό που δόθηκε αργότερα από τον Scollnik. Στην έρευνα των Scollnik – Sun (2012) πραγματοποιείται η σύγκριση μεταξύ της WGPD και των LP, LGPD, WP. Το σεντ που χρησιμοποιείται για την προσαρμογή των κατανομών είναι το "Danish fire insurance". Η κατανομή WGPD παρουσιάζει καλύτερα αποτελέσματα σε σχέση με όλες τις προηγούμενες κατανομές αν συγκριθεί ως προς τις τιμές των LL, AIC, BIC. Συνεπώς φαίνεται πως η χρήση της κατανομής GPD ως δεξί άκρο των συγκολλημένων κατανομών μπορεί να περιγράψει καλύτερη την συμπεριφορά των δεδομένων στην ουρά (ακραίων τιμών) σε σχέση με την απλή μορφή της Pareto. Επίσης, η Weibull υπερτερεί της Lognormal όσον αφορά τις ζημιές που παρουσιάζουν μικρότερη σφοδρότητα αλλά μεγαλύτερη συχνότητα.

**Πρόταση 6.2.** Οι ροπές  $k$  – τάξης της κατανομής WGPD( $\alpha, \theta, \lambda, \tau$ ) θα υπολογίζονται μέσω του παρακάτω τύπου:

$$E(X^k) = r \frac{\varphi^k}{1 - e^{-(\theta/\varphi)^\tau}} \Gamma\left(\frac{k}{\tau} + 1; (\theta/\varphi)^\tau\right) + (1 - r) \alpha \sum_{m=0}^k \binom{k}{m} \frac{(-\gamma)^{k-m} (\gamma + \theta)^m}{\alpha - m}, \quad \alpha > k$$

**Απόδειξη.** Πράγματι, αν ισχύει ότι η τ.μ  $X \sim WGPD(\alpha, \theta, \lambda, \tau)$  τότε είναι η διακριτή μίξη των τ.μ  $Y_1^*, Y_2^*$  (με βάρη  $r, 1 - r$  αντίστοιχα) για τις οποίες ισχύει ότι

$$Y_1^* = X_1 | X_1 < \theta$$

και

$$Y_2^* = X_2 | X_2 > \theta$$

όπου  $X_1 \sim Weibull(\tau, \varphi)$  και  $X_2 \sim GPD(\alpha, \theta, \lambda)$ . Τότε από την Πρόταση 2.4 θα ισχύει ότι:

$$E(X^k) = rE(Y_1^{*k}) + (1-r)E(Y_2^{*k})$$

Συνεπώς, θα υπολογίσουμε αρχικά την ποσότητα  $E(Y_1^{*k})$ :

$$E(Y_1^{*k}) = \int_0^{\infty} x^k f_{X_1|X_1 < \theta}(x) dx = \frac{\int_0^{\theta} x^k f_{X_1}(x) dx}{P(X_1 < \theta)} = \frac{1}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} \int_0^{\theta} \frac{\tau}{\varphi^\tau} x^{k+\tau-1} e^{-\left(\frac{x}{\varphi}\right)^\tau} dx$$

Για τον υπολογισμό του ολοκληρώματος θα χρησιμοποιήσουμε την μη – πλήρη συνάρτηση Γάμμα:

$$\Gamma(\alpha; \beta) = \int_0^{\beta} x^{\alpha-1} e^{-x} dx$$

Αν κάνουμε αλλαγή μεταβλητής στο ολοκλήρωμα ως εξής:

$$y = \left(\frac{x}{\varphi}\right)^\tau$$

$$dx = \frac{\varphi}{\tau} y^{\frac{1}{\tau}-1} dy$$

Τότε θα έχουμε:

$$E(Y_1^{*k}) = \frac{\varphi^k}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} \int_0^{\left(\frac{\theta}{\varphi}\right)^\tau} y^{\frac{k}{\tau}} e^{-y} dy$$

Από τον τύπο της μη – πλήρους συνάρτησης Γάμμα προκύπτει ότι:

$$E(Y_1^{*k}) = \frac{\varphi^k}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} \Gamma\left(\frac{k}{\tau} + 1; (\theta/\varphi)^\tau\right)$$

Τα βήματα για τον υπολογισμό της μέσης τιμής  $E(Y_2^{*k})$  έχουν ήδη παρουσιασθεί στην απόδειξη της πρότασης 4.2. Δηλαδή ισχύει ότι:

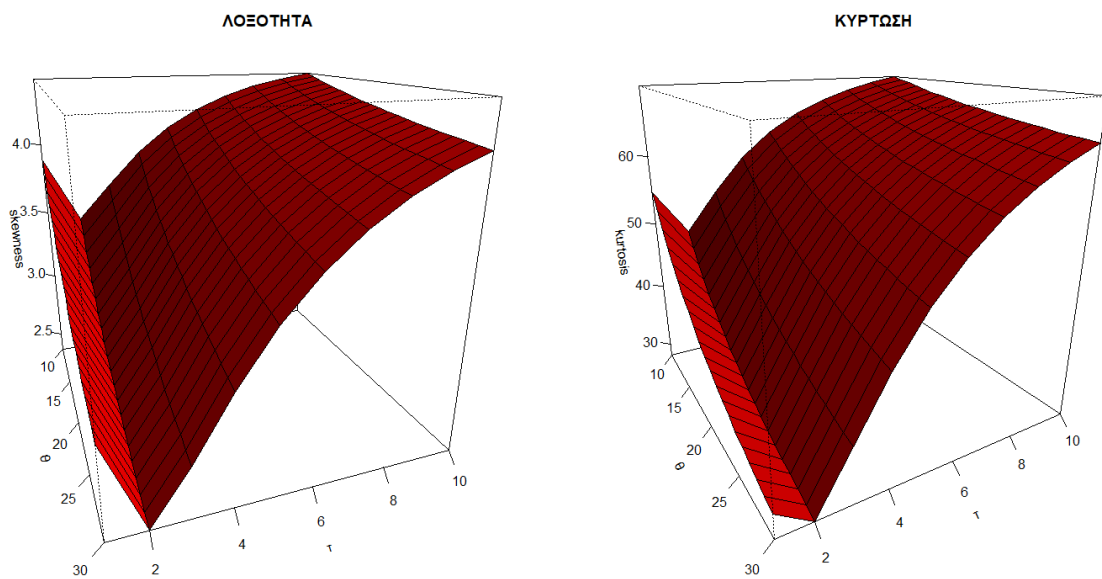
$$E(Y_2^{*k}) = \alpha \sum_{m=0}^k \binom{k}{m} \frac{(-\gamma)^{k-m} (\gamma + \theta)^m}{\alpha - m}$$

Συνεπώς, για την ροπή τάξης – κ της τ.μ X θα ισχύει:

$$E(X^k) = r \frac{\varphi^k}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} \Gamma\left(\frac{k}{\tau} + 1; (\theta/\varphi)^\tau\right) + (1-r) \alpha \sum_{m=0}^k \binom{k}{m} \frac{(-\gamma)^{k-m} (\gamma + \theta)^m}{\alpha - m}, \quad \alpha > k$$

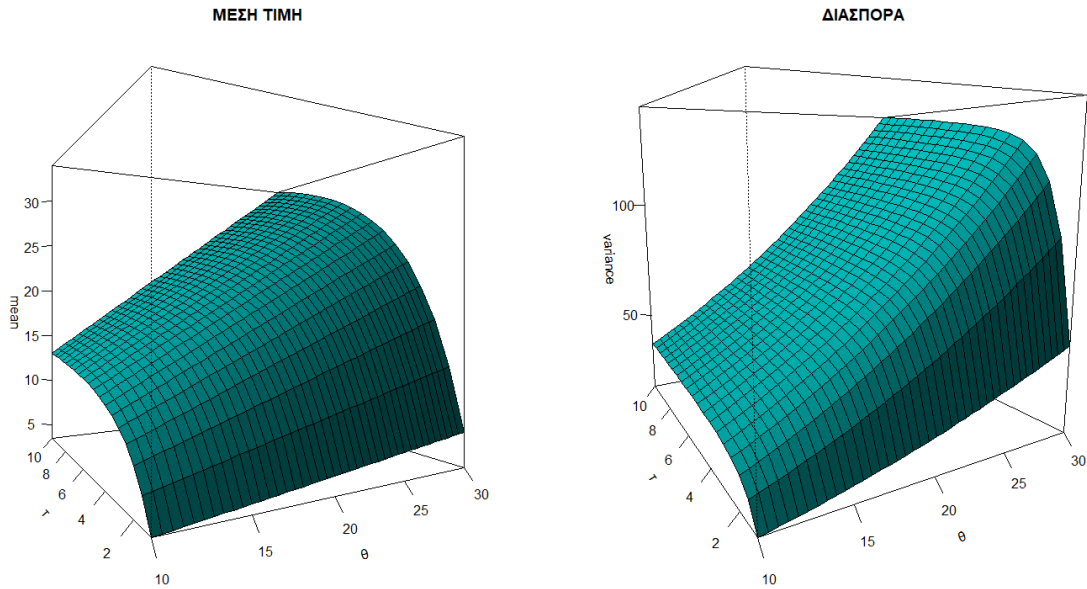
Από τον παραπάνω κλειστό τύπο για τις ροπές της κατανομής  $WGPD(\alpha, \theta, \lambda, \tau)$  μπορούμε να υπολογίσουμε την μέση τιμή, την διασπορά καθώς και τους συντελεστές κύρτωσης και λοξότητας της κατανομής μέσω των  $\kappa$ -ροπών, για  $\kappa = 1, 2, 3, 4$ .

Στο σχήμα 6.4 απεικονίζεται η λοξότητα και η κύρτωση της κατανομής  $WGPD(5, \theta, 10, \tau)$  για επιλεγμένες τιμές των παραμέτρων  $\theta, \tau$ .



**Σχήμα 6.4** Λοξότητα και Κύρτωση της κατανομής  $WGPD(5, \theta, 10, \tau)$  στις 3 διαστάσεις

Στο σχήμα 6.5 απεικονίζεται η μέση τιμή και η διασπορά της κατανομής  $WGPD(5, \theta, 10, \tau)$  για επιλεγμένες τιμές των παραμέτρων  $\theta, \tau$ .



**Σχήμα 6.5** Μέση τιμή και Διασπορά της κατανομής  $WGPD(5, \theta, 10, \tau)$  στις 3 διαστάσεις

**Πρόταση 6.3.** Το Value at Risk για την κατανομή  $WGPD(\alpha, \theta, \lambda, \tau)$  θα δίνεται μέσω του παρακάτω τύπου:

$$VaR_p(X) = \begin{cases} \varphi \left[ \ln \left( 1 - \frac{p}{r} (1 - e^{-(\theta/\varphi)^\tau}) \right)^{-1} \right]^{1/\tau} & , 0 \leq p < r \\ (\lambda + \theta) \left( \frac{1-p}{1-r} \right)^{-\frac{1}{\alpha}} - \lambda & , r \leq p < 1 \end{cases}$$

**Απόδειξη.** Πράγματι, έστω ότι:

$$F(x) = p$$

Θέλουμε να λύσουμε την παραπάνω εξίσωση ως προς  $x$  δηλαδή να έχουμε:

$$x = F^{-1}(p)$$

Οπότε θα διακρίνουμε περιπτώσεις για το  $p$ :

Αν  $0 \leq p < r$  τότε:

$$\begin{aligned} F(x) = p &\Rightarrow r \frac{1 - e^{-\left(\frac{x}{\varphi}\right)^\tau}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}} = p \Rightarrow 1 - e^{-\left(\frac{x}{\varphi}\right)^\tau} = \frac{p}{r} \left[ 1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau} \right] \\ &\Rightarrow \left(\frac{x}{\varphi}\right)^\tau = \ln \left( 1 - \frac{p}{r} \left( 1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau} \right) \right)^{-1} \\ &\Rightarrow x = \varphi \left[ \ln \left( 1 - \frac{p}{r} \left( 1 - e^{-\left(\theta/\varphi\right)^\tau} \right) \right)^{-1} \right]^{1/\tau} \end{aligned}$$

Άρα θα είναι:

$$VaR_p(X) = \varphi \left[ \ln \left( 1 - \frac{p}{r} (1 - e^{-(\theta/\varphi)^\tau}) \right)^{-1} \right]^{1/\tau}$$

Αν  $r \leq p < 1$  τότε:

$$F(x) = p \Rightarrow r + (1-r) \left[ 1 - \left( \frac{\lambda+x}{\lambda+\theta} \right)^{-\alpha} \right] = p \Rightarrow \left( \frac{\lambda+x}{\lambda+\theta} \right)^{-\alpha} = \frac{1-p}{1-r}$$

$$\Rightarrow \lambda+x = (\lambda+\theta) \left( \frac{1-p}{1-r} \right)^{-\frac{1}{\alpha}}$$

$$\Rightarrow x = (\lambda+\theta) \left( \frac{1-p}{1-r} \right)^{-\frac{1}{\alpha}} - \lambda$$

Δηλαδή έχουμε:

$$VaR_p(X) = (\lambda+\theta) \left( \frac{1-p}{1-r} \right)^{-\frac{1}{\alpha}} - \lambda$$

**Πρόταση 6.4.** Το μέτρο κινδύνου  $TVaR_p(X)$  για την κατανομή  $WGPD(\alpha, \theta, \lambda, \tau)$  δίνεται από τον παρακάτω τύπο:

Αν  $VaR_p(X) > \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \alpha(\lambda+\theta)^\alpha \frac{1-r}{1-p} \left[ \frac{(VaR_p(X) + \lambda)^{1-\alpha}}{\alpha-1} - \frac{\lambda}{\alpha} (VaR_p(X) + \lambda)^{-\alpha} \right]$$

Αν  $VaR_p(X) \leq \theta$  και  $\alpha > 1$  τότε:

$$TVaR_p(X) = \frac{r\varphi}{(1-p) \left( 1 - e^{-(\frac{\theta}{\varphi})^\tau} \right)} \left[ \Gamma \left( \frac{1}{\tau} + 1; \left( \frac{\theta}{\varphi} \right)^\tau \right) - \Gamma \left( \frac{1}{\tau} + 1; \left( \frac{VaR_p(X)}{\varphi} \right)^\tau \right) \right] \\ + \alpha(\lambda+\theta)^\alpha \frac{1-r}{1-p} \left[ \frac{(\lambda+\theta)^{1-\alpha}}{\alpha-1} - \frac{\lambda}{\alpha} (\lambda+\theta)^{-\alpha} \right]$$

**Απόδειξη.** Πράγματι, από Πρόταση 2.6 θα έχουμε:

Αν  $VaR_p(X) > \theta$  τότε:

$$TVaR_p(X) = \int_{VaR_p(X)}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Το παραπάνω ολοκλήρωμα έχει ήδη υπολογισθεί στην απόδειξη της πρότασης 4.4.

Αν  $VaR_p(X) \leq \theta$  τότε:

$$TVaR_p(X) = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx + \int_{\theta}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Για ευκολία θα θέσουμε

$$A_1 = \int_{VaR_p(X)}^{\theta} \frac{xr}{1-p} f_{Y_1}^*(x) dx$$

και

$$B_1 = \int_{\theta}^{\infty} \frac{x(1-r)}{1-p} f_{Y_2}^*(x) dx$$

Έχουμε παρουσιάσει τα βήματα για τον υπολογισμό του ολοκληρώματος  $B_1$  στην απόδειξη της πρότασης 4.4 συνεπώς μένει για την ολοκλήρωση της παρούσας απόδειξης να υπολογίσουμε το ολοκλήρωμα  $A_1$ .

Θα είναι:

$$A_1 = \frac{r}{(1-p) \left[ 1 - e^{-\left(\frac{\theta}{\varphi}\right)^{\tau}} \right]} \int_{VaR_p(X)}^{\theta} \frac{\tau}{\varphi^{\tau}} x^{\tau} e^{-\left(\frac{x}{\varphi}\right)^{\tau}} dx$$

Στο παραπάνω ολοκλήρωμα κάνουμε αλλαγή μεταβλητής ως εξής:

$$y = \left(\frac{x}{\varphi}\right)^{\tau}$$

$$dx = \frac{\varphi}{\tau} y^{\frac{1}{\tau}-1} dy$$

Άρα θα έχουμε:

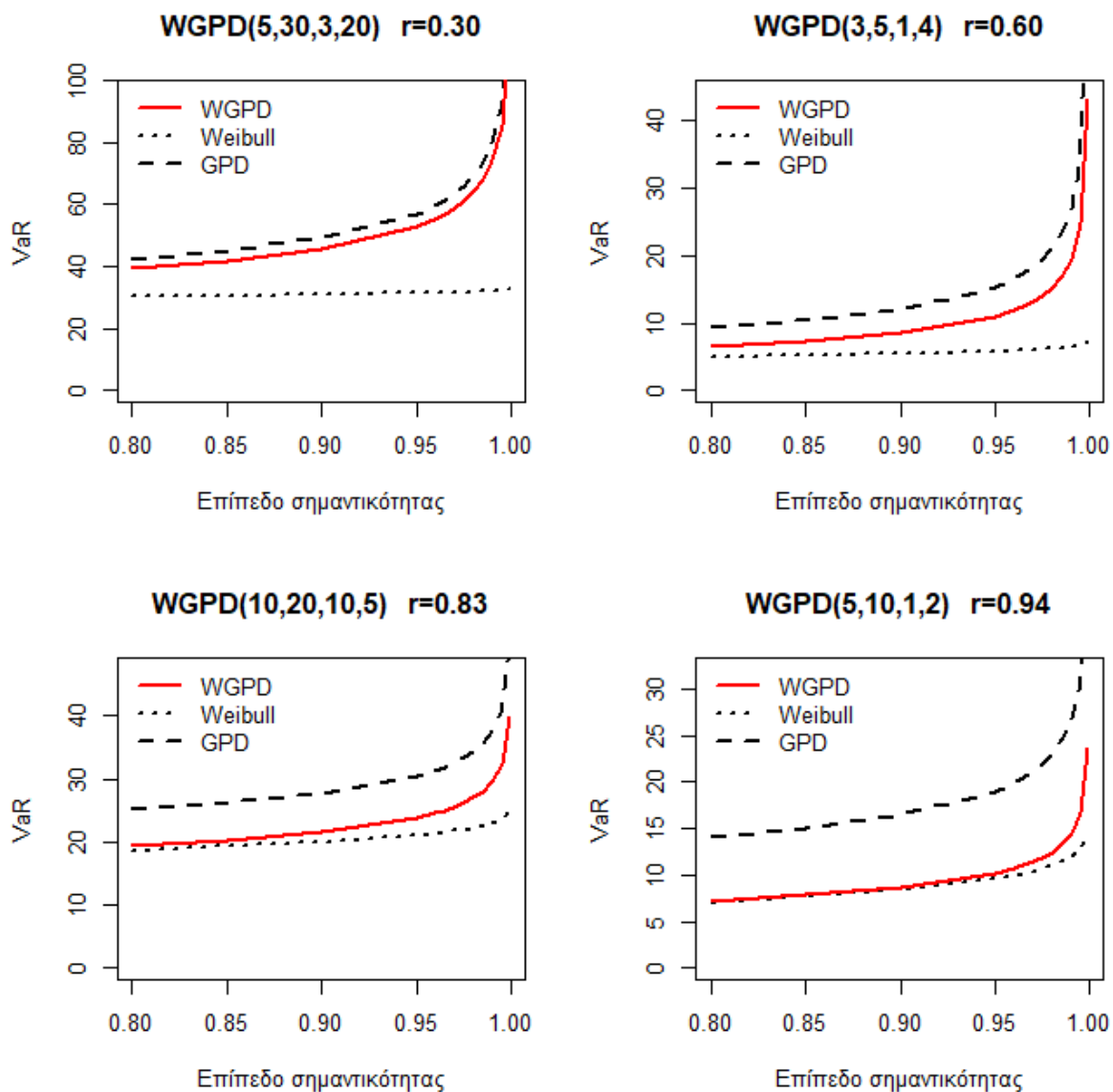
$$\begin{aligned} A_1 &= \frac{r\varphi}{(1-p) \left[ 1 - e^{-\left(\frac{\theta}{\varphi}\right)^{\tau}} \right]} \int_{\left(\frac{VaR_p(X)}{\varphi}\right)^{\tau}}^{\left(\frac{\theta}{\varphi}\right)^{\tau}} y^{\frac{1}{\tau}} e^{-y} dy = \\ &= \frac{r\varphi}{(1-p) \left[ 1 - e^{-\left(\frac{\theta}{\varphi}\right)^{\tau}} \right]} \left[ \int_0^{\left(\frac{\theta}{\varphi}\right)^{\tau}} y^{\frac{1}{\tau}} e^{-y} dy - \int_0^{\left(\frac{VaR_p(X)}{\varphi}\right)^{\tau}} y^{\frac{1}{\tau}} e^{-y} dy \right] = \end{aligned}$$



$$= \frac{r\varphi}{(1-p) \left[ 1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau} \right]} \left[ \Gamma\left(\frac{1}{\tau} + 1; \left(\frac{\theta}{\varphi}\right)^\tau\right) - \Gamma\left(\frac{1}{\tau} + 1; \left(\frac{VaR_p(X)}{\varphi}\right)^\tau\right) \right]$$

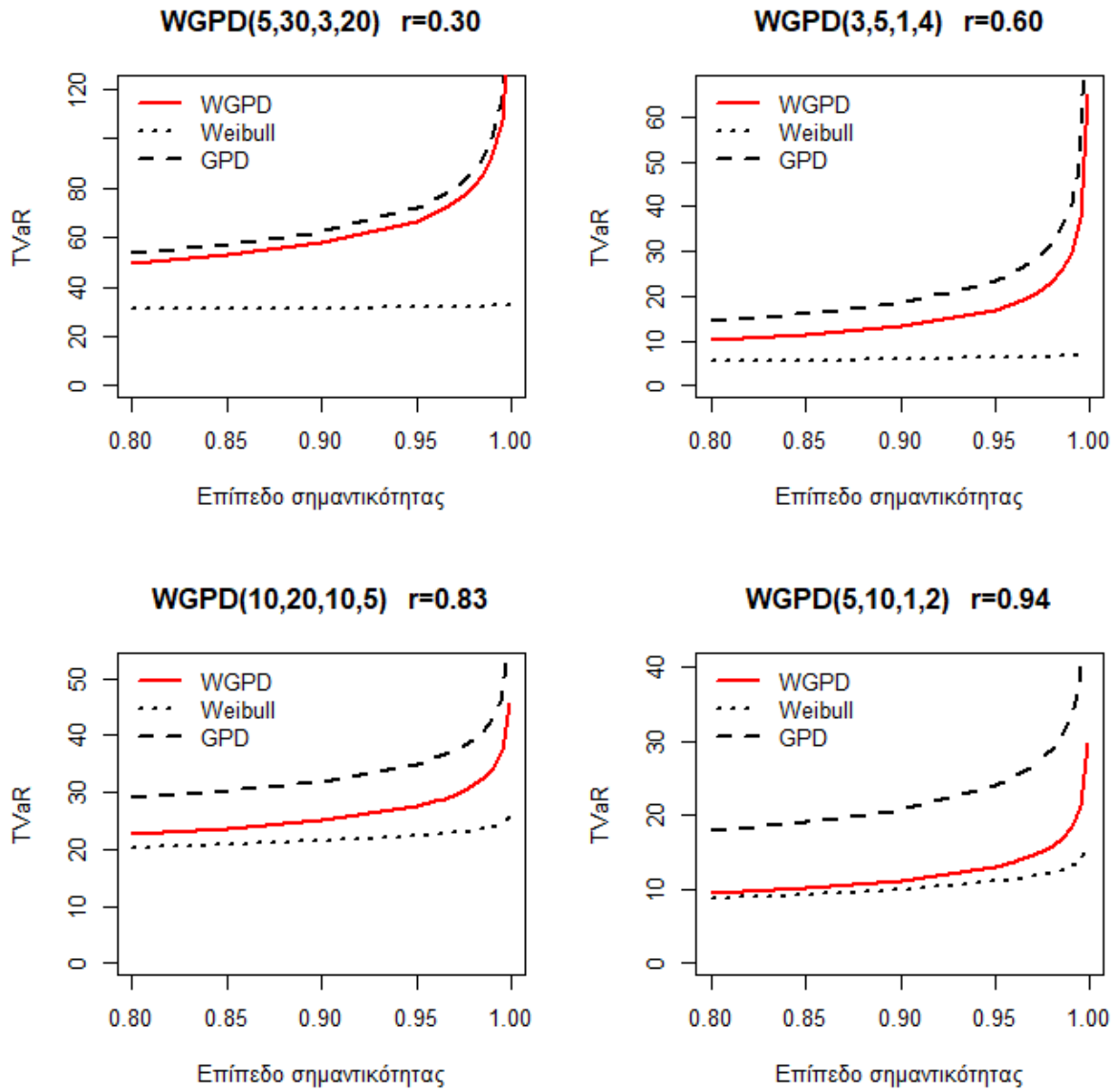
### 6.3 Αριθμητική εφαρμογή για σύγκριση $VaR_p - TVaR_p$

Στην παρακάτω εφαρμογή θα συγκρίνουμε γραφικά τα μεγέθη  $VaR_p - TVaR_p$  των κατανομών  $WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$ ,  $Weibull(\tau, \varphi)$ ,  $GPD(\alpha, \theta, \lambda)$  για τέσσερα διαφορετικά σετ παραμέτρων ώστε να παρατηρήσουμε την συμπεριφορά στις ουρές των κατανομών. Για 1<sup>ο</sup> σετ επιλέχθηκαν  $\alpha = 5, \theta = 30, \lambda = 3, \tau = 20$ , για 2<sup>ο</sup> σετ έχουμε  $\alpha = 3, \theta = 5, \lambda = 1, \tau = 4$ , για 3<sup>ο</sup> σετ είναι  $\alpha = 10, \theta = 20, \lambda = 10, \tau = 5$  ενώ για το 4<sup>ο</sup> σετ θα είναι  $\alpha = 5, \theta = 10, \lambda = 1, \tau = 2$ . Οι παράμετροι  $\varphi, r$  προκύπτουν από τις σχέσεις 6.1 και 6.2.



**Σχήμα 6.6** Διαγράμματα σύγκρισης των  $VaR_p$  για τα τέσσερα σετ παραμέτρων.

Παρατηρούμε πως η κατανομή WGPD ενσωματώνει τα χαρακτηριστικά των δύο συνιστωσών της καθώς η ουρά της είναι πιο βαριά από αυτή της Weibull αλλά πιο ελαφριά από την ουρά της GPD ενώ ταυτόχρονα ανάλογα τις τιμές των παραμέτρων  $\alpha, \theta, \lambda, \tau$  το βάρος μίξης μπορεί να μεταβληθεί σημαντικά με αποτέλεσμα η κατανομή WGPD να παρομοιάζει την συμπεριφορά της είτε με την GPD (1<sup>ο</sup> σετ :  $r = 0.30$ ) είτε με την Weibull (4<sup>ο</sup> σετ :  $r = 0.94$ ).



**Σχήμα 6.7** Διαγράμματα σύγκρισης των  $TVaR_p$  για τα τέσσερα σεί παραμέτρων.

#### 6.4 Εκτιμητές μέγιστης πιθανοφάνειας για την $WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$

Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την συγκολλημένη κατανομή  $WGPD(\alpha, \theta, \lambda, \tau, \varphi, r)$  με συνάρτηση πυκνότητας πιθανότητας (σ.π.π):

$$f(x; \alpha, \theta, \lambda, \tau, \varphi, r) = \begin{cases} r f_{Y_1}^*(x; \tau, \varphi) & , 0 < x \leq \theta \\ (1-r) f_{Y_2}^*(x; \alpha, \theta, \lambda) & , \theta < x < \infty \end{cases}$$

όπου

$$f_{Y_1}^*(x; \tau, \varphi) = \frac{\left(\frac{\tau}{x}\right) \left(\frac{x}{\varphi}\right)^\tau e^{-\left(\frac{x}{\varphi}\right)^\tau}}{1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}}, \quad x \leq \theta$$

και

$$f_{Y_2}^*(x; \alpha, \theta, \lambda) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta$$

Χωρίς βλάβη της γενικότητας υποθέτουμε ότι  $x_1 \leq x_2 \leq \dots \leq x_n$  και ότι υπάρχει θετικός ακέραιος αριθμός  $m$  τέτοιος ώστε:

$$x_m \leq \theta \leq x_{m+1}$$

Έχει αποδειχθεί στην πρόταση 6.1. ότι:

$$WGPD(\alpha, \theta, \lambda, \tau, \varphi, r) \equiv WGPD(\alpha, \theta, \lambda, \tau)$$

Συνεπώς, η συνάρτηση πιθανοφάνειας θα εξαρτάται μόνο από τις παραμέτρους  $\alpha, \theta, \lambda, \tau$  και θα είναι η εξής:

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \alpha, \theta, \lambda, \tau) &= \prod_{i=1}^n f(x_i; \alpha, \theta, \lambda, \tau) = \prod_{i=1}^m r f_{Y_1}^*(x_i; \tau, \varphi) \prod_{i=m+1}^n (1-r) f_{Y_2}^*(x_i; \alpha, \theta, \lambda) \\ &= r^m (1-r)^{n-m} \prod_{i=1}^m f_{Y_1}^*(x_i; \tau, \varphi) \prod_{i=m+1}^n f_{Y_2}^*(x_i; \alpha, \theta, \lambda) \end{aligned}$$

Λογαριθμίζοντας την συνάρτηση πιθανοφάνειας έχουμε:

$$\begin{aligned} \log(L(\alpha, \theta, \lambda, \tau)) &= l(\alpha, \theta, \lambda, \tau) \\ &= m \ln(r) + (n-m) \ln(1-r) + \sum_{i=1}^m \ln f_{Y_1}^*(x_i, \tau, \varphi) + \sum_{i=m+1}^n \ln f_{Y_2}^*(x_i, \alpha, \theta, \lambda) \end{aligned}$$

Συνεπώς θα είναι:

$$\begin{aligned} l(\alpha, \theta, \lambda, \tau) &= m \ln(r) + (n-m) \ln(1-r) + m \ln\left(\frac{\tau}{\varphi^\tau}\right) + (\tau-1) \sum_{i=1}^m \ln(x_i) - \sum_{i=1}^m \left(\frac{x_i}{\varphi}\right)^\tau \\ &\quad - m \ln\left(1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}\right) + \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) + (n-m) \ln[\alpha(\lambda + \theta)^\alpha] \\ &\quad - (\alpha+1) \sum_{i=m+1}^n \ln(\lambda + x_i) + \\ &\quad + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) \end{aligned} \tag{6.3}$$

Για τις παραμέτρους  $\varphi, r$  ισχύει ότι υπολογίζονται συναρτήσει των παραμέτρων  $\alpha, \theta, \lambda, \tau$  από τις σχέσεις 6.1 και 6.2. Συνεπώς είναι:

$$\varphi = \varphi(\alpha, \theta, \lambda, \tau)$$

και

$$r = r(\alpha, \theta, \lambda, \tau)$$

Η παραπάνω συνάρτηση επηρεάζεται από την τιμή του αγνώστου ακεραίου  $m$ . Για τον λόγο αυτόν έχει προταθεί ο αλγόριθμος της ενότητας 3.4 (βλ. Cooray – Ananda). Η παραπάνω συνάρτηση θα πρέπει να μεγιστοποιηθεί ως προς τις παραμέτρους  $\alpha, \theta, \lambda, \tau$  συνεπώς θα πρέπει να επιλυθεί το παρακάτω σύστημα εξισώσεων.

$$\begin{cases} \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \alpha} = 0 \\ \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \theta} = 0 \\ \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \lambda} = 0 \\ \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \tau} = 0 \end{cases}$$

Για την 1<sup>η</sup> εξίσωση έχουμε:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \alpha} &= \frac{m}{r} \frac{\partial r}{\partial \alpha} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \alpha} + m \frac{\partial \ln\left(\frac{\tau}{\varphi^\tau}\right)}{\partial \alpha} - \frac{\partial\left(\frac{1}{\varphi^\tau}\right)}{\partial \alpha} \sum_{i=1}^m x_i^\tau - m \frac{\partial \ln\left(1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}\right)}{\partial \alpha} \\ &+ (n-m) \left[ \frac{1}{\alpha} + \ln(\lambda + \theta) \right] - \sum_{i=m+1}^n \ln(\lambda + x_i) = 0 \end{aligned}$$

Για την 2<sup>η</sup> εξίσωση θα είναι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \theta} &= \frac{m}{r} \frac{\partial r}{\partial \theta} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \theta} + m \frac{\partial \ln\left(\frac{\tau}{\varphi^\tau}\right)}{\partial \theta} - \frac{\partial\left(\frac{1}{\varphi^\tau}\right)}{\partial \theta} \sum_{i=1}^m x_i^\tau - m \frac{\partial \ln\left(1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}\right)}{\partial \theta} \\ &+ \sum_{i=1}^m \ln(I_{\{x_i \leq \theta\}}) + \frac{\alpha(n-m)}{\lambda + \theta} + \sum_{i=m+1}^n \ln(I_{\{x_i > \theta\}}) = 0 \end{aligned}$$

Για την 3<sup>η</sup> εξίσωση προκύπτει ότι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \lambda} &= \frac{m}{r} \frac{\partial r}{\partial \lambda} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \lambda} + m \frac{\partial \ln\left(\frac{\tau}{\varphi^\tau}\right)}{\partial \lambda} - \frac{\partial\left(\frac{1}{\varphi^\tau}\right)}{\partial \lambda} \sum_{i=1}^m x_i^\tau - m \frac{\partial \ln\left(1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}\right)}{\partial \lambda} \\ &+ \frac{\alpha(n-m)}{\lambda + \theta} - (\alpha + 1) \sum_{i=m+1}^n \frac{1}{\lambda + x_i} = 0 \end{aligned}$$

Τέλος για την 4<sup>η</sup> εξίσωση θα ισχύει ότι:

$$\begin{aligned} \frac{\partial l(\alpha, \theta, \lambda, \tau)}{\partial \tau} &= \frac{m}{r} \frac{\partial r}{\partial \tau} - \frac{(n-m)}{1-r} \frac{\partial r}{\partial \tau} + m \frac{\partial \ln\left(\frac{\tau}{\varphi^\tau}\right)}{\partial \tau} + \sum_{i=1}^m \ln(x_i) - \frac{\partial\left(\frac{1}{\varphi^\tau}\right)}{\partial \tau} \sum_{i=1}^m x_i^\tau \\ &\quad - \frac{1}{\varphi^\tau} \sum_{i=1}^m \frac{\partial(x_i^\tau)}{\partial \tau} - m \frac{\partial \ln\left(1 - e^{-\left(\frac{\theta}{\varphi}\right)^\tau}\right)}{\partial \tau} = 0 \end{aligned}$$

Η εξαγωγή αναλυτικών τύπων για τον υπολογισμό των ε.μ.π είναι ανέφικτη συνεπώς θα πρέπει να καταφύγουμε σε αριθμητικές τεχνικές για την μεγιστοποίηση της συνάρτησης που δίνεται στην σχέση 6.3. Για την εκτίμηση των ε.μ.π χρησιμοποιείται το στατιστικό πακέτο R και συγκεκριμένα η συνάρτηση “fitdistr”. Στο κεφάλαιο 7 θα πραγματοποιηθεί μελέτη προσομοίωσης με σκοπό να ελεγχθεί η απόδοση των ε.μ.π ως προς την μέση μεροληψία και το μέσο τετρ. σφάλμα.

# ΚΕΦΑΛΑΙΟ 7: ΜΕΛΕΤΗ ΠΡΟΣΟΜΟΙΩΣΗΣ ΓΙΑ ΤΙΣ ΣΥΓΚΟΛΛΗΜΕΝΕΣ ΚΑΤΑΝΟΜΕΣ

## 7.1 Σχεδιασμός προσομοίωσης

Σε αυτό το κεφάλαιο θα παρουσιάσουμε δύο μεθόδους προσομοίωσης δείγματος από τις συγκολλημένες κατανομές LP, LGPD, GGPD, WGPD. Στην συνέχεια θα πραγματοποιηθεί μελέτη προσομοίωσης για κάθε κατανομή ξεχωριστά με σκοπό να ελεγχθεί η απόδοση των εκτιμητών μέγιστης πιθανοφάνειας (ε.μ.π). Όπως ήδη έχουμε αναφέρει, η εξαγωγή αναλυτικών τύπων για την εύρεση των ε.μ.π είναι αδύνατη συνεπώς είναι αναγκαία η χρήση αριθμητικών μεθόδων βελτιστοποίησης. Για την εκτίμηση των ε.μ.π θα χρησιμοποιήσουμε το στατιστικό πακέτο R και συγκεκριμένα την συνάρτηση “fitdistr” του πακέτου δεδομένων “MASS”. Η μέθοδος εκτίμησης που εισάγουμε στην συνάρτηση “fitdistr” είναι η “Nelder – Mead” η οποία προτείνεται σε περιπτώσεις όπου το πλήθος των παραμέτρων προς εκτίμηση είναι  $\geq 2$ .

Για κάθε ένα από τα δειγματικά μεγέθη  $n = 100, 400, 700, 1000, 1500$  θα ακολουθήσουμε την παρακάτω διαδικασία για την μελέτη προσομοίωσης:

Προσομοιώνουμε  $N = 1000$  δείγματα από την συγκολλημένη κατανομή που μελετάμε και έπειτα για κάθε δείγμα υπολογίζονται μέσω του πακέτου R οι εκτιμητές μέγιστης πιθανοφάνειας. Αν υποθέσουμε ότι οι παράμετροι της κατανομής είναι οι  $\delta_1, \delta_2, \dots, \delta_s, s \in \mathbb{N}$  και ότι οι εκτιμητές είναι αντίστοιχα οι  $\widehat{\delta}_1, \widehat{\delta}_2, \dots, \widehat{\delta}_s, s \in \mathbb{N}$  τότε υπολογίζουμε τα παρακάτω μεγέθη:

1. Μέση μεροληψία (Average bias) για τους εκτιμητές  $\widehat{\delta}_i, 1 \leq i \leq s$ :

$$\frac{1}{N} \sum_{k=1}^N (\widehat{\delta}_{ik} - \delta_i) \quad (7.1)$$

2. Μέση απόλυτη μεροληψία (Average absolute bias) για τους εκτιμητές  $\widehat{\delta}_i, 1 \leq i \leq s$ :

$$\frac{1}{N} \sum_{k=1}^N |\widehat{\delta}_{ik} - \delta_i| \quad (7.2)$$

3. Μέσο τετραγωνικό σφάλμα (Mean squared error - MSE για τους εκτιμητές  $\widehat{\delta}_i, 1 \leq i \leq s$ :

$$\frac{1}{N} \sum_{k=1}^N (\widehat{\delta}_{ik} - \delta_i)^2 \quad (7.3)$$

4. Τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Square root of mean squared error – RMSE) για τους εκτιμητές  $\widehat{\delta}_i$ ,  $1 \leq i \leq s$ :

$$\sqrt{\frac{1}{N} \sum_{k=1}^N (\widehat{\delta}_{ik} - \delta_i)^2} \quad (7.4)$$

Για την καλύτερη κατανόηση της διαδικασίας αναφέρουμε πως η ποσότητα  $\widehat{\delta}_{ik}$  αντιπροσωπεύει τον εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου  $\delta_i$  για το  $\kappa$ -οστό δείγμα,  $1 \leq \kappa \leq N$ .

Στην συνέχεια θα περιγράψουμε τις δύο μεθόδους με τις οποίες μπορούμε να προσομοιώνουμε δείγματα από τις συγκολλημένες κατανομές LP, LGPD, GGPD, WGPD .

### 1. Μέθοδος αντιστροφής (The inverse method)

Μια από τις βασικότερες μεθόδους στην προσομοίωση για την παραγωγή δείγματος από κάποια κατανομή  $F$  είναι η μέθοδος της αντιστροφής. Θεωρούμε πως με  $F$  συμβολίζουμε ταυτόχρονα και την συνάρτηση κατανομής. Σύμφωνα με αυτή αν επιθυμούμε να παράξουμε παρατήρηση  $x$  από την κατανομή  $F$  τότε ακολουθούμε την εξής διαδικασία:

- Παράγουμε τυχαία παρατήρηση  $u \sim U(0,1)$
- Θέτουμε  $x = F^{-1}(u)$

Η παραπάνω διαδικασία προϋποθέτει την αντιστροφή της συνάρτησης  $F$  κάτι το οποίο στις περισσότερες περιπτώσεις είναι ανέφικτο. Επομένως η μέθοδος της αντιστροφής συχνά απαιτεί την χρήση κατάλληλων αριθμητικών μεθόδων για την αντιστροφή συναρτήσεων. Στις περιπτώσεις όπου η συνάρτηση κατανομής δεν ήταν δυνατό να αντιστραφεί (π.χ κατανομή  $GGPD$ ) εφαρμόστηκε η συνάρτηση “inverse” του πακέτου δεδομένων “GoFKernel”. Για τις κατανομές που βρίσκονται εντός των πλαισίων της μελέτης μας έχουμε ήδη παρουσιάσει του αναλυτικούς τύπους για την εύρεση των ποσοστιαίων σημείων (βλ. Πρόταση 3.3, Πρόταση 4.3, Πρόταση 6.3).

Οπότε π.χ, για την κατανομή LGPD εργαζόμαστε ως εξής:

- Παράγουμε τυχαία παρατήρηση  $u \sim U(0,1)$



- Αν  $0 \leq u < r$  θέτουμε  $x = e^{\mu + \sigma \Phi^{-1}\left[\frac{u}{r} \Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)\right]}$
- Αλλιώς θέτουμε  $x = (\lambda + \theta) \left(\frac{1-u}{1-r}\right)^{-\frac{1}{a}} - \lambda$

Όμοια μπορούμε να προσομοιώσουμε δείγματα και από τις υπόλοιπες συγκολλημένες κατανομές.

## 2. Μέθοδος διακριτής μίξης (The discrete mixture method)

Έστω ότι θέλουμε να προσομοιώσουμε παρατηρήσεις από την τ.μ  $X \sim F$  με σ.π.π  $f_X(x)$ . Αν υπάρχουν τ.μ  $X_1, X_2$  με σ.π.π  $f_{X_1}(x), f_{X_2}(x)$  αντίστοιχα για τις οποίες ισχύει ότι:

$$f_X(x) = r * f_{X_1}(x) + (1 - r) * f_{X_2}(x) , \quad 0 \leq r \leq 1$$

Τότε γνωρίζουμε πως η τ.μ  $X$  είναι η διακριτή μίξη των τ.μ  $X_1, X_2$  με βάρη  $r$  και  $1 - r$  αντίστοιχα.

Σε αυτή την περίπτωση είναι εύκολο να παράγουμε παρατήρηση  $x$  από την κατανομή  $F$  χωρίς να χρειάζεται η αντιστροφή της συνάρτησης κατανομής της  $F$ . Η διαδικασία είναι η ακόλουθη:

- Παράγουμε τυχαία παρατήρηση  $u \sim U(0,1)$
- Αν  $0 < u \leq r$  τότε παράγουμε παρατήρηση  $x_1$  από την τ.μ  $X_1$  και θέτουμε  $x = x_1$ .
- Αν  $r < u \leq 1$  τότε παράγουμε παρατήρηση  $x_2$  από την τ.μ  $X_2$  και θέτουμε  $x = x_2$ .

Είναι προφανές πως στις περιπτώσεις των κατανομών που μελετάμε η μέθοδος της διακριτής μίξης είναι προτιμότερη από την μέθοδο της αντιστροφής διότι έχει αποδειχθεί στην πρόταση 2.1 πως κάθε συγκολλημένη κατανομή είναι διακριτή μίξη από περικομμένες κατανομές συνεπώς δεν είναι αναγκαία η χρήση αριθμητικών μεθόδων για την αντιστροφή συναρτήσεων (βλ. μέθοδο αντιστροφής).

Οπότε π.χ για την κατανομή LGPD εργαζόμαστε ως εξής:

- Παράγουμε τυχαία παρατήρηση  $u \sim U(0,1)$
- Αν  $0 < u \leq r$ , τότε παράγουμε τυχαία παρατήρηση  $x_1$  από την περικομμένη (αριστερά) στο σημείο  $\theta$  κατανομή Lognormal( $\mu, \sigma^2$ ) με σ.π.π:

$$f_{X_1}(x) = \frac{\frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}}{\Phi\left(\frac{\ln \theta - \mu}{\sigma}\right)}, \quad x \leq \theta, \sigma > 0, \mu \in \mathbb{R}$$

και θέτουμε  $x = x_1$

- Αν  $r < u \leq 1$ , τότε παράγουμε τυχαία παρατήρηση  $x_2$  από την περικομμένη (δεξιά) στο σημείο  $\theta$  κατανομή  $GPD(\alpha, \theta, \lambda)$  με σ.π.π:

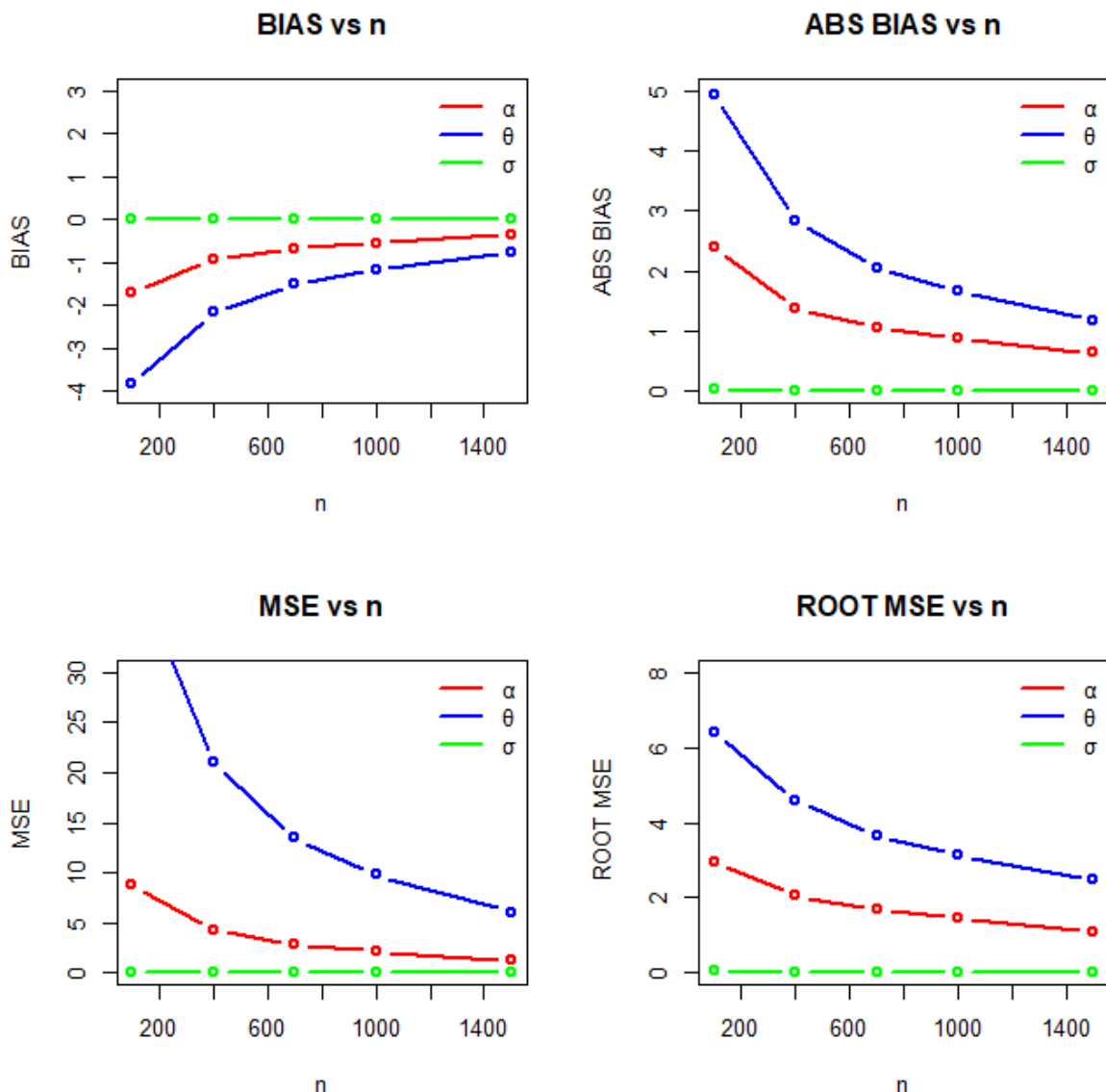
$$f_{x_2}(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta, \alpha > 0, \theta > 0, \lambda > -\theta$$

και θέτουμε  $x = x_2$

Με όμοιο τρόπο μπορούμε να παράξουμε δείγμα από τις υπόλοιπες συγκολλημένες κατανομές.

## 7.2 Μελέτη προσομοίωσης για την συγκολλημένη $LP(\alpha, \theta, \sigma)$

Για τις παραμέτρους  $\alpha = 4, \theta = 5, \sigma = 0.5$  προσομοιώσαμε δείγματα από την κατανομή LP ακολουθώντας τα βήματα της ενότητας 7.1 και χρησιμοποιώντας την μέθοδο της διακριτής μίξης. Στο σχήμα 7.1 παρουσιάζονται γραφικά τα αποτελέσματα της προσομοίωσης.



**Σχήμα 7.1** Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους  $\alpha, \theta, \sigma$ .

Τα μέτρα των σχέσεων 7.1 – 7.4 χρησιμοποιούνται με σκοπό να ελεγχθεί η ποιότητα των ε.μ.π. των παραμέτρων για διαφορετικά μεγέθη δείγματος. Συμπεραίνουμε μέσω του σχήματος 7.1 πως η απόδοση των ε.μ.π. βελτιώνεται όσο το μέγεθος του δείγματος αυξάνεται παρότι υπάρχουν ενδείξεις αρνητικής μεροληψίας για τους ε.μ.π. των  $\alpha, \theta$ . Στον πίνακα 7.1 συνοψίζονται μερικές από τις τιμές της μεροληψίας και του μέσου τετραγωνικού σφάλματος όπως αυτές προκύπτουν από την μελέτη προσομοίωσης για επιλεγμένα μεγέθη δείγματος.

<i>LP</i>	<i>ΜΕΡΟΛΗΨΙΑ</i>			<i>ΜΤΣ</i>		
	$\alpha$	$\theta$	$\sigma$	$\alpha$	$\theta$	$\sigma$
<i>n</i>						
100	- 1,704	- 3,840	0,010	8,75	41,08	0,001
700	- 0,670	- 1,500	0,001	2,83	13,44	0,0002
1500	- 0,348	- 0,748	0,0002	1,23	6,07	0,0001

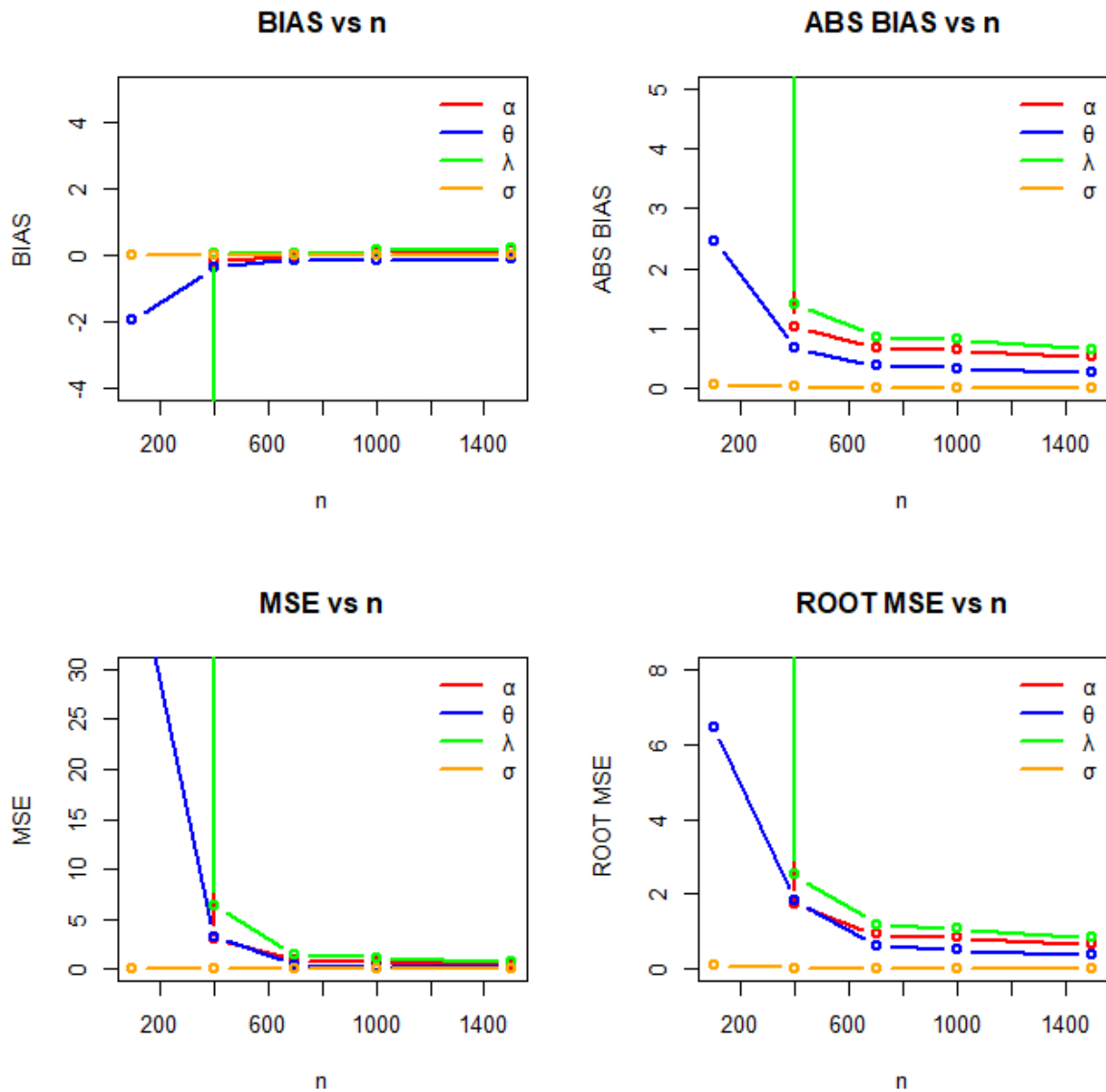
**Πίνακας 7.1** Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων  $\alpha, \theta, \sigma$  για επιλεγμένες τιμές του  $n$ .

### 7.3 Μελέτη προσομοίωσης για την συγκολλημένη $LGPD(\alpha, \theta, \lambda, \sigma)$

Για τις παραμέτρους  $\alpha = 3, \theta = 2, \lambda = 1, \sigma = 0.5$  προσομοιώσαμε δείγματα από την κατανομή  $LGPD$  ακολουθώντας τα βήματα της ενότητας 7.1 και χρησιμοποιώντας την μέθοδο της διακριτής μίξης. Στο σχήμα 7.2 παρουσιάζονται γραφικά τα αποτελέσματα της προσομοίωσης.

Παρατηρούμε στο σχήμα 7.2 πως οι ε.μ.π. των παραμέτρων  $\theta, \lambda$  παρουσιάζουν υψηλές τιμές μεροληψίας και μέσου τετραγωνικού σφάλματος για μικρά μεγέθη δείγματος ( $n < 400$ ). Ωστόσο φαίνεται πως οι ε.μ.π. βελτιώνουν την απόδοσή τους καθώς το  $n$  αυξάνεται.

Στον πίνακα 7.2 συνοψίζονται μερικές από τις τιμές της μεροληψίας και του μέσου τετραγωνικού σφάλματος όπως αυτές προκύπτουν από την μελέτη προσομοίωσης για επιλεγμένα μεγέθη δείγματος. Πράγματι, συμπεραίνουμε πως οι τιμές αυτές συγκλίνουν στο μηδέν.



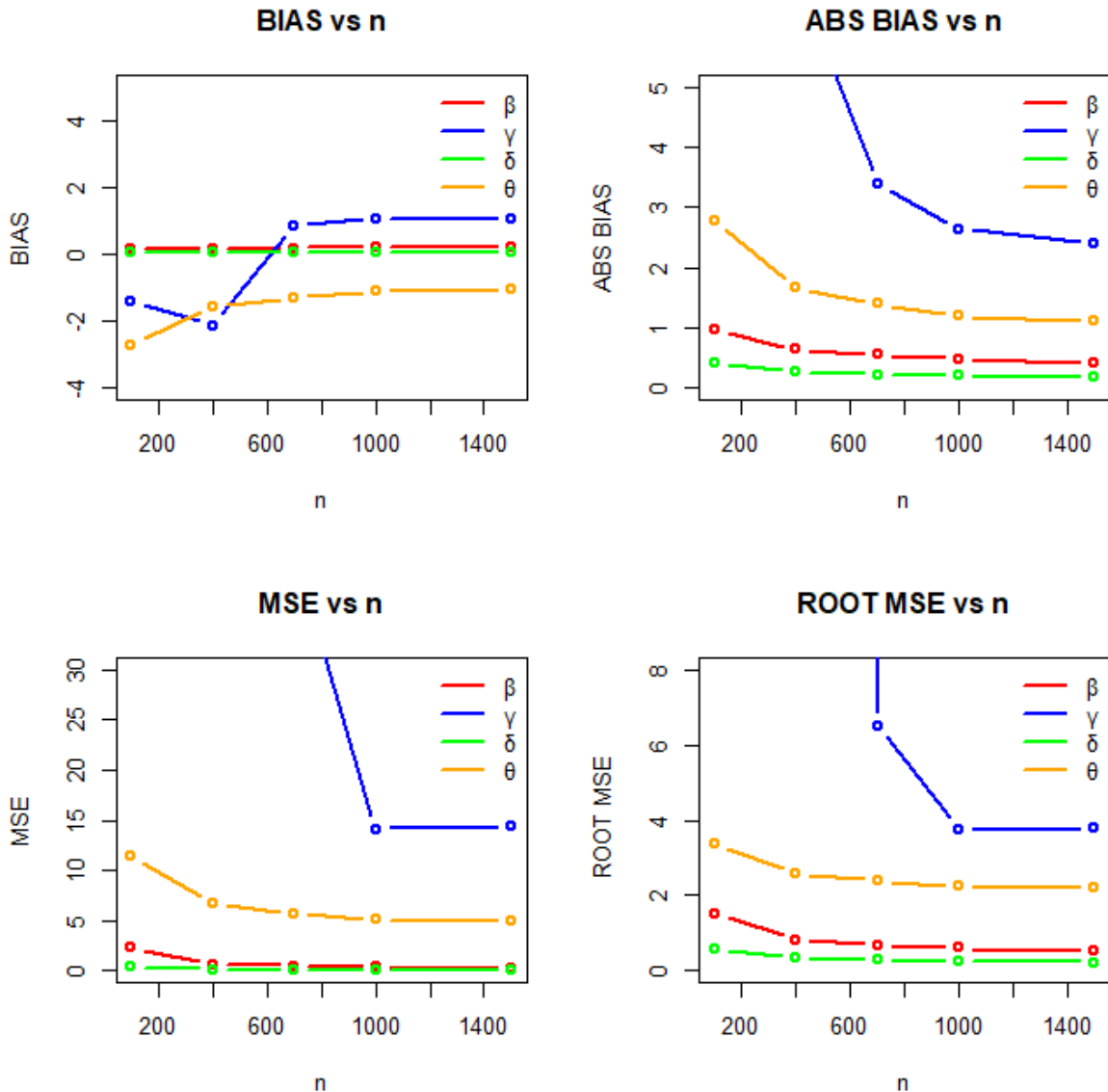
Σχήμα 7.2 Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους  $\alpha, \theta, \lambda, \sigma$ .

<i>L</i> GPD	<i>ΜΕΡΟΛΗΨΙΑ</i>				<i>ΜΤΣ</i>			
	$\alpha$	$\theta$	$\lambda$	$\sigma$	$\alpha$	$\theta$	$\lambda$	$\sigma$
$n$								
100	-2666	-1,95	-3035,00	0,021	1690894	41,76	2388942	0,008
700	-0,04	-0,13	0,036	0,0026	0,85	0,39	1,43	0,0008
1500	0,14	-0,12	0,21	-0,0009	0,44	0,14	0,7	0,0003

Πίνακας 7.2 Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων  $\alpha, \theta, \lambda, \sigma$  για επιλεγμένες τιμές του  $n$ .

#### 7.4 Μελέτη προσομοίωσης για την συγκολλημένη $GGPD(\beta, \gamma, \delta, \theta)$

Για τις παραμέτρους  $\beta = 5, \gamma = 2, \delta = 3, \theta = 1$  προσομοιώσαμε δείγματα από την κατανομή  $GGPD$  ακολουθώντας τα βήματα της ενότητας 7.1 και χρησιμοποιώντας την μέθοδο της διακριτής μίξης. Στο σχήμα 7.3 παρουσιάζονται γραφικά τα αποτελέσματα της προσομοίωσης.



**Σχήμα 7.3** Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους  $\beta, \gamma, \delta, \theta$ .

Βάσει του σχήματος 7.3 συμπεραίνουμε πως οι εκτιμητές των παραμέτρων παρουσιάζουν βελτίωση για μεγάλα μεγέθη δείγματος ( $n > 1000$ ). Ωστόσο για τους εκτιμητές των  $\theta, \gamma$  βλέπουμε πως υπάρχουν ενδείξεις αρνητικής και θετικής μεροληψίας αντίστοιχα όπως επίσης ότι παρουσιάζουν υψηλές τιμές στο μέσο τετραγωνικό σφάλμα. Στον πίνακα 7.3 συνοψίζουμε ορισμένες από τις τιμές των παραπάνω μέτρων για επιλεγμένα μεγέθη δείγματος.

<i>GGPD</i>	<i>ΜΕΡΟΛΗΨΙΑ</i>				<i>ΜΤΣ</i>			
<i>n</i>	$\beta$	$\gamma$	$\Delta$	$\theta$	$\beta$	$\gamma$	$\delta$	$\theta$
100	0,16	-1,4	0,07	-2,7	2,28	9726	0,35	11,4
700	0,18	0,83	0,07	-1,3	0,47	42,1	0,07	5,7
1500	0,19	1,05	0,08	-1,03	0,29	14,3	0,05	4,9

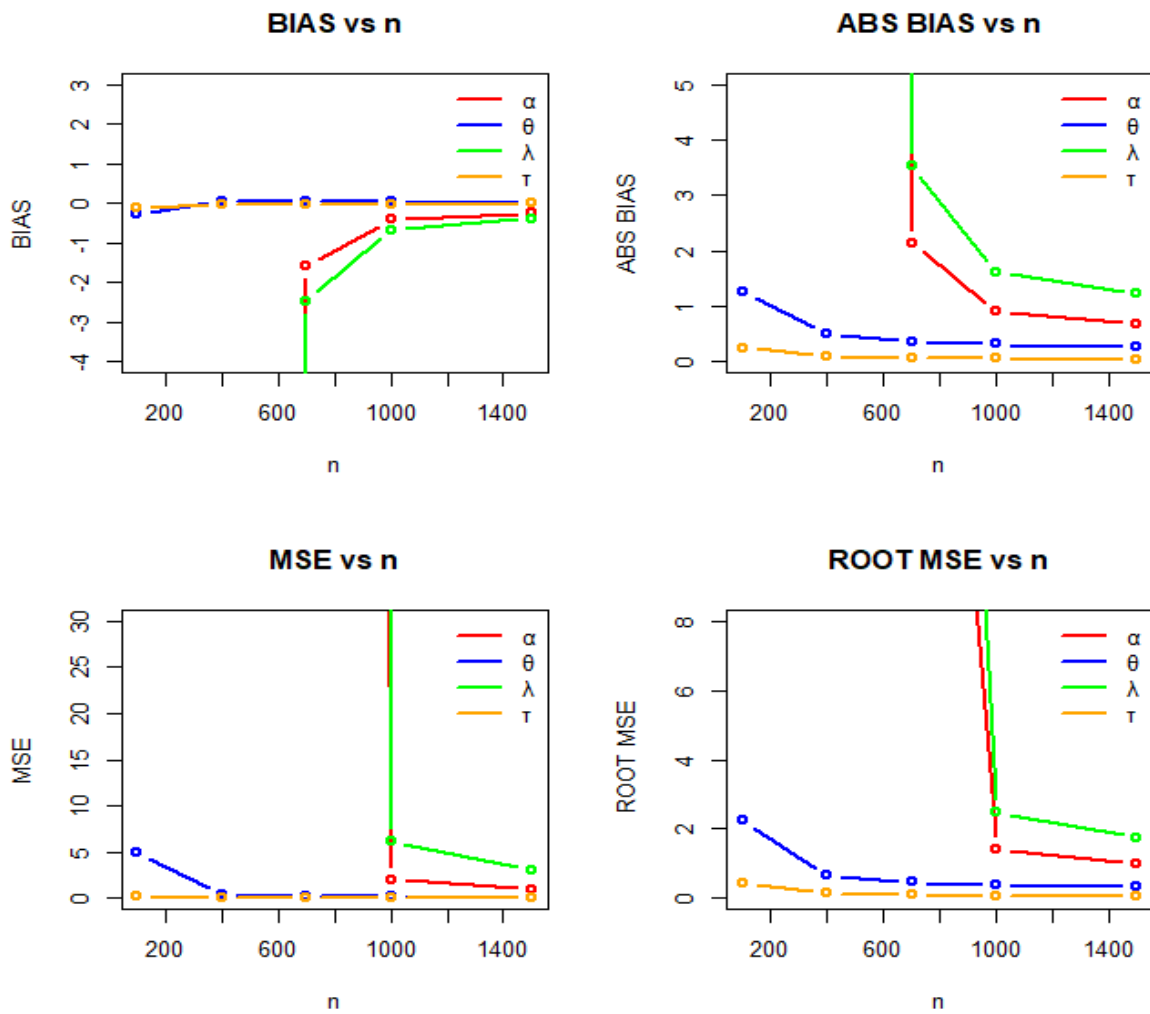
**Πίνακας 7.3** Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων  $\beta, \gamma, \delta, \theta$  για επιλεγμένες τιμές του  $n$ .

### 7.5 Μελέτη προσομοίωσης για την συγκολλημένη *WGPD* ( $\alpha, \theta, \lambda, \tau$ )

Για τις παραμέτρους  $\alpha = 3, \theta = 4, \lambda = 0.5, \tau = 2$  προσομοιώσαμε δείγματα από την κατανομή *WGPD* ακολουθώντας τα βήματα της ενότητας 7.1 και χρησιμοποιώντας την μέθοδο της διακριτής μίξης. Στο σχήμα 7.4 παρουσιάζονται γραφικά τα αποτελέσματα της προσομοίωσης.

Παρατηρούμε μέσω του σχήματος 7.4 πως οι αποδόσεις των ε.μ.π των  $\theta, \tau$  είναι ικανοποιητικές για μεγέθη δείγματος  $n \geq 400$ . Σχετικά με τις ε.μ.π. των  $\alpha, \lambda$  βλέπουμε πως για μεγέθη  $n < 1000$  αποδίδουν υψηλές τιμές μέσου τετραγωνικού σφάλματος καθώς επίσης υπάρχουν ενδείξεις αρνητικής μεροληψίας για τους δύο εκτιμητές. Για μεγέθη  $n \geq 1000$  οι παραπάνω τιμές βελτιώνονται σημαντικά εφόσον φαίνεται να συγκλίνουν στο μηδέν.

Στον πίνακα 7.4 συνοψίζουμε ορισμένες από τις τιμές των παραπάνω μέτρων για επιλεγμένα μεγέθη δείγματος.



Σχήμα 7.4 Γραφικά αποτελέσματα της προσομοίωσης για τις παραμέτρους  $\alpha, \theta, \lambda, \tau$ .

<i>W</i> GPD	<i>ΜΕΡΟΛΗΨΙΑ</i>				<i>ΜΤΣ</i>			
	$\alpha$	$\theta$	$\lambda$	$\tau$	$\alpha$	$\theta$	$\lambda$	$\tau$
$n$								
100	-34087	-0,27	-58033	-0,11	140918	4,9	4127234	0,2
700	-1,5	0,04	-2,4	-0,01	910	0,2	2155	0,009
1500	-0,24	0,004	-0,37	-0,005	1,02	0,1	3,07	0,004

Πίνακας 7.4 Αριθμητικά αποτελέσματα της μεροληψίας και του μ.τ.σ. των παραμέτρων  $\alpha, \theta, \lambda, \tau$  για επιλεγμένες τιμές του  $n$ .

## ΚΕΦΑΛΑΙΟ 8: ΕΦΑΡΜΟΓΗ ΣΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ

### 8.1 Εισαγωγή

Στο κεφάλαιο αυτό θα προσαρμόσουμε τις κατανομές LP, LGPD, GGPD, WGPD σε δύο σετ δεδομένων. Το 1<sup>ο</sup> σετ δεδομένων αποτελείται από 1823 παρατηρήσεις που αφορούν τα ύψη των ασφαλιστικών αποζημιώσεων που προκλήθηκαν από πυρκαγιές στο Βέλγιο. Τα δεδομένα συλλέχθηκαν από την Aon Belgium. Το 2<sup>ο</sup> σετ δεδομένων αποτελείται από 1500 παρατηρήσεις που αφορούν έξοδα ALAE. Τα ALAE είναι η κατηγορία των ασφαλιστικών εξόδων που αφορούν τον διακανονισμό των ζημιών όπως π.χ, τα έξοδα δικηγόρων ή εμπειρογνομόνων. Τα δεδομένα συλλέχθηκαν από την Insurance Services Office, Inc. Τα δύο σετ δεδομένων βρίσκονται στο στατιστικό πακέτο R και συγκεκριμένα στο πακέτο “CASdatasets” με ονομασίες “beaonre” και “lossalae” αντίστοιχα. Στην συνέχεια θα συγκρίνουμε τις συγκολλημένες κατανομές με τις βασικές κατανομές Lognormal, Gamma, Weibull, Pareto πάνω στα δύο σετ δεδομένων ως προς τα παρακάτω μέτρα:

- AIC ( Akaike information criterion ):

$$AIC = 2k - 2\ln(\hat{L})$$

όπου  $\hat{L}$  είναι η συνάρτηση πιθανοφάνειας υπολογισμένη πάνω στους ε.μ.π των παραμέτρων και  $k$  είναι το πλήθος των παραμέτρων . Η τιμή του κριτηρίου AIC δεν μας υποδεικνύει την καλή προσαρμογή του μοντέλου στα δεδομένα, όμως αποτελεί ένα σχετικό μέτρο σύγκρισης μεταξύ των κατανομών. Το μέγεθος αυτό ερμηνεύεται και ως “μέτρο απώλειας πληροφορίας” συνεπώς προτιμάται η κατανομή με το μικρότερο AIC.

- BIC ( Bayesian information criterion ):

$$BIC = k\ln(n) - 2\ln(\hat{L})$$

όπου  $\hat{L}$  είναι η συνάρτηση πιθανοφάνειας υπολογισμένη πάνω στους ε.μ.π των παραμέτρων και  $k$  είναι το πλήθος των παραμέτρων. Η λογική με την οποία συγκρίνουμε τα μοντέλα μέσω του κριτηρίου BIC είναι παρόμοια με αυτή που είδαμε για το AIC. Η κύρια διαφορά έγκειται στο ότι το κριτήριο BIC ορίζει πιο αυστηρή “ποινή” στα μοντέλα με πολλές παραμέτρους και αυτό φαίνεται από τον ορισμό του διότι όσο το  $k$  αυξάνεται τόσο αυξάνεται και ο όρος  $k \log(n)$  σε σχέση με τον όρο  $2k$  του κριτηρίου AIC.



Για τον έλεγχο καλής προσαρμογής των κατανομών στα δεδομένα θα χρησιμοποιήσουμε τα τέστ καλής προσαρμογής Kolmogorov – Smirnov (K – S) και Anderson – Darling (A – D). Αν θεωρήσουμε ότι  $F_n, F$  είναι η εμπειρική συνάρτηση κατανομής και η συνάρτηση κατανομής του εξεταζόμενου μοντέλου αντίστοιχα τότε τα στατιστικά των παραπάνω ελέγχων ορίζονται ως εξής:

- Για K – S:

$$D_n = \sup_x |F_n(x) - F(x)|$$

- Για A – D:

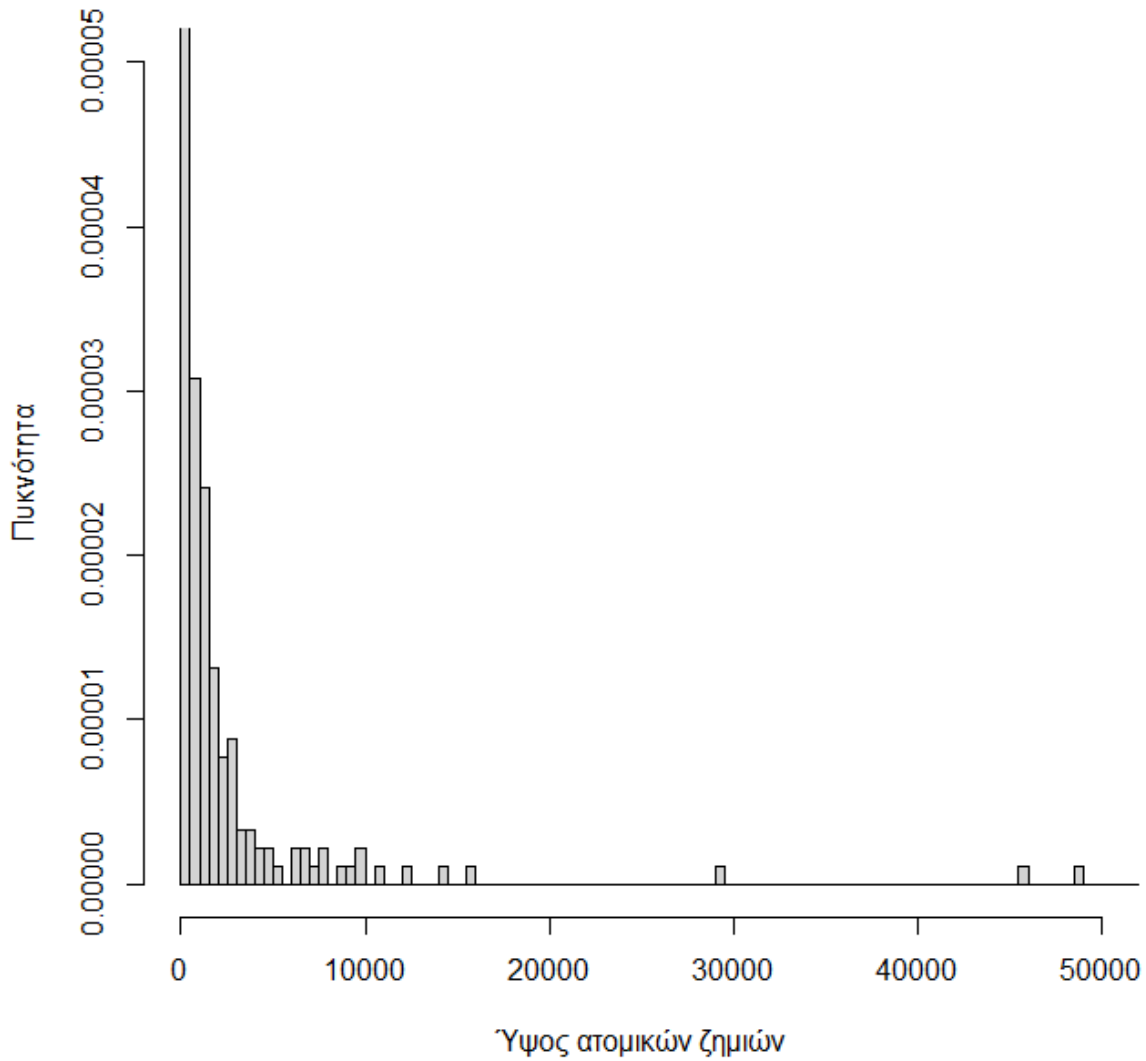
$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x)$$

Τα παραπάνω στατιστικά δηλώνουν την έννοια της “απόστασης” μεταξύ της εμπειρικής και της θεωρητικής κατανομής των δεδομένων που μελετάμε. Συνεπώς, θα θέλαμε οι τιμές των παραπάνω συναρτήσεων να είναι μικρές ώστε να μην απορρίπτουμε την υπόθεση ότι τα δεδομένα προέρχονται από την προτεινόμενη κατανομή  $F$ .

## 8.2 Προσαρμογή στο σετ δεδομένων “beaonre”

Στο σχήμα 8.1 παρουσιάζεται το ιστόγραμμα των δεδομένων. Παρατηρούμε πως οι ζημιές ακολουθούν κάποια λοξή κατανομή η οποία έχει βαριά ουρά. Τα περιγραφικά μέτρα του δείγματος συνοψίζονται στον πίνακα 8.1 και επαληθεύουν τις παρατηρήσεις μας καθώς ο συντελεστής κύρτωσης ισούται με 1289,86 και ο συντελεστής λοξότητας ισούται με 33.93. Η μέση τιμή είναι ίση με 363,4 ενώ η τιμή της τυπικής απόκλισης (4868,2) δηλώνει την υψηλή μεταβλητότητα των δεδομένων.

## Ιστόγραμμα δεδομένων



Σχήμα 8.1 Ιστόγραμμα των δεδομένων “beaopre”.

Περιγραφικά μέτρα	
Μέση τιμή	363.464
Διασπορά	23699945
Τυπική απόκλιση	4868.25
Διάμεσος	9.073
Εύρος	190541.54
Μέγιστη τιμή	190541.69
Ελάχιστη τιμή	0.153
Συντελεστής κύρτωσης	1289.86
Συντελεστής λοξότητας	33.93

Πίνακας 8.1 Περιγραφικά μέτρα των δεδομένων “beaopre”.

Στον πίνακα 8.2 παρουσιάζονται τα αποτελέσματα της ανάλυσης των δεδομένων. Παρατηρούμε πως οι συγκολλημένες κατανομές LGPD, WGPLD περνούν τα τεστ καλής

προσαρμογής των K – S και A – D σε επίπεδο 5%. Η κατανομή GGPD επιτυγχάνει την καλή προσαρμογή βάσει του κριτηρίου A - D σε επίπεδο 5% ενώ παράλληλα αποτυγχάνει στο τεστ K – S ( p-value=0.031). Επίσης, Θα πρέπει να λάβουμε υπόψη πως το κριτήριο A – D δίνει περισσότερο βάρος στην ουρά της κατανομής σε σχέση με το K – S. Τέλος, η κατανομή LP αποτυγχάνει την καλή προσαρμογή βάσει των δύο κριτηρίων σε επίπεδο 5%. Διαπιστώνουμε πως καμία από τις βασικές ζημιοκατανομές δεν περνά τα τεστ καλής προσαρμογής συνεπώς μέσα από την εφαρμογή αυτή αναγνωρίζουμε την ανάγκη για την χρήση πιο ευέλικτων μοντέλων.

Κατανομή	Παράμετροι (Ε.Μ.Π)	AIC	BIC	K-S (p-value)	A-D (p-value)
LP	$\alpha=0.4889$	16820	16837	0.04565 (0.00101)	3.4749 (0.01582)
	$\theta=7.1524$				
	$\sigma=0.9538$				
LGPD	$\alpha=0.578$	16789	16811	0.02961 (0.08177)	0.8132 (0.47158)
	$\theta=2.868$				
	$\lambda=2.528$				
	$\sigma=0.805$				
GGPD	$\beta=0.5022$	16814	16836	0.03375 (0.03146)	1.9074 (0.10333)
	$\gamma=1.4834$				
	$\delta=2.0246$				
	$\theta=4.3314$				
WGPD	$\alpha=0.5809$	16790	16812	0.02744 (0.12835)	0.7380 (0.52788)
	$\theta=1.8512$				
	$\lambda=2.6890$				
	$\tau=2.6179$				
Gamma	$\delta=0.2206$	19350	19361	0.28541 (0.00000)	Inf (0.00000)
	$\beta=1653.56$				
Lognormal	$\mu=2.606$	17160	17171	0.08164 (0.00000)	29.7923 (0.00000)
	$\sigma=1.974$				
Weibull	$\tau=0.4106$	17999	18010	0.16607 (0.00000)	96.2923 (0.00000)
	$\phi=39.4249$				
Pareto	$\alpha=0.223$	18622	18633	0.31994 (0.00000)	Inf (0.00000)
	$\theta=0.153$				

Πίνακας 8.2 Αποτελέσματα της ανάλυσης των δεδομένων “beaonre”.

Βάσει των κριτηρίων AIC, BIC μπορούμε να πραγματοποιήσουμε την σύγκριση μεταξύ των υποψήφιων κατανομών για την περιγραφή του συνόλου των δεδομένων. Παρατηρούμε στον

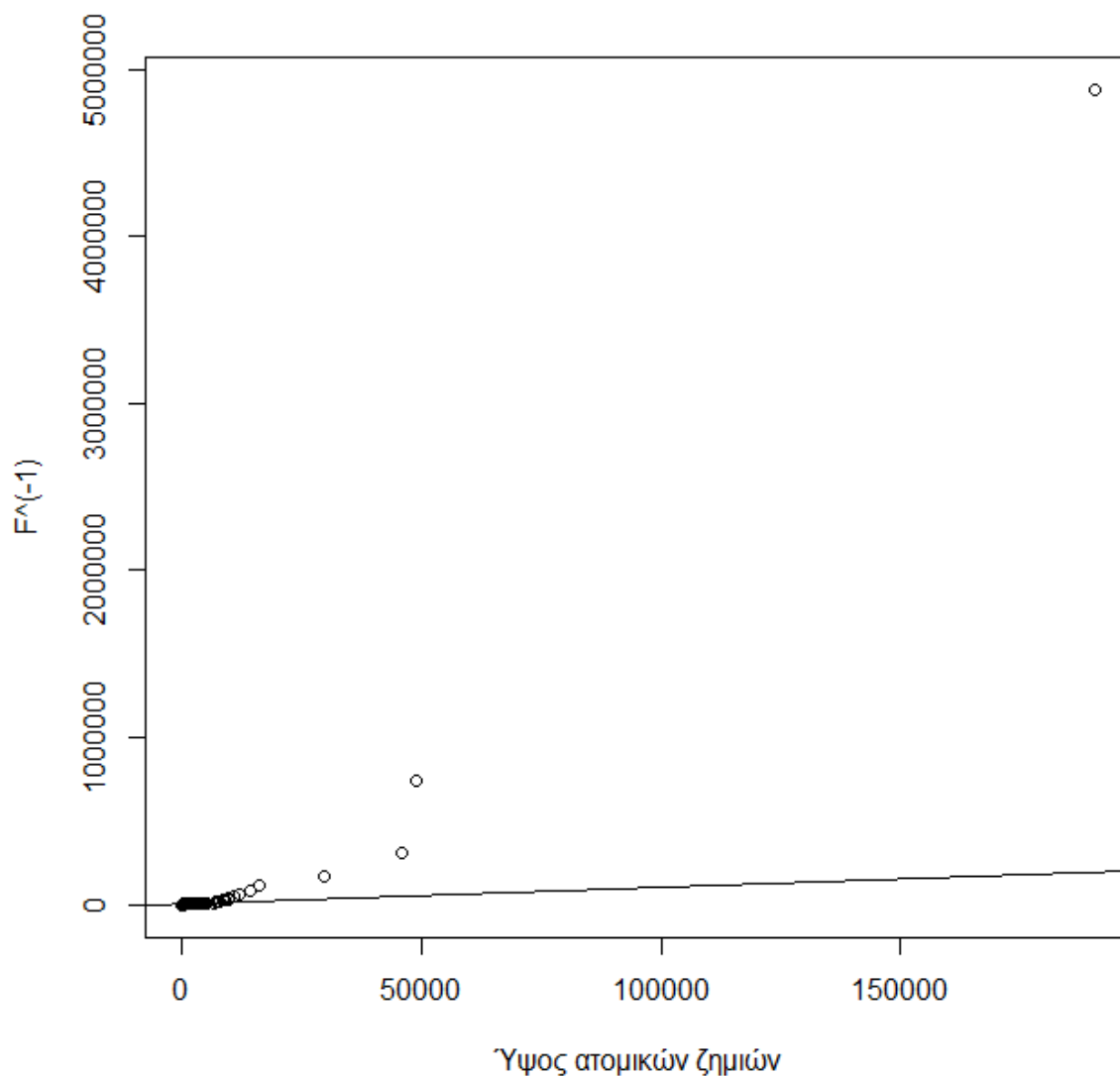
πίνακα 8.2 πως οι τιμές των AIC και BIC των συγκολλημένων κατανομών είναι πολύ μικρότερες σε σχέση με τις αντίστοιχες των απλών ζημιοκατανομών. Το γεγονός αυτό δείχνει την υπεροχή των συγκολλημένων κατανομών για την περιγραφή των δεδομένων. Για την επιλογή του βέλτιστου μοντέλου είναι προφανές ότι θα πρέπει να διαλέξουμε μεταξύ των κατανομών LGPD, WGPD οι οποίες παρουσιάζουν μικρότερες τιμές στα κριτήρια AIC, BIC. Ωστόσο η διαφορά μεταξύ αυτών των τιμών θεωρείται ασήμαντη διότι είναι  $\leq 1$ . Συνεπώς, καταλήγουμε στο ότι για την περιγραφή των δεδομένων, οι κατανομές LGPD και WGPD αποτελούν τις βέλτιστες επιλογές.

Κατανομή	Επίπεδο σημαντικότητας						
	80%	85%	90%	95%	97.5%	99%	99.5%
LP	52,59	94,72	217,1	896,3	3700,1	24112	99541
LGPD	54,45	91,2	186,5	624,6	2077,9	10151	33682
GGPD	53,4	91,62	194,6	699	2500,9	13468	48116
WGPD	54,84	91,71	187	622,9	2060,2	9986	32935
Gamma	503,9	732,03	1101,8	1830,3	2643,8	3806	4730
Weibull	125,63	187,52	300,5	570,5	947,1	1626	2287
Lognormal	71,33	104,79	170	348,4	649	1338	2190
Pareto	208,19	756,17	4657,1	104173,6	2330248,1	141753925	31708781114
Δειγματικά	57,38	90	205,5	687,7	1936	6154	9492

Πίνακας 8.3 Θεωρητικά – Δειγματικά  $VaR_p$  ανά επίπεδο σημαντικότητας.

Στον πίνακα 8.3 παρουσιάζονται τα  $VaR$  των κατανομών σε σχέση με τα αντίστοιχα του δείγματος. Σκοπός είναι να ελέγξουμε αν οι συγκολλημένες κατανομές μοντελοποιούν την ουρά των δεδομένων με μεγαλύτερη αποτελεσματικότητα. Παρατηρούμε πως η κατανομή Pareto υπερεκτιμά την ουρά των δεδομένων κάτι το οποίο μπορεί να οδηγήσει σε υπερβολικά υψηλά ασφάλιστρα κατά την τιμολόγηση. Από την άλλη, οι κατανομές Lognormal, Weibull παρουσιάζουν ουρά που φθίνει αρκετά γρήγορα στο μηδέν. Οι συγκολλημένες κατανομές LGPD, WGPD φαίνεται να συλλαμβάνουν τα χαρακτηριστικά της ουράς των δεδομένων καθώς μέχρι το επίπεδο 97.5% οι διαφορές είναι μικρές. Από το επίπεδο 97.5% και μετά παρουσιάζουν αρκετά πιο βαριά ουρά σε σχέση με αυτή του δείγματος όμως φαίνεται ότι συγκριτικά με τις υπόλοιπες κατανομές εμφανίζουν τα καλύτερα αποτελέσματα. Συνεπώς, από τον πίνακα 8.3 μπορούμε να επαληθεύσουμε την επιλογή της βέλτιστης κατανομής που έγινε με την χρήση των κριτηρίων AIC και BIC.

Στο σχήμα 8.2 παρατηρούμε τα ποσοστιαία σημεία της κατανομής LGPD σε σχέση με τα αντίστοιχα δειγματικά.

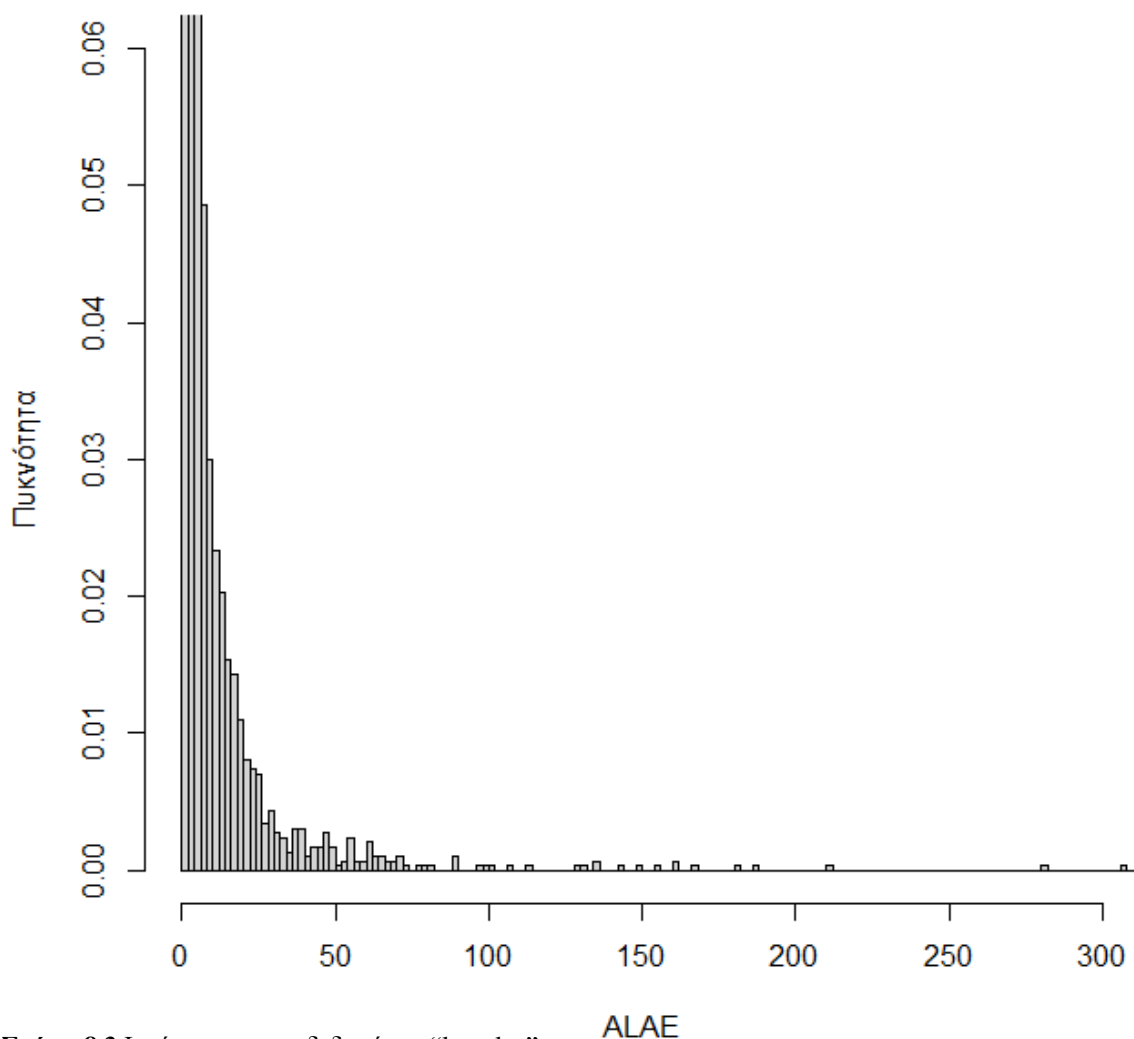


Σχήμα 8.2 Q-Q plot των δεδομένων έναντι της κατανομής LGPD.

### 8.3 Προσαρμογή στο σετ δεδομένων "lossalae"

Στο σχήμα 8.3 παρουσιάζεται το ιστόγραμμα των δεδομένων. Τα δεδομένα παρουσιάζουν βαριά ουρά και υψηλή θετική ασυμμετρία. Οι τιμές των συντελεστών κύρτωσης και λοξότητας είναι 125,13 και 9,25 αντίστοιχα. Η μέση τιμή ισούται με 12,58 ενώ η τιμή της τυπικής απόκλισης είναι 28,14.

## Ιστόγραμμα δεδομένων



Σχήμα 8.3 Ιστόγραμμα των δεδομένων “lossalae”

ALAE

Περιγραφικά μέτρα	
Μέση τιμή	12.588
Διασπορά	792.177
Τυπική απόκλιση	28.145
Διάμεσος	5.471
Εύρος	501.848
Μέγιστη τιμή	501.863
Ελάχιστη τιμή	0.015
Συντελεστής κύρτωσης	125.136
Συντελεστής λοξότητας	9.259

Πίνακας 8.4 Περιγραφικά μέτρα των δεδομένων “lossalae”.

Στον πίνακα 8.5 παρουσιάζονται τα αποτελέσματα της ανάλυσης των δεδομένων. Παρατηρούμε πως λόγω των χαμηλών  $p$  – values ( $<0.05$ ) οι κατανομές LP, Gamma, Lognormal, Weibull, Pareto δεν θεωρούνται ιδανικές για την περιγραφή του συνόλου των δεδομένων. Οι συγκολλημένες κατανομές LGPD, GGPD, WGPD περνούν τα τεστ καλής προσαρμογής K – S και A – D σε επίπεδο 5% συνεπώς μπορούμε να θεωρήσουμε ότι κάθε μία από τις παραπάνω κατανομές είναι πιθανό να έχει παράξει τα δεδομένα.

Κατανομή	Παράμετροι (Ε.Μ.Π)	AIC	BIC	K-S (p-value)	A-D (p-value)
LP	$\alpha=2.155$	10177.55	10193.49	0.049 (0.001)	6.285 (0.000)
	$\theta=408.635$				
	$\sigma=1.428$				
LGPD	$\alpha=2.197$	10109.51	10130.77	0.029 (0.151)	1.228 (0.257)
	$\theta=0.567$				
	$\lambda=14.799$				
	$\sigma=2.915$				
GGPD	$\beta=0.385$	10109.86	10131.11	0.028 (0.179)	1.120 (0.299)
	$\gamma=14.652$				
	$\delta=1.132$				
	$\theta=0.743$				
WGPD	$\alpha=1.381$	10109.96	10131.22	0.019 (0.600)	0.568 (0.678)
	$\theta=16.867$				
	$\lambda=1.347$				
	$\tau=0.979$				
Gamma	$\delta=0.663$	10404.08	10414.71	0.099 (0.000)	25.847 (0.000)
	$\beta=18.991$				
Lognormal	$\mu=1.614$	10175.29	10185.92	0.049 (0.001)	6.182 (0.000)
	$\sigma=1.429$				
Weibull	$\tau=0.741$	10271.06	10281.68	0.060 (0.000)	14.061 (0.000)
	$\phi=9.979$				
Pareto	$\alpha=0.172$	13127.43	13138.06	0.397 (0.000)	Inf (0.000)
	$\theta=0.015$				

Πίνακας 8.5 Αποτελέσματα της ανάλυσης των δεδομένων “Iossalae”.

Με σκοπό την σύγκριση των παραπάνω μοντέλων θα χρησιμοποιήσουμε τις τιμές των κριτηρίων AIC και BIC. Παρατηρούμε πως οι κατανομές LGPD, GGPD, WGPD παρουσιάζουν τις μικρότερες τιμές στα παραπάνω κριτήρια ωστόσο οι διαφορές μεταξύ των AIC και BIC για τις τρεις κατανομές θεωρούνται μη σημαντικές εφόσον είναι  $<1$ . Καταλήγουμε στο ότι για την περιγραφή του συνόλου των δεδομένων μπορεί να

χρησιμοποιηθεί οποιαδήποτε από τις κατανομές LGPD, GGPD, WGPD. Μέσα από την εφαρμογή αυτή διαπιστώνουμε την χρησιμότητα των συγκολλημένων κατανομών διότι η ευελιξία και η προσαρμοστικότητα τους προσδίδει καλύτερα αποτελέσματα κατά την μοντελοποίηση των δεδομένων σε σχέση με τις βασικές ζημιοκατανομές.

Κατανομή	Επίπεδο σημαντικότητας						
	80%	85%	90%	95%	97.5%	99%	99.5%
LP	16,68	22,03	31,28	52,60	82,61	139,83	200,57
LGPD	16,04	20,35	27,47	43,15	64,64	105,75	150,45
GGPD	16,01	20,31	27,43	43,12	64,64	105,88	150,80
WGPD	14,87	18,63	25,45	42,91	71,75	140,54	232,98
Gamma	20,73	25,34	32,01	43,70	55,67	71,75	84,07
Weibull	18,95	23,66	30,72	43,81	58,00	78,22	94,50
Lognormal	16,74	22,11	31,39	52,78	82,82	139,83	199,76
Pareto	173,72	925,20	9772,76	549773,10	30927847,30	636712689,00	358186869,00
Δειγματικά	15,71	19,22	25,92	45,96	64,86	131,70	163,76

Πίνακας 8.6 Θεωρητικά – Δειγματικά  $VaR_p$  ανά επίπεδο σημαντικότητας.

Στους πίνακες 8.6 και 8.7 παρουσιάζεται η σύγκριση μεταξύ των θεωρητικών και δειγματικών  $VaR$  και  $TVaR$ . Σκοπός είναι να επαληθεύσουμε τα ευρήματά μας μέσα από την μελέτη της ουράς των δεδομένων. Παρατηρούμε πως οι κατανομές LP, Lognormal, Pareto τείνουν να υπερεκτιμούν την ουρά των δεδομένων ενώ οι κατανομές Gamma και Weibull εμφανίζουν πιο ελαφριά ουρά απ' ό,τι αυτή των δεδομένων με αποτέλεσμα την υποεκτίμηση των ζημιών που παρουσιάζουν υψηλή σφοδρότητα, δηλαδή αυτές που εμφανίζονται στην ουρά της κατανομής.

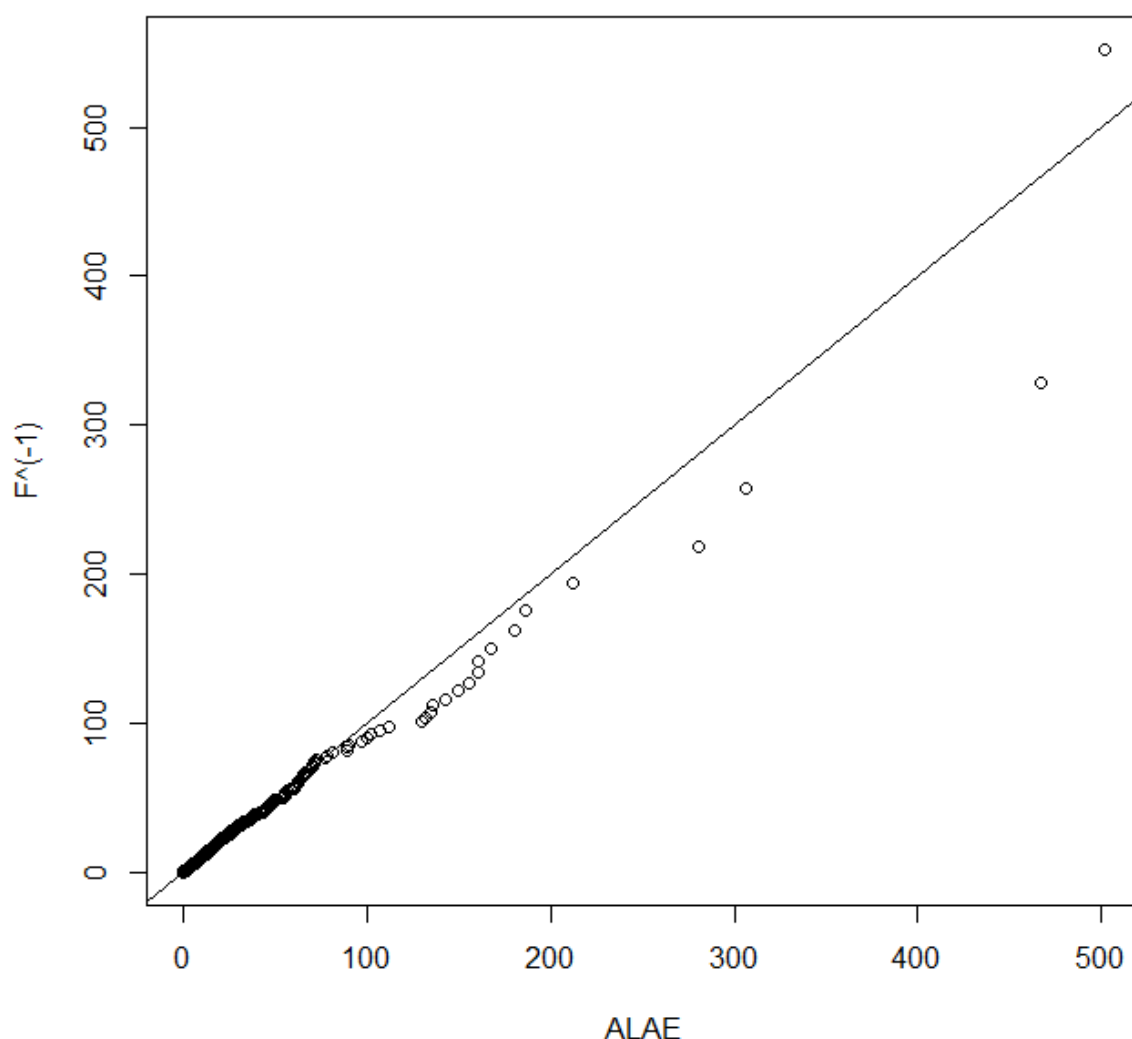
Κατανομή	Επίπεδο σημαντικότητας						
	80%	85%	90%	95%	97.5%	99%	99.5%
LP	51,01	61,63	79,36	118,63	172,18	272,89	380,50
LGPD	41,79	49,71	62,78	91,55	130,99	206,42	288,47
GGPD	41,81	49,74	62,85	91,73	131,37	207,31	290,02
WGPD	57,40	71,00	95,68	158,90	263,30	512,33	846,99
Gamma	37,42	42,25	49,16	61,16	73,34	89,63	102,07
Weibull	37,49	42,93	50,94	65,47	80,94	102,67	119,98
Lognormal	50,39	60,78	78,04	115,85	166,43	258,26	351,51
Pareto	NA	NA	NA	NA	NA	NA	NA
Δειγματικά	42,86	51,39	65,88	97,64	139,76	222,68	287,64

Πίνακας 8.7 Θεωρητικά – Δειγματικά  $TVaR_p$  ανά επίπεδο σημαντικότητας.



Συμπεραίνουμε από τους πίνακες 8.6 και 8.7 πως οι κατανομές LGPD, GGPD, WGPD συλλαμβάνουν επαρκώς τα χαρακτηριστικά της ουράς των δεδομένων εφόσον οι αποκλίσεις μεταξύ των θεωρητικών και των δειγματικών ποσοστιαίων σημείων είναι αρκετά μικρές. Επομένως, τα παραπάνω ευρήματα είναι συνεπή ως προς την επιλογή των βέλτιστων μοντέλων που πραγματοποιήθηκε βάσει των κριτηρίων AIC και BIC.

Στο σχήμα 8.4 παρουσιάζουμε τα ποσοστιαία σημεία της κατανομής LGPD σε σχέση με τα αντίστοιχα δειγματικά.



Σχήμα 8.4 Q-Q plot των δεδομένων έναντι της κατανομής LGPD.

## ΚΕΦΑΛΑΙΟ 9: ΣΥΜΠΕΡΑΣΜΑΤΑ

Με σκοπό την μοντελοποίηση του ύψους των ατομικών ζημιών ενός χαρτοφυλακίου παρουσιάσαμε την έννοια των συγκολλημένων κατανομών. Η έννοια αυτή εμφανίσθηκε για πρώτη φορά στην μελέτη των Cooray – Ananda (2005) και έπειτα γενικεύθηκε από τον Scollnik (2007). Η έρευνα πάνω στις συγκολλημένες κατανομές στηρίζεται έως και σήμερα στον γενικότερο ορισμό που δόθηκε από τον Scollnik. Με την βοήθεια των σύγχρονων συστημάτων και αριθμητικών μεθόδων μπορούμε να γενικεύσουμε το πλαίσιο της έρευνας των συγκολλημένων κατανομών και να χρησιμοποιήσουμε πιο σύνθετες κατανομές ως συνιστώσες π.χ, Burr, Log – Logistic (Grün & Miljkovic, 2019). Γενικά, για την κατασκευή ενός μοντέλου συγκολλημένης κατανομής αρκεί να θεωρήσουμε δύο από τις ήδη γνωστές ζημιοκατανομές, να περιορίσουμε τις συναρτήσεις πυκνότητάς τους σε κατάλληλα διαστήματα (έως ένα κατώφλι  $\theta$ ) και να εφαρμόσουμε πάνω σε αυτές την έννοια της διακριτής μίξης. Μελετήσαμε τις περιπτώσεις των Lognormal – Pareto, Lognormal – GPD, Gamma – GPD και Weibull – GPD και καταλήξαμε στους κλειστούς τύπους για τον υπολογισμό των ροπών και των μέτρων κινδύνου VaR και TVaR. Για τον υπολογισμό των ε.μ.π. εφαρμόσαμε τις αριθμητικές μεθόδους του πακέτου R και διαπιστώσαμε μέσω της μελέτης προσομοίωσης πως η απόδοσή τους είναι ικανοποιητική ως προς τα μέτρα της μεροληψίας και του μέσου τετραγωνικού σφάλματος. Στην τελική ανάλυση εφαρμόσαμε τα νέα μοντέλα σε δύο σειτ πραγματικών δεδομένων και διαπιστώσαμε την υπεροχή των συγκολλημένων κατανομών έναντι των βασικών ζημιοκατανομών που τις απαρτίζουν (π.χ Lognormal, Pareto). Καταλήγουμε στο ότι η χρήση αυτών των κατανομών μπορεί να προσφέρει πολλά πλεονεκτήματα κατά την μελέτη των ζημιών εφόσον προσφέρει μεγαλύτερη ευελιξία και προσαρμοστικότητα στην ιδιάζουσα συμπεριφορά των δεδομένων. Συνεπώς, ο αναλογιστής με την χρήση των συγκολλημένων κατανομών είναι σε θέση να πραγματοποιήσει μία πιο λεπτομερή ανάλυση στα δεδομένα που διαθέτει και έτσι να εκτιμήσει καλύτερα τα ασφάλιστρα καθώς και τα υπόλοιπα σχετικά μεγέθη που συνδέονται με την μελέτη των ζημιών του χαρτοφυλακίου.

## ΠΑΡΑΡΤΗΜΑ

Για τους παρακάτω κώδικες θα χρειαστεί να φορτώσουμε στο λογισμικό R τα εξής πακέτα:

1. Actuar
2. BBmisc
3. CASdatasets
4. DescTools
5. e1071
6. ggplot2
7. GoFKernel
8. GofTest
9. MASS
10. purrr
11. rootSolve
12. stats4
13. truncdist
14. EnvStats
15. pracma

Για να ορίσουμε μία νέα κατανομή στο R χρειάζεται να ορίσουμε την συνάρτηση κατανομής, την συνάρτηση πυκνότητας και την συνάρτηση ποσοστιαίων σημείων. Θα πρέπει δηλαδή να ορίσουμε τα  $p$ ,  $d$ ,  $q$  για την κατανομή.

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο ορίζουμε την συγκολλημένη κατανομή LP:](#)

```
rlp=function(alphalp,sigmalp) {  
  A=sqrt(2*pi)*alphalp*sigmalp*pnorm(alphalp*sigmalp,0,1)*exp(alphalp^2*sigmalp^2/2)  
  return(A/(A+1))  
}  
milp=function(alphalp,thetalp,sigmalp) {  
  return(log(thetalp)-alphalp*sigmalp^2)  
}
```

```
dlp=function(x,alphalp,thetalp,sigmalp) {  
  d=rlp(alphalp,sigmalp)*dtrunc(x,meanlog=milp(alphalp,thetalp,sigmalp),sdlog=sigmalp,spec="lnorm",  
  a=0,b=thetalp)+(1-rlp(alphalp,sigmalp))*dpareto1(x,shape=alphalp,min=thetalp+0.0001)  
  return(d)  
}
```

```
plp=function(x,alphalp,thetalp,sigmalp) {  
  d=rlp(alphalp,sigmalp)*ptrunc(x,meanlog=milp(alphalp,thetalp,sigmalp),sdlog=sigmalp,spec="lnorm",  
  a=0,b=thetalp)+(1-rlp(alphalp,sigmalp))*ppareto1(x,shape=alphalp,min=thetalp)  
  return(d)  
}
```

```

qlp=function(q,alphalp,thetalp,sigmalp) {
d=inverse( function(x) { plp(x,alphalp,thetalp,sigmalp) },lower=0.7)
return(d(q))
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο ορίζουμε την συγκολλημένη κατανομή LGPD:](#)

```

rlgpd=function(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd) {
par=((alphalgpd*thetalgpd-lambdalgpd)*sigmalgpd)/(lambdalgpd+thetalgpd)

```

```

A=sqrt(2*pi)*alphalgpd*thetalgpd*sigmalgpd*pnorm(par,0,1)*exp(par^2/2)
return(A/(A+lambdalgpd+thetalgpd))
}

```

```

milgpd=function(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd) {
par=((alphalgpd*thetalgpd-lambdalgpd)*sigmalgpd)/(lambdalgpd+thetalgpd)

```

```

return(log(thetalgpd)-sigmalgpd*par)
}

```

```

dlgpd=function(x,alphalgpd,thetalgpd,lambdalgpd,sigmalgpd) {
d=rlgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd)*dtrunc(x,meanlog=milgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd),sdlog=sigmalgpd,"lnorm",a=0,b=thetalgpd)+(1-rlgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd))*dpareto2(x,min=thetalgpd+0.0001,shape=alphalgpd,scale=lambdalgpd+thetalgpd)
return(d)
}

```

```

plgpd=function(x,alphalgpd,thetalgpd,lambdalgpd,sigmalgpd) {
d=rlgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd)*ptrunc(x,meanlog=milgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd),sdlog=sigmalgpd,"lnorm",a=0,b=thetalgpd)+(1-rlgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd))*ppareto2(x,min=thetalgpd,shape=alphalgpd,scale=lambdalgpd+thetalgpd)
return(d)
}

```

```

qlgpd=function(x,alphalgpd,thetalgpd,lambdalgpd,sigmalgpd) {
d=inverse( function(x) { plgpd(x,alphalp,thetalp,lambdalgpd,sigmalp) },lower=0.7)
return(d(q))
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο ορίζουμε την συγκολλημένη κατανομή GGPD:](#)

```

alphagppd=function(betagppd,gammagppd,deltagppd,thetagppd) {
alphagppd=(gammagppd+thetagppd)*(betagppd*thetagppd-deltagppd+1)/thetagppd-1
return(alphagppd)
}

```

```

rggpd=function(betagppd,gammagppd,deltagppd,thetagppd) {
ginc=gammainc(betagppd*thetagppd,deltagppd)
A=alphagppd(betagppd,gammagppd,deltagppd,thetagppd)*(betagppd*thetagppd-
deltagppd+1)*ginc[1]
rggpd=A/(A+(alphagppd(betagppd,gammagppd,deltagppd,thetagppd)+1)*(betagppd*thetagppd)^(d
eltagppd)*exp(-betagppd*thetagppd))
return(rggpd)
}

```

```

dggpd=function(x,betagppd,gammagppd,deltagppd,thetagppd) {
d=rggpd(betagppd,gammagppd,deltagppd,thetagppd)*dtrunc(x,shape=deltagppd,scale=1/betagppd,
"gamma",a=0,b=thetagppd)+(1-
rggpd(betagppd,gammagppd,deltagppd,thetagppd))*dpareto2(x,shape=alphagppd(betagppd,gamma
ggpd,deltagppd,thetagppd),min=thetagppd+0.0001,scale=gammagppd+thetagppd)
return(d)
}

```

```

pggpd=function(x,betagppd,gammagppd,deltagppd,thetagppd) {
d=rggpd(betagppd,gammagppd,deltagppd,thetagppd)*ptrunc(x,shape=deltagppd,scale=1/betagppd,
"gamma",a=0,b=thetagppd)+(1-
rggpd(betagppd,gammagppd,deltagppd,thetagppd))*ppareto2(x,shape=alphagppd(betagppd,gamma
ggpd,deltagppd,thetagppd),min=thetagppd+0.0001,scale=gammagppd+thetagppd)
return(d)
}

```

```

qggpd=function(x,betagppd,gammagppd,deltagppd,thetagppd) {
d=inverse( function(x) { pggpd(x,betagppd,gammagppd,deltagppd,thetagppd) },lower=0.7)
return(d(q))
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο ορίζουμε την συγκολλημένη κατανομή WGPD:](#)

```

phiwgpdp=function(alphawgpdp,thetawgpdp,lambdawgpdp,tauwgpdp) {
A=(alphawgpdp*thetawgpdp-lambdawgpdp)/((lambdawgpdp+thetawgpdp)*tauwgpdp)
phiwgpdp=thetawgpdp/((A+1)^(1/tauwgpdp))
return(phiwgpdp)
}

```

```

rwwgpdp=function(alphawgpdp,thetawgpdp,lambdawgpdp,tauwgpdp) {
a_1=alphawgpdp/tauwgpdp
a_2=(lambdawgpdp+thetawgpdp)/thetawgpdp
}

```

```

a_3=(thetawgpd/phiwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd))^(tauwgpd)
A=a_3/(exp(a_3)-1)
rwgpd=a_1/(a_2*A+a_1)
return(rwgpd)
}

```

```

dwgpd=function(x,alphawgpd,thetawgpd,lambdawgpd,tauwgpd) {
d=rwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd)*dtrunc(x,shape=tauwgpd,scale=phiwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd),"weibull",a=0,b=thetawgpd)+(1-rwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd))*dpareto2(x,shape=alphawgpd,min=thetawgpd+0.0001,scale=lambdawgpd+thetawgpd)
return(d)
}

```

```

pwgpd=function(x,alphawgpd,thetawgpd,lambdawgpd,tauwgpd) {
d=rwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd)*ptrunc(x,shape=tauwgpd,scale=phiwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd),"weibull",a=0,b=thetawgpd)+(1-rwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd))*ppareto2(x,shape=alphawgpd,min=thetawgpd+0.0001,scale=lambdawgpd+thetawgpd)
return(d)
}

```

```

qwgpd=function(x,alphawgpd,thetawgpd,lambdawgpd,tauwgpd) {
d=inverse( function(x) { pwgpd(x,alphawgpd,thetawgpd,lambdawgpd,tauwgpd) },lower=0.7)
return(d(q))
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο προσομοιώνουμε δείγμα από την συγκολλημένη κατανομή LP με την μέθοδο της διακριτής μίξης που αναφέραμε στην ενότητα 7.1.](#)

```

simlp=function(n,alphalp,thetalp,sigmalp) {
x=numeric(n)
for (i in 1:n) {
u=runif(1,0,1)
if (u<rIp(alphalp,sigmalp)) {
x[i]=rtrunc(1,meanlog=miIp(alphalp,thetalp,sigmalp),sdlog=sigmalp,"lnorm",a=0,b=thetalp) }
else {
x[i]=rpareto1(1,shape=alphalp,min=thetalp) }
}
return(x)
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο προσομοιώνουμε δείγμα από την συγκολλημένη κατανομή LGPD με την μέθοδο της διακριτής μίξης που αναφέραμε στην ενότητα 7.1.](#)

```

simlpgpd=function(n,alphalpgpd,thetalpgpd,lambdalpgpd,sigmalpgpd) {
x=numeric(n)

```

```

for (i in 1:n) {
  u=runif(1,0,1)
  if (u<rlgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd)) {

x[i]=rtrunc(1,meanlog=milgpd(alphalgpd,thetalgpd,lambdalgpd,sigmalgpd),sdlog=sigmalgpd,"lnorm",
a=0,b=thetalgpd) }
  else {
    x[i]=rpareto2(1,min=thetalgpd,shape=alphalgpd,scale=lambdalgpd+thetalgpd) }
}
return(x)
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο προσομοιώνουμε δείγμα από την συγκολλημένη κατανομή GGPD με την μέθοδο της διακριτής μίξης που αναφέραμε στην ενότητα 7.1.](#)

```

simggpd=function(n,betaggpd,gammaggpd,deltaggpd,thetaggpd) {
x=numeric(n)
for (i in 1:n) {
  u=runif(1,0,1)
  if (u<rggpd(betaggpd,gammaggpd,deltaggpd,thetaggpd)) {
    x[i]=rtrunc(1,shape=deltaggpd,scale=1/betaggpd,"gamma",a=0,b=thetaggpd) }
  else {

x[i]=rpareto2(1,min=thetaggpd,shape=alphaggpd(betaggpd,gammaggpd,deltaggpd,thetaggpd),scale
=gammaggpd+thetaggpd) }
}
return(x)
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο προσομοιώνουμε δείγμα από την συγκολλημένη κατανομή WGPD με την μέθοδο της διακριτής μίξης που αναφέραμε στην ενότητα 7.1.](#)

```

simwgpd=function(n,alphawgpd,thetawgpd,lambdawgpd,tauwgpd) {
x=numeric(n)
for (i in 1:n) {
  u=runif(1,0,1)
  if (u<rwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd)) {

x[i]=rtrunc(1,shape=tauwgpd,scale=phiwgpd(alphawgpd,thetawgpd,lambdawgpd,tauwgpd),"weibull",a
=0,b=thetawgpd) }
  else {
    x[i]=rpareto2(1,shape=alphawgpd,min=thetawgpd,scale=lambdawgpd+thetawgpd) }
}
return(x)
}

```

- Κατά την εφαρμογή της συνάρτησης “fitdistr” απαιτείται να εισάγουμε κάποιες αρχικές τιμές στον αλγόριθμο. Σε αρκετές από τις περιπτώσεις που μελετήσαμε στην παραπάνω διπλωματική εργασία η επιλογή των αρχικών τιμών (start values) δεν αποτελούσε απλή διαδικασία διότι ο αλγόριθμος μας οδηγούσε σε σφάλμα. Παρακάτω παρουσιάζεται ο αλγόριθμος “SP” μέσω του οποίου πραγματοποιούνται πολλαπλές επαναλήψεις της συνάρτησης “fitdistr” με διαφορετικές επιλογές αρχικών τιμών δίχως να απαιτείται να εκτελέσουμε την διαδικασία χειροκίνητα. Ο αλγόριθμος είναι φτιαγμένος ώστε να προσαρμόζεται στην κατάλληλη συνάρτηση πυκνότητας “F” και συνάρτηση κατανομής “F” που θα ορίσει ο χρήστης π.χ. αν θέλουμε να εφαρμόσουμε την κατανομή “LP” θα πρέπει, εσωτερικά του αλγορίθμου, όπου “F” να θέσουμε “dlp” και όπου “F” να θέσουμε “rlp”. Ως input ο αλγόριθμος απαιτεί το σύνολο δεδομένων “x” και ως outputs λαμβάνουμε τα εξής:

1. Εκτιμητές μέγιστης πιθανοφάνειας (E.M.Π.)
2. Κριτήριο AIC
3. Κριτήριο BIC
4. Στατιστικό K – S και p – value του ελέγχου
5. Στατιστικό A – D και p – value του ελέγχου
6. Αρχικές τιμές (start values) του αλγορίθμου

```

SP_f=function(x) {
fit=function(i,j,k) {fitdistr(x,f,list(alpha=p=i,theta=p=j,sigma=p=k),method="Nelder-Mead")}
i=1
j=1
k=1
for (m in 1:5000) {
  if(is.error(try(fit(i+0.001,j,k),silent=T))==FALSE) {
    MLE=fit(i+0.001,j,k)
    A=AIC(MLE)
    B=BIC(MLE)
    n=c(i+0.001,j,k)
    break }
  else if (is.error(try(fit(i,j+0.001,k),silent=T))==FALSE) {
    MLE=fit(i,j+0.001,k)
    A=AIC(MLE)
    B=BIC(MLE)
    n=c(i,j+0.001,k)
    break }
  else if (is.error(try(fit(i,j,k+0.001),silent=T))==FALSE) {
    MLE=fit(i,j,k+0.001)
    A=AIC(MLE)
    B=BIC(MLE)
    n=c(i,j,k+0.001)
    break }
  else {
    i=i+0.001
    j=j+0.001
    k=k+0.001 } }

```



```

ks=ks.test(x,F,MLE$estimate[1],MLE$estimate[2],MLE$estimate[3])
ad=ad.test(x,F,MLE$estimate[1],MLE$estimate[2],MLE$estimate[3])
return(list("MLE"=MLE,"AIC"=A,"BIC"=B,c(ks[1],ks[2]),c(ad[1],ad[2]),"Start values"=n))
}

```

- [Παρακάτω παρουσιάζονται οι κώδικες για τον υπολογισμό των VaR – TVaR της κατανομής LP.](#)

```

LP_TVAR=function(alpha,theta,sigma,p) {
v=LP_VAR(alpha,theta,sigma,p)
if (v>theta) {
A=(1-rlp(alpha,sigma))*alpha*theta^alpha/((1-p)^(alpha-1))
tvar=A*v^(1-alpha) }
else {
A=(1-rlp(alpha,sigma))*alpha*theta/((1-p)^(alpha-1))
B=rlp(alpha,sigma)*exp(milp(alpha,theta,sigma)+sigma^2/2)/((1-p)*pnorm((log(theta)-
milp(alpha,theta,sigma))/sigma))
C=pnorm((log(theta)-milp(alpha,theta,sigma)-sigma^2)/sigma)-pnorm((log(v)-milp(alpha,theta,sigma)-
sigma^2)/sigma)
tvar=B*C+A}
return(tvar)
}

```

```

LP_VAR=function(alpha,theta,sigma,p) {
if(p<rlp(alpha,sigma)) {
A=p*pnorm((log(theta)-milp(alpha,theta,sigma))/sigma)/rlp(alpha,sigma)
var=exp(milp(alpha,theta,sigma)+sigma*qnorm(A,mean=0,sd=1))
}
else {
A=((1-p)/(1-rlp(alpha,sigma)))^(-1/alpha)
var=theta*A
}
return(var)
}

```

- [Παρακάτω παρουσιάζονται οι κώδικες για τον υπολογισμό των VaR – TVaR της κατανομής LGPD.](#)

```

LGPD_VAR=function(alpha,theta,lambda,sigma,p) {
if(p<rlgpd(alpha,theta,lambda,sigma)) {
A=p*pnorm((log(theta)-milgpd(alpha,theta,lambda,sigma))/sigma)/rlgpd(alpha,theta,lambda,sigma)
var=exp(milgpd(alpha,theta,lambda,sigma)+sigma*qnorm(A,mean=0,sd=1))
}
else {
A=((1-p)/(1-rlgpd(alpha,theta,lambda,sigma)))^(-1/alpha)
var=(lambda+theta)*A-lambda
}
}

```

```
return(var)
}
```

```
LGPD_TVARE=function(alpha,theta,lambda,sigma,p) {
v=LGPD_VAR(alpha,theta,lambda,sigma,p)
if (v>theta) {
A=((v+lambda)^(1-alpha))/(alpha-1)-lambda*(v+lambda)^(-alpha)/alpha
tvar=alpha*(lambda+theta)^alpha*(1-rlgpd(alpha,theta,lambda,sigma))*A/(1-p)
}
else {
A=((lambda+theta)^(1-alpha))/(alpha-1)-lambda*(lambda+theta)^(-
alpha)/alpha*alpha*(lambda+theta)^alpha*(1-rlgpd(alpha,theta,lambda,sigma))/(1-p)
B=rlgpd(alpha,theta,lambda,sigma)*exp(milgpd(alpha,theta,lambda,sigma)+sigma^2/2)/((1-
p)*pnorm((log(theta)-milgpd(alpha,theta,lambda,sigma))/sigma))
C=pnorm((log(theta)-milgpd(alpha,theta,lambda,sigma)-sigma^2)/sigma)-pnorm((log(v)-
milgpd(alpha,theta,lambda,sigma)-sigma^2)/sigma)
tvar=B*C+A}
return(tvar)
}
```

- [Παρακάτω παρουσιάζονται οι κώδικες για τον υπολογισμό των VaR – TVaR της κατανομής GGPD.](#)

```
GGPD_TVARE=function(beta,gamma,delta,theta,p) {
v=GGPD_VAR(beta,gamma,delta,theta,p)
if (v>theta) {
A=(v+gamma)^(1-alphagpd(beta,gamma,delta,theta))/(alphagpd(beta,gamma,delta,theta)-1)-
gamma*(v+gamma)^(-alphagpd(beta,gamma,delta,theta))/alphagpd(beta,gamma,delta,theta)
B=alphagpd(beta,gamma,delta,theta)*(gamma+theta)^alphagpd(beta,gamma,delta,theta)*(1-
rggpd(beta,gamma,delta,theta))/(1-p)
tvar=A*B
}
else {
A=(theta+gamma)^(1-alphagpd(beta,gamma,delta,theta))/(alphagpd(beta,gamma,delta,theta)-1)-
gamma*(theta+gamma)^(-alphagpd(beta,gamma,delta,theta))/alphagpd(beta,gamma,delta,theta)
B=alphagpd(beta,gamma,delta,theta)*(gamma+theta)^alphagpd(beta,gamma,delta,theta)*(1-
rggpd(beta,gamma,delta,tau))/(1-p)
C=gammainc(beta*theta,delta+1)[[1]]-gammainc(beta*v,delta+1)[[1]]
D=beta*(1-p)*gammainc(beta*theta,delta)
tvar=A*B+C*D^(-1)*rggpd(beta,gamma,delta,theta)
}
return(tvar)
}
```

```
GGPD_VAR=function(beta,gamma,delta,theta,p) {
if (p<rggpd(beta,gamma,delta,theta)) {
var=uniroot.all(function(x) {gammainc(beta*x,delta)[[1]]-
p*gammainc(beta*theta,delta)[[1]]/rggpd(beta,gamma,delta,theta)},c(-100,100))
```

```

}
else {
A=((1-p)/(1-rggpd(beta,gamma,delta,theta)))^(-1/alphaggpd(beta,gamma,delta,theta))
var=(gamma+theta)*A-gamma
}
return(var)
}

```

- [Παρακάτω παρουσιάζονται οι κώδικες για τον υπολογισμό των VaR – TVaR της κατανομής WGPD.](#)

```

WGPD_TVAR=function(alpha,theta,lambda,tau,p) {
v=WGPD_VAR(alpha,theta,lambda,tau,p)
if (v>theta) {
A=(v+lambda)^(1-alpha)/(alpha-1)-lambda*(v+lambda)^(-alpha)/alpha
B=alpha*(lambda+theta)^alpha*(1-rwgpd(alpha,theta,lambda,tau))/(1-p)
tvar=A*B
}
else {
A=(theta+lambda)^(1-alpha)/(alpha-1)-lambda*(theta+lambda)^(-alpha)/alpha
B=alpha*(lambda+theta)^alpha*(1-rwgpd(alpha,theta,lambda,tau))/(1-p)
C=gammainc((theta/phiwgpd(alpha,theta,lambda,tau))^tau,1/tau+1)[[1]]-
gammainc((v/phiwgpd(alpha,theta,lambda,tau))^tau,1/tau+1)[[1]]
D=(1-p)*(1-exp(-(theta/phiwgpd(alpha,theta,lambda,tau))^tau))
tvar=B*A+C*D^(-1)*rwgpd(alpha,theta,lambda,tau)*phiwgpd(alpha,theta,lambda,tau)
}
return(tvar)
}

```

```

WGPD_VAR=function(alpha,theta,lambda,tau,p) {
if (p<rwgpd(alpha,theta,lambda,tau)) {
A=1-exp(-(theta/phiwgpd(alpha,theta,lambda,tau))^tau)
var=phiwgpd(alpha,theta,lambda,tau)*((-log(1-p*A/rwgpd(alpha,theta,lambda,tau)))^(1/tau))
}
else {
A=((1-p)/(1-rwgpd(alpha,theta,lambda,tau)))^(-1/alpha)
var=(lambda+theta)*A-lambda
}
return(var)
}

```

- [Παρακάτω παρουσιάζεται ο κώδικας για την μελέτη προσομοίωσης για την κατανομή LP που πραγματοποιήθηκε στην ενότητα 7.2.](#)

```

SIMULATION_LP=function(alpha,theta,sigma) {
size=c(100,400,700,1000,1500)
A=c()
B=c()
C=c()

```

```

bias_alpha=c()
abs_bias_alpha=c()
mse_alpha=c()
bias_theta=c()
abs_bias_theta=c()
mse_theta=c()
bias_sigma=c()
abs_bias_sigma=c()
mse_sigma=c()

for (i in 1:length(size)) {
for (j in 1:1000) {
  sample=simlp(size[i],alpha,theta,sigma)
  SP=SP_LP(sample)
  A[j]=alpha-SP[[1]]$estimate[1]
  B[j]=theta-SP[[1]]$estimate[2]
  C[j]=sigma-SP[[1]]$estimate[3]
}
bias_alpha[i]=mean(A)
abs_bias_alpha[i]=mean(abs(A))
mse_alpha[i]=mean(A^2)
bias_theta[i]=mean(B)
abs_bias_theta[i]=mean(abs(B))
mse_theta[i]=mean(B^2)
bias_sigma[i]=mean(C)
abs_bias_sigma[i]=mean(abs(C))
mse_sigma[i]=mean(C^2)
}
return(list("sample size"=size,"Bias for  $\alpha$ "=bias_alpha,"Bias for  $\theta$ "=bias_theta,"Bias for  $\sigma$ "=bias_sigma,"Abs.bias for  $\alpha$ "=abs_bias_alpha,"Abs.bias for  $\theta$ "=abs_bias_theta,"Abs.bias for  $\sigma$ "=abs_bias_sigma,"MSE for  $\alpha$ "=mse_alpha,"MSE for  $\theta$ "=mse_theta,"MSE for  $\sigma$ "=mse_sigma,"Square Root MSE of  $\alpha$ "=sqrt(mse_alpha),"Square Root MSE of  $\theta$ "=sqrt(mse_theta),"Square Root MSE of  $\sigma$ "=sqrt(mse_sigma)))
}

SLP=SIMULATION_LP(4,5,0.5)
SLP

par(mfrow=c(2,2))
plot(SLP[[1]],SLP[[2]],type="b",col="red",lwd=2,xlab="n",ylab="BIAS",main="BIAS vs n",ylim=c(-4,3))
lines(SLP[[1]],SLP[[3]],type="b",col="blue",lwd=2)
lines(SLP[[1]],SLP[[4]],type="b",col="green",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ ", " $\theta$ ", " $\sigma$ "),col=c("red", "blue", "green"),lwd=2)

plot(SLP[[1]],SLP[[5]],type="b",col="red",lwd=2,xlab="n",ylab="ABS BIAS",main="ABS BIAS vs n",ylim=c(0,5))
lines(SLP[[1]],SLP[[6]],type="b",col="blue",lwd=2)
lines(SLP[[1]],SLP[[7]],type="b",col="green",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ ", " $\theta$ ", " $\sigma$ "),col=c("red", "blue", "green"),lwd=2)

```

```

plot(SLP[[1]],SLP[[8]],type="b",col="red",lwd=2,xlab="n",ylab="MSE",main="MSE vs n",ylim=c(0,30))
lines(SLP[[1]],SLP[[9]],type="b",col="blue",lwd=2)
lines(SLP[[1]],SLP[[10]],type="b",col="green",lwd=2)
legend("topright",bty="n",legend=c("α","θ","σ"),col=c("red","blue","green"),lwd=2)

```

```

plot(SLP[[1]],SLP[[11]],type="b",col="red",lwd=2,xlab="n",ylab=" ROOT MSE",main="ROOT MSE vs
n",ylim=c(0,8))
lines(SLP[[1]],SLP[[12]],type="b",col="blue",lwd=2)
lines(SLP[[1]],SLP[[13]],type="b",col="green",lwd=2)
legend("topright",bty="n",legend=c("α","θ","σ"),col=c("red","blue","green"),lwd=2)

```

- [Παρακάτω παρουσιάζεται ο κώδικας για την μελέτη προσομοίωσης για την κατανομή LGPD που πραγματοποιήθηκε στην ενότητα 7.3.](#)

```

SIMULATION_LGPD=function(alpha,theta,lambda,sigma) {
size=c(100,400,700,1000,1500)
A=c()
B=c()
C=c()
D=c()
bias_alpha=c()
abs_bias_alpha=c()
mse_alpha=c()
bias_theta=c()
abs_bias_theta=c()
mse_theta=c()
bias_lambda=c()
abs_bias_lambda=c()
mse_lambda=c()
bias_sigma=c()
abs_bias_sigma=c()
mse_sigma=c()

for (i in 1:length(size)) {
for (j in 1:1000) {
sample=simlgpd(size[i],alpha,theta,lambda,sigma)
SP=SP_LGPD(sample)
A[j]=alpha-SP[[1]]$estimate[1]
B[j]=theta-SP[[1]]$estimate[2]
C[j]=lambda-SP[[1]]$estimate[3]
D[j]=sigma-SP[[1]]$estimate[4]
}
bias_alpha[i]=mean(A)
abs_bias_alpha[i]=mean(abs(A))
mse_alpha[i]=mean(A^2)
bias_theta[i]=mean(B)
abs_bias_theta[i]=mean(abs(B))
mse_theta[i]=mean(B^2)
bias_lambda[i]=mean(C)

```

```

abs_bias_lambda[i]=mean(abs(C))
mse_lambda[i]=mean(C^2)
bias_sigma[i]=mean(D)
abs_bias_sigma[i]=mean(abs(D))
mse_sigma[i]=mean(D^2)
}
return(list("sample size"=size,"Bias for  $\alpha$ "=bias_alpha,"Bias for  $\theta$ "=bias_theta,"Bias for
 $\lambda$ "=bias_lambda,"Bias for  $\sigma$ "=bias_sigma,"Abs.bias for  $\alpha$ "=abs_bias_alpha,"Abs.bias for
 $\theta$ "=abs_bias_theta,"Abs.bias for  $\lambda$ "=abs_bias_lambda,"Abs.bias for  $\sigma$ "=abs_bias_sigma,"MSE for
 $\alpha$ "=mse_alpha,"MSE for  $\theta$ "=mse_theta,"MSE for  $\lambda$ "=mse_lambda,"MSE for  $\sigma$ "=mse_sigma,"Square Root
MSE of  $\alpha$ "=sqrt(mse_alpha),"Square Root MSE of  $\theta$ "=sqrt(mse_theta),"Square Root MSE of
 $\lambda$ "=sqrt(mse_lambda),"Square Root MSE of  $\sigma$ "=sqrt(mse_sigma)))
}

```

```

SLGPD=SIMULATION_LGPD(3,2,1,0.5)
SLGPD

```

```

par(mfrow=c(2,2))
plot(SLGPD[[1]],SLGPD[[2]],type="b",col="red",lwd=2,xlab="n",ylab="BIAS",main="BIAS vs n",ylim=c(-
4,5))

```

```

lines(SLGPD[[1]],SLGPD[[3]],type="b",col="blue",lwd=2)
lines(SLGPD[[1]],SLGPD[[4]],type="b",col="green",lwd=2)
lines(SLGPD[[1]],SLGPD[[5]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ "," $\theta$ "," $\lambda$ "," $\sigma$ "),col=c("red","blue","green","orange"),lwd=2)

```

```

plot(SLGPD[[1]],SLGPD[[6]],type="b",col="red",lwd=2,xlab="n",ylab="ABS BIAS",main="ABS BIAS vs
n",ylim=c(0,5))

```

```

lines(SLGPD[[1]],SLGPD[[7]],type="b",col="blue",lwd=2)
lines(SLGPD[[1]],SLGPD[[8]],type="b",col="green",lwd=2)
lines(SLGPD[[1]],SLGPD[[9]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ "," $\theta$ "," $\lambda$ "," $\sigma$ "),col=c("red","blue","green","orange"),lwd=2)

```

```

plot(SLGPD[[1]],SLGPD[[10]],type="b",col="red",lwd=2,xlab="n",ylab="MSE",main="MSE vs
n",ylim=c(0,30))

```

```

lines(SLGPD[[1]],SLGPD[[11]],type="b",col="blue",lwd=2)
lines(SLGPD[[1]],SLGPD[[12]],type="b",col="green",lwd=2)
lines(SLGPD[[1]],SLGPD[[13]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ "," $\theta$ "," $\lambda$ "," $\sigma$ "),col=c("red","blue","green","orange"),lwd=2)

```

```

plot(SLGPD[[1]],SLGPD[[14]],type="b",col="red",lwd=2,xlab="n",ylab=" ROOT MSE",main="ROOT MSE
vs n",ylim=c(0,8))

```

```

lines(SLGPD[[1]],SLGPD[[15]],type="b",col="blue",lwd=2)
lines(SLGPD[[1]],SLGPD[[16]],type="b",col="green",lwd=2)
lines(SLGPD[[1]],SLGPD[[17]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ "," $\theta$ "," $\lambda$ "," $\sigma$ "),col=c("red","blue","green","orange"),lwd=2)

```

- [Παρακάτω παρουσιάζεται ο κώδικας για την μελέτη προσομοίωσης για την κατανομή GGPD που πραγματοποιήθηκε στην ενότητα 7.4.](#)

```

SIMULATION_GGPD=function(beta,gamma,delta,theta) {
size=c(100,400,700,1000,1500)
A=c()
B=c()
C=c()
D=c()
bias_beta=c()
abs_bias_beta=c()
mse_beta=c()
bias_gamma=c()
abs_bias_gamma=c()
mse_gamma=c()
bias_delta=c()
abs_bias_delta=c()
mse_delta=c()
bias_theta=c()
abs_bias_theta=c()
mse_theta=c()

for (i in 1:length(size)) {
for (j in 1:1000) {
sample=simggpd(size[i],beta,gamma,delta,theta)
SP=SP_GGPD(sample)
A[j]=beta-SP[[1]]$estimate[1]
B[j]=gamma-SP[[1]]$estimate[2]
C[j]=delta-SP[[1]]$estimate[3]
D[j]=theta-SP[[1]]$estimate[4]
}
bias_beta[i]=mean(A)
abs_bias_beta[i]=mean(abs(A))
mse_beta[i]=mean(A^2)
bias_gamma[i]=mean(B)
abs_bias_gamma[i]=mean(abs(B))
mse_gamma[i]=mean(B^2)
bias_delta[i]=mean(C)
abs_bias_delta[i]=mean(abs(C))
mse_delta[i]=mean(C^2)
bias_theta[i]=mean(D)
abs_bias_theta[i]=mean(abs(D))
mse_theta[i]=mean(D^2)
}
return(list("sample size"=size,"Bias for  $\beta$ "=bias_beta,"Bias for  $\gamma$ "=bias_gamma,"Bias for  $\delta$ "=bias_delta,"Bias for  $\theta$ "=bias_theta,"Abs.bias for  $\beta$ "=abs_bias_beta,"Abs.bias for  $\gamma$ "=abs_bias_gamma,"Abs.bias for  $\delta$ "=abs_bias_delta,"Abs.bias for  $\theta$ "=abs_bias_theta,"MSE for  $\beta$ "=mse_beta,"MSE for  $\gamma$ "=mse_gamma,"MSE for  $\delta$ "=mse_delta,"MSE for  $\theta$ "=mse_theta,"Square Root MSE of  $\beta$ "=sqrt(mse_beta),"Square Root MSE of  $\gamma$ "=sqrt(mse_gamma),"Square Root MSE of  $\delta$ "=sqrt(mse_delta),"Square Root MSE of  $\theta$ "=sqrt(mse_theta)))
}

```

```
SGGPD=SIMULATION_GGPD(5,2,3,1)
SGGPD
```

```
par(mfrow=c(2,2))
plot(SGGPD[[1]],SGGPD[[2]],type="b",col="red",lwd=2,xlab="n",ylab="BIAS",main="BIAS vs n",ylim=c(-4,5))
lines(SGGPD[[1]],SGGPD[[3]],type="b",col="blue",lwd=2)
lines(SGGPD[[1]],SGGPD[[4]],type="b",col="green",lwd=2)
lines(SGGPD[[1]],SGGPD[[5]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c("β", "γ", "δ", "θ"),col=c("red", "blue", "green", "orange"),lwd=2)
```

```
plot(SGGPD[[1]],SGGPD[[6]],type="b",col="red",lwd=2,xlab="n",ylab="ABS BIAS",main="ABS BIAS vs n",ylim=c(0,5))
lines(SGGPD[[1]],SGGPD[[7]],type="b",col="blue",lwd=2)
lines(SGGPD[[1]],SGGPD[[8]],type="b",col="green",lwd=2)
lines(SGGPD[[1]],SGGPD[[9]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c("β", "γ", "δ", "θ"),col=c("red", "blue", "green", "orange"),lwd=2)
```

```
plot(SGGPD[[1]],SGGPD[[10]],type="b",col="red",lwd=2,xlab="n",ylab="MSE",main="MSE vs n",ylim=c(0,30))
lines(SGGPD[[1]],SGGPD[[11]],type="b",col="blue",lwd=2)
lines(SGGPD[[1]],SGGPD[[12]],type="b",col="green",lwd=2)
lines(SGGPD[[1]],SGGPD[[13]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c("β", "γ", "δ", "θ"),col=c("red", "blue", "green", "orange"),lwd=2)
```

```
plot(SGGPD[[1]],SGGPD[[14]],type="b",col="red",lwd=2,xlab="n",ylab="ROOT MSE",main="ROOT MSE vs n",ylim=c(0,8))
lines(SGGPD[[1]],SGGPD[[15]],type="b",col="blue",lwd=2)
lines(SGGPD[[1]],SGGPD[[16]],type="b",col="green",lwd=2)
lines(SGGPD[[1]],SGGPD[[17]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c("β", "γ", "δ", "θ"),col=c("red", "blue", "green", "orange"),lwd=2)
```

- [Παρακάτω παρουσιάζεται ο κώδικας για την μελέτη προσομοίωσης για την κατανομή WGPD που πραγματοποιήθηκε στην ενότητα 7.5.](#)

```
SIMULATION_WGPD=function(alpha,theta,lambda,tau) {
size=c(100,400,700,1000,1500)
A=c()
B=c()
C=c()
D=c()
bias_alpha=c()
abs_bias_alpha=c()
mse_alpha=c()
bias_theta=c()
abs_bias_theta=c()
mse_theta=c()
bias_lambda=c()
abs_bias_lambda=c()
mse_lambda=c()
}
```



```

bias_tau=c()
abs_bias_tau=c()
mse_tau=c()

for (i in 1:length(size)) {
for (j in 1:1000) {
    sample=simwgpd(size[i],alpha,theta,lambda,tau)
    SP=SP_WGPD(sample)
    A[j]=alpha-SP[[1]]$estimate[1]
    B[j]=theta-SP[[1]]$estimate[2]
    C[j]=lambda-SP[[1]]$estimate[3]
    D[j]=tau-SP[[1]]$estimate[4]
}
bias_alpha[i]=mean(A)
abs_bias_alpha[i]=mean(abs(A))
mse_alpha[i]=mean(A^2)
bias_theta[i]=mean(B)
abs_bias_theta[i]=mean(abs(B))
mse_theta[i]=mean(B^2)
bias_lambda[i]=mean(C)
abs_bias_lambda[i]=mean(abs(C))
mse_lambda[i]=mean(C^2)
bias_tau[i]=mean(D)
abs_bias_tau[i]=mean(abs(D))
mse_tau[i]=mean(D^2)
}
return(list("sample size"=size,"Bias for  $\alpha$ "=bias_alpha,"Bias for  $\theta$ "=bias_theta,"Bias for
 $\lambda$ "=bias_lambda,"Bias for  $\tau$ "=bias_tau,"Abs.bias for  $\alpha$ "=abs_bias_alpha,"Abs.bias for
 $\theta$ "=abs_bias_theta,"Abs.bias for  $\lambda$ "=abs_bias_lambda,"Abs.bias for  $\tau$ "=abs_bias_tau,"MSE for
 $\alpha$ "=mse_alpha,"MSE for  $\theta$ "=mse_theta,"MSE for  $\lambda$ "=mse_lambda,"MSE for  $\tau$ "=mse_tau,"Square Root
MSE of  $\alpha$ "=sqrt(mse_alpha),"Square Root MSE of  $\theta$ "=sqrt(mse_theta),"Square Root MSE of
 $\lambda$ "=sqrt(mse_lambda),"Square Root MSE of  $\tau$ "=sqrt(mse_tau)))
}

SWGPD=SIMULATION_WGPD(3,4,0.5,2)
SWGPD

par(mfrow=c(2,2))
plot(SWGPD[[1]],SWGPD[[2]],type="b",col="red",lwd=2,xlab="n",ylab="BIAS",main="BIAS vs
n",ylim=c(-4,3))
lines(SWGPD[[1]],SWGPD[[3]],type="b",col="blue",lwd=2)
lines(SWGPD[[1]],SWGPD[[4]],type="b",col="green",lwd=2)
lines(SWGPD[[1]],SWGPD[[5]],type="b",col="orange",lwd=2)
legend("topright",bty="n",legend=c(" $\alpha$ "," $\theta$ "," $\lambda$ "," $\tau$ "),col=c("red","blue","green","orange"),lwd=2)

plot(SWGPD[[1]],SWGPD[[6]],type="b",col="red",lwd=2,xlab="n",ylab="ABS BIAS",main="ABS BIAS vs
n",ylim=c(0,5))
lines(SWGPD[[1]],SWGPD[[7]],type="b",col="blue",lwd=2)
lines(SWGPD[[1]],SWGPD[[8]],type="b",col="green",lwd=2)
lines(SWGPD[[1]],SWGPD[[9]],type="b",col="orange",lwd=2)

```

```
legend("topright",bty="n",legend=c("α","θ","λ","τ"),col=c("red","blue","green","orange"),lwd=2)
```

```
plot(SWGPDP[[1]],SWGPDP[[10]],type="b",col="red",lwd=2,xlab="n",ylab="MSE",main="MSE vs n",ylim=c(0,30))
```

```
lines(SWGPDP[[1]],SWGPDP[[11]],type="b",col="blue",lwd=2)
```

```
lines(SWGPDP[[1]],SWGPDP[[12]],type="b",col="green",lwd=2)
```

```
lines(SWGPDP[[1]],SWGPDP[[13]],type="b",col="orange",lwd=2)
```

```
legend("topright",bty="n",legend=c("α","θ","λ","τ"),col=c("red","blue","green","orange"),lwd=2)
```

```
plot(SWGPDP[[1]],SWGPDP[[14]],type="b",col="red",lwd=2,xlab="n",ylab="ROOT MSE",main="ROOT MSE vs n",ylim=c(0,8))
```

```
lines(SWGPDP[[1]],SWGPDP[[15]],type="b",col="blue",lwd=2)
```

```
lines(SWGPDP[[1]],SWGPDP[[16]],type="b",col="green",lwd=2)
```

```
lines(SWGPDP[[1]],SWGPDP[[17]],type="b",col="orange",lwd=2)
```

```
legend("topright",bty="n",legend=c("α","θ","λ","τ"),col=c("red","blue","green","orange"),lwd=2)
```

- Παρακάτω παρουσιάζεται ο κώδικας “ANALYSIS” που εφαρμόστηκε για την ανάλυση των δεδομένων στο κεφάλαιο 8. Το input που απαιτεί είναι το σύνολο δεδομένων “x” και στην συνέχεια προσαρμόζει στα δεδομένα τις κατανομές LP, LGPD, GGPDP, WGPDP, Gamma, Weibull, Lognormal, Pareto. Στην συνέχεια τα outputs που λαμβάνουμε είναι τα παρακάτω:

1. Περιγραφικά μέτρα του συνόλου των δεδομένων
2. Εκτιμητές μέγιστης πιθανοφάνειας των κατανομών
3. Εμπειρικά και θεωρητικά VaR
4. Εμπειρικά και θεωρητικά TVaR (όπου υπολογίζονται)
5. AIC
6. BIC
7. Έλεγχος υποθέσεων K – S (τιμή στατιστικού και p – value)
8. Έλεγχος υποθέσεων A – D (τιμή στατιστικού και p – value)

```
ANALYSIS=function(x) {
options(scipen=100)
p=c(0.8,0.85,0.9,0.95,0.975,0.99,0.995)
sample_var=c()
LP=SP_LP(x)
LGPD=SP_LGPD(x)
GGPD=SP_GGPD(x)
WGPDP=SP_WGPDP(x)
GAMMA=SP_GAMMA(x)
WEIBULL=SP_WEIBULL(x)
LOGNORMAL=SP_LNORM(x)
PARETO=epareto(as.numeric(x),method="mle")
PARETO_KS=ks.test(x,ppareto1,shape=PARETO$parameters[[2]],min=PARETO$parameters[[1]])
PARETO_AD=ad.test(x,ppareto1,shape=PARETO$parameters[[2]],min=PARETO$parameters[[1]])
PARETOAIC=-
2*(length(x)*log(PARETO$parameters[2])+length(x)*PARETO$parameters[2]*log(PARETO$parameters[1
]))-(PARETO$parameters[2]+1)*sum(log(x))+4
```

```

PARETOBIC=-
2*(length(x)*log(PARETO$parameters[2])+length(x)*PARETO$parameters[2]*log(PARETO$parameters[1
]))-(PARETO$parameters[2]+1)*sum(log(x))+2*log(length(x))
varLP=c()
varLGPD=c()
varGGPD=c()
varWGPD=c()
varGAMMA=c()
varWEIBULL=c()
varLOGNORMAL=c()
varPARETO=c()
samplevar=c()
TvarLP=c()
TvarLGPD=c()
TvarGGPD=c()
TvarWGPD=c()
TvarGAMMA=c()
TvarWEIBULL=c()
TvarLOGNORMAL=c()
TvarPARETO=c()
for(i in 1:length(p)) {
sample_var[i]=quantile(x,p[i])
varLP[i]=LP_VAR(LP[[1]]$estimate[[1]],LP[[1]]$estimate[[2]],LP[[1]]$estimate[[3]],p[i])
varLGPD[i]=LGPD_VAR(LGPD[[1]]$estimate[[1]],LGPD[[1]]$estimate[[2]],LGPD[[1]]$estimate[[3]],LGPD[[1
]]$estimate[[4]],p[i])
varGGPD[i]=GGPD_VAR(GGPD[[1]]$estimate[[1]],GGPD[[1]]$estimate[[2]],GGPD[[1]]$estimate[[3]],GGPD
[[1]]$estimate[[4]],p[i])
varWGPD[i]=WGPD_VAR(WGPD[[1]]$estimate[[1]],WGPD[[1]]$estimate[[2]],WGPD[[1]]$estimate[[3]],W
GPD[[1]]$estimate[[4]],p[i])
varGAMMA[i]=qgamma(p[i],shape=GAMMA[[1]]$estimate[[1]],scale=GAMMA[[1]]$estimate[[2]])
varWEIBULL[i]=qweibull(p[i],shape=WEIBULL[[1]]$estimate[[1]],scale=WEIBULL[[1]]$estimate[[2]])
varLOGNORMAL[i]=qlnorm(p[i],meanlog=LOGNORMAL[[1]]$estimate[[1]],sdlog=LOGNORMAL[[1]]$est
imate[[2]])
varPARETO[i]=qpareto1(p[i],shape=PARETO$parameters[[2]],min=PARETO$parameters[[1]])
samplevar[i]=quantile(x,p[i])
if(LP[[1]]$estimate[[1]]>1) {
TvarLP[i]=LP_TVAR(LP[[1]]$estimate[[1]],LP[[1]]$estimate[[2]],LP[[1]]$estimate[[3]],p[i])}
else {TvarLP[i]="NA"}

if(LGPD[[1]]$estimate[[1]]>1) {
TvarLGPD[i]=LGPD_TVAR(LGPD[[1]]$estimate[[1]],LGPD[[1]]$estimate[[2]],LGPD[[1]]$estimate[[3]],LGPD[
[[1]]$estimate[[4]],p[i])}
else { TvarLGPD[i]="NA"}

if(alphaggpd(GGPD[[1]]$estimate[[1]],GGPD[[1]]$estimate[[2]],GGPD[[1]]$estimate[[3]],GGPD[[1]]$estim
ate[[4]])>1) {
TvarGGPD[i]=GGPD_TVAR(GGPD[[1]]$estimate[[1]],GGPD[[1]]$estimate[[2]],GGPD[[1]]$estimate[[3]],GG
PD[[1]]$estimate[[4]],p[i])}
else {TvarGGPD[i]="NA"}

```

```

if(WGPD[[1]]$estimate[[1]] > 1) {
TvarWGPD[i]=WGPD_TVAR(WGPD[[1]]$estimate[[1]],WGPD[[1]]$estimate[[2]],WGPD[[1]]$estimate[[3]],
WGPD[[1]]$estimate[[4]],p[i])
else {TvarWGPD[i]="NA"}

TvarGAMMA[i]=integrate(function(x)
{x*dgamma(x,shape=GAMMA[[1]]$estimate[[1]],scale=GAMMA[[1]]$estimate[[2]]},lower=varGAMMA[i]
,upper=Inf,subdivisions=100000L)[[1]]/(1-p[i])
TvarWEIBULL[i]=integrate(function(x)
{x*dweibull(x,shape=WEIBULL[[1]]$estimate[[1]],scale=WEIBULL[[1]]$estimate[[2]]},lower=varWEIBULL[i]
,upper=Inf,subdivisions=100000L)[[1]]/(1-p[i])
TvarLOGNORMAL[i]=integrate(function(x)
{x*dlnorm(x,meanlog=LOGNORMAL[[1]]$estimate[[1]],sdlog=LOGNORMAL[[1]]$estimate[[2]]},lower=v
arLOGNORMAL[i],upper=Inf,subdivisions=100000L)[[1]]/(1-p[i])
if(PARETO$parameters[[2]] > 1) {
TvarPARETO[i]=integrate(function(x)
{x*dpareto1(x,shape=PARETO$parameters[[2]],min=PARETO$parameters[[1]]},lower=varPARETO[i],up
per=Inf,subdivisions=100000L)[[1]]/(1-p[i])
}
else {TvarPARETO[i]="NA" }
}

m1=matrix(c(mean(x),sd(x),var(x),median(x),max(x)-
min(x),max(x),min(x),skewness(as.numeric(x)),kurtosis(as.numeric(x))),nrow=9)
dimnames(m1)=list(c("mean","sd","variance","median","range","maximum","minimum","skewness","kurt
osis"),"Value")

m2=matrix(c(LP[[1]]$estimate[[1]],LP[[1]]$estimate[[2]],LP[[1]]$estimate[[3]]),nrow=3)
dimnames(m2)=list(c("α","θ","σ"),"MLE")

m3=matrix(c(LGPD[[1]]$estimate[[1]],LGPD[[1]]$estimate[[2]],LGPD[[1]]$estimate[[3]],LGPD[[1]]$estim
ate[[4]]),nrow=4)
dimnames(m3)=list(c("α","θ","λ","σ"),"MLE")

m4=matrix(c(GGPD[[1]]$estimate[[1]],GGPD[[1]]$estimate[[2]],GGPD[[1]]$estimate[[3]],GGPD[[1]]$estim
ate[[4]]),nrow=4)
dimnames(m4)=list(c("β","γ","δ","θ"),"MLE")

m5=matrix(c(WGPD[[1]]$estimate[[1]],WGPD[[1]]$estimate[[2]],WGPD[[1]]$estimate[[3]],WGPD[[1]]$esti
mate[[4]]),nrow=4)
dimnames(m5)=list(c("α","θ","λ","τ"),"MLE")

m6=matrix(c(GAMMA[[1]]$estimate[[1]],GAMMA[[1]]$estimate[[2]]),nrow=2)
dimnames(m6)=list(c("δ","β"),"MLE")

m7=matrix(c(WEIBULL[[1]]$estimate[[1]],WEIBULL[[1]]$estimate[[2]]),nrow=2)
dimnames(m7)=list(c("τ","φ"),"MLE")

m8=matrix(c(LOGNORMAL[[1]]$estimate[[1]],LOGNORMAL[[1]]$estimate[[2]]),nrow=2)
dimnames(m8)=list(c("μ","σ"),"MLE")

```

```
m9=matrix(c(PARETO$parameters[[2]],PARETO$parameters[[1]]),nrow=2)
dimnames(m9)=list(c("α","θ"),"MLE")
```

```
m10=matrix(c(LP[[2]],LGPD[[2]],GGPD[[2]],WGPD[[2]],GAMMA[[2]],WEIBULL[[2]],LOGNORMAL[[2]],PARETOAIC,LP[[3]],LGPD[[3]],GGPD[[3]],WGPD[[3]],GAMMA[[3]],WEIBULL[[3]],LOGNORMAL[[3]],PARETOBIC,LP[[4]]$statistic,LGPD[[4]]$statistic,GGPD[[4]]$statistic,WGPD[[4]]$statistic,GAMMA[[4]]$statistic,WEIBULL[[4]]$statistic,LOGNORMAL[[4]]$statistic,PARETO_KS[[1]],round(LP[[4]]$p.value,digits=5),round(LGPD[[4]]$p.value,digits=5),round(GGPD[[4]]$p.value,digits=5),round(WGPD[[4]]$p.value,digits=5),round(GAMMA[[4]]$p.value,digits=5),round(WEIBULL[[4]]$p.value,digits=5),round(LOGNORMAL[[4]]$p.value,digits=5),round(PARETO_KS[[2]],digits=5),LP[[5]]$statistic,LGPD[[5]]$statistic,GGPD[[5]]$statistic,WGPD[[5]]$statistic,GAMMA[[5]]$statistic,WEIBULL[[5]]$statistic,LOGNORMAL[[5]]$statistic,PARETO_AD[[1]],round(LP[[5]]$p.value,digits=5),round(LGPD[[5]]$p.value,digits=5),round(GGPD[[5]]$p.value,digits=5),round(WGPD[[5]]$p.value,digits=5),round(GAMMA[[5]]$p.value,digits=5),round(WEIBULL[[5]]$p.value,digits=5),round(LOGNORMAL[[5]]$p.value,digits=5),round(PARETO_AD[[2]],digits=5)),ncol=6)
dimnames(m10)=list(c("LP","LGPD","GGPD","WGPD","Gamma","Weibull","Lognormal","Pareto"),c("AIC","BIC","K-S Statistic","K-S P-value","A-D Statistic","A-D P-value"))
```

```
m11=matrix(c(varLP[1],varLGPD[1],varGGPD[1],varWGPD[1],varGAMMA[1],varWEIBULL[1],varLOGNORMAL[1],varPARETO[1],samplevar[1],varLP[2],varLGPD[2],varGGPD[2],varWGPD[2],varGAMMA[2],varWEIBULL[2],varLOGNORMAL[2],varPARETO[2],samplevar[2],varLP[3],varLGPD[3],varGGPD[3],varWGPD[3],varGAMMA[3],varWEIBULL[3],varLOGNORMAL[3],varPARETO[3],samplevar[3],varLP[4],varLGPD[4],varGGPD[4],varWGPD[4],varGAMMA[4],varWEIBULL[4],varLOGNORMAL[4],varPARETO[4],samplevar[4],varLP[5],varLGPD[5],varGGPD[5],varWGPD[5],varGAMMA[5],varWEIBULL[5],varLOGNORMAL[5],varPARETO[5],samplevar[5],varLP[6],varLGPD[6],varGGPD[6],varWGPD[6],varGAMMA[6],varWEIBULL[6],varLOGNORMAL[6],varPARETO[6],samplevar[6],varLP[7],varLGPD[7],varGGPD[7],varWGPD[7],varGAMMA[7],varWEIBULL[7],varLOGNORMAL[7],varPARETO[7],samplevar[7]),ncol=7)
dimnames(m11)=list(c("LP","LGPD","GGPD","WGPD","Gamma","Weibull","Lognormal","Pareto","Empirical"),c("80%","85%","90%","95%","97.5%","99%","99.5%"))
```

```
m12=matrix(as.numeric(c(TvarLP[1],TvarLGPD[1],TvarGGPD[1],TvarWGPD[1],TvarGAMMA[1],TvarWEIBULL[1],TvarLOGNORMAL[1],TvarPARETO[1],mean(x[x>quantile(x,p[1])]),TvarLP[2],TvarLGPD[2],TvarGGPD[2],TvarWGPD[2],TvarGAMMA[2],TvarWEIBULL[2],TvarLOGNORMAL[2],TvarPARETO[2],mean(x[x>quantile(x,p[2])]),TvarLP[3],TvarLGPD[3],TvarGGPD[3],TvarWGPD[3],TvarGAMMA[3],TvarWEIBULL[3],TvarLOGNORMAL[3],TvarPARETO[3],mean(x[x>quantile(x,p[3])]),TvarLP[4],TvarLGPD[4],TvarGGPD[4],TvarWGPD[4],TvarGAMMA[4],TvarWEIBULL[4],TvarLOGNORMAL[4],TvarPARETO[4],mean(x[x>quantile(x,p[4])]),TvarLP[5],TvarLGPD[5],TvarGGPD[5],TvarWGPD[5],TvarGAMMA[5],TvarWEIBULL[5],TvarLOGNORMAL[5],TvarPARETO[5],mean(x[x>quantile(x,p[5])]),TvarLP[6],TvarLGPD[6],TvarGGPD[6],TvarWGPD[6],TvarGAMMA[6],TvarWEIBULL[6],TvarLOGNORMAL[6],TvarPARETO[6],mean(x[x>quantile(x,p[6])]),TvarLP[7],TvarLGPD[7],TvarGGPD[7],TvarWGPD[7],TvarGAMMA[7],TvarWEIBULL[7],TvarLOGNORMAL[7],TvarPARETO[7],mean(x[x>quantile(x,p[7])])),ncol=7)
dimnames(m12)=list(c("LP","LGPD","GGPD","WGPD","Gamma","Weibull","Lognormal","Pareto","Empirical"),c("80%","85%","90%","95%","97.5%","99%","99.5%"))
```

```
return(list("Descriptive
statistics"=m1,"LP"=m2,"LGPD"=m3,"GGPD"=m4,"WGPD"=m5,"Gamma"=m6,"Weibull"=m7,"Lognormal"=m8,"Pareto"=m9,"Goodness of fit"=m10,"VaR"=m11,"TVaR"=m12))
}
```

- Παρακάτω παρουσιάζονται τα outputs του αλγορίθμου “ANALYSIS” για τα δεδομένα “beaonre” (συμβολ. “b”) που μελετήθηκαν στην ενότητα 8.2 (βλ. πίνακες 8.1, 8.2, 8.3).

```

>
>
> ANALYSIS(b)
$`Descriptive statistics`
              Value
mean          363.464
sd            4868.259
variance 23699945.073
median         9.073
range        190541.545
maximum      190541.698
minimum         0.153
skewness      33.931
kurtosis     1289.860

$LP
      MLE
α 0.4889
θ 7.1524
σ 0.9538

$LGPD
      MLE
α 0.578
θ 2.868
λ 2.528
σ 0.805

$GGPD
      MLE
β 0.5022
γ 1.4834
δ 2.0246
θ 4.3314

$WGPD
      MLE
α 0.5809
θ 1.8512
λ 2.6890
τ 2.6179

$Gamma
      MLE
δ 0.2206
β 1653.5687

$Weibull
      MLE
τ 0.4106
φ 39.4249

$Lognormal
      MLE
μ 2.606
σ 1.974

$Pareto
      MLE
α 0.223
θ 0.153

```

```

$`Goodness of fit`
      AIC   BIC K-S Statistic K-S P-value A-D Statistic A-D P-value
LP      16820 16837      0.04564 0.00101      3.4749 0.01582
LGPD    16789 16811      0.02961 0.08177      0.8132 0.47158
GGPD    16814 16836      0.03375 0.03146      1.9074 0.10333
WGPD    16790 16812      0.02744 0.12835      0.7380 0.52788
Gamma   19350 19361      0.28541 0.00000      Inf 0.00000
Weibull 17999 18010      0.16607 0.00000      96.2923 0.00000
Lognormal 17160 17171      0.08164 0.00000      29.7923 0.00000
Pareto  18622 18633      0.31994 0.00000      Inf 0.00000

```

```

$VaR
      80%   85%   90%   95%   97.5%   99%   99.5%
LP      52.59  94.72 217.1  896.3 3700.1 24112 99541
LGPD    54.45  91.20 186.5  624.6 2077.9 10151 33682
GGPD    53.40  91.62 194.6  699.0 2500.9 13468 48116
WGPD    54.84  91.71 187.0  622.9 2060.2  9986 32935
Gamma   503.90 732.03 1101.8 1830.3 2643.8  3806  4730
Weibull 125.63 187.52 300.5  570.5  947.1  1626  2287
Lognormal 71.33 104.79 170.0  348.4  649.0  1338  2190
Pareto  208.19 756.17 4657.1 104173.6 2330248.1 141753925 3170878114
Empirical 57.38  90.00 205.5  687.7 1936.0  6154  9492

```

```

$TVaR
      80%   85%   90%   95%   97.5%   99%   99.5%
LP      NA    NA    NA    NA    NA    NA    NA
LGPD    NA    NA    NA    NA    NA    NA    NA
GGPD    NA    NA    NA    NA    NA    NA    NA
WGPD    NA    NA    NA    NA    NA    NA    NA
Gamma   1497.9 1793.9 2240.8 3066  3947 5172  6131
Weibull 521.9  644.6  848.1 1285 1841 2776  3645
Lognormal 414.3 523.6  718.7 1197 1924 3447  5208
Pareto  NA    NA    NA    NA    NA    NA    NA

```

- Παρακάτω παρουσιάζονται τα outputs του αλγορίθμου “ANALYSIS” για τα δεδομένα “Iossalae” (συμβολ. “I”) που μελετήθηκαν στην ενότητα 8.3 (βλ. πίνακες 8.4, 8.5, 8.6, 8.7).

```

>
>
> ANALYSIS(1)
$`Descriptive statistics`
      Value
mean    12.588163
sd      28.145642
variance 792.177165
median   5.471000
range    501.848000
maximum  501.863000
minimum   0.015000
skewness 9.259366
kurtosis 125.136494

$LP
      MLE
α    2.155587
θ  408.635886
σ    1.428976

```

\$LGPD  
MLE  
 $\alpha$  2.1974379  
 $\theta$  0.5675039  
 $\lambda$  14.7993922  
 $\sigma$  2.9152304

\$GGPD  
MLE  
 $\beta$  0.3853078  
 $\gamma$  14.6527083  
 $\delta$  1.1324753  
 $\theta$  0.7433494

\$WGPD  
MLE  
 $\alpha$  1.3816500  
 $\theta$  16.8677436  
 $\lambda$  1.3478338  
 $\tau$  0.9794002

\$Gamma  
MLE  
 $\delta$  0.6630515  
 $\beta$  18.9919855

\$Weibull  
MLE  
 $\tau$  0.7416908  
 $\phi$  9.9798699

\$Lognormal  
MLE  
 $\mu$  1.61467  
 $\sigma$  1.42962

\$Pareto  
MLE  
 $\alpha$  0.1720008  
 $\theta$  0.0150000



```

`Goodness of fit`
      AIC      BIC K-S Statistic K-S P-value A-D Statistic A-D P-value
LP      10177.55 10193.49  0.04992248  0.00113  6.285963  0.00071
LGPD    10109.51 10130.77  0.02931978  0.15164  1.228130  0.25707
GGPD    10109.86 10131.11  0.02835151  0.17925  1.120064  0.29998
WGPD    10109.96 10131.22  0.01978141  0.60009  0.568188  0.67854
Gamma   10404.08 10414.71  0.09964149  0.00000  25.847783  0.00000
Weibull 10271.06 10281.68  0.06039543  0.00004  14.061407  0.00000
Lognormal 10175.29 10185.92  0.04921631  0.00140  6.182634  0.00079
Pareto  13127.43 13138.06  0.39706599  0.00000  Inf      0.00000

```

```

$VaR
      80%      85%      90%      95%      97.5%
LP      16.68000  22.03643  31.28575  52.60741  82.61071
LGPD    16.04008  20.35370  27.47704  43.15544  64.64825
GGPD    16.01059  20.31841  27.43685  43.12094  64.64949
WGPD    14.87279  18.63867  25.45538  42.91753  71.75617
Gamma   20.73065  25.34806  32.01259  43.70909  55.67048
Weibull 18.95735  23.66312  30.72485  43.81068  58.00322
Lognormal 16.74112  22.11763  31.39978  52.78277  82.82027
Pareto  173.72056  925.20040  9772.76335  549773.16517  30927847.36967
Empirical 15.71440  19.22020  25.92470  45.96570  64.86325

      99%      99.5%
LP      139.83093  200.57925
LGPD    105.75242  150.45960
GGPD    105.88518  150.80196
WGPD    140.54457  232.98671
Gamma   71.75661  84.07395
Weibull 78.22754  94.50591
Lognormal 139.83490  199.76187
Pareto  6367126894.44486  358186869147.50238
Empirical 131.70865  163.76856

```

---

R Console

```

$TVaR
      80%      85%      90%      95%      97.5%      99%      99.5%
LP      51.01415  61.63467  79.36520  118.63401  172.18641  272.89614  380.5059
LGPD    41.79463  49.71062  62.78278  91.55447  130.99628  206.42721  288.4701
GGPD    41.81218  49.74481  62.85304  91.73448  131.37820  207.31161  290.0235
WGPD    57.40648  71.00735  95.68520  158.90170  263.30341  512.33127  846.9905
Gamma   37.42267  42.25692  49.16417  61.16449  73.34262  89.63565  102.0708
Weibull 37.49728  42.93791  50.94874  65.47862  80.94860  102.67570  119.9832
Lognormal 50.39612  60.78715  78.04587  115.85406  166.43102  258.26316  351.5185
Pareto   NA      NA      NA      NA      NA      NA      NA
Empirical 42.86115  51.39320  65.88165  97.64420  139.76205  222.68033  287.6418

```

## ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Χατζηκωνσταντινίδης Ευ., Θεωρία Κινδύνου I, ΜΑΕ, ΠΑ.ΠΕΙ
2. Αντζουλάκος Δ., Ανάλυση Δεδομένων με τη Χρήση Στατιστικών Πακέτων – Εισαγωγή στο R, ΜΑΕ,ΠΑ.ΠΕΙ (2013).
3. Cooray, K., Ananda, M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.* 5:321–334.
4. Roxana Ciumara, An actuarial model based on the composite Weibull-Pareto distribution. *Math. Rep. (Bucur.)* 8(58) (2006), 401–414.
5. Teodorescu, S., Vernic, R. (2006), “A composite Exponential–Pareto distribution”. *The Annals of the “Ovidius” University of Constanta, Mathematics Series*, vol.XIV (1): 99-108.
6. Klugman, S. A., H. H. Panjer, and G. E. Willmot. 2008. *Loss Models: From Data to Decisions*. 3rd ed. Hoboken, NJ: Wiley.
7. Sheldon M. Ross. 2006. *Simulation*, Fourth Edition. Academic Press, Inc., USA.
8. Scollnik, D. P. M. 2007. On Composite Lognormal-Pareto Models. *Scandinavian Actuarial Journal* 1:20–33.
9. S. Teodorescu and R. Vernic, Some composite Exponential-Pareto models for actuarial prediction. *Rom. J. Econ. Forecast.* 12, 2009, 82–100;
10. Cooray, K (2009). The Weibull–Pareto Composite Family with Applications to the Analysis of Unimodal Failure Rate Data, *Communications in Statistics - Theory and Methods*, 38:11, 1901-1915.
11. Pigeon, M., Denuit, M., 2010. Composite lognormal-Pareto model with random threshold. *Scandinavian Actuarial Journal* 3, 177–192.
12. Scollnik, D. P. M., and C. Sun. 2012. Modeling with Weibull–Pareto Models. *North American Actuarial Journal* 16(2):260–272.
13. S. Nadarajah and S.A.A. Bakar, New composite models for the Danish fire insurance data. *Scand. Actuar. J.* 2012, 1–8;
14. S. Teodorescu and R. Vernic, On composite Pareto model. *Math Reports* 15 (65), 1, 2013, 11 -29.
15. Enrique Calderín-Ojeda (2015) On the Composite Weibull–Burr Model to describe claim data, *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1:1, 59-69.

16. Grün B & Miljkovic T (2019). Extending composite loss models using a general framework of advanced computational tools, *Scandinavian Actuarial Journal*, 2019:8, 642-660.
17. CIOCHINĂ, Ștefănuț. (2019). Composite models used in actuarial practice. *Review of the Air Force Academy*. 17. 65-79. 10.19062/1842-9238.2019.17.2.8.
18. M. Arif, D. M. Khan, S. K. Khosa, M. Aamir, A. Aslam *et al.*, Modelling insurance losses with a new family of heavy-tailed distributions, *Computers, Materials & Continua*, vol. 66, no.1, pp. 537–550, 2021.

