

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

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

**ΤΑΞΙΝΟΜΗΣΗ ΟΥΡΩΝ
ΣΥΝΑΡΤΗΣΕΩΝ ΚΑΤΑΝΟΜΩΝ
ΠΙΘΑΝΟΤΗΤΑΣ**

Δημήτριος Δ. Κονιδάρης

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

Πειραιάς
Ιούνιος 2018

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

Τα μέλη της Επιτροπής ήταν:

- Γεώργιος Πιτσέλης (Επιβλέπων)
- Δημήτριος Αντζουλάκος
- Βασίλειος Σεβρόγλου

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

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

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

TAIL ORDERING

By

Dimitrios D. Konidaris

MSc Dissertation

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

Piraeus, Greece
June 2018

*Στον πατέρα μου
Διονύσιο*

Ευχαριστίες

Ευχαριστώ θερμώς τον επιβλέποντα Επίκουρο Καθηγητή κύριο Πιτσέλη για την υπομονή και τη συμπαράσταση που έδειξε προς εμένα.

ΠΕΡΙΛΗΨΗ

Μεθοδολογίες εισάγονται για τον έλεγχο υπόθεσης σε σχέση με τη βαρύτητα της ουράς κατανομής βασισμένες στα αποτελέσματα του Rojo (1996). Χρησιμοποιείται η ασυμπτωτική συμπεριφορά της διαφοράς ακραίων τιμών και τα τεστ αποδεικνύονται θεωρητικά συνεπή και έχοντα υψηλά επίπεδα ισχύος ανά σημείο σύμφωνα με τους Lehmann (2005) και Lehmann and Loh (1990).

Τα αποτελέσματα προσομοίωσης υποδεικνύουν ότι το καινούριο στατιστικό τεστ παρουσιάζει μέτριο έλεγχο της πιθανότητας του σφάλματος τύπου I και έχει επαρκείς ιδιότητες ισχύος για συγκεκριμένα μεγέθη δείγματος. Γίνεται σύγκριση με το Bryson τεστ (1974) και αποδεικνύεται, ότι το Bryson τεστ δεν είναι πάντοτε ικανό να ελέγξει το σφάλμα τύπου I. Εξετάζονται επίσης τα λειτουργικά χαρακτηριστικά του τεστ μέσω της δημιουργίας υποδειγμάτων, και φαίνεται ότι η ισχύς εν γένει βελτιώνεται μέσω αυτής της λογικής. Η μεθοδολογία εφαρμόζεται και σε συγκεκριμένο σύνολο δεδομένων.

ABSTRACT

Methodologies to test hypotheses about the tail-heaviness of an underlying distribution are introduced based on results of Rojo (1996) using the limiting behavior of the extreme spacings. The tests are consistent and have point-wise robust levels in the sense of Lehmann (2005) and Lehmann and Loh (1990).

Simulation results based on these new methodologies indicate that the tests exhibit medium control of the probability of Type I error and have adequate power properties for finite sample sizes. The tests are compared with a test proposed by Bryson (1974) and it is seen that Bryson's test can not always control the Type I error. The operating characteristics of the tests are also explored when the data is blocked. It turns out that the power generally increases substantially by blocking. The methodology is illustrated by analyzing one particular data set.

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Πινάκων.....	
Κεφάλαιο 1: Εισαγωγή	1
Κεφάλαιο 2: Ταξινόμηση βάσει της συνάρτησης πυκνότητας ποσοστημορίου.....	4
Κεφάλαιο 3: Βελτιώσεις και ταξινόμηση βάσει της ασυμπτωτικής συμπεριφοράς της διαφοράς ακραίων τιμών (Extreme Spacing).....	8
Κεφάλαιο 4: Ταξινόμηση ουράς με χρήση της κατανομής του υπολειπόμενου χρόνου ζωής.....	14
Κεφάλαιο 5: Έλεγχος διαφοράς ακραίων τιμών(ES) κατανομής μεσαίας ουράς.....	19
Κεφάλαιο 6: Αποτελέσματα προσομοίωσης.....	25
Κεφάλαιο 7: Ισχυρότερα επίπεδα ελέγχου μέσω διάσπασης και ομαδοποίησης δεδομένων.....	30
Κεφάλαιο 8: Έλεγχος BRYSON.....	33
Κεφάλαιο 9: Στατιστική ανάλυση ουράς πραγματικών δεδομένων.....	37
Παράρτημα.....	41
Αποδείξεις Θεωρημάτων-Λημμάτων.....	41
Κώδικας.....	55
Βιβλιογραφία.....	83

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Συναρτήσεις πυκνότητας ποσοστημορίου για διάφορες συνήθεις κατανομές	6
Πίνακας 2: Κανόνας ταξινόμησης RP.....	9
Πίνακας 3: Ταξινόμηση ουρών συνήθων κατανομών με 3 τρόπους.....	18
Πίνακας 4: Πιθανότητα σφάλματος τύπου I – εκθετική, λογιστική και γάμμα κατανομή	26
Πίνακας 5: Πιθανότητες ταξινόμησης στην περίπτωση της Pareto(γ)	27
Πίνακας 6: Πιθανότητες ταξινόμησης στην περίπτωση της Weibull(γ).....	27
Πίνακας 7: Πιθανότητες ταξινόμησης στην περίπτωση ES short κατανομών...28	
Πίνακας 8: Πιθανότητες ταξινόμησης στην περίπτωση ES long κατανομών...29	
Πίνακας 9: Πιθανότητα σφάλματος τύπου I για δείγματα μεγέθους 500, 5000 και 20000 και διάφορους αριθμούς blocks.....	31
Πίνακας 10: Πιθανότητες ταξινόμησης στην περίπτωση ES long ή short κατανομών.....	31
Πίνακας 11: Πιθανότητες ταξινόμησης στην περίπτωση ES long ή short κατανομών.....	32
Πίνακας 12: Πιθανότητες ταξινόμησης στην περίπτωση ES long ή short κατανομών.....	32
Πίνακας 13: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape=2 , Scale=1.....	34
Πίνακας 14: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape=1 , Scale=1.....	34

Πίνακας 15: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape=1/2 , Scale=1.....	34
Πίνακας 16: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Logistic(0.7,1) κατανομή.....	35
Πίνακας 17: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Weibull(2) κατανομή.....	35
Πίνακας 18: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Weibull(1/2) κατανομή.....	35
Πίνακας 19: Υδρολογικό Έτος(Φράγμα Μαραθώνα) Παροχή (hm ³).....	37

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η συμπεριφορά της ουράς μία συνάρτησης κατανομής διαδραματίζει σπουδαίο ρόλο σε ποικίλες εφαρμογές, όπως η υδρολογία, διαστημική μηχανική, μετεωρολογία, ασφαλιστική επιστήμη και χρηματοοικονομικά. Η πιο συνηθισμένη προσέγγιση στην ταξινόμηση ουρών ορίζει ως βάρος ουράς το ρυθμό, με τον οποίο η συνάρτηση πυκνότητας πιθανότητας τείνει στο μηδέν στην περιοχή του ∞ , βλέπε Roho (1992). Με μεγαλύτερη ακρίβεια, μία πυκνότητα πιθανότητας f λέγεται ότι έχει ελαφρύτερη ουρά από την g εάν:

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0.$$

Εν γένει λοιπόν, μία αρχική προσέγγιση του θέματος είναι η παρακάτω (Πιτσέλης (2015)):

$$\lim_{x \rightarrow \infty} \frac{S_1(x)}{S_2(x)} = \lim_{x \rightarrow \infty} \frac{f_1(x)}{f_2(x)} = c,$$

όπου $f_1(x), f_2(x)$ συναρτήσεις πυκνότητας πιθανότητας και $S_1(x), S_2(x)$ οι αντίστοιχες συναρτήσεις επιβίωσης.

- (i) Αν $c=0$, τότε η $f_1(x)$ έχει πιο ελαφριά ουρά από την $f_2(x)$.
- (ii) Αν $0 < c < \infty$, τότε η $f_1(x)$ και η $f_2(x)$ έχουν όμοιες ουρές.
- (iii) Αν $c=\infty$, τότε η $f_1(x)$ έχει πιο βαριά ουρά από την $f_2(x)$.

Στην κλασική θεωρία ακραίων τιμών, το θεώρημα Fischer-Tippett ταξινομεί την δεξιά ουρά μίας συνάρτησης κατανομής ανάλογα με την ασυμπτωτική κατανομή του κανονικοποιημένου μεγίστου. Δηλαδή η συνάρτηση κατανομής F έχει κοντή, μεσαία ή μακριά ουρά αναλόγως αν η F βρίσκεται στην περιοχή έλξης της Weibull, Gumbel ή Frechet αντίστοιχα. Είναι ευρέως γνωστό, πάντως, ότι η οριακή κατανομή του κανονικοποιημένου μεγίστου δεν υπάρχει για όλες τις συναρτήσεις κατανομών. Για παράδειγμα, κάθε συνάρτηση κατανομής η οποία παρουσιάζει θετική μάζα πιθανότητας στο δεξί άκρο του στηρίγματος της F , δεν μπορεί να ταξινομηθεί από το παραπάνω θεώρημα. Ένα άλλο μειονέκτημα αυτού του θεωρήματος είναι ότι η κατηγορία των συναρτήσεων κατανομών μεσαίας ουράς καταλήγει να είναι μεγάλη. Για παράδειγμα, ο Schuster (1984) επιχειρηματολογεί λέγοντας ότι «ο στατιστικός θεωρεί την κανονική κατανομή κοντότερη από την εκθετική και τη δεύτερη κοντότερη από την Lognormal». Παρόλα αυτά, το θεώρημα Fischer-Tippett κατατάσσει και τις τρεις συναρτήσεις κατανομών στην περιοχή έλξης της Gumbel (μεσαία ουρά). Συνεπώς υπάρχει ανάγκη ταξινόμησης ουρών με εναλλακτικά σχήματα!

Ο σκοπός αυτής της διπλωματικής εργασίας είναι να παρουσιάσει μεθοδολογίες για έλεγχο υποθέσεων σχετικά με τη βαρύτητα της ουράς μίας συνάρτησης κατανομής βασισμένη στα αποτελέσματα του Rojo (1996). Υπάρχει επικέντρωση στη δεξιά ουρά της συνάρτησης κατανομής, αλλά ανάλογα αποτελέσματα ισχύουν και για την αριστερή ουρά λαμβάνοντας υπόψη τις τυχαίες μεταβλητές $\{-X_i, \dots, i = 1, \dots, n\}$. Μάλιστα, όταν η συνάρτηση πυκνότητας πιθανότητας είναι συμμετρική γύρω από το μηδέν, κάποιος θα μπορούσε να μελετήσει τις μεταβλητές $\{|X_i|, \dots, i = 1, \dots, n\}$ διπλασιάζοντας το μέγεθος του δείγματος. Το Θεώρημα 3.1 και το Πρόσιμα 4.2 του Rojo (1996) προσφέρουν τα απαιτούμενα εργαλεία για την ανάπτυξη μεθοδολογιών για έλεγχο της υπόθεσης ότι τα δεδομένα μας προέρχονται από μία κατανομή με μεσαία ουρά έναντι κοντής ή μακριάς ουράς. Βασιζόμαστε στην ασυμπτωτική κατανομή της διαφοράς ακραίων τιμών (Extreme Spacing) $X_{(n)} - X_{(n-1)}$, όπου

$X_{(k)}$ αναπαριστά το στατιστικό k τάξεως ενός τυχαίου δείγματος $\{X_1, \dots, X_n\}$ από την F . Επίσης η κατανομή F θεωρείται συνεχής και γνησίως αύξουσα σε όλη την εργασία.

Τα Κεφάλαια 2,3 παρέχουν το εισαγωγικό υλικό και μία σύντομη επισκόπηση των σχημάτων ταξινόμησης των Parzen (1979) και Schuster (1984). Το Κεφάλαιο 4 πραγματεύεται τα πιο σημαντικά αποτελέσματα από τον Rojo (1996) και το Κεφάλαιο 5 παρουσιάζει ένα καινούριο τεστ για τη βαρύτητα των ουρών. Το τεστ αυτό είναι συνεπές έναντι των εναλλακτικών υποθέσεων για κοντές ή μακριές ουρές και το πραγματικό επίπεδο σημαντικότητας είναι point-wise ισχυρό (Lehmann and Loh (1990)). Τα αποτελέσματα προσομοίωσης υποδεικνύουν, ότι για μικρά μεγέθη δείγματος υπάρχει καλός έλεγχος της πιθανότητας σφάλματος τύπου I του τεστ και υπάρχουν καλές ιδιότητες ισχύος. Έπειτα παρουσιάζεται μία μελέτη προσομοίωσης, όπου τα δεδομένα δομούνται σε blocks ώστε να αυξηθεί η ισχύς του ελέγχου. Μία σύγκριση που πραγματοποιείται στο Κεφάλαιο 8 με ένα τεστ προτεινόμενο από τον Bryson (1974) καταλήγει, ότι η πιθανότητα σφάλματος τύπου I είναι κοντά στο 1 για gamma κατανομές (κατανομές με μεσαία ουρά). Τέλος, υπάρχει και πρακτική εφαρμογή των παραπάνω στο Κεφάλαιο 9.

ΚΕΦΑΛΑΙΟ 2: ΤΑΞΙΝΟΜΗΣΗ ΒΑΣΕΙ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ ΠΥΚΝΟΤΗΤΑΣ ΠΟΣΟΣΤΗΜΟΡΙΟΥ

Ο Parzen (1979) ισχυρίσθηκε ότι πολλές κατανομές έχουν συνάρτηση πυκνότητας ποσοστημορίου της μορφής:

$$f[Q(u)] \sim (1 - u)^\alpha, \alpha > 0, \quad (1)$$

όπου f αναπαριστά την πυκνότητα πιθανότητας, Q είναι η συνάρτηση ποσοστημορίου (η αριστερά συνεχής αντίστροφη συνάρτηση της F), και $g_1(u) \sim g_2(u)$ σημαίνει ότι $g_1(u)/g_2(u)$ τείνει προς μία θετική σταθερά καθώς $u \rightarrow 1$. Η παράμετρος α ονομάζεται εκθέτης ουράς (tail exponent), και ο Parzen (1979) όρισε μία κατανομή να έχει κοντή, μεσαία ή μακριά ουρά ανάλογα με τον αν $\alpha < 1$, $\alpha = 1$ ή $\alpha > 1$ αντίστοιχα. Όταν $\alpha = 1$, η σχέση (1) μπορεί να γραφτεί σε πολλές περιπτώσεις (με μεγαλύτερη ακρίβεια) ως:

$$f[Q(u)] \sim (1 - u)\{\ln(1 - u)^{-1}\}^{1-\beta}, 0 \leq \beta \leq 1, \quad (2)$$

όπου β είναι παράμετρος σχήματος.

Στη συνέχεια παραθέτουμε τον ορισμό των συναρτήσεων βραδείας, μεσαίας και ταχείας κύμανσης.

Ορισμός 1 (Μπούτσικας 2015): Μια θετική συνάρτηση h στο $(0, \infty)$ καλείται **ομαλής κύμανσης** (regularly varying) στο ∞ με δείκτη α (συμβ. $h \in \mathcal{R}_\alpha$) αν

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

- Αν $\alpha = 0$ τότε καλείται **βραδείας κύμανσης** (slowly varying).
- Αν $\alpha = \infty$ ($t^\alpha = \infty$ αν $t \in (0, 1)$ ή 0 αν $t \in (1, \infty)$) καλείται **ταχείας κύμανσης** (rapidly varying).

Παραδείγματα:

- Συναρτήσεις **βραδείας κύμανσης**: οι σταθερές συναρτήσεις, οι συναρτήσεις που συγκλίνουν σε μια σταθερά (στο ∞), και συναρτήσεις που μεταβάλλονται «πολύ αργά» στο ∞ όπως η λογαριθμική συνάρτηση, κ.ο.κ.
 - Συναρτήσεις **ομαλής κύμανσης** είναι π.χ.: οι συναρτήσεις $x^a, x^a \ln(1+x)$ κ.ο.κ.
 - Συναρτήσεις **ταχείας κύμανσης** είναι π.χ. οι εκθετικές συναρτήσεις, π.χ. e^{-x} .
- Κάθε συνάρτηση $h \in \mathcal{R}_\alpha$, $\alpha \in \mathbb{R}$, μπορεί να γραφεί ως $h(x) = x^\alpha L(x)$, όπου $L \in \mathcal{R}_0$.

Έστω ότι L αναπαριστά μία συνάρτηση βραδείας κύμανσης από τα αριστερά στο 1. Τότε η ακριβής σχέση της (1) είναι:

$$f[Q(u)] = L(u) (1 - u)^\alpha, \alpha > 0, \tag{3}$$

Αποδεικνύεται ότι η σχέση (1) εφαρμόζεται σε πολλές περιπτώσεις συνηθισμένων κατανομών αλλά όχι σε όλες. Ένα παράδειγμα από τον Parzen (1979) κατανομής που δεν ακολουθεί την (1) είναι η $1 - F(x) = \exp(-x - 0.75 \sin x)$. Επίσης ο Parzen (1979) αναφέρει ότι η Lognormal κατανομή είναι

ένα παράδειγμα με $\alpha = 1$, αλλά δεν είναι δυνατή η κατασκευή της συνάρτησης πυκνότητας ποσοστημορίου στο πρότυπο της (2).

Ο Πίνακας 1 παρουσιάζει την συνάρτηση πυκνότητας ποσοστημορίου και την αντίστοιχη ταξινόμηση για πολλές συνήθεις κατανομές.

Κατανομή	Πυκνότητα ποσοστημορίου	Ταξινόμηση
Uniform(0,1)	1	Short
Exponential(γ)	$\frac{1}{\gamma}(1-u)$	Medium
Logistic	$u(1-u)$	Medium
Weibull(γ)	$\gamma(1-u)\left\{\log\frac{1}{1-u}\right\}^{1-\frac{1}{\gamma}}$	Medium
Extreme Value	$(1-u)\left\{\log\frac{1}{1-u}\right\}$	Medium
Normal	$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(\Phi^{-1}(u))^2\right\} \sim$ $\sim(1-u)\left\{2\log\frac{1}{1-u}\right\}^{\frac{1}{2}}$	Medium
Cauchy	$\frac{1}{\pi}\sin^2\pi u \sim(1-u)^2$	Long
Pareto(γ)	$\frac{1}{\gamma}(1-u)^{1+\gamma}$	Long
Burr(γ,τ)	$(1-u)^{1+\frac{1}{\gamma r}}$	Long

Πίνακας 1: Συναρτήσεις πυκνότητας ποσοστημορίου για διάφορες συνήθεις κατανομές και ταξινόμηση βάσει Parzen (1979)

Παρόλο που το σχήμα ταξινόμησης βάσει της (3) έχει θεωρητικό ενδιαφέρον, ένα κύριο μειονέκτημά του είναι ότι δεν προσφέρεται για ταξινόμηση βασισμένη σε πραγματικά δεδομένα, διότι είναι αρκετά δύσκολο να εκτιμηθεί ο

εκθέτης ουράς α της (1). Παρόλα αυτά, αφού το σχήμα ταξινόμησης βασισμένο στην (1) δίνει τελικά τα ίδια αποτελέσματα με αυτά από το Θεώρημα Fischer-Tippett, είναι εφικτό να χρησιμοποιηθούν μεθοδολογίες βασισμένες στην ασυμπτωτική συμπεριφορά του κανονικοποιημένου μεγίστου με σκοπό την ταξινόμηση από τυχαίο δείγμα. Σε αυτή την περίπτωση, όμως, η πρόκληση είναι να βρεθούν οι κατάλληλες ακολουθίες, ώστε να κανονικοποιηθεί το μέγιστο. Εφόσον η F είναι άγνωστη, αυτές οι ακολουθίες θα πρέπει να εκτιμηθούν από τα δεδομένα, πράγμα που περιπλέκει την ανάλυση. Τέλος σοβαρό μειονέκτημα της ταξινόμησης βασισμένη στις (1),(3) είναι και το μεγάλο προκύπτον μέγεθος της κατηγορίας μεσαίας ουράς.

ΚΕΦΑΛΑΙΟ 3: ΒΕΛΤΙΩΣΕΙΣ ΚΑΙ ΤΑΞΙΝΟΜΗΣΗ ΒΑΣΕΙ ΤΗΣ ΑΣΥΜΠΤΩΤΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΗΣ ΔΙΑΦΟΡΑΣ ΑΚΡΑΙΩΝ ΤΙΜΩΝ (EXTREME SPACING)

Εφόσον το σχήμα ταξινόμησης βασισμένο στην (1) δεν είναι αρκετά ευαίσθητο ώστε να διακρίνει π.χ. μεταξύ της κανονικής, εκθετικής και Lognormal κατανομής, ο Schuster (1984) βελτίωσε την προσέγγιση του Parzen (1979) για την κατηγορία της κατανομής μεσαίας ουράς. Μέσω αυτής της αλλαγής, η κανονική, εκθετική και Lognormal κατανομή ταξινομούνται σε ξεχωριστές κατηγορίες. Οι παρακάτω ορισμοί δίνουν τη λεγόμενη βελτιωμένη (refined) ταξινόμηση Parzen (RP) κάνοντας χρήση της οριακής τιμής της συνάρτησης έντασης αποτυχίας (στις ασφαλίσσεις ζωής: ένταση θνησιμότητας) $r_F(x) = f(x)/\{1 - F(x)\}$.

Ορισμός 2 (Schuster (1984)):

$$\alpha = \lim_{u \rightarrow 1^-} -(1 - u)f'Q(u)/[fQ(u)]^2, \quad (\text{tail exponent}) \quad (4)$$

$$c = \lim_{u \rightarrow 1^-} (1 - u)/fQ(u) = \lim_{u \rightarrow 1^-} 1/r_F(Q(u)) \quad (5)$$

Μία κατανομή ανήκει σε μία από τις ακόλουθες κατηγορίες όταν ικανοποιούνται οι αντίστοιχες συνθήκες:

Ταξινόμηση	Πρώτη συνθήκη	Δεύτερη συνθήκη
Short	$\alpha < 1$	
Medium-Short	$\alpha = 1$	$c = 0$
Medium-Medium	$\alpha = 1$	$0 < c < \infty$
Medium-Long	$\alpha = 1$	$c = \infty$
Long	$\alpha > 1$	

Πίνακας 2: Κανόνες ταξινόμησης RP βάσει Schuster (1984)

Η RP μέθοδος ταξινόμησης κατατάσσει την κανονική, εκθετική και Lognormal κατανομή στις κατηγορίες medium-short, medium-medium, και medium-long αντίστοιχα. Δυστυχώς, όπως και με το σχήμα ταξινόμησης του Parzen, η RP ταξινόμηση δεν μπορεί να χρησιμοποιηθεί εύκολα με πραγματικά δεδομένα διότι απαιτεί την εκτίμηση των ποσοτήτων $fQ(u)$ και $f'Q(u)$ για $u \rightarrow 1^-$.

Προς αντιμετώπιση του προαναφερθέντος προβλήματος, ο Schuster(1984) πρόσφερε ένα νέο σχήμα ταξινόμησης βάσει συμπεριφοράς ουράς μέσω της ασυμπτωτικής συμπεριφοράς της διαφοράς ακραίων τιμών (ES: Extreme Spacing). Πρόκειται για τη διαφορά μεταξύ της μέγιστης και της δεύτερης μεγαλύτερης παρατήρησης. Όταν η συνάρτηση ποσοστημορίου Q είναι διαφορίσιμη σε ανοιχτό από αριστερά διάστημα του 1, και αν η ποσότητα c ορισμένη στη σχέση (5) υπάρχει, ο Schuster κατηγοριοποίησε τις κατανομές βάσει του ES σύμφωνα με το Θεώρημα 1.

Θεώρημα 1 (Schuster (1984)): Έστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από την κατανομή $F(x)$. Ορίζουμε $S_n = X_{(n)} - X_{(n-1)}$, και υποθέτουμε ότι το c στην (5) υπάρχει. Τότε,

(i) $c = 0$ αν και μόνο αν $S_n = o_p(1)$,

(ii) $c = \alpha$, $0 < \alpha < \infty$ αν και μόνο αν $S_n = O_p(1)$, $S_n \neq o_p(1)$,

(iii) $c = \infty$ αν και μόνο αν $S_n \xrightarrow{P} \infty$,

όπου $o_p(1)$ υποδηλώνει ότι η ακολουθία των τυχαίων μεταβλητών συγκλίνει στο μηδέν κατά πιθανότητα, και $O_p(1)$ σημαίνει ότι η ακολουθία φράσσεται κατά πιθανότητα.

Η κατανομή F λέγεται τότε ES short, ES medium ή ES long, όταν ισχύουν τα (i), (ii) η (iii) αντίστοιχα. Το Θεώρημα 1 πραγματοποιεί τη σύνδεση μεταξύ της ασυμπτωτικής συμπεριφοράς της ES και της οριακής συμπεριφοράς της συνάρτησης έντασης αποτυχίας, όταν η ποσότητα c στην (5) υπάρχει. Η ένταση αποτυχίας τείνει στο μηδέν (π.χ. Pareto κατανομή $\bar{F}(x) = x^{-\alpha}$ με $\alpha > 0$), αν και μόνο αν η S_n συγκλίνει στο ∞ κατά πιθανότητα, και η ένταση αποτυχίας τείνει στο ∞ (π.χ. $\bar{F}(x) = e^{-e^x}$) αν και μόνο αν η S_n τείνει στο μηδέν κατά πιθανότητα. Διαφορετικά η ένταση αποτυχίας συγκλίνει σε μια θετική σταθερή ποσότητα (π.χ. $\bar{F}(x) = e^{-x}$) αν και μόνο αν η S_n δεν συγκλίνει στο μηδέν αλλά παραμένει φραγμένη κατά πιθανότητα.

Ο Schuster (1984) έκανε επίσης και μία σύνδεση μεταξύ της RP και της ES μεθόδου ταξινόμησης. Χρησιμοποιώντας την συνάρτηση πυκνότητας ποσοστημορίου (3) και ιδιότητες των συναρτήσεων βραδείας κύμανσης, έπεται ότι:

$$\lim_{u \rightarrow 1^-} \frac{fQ(u)}{1-u} = \lim_{u \rightarrow 1^-} L(u)(1-u)^{\alpha-1} = \begin{cases} 0 & \text{if } \alpha > 1, \\ 0 \leq \lim_{u \rightarrow 1^-} L(u) \leq \infty & \text{if } \alpha = 1, \\ \infty & \text{if } \alpha < 1. \end{cases} \quad (6)$$

Επομένως η κατηγορία ES short περιλαμβάνει τις RP short και RP medium-short κατηγορίες, η κατηγορία ES medium αντιστοιχεί στην RP medium-medium και η ES long περιλαμβάνει τις RP medium-long και RP long κατηγορίες.

Συνεπώς η εξίσωση (6) παρέχει μία απλή, διαισθητική ερμηνεία της ES ταξινόμησης. Εφόσον οι απαραίτητες υποθέσεις ικανοποιούνται, η κατανομή είναι:

- ES short εάν $1-F(x) \rightarrow 0$ γρηγορότερα από $f(x) \rightarrow 0$.
- ES medium εάν $1-F(x) \rightarrow 0$ με τον ίδιο ρυθμό με $f(x) \rightarrow 0$.
- ES long εάν $1-F(x) \rightarrow 0$ αργότερα από $f(x) \rightarrow 0$.

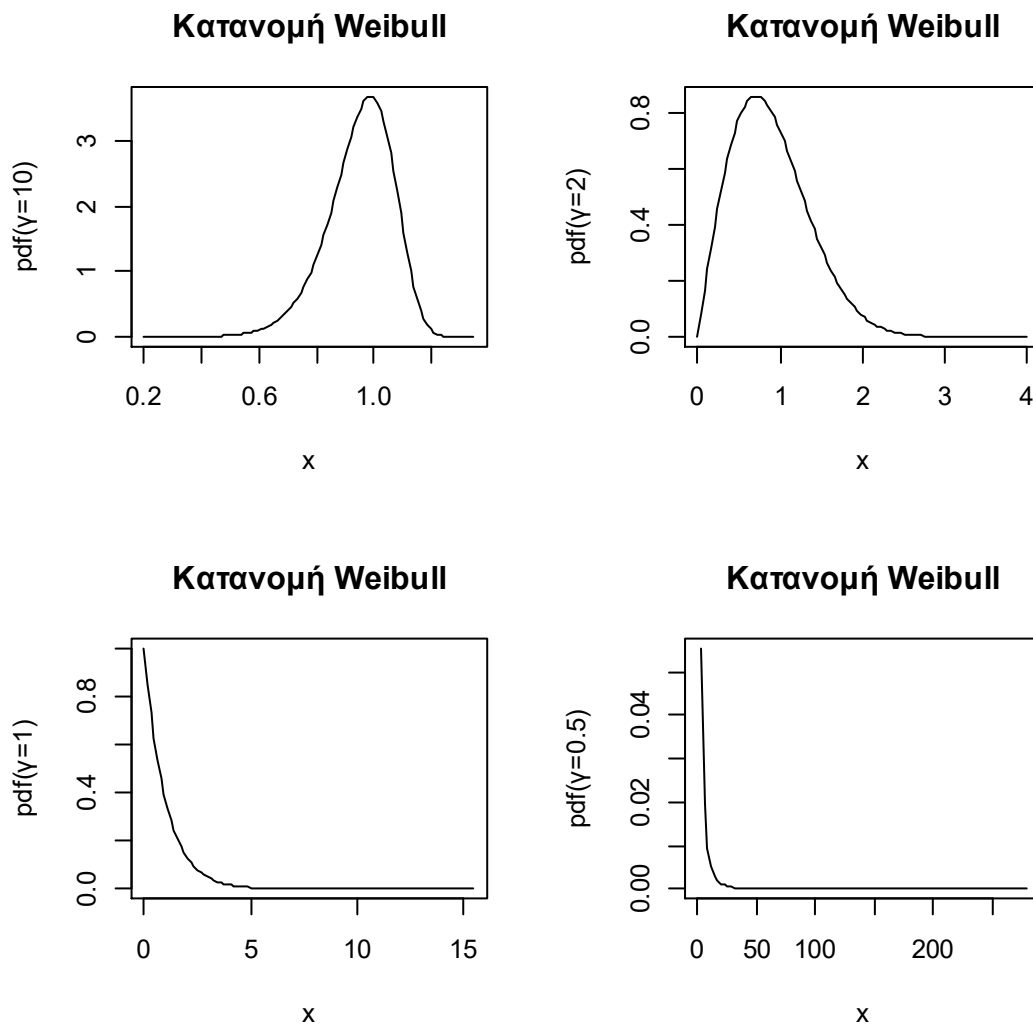
Η κατανομή Weibull αποτελεί ένα παράδειγμα όπου εμφανίζεται κοντή, μεσαία ή μακριά ουρά αναλόγως με την τιμή της παραμέτρου σχήματος γ .

Παράδειγμα 1 (Schuster (1984)): Για την κατανομή Weibull $\bar{F}(x) = e^{-x^\gamma}$,

$$\lim_{u \rightarrow 1^-} \frac{1-u}{fQ(u)} = \gamma^{-1} [-\ln(1-u)]^{\frac{1}{\gamma}-1} = \begin{cases} \infty, & 0 < \gamma < 1, \\ 1, & \gamma = 1 \\ 0, & \gamma > 1. \end{cases} \quad (7)$$

Άρα η κατανομή Weibull είναι ES short για $\gamma > 1$, ES medium για $\gamma = 1$ και ES long για $\gamma < 1$.

Στο Γράφημα 1 απεικονίζεται η μορφή της συνάρτησης πυκνότητας πιθανότητας Weibull για διάφορες παραμέτρους σχήματος γ , από τα οποία συμπεραίνουμε οπτικά την ουσία του παραδείγματος 1.



Γράφημα 1: Συνάρτηση πυκνότητας πιθανότητας της Weibull(γ) (Βλέπε Παράρτημα_Κώδικας_Graph1_Chapter 3, Weibull)

Στον Πίνακα 3 δίνονται αντίστοιχα παραδείγματα κατανομών ταξινομημένων σύμφωνα με την ασυμπτωτική συμπεριφορά της ποσότητας ES.

Υπάρχει, όμως, μία έλλειψη ακρίβειας στο Θεώρημα 1 και την πληροφορία που παρέχεται από την (6), με την έννοια ότι η περίπτωση (iii) του Θεωρήματος 1 και η περίπτωση $\alpha = 1$ στην (6) περιλαμβάνουν κατανομές με ουρές medium, short και long. Η σύνδεση μεταξύ της ασυμπτωτικής συμπεριφοράς της έντασης αποτυχίας και της βαρύτητας της ουράς της κατανομής F θα γίνει ακριβής στο Λήμμα 5 παρακάτω.

Η μέθοδος ταξινόμησης ES προτείνει την πιθανότητα χρήσης της ασυμπτωτικής συμπεριφοράς της S_n για να γίνει η διαφοροποίηση μεταξύ των κατανομών με κοντή, μεσαία και μακριά ουρά, αλλά χρειάζονται πιο εξειδικευμένα αποτελέσματα αυτής της ασυμπτωτικής συμπεριφοράς της S_n .

ΚΕΦΑΛΑΙΟ 4: ΤΑΞΙΝΟΜΗΣΗ ΟΥΡΑΣ ΜΕ ΧΡΗΣΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΤΟΥ ΥΠΟΛΕΙΠΟΜΕΝΟΥ ΧΡΟΝΟΥ ΖΩΗΣ

Ο Rojo (1996) πρότεινε ένα σχήμα ταξινόμησης βασισμένο στην ασυμπτωτική συμπεριφορά της κατανομής της υπολειπόμενης ζωής. Αυτή η προσέγγιση παρακάμπτει πολλές από τις απαιτούμενες τεχνικές υποθέσεις προηγούμενων προσεγγίσεων και παρέχει έναν πιο ακριβή χαρακτηρισμό κατάταξης.

Ορισμός 3 (Rojo (1996)): Ορίζουμε $h(t) = \lim_{x \rightarrow \infty} H_x(t) = \lim_{x \rightarrow \infty} \frac{\bar{F}(x+t)}{\bar{F}(x)}$, $t > 0$ όταν υπάρχει το όριο. Η συνάρτηση κατανομής F θεωρείται ότι έχει κοντή ουρά εάν $h(t) = 0$, μεσαία ουρά αν $0 < h(t) < 1$ και μακριά ουρά εάν $h(t)=1, \forall t$.

Παρόλο που το όριο $h(t)$ υπάρχει για μία αρκετά μεγάλη ομάδα των συναρτήσεων κατανομής, αυτό δεν ισχύει πάντα. Το όριο δεν υπάρχει, για παράδειγμα, όταν υπάρχει μία ταλαντωτική συμπεριφορά στην ουρά της πυκνότητας της υπολειπόμενης ζωής. Παραδείγματα όπου το όριο $h(t)$ δεν υπάρχει είναι: $\bar{F}(x) = \exp(-x - 0.75\sin(x))$, $\bar{F}(x) = c \left(1 + (1+x)^{-1/2} + \sin((1+x)^{1/2})\right) e^{-x}$, $x > 0$, όπου $c = (2 + \sin(1))^{-1}$.

Τα ακόλουθα αποτελέσματα είναι απόρροια του Ορισμού 3. Το Θεώρημα 3 συνδυάζει το Θεώρημα 4.1 και το Πρόσχημα 4.2 του Rojo (1996). Από εδώ και στο έξης, ο συμβολισμός $Exp(\theta)$ θα υποδηλώνει την εκθετική κατανομή με παράμετρο θ , και άρα μέση τιμή $1/\theta$.

Θεώρημα 3 (Rojo (1996)): Έστω $F(x)$ μία συνάρτηση πυκνότητας πιθανότητας και $h(t)$ το όριο του Ορισμού 2. Τότε:

Η F έχει κοντή ουρά αν και μόνο αν $S_n = X_{(n)} - X_{(n-1)} \xrightarrow{a.s.} 0$,

Η F έχει κοντή ουρά αν και μόνο αν $S_n = X_{(n)} - X_{(n-1)} \xrightarrow{a.s.} \text{Exp}(\theta)$,

Η F έχει μακριά ουρά αν και μόνο αν $S_n = X_{(n)} - X_{(n-1)} \xrightarrow{a.s.} \infty$.

Σημειώστε ότι τα αποτελέσματα του Θεωρήματος 3 παρέχουν έναν πιο ακριβή χαρακτηρισμό της ουράς των κατανομών σε σχέση με το Θεώρημα 1, ο οποίος βασίζεται στην ασυμπτωτική συμπεριφορά του ES. Συγκεκριμένα, το Θεώρημα 3 σκιαγραφεί την ασυμπτωτική κατανομή του ES για τη μεσαία ουρά. Αυτό είναι και το αποτέλεσμα που θα οδηγήσει σε μία μεθοδολογία για έλεγχο της υπόθεσης της μεσαίας ουράς έναντι των κοντών ή μακριών ουρών στο επόμενο κεφάλαιο.

Παράδειγμα 2 (Rojo (1996)): Για την κατανομή Weibull(γ) έχουμε

$$h(t) = \lim_{x \rightarrow \infty} \frac{\bar{F}(x+t)}{\bar{F}(x)} = \lim_{x \rightarrow \infty} e^{-(x+t)^\gamma + x^\gamma} = \begin{cases} 1, & 0 < \gamma < 1, \\ e^{-t}, & \gamma = 1 \\ 0, & \gamma > 1. \end{cases} \quad (8)$$

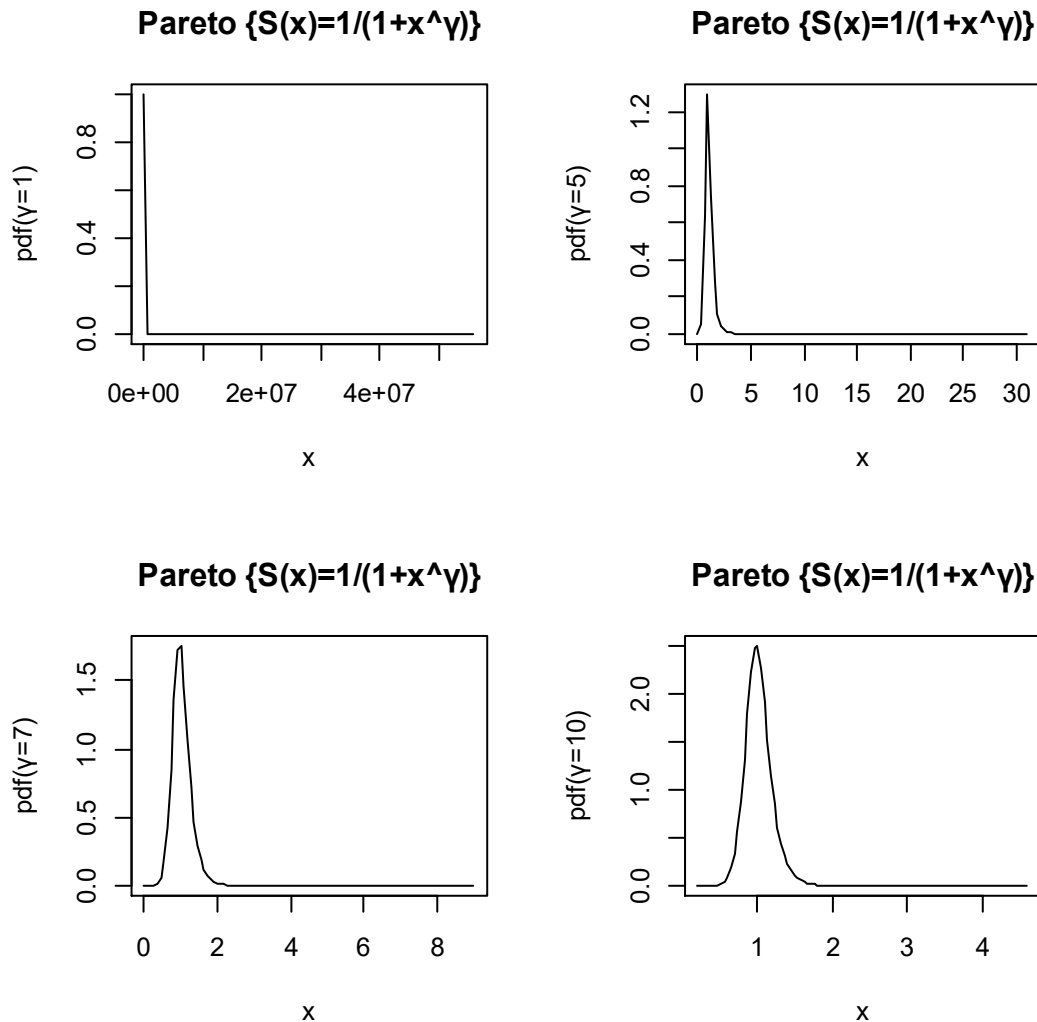
Επομένως, για $\gamma=1$ η κατανομή Weibull(1) έχει μεσαία ουρά. Για $0 < \gamma < 1$ έχει μακριά ουρά και για $\gamma > 1$ έχει κοντή ουρά.

Παράδειγμα 3 (Rojo (1996)): Για την κατανομή Pareto(γ) έχουμε

$$h(t) = \lim_{x \rightarrow \infty} \frac{\bar{F}(x+t)}{\bar{F}(x)} = \lim_{x \rightarrow \infty} \frac{(x+t)^{-\gamma}}{x^{-\gamma}} = 1, \quad \forall t > 0. \quad (9)$$

Επομένως η κατανομή Pareto(γ) είναι ES long από τον Ορισμό 3. Στο Γράφημα 2 απεικονίζεται η μορφή της συνάρτησης πυκνότητας πιθανότητας Pareto για

διάφορες παραμέτρους σχήματος γ , από τα οποία συμπεραίνουμε οπτικά την ουσία του παραδείγματος 3.



**Γράφημα 2: Συνάρτηση πυκνότητας πιθανότητας της Pareto(γ)
(Βλέπε Παράρτημα_Κώδικας_Graph2_Chapter 3, Pareto)**

Είναι δυνατό να βελτιώσουμε την ταξινόμηση που επιβάλλει ο Ορισμός 3 υποδιαιρώντας την κλάση των κοντών και μακριών ουρών κατανομών σε τρεις υποκλάσεις. Αυτό μπορεί να γίνει με τη βοήθεια των συναρτήσεων $M(x) = \bar{F}^{-1}(e^{-\bar{F}(x)})$, $N(x) = \bar{F}^{-1}(-1/\ln \bar{F}(x))$ και για $0 < \delta_1 < \infty$, $-\infty < \delta_2 < 0$ σύμφωνα με τα παρακάτω (Rojo (1996)) $\{V \sim U(0,1)\}$:

- Η F έχει weakly short tail αν και μόνο αν $M(X_{(n)}) - M(X_{(n-1)}) \xrightarrow{a.s.} -\infty$.
- Η F έχει moderately short tail αν και μόνο αν $M(X_{(n)}) - M(X_{(n-1)}) \xrightarrow{a.s.} \frac{1}{\delta_1} \ln V$.
- Η F έχει super short tail αν και μόνο αν $M(X_{(n)}) - M(X_{(n-1)}) \xrightarrow{a.s.} 0$.
- Η F έχει weakly long tail αν και μόνο αν $N(X_{(n)}) - N(X_{(n-1)}) \xrightarrow{a.s.} 0$.
- Η F έχει moderately long tail αν και μόνο αν $N(X_{(n)}) - N(X_{(n-1)}) \xrightarrow{a.s.} -\frac{1}{\delta_2} \ln V$.
- Η F έχει super long tail αν και μόνο αν $N(X_{(n)}) - N(X_{(n-1)}) \xrightarrow{a.s.} -\infty$.

Ο Πίνακας 3 συνοψίζει την ταξινόμηση διαφόρων κατανομών σύμφωνα με τα σχήματα που έχουν ήδη αναφερθεί. Παρατηρούμε ότι το σχήμα ταξινόμησης βάσει της ασυμπτωτικής συμπεριφοράς της ES είναι σταθερό, και αποτελεί τη βάση αναφοράς για τα άλλα δύο.

Κατανομή	Ταξινόμηση RP SCHUSTER (1984)	Ταξινόμηση ES SCHUSTER (1984)	Ταξινόμηση ROJO (1996)
Exponential	Medium-Medium	Medium	Medium
Normal	Medium-Short	Short	Weakly-Short
Lognormal	Medium-Long	Long	Weakly-Long
Uniform	Short	Short	Super-Short
Cauchy	Long	Long	Weakly-Long
Extreme Value	Medium-Short	Short	Moderately- Short
Pareto($\gamma < 1$)	Long	Long	Super-Long
Pareto($\gamma = 1$)	Long	Long	Moderately- Long
Pareto($\gamma > 1$)	Long	Long	Weakly-Long
Weibull($\gamma < 1$)	Medium-Long	Long	Weakly-Long
Weibull($\gamma = 1$)	Medium-Medium	Medium	Medium
Weibull($\gamma > 1$)	Medium-Short	Short	Weakly-Short
Logistic	Medium-Medium	Medium	Medium
Standard Extreme Value	Medium-Short	Short	Moderately- Short

Πίνακας 3: Ταξινόμηση συρών συνήθων κατανομών με 3 τρόπους

ΚΕΦΑΛΑΙΟ 5: ΕΛΕΓΧΟΣ ΔΙΑΦΟΡΑΣ ΑΚΡΑΙΩΝ ΤΙΜΩΝ(ΕΣ) ΚΑΤΑΝΟΜΗΣ ΜΕΣΑΙΑΣ ΟΥΡΑΣ

Έστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από μία κατανομή F , και έστω \bar{F}_n η εμπειρική συνάρτηση επιβίωσης. Θεωρούμε τη στατιστική συνάρτηση:

$$T_n = \frac{-\ln \bar{F}_n(\ln X_{(n)})(X_{(n)} - X_{(n-1)})}{\ln X_{(n)}}. \quad (10)$$

Αυτό το κεφάλαιο εξετάζει τα λειτουργικά χαρακτηριστικά αυτής της συνάρτησης στο πλαίσιο του ελέγχου υπόθεσης για την ουρά της κατανομής F . Η διαίσθηση που μας οδήγησε να κάνουμε την επιλογή της (10) ως στατιστικής συνάρτησης ελέγχου προέρχεται από συγκεκριμένο αποτέλεσμα του Rojo (1996). Αφού για μία κατανομή μεσαίας ουράς ισχύει $S_n = X_{(n)} - X_{(n-1)} \xrightarrow{a.s.} \text{Exp}(\theta)$ για κάποιο $\theta > 0$ αλλά άγνωστο, υπάρχει η ανάγκη να εκτιμηθεί το θ ώστε να κατασκευάσουμε μία στατιστική συνάρτηση ελέγχου της οποίας η ασυμπτωτική κατανομή να μην εξαρτάται από το θ . Από τον Rojo (1996) έχουμε, ότι για μία κατανομή μεσαίας ουράς ισχύει:

$$\bar{F}(\ln y) = y^{-\theta} l(y), \theta > 0, \quad (11)$$

όπου $l(y)$ είναι μία άγνωστη συνάρτηση βραδείας κύμανσης. Επομένως

$$\frac{-\ln \bar{F}(\ln y)}{\ln y} = \theta - \frac{\ln l(y)}{\ln y} \quad (12)$$

Για μία συνάρτηση βραδείας κύμανσης $l(y)$ είναι:

$$\lim_{y \rightarrow \infty} \frac{\ln l(y)}{\ln y} = 0.$$

Άρα

$$\theta = \lim_{y \rightarrow \infty} \frac{-\ln \bar{F}(\ln y)}{\ln y}, \quad (13)$$

και το θ μπορεί να εκτιμηθεί ως κάτωθι:

$$\widehat{\theta}_n = \frac{-\ln \bar{F}_n(\ln X_{(n)})}{\ln X_{(n)}}. \quad (14)$$

Η συνέπεια του εκτιμητή $\widehat{\theta}_n$ είναι άμεση συνέπεια του Θεωρήματος 4.

Θεώρημα 4 (Rojo (2010)): Έστω ότι $\frac{\bar{F}(y)}{(\bar{F}(\ln y))^\delta} \rightarrow 0$ καθώς $y \rightarrow \infty$ για κάποιο $\delta > 2$. Τότε

$$\frac{-\ln \bar{F}_n(\ln X_{(n)})}{-\ln \bar{F}(\ln X_{(n)})} \xrightarrow{P} 1. \quad (15)$$

Απόδειξη (Rojo (2010)): Βλέπε Παράρτημα, Αποδείξεις Θεωρημάτων-Λημμάτων.

Η συνέπεια του εκτιμητή $\widehat{\theta}_n$ έπεται από το Θεώρημα 4 πολλαπλασιάζοντας και διαιρώντας την έκφραση (14) με το $-\ln \bar{F}(\ln X_{(n)})$ και μετά κάνοντας χρήση της (15).

Αφού η θ είναι η παράμετρος κλίμακας της ES στην περίπτωση της μεσαίας ουράς του Θεωρήματος 3, για τη στατιστική συνάρτηση (10) ισχύει κάτω από την υπόθεση της ES medium-tailed κατανομής:

$$\frac{-\ln \bar{F}_n(\ln X_{(n)})}{\ln X_{(n)}} (X_{(n)} - X_{(n-1)}) \xrightarrow{D} \text{Exp}(1) \quad (16)$$

Με σκοπό τη χρήση της (16) ως βάση για τον έλεγχο υπόθεσης μεσαίας ουράς, θα πρέπει πρώτα να επιβεβαιωθεί ότι οι υποθέσεις του Θεωρήματος 4 ισχύουν για κατανομές με μεσαία ουρά. Αποδεικνύεται ότι όντως $\frac{\bar{F}(y)}{(\bar{F}(\ln y))^\delta} \rightarrow 0$ καθώς $y \rightarrow \infty$ για κάποιο $\delta > 2$ ικανοποιείται για όλες τις κατανομές με μεσαία ουρά και για τις περισσότερες κατανομές με κοντές ή μακριές ουρές. Αυτό προκύπτει από τα επόμενα Λήμματα 5,6, για τα οποία χρησιμοποιούμε τους εξής συμβολισμούς:

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

$$f \in R_{-\infty} \text{ αν και μόνο αν } \lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \begin{cases} 0, & \lambda > 1 \\ \infty, & 0 < \lambda < 1 \end{cases}$$

Λήμμα 5 (Rojo (2010)): Υποθέτουμε ότι η F έχει πυκνότητα f και $r_F(x) = f(x)/\bar{F}(x)$ η ένταση αποτυχίας. Τότε:

- (i) Η κατανομή F έχει κοντή ουρά αν και μόνο αν $r_F(t) \rightarrow \infty$ καθώς $t \rightarrow \infty$.
- (ii) Η κατανομή F έχει μεσαία ουρά αν και μόνο αν $r_F(t) \rightarrow \theta$ καθώς $t \rightarrow \infty$ για κάποιο $0 < \theta < \infty$.
- (iii) Η κατανομή F έχει μακριά ουρά αν και μόνο αν $r_F(t) \rightarrow 0$ καθώς $t \rightarrow \infty$.

Απόδειξη (Rojo (2010)): Βλέπε Παράρτημα, Αποδείξεις Θεωρημάτων-Λημμάτων.

Λήμμα 6 (Rojo (2010)): Έστω ότι η F έχει κοντή, μεσαία ή μακριά ουρά έτσι ώστε $\bar{F}(\ln x) \in R_{-\infty}$, $\bar{F}(\ln x) = x^{-\theta} l_1(x)$, $\bar{F}(\ln x) = l_2(x)$ αντίστοιχα, για κάποιες συναρτήσεις $l_i(x)$ βραδείας κύμανσης και κάποιο $\theta > 0$.

(i) Εάν η F έχει μακριά ουρά, τότε χωρίς περιορισμό της γενικότητας, $\bar{F}(\ln x) = \exp\left\{\int_1^x \frac{\varepsilon(t)}{t} dt\right\}$, με $\varepsilon(t) \rightarrow 0$ καθώς $t \rightarrow \infty$, έτσι ώστε να είναι η $\bar{F}(\ln x)$ μία συνάρτηση βραδείας κύμανσης και $r_F(\ln x) = -\varepsilon(x)$.

(ii) Εάν η F έχει κοντή ουρά, τότε χωρίς περιορισμό της γενικότητας, $\bar{F}(\ln x) = \exp\left\{\int_1^x \frac{z(u)}{u} du\right\}$, με $z(u) \rightarrow -\infty$ καθώς $u \rightarrow \infty$, έτσι ώστε να είναι η $\bar{F}(\ln x)$ μία συνάρτηση ταχείας κύμανσης και $r_F(\ln x) = -z(x)$.

(iii) Έστω ότι η F έχει κοντή ή μακριά ουρά με $\frac{z(x)}{z(e^x)} = o(x)$ ή $\frac{\varepsilon(x)}{\varepsilon(e^x)} = o(x)$ αντίστοιχα. Τότε (και για μεσαία ουρά) ισχύει:

$$\frac{\bar{F}(y)}{(\bar{F}(\ln y))^\delta} \rightarrow 0 \quad (17)$$

καθώς $y \rightarrow \infty$ για όλα τα $\delta > 0$.

Απόδειξη (Rojo (2010)): Βλέπε Παράρτημα, Αποδείξεις Θεωρημάτων-Λημμάτων.

Με βάση τα παραπάνω Θεωρήματα και Λήμματα μπορεί να στηθεί έλεγχος υπόθεσης μίας κατανομής με ES medium ουρά και επίπεδο σημαντικότητας α . Η μηδενική υπόθεση είναι ότι η F έχει ES medium ουρά έναντι της πρώτης

εναλλακτικής ότι η F έχει ES short ουρά ή της δεύτερης εναλλακτικής ότι η F έχει ES long ούρα. Οι κανόνες απόφασης σε επίπεδο σημαντικότητας α είναι: απορρίπτουμε την H_0 έναντι της πρώτης εναλλακτικής εάν $T_n < -\ln(1 - \alpha)$ και απορρίπτουμε τη δεύτερη εναλλακτική εάν $T_n > -\ln(\alpha)$. Διαφορετικά δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση!!

Είναι φανερό ότι, λόγω της ασυμπτωτικής κατανομής της T_n τα ασυμπτωτικά επίπεδα του ελέγχου αναμένουμε να είναι ίσα με α . Άρα ο έλεγχος είναι point-wise (σε κάθε σημείο) ισχυρός σύμφωνα με τους Lehmann and Loh (1990). Δηλαδή αναμένουμε να ισχύει:

« Η αληθινή στάθμη εμπιστοσύνης $a_n(F)$ του ελέγχου σε ονομαστικό επίπεδο α και με δείγμα μεγέθους n τείνει στο α καθώς $n \rightarrow \infty$ για κάθε κατανομή F ».

Τα τρία επόμενα θεωρήματα αποδεικνύουν την συνέπεια του ελέγχου έναντι των εναλλακτικών της κοντής ή μακριάς ουράς.

Θεώρημα 7 (Rojo (2010)): Έστω ότι η F έχει κοντή ουρά έτσι ώστε $-\ln \bar{F}(\ln x) = h(x)$ με h συνάρτηση ομαλής κύμανσης με δείκτη $0 \leq \gamma \leq \infty$. Όταν $\gamma=0$, υποθέτουμε επίσης ότι $\frac{r_F(\ln x)}{r_F(x)} \rightarrow 0$ όταν $x \rightarrow \infty$. Κάτω από τις υποθέσεις του Θεωρήματος 4, ο έλεγχος που ορίζεται από τη στατιστική συνάρτηση T_n ο οποίος απορρίπτει την H_0 όταν $T_n < -\ln(1 - \alpha)$ είναι συνεπής έναντι της τάξης των εναλλακτικών κοντών ουρών.

Απόδειξη (Rojo (2010)): Βλέπε Παράρτημα, Αποδείξεις Θεωρημάτων-Λημμάτων.

Θεώρημα 8 (Rojo (2010)): Έστω ότι η F έχει μακριά ουρά και έστω ότι $r_F(x)$ τελικά είναι φθίνουσα με

$$\frac{r_F(x)}{r_F(\ln x)} = o(1) \quad (18).$$

Κάτω από τις υποθέσεις του Θεωρήματος 4, ο έλεγχος που ορίζεται από τη στατιστική συνάρτηση T_n ο οποίος απορρίπτει την H_0 όταν $T_n > -\ln(a)$ είναι συνεπής έναντι της τάξης των εναλλακτικών μακριών ουρών.

Απόδειξη (Rojo (2010)): Βλέπε Παράρτημα, Αποδείξεις Θεωρημάτων-Λημμάτων.

Θεώρημα 9 (Rojo (2010)): Έστω ότι η F έχει μακριά ουρά με \bar{F} να είναι ομαλής κύμανσης με δείκτη $\alpha \geq 0$, έστι ώστε $\bar{F}(x) = c(x) \exp\left(\int_1^x \frac{\varepsilon(t)}{t} dt\right) \equiv L(x)$, όπου $\varepsilon(t) \rightarrow -\alpha$, και υποθέτουμε επίσης ότι $c(x) \rightarrow c > 0$ με $c(x)$ μη φθίνουσα και $-L'/L$ τελικά μη αύξουσα. Τότε, όταν το Θεώρημα 4 ισχύει, ο έλεγχος που ορίζεται από τη στατιστική συνάρτηση T_n ο οποίος απορρίπτει την H_0 όταν $T_n > -\ln(a)$ είναι συνεπής έναντι της τάξης των εναλλακτικών μακριών ουρών.

Απόδειξη (Rojo (2010)): Βλέπε Παράρτημα, Αποδείξεις Θεωρημάτων-Λημμάτων.

ΚΕΦΑΛΑΙΟ 6: ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ

Στο Κεφάλαιο 5 αναφερθήκαμε στη συνέπεια του ελέγχου έναντι των εναλλακτικών ουρών κατανομών, καθώς και στις ασυμπτωτικές ιδιότητες του σφάλματος τύπου I. Σε αυτό το κεφάλαιο θα μελετήσουμε το σφάλμα τύπου I, και την ισχύ του ελέγχου για πεπερασμένα δείγματα από διάφορες κατανομές μέσα από την τεχνική της προσομοίωσης.

Ο Πίνακας 4 δίνει τις πιθανότητες απόρριψης καθώς παίρνουμε δείγμα από ποικίλες εκθετικές αλλά και λογιστική κατανομή και κατανομή gamma με $scale=1$ και $shape=0.7$. Οι τιμές που δίνονται είναι πιθανότητες σφαλμάτων τύπου I, αφού όλες οι κατανομές είναι ES medium. Η κάθε προσομοίωση τρέχει 10.000 φορές για κάθε επιλεγμένη κατανομή.

Ο Πίνακας 4 δείχνει μέτρια απόδοση κάτω από ποικίλες ES medium-tailed κατανομές της στατιστικής συνάρτησης

$$T_n = \frac{-\ln \bar{F}_n(\ln X_{(n)})(X_{(n)} - X_{(n-1)})}{\ln X_{(n)}}$$

Κάποιες από τις τιμές είναι κοντά στο επιθυμητό $\alpha=0.05$ εκτός από τις περιπτώσεις $\text{Exp}(100)$, $\text{Lgis}(0.7,1)$ και $\text{Gamma}(0.7,1)$, καθώς και από μεμονωμένες όπως τις $\text{Exp}(1)$ και $\text{Lgis}(0.7,1)$ για μικρά δείγματα ($n=10$).

Οι περιπτώσεις $\text{Exp}(100)$, $\text{Lgis}(0.7,1)$ και $\text{Gamma}(0.7)$ απεικονίζουν το γεγονός ότι η σύγκλιση του όρου $\frac{\ln(y)}{\ln y}$ στις σχέσεις (12) και (13) μπορεί να είναι αρκετά αργή. Όταν αυτό συμβαίνει, τότε ο εκτιμητής $\hat{\theta}_n = \frac{-\ln \bar{F}_n(\ln X_{(n)})}{\ln X_{(n)}}$ δεν

συγκλίνει αρκετά γρήγορα στο θ και αυτό αντανakλάται στην πιθανότητα σφάλματος τύπου I.

n	E(100)	E(100)	E(1)	E(1)	E(0.01)	E(0.01)	Lgis	Lgis	Gamma	Gamma
	S	L	S	L	S	L	S	L	S	L
10	1	0	0.0717	0.0634	0.5885	0.1157	0.058	0.057	0.1018	0.1118
50	1	0	0.0503	0.0587	0.111	0.07	0.0629	0.0272	0.0391	0.0955
100	1	0	0.048	0.053	0.0625	0.0611	0.0662	0.0219	0.0376	0.0913
250	1	0	0.0493	0.0526	0.053	0.0495	0.0629	0.0242	0.0384	0.0955
500	1	0	0.0493	0.0498	0.0513	0.0525	0.0648	0.0226	0.0366	0.0898
1000	1	0	0.0484	0.05	0.0538	0.0482	0.0623	0.0201	0.0387	0.0857
2500	1	0	0.0481	0.052	0.0481	0.0491	0.0633	0.0208	0.0403	0.086
5000	1	0	0.0465	0.0497	0.0486	0.0517	0.0658	0.0224	0.0429	0.0854
10000	1	0	0.0505	0.048	0.0495	0.0514	0.0625	0.0218	0.0393	0.0797
20000	1	0	0.0498	0.0517	0.0503	0.0499	0.0579	0.0226	0.0439	0.0845

Πίνακας 4: Πιθανότητα σφάλματος τύπου I – εκθετική, λογιστική και γάμμα κατανομή (Βλέπε Παράρτημα_Κώδικας Table 4)

Τώρα στρέφουμε την προσοχή μας στην ισχύ αυτού του στατιστικού ελέγχου με δείγματα από διάφορες κατανομές με κοντές και μακριές ουρές. Εκτός από το να ανιχνεύσουμε τη δυνατότητα εντόπισης κοντών ή μακριών ουρών, είναι εξίσου σημαντικό να μελετήσουμε την πιθανότητα σοβαρού λάθους κατάταξης. Ο Πίνακας 5 δίνει πιθανότητες κατάταξεις για διάφορες κατανομές Pareto(γ) {μακριά ουρά} με συνάρτηση επιβίωσης της μορφής $\bar{F}(x) = \frac{1}{1+x^\gamma}$, $x > 0$. Παρατηρούμε μεγάλη ισχύ στις περιπτώσεις Pareto(1), Pareto(2) και μειωμένη ισχύ ελέγχου όσο αυξάνεται η παράμετρος γ , πράγμα αναμενόμενο αφού τότε η $\bar{F}(\ln x)$ τείνει να γίνει ταχείας κύμανσης (Rojo (1996)).

n	P(1) S	P(1) L	P(2) S	P(2) L	P(5) S	P(5) L	P(10)S	P(10)L
10	0.0159	0.5231	0.0753	0.1879	0.6593	0.0074	0.992	0
50	0.0056	0.8062	0.023	0.3711	0.3387	0.0281	0.9773	0.0001
100	0.002	0.8788	0.0188	0.463	0.2359	0.0447	0.96	0.0004
250	0.0008	0.9462	0.0106	0.5767	0.126	0.0683	0.9151	0.0014
500	0.0006	0.9694	0.0078	0.6666	0.0874	0.1003	0.8658	0.0022
1000	0.0003	0.9812	0.0049	0.7478	0.0645	0.1346	0.7887	0.0026
2500	0.0001	0.9934	0.0044	0.818	0.0399	0.1766	0.6668	0.0055
5000	0.0001	0.9959	0.0032	0.8608	0.032	0.2366	0.5358	0.0072
10000	0	0.998	0.0023	0.894	0.0278	0.2724	0.4065	0.0103
20000	0	0.999	0.0017	0.9205	0.022	0.3262	0.2944	0.0179

**Πίνακας 5: Πιθανότητες ταξινόμησης στην περίπτωση της Pareto(γ)
(Βλέπε Παράρτημα_Κώδικας Table 5)**

Ο Πίνακας 6 παρέχει τις πιθανότητες ταξινόμησης για Weibull(γ) κατανομές με συνάρτηση επιβίωσης $\bar{F}(x) = e^{-x^\gamma}$, $x > 0$. Όπως είδαμε προηγουμένως, η κατανομή Weibull έχει μακριά ουρά για $0 < \gamma < 1$ και κοντή ουρά για $\gamma > 1$. Συμπεραίνουμε ότι η ισχύς του ελέγχου μειώνεται όσο πλησιάζουμε στο $\gamma = 1$ (όριο).

n	W(5) S	W(5) L	W(2) S	W(2) L	W(1/2) S	W(1/2) L
10	0.9945	0	0.3819	0.0006	0.0254	0.3991
50	1	0	0.2725	0	0.0103	0.515
100	1	0	0.259	0	0.0112	0.5429
250	1	0	0.2497	0	0.0093	0.5871
500	1	0	0.2416	0	0.0065	0.6226
1000	1	0	0.2504	0	0.0075	0.6471
2500	1	0	0.2414	0	0.0053	0.6634
5000	1	0	0.2413	0	0.0075	0.6817
10000	1	0	0.242	0	0.0051	0.6958
20000	1	0	0.2458	0	0.0059	0.7171

**Πίνακας 6: Πιθανότητες ταξινόμησης στην περίπτωση της Weibull(γ)
(Βλέπε Παράρτημα_Κώδικας Table 6)**

Ο Πίνακας 7 απεικονίζει την ισχύ του ελέγχου έναντι των κατανομών $U(0,1)$, Inverse Weibull($\mu=0$, $\sigma=0$, $\xi=-1$) και τυποποιημένης κανονικής. Στις 2 πρώτες περιπτώσεις υπάρχει τελειότητα στον εντοπισμό κοντής ουράς μέσα σε δείγμα οποιουδήποτε μεγέθους. Στην περίπτωση όμως της κανονικής κατανομής $N(0,1)$ για δείγματα μεγαλύτερα του 10 απαιτούνται δείγματα μεγέθους 5000, ώστε να έχουμε 500 (10%) περιπτώσεις, όπου η προσομοίωση απορρίφθηκε ως ES medium και κατατάχθηκε στην κατηγορία ES short.

n	U(0,1) S	U(0,1) L	Inverse Weibull S	Inverse Weibull L	Normal S	Normal L
10	1	0	1	0	0.1919	0.1480
50	1	0	1	0	0.0508	0.0148
100	1	0	1	0	0.0629	0.0066
250	1	0	1	0	0.0698	0.0045
500	1	0	1	0	0.084	0.0013
1000	1	0	1	0	0.0856	0.0013
2500	1	0	1	0	0.0905	0.0009
5000	1	0	1	0	0.1007	0.0003
10000	1	0	1	0	0.0996	0.0008
20000	1	0	1	0	0.1085	0.0009

Πίνακας 7: Πιθανότητες ταξινόμησης στην περίπτωση ES short κατανομών (Βλέπε Παράρτημα_Κώδικας Table 7)

Τέλος, στον Πίνακα 8 φαίνεται το επίπεδο ισχύος στην περίπτωση συνήθων κατανομών μακριάς ουράς. Το ποσοστό επιτυχούς κατάταξης για την Lognormal είναι σχετικά μικρό, ενώ μεγαλώνει καθώς πηγαίνουμε στην Student και στην Cauchy.

n	Lnorm S	Lnorm L	t(3) S	t(3) L	Cauchy S	Cauchy L
10	0.0555	0.1767	0.1349	0.2297	0.0165	0.4742
50	0.0242	0.2863	0.0226	0.2559	0.0061	0.7214
100	0.021	0.3354	0.0264	0.2985	0.0039	0.8257
250	0.0201	0.4047	0.0192	0.3753	0.0012	0.9036
500	0.0157	0.4437	0.0129	0.4321	0.0016	0.943
1000	0.013	0.4948	0.0118	0.496	0.0006	0.9688
2500	0.0109	0.5463	0.0102	0.582	0.0001	0.9846
5000	0.0086	0.5845	0.0083	0.6354	0.0003	0.9917
10000	0.0082	0.616	0.0073	0.6899	0.0001	0.9949
20000	0.0076	0.6614	0.0049	0.7234	0.0002	0.9978

**Πίνακας 8: Πιθανότητες ταξινόμησης στην περίπτωση ES long κατανομών
(Βλέπε Παράρτημα_Κώδικας Table 8)**

ΚΕΦΑΛΑΙΟ 7: ΙΣΧΥΡΟΤΕΡΑ ΕΠΙΠΕΔΑ ΕΛΕΓΧΟΥ ΜΕΣΩ ΔΙΑΣΠΑΣΗΣ ΚΑΙ ΟΜΑΔΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ

Στο Κεφάλαιο 6 έγινε παρουσίαση της μεθόδου ES (Extreme Spacing) με σκοπό την κατάταξη των ουρών συναρτήσεων κατανομών. Είδαμε, όμως, ότι υπάρχει αδυναμία διαχωρισμού μεταξύ Lognormal και Normal κατανομής με την εκθετική (Exponential). Σε αυτό το κεφάλαιο θα μελετήσουμε την τεχνική της διάσπασης του δείγματος σε υποδείγματα ίδιου μεγέθους, την εύρεση της στατιστικής συνάρτησης ελέγχου για κάθε υποδείγμα και τελικά τον συνδυασμό όλων των στατιστικών συναρτήσεων ελέγχου σε μία τελική (Rojo (2010)).

Κάθε ομάδα δεδομένων (υποδείγμα) έχει μέγεθος $m=n/k$, όπου k ο αριθμός των διαφορετικών υποδειγμάτων. Κάτω από τη μηδενική υπόθεση της ES medium-tailed περίπτωσης, το άθροισμα των στατιστικών συναρτήσεων της κάθε ομάδας ακολουθεί ασυμπτωτικά την κατανομή $\text{Gamma}(k,1)$. Άρα, αν $\sum_{j=1}^k TS_j < q_{\text{gamma}}(\alpha, k, 1)$ απορρίπτουμε την μηδενική έναντι της ES short-tailed υπόθεσης. Εάν, όμως, $\sum_{j=1}^k TS_j > q_{\text{gamma}}(1 - \alpha, k, 1)$ απορρίπτουμε την μηδενική έναντι της ES long-tailed υπόθεσης.

Ο Πίνακας 9 απεικονίζει την πιθανότητα σφάλματος τύπου I με δείγματα από τις $\text{Exp}(1)$ και $\text{Lgis}(0.7,1)$ χρησιμοποιώντας διάφορα μεγέθη δείγματος και διάφορους αριθμούς ομάδων δεδομένων (blocks). Η κατάσταση βελτιώνεται για την $\text{Exp}(1)$ για 5 ή 10 blocks, ενώ δεν υπάρχει βελτίωση για τη λογιστική κατανομή.

blocks	E500S	E500L	L500S	L500L	E5000S	E5000L	L5000S	L5000L
1	0.0505	0.049	0.0637	0.0208	0.0485	0.0511	0.0726	0.0181
5	0.0516	0.0596	0.1109	0.0109	0.0503	0.0522	0.1187	0.0094
10	0.0477	0.0638	0.1435	0.0098	0.051	0.0497	0.1689	0.0037
25	0.0473	0.0975	0.1736	0.0115	0.0471	0.0596	0.3083	0.0009
50	0.0426	0.1399	0.0767	0.0932	0.0473	0.0618	0.4914	0.0004

blocks	E20000S	E2000L	L20000S	L20000L
1	0.0512	0.0483	0.0609	0.0214
5	0.0473	0.0507	0.1177	0.0099
10	0.0515	0.0484	0.1743	0.0039
25	0.0471	0.0533	0.3137	0.0007
50	0.0451	0.0555	0.5241	0
100	0.0439	0.0594	0.7871	0

Πίνακας 9: Πιθανότητα σφάλματος τύπου I για δείγματα μεγέθους 500, 5000 και 20000 και για διάφορους αριθμούς blocks (Βλέπε Παράρτημα_Κώδικας Table 9)

Οι Πίνακες 10-12 συνοψίζουν τη βελτίωση της ισχύος του ελέγχου εντοπισμού μία κοντής ή μακρίας ουράς για 5 ή 10 blocks σε όλα τα μεγέθη δείγματος με εξαίρεση την κατανομή Pareto($\gamma=5$). Στην περίπτωση της Pareto($\gamma=5$) είμαστε αναγκασμένοι να χρησιμοποιήσουμε μεγάλο δείγμα ($n=20000$) για να διακρίνουμε αυτή τη βελτίωση, διότι έχουμε σχετικά μεγάλη τιμή για την παράμετρο γ .

$N=500$

blocks	NormS	NormL	LnormL	Par(5)L	Weib(2)S
1	0.1932	0	0.4436	0.095	0.2471
5	0.8524	0	0.7563	0.0433	0.9544
10	0.9885	0	0.864	0.0113	0.9995
25	0.9998	0	0.9398	0.0001	1

Πίνακας 10: Πιθανότητες ταξινόμησης στην περίπτωση ES long ή short κατανομών (Βλέπε Παράρτημα_Κώδικας Tables 10-12)

N=5000

blocks	NormS	LnormL	Par(5)L	Weib(2)S
1	0.2191	0.579	0.2315	0.2406
5	0.2248	0.936	0.253	0.9681
10	0.9994	0.9874	0.2052	1
20	1	0.9996	0.0979	1
50	1	1	0.0051	1

Πίνακας 11: Πιθανότητες ταξινόμησης στην περίπτωση ES long ή short κατανομών (Βλέπε Παράρτημα_Κώδικας Tables 10-12)

N=20000

blocks	NormS	LnormL	Par(5)L	Weib(2)S
1	0.225	0.6538	0.3267	0.2417
5	0.9615	0.9728	0.4862	0.9701
10	0.9998	0.9975	0.5233	0.9999
20	1	1	0.4871	1
50	1	1	0.2549	1

Πίνακας 12: Πιθανότητες ταξινόμησης στην περίπτωση ES long ή short κατανομών (Βλέπε Παράρτημα_Κώδικας Tables 10-12)

Παρατηρούμε επίσης ότι η ισχύς του ελέγχου δεν αυξάνεται υπερβολικά μετά τα 10 blocks, και αυτό οφείλεται στο ότι πλέον το μέγεθος του υποδείγματος μικραίνει υπερβολικά με αποτέλεσμα να έχουμε σημαντικό βαθμό μεροληψίας

στην εκτίμηση της ποσότητας $\frac{-\ln \bar{F}_n(\ln X_{(n)})}{\ln (n)}$.

ΚΕΦΑΛΑΙΟ 8: ΕΛΕΓΧΟΣ BRYSON

Ο Bryson (1974) πρότεινε μία διαδικασία για έλεγχο της υπόθεσης εκθετικής κατανομής έναντι κατανομών μακριών ουρών με γραμμική μέση τιμή υπολειπόμενου χρόνου ζωής, αλλά και έναντι κατανομών κοντών ουρών. Αυτό το τεστ βασίζεται στη στατιστική συνάρτηση:

$$T^* = \frac{\bar{X}X_{(n)}}{(n-1)\bar{X}_{GA}^2} \quad (19)$$

όπου:

$$\bar{X}_{GA} = \left(\prod_{i=1}^n (X_i + A_n) \right)^{\frac{1}{n}}$$

$$A_n = \frac{X_{(n)}}{n-1}.$$

Από την (19) έπεται ότι η ασυμπτωτική συμπεριφορά της νέας στατιστικής συνάρτησης εξαρτάται από τη συμπεριφορά του $X_{(n)}$. Μειονέκτημα σε αυτή την περίπτωση είναι ότι αυτή η ασυμπτωτική συμπεριφορά είναι άγνωστη, αλλά μέσω προσομοίωσης παρουσιάζουμε στους Πίνακες 13-16 τα τεταρτημόρια του Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape = 2,1,1/2, Scale = 1 και από τη Logistic(0.7,1) κατανομή. Σκοπός μας είναι να εξετάσουμε την ισχύ του ελέγχου έναντι μακριών (Weibull(1/2)) και κοντών ουρών (Weibull(2)). Αυτός είναι και ο λόγος που μέσω προσομοίωσης βρήκαμε

και καταγράψαμε και τα τεταρτημόρια των (Weibull(1/2, 2)) κατανομών στους Πίνακες 17,18.

n	2.5%	5%	95%	97.5%
50	0.06105407	0.06426653	0.12392225	0.13210505
100	0.03848229	0.04053895	0.07534090	0.08091435
300	0.01692806	0.01779872	0.03027503	0.03309065
500	0.01159142	0.01198328	0.01955472	0.02088534

**Πίνακας 13: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape = 2 , Scale = 1
(Βλέπε Παράρτημα_Κώδικας Bryson, Quantiles_Tables 13-15)**

n	2.5%	5%	95%	97.5%
50	0.1013047	0.1079076	0.2340258	0.2530156
100	0.07545713	0.07955459	0.16455284	0.17792716
300	0.03784540	0.03979158	0.07639708	0.08209742
500	0.02740529	0.02860571	0.05149976	0.05428714

**Πίνακας 14: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape = 1 , Scale = 1
(Βλέπε Παράρτημα_Κώδικας Bryson, Quantiles_Tables 13-15)**

n	2.5%	5%	95%	97.5%
50	0.2166569	0.2297596	0.4569863	0.4806051
100	0.1749157	0.1900029	0.3875425	0.4084017
300	0.1195284	0.1268536	0.2443472	0.2536978
500	0.1296872	0.1770937	Inf	Inf

**Πίνακας 15: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Gamma κατανομή με Shape = 1/2 , Scale = 1
(Βλέπε Παράρτημα_Κώδικας Bryson, Quantiles_Tables 13-15)**

n	2.5%	5%	95%	97.5%
50	0.01666583	0.02378389	0.14703876	0.16938170
100	0.01634496	0.02026398	0.07640472	0.08347893
300	0.01012007	0.01097531	0.02463865	0.02775749
500	0.007185861	0.007623876	0.015818078	0.017160421

Πίνακας 16: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Logistic(0.7,1) κατανομή (Βλέπε Παράρτημα_Κώδικας Bryson, Quantiles_Logistic_Table 16)

n	2.5%	5%	95%	97.5%
50	0.04457463	0.04681970	0.07388387	0.07771579
100	0.02617748	0.02706633	0.04116554	0.04296127
300	0.01064480	0.01091023	0.01522705	0.01592211
500	0.006791429	0.006970442	0.009447375	0.009762864

Πίνακας 17: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Weibull(2) κατανομή (Βλέπε Παράρτημα_Κώδικας Bryson, Quantiles_Weibull_Tables 17-18)

n	2.5%	5%	95%	97.5%
50	0.3452717	0.3709733	0.6914648	0.7193314
100	0.3437794	0.3666789	0.6942977	0.7284329
300	0.3121409	0.3314666	0.6208897	0.6483844
500	0.2710508	0.2899162	0.5567837	0.5778312

Πίνακας 18: Τεταρτημόρια Bryson τεστ για διάφορα μεγέθη δείγματος από τη Weibull(1/2) κατανομή (Βλέπε Παράρτημα_Κώδικας Bryson, Quantiles_Weibull_Tables 17-18)

Συμπεραίνουμε ότι το Bryson τεστ θα απορρίψει (με πιθανότητα κοντά στο 1) την υπόθεση μεσαίας ουράς έναντι κοντής ουράς, όταν η κατανομή μεσαίας ουράς είναι η Gamma(2,1) και η κατανομή κοντής ουράς είναι η Weibull(2). Αυτό ισχύει, διότι το 97,5% εκατοστημόριο της Gamma(2,1) είναι μεγαλύτερο του 2,5% εκατοστημορίου της Weibull(2) για όλα τα μεγέθη δείγματος. Άρα θα έχουμε ένα τεράστιο σφάλμα τύπου I και ένα ελάχιστο σφάλμα τύπου II χωρίς να μπορούμε να το χειριστούμε διαφορετικά!!

Το ίδιο ισχύει για την υπόθεση μεσαίας ουράς έναντι μακριάς ουράς, όταν η κατανομή μεσαίας ουράς είναι η $\text{Gamma}(1/2,1)$ και η κατανομή μακριάς ουράς είναι η $\text{Weibull}(1/2)$. Αυτό ισχύει, διότι το 97,5% εκατοστημόριο της $\text{Gamma}(1/2,1)$ είναι μεγαλύτερο του 2,5% εκατοστημορίου της $\text{Weibull}(1/2)$ για $n \leq 100$. Άρα θα έχουμε ένα τεράστιο σφάλμα τύπου I και ένα ελάχιστο σφάλμα τύπου II χωρίς να μπορούμε να το χειριστούμε διαφορετικά!!

Στην περίπτωση, όμως, της $\text{Logistic}(0.7,1)$ το 97,5% εκατοστημόριο της $\text{Logistic}(0.7,1)$ είναι μικρότερο του 2,5% εκατοστημορίου της $\text{Weibull}(1/2)$ για όλα τα μεγέθη δείγματος. Άρα έχουμε την άνεση να διαχειριστούμε το trade-off μεταξύ των στατιστικών σφαλμάτων τύπου I, II.

ΚΕΦΑΛΑΙΟ 9: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΟΥΡΑΣ ΠΡΑΓΜΑΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Ο Πίνακας 19 απεικονίζει την ετήσια παροχή του φράγματος Μαραθώνα σε κυβικά εκατόμετρα από το 1994 έως και το 2013. Σκοπός της ανάλυσής μας είναι να δούμε τη μορφή της δεξιάς αλλά και της αριστερής ουράς των δεδομένων μας, ώστε να προσφέρουμε καλύτερη διαχείριση του συγκεκριμένου υδάτινου πόρου.

Έτη	Παροχή (hm ³)
1994-1995	27,40
1995-1996	17,35
1996-1997	16,20
1997-1998	43,91
1998-1999	23,59
1999-2000	5,52
2000-2001	2,69
2001-2002	58,16
2002-2003	61,84
2003-2004	24,76
2004-2005	15,79
2005-2006	19,15
2006-2007	8,05
2007-2008	5,37
2008-2009	14,43
2009-2010	15,34
2010-2011	10,41
2011-2012	6,05
2012-2013	30,16

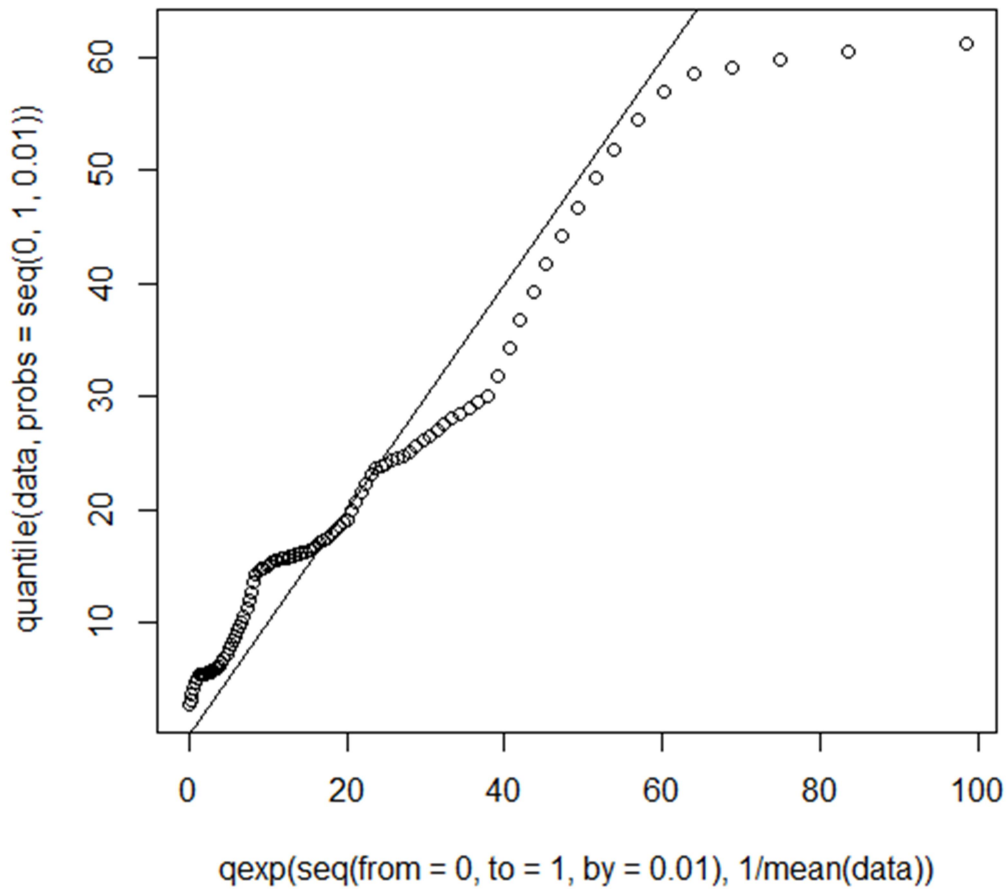
**Πίνακας 19: Υδρολογικό Έτος(Φράγμα Μαραθώνα) Παροχή (hm³)
(Βλέπε Σπόρος Γεωργιάς (2014), «Εκτίμηση Οικολογικής Παροχής από το Φράγμα του Μαραθώνα και Διαχείριση Διατιθέμενων Υδάτων», Διπλωματική Εργασία)**

Και πάλι χρησιμοποιούμε τη στατιστική συνάρτηση (20). Όμως δεν παίρνουμε ως δεδομένο την ασυμπτωτική της συμπεριφορά, λόγω του μικρού δείγματος που έχουμε. Αντιθέτως θα εφαρμόσουμε μέθοδο Bootstrap (Καρλής (2008)) για τον έλεγχο της υπόθεσης μεσαίας ουράς έναντι κοντής ή μακριάς.

$$T_n = \frac{-\ln \bar{F}_n(\ln X_{(n)})(X_{(n)} - X_{(n-1)})}{\ln X_{(n)}}. \quad (20)$$

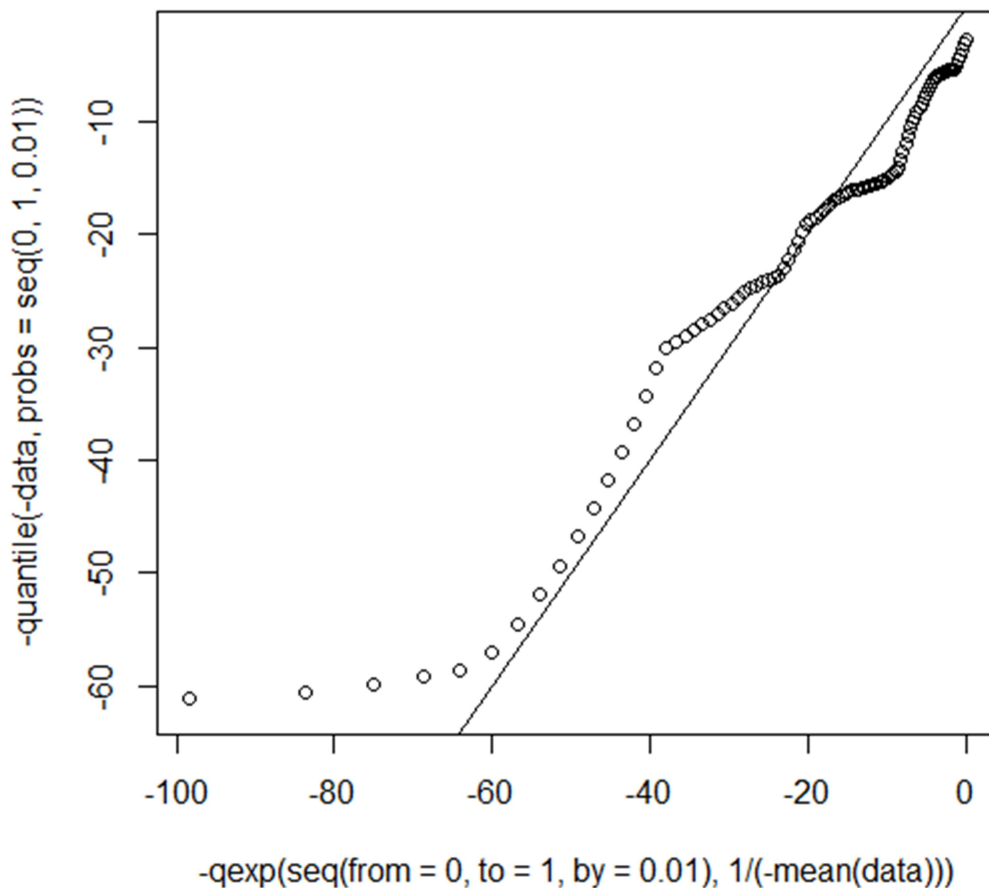
Με άλλα λόγια, προσπαθούμε να αναπαράγουμε την κατανομή, από την οποία προήλθαν τα δεδομένα μας μέσω των ιδίων των δεδομένων με τη λογική των διατάξεων με επανάληψη. Αυτό προσδίδει ισάξια πιθανότητα εμφάνισης ανά δεδομένο σε κάθε δείγμα.

Ο έλεγχος Bootstrap (Βλέπε Παράρτημα_Κώδικας_Realdata_Chapter 9, Bootstrap(right tail)) φανερώνει μέσω των τιμών p-value(0.0271,0.5717), ότι απορρίπτουμε την υπόθεση μεσαίας δεξιάς ουράς έναντι της κοντής δεξιάς ουράς, ενώ δεν μπορούμε να απορρίψουμε την υπόθεση μεσαίας δεξιάς ουράς έναντι μακριάς δεξιάς ουράς ($\alpha=0.05$). Αυτό το αποτέλεσμα συμφωνεί και διαισθητικά με το Γράφημα 3.



Γράφημα 3: Εκθετικό Q-Q Plot
(Βλέπε Παράρτημα_Κώδικας_Realdata_Chapter 9, right tail)

Με σκοπό την ανάλυση και της αριστερής ουράς των δεδομένων μας, βάζουμε αρνητικό πρόσημο στις μετρήσεις μας και προκύπτει το Γράφημα 4.



**Γράφημα 4: Εκθετικό Q-Q Plot(αρνητικές τιμές)
(Βλέπε Παράρτημα_Κώδικας_Realdata_Chapter 9, left tail)**

Έπειτα μεταφέρουμε τις αρνητικές μετρήσεις μας στον θετικό ημιάξονα προσθέτοντας σε όλες τις αρνητικές μετρήσεις την απόλυτη τιμή της μικρότερης αλγεβρικά τιμής.

Ο έλεγχος Bootstrap (Βλέπε Παράρτημα_Κώδικας_Realdata_Chapter 9, Bootstrap(left tail)) φανερώνει μέσω των τιμών p-value(0.3151,0.2801), ότι δεν μπορούμε να απορρίψουμε την υπόθεση μεσαίας αριστερής ουράς έναντι της κοντής αριστερής ουράς, ενώ δεν μπορούμε να απορρίψουμε επίσης την υπόθεση μεσαίας αριστερής ουράς έναντι μακριάς αριστερής ουράς ($\alpha=0.05$). Αυτό το αποτέλεσμα και πάλι συμφωνεί και διαισθητικά με το Γράφημα 4.

ΠΑΡΑΡΤΗΜΑ

Αποδείξεις Θεωρημάτων-Λημμάτων

Απόδειξη Θεωρήματος 4: Υποθέτουμε ότι $\bar{F}(y)/(\bar{F}(\ln y))^\delta \rightarrow 0$ για κάποιο

$\delta > 2$, και έστω $0 < \gamma < 1/2$, $\varepsilon_n = n^{\gamma - \frac{1}{2}}$. Ορίζουμε $A_n = \{\|F_n - F\| \geq \varepsilon_n\}$ και $B_n = \{\bar{F}(\ln X_{(n)}) \leq \varepsilon_n\}$, όπου $\|F_n - F\| = \sup_x |F_n(x) - F(x)|$. Έστω

$$Z_n = \frac{-\ln \bar{F}_n(\ln X_{(n)})}{-\ln \bar{F}(\ln X_{(n)})}$$

και για $\varepsilon > 0$,

$$\begin{aligned} P(Z_n > 1 + \varepsilon) &= P(Z_n > 1 + \varepsilon, A_n) + P(Z_n > 1 + \varepsilon, A_n^c) \leq \\ &\leq 2e^{-2n\varepsilon_n^2} + P(Z_n > 1 + \varepsilon, A_n^c). \end{aligned}$$

Άρα είναι αρκετό να δείξουμε, ότι ο δεύτερος όρος της τελευταίας έκφρασης τείνει στο μηδέν όταν $n \rightarrow \infty$. Ο δεύτερος όρος φράσσεται από πάνω ως εξής:

$$\begin{aligned} P((Z_n > 1 + \varepsilon) \cap A_n^c \cap B_n) + P\left(\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) - \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} > 1 + \varepsilon\right) \cap A_n^c \cap B_n^c\right) &\leq \\ \leq P(B_n) + P\left(\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) - \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} > 1 + \varepsilon\right) \cap A_n^c \cap B_n^c\right). \end{aligned}$$

Προσέχουμε τώρα ότι :

$$\begin{aligned} P(B_n) &= P(\ln X_{(n)} \geq \bar{F}^{-1}(\varepsilon_n)) = 1 - P\left(X_{(n)} < \exp\left\{\bar{F}^{-1}\left(n^{\gamma - \frac{1}{2}}\right)\right\}\right) = \\ &= 1 - \left\{1 - \bar{F}\left(\exp\left\{\bar{F}^{-1}\left(n^{\gamma - \frac{1}{2}}\right)\right\}\right)\right\}^n. \end{aligned}$$

Αλλά $n\bar{F}\left(\exp\left\{\bar{F}^{-1}\left(n^{\gamma - \frac{1}{2}}\right)\right\}\right) = o(1)$ ως συνέπεια της υπόθεσης ότι:

$\bar{F}(y)/(\bar{F}(\ln y))^\delta \rightarrow 0$ για κάποιο $\delta > 2$ όταν $y \rightarrow \infty$. Για να το δούμε αυτό, θέτουμε $k = \gamma - \frac{1}{2}$ και $u = \bar{F}^{-1}(n^k)$, έτσι ώστε $u \rightarrow \infty$ καθώς $n \rightarrow \infty$, και $n\bar{F}(\exp\{\bar{F}^{-1}(n^k)\}) = (\bar{F}(u))^{1/k} \bar{F}(e^u)$. Τελικά θέτοντας $y = e^u$, τότε η τελευταία έκφραση είναι ισοδύναμη με $\bar{F}(y)/(\bar{F}(\ln y))^{-1/k} \rightarrow \infty$ καθώς $y \rightarrow \infty$ αφού $-1/k > 2$. Επομένως, $P(B_n) \rightarrow 0$ καθώς $n \rightarrow \infty$.

Απομένει να αποδείξουμε ότι:

$$P\left(\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) - \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} > 1 + \varepsilon\right) \cap A_n^c \cap B_n^c\right) \rightarrow 0$$

για $n \rightarrow \infty$.

Γράφουμε (Taylor):

$$-\ln(\bar{F}(\ln X_{(n)}) - \varepsilon_n) = -\ln(\bar{F}(\ln X_{(n)})) + \frac{\varepsilon_n}{1 - \xi_n},$$

όπου $\bar{F}(\ln X_{(n)}) < 1 - \xi_n < \bar{F}(\ln X_{(n)}) + \varepsilon_n$, έτσι ώστε για $0 < \alpha < 1$ θέτοντας $C_n = A_n^c \cap B_n^c$ να ισχύει:

$$\begin{aligned} & P\left(\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) - \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} > 1 + \varepsilon\right) \cap C_n\right) = \\ & = P\left(\left(1 + \frac{\varepsilon_n}{(1 - \xi_n)(-\ln \bar{F}(\ln X_{(n)}))} > 1 + \varepsilon\right) \cap C_n\right) \leq \\ & \leq P\left(\left(\frac{\varepsilon_n}{(1 - \xi_n)(-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon\right)\right) \leq \\ & \leq P\left(\left\{\frac{\varepsilon_n}{\bar{F}(\ln X_{(n)})(-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon\right\} \cap \left\{n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) > 1 + \alpha\right\}\right) \\ & + P\left(n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) < 1 + \alpha\right) = \end{aligned}$$

$$= P \left(\left\{ \frac{1}{\left(n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) \right) (-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon \right\} \cap \right. \\ \left. \cap \left\{ n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) > 1 + \alpha \right\} \right) +$$

$$+ P \left(n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) < 1 + \alpha \right) \leq$$

$$\leq P \left(\frac{1}{a(-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon \right) + P \left(n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) < 1 + \alpha \right).$$

Αφού $\alpha > 0$ και $-\ln \bar{F}(\ln X_{(n)}) \xrightarrow{a.s.} \infty$, ο πρώτος όρος στη δεξιά πλευρά της τελευταίας ανισότητας τείνει στο μηδέν. Για τον δεύτερο όρο παρατηρούμε ότι:

$$P \left(n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) < 1 + \alpha \right) = P \left(X_{(n)} > \exp \left\{ \bar{F}^{-1} \left(cn^{\gamma-\frac{1}{2}} \right) \right\} \right) = \\ = 1 - \left\{ 1 - \bar{F} \left(\exp \left\{ \bar{F}^{-1} \left(cn^{\gamma-\frac{1}{2}} \right) \right\} \right) \right\}^n$$

όπου $c = 1 + \alpha$.

Όπως και προηγουμένως, $\bar{F}(y)/(\bar{F}(\ln y))^\gamma \rightarrow 0$ για κάποιο $\gamma > 2$ υποδηλώνει

ότι $n \bar{F} \left(\exp \left\{ \bar{F}^{-1} \left(cn^{\gamma-\frac{1}{2}} \right) \right\} \right) \rightarrow 0$ όταν $n \rightarrow \infty$. Επομένως

$$P \left(n^{\frac{1}{2}-\delta} \bar{F}(\ln X_{(n)}) < 1 + \alpha \right) \rightarrow 0$$

και

$$P \left(\frac{-\ln \bar{F}_n(\ln X_{(n)})}{-\ln \bar{F}(\ln X_{(n)})} > 1 + \varepsilon \right) \rightarrow 0.$$

Την παρακάτω περίπτωση τη χειριζόμαστε με τον ίδιο τρόπο:

$$P\left(\frac{-\ln \bar{F}_n(\ln X_{(n)})}{-\ln \bar{F}(\ln X_{(n)})} < 1 - \varepsilon\right) = P(Z_n < 1 - \varepsilon).$$

Θεωρούμε τώρα την πιθανότητα:

$$\begin{aligned} P(Z_n < 1 - \varepsilon) &= P((Z_n < 1 - \varepsilon) \cap A_n) + P((Z_n < 1 - \varepsilon) \cap A_n^c) \leq \\ &\leq 2e^{-2n\varepsilon_n^2} + P\left(\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) + \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} < 1 - \varepsilon\right) \cap A_n^c\right) \leq \\ &\leq 2e^{-2n\varepsilon_n^2} + P\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) + \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} < 1 - \varepsilon\right). \end{aligned}$$

Όπως και πριν, γράφουμε:

$$-\ln(\bar{F}(\ln X_{(n)}) + \varepsilon_n) = -\ln(\bar{F}(\ln X_{(n)})) - \frac{\varepsilon_n}{\xi_n},$$

όπου $\bar{F}(\ln X_{(n)}) < \xi_n < \bar{F}(\ln X_{(n)}) + \varepsilon_n$. Τότε:

$$\begin{aligned} P\left(\frac{-\ln(\bar{F}(\ln X_{(n)}) + \varepsilon_n)}{-\ln \bar{F}(\ln X_{(n)})} < 1 - \varepsilon\right) &= \\ &= P\left(\left(1 - \frac{\varepsilon_n}{\xi_n(-\ln \bar{F}(\ln X_{(n)}))} < 1 - \varepsilon\right)\right) = \\ &= P\left(\frac{\varepsilon_n}{\xi_n(-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon\right) \leq \\ &\leq P\left(\frac{\varepsilon_n}{\bar{F}(\ln X_{(n)})(-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon\right) = \\ &= P\left(\frac{1}{n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)})(-\ln \bar{F}(\ln X_{(n)}))} > \varepsilon\right) \leq \\ &\leq P\left(\frac{1}{-\ln \bar{F}(\ln X_{(n)})} > \varepsilon, n^{\frac{1}{2}-\gamma} \bar{F}(\ln X_{(n)}) > 1\right) + \end{aligned}$$

$$\begin{aligned}
& +P\left(n^{\frac{1}{2}-\gamma}\bar{F}(\ln X_{(n)}) < 1\right) \leq \\
& \leq P\left(\frac{1}{-\ln\bar{F}(\ln X_{(n)})} > \varepsilon\right) + P\left(n^{\frac{1}{2}-\gamma}\bar{F}(\ln X_{(n)}) < 1\right).
\end{aligned}$$

Όμοια επιχειρήματα με προηγουμένως οδηγούν στο αποτέλεσμα, ότι και οι δύο όροι στη δεξιά πλευρά της παραπάνω ανισότητας τείνουν στο μηδέν καθώς $n \rightarrow \infty$. Συνεπώς:

$$\frac{-\ln\bar{F}_n(\ln X_{(n)})}{-\ln\bar{F}(\ln X_{(n)})} \xrightarrow{P} 1.$$

Απόδειξη Λήμματος 5: Χωρίς απώλεια της γενικότητας υποθέτουμε, ότι η F είναι μία συνάρτηση κατανομής ζωής. Η απόδειξη προκύπτει εύκολα γράφοντας, μετά από ανάπτυγμα Taylor ενός βήματος:

$$-\ln\left\{\frac{\bar{F}(t+x)}{\bar{F}(x)}\right\} = \int_x^{t+x} r_F(u)du = t * r_f(\xi) \quad (21)$$

όπου $x < \xi < x+t$.

Για την (i) περίπτωση είναι $\frac{\bar{F}(t+x)}{\bar{F}(x)} \rightarrow 0$, για όλα τα t καθώς $x \rightarrow \infty$. Άρα η (21) είναι ισοδύναμη με το: $t * r_f(\xi) \rightarrow \infty$, για $x \rightarrow \infty$ και για όλα τα $t > 0$.

Στη (ii) περίπτωση, $\bar{F}(\ln x) = x^{-\theta}l(x)$ για κάποιο $\theta > 0$ και για κάποια συνάρτηση βραδείας κύμανσης $l(x)$. Χρησιμοποιώντας την (12) είναι προφανές ότι, από κανόνα L'Hopital,

$$\lim_{y \rightarrow \infty} r_F(y) = \lim_{y \rightarrow \infty} \frac{-\ln\bar{F}(\ln y)}{\ln y} = \theta - \lim_{y \rightarrow \infty} \frac{-\ln l(y)}{\ln y}.$$

Το αποτέλεσμα προκύπτει άμεσα αφού για συνάρτηση βραδείας κύμανσης ισχύει $\ln l(y)/\ln y \rightarrow 0$. Το αντίστροφο προκύπτει άμεσα από την (21) με το να πάρουμε το όριο για $x \rightarrow \infty$.

Η περίπτωση (iii) επίσης προκύπτει άμεσα από την (21), αφού όταν η F έχει μακριά ουρά τότε αυτό ισοδυναμεί με τις εκφράσεις στην (21) να συγκλίνουν στο μηδέν.

Απόδειξη Λήμματος 6:

(i) περίπτωση: έστω $\bar{F}(\ln x) = c(x) \exp \int_1^x \frac{\varepsilon(t)}{t} dt$, με $c(x) \rightarrow c > 0$ και $\varepsilon(x) \rightarrow 0$ για $x \rightarrow \infty$. Αφού η F έχει μακριά ουρά, $r_F(x) \rightarrow 0$ για $x \rightarrow \infty$. Συνεπώς,

$$\frac{r_F(\ln x)}{x} = -\frac{c'(x)}{c(x)} - \frac{\varepsilon(x)}{x}.$$

Άρα: $r_F(\ln x) = -\frac{xc'(x)}{c(x)} - \varepsilon(x) \rightarrow 0$, και προκύπτει ότι $\frac{xc'(x)}{c(x)} \rightarrow 0$. Στη συνέχεια γράφοντας $\varepsilon^*(t) = \varepsilon(t) + tc'(t)/c(t)$ παίρνουμε ότι $\bar{F}(\ln x) = \exp \int_1^x \frac{\varepsilon^*(t)}{t} dt$ και προκύπτει το αποτέλεσμα αφού $r_F(\ln x) = -\varepsilon^*(x)$.

Η απόδειξη της (ii) περίπτωσης είναι παρόμοια της πρώτης με το να γράψουμε για την F με την κοντή ουρά:

$$\bar{F}(\ln x) = c(x) \exp \int_1^x \frac{z(t)}{t} dt$$

με $c(x) \rightarrow c > 0$ και $z(t) \rightarrow -\infty$ για $x \rightarrow \infty$ και χρησιμοποιώντας από το Λήμμα 5, ότι $r_F(x) \rightarrow \infty$.

Για να αποδείξουμε την (iii) περίπτωση, επισημαίνουμε ότι η (17) ισχύει αν και μόνο αν

$$-\ln \bar{F}(y) + \delta \ln \bar{F}(\ln y) \rightarrow \infty, \text{ καθώς } y \rightarrow \infty. \quad (22)$$

$$-\ln\bar{F}(y) + \delta\ln\bar{F}(\ln y) = -\ln\bar{F}(y) \left(1 - \delta \left(\frac{\ln\bar{F}(\ln y)}{\ln\bar{F}(y)} \right) \right)$$

Παρατηρούμε ότι:

$$\lim_{y \rightarrow \infty} \frac{\ln\bar{F}(\ln y)}{\ln\bar{F}(y)} = \lim_{y \rightarrow \infty} \frac{r_F(\ln y)}{y r_F(y)} = \lim_{y \rightarrow \infty} \frac{\varepsilon(y)}{y \varepsilon(e^y)} = \lim_{y \rightarrow \infty} \frac{z(y)}{y z(e^y)}, \quad (23)$$

όπου το τρίτο και τέταρτο όριο αντιστοιχούν στις περιπτώσεις των κοντών και μακρυνών ουρών αντιστοίχως. Το αποτέλεσμα για τις κατανομές μεσαίας ουράς προκύπτει άμεσα από το Λήμμα 5, αφού σε αυτή την περίπτωση ισχύει $\lim_{y \rightarrow \infty} \frac{r_F(\ln y)}{y r_F(y)} = 0$. Για τις κατανομές κοντής και μακριάς ουράς τα αποτελέσματα προκύπτουν από τις υποθέσεις.

Απόδειξη Θεωρήματος 7: Αφού η F έχει κοντή ουρά, $X_{(n)} - X_{(n-1)} \xrightarrow{a.s.} 0$ και η $h(x) = -\ln\bar{F}(\ln x)$ είναι συνάρτηση ομαλής κύμανσης με δείκτη γ . Αυτό σημαίνει ότι:

$$h(x) = x^\gamma l(x), \quad 0 \leq \gamma \leq \infty \quad (24)$$

για κάποια συνάρτηση βραδείας κύμανσης $l(x)$. Η περίπτωση $\gamma = \infty$ αντιστοιχεί στην περίπτωση που η $h(x)$ είναι ταχείας κύμανσης. Προκύπτει ότι:

$$\bar{F}^{-1}(u) = \ln h^{-1}(-\ln u) \quad (25)$$

με h^{-1} ομαλής κύμανσης με δείκτη $1/\gamma$. Άρα, όταν $\gamma = 0$, η h^{-1} είναι ταχείας κύμανσης. Κάτω από τις υποθέσεις του Θεωρήματος 4, $\frac{\ln\bar{F}_n(\ln X_{(n)})}{\ln\bar{F}(X_{(n)})} \rightarrow 1$, και είναι αρκετό να λάβουμε υπόψη την συμπεριφορά της έκφρασης:

$$\frac{-\ln\bar{F}(\ln X_{(n)})}{\ln X_{(n)}} (X_{(n)} - X_{(n-1)}). \quad (26)$$

Αποδεικνύουμε, ότι η (26) συγκλίνει στο μηδέν κατά πιθανότητα. Οι περιπτώσεις όπου $\gamma=0$ ή ∞ προκύπτουν άμεσα από ιδιότητες των συναρτήσεων βραδείας και ταχείας κύμανσης. Η περίπτωση όπου $0 < \gamma < \infty$ παρουσιάζει τις περισσότερες τεχνικές προκλήσεις και θα εξεταστεί πρώτη. Πρέπει να θυμηθούμε, ότι μία θετική συνάρτηση g ορισμένη σε μία γειτονιά του ∞ , είναι λείας κύμανσης με δείκτη $n \in R, g \in SR_n$, εάν $H'(x) \rightarrow n, H^{(n)}(x) \rightarrow 0$ για $n > 1$ καθώς $x \rightarrow \infty$ με $H(x) = \ln(g(e^x))$.

Το ακόλουθο Θεώρημα (βλέπε Bingham, Goldie και Teugels (1987)) θα μας επιτρέψει να υποθέσουμε, χωρίς απώλεια της γενικότητας, ότι η $h(x)$ είναι λείας κύμανσης με αποτέλεσμα να ισχύει:

$$\lim_{t \rightarrow \infty} \frac{h'(t)t}{h(t)} = \gamma .$$

.

Θεώρημα: Έστω $g \in R_n$. Τότε υπάρχουν $g_1, g_2 \in SR_n$ με $g_1 \sim g_2$ και $g_1 \leq g \leq g_2$ σε μία γειτονιά του ∞ . Συγκεκριμένα, εάν $g \in R_n$, υπάρχει $g^* \in SR_n$ με $g^* \sim g$.

Έστω τότε $0 < \gamma < \infty$ και ότι η ακολουθία $U_{(1)}, U_{(2)}, \dots \dots \dots, U_{(n)}$ παριστάνει τα στατιστικά τάξεως από μία ομοιόμορφη κατανομή στο $(0,1)$. Αφού $X_{(n)} = \bar{F}^{(-1)}(1 - U_{(n)}) = \bar{F}^{(-1)}(U_{(1)})$ και $X_{(n-1)} = \bar{F}^{(-1)}(1 - U_{(n-1)}) = \bar{F}^{(-1)}(U_{(2)})$, όπου η ισότητα φανερώνει ισότητα σε κατανομή, η έκφραση (26) έχει την ίδια κατανομή με:

$$\frac{(\bar{F}^{-1}(U_{(1)}))^{\gamma} l(\bar{F}^{-1}(U_{(1)}))}{\ln \bar{F}^{-1}(U_{(1)})} (\bar{F}^{(-1)}(U_{(1)}) - \bar{F}^{(-1)}(U_{(2)})), \quad (27)$$

όπου η (27) προκύπτει από την (24) και το γεγονός ότι $h(x) = -\ln\bar{F}(\ln x)$.
Χρησιμοποιώντας ανάπτυγμα Taylor ενός βήματος, παίρνουμε

$$\begin{aligned} \bar{F}^{(-1)}(U_{(1)}) - \bar{F}^{(-1)}(U_{(2)}) &= \ln h^{-1}(-\ln U_{(1)}) - \ln h^{-1}(-\ln U_{(2)}) = \\ &= \frac{U_{(2)} - U_{(1)}}{\xi_n h^{-1}(-\ln \xi_n) h'(h^{-1}(-\ln \xi_n))} \end{aligned}$$

με $U_{(1)} < \xi_n < U_{(2)}$.

Συνεπώς, η (27) φράσσεται από πάνω από:

$$\begin{aligned} &\frac{(\bar{F}^{-1}(U_{(1)}))^{\gamma} l(\bar{F}^{-1}(U_{(1)}))}{\ln \bar{F}^{-1}(U_{(1)})} \frac{U_{(2)} - U_{(1)}}{U_{(1)}} \frac{1}{h^{-1}(-\ln \xi_n) h'(h^{-1}(-\ln \xi_n))} = \\ &= \frac{(\ln h^{-1}(-\ln U_{(1)}))^{\gamma} l(\ln h^{-1}(-\ln U_{(1)}))}{\ln \ln h^{-1}(-\ln U_{(1)}) h^{-1}(-\ln \xi_n) h'(h^{-1}(-\ln \xi_n))} \frac{U_{(2)} - U_{(1)}}{U_{(1)}} . \end{aligned} \quad (28)$$

Αφού

$$\frac{U_{(2)} - U_{(1)}}{U_{(1)}} = \frac{1}{V} - 1 \text{ (ισότητα κατανομών) ,}$$

όπου $V \sim U(0,1)$, και επειδή $\ln \ln h^{-1}(-\ln U_{(1)}) \xrightarrow{a.s.} \infty$ και από το παραπάνω Θεώρημα είναι:

$$\frac{h'(h^{-1}(-\ln \xi_n)) h^{-1}(-\ln \xi_n)}{h(h^{-1}(-\ln \xi_n))} \rightarrow \gamma > 0 . \quad (29)$$

Τότε για να δείξουμε ότι:

$$\frac{-\ln \bar{F}(\ln X_{(n)})}{\ln X_{(n)}} (X_{(n)} - X_{(n-1)}) \xrightarrow{P} 0 ,$$

είναι αρκετό να δείξουμε ότι:

$$\frac{(\ln h^{-1}(-\ln U_{(1)}))^{\gamma} l(\ln h^{-1}(-\ln U_{(1)}))}{-\ln \xi_n} \xrightarrow{P} 0. \quad (30)$$

Για να επιβεβαιώσουμε την (29), θέτουμε $h(x) = -\ln \bar{F}(\ln x)$ έτσι ώστε $h'(x) = \frac{r_F(\ln x)}{x}$ και $h^{-1}(t) = \exp\{\bar{F}^{-1}(e^{-t})\}$. Θέτοντας $t = \bar{F}^{-1}(\xi_n)$, προκύπτει ότι:

$$\begin{aligned} \frac{h'(h^{-1}(-\ln \xi_n))h^{-1}(-\ln \xi_n)}{h(h^{-1}(-\ln \xi_n))} &= \frac{r_F(t)}{-\ln \bar{F}(t)} = \frac{d}{dt} \ln(-\ln \bar{F}(t))|_{t=\bar{F}^{-1}(\xi_n)} = \\ &= \frac{d}{dt} \ln h(e^t). \end{aligned}$$

Άρα η (29) προκύπτει για $n=\gamma$. Για να αποδείξουμε την (30) την ξαναγράφουμε ως:

$$\frac{(\ln h^{-1}(-\ln U_{(1)}))^{\gamma} l(\ln h^{-1}(-\ln U_{(1)}))}{-\ln U_{(1)}} \frac{-\ln U_{(1)}}{-\ln \xi_n},$$

η οποία φράσσεται από πάνω από:

$$\frac{(\ln h^{-1}(-\ln U_{(1)}))^{\gamma} l(\ln h^{-1}(-\ln U_{(1)}))}{-\ln U_{(1)}} \frac{-\ln U_{(1)}}{-\ln U_{(2)}}.$$

Παρατηρούμε τώρα, ότι $P\left(\frac{-\ln U_{(1)}}{-\ln U_{(2)}} > 2\right) = o(1)$ και στην πραγματικότητα

$$P\left(\frac{-\ln U_{(1)}}{-\ln U_{(2)}} > 3\right) = 0. \text{ Γράφοντας λοιπόν } t_n = -\ln U_{(1)},$$

$$\frac{(\ln h^{-1}(t_n))^{\gamma} l(\ln h^{-1}(t_n))}{t_n} \xrightarrow{a.s.} 0,$$

αφού η h^{-1} είναι $R_{\frac{1}{\gamma}}$ με αποτέλεσμα η $\ln h^{-1}$ να είναι βραδείας κύμανσης, και

έτσι η $l(\ln h^{-1}(t_n))$ και η $(\ln h^{-1}(t_n))^{\gamma}$ είναι επίσης βραδείας κύμανσης.

Συνεπάγεται λοιπόν ότι η (30) είναι σωστή αφού $l(x)/x \rightarrow 0$ για συνάρτηση l βραδείας κύμανσης.

Στην περίπτωση που $\gamma=0$, αφού $\frac{-\ln U_{(1)}}{-\ln \xi_n} \xrightarrow{P} 1$ έπεται από την (28) ότι για να δείξουμε ότι η (26) συγκλίνει στο μηδέν κατά πιθανότητα, είναι αρκετό να δείξουμε ότι:

$$\frac{h\left(\ln h^{-1}(-\ln U_{(1)})\right)}{\ln \ln h^{-1}(-\ln U_{(1)}) h^{-1}(-\ln U_{(1)}) h'\left(h^{-1}(-\ln U_{(1)})\right)} \rightarrow 0. \quad (31)$$

Αυτό προκύπτει άμεσα από τις υποθέσεις στην περίπτωση που $\gamma=0$ θέτοντας $y = \ln h^{-1}(-\ln U_{(1)})$, αφού σε αυτή την περίπτωση $r_F(x) = e^x h'(e^x)$ και $-\frac{\ln \bar{F}(\ln x)}{\ln x} \sim r_F(\ln x)$.

Τέλος, έστω ότι $\gamma=\infty$. Αυτό σημαίνει ότι η $-\ln \bar{F}(\ln x)$ έχει ταχεία κύμανση. Η σχέση (31) είναι ισοδύναμη με την:

$$\frac{h(x)}{e^x h'(e^x) \ln x} \rightarrow 0 \text{ για } x \rightarrow \infty. \quad (32)$$

Θυμόμαστε, ότι η συνάρτηση ταχείας κύμανσης $h(x)$ μπορεί να γραφτεί ως $c(x) \exp\left(\int_1^x \frac{z(t)}{t} dt\right)$ με $c(x) \rightarrow c$, $z(t) > 0$ και $z(t) \rightarrow \infty$ για $t \rightarrow \infty$.

Υποθέτουμε χωρίς απώλεια της γενικότητας ότι $c(x) = c$, οπότε είναι προφανές ότι $e^x h'(e^x) = c * z(e^x) h(e^x)$. Άρα είναι:

$$\frac{h(x)}{e^x h'(e^x) \ln x} = \frac{h(x)}{z(e^x) h(e^x) \ln x}. \quad (33)$$

Αφού $z(x) \rightarrow \infty$ για $x \rightarrow \infty$ ενώ η h είναι μη φθίνουσα, η (32) ισχύει.

Απόδειξη Θεωρήματος 8: Για F με μακριά ουρά, $\bar{F}(\ln x) = L(x)$ για κάποια L βραδείας κύμανσης. Συνεπώς $-\ln \bar{F}(\ln x) = -\ln L(x)$ και άρα $r_F(\ln x) = -xL'(x)/L(x)$, όπου $L(x) = c(x) \exp\left(\int_1^x \frac{\varepsilon(t)}{t} dt\right)$. Κάτω από τις προϋποθέσεις του Θεωρήματος 4 λαμβάνουμε υπόψη τον όρο $\frac{-\ln \bar{F}(\ln X_{(n)})}{\ln X_{(n)}}$ αντί για τον όρο

$\frac{-\ln \bar{F}_n(\ln X_{(n)})}{\ln X_{(n)}}$ και γράφουμε:

$$\frac{-\ln \bar{F}(\ln X_{(n)})}{\ln X_{(n)}} = \frac{-\ln \bar{F}(\ln X_{(n-1)})}{\ln X_{(n-1)}} + (X_{(n)} - X_{(n-1)}) \left\{ \frac{r_F(\ln \xi_n)}{\xi_n \ln \xi_n} + \frac{\ln \bar{F}(\ln \xi_n)}{\xi_n (\ln \xi_n)^2} \right\}$$

για κάποιο ξ_n με $X_{(n-1)} < \xi_n < X_{(n)}$. Παρατηρήστε ότι, σχεδόν βέβαια, για $n \rightarrow \infty$:

$$\frac{\ln \bar{F}(\ln \xi_n)}{\xi_n (\ln \xi_n)^2} \sim \frac{-r_F(\ln \xi_n)}{\xi_n (2 \ln \xi_n + (\ln \xi_n)^2)}.$$

Συνεπώς, σχεδόν βέβαια, για αρκετά μεγάλο n , ισχύει

$$T_n \geq \frac{-\ln \bar{F}(\ln X_{(n-1)})}{\ln X_{(n-1)}} (X_{(n)} - X_{(n-1)}).$$

Αφού $\bar{F}^{-1}(u) = \ln(L^{-1}(u))$, μπορούμε να γράψουμε χρησιμοποιώντας το πρώτο βήμα του αναπτύγματος Taylor,

$$\begin{aligned} & \frac{-\ln \bar{F}(\ln X_{(n-1)})}{\ln X_{(n-1)}} (X_{(n)} - X_{(n-1)}) = \\ & = \frac{-\ln L(X_{(n-1)})}{\ln X_{(n-1)}} \left(\ln L^{-1}(U_{(1)}) - \ln L^{-1}(U_{(2)}) \right) = \\ & = \frac{-\ln L(\ln L^{-1}(U_{(2)}))}{\ln \ln L^{-1}(U_{(2)})} \frac{(U_{(1)} - U_{(2)})}{L'(L^{-1}(\psi_n))L^{-1}(\psi_n)}, \end{aligned}$$

όπου $U_{(1)}, U_{(2)}$ είναι τα στατιστικά πρώτης και δευτέρας τάξης από μία $U \sim (0,1)$ και $U_{(1)} < \psi_n < U_{(2)}$. Γράφοντας $U_{(1)} - U_{(2)} = -(1 - V)U_{(2)}$, όπου $V \sim U(0,1)$ με την τυχαία μεταβλητή V να είναι ανεξάρτητη από τη $U_{(2)}$ και αφού $\psi_n < U_{(2)}$ και $L'(x) < 0$, έπεται ότι T_n είναι το λιγότερο ίσο με

$$\frac{-\ln L(\ln L^{-1}(U_{(2)}))}{\ln \ln L^{-1}(U_{(2)})} \frac{(1 - V)LL^{-1}(\psi_n)}{-L'(L^{-1}(\psi_n))L^{-1}(\psi_n)} \quad (34)$$

Αφού το r_F τελικά μειώνεται, $-xL'(x)/L(x)$ τελικά μειώνεται και εφόσον $L^{-1}(\psi_n) \geq L^{-1}(U_{(2)})$, έπεται ότι

$$\frac{-L^{-1}(U_{(2)})L'L^{-1}(U_{(2)})}{LL^{-1}(U_{(2)})} \geq \frac{-L^{-1}(\psi_n)L'L^{-1}(\psi_n)}{LL^{-1}(\psi_n)}$$

και έτσι η (34) υποδηλώνει ότι, σχεδόν βέβαια, για μεγάλα n , θέτοντας $Y_n = L^{-1}(U_{(2)})$:

$$T_n \geq \frac{-\ln L(\ln Y_n)}{\ln \ln Y_n} \frac{(1 - V)L(Y_n)}{-L'(Y_n)Y_n} \quad (35)$$

με V ανεξάρτητο από το Y_n και $Y_n \rightarrow \infty$ σχεδόν βέβαια. Τώρα:

$$\frac{L(y)}{-L'(y)y} = \frac{1}{r_F \ln y}, \text{ ενώ } \frac{-\ln L(\ln y)}{\ln \ln y} \sim \frac{-L'(\ln y) \ln y}{L(\ln y)} = r_F(\ln \ln y).$$

Συνεπώς η δεξιά πλευρά της (35) είναι ασυμπτωτικά ισοδύναμη, σχεδόν βέβαια, με την έκφραση:

$$(1 - V) \frac{r_F(\ln \ln Y_n)}{r_F(\ln Y_n)}.$$

Το αποτέλεσμα έπεται από την υπόθεση ότι $r_F(y) = o(r_F(\ln y))$.

Απόδειξη Θεωρήματος 9: Από την απόδειξη του προηγούμενου Θεωρήματος μπορούμε να γράψουμε ότι, σχεδόν βέβαια, για αρκετά μεγάλο n , ισχύει

$$T_n \geq \frac{-\ln \bar{F}(\ln X_{(n-1)})}{\ln X_{(n-1)}} (X_{(n)} - X_{(n-1)}). \quad (36)$$

Αφού $\bar{F}^{-1}(u) = L^{-1}(u)$, μπορούμε να γράψουμε χρησιμοποιώντας το πρώτο βήμα του αναπτύγματος Taylor,

$$\frac{-\ln \bar{F}(\ln X_{(n-1)})}{\ln X_{(n-1)}} (X_{(n)} - X_{(n-1)}) = \quad (37)$$

$$= \frac{-\ln L(\ln L^{-1}(U_{(2)}))}{\ln L^{-1}(U_{(2)})} \frac{(U_{(1)} - U_{(2)})}{L'(L^{-1}(\psi_n))} = \quad (38)$$

$$= \frac{-\ln L(\ln L^{-1}(U_{(2)}))}{\ln L^{-1}(U_{(2)})} \frac{(1-V)U_{(2)}}{-L'(L^{-1}(\psi_n))}, \quad (39)$$

όπου $U_{(1)}, U_{(2)}$ είναι τα στατιστικά πρώτης και δευτέρας τάξης από μία $U \sim (0,1)$ και $U_{(1)} < \psi_n < U_{(2)}$, με V ανεξάρτητη του $U_{(2)}$.

Αλλά, αφού L^{-1} είναι φθίνουσα και άρα $L^{-1}(\psi_n) > L^{-1}(U_{(2)})$, και εφόσον τελικά η $-L'/L$ είναι φθίνουσα, τότε:

$$\frac{U_{(2)}}{-L'(L^{-1}(\psi_n))} \geq \frac{L(L^{-1}(\psi_n))}{-L'(L^{-1}(\psi_n))} \geq \frac{L(L^{-1}(U_{(2)}))}{-L'(L^{-1}(U_{(2)}))}.$$

Έπεται από την (37) ότι, θέτοντας $Y_n = L^{-1}(U_{(2)})$:

$$\begin{aligned} \frac{-\ln \bar{F}(\ln X_{(n-1)})}{\ln X_{(n-1)}} (X_{(n)} - X_{(n-1)}) &\geq \frac{-\ln L(\ln Y_n)L(Y_n)}{-\ln Y_n L'(Y_n)} = \\ &= \frac{-\ln L(\ln Y_n)}{-\ln Y_n \left(\frac{-c'(Y_n)}{c(Y_n)} - \frac{\varepsilon(Y_n)}{Y_n} \right)} \geq \frac{-\ln L(\ln Y_n)Y_n}{\ln Y_n (-\varepsilon(Y_n))} \rightarrow \infty, \end{aligned}$$

αφού $-\varepsilon(Y_n) \rightarrow \alpha \geq 0$ και $\frac{-\ln L(\ln Y_n) Y_n}{\ln n} \rightarrow \infty$, σχεδόν βέβαια.

Κώδικας

```
#####Graph1_Chapter3#####  
#####Weibull#####  
  
n<-10000000  
  
g<-c(0.1,0.2,0.5,0.8,1,1.5,2,3,5,10)  
  
par(mfrow=c(2,2))  
  
gamma<-g[10]  
  
z<-runif(n,0,1)  
  
x<-(-log(1-z))^(1/gamma)  
  
h <- hist(x,plot=FALSE) #generate hist  
  
xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf  
  
plot(x=xlines,dweibull(xlines,gamma,1,log=FALSE),type="l",main=  
"Κατανομή Weibull",ylab = "pdf(γ=10)",xlab="x")  
  
gamma<-g[7]  
  
z<-runif(n,0,1)  
  
x<-(-log(1-z))^(1/gamma)
```

```

h <- hist(x,plot=FALSE) #generate hist

xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf

plot(x=xlines,dweibull(xlines,gamma,1,log=FALSE),type="l",main =
"Κατανομή Weibull",ylab = "pdf( $\gamma=2$ )",xlab="x" )

```

```

gamma<-g[5]

```

```

z<-runif(n,0,1)

```

```

x<-(-log(1-z))^(1/gamma)

```

```

h <- hist(x,plot=FALSE) #generate hist

```

```

xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf

```

```

plot(x=xlines,dweibull(xlines,gamma,1,log=FALSE),type="l",main=
"Κατανομή Weibull",ylab = "pdf( $\gamma=1$ )",xlab="x" )

```

```

gamma<-g[3]

```

```

z<-runif(n,0,1)

```

```

x<-(-log(1-z))^(1/gamma)

```

```

h <- hist(x,plot=FALSE) #generate hist

```

```

xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf

```

```

plot(x=xlines,dweibull(xlines,gamma,1,log=FALSE),type="l",main=
"Κατανομή Weibull",ylab = "pdf( $\gamma=0.5$ )",xlab="x" )

```

```
#####Graph2_Chapter3#####
```

```
#####Pareto#####
```

```
n<-10000000
```

```
g<-1:10
```

```
par(mfrow=c(2,2))
```

```
gamma<-g[1]
```

```
z<-runif(n,0,1)
```

```
x<-(z/(1-z))^(1/gamma)
```

```
h <- hist(x,plot=FALSE) #generate hist
```

```
xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf
```

```
plot(x=xlines,(gamma*xlines^(gamma-1))/((1+xlines^gamma)^2),type="l",main  
= "Pareto {S(x)=1/(1+x^γ)}",ylab = "pdf(γ=1)",xlab="x")
```

```
gamma<-g[5]
```

```
z<-runif(n,0,1)
```

```
x<-(z/(1-z))^(1/gamma)
```

```
h <- hist(x,plot=FALSE) #generate hist
```

```
xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf
```

```
plot(x=xlines,(gamma*xlines^(gamma-1))/((1+xlines^gamma)^2),type="l",main
= "Pareto {S(x)=1/(1+x^γ)}",ylab = "pdf(γ=5)",xlab="x")
```

```
gamma<-g[7]
```

```
z<-runif(n,0,1)
```

```
x<-(z/(1-z))^(1/gamma)
```

```
h <- hist(x,plot=FALSE) #generate hist
```

```
xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf
```

```
plot(x=xlines,(gamma*xlines^(gamma-1))/((1+xlines^gamma)^2),type="l",main
= "Pareto {S(x)=1/(1+x^γ)}",ylab = "pdf(γ=7)",xlab="x")
```

```
gamma<-g[10]
```

```
z<-runif(n,0,1)
```

```
x<-(z/(1-z))^(1/gamma)
```

```
h <- hist(x,plot=FALSE) #generate hist
```

```
xlines <-seq(min(h$breaks),max(h$breaks),length.out=100) #seq of x for pdf
```

```
plot(x=xlines,(gamma*xlines^(gamma-1))/((1+xlines^gamma)^2),type="l",main
= "Pareto {S(x)=1/(1+x^γ)}",ylab = "pdf(γ=10)",xlab="x")
```

```
#####Table 4#####
```

```
#####Exponential#####
```

```
B<-10000
```

```
K<-NULL
```

```
for (i in 1:B) {
```

```
  n<-10000
```

```
  x<-rexp(n,100)
```

```
  xn<-max(x)
```

```
  F<-sum(x<=log(xn))/n
```

```
  S<-1-F
```

```
  xn1<-x[rank(x)==n-1]
```

```
  Tn<--log(S)*(xn-xn1)/log(xn)
```

```
  K<-c(K,Tn)}
```

```
sum(K>(-log(0.05)))/B ##### Long #####
```

```
sum(K<(-log(0.95)))/B ##### Short #####
```

```
#####Logistic#####
```

```
B<-10000
```

```
K<-NULL
```

```

for (i in 1:B) {

n<-500

x<-rlogis(n,0.7,1)

xn<-max(x)

F<-sum(x<=log(xn))/n

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

sum(K>(-log(0.05)))/B ##### Long #####

sum(K<(-log(0.95)))/B ##### Short #####

#####Gamma#####

B<-10000

K<-NULL

for (i in 1:B) {

n<-100

x<-rgamma(n,0.7,1)

xn<-max(x)

F<-sum(x<=log(xn))/n

S<-1-F

```

```

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

sum(K>(-log(0.05)))/B ##### Long #####

sum(K<(-log(0.95)))/B ##### Short #####

##### Table 5#####

B<-10000

K<-NULL

g<-5

for (i in 1:B) {

n<-5000

y<-runif(n,0,1)

x<-(y/(1-y))^(1/g)

xn<-max(x)

F<-sum(x<=log(xn))/n

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

```

```
sum(K>(-log(0.05)))/B ##### Long #####
```

```
sum(K<(-log(0.95)))/B ##### Short #####
```

```
##### Table 6#####
```

```
B<-10000
```

```
K<-NULL
```

```
g<-5
```

```
for (i in 1:B) {
```

```
  n<-5000
```

```
  y<-runif(n,0,1)
```

```
  x<-(-log(1-y))^(1/g)
```

```
  xn<-max(x)
```

```
  F<-sum(x<=log(xn))/n
```

```
  S<-1-F
```

```
  xn1<-x[rank(x)==n-1]
```

```
  Tn<-log(S)*(xn-xn1)/log(xn)
```

```
  K<-c(K,Tn)}
```

```
sum(K>(-log(0.05)))/B ##### Long #####
```

```
sum(K<(-log(0.95)))/B ##### Short #####
```



```
##### Table 7#####
```

```
#####Inverse Weibull#####
```

```
B<-10000
```

```
K<-NULL
```

```
for (i in 1:B) {
```

```
  n<-10000
```

```
  y<-runif(n,0,1)
```

```
  x<-1+log(y)
```

```
  xn<-max(x)
```

```
  if (xn>0){
```

```
    F<-sum(x<=log(xn))/n
```

```
    S<-1-F
```

```
    xn1<-x[rank(x)==n-1]
```

```
    Tn<--log(S)*(xn-xn1)/log(xn)
```

```
    K<-c(K,Tn)} }
```

```
sum(K>(-log(0.05)))/B ##### Long #####
```

```
sum(K<(-log(0.95)))/B ##### Short #####
```

```
#####Uniform#####
```

```
B<-10000
```

```
K<-NULL
```

```

for (i in 1:B) {

n<-10

x<-runif(n,0,1)

xn<-max(x)

F<-sum(x<=log(xn))/n

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

sum(K>(-log(0.05)))/B ##### Long #####

sum(K<(-log(0.95)))/B ##### Short #####

#####Normal#####

B<-10000

K<-NULL

for (i in 1:B) {

n<-10

x<-rnorm(n,0,1)

xn<-max(x)

if (xn>0) {

F<-sum(x<=log(xn))/n

```

```

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)} }

sum(K>(-log(0.05)))/length(K) ##### Long #####

sum(K<(-log(0.95)))/length(K) ##### Short #####

##### Table 8#####

#####LogNormal#####

B<-10000

K<-NULL

for (i in 1:B) {

n<-10

y<-rnorm(n,0,1)

x<-exp(y)

xn<-max(x)

F<-sum(x<=log(xn))/n

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

```

```

K<-c(K,Tn)}

sum(K>(-log(0.05)))/B ##### Long #####

sum(K<(-log(0.95)))/B ##### Short #####

#####Student#####

B<-10000

K<-NULL

for (i in 1:B) {

n<-10

x<-rt(n,3)

xn<-max(x)

if (xn>0) {

F<-sum(x<=log(xn))/n

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}}

sum(K>(-log(0.05)))/length(K) ##### Long #####

sum(K<(-log(0.95)))/length(K) ##### Short #####

#####Cauchy#####

B<-10000

```

```

K<-NULL

for (i in 1:B) {

n<-50

x<-rcauchy(n,0,1)

xn<-max(x)

F<-sum(x<=log(xn))/n

S<-1-F

xn1<-x[rank(x)==n-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

sum(K>(-log(0.05)))/B ##### Long #####

sum(K<(-log(0.95)))/B ##### Short #####

##### Table 9#####

#####Exponential#####

k<-100

n<-20000

B<-10000

H<-NULL

for (i in 1:B) {

```

```

K<-NULL

for (j in 1:k) {

x<-rexp(n/k,1)

xn<-max(x)

F<-sum(x<=log(xn))/(n/k)

S<-1-F

xn1<-x[rank(x)==(n/k)-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

G<-sum(K)

H<-c(H,G)}

sum(H>(qgamma(0.95,k,1)))/B ##### Long #####

sum(H<(qgamma(0.05,k,1)))/B ##### Short #####

#####Logistic#####

k<-50

n<-500

B<-10000

H<-NULL

for (i in 1:B) {

K<-NULL

```

```

for (j in 1:k) {
x<-rlogis(n/k,0.7,1)

xn<-max(x)

if (xn>0) {

F<-sum(x<=log(xn))/(n/k)

S<-1-F

xn1<-x[rank(x)==(n/k)-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}}

G<-sum(K)

H<-c(H,G)}

sum(H>(qgamma(0.95,k,1)))/B ##### Long #####

sum(H<(qgamma(0.05,k,1)))/B ##### Short #####

#####Tables 10-12#####

#####Normal#####

k<-50

n<-20000

B<-10000

H<-NULL

```

```

for (i in 1:B) {

K<-NULL

for (j in 1:k) {

x<-rnorm(n/k,1)

xn<-max(x)

if (xn>0) {

F<-sum(x<=log(xn))/(n/k)

S<-1-F

xn1<-x[rank(x)==(n/k)-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}}

G<-sum(K)

H<-c(H,G)}

sum(H>(qgamma(0.95,k,1)))/B ##### Long #####

sum(H<(qgamma(0.05,k,1)))/B ##### Short #####

#####LogNormal#####

k<-50

n<-20000

B<-10000

H<-NULL

```



```

for (i in 1:B) {

K<-NULL

for (j in 1:k) {

y<-rnorm(n/k,0,1)

x<-exp(y)

xn<-max(x)

F<-sum(x<=log(xn))/(n/k)

S<-1-F

xn1<-x[rank(x)==(n/k)-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

G<-sum(K)

H<-c(H,G)}

sum(H>(qgamma(0.95,k,1)))/B ##### Long #####

sum(H<(qgamma(0.05,k,1)))/B ##### Short #####

#####Pareto#####

g<-5

k<-50

n<-20000

B<-10000

```

```

H<-NULL

for (i in 1:B) {

K<-NULL

for (j in 1:k) {

y<-runif(n/k,0,1)

x<-(y/(1-y))^(1/g)

xn<-max(x)

F<-sum(x<=log(xn))/(n/k)

S<-1-F

xn1<-x[rank(x)==(n/k)-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

G<-sum(K)

H<-c(H,G)}

sum(H>(qgamma(0.95,k,1)))/B ##### Long #####

sum(H<(qgamma(0.05,k,1)))/B ##### Short #####

#####Weibull#####

g<-2

k<-50

n<-20000

```

```

B<-10000

H<-NULL

for (i in 1:B) {

K<-NULL

for (j in 1:k) {

y<-runif(n/k,0,1)

x<-(-log(1-y))^(1/g)

xn<-max(x)

F<-sum(x<=log(xn))/(n/k)

S<-1-F

xn1<-x[rank(x)==(n/k)-1]

Tn<--log(S)*(xn-xn1)/log(xn)

K<-c(K,Tn)}

G<-sum(K)

H<-c(H,G)}

sum(H>(qgamma(0.95,k,1)))/B ##### Long #####

sum(H<(qgamma(0.05,k,1)))/B ##### Short #####

```

```
#####Realdata_Chapter 9#####
```

```
##### Right tail#####
```

```

data<-
c(27.40,17.35,16.20,43.91,23.59,5.52,2.69,58.16,61.84,24.76,15.79,19.15,8.05,5
.37,14.43,15.34,10.41,6.05,30.16)

plot(data)

B<-10000

Sn<-bootsample<-NULL

for (i in 1:B) {

Bootsample<- sample(data,replace=T)

Bootsample<-sort(Bootsample)

d<-sort(rank(Bootsample,ties.method="max"))

t<- max(Bootsample)- Bootsample[length(data)-sum(d==length(data))]

Sn<-c(Sn,t)

bootsample<-c(bootsample,Bootsample)

}

windows()

hist(Sn,freq=FALSE)

windows()

hist(bootsample,freq=FALSE)

windows()

```

```

dataaxis<-seq(from=min(data),to=max(data),by=0.01)

ecdf<-rep(NA,length(dataaxis))

for (i in 1:length(dataaxis)) {

ecdf[i]<-sum(data<=dataaxis[i])/length(data)}

plot(dataaxis,ecdf)

windows()

plot(qexp(seq(from=0,to=1,by=0.01),1/mean(data)),quantile(data, probs = seq(0,
1, 0.01)))

abline(0,1)

#####Bootstrap_right tail#####

data<-
c(27.40,17.35,16.20,43.91,23.59,5.52,2.69,58.16,61.84,24.76,15.79,19.15,8.05,5
.37,14.43,15.34,10.41,6.05,30.16)

xnobs<-max(data)

Fobs<-sum(data<=log(xnobs))/length(data)

Sobs<-1-Fobs

xn1obs<-data[rank(data)==length(data)-1]

Tnobs<--log(Sobs)*(xnobs-xn1obs)/log(xnobs)

B<-10000

```

```

K<-NULL

for (i in 1:B) {

bootsample<- sort(sample(data,replace=T))

xn<-max(bootsample)

F<-sum(data<=log(xn))/length(data)

S<-1-F

e<-sort(rank(bootsample,ties.method="max"))

z<-max(bootsample)- bootsample[length(data)-sum(e==length(data))]

Tn<--log(S)*(z)/log(xn)

K<-c(K,Tn)}

X<-hist(K,freq=FALSE)

sum(K>Tnobs)/B ##### Long #####

sum(K<Tnobs)/B ##### Short #####

##### Left tail#####

data<--
c(27.40,17.35,16.20,43.91,23.59,5.52,2.69,58.16,61.84,24.76,15.79,19.15,8.05,5
.37,14.43,15.34,10.41,6.05,30.16)

plot(data)

B<-10000

Sn<-bootsample<-NULL

for (i in 1:B) {

```

```

Bootsample<- sample(data,replace=T)

Bootsample<-sort(Bootsample)

d<-sort(rank(Bootsample,ties.method="max"))

t<- max(Bootsample)- Bootsample[length(data)-sum(d==length(data))]

Sn<-c(Sn,t)

bootsample<-c(bootsample,Bootsample)

}

windows()

hist(Sn,freq=FALSE)

windows()

hist(bootsample,freq=FALSE)

windows()

dataaxis<-seq(from=min(data),to=max(data),by=0.01)

ecdf<-rep(NA,length(dataaxis))

for (i in 1:length(dataaxis)) {

ecdf[i]<-sum(data<=dataaxis[i])/length(data)}

plot(dataaxis,ecdf)

windows()

plot(-qexp(seq(from=0,to=1,by=0.01),1/(-mean(data))),-quantile(-data, probs =
seq(0, 1, 0.01)))

```

```
abline(0,1)
```

```
#####Bootstrap_ left tail#####
```

```
data<--
```

```
c(27.40,17.35,16.20,43.91,23.59,5.52,2.69,58.16,61.84,24.76,15.79,19.15,8.05,5  
.37,14.43,15.34,10.41,6.05,30.16)
```

```
data<--min(data)+data
```

```
xnobs<-max(data)
```

```
Fobs<-sum(data<=log(xnobs))/length(data)
```

```
Sobs<-1-Fobs
```

```
xn1obs<-data[rank(data)==length(data)-1]
```

```
Tnobs<--log(Sobs)*(xnobs-xn1obs)/(log(xnobs))
```

```
B<-10000
```

```
K<-NULL
```

```
for (i in 1:B) {
```

```
  bootsample<- sort(sample(data,replace=T))
```

```
  xn<-max(bootsample)
```

```
  F<-sum(data<=log(xn))/length(data)
```

```
  S<-1-F
```

```
  e<-sort(rank(bootsample,ties.method="max"))
```

```
  z<-max(bootsample)- bootsample[length(data)-sum(e==length(data))]
```



```

Tn<--log(S)*(z)/log(xn)

K<-c(K,Tn)}

hist(K,freq=FALSE)

sum(K>Tnobs)/B ##### Long #####

sum(K<Tnobs)/B ##### Short #####

#####Bryson test#####

#####Quantiles_ Tables 13-15#####

n<-5000

shape<-1/2

B<-1000

Br<-NULL

for (i in 1:B) {

x<-rgamma(n,shape,1)

y<-rep(NA,n)

xn<-max(x)

for (i in 1:n) {

y[i]<-(x[i]+(xn/(n-1)))}

a<-prod(y[1:n])

```

```

xga<-a^(1/n)

T<-((mean(x)*xn)/((n-1)*xga^2))

Br<-c(Br,T)

}

hist(Br)

quantile(Br, probs=c(0.025,0.05,0.95,0.975),na.rm=TRUE)

#####Quantiles_logistic_Table 16#####

n<-20000

B<-1000

Br<-NULL

for (i in 1:B) {

x<-rlogis(n,0.7,1)

y<-rep(NA,n)

xn<-max(x)

for (i in 1:n) {

y[i]<-(x[i]+(xn/(n-1)))}

a<-prod(y[1:n])

xga<-a^(1/n)

T<-((mean(x)*xn)/((n-1)*xga^2))

```

```

Br<-c(Br,T)

}

hist(Br)

quantile(Br, probs=c(0.025,0.05,0.95,0.975),na.rm=TRUE)

#####Quantiles_Weibull_Tables 17-18#####

n<-50

B<-1000

g<-2

Br<-NULL

for (i in 1:B) {

z<-runif(n,0,1)

x<-(-log(1-z))^(1/g)

y<-rep(NA,n)

xn<-max(x)

for (i in 1:n) {

y[i]<-(x[i]+(xn/(n-1)))}

a<-prod(y[1:n])

xga<-a^(1/n)

T<-((mean(x)*xn)/((n-1)*xga^2)

```

```
Br<-c(Br,T)
```

```
}
```

```
hist(Br)
```

```
quantile(Br, probs=c(0.025,0.05,0.95,0.975))
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική βιβλιογραφία

- [1] Σπύρος Γεωργαλάς (2014), «Εκτίμηση Οικολογικής Παροχής από το Φράγμα του Μαραθώνα και Διαχείριση Διατιθέμενων Υδάτων», Διπλωματική Εργασία.
- [2] Δημήτριος Καρλής (2008), «Υπολογιστική Στατιστική», Σημειώσεις Προπτυχιακού Μαθήματος.
- [3] Μιχαήλ Μπούτσικας (2015), «Θεωρία ακραίων Τιμών», Σημειώσεις Μεταπτυχιακού Αναλογιστικής Επιστήμης και Διοικητικής Κινδύνου.
- [4] Γεώργιος Πιτσέλης (2015), «Ζημιοκατανομές», Σημειώσεις Μεταπτυχιακού Αναλογιστικής Επιστήμης και Διοικητικής Κινδύνου.
- [5] Ιωάννης Χαλικιάς (2010), «Στατιστική. Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις» Γ' έκδοση, έκδ. Rosili.

Ξενόγλωσση βιβλιογραφία

- [1] Beirlant, J. G, Goegebeur, Y, Segers J, Teugels, J (2004), “Statistics of Extremes, Theory and Applications”, Wiley, England.
- [2] Bingham, N, Goldie, C, Teugels, J (1987), “Regular Variation”, Cambridge University Press, New York.
- [3] Bryson, M.C (1974), “Heavy-tailed distributions: properties and tests”, Technometrics, 61-68.
- [4] Heyde, C.C, Kou, S.G (2004), “On the controversy over tailweight of distributions”, Operations Research Letters, 32: 399 – 408.
- [5] Klugman, AS, Panjer, HH, Willmot, EG (2008), “Loss models, from data to decisions”, 3rd edition Wiley, New York.

[6] Lehmann, E (1988), “Comparing location experiments”, *The Annals of Statistics*, 16:521-533.

[7] Lehmann. E. L, Loh. W.Y (1990), “Pointwise versus uniform robustness of some large-sample tests and confidence intervals”, *Scand. J. Statist.*, 17:177–187.

[8] McNeil, AJ (1997), “Estimating the tails of loss severity distributions using extreme value theory”, *ASTIN Bull* 27:117–137.

[9] Rojo, J (1988), “On the concept of tail heaviness”, Technical Report 175, University of California at Berkeley.

[10] Rojo, J (1992), “A pure-tail ordering based on the ratio of the quantile functions”, *The Annals of Statistics*, 20:570-579.

[11] Rojo, J (1993), “On the preservation of some pure-tail orderings by reliability operations”, *Statistics and Probability Letters*, 17:189-198.

[12] Rojo, J (1996), “On tail categorization of probability laws”, *Journal of the American Statistical Association*, 91:378-384.

[13] Rojo, J, Ott, C (2010), “Testing for tail behavior using extreme spacings” Rice University and Mesa State College.

[14] Schuster, E (1984), “Classification of probability laws by tail behavior”, *Journal of the American Statistical Association*, 79:936-939.