

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

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

**Η επίδραση της λογοκρισίας στα δεδομένα στην  
εκτίμηση των παραμέτρων στις κατανομές απώλειας**

*Μπάβα Αναστασία*

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

*Πειραιάς  
Νοέμβριος 2015*

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

Τα μέλη της Επιτροπής ήταν:

- Αναπλ. Καθηγητής: Κωνσταντίνος Πολίτης
- Επίκουρος Καθηγητής: Γεώργιος Τζαβελάς (Επιβλέπων)
- Επίκουρος Καθηγητής: 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**

**The effects of censoring at the data at the  
estimation of  
parameters for the loss distributions**

By

Bava Anastasia

MSc Dissertation

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

Piraeus, Greece

November 2015



## **Ευχαριστίες**

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κύριο Τζαβελά Γιώργο για την υποστήριξη και τον καθοδήγηση που μου παρείχε καθόλη τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Επίσης, ένα μεγάλο ευχαριστώ στην οικογένεια μου και σε όλους όσους με βοήθησαν σε αυτή την προσπάθεια.

## Περίληψη

Η λογοκρισία σε ένα δείγμα μπορεί να έχει σημαντικές επιπτώσεις στην εκτίμηση παραμέτρων. Στην εργασία μελετώνται η λογοκρισία τύπου I, τύπου II και η προοδευτική λογοκρισία ως προς την επίδραση που έχουν στην αποτελεσματικότητα του εκτιμητή μέγιστης πιθανοφάνειας για κατανομές απώλειας. Θα χρησιμοποιηθούν προσομοιωμένα δείγματα από διάφορες κατανομές απώλειας.

Στο πρώτο κεφάλαιο γίνεται μία περιληπτική αναφορά στις βασικές έννοιες όπως κατανομές απώλειας, πληροφοριακός αριθμός Fisher, εκτιμητής μέγιστης πιθανοφάνειας. Στο δεύτερο κεφάλαιο γίνεται αναφορά στη λογοκρισία και αναλύονται τα είδη της (λογοκρισία τύπου I, λογοκρισία τύπου II και τυχαία λογοκρισία). Στα τρίτο, τέταρτο και πέμπτο κεφάλαιο παρουσιάζεται μία ανάλυση των κατανομών που χρησιμοποιήθηκαν στην παρούσα εργασία. Πρόκειται για τις κατανομές Weibull, Pareto και Gamma αντίστοιχα. Στα επιμέρους αυτά κεφάλαια παρουσιάζονται επίσης και αποτελέσματα από την εφαρμογή των κατανομών στο πρόγραμμα Mathematica. Οι αλγόριθμοι παρατίθενται στο Παράρτημα.

Τέλος, στο έκτο κεφάλαιο γίνεται έλεγχος καλής προσαρμογής σε πραγματικά στοιχεία των κατανομών Weibull, Pareto και Gamma, από τις οποίες καλύτερη προσαρμογή παρουσιάζει η Pareto. Στη συνέχεια για την κατανομή αυτή ελέγχεται η επίπτωση του βαθμού λογοκρισίας στα διάφορα στατιστικά χαρακτηριστικά της.

## **Abstract**

Censoring in a sample can have a significant impact on parameter estimation. This paper studied the censoring type I, type II and progressive censoring and the effect they have on the efficiency of the maximum likelihood estimator for loss distributions. Simulated samples will be used from different loss distributions.

The first chapter is a brief reference to basic concepts such as loss distributions, the Fisher information and maximum likelihood estimators. The second chapter refers to censoring and its types (censored type I, type II and random censoring). In third, fourth and fifth chapter presents an analysis of Weibull, Pareto and Gamma distribution used in this work. The individual chapters are presented also results from the application of the distributions in Mathematica program. The algorithms listed in the Annex.

Finally, the sixth chapter presents the goodness of fit in Weibull, Pareto and Gamma distribution samples from which better adaptation presented by Pareto. Then the distribution is controlled by the effect of censoring degree in various statistical characteristics.

## Περιεχόμενα

|   |    |
|---|----|
| ΚΕΦΑΛΑΙΟ 1 .....  | 5  |
| Εισαγωγή.....   | 5  |
| 1.1 Βασικοί Ορισμοί Εκτιμητικής.....  | 5  |
| 1.2 Πληροφοριακός αριθμός του Fisher .....  | 6  |
| 1.3 Ο εκτιμητής μέγιστης πιθανοφάνειας (ΕΜΠ) .....                                | 10 |
| 1.4 Κατανομές απώλειας .....  | 12 |
| 1.5 Μεθοδολογία .....   | 13 |
| ΚΕΦΑΛΑΙΟ 2 .....  | 14 |
| Τι είναι η λογοκρισία .....   | 14 |
| 2.1 Εισαγωγή.....   | 14 |
| 2.2 Δεξιά λογοκρισία τύπου I.....   | 18 |
| 2.3 Δεξιά λογοκρισία τύπου II.....  | 19 |
| 2.4 Τυχαία λογοκρισία (Random Censoring) .....                                    | 21 |
| ΚΕΦΑΛΑΙΟ 3 .....  | 26 |
| Η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων για την κατανομή Weibull..... | 26 |
| 3.1 Εισαγωγή.....   | 26 |
| 3.2 Εκτιμητής μέγιστης πιθανοφάνειας για πλήρες δείγμα .....                      | 28 |
| 3.3 Εκτιμητής μέγιστης πιθανοφάνειας για λογοκριμένο δείγμα τύπου I και II .....  | 29 |
| 3.4 Εκτιμητής μέγιστης πιθανοφάνειας για τυχαία λογοκριμένο δείγμα .....          | 36 |
| ΚΕΦΑΛΑΙΟ 4 .....  | 45 |
| Η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων για την κατανομή Pareto ..... | 45 |
| 4.1 Εισαγωγή.....   | 45 |
| 4.2 Εκτιμητής μέγιστης πιθανοφάνειας για πλήρες δείγμα .....                      | 47 |
| 4.3 Εκτιμητής μέγιστης πιθανοφάνειας για δεξιά λογοκριμένο δείγμα .....           | 48 |
| 4.4 Εκτιμητής μέγιστης πιθανοφάνειας τυχαία λογοκριμένο δείγμα .....              | 52 |
| Πίνακας αποτελεσμάτων προσομοίωσης.....   | 53 |
| ΚΕΦΑΛΑΙΟ 5 .....  | 59 |
| Η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων για την κατανομή Gamma.....   | 59 |
| 5.1 Εισαγωγικά.....   | 59 |
| 5.2 Εκτιμητής Μέγιστης Πιθανοφάνειας για πλήρες δείγμα .....                      | 61 |
| 5.3 Εκτιμητής Μέγιστης Πιθανοφάνειας για λογοκριμένο δείγμα τύπου I και II.....   | 62 |



|   |     |
|---|-----|
| 5.4 Εκτιμητής Μέγιστης Πιθανοφάνειας για τυχαία λογοκριμένο δείγμα.....                                       | 69  |
| 5.5 Σχέση μεταξύ $p$ και $L$ .....  | 73  |
| ΚΕΦΑΛΑΙΟ 6 .....  | 77  |
| 6.1 Ανάλυση Δεδομένων.....  | 77  |
| Παράρτημα.....  | 90  |
| Π1. Πρόγραμμα εκτίμησης για Weibull .....   | 90  |
| Π2. Πρόγραμμα εκτίμησης για Weibull για μικρά δείγματα .....  | 92  |
| Π3. Πρόγραμμα εκτίμησης για Pareto .....  | 95  |
| Π4. Πρόγραμμα εκτίμησης για Pareto με λογοκρισία τύπου I και II .....   | 97  |
| Π5. Πρόγραμμα εκτίμησης για Gamma .....   | 99  |
| Π6. Πρόγραμμα εκτίμησης της παραμέτρου κλίμακας (scale parameter) για Gamma με λογοκρισία τύπου I και II..... | 101 |
| Π7. Πρόγραμμα εκτίμησης της παραμέτρου σχήματος (shape parameter) για Gamma με λογοκρισία τύπου I και II..... | 103 |

## Κατάλογος διαγραμμάτων

|  |    |
|--|----|
| Διάγραμμα 1: Θετική Ασυμμετρία .....   | 12 |
| Διάγραμμα 2: Λογοκρισία Τύπου I.....   | 15 |
| Διάγραμμα 3: Λογοκρισία Τύπου II.....  | 16 |
| Διάγραμμα 4: Προοδευτική Λογοκρισία Τύπου II .....   | 17 |
| Διάγραμμα 5: Περιοχή Ολοκλήρωσης .....   | 24 |
| Διάγραμμα 6: Κατανομή Weibull για διάφορες τιμές παραμέτρων .....  | 27 |
| Διάγραμμα 7: Αθροιστική συνάρτηση κατανομής Weibull .....  | 28 |
| Διάγραμμα 8: Λόγος λογοκριμένου δείγματος ( $\rho$ ) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο μορφής .....                                     | 31 |
| Διάγραμμα 9: Λόγος λογοκριμένου δείγματος ( $\rho$ ) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο κλίμακας.....                                    | 31 |
| Διάγραμμα 10: Σχέση ποσοστού λογοκρισίας $\rho$ με το άνω άκρο $L$ .....   | 37 |
| Διάγραμμα 11: Λόγος $I_p / I$ .....  | 40 |
| Διάγραμμα 12: Συνάρτηση πυκνότητας πιθανότητας.....  | 46 |
| Διάγραμμα 13; Αθροιστική συνάρτηση κατανομής .....   | 46 |
| Διάγραμμα 14: Λόγος $I_p / I_0$ .....  | 50 |
| Διάγραμμα 15: Gamma Distribution.....  | 61 |
| Διάγραμμα 16: Λόγος λογοκριμένου δείγματος ( $\rho$ ) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο κλίμακας, για διάφορες τιμές της $\gamma$ ..... | 64 |
| Διάγραμμα 17: Λόγος λογοκριμένου δείγματος ( $\rho$ ) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο μορφής, για διάφορες τιμές της $\gamma$ .....   | 65 |
| Διάγραμμα 18: Σχέση μεταξύ $\rho$ και $L$ .....  | 74 |
| Διάγραμμα 19: Σχέση μεταξύ $\rho$ και $L$ .....  | 75 |
| Διάγραμμα 20: Σχέση μεταξύ $\rho$ και $L$ .....  | 76 |

## Κατάλογος πινάκων

|   |    |
|---|----|
| Πίνακας 1: Weibull Distribution $W(2, 2)$ .....                         | 33 |
| Πίνακας 2: Weibull Distribution $W(2, 1)$ .....                         | 34 |
| Πίνακας 3: Weibull Distribution $W(3, 0.5)$ .....                       | 35 |
| Πίνακας 4: Τυχαία λογοκριμένο δείγμα $W(2, 2)$ .....                    | 41 |
| Πίνακας 5: Τυχαία λογοκριμένο δείγμα $W(2, 1)$ .....                    | 42 |
| Πίνακας 6: Τυχαία λογοκριμένο δείγμα $W(3, 0.5)$ .....                  | 43 |
| Πίνακας 7: Pareto Distribution $Pa(2)$ , $Pa(3)$ , $Pa(4)$ .....        | 51 |
| Πίνακας 8: Random Pareto Distribution $Pa(2)$ , $Pa(3)$ , $Pa(4)$ ..... | 54 |
| Πίνακας 9: Gamma Distribution $Ga(2,2)$ .....                           | 66 |
| Πίνακας 10: Gamma Distribution $Ga(2,1)$ .....                          | 67 |

|  |    |
|--|----|
| Πίνακας 11: Gamma Distribution $Ga(3, 0.5)$ .....            | 68 |
| Πίνακας 12: Random Gamma Distribution $Ga(2, 2)$ .....       | 70 |
| Πίνακας 13: Random Gamma Distribution $Ga(2, 1)$ .....       | 71 |
| Πίνακας 14: Random Gamma Distribution $Ga(3, 0.5)$ .....     | 72 |
| Πίνακας 15: Περιγραφικά στατιστικά 1 .....                   | 78 |
| Πίνακας 16: Περιγραφικά στατιστικά 2 .....                   | 78 |
| Πίνακας 17: Εκτιμήσεις παραμέτρων στην κατανομή Weibull..... | 80 |

## Κατάλογος γραφημάτων

|  |    |
|--|----|
| Γράφημα 1: Εκτίμηση μέσης τιμής στην περίπτωση που υπάρχει λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες) .....                               | 84 |
| Γράφημα 2: Εκτίμηση διαμέσου στην περίπτωση που υπάρχει λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες) .....                                  | 84 |
| Γράφημα 3: Εκτίμηση μέσης τιμής, διαμέσου στην περίπτωση που υπάρχει λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες) .....                     | 85 |
| Γράφημα 4: Διάγραμμα διασποράς μέσης τιμής και διαμέσου καθώς το ποσοστό λογοκρισίας αλλάζει .....   | 85 |
| Γράφημα 5: Εκτίμηση μέσης τιμής στην περίπτωση που υπάρχει τυχαία λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες).....                         | 86 |
| Γράφημα 6: Εκτίμηση διαμέσου στην περίπτωση που υπάρχει τυχαία λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες).....                            | 86 |
| Γράφημα 7: Εκτίμηση μέσης τιμής και διαμέσου στην περίπτωση που υπάρχει τυχαία λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες).....            | 87 |
| Γράφημα 8: Διάγραμμα διασποράς μέσης τιμής και διαμέσου στην περίπτωση που υπάρχει τυχαία λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)..... | 87 |

## Κατάλογος εικόνων

|   |    |
|---|----|
| Εικόνα 1: Ιστόγραμμα συχνοτήτων .....                               | 77 |
| Εικόνα 2: Θηκόγραμμα δεδομένων .....                                | 78 |
| Εικόνα 3: Γράφημα λοξότητας- κύρτωσης.....                          | 79 |
| Εικόνα 4: Γραφήματα για την προσαρμογή στην κατανομή Weibull .....  | 81 |
| Εικόνα 5: Γραφήματα για την προσαρμογή στην κατανομή LogNormal..... | 82 |



# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

### 1.1 Βασικοί Ορισμοί Εκτιμητικής

Όταν λέμε παράμετρο (πεπερασμένης διάστασης) του πληθυσμού εννοούμε ένα αριθμητικό χαρακτηριστικό του πληθυσμού π.χ η μέση τιμή του πληθυσμού, η διακύμανση, κάποιο ποσοστό κ.λ.π.

Βασικός σκοπός της εκτιμητικής είναι η εκτίμηση αγνώστων παραμέτρων του πληθυσμού. Η εκτίμηση αυτή γίνεται με δύο ειδών εκτιμητές: εκτιμητές κατά σημείο και εκτιμητές κατά διάστημα.

**Ορισμός:** Εκτιμητής (estimator) μιας άγνωστης παραμέτρου είναι μια στατιστική συνάρτηση. Μία συνάρτηση δηλαδή των δεδομένων μόνο, χωρίς να εμπλέκει άγνωστες παραμέτρους.

Για κάθε δείγμα ο εκτιμητής θα δίνει μια εκτίμηση κάποιας άγνωστης παραμέτρου που ενδιαφέρει. Η αριθμητική τιμή που η εκτιμητήρια παίρνει για κάποιο συγκεκριμένο δείγμα ονομάζεται εκτίμηση (estimate). Π.χ. ένας εκτιμητής για την άγνωστη παράμετρο μέση τιμή  $\mu$  του πληθυσμού είναι ο αριθμητικός μέσος  $\bar{X}$  και ένα δείγμα μπορεί να δώσει μια εκτίμηση  $\bar{x} = 3.45$ .

Οποιαδήποτε στατιστική συνάρτηση μπορεί να θεωρηθεί ως εκτιμητής οποιασδήποτε παραμέτρου του πληθυσμού. Για να θεωρηθεί ένας εκτιμητής “καλός” πρέπει να ικανοποιεί κάποια κριτήρια.

Τα κριτήρια για την αξιολόγηση ενός εκτιμητή είναι τα εξής:

- ✓ Για μικρά δείγματα
  - Αμεροληψία
  - Μικρό μέσο τετραγωνικό σφάλμα.

Αν ο εκτιμητής  $\hat{\theta}_n$  του  $\theta$  είναι αμερόληπτος τότε από την σχέση

$MT\sigma(\hat{\theta}_n, \theta) = E_{\theta}(\hat{\theta} - \theta)^2 = Var(\hat{\theta}_n) + (E_{\theta}(\hat{\theta}_n) - \theta)^2$  το μικρό μέσο τετραγωνικό σφάλμα μεταφράζεται ως μικρή διακύμανση.

✓ Για μεγάλα δείγματα

- Η συνέπεια
- Η ασυμπτωτική κανονικότητα
- Η μικρή ασυμπτωτική διακύμανση
- Η ασυμπτωτική αποτελεσματικότητα

Τα κριτήρια για μικρά δείγματα είναι δύσχρηστα γιατί τις περισσότερες φορές οι εκτιμητές είτε δεν βρίσκονται σε κλειστή μορφή είτε δεν είναι γνωστή η κατανομή τους. Στην περίπτωση αυτή μπορεί κάποιος να καταφύγει σε προσομοιώσεις με μεγάλο αριθμό δειγμάτων για να υπάρξει μια προσέγγιση της διακύμανσης ή του ΜΤΣ. Στις περισσότερες των περιπτώσεων η ασυμπτωτική κατανομή είναι γνωστή. Έτσι για μεγάλα δείγματα η σύγκριση είναι ευκολότερη.

Η ασυμπτωτική αποτελεσματικότητα σχετίζεται με την περίφημη ανισότητα των Cramer-Rao η οποία με την σειρά της εκφράζεται με την βοήθεια του πληροφοριακού αριθμού Fisher.

## 1.2 Πληροφοριακός αριθμός του Fisher

Έστω  $X$  μια τυχαία μεταβλητή (τ.μ.) με πυκνότητα πιθανότητας  $f(x; \theta)$  όπου  $\theta$  είναι μια άγνωστη παράμετρος πεπερασμένης διάστασης.

Η βασική υπόθεση είναι ότι ο παραμετρικός χώρος είναι ένα ανοικτό διάστημα.

**Ορισμός 1.1.** Αν  $\dim(\theta)=1$  τότε ως πληροφοριακός αριθμός του Fisher  $I(\theta)$  ορίζεται η ποσότητα

$$I(\theta) = E\left[\left(\frac{\partial \log f(x; \theta)}{\partial \theta}\right)^2\right] \quad (1.1)$$

δεδομένου ότι  $0 < I(\theta) < \infty$ .

Η πληροφορία Fisher (ή απλά πληροφορία) είναι ένας τρόπος μέτρησης της ποσότητας των πληροφοριών, την οποία η παρατηρηθείσα τυχαία μεταβλητή  $X$  μεταφέρει για την παράμετρο  $\theta$ .

Είναι ένα χρήσιμο εργαλείο για τη σύγκριση εκτιμητών για μεγάλα δείγματα, επειδή εμφανίζεται ως ασυμπτωτική διακύμανση στον Εκτιμητή Μέγιστης Πιθανοφάνειας.

Ο Efron (1975) παρατηρεί ότι για τις εκθετικές οικογένειες κατανομών τυπικής μορφής

$$f(x; \theta) = \exp\{\theta x - k(\theta)\} h(x)$$

η πληροφορία του Fisher είναι η καμπυλότητα της επιφάνειας  $k(\theta)$ .

Επιπλέον η  $I(\theta)$  σχετίζεται με το μέτρο απόκλισης των Kullback-Leibler (Kullback-Leibler divergence measure). Μπορεί να αποδειχθεί ότι η  $I(\theta)$  είναι ο ρυθμός μεταβολής (η παράγωγος) του μέτρου Kullback-Leibler στο σημείο  $\theta$ . Αν η παράγωγος ως προς  $\theta$  μπορεί να εναλλάσσεται με το ολοκλήρωμα ως προς  $x$  τότε ισχύει ότι

$$I(\theta) = E \left[ \left( \frac{\partial \log f(x; \theta)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \right]$$

Αν το  $\theta$  είναι  $k$ -διάστατη παράμετρος  $\theta = (\theta_1, \dots, \theta_k)'$ , η αντίστοιχη ποσότητα λέγεται πίνακας πληροφορίας του Fisher και ορίζεται ως εξής:

$$I_X(\theta) = \left[ E_\theta \left[ \frac{\partial \ln f(X, \theta)}{\partial \theta_i} \cdot \frac{\partial \ln f(X, \theta)}{\partial \theta_j} \right] \right]_{k \times k} \quad (1.2)$$

όπου  $[\cdot]_{k \times k}$  συμβολίζει έναν πίνακα διαστάσεων  $k \times k$ . Κατ' αναλογία με την μονοδιάστατη περίπτωση, αν η παράγωγος ως προς  $\theta_i$   $1 \leq i \leq k$ , μπορεί να εναλλάσσεται με το ολοκλήρωμα ως προς  $x$  τότε ισχύει ότι

$$I_X(\theta) = \left[ E_\theta \left[ \frac{\partial \ln f(X, \theta)}{\partial \theta_i} \cdot \frac{\partial \ln f(X, \theta)}{\partial \theta_j} \right] \right]_{k \times k} = - \left[ E_\theta \left[ \frac{\partial^2 \ln f(X, \theta)}{\partial \theta_i \partial \theta_j} \right] \right]_{k \times k} .$$

Οι κατανομές οι οποίες θα μελετηθούν ικανοποιούν κάποιες συνθήκες (ομαλότητας) οι οποίες είναι αναγκαίες για να εξαχθούν τα αποτελέσματα τα οποία θα χρησιμοποιηθούν στη μελέτη αυτή.



## Συνθήκες Ομαλότητας

Έστω  $X$  μία τυχαία μεταβλητή με π.π.  $f(x; \theta)$ . Λέμε ότι η  $f(x; \theta)$  ικανοποιεί τις συνθήκες ομαλότητας όταν ικανοποιούνται τα εξής:

1. Ο παραμετρικός χώρος  $\Theta$  είναι ανοιχτό υποσύνολο του  $R$ .
2. Το στήριγμα της  $f(x; \theta)$ ,  $S = \{x: f(x; \theta) > 0\}$  είναι ανεξάρτητο του  $\theta$ .
3. Για κάθε  $x$  και  $\theta$  η παράγωγος  $\frac{\partial}{\partial \theta} f(x; \theta)$  υπάρχει και είναι πεπερασμένη.
4. Για όλα τα  $\theta \in \Theta$  ισχύει η αντιμετάθεση της παραγωγίσης ως προς  $\theta$  με την ολοκλήρωση ως προς  $x$ . Δηλαδή

$$\int_S \frac{\partial}{\partial \theta} f(x; \theta) dx = \frac{\partial}{\partial \theta} \int_S f(x; \theta) dx$$

5. Για κάθε  $\theta \in \Theta$  ισχύει ότι  $0 < I(\theta) < \infty$  όπου  $I(\theta)$  είναι ο πληροφοριακός αριθμός του Fisher.

## Ανισότητα Cramer-Rao

Έστω  $X_1, X_2, \dots, X_n$  τυχαίο δείγμα από μία μονοδιάστατη τυχαία μεταβλητή  $X$  με π.π.  $f(x, \theta)$  η οποία ικανοποιεί τις συνθήκες ομαλότητας. Αν  $\hat{\theta}_n$  είναι αμερόληπτος εκτιμητής της ποσότητας  $g(\theta)$  τότε ισχύει ότι

$$\text{Var}_{\theta}(\hat{\theta}_n) \geq \frac{(g'(\theta))^2}{I(\theta)} \quad \forall \theta \in \Theta \quad (1.3)$$

Η ανισότητα αυτή δίνει ένα κάτω φράγμα της διακύμανση που μπορεί να έχει ένας αμερόληπτος εκτιμητής της ποσότητας που ενδιαφέρει. Αν ο εκτιμητής αυτός πιάσει το κάτω φράγμα τότε ο εκτιμητής αυτός είναι Αμερόληπτος Εκτιμητής Ελαχίστης Διασποράς (ΑΟΕΔ).

Υπάρχουν διάφορες γενικεύσεις της παραπάνω ανισότητας στην περίπτωση που δεν ισχύουν οι συνθήκες ομαλότητας. Αξίζει να σημειωθεί η περίπτωση κατά την οποία η συνάρτηση  $g(\theta)$  δεν είναι παραγωγίσιμη. Τότε προκύπτει η ανισότητα Chapman and Robins

$$\text{Var}_\theta(\hat{\theta}_n) \geq \sup_{m \neq \theta} \frac{(g(m) - g(\theta))^2}{\text{Var}_\theta \left\{ \frac{f(X; m)}{f(X; \theta)} \right\}} \quad \forall \theta \in \Theta$$

**Ορισμός 1.2** Ένας εκτιμητής  $\hat{\theta}_n$  της ποσότητας  $g(\theta)$  λέγεται ασυμπτωτικά αποτελεσματικός όταν η ασυμπτωτική διακύμανσή του είναι το κάτω φράγμα των Cramer-Rao δηλαδή

$$\lim_{n \rightarrow \infty} \text{Var}_\theta(\hat{\theta}_n) = \frac{(g'(\theta))^2}{I(\theta)} \quad \forall \theta \in \Theta \quad (1.4)$$

Για την πολυδιάστατη περίπτωση όπου  $\hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n}, \dots, \hat{\theta}_{k,n})$  είναι αμερόληπτος εκτιμητής του  $g(\theta) = (g_1(\theta), g_2(\theta), \dots, g_k(\theta))$ , το κάτω φράγμα των Cramer Rao είναι ο πίνακας

$$G(\theta)I^{-1}(\theta)G^t(\theta) \quad (1.5)$$

όπου

$$G(\theta) = \left[ \frac{\partial g_i(\theta)}{\partial \theta_j} \right]_{k \times k} \quad 1 \leq i, j \leq k \quad (1.6)$$

Στις περισσότερες των περιπτώσεων ισχύει ότι  $g(\theta) = (\theta_1, \theta_2, \dots, \theta_k)$ . Στην περίπτωση αυτή το κάτω φράγμα (1.5) είναι ο πίνακας  $I^{-1}(\theta)$ .

Υπάρχουν πολλές μέθοδοι εκτίμησης: Οι πιο γνωστές είναι οι

- Μέθοδος ελαχίστων τετραγώνων
- Μέθοδος των ροπών
- Μέθοδος Εκτιμητή Μέγιστης Πιθανοφάνειας

Η πιο δημοφιλής μέθοδος εκτίμησης είναι αυτή της μέγιστης πιθανοφάνειας η οποία χρησιμοποιείται στην εργασία αυτή.

### 1.3 Ο εκτιμητής μέγιστης πιθανοφάνειας (EMΠ)

Έστω  $X = (X_1, X_2, \dots, X_n)$  είναι ένα τυχαίο δείγμα που προέρχεται από μία κατανομή με συνάρτηση πυκνότητας πιθανότητας  $f(x; \theta)$ . Η από κοινού συνάρτηση πυκνότητας πιθανότητας των  $X_1, X_2, \dots, X_n$ , λόγω της ανεξαρτησίας δίδεται από την σχέση

$$f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta). \quad (1.7)$$

Για παρατηρηθέν δείγμα  $x = (x_1, x_2, \dots, x_n)$  παραπάνω από κοινού κατανομή είναι πλέον μία συνάρτηση της παραμέτρου  $\theta$ . Η συνάρτηση αυτή λέγεται συνάρτηση πιθανοφάνειας και συμβολίζεται ως

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (1.8)$$

**Ορισμός:** Εκτιμητής Μέγιστης Πιθανοφάνειας (EMΠ) της παραμέτρου  $\theta$  είναι η τιμή που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας.

Στην πραγματικότητα ο EMΠ είναι η τιμή του  $\theta$  η οποία μεγιστοποιεί την πιθανότητα των παρατηρούμενων τιμών. Μία άλλη ερμηνεία είναι ότι ο EMΠ είναι η τιμή η οποία ελαχιστοποιεί την απόσταση Kullback- Leibler μεταξύ της θεωρητικής και της Εμπειρικής συνάρτησης κατανομής των δεδομένων. Δηλαδή, είναι η τιμή για την οποία τα δεδομένα προσαρμόζονται με τον «βέλτιστο τρόπο» στην κατανομή.

Για την εύρεση του EMΠ είναι πιο εύκολο αντί να μεγιστοποιηθεί την λογαριθμική συνάρτηση πιθανοφάνειας

$$l(\theta) = \ln L(\theta) = \ln f(x_1, x_2, \dots, x_n) = \ln \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \ln f(x_i; \theta) \quad (1.9)$$

Ο εκτιμητής Μέγιστης Πιθανοφάνειας  $\hat{\theta}$  ικανοποιεί τη σχέση

$$\frac{d}{d\theta} l(\hat{\theta}) = 0. \quad (1.10)$$

Αν η εξίσωση έχει πολλές ρίζες τότε θα πρέπει με το κριτήριο της δεύτερης παραγώγου να βρεθεί η ρίζα η οποία μεγιστοποιεί την συνάρτηση πιθανοφάνειας. Το πιο σοβαρό πρόβλημα υπάρχει στην περίπτωση κατά την οποία υπάρχουν δυο ολικά μέγιστα. Στην περίπτωση αυτή είναι δύσκολη η επιλογή της ρίζας η οποία θα μπορεί να θεωρηθεί ως εκτίμηση της  $\theta$ .

Η εκτίμηση με την μέθοδο της πιθανοφάνειας είναι η πιο δημοφιλής μέθοδος κυρίως για δυο λόγους:

- 1) Έχει την ιδιότητα του αναλλοίωτου δηλαδή αν  $\hat{\theta}_n$  είναι ο ΕΜΠ του  $\theta$  τότε ο  $g(\hat{\theta}_n)$  είναι ο ΕΜΠ του  $g(\theta)$ .
- 2) Κάτω από κάποιες συνθήκες ομαλότητας, ο  $\hat{\theta}_n$  είναι ασυμπτωτικά κανονικός και ασυμπτωτικά αποτελεσματικός. Δηλαδή αν ισχύουν οι συνθήκες ομαλότητας και  $\hat{\theta}_n$  είναι ο εκτιμητής μέγιστης πιθανοφάνειας του  $\theta$  τότε

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, I^{-1}(\theta)) \quad n \rightarrow \infty \quad (1.11)$$

## 1.4 Κατανομές απώλειας

Ως κατανομή απώλειας (loss distribution) μπορεί να χαρακτηριστεί οποιαδήποτε θετική κατανομή. Με άλλα λόγια κατανομές που έχουν πεδίο τιμών στο  $(0, +\infty)$  χαρακτηρίζονται ως κατανομές απώλειας. Ένα κοινό χαρακτηριστικό των κατανομών απώλειας είναι ότι παρουσιάζουν θετική ασυμμετρία. Μία καμπύλη συχνότητας παρουσιάζει θετική ασυμμετρία όταν οι περισσότερες παρατηρήσεις βρίσκονται δεξιά της κορυφής ( $\gamma > 0$ ).



Διάγραμμα 1: Θετική Ασυμμετρία

Ένα άλλο χαρακτηριστικό αυτών των κατανομών είναι ότι έχουν βαριές ουρές. Όπως φαίνεται και από το παραπάνω σχεδιάγραμμα θετική ασυμμετρία έχει ως αποτέλεσμα μία ουρά. Στην περίπτωση δε, των κατανομών απώλειας η ουρά αυτή τείνει να γίνει όλο και πιο βαριά.

Η έννοια της κατανομής ζημιών είναι συχνά συνδεδεμένη με το αναλογιστικό θέμα των «απροόπτων ζωής». Σε κάθε περίπτωση η έννοια αυτή μπορεί να διευρυνθεί και να συνδεθεί έτσι, με τα διάφορα είδη ασφαλιστηρίων συμβολαίων. Το μόνο που παρατηρείται είναι η κατανομή του μεγέθους των απωλειών, αφού είναι δεδομένο ότι έχει συμβεί μία τέτοια απώλεια.

Οι ζημιές που προκαλούνται από απρόσμενα γεγονότα είναι προβλήματα, τόσο ατομικά όσο και κοινωνικά. Η ασφάλιση είναι ένας μηχανισμός για να διαμεριστεί το κόστος των ζημιών αυτών. Σαφώς, ο αναλογιστής θέλει να γνωρίζει την πιθανότητα εμφάνισης των ζημιών. Είναι σημαντικό να είναι γνωστός ο αναμενόμενος αριθμός ζημιών. Για παράδειγμα, θα μπορούσε να παρατηρηθεί ο αριθμός των ζημιών που συμβαίνουν κατά τη διάρκεια του επόμενου έτους για μια συγκεκριμένη ομάδα ασφαλισμένων αυτοκινήτων.

Όσον αφορά στις κατανομές απώλειας, θα πρέπει να δοθεί προσοχή στο σκοπό για τον οποίο τα δεδομένα πρέπει να χρησιμοποιούνται. Εάν το σύνολο δεδομένων πρόκειται να χρησιμοποιηθεί για αποτίμηση των απαλλαγών ή των κρατήσεων, τότε θα πρέπει να γίνουν tests καλής προσαρμογής στα χαμηλότερα ποσά ζημίας. Από την άλλη, εάν το πρόβλημα είναι να αυξηθούν τα όρια, τότε θα πρέπει να δοθεί έμφαση στο μοντέλο και τις παραμέτρους που έχουν επιλεγεί και θα πρέπει να γίνουν tests καλής προσαρμογής στα υψηλότερα ποσά ζημίας.

## 1.5 Μεθοδολογία

Η επίδραση των διαφόρων ειδών λογοκρισίας στην εκτίμηση παραμέτρων θα γίνει σε σχέση με το ποσοστό λογοκρισίας ( $p$ ) και λαμβάνοντας υπόψη το μέγεθος του δείγματος. Έτσι όταν το δείγμα είναι μεγάλο θα συγκριθεί η ασυμπτωτική διακύμανση των εκτιμητών ή ισοδύναμα την απώλεια πληροφορίας σε σχέση πάντα με το  $p$ . Πιο συγκεκριμένα θα υπολογίζεται ο λόγος  $I_p(\theta)/I(\theta)$  ως συνάρτηση του  $p$  και για διάφορες τιμές της  $\theta$ . Όταν ο λόγος δεν μπορεί να βρεθεί σε κλειστή μορφή, τα αποτελέσματα παρουσιάζονται γραφικά. Για να μελετηθεί η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων στην περίπτωση μικρών δειγμάτων θα χρησιμοποιηθούν προσομοιωμένα δείγματα. Για τις διάφορες κατανομές που μελετώνται ελήφθησαν 5.000 δείγματα για διάφορες επιλογές των παραμέτρων και για διάφορες τιμές του ποσοστού  $p$  λογοκριμένου δείγματος. Υπολογίζονται οι εκτιμητές μέγιστης πιθανοφάνειας των αγνώστων παραμέτρων και παρατίθενται σε πίνακα η μέση τιμή και η διακύμανση των εκτιμητών αυτών. Η παραπάνω διαδικασία γίνεται με την βοήθεια κατάλληλων προγραμμάτων στο Mathematica, τα οποία παρατίθενται στο παραρτήματα.

## ΚΕΦΑΛΑΙΟ 2

### Τι είναι η λογοκρισία

#### 2.1 Εισαγωγή

Στη στατιστική, λογοκρισία (censoring) είναι μία κατάσταση κατά την οποία οι παρατηρήσεις είναι μερικώς γνωστές.

Μια εταιρεία κατασκευής ανταλλακτικών αεροπλάνου καταγράφει την διάρκεια ζωής ενός ανταλλακτικού. Αν υπάρξει πρόβλημα, το ανταλλακτικό αντικαθίσταται αμέσως και καταγράφεται η διάρκεια ζωής του μέχρι την στιγμή της αντικατάστασης. Σε διαφορετική περίπτωση αφήνεται να λειτουργήσει μέχρι κάποιες ώρες πτήσεις ( $T$ ) και στη συνέχεια αντικαθίσταται. Αυτό που καταγράφεται λοιπόν είναι το  $\min\{X, T\}$  όπου  $X$  είναι ο χρόνος ζωής του ανταλλακτικού είτε ο χρόνος αντικατάστασης  $T$ .

Σε μια κλινική μελέτη δοκιμής ενός νέου φαρμάκου που υπόσχεται μακροζωία στους καρκινοπαθείς, οι ασθενείς πολλές φορές αποχωρούν από το πείραμα. Και στην περίπτωση αυτή έχουμε την περίπτωση της λογοκρισίας.

Λογοκρισία συμβαίνει επίσης όταν μια τιμή βρίσκεται εκτός του εύρους του οργάνου μετρήσεων. Για παράδειγμα, μια ζυγαριά μπάνιου μπορεί να μετρήσει μόνο έως  $140\text{ kg}$ . Εάν ένα άτομο  $160\text{ kg}$ , ζυγίζεται με τη χρήση της ζυγαριάς, ο παρατηρητής θα γνωρίζει μόνο ότι το βάρος του ατόμου είναι τουλάχιστον  $140\text{ kg}$ .

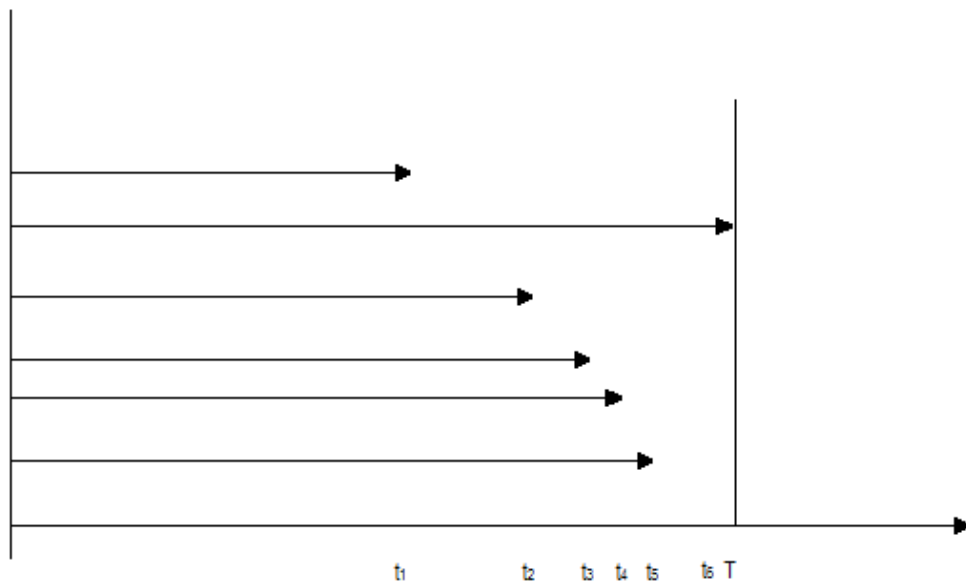
Το πρόβλημα των λογοκριμένων δεδομένων, όπου η παρατηρούμενη τιμή κάποιας μεταβλητής είναι εν μέρει γνωστή, σχετίζεται με το πρόβλημα των ελλιπών δεδομένων, όπου οι παρατηρούμενες τιμές κάποιων μεταβλητών είναι άγνωστες.

Η λογοκρισία, δεν θα πρέπει να συγχέεται με την έννοια της περικοπής των δεδομένων. Με τη λογοκρισία, έχω το ακριβές αποτέλεσμα είτε λαμβάνοντας υπόψη όλες τις παρατηρήσεις είτε λαμβάνοντας υπόψη μόνο τις παρατηρήσεις που βρίσκονται εντός ενός διαστήματος.

Με την περικοπή (truncated), οι παρατηρήσεις ποτέ δεν οδηγούν σε τιμές εκτός ενός συγκεκριμένου εύρους. Παρατηρήσεις που βρίσκονται εκτός αυτού του εύρους είτε δεν παρατηρούνται ή ακόμα και αν παρατηρηθούν, δεν καταγράφονται. Να σημειωθεί πως σε μία στατιστική μελέτη η περικοπή δεν ταυτίζεται με τη στρωγγυλοποίηση.

Παρακάτω παρουσιάζονται συνοπτικά οι κυριότερες μορφές λογοκρισίας.

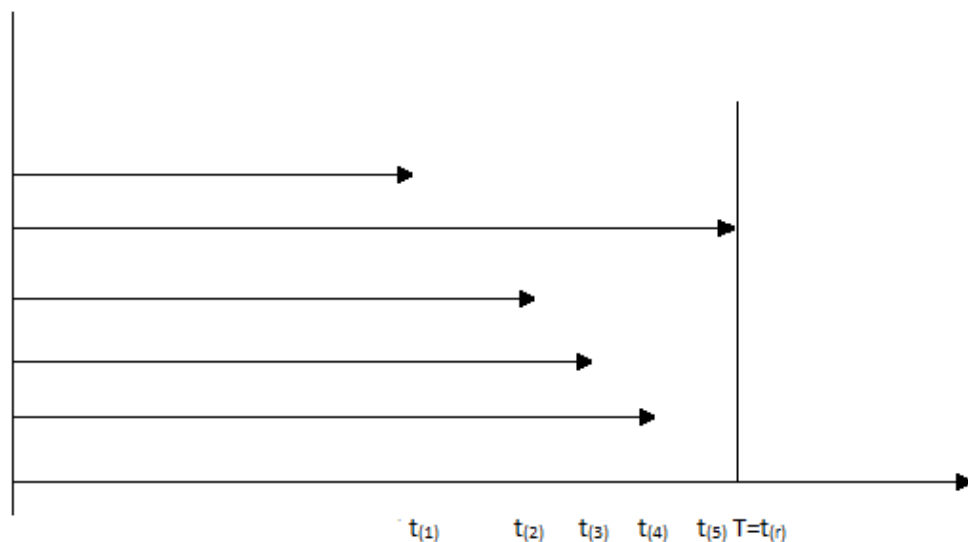
- **Αριστερή λογοκρισία** - ένα σημείο είναι κάτω από μια ορισμένη τιμή, αλλά είναι άγνωστο κατά πόσο.
- **Διάστημα λογοκρισίας** - ένα σημείο βρίσκεται σε ένα διάστημα μεταξύ δύο τιμών χωρίς να είναι γνωστή η ακριβής τιμή της.
- **Δεξιά λογοκρισία** - ένα σημείο είναι πάνω από μια ορισμένη τιμή, αλλά είναι άγνωστο κατά πόσο.
- **Λογοκρισία Τύπου I** συμβαίνει σε ένα πείραμα που έχει ένα συγκεκριμένο αριθμό ατόμων ή αντικειμένων και σταματά σε ένα προκαθορισμένο χρονικό διάστημα. Ένα δείγμα εμφανίζει **λογοκρισία Τύπου I**, όταν τα όρια λογοκρισίας ( $T$ ) είναι γνωστά. Το πλήθος των μη λογοκριμένων παρατηρήσεων θα συμβολίζεται με  $r$  και των λογοκριμένων με  $n-r$ . Το  $n$  είναι γνωστό εκ των προτέρων και προκαθορισμένο και το  $r$  άγνωστο.



Διάγραμμα 2: Λογοκρισία Τύπου I



- Λογοκρισία Τύπου II** συμβαίνει σε ένα πείραμα που έχει ένα συγκεκριμένο αριθμό ατόμων ή αντικειμένων και σταματά όταν ένας προκαθορισμένος αριθμός παρατηρήσεων έχουν αποτύχει, οι υπόλοιπες παρατηρήσεις είναι δεξιά λογοκριμένες. Ένα δείγμα εμφανίζει **λογοκρισία Τύπου II**, όταν αυτό που είναι γνωστό εκ των προτέρων είναι το  $r$  και όχι το μέγεθος του δείγματος. Επιπλέον το  $T$  είναι άγνωστο αλλά ορίζεται ως  $T = x_{(r)}$ . Τα λογοκριμένα δείγματα Τύπου II συνήθως ολοκληρώνονται μετά από ένα προκαθορισμένο αριθμό αποτυχιών και γι αυτό πολλές φορές ονομάζονται failure-censored δείγματα. Ένα μειονέκτημα της λογοκρισίας τύπου II είναι ότι δεν είναι γνωστός ο χρόνος λήξης του πειράματος.



Διάγραμμα 3: Λογοκρισία Τύπου II

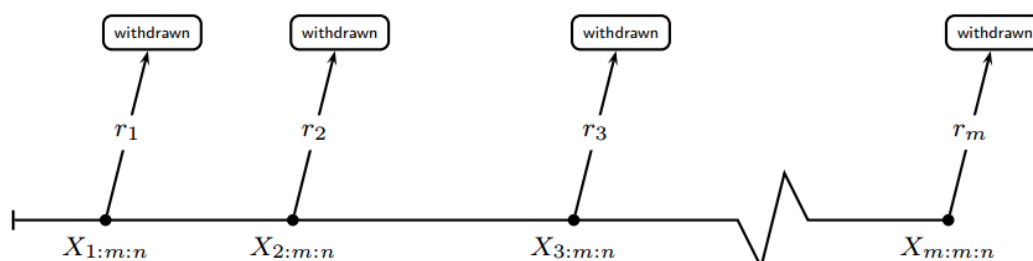
- Τυχαία λογοκρισία** - ένα δείγμα είναι Τυχαία λογοκριμένο, όταν τα όρια λογοκρισίας ( $T$ ) είναι τυχαίες μεταβλητές. Μπορεί να θεωρηθεί ότι η τυχαία λογοκρισία είναι μια γενίκευση της Λογοκρισίας τύπου I στην οποία καταγράφονται την τιμή το  $\min\{X, T\}$ . Στην περίπτωση της τυχαίας λογοκρισίας στην παρατήρηση  $i$  καταγράφεται το  $\min\{X_i, T_i\}$  όπου  $X_i$  είναι η διάρκεια ζωής του αντικειμένου και  $T_i$  το όριο της

λογοκρισίας το οποίο είναι στοχαστικό. Για παράδειγμα, σε μια κλινική δοκιμή το όριο λογοκρισίας διαφέρει από ασθενή σε ασθενή αφού μπορεί να αποχωρήσει από το πείραμα όποτε το επιθυμήσει.

- **Προοδευτική λογοκρισία (Progressive censoring)**

Η προοδευτική λογοκρισία είναι πολύ χρήσιμη στα πειράματα ζωής και σε κλινικές μελέτες και βασίζεται σε καταγεγραμμένες αξίες. Κατά την επέκταση του μοντέλου της προοδευτικής λογοκρισίας τύπου II, επιτρέπεται να επιλέξει τον επόμενο αριθμό λογοκρισίας λαμβάνοντας υπόψη τόσο τον προηγούμενο αριθμό λογοκρισίας όσο και τις προηγούμενες αποτυχίες. Έχοντας ορισμένα αποτελέσματα, φαίνεται ότι οι εκτιμητές μέγιστης πιθανοφάνειας συμπίπτουν με εκείνους της προοδευτικής λογοκρισίας τύπου II.

Στην προοδευτική λογοκρισία τύπου II, ο αριθμός των αποτυχιών ( $r_1, \dots, r_m$ ) είναι γνωστός εκ των προτέρων. Επιπλέον γνωστά είναι, ο αριθμός των μονάδων ( $n$ ) και ο αριθμός των παρατηρούμενων αποτυχιών ( $m$ ). Ξεκινώντας όλες τις  $n$  μονάδες την ίδια στιγμή, το πρώτο βήμα της προοδευτικής λογοκρισίας γίνεται κατά την παρατήρηση της πρώτης αποτυχίας στον χρόνο  $X_{1:m:n}$ . Τη συγκεκριμένη στιγμή, οι  $r_1$  μονάδες επιλέγονται τυχαία από τις μονάδες που εξακολουθούν να λειτουργούν, ενώ οι υπόλοιπες μονάδες ( $n - r_1 - 1$ ) συνεχίζουν να λειτουργούν κανονικά. Κατά την επόμενη αποτυχία στο χρόνο  $X_{2:m:n}$ , επιλέγονται τυχαία οι  $r_2$  μονάδες. Η διαδικασία συνεχίζεται έως την  $m$ -ιοστή αποτυχία. Το σύνολο των αποτυχιών  $X_{1:m:n}, \dots, X_{m:m:n}$ , ονομάζεται προοδευτική λογοκρισία τύπου II και απεικονίζεται στο παρακάτω διάγραμμα.



Διάγραμμα 4: Προοδευτική Λογοκρισία Τύπου II

Στην εργασία αυτή θα γίνει μελέτη σχετικά με τη δεξιά λογοκρισία τύπου I τύπου II και την τυχαία λογοκρισία οι οποίες θα αναπτυχθούν λεπτομερώς στη συνέχεια.

## 2.2 Δεξιά λογοκρισία τύπου I

Στην Λογοκρισία τύπου I ο χρόνος λογοκρισίας  $T$  είναι σταθερά και γνωστή εκ των προτέρων.

Αν  $X_1, X_2, \dots, X_n$  είναι ένα δείγμα από μια κατανομή  $f(x, \theta)$ , τότε αυτό που καταγράφεται είναι η

$$Y_i = \min\{X_i, T\} \quad (2.1)$$

η κατανομή της οποίας είναι η

$$f(y, \theta) = \begin{cases} f(y; \theta) & x < T \\ \bar{F}(T; \theta) & x > T \end{cases} \quad (2.2)$$

Η συνάρτηση πιθανοφάνειας μιας παρατήρησης είναι

$$L(\theta) = f(x; \theta)^t \bar{F}(T; \theta)^{1-t} \\ t = 1 \quad \text{όταν } x < T \quad \text{και } t = 0 \quad \text{όταν } x > T \quad (2.3)$$

Για δείγμα  $y_1, y_2, \dots, y_n$  η συνάρτηση πιθανοφάνειας είναι η

$$L(\theta) = \left[ \prod_{i=1}^r f(x_i; \theta) \right] (\bar{F}(T; \theta))^{n-r} \quad (2.4)$$

όπου  $r$  είναι ο αριθμός των  $x_i < T$ .

Για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας λύνεται το σύστημα των εξισώσεων

$$\frac{\partial \log L(\theta)}{\partial \theta} = \sum_{i=1}^r \frac{\partial \log f(x_i; \theta)}{\partial \theta} + (n-r) \frac{\partial \log \bar{F}(T; \theta)}{\partial \theta} = 0 \quad (2.5)$$

Ο πληροφοριακός αριθμός για μια παρατήρηση είναι

$$I(\theta) = E \left[ \left( \frac{\partial \log L(\theta)}{\partial \theta} \right)^2 \right] = \int_0^T \left( \frac{\partial \log L(\theta)}{\partial \theta} \right)^2 f(x; \theta) dx + \left( \frac{\partial \log \bar{F}(T; \theta)}{\partial \theta} \right)^2 \bar{F}(T; \theta) \quad (2.6)$$

Οι εκτιμητές μέγιστης πιθανοφάνειας καθώς και οι έλεγχοι υποθέσεων που βασίζονται στην συνάρτηση πιθανοφάνειας λόγω της πολυπλοκότητας των μαθηματικών τύπων δεν είναι ελκυστικοί. Όμως οι ασυμπτωτικές τους ιδιότητες είναι γνωστές και για μεγάλα δείγματα μπορούν να χρησιμοποιηθούν απλοποιημένοι τύποι που δίνουν βέλτιστα αποτελέσματα. Για μικρά δείγματα, η συμπεριφορά των εκτιμητών μελετάται με μεγάλο αριθμό δειγμάτων προσομοίωσης.

Χαρακτηριστικά παραδείγματα Λογοκρισίας Τύπου I (όχι απαραίτητα δεξιάς Λογοκρισίας Τύπου I) είναι η περίπτωση της Αντασφάλειας και της απαλλαγής αποζημίωσης. Και στις δύο περιπτώσεις εμφανίζονται όλα τα δεδομένα, με την ιδιαιτερότητα ότι πάνω (στην περίπτωση της Αντασφάλειας) ή κάτω (στην περίπτωση της απαλλαγής) από το όριο λογοκρισίας  $T$ , οι καταγραφές είναι ίσες με αυτό.

### 2.3 Δεξιά λογοκρισία τύπου II

Η λογοκρισία τύπου II αναφέρεται στην περίπτωση όπου από ένα τυχαίο δείγμα μεγέθους  $n$ , παρατηρούνται μόνο οι  $r$  μικρότερες διάρκειες ζωής  $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(r)}$ . Οι υπόλοιπες  $n-r$  θεωρούνται λογοκριμένες στο σημείο  $T = t_{(r)}$ . Αυτού του τύπου λογοκρισία εμφανίζεται όταν  $n$  μονάδες μπαίνουν ταυτόχρονα στη διαδικασία μελέτης η οποία τερματίζεται όταν παρατηρηθεί η  $r$  αποτυχία. Αν και πολλά πειράματα τυποποιούνται ως λογοκρισία τύπου II αυτά έχουν το μειονέκτημα ο χρόνος τερματισμού του πειράματος  $t_{(r)}$  είναι τυχαίος και έτσι ο προγραμματισμός του πειράματος δύσκολος. Ενώ όμως δεν μπορεί να ελεγχθεί η διάρκεια του πειράματος, μπορεί να ελεγχθεί το κόστος αυτού αφού εκ των προτέρων αποφασίζεται να «θυσιαστούν»  $r$  μονάδες των οποίων το κόστος είναι ελεγχόμενο. Για τον λόγο αυτό πειράματα με λογοκρισία τύπου II χρησιμοποιούνται κυρίως από μηχανικούς για το έλεγχο ακριβών μηχανικών συστημάτων.

Στην περίπτωση αυτή το δείγμα αποτελείται από τις  $r$  μικρότερες παρατηρήσεις από ένα τυχαίο δείγμα  $T_1, T_2, \dots, T_n$ . Για τις συνεχείς κατανομές μπορεί να αγνοηθεί η πιθανότητα της ισότητας μεταξύ τους (ties). Αν η  $T_i$  έχει π.π.  $f(t)$  και συνάρτηση επιβίωσης

$S(t)$  τότε η από κοινού κατανομή των  $T_{(1)} < T_{(2)} < \dots < T_{(r)}$  είναι

$$f(t_{(1)}, t_{(2)}, \dots, t_{(r)}; \theta) = \frac{n!}{(n-r)!} \{\prod_{i=1}^r f(t_{(i)})\} S(t_{(r)})^{n-r} \quad (2.7)$$

Ο συντελεστής  $\frac{n!}{(n-r)!}$  δηλώνει το αριθμό των διατεταγμένων  $r$ -άδων χωρίς επανατοποθέτηση από ένα σύνολο  $n$  αντικειμένων.

Η συνάρτηση πιθανοφάνειας είναι

$$L(\theta) = \frac{n!}{(n-r)!} \{\prod_{i=1}^r f(t_{(i)})\} S(t_{(r)})^{n-r} \quad (2.8)$$

Οι εκτιμητές μέγιστης πιθανοφάνειας βρίσκονται από την λύση του συστήματος

$$\frac{\partial \log L(\theta)}{\partial \theta} = \sum_{i=1}^r \frac{\partial \log f(x_{(i)}; \theta)}{\partial \theta} + (n-r) \frac{\partial \log \bar{F}(t_{(r)}; \theta)}{\partial \theta} = 0 \quad (2.9)$$

Συγκρίνοντας τις (2.5) και (2.9) παρατηρείται ότι το σύστημα ή η εξίσωση που πρέπει να λυθεί είναι το ίδιο για  $T = t_{(r)}$  και στους δυο τύπου λογοκρισίας.

Σημαντική λεπτομέρεια : Στην λογοκρισία τύπου II δεν είναι εφικτό να γραφεί η (2.9) για μια παρατήρηση αφού απαιτείται να υπάρχουν τουλάχιστον  $r$  παρατηρήσεις.

Τότε πως ορίζεται ο πληροφοριακός αριθμός για μια παρατήρηση;

Από την (2.9) ορίζεται ο πληροφοριακός αριθμός

$$I_{n,r}(\theta) = E_{\theta} \left[ \left\{ \frac{\partial}{\partial \theta} \log L(\theta) \right\}^2 \right] \quad (2.10)$$

Για  $t_{ij} = E \left[ \frac{\partial \log f(x_i; \theta)}{\partial \theta} \frac{\partial \log f(x_j; \theta)}{\partial \theta} \right]$  η (2.10) γράφεται ως

$$I_{n,r}(\theta) = \sum_{i=1}^r t_{ii} + \frac{2}{n-r-1} \sum_{i=r+1}^{n-1} \sum_{j=i+1}^n t_{ij} \quad (2.11)$$

Η πληροφορία ανά μονάδα ορίζεται ως το όριο

$$\lim_{n \rightarrow \infty} \frac{I_{n,r}(\theta)}{n}$$

Αποδεικνύεται ότι αν  $n/r \rightarrow p$  καθώς  $n \rightarrow \infty$  τότε

$$\frac{I_{nr}(\theta)}{n} \rightarrow \int_0^{\xi_p} \left( \frac{\partial \log f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) dx + \left[ \frac{\partial \log \bar{F}(\xi_p; \theta)}{\partial \theta} \right]^2 \bar{F}(T; \theta). \quad (2.12)$$

Δηλαδή ασυμπτωτικά οι δυο τύποι είναι ίδιοι για  $T = \xi_p = F^{-1}(p)$ .

Εδώ  $p$  είναι το ποσοστό του δείγματος που δεν λογοκρίνεται.

## 2.4 Τυχαία λογοκρισία (Random Censoring)

Στην λογοκρισία τύπου I θεωρείται γνωστός και προκαθορισμένος ο χρόνος λογοκρισίας  $T$ . Στην τυχαία λογοκρισία τα άτομα ξεκινούν από τυχαίες στιγμές, έτσι ώστε τόσο οι ζωές όσο και οι φορές που θα λογοκριθούν οι τιμές να είναι τυχαίες.

Ορίζονται:

$T_i$  = η διάρκεια ζωής του  $i$  ατόμου

$C_i$  = η στιγμή της λογοκρισίας του  $i$  ατόμου

Επίσης γίνεται η υπόθεση:

- $T_i$  και  $C_i$  ανεξάρτητες και τυχαίες μεταβλητές
- $T_1, \dots, T_n$  είναι ανεξάρτητες παρατηρήσεις από την ίδια κατανομή με συνάρτηση πυκνότητας πιθανότητας  $f(t)$  και συνάρτηση επιβίωσης  $S(t)$ .
- $C_1, \dots, C_n$  είναι ανεξάρτητες παρατηρήσεις από την ίδια κατανομή με συνάρτηση πυκνότητας πιθανότητας  $g(\ell)$  και της συνάρτηση επιβίωσης  $G(\ell)$

Αυτό σημαίνει:

$$Pr(T) = f(t)$$

$$Pr(T > t) = S(t)$$

$$Pr(C) = g(c)$$

$$Pr(C > c) = G(c).$$

Ορίζονται όπως πριν:

$$t_i = \min(T_i, C_i)$$

$$\delta_i = \begin{cases} 1, & T_i \leq C_i \\ 0, & T_i > C_i \end{cases}$$

Στο πείραμά λοιπόν παρατηρούνται τα ζεύγη  $(t_i, \delta_i) \quad i=1, \dots, n$ .

Η συνάρτηση πυκνότητας πιθανότητας για  $(t_i, \delta_i)$  είναι:

$$Pr(t_i = t, \delta_i = 0) = Pr(C_i = t, T_i > C_i) = g(t)S(t) \quad (2.13)$$

$$Pr(t_i = t, \delta_i = 1) = Pr(T_i = t, T_i \leq C_i) = f(t)G(t) \quad (2.14)$$

Από τις δύο τελευταίες σχέσεις προκύπτει:

$$Pr(t_i = t, \delta_i) = [f(t)G(t)]^{\delta_i} [g(t)S(t)]^{1-\delta_i} \quad (2.15)$$

Έτσι, για ανεξάρτητες  $n$  και παρατηρήσεις  $(t_1, \delta_1), \dots, (t_n, \delta_n)$  η συνάρτηση πιθανότητας είναι:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(t_i; \theta)G(t_i; \theta)]^{\delta_i} [g(t_i; \theta)S(t_i; \theta)]^{1-\delta_i} \\ &= \underbrace{\left( \prod_{i=1}^n G(t_i; \theta)^{\delta_i} g(t_i; \theta)^{1-\delta_i} \right)}_{\text{Χρόνος λογοκρισίας}} \underbrace{\left( \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i} \right)}_{\text{Διάρκεια ζωής}} \end{aligned} \quad (2.16)$$

Ενδέχεται το  $G$  και το  $g$ , τα οποία εκφράζουν τον χρόνο λογοκρισίας, να μην εξαρτώνται από την άγνωστη παράμετρο  $\theta$ . Σε αυτή την περίπτωση στη συνάρτηση πιθανοφάνειας του τυχαία λογοκριμένου δείγματος μπορεί να αγνοηθεί η πρώτη παρένθεση που αφορά τον χρόνο λογοκρισίας και τελικά η συνάρτηση πιθανοφάνειας γίνεται

$$L(\theta) \propto \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i} \quad (2.17)$$

Παρατηρείται δηλαδή ότι συμπίπτει με τη συνάρτηση πιθανότητας της λογοκρισίας Τύπου I και τύπου II. Δηλαδή η παράγωγος της λογαριθμικής συνάρτησης πιθανοφάνειας θα δώσει τις ίδιες εξισώσεις για την εκτίμηση των αγνώστων παραμέτρων.

Η πληροφορία του Fisher γράφεται ως

$$I^{T,C}(\theta) = \int \left( \frac{\partial}{\partial \theta} \log(f\bar{G}) \right)^2 f\bar{G} dx + \int \left( \frac{\partial}{\partial \theta} \log(g\bar{F}) \right)^2 g\bar{F} dx. \quad (2.18)$$

Η πολυπλοκότητα του παραπάνω τύπου έγκειται στο ότι εμπλέκονται δυο κατανομές: του χρόνου ζωής και του χρόνου λογοκρισίας. Στην περίπτωση που ο χρόνος λογοκρισίας είναι ανεξάρτητος της παραμέτρου  $\theta$ , ο πληροφοριακός αριθμός γίνεται

$$I^{T,C}(\theta) = \int \left( \frac{\partial}{\partial \theta} \log f \right)^2 f \bar{G} dx + \int \left( \frac{\partial}{\partial \theta} \log \bar{F} \right)^2 g \bar{F} dx. \quad (2.19)$$

Ακόμα και στην περίπτωση αυτή ο υπολογισμός του πληροφοριακού αριθμού είναι δύσκολος και απαιτεί την υπόθεση μιας κατανομής για τον χρόνο λογοκρισίας. Συνήθως η κατανομή αυτή λαμβάνεται μάλλον αυθαίρετα και μάλιστα κάτω από την υπόθεση ότι είναι της ίδιας οικογένειας με την κατανομή του χρόνου ζωής. Μια άλλη μεθοδολογία είναι να γίνει η υπόθεση ότι ο χρόνος λογοκρισίας ακολουθεί την ομοιόμορφη κατανομή.

Οι Zheng and Gastwirth (2001) έγραψαν την παραπάνω σχέση στην μορφή

$$I^{T,C}(\theta) = \int \left( \frac{\partial}{\partial \theta} \log \lambda_T \right)^2 f \bar{G} dx + \int \left( \frac{\partial}{\partial \theta} \log \lambda_C \right)^2 g \bar{F} dx. \quad (2.20)$$

Οι ίδιοι συγγραφείς γράφουν την πληροφορία της τ.μ.  $T$  ως εξής

$$I^T(\theta) = \int \left( \frac{\partial}{\partial \theta} \log \lambda_T \right)^2 f \bar{G} dx + \int \left( \frac{\partial}{\partial \theta} \log \lambda_T \right)^2 f G dx = I_R^T(\theta) + I_L^T(\theta) \quad (2.21)$$

και ερμηνεύουν την  $I_R^T(\theta)$  ως την πληροφορία της  $T$  που διατηρείται μετά την λογοκρισία και την  $I_L^T(\theta)$  την πληροφορία που χάνεται με αυτήν .

Παρομοίως

$$I^C(\theta) = \int \left( \frac{\partial}{\partial \theta} \log \lambda_C \right)^2 g \bar{F} dx + \int \left( \frac{\partial}{\partial \theta} \log \lambda_C \right)^2 g F dx = I_R^C(\theta) + I_L^C(\theta) \quad (2.22)$$

Επομένως ο πληροφοριακός αριθμός  $I^{T,C}(\theta)$  γράφεται ως

$$I^{T,C}(\theta) = I_R^T(\theta) + I_R^C(\theta) \quad (2.23)$$

δηλαδή είναι το άθροισμα των πληροφοριακών αριθμών των τ.μ.  $T$  και  $C$  που διατηρούνται μετά την λογοκρισία.

Το ενδιαφέρον με την τύπο της λογοκρισίας αυτής είναι ότι σε αντίθεση με την λογοκρισία τύπου I και II ο πληροφοριακός αριθμός του λογοκριμένου δείγματος  $I^{T,C}(\theta)$  δεν είναι πάντα μικρότερος από τον πληροφοριακό αριθμό  $I^T(\theta)$  του μη λογοκριμένου δείγματος. Αυτό



εξαρτάται από την σχέση των  $I_R^C(\theta)$  και  $I_L^T(\theta)$ . Δηλαδή αν η πληροφορία που χάνεται από την τ.μ.  $T$  λόγω της λογοκρισίας είναι μεγαλύτερη ίση ή μεγαλύτερη από την πληροφορία που διατηρείται από την τ.μ.  $C$  τότε η  $I^{T,C}(\theta)$  είναι μικρότερη ίση ή μεγαλύτερη από την  $I^T(\theta)$ .

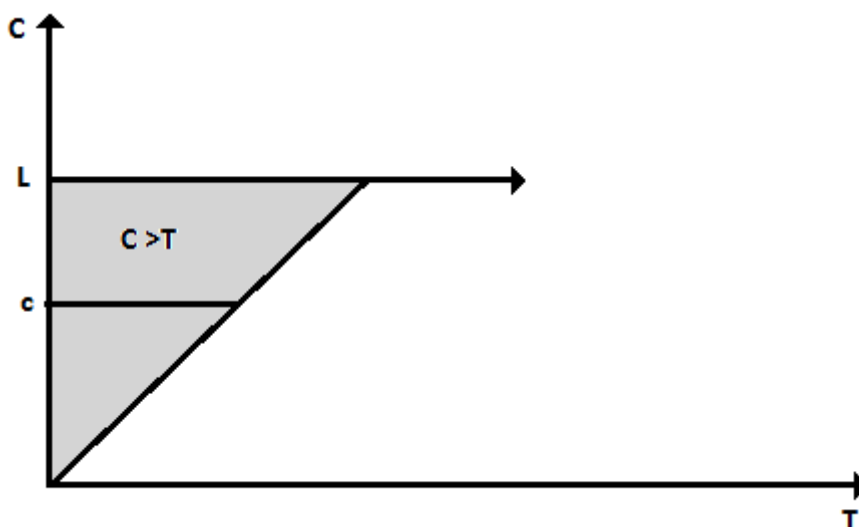
Ένα άμεσο αποτέλεσμα είναι ότι αν ο χρόνος λογοκρισίας είναι ανεξάρτητο του  $\theta$  τότε έχουμε πάντα απώλεια πληροφορίας

$$I^{T,C}(\theta) = I_R^T(\theta) < I^T(\theta). \quad (2.24)$$

Στην εργασία αυτή θα μελετηθεί η δεξιά λογοκρισία τύπου I τύπου II και τυχαία λογοκρισία για την οποία θα γίνει η υπόθεση ότι οι χρόνοι λογοκρισίας ακολουθούν την ομοιόμορφη κατανομή  $C \sim U[0, L]$ . Προφανώς υπάρχει σχέση μεταξύ του ποσοστού  $p$  και του  $L$ . Αυτό μπορεί να βρεθεί θεωρώντας ότι αν  $p$  είναι το ποσοστό του λογοκριμένου δείγματος και λογοκρισία υπάρχει όταν  $T_i > C_i$   $i = 1, 2, \dots, n$  τότε προκύπτει η σχέση

$$1 - p = P(C_i > T_i) \quad (2.25)$$

Λύνοντας την σχέση αυτή ως προς  $L$  θα προκύψει το άνω φράγμα της ομοιόμορφης η οποία θα δώσει το επιθυμητό ποσοστό λογοκρισίας. Αν  $f$  η πυκνότητα πιθανότητας του χρόνου  $t$  και  $g$  της ομοιόμορφης  $[0, L]$  τότε ολοκληρώνοντας στο παρακάτω σχήμα



Διάγραμμα 5: Περιοχή Ολοκλήρωσης

$$\begin{aligned}1 - p = P(C > T) &= \int_0^L \int_0^c f(t)g(c) dt dc \\ &= \frac{1}{L} \int_0^L \int_0^c f(t) dt dc \\ &= \frac{1}{L} \int_0^L F(c) dc\end{aligned}\tag{2.26}$$

## ΚΕΦΑΛΑΙΟ 3

Η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων για την κατανομή Weibull

### 3.1 Εισαγωγή

Η κατανομή Weibull πήρε το όνομα της από τον σουηδό φυσικό Waloddi Weibull, ο οποίος το 1939 τη χρησιμοποίησε για να παρουσιάσει την αντοχή των υλικών στη θραύση. Η ταύτιση μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών από τη νέα του εφαρμογή ήταν εντυπωσιακή. Έκτοτε, χρησιμοποίησε την κατανομή αυτή για να μοντελοποιήσει δεδομένα και να λύσει προβλήματα που απασχολούσαν την καθημερινότητα. Έτσι μπορούσε να βρει την αντοχή του χάλυβα στον αέρα, τη δύναμη των ινών του ινδικού βαμβακιού, τη διάρκεια ζωής του ST-37 χάλυβα, το ανάστημα ενός ενήλικα γεννημένου στις Βρετανικές Νήσους ή το μήκος των φασολιών *Phaseolus vulgaris*.

Νωρίτερα το 1933, οι Rosen και Rammler (1933) είχαν χρησιμοποιήσει αυτή την κατανομή προκειμένου να περιγράψουν τη συμπεριφορά του κονιοποιημένου άνθρακα.

Λόγω της φύσης της συνάρτησης κινδύνου που περιγράφηκε παραπάνω, η κατανομή Weibull είναι η πλέον κατάλληλη κατανομή για τις περιπτώσεις όπου οι συνθήκες τυχαιότητας δεν ικανοποιούνται.

Μια τυχαία μεταβλητή  $X$  ακολουθεί την κατανομή Weibull με παράμετρο  $\theta$  και  $\gamma$ , εάν η συνάρτηση κατανομής είναι

$$F(x; \theta, \gamma) = 1 - \exp(-(x/\theta)^\gamma) \quad (3.1)$$

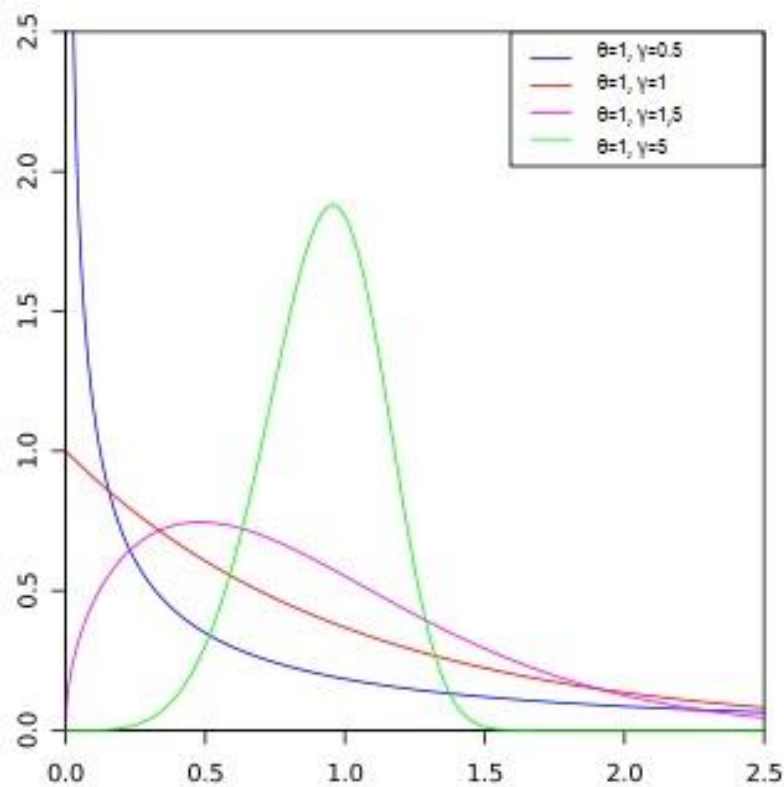
όπου  $\theta$  είναι η παράμετρος κλίμακας (scale parameter) και  $\gamma$  η παράμετρος μορφής (shape parameter) και συμβολίζεται  $X \sim W(\theta, \gamma)$ .

Η συνάρτηση πυκνότητας πιθανότητας της  $W(\theta, \gamma)$  κατανομής είναι η

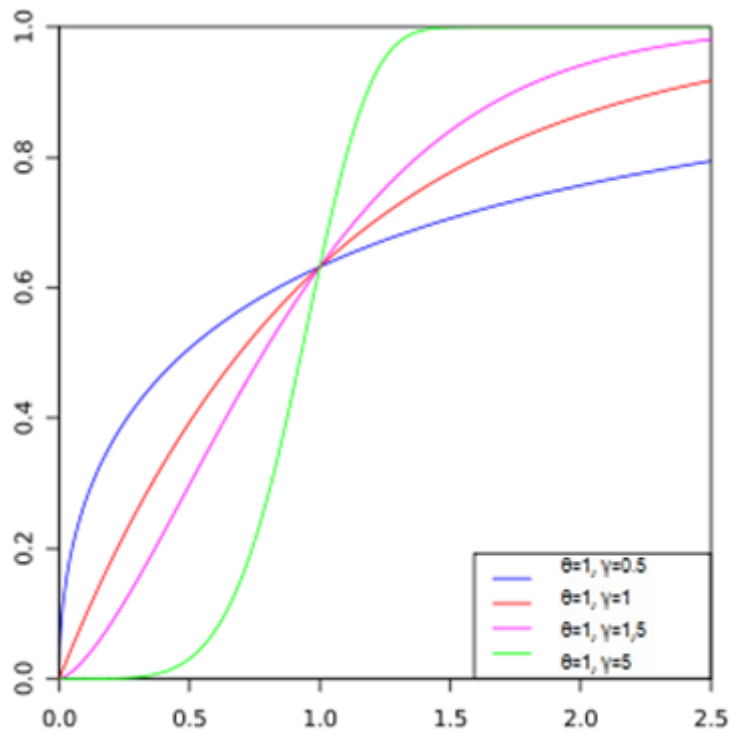
$$f(x; \theta, \gamma) = \gamma \frac{x^{\gamma-1}}{\theta^\gamma} \exp[-(x/\theta)^\gamma], x > 0, \theta > 0, \gamma > 0. \quad (3.2)$$

Η κατανομή Weibull είναι γενίκευση της εκθετικής κατανομής, με παράμετρο  $\lambda$  και  $\gamma=1$ . Σε αντίθεση όμως με την εκθετική δεν χαρακτηρίζεται από σταθερή συνάρτηση κινδύνου και γι αυτό έχει ευρύτερες εφαρμογές. Η κατανομή της ουράς καθορίζει την κατανομή Weibull, μία κατά τ' άλλα πολύ ευέλικτη κατανομή, που μπορεί να χρησιμοποιηθεί ως μοντέλο για τις απώλειες στον τομέα των ασφαλειών, συνήθως με  $\gamma < 1$ .

Γενικά, ο ρυθμός κινδύνου αυξάνεται όταν  $\gamma > 1$  και μειώνεται όταν  $\gamma < 1$ , καθώς ο χρόνος  $t$  αυξάνεται.



Διάγραμμα 6: Κατανομή Weibull για διάφορες τιμές παραμέτρων (πηγή Wikipedia)



Διάγραμμα 7: Αθροιστική συνάρτηση κατανομής Weibull (πηγή Wikipedia)

### 3.2 Εκτιμητής μέγιστης πιθανοφάνειας για πλήρες δείγμα

Έστω  $x_1, x_2, \dots, x_n$  ένα τυχαίο δείγμα από την (3.2). Όπως έχουμε πει ο ΕΜΠ για τις παραμέτρους  $\theta$  και  $\gamma$  είναι η λύση του συστήματος

$$\begin{cases} \frac{\partial \ln L}{\partial \gamma} = \frac{1}{\gamma} + \frac{\sum \ln x_i}{n} - \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\theta}\right)^\gamma \ln \frac{x_i}{\theta} = 0 \\ \frac{\partial \ln L}{\partial \theta} = \frac{-n\gamma}{\theta} + \gamma \frac{\sum x_i^\gamma}{\theta^{\gamma+1}} = 0 \end{cases} \quad (3.3)$$

Το παραπάνω σύστημα μπορεί να γραφεί στην μορφή

$$\frac{1}{\gamma} - \sum_{i=1}^n \frac{\ln x_i}{n} + \frac{\sum_{i=1}^n x_i^\gamma \ln x_i}{\sum_{i=1}^n x_i^\gamma} = 0 \quad (3.4)$$

$$\theta = \left( \frac{\sum_{i=1}^n x_i^\gamma}{n} \right)^{1/\gamma} \quad (3.5)$$

Μπορεί να αποδειχθεί ότι η (3.4) έχει μοναδική λύση (τον ΕΜΠ  $\hat{\gamma}$ ) του  $\gamma$  η οποία μπορεί να βρεθεί με αριθμητικές μεθόδους. Στη συνέχεια από την (3.5) προκύπτει ο ΕΜΠ  $\hat{\theta}$  του  $\theta$ .

Ο πληροφοριακός αριθμός του Fisher είναι (δες Bain L., and Engelhardt (1991))

$$I(\theta, \gamma) = \begin{bmatrix} \frac{1.8237}{\gamma^2} & -\frac{1.4228}{\theta} \\ -\frac{0.4228}{\theta} & \left(\frac{\gamma}{\theta}\right)^2 \end{bmatrix} \quad (3.6)$$

### 3.3 Εκτιμητής μέγιστης πιθανοφάνειας για λογοκριμένο δείγμα τύπου I και II

Έστω  $x_1, x_2, \dots, x_n$  ένα λογοκριμένο δείγμα με ποσοστό λογοκρισίας  $p = 1 - q = \bar{F}(T)$ . Το δείγμα είναι της μορφής

$$\underbrace{x_1, x_2, \dots, x_n}_r, \underbrace{T, T, \dots, T}_{n-r}$$

Οι εξισώσεις για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας για ένα λογοκριμένο δείγμα με ποσοστό λογοκρισίας  $p$  δίδονται από το σύστημα (Qian (2012))

$$\begin{cases} \frac{\partial \ln L}{\partial \gamma} = \frac{1}{\gamma} + \frac{\sum \ln x_i}{r} - \left(\frac{\sum x_i}{\theta}\right)^\gamma \ln \frac{x_i}{\theta} - (n-r) \frac{1}{\bar{F}(T)} f(T) = 0 \\ \frac{\partial \ln L}{\partial \theta} = -r \frac{\gamma}{\theta} + \gamma \frac{\sum x_i^\gamma}{\theta^{\gamma+1}} - (n-r) \frac{1}{\bar{F}(T)} f(T) = 0 \end{cases}$$

Αν γίνει η υπόθεση  $p=0$  ή ισοδύναμα  $r = n$  προκύπτει το σύστημα εξισώσεων (3.3) για πλήρες δείγμα. Χρησιμοποιώντας την σχέση

$$\bar{F}(t) = e^{-\left(\frac{t}{\theta}\right)^\gamma}$$

καταλήγει στο σύστημα

$$\begin{cases} \frac{\partial \ln L}{\partial \gamma} = \frac{1}{\gamma} + \frac{\sum \ln x_i}{r} - \left(\frac{\sum x_i}{\theta}\right)^\gamma \ln \frac{x_i}{\theta} - (n-r) \left(\frac{T}{\theta}\right)^\gamma \ln \frac{T}{\theta} = 0 \\ \frac{\partial \ln L}{\partial \theta} = -r \frac{\gamma}{\theta} + \gamma \frac{\sum x_i^\gamma}{\theta^{\gamma+1}} - (n-r) \frac{\gamma}{\theta} \left(\frac{T}{\theta}\right)^\gamma = 0. \end{cases} \quad (3.7)$$

Ο πληροφοριακός πίνακας του Fisher ως συνάρτηση του  $p$  και των παραμέτρων  $\theta$  και  $\gamma$  είναι

$$I(p, \theta, \gamma) = \begin{bmatrix} I_{11}(p, \theta, \gamma) & I_{12}(p, \theta, \gamma) \\ I_{21}(p, \theta, \gamma) & I_{22}(p, \theta, \gamma) \end{bmatrix}$$

όπου

$$I_{11}(p, \theta, \gamma) = \frac{1}{\gamma^2} \int_0^{1-p} \{1 + \ln[-\ln(1-x)]\}^2 dx$$

$$I_{12}(p, \theta, \gamma) = -\frac{1}{\theta^2} \int_0^{1-p} 1 + \ln[-\ln(1-x)] dx$$

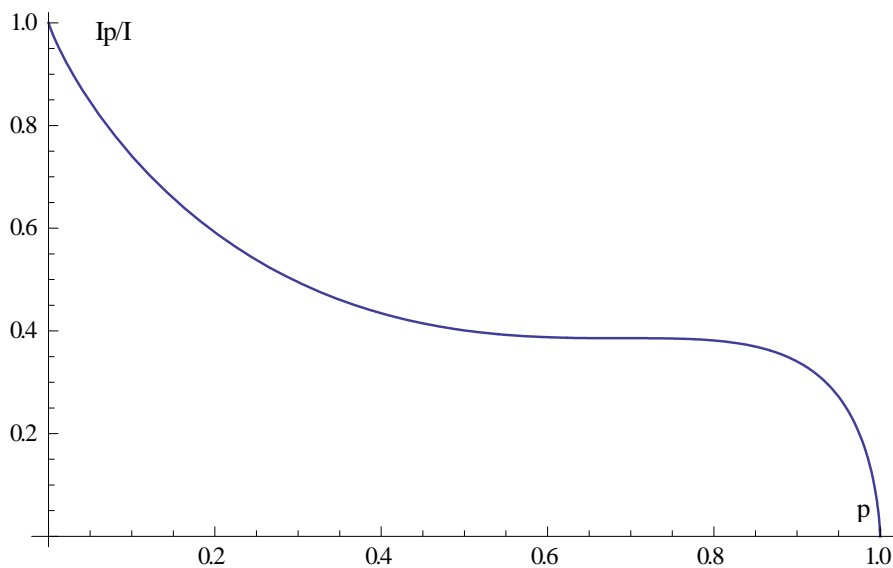
$$I_{22}(p, \theta, \gamma) = \left(\frac{\gamma}{\theta}\right)^2 (1-p)$$

όπου  $p$  είναι το ποσοστό λογοκρισίας.

Για  $p=0$  προκύπτει ο πληροφοριακός πίνακας του Fisher που αντιστοιχεί στη μη λογοκρισία.

Για να μελετηθεί η επίδραση του ποσοστού της λογοκρισίας στον πληροφοριακό αριθμό του Fisher όταν εκτιμάται η παράμετρος  $\gamma$ , μελετάται ο λόγος

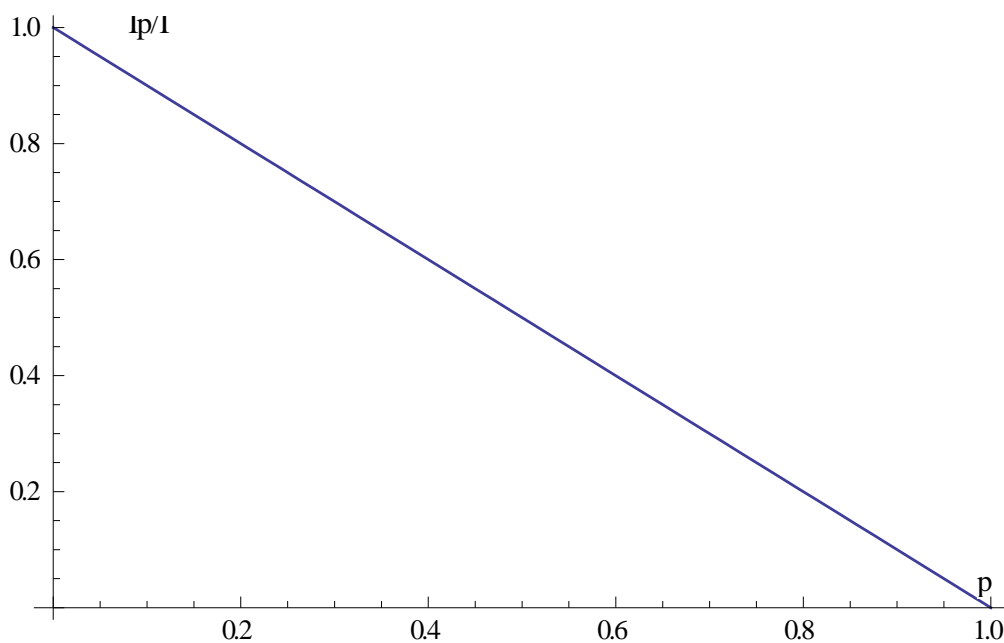
$$\frac{I_{11}(p, \theta, \gamma)}{I_{11}(0, \theta, \gamma)} = \frac{\int_0^{1-p} \{1 + \ln[-\ln(1-x)]\}^2 dx}{1.8237} \quad (3.8)$$



Διάγραμμα 8: Λόγος λογοκρίμένου δείγματος ( $p$ ) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο μορφής

Παρομοίως για να μελετηθεί η επίδραση του ποσοστού της λογοκρισίας στον πληροφοριακό αριθμό του Fisher, όταν εκτιμάται η παράμετρος θέσης  $\theta$  μελετάται ο λόγος

$$\frac{I_{22}(p,\theta,\gamma)}{I_{22}(0,\theta,\gamma)} = \frac{\left(\frac{\gamma}{\theta}\right)^2 (1-p)}{\left(\frac{\gamma}{\theta}\right)^2} = (1-p) \quad (3.9)$$



Διάγραμμα 9: Λόγος λογοκρίμένου δείγματος ( $p$ ) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο κλίμακας



Από το διάγραμμα προκύπτει πως η σχέση του λόγου των πληροφοριών ενός λογοκριμένου (με ποσοστό  $p$ ) δείγματος προς το μη λογοκριμένο δείγμα είναι γραμμική ως προς  $p$  και πιο συγκεκριμένα είναι ίσος με  $1-p$ . Με άλλα λόγια όσο αυξάνει το επίπεδο λογοκρισίας  $p$  η απώλεια της πληροφορίας φθίνει σταθερά.

### **Συγκριτική μελέτη των ΕΜΠ για μικρά δείγματα**

Για να μελετήσουμε την επίδραση του ποσοστού λογοκρισίας στην αποτελεσματικότητα του ΕΜΠ θα χρησιμοποιηθούν προσομοιωμένα δείγματα για διάφορες τιμές των  $n$ ,  $p$ ,  $\theta$  και  $\gamma$ . Κατασκευάστηκε το πρόγραμμα στο Mathematica το οποίο παράγει προσομοιωμένα δείγματα και δίνει τους ΕΜΠ (βλ. Παράρτημα Π1).

Τα αποτελέσματα παρουσιάζονται στους πίνακες που ακολουθούν

|     |                 | $\theta = 2$ |                    | $\gamma = 2$ |                    |
|-----|-----------------|--------------|--------------------|--------------|--------------------|
| n   | Censoring level | Mean         | Standard Deviation | Mean         | Standard Deviation |
| 50  | 0               | 1,993370     | 0,140519           | 2,035180     | 0,220560           |
|     | 10              | 1,996220     | 0,152064           | 2,015220     | 0,244941           |
|     | 20              | 2,007740     | 0,159706           | 2,028480     | 0,271623           |
|     | 50              | 2,014210     | 0,206257           | 2,062390     | 0,358233           |
|     | 60              | 2,027290     | 0,238536           | 2,064620     | 0,357784           |
|     | 70              | 2,042610     | 0,278605           | 2,073300     | 0,363570           |
| 80  | 0               | 1,995380     | 0,111041           | 2,022650     | 0,169615           |
|     | 10              | 2,002590     | 0,119374           | 2,010600     | 0,193889           |
|     | 20              | 2,005490     | 0,127097           | 2,018610     | 0,218220           |
|     | 50              | 2,012290     | 0,161958           | 2,042990     | 0,269053           |
|     | 60              | 2,015090     | 0,182952           | 2,045000     | 0,275592           |
|     | 70              | 2,026330     | 0,218903           | 2,043040     | 0,288523           |
| 100 | 0               | 1,997080     | 0,099863           | 2,017100     | 0,152706           |
|     | 10              | 2,000610     | 0,106116           | 1,998610     | 0,106116           |
|     | 20              | 2,001630     | 0,113598           | 2,016330     | 0,192359           |
|     | 50              | 2,008930     | 0,143415           | 2,032980     | 0,242577           |
|     | 60              | 2,015400     | 0,162478           | 2,032940     | 0,245545           |
|     | 70              | 2,021090     | 0,191413           | 2,035790     | 0,247699           |
| 150 | 0               | 1,999120     | 0,080795           | 2,010750     | 0,120838           |
|     | 10              | 2,000700     | 0,086037           | 2,007750     | 0,141682           |
|     | 20              | 2,002920     | 0,091963           | 2,007880     | 0,156420           |
|     | 50              | 2,004650     | 0,117939           | 2,023220     | 0,195513           |
|     | 60              | 2,007270     | 0,132069           | 2,024210     | 0,202402           |
|     | 70              | 2,010370     | 0,154710           | 2,020600     | 0,200421           |
| 200 | 0               | 1,997170     | 0,070779           | 2,009640     | 0,106419           |
|     | 10              | 1,999940     | 0,077144           | 2,002720     | 0,119707           |
|     | 20              | 2,000070     | 0,080597           | 2,007400     | 0,135291           |
|     | 50              | 2,005110     | 0,100128           | 2,019010     | 0,169041           |
|     | 60              | 2,006990     | 0,113024           | 2,019670     | 0,169690           |
|     | 70              | 2,010390     | 0,130069           | 2,014330     | 0,172494           |
| 250 | 0               | 1,997880     | 0,061282           | 2,009140     | 0,092970           |
|     | 10              | 2,002130     | 0,066279           | 2,005230     | 0,110066           |
|     | 20              | 1,999870     | 0,071591           | 2,006530     | 0,119941           |
|     | 50              | 2,004600     | 0,090542           | 2,012170     | 0,149147           |
|     | 60              | 2,004410     | 0,101917           | 2,010810     | 0,153313           |
|     | 70              | 2,007930     | 0,117310           | 2,013110     | 0,153389           |
| 300 | 0               | 1,999430     | 0,056948           | 2,003280     | 0,086511           |
|     | 10              | 2,000040     | 0,061247           | 2,003630     | 0,099712           |
|     | 20              | 2,002270     | 0,064593           | 2,001710     | 0,110594           |
|     | 50              | 2,003030     | 0,082216           | 2,012800     | 0,136723           |
|     | 60              | 2,003850     | 0,091327           | 2,011030     | 0,141071           |
|     | 70              | 2,006720     | 0,106609           | 2,010840     | 0,139918           |

Πίνακας 1: Weibull Distribution W(2, 2)

|     |                 | $\theta = 2$ |                    | $\gamma = 1$ |                    |
|-----|-----------------|--------------|--------------------|--------------|--------------------|
| n   | Censoring level | Mean         | Standard Deviation | Mean         | Standard Deviation |
| 50  | 0               | 1,991140     | 0,282873           | 1,015670     | 0,108255           |
|     | 10              | 2,014320     | 0,304557           | 1,010920     | 0,123565           |
|     | 20              | 2,015920     | 0,323096           | 1,016510     | 0,140002           |
|     | 50              | 2,066080     | 0,442357           | 1,033940     | 0,179736           |
|     | 60              | 2,080770     | 0,498213           | 1,032900     | 0,182062           |
|     | 70              | 2,130170     | 0,638820           | 1,035320     | 0,184643           |
| 80  | 0               | 2,001290     | 0,221134           | 1,010500     | 0,085927           |
|     | 10              | 2,014540     | 0,235224           | 1,006400     | 0,097137           |
|     | 20              | 2,019580     | 0,252536           | 1,007310     | 0,109165           |
|     | 50              | 2,035240     | 0,322015           | 1,020380     | 0,139105           |
|     | 60              | 2,047540     | 0,373056           | 1,021100     | 0,141158           |
|     | 70              | 2,070460     | 0,450484           | 1,020430     | 0,138895           |
| 100 | 0               | 2,000490     | 0,200791           | 1,009270     | 0,075904           |
|     | 10              | 2,005470     | 0,212665           | 1,006770     | 0,086778           |
|     | 20              | 2,010390     | 0,227043           | 1,008220     | 0,098418           |
|     | 50              | 2,032140     | 0,298922           | 1,016690     | 0,123835           |
|     | 60              | 2,040680     | 0,339290           | 1,018390     | 0,123832           |
|     | 70              | 2,063710     | 0,397080           | 1,019000     | 0,122202           |
| 150 | 0               | 1,997940     | 0,166066           | 1,006710     | 0,061612           |
|     | 10              | 2,002300     | 0,173712           | 1,005060     | 0,071120           |
|     | 20              | 2,007300     | 0,185601           | 1,005850     | 0,078395           |
|     | 50              | 2,018500     | 0,240289           | 1,008640     | 0,097401           |
|     | 60              | 2,022370     | 0,266377           | 1,009820     | 0,099761           |
|     | 70              | 2,036600     | 0,316650           | 1,011580     | 0,098575           |
| 200 | 0               | 1,998560     | 0,138462           | 1,003590     | 0,053164           |
|     | 10              | 2,003680     | 0,149005           | 1,003220     | 0,061721           |
|     | 20              | 2,004230     | 0,156263           | 1,004840     | 0,068042           |
|     | 50              | 2,011760     | 0,205195           | 1,008610     | 0,084017           |
|     | 60              | 2,018520     | 0,230454           | 1,007980     | 0,085487           |
|     | 70              | 2,022600     | 0,269491           | 1,008140     | 0,086469           |
| 250 | 0               | 1,998400     | 0,124825           | 1,003450     | 0,047131           |
|     | 10              | 2,002480     | 0,134662           | 1,002010     | 0,054704           |
|     | 20              | 2,009310     | 0,140449           | 1,002180     | 0,061748           |
|     | 50              | 2,010280     | 0,184026           | 1,005170     | 0,075703           |
|     | 60              | 2,015350     | 0,203709           | 1,007030     | 0,077031           |
|     | 70              | 2,023170     | 0,238673           | 1,006610     | 0,076885           |
| 300 | 0               | 1,999120     | 0,116650           | 1,003230     | 0,043579           |
|     | 10              | 2,001000     | 0,123514           | 1,003430     | 0,049301           |
|     | 20              | 2,004310     | 0,130370           | 1,001980     | 0,055509           |
|     | 50              | 2,009060     | 0,166697           | 1,005510     | 0,069731           |
|     | 60              | 2,008150     | 0,184239           | 1,004170     | 0,069352           |
|     | 70              | 2,012650     | 0,215361           | 1,005060     | 0,070065           |

Πίνακας 2: Weibull Distribution  $W(2, 1)$

| n   | Censoring level | $\theta = 3$ |                    | $\gamma = 0,5$ |                    |
|-----|-----------------|--------------|--------------------|----------------|--------------------|
|     |                 | Mean         | Standard Deviation | Mean           | Standard Deviation |
| 50  | 0               | 3,067490     | 0,884238           | 0,507442       | 0,054917           |
|     | 10              | 3,087480     | 0,947215           | 0,505068       | 0,061925           |
|     | 20              | 3,153400     | 1,052630           | 0,500293       | 0,102165           |
|     | 50              | 3,294080     | 1,462590           | 0,488568       | 0,167213           |
|     | 60              | 3,495480     | 1,847230           | 0,490720       | 0,168974           |
|     | 70              | 3,653550     | 2,552440           | 0,484063       | 0,177480           |
| 80  | 0               | 3,055230     | 0,691445           | 0,505675       | 0,041978           |
|     | 10              | 3,082810     | 0,733047           | 0,503080       | 0,048277           |
|     | 20              | 3,087830     | 0,799376           | 0,502365       | 0,068302           |
|     | 50              | 3,192410     | 1,090420           | 0,485281       | 0,147996           |
|     | 60              | 3,292750     | 1,331190           | 0,485640       | 0,149909           |
|     | 70              | 3,411610     | 1,604640           | 0,482002       | 0,157252           |
| 100 | 0               | 3,024020     | 0,607534           | 0,504731       | 0,038296           |
|     | 10              | 3,041360     | 0,655653           | 0,503891       | 0,044465           |
|     | 20              | 3,082870     | 0,703801           | 0,502019       | 0,060835           |
|     | 50              | 3,137900     | 0,933935           | 0,483020       | 0,144266           |
|     | 60              | 3,217980     | 1,127170           | 0,484311       | 0,142495           |
|     | 70              | 3,342480     | 1,381500           | 0,481208       | 0,151963           |
| 150 | 0               | 3,024210     | 0,488835           | 0,502815       | 0,031018           |
|     | 10              | 3,017630     | 0,523297           | 0,503197       | 0,036645           |
|     | 20              | 3,040790     | 0,560716           | 0,502862       | 0,039133           |
|     | 50              | 3,079360     | 0,750367           | 0,483848       | 0,133153           |
|     | 60              | 3,138710     | 0,843372           | 0,482902       | 0,135535           |
|     | 70              | 3,187860     | 1,014280           | 0,472100       | 0,153327           |
| 200 | 0               | 3,014340     | 0,421870           | 0,502628       | 0,026418           |
|     | 10              | 3,018670     | 0,454223           | 0,502312       | 0,030373           |
|     | 20              | 3,030040     | 0,484671           | 0,501870       | 0,035609           |
|     | 50              | 3,068400     | 0,620367           | 0,484735       | 0,124004           |
|     | 60              | 3,108440     | 0,699113           | 0,484437       | 0,125741           |
|     | 70              | 3,146370     | 0,850621           | 0,467074       | 0,157569           |
| 250 | 0               | 3,015340     | 0,381953           | 0,501578       | 0,023411           |
|     | 10              | 3,035140     | 0,404464           | 0,501249       | 0,026783           |
|     | 20              | 3,044660     | 0,429533           | 0,502013       | 0,031149           |
|     | 50              | 3,059980     | 0,557521           | 0,486866       | 0,114853           |
|     | 60              | 3,078070     | 0,634066           | 0,483551       | 0,121236           |
|     | 70              | 3,114500     | 0,761324           | 0,470709       | 0,148656           |
| 300 | 0               | 3,007720     | 0,351086           | 0,500978       | 0,021335           |
|     | 10              | 3,014170     | 0,379675           | 0,500702       | 0,024977           |
|     | 20              | 3,027780     | 0,392873           | 0,501451       | 0,027934           |
|     | 50              | 3,048340     | 0,497845           | 0,487345       | 0,109593           |
|     | 60              | 3,052000     | 0,571459           | 0,481357       | 0,122221           |
|     | 70              | 3,104430     | 0,679829           | 0,467351       | 0,151815           |

Πίνακας 3: Weibull Distribution  $W(3, 0.5)$

Από τους παραπάνω πίνακες παρατηρείται ότι

- 1) Όταν το ποσοστό λογοκρισίας αυξάνει, τότε αυξάνει τόσο η μεροληψία όσο και το τυπικό σφάλμα.
- 2) Όταν αυξάνει το δείγμα οι επιπτώσεις της λογοκρισίας ελαττώνονται.

### 3.4 Εκτιμητής μέγιστης πιθανοφάνειας για τυχαία λογοκριμένο δείγμα

Έστω  $t_1, t_2, \dots, t_n$  ένα τυχαία λογοκριμένο δείγμα με ποσοστό λογοκρισίας

$$p = 1 - q = \bar{F}(T).$$

Το δείγμα είναι της μορφής

$$\underbrace{t_1, t_2, \dots, t_{n_1}}_{n_1}, \underbrace{c_1, c_2, \dots, c_{n-n_1}}_{n-n_1}$$

Από την (3.11) η συνάρτηση πιθανοφάνειας είναι

$$L(\theta) \propto \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(c_i; \theta)^{1-\delta_i} = \prod_{i=1}^r \left( \gamma \frac{t_i^{\gamma-1}}{\theta^\gamma} \right) e^{-\sum_{i=1}^r \left( \frac{t_i}{\theta} \right)^\gamma - \sum_{i=1}^{n-r} \left( \frac{c_i}{\theta} \right)^\gamma} \quad (3.10)$$

Λογαριθμίζοντας και παίρνοντας την παράγωγο ως προς  $\theta$  και  $\gamma$  προκύπτει το σύστημα

$$\begin{cases} \frac{1}{\gamma} + \frac{\sum_{i=1}^r \ln t_i}{r} - \sum_{i=1}^r \left( \frac{t_i}{\theta} \right)^\gamma \ln \frac{t_i}{\theta} - \sum_{i=1}^{n-r} \left( \frac{c_i}{\theta} \right)^\gamma \ln \frac{t_i}{\theta} = 0 \\ -r \frac{\gamma}{\theta} + \gamma \frac{\sum_{i=1}^r t_i^\gamma}{\theta^{\gamma+1}} + \gamma \frac{\sum_{i=1}^{n-r} c_i^\gamma}{\theta^{\gamma+1}} = 0. \end{cases}$$

Για να υπολογιστεί ο πληροφοριακός αριθμός του Fisher σε σχέση με το ποσοστό λογοκρισίας  $p$  θα πρέπει να υπολογιστεί η σχέση μεταξύ του  $p$  και  $L$ . Από την σχέση (2.26) προκύπτει ότι

$$1 - p = 1 - \frac{1}{L} \int_0^L e^{(-\frac{c}{\theta})^\gamma} dc \quad (3.11)$$

Γίνεται ο μετασχηματισμός:

$$\left(\frac{c}{\theta}\right)^\gamma = w \Leftrightarrow \frac{c}{\theta} = w^{1/\gamma} \Leftrightarrow c = \theta w^{1/\gamma}$$

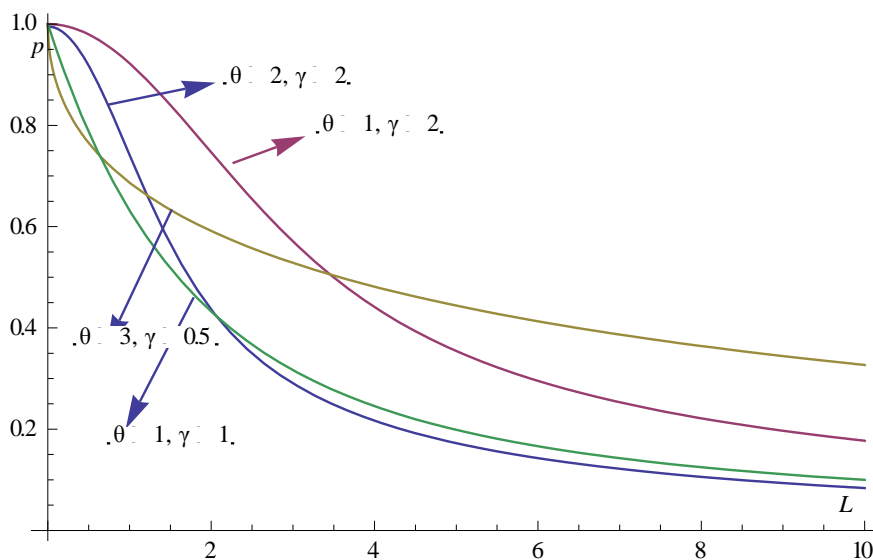
$$dc = \theta \frac{1}{\gamma} w^{(1/\gamma)-1} dw.$$

Άρα:

$$\begin{aligned} 1 - p = P(C > T) &= 1 - \frac{1}{L} \int_0^{\left(\frac{L}{\theta}\right)^\gamma} \theta \frac{1}{\gamma} w^{1/\gamma-1} e^{-w} dw \\ &= 1 - \frac{\theta}{L\gamma} \left[ G\left(\frac{1}{\gamma}\right) - G\left(\frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right) \right] \end{aligned}$$

Επομένως, η τιμή  $L$  είναι η λύση της εξίσωσης

$$\frac{\theta}{L\gamma} \left[ G\left(\frac{1}{\gamma}\right) - G\left(\frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right) \right] = p \quad (3.12)$$



Διάγραμμα 10: Σχέση ποσοστού λογοκρισίας  $p$  με το άνω άκρο  $L$

Όπως αναμένεται όταν το ποσοστό λογοκρισίας  $p$  μικραίνει το  $L$  τείνει στο άπειρο, διότι το σύνολο των παρατηρήσεων που λαμβάνεται υπόψη αυξάνεται συνεχώς.

Αν υπολογιστεί ο πληροφοριακός αριθμός στην περίπτωση αυτή προκύπτει ότι:

$$f(x; \theta, \gamma) = \gamma \frac{x^{\gamma-1}}{\theta^\gamma} e^{-\left(\frac{x}{\theta}\right)^\gamma}$$

$$\bar{F}(x; \theta, \gamma) = e^{-\left(\frac{x}{\theta}\right)^\gamma}$$

$$g(x) = \frac{1}{L} I_{[0,L]}^{(x)}$$

$$\bar{G}(x) = \frac{1}{L}(L-x), 0 < x < L.$$

$$\begin{aligned} I_p(\theta, \gamma) &= \int_0^L \left[ \left(-1 + \left(\frac{x}{\theta}\right)^\gamma\right)^2 \gamma \frac{x^{\gamma-1}}{\theta^\gamma} \left(1 - \frac{x}{L}\right) + \left(\frac{\gamma x^\gamma}{\theta^{\gamma+1}}\right)^2 \frac{1}{L} \right] e^{-\left(\frac{x}{\theta}\right)^\gamma} dx \\ &= \left(\frac{\gamma}{\theta}\right)^2 \int_0^L \left[ \left(-1 + \left(\frac{x}{\theta}\right)^\gamma\right)^2 \left(1 - \frac{x}{L}\right) \gamma \frac{x^{\gamma-1}}{\theta^\gamma} + \left(\frac{x}{\theta}\right)^{2\gamma} \frac{1}{L} \right] e^{-\left(\frac{x}{\theta}\right)^\gamma} dx. \end{aligned}$$

Γίνεται ο μετασχηματισμός:

$$u = \left(\frac{x}{\theta}\right)^\gamma \Leftrightarrow x = \theta u^{\frac{1}{\gamma}} \Leftrightarrow dx = \frac{\theta}{\gamma} u^{\frac{1}{\gamma}-1} \quad (3.13)$$

οπότε για τον πληροφοριακό αριθμό προκύπτει

$$\begin{aligned} I(\theta) &= \left(\frac{\gamma}{\theta}\right)^2 \int_0^{\left(\frac{L}{\theta}\right)^\gamma} \left[ (-1+u)^2 \left(1 - \frac{\theta}{L} u^{\frac{1}{\gamma}}\right) + \frac{\theta}{L} \left(u^{\frac{1}{\gamma}+1}\right)^{2\gamma} \right] e^u du \\ &= \left(\frac{\gamma}{\theta}\right)^2 \int_0^{\left(\frac{L}{\theta}\right)^\gamma} \left[ 1 - 2u + u^2 - \frac{\theta}{L} u^{\frac{1}{\gamma}} + \frac{\theta}{L} \left(2 + \frac{1}{\gamma}\right) u^{\frac{1}{\gamma}+1} - \frac{\theta}{L} u^{\frac{1}{\gamma}+2} \right] e^u du \\ &= \left(\frac{\gamma}{\theta}\right)^2 \left[ 1 - e^{-\left(\frac{L}{\theta}\right)^\gamma} - \left(\frac{L}{\theta}\right)^\gamma e^{-\left(\frac{L}{\theta}\right)^\gamma} - \frac{\theta}{L\gamma} G\left(\frac{1}{\gamma}\right) + \frac{\theta}{\gamma} G\left(1 + \frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right) \right. \\ &\quad \left. - \frac{(\theta + 2\gamma\theta)}{L\gamma} G\left(2 + \frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right) + \frac{\theta}{L} G\left(3 + \frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right) \right] \end{aligned}$$

Από τη σχέση

$$G\left(3 + \frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right) = \left(\frac{L}{\theta}\right)^{1+2\gamma} e^{-\left(\frac{L}{\theta}\right)^\gamma} + \frac{2\gamma + 1}{\gamma} G\left(2 + \frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right)$$

παίρνουμε

$$I_p(\theta, \gamma) = \left(\frac{\gamma}{\theta}\right)^2 \left[1 - \left(\frac{L}{\theta}\right) \left(G\left(\frac{1}{\gamma}\right) - G\left(\frac{1}{\gamma}, \left(\frac{L}{\theta}\right)^\gamma\right)\right)\right]. \quad (3.14)$$

Από την (3.12) και την (3.14) προκύπτει ότι

$$I_p(\theta, \gamma) = \left(\frac{\gamma}{\theta}\right)^2 (1 - p).$$

Έτσι λοιπόν στην περίπτωση της τυχαίας λογοκρισίας με χρόνους λογοκρισίας από την ομοιόμορφη κατανομή η σχετική ασυμπτωτική αποδοτικότητα για την παράμετρο κλίμακας  $\theta$  σε σχέση με τη μη λογοκρισία είναι ίδια με αυτή από την λογοκρισία τύπου I και II (3.9).

Δεν ισχύει όμως το ίδιο και για την παράμετρο μορφής. Στην περίπτωση αυτή ο πληροφοριακός αριθμός του Fisher (3.13) είναι

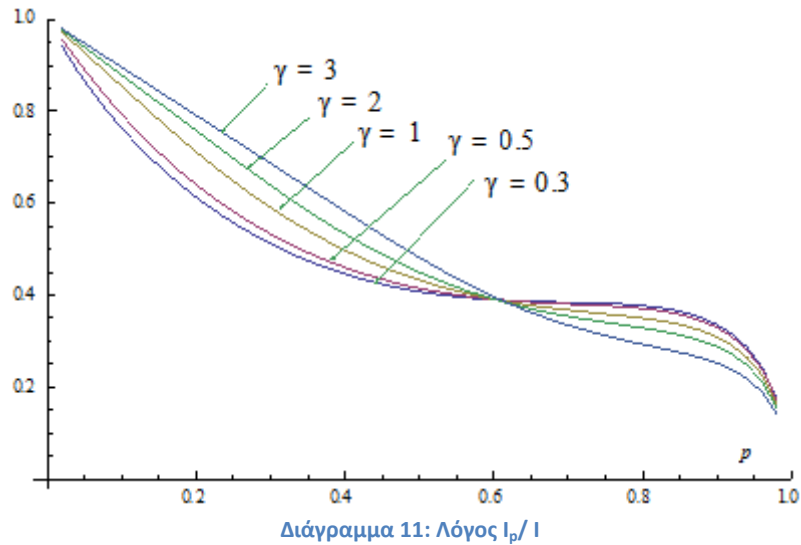
$$I_p(\gamma) = \frac{1}{\gamma^2} \int_0^L \left( \left(1 + \ln\left(\frac{x}{\theta}\right)^\gamma - \left(\frac{x}{\theta}\right)^\gamma \ln\left(\frac{x}{\theta}\right)^\gamma\right)^2 \gamma \frac{x^{\gamma-1}}{\theta^\gamma} \left(1 - \frac{x}{L}\right) + \left(\frac{x}{\theta}\right)^{2\gamma} \ln^2\left(\frac{x}{\theta}\right)^\gamma \frac{1}{L} \right) e^{-\left(\frac{x}{\theta}\right)^\gamma} dx$$

Με διαδικασία ανάλογη της παραμέτρου θέσης και με την βοήθεια του μετασχηματισμού (3.14) ο παραπάνω μετασχηματισμός γράφεται στην μορφή

$$I_p(\gamma) = \frac{1}{\gamma^2} \int_0^{\left(\frac{L}{\theta}\right)^\gamma} \left( (1 + \ln u)^2 \left(1 - \frac{\theta}{L} u^{1/\gamma}\right) \right) e^{-u} du \quad (3.15)$$

Σε αντίθεση με την παράμετρο θέσης ο πληροφοριακός αριθμός (3.15) δεν μπορεί να εκφρασθεί άμεσα σε σχέση με το  $p$ . Η γραφική παράσταση του λόγου βρίσκεται γραφικά στο παρακάτω διάγραμμα.





Από το παραπάνω σχήμα εξάγεται το συμπέρασμα ότι για ποσοστό λογοκρισίας περί το  $p_0 = 0,6$  η απώλεια πληροφορίας είναι η ίδια (περίπου 60%), ανεξαρτήτως της τιμής των παραμέτρων. Επιπλέον για  $p < p_0$  ο λόγος των πληροφοριών είναι μια αύξουσα συνάρτηση σε σχέση με την παράμετρο  $\gamma$  και φθίνουσα όταν  $p > p_0$ . Επειδή η ασυμμετρία είναι μια αύξουσα συνάρτηση ως προς  $\gamma$  μπορούμε να πούμε ότι όταν αυξάνει η ασυμμετρία το ποσοστό απώλειας της πληροφορίας ελαττώνεται για  $p < p_0$  και αυξάνει για  $p > p_0$ .

Για την μελέτη των ΕΜΠ για μικρά δείγματα χρησιμοποιήθηκαν προσομοιωμένα δείγματα τα οποία παρήχθησαν με τη βοήθεια του Mathematica. Το αποτέλεσμα είναι ένα δείγμα δισδιάστατο όπου το δεύτερο στοιχείο είναι 1 όταν το πρώτο στοιχείο είναι λογοκριμένο και 0 όταν δεν είναι. (βλ. Παράρτημα Π2)

Οι παρακάτω πίνακες δείχνουν την συμπεριφορά του ΕΜΠ για την τυχαία λογοκρισία για μικρά δείγματα.

| n   | Censoring level | $\theta=2$ |                    | $\gamma=2$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|
|     |                 | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 1,9822     | 0,1379             | 2,0325     | 0,2152             |
|     | 10              | 2,1467     | 0,5553             | 2,1212     | 0,4496             |
|     | 20              | 2,0949     | 0,3902             | 2,0980     | 0,4219             |
|     | 50              | 2,0251     | 0,1886             | 2,0832     | 0,3382             |
|     | 60              | 2,0193     | 0,1910             | 2,0314     | 0,2649             |
|     | 70              | 1,9982     | 0,1723             | 2,0116     | 0,2750             |
| 80  | 0               | 1,9972     | 0,1096             | 2,0176     | 0,1708             |
|     | 10              | 2,1050     | 0,4619             | 2,0917     | 0,3329             |
|     | 20              | 2,0223     | 0,2495             | 2,0424     | 0,3051             |
|     | 50              | 2,0081     | 0,1316             | 2,0423     | 0,2719             |
|     | 60              | 2,0079     | 0,1404             | 2,0403     | 0,2217             |
|     | 70              | 1,9856     | 0,1318             | 2,0332     | 0,2280             |
| 100 | 0               | 1,9968     | 0,0974             | 2,0183     | 0,1538             |
|     | 10              | 2,0581     | 0,3526             | 2,0421     | 0,3271             |
|     | 20              | 2,0249     | 0,2557             | 2,0408     | 0,2986             |
|     | 50              | 2,0146     | 0,1499             | 2,0237     | 0,2157             |
|     | 60              | 2,0036     | 0,1296             | 2,0212     | 0,1743             |
|     | 70              | 2,0013     | 0,1366             | 2,0170     | 0,1863             |
| 150 | 0               | 1,9930     | 0,0808             | 2,0118     | 0,1223             |
|     | 10              | 2,0309     | 0,2857             | 2,0429     | 0,2622             |
|     | 20              | 2,0229     | 0,2071             | 2,0340     | 0,2475             |
|     | 50              | 2,0014     | 0,1155             | 2,0256     | 0,1822             |
|     | 60              | 1,9998     | 0,1152             | 2,0185     | 0,1598             |
|     | 70              | 1,9971     | 0,0983             | 2,0010     | 0,1367             |
| 200 | 0               | 2,0001     | 0,0678             | 2,0122     | 0,0979             |
|     | 10              | 2,0851     | 0,2536             | 2,0716     | 0,2085             |
|     | 20              | 2,0107     | 0,1567             | 2,0162     | 0,1863             |
|     | 50              | 2,0025     | 0,0974             | 2,0108     | 0,1488             |
|     | 60              | 1,9943     | 0,0973             | 2,0098     | 0,1472             |
|     | 70              | 1,9884     | 0,0871             | 2,0078     | 0,1284             |
| 250 | 0               | 1,9957     | 0,0656             | 1,9985     | 0,0919             |
|     | 10              | 2,0064     | 0,2096             | 2,0241     | 0,1811             |
|     | 20              | 2,0033     | 0,1428             | 2,0170     | 0,1656             |
|     | 50              | 2,0021     | 0,0964             | 2,0048     | 0,1397             |
|     | 60              | 1,9966     | 0,0767             | 2,0014     | 0,1422             |
|     | 70              | 1,9921     | 0,0771             | 1,9960     | 0,1174             |
| 300 | 0               | 1,9996     | 0,0570             | 1,9900     | 0,0914             |
|     | 10              | 2,0065     | 0,1767             | 2,0184     | 0,1578             |
|     | 20              | 2,0053     | 0,1279             | 2,0136     | 0,1515             |
|     | 50              | 2,0032     | 0,0833             | 2,0057     | 0,1327             |
|     | 60              | 2,0018     | 0,0697             | 2,0050     | 0,1171             |
|     | 70              | 1,9966     | 0,0758             | 2,0045     | 0,1077             |

Πίνακας 4: Τυχαία λογοκρίμενο δείγμα  $W(2, 2)$

|     |                 | $\theta=2$ |                    | $\gamma=1$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|
| n   | Censoring level | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 1,9824     | 0,2705             | 0,9973     | 0,0968             |
|     | 10              | 2,0305     | 0,4561             | 1,0174     | 0,1600             |
|     | 20              | 2,0144     | 0,3819             | 1,0104     | 0,1538             |
|     | 50              | 2,0122     | 0,3267             | 1,0060     | 0,1400             |
|     | 60              | 2,0028     | 0,3314             | 1,0181     | 0,1174             |
|     | 70              | 1,9734     | 0,3368             | 1,0087     | 0,1147             |
| 80  | 0               | 2,0112     | 0,2399             | 1,0032     | 0,0877             |
|     | 10              | 2,0681     | 0,3282             | 1,0272     | 0,1215             |
|     | 20              | 2,0360     | 0,3174             | 1,0153     | 0,1037             |
|     | 50              | 1,9795     | 0,2634             | 1,0060     | 0,1002             |
|     | 60              | 1,9546     | 0,2543             | 1,0171     | 0,0940             |
|     | 70              | 1,8697     | 0,2430             | 1,0146     | 0,0957             |
| 100 | 0               | 2,0090     | 0,2042             | 1,0089     | 0,0746             |
|     | 10              | 2,0372     | 0,2584             | 1,0230     | 0,1106             |
|     | 20              | 2,0154     | 0,2644             | 1,0066     | 0,1115             |
|     | 50              | 2,0133     | 0,2634             | 1,0003     | 0,0921             |
|     | 60              | 1,9933     | 0,2126             | 1,0144     | 0,0895             |
|     | 70              | 1,8572     | 0,2325             | 1,0042     | 0,0889             |
| 150 | 0               | 2,0012     | 0,1728             | 1,0037     | 0,0640             |
|     | 10              | 2,0578     | 0,2243             | 1,0072     | 0,0849             |
|     | 20              | 2,0313     | 0,2274             | 1,0031     | 0,0909             |
|     | 50              | 2,0185     | 0,2095             | 0,9949     | 0,0815             |
|     | 60              | 1,9908     | 0,1876             | 0,9954     | 0,0735             |
|     | 70              | 1,8746     | 0,1987             | 0,9940     | 0,0683             |
| 200 | 0               | 1,9994     | 0,1335             | 1,0070     | 0,0549             |
|     | 10              | 2,0290     | 0,2146             | 1,0100     | 0,0798             |
|     | 20              | 2,0102     | 0,2059             | 1,0090     | 0,0714             |
|     | 50              | 2,0051     | 0,1687             | 1,0069     | 0,0714             |
|     | 60              | 2,0035     | 0,1507             | 1,0005     | 0,0613             |
|     | 70              | 1,9915     | 0,1522             | 0,9979     | 0,0600             |
| 250 | 0               | 2,0004     | 0,1358             | 1,0013     | 0,0487             |
|     | 10              | 1,9947     | 0,1580             | 1,0430     | 0,0651             |
|     | 20              | 1,9916     | 0,1767             | 1,0043     | 0,0702             |
|     | 50              | 1,9898     | 0,1431             | 1,0034     | 0,0633             |
|     | 60              | 1,9886     | 0,1329             | 0,9973     | 0,0609             |
|     | 70              | 1,8886     | 0,1502             | 0,9964     | 0,0535             |
| 300 | 0               | 1,9960     | 0,1253             | 1,0051     | 0,0404             |
|     | 10              | 1,9926     | 0,1624             | 1,0054     | 0,0625             |
|     | 20              | 1,9913     | 0,1499             | 1,0042     | 0,0645             |
|     | 50              | 1,8942     | 0,1262             | 1,0019     | 0,0575             |
|     | 60              | 1,8655     | 0,1431             | 1,0028     | 0,0512             |
|     | 70              | 1,7534     | 0,1282             | 1,0047     | 0,0556             |

Πίνακας 5: Τυχαία λογοκρίμενο δείγμα  $W(2, 1)$

|     |                 | $\theta=3$ |                    | $\gamma=0,5$ |                    |
|-----|-----------------|------------|--------------------|--------------|--------------------|
| n   | Censoring level | Mean       | Standard Deviation | Mean         | Standard Deviation |
| 50  | 0               | 3,1697     | 0,8403             | 0,5118       | 0,0512             |
|     | 10              | 3,0825     | 1,1436             | 0,4977       | 0,0952             |
|     | 20              | 3,0593     | 1,0576             | 0,5091       | 0,0701             |
|     | 50              | 3,0439     | 0,9816             | 0,5095       | 0,0634             |
|     | 60              | 3,0114     | 1,0055             | 0,5076       | 0,0567             |
|     | 70              | 3,0018     | 0,9159             | 0,5142       | 0,0615             |
| 80  | 0               | 3,0813     | 0,6926             | 0,5025       | 0,0466             |
|     | 10              | 3,3087     | 0,7975             | 0,5023       | 0,0733             |
|     | 20              | 3,0397     | 0,8506             | 0,5075       | 0,0507             |
|     | 50              | 3,0292     | 0,8192             | 0,5046       | 0,0466             |
|     | 60              | 3,0166     | 0,7004             | 0,4953       | 0,0507             |
|     | 70              | 3,0093     | 0,7787             | 0,5058       | 0,0434             |
| 100 | 0               | 3,1009     | 0,5733             | 0,5047       | 0,0352             |
|     | 10              | 3,0886     | 0,7955             | 0,5061       | 0,0478             |
|     | 20              | 3,0772     | 0,7292             | 0,5026       | 0,0481             |
|     | 50              | 3,0767     | 0,6715             | 0,5039       | 0,0429             |
|     | 60              | 3,0695     | 0,6524             | 0,4969       | 0,0451             |
|     | 70              | 3,0103     | 0,6271             | 0,5022       | 0,0432             |
| 150 | 0               | 3,0837     | 0,5097             | 0,5033       | 0,0284             |
|     | 10              | 3,0709     | 0,5554             | 0,5023       | 0,0604             |
|     | 20              | 3,0609     | 0,5374             | 0,5024       | 0,0401             |
|     | 50              | 3,0597     | 0,5182             | 0,5052       | 0,0401             |
|     | 60              | 3,0524     | 0,5224             | 0,5049       | 0,0337             |
|     | 70              | 3,0407     | 0,5498             | 0,5009       | 0,0334             |
| 200 | 0               | 3,0985     | 0,4399             | 0,5019       | 0,0266             |
|     | 10              | 3,0850     | 0,4916             | 0,5029       | 0,0341             |
|     | 20              | 3,0703     | 0,5033             | 0,5019       | 0,0333             |
|     | 50              | 3,0487     | 0,5000             | 0,4986       | 0,0307             |
|     | 60              | 3,0342     | 0,4685             | 0,4994       | 0,0293             |
|     | 70              | 2,9863     | 0,4573             | 0,5053       | 0,0301             |
| 250 | 0               | 3,0180     | 0,4008             | 0,5025       | 0,0230             |
|     | 10              | 3,0121     | 0,4644             | 0,4996       | 0,0322             |
|     | 20              | 3,0099     | 0,4685             | 0,5015       | 0,0301             |
|     | 50              | 3,0015     | 0,4260             | 0,5030       | 0,0275             |
|     | 60              | 2,9881     | 0,4307             | 0,5013       | 0,0250             |
|     | 70              | 2,9878     | 0,3924             | 0,5029       | 0,0257             |
| 300 | 0               | 3,0222     | 0,3601             | 0,4988       | 0,0199             |
|     | 10              | 3,0197     | 0,3563             | 0,5013       | 0,0274             |
|     | 20              | 3,0128     | 0,4009             | 0,4995       | 0,0275             |
|     | 50              | 2,9915     | 0,3648             | 0,5065       | 0,0252             |
|     | 60              | 2,9815     | 0,3929             | 0,5019       | 0,0238             |
|     | 70              | 2,9799     | 0,3716             | 0,5004       | 0,0247             |

Πίνακας 6: Τυχαία λογοκρίμενο δείγμα  $W(3, 0.5)$

Συμπερασματικά από τους παραπάνω πίνακες προκύπτει ότι:

1. Όταν αυξάνει το δείγμα οι επιπτώσεις της λογοκρισίας ελαττώνονται.
2. Εν αντιθέσει με την λογοκρισία τύπου I και II παρατηρούμε ότι η τυχαία λογοκρισία επηρεάζει λιγότερο την εκτίμηση των παραμέτρων. Αυτό φαίνεται και από τη συμπεριφορά του  $\theta$ . Στην περίπτωση της λογοκρισίας τύπου I και II όσο αυξάνει το  $p$  τόσο πιο πολύ επηρεάζεται το  $\theta$ , ενώ στην περίπτωση της random λογοκρισίας δείχνει να υπάρχει μία σταθερότητα και να μην επηρεάζεται τόσο η παράμετρος  $\theta$ .

## ΚΕΦΑΛΑΙΟ 4

Η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων για την κατανομή Pareto

### 4.1 Εισαγωγή

Η κατανομή Pareto πήρε το όνομά της από τον Ελβετό καθηγητή οικονομικών Vilfredo Pareto. Ο νόμος του Pareto, όπως διατυπώθηκε από τον ίδιο, ασχολήθηκε με την κατανομή του εισοδήματος ενός πληθυσμού. Θεωρήθηκε από τον Pareto ότι ο νόμος αυτός ήταν καθολικός και αναπόφευκτος, ανεξάρτητα από τη φορολογία, τις κοινωνικές και πολιτικές συνθήκες. Σαφέστατα υπήρξε και έντονος αντίλογος και διαμαρτυρίες από διάφορους γνωστούς οικονομολόγους τις επόμενες δεκαετίες όπως από τους Pigou (1932), Shirras (1935) και Hayakawa (1951). Πιο πρόσφατα έχουν γίνει προσπάθειες να εξηγηθούν πολλά εμπειρικά φαινόμενα με τη χρήση της κατανομής Pareto ή κάποιας σχετικής κατανομής.

Ο Harris (1968) είχε επισημάνει ότι η κατανομή Pareto είναι ένα μίγμα εκθετικής κατανομής με παράμετρο  $\theta^{-1}$ .

Η κατανομή Pareto είναι λοιπόν μία ασύμμετρη κατανομή με εφαρμογή στις οικονομικές επιστήμες. Η συνάρτηση πυκνότητας πιθανότητας της Pareto είναι η εξής:

$$f(x; \gamma, \theta) = \theta \gamma^\theta x^{-(\theta+1)}, 0 \leq \gamma \leq x, \theta > 0 \quad (4.1)$$

$= 0$  αλλού.

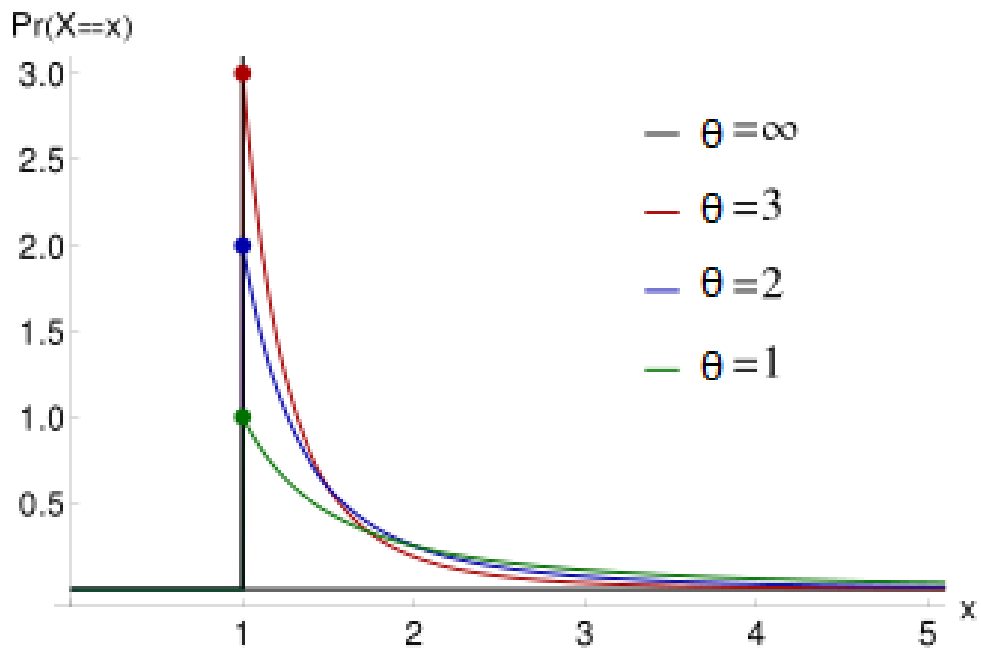
με συνάρτηση κατανομής

$$F(x; \gamma, \theta) = 1 - \left(\frac{\gamma}{x}\right)^\theta \quad (4.2)$$

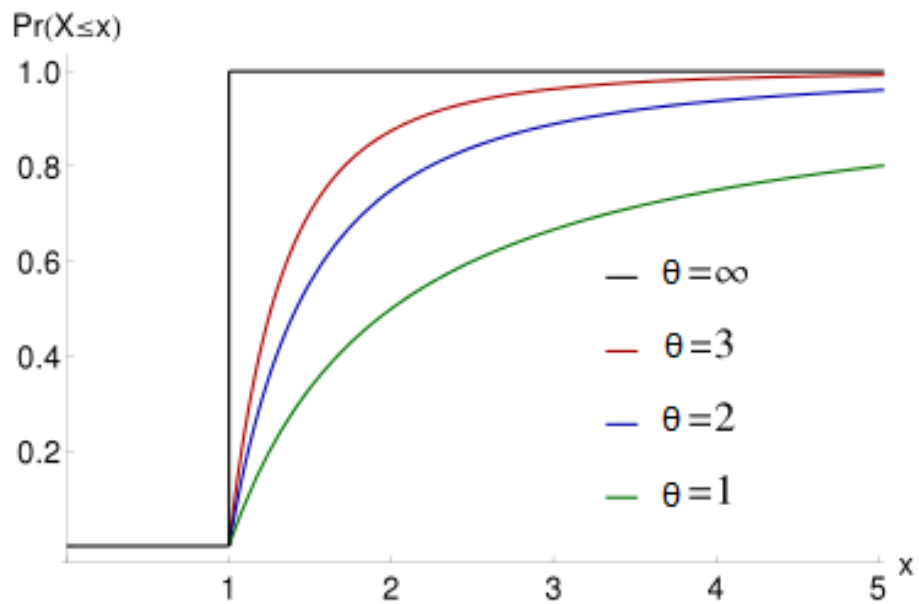
και συνάρτηση επιβίωσης

$$S(x; \gamma, \theta) = \bar{F}(x; \gamma, \theta) = \left(\frac{\gamma}{x}\right)^\theta. \quad (4.3)$$

Ο συμβολισμός  $X \sim \text{Pa}(\gamma, \theta)$  σημαίνει ότι η τ.μ.  $X$  ακολουθεί την Pareto όπου  $\gamma$  είναι η παράμετρος θέσης, ενώ  $\theta$  είναι η παράμετρος μορφής.



Διάγραμμα 12: Συνάρτηση πυκνότητας πιθανότητας (πηγή Wikipedia)



Διάγραμμα 13; Αθροιστική συνάρτηση κατανομής (πηγή Wikipedia)

Για την κατανομή Pareto επίσης ισχύουν:

$$E(X) = \frac{\theta\gamma}{\theta - 1}, \theta > 1$$

$$V(X) = \frac{\theta\gamma^2}{(\theta - 1)^2(\theta - 2)}, \theta > 2$$

$$\theta_3(X) = 2 \left( \frac{\theta + 1}{\theta - 3} \right) \sqrt{\frac{\theta - 2}{\theta}}, \theta > 3$$

$$\theta_4(X) = \frac{3(\theta - 2)(3\theta^2 + \theta + 2)}{\theta(\theta - 3)(\theta - 4)}, \theta > 4$$

$$Mo(X) = \gamma, Me(X) = 2^{1/\theta}\gamma, MD(X) = 2\gamma(\theta - 1)^{-1}(1 - \theta^{-1})^{\theta-1}, \theta > 1$$

όπου  $Mo$  είναι η διάμεσος,  $Me$  ο μέσος,  $MD$  η μέση απόκλιση της Pareto,  $\theta_3(X)$  ο συντελεστής ασυμμετρίας και  $\theta_4(X)$  ο συντελεστής κύρτωσης.

Η συνάρτηση κινδύνου αυτής της κατανομής είναι:

$$H(x) = \frac{\theta}{x}$$

Όταν η παράμετρος θέσης είναι άγνωστη τότε δεν ισχύουν οι συνθήκες κανονικότητας γιατί το στήριγμα της κατανομής εξαρτάται από την άγνωστη παράμετρο και τα ασυμπτωτικά θεωρήματα δεν ισχύουν πλέον. Για τον λόγο αυτό θα μελετηθεί η κατανομή αυτή στην περίπτωση που η παράμετρος  $\gamma$  είναι γνωστή.

## 4.2 Εκτιμητής μέγιστης πιθανοφάνειας για πλήρες δείγμα

Έστω  $\underline{x} = (x_1, x_2, \dots, x_n)$  ένα πλήρες δείγμα από την Pareto  $P(\theta)$ . Η από κοινού κατανομή του δείγματος είναι

$$f(\underline{x}; \theta) = \prod_{i=1}^n \theta x_i^{-(\theta+1)} = \theta^n (\prod_{i=1}^n x_i)^{-(\theta+1)}$$

με λογαριθμική συνάρτηση πιθανοφάνειας την

$$\log[L(\theta | \underline{x})] = n \ln \theta - (\theta + 1) \sum_{i=1}^n \ln x_i.$$



Ο εκτιμητής μέγιστης πιθανοφάνειας είναι η λύση της εξίσωσης

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n \ln x_i = 0.$$

Δηλαδή

$$\hat{\theta} = [\overline{\ln x.}]^{-1}.$$

Ο πληροφοριακός αριθμός είναι

$$I(\theta) = -E\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right) = \frac{1}{\theta^2} \quad (4.4)$$

### 4.3 Εκτιμητής μέγιστης πιθανοφάνειας για δεξιά λογοκριμένο δείγμα

Έστω ένα δείγμα όπου  $n$  το μέγεθος του πλήρους δείγματος,  $n_1$  ο αριθμός των πλήρων παρατηρήσεων μικρότερο ή ίσο του  $T$ , ενώ  $n-n_1$  είναι ο αριθμός των λογοκριμένων παρατηρήσεων πάνω από το όριο  $T$ . Η λογαριθμική συνάρτηση πιθανότητας, ενός δείγματος που ακολουθεί την κατανομή Pareto είναι:

$$\ln L = n \ln \theta - (\theta + 1) \sum_{i=1}^{n_1} \ln x_i + (n - n_1) \theta \ln \left(\frac{1}{T}\right) + const$$

Ο εκτιμητής μέγιστης πιθανοφάνειας είναι η λύση της εξίσωσης

$$\frac{\partial \ln L}{\partial \theta} = \frac{n_1}{\theta} - \sum_{i=1}^{n_1} \ln x_i + (n - n_1) \ln \left(\frac{1}{T}\right) = 0.$$

Άρα

$$\hat{\theta} = n_1 \left[ \sum_{i=1}^{n_1} \ln x_i + (n - n_1) \ln T \right]^{-1}.$$

Για την εύρεση του Πληροφοριακού αριθμού στην περίπτωση της λογοκρισίας τύπου I και II χρησιμοποιείται ο τύπος

$$\int_0^T \left(\frac{\partial \log f}{\partial \theta}\right)^2 f(t;\theta) dt + \left(\frac{\partial \log \bar{F}(T;\theta)}{\partial \theta}\right)^2 \bar{F}(T;\theta)$$

όπου

$$\left(\frac{\partial \log \bar{F}(T_i; \theta)}{\partial \theta}\right)^2 \bar{F}(T_i; \theta) = \left(\ln \frac{\gamma}{T}\right)^2 \left(\frac{T}{\gamma}\right)^{-\theta} \quad (4.5)$$

Επιπλέον,

$$\int_{\gamma}^T \left(\frac{\partial \log f}{\partial \theta}\right)^2 f(t; \theta) dt = \int_{\gamma}^T \left(\frac{1}{\theta} + \ln \frac{\gamma}{t}\right)^2 \frac{\theta}{\gamma} \left(\frac{\gamma}{t}\right)^{\theta+1} dt = \int_{\gamma}^T \left(\frac{1}{\theta} - \ln \frac{t}{\gamma}\right)^2 \theta \left(\frac{t}{\gamma}\right)^{-(\theta+1)} d\frac{t}{\gamma}$$

Με τη βοήθεια του μετασχηματισμού  $w = t/\gamma$  έχω

$$\int_1^{\frac{T}{\gamma}} \left(\frac{1}{\theta} - \ln w\right)^2 \theta w^{-(\theta+1)} dw.$$

Με παραγοντική ολοκλήρωση προκύπτει:

$$\frac{1}{\theta^2} \left[ -\left(\frac{T}{\gamma}\right)^{-\theta} + 1 - \theta^2 \ln \left(\frac{T}{\gamma}\right)^2 \left(\frac{T}{\gamma}\right)^{-\theta} \right] \quad (4.6)$$

Από τις (4.5) και (4.6) συνεπάγεται ότι:

$$\begin{aligned} \left(\ln \frac{\gamma}{T}\right)^2 \left(\frac{T}{\gamma}\right)^{-\theta} + \frac{1}{\theta^2} \left[ -\left(\frac{T}{\gamma}\right)^{-\theta} + 1 - \theta^2 \left(\ln \frac{T}{\gamma}\right)^2 \left(\frac{T}{\gamma}\right)^{-\theta} \right] &= \frac{1}{\theta^2} \left( 1 - \left(\frac{T}{\gamma}\right)^{-\theta} \right) \\ &= \frac{1}{\theta^2} \left( 1 - \left(\frac{\gamma}{T}\right)^{\theta} \right). \end{aligned}$$

Δηλαδή

$$I_p(\gamma) = \frac{1}{\theta^2} \left( 1 - \left(\frac{\gamma}{T}\right)^{\theta} \right)$$

Στη συνέχεια, εκφράζεται σε σχέση με το ποσοστό λογοκρίσιμης  $p$ , το οποίο σε σχέση με το  $T$  δίνεται από τη σχέση:

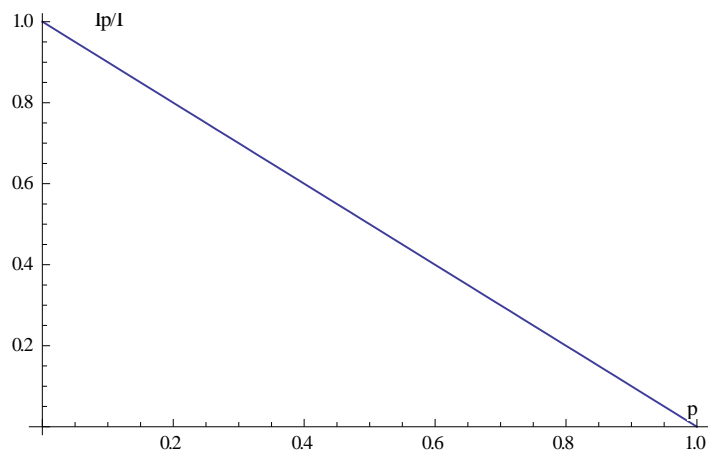
$$\bar{F}(T_i; \theta) = \left(\frac{\gamma}{T}\right)^{\theta} = p$$

Τελικά:

$$I_p(\gamma) = 1 - \left(\frac{\gamma}{T}\right)^\theta = 1 - p$$

Τελικά ο λόγος  $I_p/I_0$  είναι ανεξάρτητος του  $\theta$  και  $\gamma$  και ισούται με

$$\frac{I_p}{I_0} = 1 - p$$



Διάγραμμα 14: Λόγος  $I_p/I_0$

Όπως και στην περίπτωση των άλλων κατανομών το πρόγραμμα του Mathematica με το οποίο έγιναν οι προσομοιώσεις και οι υπολογισμοί των ΕΜΠ παρουσιάζονται παρακάτω.

(βλ. Παράρτημα Π3)

|     |               | $\theta=2$ |                    | $\theta=3$ |                    | $\theta=4$ |                    |
|-----|---------------|------------|--------------------|------------|--------------------|------------|--------------------|
| n   | Censore level | Mean       | Standard Deviation | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0             | 2,0386     | 0,2945             | 3,0577     | 0,4440             | 4,0866     | 0,5943             |
|     | 10            | 2,2516     | 0,2628             | 3,3813     | 0,4018             | 4,4992     | 0,5286             |
|     | 20            | 2,5228     | 0,2603             | 3,7851     | 0,3927             | 5,0574     | 0,5132             |
|     | 50            | 4,0137     | 0,2722             | 6,0280     | 0,4184             | 8,0372     | 0,5578             |
|     | 60            | 5,0156     | 0,3001             | 7,5224     | 0,4532             | 10,0347    | 0,5924             |
|     | 70            | 6,6820     | 0,3312             | 10,0229    | 0,5040             | 13,3649    | 0,6637             |
| 80  | 0             | 2,0211     | 0,2320             | 3,0391     | 0,3428             | 4,0507     | 0,4599             |
|     | 10            | 2,2451     | 0,2048             | 3,3591     | 0,3128             | 4,4852     | 0,4131             |
|     | 20            | 2,5159     | 0,1980             | 3,7735     | 0,3032             | 5,0335     | 0,4004             |
|     | 50            | 4,0097     | 0,2154             | 6,0190     | 0,3328             | 8,0355     | 0,4344             |
|     | 60            | 5,0111     | 0,2372             | 7,5242     | 0,3477             | 10,0263    | 0,4698             |
|     | 70            | 6,6829     | 0,2609             | 10,0016    | 0,3909             | 13,3671    | 0,5206             |
| 100 | 0             | 2,0213     | 0,2052             | 3,0242     | 0,3051             | 4,0359     | 0,4017             |
|     | 10            | 2,2394     | 0,1802             | 3,3562     | 0,2743             | 4,4796     | 0,3696             |
|     | 20            | 2,5069     | 0,1791             | 3,7698     | 0,2692             | 5,0179     | 0,3520             |
|     | 50            | 4,0105     | 0,1939             | 6,0119     | 0,2917             | 8,0313     | 0,3859             |
|     | 60            | 5,0067     | 0,2084             | 7,5144     | 0,3144             | 10,0212    | 0,4212             |
|     | 70            | 6,6730     | 0,2274             | 10,0227    | 0,3562             | 13,3604    | 0,4749             |
| 150 | 0             | 2,0168     | 0,1651             | 3,0223     | 0,2455             | 4,0232     | 0,3365             |
|     | 10            | 2,2324     | 0,1451             | 3,3459     | 0,2209             | 4,4703     | 0,2998             |
|     | 20            | 2,5084     | 0,1443             | 3,7639     | 0,2193             | 5,0190     | 0,2905             |
|     | 50            | 4,0075     | 0,1594             | 6,0029     | 0,2349             | 8,0145     | 0,3085             |
|     | 60            | 5,0076     | 0,1700             | 7,5085     | 0,2556             | 10,0198    | 0,3354             |
|     | 70            | 6,6690     | 0,1871             | 10,0127    | 0,2846             | 13,3430    | 0,3710             |
| 200 | 0             | 2,0096     | 0,1433             | 3,0145     | 0,2152             | 4,0208     | 0,2864             |
|     | 10            | 2,2315     | 0,1300             | 3,3474     | 0,1936             | 4,4595     | 0,2575             |
|     | 20            | 2,5043     | 0,1236             | 3,7599     | 0,1892             | 5,0052     | 0,2532             |
|     | 50            | 4,0056     | 0,1368             | 6,0068     | 0,2040             | 8,0055     | 0,2711             |
|     | 60            | 5,0041     | 0,1479             | 7,5089     | 0,2224             | 10,0067    | 0,2905             |
|     | 70            | 6,6685     | 0,1600             | 9,9947     | 0,2426             | 13,3438    | 0,3284             |
| 250 | 0             | 2,0091     | 0,1281             | 3,0131     | 0,1881             | 4,0179     | 0,2570             |
|     | 10            | 2,2293     | 0,1160             | 3,3429     | 0,1689             | 4,4536     | 0,2280             |
|     | 20            | 2,5063     | 0,1101             | 3,7557     | 0,1659             | 5,0164     | 0,2227             |
|     | 50            | 4,0025     | 0,1217             | 6,0067     | 0,1809             | 8,0111     | 0,2426             |
|     | 60            | 5,0049     | 0,1288             | 7,5065     | 0,1946             | 10,0107    | 0,2621             |
|     | 70            | 6,6693     | 0,1450             | 10,0081    | 0,2193             | 13,3356    | 0,2914             |
| 300 | 0             | 2,0083     | 0,1178             | 3,0098     | 0,1755             | 4,0145     | 0,2336             |
|     | 10            | 2,2292     | 0,1047             | 3,3405     | 0,1594             | 4,4509     | 0,2108             |
|     | 20            | 2,5043     | 0,1014             | 3,7557     | 0,1562             | 5,0100     | 0,2063             |
|     | 50            | 4,0050     | 0,1102             | 6,0038     | 0,1662             | 8,0043     | 0,2211             |
|     | 60            | 5,0025     | 0,1187             | 7,5083     | 0,1788             | 10,0055    | 0,2410             |
|     | 70            | 6,6681     | 0,1341             | 10,0029    | 0,1998             | 13,3327    | 0,2661             |

Πίνακας 7: Pareto Distribution Pa(2), Pa(3), Pa(4)

#### 4.4 Εκτιμητής μέγιστης πιθανοφάνειας τυχαία λογοκριμένο δείγμα

Έστω  $t_1, t_2, \dots, t_n$  ένα τυχαία λογοκριμένο δείγμα με ποσοστό λογοκρισίας

$$p = 1 - q = \bar{F}(T).$$

Το δείγμα είναι της μορφής

$$\underbrace{t_1, t_2, \dots, t_{n_1}}_{n_1}, \underbrace{c_1, c_2, \dots, c_{n-n_1}}_{n-n_1}$$

Οι χρόνοι της λογοκρισίας έχουν ως εξής:

$$g_L(c) = \frac{1}{L-1} I_{[1,T]}^{(c)}, U[1, L]$$

με συνάρτηση επιβίωσης

$$\bar{G}(c) = \frac{L-c}{L-1}$$

και

$$f(t; \theta) = \theta t^{-\theta-1} I_{[1,\infty]}^{(t)}$$

Με την αντίστοιχη συνάρτηση επιβίωσης

$$\bar{F}(t) = -[x^{-\theta}]_c^\infty = c^{-\theta}$$

Από την (2.17) η συνάρτηση πιθανοφάνειας είναι

$$L(\theta) \propto \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(c_i; \theta)^{1-\delta_i} = \prod_{i=1}^{n_1} \theta t_i^{-(\theta+1)} \prod_{i=1}^{n-n_1} c_i^{-\theta} \quad (4.7)$$

Στη συνέχεια ακολουθώντας τη γνωστή διαδικασία , αφού χρησιμοποιηθεί ο λογάριθμος και γίνει παραγωγίσιμη, γίνεται επίλυση ως προς  $\theta$  και προκύπτει ο ΕΜΠ η στατιστική συνάρτηση

$$\hat{\theta} = n_1 \left[ \sum_{i=1}^{n_1} \ln x_i + \sum_{i=1}^{n-n_1} \ln c_i \right]^{-1} \quad (4.8)$$

### Πίνακας αποτελεσμάτων προσομοίωσης

Όπως και στην περίπτωση των άλλων κατανομών το πρόγραμμα του Mathematica με το οποίο έγιναν οι προσομοιώσεις και οι υπολογισμοί των ΕΜΠ παρουσιάζονται παρακάτω. (βλ. Παράρτημα Π4)

|     |                 | $\theta=2$ |                    | $\theta=3$ |                    | $\theta=4$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|------------|--------------------|
| n   | Censoring level | Mean       | Standard Deviation | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 2,0343     | 0,3014             | 3,0647     | 0,4525             | 4,0926     | 0,6007             |
|     | 10              | 2,0416     | 0,3315             | 3,0695     | 0,4736             | 4,0824     | 0,5969             |
|     | 20              | 2,0476     | 0,3433             | 3,0453     | 0,4927             | 4,0953     | 0,7070             |
|     | 50              | 2,0496     | 0,3878             | 3,0941     | 0,6500             | 4,0967     | 0,8586             |
|     | 60              | 2,0531     | 0,4293             | 3,1484     | 0,6783             | 4,0999     | 0,9418             |
|     | 70              | 2,0714     | 0,5450             | 3,1491     | 0,7835             | 4,1152     | 1,1449             |
| 80  | 0               | 2,0430     | 0,2400             | 3,0444     | 0,3372             | 4,0284     | 0,4443             |
|     | 10              | 2,0349     | 0,2440             | 3,0548     | 0,3714             | 4,0438     | 0,4777             |
|     | 20              | 2,1188     | 0,2538             | 3,0809     | 0,3751             | 4,0728     | 0,5269             |
|     | 50              | 2,2006     | 0,3178             | 3,1109     | 0,4918             | 4,0809     | 0,6281             |
|     | 60              | 2,2340     | 0,3495             | 3,1809     | 0,5283             | 4,0868     | 0,7132             |
|     | 70              | 2,2414     | 0,3972             | 3,1909     | 0,6330             | 4,0947     | 0,8167             |
| 100 | 0               | 2,0204     | 0,2084             | 3,0298     | 0,3208             | 4,0552     | 0,4077             |
|     | 10              | 2,0090     | 0,2103             | 3,0400     | 0,3280             | 4,0566     | 0,4189             |
|     | 20              | 2,2021     | 0,2383             | 3,0466     | 0,3098             | 4,0541     | 0,4593             |
|     | 50              | 2,2503     | 0,2925             | 3,0496     | 0,4371             | 4,0560     | 0,5599             |
|     | 60              | 2,2840     | 0,3281             | 3,0511     | 0,4755             | 4,0588     | 0,6455             |
|     | 70              | 2,3019     | 0,3580             | 3,0526     | 0,5715             | 4,0671     | 0,7106             |
| 150 | 0               | 2,0015     | 0,1701             | 3,0443     | 0,2551             | 4,0199     | 0,3155             |
|     | 10              | 2,0173     | 0,1770             | 3,0551     | 0,2604             | 4,0408     | 0,3304             |
|     | 20              | 2,0180     | 0,1906             | 3,0609     | 0,2783             | 4,0506     | 0,3684             |
|     | 50              | 2,0241     | 0,2350             | 3,0836     | 0,3646             | 4,0658     | 0,4415             |
|     | 60              | 2,0329     | 0,2485             | 3,0858     | 0,4012             | 4,0776     | 0,5212             |
|     | 70              | 2,0437     | 0,3020             | 3,0927     | 0,4574             | 4,0786     | 0,5744             |
| 200 | 0               | 1,9974     | 0,1404             | 3,0198     | 0,2125             | 4,0067     | 0,2740             |
|     | 10              | 2,0097     | 0,1540             | 3,0215     | 0,2283             | 4,0338     | 0,3025             |
|     | 20              | 2,0126     | 0,1502             | 3,0211     | 0,2507             | 4,0458     | 0,3058             |
|     | 50              | 2,0160     | 0,2084             | 3,0214     | 0,2837             | 4,0478     | 0,4038             |
|     | 60              | 2,0190     | 0,2322             | 3,0219     | 0,3467             | 4,0531     | 0,4276             |
|     | 70              | 2,0271     | 0,2621             | 3,0399     | 0,3989             | 4,0626     | 0,5189             |
| 250 | 0               | 1,9989     | 0,1260             | 3,0273     | 0,1953             | 4,0398     | 0,2643             |
|     | 10              | 2,0201     | 0,1358             | 3,0138     | 0,2064             | 3,9937     | 0,2589             |
|     | 20              | 2,0297     | 0,1367             | 3,0196     | 0,2134             | 4,0153     | 0,2735             |
|     | 50              | 2,0368     | 0,1828             | 3,0151     | 0,2616             | 4,0199     | 0,3580             |
|     | 60              | 2,0391     | 0,2052             | 3,0190     | 0,2722             | 4,0248     | 0,4006             |
|     | 70              | 2,0573     | 0,2267             | 3,0225     | 0,3614             | 4,0343     | 0,4693             |
| 300 | 0               | 2,0120     | 0,1126             | 3,0130     | 0,1666             | 4,0085     | 0,2435             |
|     | 10              | 2,0154     | 0,1232             | 3,0191     | 0,1705             | 4,0245     | 0,2581             |
|     | 20              | 2,0242     | 0,1334             | 3,0165     | 0,1917             | 4,0345     | 0,2643             |
|     | 50              | 2,0311     | 0,1648             | 3,0245     | 0,2402             | 4,0410     | 0,3080             |
|     | 60              | 2,0480     | 0,1965             | 3,0252     | 0,2965             | 4,0440     | 0,3607             |
|     | 70              | 2,0580     | 0,2072             | 3,0452     | 0,3265             | 4,0511     | 0,4086             |

Πίνακας 8: Random Pareto Distribution Pa(2), Pa(3), Pa(4)

Συμπερασματικά από τους παραπάνω πίνακες προκύπτει ότι η λογοκρισία τύπου I και II έχουν μεγαλύτερη επίπτωση στην εκτίμηση των παραμέτρων σε σχέση με την τυχαία λογοκρισία.

Για να εκφραστεί ο πληροφοριακός αριθμός του Fisher σε σχέση με το ποσοστό  $p$  θα βρεθεί η σχέση μεταξύ του  $p$  και του  $L$  από τη σχέση

$$1 - p = \int_1^L \int_1^c f(t)g(t)dt dc$$

όπου

$$\int_1^c f(x)g(x)dx = \frac{1}{L-1} \int_1^c \theta x^{-\theta-1} dx = -[x^{-\theta}]_1^c = \frac{-c^{-\theta} + 1}{L-1} = \frac{1 - c^{-\theta}}{L-1}$$

άρα:

$$1 - p = \frac{1}{L-1} \int_1^L (1 - c^{-\theta}) dc = \frac{1}{L-1} \left[ L - 1 - \left[ \frac{1}{-\theta+1} c^{-\theta+1} \right]_1^L \right] = \frac{1}{L-1} \left[ L - 1 - \frac{L^{-\theta+1}}{-\theta+1} + \frac{1}{-\theta+1} \right].$$

Μετά από κάποιες πράξεις προκύπτει

$$p = \frac{L^{-\theta+1} - 1}{(L-1)(-\theta+1)} \quad (4.9)$$

Στη συνέχεια, για την εύρεση του πληροφοριακού αριθμού υπάρχουν οι ποσότητες

$$\log f = \log \theta + (-\theta + 1) \log x$$

$$\frac{\partial \log f}{\partial \theta} = \frac{1}{\theta} - x$$

$$\log \bar{F} = -\theta \log x$$

$$\frac{\partial \log \bar{F}}{\partial \theta} = -\log x$$

$$I_p(\theta) = \int_1^L \left( \frac{1}{\theta} - \log x \right)^2 \theta x^{-\theta-1} \frac{L-x}{L-1} dx + \int_1^L \log^2 x \frac{1}{L-1} x^{-\theta} dx$$



$$\begin{aligned}
&= \frac{1}{L-1} \int_1^L \left( \frac{1}{\theta^2} - \frac{2}{\theta} \log x + \log^2 x \right) \theta x^{-\theta-1} (L-x) + \log^2 x x^{-\theta} dx \\
&= \frac{1}{L-1} \int_1^L \left[ \left( \frac{1}{\theta^2} - \frac{2}{\theta} \log x + \log^2 x \right) (\theta L x^{-\theta-1} - \theta x^{-\theta}) + \log^2 x x^{-\theta} \right] dx \quad (4.8)
\end{aligned}$$

Η ποσότητα κάτω από το ολοκλήρωμα μπορεί να γραφεί ως εξής

$$\begin{aligned}
&\int_1^L \left( \frac{L}{\theta} x^{-\theta-1} - \frac{1}{\theta} x^{-\theta} \right) dx + \int_1^L (-\theta+1) \log^2 x x^{-\theta} dx + \\
&\int_1^L (-2L \log x x^{-\theta-1} + \theta L \log^2 x x^{-\theta-1}) dx + \int_1^L 2 \log x x^{-\theta} dx \quad (4.9)
\end{aligned}$$

Ας υπολογιστούν τα ολοκληρώματα

$$\begin{aligned}
\int_1^L \left( \frac{L}{\theta} x^{-\theta-1} - \frac{1}{\theta} x^{-\theta} \right) dx &= \frac{L}{\theta} \int_1^L x^{-\theta-1} dx - \frac{1}{\theta} \int_1^L x^{-\theta} dx \\
&= -\frac{L}{\theta^2} [x^{-\theta}]_1^L - \left[ \frac{1}{\theta(-\theta+1)} x^{-\theta+1} \right]_1^L \\
&= -\frac{L}{\theta^2} [L^{-\theta} - 1] - \frac{1}{\theta(-\theta+1)} [L^{-\theta+1} - 1] \\
&= -\frac{L^{-\theta+1}}{\theta^2} + \frac{L}{\theta^2} - \frac{L^{-\theta+1}}{\theta(-\theta+1)} + \frac{1}{\theta(-\theta+1)}
\end{aligned}$$

Όμως:

$$-\frac{L^{-\theta+1}}{\theta^2} - \frac{L^{-\theta+1}}{\theta(-\theta+1)} = -\frac{1}{\theta^2} - \frac{1}{\theta(-\theta+1)} = \frac{-\theta-1+\theta}{\theta^2(-\theta+1)} = -\frac{1}{\theta^2(-\theta+1)} L^{-\theta+1}$$

και

$$\frac{L}{\theta^2} + \frac{1}{\theta(-\theta+1)} = \frac{1}{\theta} \left[ \frac{L}{\theta} + \frac{1}{-\theta+1} \right] = \frac{L(-\theta+1)+\theta}{\theta^2(-\theta+1)}.$$

Συνολικά:

$$\frac{L}{\theta} \int_1^L x^{-\theta-1} dx - \frac{1}{\theta} \int_1^L x^{-\theta} dx = \frac{1}{L-1} \left( \frac{L(-\theta+1)+\theta}{\theta^2(-\theta+1)} - \frac{1}{\theta^2(-\theta+1)} L^{-\theta+1} \right) \quad (4.10)$$

Στη συνέχεια γίνεται ο υπολογισμός του ολοκληρώματος

$$\begin{aligned} \int_0^L \log^2 x x^{-\theta-1} dx &= \frac{1}{-\theta} \int_1^L \log^2 dx^{-\theta} = \frac{1}{-\theta} \{ [\log^2 x x^{-\theta}]_1^L - \int_1^L 2x^{-\theta} \log x \frac{1}{x} dx \} \\ &= \frac{1}{-\theta} \{ \log^2 L L^{-\theta} - 2 \int_1^L x^{-\theta-1} \log x dx + \int_0^L \log^2 x x^{-\theta-1} dx \} \\ &= \frac{1}{-\theta} \log^2 L L^{-\theta} + \frac{2}{\theta} \int_1^L x^{-\theta-1} \log x dx. \end{aligned}$$

$$\begin{aligned} &\int_1^L (-2L \log x x^{\theta-1} + \theta L \log^2 x x^{\theta-1}) dx \\ &= \int_1^L (-2L \log x x^{-\theta-1} dx + \theta L \int_0^L \log^2 x x^{-\theta-1} dx \\ &= \int_1^L (-2L \log x x^{-\theta-1} dx - L \log^2 L L^{-\theta} + 2L \int_1^L x^{-\theta-1} \log x dx \\ &= -L \log^2 L L^{-\theta} \end{aligned}$$

Δηλαδή

$$\int_1^L (-2L \log x x^{-\theta-1} + \theta L \log^2 x x^{-\theta-1}) dx = -L \log^2 L L^{-\theta} \quad (4.11)$$

Για το ολοκλήρωμα

$$(1-\theta) \int_1^L x^{-\theta} \log^2 x dx$$

Υπολογίζεται πρώτα το ολοκλήρωμα

$$\begin{aligned}
\int_1^L x^{-\theta} \log^2 x \, dx &= \frac{1}{-\theta + 1} \int_1^L x^{-\theta} \log^2 x \, dx^{-\theta+1} \\
&= \frac{1}{-\theta + 1} \left\{ [\log^2 x x^{-\theta+1}]_1^L - 2 \int_1^L \frac{x^{-\theta+1}}{x} \log x \, dx \right\} \\
&= \frac{1}{-\theta + 1} \left\{ \log^2 L L^{-\theta+1} - 2 \int_1^L x^{-\theta} \log x \, dx \right\} \\
&= \frac{1}{-\theta + 1} \log^2 L L^{-\theta+1} - \frac{2}{-\theta + 1} \int_1^L x^{-\theta} \log x \, dx.
\end{aligned}$$

Τελικά

$$(1 - \theta) \int_1^L x^{-\theta} \log^2 x \, dx = \log^2 L L^{-\theta+1} - 2 \int_1^L x^{-\theta} \log x \, dx \quad (4.12)$$

Το ολοκλήρωμα (4.8) από τα (4.9), (4.10) και (4.11) (4.12) γίνεται

$$I_p(\theta) = \frac{1}{\theta^2} \left( 1 - \frac{L^{-\theta+1} - 1}{(-\theta + 1)(L - 1)} \right) = \frac{1 - p}{\theta^2}$$

Παρατηρούμε ότι ο λόγος των πληροφοριακών αριθμών για την παράμετρο μορφής για την Pareto είναι ίδιος με το λόγο των πληροφοριακών αριθμών για την παράμετρο κλίμακας για την Weibull.

## ΚΕΦΑΛΑΙΟ 5

### Η επίδραση της λογοκρισίας στην εκτίμηση παραμέτρων για την κατανομή Gamma

#### 5.1 Εισαγωγικά

Ο Lancaster το 1966 χρησιμοποιώντας τη μέθοδο Laplace, παρατήρησε την κατανομή Gamma. Η εφαρμογή της κατανομής Gamma δίνει χρήσιμες πληροφορίες για πολλές «φυσικές» καταστάσεις. Οι τομείς που μπορεί να εφαρμοστεί η κατανομή Gamma ποικίλουν. Έχει χρησιμοποιηθεί για να κάνει ρεαλιστικές προσαρμογές της εκθετικής κατανομής στη διάρκεια ζωής του πληθυσμού και σε ποικιλία διεργασιών σε σχέση με το χρόνο. Για παράδειγμα σε θέματα που σχετίζονται με μετεωρολογικές διεργασίες ή ακόμα και διεργασίες καθίζησης των πετρωμάτων σε σχέση με το χρόνο. Αποτελεί την ιδανική περίπτωση για τη μέτρηση και τον υπολογισμό του εισοδήματος (Salem and Mount (1974)). Ιδιαίτερο ενδιαφέρον παρουσιάζουν η εφαρμογή της κατανομής Gamma στη στατιστική οικολογία από τους Dennis and Patil (1984). Μεταξύ άλλων ξεχωρίζει και η έρευνα των Costantino και Desharnais (1981) για την εμπειρική εφαρμογή και μελέτη του πληθυσμού των σκαθαριών σίτου (*Tribolium*). Η κατανομή Gamma χρησιμοποιείται για να διαμορφώσει το μέγεθος των ασφαλιστικών απαιτήσεων. Αυτό σημαίνει, ότι το σύνολο των απαιτήσεων συσσωρεύεται και διαμορφώνεται η κατανομή Gamma. Η κατανομή χρησιμοποιείται επίσης για τη μοντελοποίηση σφαλμάτων σε πολυεπίπεδα μοντέλα Poisson παλινδρόμησης, επειδή ο συνδυασμός της κατανομής Poisson και της κατανομής Gamma δίνει την μία αρνητική διωνυμική κατανομή.

Μια τυχαία μεταβλητή  $X$  ακολουθεί την κατανομή Gamma με παράμετρο κλίμακας  $\theta$  και παράμετρο μορφής  $\gamma$  εάν η συνάρτηση πυκνότητας πιθανότητας είναι

$$f(x; \gamma, \theta) = \frac{1}{\theta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-\frac{x}{\theta}}, x > 0, \gamma > 0, \theta > 0 \quad (5.1)$$

Συμβολίζεται  $X \sim Ga(\gamma, \theta)$ .

Στην βιβλιογραφία υπάρχει και η 3-παραμετρική κατανομή Gamma με μια επιπλέον παράμετρο θέσης  $\alpha$

$$f(x) = \frac{1}{\theta^\gamma \Gamma(\gamma)} (x - \alpha)^{\gamma-1} e^{-\frac{x-\alpha}{\theta}}, x > \alpha, \gamma > 0, \theta > 0.$$

Η περίπτωση αυτή δεν είναι θέμα της παρούσας μελέτης επειδή το στήριγμα εξαρτάται από την άγνωστη παράμετρο, οι συνθήκες κανονικότητας δεν ισχύουν και η μελέτη τους είναι εξαιρετικά δύσκολη.

Η εκθετική κατανομή είναι ειδική περίπτωση της κατανομής Gamma για  $\gamma=1$ . Δηλαδή  $Ga(1, \theta) \equiv Exp(\theta)$ .

Η μέση τιμή, διακύμανση, συντελεστής ασυμμετρίας και η ροπογεννήτρια του X είναι

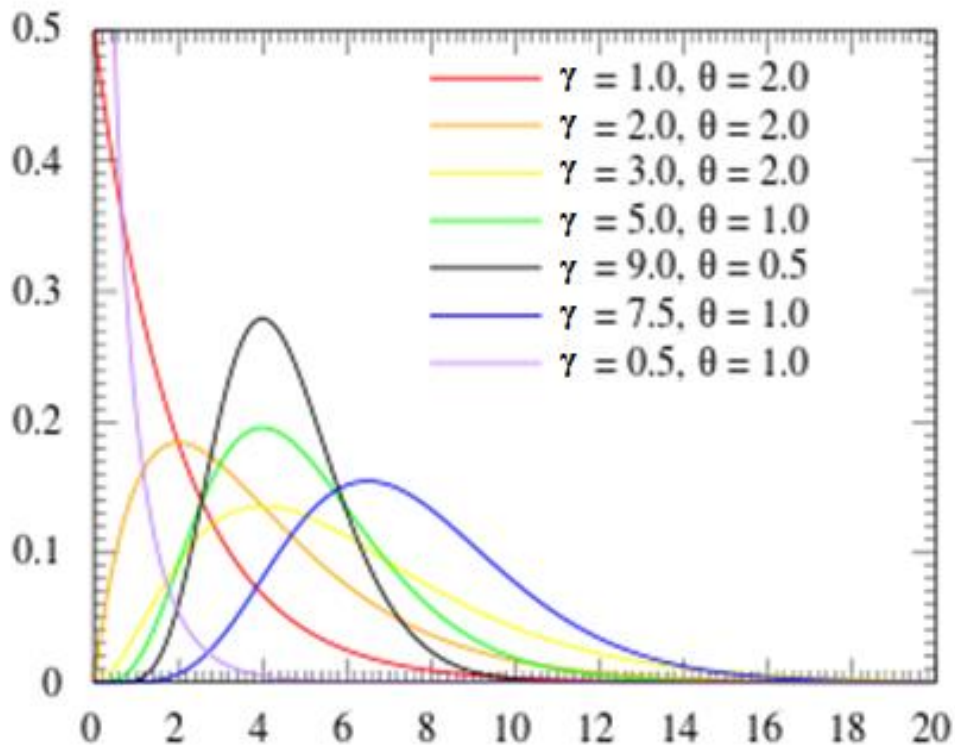
$$\begin{aligned} E(X) &= \theta\gamma \\ Var(X) &= \theta\gamma^2 \\ \beta_1 &= \frac{2}{\sqrt{\gamma}} \\ M(t) &= \left(\frac{\theta}{\theta-t}\right)^\gamma, t < \theta \end{aligned}$$

Η ροπογεννήτρια μπορεί να χρησιμοποιηθεί για τον υπολογισμό ροπών.

Επειδή η Gamma δεν είναι απλού τύπου κατανομή, η εκτίμηση των πιθανοτήτων μπορεί να ληφθεί από τον πίνακα της κατανομής  $\chi^2$ .

Εάν το  $X \sim Ga(n, \lambda)$ , τότε  $2\lambda X \sim \chi_{2n}^2$ .

Στη θεωρία πιθανοτήτων και της στατιστικής, η κατανομή Gamma είναι μια οικογένεια δύο παραμέτρων των συνεχών κατανομών. Η εκθετική κατανομή και η  $\chi^2$  κατανομή είναι ειδικές περιπτώσεις της κατανομής Gamma.



Διάγραμμα 15: Gamma Distribution (πηγή Wikipedia)

## 5.2 Εκτιμητής Μέγιστης Πιθανοφάνειας για πλήρες δείγμα

Έστω  $x_1, x_2, \dots, x_n$  ένα τυχαίο δείγμα από την (5.1). Όπως έχουμε πει ο ΕΜΠ για τις παραμέτρους  $\theta$  και  $\gamma$  είναι η λύση του συστήματος

$$\begin{aligned} \frac{\partial L(\theta; \underline{x})}{\partial \gamma} &= -n \frac{\Gamma'(\gamma)}{\Gamma(\gamma)} - n \log(\theta) + \sum \log x_i = 0 \\ \frac{\partial L(\theta; \underline{x})}{\partial \theta} &= -n \frac{\gamma}{\theta} + \frac{\sum x_i}{\theta^2} = 0 \end{aligned}$$

Το παραπάνω σύστημα γράφεται στη μορφή

$$-n \Psi(\gamma) + n \log \gamma - n \log \bar{x} + \sum \log x_i = 0 \quad (5.2)$$

$$\theta = \frac{\sum x_i}{n\gamma} = \frac{\bar{x}}{\gamma} \quad (5.3)$$

όπου  $\Psi(\gamma)$  είναι η δίγαμμα συνάρτηση  $\Psi(\gamma) = \log'(\Gamma(\gamma))$ .

Υπολογίζεται η ρίζα  $\hat{\gamma}$  της (5.2) (EMΠ του  $\gamma$ ) και στη συνέχεια υπολογίζεται από την (5.3) τον EMΠ  $\hat{\theta}$  του  $\theta$ . Μπορεί να αποδειχθεί ότι (5.2) έχει πάντα μία και μοναδική λύση.

Ο πληροφοριακός πίνακας των EMΠ είναι

$$I(\gamma, \theta) = \begin{bmatrix} \psi'(\gamma) & \theta^{-1} \\ \theta^{-1} & \gamma\theta^{-2} \end{bmatrix} \quad (5.4)$$

Αν εκτιμάται μόνο το  $\gamma$  ενώ το  $\theta$  είναι γνωστό, τότε η πληροφορία είναι

$$I(\gamma) = \psi'(\gamma) \quad (5.5)$$

Ενώ όταν εκτιμάται το  $\theta$  ενώ το  $\gamma$  είναι γνωστό, η πληροφορία είναι

$$I(\gamma, \theta) = \gamma\theta^{-2} \quad (5.6)$$

### 5.3 Εκτιμητής Μέγιστης Πιθανοφάνειας για λογοκριμένο δείγμα τύπου I και II

Έστω  $\underbrace{x_1, x_2, \dots, x_{n_1}}_{n_1}, \underbrace{T, T, \dots, T}_{n-n_1}$  ένα λογοκριμένο δείγμα από την (5.1).

Λογαριθμίζοντας τον εκτιμητή μέγιστης πιθανοφάνειας ενός λογοκριμένου δείγματος τύπου I η τύπου II προκύπτει

$$\ln L = c - n_1 \ln \Gamma(\gamma) - n_1 \gamma \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n_1} x_i + (\gamma - 1) \sum_{i=1}^{n_1} \ln x_i + (n - n_1) \ln[1 - F(T)] \quad (5.7)$$

όπου  $n_1$  είναι ο αριθμός των μη λογοκριμένων παρατηρήσεων και  $n - n_1$  είναι το πλήθος των λογοκριμένων τιμών που έχει αντικατασταθεί με  $T = \bar{F}^{-1}(p, \theta, \gamma)$ .

Παραγωγίζοντας την (5.7) ως προς την  $\gamma$  και  $\theta$  αντίστοιχα προκύπτει

$$\begin{cases} \frac{\partial \ln L}{\partial \gamma} = -n_1 \Psi(\gamma) - n_1 \ln \theta + \sum_{i=1}^{n_1} x_i - \frac{n - n_1}{1 - F(T)} \frac{\partial F(T)}{\partial \gamma} = 0 \\ \frac{\partial \ln L}{\partial \theta} = -\frac{n_1 \gamma}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i - \frac{n - n_1}{1 - F(T)} \frac{\partial F(T)}{\partial \theta} = 0 \end{cases}$$

Η επίλυση των εξισώσεων αυτών γίνεται με την βοήθεια του Mathematica.

Ο ασυμπτωτικός πίνακας συνδιακύμανσης των εκτιμητών μέγιστης πιθανοφάνειας είναι ο αντίστροφος πίνακας του πίνακα πληροφορίας του Fisher ο οποίος είναι ο εξής (Harter and Moore, 1965).

$$I(p, \gamma, \theta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

όπου

$$I_{11}(\gamma, p) = \Psi'(\gamma) - \{[\Gamma(\gamma) - \Gamma(\gamma, z_p)][\Gamma''(\gamma) - \Gamma''(\gamma, z_p)] - [\Gamma'(\gamma) - \Gamma'(\gamma, z_p)]^2\} / p[\Gamma(\gamma)]^2$$

$$I_{22}(\gamma, p) = \theta^{-2} \{-\gamma(1-p) + 2\Gamma(\gamma+1, z_p) / \Gamma(\gamma) - z_p f(z_p) \frac{p(z_p - \gamma - 1) - z_p f(z_p)}{p}\}$$

$$I_{12}(\gamma, p) = I_{21}(\gamma, p) = \theta^{-1} \{1 - p - z_p f(z_p) \{p \ln z_p - \frac{\Gamma'(\gamma) - \Gamma'(\gamma, z_p)}{\Gamma(\gamma)}\} / p$$

Αξίζει να σημειωθεί ότι:

Όταν  $p \rightarrow 0$ , δηλαδή όταν δεν υπάρχει λογοκρισία τότε:

$$I_{11}(\gamma, p) \rightarrow (\ln[\Gamma(\gamma)])'' = \Psi'(\gamma)$$

$$I_{12}(\gamma, p) \rightarrow \frac{1}{\theta}$$

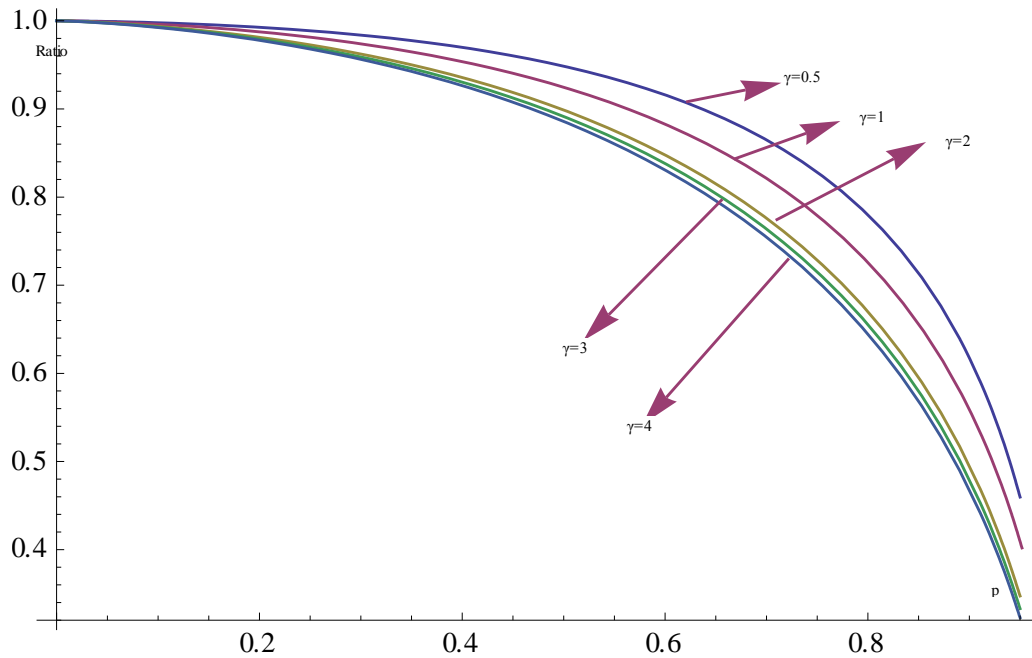
$$I_{22}(\gamma, p) \rightarrow \frac{\gamma}{\theta^2}.$$

Δηλαδή προκύπτει ο γνωστός πίνακας πληροφορίας για μη λογοκριμένα δείγματα ο οποίος είναι ο (5.4).

Για την μελέτη της επίδρασης του βαθμού λογοκρισίας τύπου I και II στον πληροφοριακό αριθμό του Fisher, γίνεται η γραφική παράσταση του λόγου



$$\frac{I_{11}(\gamma, p)}{I(\gamma, 0)} = \frac{\Psi'(\gamma) - \{[\Gamma(\gamma) - \Gamma(\gamma, z_p)][\Gamma''(\gamma) - \Gamma''(\gamma, z_p)] - [\Gamma'(\gamma) - \Gamma'(\gamma, z_p)]^2\} / p[\Gamma(\gamma)]^2}{\Psi'(\gamma)}$$

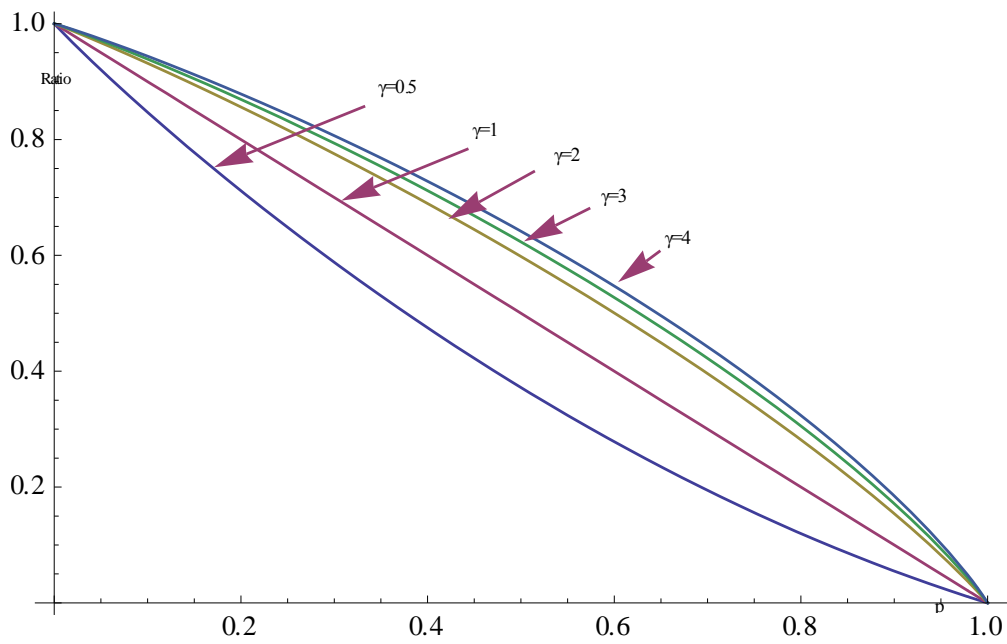


Διάγραμμα 16: Λόγος λογοκρίμενου δείγματος (p) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο κλίμακας, για διάφορες τιμές της  $\gamma$

Παρατηρεί κανείς ότι όταν μεγαλώνει η ασυμμετρία της κατανομής, μικραίνει ο λόγος των πληροφοριών δηλαδή αυξάνει η απώλεια της πληροφορίας.

Με ανάλογο τρόπο προκύπτει και η γραφική παράσταση του λόγου

$$\frac{I_{22}(\gamma, p)}{I_{22}(\gamma, 0)} = \frac{1}{\gamma} \left\{ -\gamma(1-p) + 2\Gamma(\gamma+1, z_p) / \Gamma(\gamma) - z_p f(z_p) \frac{p(z_p - \gamma - 1) - z_p f(z_p)}{p} \right\}$$



Διάγραμμα 17: Λόγος λογοκρινμένου δείγματος (p) προς τον πληροφοριακό αριθμό πλήρους δείγματος για την παράμετρο μορφής, για διάφορες τιμές της  $\gamma$

Εν αντιθέσει με την παράμετρο μορφής, εδώ παρατηρείται ότι όταν αυξάνει η ασυμμετρία τότε ελαττώνεται η πληροφορία.

(βλ. Παράρτημα Π5)

|     |                 | $\theta=2$ |                    | $\gamma=2$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|
| n   | Censoring level | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 2,0454     | 0,1833             | 2,0477     | 0,1665             |
|     | 10              | 1,6796     | 0,1221             | 2,0303     | 0,1676             |
|     | 20              | 1,4090     | 0,1400             | 1,9875     | 0,1855             |
|     | 50              | 0,9674     | 0,0980             | 1,9707     | 0,1818             |
|     | 60              | 0,8534     | 0,1142             | 1,9507     | 0,1754             |
|     | 70              | 0,8465     | 0,2580             | 1,9207     | 0,1879             |
| 80  | 0               | 1,9469     | 0,1314             | 1,9680     | 0,1165             |
|     | 10              | 1,6275     | 0,0930             | 1,9676     | 0,1164             |
|     | 20              | 1,4421     | 0,0937             | 1,9576     | 0,1346             |
|     | 50              | 0,9977     | 0,0802             | 1,9546     | 0,1414             |
|     | 60              | 0,8532     | 0,0753             | 1,9523     | 0,1327             |
|     | 70              | 0,8188     | 0,2351             | 1,9519     | 0,1702             |
| 100 | 0               | 2,0395     | 0,1172             | 2,0344     | 0,1211             |
|     | 10              | 1,6493     | 0,0998             | 2,0350     | 0,1262             |
|     | 20              | 1,4257     | 0,0951             | 1,9947     | 0,1168             |
|     | 50              | 0,9817     | 0,0703             | 1,9850     | 0,1122             |
|     | 60              | 0,8380     | 0,0592             | 1,9834     | 0,0800             |
|     | 70              | 0,7680     | 0,2133             | 1,9814     | 0,1508             |
| 150 | 0               | 2,0022     | 0,1235             | 2,2089     | 0,0980             |
|     | 10              | 1,6591     | 0,0855             | 2,0079     | 0,0980             |
|     | 20              | 1,4479     | 0,0735             | 2,0024     | 0,1057             |
|     | 50              | 0,9783     | 0,0500             | 2,0007     | 0,1070             |
|     | 60              | 0,8580     | 0,0698             | 1,9912     | 0,1145             |
|     | 70              | 0,7462     | 0,0796             | 1,9872     | 0,1016             |
| 200 | 0               | 2,0045     | 0,0866             | 2,0192     | 0,0759             |
|     | 10              | 1,6780     | 0,0657             | 2,0172     | 0,0781             |
|     | 20              | 1,4347     | 0,0665             | 1,9930     | 0,0924             |
|     | 50              | 0,9788     | 0,0517             | 1,9911     | 0,0876             |
|     | 60              | 0,8451     | 0,0464             | 1,9871     | 0,0935             |
|     | 70              | 0,7446     | 0,0653             | 1,9730     | 0,1013             |
| 250 | 0               | 1,9876     | 0,0792             | 1,9900     | 0,0712             |
|     | 10              | 1,6580     | 0,0641             | 1,9891     | 0,0720             |
|     | 20              | 1,4429     | 0,0639             | 1,9882     | 0,0866             |
|     | 50              | 0,9806     | 0,0438             | 1,9843     | 0,0799             |
|     | 60              | 0,8316     | 0,0453             | 1,9842     | 0,0921             |
|     | 70              | 0,7419     | 0,0557             | 1,9821     | 0,0747             |
| 300 | 0               | 1,9999     | 0,0813             | 2,0013     | 0,0650             |
|     | 10              | 1,6666     | 0,0473             | 2,0012     | 0,0646             |
|     | 20              | 1,4402     | 0,0570             | 1,9902     | 0,0722             |
|     | 50              | 0,9742     | 0,0418             | 1,9880     | 0,0739             |
|     | 60              | 0,8453     | 0,0391             | 1,9874     | 0,0768             |
|     | 70              | 0,7525     | 0,0591             | 1,9856     | 0,0666             |

Πίνακας 9: Gamma Distribution Ga(2,2)

| n   | Censoring level | $\theta=2$ |                    | $\gamma=1$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|
|     |                 | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 2,0498     | 0,2980             | 1,0278     | 0,1206             |
|     | 10              | 1,5043     | 0,1933             | 1,0250     | 0,1217             |
|     | 20              | 1,1947     | 0,1503             | 1,0247     | 0,1259             |
|     | 50              | 0,6328     | 0,1069             | 1,0228     | 0,1100             |
|     | 60              | 0,5244     | 0,1918             | 1,0074     | 0,1115             |
|     | 70              | 0,4237     | 0,1773             | 1,0055     | 0,0957             |
| 80  | 0               | 1,9980     | 0,2012             | 1,0234     | 0,0930             |
|     | 10              | 1,4699     | 0,1572             | 1,0179     | 0,0932             |
|     | 20              | 1,2062     | 0,1261             | 1,0117     | 0,0938             |
|     | 50              | 0,6197     | 0,0756             | 0,9993     | 0,0998             |
|     | 60              | 0,4871     | 0,0641             | 0,9987     | 0,0730             |
|     | 70              | 0,4622     | 0,2417             | 0,9943     | 0,0916             |
| 100 | 0               | 2,0218     | 0,1948             | 1,0234     | 0,0765             |
|     | 10              | 1,4738     | 0,1234             | 1,0098     | 0,0760             |
|     | 20              | 1,1924     | 0,1017             | 1,0015     | 0,0708             |
|     | 50              | 0,6296     | 0,0590             | 1,0010     | 0,0841             |
|     | 60              | 0,5092     | 0,0663             | 0,9936     | 0,1040             |
|     | 70              | 0,4074     | 0,1613             | 0,9859     | 0,0798             |
| 150 | 0               | 1,9751     | 0,1393             | 1,0129     | 0,0583             |
|     | 10              | 1,4712     | 0,0947             | 1,0082     | 0,0589             |
|     | 20              | 0,1200     | 0,0811             | 1,0060     | 0,0711             |
|     | 50              | 0,6284     | 0,0533             | 1,0025     | 0,0685             |
|     | 60              | 0,4885     | 0,0525             | 0,9926     | 0,0657             |
|     | 70              | 0,3911     | 0,0717             | 0,9918     | 0,0712             |
| 200 | 0               | 1,9986     | 0,1190             | 1,0051     | 0,0566             |
|     | 10              | 1,4962     | 0,0893             | 1,0043     | 0,0568             |
|     | 20              | 1,1912     | 0,0775             | 1,0034     | 0,0520             |
|     | 50              | 0,6393     | 0,0504             | 0,9994     | 0,0685             |
|     | 60              | 0,4941     | 0,0421             | 0,9812     | 0,0610             |
|     | 70              | 0,3663     | 0,0387             | 0,1006     | 0,0557             |
| 250 | 0               | 2,0083     | 0,1259             | 1,0123     | 0,0510             |
|     | 10              | 1,5032     | 0,0718             | 1,0117     | 0,0517             |
|     | 20              | 1,1987     | 0,0623             | 1,0064     | 0,0445             |
|     | 50              | 0,6352     | 0,0419             | 0,9977     | 0,0571             |
|     | 60              | 0,4942     | 0,0379             | 0,9958     | 0,0431             |
|     | 70              | 0,3668     | 0,0372             | 0,9853     | 0,0553             |
| 300 | 0               | 1,9801     | 0,1061             | 0,9995     | 0,0456             |
|     | 10              | 1,4879     | 0,0657             | 0,9990     | 0,0456             |
|     | 20              | 1,1832     | 0,0487             | 0,9980     | 0,0382             |
|     | 50              | 0,6255     | 0,0351             | 0,9972     | 0,0397             |
|     | 60              | 0,4963     | 0,0358             | 0,9966     | 0,0454             |
|     | 70              | 0,3797     | 0,0374             | 0,9955     | 0,0546             |

Πίνακας 10: Gamma Distribution Ga(2,1)

| n   | Censoring level | $\theta=3$ |                    | $\gamma=0,5$ |                    |
|-----|-----------------|------------|--------------------|--------------|--------------------|
|     |                 | Mean       | Standard Deviation | Mean         | Standard Deviation |
| 50  | 0               | 2,9600     | 0,5855             | 0,5253       | 0,0725             |
|     | 10              | 1,8832     | 0,3258             | 0,5082       | 0,0731             |
|     | 20              | 1,2375     | 0,2273             | 0,4980       | 0,0569             |
|     | 50              | 1,4729     | 2,0073             | 0,4976       | 0,0566             |
|     | 60              | 2,4386     | 3,9391             | 0,4972       | 0,0723             |
|     | 70              | 4,4238     | 6,0135             | 0,4911       | 0,0818             |
| 80  | 0               | 2,9310     | 0,4886             | 0,5146       | 0,0486             |
|     | 10              | 1,8406     | 0,2557             | 0,5110       | 0,0483             |
|     | 20              | 1,3227     | 0,1227             | 0,5046       | 0,0421             |
|     | 50              | 1,2867     | 1,0425             | 0,4989       | 0,0516             |
|     | 60              | 2,6481     | 4,5764             | 0,4933       | 0,0515             |
|     | 70              | 6,4089     | 7,1629             | 0,4933       | 0,0502             |
| 100 | 0               | 3,1392     | 0,4337             | 0,5245       | 0,0474             |
|     | 10              | 1,9423     | 0,1990             | 0,5245       | 0,0473             |
|     | 20              | 1,3073     | 0,1557             | 0,5093       | 0,0519             |
|     | 50              | 1,1584     | 1,0648             | 0,5045       | 0,0544             |
|     | 60              | 2,6395     | 4,9979             | 0,4962       | 0,0510             |
|     | 70              | 4,9822     | 6,4711             | 0,4954       | 0,0475             |
| 150 | 0               | 2,9920     | 0,3375             | 1,7094       | 0,0335             |
|     | 10              | 2,9394     | 0,3158             | 1,1455       | 1,9420             |
|     | 20              | 2,5618     | 0,2722             | 1,0867       | 0,1701             |
|     | 50              | 0,8954     | 0,0895             | 1,0534       | 0,0358             |
|     | 60              | 0,7793     | 0,8089             | 0,5014       | 0,0275             |
|     | 70              | 2,0449     | 3,8830             | 0,0380       | 0,0270             |
| 200 | 0               | 3,0180     | 0,3280             | 1,9648       | 0,0333             |
|     | 10              | 2,9717     | 0,2929             | 1,7178       | 0,2216             |
|     | 20              | 2,5796     | 0,2478             | 1,1423       | 0,1588             |
|     | 50              | 0,8825     | 0,0715             | 1,0861       | 0,0255             |
|     | 60              | 0,6747     | 0,6162             | 1,0518       | 0,0262             |
|     | 70              | 1,5456     | 3,0201             | 0,5024       | 0,0235             |
| 250 | 0               | 2,9547     | 0,2958             | 1,9658       | 0,0298             |
|     | 10              | 2,9800     | 0,2503             | 1,7086       | 0,2008             |
|     | 20              | 2,5670     | 0,2218             | 1,1345       | 0,1378             |
|     | 50              | 0,8595     | 0,0724             | 1,0816       | 0,0278             |
|     | 60              | 0,6683     | 0,6112             | 1,0503       | 0,0260             |
|     | 70              | 1,5599     | 3,1780             | 0,4954       | 0,0245             |
| 300 | 0               | 2,9780     | 0,2535             | 1,9637       | 0,0250             |
|     | 10              | 2,9788     | 0,2363             | 1,7316       | 0,1786             |
|     | 20              | 2,6067     | 0,1960             | 1,1380       | 0,1267             |
|     | 50              | 0,8637     | 0,0680             | 1,0864       | 0,0240             |
|     | 60              | 0,6117     | 0,4747             | 1,0487       | 0,0235             |
|     | 70              | 1,0596     | 2,4978             | 0,4983       | 0,0209             |

Πίνακας 11: Gamma Distribution Ga(3, 0.5)

## 5.4 Εκτιμητής Μέγιστης Πιθανοφάνειας για τυχαία λογοκριμένο δείγμα

Έστω  $t_1, t_2, \dots, t_n$  ένα τυχαία λογοκριμένο δείγμα με ποσοστό λογοκρισίας

$$p = 1 - q = \bar{F}(T).$$

Το δείγμα είναι της μορφής

$$\underbrace{t_1, t_2, \dots, t_{n_1}}_{n_1} \underbrace{c_1, c_2, \dots, c_{n-n_1}}_{n-n_1}$$

Από την (5.11) η συνάρτηση πιθανοφάνειας είναι

$$L(\theta) \propto \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(c_i; \theta)^{1-\delta_i} = \left( \prod_{i=1}^{n_1} \frac{t_i^{\gamma-1}}{\Gamma(\gamma)\theta^\gamma} e^{-\sum_{i=1}^{n_1} \frac{t_i}{\theta}} \right) \prod_{i=1}^{n-n_1} \left( \Gamma(\gamma) - \Gamma\left(\gamma, \frac{c_i}{\theta}\right) \right) \quad (5.10)$$

Λογαθμίζοντας και παίρνοντας την παράγωγο ως προς  $\theta$  και  $\gamma$  προκύπτει το σύστημα

$$\begin{cases} \frac{\partial \ln L}{\partial \gamma} = -n_1 \Psi(\gamma) - n_1 \ln \theta + \sum_{i=1}^{n_1} \ln x_i - \sum_{i=1}^{n-n_1} \frac{\frac{\partial \Gamma(\gamma, \frac{c_i}{\theta})}{\partial \gamma}}{\Gamma(\gamma) - \Gamma(\gamma, \frac{c_i}{\theta})} = 0 \\ \frac{\partial \ln L}{\partial \theta} = -\frac{n_1 \gamma}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n_1} \ln x_i - \sum_{i=1}^{n-n_1} \frac{\frac{\partial \Gamma(\gamma, \frac{c_i}{\theta})}{\partial \theta}}{\Gamma(\gamma) - \Gamma(\gamma, \frac{c_i}{\theta})} = 0 \end{cases}$$

Η πολυπλοκότητα των συναρτήσεων αυτών κάνει την επίλυσή τους ιδιαίτερα δύσκολη.

(βλ. Παράρτημα Π6 και Παράρτημα Π7)

| n   | Censoring level | $\theta=2$ |                    | $\gamma=2$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|
|     |                 | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 1,9807     | 0,1990             | 2,0114     | 0,1676             |
|     | 10              | 2,3925     | 0,3796             | 0,5687     | 0,7772             |
|     | 20              | 1,2881     | 0,3144             | 0,7403     | 0,6416             |
|     | 50              | 1,0079     | 0,1654             | 0,4898     | 1,3097             |
|     | 60              | 0,8493     | 0,1574             | 0,3724     | 1,2807             |
|     | 70              | 0,1561     | 0,1109             | 0,0264     | 1,1523             |
| 80  | 0               | 2,0087     | 0,1688             | 1,9977     | 0,1341             |
|     | 10              | 3,3809     | 0,3098             | 1,1812     | 0,2734             |
|     | 20              | 2,8025     | 0,2950             | 1,0300     | 0,5067             |
|     | 50              | 1,1240     | 0,1439             | 0,9588     | 0,7180             |
|     | 60              | 0,8343     | 0,1172             | 0,9419     | 0,2336             |
|     | 70              | 0,1464     | 0,0730             | 0,5966     | 0,0227             |
| 100 | 0               | 1,9927     | 0,1240             | 2,0207     | 0,1394             |
|     | 10              | 3,3504     | 0,2758             | 1,3170     | 0,0562             |
|     | 20              | 1,4577     | 0,2573             | 1,2373     | 0,0567             |
|     | 50              | 1,1187     | 0,1198             | 0,7587     | 0,8806             |
|     | 60              | 0,8490     | 0,1074             | 0,1875     | 0,8298             |
|     | 70              | 0,2790     | 0,0732             | 0,0470     | 0,5979             |
| 150 | 0               | 1,9989     | 0,1239             | 2,0612     | 0,1125             |
|     | 10              | 2,7805     | 0,2262             | 1,3938     | 0,0584             |
|     | 20              | 2,4018     | 0,1911             | 1,1828     | 0,0515             |
|     | 50              | 1,1429     | 0,1184             | 0,9585     | 0,0295             |
|     | 60              | 0,8414     | 0,0716             | 0,3135     | 0,5334             |
|     | 70              | 0,1485     | 0,0582             | 0,2015     | 0,8754             |
| 200 | 0               | 1,9785     | 0,1067             | 2,0013     | 0,0854             |
|     | 10              | 3,3884     | 0,1827             | 1,5656     | 0,0584             |
|     | 20              | 2,8299     | 0,1615             | 1,4619     | 0,0436             |
|     | 50              | 1,4678     | 0,0831             | 1,2438     | 0,0266             |
|     | 60              | 1,1294     | 0,0688             | 1,1857     | 0,0292             |
|     | 70              | 0,8392     | 0,0517             | 1,1500     | 0,0204             |
| 250 | 0               | 1,9964     | 0,0848             | 2,0001     | 0,0693             |
|     | 10              | 3,3915     | 0,1655             | 1,5931     | 0,0559             |
|     | 20              | 2,8217     | 0,1450             | 1,5084     | 0,0479             |
|     | 50              | 1,4778     | 0,0814             | 1,2827     | 0,0286             |
|     | 60              | 1,1383     | 0,0576             | 1,2283     | 0,0266             |
|     | 70              | 1,1244     | 0,0645             | 1,1844     | 0,0210             |
| 300 | 0               | 1,9900     | 0,0831             | 2,0044     | 0,0578             |
|     | 10              | 3,4023     | 0,1525             | 1,6558     | 0,0502             |
|     | 20              | 2,8407     | 0,1294             | 1,5377     | 0,0515             |
|     | 50              | 1,4617     | 0,0741             | 1,3069     | 0,0242             |
|     | 60              | 1,1353     | 0,0623             | 1,2514     | 0,0199             |
|     | 70              | 0,8441     | 0,0376             | 1,2119     | 0,0194             |

Πίνακας 12: Random Gamma Distribution  $Ga(2, 2)$

| n   | Censoring level | $\theta=2$ |                    | $\gamma=1$ |                    |
|-----|-----------------|------------|--------------------|------------|--------------------|
|     |                 | Mean       | Standard Deviation | Mean       | Standard Deviation |
| 50  | 0               | 2,0317     | 0,2960             | 0,9504     | 0,7550             |
|     | 10              | 1,7496     | 0,2834             | 0,2752     | 0,8150             |
|     | 20              | 1,4439     | 0,2121             | 0,3552     | 0,9550             |
|     | 50              | 0,7608     | 0,0906             | 0,4352     | 1,0824             |
|     | 60              | 0,5950     | 0,0744             | 0,5152     | 1,0157             |
|     | 70              | 0,4574     | 0,0527             | 0,5952     | 0,9949             |
| 80  | 0               | 2,0129     | 0,2356             | 0,9886     | 0,6895             |
|     | 10              | 1,7644     | 0,1930             | 0,2586     | 0,7695             |
|     | 20              | 1,4615     | 0,1363             | 0,2560     | 0,8294             |
|     | 50              | 0,7675     | 0,0728             | 0,1862     | 0,9900             |
|     | 60              | 0,6011     | 0,0552             | 0,0340     | 1,0535             |
|     | 70              | 0,4629     | 0,0422             | 0,0279     | 0,9275             |
| 100 | 0               | 2,0056     | 0,1858             | 0,9302     | 0,7206             |
|     | 10              | 1,7415     | 0,1745             | 0,3002     | 0,8106             |
|     | 20              | 1,4489     | 0,1340             | 0,0467     | 0,8280             |
|     | 50              | 0,7640     | 0,0642             | 0,0934     | 0,9711             |
|     | 60              | 0,6116     | 0,0497             | 0,1559     | 0,8870             |
|     | 70              | 0,4611     | 0,0340             | 0,1219     | 0,9487             |
| 150 | 0               | 2,0256     | 0,1616             | 0,9773     | 0,6762             |
|     | 10              | 1,4562     | 0,1202             | 0,1773     | 0,7862             |
|     | 20              | 1,4426     | 0,1116             | 0,0013     | 0,7447             |
|     | 50              | 0,7741     | 0,0525             | 0,0056     | 0,9620             |
|     | 60              | 0,5998     | 0,0338             | 0,0762     | 1,0853             |
|     | 70              | 0,4608     | 0,0308             | 0,1268     | 0,9528             |
| 200 | 0               | 0,1993     | 0,1345             | 0,4240     | 0,6538             |
|     | 10              | 1,7224     | 0,1234             | 0,4240     | 0,7838             |
|     | 20              | 1,4589     | 0,0928             | 0,3006     | 0,8413             |
|     | 50              | 0,7711     | 0,0430             | 0,1078     | 0,9622             |
|     | 60              | 0,5964     | 0,0376             | 0,0164     | 1,0151             |
|     | 70              | 0,4606     | 0,0233             | 0,2085     | 0,8686             |
| 250 | 0               | 1,9730     | 0,1418             | 0,9285     | 0,6683             |
|     | 10              | 1,7401     | 0,1149             | 0,4285     | 0,7883             |
|     | 20              | 1,4573     | 0,0830             | 0,0277     | 0,8356             |
|     | 50              | 0,7747     | 0,0396             | 0,3238     | 0,9606             |
|     | 60              | 0,6090     | 0,0299             | 0,1092     | 0,9371             |
|     | 70              | 0,4607     | 0,0226             | 0,2467     | 0,9414             |
| 300 | 0               | 1,9887     | 0,1123             | 0,9788     | 0,6404             |
|     | 10              | 1,7481     | 0,0999             | 0,4788     | 0,7304             |
|     | 20              | 1,4731     | 0,0902             | 0,1010     | 0,7748             |
|     | 50              | 0,7721     | 0,0318             | 0,0146     | 0,8991             |
|     | 60              | 0,6039     | 0,0284             | 0,0518     | 0,9574             |
|     | 70              | 0,4649     | 0,0185             | 0,1634     | 0,7812             |

Πίνακας 13: Random Gamma Distribution  $Ga(2, 1)$



| n   | Censoring level | $\theta=3$ |                    | $\gamma=0,5$ |                    |
|-----|-----------------|------------|--------------------|--------------|--------------------|
|     |                 | Mean       | Standard Deviation | Mean         | Standard Deviation |
| 50  | 0               | 0,3035     | 0,5784             | 0,0737       | 2,0272             |
|     | 10              | 1,2050     | 0,2096             | 0,0536       | 2,8510             |
|     | 20              | 0,8844     | 0,1768             | 0,8080       | 2,8287             |
|     | 50              | 0,3793     | 0,0633             | 0,4948       | 2,8812             |
|     | 60              | 0,2922     | 0,0525             | 0,3641       | 2,8684             |
|     | 70              | 0,2290     | 0,0405             | 0,2895       | 2,7524             |
| 80  | 0               | 3,0713     | 0,4977             | 0,5722       | 2,2834             |
|     | 10              | 1,2029     | 0,1935             | 0,8176       | 2,5297             |
|     | 20              | 0,9053     | 0,1156             | 0,2903       | 2,4258             |
|     | 50              | 0,3822     | 0,0494             | 0,1713       | 2,7067             |
|     | 60              | 0,2909     | 0,0349             | 0,1281       | 2,7326             |
|     | 70              | 0,2267     | 0,0318             | 0,1180       | 2,7662             |
| 100 | 0               | 3,0011     | 0,3783             | 0,1320       | 1,9572             |
|     | 10              | 1,2123     | 0,1848             | 0,5689       | 2,3453             |
|     | 20              | 0,9402     | 0,1119             | 0,4159       | 2,6698             |
|     | 50              | 0,3829     | 0,0495             | 0,1520       | 0,2607             |
|     | 60              | 0,2955     | 0,0331             | 0,1482       | 2,7153             |
|     | 70              | 0,2262     | 0,0260             | 0,1399       | 2,3118             |
| 150 | 0               | 2,9302     | 0,3264             | 0,5725       | 1,9001             |
|     | 10              | 1,1908     | 0,1393             | 0,5714       | 0,2191             |
|     | 20              | 0,8952     | 0,0868             | 0,3807       | 2,3825             |
|     | 50              | 0,3875     | 0,0330             | 0,2793       | 2,2496             |
|     | 60              | 0,2867     | 0,0296             | 0,2241       | 2,3119             |
|     | 70              | 0,2314     | 0,0234             | 0,1544       | 2,3455             |
| 200 | 0               | 3,0108     | 0,3090             | 0,1277       | 2,0294             |
|     | 10              | 1,1742     | 0,1089             | 0,7593       | 2,4518             |
|     | 20              | 0,9108     | 0,0886             | 0,6961       | 0,2128             |
|     | 50              | 0,3774     | 0,0325             | 0,3219       | 2,4435             |
|     | 60              | 0,2952     | 0,0219             | 0,2912       | 0,2043             |
|     | 70              | 0,2227     | 0,0220             | 0,1201       | 2,1952             |
| 250 | 0               | 0,9895     | 0,2646             | 0,1803       | 2,1756             |
|     | 10              | 1,1953     | 0,1010             | 0,1654       | 1,9239             |
|     | 20              | 0,9074     | 0,0669             | 0,1572       | 2,0026             |
|     | 50              | 0,3826     | 0,0283             | 0,1402       | 1,8252             |
|     | 60              | 0,2885     | 0,0204             | 0,1401       | 2,0185             |
|     | 70              | 0,2276     | 0,0146             | 0,1151       | 1,7899             |
| 300 | 0               | 2,9742     | 0,2011             | 0,1957       | 2,3519             |
|     | 10              | 1,1920     | 0,0954             | 0,1783       | 2,2019             |
|     | 20              | 0,9016     | 0,0691             | 0,1657       | 2,2393             |
|     | 50              | 0,3801     | 0,0246             | 0,1253       | 0,1936             |
|     | 60              | 0,2916     | 0,0178             | 0,1191       | 1,7524             |
|     | 70              | 0,2305     | 0,0162             | 0,1052       | 1,7816             |

Πίνακας 14: Random Gamma Distribution Ga(3, 0.5)

## 5.5 Σχέση μεταξύ $p$ και $L$

Από την (2.26) έχουμε ότι

$$\begin{aligned}
 1-p &= \frac{1}{L} \int_0^L \int_0^y f(x; \theta) dx dy \\
 &= \frac{1}{L} \int_0^L \int_0^y \frac{x^{\gamma-1}}{\Gamma(\gamma)\theta^\gamma} e^{-\frac{x}{\theta}} dx dy \\
 &= \frac{1}{L} \int_0^{\frac{L}{\theta}} \frac{w^{\gamma-1}}{\Gamma(\gamma)} e^{-w} dw dy \\
 &= \frac{1}{L/\theta} \left[ \frac{L}{\theta} - \int_0^{\frac{L}{\theta}} \frac{\Gamma(\gamma, y)}{\Gamma(\gamma)} dy \right].
 \end{aligned} \tag{5.8}$$

Έτσι προκύπτει η σχέση

$$p = \frac{1}{L/\theta} \int_0^{\frac{L}{\theta}} \frac{\Gamma(\gamma, y)}{\Gamma(\gamma)} dy. \tag{5.9}$$

Δυστυχώς η παράσταση αυτή είναι δύσκολο να διαχειρισθεί από τα διάφορα υπολογιστικά πακέτα. Για τον λόγο αυτό η (5.8) με αλλαγή της σειράς των ολοκληρωμάτων μπορεί να γραφτεί ως

$$\begin{aligned}
 1-p &= \frac{1}{L} \int_0^L \int_0^y f(x; \theta) dx dy \\
 &= \frac{1}{L} \int_0^L \int_0^{L-x} f(x; \theta) dy dx \\
 &= \frac{1}{L} \int_0^L (L-x) f(x; \theta) dx.
 \end{aligned}$$

Αντικαθιστώντας την πυκνότητα πιθανότητα της Gamma κατανομής και με παραγοντική ολοκλήρωση προκύπτει

$$1-p = \left(\frac{L}{\theta}\right)^{\gamma-1} e^{-L/\theta} \frac{1}{\Gamma(\gamma)} + (1-\gamma \frac{\theta}{L}) \left(1 - \frac{\Gamma(\gamma, L/\theta)}{\Gamma(\gamma)}\right)$$

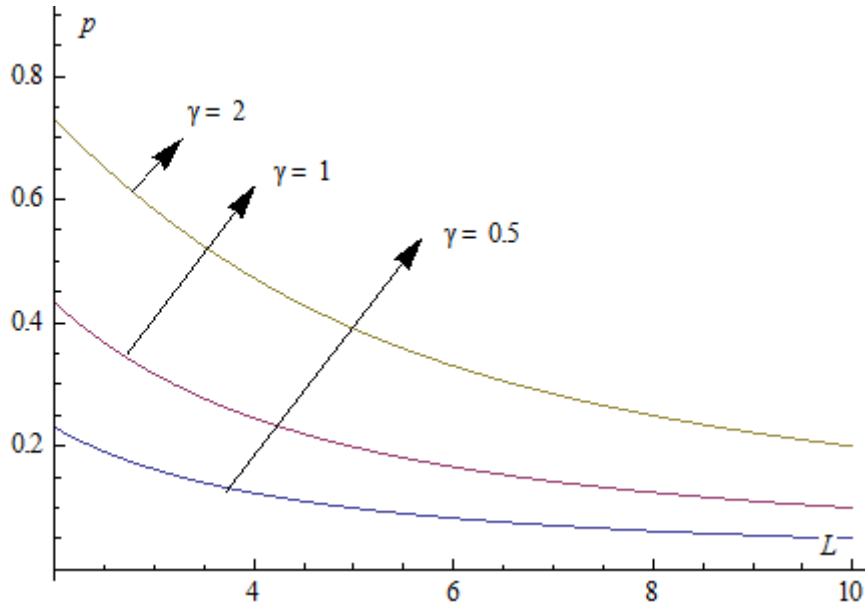
Δηλαδή

$$p = 1 - \left(\frac{L}{\theta}\right)^{\gamma-1} e^{-L/\theta} \frac{1}{\Gamma(\gamma)} - (1-\gamma \frac{\theta}{L}) \left(1 - \frac{\Gamma(\gamma, L/\theta)}{\Gamma(\gamma)}\right) \tag{5.10}$$

Η μορφή αυτή είναι πιο εύχρηστη γιατί δεν περιλαμβάνει ολοκληρώματα.

Από την παραπάνω σχέση παρατηρείται ότι το ποσοστό  $p$  εξαρτάται από την παράμετρο μορφής  $\gamma$  και το λόγο  $L/\theta$ . Για το λόγο αυτό θα θεωρηθεί  $\theta=1$  και θα βρεθεί το  $L$  για διάφορες τιμές του  $\gamma$ . Αν  $\theta \neq 1$  τότε το  $L$  είναι το  $L$  που βρέθηκε για  $\theta=1$  πολλαπλασιασμένο με το  $\theta$ .

Ο παρακάτω πίνακας δίνει κάποιες τιμές του  $L/\theta=0.1(0.1) 0.9$  για  $\gamma=0.5, 1, 2$ .



Διάγραμμα 18: Σχέση μεταξύ  $p$  και  $L$

Σχετικά με τον πληροφοριακό αριθμό

**Για την  $\theta$**

Από την σχέση (3.13) έχουμε

$$I^{T,C}(\theta) = \underbrace{\int \left(\frac{\partial}{\partial \theta} \log f\right)^2 f \bar{G} dx}_I + \underbrace{\int \left(\frac{\partial}{\partial \theta} \log \bar{F}\right)^2 g \bar{F} dx}_{II} \quad (5.11)$$

Αντικαθιστώντας τις αντίστοιχες ποσότητες προκύπτει:

Για το πρώτο ολοκλήρωμα προκύπτει

$$I = \frac{1}{\Gamma(\gamma)\theta^2} \int_0^{L/\theta} (x-\gamma)^2 \left(1 - \frac{x}{L/\theta}\right) x^{\gamma-1} e^{-x} dx$$

Το ολοκλήρωμα αυτό σε σχέση με τη συνάρτηση Gamma γράφεται

$$I = \frac{1}{\Gamma(\gamma)\theta^2} \left( \left(\frac{L}{\theta}\right)^\gamma e^{-L/\theta} + (\gamma+2)\left(\frac{L}{\theta}\right)^{\gamma-1} e^{-L/\theta} + \gamma \left(1 - \frac{(\gamma+2)}{L/\theta}\right) (\Gamma(\gamma) - \Gamma(\gamma, \frac{L}{\theta})) \right)$$

Το δεύτερο ολοκλήρωμα είναι

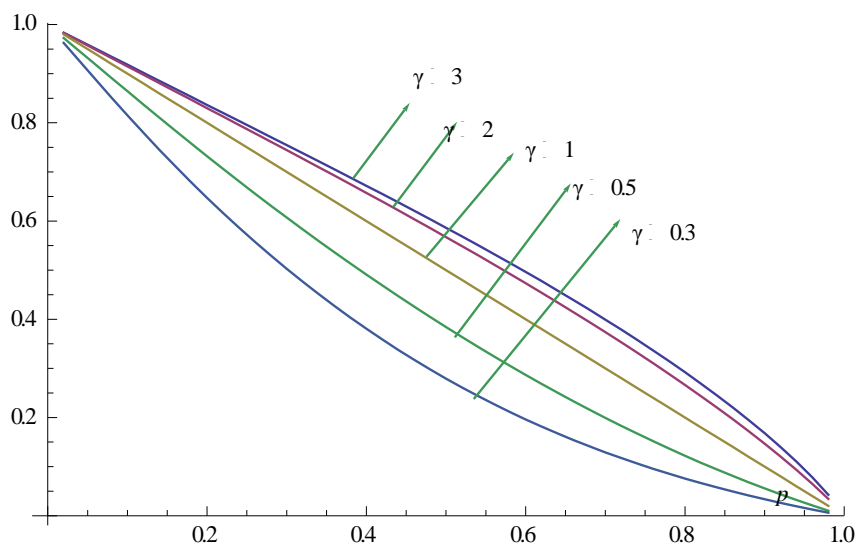
$$II = \frac{1}{L} \int_0^L \left( \frac{\partial}{\partial \theta} \log \bar{F} \right)^2 \bar{F} dx.$$

Αντικαθιστώντας τις απαραίτητες ποσότητες και παραγωγίζοντας προκύπτει η παράσταση

$$II = \frac{1}{\theta^2} \frac{1}{L/\theta} \int_0^{L/\theta} \frac{y^{2\gamma} e^{-2y}}{G(\gamma, y)} dy.$$

Έτσι, ο λόγος των πληροφοριών για το λογοκριμένο δείγμα προς το μη-λογοκριμένο δείγμα είναι

$$\frac{I_p(\theta)}{I(\theta)} = \frac{\left(\frac{L}{\theta}\right)^\gamma e^{-L/\theta} + (\gamma+2)\left(\frac{L}{\theta}\right)^{\gamma-1} e^{-L/\theta} + \frac{\gamma}{L/\theta} (1 - (\gamma+2))(\Gamma(\gamma) - \Gamma(\gamma, \frac{L}{\theta})) + \frac{1}{L/\theta} \int_0^{L/\theta} \frac{y^{2\gamma} e^{-2y}}{G(\gamma, y)} dy}{\Gamma(\gamma)\gamma}$$



Διάγραμμα 19: Σχέση μεταξύ  $p$  και  $L$

Παρατηρείται ότι για  $\gamma=1$  έχουμε την περίπτωση της εκθετικής κατανομής

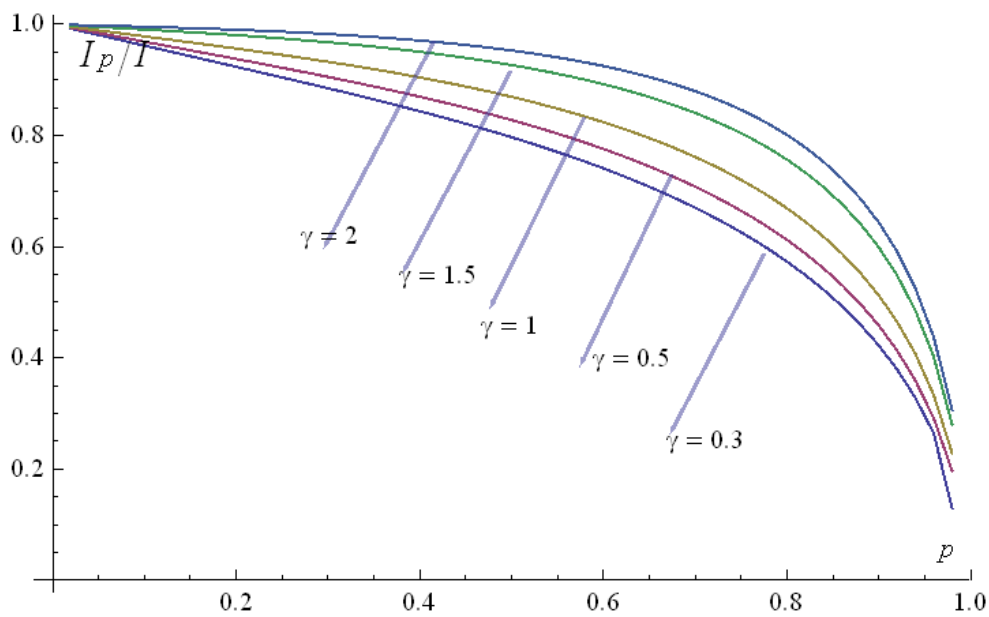
**Για την περίπτωση κατά την οποία η άγνωστη παράμετρος είναι η  $\gamma$**

Από την (4.13) ο πληροφοριακός αριθμός είναι

$$I_p(\gamma) = \int_0^L \left( \log \frac{x}{\theta} - \Psi(\gamma) \right)^2 \left( 1 - \frac{x}{L} \right) \frac{x^{\gamma-1}}{\theta^\gamma} \frac{1}{\Gamma(\gamma)} e^{-\frac{x}{\theta}} dx + \frac{\theta}{L} \int_0^{L/\theta} \frac{\left( \int_y^\infty w^{\gamma-1} \log(w) e^{-w} dw \right)^2}{\Gamma(\gamma, y)} dy$$

Μετά από στοιχειώδεις πράξεις στο πρώτο ολοκλήρωμα προκύπτει

$$I_p(\gamma) = \int_0^{L/\theta} (\log w - \Psi(\gamma))^2 \left( 1 - \frac{w}{L/\theta} \right) w^{\gamma-1} \frac{1}{\Gamma(\gamma)} e^{-w} dw + \frac{1}{\Gamma(\gamma)} \frac{L}{\theta} \int_0^{L/\theta} \frac{\left( \int_y^\infty w^{\gamma-1} (\log(w) - \Psi(\gamma)) e^{-w} dw \right)^2}{\Gamma(\gamma, y)} dy$$

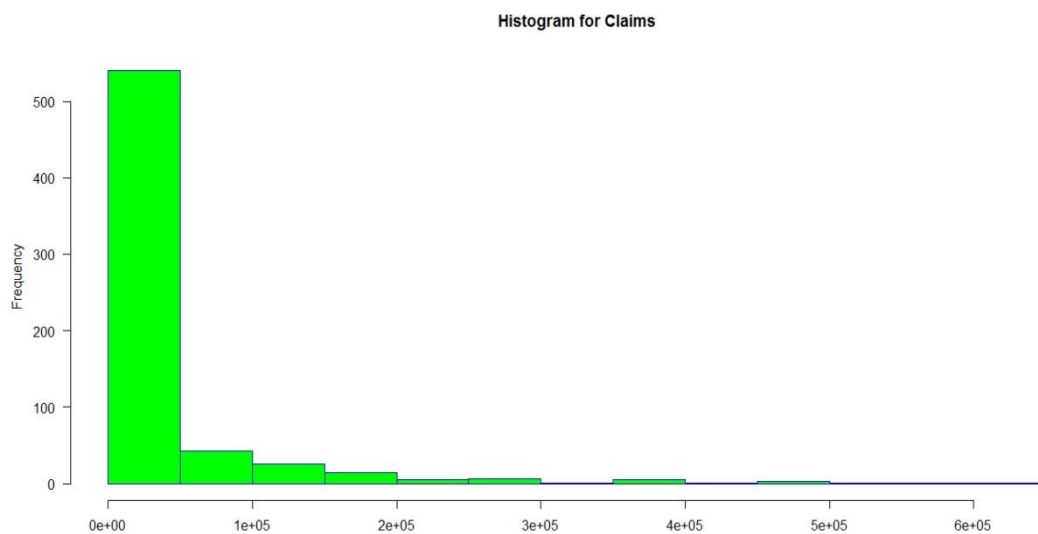


Διάγραμμα 20: Σχέση μεταξύ  $p$  και  $L$

## ΚΕΦΑΛΑΙΟ 6

### 6.1 Ανάλυση Δεδομένων

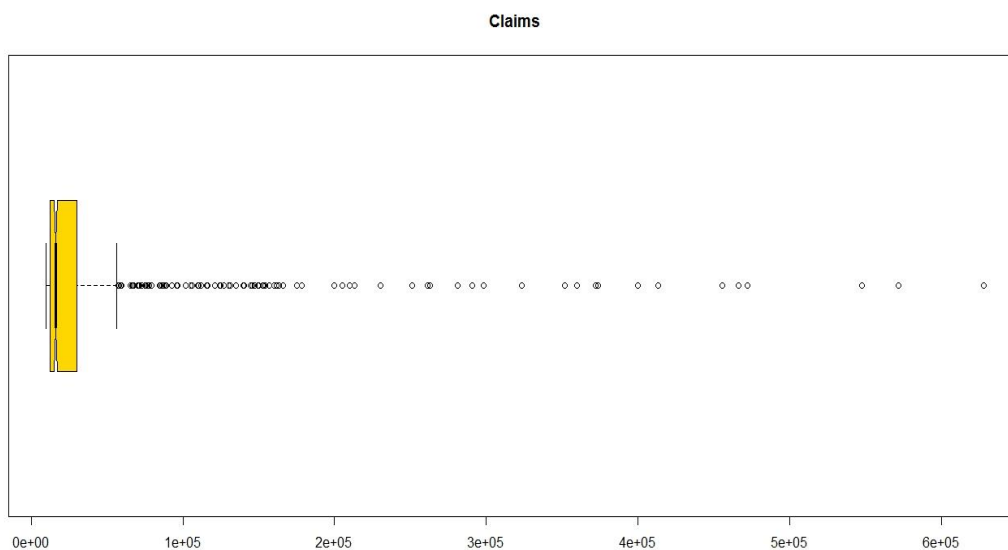
Το δείγμα που χρησιμοποιήθηκε για την ανάλυση είναι ζημιές (claims) ασφαλιστικής εταιρίας για την περίοδο 01/03/2014- 01/04/2014. Το σύνολο των στοιχείων είναι 647 καταγραφές. Πρώτο βήμα για την ανάλυση των δεδομένων είναι η γραφική απεικόνισή τους. Για το λόγο αυτό παρουσιάζεται αρχικά ένα ιστόγραμμα συχνοτήτων των δεδομένων όπως φαίνεται ακολούθως:



Εικόνα 1: Ιστόγραμμα συχνοτήτων

Η πρώτη εικόνα που μπορεί να λάβει κανείς από το παραπάνω σχήμα είναι η ύπαρξη αρκετών τιμών στην αρχή αλλά και η ύπαρξη αρκετών τιμών δεξιά. Αυτό οδηγεί στο συμπέρασμα ότι πρόκειται για δεδομένα όπου η κατανομή που θα προσαρμοστεί θα πρέπει να έχει δεξιά βαριά ουρά.

Στο ίδιο συμπέρασμα καταλήγει κανείς αν παρατηρήσει προσεκτικά το θηκόγραμμα των δεδομένων.



Εικόνα 2: Θηκόγραμμα δεδομένων

Για να γίνει ακριβέστερη μελέτη των δεδομένων θα παρουσιαστούν στην συνέχεια κάποια περιγραφικά στατιστικά για τα δεδομένα.

| Ελάχιστη τιμή | Πρώτο τεταρτημόριο | Διάμεσος | Μέση τιμή | Πρώτο τεταρτημόριο | Μέγιστη τιμή |
|---------------|--------------------|----------|-----------|--------------------|--------------|
| 10000         | 12320              | 16120    | 40910     | 30090              | 627900       |

Πίνακας 15: Περιγραφικά στατιστικά 1

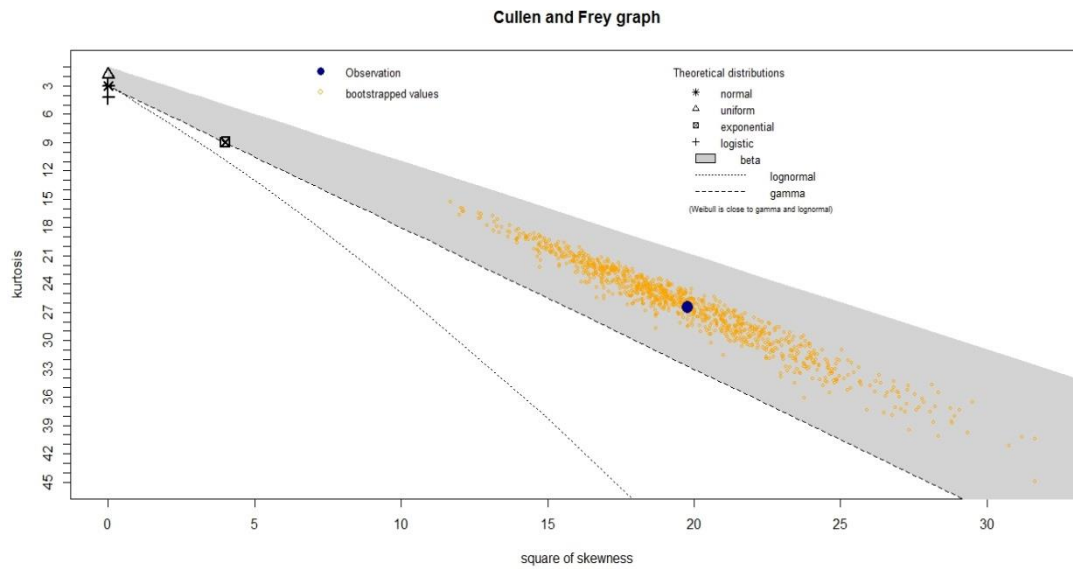
Επίσης, χρησιμοποιώντας την μεθοδολογία Bootstrap προέκυψαν τα ακόλουθα περιγραφικά μεγέθη:

**Εκτιμηθείσα λοξότητα: 4.446845**

**Εκτιμηθείσα κύρτωση: 26.43185**

Πίνακας 16: Περιγραφικά στατιστικά 2

Χρησιμοποιώντας τις τιμές της λοξότητας και κύρτωσης, δημιουργείται ένα γράφημα με τις βασικές κατανομές και την προσαρμογή τους στα δεδομένα:



**Εικόνα 3: Γράφημα λοξότητας- κύρτωσης**

Από το εν λόγω γράφημα, είναι φανερό ότι τα δεδομένα δεν μπορούν να περιγραφούν από κάποια συνήθη κατανομή. Παρόλα αυτά θα γίνει προσαρμογή των δεδομένων στις ακόλουθες τέσσερις κατανομές και θα δοθούν τα αντίστοιχα στατιστικά αποτελέσματα:

- Η 2-παραμετρική Weibull
- η 2-παραμετρική LogNormal
- η κατανομή Pareto
- η 2-παραμετρική Gamma



## Κατανομή Weibull

Στο πρώτο βήμα παρουσιάζονται οι εκτιμητές μεγίστης πιθανοφάνειας για τις δύο παραμέτρους της κατανομής όπως φαίνονται στον παρακάτω πίνακα:

|                            | Εκτίμηση    | Τυπικό σφάλμα |
|----------------------------|-------------|---------------|
| <b>Παράμετρος μορφής</b>   | 0.88098     | 0.02255       |
| <b>Παράμετρος κλίμακας</b> | 37476.94757 | 1558.92525    |

Πίνακας 17: Εκτιμήσεις παραμέτρων στην κατανομή Weibull

Για την σύγκριση των μοντέλων, βασικό ρόλο παίζουν κάποιες στατιστικές ποσότητες τις οποίες θα παρουσιάζονται σε κάθε μοντέλο. Αυτές είναι η τιμή του λογαρίθμου της συνάρτησης πιθανοφάνειας και οι τιμές των κριτηρίων AIC και BIC. Για το μοντέλο της κατανομής Weibull οι τιμές αυτών των ποσοτήτων είναι οι ακόλουθες:

|                       |                  |
|-----------------------|------------------|
| <b>Loglikelihood:</b> | <b>-7492.964</b> |
| <b>AIC:</b>           | 14989.93         |
| <b>BIC:</b>           | 14998.87         |

Πίνακας 18: Τιμές στατιστικών ποσοτήτων για την Weibull

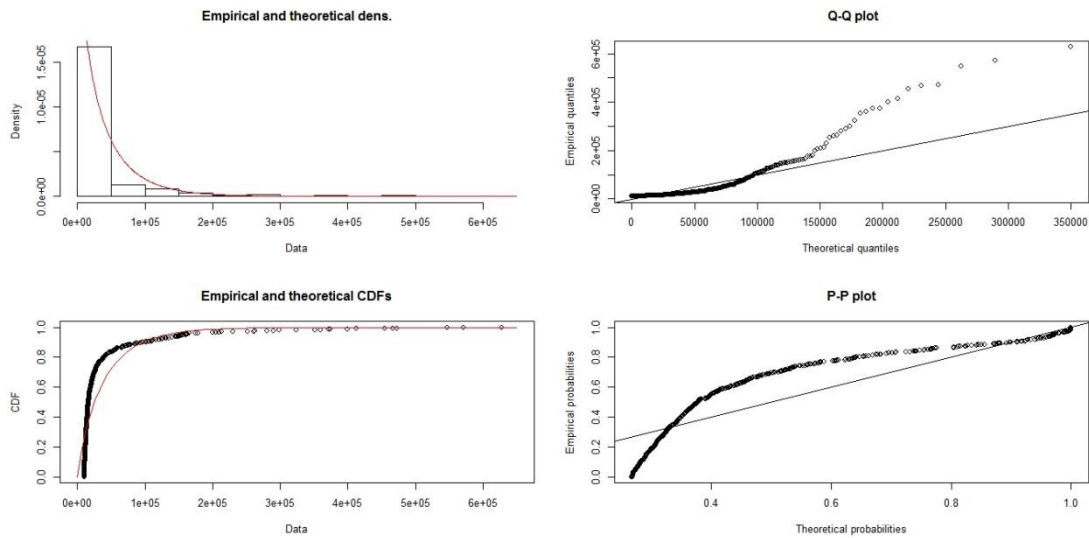
Ενώ ο πίνακας συσχετίσεων των παραμέτρων είναι ο ακόλουθος:

|              |                 |                 |
|--------------|-----------------|-----------------|
|              | <b>shape</b>    | <b>scale</b>    |
| <b>shape</b> | <b>1.000000</b> | <b>0.308455</b> |
| <b>scale</b> | <b>0.308455</b> | <b>1.000000</b> |

Πίνακας 19: Συσχετίσεις των παραμέτρων στην κατανομή Weibull

Όπως φαίνεται και στο σχήμα που ακολουθεί, η κατανομή Weibull δεν θα ήταν και η καλύτερη επιλογή για την προσαρμογή των δεδομένων. Το QQ-Plot όπως και το PP-Plot δείχνουν ξεκάθαρα την καθόλου καλή προσαρμογή των δεδομένων σε αυτή την κατανομή. Στο ίδιο συμπέρασμα μπορεί να καταλήξει κανείς παρατηρώντας το γράφημα με την εμπειρική συνάρτηση κατανομής αλλά και την προσαρμοσμένη σε αυτά Weibull κατανομή

βάσει των εκτιμήσεων των παραμέτρων που προέκυψαν με την μέθοδο μεγίστης πιθανοφάνειας .



Εικόνα 4: Γραφήματα για την προσαρμογή στην κατανομή Weibull

### Κατανομή LogNormal

Εν συνεχεία γίνεται η προσαρμογή των δεδομένων στην κατανομή LogNormal. Οι εκτιμήσεις των παραμέτρων που προκύπτουν μετά την προσαρμογή της κατανομής είναι οι ακόλουθες:

I

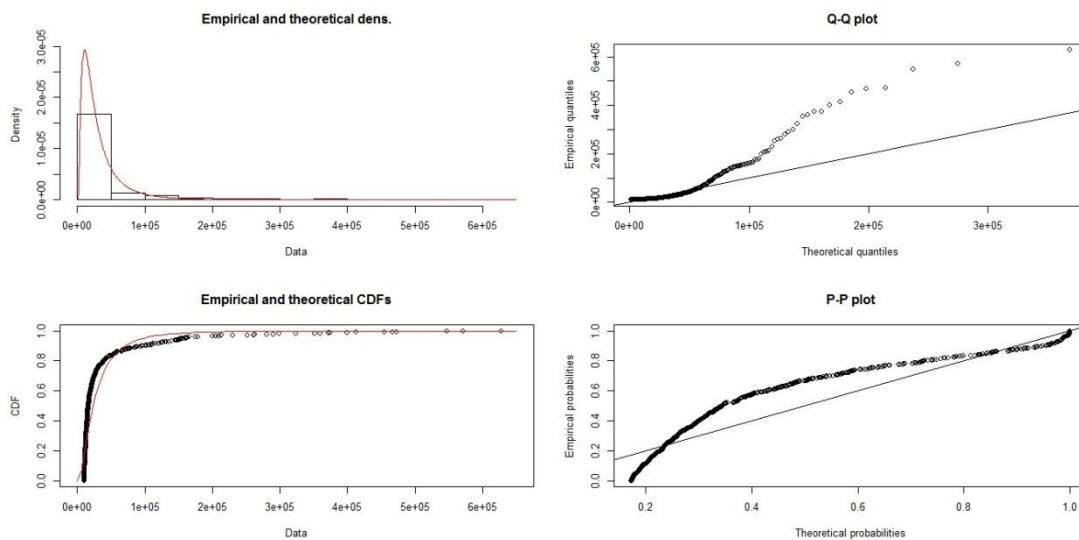
|                        | Εκτίμηση    | Τυπικό σφάλμα |
|------------------------|-------------|---------------|
| <b>Μέση τιμή</b>       | 10.03554329 | 0.03458193    |
| <b>Τυπική απόκλιση</b> | 0.87895269  | 0.02445312    |

Πίνακας 20: Εκτιμήσεις παραμέτρων στην κατανομή LogNormal

Η τιμή του λογαρίθμου της συνάρτησης πιθανοφάνειας και οι τιμές των κριτηρίων AIC και BIC για το μοντέλο της κατανομής LogNormal είναι οι ακόλουθες:

|                       |                  |
|-----------------------|------------------|
| <b>Loglikelihood:</b> | <b>-7316.246</b> |
| <b>AIC:</b>           | 14636.49         |
| <b>BIC:</b>           | 14645.43         |

Πίνακας 21: Τιμές στατιστικών ποσοτήτων για την LogNormal



Εικόνα 5: Γραφήματα για την προσαρμογή στην κατανομή LogNormal

Η επιλογή της κατανομής LogNormal όπως φαίνεται και από τις εκτιμήσεις των στατιστικών ποσοτήτων αλλά και από τα γραφήματα, δεν είναι επίσης μία καλή επιλογή.

### Κατανομή Pareto

Η κατανομή αυτή έχει αρχικά συνάρτηση πιθανότητας πυκνότητας την παρακάτω :

$$f(x|a, b) = \frac{ba^b}{x^{b+1}}, x \geq a > 0, b > 0$$

Έστω τώρα ότι έχουμε ένα τυχαίο δείγμα  $X_1, X_2, \dots, X_n$  από την παραπάνω κατανομή, μεγέθους  $n$  και έστω επίσης ορίζουμε τα ακόλουθα:

$$X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$$

$$T = \left( \prod_{i=1}^n X_i \right)^{\frac{1}{n}}$$

Θα έχουμε λοιπόν τα παρακάτω :

$$\hat{a} = X_{(1)}$$

$$\hat{b} = \frac{1}{Ln\left(\frac{T}{\hat{a}}\right)}$$

Από τα παραπάνω προκύπτει ότι οι εκτιμητές των παραμέτρων στη κατανομή Pareto είναι οι ακόλουθες:

| Εκτίμηση |          |
|----------|----------|
| <b>a</b> | 1        |
| <b>b</b> | 1.211823 |

Πίνακας 22: Εκτιμήσεις παραμέτρων στην κατανομή Pareto

Να παρατηρήσουμε ότι η εκτίμηση των παραμέτρων έγινε χρησιμοποιώντας τα δεδομένα διαιρώντας τα με το 10.000.

Το εν λόγω μοντέλο θα χρησιμοποιηθεί στην συνέχεια για την μοντελοποίηση των ζημιών αυτού του χαρτοφυλακίου.

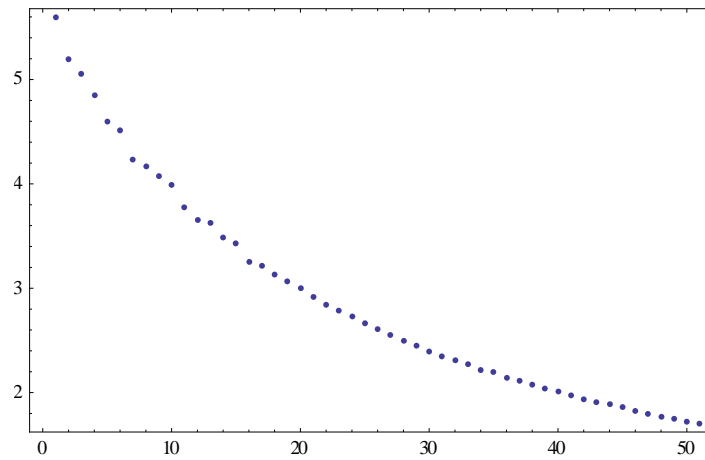
Επίσης, να σημειωθεί ότι έγινε έλεγχος καλής προσαρμογής των δεδομένων στην κατανομή Pareto με τις εκτιμήσεις που δίνει ο Πίνακας 22 και προέκυψε τιμή  $p\text{-value} = 0,0134$ , η οποία οριακά σε επίπεδο σημαντικότητας 1% οδηγεί στην μη απόρριψη της υπόθεσης ότι τα δεδομένα προέρχονται από την εν λόγω κατανομή.

### Περιπτώσεις λογοκρισίας των δεδομένων

Στην ενότητα αυτή, θα γίνει είναι μελέτη των δεδομένων βάσει της υπόθεσης ότι τα δεδομένα που μελετώνται προέρχονται από ένα πληθυσμό από την κατανομή Pareto(1, 1.211823). Αυτό που θα μελετηθεί είναι πόσο επηρεάζονται κάποια περιγραφικά χαρακτηριστικά στην περίπτωση που λογοκριθεί ένα ποσοστό από τα δεδομένα.

Αρχικά παρουσιάζεται ένα γράφημα το οποίο δείχνει την εκτίμηση της μέσης τιμής στην περίπτωση που θεωρήσουμε ότι υπάρχει λογοκρισία στα δεδομένα, η οποία βασίζεται σε ένα ποσοστό  $p$  το οποίο αλλάζει συνεχώς. Συγκεκριμένα, για διαφορετικές τιμές της παραμέτρου  $p$  υπολογίζεται η μέση τιμή. Είναι λογικό το γεγονός ότι στην περίπτωση αυτή η μέση τιμή θα μειώνεται συνεχώς καθώς μεγαλύτερες τιμές του ποσοστού λογοκρισίας  $p$  οδηγούν ταυτόχρονα σε μικρότερο όριο για το οποίο πάνω από αυτό έχουμε λογοκριμένα δεδομένα. Για τιμές λογοκρισίας  $p$  από 0 μέχρι και 0.5, το γράφημα με τις εκτιμήσεις της μέσης τιμής είναι αυτό που ακολουθεί. Να τονιστεί ότι η περίπτωση λογοκρισίας με ποσοστό  $p=0$  στην πραγματικότητα αντιστοιχεί στην περίπτωση που τα δεδομένα δεν λογοκρίνονται καθόλου

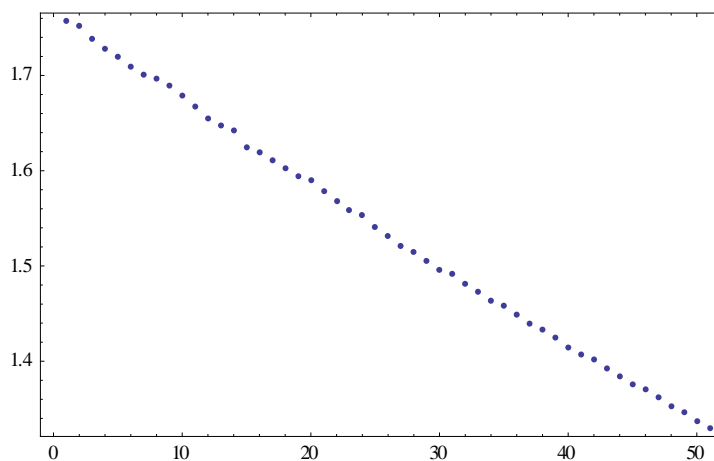
φυσικά και η μέση τιμή αντιστοιχεί στην αληθινή εκτίμηση της μέσης τιμής των δεδομένων (σε δεκάδες χιλιάδες).



Γράφημα 1: Εκτίμηση μέσης τιμής στην περίπτωση που υπάρχει λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)

Η εκτίμηση για τη μέση τιμή φαντάζει αρκετά πιο ομαλή σε περιπτώσεις που το ποσοστό λογοκρισίας αυξάνεται αρκετά.

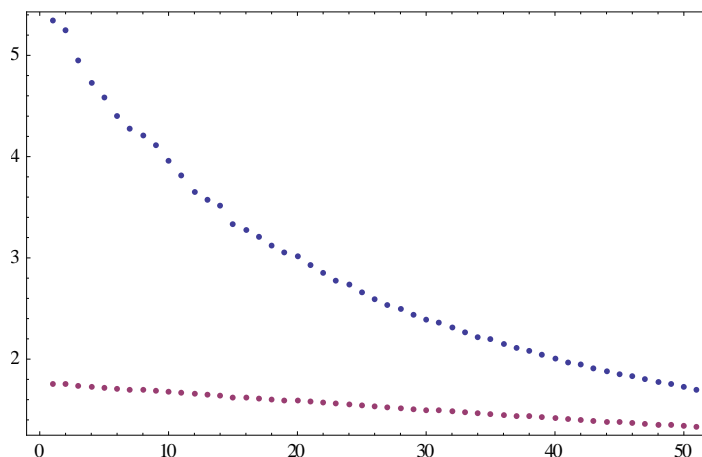
Επίσης, επειδή η κατανομή Pareto δεν είναι συμμετρική, κατασκευάζεται και το αντίστοιχο γράφημα που παρουσιάζει τις τιμές της διαμέσου της εν λόγω κατανομής για τα αντίστοιχα ποσοστά λογοκρισίας.



Γράφημα 2: Εκτίμηση διαμέσου στην περίπτωση που υπάρχει λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)

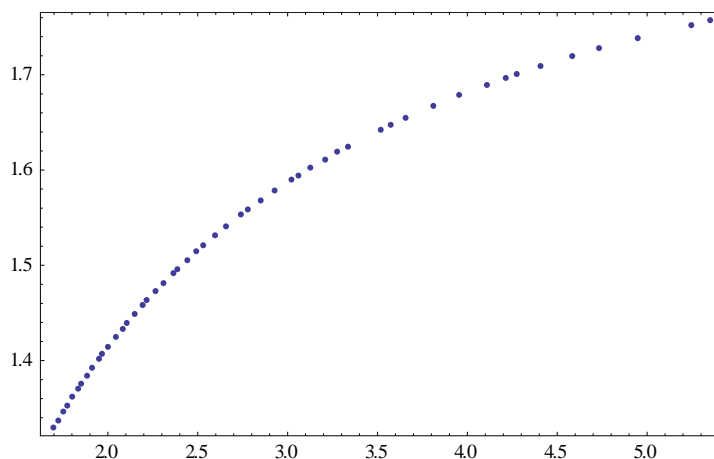
Από το Γράφημα 2, παρατηρείται ότι η διάμεσος επηρεάζεται γραμμικά από το ποσοστό λογοκρισίας σε αντίθεση με τη μέση τιμή όπου παρατηρείται μία καμπύλη. Αυτό πιθανόν και

να είναι αποτέλεσμα του ότι η διάμεσος είναι ένα μέτρο θέσης το οποίο επηρεάζεται λιγότερο από τις ακραίες παρατηρήσεις. Αυτό φαίνεται στο Γράφημα 3 όπου παρουσιάζονται και οι δύο εκτιμήσεις μαζί.



Γράφημα 3: Εκτίμηση μέσης τιμής, διαμέσου στην περίπτωση που υπάρχει λογοκρίσια (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρίσιας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)

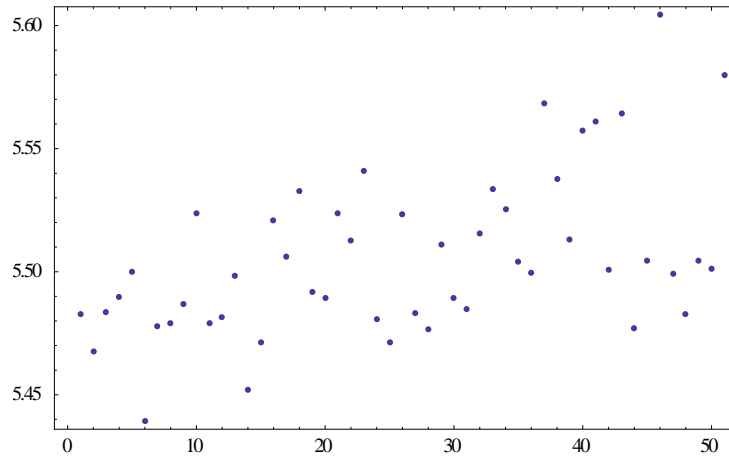
Κάνοντας το διάγραμμα διασποράς των δύο εκτιμήσεων (μέσης τιμής και διαμέσου) φαίνεται όντως ότι οι δύο ποσότητες δεν επηρεάζονται γραμμικά (δες Γράφημα 4).



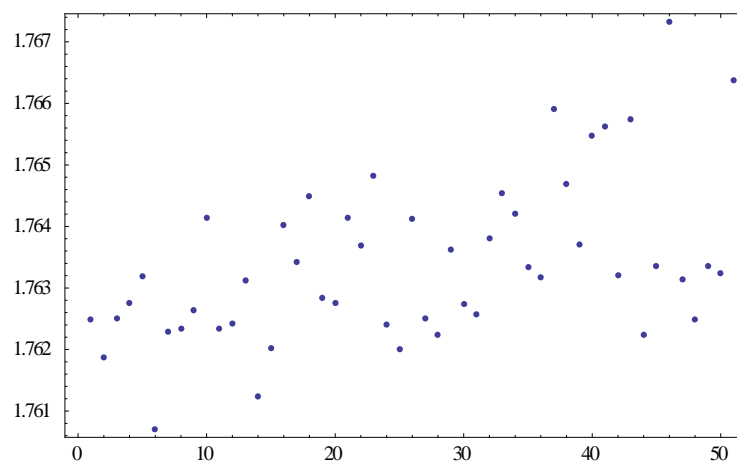
Γράφημα 4: Διάγραμμα διασποράς μέσης τιμής και διαμέσου καθώς το ποσοστό λογοκρίσιας αλλάζει

Εν συνεχεία θα παρουσιαστεί η περίπτωση που χρησιμοποιηθεί τυχαία λογοκρίσια βάσει της ομοιόμορφης κατανομής. Στην περίπτωση αυτή, υπάρχει αρκετά μεγάλη διαφοροποίηση. Όπως φαίνεται από το Γράφημα 5 φαίνεται ξεκάθαρα ότι δεν υπάρχει κάποια γραμμική ή καμπυλοειδής σχέση που προκύπτει για τα διαφορετικά ποσοστά λογοκρίσιας. Στο ίδιο συμπέρασμα καταλήγουμε αν γίνει το αντίστοιχο γράφημα με την εκτίμηση της διαμέσου

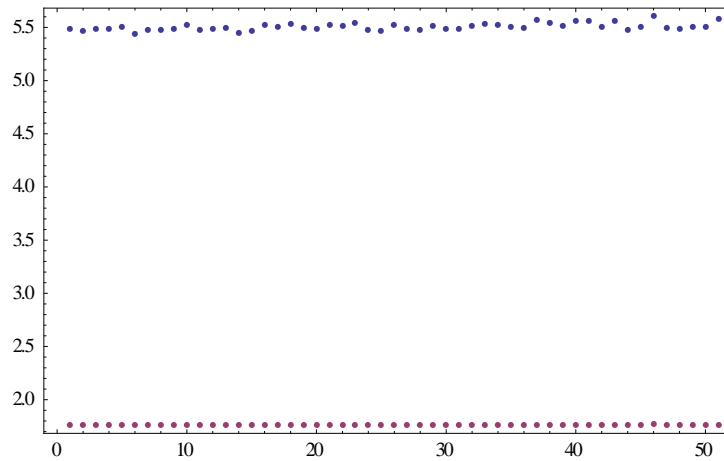
(δες Γράφημα 6). Αυτό στην πραγματικότητα σημαίνει ότι η εκτιμήσεις είναι σταθερές όπως φαίνεται καλύτερα και από το Γράφημα 7.



Γράφημα 5: Εκτίμηση μέσης τιμής στην περίπτωση που υπάρχει τυχαία λογοκρίσιμια (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρίσιμης και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)

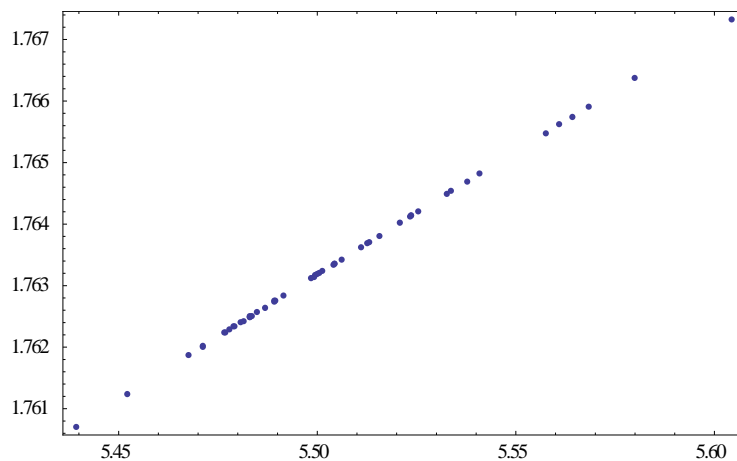


Γράφημα 6: Εκτίμηση διαμέσου στην περίπτωση που υπάρχει τυχαία λογοκρίσιμια (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρίσιμης και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)



Γράφημα 7: Εκτίμηση μέσης τιμής και διαμέσου στην περίπτωση που υπάρχει τυχαία λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)

Δημιουργώντας το διάγραμμα διασποράς για τις μέσες τιμές και διακυμάνσεις προκύπτει ότι η μέση τιμή και η διάμεσος συσχετίζονται γραμμικά όπως φαίνεται στο Γράφημα 8.



Γράφημα 8: Διάγραμμα διασποράς μέσης τιμής και διαμέσου στην περίπτωση που υπάρχει τυχαία λογοκρισία (στον οριζόντιο άξονα αντιστοιχεί το ποσοστό λογοκρισίας και στον κάθετο άξονα η εκτίμηση της μέσης τιμής σε δεκάδες χιλιάδες)



## Γιατί όχι η κατανομή Gamma;

Μία από τις αρκετά σημαντικές κατανομές που χρησιμοποιούνται για την μελέτη ζημιών είναι και η κατανομή Gamma. Η εν λόγω κατανομή στην πραγματικότητα μελετά ζημιές οι οποίες μπορούν να λάβουν τιμές στο σύνολο των θετικών πραγματικών αριθμών δίνοντας μεγάλη μάζα στην στις πρώτες τιμές αλλά ταυτόχρονα θεωρείται από τις κατανομές με βαριά δεξιά ουρά.

Είναι γνωστό όμως ότι η κατανομή Gamma δεν έχει τόσο βαριά ουρά όσο έχουν για παράδειγμα η κατανομή Pareto ή η κατανομή LogNormal. Αυτό αποτελεί το πρώτο βασικό μειονέκτημα της κατανομής αυτής στην περίπτωση που τα δεδομένα που μελετώνται έχουν μεγάλες τιμές (θεωρητικά ακραίες τιμές). Παρόλα αυτά, όπως παρουσιάστηκε και από το θηκόγραμμα των δεδομένων που μελετώνται, εμφανίζονται αρκετές παρατηρήσεις οι οποίες θεωρούνται ακραίες κάτι το οποίο είναι λογικό διότι τα δεδομένα αποτελούν αποζημιώσεις μίας ασφαλιστικής εταιρίας. Βιβλιογραφικά, δεδομένα με ακραίες τιμές είναι πολύ δύσκολο να μοντελοποιηθούν από την κατανομή Gamma.

Επίσης, ένα επιπλέον μειονέκτημα της κατανομής Gamma είναι ότι έχει πεδίο ορισμού το σύνολο όλων των θετικών αριθμών. Τα δεδομένα που μελετώνται εδώ, όπως παρουσιάστηκε και στην βασική ανάλυση, έχουν ελάχιστη τιμή το ποσό των 10.000 και αυτό φυσικά αποτελεί τροχοπέδη για την εκτίμηση των παραμέτρων της κατανομής Gamma χρησιμοποιώντας την μέθοδο της μεγίστης πιθανοφάνειας. Συγκεκριμένα, μετά από τη χρήση του αλγορίθμου Newton Raphson, παρατηρήθηκε η αδυναμία σύγκλισης του αλγορίθμου και συνεπώς είναι αδύνατη η εκτίμηση των παραμέτρων της κατανομής. Εναλλακτική επιλογή θα ήταν η εκτίμηση των παραμέτρων με την μέθοδο των ροπών αλλά φυσικά αυτό δεν αποτελεί αξιόπιστη μέθοδο για τα δεδομένα που μελετώνται εδώ.

Επομένως, η επιλογή της μετατοπισμένης κατανομής Pareto αποτελεί ιδανική λύση στο κυρίαρχο πρόβλημα που είναι η ελάχιστη τιμή 10.000 η οποία αποτελεί την μικρότερη αποζημίωση. Η μετατοπισμένη κατανομή Pareto που χρησιμοποιήθηκε, έχει ως πρώτη παράμετρο την ελάχιστη τιμή των δεδομένων (κάτι άκρως λογικό) κάτι το οποίο επιβεβαιώνεται πλήρως μαθηματικά καθώς ο εκτιμητής μεγίστης πιθανοφάνειας είναι η ελάχιστη παρατήρηση από τα δεδομένα.



## Παράρτημα

### Π1. Πρόγραμμα εκτίμησης για Weibull

```
γ = shape;
θ = scale;
Clear[Qγ, Qθ, Qγc, Qθc];
iterations = 5000; (*ο αριθμός των δειγμάτων*)
n = 300; (* το μέγεθος του δείγματος*)
γ = 2;
θ = 2;
ρ = 0.7; (* το ποσοστό του δείγματος που θα λογοκρίνουμε*)
Qγ = Table[0, {iterations}];
Qθ = Table[0, {iterations}];
Qγc = Table[0, {iterations}];
Qθc = Table[0, {iterations}];
Do[
  ls = RandomReal[WeibullDistribution[γ, θ], n];
  ls1 := ls/ θ; (*από εδώ ξεκινά ο αλγόριθμος για την
  εύρεση των ΕΜΠ για το πλήρες δείγμα*)
  g0 := Apply[Plus, Log[ls1]]/Length[ls];
  g1[x_] := Apply[Plus, ls1^x]/Length[ls];
  g2[x_] := Apply[Plus, ls1^x*Log[ls1]]/Length[ls];
  G[x_] = 1/x - g2[x] + g0;
  x2 = x /. FindRoot[G[x] == 0, {x, 1}];
  Qγ[[k]] = x2;
  Qθ[[k]] =
  Mean[(ls)^γ]^(1/ γ); (*εδώ τελειώνει ο αλγόριθμος για την
  εύρεση των ΕΜΠ για το πλήρες δείγμα*)
  Tq = InverseCDF[WeibullDistribution[γ, θ], 1 - ρ]; (*από εδώ ξεκινά ο αλγόριθμος για την
  εύρεση των ΕΜΠ για το λογοκρίμενο δείγμα*)
  lsc := Select[ls, # < Tq &];(*το κομμάτι που κρατάμε*)
  lsT := ConstantArray[Tq,
```

```

Length[Select[ls, # > Tq &]]; (*το κομμάτι που συμπληρώνουμε*)
lsct = Join[lsc, lsT]; (*το συνολικό censored δείγμα*)
lsct1 :=
lsct/θ; (*από δω και κάτω είναι ο αλγόριθμος για
την εύρεση των ΕΜΠ *)

gct1 := Apply[Plus, Log[lsc/θ]];
gct2[x_] := Apply[Plus, lsct1^x*Log[lsct1]];
Gc[x_] := Length[lsc]/x - gct2[x] + gct1;
γ2 = x /. FindRoot[Gc[x] == 0, {x, 1}]; (*ο εκτιμητής του shape*)
Qγc[[k]] = γ2;
lsct2 := Apply[Plus, lsct^γ];
Qθc[[k]] = (lsct2/Length[lsc])^(1/γ),
{k, 1, iterations}]
Qγ;
Qθ;
Qγc;
Qθc;
Print["θ=", Mean[Qθ]]
Print["sdθ=", StandardDeviation[Qθ]]
Print["θc=", Mean[Qθc]]
Print["sdθc=", StandardDeviation[Qθc]]
Print["γ=", Mean[Qγ]]
Print["sdγ=", StandardDeviation[Qγ]]
Print["γc=", Mean[Qγc]]
Print["sdγc=", StandardDeviation[Qγc]]

```

## Π2. Πρόγραμμα εκτίμησης για Weibull για μικρά δείγματα

```
WeibSample[θ_, γ_, n_, Upper_] :=  
Module[{RanCen, Init}, Clear[data, MeanUnCen]; SeedRandom[];  
    (*Δημιουργούμε τυχαίο δείγμα από την ομοιόμορφη κατανομή μεγέθους n*);  
    RanCen =  
    If[Upper > 0, Table[Random[Real, {0, Upper}], {i, 1, n}],  
    Table[0, {i, 1, n}];  
    (*Δημιουργούμε τυχαίο δείγμα από την κατανομή Weibull μεγέθους n*);  
    Init = Table[θ Power[-Log[Random[]], (γ)^-1], {i, 1, n}];  
    (*Δημιουργούμε το λογοκριμμένο δείγμα *);  
    data = Table[  
    If[Init[[i]] > RanCen[[i]] && Upper > 0, {RanCen[[i]], 1}, {Init[[i]], 0}], {i, 1, n}];  
    Is = Init;  
    ]  
γ = shape;  
θ = scale;  
Clear[Qγ, Qθ, Qγc, Qθc];  
iterations = 5000; (*ο αριθμός των δειγμάτων*)  
n = 50; (* το μέγεθος του δείγματος*)  
γ = 2;  
θ = 2;  
ρ = 0.5; (* το ποσοστό του δείγματος που θα λογοκρίνουμε*)  
Qγ = Table[0, {iterations}];  
Qθ = Table[0, {iterations}];  
Qγc = Table[0, {iterations}];  
Qθc = Table[0, {iterations}];  
  
    (*Βρίσκουμε το πάνω όριο της ομοιόμορφης*)  
Upper = Last[ Last[FindRoot[(θ (Gamma[1/θ] - Gamma[1/γ, (u/θ)^γ ]))/(u γ) == 1 - ρ, {u, 2}]]]  
    (* datunc είναι το κομμάτι του δείγματος που δεν  
    λογοκρίνεται*)  
    (*datc αυτό που λογοκρίναμε*)
```

```

Do[
WeibSample[ $\theta$ ,  $\gamma$ , n, Upper]
data;
Isct = data[[All, 1]];          (*διαλέγει το συνολικό δείγμα*)
datc = {}; Isc = {};
Do[
  If[data[[j, 2]] == 1, AppendTo[Isct, data[[j, 1]]],
  AppendTo[datc, data[[j, 1]]] ]
, {j, 1, n}];
Is = init;
WeibSample[ $\theta$ ,  $\gamma$ , n, Upper]
data;
Is1 := Is/  $\theta$ ;              (*από εδώ ξεκινά ο αλγόριθμος για την ευρεση των ΕΜΠ για
                             το πλήρες δείγμα*)
g0 := Apply[Plus, Log[Is1]]/Length[Is];
g1[x_] := Apply[Plus, Is1^x]/Length[Is];
g2[x_] := Apply[Plus, Is1^x*Log[Is1]]/Length[Is];
G[x_] = 1/x - g2[x] + g0;
x2 = x /. FindRoot[G[x] == 0, {x, 1}];
Q $\gamma$ [[k]] = x2;
Q $\theta$ [[k]] = Mean[(Is)^ $\gamma$ ^(1/ $\gamma$ )];
(*εδώ τελειώνει ο αλγόριθμος για την ευρεση των ΕΜΠ για το πλήρες δείγμα*)
(*από εδώ ξεκινά ο αλγόριθμος για την ευρεση των ΕΜΠ για το
λογοκριμμένο δείγμα*)
(* Isc είναι το κομμάτι του δείγματος που δεν λογοκρίνεται*)
(*datc αυτό που λογοκρίναμε*)
Isct1 := Isct/ $\theta$ ;          (*απο δω και κατω είναι ο αλγοριθμος για την ευρευση των ΕΜΠ *)
gct1 := Apply[Plus, Log[Isct/ $\theta$ ]];
gct2[x_] := Apply[Plus, Isct1^x*Log[Isct1]];
Gc[x_] := Length[Isct]/x - gct2[x] + gct1;
 $\gamma$ 2 = x /. FindRoot[Gc[x] == 0, {x, 1}];          (*ο εκτιμητής του shape*)
Q $\gamma$ c[[k]] =  $\gamma$ 2;

```

```
Isct2 := Apply[Plus, Isct^γ];
Qθc[[k]] = (Isct2/Length[Isct])^(1/γ);,
{k, 1, iterations}
Qγ;
Qθ;
Qγc;
Qθc;
Print["θ=", Mean[Qθ]]
Print["sdθ=", StandardDeviation[Qθ]]
Print["θc=", Mean[Qθc]]
Print["sdθc=", StandardDeviation[Qθc]]
Print["γ=", Mean[Qγ]]
Print["sdγ=", StandardDeviation[Qγ]]
Print["γc=", Mean[Qγc]]
Print["sdγc=", StandardDeviation[Qγc]]
```

### Π3. Πρόγραμμα εκτίμησης για Pareto

```
θ = scale;

Clear[Qθ, Qθc];

iterations = 5000;          (*ο αριθμός των δειγμάτων*)
n = 30;                    (* το μέγεθος του δείγματος*)
θ = 2;
ρ = 0.4;                  (* το ποσοστό του δείγματος που θα λογοκρίνουμε*)

Qθ = Table[0, {iterations}];
Qθc = Table[0, {iterations}];

Do[
  Is = RandomReal[ParetoDistribution[1, θ], n];
  Qθ[[k]] =
    1/Mean[Log[
      Is]];                (*εδώ τελειώνει ο αλγόριθμος για την εύρεση των ΕΜΠ για
                           το πλήρες δείγμα*)
  Tq = InverseCDF[ParetoDistribution[1, θ],
    1 - ρ];                (*από εδώ ξεκινά ο αλγόριθμος για την εύρεση των ΕΜΠ για
                           το λογοκριμένο δείγμα*)
  Isc := Select[Is, # < Tq &];      (*το κομμάτι που κρατάμε*)
  IsT := ConstantArray[Tq,
    Length[Select[Is, # > Tq &]]];  (*το κομμάτι που συμπληρώνουμε*)
  Isct = Join[Isc, IsT];          (*το συνολικό censored δείγμα*)
  Qθc[[k]] = 1/Mean[Log[Isct]],
  {k, 1, iterations}]

Qγ;
Qθ;
Qγc;
Qθc;

Print["θ=", Mean[Qθ]]
Print["sdθ=", StandardDeviation[Qθ]]
```



```
Print[" $\theta_c$ =", Mean[Q $\theta_c$ ]]
```

```
Print["sd $\theta_c$ =", StandardDeviation[Q $\theta_c$ ]]
```

#### Π4. Πρόγραμμα εκτίμησης για Pareto με λογοκρισία τύπου I και II

```
parSample[θ_,n_,Upper_]:=Module[{RanCen,Init},Clear[data,MeanUnCen];SeedRandom[];

      (*Δημιουργούμε τυχαίο δείγμα από την ομοιόμορφη κατανομή
      μεγέθους n*);

RanCen=If[Upper>0,Table[Random[Real,{1,Upper}],{i,1,n}],Table[0,{i,1,n}]];

      (*Δημιουργούμε τυχαίο δείγμα από την κατανομή Weibull
      μεγέθους n*);

Init=RandomReal[ParetoDistribution[1,θ],n];

      (*Δημιουργούμε το λογοκριμένο δείγμα *);

data=Table[If[Init[[i]]>RanCen[[i]]&&Upper>0,{RanCen[[i]],1},{Init[[i]],0}],{i,1,n}];

Is=Init;

]

θ=parameter;

Clear[Qθ,Qθc];

iterations=5000;      (*ο αριθμός των δειγμάτων*)

n=50;      (* το μέγεθος του δείγματος*)

θ=2;

p=0.1      (* το ποσοστό του δείγματος που θα λογοκρίνουμε*)

Qθ=Table[0,{iterations}];

Qθc=Table[0,{iterations}];

      (*Βρίσκουμε το πάνω όριο της ομοιόμορφης*)

Upper=Last[Last[FindRoot[(u^(-θ+1)-1)/((u-1)*(-θ+1))=p,{u,2}]]]

      (* datunc είναι το κομμάτι του δείγματος που δεν λογοκρίνεται*)

      (* datc αυτό που λογοκρίναμε*)

Do[

parSample[θ,n,Upper]

data;

Isct=data[[All,1]];      (*διαλέγει το συνολικό δείγμα*)

datc={};Isct={};
```

```

Do[
  If[data[[j,2]]==1,AppendTo[Isct,data[[j,1]]], AppendTo[datc,data[[j,1]]] ]
, {j,1,n}];
Is=init;
parSample[θ,n,Upper]
data;
Qθ[[k]]=1/Mean[Log[Is]];
      (*εδώ τελειώνει ο αλγόριθμος για την εύρεση των ΕΜΠ για το πλήρες
      δείγμα*)
      (*από εδώ ξεκινά ο αλγόριθμος για την εύρεση των ΕΜΠ για το λογοκριμένο
      δείγμα*)
      (* Isct είναι το κομμάτι του δείγματος που λογοκρίνεται*)
      (* datc αυτό που δεν λογοκρίναμε*)
      (*από δω και κάτω είναι ο αλγόριθμος για την εύρεση των ΕΜΠ *)
Qθc[[k]]=1/(Apply[Plus,Log[Isct]]/Length[datc]);,
{k,1,iterations}]
Qθ;
Qθc;
Print["θ=",Mean[Qθ]]
Print["sdθ=",StandardDeviation[Qθ]]
Print["θc=",Mean[Qθc]]
Print["sdθc=",StandardDeviation[Qθc]]

```

## Π5. Πρόγραμμα εκτίμησης για Gamma

```
"γ=shape";
"θ=scale";
Clear[Qγ, Qθ, Qγc, Qθc];
iterations = 5000;          (*ο αριθμός των δειγμάτων*)
n = 100;                   (* το μέγεθος του δείγματος*)
γ = 2;
θ = 1;
ρ = 0.3;                   (* το ποσοστό του δείγματος που θα λογοκρίνουμε*)
Qγ = Table[0, {iterations}];
Qθ = Table[0, {iterations}];
Qγc = Table[0, {iterations}];
Qθc = Table[0, {iterations}];
Do[
  Is = RandomReal[GammaDistribution[γ, θ], n];
  Is1 := Is/θ;
  (*από εδώ ξεκινά ο αλγόριθμος για την εύρεση των ΕΜΠ για το πλήρες δείγμα*)
  g0 := Mean[Log[Is1]];
  g1[x_] := Apply[Plus, Is1^x]/Length[Is];
  g2[x_] := Apply[Plus, Is1^x*Log[Is1]]/Length[Is];
  G[x_] := PolyGamma[x] - g0;
  x2 = x /. FindRoot[G[x] == 0, {x, 1}];
  Qγ[[k]] = x2;
  Qθ[[k]] = Mean[Is]/γ;
  (*εδώ τελειώνει ο αλγόριθμος για την εύρεση των ΕΜΠ για το πλήρες δείγμα*)
  Tq = InverseCDF[GammaDistribution[γ, θ], 1 - ρ];
  (*από εδώ ξεκινά ο αλγόριθμος για την εύρεση των ΕΜΠ για το λογοκριμένο
  δείγμα*)
  Isc := Select[Is, # < Tq &];          (*το κομμάτι που κρατάμε*)
  IsT := ConstantArray[Tq, Length[Select[Is, # > Tq &]]];          (*το κομμάτι που συμπληρώνουμε*)
  Isct := Join[Isc, IsT];              (*το συνολικό censored δείγμα*)
  Gθ[x_] :=
    -Length[Isc]*γ + Apply[Plus, Isc]/x + (n - Length[Isc])*Tq^γ/x^γ*Exp[-Tq/x]*
```

```

Gamma[γ, Tq/x] // N;
γ2 = x /. FindRoot[Gθ[x] == 0, {x, 1}]; (*ο εκτιμητής του scale*)
Qθc[[k]] = γ2;
G1[x_] = D[Gamma[x, Tq/θ], x];
Gγ[x_] := -Length[lsc]*Log[θ] + Apply[Plus, Log[lsc]] - n*PolyGamma[x] + (n -
Length[lsc])*G1[x]/Gamma[x, Tq/θ];
γ1 = x /. FindRoot[Gγ[x] == 0, {x, 1}]; (*ο εκτιμητής του shape*)
Qγc[[k]] = γ1;;
{k, 1, iterations}
Print["θ=", Mean[Qθ]]
Print["sdθ=", StandardDeviation[Qθ]]
Print["θc=", Mean[Qθc]]
Print["sdθc=", StandardDeviation[Qθc]]
Print["γ=", Mean[Qγ]]
Print["sdγ=", StandardDeviation[Qγ]]
Print["γc=", Mean[Qγc]]
Print["sdγc=", StandardDeviation[Qγc]]

```

## Π6. Πρόγραμμα εκτίμησης της παραμέτρου κλίμακας (scale parameter) για Gamma με λογοκρισία τύπου I και II

```

GammaSample[α_, θ_, n_, Upper_] :=
Module[{RanCen, Init}, Clear[data, MeanUnCen]; SeedRandom[];
    (*Δημιουργούμε τυχαίο δείγμα από την ομοιόμορφη κατανομή μεγέθους n*);
RanCen = If[Upper > 0, Table[Random[Real, {1, Upper}], {i, 1, n}],
Table[0, {i, 1, n}]];
    (*Δημιουργούμε τυχαίο δείγμα από την κατανομή Gamma μεγέθους n*);
Init = RandomReal[GammaDistribution[α, θ], n];
    (*Δημιουργούμε το λογοκριμμένο δείγμα *);
data = Table[If[Init[[i]] > RanCen[[i]] && Upper > 0, {RanCen[[i]], 1}, {Init[[i]], 0}], {i, 1, n}];
Is = Init;
]
Clear[Qθ, Qθc, g];
iterations = 5000; (*ο αριθμός των δειγμάτων*)
n = 50; (* το μέγεθος του δείγματος*)
θ = 2;
α = 2;
ρ = 0.3; (* το ποσοστό του δείγματος που θα λογοκρίνουμε*)
Qθ = Table[0, {iterations}];
Qθc = Table[0, {iterations}]; (*Βρίσκουμε το πάνω όριο της ομοιόμορφης*)
Upper = θ*
Last[Last[
FindRoot[
1/Gamma[α]*(x^(α - 1)*
Exp[-x] + (1 - α/x)*(Gamma[α] - Gamma[α, x])) == 1 - ρ, {x, 2}]]]

Do[
GammaSample[α, θ, n, Upper];
data;
Isct = data[[All, 1]]; (*διαλέγει το συνολικό δείγμα*)
datc = {}; Isc = {};
Do[

```

```

If[data[[j, 2]] == 1, AppendTo[lsc, data[[j, 1]], AppendTo[datc, data[[j, 1]] ] , {j, 1, n}];
ls = Init;
GammaSample[α, θ, n, Upper];
data;
Qθ[[k]] = Mean[ls]/α;
(*εδώ τελειώνει ο αλγόριθμος για την εύρεση των ΕΜΠ για το πλήρες δείγμα*)
(*από εδώ ξεκινά ο αλγόριθμος για την εύρεση των ΕΜΠ για το λογοκριμένο
δείγμα*)
(* lsc είναι το κομμάτι του δείγματος που λογοκρίνεται*)
(*datc αυτό που δεν λογοκρίναμε*)
(*από δω και κατω είναι ο αλγόριθμος για την εύρεση των ΕΜΠ *)
modf[x_] = (lsc/x)^α*Exp[-lsc/x]/(Gamma[α] - Gamma[α, lsc/x]);
h[x_] = -Length[datc] + Apply[Plus, datc]/x - Apply[Plus, modf[x]];
Qθc[[k]] = Last[Last[FindRoot[h[x] == 0, {x, 3}]]];,
{k, 1, iterations}]
Qθ;
Qθc;
Print["θ=", Mean[Qθ]]
Print["sdθ=", StandardDeviation[Qθ]]
Print["θc=", Mean[Qθc]]
Print["sdθc=", StandardDeviation[Qθc]]

```

## Π7. Πρόγραμμα εκτίμησης της παραμέτρου σχήματος (shape parameter) για Gamma με λογοκρισία τύπου I και II

```

GammaSample[α_, θ_, n_, Upper_] :=
Module[{RanCen, Init}, Clear[data, MeanUnCen]; SeedRandom[];
    (*Δημιουργούμε τυχαίο δείγμα από την ομοιόμορφη κατανομή μεγέθους n*);
RanCen = If[Upper > 0, Table[Random[Real, {1, Upper}], {i, 1, n}], Table[0, {i, 1, n}]];
    (*Δημιουργούμε τυχαίο δείγμα από την κατανομή Gamma μεγέθους n*);
Init = RandomReal[GammaDistribution[α, θ], n];
    (*Δημιουργούμε το λογοκρινόμενο δείγμα *);
data = Table[ If[Init[[i]] > RanCen[[i]] && Upper > 0, {RanCen[[i]], 1}, {Init[[i]], 0}], {i, 1, n}];
Is = Init;
]
Clear[Qγ, Qγc, g];
iterations = 5000;          (*ο αριθμός των δειγμάτων*)
n = 500;                   (* το μέγεθος του δείγματος*)
θ = 2;
α = 3;
p = 0.1                    (* το ποσοστό του δείγματος που θα λογοκρίνουμε*);
Qα = Table[0, {iterations}];
Qαc = Table[0, {iterations}];
                                (*Βρίσκουμε το πάνω όριο της ομοιόμορφης*)
Upper = θ*
Last[Last[
    FindRoot[
        1/Gamma[α]*(x^(α - 1)*Exp[-x] + (1 - α/x)*(Gamma[α] - Gamma[α, x])) == 1 - p, {x, 2}]]]

Do[
GammaSample[α, θ, n, Upper];
data;
Isct = data[[All, 1]];      (*διαλέγει το συνολικό δείγμα*)
datc = {}; Isc = {};
Do[
    If[data[[j, 2]] == 1, AppendTo[Isc, data[[j, 1]]],

```



```

AppendTo[datc, data[[j, 1]] ], {j, 1, n}); ls = Init;
GammaSample[α, θ, n, Upper];
data;
Qα[[k]] =
Last[Last[FindRoot[PolyGamma[x] == Mean[Log[ls]] - Log[θ], {x, 3}]];
(*εδώ τελειώνει ο αλγόριθμος για την εύρεση των ΕΜΠ για το πλήρες δείγμα*)
(*από εδώ ξεκινά ο αλγόριθμος για την εύρεση των ΕΜΠ για το λογοκρινμένο
δείγμα*)
(* lsc είναι το κομμάτι του δείγματος που λογοκρίνεται*)
(*datc αυτό που δεν λογοκρίναμε*)
(*από δω και κάτω είναι ο αλγόριθμος για την εύρεση των ΕΜΠ *)
mod2f[x_] = (PolyGamma[x] - D[Gamma[x, lsc/θ], x])/(Gamma[x] - Gamma[x, lsc/θ]);
h2[x_] = -n*PolyGamma[x] + Apply[Plus, Log[datc/θ]] - Apply[Plus, modf[x]];
Qαc[[k]] = Last[Last[FindRoot[h2[x] == 0, {x, 3}]]];, {k, 1, iterations}
Qα;
Qαc;
Print["α=", Mean[Qα]]
Print["sdα=", StandardDeviation[Qα]]
Print["αc=", Mean[Qαc]]
Print["sdαc=", StandardDeviation[Qαc]]

```

## ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Arnold, B. C. (1983) Pareto Distributions. International Co- op Publishing Co., Fairland, Md.
- Bain L., and Engelhardt (1991), Statistical Analysis of Reliability and Life- testing Models, second edition NY Dekker
- Bowman, K. O. and Shenton, L. R. (1987) Properties of Estimators for the Gamma Distribution. Marcel Dekker, New York.
- Cohen, A. C. (1963) Progressively censored samples in life testing. *Technometrics*, 5, 237- 339
- Cohen, A. C. (1965a) Maximum Likelihood estimation in the Weibull distribution based on complete and censored samples. *Technometrics*, 7, 579- 588.
- Costantino, R. F., and Desharnais, R. A. (1981). Gamma distributions of adult numbers for Tribolium populations in the regions of their steady states, *Journal of Animal Ecology*, 50, 667-681.
- Dennis, B., and Patil, G. P. (1984). The gamma distribution and the weighted multimodal gamma distributions as models of population abundance, *Mathematical Biosciences*, 68, 187-212.
- Efron , (1975) Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency). *Ann. Statistics, Volume 3, Number 6 (1975), 1189-1242*
- Harter, H. L. and Moore, A. H. (1965) Maximum likelihood estimation of the parameters of the gamma and Weibull populations from censored samples. *Technometrics*, 7, 639- 643.
- Harter, H. L. and Moore, A. H. (1967) Asymptotic variances and covariances of maximum likelihood estimators of parameters of Weibull and gamma populations from censored samples. *Ann. Mth. Statist.*, 38, 557- 571.
- Hayakawa, M. (1951). The application of Pareto's law of income to Japanese data, *Econometrics*, 19, 174-183.
- Lawless, J. F. (1982) *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- Lehman, E. H. (1963) Shapes, moments and estimators of the Weibull distribution. *IEEE Trans. Reliab.*, R-12, 3, 32-38.
- Pigou, A. C. (1932). *The Economics of Welfare*, London: Macmillan
- Qian, L.F. (2012), The Fisher information matrix for a three-parameter exponentiated Weibull distribution under type II censoring. *Statistical Methodology*. 9 (2012), 320- 329.
- Rohatgi V. K. (2003) *Statistical inference*. Dover Publication, New York.

Rosen P., and Rammler, B. (1933). The laws governing the fineness of powdered coal, *Journal of the Institute of Fuels*, 6, 29-36.

Salem, A. B., and Mount, T. D. (1974). A convenient descriptive model of income distribution: the gamma density, *Econometrica*, 42, 1115-1127.

Shirras, G. F. (1935). The Pareto law and the distribution of income, *Economic Journal*, 45, 663-681.