

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



## ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

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

### Η ΚΑΤΑΝΟΜΗ MAXWELL ΚΑΙ ΣΥΝΑΦΕΙΣ ΚΑΤΑΝΟΜΕΣ ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ

Τσιότρας Ζώης

Διπλωματική Εργασία

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

Πειραιάς

Οκτώβριος 2017



# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

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

**Η ΚΑΤΑΝΟΜΗ MAXWELL  
ΚΑΙ ΣΥΝΑΦΕΙΣ ΚΑΤΑΝΟΜΕΣ ΣΤΗΝ  
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ**

**Τσιότρας Ζώης**

**Διπλωματική Εργασία**

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

**Πειραιάς**

**Οκτώβριος 2017**

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

Τα μέλη της Επιτροπής ήταν  
Κ. Πολίτης (επιβλέπων)  
Ε. Χατζηκωνσταντινίδης  
Β. Σεβρόγλου

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 MAXWELL AND RELATED  
DISTRIBUTIONS IN ACTUARIAL SCIENCE**

**By**

**Tsiotras Zois**

MSc Dissertation

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

**Piraeus  
October 2017**



## **Ευχαριστίες**

Με την ολοκλήρωση της διπλωματικής μου εργασίας, θα ήθελα να ευχαριστήσω τους ανθρώπους που συνέβαλαν στην διεκπεραίωση της.

Ξεκινώ με τον επιβλέποντα μου κ. Κωνσταντίνο Πολίτη, Αναπληρωτή Καθηγητή του Πανεπιστημίου Πειραιώς. Είναι από τους καθηγητές που αξίζουν καθώς έχει την διάθεση να μεταδώσει τις γνώσεις του και να αφιερώσει χρόνο στους φοιτητές του. Τον ευχαριστώ ολόψυχα για την υπομονή του αλλά ταυτόχρονα για όσα μου προσέφερε και του εύχομαι καλή συνέχεια και προσωπική ευτυχία σε όλους τους τομείς στον ίδιο αλλά και την οικογένεια του.

Ευχαριστώ τα μέλη της συμβουλευτικής επιτροπής, κύριο Ευστάθιο Χατζηκωνσταντινίδη, Αναπληρωτή καθηγητή του Πανεπιστημίου Πειραιώς και τον κύριο Βασίλειο Σεβρόγλου, Επίκουρο καθηγητή του Πανεπιστημίου Πειραιώς. Ευχαριστώ όμως και όλους τους καθηγητές για τις γνώσεις που μου προσέφεραν κατά την διάρκεια του μεταπτυχιακού προγράμματος.

Στην συνέχεια θα ήθελα να ευχαριστήσω εκ βάθους καρδιάς την οικογένεια μου για την στήριξη τους και την αγάπη τους σε μια δύσκολη περίοδο για την χώρα, καθώς μου προσέφεραν την δυνατότητα να συνεχίσω τις πανεπιστημιακές μου σπουδές και να κάνω το όνειρο μου πραγματικότητα.

Τέλος θα ήθελα να ευχαριστήσω την κοπέλα μου Ευαγγελία Νομικού για την πολύτιμη στήριξη της και την επιμονή να συνεχίσω την προσπάθεια μου αλλά και για την αγάπη και την υπομονή που μου έδειξε όλο αυτό το διάστημα .

*Για την Έβελιν και την οικογένεια μου  
Σωτήρη, Φανή και Αναστασία.*



# Περίληψη

Στην παρούσα διπλωματική εργασία μελετάμε διεξοδικά διάφορες συνεχείς κατανομές στην αναλογιστική επιστήμη. Πιο συγκεκριμένα παρουσιάζονται αναλυτικά με την χρήση προγραμμάτων (Mathematica ,R) και οι πιο δημοφιλείς κατανομές στον ασφαλιστικό χώρο και μας επιτρέπεται να κατανοήσουμε καλύτερα την εφαρμογή τους. Σε ευρύτερο πλαίσιο αναλύεται η κατανομή Maxwell Boltzmann η οποία είναι στα αρχικά στάδια της εφαρμογής της στον αναλογισμό σε χαρτοφυλάκια ζωής. Παρουσιάζονται αναλυτικά όλοι τύποι και οι εφαρμογές της για διάφορες τιμές των παραμέτρων της για καλύτερη κατανόηση και ένα αριθμητικό παράδειγμα με πραγματικά ασφαλιστικά δεδομένα ώστε να διαπιστώσουμε την εφαρμογής της μέσα στον ασφαλιστικό χώρο με τον έλεγχο του Kolmogorov-Smirnov. Τέλος δίνεται και προσέγγιση του υπολειπόμενου χρόνου ζωής και της βαθμίδας αποτυχίας καθώς και όλων των μέτρων της κατανομής τόσο θεωρητικά όσο και εμπειρικά.

# ***Abstract***

In the present thesis we examine, in detail, various continuous distributions which are used as models in actuarial science. More specifically, we use the Mathematica, R and other programs to show the most popular distributions in the insurance sector and we are capable of understanding their applications. In a wider context, we analyzed the Maxwell Boltzmann distribution, which is in an initial stage of its application in actuarial science in life portfolios. Moreover, we present in detail all the types and the applications for various parameter values of Maxwell, for better understanding through an example with real data from an insurance company. What is more, we checked how the Maxwell distribution fits a real data set from insurance using the Kolmogorov - Smirnov test. To conclude, we study the functions of mean residual life and failure rate, and compare the results we obtain theoretically with those using the real data set.

## Βοηθητικοί Πίνακες- Συντομογραφίες

Σύμβολο	Περιγραφή
$X$	τ.μ. που περιγράφει την συνολική διάρκεια ζωής
$F(x)$	Συνάρτηση κατανομής της τ.μ. $X$
$S(x)=1-F(x)$	Συνάρτηση αξιοπιστίας της τ.μ. $X$
$f(x)=F'(x)$	Συνάρτηση αξιοπιστίας της τ.μ. $X$
$\Gamma(x)$	Συνάρτηση Γάμμα
$\mu=E(x)$	Μέση τιμή της $X$
$T_x=X-x   X>x$	τ.μ. υπολειπόμενου χρόνου ζωής ηλικίας $x$
$\log$	Φυσικός Λογάριθμος
$\Phi$	Η αθροιστική συνάρτηση κατανομής της τυπικής κατανομής

Πίνακας 1:Χρήσιμες Έννοιες

Συντομογραφία	Ανάπτυξη στα Αγγλικά
MRL	Mean Residual Life
FR	Failure Rate
IFR	Increasing Failure Rate
DFR	Decreasing Failure Rate

Πίνακας 2 :Συντομογραφίες



# Περιεχόμενα

1.	Εισαγωγικές έννοιες στην αναλογιστική επιστήμη .....	14
1.1	Πρόλογος και στόχος αυτής της εργασίας .....	14
1.2	Περιγραφή της εργασίας.....	14
2.	Συνεχείς κατανομές στον αναλογισμό.....	16
2.1	Βαθμίδα αποτυχίας ( Failure Rate).....	16
2.2	Ο υπολειπόμενος χρόνος ζωής .....	21
2.3	Η εκθετική κατανομή (Exponential Distribution).....	21
2.4	Η κατανομή Γάμμα (Gamma Distribution) .....	25
2.5	Κατανομή Weibull .....	30
2.5.1	Αντίστροφη και γενικευμένη κατανομή Weibull.....	37
2.6	Κατανομή Pareto .....	38
2.7	Rayleigh distribution.....	44
3.	Η κατανομή Maxwell–Boltzmann στην αναλογιστική επιστήμη .....	51
3.1	Εισαγωγή.....	51
3.2	Η κατανομή Maxwell –Boltzmann .....	52
3.3	Αντίστροφη Maxwell-Boltzmann κατανομή (Inverse Maxwell –Boltzmann) .....	61
4.	Εφαρμογή της κατανομής Maxwell–Boltzmann σε πραγματικά αναλογιστικά δεδομένα .....	62
4.1	Εισαγωγή.....	62
4.2	Έλεγχος προσαρμογής των δεδομένων σε άλλες συνεχείς κατανομές .....	69
4.3	Υπολογισμός άλλων μέτρων για την κατανομή Maxwell- Boltzmann και περαιτέρω ανάλυση αυτής .....	70
4.4	Υπολογισμός της παραμέτρου με την μέθοδο των ροπών .....	75
4.5	Υπολογισμός Μέσου Υπολειπόμενου χρόνου ζωής .....	76
4.6	Υπολογισμός βαθμίδας αποτυχίας για τα δεδομένα .....	79
4.7	Παρουσίαση Γραφικών παραστάσεων για $\alpha=6,49$ τιμή παραμέτρου.....	81
5.	Παράρτημα : .....	82
6.	Βιβλιογραφία.....	89

# 1. Εισαγωγικές έννοιες στην αναλογιστική επιστήμη

## 1.1 Πρόλογος και στόχος αυτής της εργασίας

Όπως δηλώνει και ο τίτλος της εργασίας στόχος είναι να μελετηθούν κάποιες συνεχείς κατανομές που έχουν πρωταγωνιστικό ρόλο στον κλάδο ασφαλίσεων ζωής. Έτσι αναλύονται διεξοδικά οι κατανομές αυτές και περιγράφονται οι εφαρμογές στην αναλογιστική επιστήμη. Με την ανάλυση και την δημιουργία όλων των γραφικών παραστάσεων και των τύπων γίνεται ευκολότερη η κατανόηση των κατανομών στον αναλογισμό. Κατά κύριο λόγο περιγράφεται και αναλύεται η κατανομή η Maxwell-Boltzmann και σκοπός είναι η ανάλυση της και η εφαρμογή της στον κλάδο ασφαλίσεων ζωής. Έτσι περιγράφονται πραγματικά δεδομένα ασφαλιστικής εταιρίας στο κατά πόσο προσαρμόζονται στην εν λόγω κατανομή και όχι μόνο βρίσκοντας πληροφορίες σχετικά με την προσαρμογή της αλλά και άλλα στατιστικά μέτρα θέσης και μεγέθη αλλά και η βαθμίδα αποτυχίας και ο υπολειπόμενος χρόνος ζωής (MRL-Mean Residual Life).

## 1.2 Περιγραφή της εργασίας

### Κεφάλαιο 1

Περιγράφονται συνοπτικά θέματα που θα αναλυθούν στα κεφάλαια που ακολουθούν.

### Κεφάλαιο 2

Στο κεφάλαιο αυτό πραγματοποιείται μια αναλυτική περιγραφή της βαθμίδας αποτυχίας και του υπολειπόμενου χρόνου ζωής με εκτενή αναφορά στους ορισμούς αυτών, των τύπων και των ιδιοτήτων τους. Έπειτα στο κεφάλαιο αυτό αναφέρονται διάφορες συνεχείς κατανομές η οποίες έχουν εκτενή εφαρμογή στον αναλογισμό. Δίνονται μια σειρά από σχέσεις αυτών των κατανομών και οι γραφικές παραστάσεις τους για να εξεταστεί η ουρά ως προς το βάρος της και τέλος μερικά παραδείγματα για καλύτερη κατανόηση.

### Κεφάλαιο 3

Στο συγκεκριμένο κεφάλαιο της εργασίας θα παρουσιάζεται διεξοδικά η κατανομή Maxwell-Boltzmann. Πρόκειται μια εκτενή περιγραφή της εν λόγω κατανομής σχετικά με τις ιδιότητες και τις εφαρμογές της στην αναλογιστική επιστήμη. Πραγματοποιείται διεξοδική περιγραφή των τύπων της κατανομής αλλά και γραφικές παραστάσεις αυτής.

## **Κεφάλαιο 4**

Στο κεφάλαιο αυτό πραγματοποιείται υπολογισμός ενός αριθμητικού παραδείγματος με πραγματικά δεδομένα όπου εξετάζεται η προσαρμογή τους στην κατανομή Maxwell-Boltzmann αλλά και σε άλλες κατανομές και εύρεση όλων των μέτρων θέσης για καλύτερη περιγραφή της κατανομής.

### **Παράρτημα**

Στην συγκεκριμένη ενότητα υπάρχουν κάποια αριθμητικά δεδομένα , ο πίνακας αποτελεσμάτων του Kolmogorov smirnov και κάποιοι από τους κώδικες που χρησιμοποιήθηκαν σε πρόγραμμα R και Mathematica για την εξαγωγή των αποτελεσμάτων της εργασίας.

## 2. Συνεχείς κατανομές στον αναλογισμό

### Εισαγωγή

Στο δεύτερο κεφάλαιο της διπλωματικής θα αναλύσουμε διεξοδικά συνεχείς κατανομές οι οποίες έχουν μεγάλη δραστηριοποίηση στον ασφαλιστικό χώρο και την αναλογιστική επιστήμη. Σε πρώτο στάδιο αναλύεται η βαθμίδα αποτυχίας με μια γενική προσέγγιση και καταγράφεται ο γενικός τύπος που προσαρμόζεται σε όλες τις κατανομές ώστε να υπολογιστεί. Στην συνέχεια αναλύουμε κατανομές όπως η εκθετική, η Γάμμα, Pareto και Weibull και όλα τα μέτρα αυτών όπως μέση τιμή, διακύμανση, οι ροπές τους, οι ροπογεννήτριες (εφόσον υπάρχουν) αλλά και η βαθμίδα αποτυχίας τους. Τέλος παρουσιάζονται γραφικές παραστάσεις των διάφορων συναρτήσεων κάθε κατανομής όπως συνάρτηση κατανομής, πυκνότητας και επιβίωσης ώστε να αναλυθεί η ουρά της κάθε κατανομής και πόσο γρήγορα συγκλίνει στο μηδέν.

### 2.1 Βαθμίδα αποτυχίας ( Failure Rate)

Η συνάρτηση «βαθμίδα αποτυχίας» είναι καθοριστικής σημασίας για την μελέτη των φαινομένων στα πλαίσια της επιβίωσης. Η «βαθμίδα αποτυχίας» ή διαφορετικά το «ποσοστό αποτυχίας» είναι μια συνάρτηση η οποία εκφράζει τον ρυθμό, με τον οποίο ένα σύστημα θα αποτύχει. Η αποτυχία αυτή είναι εκφρασμένη ανά μονάδα χρόνου, αλλιώς εκφράζει διαισθητικά την πιθανότητα αποτυχίας ενός συστήματος ανά μονάδα χρόνου δεδομένου ότι έχει ζήσει κάποια χρόνια  $t$ . Ο συμβολισμός αυτής είναι το ελληνικό γράμμα  $\lambda(t)$  και χρησιμοποιείται σε πολλούς κλάδους της σύγχρονης επιστήμης. Το «ποσοστό αποτυχίας» ενός συστήματος εξαρτάται από το χρόνο με το ρυθμό μεταβολής κατά την διάρκεια του κύκλου ζωής ενός συστήματος.

Πιο αναλυτικά, ένας διαφορετικός στατιστικός ορισμός της «βαθμίδα αποτυχίας» είναι ο συνολικός αριθμός αποτυχιών του πληθυσμού, διαιρούμενος με το συνολικό χρόνο που δαπανάται από τον εν λόγω πληθυσμό σε ένα χρονικό διάστημα βάσει ορισμένων συνθηκών.

Η βαθμίδα αποτυχίας έχει πολλαπλές ονομασίες ανάλογα τον κλάδο με τον οποίο ασχολούμαστε. Πιο συγκεκριμένα, αναφέρουμε ότι στην δημογραφία και στην ασφαλιστική επιστήμη ονομάζεται «ένταση θνησιμότητας» (force of mortality), στην επιδημιολογία ορίζεται ως "age specific failure rate" ενώ στη θεωρία αξιοπιστίας (reliability theory) ως «failure rate» ή «hazard rate».

Αν θεωρήσουμε μια συνεχή τυχαία μεταβλητή η οποία είναι θετική και την συμβολίσουμε με  $T$  ( $T > 0$ ), εφόσον αυτή η μεταβλητή παριστάνει τον χρόνο ζωής μιας μονάδας ή ενός συστήματος τότε έχει ως συνάρτηση κατανομής μια συνάρτηση  $F(t)$  και επιπροσθέτως μια συνάρτηση πυκνότητας πιθανότητάς που συμβολίζεται ως  $f(t)$ .

Πιο αναλυτικά ισχύει το εξής:



$F(t) = P[T \leq t]$  και  $f(t) = F'(t)$  με  $F(-\infty) = F(0) = 0$ ,  $\lim_{t \rightarrow \infty} F(t) = 1$  και η  $F(t)$  αύξουσα συνάρτηση και συνεχείς από δεξιά .

Αν υποθέσουμε πως ένα σύστημα ή μια μηχανή που έχει τεθεί σε λειτουργία δεν έχει αποτύχει για το χρονικό διάστημα  $[0, t]$ , τότε η πιθανότητα το εν λόγω σύστημα να μην αποτύχει για  $z$  επιπλέον χρονικές στιγμές μετά την χρονική στιγμή  $t$ , ισούται με την πιθανότητα να λειτουργήσει στο διάστημα  $(t, t+z]$  με  $z \geq 0$ , η οποία δίνεται από τη σχέση

$$P[T > t + z / T > t] = \frac{P[T > t+z]}{P[T > t]} = \frac{S(t+z)}{S(t)} \text{ όπου } S(t) > 0.$$

Αντίστοιχα η δεσμευμένη πιθανότητα αποτυχίας του εν λόγω συστήματος στο ίδιο διάστημα  $(t, t+z]$  θα είναι ίση με

$$P[T \leq t + z / T > t] = \frac{P[t < T \leq t + z]}{P[T > t]} = \frac{S(t) - S(t + z)}{S(t)}$$

Όπου  $S(t)$  ορίζεται ως η συνάρτηση επιβίωσης ή συνάρτηση αξιοπιστίας δηλαδή η πιθανότητα το σύστημα να επιζήσει μέχρι την χρονική στιγμή  $t$ .

Ορίζουμε ως βαθμίδα αποτυχίας (failure rate) το λόγο που εκφράζει τον δεσμευμένο ρυθμό αποτυχίας τη μονάδας στο διάστημα  $(t, t+z)$  όταν  $z \rightarrow 0$  δοθέντος  $T > t$ .

$$\lambda(t) = -\frac{S'(t)}{S(t)} = \lim_{z \rightarrow 0} \frac{P[T \leq t+z | T > t]}{z} = (-\ln S(t))' = \frac{f(t)}{S(t)}, \quad t \geq 0 \quad (2.1.1)$$

Εδώ η  $f$  είναι η συνάρτηση πυκνότητας πιθανότητας της τ.μ  $T$  και  $S'(t)$  η παράγωγος της συνάρτησης αξιοπιστίας για την οποία ισχύει η σχέση  $S'(t) = (1-F(t))' = -F'(t) = -f(t)$ . Η βαθμίδα αποτυχίας, μας δείχνει την πιθανότητα ένα σύστημα να αποτύχει σε ένα πολύ μικρό διάστημα  $z$  η οποία στιγμή βρίσκεται αμέσως μετά την χρονική στιγμή  $t$ . Στην αναλογιστική επιστήμη, την συνάρτηση κινδύνου την ονομάζουμε επίσης «ένταση ή ισχύ θνησιμότητας» (force of mortality), για άτομα ηλικίας ακριβώς  $x$ , και την συμβολίζουμε ως  $\lambda(t) = \mu(t)$ . Από τον παραπάνω ορισμό και τον τύπο που έχει δοθεί καταλήγουμε στον τύπο της έντασης θνησιμότητας

$$\lambda(t) = \frac{f(t)}{P(T > t)} = \frac{f(t)}{S(t)}, \quad (2.1.2)$$

η οποία καλείται (στιγμιαία) βαθμίδα αποτυχίας ή **ένταση αποτυχίας ή ένταση θνησιμότητας (failure rate, intensity of failures, hazard rate, force of mortality)** και βάσει του τύπου αυτού της έντασης θνησιμότητας εφόσον

$$\lambda(t) = -\frac{\left(\frac{d}{dt}\right)S(t)}{S(t)} \text{ ή } \lambda(t) = -\frac{d}{dt} \ln(S(t)) \quad (2.1.3)$$

ολοκληρώνοντας καταλήγουμε στον τύπο της συνάρτησης επιβίωσης

$$S(t) = e^{-\int_0^t \lambda(y) dy}. \quad (2.1.4)$$

Η συνάρτηση επιβίωσης συνδέεται με την αθροιστική συνάρτηση κατανομής  $F(t)$  σύμφωνα με την γνωστή σχέση

$$S(t) = P[T > t] = 1 - P[T \leq t] = 1 - F(t),$$

$$S(t) + F(t) = 1,$$

όπου η συνάρτηση επιβίωσης μιας κατανομής είναι πάντα μια φθίνουσα συνάρτηση στο διάστημα  $[0, \infty)$ .

Παρατηρούμε ότι η «ένταση θνησιμότητας» έχει ερμηνεία δεσμευμένης πυκνότητας πιθανότητας. Η συνάρτηση κινδύνου είναι ρυθμός και όχι πιθανότητα γι' αυτό μπορεί και να πάρει τιμές μεγαλύτερες της μονάδας. Άρα βάσει αυτού συμπεραίνουμε πως δεν είναι στην πραγματικότητα μια πιθανότητα διότι μπορεί να ξεπερνάει η τιμή της το 1. Επίσης μια εσφαλμένη έκφραση του μέτρου, μπορεί να δώσει λανθασμένη αντίληψη του αυτού και μη σταθερά εκτιμώμενα ποσοστά.

Όσον αφορά την μονοτονία της έντασης θνησιμότητας ή αλλιώς της βαθμίδα αποτυχίας, αν φθίνει ή όχι εξαρτάται από πολλούς παράγοντες. Αρχικά θα ορίσουμε την αύξουσα και την φθίνουσα βαθμίδα αποτυχίας .

- Ορίζουμε ότι μια τυχαία μεταβλητή είναι IFR (Increasing Failure Rate) αν η αντίστοιχη βαθμίδα αποτυχίας  $\lambda(t)$  είναι αύξουσα συνάρτηση του  $t$ .
- Αναλόγως ορίζουμε μια τυχαία μεταβλητή ως DFR ( Decreasing Failure Rate ) αν η αντίστοιχη βαθμίδα αποτυχίας  $\lambda(t)$  είναι φθίνουσα συνάρτηση του  $t$ .

Υπάρχουν φυσικά περιπτώσεις όπου η βαθμίδα αποτυχίας δεν είναι ούτε αύξουσα ούτε φθίνουσα σε όλο το πεδίο τιμών της τ.μ.  $T$ . Στην περίπτωση αυτή τότε το σύστημα που έχει τεθεί σε λειτουργία περνά από κάποια στάδια. Πιο συγκεκριμένα , το πρώτο διάστημα μπορεί να χαρακτηριστεί ως αρχική περίοδος που ξεκινά με μεγάλη βαθμίδα αποτυχίας που σταδιακά αρχίζει και αυξάνεται με το πέρασμα του χρόνου λόγω φθοράς. Εφόσον το σύστημά μας συνεχίζει και βρίσκεται σε λειτουργία μετά την αρχική περίοδο ακολουθεί η χρήσιμη περίοδος λειτουργίας του συστήματος με σταθερή βαθμίδα αποτυχίας που αν σε αυτή την χρονική περίοδο το σύστημα αποτύχει λέγεται ότι έχει υποστεί τυχαία αποτυχία. Εάν περάσει το σύστημα τα πρώτα διαστήματα τότε περνάει στο χρονικό διάστημα φθοράς όπου η βαθμίδα αποτυχίας είναι μεγάλη όπως είναι λογικό αφού οι αποτυχίες λειτουργίας του συστήματος αυξάνονται ολοένα και περισσότερο και λόγω φθοράς του.

Πολλές φορές μπορεί να συναντήσουμε το γεγονός η βαθμίδα αποτυχίας να παίρνει σταθερή τιμή π.χ.  $\lambda(t)=\lambda>0, t \geq 0$  και τότε είναι και IFR και DFR. Όταν συμβαίνει αυτό, ο χρόνος ζωής μπορεί τετριμμένα να θεωρηθεί IFR και DFR ταυτόχρονα. Σε αυτή την περίπτωση μπορούμε να πούμε ότι η μονάδα παραμένει «αγέραστη» (άφθαρτη) διότι

όσο και αν περνά ο χρόνος, η πιθανότητα να αποτύχει (δεδομένου ότι λειτουργεί) παραμένει σταθερή. Αυτό ισχύει από τις συνεχείς κατανομές μόνο για την εκθετική κατανομή. Αυτό δεν σημαίνει καθόλου ότι η μονάδα είναι «αθάνατη» (αυτό συμβαίνει μόνο στην ιδιόζουσα περίπτωση  $\lambda(t)=0$ )

**(Μπούτσικας 2008)**

### **Παράδειγμα 2.1**

Αν  $\lambda(t) = \lambda$ , η αξιοπιστία της μονάδας θα είναι ίση με:

$$S(t) = e^{-\int_0^t \lambda(s) ds} = e^{-\int_0^t \lambda ds} = e^{-\lambda t}$$

και επομένως η τ.μ.  $T$  θα έχει συνάρτηση κατανομής την ακόλουθη

$$F(t) = 1 - S(t) = 1 - e^{-\lambda t}, t \geq 0.$$

Επομένως ότι η αντίστοιχη συνάρτηση πυκνότητας πιθανότητας θα είναι η

$$f(t) = \lambda e^{-\lambda t}, t \geq 0.$$

Άρα δείξαμε ότι χρόνος ζωής μιας μονάδας (με  $\lambda(t) = \lambda, t \geq 0$ ) ακολουθεί εκθετική κατανομή με παράμετρο  $\lambda$ .

Αντίστροφα, αν μία μονάδα έχει χρόνο ζωής που ακολουθεί εκθετική κατανομή, τότε η μονάδα αυτή μπορεί να θεωρηθεί «αγέραστη» δηλαδή η πιθανότητα αποτυχίας της, δεδομένου ότι λειτουργεί, παραμένει σταθερή όσος χρόνος και αν έχει περάσει από την έναρξη λειτουργίας της. Η εκθετική κατανομή τετριμμένα είναι IFR και DFR ταυτόχρονα (ακριβέστερα, ένας χρόνος ζωής  $T$  που ακολουθεί την εκθετική κατανομή είναι και IFR και DFR).

### **Εφαρμογή 1.2:**

Όπως θα δούμε αργότερα από την παράγραφο 2.7 όπου παρουσιάζεται η κατανομή Rayleigh αναλύονται όλες οι συνάρτησης της κατανομής και έτσι έχουμε δυνατότητα να βρούμε την βαθμίδα αποτυχίας (failure rate). Πιο συγκεκριμένα έχουμε την συνάρτηση κατανομής και συνάρτηση επιβίωσης αντίστοιχα

$$F(x) = 1 - (e^{-\frac{x^2}{2\sigma^2}}), \text{ για κάθε } x \in (0, \infty)$$

$$S(x) = (e^{-\frac{x^2}{2\sigma^2}}), \text{ για κάθε } x \in (0, \infty)$$

Επίσης η συνάρτηση πυκνότητας πιθανότητας της Rayleigh περιγράφεται από τον τύπο

$$f(x) = \frac{x}{\sigma^2} * e^{-\left(\frac{x^2}{2\sigma^2}\right)}$$

Έτσι έχουμε την δυνατότητα να υπολογίσουμε την βαθμίδα αποτυχίας .

Άρα, βάσει του τύπου της έντασης θνησιμότητας ή βαθμίδας αποτυχίας :

$$\lambda(t) = -\frac{S'(t)}{S(t)} = \lim_{z \rightarrow \infty} \frac{P[T \leq t+z | T > t]}{z} = (-\ln S(t))' = \frac{f(t)}{S(t)}, \quad t \geq 0$$

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\left(\frac{t}{\sigma^2} \left( e^{-\left(\frac{t^2}{2\sigma^2}\right)} \right)\right)}{e^{-\left(\frac{t^2}{2\sigma^2}\right)}} = \frac{t}{\sigma^2}, \quad t \geq 0$$

**Σχόλιο:** Η βαθμίδα αποτυχίας είναι αύξουσα και γραμμική συνάρτηση του χρόνου.

### Εφαρμογή 2.3:

Ας δούμε τώρα μια εφαρμογή όπου η βαθμίδα αποτυχίας είναι της γενικότερης μορφής

$$\lambda(t) = cb^t$$

Αν υπάρχει μονάδα με τέτοια βαθμίδα αποτυχίας θα έχει αξιοπιστία της μορφής

$$S(t) = \exp\left(-\int_0^t \lambda(s) ds\right) = \exp\left(-\int_0^t cb^s ds\right) = \exp\left(\frac{-ct^{b+1}}{b+1}\right)$$

Και η συνάρτηση κατανομής της μορφής  $F(t) = 1 - S(t) = 1 - \exp\left(\frac{-c*t^{b+1}}{b+1}\right)$ .

Η συνάρτηση αυτή μπορεί πράγματι να θεωρηθεί συνάρτηση κατανομής χρόνου ζωής όταν είναι  $c > 0$  και  $b > -1$  (ώστε  $F$ : αύξουσα με  $F(0) = 0$ ,  $F(\infty) = 1$ ). Η συγκεκριμένη κατανομή, η οποία είναι γνωστή ως κατανομή Weibull, είναι αρκετά σημαντική στην θεωρία αξιοπιστίας διότι αντιστοιχεί σε βαθμίδα αποτυχίας με πολύ απλή και γενική μορφή που εμφανίζεται αρκετές φορές στην πράξη. Αξίζει να παρατηρήσουμε ότι όταν  $b > 0$  τότε η  $T \in \text{IFR}$  ενώ όταν  $-1 < b < 0$  τότε  $T \in \text{DFR}$ . Αν  $b = 0$  τότε προκύπτει και πάλι η εκθετική κατανομή η οποία μπορεί να θεωρηθεί ως υπό- περίπτωση της κατανομής Weibull. Θα ασχοληθούμε περισσότερο με αυτή την κατανομή σε επόμενο κεφάλαιο όταν εξετάσουμε αναλυτικά αρκετές κατανομές χρόνων ζωής. Συνήθως ως Weibull ορίζεται η κατανομή με σ.κ.  $F(t) = 1 - e^{-(\lambda t)^a}$ ,  $\lambda > 0, a > 0$  η οποία προφανώς είναι η ίδια με την παραπάνω αν θέσουμε  $a = b+1$ ,  $\lambda = \left[\frac{c}{b+1}\right]^{\frac{1}{b+1}}$  **(Μπούτσικας 2011)**

## 2.2 Ο υπολειπόμενος χρόνος ζωής

Ο «υπολειπόμενος χρόνος ζωής» (Residual Life) είναι μία σημαντική έννοια στην θεωρία αξιοπιστίας, στη στατιστική, στην ανάλυση επιβίωσης αλλά και αλλού. Πολλά χρήσιμα αποτελέσματα έχουν προκύψει από την μελέτη αυτής της ποσότητας. Έστω μια μονάδα ηλικίας  $t > 0$  της οποίας ο χρόνος περιγράφεται από τη μη αρνητική τυχαία μεταβλητή  $T$ . Ο χρόνος που απομένει μέχρι να καταστραφεί η μονάδα δίνεται από την τυχαία μεταβλητή  $T-t$ . Με τον όρο «υπολειπόμενος χρόνος ζωής» μια μονάδας με ηλικία  $t$  αναφερόμαστε στην δεσμευμένη τυχαία μεταβλητή

$$T - t/T > t$$

Η μέση τιμή αυτής της τυχαίας μεταβλητής ονομάζεται αναμενόμενος υπολειπόμενος χρόνος ζωής και συμβολίζεται με MRL (Mean Residual Life). Η συνάρτηση αυτή αποτελεί συχνά το βασικό κριτήριο στην εύρεση ενός βέλτιστου χρονικού σημείου για την αντικατάσταση εξαρτημάτων για παράδειγμα.

Έστω  $S(x)$  μια συνάρτηση αξιοπιστίας μιας συνεχούς τ.μ.  $X$  ορισμένης στο  $[0, \infty)$  με πεπερασμένη πρώτη ροπή  $\mu$  και  $F(0) = 0$ . Τότε η συνάρτηση του υπολειπόμενου χρόνου ζωής της τ.μ.  $X$  δίνεται από τη σχέση

$$\mu(x) = E(X - x | X > x) = \frac{\int_x^\infty S(t) dt}{S(x)}, x \geq 0 \quad (2.2.1)$$

Προφανώς  $\mu(0) = \mu = E(X)$ . Αν επιπλέον η  $F$  έχει πυκνότητα έστω  $f$ , τότε μπορούμε να γράψουμε εναλλακτικά

$$\mu(x) = \frac{\int_x^\infty t f(t) dt}{S(x)} - x \quad (2.2.2)$$

Στην μελέτη για αντικατάσταση βιομηχανικών μηχανημάτων η συνάρτηση MRL μπορεί να φανεί πολύ πιο χρήσιμη από τη συνάρτηση FR. Η πρώτη συνοψίζει ολόκληρο τον αναμενόμενο υπολειπόμενο χρόνο ζωής ενός αντικειμένου ενώ η δεύτερη αναφέρεται στον κίνδυνο στιγμιαίας καταστροφής. Στην μελέτη ανθρώπινης διάρκειας ζωής, η συνάρτηση MRL μπορεί να αναφέρεται και σαν προσδόκιμο ζωής. Προφανώς η μελέτη αυτής της συνάρτησης είναι ζωτικής σημασίας και στην αναλογιστική επιστήμη.

Συγκεκριμένα, για ένα άτομο ηλικίας  $x$ , η τιμή  $\mu(x)$  εκφράζει τον υπολειπόμενο προσδοκώμενο χρόνο ζωής για το άτομο αυτό.

## 2.3 Η εκθετική κατανομή (Exponential Distribution)

Στην θεωρία πιθανοτήτων και στην επιστήμη της στατιστικής η εκθετική κατανομή ή διαφορετικά αρνητική εκθετική κατανομή ορίζεται ως μια οικογένεια συνεχών κατανομών πιθανότητας. Η κατανομή αυτή περιγράφει ουσιαστικά τον χρόνο μεταξύ

γεγονότων σε μια διαδικασία Poisson , δηλαδή μια διαδικασία που συμβαίνουν συνεχώς γεγονότα ανεξάρτητα με σταθερό μέσο ρυθμό . Η εκθετική κατανομή ανήκει στην ευρύτερη εκθετική οικογένεια, η οποία είναι μια κλάση κατανομών πιθανότητας που περιλαμβάνει ακόμα την κανονική κατανομή, την Διωνυμική κατανομή, την Gamma, την Poisson , και άλλες.

Η συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής με παράμετρο  $\theta$  είναι:

$$f(x) = \theta e^{-\theta x}, \quad x \geq 0 (\lambda > 0) \quad (2.3.1)$$

Η συνεχής κατανομή της εκθετικής κατανομής θα είναι :

$$F(x) = \int_{-\infty}^x f(t)dt = \int_0^x \theta e^{-\theta t} dt = 1 - e^{-\theta x}, \quad \theta \geq 0, \quad (2.3.2)$$

Ενώ  $F(x) = 0$  αν  $x < 0$ .

Η μέση τιμή και η διασπορά της εκθετικής κατανομής θα είναι:

$$E(X) = \frac{1}{\theta} \quad (2.3.3)$$

$$Var(X) = \frac{1}{\theta^2} \quad (2.3.4)$$

Τέλος η συνάρτηση επιβίωσης (survival function) ισούται με:

$$S(x) = e^{-\theta x}, \theta \geq 0 \quad (2.3.5)$$

Αναλυτικότερα η μέση τιμή της εκθετικής κατανομής αλλά και η διακύμανση προκύπτει με ολοκλήρωση κατά παράγοντες .Συγκεκριμένα έχουμε,

$$E(X) = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \theta e^{-\theta x} = 1 / \theta ,$$

$$E(X^2) = \int_0^{\infty} x^2 f(x) dx = \int_0^{\infty} x^2 \theta e^{-\theta x} = 2 / \theta^2 .$$

Άρα βάσει του γνωστού τύπου της διακύμανσης  $Var(X) = E(X^2) - E^2(X)$  προκύπτει με αντικατάσταση ότι η διακύμανση της εκθετικής κατανομής είναι

$$Var(X) = \frac{2}{\theta^2} - \left(\frac{1}{\theta}\right)^2 \Rightarrow Var(X) = \frac{1}{\theta^2} \quad (2.3.6)$$

Η ροπογεννήτρια συνάρτηση μίας συνεχούς τυχαιάς μεταβλητής  $X$  έχουμε δει ότι υπολογίζεται από την σχέση:

$$M_x(t) = E(e^{tx})$$

Επομένως για την ροπογεννήτρια της Εκθετικής κατανομής με παράμετρο  $\lambda$  θα ισχύει:

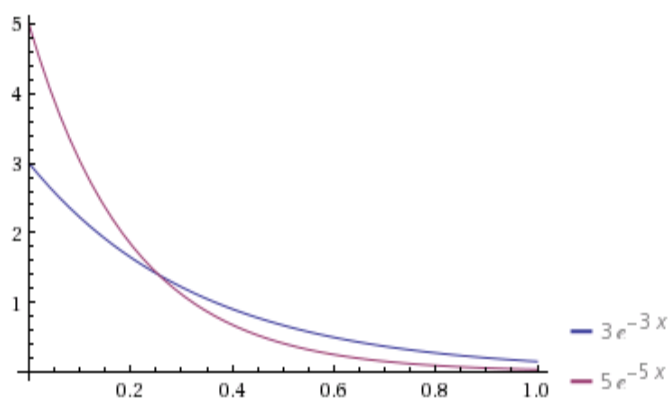
$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \Rightarrow$$

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} \theta e^{-\theta x} dx = \theta \int_{-\infty}^{\infty} e^{-(\theta-x)x} dx = \frac{\theta}{\theta-t} \quad (2.3.7)$$

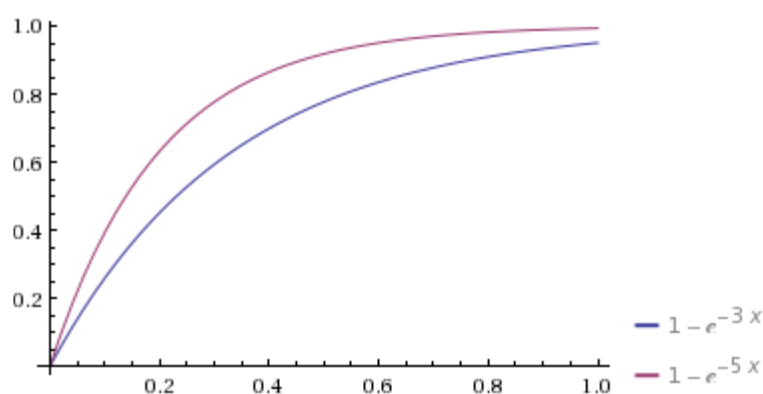
Η οποία είναι πεπερασμένη για  $t < \theta$ .

### Οι γραφικές παραστάσεις για την εκθετική κατανομή

Στο παρακάτω πρώτο σχήμα βλέπουμε τις γραφικές παραστάσεις των συναρτήσεων πυκνότητας πιθανότητας εκθετικής κατανομής με παραμέτρους 3 και 5, ενώ στο δεύτερο παρατηρούμε τις συναρτήσεις κατανομής τους

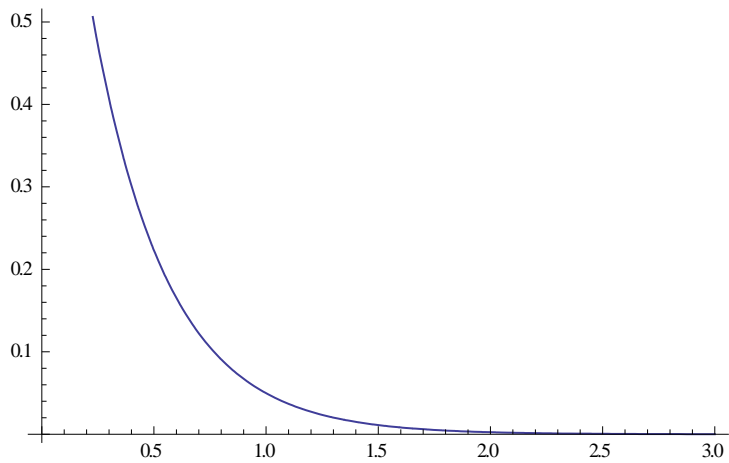


**Σχήμα 2.3.1 :** Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας  $f(x)$  για  $\theta=3$  (μπλε χρώμα),  $\theta=5$  (ροζ χρώμα).

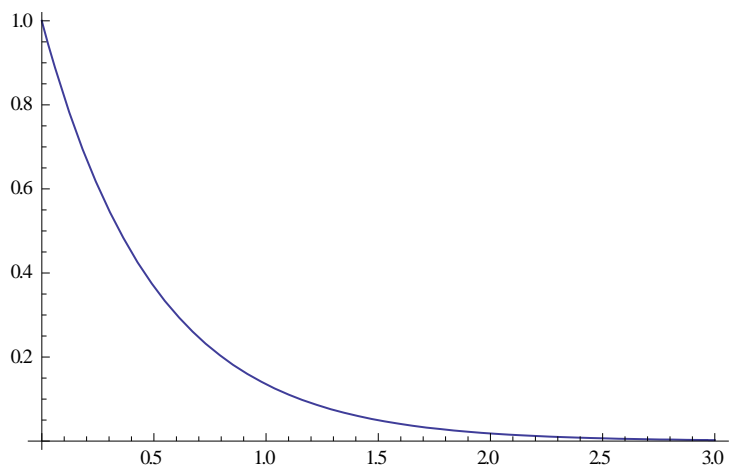


**Σχήμα 2.3.2:** Διάγραμμα της συνάρτησης κατανομής  $F(x)$  για  $\theta=3$  (μπλε χρώμα)  $\theta=5$  (ροζ χρώμα).

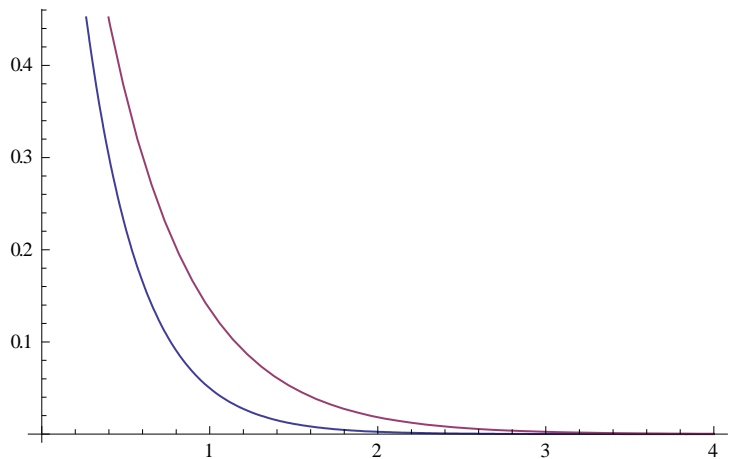
Έπειτα στα παρακάτω δυο σχήματα παρουσιάζονται ξεχωριστά οι γραφικές παραστάσεις της συνάρτησης επιβίωσης εκθετικής κατανομής. Αρχικά στην πρώτη γραφική παράσταση βλέπουμε την συνάρτηση επιβίωσης με παράμετρο  $\lambda=3$  ενώ στην δεύτερη για παράμετρο  $\theta=5$ .



**Σχήμα 2.3.3 :** Διάγραμμα της συνάρτησης επιβίωσης  $S(x)$  με παράμετρο  $\theta=3$ .



**Σχήμα 2.3.4:** Διάγραμμα της συνάρτησης επιβίωσης  $S(x)$  με παράμετρο  $\theta=2$ .



**Σχήμα 2.3.5:** Γραφική παράσταση συνάρτησης επιβίωσης με παραμέτρους  $\theta=2$  (κόκκινο) και  $\theta=3$  (μπλε).

**Σχολιασμός Σχεδιαγραμμάτων:** Παρατηρούμε πως η συνάρτηση επιβίωσης είναι μια γνησίως φθίνουσα συνάρτηση όπου για την παράμετρο  $\lambda=2$  συγκλίνει πιο αργά στο μηδέν ενώ για  $\lambda=3$  συγκλίνει πιο γρήγορα στο μηδέν έχοντας πιο ελαφριά ουρά.



## Υπολογισμός βαθμίδας αποτυχίας για την εκθετική κατανομή

Γνωρίζουμε από την παράγραφο 2.1 όπως έχει ορισθεί ήδη πως για την βαθμίδα αποτυχίας ισχύει ότι

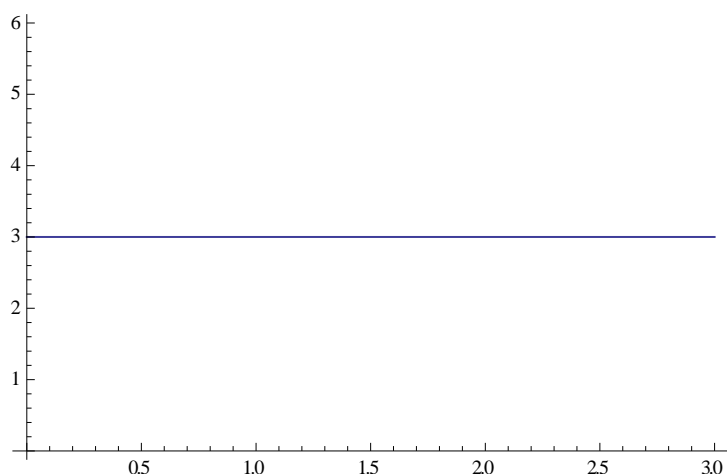
$$\lambda(t) = -\frac{S'(t)}{S(t)} = \lim_{z \rightarrow \infty} \frac{P[T \leq t+z \mid T > t]}{z} = (-\ln S(t))' = \frac{f(t)}{S(t)}, \quad t \geq 0$$

Έτσι από τις σχέσεις (2.3.1), (2.3.2), (2.3.5) καταλήγουμε με απλές πράξεις ότι :

$$\lambda(t) = \frac{F'(t)}{S(t)} = \theta * \frac{e^{-\theta t}}{e^{-\theta t}} = \theta$$

Άρα, αναφέρουμε η βαθμίδα αποτυχίας, δεδομένου ότι ένα σύστημα λειτουργεί, παραμένει σταθερή όσος χρόνος και αν έχει περάσει από την έναρξη λειτουργίας της. Η εκθετική κατανομή τετριμμένα είναι IFR και DFR ταυτόχρονα (ακριβέστερα, ένας χρόνος ζωής  $T$  που ακολουθεί την εκθετική κατανομή είναι και IFR και DFR).

Παρακάτω παρατηρούμε το σχεδιάγραμμα της βαθμίδας αξιοπιστίας για την εκθετική κατανομή για παράμετρο  $\theta=3$ .



**Σχήμα 2.3.6:** Σχεδιάγραμμα βαθμίδας αποτυχίας εκθετικής κατανομής με  $\theta=3$ .

Εύλογα παρατηρούμε πως η βαθμίδα αποτυχίας παραμένει σταθερή σε όλο το χρονικό διάστημα για όλη την διάρκεια λειτουργίας του συστήματος.

## 2.4 Η κατανομή Γάμμα (Gamma Distribution)

Στην θεωρία πιθανοτήτων και στην στατιστική επιστήμη όπως και στον αναλογισμό χρησιμοποιείται εκτενώς «η κατανομή Γάμμα» η οποία είναι συνεχής κατανομή με 2 παραμέτρους. Όσον αφορά την επιστήμη του αναλογισμού, η κατανομή αυτή χρησιμοποιείται για την περιγραφή των συνολικών ζημιών. Εκτενέστερα οι κατανομές Γάμμα διαδραματίζουν εξέχοντα ρόλο στην αναλογιστική επιστήμη. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι οι περισσότερες συνολικές κατανομές ασφαλιστικών αποζημιώσεων έχουν περίπου το ίδιο σχήμα όπως οι κατανομές Γάμμα : είναι στραμμένες προς τα δεξιά, είναι μη αρνητικές και μονότροπες. Επιπλέον, οι κατανομές Γάμμα είναι

καλά μελετημένες και έχουν υποστεί αρκετή ανάλυση. Σαν αποτέλεσμα υπάρχουν πολλά παραδείγματα εφαρμογής της κατανομής Γάμμα για την μοντελοποίηση ασφαλιστικών χαρτοφυλακίων π.χ. Hurlimann (2001), Melnick και Tenenbein (2000) και Rioux και Klugman (2004). Επίσης οι Herzog (1999) και Hossack et al. (1983) σημειώνουν ότι οι κατανομές Γάμμα παρέχουν ένα βολικό μοντέλο για το μέσο ποσοστό των αξιώσεων που υποβάλλονται από διάφορους ασφαλισμένους της ασφαλιστικής εταιρίας. Ο Bowers et al. (1997) χρησιμοποίησε μεταφρασμένες(translated) κατανομές Γάμμα ως μοντέλο για τις συνολικές ασφαλιστικές απαιτήσεις. Η συνηθισμένη εκθετική κατανομή που αναλύσαμε στην προηγούμενη παράγραφο αλλά και η  $\chi$ - τετράγωνο (chi-square) είναι ειδικές περιπτώσεις της κατανομής Γάμμα.

Η κατανομή Γάμμα με παραμέτρους  $(\alpha, \theta)$  έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{\theta^\alpha}{\Gamma(\alpha)} * x^{\alpha-1} * e^{-\theta x} \text{ για } x \geq 0, \quad (2.4.1)$$

όπου  $\theta > 0$  και  $\alpha > 0$  η παράμετρος κλίμακας (scale parameter) και παράμετρος μορφής (shape parameter) αντίστοιχα.

Εναλλακτικά χρησιμοποιώντας διαφορετική παραμετροποίηση η συνάρτηση πυκνότητας πιθανότητας μπορεί να οριστεί από τον τύπο

$$f(x) = x^{\theta-1} * \frac{e^{-\frac{x}{\theta}}}{\theta^\alpha * \Gamma(\alpha)}, \quad x > 0, \text{ όπου } \alpha, \theta > 0$$

Αν τις δυο παραπάνω σχέσεις θέσουμε  $\alpha=1$  και  $\lambda=\theta$  τότε παίρνουμε την εκθετική συνάρτηση πυκνότητας πιθανότητας.

Η  $\Gamma(\alpha)$  είναι η γνωστή ως Γάμμα συνάρτηση η οποία ορίζεται ως:

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, x > 0$  όπου  $\alpha > 0$  και ικανοποιεί την σχέση  $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ . Αν  $\alpha$  είναι θετικός ακέραιος τότε έχουμε  $\Gamma(\alpha) = (\alpha-1)!$

Η συνάρτηση κατανομής της εν λόγω κατανομής υπολογίζεται από τον τύπο

$$F(x) = \int_0^x f(t) dt \quad (2.4.2)$$

Η συνάρτηση της μέσης τιμής για την σχέση (2.4.1) υπολογίζεται από τον τύπο

$$E(X) = \alpha/\theta \quad (2.4.3)$$

και η δεύτερη ροπή της κατανομής Γάμμα μπορεί εύκολα να βρεθεί ότι ισούται με

$$E(X^2) = \frac{\alpha(\alpha+1)}{\theta^2} \quad (2.4.4)$$

Άρα βάσει του τύπου  $Var(X) = E^2(X) - E(X^2)$ , προκύπτει εύκολα με αντικατάσταση ότι η διακύμανση της κατανομής υπολογίζεται από τον τύπο

$$V(X) = \frac{\alpha}{\theta^2}. \quad (2.4.5)$$

Η συνάρτηση επιβίωσης της κατανομής είναι ίση με

$$S(x) = \int_x^{\infty} f(t)dt = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_x^{\infty} y^{\alpha-1} e^{-\theta y} dy = \frac{\Gamma(\alpha, \theta x)}{\Gamma(\alpha)}$$

που με την συνάρτηση  $\Gamma(\alpha, \beta) = \int_x^{\infty} x^{\alpha-1} e^{-x} dx$  συμβολίζουμε την μη πλήρη συνάρτηση Γάμμα.

Στην περίπτωση που η παράμετρος  $\alpha$  είναι ένας θετικός ακέραιος τότε η κατανομή Γάμμα ταυτίζεται με την κατανομή του αθροίσματος  $\alpha$  ανεξάρτητων, εκθετικών τυχαίων μεταβλητών και έχουν βαθμίδα αποτυχίας ίση με  $\theta$

. Έτσι προκύπτει η κατανομή Erlang η οποία δημιουργήθηκε από τον Δανό μαθηματικό A.K Erlang στην προσπάθεια του να εξετάσει το τηλεπικοινωνιακό σύστημα της πατρίδας του προκειμένου να καταγράψει έναν μικρό αριθμό τηλεφωνημάτων που είναι σε αναμονή για ένα πολύ μικρό χρονικό διάστημα μέχρι να απαντηθούν. Πλέον η κατανομή χρησιμοποιείται στις στοχαστικές διαδικασίες και σε πολλούς κλάδους των εφαρμοσμένων πιθανοτήτων.

Για την κατανομή Erlang η συνάρτηση πυκνότητας δίνεται από τον τύπο

$$f(x) = \frac{\theta^\alpha}{(\alpha-1)!} x^{\alpha-1} e^{-\theta x}, x > 0, \alpha \in \mathbb{N} \quad (2.4.6)$$

Η συνάρτηση επιβίωσης υπολογίζεται από τον τύπο

$$S(x) = e^{-\theta x} \sum_{n=0}^{\alpha-1} \frac{(\theta x)^n}{n!} e^{-\theta x} \quad x \geq 0, \alpha \in \mathbb{N} \quad (2.4.7)$$

Η ροπογεννήτρια συνάρτηση μίας συνεχούς τυχαίας μεταβλητής  $X$  έχουμε δει ότι υπολογίζεται από την σχέση:

$$M_x(t) = E(e^{tx})$$

Επομένως για την ροπογεννήτρια της κατανομής Erlang με παράμετρο  $\theta > 0, \alpha \in \mathbb{N}$  θα ισχύει:

$$\begin{aligned} M_x(t) &= E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \Rightarrow \\ M_x(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{\theta^\alpha}{\Gamma(\alpha)} * x^{\alpha-1} * e^{-\theta x} dx = \\ &= \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \int_{-\infty}^{\infty} e^{-(x-t)} = \frac{\theta}{(\theta-t)^\alpha}, \quad t < \frac{1}{\theta} \end{aligned}$$

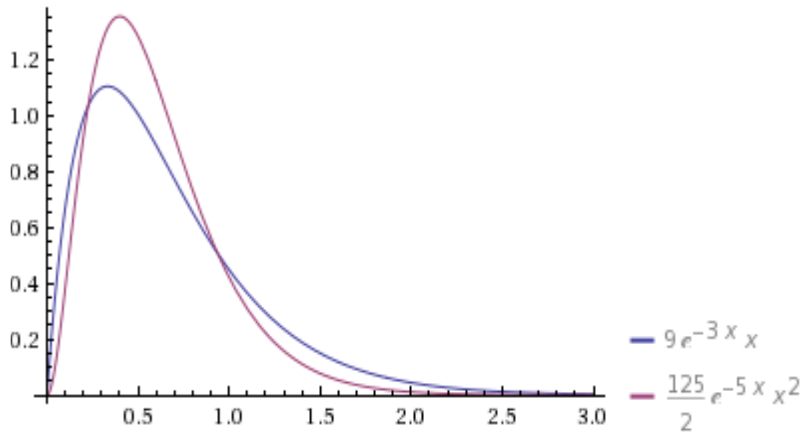
Παρακάτω παρουσιάζονται οι γραφικές παραστάσεις για την συνάρτηση πυκνότητας, συνάρτησης κατανομής και συνάρτησης επιβίωσης της Γάμμα κατανομής με παραμέτρους  $\alpha=2$  και  $\theta=2$ .

<https://eclass.uoa.gr/modules/document/file.php/MATH265/Lecture14.pdf>

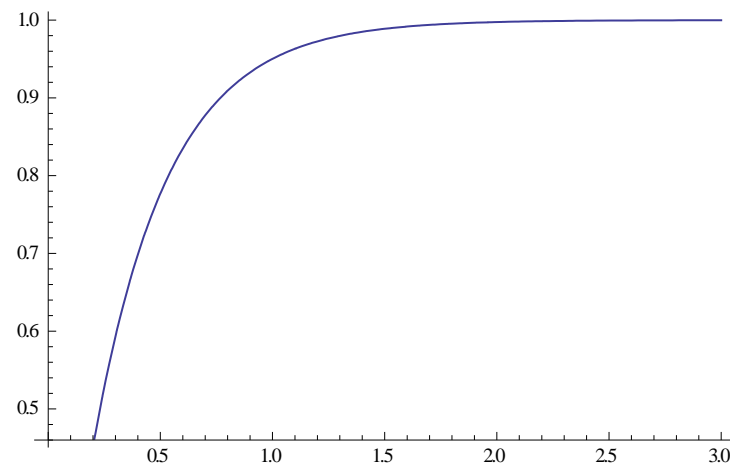
(Χαράλαμπος Α. Χαραλαμπίδης 2009)

### Γραφικές παραστάσεις της εν λόγω κατανομής

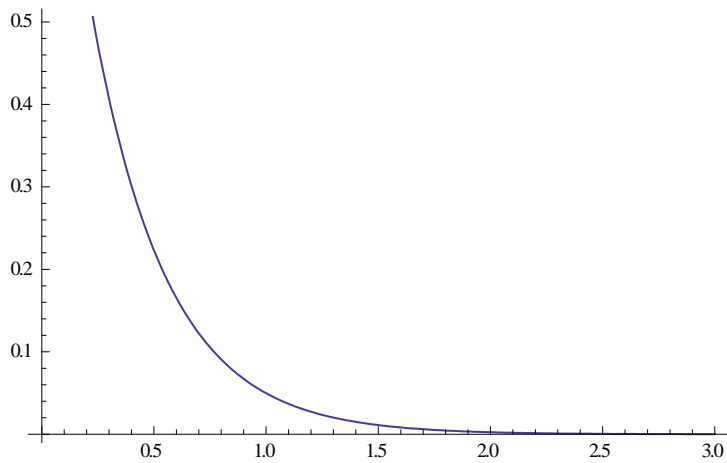
Παρακάτω παρουσιάζονται οι γραφικές παραστάσεις για την συνάρτηση πυκνότητας, συνάρτησης κατανομής και συνάρτησης επιβίωσης της Γάμμα κατανομής με παραμέτρους  $\alpha=2$  και  $\theta=3$  και  $\alpha=3$  και  $\theta=5$



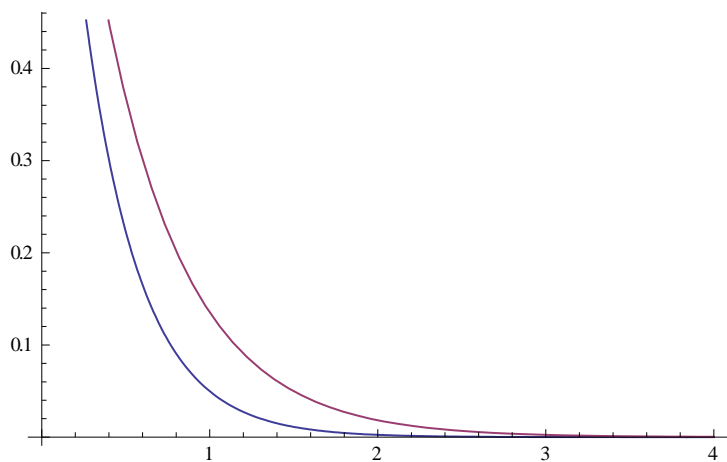
**Σχήμα 2.4.1:** Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας για τις παραμέτρους  $\alpha=2$  και  $\theta=3$  και  $\alpha=3$  και  $\theta=5$ .



**Σχήμα 2.4.2:** Η γραφική παράσταση της συνάρτησης κατανομής με παραμέτρους  $\alpha=2$  και  $\theta=2$ .



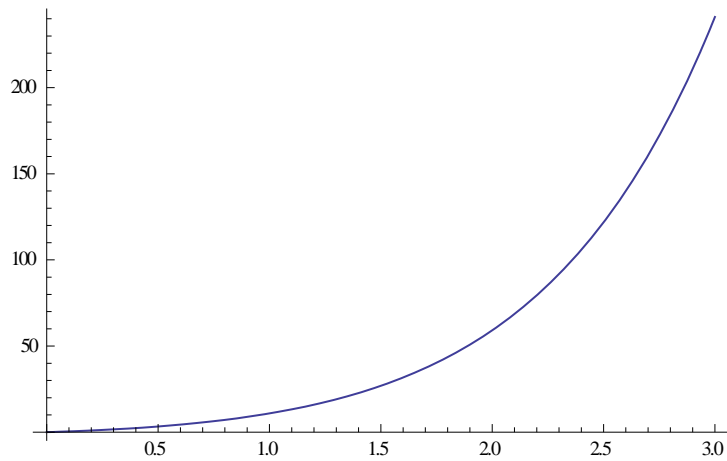
**Σχήμα 2.4.3:** Η γραφική παράσταση της συνάρτησης επιβίωσης με παραμέτρους  $\alpha=2$  και  $\theta=2$ .



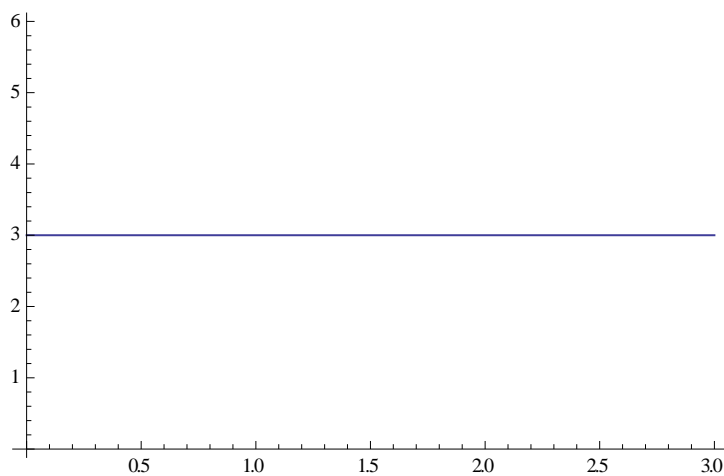
**Σχήμα 2.4.4:** Γραφική παράσταση συνάρτησης επιβίωσης για  $\alpha=2$  και  $\theta=2$  (μπλε) και  $\alpha=2$  και  $\theta=3$  (κόκκινο)

Έτσι παρατηρούμε από την προηγούμενη γραφική παράσταση ότι για  $\theta= 2$  η γραφική παράσταση συγκλίνει πιο γρήγορα στο μηδέν και έχει πιο ελαφριά ουρά.

Στην συνέχεια παρουσιάζουμε την γραφική παράσταση της βαθμίδας αποτυχίας για  $\theta=2, \theta=1$  .



**Σχήμα 2.4.5:** η γραφική παράσταση για την βαθμίδα αποτυχίας, η οποία είναι αύξουσα συνάρτηση στον χρόνο(IFR).



**Σχήμα 2.4.6:** Η γραφική παράσταση για  $\theta=1$  έχει ήδη σχολιαστεί στην προηγούμενη παράγραφο.

**Σχόλιο:** Άξιο αναφοράς είναι το γεγονός ότι για  $\theta < 1$ , τότε η βαθμίδα αποτυχίας  $\lambda(t)$  είναι φθίνουσα συνάρτηση του  $t$ .

(Marshall and Olkin, ,2007)

## 2.5 Κατανομή Weibull

Η κατανομή Weibull θεωρείται μια από τις πιο δημοφιλείς παραμετρικές κατανομές που μελετάει διεξοδικά τον χρόνο που ένα σύστημα αποτυγχάνει, αλλά ταυτόχρονα πρόκειται για μια κατανομή που έχει θεωρηθεί ως περιγραφικό μοντέλο για την κατανομή του εισοδήματος. Στην επιστήμη πιθανοτήτων και στατιστικής, είναι μια συνεχής κατανομή με βάρος ουράς εξαρτώμενο από τις τιμές των παραμέτρων, κάτι το οποίο προσδιορίζει πόσο αργά ή μη συγκλίνει στο μηδέν. Η εν λόγω κατανομή πήρε το όνομά της από τον Σουηδό μαθηματικό καθηγητή Wallodi Weibull που γεννήθηκε το 1887, ο οποίος την πρότεινε ως μία κατανομή με διάφορες εφαρμογές και την ανέλυσε διεξοδικά το 1951, παρόλο που εντοπίστηκε για πρώτη φορά από

τους Rossin & Rammler (1933) για να περιγράψει μια κατανομή σωματιδίων. Οι ομοιότητες που πρόσεξε ανάμεσα στις παρατηρήσεις και στις προβλέψεις μέσω της κατανομής ήταν εντυπωσιακές. Μελέτησε θέματα όπως η αντοχή του ατσαλιού και της ίνας του ινδικού βαμβακιού, το ύψος των ενήλικων ανδρών στις Βρετανικές νήσους καθώς και τον πολλαπλασιασμό μερικών φυτικών ειδών. Η κατανομή αυτή εφαρμόστηκε για πρώτη φορά το 1933 από τον Ροσίν (Rosin) και τον Ράμμερ (Rammler), όπως έχει ήδη αναφερθεί στην περιγραφή των «νόμων που διέπουν την λεπτότητα του κονιοροποιημένου άνθρακα» ("Laws governing the fineness of powdered coal"). Επίσης προτάθηκε να εφαρμοστεί για την δύναμη των υλικών και την ρήξη των στερεών. Αργότερα, σε κάποια επιστημονικά άρθρα (1951, 1952) προτάθηκε από τον Weibull η εφαρμογή της κατανομής σε ανάλυση σπουδαίων εφαρμογών και δεδομένων και σε αυτά τα άρθρα αναλύθηκε ότι η κατανομή αυτή έχει ως αποτέλεσμα η συνάρτηση κινδύνου μπορεί να είναι μια μη αρνητική αύξουσα συνάρτηση με ορισθέντα τύπο  $\lambda(x) = (\theta x)^{\alpha-1}$ . Επιπλέον, έδωσε έναν αριθμό από δεδομένα, όπου καθένα από αυτά είχε καλή εφαρμογή στην εν λόγω κατανομή. Το 1958 χρησιμοποιήθηκε από τους Lieblein, Zelen και Kao για να περιγράψει την αποτυχία ρουλεμάν και σε κενό σωλήνα αντίστοιχα. Τέλος, η Weibull εφαρμόστηκε σε εκτενή βαθμό σε ιατρικές μελέτες και σε άλλα επιστημονικά πεδία συμπεριλαμβανομένης της μηχανικής επιστήμης αλλά παράλληλα πέρα από την εφαρμογή της για την περιγραφή χρόνου ζωής το μοντέλο της Weibull βρίσκεται στην οικονομία, μετεωρολογία και την ασύρματη επικοινωνία.

Η κατανομή Weibull συμπεριλαμβάνει και την εκθετική κατανομή, ως μια ειδική περίπτωση και μερικές φορές θεωρείται γενίκευσή της. Το εκθετικό μοντέλο αναπτύχθηκε και χρησιμοποιήθηκε εκτενώς στην δεκαετία του 1950 και η κατανομή Weibull άρχισε να θεωρείται ανταγωνιστικό μοντέλο την επόμενη δεκαετία, ειδικά σε προβλήματα στα οποία ο χρόνος ζωής ήταν η υπό μελέτη μεταβλητή.

Σε αντίθεση όμως με την εκθετική, δεν χαρακτηρίζεται από σταθερή συνάρτηση κινδύνου και γι' αυτό έχει ευρύτερες εφαρμογές. Όταν ο χρόνος επιβίωσης  $X$  ακολουθεί την κατανομή Weibull με παραμέτρους  $\alpha$  και  $\theta$ , όπου έχουμε παράμετρο σχήματος  $\alpha$  και παράμετρο κλίμακας  $\theta$  αντίστοιχα, η συνάρτηση πυκνότητας περιγράφεται από την συνάρτηση

$$f(x) = \alpha * \theta * (\theta x)^{\alpha-1} * e^{-(\theta x)^\alpha}, \quad x \geq 0 \quad (2.5.1)$$

Και η αθροιστική συνάρτηση κατανομής είναι η

$$F(x) = 1 - e^{-(\theta x)^\alpha} \quad \theta, \alpha, x \geq 0 \quad (2.5.2)$$

Αυτό έχει ως αποτέλεσμα από την στιγμή που ισχύει η σχέση  $F(x) = 1 - S(x)$ , τότε η συνάρτηση επιβίωσης της εν λόγω κατανομής είναι

$$S(x) = e^{-(\theta x)^\alpha}, \quad \lambda, \alpha, x \geq 0 \quad (2.5.3)$$

Τέλος είναι σημαντικό να προσδιοριστεί ότι η βαθμίδα αποτυχίας μπορεί να περιγραφεί από την σχέση

$$\lambda(x) = a\theta(\theta x)^{a-1}, \quad x \geq 0 \quad (2.5.4)$$

Παρατηρούμε όμως  $a > 1$  η κατανομή αυτή είναι IFR ενώ για  $a < 1$  είναι DFR, ενώ για  $a = 1$  όπως ήδη αναφέραμε είναι η περίπτωση της εκθετικής κατανομής με σταθερή ένταση κινδύνου  $\lambda(x) = \theta$  για κάθε  $x \geq 0$ .

Εάν  $X \sim \text{Weibull}(\alpha, \theta)$  με  $r, \lambda > 0$  τότε η μέση τιμή της κατανομής είναι

$$E(X) = \frac{\Gamma(1 + \frac{1}{a})}{\theta} \quad (2.5.5)$$

και η διακύμανση της κατανομής είναι

$$\text{Var}(X) = 1/\theta^2 * [\Gamma(1 + \frac{2}{a}) - \Gamma^2(1 + \frac{1}{a})] \quad (2.5.6)$$

$\Gamma(r)$  με  $r > 0$  είναι η συνάρτηση  $\Gamma$  που ορίζεται ως το γενικευμένο ολοκλήρωμα.

(Μιχαήλ Μπούτσικας, 2015)

([http://www.unipi.gr/faculty/mbouts/prob\\_intro/prob\\_intro4.pdf](http://www.unipi.gr/faculty/mbouts/prob_intro/prob_intro4.pdf))

Η διακύμανση για την κατανομή Weibull μπορεί να βρεθεί ως συνέπεια ενός γενικότερου αποτελέσματος, συγκεκριμένα:

Γνωρίζουμε πως οι ροπές της κατανομής αυτής προέρχονται από τον τύπο

$$E(X^k) = \theta^{-k} \Gamma(1 + \frac{k}{a})$$

Επομένως η μέση τιμή της κατανομής προκύπτει ότι ισούται με

$$E(X) = \frac{\Gamma(1 + \frac{1}{a})}{\theta}, \quad (2.5.7)$$

ενώ η δεύτερη ροπή ισούται με τον τύπο

$$E(X^2) = \frac{\Gamma(2 + \frac{1}{a})}{\theta^2} \quad (2.5.8)$$

Άρα με αντικατάσταση στον γνωστό τύπο  $\text{Var}(X) = E(X^2) - E^2(X)$  προκύπτει ότι

$$\text{Var}(X) = 1/\theta^2 * [\Gamma(1 + \frac{2}{a}) - \Gamma^2(1 + \frac{1}{a})]$$

Σχετικά με την ροπογεννήτρια συνάρτηση μπορούμε να σχολιάσουμε τα εξής:

- i. Για  $a > 1$ , η ροπογεννήτρια υπάρχει, δεν γνωρίζουμε όμως το γενικό τύπο και δεν μπορούμε να την υπολογίσουμε.
- ii. Για  $a < 1$ , η ροπογεννήτρια δεν υπάρχει.



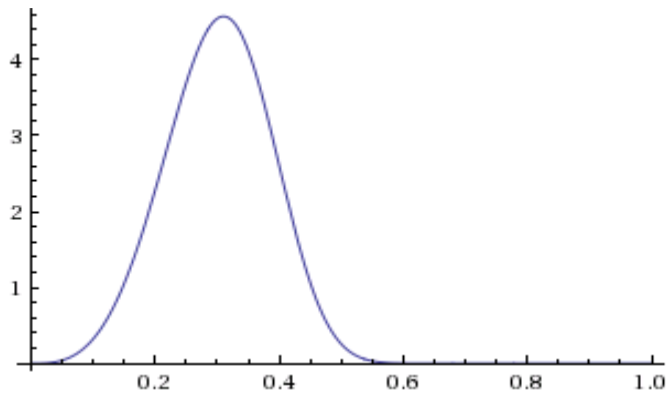
Η κατανομή Weibull σχετίζεται όμως με μια σειρά από άλλες κατανομές πιθανότητας. Ειδικότερα, παρεμβάλλεται μεταξύ της εκθετικής κατανομής ( $\alpha=1$ ) και της κατανομής Rayleigh ( $\alpha=2$ ). Εάν η ποσότητα  $X$  είναι έτοιμη να «αποτύχει», η κατανομή Weibull δίνει μια κατανομή για την οποία το ποσοστό αποτυχίας είναι ανάλογο με μια δύναμη του χρόνου.

Η παράμετρος σχήματος,  $\alpha$ , είναι αυτή η δύναμη συν ένα, και έτσι αυτή η παράμετρος μπορεί να ερμηνευθεί ως εξής :

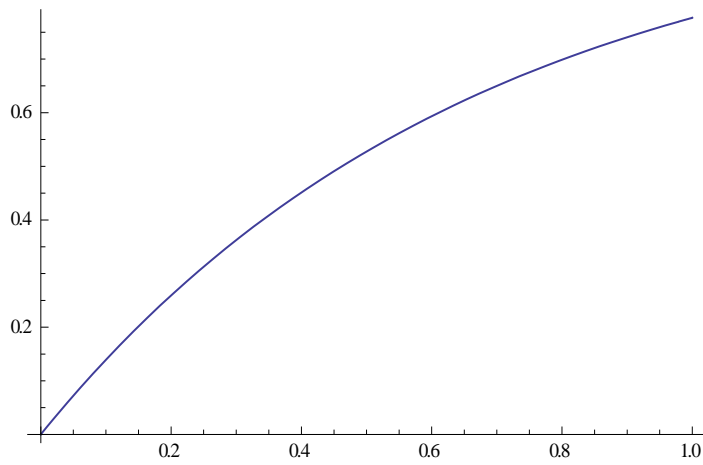
- ✓ Μια τιμή του  $\alpha < 1$  δείχνει ότι το ποσοστό αποτυχίας μειώνεται με την πάροδο του χρόνου. Αυτό συμβαίνει όταν υπάρχει σημαντική «θνησιμότητα», ή ελαττωματικά στοιχεία που παραλείπονται νωρίς και το ποσοστό αποτυχίας μειώνεται με την πάροδο του χρόνου, καθώς τα ελαττωματικά στοιχεία εξαλείφονται από τον πληθυσμό. Έτσι η κατανομή έχει πιο βαριά ουρά από την εκθετική κατανομή και συγκλίνει πιο αργά στο μηδέν αλλά πιο γρήγορα από την κατανομή Pareto.
- ✓ Η τιμή του  $\alpha=1$  δείχνει ότι το ποσοστό αποτυχίας είναι σταθερό με την πάροδο του χρόνου. Αυτό μπορεί να υποδηλώνει ότι τυχαία εξωτερικά γεγονότα προκαλούν θνησιμότητα, ή αποτυχία. Έτσι παίρνουμε την εκθετική κατανομή.
- ✓ Μια τιμή του  $\alpha > 1$  δείχνει ότι οι ποσοστό αποτυχίας αυξάνεται με το χρόνο. Αυτό συμβαίνει όταν υπάρχει μια διαδικασία «γήρανσης», ή τμήματα που είναι πιο πιθανό να αποτύχουν, με το πέρασμα του χρόνου. Έτσι η κατανομή έχει πιο ελαφριά ουρά από την εκθετική κατανομή και συγκλίνει πιο γρήγορα στο μηδέν.

### **Γραφικές παραστάσεις της εν λόγω κατανομής**

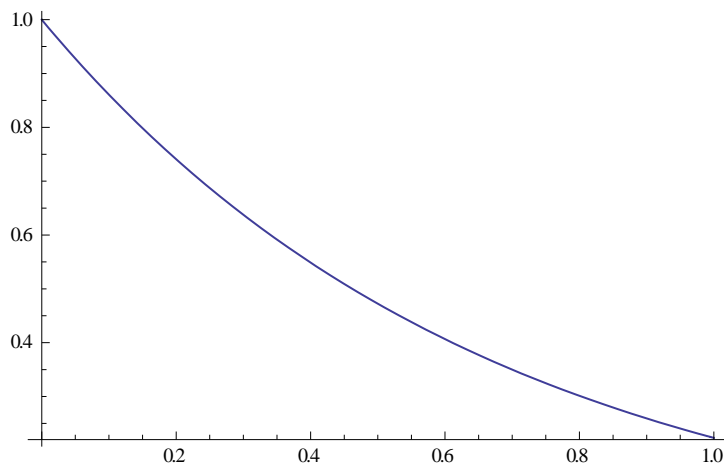
Στα παρακάτω τρία σχήματα βλέπουμε κάποιες γραφικές παραστάσεις της κατανομής Weibull με παραμέτρους  $(\theta=3, \alpha=4)$ ,  $(\theta=3, \alpha=1/2)$  και  $(\theta=3, \alpha=1)$  και αντίστοιχα παρατηρούμε ότι το πρώτο σχήμα έχει ελαφριά ουρά, το δεύτερο έχει βαριά ουρά αφού  $\alpha < 1$  και για το τρίτο σχήμα παρατηρούμε στο ότι έχουμε την εκθετική κατανομή με παράμετρο  $\theta=3$ .



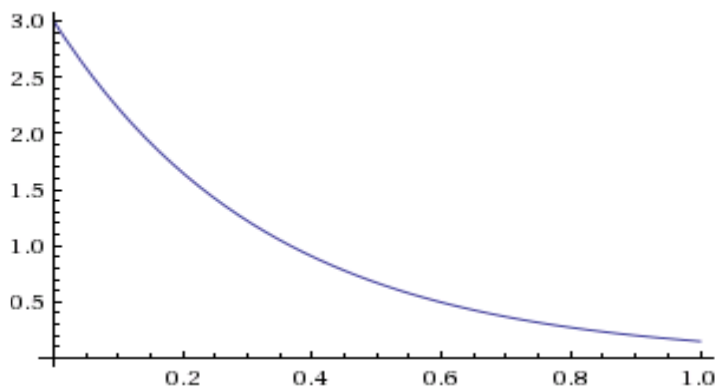
**Σχήμα 2.5.1:** Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας για  $\theta=3, \alpha=4$



**Σχήμα 2.5.2:** Διάγραμμα της συνάρτησης κατανομής για  $\theta=3, \alpha=1/2$

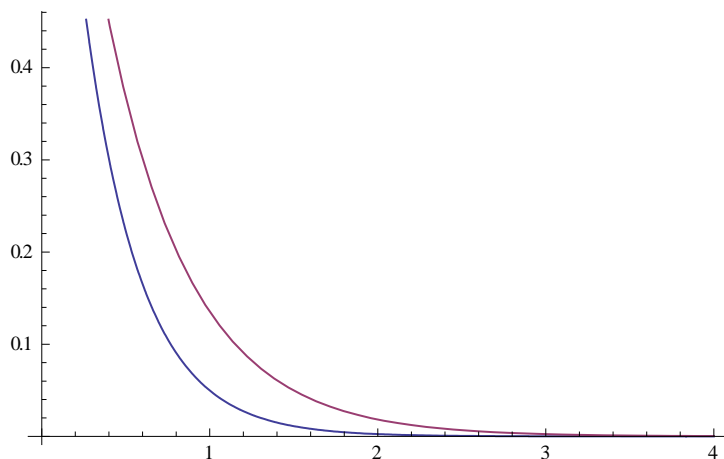


**Σχήμα 2.5.3:** Διάγραμμα της συνάρτησης επιβίωσης για  $\theta=3, \alpha=1/2$



**Σχήμα 2.5.4:** Διάγραμμα της συνάρτησης επιβίωσης για  $\theta=3, \alpha=1$

**Σχολιασμός :** Παρατηρούμε πως για τα δυο γραφήματα της συνάρτησης επιβίωσης που πρόκειται για μια γνησίως φθίνουσα συνάρτηση για  $\theta=3$  και  $\alpha=1/2$  η συνάρτηση συγκλίνει στο μηδέν πιο γρήγορα έχοντας πιο βαριά ουρά και για  $\alpha=1/2$  συγκλίνει πιο αργά έχοντας πιο ελαφριά ουρά σε σχέση με  $\alpha=1$ .



**Σχήμα 2.5.5:** Γραφική παράσταση συνάρτησης επιβίωσης για  $\alpha=2$  και  $\theta=2$  (μπλε) και  $\alpha=2$  και  $\theta=3$  (κόκκινο).

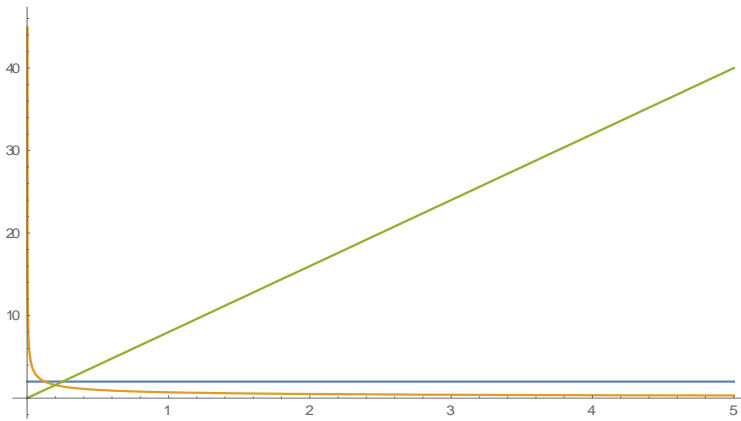
Άρα εύλογα παρατηρούμε ότι το παραπάνω παράδειγμα επαληθεύεται καθώς για  $\alpha=2$  και  $\lambda=2$  η συνάρτηση συγκλίνει πιο γρήγορα στο μηδέν.

### **Υπολογισμός βαθμίδας αποτυχίας κατανομής Weibull**

Με βάση τον ορισμό της βαθμίδας αποτυχίας που είδαμε προηγούμενη ενότητα και συγκεκριμένα από τον τύπο (2.1.1) και (2.2.2) ισχύει :

$$\lambda(t) = \frac{f(t)}{P(T>t)} = \frac{f(t)}{S(t)} = \frac{a*\lambda\theta*e^{-(\theta t)^a}}{e^{-(\theta t)^a}} = a * \theta * (\theta t)^{a-1}$$

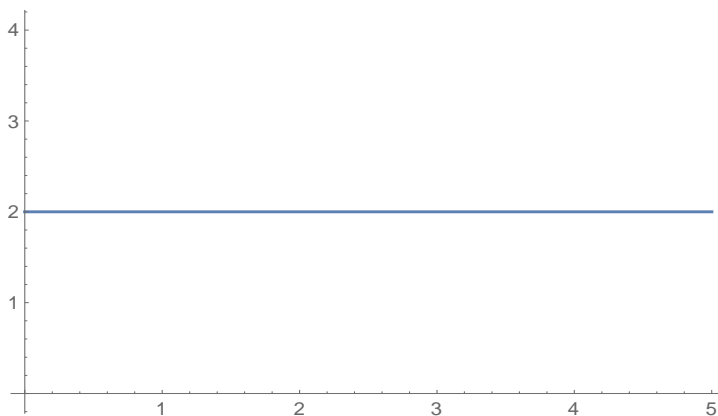
Παρακάτω παρουσιάζεται η γραφική παράσταση τις βαθμίδας αποτυχίας της κατανομής Weibull για  $\alpha=1$  και  $\alpha=2$  και  $\alpha=0,5$



**Σχήμα 2.5.6:** Σχεδιάγραμμα βαθμίδα αποτυχίας με  $\alpha=1$ (μπλε), $\alpha=2$ (πράσινο), $\alpha=0.5$ (πορτοκαλί)

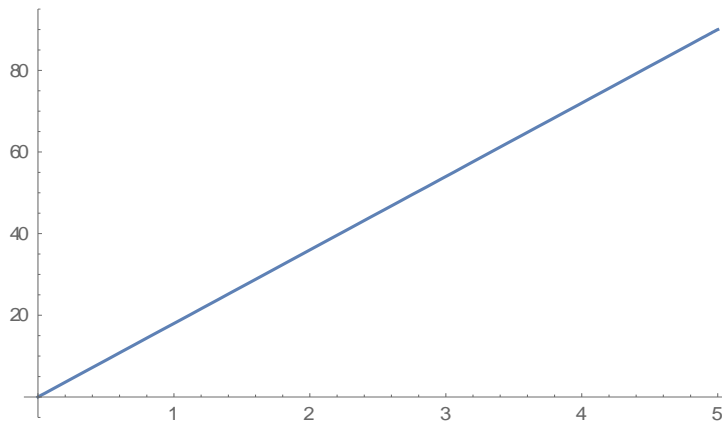
- Παρατηρούμε ότι για  $\alpha=1$  πως η βαθμίδα αποτυχίας παραμένει σταθερή δηλαδή προκύπτει όπως έχει ειπωθεί η εκθετική κατανομή

Το ανάλογο σχεδιάγραμμα είναι το εξής για  $\alpha=1, \theta=2$



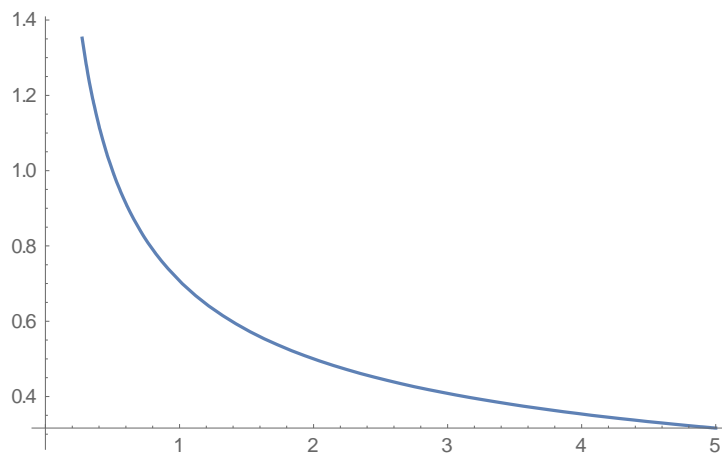
**Σχήμα 2.5.7:** Σχεδιάγραμμα βαθμίδα αποτυχίας για  $\alpha=1, \theta=2$

- Για  $\alpha>1$  είναι εύκολο να διαπιστώσουμε πως η βαθμίδα αποτυχίας αυξάνεται με το πέρασμα του χρόνου δηλαδή προκύπτει IFR χρόνος ζωής.



**Σχήμα 2.5.8:** Σχεδιάγραμμα βαθμίδα αποτυχίας με  $\alpha=2, \theta=3$

- Όταν το  $\alpha$  παίρνει τιμές  $0 < \alpha < 1$  η βαθμίδα αποτυχίας μειώνεται με τον χρόνο δηλαδή έχουμε DFR.



**Σχήμα 2.5.9:** Σχεδιάγραμμα βαθμίδα αποτυχίας με  $\alpha=2, \theta=3$

### 2.5.1 Αντίστροφη και γενικευμένη κατανομή Weibull

Ξεκινώντας την περιγραφή μπορούμε να πούμε πως αν η  $X$  ακολουθεί μια κατανομή Weibull τότε η  $1/X$  ακολουθεί αντίστροφη κατανομή Weibull. Στην προκειμένη περίπτωση η συνάρτηση κατανομής της προσεγγίζεται αν ορίσουμε τη μετατροπή

$$T = \beta^2 / X$$

Έπειτα η  $T$  έχει συνάρτηση πυκνότητας που δίνεται από τον τύπο

$$F(t) = e^{-\left(\frac{t}{\beta}\right)^{-\alpha}}, \quad \theta, \alpha > 0.$$

Έστω μια συνεχής τυχαία μεταβλητή η οποία αναφέρεται στον χρόνο επιβίωσης της συνιστώσας ενός συστήματος. Η πιθανή συμπεριφορά επιβίωσης μιας συνιστώσας ενός συστήματος προσδιορίζεται γενικά από μια συνάρτηση  $f_X(t)$  η οποία ονομάζεται Πυκνότητα Αποτυχίας και ουσιαστικά πρόκειται για μια συνάρτηση πυκνότητας-

πιθανότητας η οποία ορίζεται στον ημιάξονα των θετικών πραγματικών αριθμών και καθορίζει την πυκνότητα στιγμιαίας αποτυχίας την χρονική στιγμή  $t$ .

Αντίστοιχη θεωρείται η  $F_X(t)$  που ορίζεται σαν Αθροιστική Συνάρτηση κατανομής Αποτυχίας μέχρι την χρονική στιγμή  $t$ . Στα πλαίσια Ανάλυσης Επιβίωσης, το Μοντέλο Αποτυχίας, περιλαμβάνει την  $f_X(t)$  καθώς και τις τιμές των παραμέτρων της. Στην συνέχεια θα οριστεί η συνάρτηση επιβίωσης η οποία βάσει του γνωστού τύπου:

$$S(x) = 1 - F(x),$$

προκύπτει βάσει των παραπάνω ότι ο τύπος της συνάρτησης επιβίωσης είναι:

$$S(t) = 1 - e^{-\left(\frac{t}{\beta}\right)^{-\alpha}}, \theta, t, \alpha > 0$$

Η αντίστροφη κατανομή Weibull είναι γνωστή ως μια κατανομή 2 ακραίων τιμών ή η Frechet κατανομή (Johnson et al. 1995, Chap. 22), ενώ η αρνητική τυχαία μεταβλητή της κατανομής Weibull είναι τύπος 3 ακραίων τιμών. Η κατανομή Weibull και Frechet είναι και οι 2 ειδικές περιπτώσεις κατανομής ακραίων τιμών.

Ο Erto (1989) έχει συζητήσει τις ιδιότητες αυτής της κατανομής και την πιθανή χρήση της ως μοντέλο διάρκειας ζωής. Η εκτίμηση της μέγιστης πιθανοφάνειας και η εκτίμηση των ελάχιστων τετραγώνων των παραμέτρων της αντίστροφης κατανομής Weibull έχουν συζητηθεί από τους Calabria και Pulcini (1990).

(Chin-Dew, 2014, Chapter 2, Generalized Weibull Distributions)

## 2.6 Κατανομή Pareto

Η κατανομή Pareto είναι γνωστή στη στατιστική επιστήμη και με την ονομασία ως κατανομή νόμου δύναμης (power law distribution). Μια συναρτησιακή σχέση προκύπτει μεταξύ δυο ποσοτήτων όπου μια δύναμη μεταβάλλεται σε συνάρτηση της άλλης. Η κατανομή αυτή πήρε το όνομά της από τον μηχανικό, οικονομολόγο και κοινωνιολόγο Alfredo Pareto και χρησιμοποιείται εκτενώς σε πολλές επιστήμες όπως για παράδειγμα στην κοινωνιολογία, γεωφυσική και στον αναλογισμό. Πιο συγκεκριμένα όσο αφορά την αναλογιστική επιστήμη χρησιμοποιείται κατά κόρον σε αναλογιστικές μελέτες λόγω του βάρους στη δεξιά ουρά. Η ιδιότητα αυτή έχει την δυνατότητα να επεξηγήσει και να ερμηνεύσει λεπτομερώς κινδύνους οι οποίοι εμφανίζονται στον ασφαλιστικό κλάδο, όπως τη σφοδρότητα ατυχημάτων για μεγάλες ζημιές του κλάδου γενικών ασφαλίσεων. Στο χαρακτηριστικό της θετικής ασυμμετρίας βασίζεται και ο νόμος Pareto (80-20) ο οποίος βρίσκει πολλές εφαρμογές στον κλάδο των οικονομικών, για παράδειγμα, το 80% του πλούτου μιας χώρας κατέχεται από το 20% του πληθυσμού της.

Θα λέμε πως η τυχαία μεταβλητή  $X$  ακολουθεί κατανομή Pareto η οποία αναφέρεται επίσης και ως American Pareto με παραμέτρους  $\alpha$  (παραμέτρο μορφής),  $\theta$  (παραμέτρο σχήματος) όταν η συνάρτηση πυκνότητας πιθανότητας της είναι :

$$f(x) = a \frac{\theta^\alpha}{(\theta+x)^{a+1}}, \quad x > 0, a > 0, \theta > 0 \quad (2.6.1)$$

Η συνάρτηση κατανομής της Pareto περιγράφεται από την συνάρτηση

$$F(x) = 1 - \frac{\theta^\alpha}{(\theta+x)^\alpha}, \quad x > 0, a > 0, \theta > 0 \quad (2.6.2)$$

Έτσι προκύπτει πως η συνάρτηση επιβίωσης περιγράφεται από την σχέση

$$S(x) = \frac{\theta^\alpha}{(\theta+x)^\alpha}, \quad x > 0, a > 0, \theta > 0 \quad (2.6.3)$$

Στην συνέχεια θα αναφέρουμε σημαντικά χαρακτηριστικά μεγέθη της κατανομής Pareto .

Πιο συγκεκριμένα από το βιβλίο του Bean (2001) βλέπουμε πως οι ροπές κ-τάξης υπολογίζονται από τον τύπο

$$E(X^k) = \frac{\theta^k k!}{(\alpha - 1)(\alpha - 2) \dots (\alpha - k)}$$

Όπου σημαντική επισήμανση είναι ότι η ροπή κ-τάξης ορίζεται από τον τύπο ( $E(X^k) < \infty$ ) όταν και μόνο όταν  $\alpha > k$ .

Για  $k=1$  προκύπτει η μέση τιμή της εν λόγω κατανομής

$$E(X) = \frac{\theta}{\alpha - 1} \quad (2.6.4)$$

Για  $k=2$  προκύπτει η δεύτερη ροπή της κατανομής Pareto

$$E(X^2) = \frac{2\theta^2}{(\alpha - 1)(\alpha - 2)}$$

Για  $k=3$  προκύπτει η τρίτη ροπή της κατανομής Pareto

$$E(X^3) = \frac{6\theta^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)}$$

Έτσι, βάσει του γνωστού τύπου της διασποράς  $Var(X) = E(X)^2 - E^2(X)$ , προκύπτει η διασπορά της εν λόγω κατανομής

$$Var(X) = \frac{2\theta^2}{(\alpha - 1)(\alpha - 2)} - \frac{\theta^2}{(\alpha - 1)^2} \Rightarrow$$

$$Var(X) = \frac{\alpha\theta^2}{(\alpha - 1)(\alpha - 2)} \quad (2.6.5)$$

**Η ροπογεννήτρια της κατανομής δεν υπάρχει όπως φαίνεται στο παρακάτω παράδειγμα .**

**Παράδειγμα 2.6.1:**

Έστω  $F(x) = 1 - (1 + x)^{-a}, a > 1, x > 0$

Τότε έχουμε  $f(x) = a(1 + x)^{-a-1}$ , όπου προκύπτει η σ.π.π της κατανομής Pareto με παραμέτρους  $a$  και  $\theta=1$  μια κατανομή με βαριά ουρά όπου προφανώς συγκλίνει πολύ αργά στο μηδέν.

Βάσει του γνωστού τύπου :

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \Rightarrow M_x(t) = \int_{-\infty}^{\infty} e^{tx} a(1 + x)^{-a-1} dx$$

Εφόσον για κάθε  $t > 0$  ισχύει  $\lim_{x \rightarrow \infty} e^{tx} a(1 + x)^{-a-1} = \infty$  παρατηρούμε πως το ολοκλήρωμα δεν συγκλίνει για καμία τιμή του  $t$ , άρα και **η ροπογεννήτρια δεν υπάρχει.**

Αυτό μπορούμε επιπλέον να το διαπιστώσουμε και με την χρήση του αλγεβρικού πακέτου Mathematica . Πιο συγκεκριμένα αν ορίσουμε την συνάρτηση της ροπογεννήτριας με  $a=1$  και  $\lambda=3$  έχουμε τα εξής:

$$Mt[x] := \text{Integrate}[\text{Exp}[x * t] * a * (l^a) / (l + x)^{(a + 1)}, \{x, 0, \text{Infinity}\}]$$

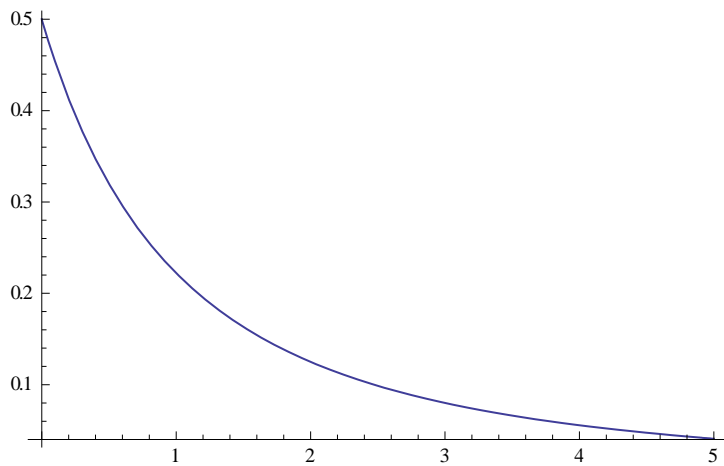
Και ακολουθεί η ένδειξη πως το αποτέλεσμα μας δεν ανήκει στο διάστημα  $[0, \infty]$ .

$$3x / (2+x)^2 \text{ does not converge on } \{0, \infty\}.$$

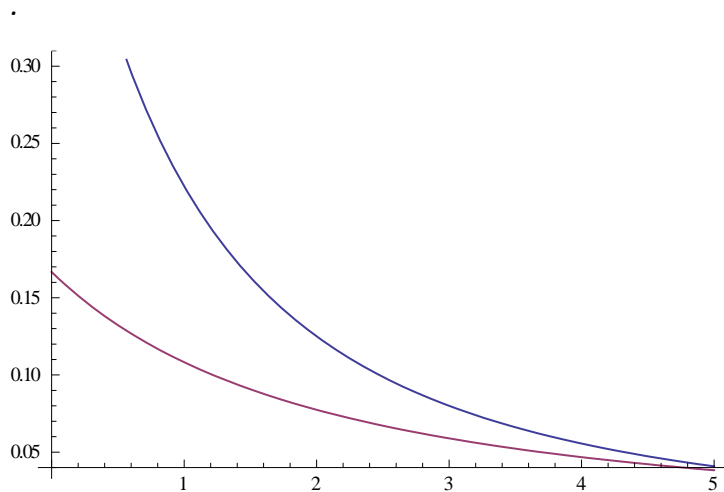
**Γραφικές παραστάσεις της εν λόγω κατανομής**

Στο παρακάτω σχήμα βλέπουμε την γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας μια συνεχούς τυχαίας μεταβλητής  $X$  που ακολουθεί την κατανομή Pareto παραμέτρους  $a=1, \theta=2$  αλλά και γραφική παράσταση συνάρτησης πυκνότητας με  $a=, \theta=2$  (μπλε και  $a=1/2, \theta=3$  (κόκκινο) .



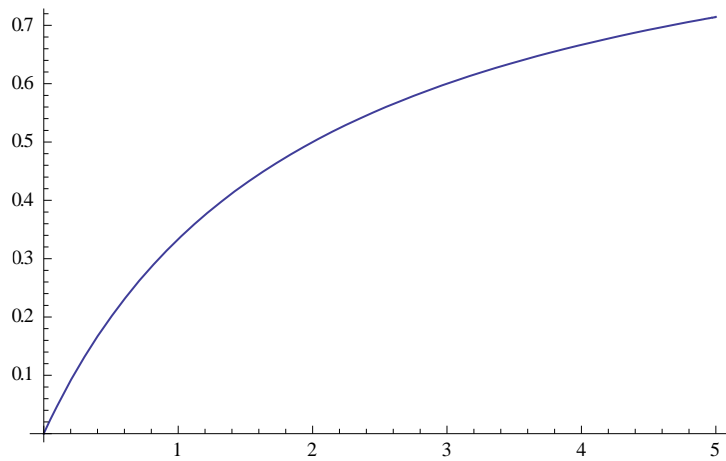


**Σχήμα 2.6.1:** Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας για  $\alpha=1, \theta=2$ .

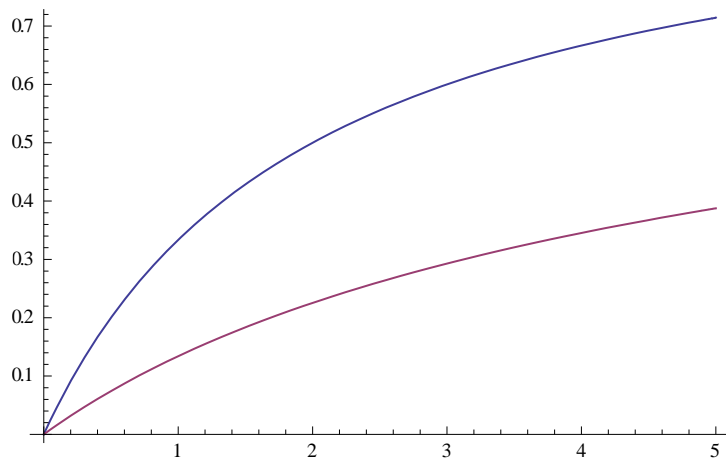


**Σχήμα 2.6.2:** Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας για  $\alpha=1, \theta=2$  (μπλε),  $\alpha=1/2, \theta=3$  (κόκκινο)

Στο παρακάτω σχήμα βλέπουμε την γραφική παράσταση της συνάρτησης κατανομής που είναι μια αύξουσα συνάρτηση για μια τυχαία μεταβλητή  $X$  που ακολουθεί την κατανομή Pareto με παραμέτρους  $\alpha=1, \theta=2$ .

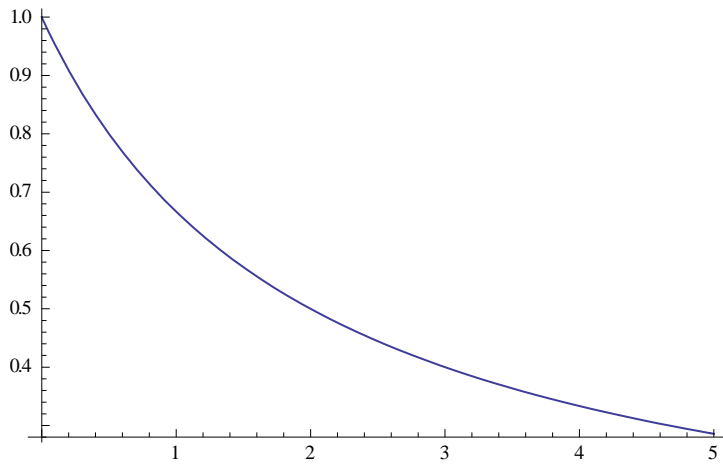


**Σχήμα 2.6.3:** Διάγραμμα της συνάρτησης κατανομής για  $\alpha=1$ ,  $\theta=2$ .



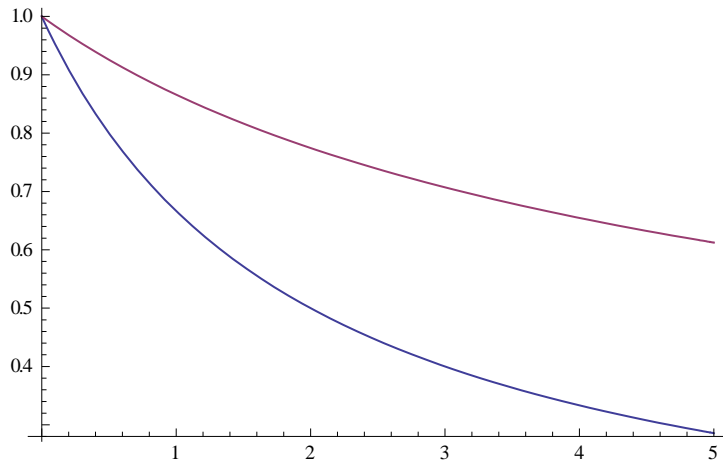
**Σχήμα 2.6.4:** Γραφική παράσταση συνάρτησης κατανομής για  $\alpha=1$ ,  $\theta=2$  (μπλε) και  $\alpha=1/2$ ,  $\theta=3$  (κόκκινο)

Στο παρακάτω σχήμα βλέπουμε την γραφική παράσταση της συνάρτησης επιβίωσης που είναι μια φθίνουσα συνάρτηση για μια τυχαία μεταβλητή  $X$  που ακολουθεί την κατανομή Pareto με παραμέτρους  $\alpha=1, \theta=2$  και παραμέτρους  $\alpha=1/2, \theta=3$  ως συνδυαστικό γράφημα για να μπορέσουμε να αξιολογήσουμε πόσο έχει πιο βαριά ουρά η κατανομή αλλά και πόσο γρήγορα συγκλίνει.



**Σχήμα 2.6.5:** Διάγραμμα της συνάρτησης επιβίωσης για  $\alpha=1, \vartheta=2$ .

Το συνδυαστικό γράφημα με διαφορετικές τιμές των παραμέτρων είναι το εξής:



**Σχήμα 2.6.6:** Γραφική παράσταση συνάρτησης επιβίωσης με παραμέτρους  $\alpha=1, \vartheta=2$  (μπλε) και  $\alpha=1/2$  και  $\vartheta=3$  (κόκκινο).

Έτσι μπορούμε να σχολιάσουμε το γεγονός ότι με παραμέτρους  $\alpha=1$  και  $\lambda=2$  η κατανομή συγκλίνει πιο γρήγορα στο μηδέν και έχει πιο ελαφριά ουρά σε σχέση με τιμές παραμέτρων  $\alpha=1/2$  και  $\theta=3$ .

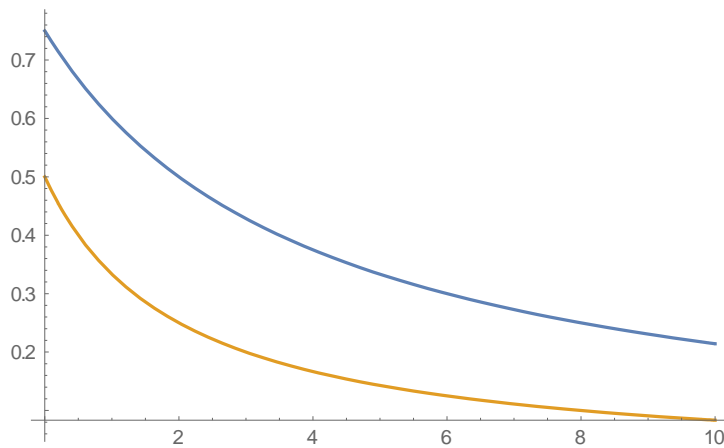
### Υπολογισμός βαθμίδας αποτυχίας για την κατανομή Pareto

Με βάση τους τύπους (2.1.1) και (2.2.2) βλέπουμε ότι για την κατανομή Pareto ισχύει :

$$\lambda(t) = \frac{f(t)}{P(T>t)} = \frac{f(t)}{S(t)} = \frac{a \frac{\theta^\alpha}{(\theta+t)^{\alpha+1}}}{\frac{\theta^\alpha}{(\theta+t)^\alpha}} = \frac{a}{(\theta+t)^1}$$

**Σχόλιο:** Παρατηρούμε εύκολα πως η παραπάνω βαθμίδα αποτυχίας είναι γνησίως φθίνουσα καθώς όσο περνάει ο χρόνος η τιμή του  $\lambda(t)$  μικραίνει.

Στο παρακάτω γράφημα παρουσιάζεται η βαθμίδα αποτυχίας για την κατανομή Pareto για τιμές παραμέτρων  $\alpha=3$  και  $\theta=4$  (μπλε) και  $\alpha=1$  και  $\theta=2$  (κίτρινο)



**Σχήμα 2.6.7:** Γραφική παράσταση βαθμίδας αποτυχίας με παραμέτρους  $\alpha=3$  και  $\vartheta=4$  (μπλε) και  $\alpha=1$  και  $\vartheta=2$  (κίτρινο)

## 2.7 Rayleigh distribution

Η κατανομή Rayleigh έχει πάρει τον όνομα της από τον Άγγλο καθηγητή πειραματικής φυσικής Lord Rayleigh. Η κατανομή αυτή στην θεωρία πιθανοτήτων και την στατιστική επιστήμη είναι συνεχής κατανομή πιθανότητας για θετικές τυχαίες μεταβλητές. Η εν λόγω κατανομή έχει εκτενή χρήση τόσο στην στατιστική επιστήμη, στην φυσική και άλλους κλάδους.

Πιο αναλυτικά η εν λόγω κατανομή βρίσκει εφαρμογή,

- ❖ στην θεωρία των επικοινωνιών με σκοπό την διαμόρφωση των πολλαπλών διάσπαρτων σημάτων ώστε να δημιουργηθεί ένας δέκτης
- ❖ στη φυσική για να μοντελοποιήσουμε την ταχύτητα του ανέμου, το ύψος των κυμάτων κλπ.
- ❖ Στην μηχανική για την μέτρηση της διάρκειας ζωής ενός αντικειμένου π.χ. πυκνωτές και αντιστάσεις
- ❖ Στην ιατρική επιστήμη

(Πηγή πληροφοριών : [www.Wikipedia.gr](http://www.Wikipedia.gr))

**Παρουσίαση Rayleigh στην Στατιστική επιστήμη:** Υποθέτουμε  $X$  και  $Y$  ανεξάρτητες τυχαίες μεταβλητές οι οποίες κατανέμονται κανονικά με μέση τιμή ίση με μηδέν και διακύμανση  $\sigma^2 = 1$ , (εφόσον υψώσουμε στο τετράγωνο και προσθέσουμε τις μεταβλητές έχουμε  $X^2 + Y^2$ , τότε προκύπτει μια κατανομή  $\chi^2$  με δυο βαθμούς ελευθερίας ( $df=2$ )).

Δίνουμε τώρα τον παρακάτω ορισμό,

- $R = \sqrt{X^2 + Y^2}$ , όπου  $X \sim N(0, \sigma^2)$  και  $Y \sim N(0, \sigma^2)$  είναι ανεξάρτητες κανονικές τυχαίες μεταβλητές. Τότε λέμε ότι η  $R$  ακολουθεί κατανομή Rayleigh με παράμετρο  $\sigma$ .

$$\triangleright \frac{R^2}{\sigma^2} \sim X_2^2$$

Αυτή η τυχαία μεταβλητή έχει ένα πολύ μεγάλο ενδιαφέρον για την ηλεκτρική μηχανική αλλά και σε άλλα πεδία εφαρμογών, όπως έχει ήδη αναφερθεί με εφαρμογή στο βιβλίο των Johnson, Kotz και Balakrishman (1994). Το  $R^2$  ακολουθεί μια κατανομή χ-τετράγωνο (chi-squared distribution) με παράμετρο  $N$  και 2 βαθμούς ελευθερίας ( $\nu=2$ ). Αυτή η κατανομή προήλθε από τον Rayleigh (1880) ως μια κατανομή πλάτους και είναι γνωστή ως Rayleigh distribution.

**Η κατανομή Rayleigh είναι μια ειδική περίπτωση της κατανομής Weibull για τιμή  $\alpha=2$ .**

Αν αντικαταστήσουμε στην συνάρτηση κατανομής της Weibull από το τύπο (2.5.2) το  $\alpha=2$  θα έχουμε το εξής:

$$F(x) = 1 - e^{-(\theta x)^\alpha} = 1 - e^{-(\theta x)^2}, x > 0$$

Έτσι μπορούμε να διαπιστώσουμε πως προκύπτει η κατανομή Rayleigh για  $\theta = 1/2$ .

Βλ. σχέση (2.7.2) αργότερα.

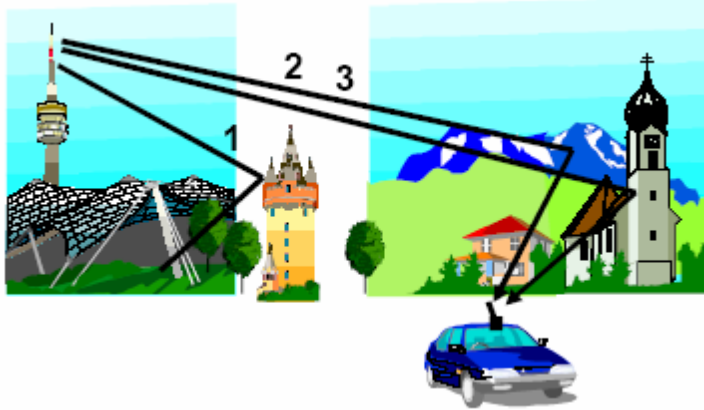
**Παράδειγμα εφαρμογής στην στατιστική επιστήμη:**

Έστω ότι ρίχνουμε ένα ζάρι σε μια επιφάνεια την οποία θεωρούμε σαν ένα καρτεσιανό σύστημα και οι συντεταγμένες υποθέτουμε  $X$  και  $Y$  του σημείου που έπεσε το ζάρι είναι ανεξάρτητες μεταξύ τους και έχουν κανονική κατανομή με  $\mu=0$  και  $\sigma=1$ .

Αν επιλέξουμε ένα ζευγάρι συντεταγμένων  $(X,Y)$ , τα οποία προέρχονται από την κανονική κατανομή με  $\mu=0$  και  $\sigma=1$  και υπολογίσουμε την απόσταση  $R = \sqrt{X^2 + Y^2}$  από το σημείο  $(x,y)$  μέχρι το αρχικό σημείο για έναν αριθμό επαναλήψεων. Τότε η  $R$  είναι μια τυχαία μεταβλητή με κατανομή Rayleigh.

Όπως αναφέραμε και νωρίτερα η κατανομή Rayleigh έχει εκτενή εφαρμογή στην επιστήμη της φυσικής. Ένα χαρακτηριστικό παράδειγμα είναι όταν η ταχύτητα του ανέμου αναλύεται σε ορθογώνια τμήματα 2-διαστάσεων του διανύσματος. Υποθέτοντας ότι κάθε συνιστώσα είναι ανεξάρτητη, κανονικά κατανεμημένη με την ίδια διακύμανση και μηδενική μέση τιμή, τότε η συνολική ταχύτητα του ανέμου (διάνυσμα μεγέθους) χαρακτηρίζεται από μια κατανομή Rayleigh. Ένα δεύτερο παράδειγμα της κατανομής προκύπτει στην περίπτωση των τυχαίων μιγαδικών αριθμών όταν υποθέσουμε ότι οι πραγματικές και φανταστικές συνιστώσες είναι ανεξάρτητα και ταυτόσημα κατανεμημένες κατά Gauss με την ίδια διακύμανση και μηδενική μέση τιμή. Σε αυτή την περίπτωση, η απόλυτη τιμή του μιγαδικού αριθμού είναι κατανεμημένη κατά Rayleigh. Ένα ακόμα παράδειγμα

για την επιστήμη της φυσικής είναι η εφαρμογή της κατανομής στους ραδιοδιαύλους όπου η κατανομή Rayleigh χρησιμοποιείται συχνά για να περιγράψει την χρονικά μεταβαλλόμενη μορφή της περιβάλλουσας του λαμβανομένου σήματος σε ένα κανάλι επίπεδων διαλείψεων ή της περιβάλλουσας ενός από τα πολλαπλά σήματα. Το μοντέλο αυτό χρησιμοποιείται όταν δεν υπάρχει δρόμος απευθείας διάδοσης όπως φαίνεται στο **Σχήμα 2.7.1**



**Σχήμα 2.7.1**

(Πηγή: [www.Wikipedia.gr](http://www.Wikipedia.gr))

Παρακάτω θα αναλυθούν η συνάρτηση πυκνότητας πιθανότητας, συνάρτηση επιβίωσης και βαθμίδα αποτυχίας της κατανομής Rayleigh. Πιο αναλυτικά, η συνάρτηση πυκνότητας πιθανότητας της κατανομής Rayleigh περιγράφεται με την ακόλουθη σχέση

$$f(x) = \frac{x}{\sigma^2} * e^{-\left(\frac{x^2}{2\sigma^2}\right)}, \quad x \geq 0 \quad (2.7.1)$$

Όπου  $\sigma$  είναι η παράμετρος κλίμακας (scale parameter). Έπειτα για την κατανομή Rayleigh η συνάρτηση επιβίωσης, που όπως γνωρίζουμε είναι μια φθίνουσα συνάρτηση και περιγράφεται από την σχέση

$$S(x) = e^{-\left(\frac{x^2}{2\sigma^2}\right)}, \quad x \geq 0 \quad (2.7.2)$$

([www.Wikipedia.gr](http://www.Wikipedia.gr))

για κάθε  $x \in (0, \infty)$  και  $\sigma$  είναι η παράμετρος κλίμακας ([scale parameter](#)).

Η βαθμίδα αποτυχίας της κατανομής Rayleigh περιγράφεται από την συνάρτηση

$$\lambda(x) = \frac{x}{\sigma^2}, \quad x \geq 0 \quad (2.7.3)$$

Ωστόσο για την κατανομή Rayleigh συζητήθηκε μια γενική μορφή της κατανομής αυτής που μπορεί να χαρακτηριστεί ως «γενικευμένη Rayleigh» και περιγράφεται από την συνάρτηση

$$f(x) = (2x^{n-1})(e^{-x^2/2\sigma^2}) / (2\sigma^2)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right), \quad x \geq 0 \quad (2.7.4)$$

Στην συγκεκριμένη περίπτωση για διαφορές τιμές του  $n$  προκύπτουν διάφορες κατανομές. Στην συνέχεια για μια τυχαία μεταβλητή  $X$  με αυτή την πυκνότητα προκύπτει

$$E(X^r) = (2^{\frac{r}{2}} \Gamma\left(\frac{n+r}{2}\right)) / \Gamma\left(\frac{n}{2}\right) \quad (2.7.5)$$

Ειδικές περιπτώσεις που προκύπτουν από τον τύπο (2.7.4) της γενικευμένης Rayleigh: Για  $n=1$  προκύπτει η κανονική κατανομή περικομμένη στο μηδέν.

Για  $n=2$  προκύπτει η συνάρτηση πυκνότητας της κατανομής Rayleigh (2.7.1).

Για  $n=3$  προκύπτει η κατανομή Maxwell-Boltzmann, με την οποία θα ασχοληθούμε στη συνέχεια της εργασίας μας.

Εφόσον αντικαταστήσουμε στον τύπο (2.7.4) για  $n=3$  προκύπτει το εξής:

$$\begin{aligned} f(x) &= (2x^{n-1})(e^{-x^2/2\sigma^2}) / (2\sigma^2)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) = \\ &= (2x^2)(e^{-x^2/2\sigma^2}) / (2\sigma^2)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right) \end{aligned}$$

Η συνάρτηση αυτή θα αναλυθεί και στο επόμενο κεφάλαιο, πρόκειται για την σ.π.π της κατανομής Maxwell-Boltzmann.

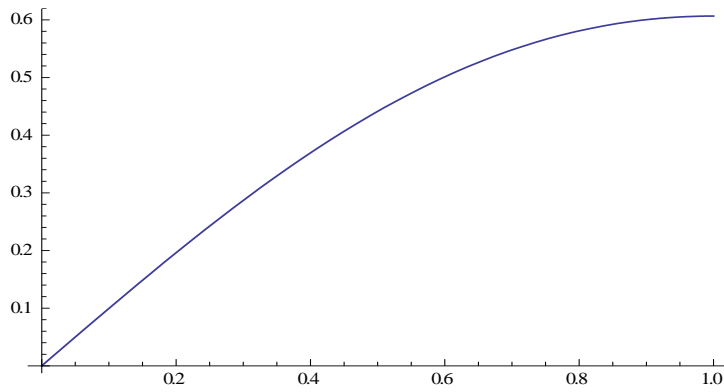
Τέλος από την συνάρτηση της πυκνότητας πιθανότητας προκύπτει ότι η μέση τιμή και η διακύμανση της κατανομής περιγράφονται από τις συναρτήσεις

$$E(X) = \sigma\sqrt{\pi/2} \quad (2.7.6)$$

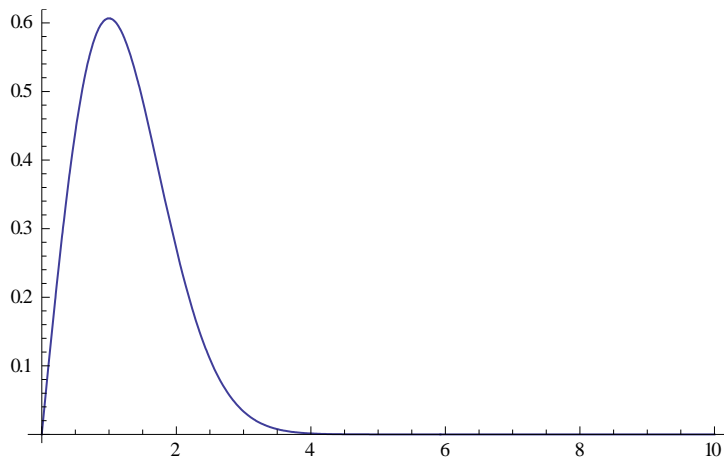
$$\text{Var}(X) = \frac{4-\pi}{2} \sigma^2 \quad (2.7.7)$$

### **Γραφικές παραστάσεις της εν λόγω κατανομής**

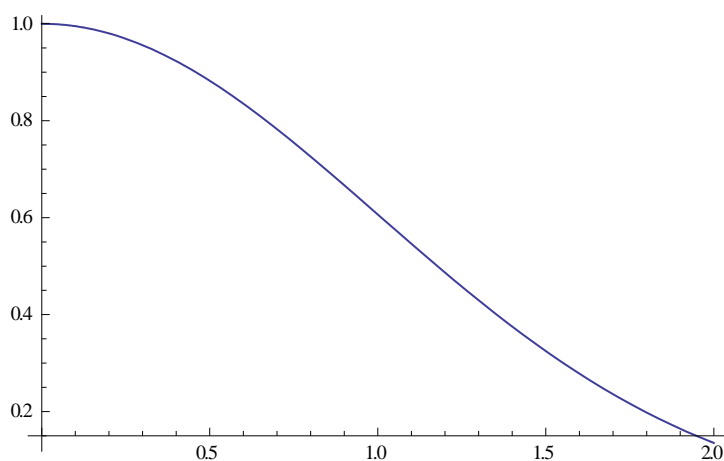
Παρακάτω παρουσιάζονται οι γραφικές παραστάσεις της συνάρτησης κατανομής, συνάρτηση πυκνότητας πιθανότητας για  $\sigma=1$  αλλά και συνάρτηση επιβίωσης για διάφορα διαστήματα για τιμές παραμέτρων  $\sigma=1$ ,  $\sigma=1/2$ .



**Σχήμα 2.7.2:** Παρατίθεται η γραφική παράσταση της γνησίως αύξουσας συνάρτησης κατανομής με τιμές παραμέτρων  $\sigma=1$  για το διάστημα  $(0,1)$ .

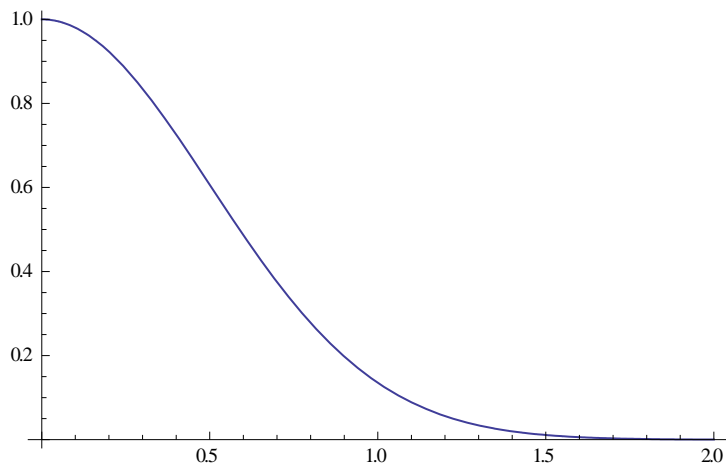


**Σχήμα 2.7.3:** Σύμφωνα με την γραφική παράσταση έχουμε μια γνησίως αύξουσα και έπειτα φθίνουσα συνάρτηση πυκνότητας πιθανότητας με ελαφριά ουρά η οποία συγκλίνει στο μηδέν σχετικά γρήγορα με τιμή παραμέτρου  $\sigma=1$  για το διάστημα  $(0,4)$ .



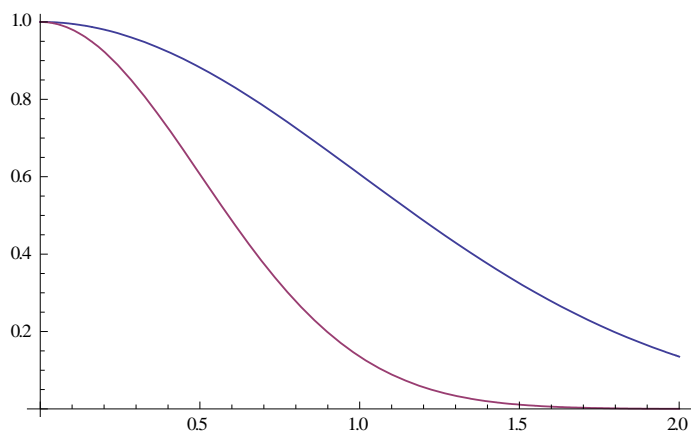
**Σχήμα 2.7.4:** Σύμφωνα με την γραφική παράσταση έχουμε μια γνησίως φθίνουσα συνάρτηση επιβίωσης με ελαφριά ουρά με τιμή παραμέτρου  $\sigma=1$  για το διάστημα  $(0,2)$



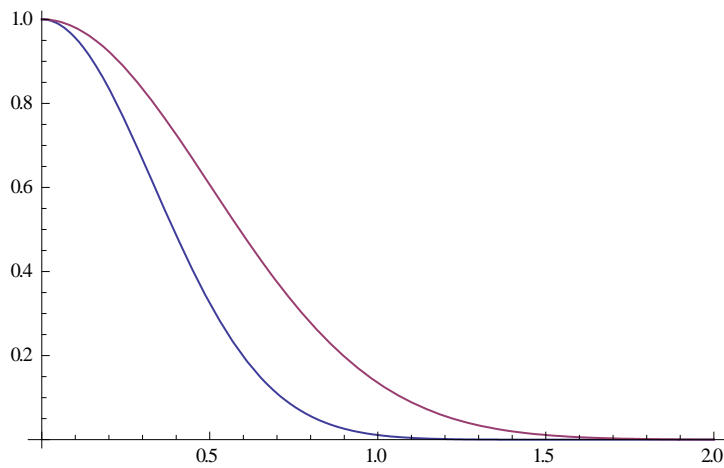


**Σχήμα 2.7.5:** Σύμφωνα με την γραφική παράσταση έχουμε μια γνησίως φθίνουσα συνάρτηση επιβίωσης με πιο βαριά ουρά με τιμή παραμέτρου  $\sigma=1/2$  για το διάστημα  $(0,2)$ . Παρατηρούμε πως συγκλίνει πιο γρήγορα στο μηδέν.

Μια πιο εκτενή ανάλυση για την συνάρτηση επιβίωσης παρατηρούμε στο παρακάτω σχήμα



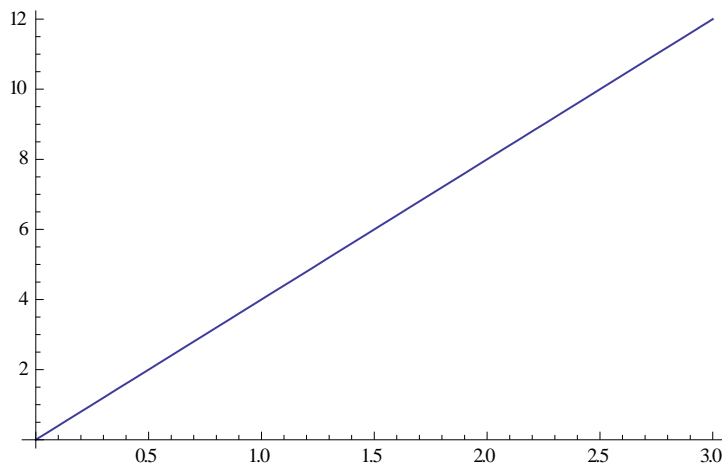
**Σχήμα 2.7.6:** Σύμφωνα με το σχήμα για παράμετρο  $\sigma=1/2$  (κόκκινη) έχουμε μια παράσταση που συγκλίνει πιο γρήγορα στο μηδέν από ότι η συνάρτηση με  $\sigma=2$  (μπλε)



**Σχήμα 2.7.7:** Σύμφωνα με το σχήμα για παράμετρο  $\sigma=1/2$  (κόκκινη) μια συνάρτηση που συγκλίνει πιο αργά στο μηδέν από ότι η παράσταση με  $\sigma=1/3$  (μπλε).

**Άρα διαπιστώνουμε πως καθώς μεγαλώνει η τιμή του  $\sigma$  έχουμε και πιο βαριά ουρά για την κατανομή.**

Έπειτα παρουσιάζουμε την γραφική παράσταση της βαθμίδας αποτυχίας για την κατανομή Rayleigh.



**Σχήμα 2.7.8:** Γραφική παράσταση βαθμίδας αποτυχίας για την κατανομή Rayleigh για το διάστημα  $(0,3)$  και  $\sigma=1/2$ .

Όπως είναι αναμενόμενο και προκύπτει από τη σχέση (2.7.3), η βαθμίδα αποτυχίας είναι μια γραμμική αύξουσα συνάρτηση. Άρα η κατανομή Rayleigh ανήκει στην οικογένεια κατανομών IFR.

## 3. Η κατανομή Maxwell–Boltzmann στην αναλογιστική επιστήμη

### 3.1 Εισαγωγή

Στο συγκεκριμένο κεφάλαιο της εργασίας θα παρουσιαστεί διεξοδικά η κατανομή Maxwell –Boltzmann. Πρόκειται για μια κατανομή με πολλές εφαρμογές σε πολλούς κλάδους επιστημών όπως στην φυσική και τον αναλογισμό η οποία είναι μια κατανομή μεγέθους ενός τρισδιάστατου τυχαίου διανύσματος του οποίου οι συντεταγμένες είναι ανεξάρτητες και ταυτόσημα κατανεμημένες. Η κατανομή έχει πολλές εφαρμογές σε ρυθμίσεις όπου τα μεγέθη των κανονικών μεταβλητών είναι σημαντικά, ιδιαίτερα στη φυσική. Όπως θα αναλυθεί και αναλυτικότερα παρακάτω η κατανομή Maxwell συνδέεται στενά με την κατανομή Rayleigh, η οποία ρυθμίζει το μέγεθος ενός δισδιάστατου τυχαίου διανύσματος του οποίου οι συντεταγμένες είναι ανεξάρτητες, κατανέμονται πανομοιότυπα, με την κανονική κατανομή και έχουν μέση τιμή 0.

Πιο αναλυτικά, η κατανομή Maxwell-Boltzmann είναι μια ειδική περίπτωση της γενικευμένης κατανομής Rayleigh για  $n=3$ . Η κατανομή αυτή έχει πάρει το όνομά της από τον James Clerk Maxwell (1831) ο οποίος ήταν Σκωτσέζος Θεωρητικός Φυσικός. Το πιο επιφανές επίτευγμά του ήταν η διατύπωση μιας σειράς εξισώσεων που ένωσαν προηγουμένως άσχετες παρατηρήσεις, πειράματα και εξισώσεις ηλεκτρισμού, μαγνητισμού, και οπτικής σε μία συνεπή θεωρία. Ο Ludwig Boltzmann (1844) ο οποίος ασχολήθηκε επίσης με αυτή την κατανομή, ήταν Αυστριακός θεωρητικός φυσικός και φιλόσοφος και υπήρξε από τους θεμελιωτές της Στατιστικής Φυσικής και Θερμοδυναμικής. Η εν λόγω κατανομή που δημιουργήθηκε από τους δυο αυτούς μεγάλους επιστήμονες είναι πρακτικά ένα στατιστικό μέσο περιγραφής των όψεων της κινητικής θεωρίας των αερίων και χρησιμοποιήθηκε τόσο στη επιστήμη της στατιστικής όσο και της φυσικής. Πιο αναλυτικά, χρησιμοποιήθηκε στη φυσική για την περιγραφή ταχυτήτων για εξατομικευμένα αέρια και σωματίδια τα οποία κινούνται ελεύθερα στο χώρο χωρίς να υπάρχει αλληλεπίδραση μεταξύ τους εκτός από τις συγκρούσεις στις οποίες ανταλλάζουν ενέργεια και ορμή μεταξύ τους ή με το θερμικό περιβάλλον τους. Πρακτικά, σε αυτό το στάδιο έχει επέλθει θερμική ισορροπία. Επιπροσθέτως, ενώ η κατανομή προήλθε αρχικά από τον Maxwell, αργότερα ο Boltzmann πραγματοποίησε και διεξήγαγε πολλές έρευνες για την φυσική προέλευση αυτής. Η κατανομή αυτή εξαρτάται από την θερμοκρασία του συστήματος και τη μάζα ενός σωματιδίου. Η κατανομή αυτή ισχύει για τα ιδανικά αέρια τα οποία δεν είναι τα κοινά αέρια αλλά μια υποκατηγορία αυτών. Στα πραγματικά αέρια υπάρχουν διάφορα εφέ που μπορούν να κάνουν την ταχύτητα της κατανομής διαφορετική από ότι στη κατανομή της Maxwell – Boltzmann. Παρόλα αυτά, τα αέρια αυτά πολλές φορές μπορούν να συμπεριφερθούν παρόμοια με τα

ιδανικά και ταυτόχρονα η κατανομή της ταχύτητας του Maxwell αποτελεί μια εξαιρετική προσέγγιση και αποτελεί την βάση της κινητικής θεωρίας των αερίων .

### 3.2 Η κατανομή Maxwell –Boltzmann

**Ορισμός 1:** για την Maxwell-Boltzmann

Αν υποθέσουμε ότι έχουμε  $Z_1, Z_2, Z_3$ , είναι τυχαίες ανεξάρτητες μεταβλητές. Τότε το μέγεθος  $R = \sqrt{Z_1^2 + Z_2^2 + Z_3^2}$  του διανύσματος  $(Z_1, Z_2, Z_3)$  ακολουθεί την Maxwell κατανομή.

Στο κομμάτι αυτό ορίζουμε τις συναρτήσεις της γενικευμένης Maxwell Boltzmann όπου με παράμετρο κλίμακας  $a$  έχουμε τον τύπο της συνάρτησης κατανομής :

$$F(x) = 2\Phi\left(\frac{x}{a}\right) - \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} - 1,$$

Όπου  $\Phi$  είναι η συνάρτηση κατανομής της  $N(0,1)$  (3.1.1)

**Για  $a=1$  προκύπτει ως ειδική περίπτωση ο τύπος της συνάρτησης κατανομής ,**

$$F(x) = 2\Phi(x) - \sqrt{\frac{2}{\pi}} x e^{-\frac{x^2}{2}} - 1, x \in [0, \infty)$$

Από τον τύπο που γνωρίζουμε ήδη  $F(x) = 1 - S(x)$  είναι προφανές ότι η συνάρτηση επιβίωσης της κατανομής δίνεται από τον τύπο:

$$S(x) = 1 - \left( 2\Phi\left(\frac{x}{a}\right) - \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} - 1 \right) \Rightarrow$$

$$S(x) = -2\Phi\left(\frac{x}{a}\right) + \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} + 2$$

όπου  $\Phi$  είναι η συνάρτηση κατανομής της  $N(0,1)$  (3.1.2)

**Για  $a=1$  προκύπτει ο τύπος της συνάρτησης επιβίωσης,**

$$S(x) = -2\Phi(x) + \sqrt{\frac{2}{\pi}} x e^{-\frac{x^2}{2}} + 2$$

Με παραγωγή προκύπτει ότι η συνάρτηση πυκνότητας πιθανότητας της συνεχούς κατανομής δίνεται από τον τύπο

$$f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3}, x \in (0, \infty) \quad (3.1.3)$$

**Για  $a=1$  προκύπτει ο τύπος της συνάρτησης πυκνότητας πιθανότητας,**

$$f(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-\frac{x^2}{2}}$$

Από την παράγραφο 2.6 γνωρίζοντας μια γενικότερη μορφή της συνάρτησης πυκνότητας πιθανότητας και αντικαθιστώντας όπου  $n=3$  προκύπτει το εξής:

$$f(x) = (2x^{n-1}) (e^{-x^2/2\sigma^2}) / ((2\sigma^2)^{n/2}) \Gamma(n/2) \Rightarrow$$

$$f(x) = (2x^2) (e^{-x^2/2\sigma^2}) / (2\sigma^3) \Gamma(3/2) \quad (3.1.4)$$

Για την μέση τιμή και την διακύμανση της εν λόγω κατανομής ισχύουν αντίστοιχα οι τύποι :

$$E(X) = 2a \sqrt{\frac{2}{\pi}} \quad (3.1.5)$$

$$Var(X) = a^2 \left(3 - \frac{8}{\pi}\right) \quad (3.1.6)$$

Για  $\alpha=1$  προκύπτουν οι αντίστοιχοι τύποι μέσης τιμής και διακύμανσης,

$$E(X) = 2 \sqrt{\frac{2}{\pi}} \cong 1,5958$$

$$Var(X) = 3 - \frac{8}{\pi}$$

(<http://www.math.uah.edu/stat/special/Maxwell.html>)

### Εφαρμογή 3.1.1

Από τον γνωστό τύπο  $E(X) = \int_0^{\infty} xf(x)dx$  , προκύπτει με χρήση του αλγεβρικού πακέτου Mathematica

$$E(X) = \int_0^{\infty} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^3 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3} dx = 2a \sqrt{\frac{2}{\pi}}$$

$$\text{Integrate}[(x^3) * \text{Exp}\left[-\frac{x^2}{2a^2}\right] * (1/a^3) * ((2/\text{Pi})^{(1/2)}), \{x, 0, \text{Infinity}\}]$$

$$2a \sqrt{\frac{2}{\pi}}$$

Επίσης η δεύτερη ροπή γύρω από το μηδέν είναι,

$$E(X^2) = \int_0^{\infty} x^2 f'(x) dx = \int_0^{\infty} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^4 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3} dx = a^2 \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right)$$

Για την δεύτερη ροπή  $E(X^2)$  η εντολή στο Mathematica είναι η εξής

$$\text{Integrate}[(x^4) * \text{Exp}\left[-\frac{x^2}{2a^2}\right] * (1/a^3) * ((2/\text{Pi})^{(1/2)}), \{x, 0, \text{Infinity}\}]$$

$$E(X^2) = \frac{3}{\left(\frac{1}{a^2}\right)^{5/2} a^3}$$

Άρα από τον τύπο  $Var(X) = E(X)^2 - E^2(X)$  προκύπτει ο παραπάνω τύπος (3.1.6).

Έπειτα παρουσιάζουμε διάφορα μεγέθη για την κατανομή Maxwell-Boltzmann και πιο συγκεκριμένα την ροπή κ-τάξης η οποία ισούται με :

$$E(X^k) = a^k * \frac{2^{\frac{k}{2}+1}}{\sqrt{\pi}} * \Gamma\left(\frac{(k+3)}{2}\right), k \in \mathbb{N}. \quad (3.1.7)$$

Η επαλήθευση και ταυτόχρονα η εύρεση της σχέσης 3.1.7 πραγματοποιείται και με την βοήθεια του Mathematica με την παρακάτω διαδικασία:

Assuming[a > 0 && k > 0, Integrate[x^k \* f[x], {x, 0, Infinity}]]

$$\frac{2^{1+\frac{k}{2}} a^k \text{Gamma}\left[\frac{3+k}{2}\right]}{\sqrt{\pi} S}$$

Δηλαδή υποθέτοντας ότι το  $a > 0$  και το  $k > 0$  επίσης ισχύσει ο παραπάνω τύπος.

Όσον αφορά την βαθμίδα αποτυχίας της κατανομής από τον τύπο (2.1.2)

καταλήγουμε στον τύπο  $\lambda(x) = \frac{f(x)}{S(x)}$  με αντικατάσταση αλλά και βάσει του προγράμματος καταλήγουμε πως η βαθμίδα αποτυχίας ισούται με

$$\lambda(t) = \frac{\left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3}}{2\Phi\left(\frac{x}{a}\right) + \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} + 2} = \frac{e^{-\frac{x^2}{2a^2}} \sqrt{\frac{2}{\pi}} x^2}{a^3 \left( \frac{e^{-\frac{t^2}{2a^2}} \sqrt{\frac{2}{\pi}} t}{a} + \Phi\left[\frac{t}{\sqrt{2}a}\right] \right)}$$

Επιπροσθέτως η συνάρτηση της βαθμίδας αποτυχίας είναι πάντα μια αύξουσα συνάρτηση όπως αναφέρεται στο άρθρο των R.C.Curta and D.M Bradley με τίτλο «Representing the Mean Residual Life in Terms of the Failure Rate». Αυτό εξηγείται από το γεγονός ότι όσο περνά το χρονικό διάστημα η πιθανότητα να θανάτου κάποιου ατόμου αυξάνεται συνεχώς.

### Απόδειξη της μονοτονίας της συνάρτησης $\lambda(t)$

Για να αποδείξουμε το παραπάνω αρκεί να παραγωγίσουμε την βαθμίδα αποτυχίας και να διαπιστώσουμε πως η παράγωγος της είναι θετική.

Έτσι με την βοήθεια του προγράμματος Mathematica γίνεται επαλήθευση με την εξής υπολογιστική διαδικασία

Αρχικά υπολογίζουμε την βαθμίδα αποτυχίας με τον τύπο  $\lambda(x) = \frac{f(x)}{S(x)}$ , άρα:

$$t[x] := f[x]/\text{tail}F[t] = \frac{e^{-\frac{x^2}{2a^2}} \sqrt{\frac{2}{\pi}} x^2}{a^3 \left( \frac{e^{-\frac{x^2}{2a^2}} \sqrt{\frac{2}{\pi}} x}{a} + \text{Erfc}\left[\frac{x}{\sqrt{2}a}\right] \right)} = \frac{2x^2}{a^2 \left( 2x + a e^{\frac{x^2}{2a^2}} \sqrt{2\pi} \text{Erfc}\left[\frac{x}{\sqrt{2}a}\right] \right)}$$

Έπειτα αν παραγωγίσουμε την παραπάνω βαθμίδα αποτυχίας τότε προκύπτει

$$(t[x])' = \frac{x(4ax + e^{\frac{x^2}{2a^2}}\sqrt{2\pi}(2a^2 - x^2)\text{Erfc}[\frac{x}{\sqrt{2a}}])}{a^3\pi(\sqrt{\frac{2}{\pi}}x + ae^{\frac{x^2}{2a^2}}\text{Erfc}[\frac{x}{\sqrt{2a}}])}$$

Όπου πρόκειται για θετική ποσότητα και επαληθεύεται το παραπάνω, όπου Erfc είναι η αντίστροφη της κανονικής κατανομής.

**Η ροπογεννήτρια** της συγκεκριμένης κατανομής υπολογίζεται από τον τύπο

$$M_x(t) = \sqrt{\frac{2}{\pi}}at + 2(1 + a^2t^2)e^{\frac{b^2t^2}{2}} \Phi(at). \quad (3.1.8)$$

Για  $a=1$  έχουμε προφανώς

$$M_x(t) = \sqrt{\frac{2}{\pi}}t + 2(1 + t^2)e^{\frac{b^2t^2}{2}} \Phi(t) \quad (3.1.9)$$

### Απόδειξη 3.2.1

Η ροπογεννήτρια συνάρτηση μίας συνεχούς τυχαίας μεταβλητής  $X$  έχουμε δει ότι υπολογίζεται από την σχέση:

$$M_x(t) = E(e^{tx})$$

Επομένως για την ροπογεννήτρια της Maxwell-Boltzmann με παράμετρο  $\kappa$  θα ισχύει:

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \Rightarrow$$

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2}} dx \Rightarrow$$

$$M_x(t) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^{\infty} x^2 e^{tx} e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} e^{\frac{t^2}{2}} \int_0^{\infty} x^2 e^{-\frac{(x-t)^2}{2}} dx$$

Αν θέσουμε  $z=x-t$  τότε προκύπτει

$$M_x(t) = \sqrt{\frac{2}{\pi}} e^{\frac{t^2}{2}} \int_{-t}^{\infty} (z+t)^2 e^{-\frac{z^2}{2}} dz = \sqrt{\frac{2}{\pi}} e^{\frac{t^2}{2}} \int_{-t}^{\infty} (z^2 + 2tz + t^2) e^{-\frac{z^2}{2}} dz$$

Χρησιμοποιώντας την αλλαγή της μεταβλητής  $z \Rightarrow 1/\sqrt{2\pi} e^{-\frac{z^2}{2}}$  και γνωρίζοντας πως  $1 - \Phi(-t) = \Phi(t)$  από την κανονική κατανομή θα προκύψει με τις ανάλογες πράξεις ο τύπος 3.1.8.

Επιπλέον με το αλγεβρικό πρόγραμμα MATHEMATICA καταλήξαμε στον παραπάνω τύπο ως εξής:

$$M[t] := \text{Integrate}[\text{Exp}[x * t] * f[x], \{x, 0, \text{Infinity}\}]$$

$$\sqrt{\frac{2}{\pi}} t + e^{\frac{t^2}{2}} (1 + t^2) (1 + \text{Erf}[\frac{t}{\sqrt{2}}])$$

Simplify[M[t]]

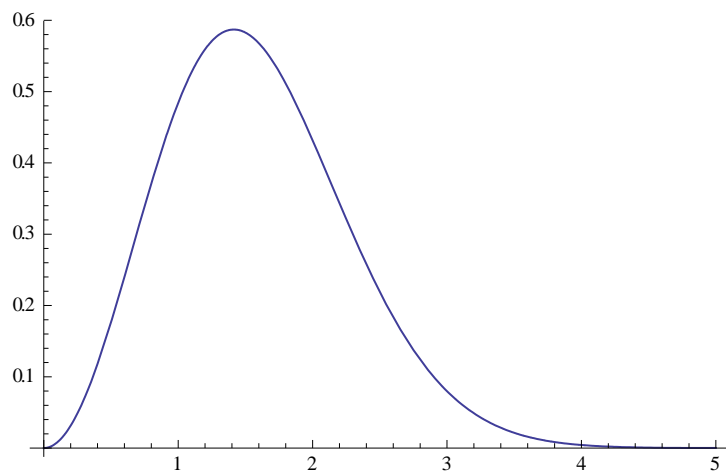
$$\sqrt{\frac{2}{\pi}} t + e^{\frac{t^2}{2}} (1 + t^2) (1 + \text{Erf}[\frac{t}{\sqrt{2}}])$$

Το παραπάνω αποτέλεσμα είναι το ίδιο με τον τύπο (3.1.9) .

### Γραφικές παραστάσεις της εν λόγω κατανομής

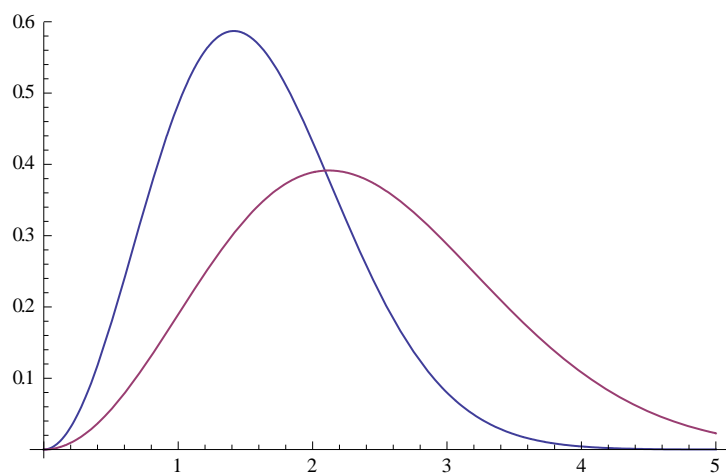
Παρακάτω παρουσιάζονται οι γραφικές παραστάσεις συνάρτησης πιθανότητας , η συνάρτηση κατανομής και επιβίωσης της κατανομής Maxwell-Boltzmann αλλά και η βαθμίδα αποτυχίας της εν λόγω κατανομής.

Πιο συγκεκριμένα έχουμε ,



**Σχήμα 3.1.1:**Συνάρτηση πυκνότητας πιθανότητας για  $\alpha=1$ .

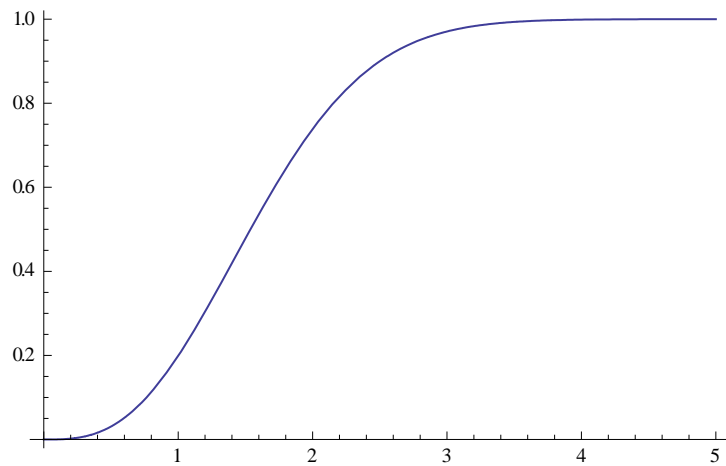
Για να δούμε την επίδραση της παραμέτρου  $\alpha$  στην κατανομή παρουσιάζουμε την συνάρτηση πυκνότητας πιθανότητας για διάφορες τιμές του  $\alpha$ .



**Σχήμα 3.1.2:**Γραφική παράσταση συνάρτησης πυκνότητας πιθανότητας για  $\alpha=1$ (μπλε), $\alpha=2$ (κόκκινο)

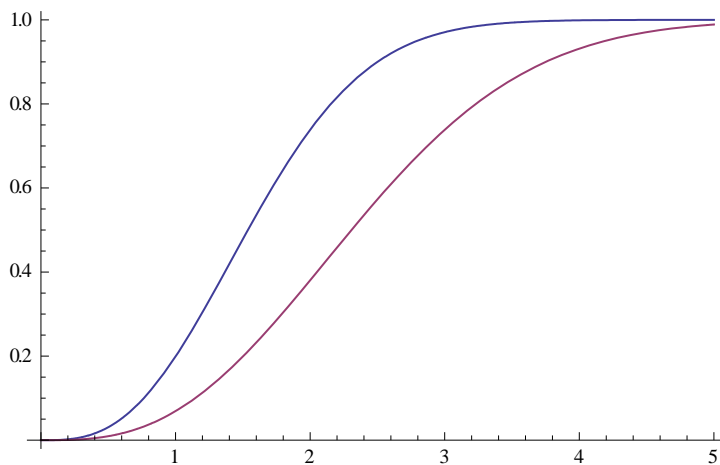


Για την συνάρτηση κατανομής παρατίθεται το παρακάτω γράφημα με παράμετρο της κατανομής  $\alpha=1$ .

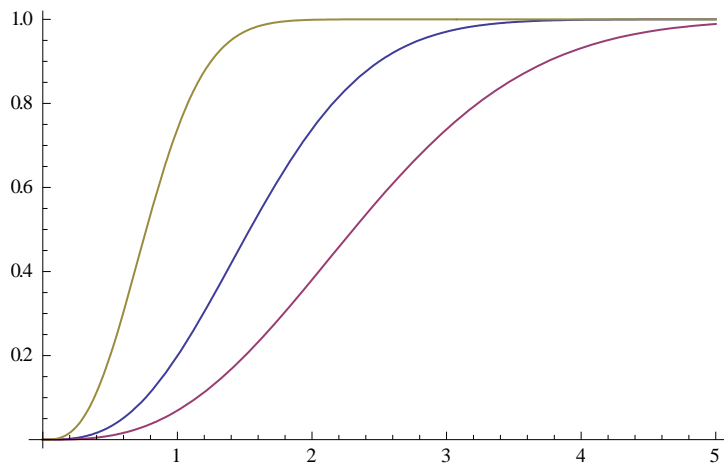


**Σχήμα 3.1.3:** Γραφική παράσταση Συνάρτησης Κατανομής με παράμετρο  $\alpha$ .

Στην συνέχεια δίνουμε για διάφορες τιμές της παραμέτρου την γραφική παράσταση που προκύπτει. Πιο συγκεκριμένα έχουμε,



**Σχήμα 3.1.4:** Γραφική παράσταση συνάρτησης κατανομής με παραμέτρους  $\alpha=1$  (μπλε),  $\alpha=3/2$  (κόκκινο).

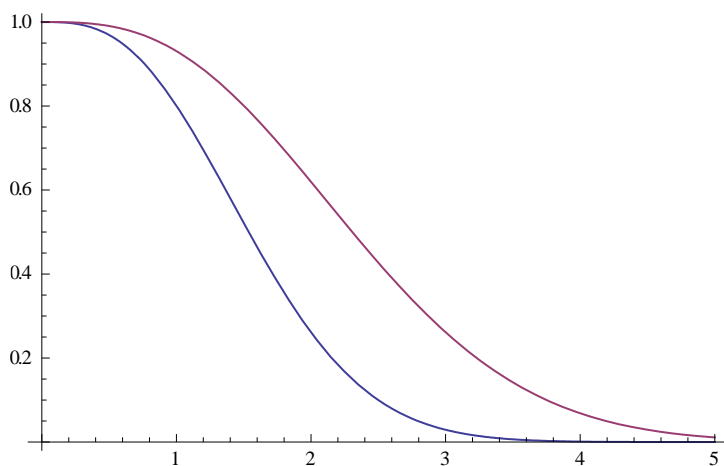


**Σχήμα 3.1.5:** Γραφική παράσταση συνάρτηση κατανομής με παραμέτρους  $\alpha=1$ (μπλε),  $\alpha=3/2$ (κόκκινο),  $\alpha=1/2$ (κίτρινο).

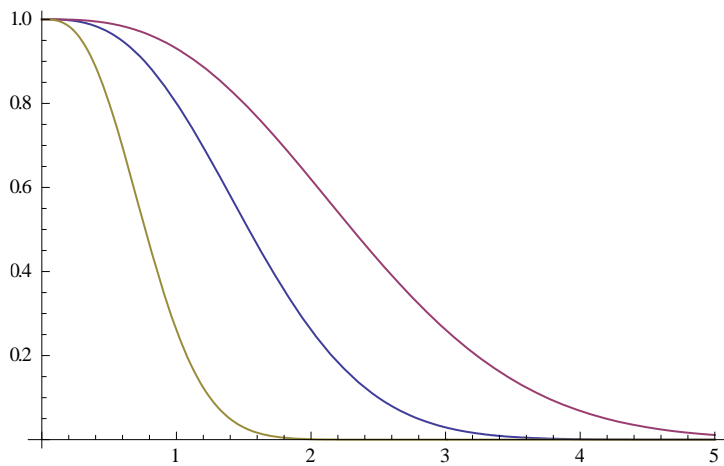
Εδώ παρουσιάζουμε δυο σχήματα συνδυαστικά για να δούμε ακριβώς πώς παρίσταται η συνάρτηση κατανομής για διάφορες τιμές των παραμέτρων . Στο πρώτο σχήμα έχουμε παραμέτρους  $\alpha=1$ (μπλε),  $\alpha=2$ (κόκκινο) και στο δεύτερο σχήμα παραμέτρους  $\alpha=1$ (μπλε),  $\alpha=1/2$ ( κίτρινο),  $\alpha=3/2$ (κόκκινο). Έτσι διαπιστώνουμε πως όσο μικρότερο είναι το  $\alpha$  τόσο μεγαλύτερες τιμές παίρνει η συνάρτηση κατανομής.

Όσον αφορά την Συνάρτηση Επιβίωσης της κατανομής Maxwell-Boltzmann παρουσιάζει ιδιαίτερο ενδιαφέρον . Εδώ θα παρουσιάσουμε κάποια σχεδιαγράμματα που μας δίνουν την δυνατότητα να διαπιστώσουμε πόσο αργά ή γρηγορά συγκλίνει στο μηδέν η συνάρτηση επιβίωσης και την ουρά ως προς το βάρος της.

Πιο συγκεκριμένα έχουμε,



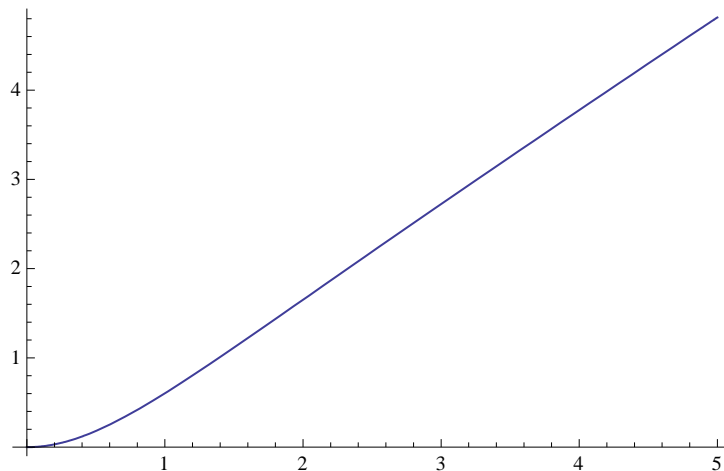
**Σχήμα 3.5.1:** Γραφική παράσταση Συνάρτησης Επιβίωσης με παράμετρο  $\alpha=1$ (μπλε),  $\alpha=3/2$ (κόκκινο).



**Σχήμα 3.1.7:** Γραφική παράσταση Συνάρτησης Επιβίωσης με παράμετρο  $\alpha=1$ (μπλε),  $\alpha=3/2$ (κόκκινο),  $\alpha=1/2$ (κίτρινο).

Εύλογα μπορεί κανείς να παρατηρήσει βάσει των παραπάνω σχημάτων , πως η συνάρτηση επιβίωσης όσο μικρότερη είναι η παράμετρος  $\alpha$  τόσο πιο γρήγορα συγκλίνει στο μηδέν και έχει αντίστοιχα πιο ελαφριά ουρά. Χαρακτηριστικό παράδειγμα είναι για  $\alpha=1/2$  που είδαμε στο παραπάνω γράφημα όπου συγκλίνει γρηγορότερα από  $\alpha=1$  και  $\alpha=3/2$ .

Όσον αφορά την παράσταση της βαθμίδας αποτυχίας για την εν λόγω κατανομή έχουμε την εξής γραφική παράσταση :



**Σχήμα 3.1.8:**Γραφική παράσταση για την βαθμίδα αποτυχίας της κατανομής Maxwell-Boltzmann.

Στο παραπάνω γράφημα καταλήξαμε με την χρήση του αλγεβρικού προγράμματος MATHEMATICA .

**Εφαρμογή 3.1.3:**

Για  $\alpha=1$  έχουμε με την χρήση του αλγεβρικού προγράμματος MATHEMATICA το εξής αποτέλεσμα.

$$l[t] := f[x]/S[x] = \frac{e^{-\frac{x^2}{2}} \sqrt{\frac{2}{\pi}} x^2}{2 + e^{-\frac{x^2}{2}} \sqrt{\frac{2}{\pi}} x - \operatorname{Erfc}\left[-\frac{x}{\sqrt{2}}\right]}$$

Άρα , με  $\text{Plot}[l[t], \{x, 0, 5\}]$  καταλήγουμε στο σχήμα 3.1.8 που προκύπτει μια σχεδόν γραμμική γραφική παράσταση .

### Μετασχηματισμός της Κατανομής Maxwell-Boltzmann

Αν υποθέσουμε  $Z_1, Z_2, Z_3$  είναι τυχαίες μεταβλητές που ακολουθούν κανονική κατανομή. Όπως είδαμε το μέγεθος  $R = \sqrt{Z_1^2 + Z_2^2 + Z_3^2}$  που είναι συνάρτηση του διανύσματος του διανύσματος  $(Z_1, Z_2, Z_3)$  που ακολουθεί την Maxwell κατανομή.

Η συνάρτηση κατανομής όπως έχουμε ήδη δει από τον τύπο (3.1.1) ισούται με

$$F(x) = 2\Phi\left(\frac{x}{a}\right) - \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} - 1, \text{ όπου } X \sim \text{MB}(a) \text{ για κάθε } x \in [0, \infty).$$

Αν όμως  $2X \sim \text{MB}(2a)$  προκύπτει τότε ο μετασχηματισμός της κατανομής αυτής ως εξής:

- $F(x) = 2\Phi\left(\frac{x}{2a}\right) - \frac{1}{2a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{4a^2}} - 1$ , για την Συνάρτηση Κατανομής.
- $f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{4a^2}} \frac{1}{2a^3}$ , για την Συνάρτηση Πυκνότητας Πιθανότητας.

(<http://www.math.uah.edu/stat/special/Maxwell.html>)

(Norman L Johnson, Samuel Kotz e.t.l. (2005)

Παρακάτω παρουσιάζεται ένας συγκεντρωτικός πίνακας των βασικών τύπων της κατανομής Maxwell-Boltzmann για καλύτερη κατανόηση:

Maxwell-Boltzmann	Scale parameter $a$	Scale parameter $a=1$
Συνάρτηση πυκνότητας	$f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3}$	$f(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-\frac{x^2}{2}}$
Συνάρτηση Κατανομής	$F(x) = 2\Phi\left(\frac{x}{a}\right) - \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} - 1,$	$F(x) = 2\Phi(x) - \sqrt{\frac{2}{\pi}} x e^{-\frac{x^2}{2}} - 1$
Συνάρτηση Επιβίωσης	$S(x) = -2\Phi\left(\frac{x}{a}\right) + \frac{1}{a} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2a^2}} + 2$	$S(x) = -2\Phi(x) + \sqrt{\frac{2}{\pi}} x e^{-\frac{x^2}{2}} + 2$
Μέση Τιμή	$E(X) = 2a \sqrt{\frac{2}{\pi}}$	$E(X) = 2 \sqrt{\frac{2}{\pi}} \cong 1,5958$
Διακύμανση	$\text{Var}(X) = a^2 \left(3 - \frac{8}{\pi}\right)$	$\text{Var}(X) = 3 - \frac{8}{\pi}$
Ροπογεννήτρια	$M_x(t) = \sqrt{\frac{2}{\pi}} a t + 2(1 + a^2 t^2) e^{\frac{b^2 t^2}{2}} \Phi(at) .$	$M_x(t) = \sqrt{\frac{2}{\pi}} t + 2(1 + t^2) e^{\frac{b^2 t^2}{2}} \Phi(t)$

### 3.3 Αντίστροφη Maxwell-Boltzmann κατανομή (Inverse Maxwell – Boltzmann)

#### Εισαγωγή

Εάν μια τυχαία μεταβλητή ακολουθεί μια κατανομή τότε ο αντίστροφος της τυχαίας μεταβλητής ακολουθεί μια αντίστροφη κατανομή. Έτσι αυτό ισχύει και στην κατανομή που αναλύουμε, την Maxwell-Boltzmann. Γενικότερα υπάρχουν πολλές κατανομές οι οποίες αντιστρέφονται και έχουν μελετηθεί διεξοδικά από επιστήμονες.

Αν μια τυχαία μεταβλητή  $X$  ακολουθεί την κατανομή Maxwell-Boltzmann βάσει του τύπου (3.1.3) η συνάρτηση πυκνότητας πιθανότητας υπολογίζεται με τον τύπο

$$f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3}, \quad x \in (0, \infty),$$

Όπου για την παράμετρο ισχύει  $a > 0$  και πρόκειται για την παράμετρο κλίμακας της κατανομής.

Η Maxwell-Boltzmann έχει κάποια μέτρα που έχουμε αναλύσει και καταγράψει στο προηγούμενο κεφάλαιο και πιο συγκεκριμένα οι τύποι της μέσης τιμής και της διακύμανσης (3.1.5) και (3.1.6) αντίστοιχα.

Αν η τυχαία μεταβλητή  $X \sim MB(a)$  τότε προκύπτει πως μια τυχαία μεταβλητή  $Y=1/X$  και ονομάζεται αντίστροφη μεταβλητή, ακολουθεί τη αντίστροφη Maxwell-Boltzmann. Η συνάρτηση πυκνότητας πιθανότητας υπολογίζεται με μετασχηματισμό ως εξής:

$$f(y) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{y^2} e^{-\frac{1}{2a^2y^2}} \frac{1}{a^3}, \quad y \in (0, \infty) \quad (3.2.1)$$

Έτσι η συνάρτηση κατανομής δίνεται από τον τύπο,

$$F(y) = 2\Phi\left(\frac{1}{ay}\right) - \frac{1}{a} * \sqrt{\frac{2}{\pi}} * \frac{1}{y} e^{-\frac{1}{2a^2y^2}} - 1 \quad (3.2.2)$$

Και η συνάρτηση επιβίωσης προφανώς από τον τύπο,

$$S(x) = -2\Phi\left(\frac{1}{ay}\right) + \frac{1}{a} * \sqrt{\frac{2}{\pi}} * \frac{1}{y} e^{-\frac{1}{2a^2y^2}} + 2 \quad (3.2.3)$$

Τα μέτρα της κατανομής όπως η μέση τιμή και η διακύμανση υπολογίζονται με τους παρακάτω τύπους αντίστοιχα:

$$E(Y^r) = \mu^r = \int_0^\infty y^r f(y) dy = \frac{2}{\sqrt{\pi}} a^{\frac{r}{2}} \Gamma\left(\frac{-r+3}{2}\right)$$

$$E(Y) = \frac{2}{\sqrt{\pi}a} \quad (3.2.4)$$

$$Var(Y) = \frac{2(\pi-2)}{\alpha\pi} \quad (3.2.5)$$

(K.S Singh and R.S. Srivastava 2014)

## 4. Εφαρμογή της κατανομής Maxwell–Boltzmann σε πραγματικά αναλογιστικά δεδομένα

### 4.1 Εισαγωγή

Στην παράγραφο αυτή της εργασίας με την βοήθεια κάποιων πραγματικών αναλογιστικών δεδομένων θα εξετάσουμε με την βοήθεια της διαδικασίας της καλής προσαρμογής (**Kolmogorov Smirnov**) αν προσαρμόζονται τα δεδομένα μας στην εν λόγω κατανομή. Τα πραγματικά αυτά δεδομένα έχουν διατεθεί από ασφαλιστική εταιρία. Πρόκειται για 168 τιμές όπου είναι ο υπολειπόμενος χρόνος ζωής ατόμων ηλικίας 60 ετών. Δηλαδή το διάστημα το οποίο θα ζήσουν τα άτομα από την ηλικία των 60 και έπειτα. Ο έλεγχος αυτός είναι ίσως ο πιο χρήσιμος από τους άλλους ελέγχους καλής προσαρμογής, (όπως ο  $X^2$  ο οποίος είναι σχεδιασμένος για δεδομένα σε ονομαστική κλίμακα) και, εν μέρει επειδή ο έλεγχος Kolmogorov στηρίζεται σε μια στατιστική συνάρτηση που δίνει την δυνατότητα ορισμού μιας ζώνης εμπιστοσύνης (confidence band) για την άγνωστη κατανομή. Υπενθυμίζεται ότι, προκειμένου να εφαρμόσουμε το κριτήριο  $X^2$  στην περίπτωση συνεχών δεδομένων, θα πρέπει να θεωρήσουμε  $k$  κελιά και να εξετάσουμε τη διαφορά μεταξύ αναμενόμενων (υπό την  $H_0$ ) και παρατηρούμενων συχνοτήτων σε αυτά. Όπως είναι φανερό, η διακριτοποίηση αυτή οδηγεί σε απώλεια πληροφορίας (διότι κρατάμε μόνο το πλήθος των παρατηρήσεων σε κάθε κατηγορία και όχι τις ακριβείς τιμές των παρατηρήσεων) και άρα σε μειωμένη ισχύ του ελέγχου. Έτσι ο έλεγχος Kolmogorov Smirnov χρησιμοποιείται όταν θέλουμε να ελέγξουμε αν ένα τυχαίο δείγμα παρατηρήσεων προέρχεται από μια γνωστή κατανομή  $F_0$ , δηλαδή αν θέλουμε να ελέγξουμε την μηδενική υπόθεση

$$H_0: F(x) = F_0(x), \text{ για κάθε } x \in \mathbb{R}$$

έναντι της μηδενικής υπόθεσης

$$H_a: F(x) \neq F_0(x) \text{ για κάποιο } x$$

όπου  $F(x)$  είναι η πραγματική συνάρτηση κατανομής του δείγματος και  $F_0(x)$  είναι η γνωστή κατανομή. Ο έλεγχος γίνεται με την χρήση του στατιστικού

$$D_n = \max |F_n(x) - F_0(x)|,$$

όπου  $F_n(x)$  είναι η εμπειρική συνάρτηση κατανομής, η οποία ορίζεται ως

$$F_n(x) = \frac{1}{n} (\text{αριθμός των } X_i \text{ που είναι μικρότερα ή ίσα του } x)$$

Στην πραγματικότητα  $F_n(x)$  είναι η σχετική συχνότητα του πλήθους των τυχαίων μεταβλητών που είναι μικρότερες ή ίσες του  $x$ . Το στατιστικό K-S υπολογίζει πόσο

μακριά βρίσκεται η εμπειρική κατανομή από την θεωρητική κατανομή. Ως κρίσιμες τιμές για τον έλεγχο της μηδενικής υπόθεσης  $H_0$ , προσεγγιστικά χρησιμοποιούνται οι τιμές

$$\frac{1,22}{\sqrt{n}} \text{ για } \alpha = 0,10, \quad \frac{1,36}{\sqrt{n}} \text{ για } \alpha = 0,05 \text{ και } \frac{1,63}{\sqrt{n}} \text{ για } \alpha = 0,01$$

όταν το  $n$  παίρνει μεγάλες τιμές.

Δηλαδή αν η τιμή της  $D_n$  είναι μεγαλύτερη ( $>$ ) της κρίσιμης τιμής για ένα συγκεκριμένο επίπεδο σημαντικότητας (π.χ.  $\alpha=0,01$ ) απορρίπτουμε την μηδενική υπόθεση.

(Γ.Πιτσέλης 2015)

*Λίγα λόγια για το P-Value ή Significance value ,*

Αν η περιοχή απόρριψης της  $H_0$  είναι της μορφής  $T(x) > c$  τότε το  $p$ -value των τιμών  $x$  του δείγματος είναι η πιθανότητα

$$p\text{-value} = P(T(X) > T(x) / H_0) = 1 - F_{T|H_0}(T(x))$$

η οποία μπορεί να θεωρηθεί ότι εκφράζει την πιθανότητα να εμφανιστεί ένα τόσο ή ακόμη και πιο «ακραίο» δείγμα από αυτό που εμφανίστηκε, δεδομένου ότι ισχύει η  $H_0$  (συνήθως η  $T(X)$  είναι συνεχείς οπότε μπορεί να θεωρηθεί ότι έχουμε  $\geq$  μέσα στην παραπάνω πιθανότητα). Διαισθητικά, αν το  $p$ -value είναι «κοντά» στο 0 τότε συμπεραίνουμε ότι είναι «απίθανο», δεδομένης της  $H_0$ , να εμφανιστεί αυτό το δείγμα, και όπως είναι φυσικό φτάνουμε στο συμπέρασμα ότι μάλλον δεν πρέπει να ισχύει η  $H_0$ . Πράγματι, αν το  $p$ -value  $< \alpha$  απορρίπτουμε την  $H_0$  διότι

$$T(x) > c = F_{T|H_0}^{-1}(1-\alpha) \Rightarrow F_{T|H_0}(T(x) | H_0) = 1 - F_{T|H_0}(T(x))$$

Επομένως αντί να εξετάζουμε αν  $T(x) > c$ , ισοδύναμα εξετάζουμε:

- αν το  $p$ -value  $< \alpha$  : απορρίπτουμε την  $H_0$ , ενώ
- αν  $p$ -value  $\geq \alpha$  : δεν απορρίπτουμε την  $H_0$ .

Αν το  $p$ -value είναι πάρα πολύ μικρό (π.χ. 0.0001) τότε απορρίπτουμε την  $H_0$  χωρίς δεύτερη σκέψη ενώ αν το  $p$ -value είναι σχετικά μικρό (π.χ. «κοντά» στο 0.05) τότε μπορεί μεν να απορρίψουμε την  $H_0$  αλλά με κάποια επιφυλακτικότητα.

Στα στατιστικά πακέτα, μετά την εισαγωγή των τιμών  $x$  του δείγματος και την επιλογή του επιθυμητού ελέγχου, εμφανίζεται η τιμή του  $p$ -value που αντιστοιχεί στο  $x$ . Σύμφωνα με τα παραπάνω, αν η τιμή αυτή είναι μικρή (μικρότερη του  $\alpha = 0.01$  ή 0.05) τότε απορρίπτουμε την  $H_0$ .

Το πλεονέκτημα από την χρήση του  $p$ -value είναι ότι δεν απορρίπτουμε ή δεχόμαστε απλώς την  $H_0$  αλλά μπορούμε να δούμε και πόσο πιθανή ήταν η εμφάνιση του δείγματος  $x$  που πήραμε (υπό την  $H_0$ ) ενώ επίσης μπορούμε να την συγκρίνουμε

άμεσα με όποιο  $a$  και αν επιλέξουμε. Ο λόγος για τον οποίο το  $p$ -value συνήθως προϋποθέτει τη χρήση  $H/Y$  είναι διότι χωρίς τον  $H/Y$  δεν είναι πάντοτε εύκολο να υπολογιστεί ή να πινακοποιηθεί για κάθε τιμή του  $T(x)$ .

Άρα συνοπτικά καταλήγουμε στο εξής συμπέρασμα:

Αυτή η πιθανότητα ονομάζεται  $P$ -Τιμή ( $P$ -Value) του δείγματος ή κρίσιμο επίπεδο (critical level) και είναι η πιθανότητα να εμφανισθεί η τιμή της στατιστικής συνάρτησης ελέγχου που εμφανίσθηκε ή κάποια πιο μακριά (πιο ακραία), προς την κατεύθυνση της  $H_1$ , δεδομένου ότι η  $H_0$  είναι αληθής. Έτσι, υπολογίζοντας την  $P$ -value του δείγματος, γνωρίζουμε πόσο πιθανή ήταν η εμφάνιση του δείγματος που πήραμε με την υπόθεση ότι η  $H_0$  είναι αληθής. Επομένως, όσο πιο μικρή είναι η  $P$ -Value τόσο ισχυρότερες ενδείξεις εναντίον της  $H_0$  προκύπτουν από το συγκεκριμένο τυχαίο δείγμα ή αλλιώς τόσο πιο σημαντική είναι η τιμή της στατιστικής συνάρτησης ελέγχου που δίνει το δείγμα.

(Μπούτσικας Μιχαήλ (2004)

Έτσι σε πρώτο στάδιο με την χρήση του τύπου της πυκνότητας πιθανότητας ( 3.1.3) θα εκτιμήσουμε την παράμετρο  $a$  με την βοήθεια της διαδικασίας της Εκτίμησης Μεγίστης Πιθανοφάνειας.

Πιο αναλυτικά έχουμε την συνάρτηση ,

$$f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3}$$

Έτσι τώρα θα παρουσιάσουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας η οποία υπολογίζεται από τον τύπο

$$f(x_1, x_2, \dots, x_n | a) = f(x_1 | a) * f(x_2 | a) * \dots * f(x_n | a)$$

Η πιθανοφάνεια για την συνάρτηση πυκνότητας υπολογίζεται από τον τύπο

$$L(a; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i, /a) = \left(\frac{2}{\pi}\right)^{\frac{n}{2}} * \prod x_i^2 * e^{-\frac{\sum x_i^2}{2a^2}} * \frac{1}{a^{3n}}$$

Η λογαριθμική-πιθανοφάνεια υπολογίζεται με τον τύπο ,

$$\text{Log}(a, x_1, x_2, \dots, x_n) = \sum_{i=1}^n \ln f\left(\frac{x_i}{a}\right) = \frac{n}{2} \log\left(\frac{2}{\pi}\right) + \sum_{i=1}^n \log x_i^2 - \frac{\sum x_i^2}{2a^2} + \log\left(\frac{1}{a^{3n}}\right)$$

Έπειτα παραγωγίζουμε ως προς την παράμετρο μας και θέτουμε την παράγωγο ίση με το μηδέν,

$$\frac{\partial \text{Log}(a, x_i)}{\partial a} = \frac{\sum x_i^2}{a^3} - \frac{3n}{a} = 0 \Rightarrow \frac{\sum x_i^2}{a^3} = \frac{3n}{a} \Rightarrow \hat{a} = \sqrt{\sum x_i^2} / \sqrt{3n}.$$



Για να διασφαλίσουμε ότι ισχύει το παραπάνω αποτέλεσμα αντιστοιχεί σε μέγιστο , θα παραγωγίσουμε για δεύτερη φορά την πρώτη παράγωγο. Πιο συγκεκριμένα θα έχουμε

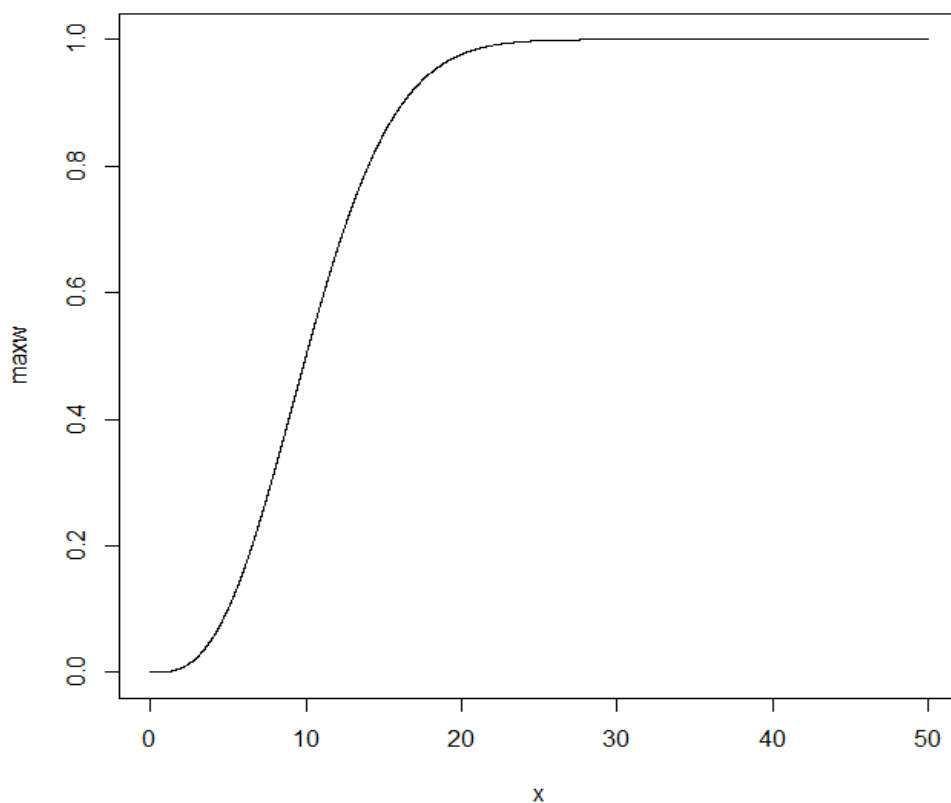
$$\frac{\partial^2 \text{Log}(a, x_i)}{\partial^2 a} = -3 \frac{\Sigma x_i^2}{a^4} + \frac{3}{a^2} < 0$$

κάτι το οποίο μας διασφαλίζει ότι το παραπάνω αποτέλεσμα του εκτιμητή δίνει πράγματι την μέγιστη πιθανοφάνεια.

Έτσι κάνοντας χρήση της μεθόδου Kolmogorov Smirnov και γνωρίζοντας τον εκτιμητή  $\hat{a} = \sqrt{\Sigma x_i^2} / \sqrt{3n}$  έχουμε την δυνατότητα να προσεγγίσουμε την γνωστή κατανομή μας. Άρα προφανώς με αντικατάσταση θα έχουμε την εξής γνωστή κατανομή η οποία υπολογίζεται με τον τύπο ,

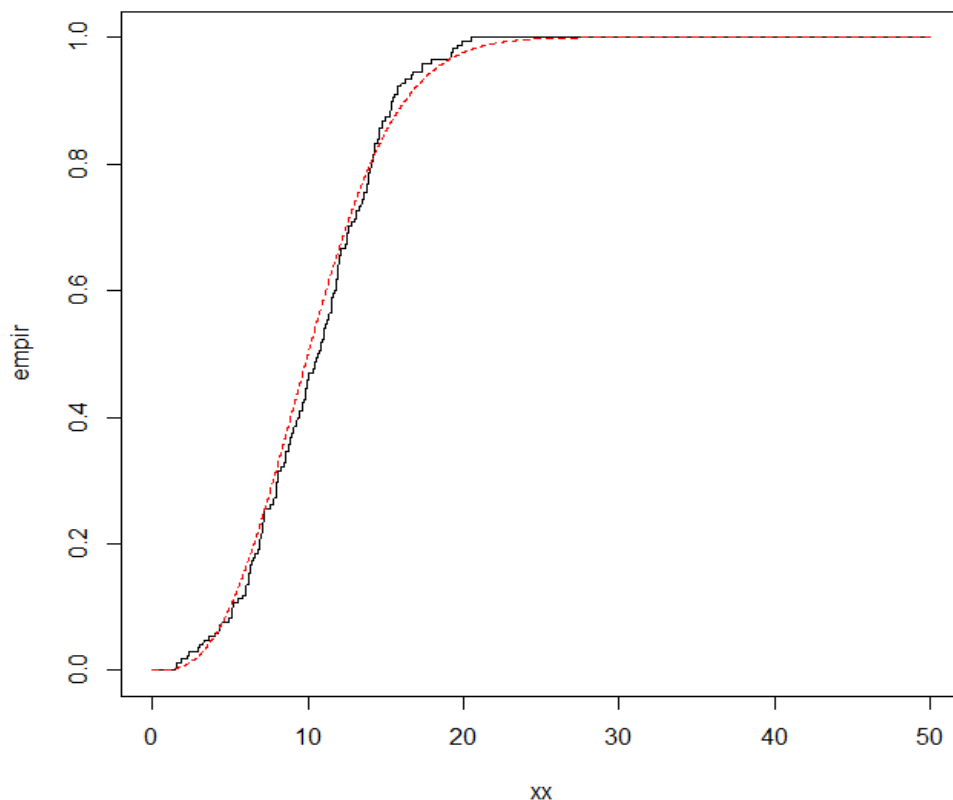
$$F_o(x) = 2\Phi\left(\frac{x}{\hat{a}}\right) - \frac{1}{\hat{a}} * \sqrt{\frac{2}{\pi}} * x e^{-\frac{x^2}{2\hat{a}^2}} - 1.$$

Η τιμή του εκτιμητή  $\hat{a} = \frac{\sqrt{21.243,42}}{\sqrt{3*168}} = \frac{145,75}{22,45} = 6,49$ . Άρα με την χρήση του προγράμματος της R πραγματοποιούμε το τεστ Kolmogorov- Smirnov. Έτσι έχουμε την δυνατότητα να δημιουργήσουμε την γραφική παράσταση της θεωρητικής κατανομής η οποία είναι η εξής:



Για την εμπειρική συνάρτηση με την χρήση του πακέτου R καταλήγουμε στις τιμές που μας δίνει το ίδιο το πρόγραμμα για την  $F_n(x)$ . Η γραφική παράσταση για την εμπειρική συνάρτηση της Maxwell-Boltzmann είναι σχεδόν ίδια με την θεωρητική όπου στα σημεία τα οποία δεν εφάπτονται οι δυο παραστάσεις υπάρχουν διαφορές από τις οποίες η μεγαλύτερη είναι και η ζητούμενη.

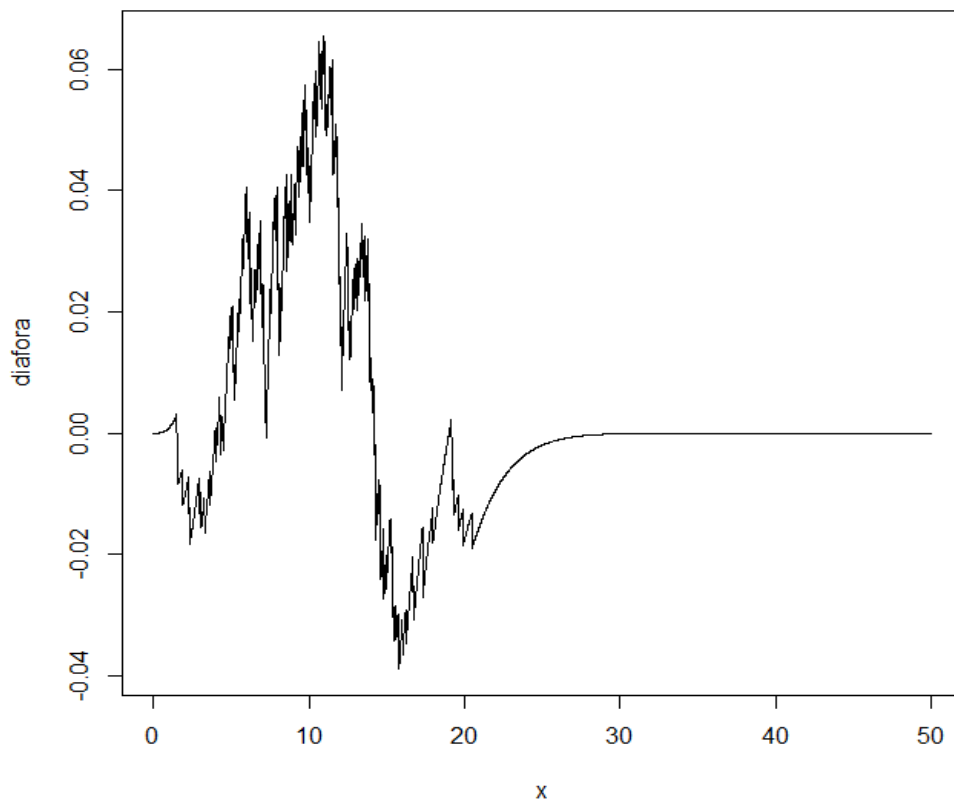
Στο παρακάτω γράφημα περιγράφεται διαγραμματικά το προαναφερθέν.



Η θεμελιώδης ιδέα του ελέγχου Kolmogorov είναι ότι η εμπειρική συνάρτηση κατανομής, ως εκτιμήτρια της αθροιστικής συνάρτησης κατανομής, δεν θα πρέπει να αποκλίνει σημαντικά από την τελευταία. Αναζητά, λοιπόν, ο έλεγχος αυτός την μέγιστη κατακόρυφη απόσταση μεταξύ των γραφημάτων της  $F_n(x)$  και της  $F_0(x)$ .

Στο Παράρτημα παρουσιάζουμε τα αποτελέσματα του στατιστικού  $D_n$  βάσει του οποίου διακρίνουμε πως το μέγιστο είναι η τιμή του πίνακα 0,0659.

Το παρακάτω γράφημα παριστάνει τις διαφορές μεταξύ εμπειρικής και θεωρητικής συνάρτησης όπου η μεγαλύτερη τιμή διαφοράς η οποία βρίσκεται και στο ψηλότερο σημείο είναι και η ζητούμενη τιμή βάση του τύπου  $D_n = \max |F_n(x) - F_0(x)|$ .



Για επίπεδο σημαντικότητας  $\alpha=0,05$  ισχύει  $1,36/\sqrt{n} = \frac{1,36}{12,96} = 0,10$

Έτσι δεν απορρίπτουμε την υπόθεση ότι τα δεδομένα μας ταιριάζουν με την κατανομή Maxwell-Boltzmann αφού  $D_{168} = 0,10 > 0,0659$

Kolmogorov-Smirnov

Dn	0,0659
P-Value	0,4733

Από την στιγμή που το Dn είναι μικρότερο από το κρίσιμο σημείο σε επίπεδο σημαντικότητας 5% ,τότε διαπιστώνουμε ότι τα δεδομένα μας προσαρμόζονται στην κατανομή μας.

Προφανώς και σε επίπεδο σημαντικότητας 1%(αφού  $1% < 5%$ ) το κρίσιμο σημείο έχει τιμή  $1,63/\sqrt{n} = \frac{1,63}{12,96} = 0,126$  , όπου και εδώ ισχύει πως τα δεδομένα μας προσαρμόζονται στην κατανομή Maxwell-Boltzmann αφού  $D_{168} = 0,126 > 0,0659$ .

Σε επίπεδο σημαντικότητας 10% το κρίσιμο σημείο έχει την τιμή  $1,22/\sqrt{n} = \frac{1,22}{12,96} = 0,941$ . όπου και εδώ ισχύει πως τα δεδομένα μας προσαρμόζονται στην κατανομή Maxwell-Boltzmann αφού  $D_{168} = 0,941 > 0,0659$ .

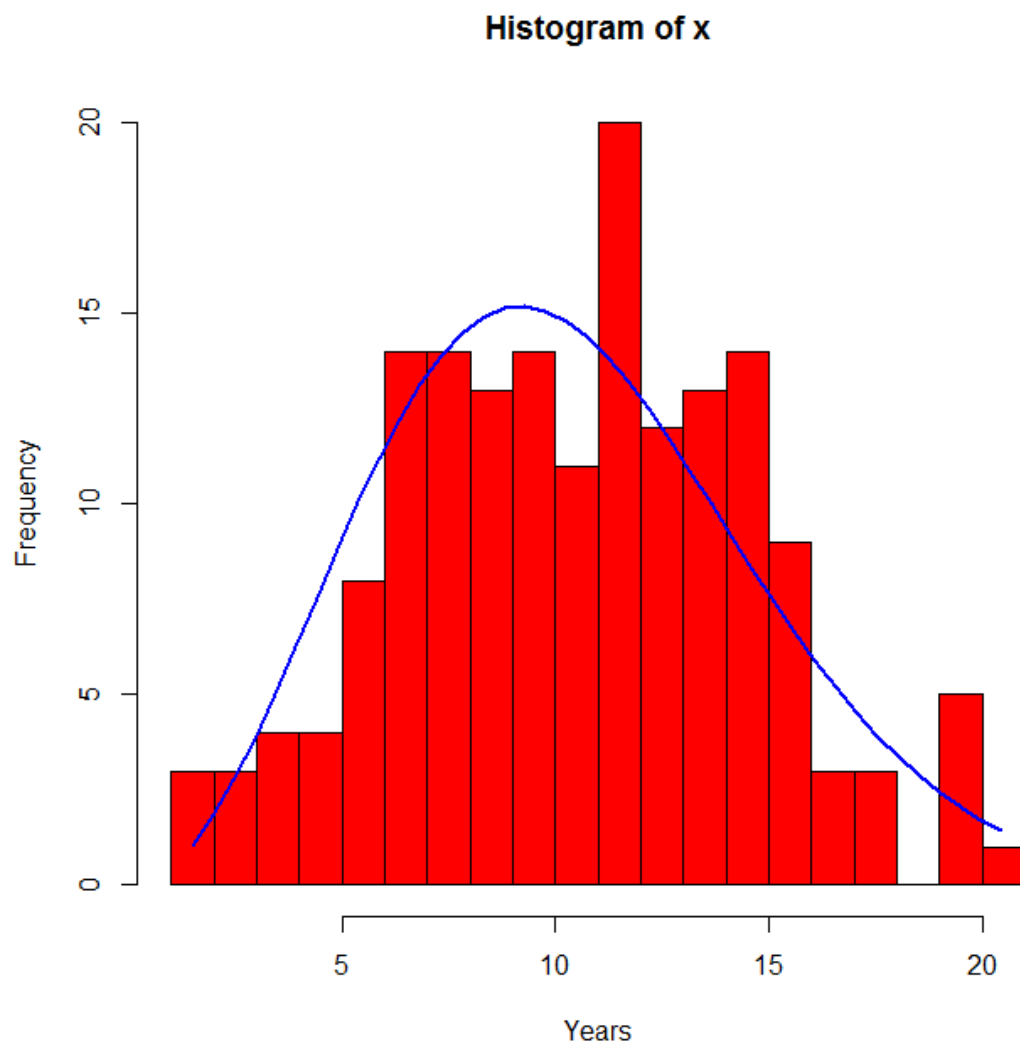
(Γ.Πιτσέλης 2015)

(Μ.Μπούτσικας 2004)

(<http://www2.stat-athens.aueb.gr/~exek/NPar-Statistics/chapter4.pdf> )

Παρακάτω παρουσιάζεται ένα γράφημα που αναπαριστά πόσο καλά προσαρμόζονται τα δεδομένα μας στην κατανομή Maxwell-Boltzmann.

Άλλες



Από το σχήμα διαπιστώνουμε ότι τα δεδομένα προσαρμόζονται πολύ καλά στην κατανομή Maxwell-Boltzmann και αυτό διαπιστώνεται από το γεγονός ότι κατανέμονται ομοιόμορφα στο ιστόγραμμα των δεδομένων.

Άλλες κατανομές που θα εξετάσουμε αν προσαρμόζονται στα δεδομένα μας είναι οι κατανομές Εκθετική , Γάμμα, Weibull, LogNormal. Στις συγκεκριμένες κατανομές εξετάζουμε το p-value ( κρίσιμο επίπεδο) που θα προκύπτει από το πακέτο της R.

## 4.2 Έλεγχος προσαρμογής των δεδομένων σε άλλες συνεχείς κατανομές

Στην παράγραφο αυτή θα εξεταστούν αν τα δεδομένα μας έχουν καλή προσαρμογή σε άλλες συνεχείς γνωστές κατανομές με την χρήση του πακέτου της R και μέσου του ελέγχου Kolmogorov –Smirnov.

*Εκθετική Κατανομή,*

Η Υπόθεση που κάνουμε σε επίπεδο σημαντικότητας  $\alpha=0,05$  είναι:

$H_0$ : Τα δεδομένα μας ακολουθούν της Εκθετική Κατανομή.

$H_1$ : Τα δεδομένα μας δεν ακολουθούν την Εκθετική Κατανομή.

Με την χρήση του πακέτου R πραγματοποιήσαμε εκτίμηση της μέγιστης πιθανοφάνειας για τα δεδομένα μας και έπειτα κάναμε χρήση του K-S τεστ όπου προέκυψαν τα παρακάτω αποτελέσματα .

	MLE	P-Value	Dn
Exponential	<b>0,0954</b>	<b>7,88E-12</b>	<b>0,31417</b>

Άρα διαπιστώνουμε ότι εφόσον  $p\text{-value} < \alpha$  , τότε απορρίπτουμε την  $H_0$  και τα δεδομένα μας φαίνεται να προσαρμόζονται ικανοποιητικά στην εν λόγω κατανομή.

*Γάμμα Κατανομή,*

Η Υπόθεση που κάνουμε σε επίπεδο σημαντικότητας  $\alpha=0,05$  είναι:

$H_0$ : Τα δεδομένα μας ακολουθούν της Γάμμα Κατανομή.

$H_1$ :: Τα δεδομένα μας δεν ακολουθούν την Γάμμα Κατανομή.

Με την χρήση και εδώ του πακέτου R εκτιμήσαμε τις παραμέτρους με την μέθοδο της μέγιστης πιθανοφάνειας και έπειτα πραγματοποιήσαμε ξανά το K-S τεστ και τα αποτελέσματα είναι τα εξής:

Gamma Distribution	
MLE	<b>a=5,23424723 , λ=0,49950779</b>
P-Value	<b>0,2407</b>
Dn	<b>0,079352</b>

Άρα διαπιστώνουμε ότι εφόσον το  $p\text{-value} > \alpha$  τότε δεχόμαστε την μηδενική υπόθεση  $H_0$  και τα δεδομένα μας προσαρμόζονται στην εν λόγω κατανομή.

*Weibull Distribution,*

Η Υπόθεση που κάνουμε σε επίπεδο σημαντικότητας  $\alpha=0,05$  είναι:

$H_0$ : Τα δεδομένα μας ακολουθούν την Weibull Κατανομή.

H1: Τα δεδομένα μας δεν ακολουθούν την Weibull Κατανομή.

Με την χρήση και εδώ του πακέτου R εκτιμήσαμε τις παραμέτρους με την μέθοδο της μεγίστης πιθανοφάνειας και έπειτα πραγματοποιήσαμε ξανά το K-S τεστ και τα αποτελέσματα είναι τα εξής:

Weibull Distribution	
MLE	a= 2,7952485 λ=11,7581978
P-Value	0,9316
Dn	0,0041733

Για επίπεδο σημαντικότητας  $\alpha=0,05$ ,

διαπιστώνουμε και εδώ πως εφόσον  $p\text{-value} > \alpha$  δεχόμαστε την  $H_0$ .

*Lognormal Distribution,*

Η Υπόθεση που κάνουμε σε επίπεδο σημαντικότητας  $\alpha=0,05$  είναι:

$H_0$ : Τα δεδομένα μας ακολουθούν της LogNormal Κατανομή.

$H_1$ : Τα δεδομένα μας δεν ακολουθούν την LogNormal Κατανομή.

Τέλος άλλη μια κατανομή που εξετάστηκε αν προσαρμόζει στα δεδομένα μας είναι η κατανομή LogNormal η οποία εκτιμήθηκε κατά τον ίδιο τρόπο. Τα αποτελέσματα είναι τα εξής:

LogNormal Distribution	
MLE	a=2,25079862 λ=0,48918958
P-Value	0,06771
Dn	0,10038

Διαπιστώνουμε και εδώ πως εφόσον  $p\text{-value} > \alpha$  δεχόμαστε την  $H_0$ .

### 4.3 Υπολογισμός άλλων μέτρων για την κατανομή Maxwell- Boltzmann και περαιτέρω ανάλυση αυτής

Ο πίνακας συχνοτήτων και το ραβδόγραμμα ή ιστόγραμμα δίνουν μια συνοπτική παρουσίαση των δεδομένων και μας επιτρέπουν να μελετήσουμε ποιοτικά την κατανομή της τυχαίας μεταβλητής που παρατηρήσαμε. Στη συνέχεια θα ορίσουμε ποσοτικά μεγέθη που περιγράφουν περιληπτικά τα βασικά χαρακτηριστικά της κατανομής της τ.μ.  $X$  και λέγονται συνοπτικά ή περιγραφικά μέτρα (summary or descriptive statistics). Κάθε τέτοιο μέτρο υπολογίζεται από τις παρατηρήσεις του δείγματος κι όπως θα δούμε στα επόμενα κεφάλαια αποτελεί εκτίμηση κάποιας παραμέτρου της κατανομής της τ.μ. που μελετάμε.

Θα ασχοληθούμε με δύο τύπους περιγραφικών μέτρων :

- τα μέτρα θέσης (measures of location) που προσδιορίζουν χαρακτηριστικές θέσεις μέσα στο εύρος των δεδομένων και
- τα μέτρα μεταβλητότητας (variability measures) που δίνουν περιληπτικά τη διασκόρπιση και μεταβλητότητα των δεδομένων

### **Μέτρα Θέσης και Τάσης**

Ως μέτρα θέσης εννοούμε κυρίως τα μέτρα κεντρικής τάσης που προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα. Τα κυριότερα μέτρα κεντρικής τάσης είναι:

- η δειγματική μέση τιμή (sample mean value) ή αριθμητικός μέσος (arithmetic mean), ή μέσος όρος (average),
- η δειγματική διάμεσος (sample median),
- η δειγματική επικρατούσα τιμή (sample mode).

Μέση τιμή: Η δειγματική μέση τιμή είναι το πιο γνωστό και χρήσιμο μμέτρο του κέντρου των δεδομένων. Έστω  $x_1, x_2, \dots, x_n$ , οι τιμές των παρατηρήσεων του δείγματος για μια τ.μ.  $X$  που μελετάμε. Η δειγματική μέση τιμή συμβολίζεται με  $\bar{x}$  και ορίζεται ως  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$ . (1.2) Η μέση τιμή είναι το κέντρο ισορροπίας των δεδομένων

Για τα δεδομένα μας η μέση τιμή του δείγματος που έχουμε ισούται με

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1760,44}{168} = 10,44$$

Διάμεσος :Η Διάμεσος ενός συνόλου είναι η τιμή που χωρίζει το σύνολο των διατεταγμένων δεδομένων σε ισοπληθή υποσύνολα ,δηλαδή η διάμεσος των δεδομένων ορίζεται ως η κεντρική τιμή όταν διατάξουμε τις παρατηρήσεις στην σειρά όπου εδώ λόγω του ότι έχουμε άρτιο αριθμό δεδομένων ισχύσει ο τύπος  $\frac{n+1}{2}$  δηλαδή το ημίαθροισμα των τιμών στις θέσεις αυτές.

Median (δείγματος) =10, 67

Επικρατούσα τιμή : Η δειγματική επικρατούσα τιμή χρησιμοποιείται επίσης για να δηλώσει την κεντρική τάση των δεδομένων και ορίζεται ως η τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα.

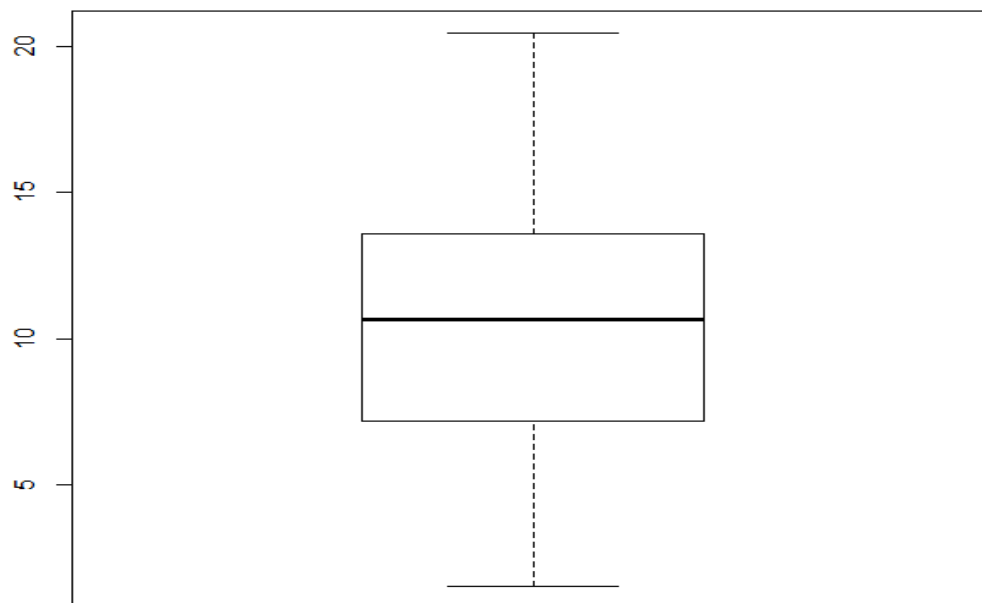
### **Μέτρα Μεταβλητότητας**

Εκτός από την κεντρική τάση μας ενδιαφέρει επίσης και η μεταβλητότητα ή διασπορά των παρατηρήσεων. Όταν τα δεδομένα είναι συγκεντρωμένα γύρω από μια κεντρική

τιμή, δηλαδή η διασπορά των δεδομένων είναι μικρή, τότε η κεντρική τιμή αντιπροσωπεύει ικανοποιητικά τα δεδομένα. Από την άλλη, όταν τα δεδομένα είναι πολύ σκορπισμένα τα μέτρα κεντρικής τιμής δε δίνουν καλή περιληπτική περιγραφή των δεδομένων. Τα κυριότερα μέτρα μεταβλητότητας είναι τα εξής :

- Εύρος δηλαδή η διαφορά μέγιστης με ελάχιστης τιμής ,όπου στο δείγμα μας είναι  $\max = 20,46$  και  $\min = 1,48$ . Άρα  $R = 18,98$ .
- Διασπορά ή διακύμανση μετράει την μεταβλητότητα των παρατηρήσεων γύρω από την μέση τιμή .Στο σύνολο των παρατηρήσε μας η διασπορά είναι  $\text{var} = 16,74$ .

Παρακάτω παρουσιάζεται ένα θηκόγραμμα για τις παρατηρήσεις μας όπου μπορούμε να παρατηρήσουμε την παραπάνω ανάλυση.



Άρα συνοπτικά έχουμε τον παρακάτω πίνακα :

mean	median	variance	minimum	maximun
10,48	10,66774109	16,7430464	1,48	20,4625594

Τα μεγέθη αυτά έχουν υπολογιστεί και με την βοήθεια του excel αλλά και του προγράμματος της R.

Όσο αφορά τα μέτρα της μέσης τιμής και της διακύμανσης έχει βρεθεί νωρίτερα η παράμετρος  $\hat{a} = 6,49$ . Έτσι αν αντικαταστήσουμε στους ανάλογους τύπους της



κατανομής για την μέση τιμή και την διακύμανση (3.1.5),(3.2.5) αντίστοιχα καταλήγουμε στο εξής :

$$E(X) = 2a \sqrt{\frac{2}{\pi}} = 12,98 * 0,80 = 10,36$$

$$Var(X) = a^2 \left(3 - \frac{8}{\pi}\right) = 42,12(3 - 2,55) = 18,95$$

Παρατηρούμε πως και οι τιμές είναι πολύ κοντά με τις τιμές που υπολογίσαμε βάσει του προγράμματος και ουσιαστικά επαληθεύουν την προσαρμογή των δεδομένων στην κατανομή μας. Άρα ουσιαστικά έχουμε σχεδόν ίδιες θεωρητικές και δειγματικές τιμές που είναι και το ζητούμενο.

(<http://users.auth.gr/dkugiu/Teach/CivilEngineer/descriptive.pdf>)

### **Ασυμμετρία και Κυρτότητα**

#### *Συντελεστής Ασυμμετρίας*

Οι παραστάσεις των δεδομένων μπορούν να έχουν μπορούν να έχουν οποιασδήποτε μορφή κι οποιοδήποτε σχήμα στην πράξη όμως επιχειρείται η ένταξη τους σε «τυπικές» μορφές οικογένειας ώστε να είναι δυνατή η μελέτη τους. Η κατανομή ενός πληθυσμού μπορεί να είναι συμμετρική ή μη συμμετρική. Στην πρώτη περίπτωση η κορυφή, η διάμεσος και η μέση τιμή συμπίπτουν. Στις άλλες περιπτώσεις ένα από τα τμήματα που στα οποία χωρίζει η κατανομή η κορυφή περιέχει περισσότερες παρατηρήσεις από το άλλο. Υπάρχουν δυο ειδών ασυμμετρίας, η θετική ασυμμετρία στην οποία οι περισσότερες παρατηρήσεις καθώς η διάμεσος και μέση τιμή βρίσκονται δεξιά της κορυφής ,αλλά και η αρνητική ασυμμετρία στην οποία οι περισσότερες παρατηρήσεις ,όπως η διάμεσος και η μέση τιμή, βρίσκονται αριστερά της κορυφής .

Ο τύπος του συντελεστή ασυμμετρίας ορίζεται ως εξής:

$$\gamma = \frac{\frac{\sum (xi - \mu)^3}{n}}{\sqrt{\frac{\sum (xi - \mu)^2}{n}}}$$

Όπου για  $\gamma > 0$  έχουμε θετική ασυμμετρία, όταν  $\gamma < 0$  έχουμε αρνητική ασυμμετρία και για  $\gamma = 0$  έχουμε συμμετρία. Έτσι αν κάνουμε αντικατάσταση στον τύπο βάσει των στοιχείων που έχουμε από τα δεδομένα μας για την κατανομή έχουμε το εξής :

$$\gamma = \frac{\frac{\sum (xi - \mu)^3}{n}}{\sqrt{\frac{\sum (xi - \mu)^2}{n}}} = \frac{\frac{523,98}{168}}{\sqrt{\frac{27796,09}{168}}} = \frac{3,12}{4,08} = 0,77$$

Άρα με αντικατάσταση προκύπτει ότι έχουμε θετική ασυμμετρία.

Με την βοήθεια του προγράμματος της R και του Excel υπολογίζουμε ότι η ασυμμετρία της κατανομής έχει την τιμή 0,046 .

Συντελεστές Κύρτωσης

Ένα πολύ σημαντικό μέτρο περιγραφής είναι το μέτρο κύρτωσης που χαρακτηρίζει το ύψος της κορυφής και δίνει πληροφορίες γύρω από την αιχμηρότητα της καμπύλης κατανομής. Οι κατανομές με βάση το συντελεστή κύρτωσης διακρίνονται σε :

- Λεπτόκυρτες
- Μεσόκυρτες
- Πλατύκυρτες

Το πρότυπο σύγκρισης της παραπάνω αναφοράς αποτελεί η κανονική κατανομή η οποία είναι μεσόκυρτη συνεπώς αν μια κατανομή έχει μεγαλύτερη κυρτότητα της κανονικής κατανομής χαρακτηρίζεται ως πλατύκυρτη, ενώ αν ισχύει το αντίστροφο λεπτόκυρτη. Για ένα σύνολο δεδομένων όπως αυτά τα οποία εξετάσαμε ότι ακολουθούν την Maxwell-Boltzmann μπορούμε θεωρητικά να εξετάσουμε την κύρτωση χρησιμοποιώντας τον τύπο

$$\alpha = \frac{\frac{\sum(x_i - \mu)^4}{n}}{\left(\sqrt{\frac{\sum(x_i - \mu)^2}{n}}\right)^4}$$

Όπου αν κάνουμε αντικατάσταση στον τύπο προκύπτει ,

$$\alpha = \frac{\frac{\sum(x_i - \mu)^4}{n}}{\left(\sqrt{\frac{\sum(x_i - \mu)^2}{n}}\right)^4} = \frac{719,42}{277} = 2,60$$

Έτσι επειδή για τις κανονικές τιμές έχουμε  $\alpha=3$  συνηθίζεται να μετράμε την κυρτότητα με την διαφορά  $\alpha-3$  , η οποία για λεπτόκυρτες κατανομές παίρνει θετικές τιμές, ενώ για πλατύκυρτες κατανομές παίρνει αρνητικές τιμές.

Άρα για  $\alpha=2,60$  έχουμε  $\alpha-3=2,6-3=-0,40$ .

Συνεπώς θεωρητικά η κατανομή Maxwell-Boltzmann είναι πλατύκυρτη κατανομή. Με την βοήθεια τώρα του Excel για το δείγμα μας καταλήγουμε πως δειγματοληπτικά η κύρτωση παίρνει την τιμή -0,37. Συμπερασματικά καταλήγουμε πως οι δυο τιμές αυτές διαφέρουν ελάχιστα όπου είναι και το ζητούμενο.

(Δ.Φουσκάκης 2005)

(Τσίμπος - Γεωργιακώδης ,1999)

#### 4.4 Υπολογισμός της παραμέτρου με την μέθοδο των ροπών

Όπως είδαμε από την προηγούμενη παράγραφο η μέση τιμή και η διακύμανση είναι πολύ κοντά είτε υπολογιστούν με την δειγματική είτε με την θεωρητική τιμή τους. Έτσι έχουμε την δυνατότητα να χρησιμοποιήσουμε την μέθοδο των ροπών για να εκτιμήσουμε την παράμετρο  $\alpha$  και να την επαληθεύσουμε.

Το σκεπτικό αυτής της μεθόδου είναι ότι εξισώνονται οι θεωρητικές ροπές με τις δειγματικές. Έτσι συνδέονται οι εκτιμώμενες παράμετροι με στατιστικές συναρτήσεις και από τη λύση των εξισώσεων που προκύπτουν, υπολογίζονται οι εκτιμητές.

Πιο συγκεκριμένα για την κατανομή μας έχουμε,

$$E(X) = 2a\sqrt{\frac{2}{\pi}} = 10,36 \text{ και } \bar{x} = \frac{\sum xi}{n} = 10,47$$

Άρα αν λύσουμε το σύστημα που προκύπτει από τις δυο παραπάνω εξισώσεις μια και έχουμε μια μόνο παράμετρο και δεν χρειάζονται και οι εξισώσεις τις διακύμανσης έχουμε το εξής:

$$10,47 = 2a\sqrt{\frac{2}{\pi}} \Rightarrow 5,235 = a * 0,80 \Rightarrow a = 6,54$$

Έτσι διαπιστώνουμε πως η παράμετρο και στις 2 περιπτώσεις είναι πολύ κοντά και επαληθεύεται.

Ας δοκιμάσουμε τώρα με τις εξισώσεις της διακύμανσης για ακόμα μεγαλύτερη ακρίβεια. Άρα έχουμε

$$Var(X) = \alpha^2 \left( 3 - \frac{8}{\pi} \right) = 18,95 \text{ και } var = 16,75$$

$$\alpha^2 \left( 3 - \frac{8}{\pi} \right) = 16,75 \Rightarrow \alpha = \sqrt{16,75/0,45} = 6,11.$$

Ισχύουν τα ίδια συμπεράσματα, καθώς η εκτιμώμενη τιμή είναι αρκετά κοντά στις προηγούμενες.

Συγκεντρωτικά αποτελέσματα μέτρων θέσης και περιγραφής τόσο θεωρητικά όσο και δειγματοληπτικά

	Μέση Τιμή	Διάμεσος	Διακύμανση	Ασυμμετρία	Κύρτωση
<b>Δειγματολογικά</b>	10,48	10,67	16,74	0,05	0,38
<b>Θεωρητικά</b>	10,36	10,67	18,95	0,77	0,40

#### 4.5 Υπολογισμός Μέσου Υπολειπόμενου χρόνου ζωής

Όπως έχουμε ήδη αναφέρει στο κεφάλαιο 2.2 ο «υπολειπόμενος χρόνος ζωής» (Residual Life) είναι μία σημαντική συνάρτηση στην θεωρία αξιοπιστίας, στη στατιστική, στην ανάλυση επιβίωσης αλλά και αλλού. Πολλά χρήσιμα αποτελέσματα έχουν προκύψει από την μελέτη αυτής της συνάρτησης. Χρησιμοποιώντας τους τύπους (2.2.1) και (2.2.2) έχουμε ,

$$\mu(x) = E(X - x | X > x) = \frac{\int_x^\infty S(t)dt}{S(x)}, x \geq 0$$

Προφανώς  $\mu(0) = \mu = E(X)$ . Αν επιπλέον η F έχει πυκνότητα έστω f , τότε μπορούμε να γράψουμε εναλλακτικά

$$\mu(x) = \frac{\int_x^\infty tf(t)dt}{S(x)} - x$$

Έχουμε την δυνατότητα να υπολογίσουμε τον υπολειπόμενο χρόνο ζωής για την κατανομή μας Maxwell-Boltzmann με αντικατάσταση στον παραπάνω τύπο.

Συγκεκριμένα με την βοήθεια του προγράμματος του Mathematica μπορώ να υπολογίσω τον τύπο όπου τα βήματα είναι τα εξής:

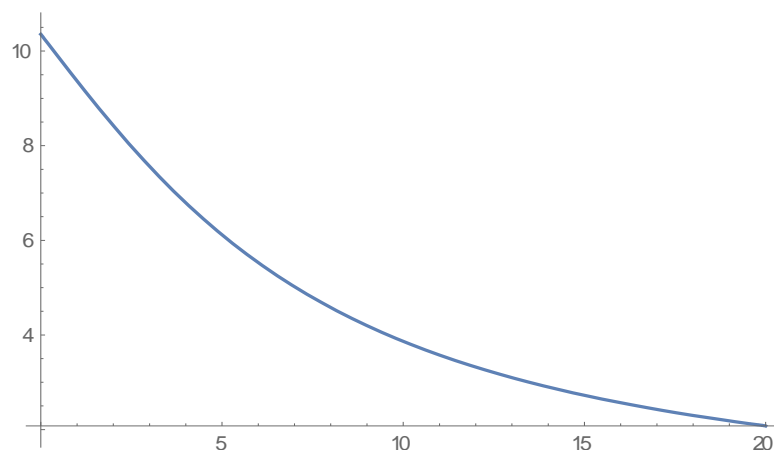
Αρχικά υπολογίζουμε τον αριθμητή του κλάσματος βάσει του τύπου που έχουμε

παραπάνω ,δηλαδή το ολοκλήρωμα  $\int_x^\infty S(t)dt = \frac{e^{-\frac{t^2}{2a^2}} \sqrt{\frac{2}{\pi}} x}{a} + \Phi\left[\frac{x}{\sqrt{2}a}\right]$

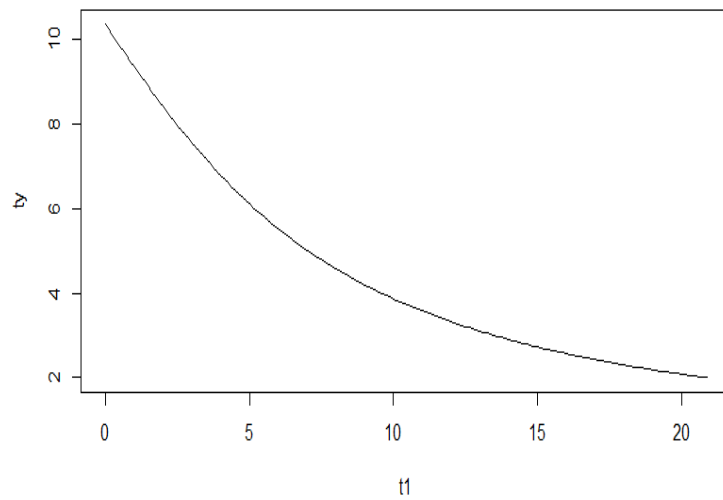
Έπειτα υπολογίζουμε το κλάσμα βάσει του τύπου  $\frac{\int_x^\infty S(t)dt}{S(x)}$  όπου έχουμε

$$\mu(x) = \frac{\int_x^\infty S(t)dt}{S(x)} = \frac{2*a*e^{-\frac{x^2}{2a^2}}*\sqrt{\frac{2}{\pi}}*x*\Phi\left(\frac{x}{\sqrt{2}*a}\right)}{-2\Phi\left(\frac{x}{a}\right)+\frac{1}{a}*\sqrt{\frac{2}{\pi}}*x*e^{-\frac{x^2}{2a^2}}+2} = \frac{2ae^{-\frac{x^2}{2a^2}}\sqrt{\frac{2}{\pi}}*x*\Phi\left[\frac{x}{\sqrt{2}a}\right]}{\frac{e^{-\frac{t^2}{2a^2}}\sqrt{\frac{2}{\pi}}x}{a}+\Phi\left[\frac{x}{\sqrt{2}a}\right]}$$

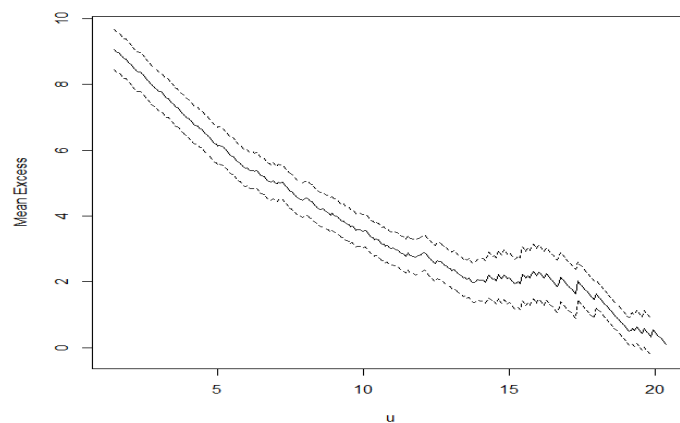
Έτσι τώρα έχουμε την δυνατότητα να δημιουργήσουμε την γραφική παράσταση του υπολειπόμενου χρόνου ζωής θεωρητικά με την βοήθεια Mathematica η οποία είναι η εξής:



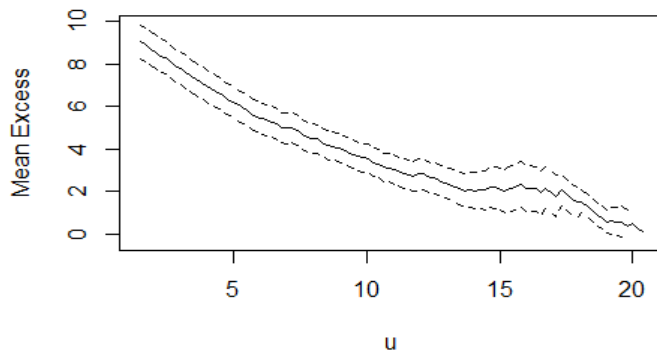
Με την βοήθεια της R βλέπουμε για επαλήθευση ότι η παραπάνω γραφική παράσταση είναι :



Εμπειρικά τώρα η κατασκευή της γραφικής παράστασης παρουσιάζεται στο παρακάτω σχεδιάγραμμα με την βοήθεια της R και παρουσιάζονται τα διαστήματα εμπιστοσύνης.



**Γραφική παράσταση υπολειπόμενου χρόνου ζωής με επίπεδο σημαντικότητας 0,95**



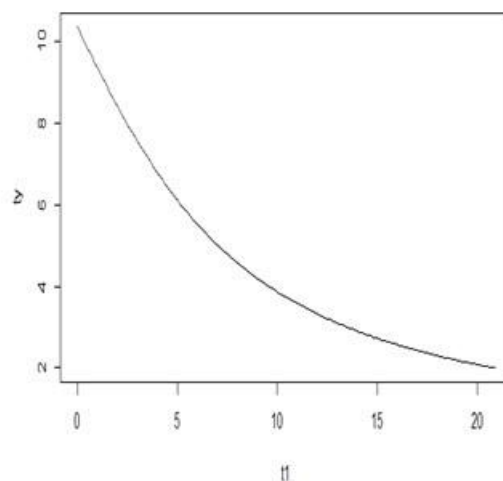
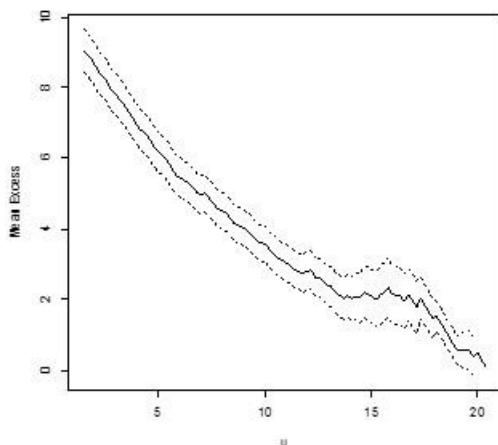
### Γραφική παράσταση υπολειπόμενου χρόνου ζωής με επίπεδο σημαντικότητας 0,99

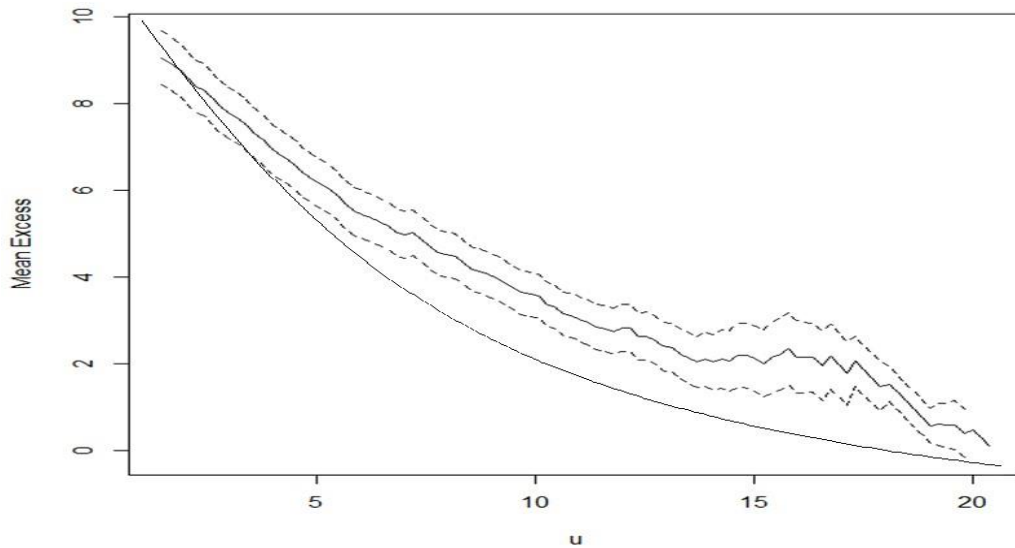
Το εν λόγω γραφήματα για να κατασκευαστούν πραγματοποιήθηκε εκτενή αναζήτηση στο διαδίκτυο και η πηγή πληροφορίας προήλθε από την ηλεκτρονική διεύθυνση <https://www.rdocumentation.org/packages/ismev/versions/1.41/topics/mrl.plot>

Είναι μια εντολή η οποία για να «τρέξει» θα πρέπει να φορτωθούν στην R τα πακέτα `ismev`, `survival` και `help=survival`. Επιπροσθέτως πρέπει να υπάρχει ένα σύνολο δεδομένων και να ορισθεί η ελάχιστη και η μέγιστη τιμή του δείγματος αλλά και το διάστημα εμπιστοσύνης. Ο τύπος που χρησιμοποιήθηκε είναι ο εξής:

```
mrl.plot(data, umin = min(data), umax = max(data) - 0.1,
conf = 0.95, nint = 100)
```

Τώρα θα πραγματοποιηθεί σύγκριση μεταξύ της θεωρητικής και εμπειρικής κατανομής και γραφική παράσταση των και των δυο. Πιο συγκεκριμένα συγκρίνοντας τα δυο αυτά σχήματα για τον υπολειπόμενο χρόνο τόσο εμπειρικά όσο και θεωρητικά έχουμε :





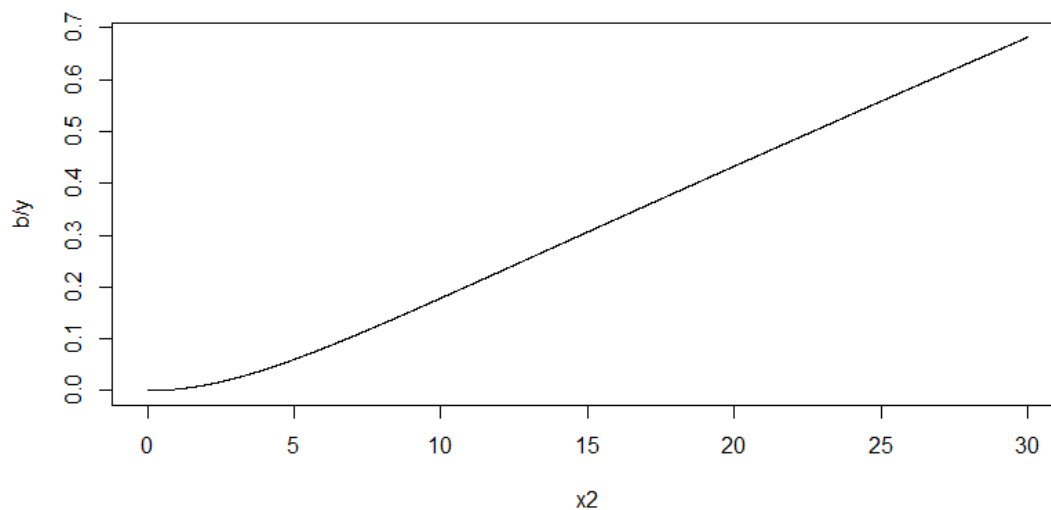
Παρατηρούμε πως η θεωρητική παράσταση φαίνεται να είναι εντός του διαστήματος εμπιστοσύνης για τις τιμές του  $u$  στο διάστημα  $[0,5]$  ωστόσο η θεωρητική βρίσκεται εκτός των  $[5-10]$ ,  $[10-15]$  και  $[15-20]$  διαστημάτων εμπιστοσύνης της εμπειρικής. Για το παράδειγμα το μηδέν είναι η ηλικία 60 ετών και εφόσον η μέγιστη τιμή τιμών είναι η 20,47 δεν υπάρχει νόημα να εξεταστεί από το 20 και έπειτα δηλαδή μετά την ηλικία των 80,5 ετών.

#### 4.6 Υπολογισμός βαθμίδας αποτυχίας για τα δεδομένα

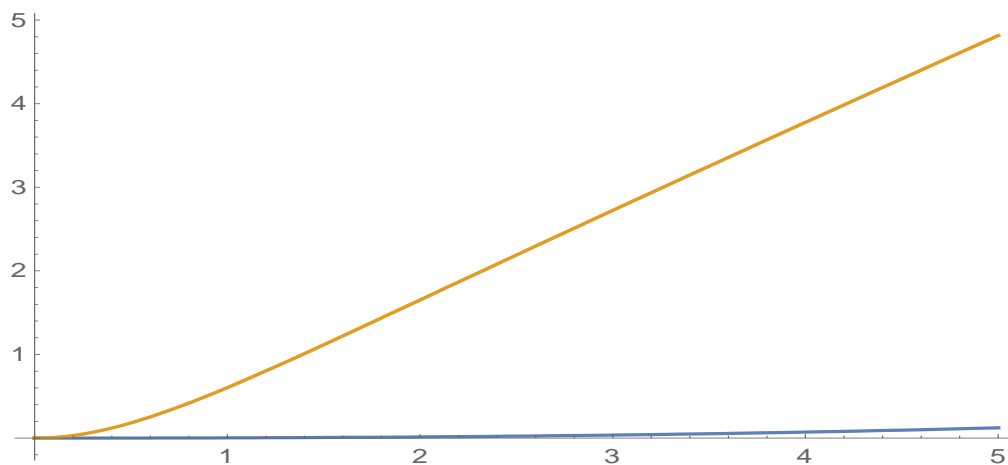
Γνωρίζουμε από το κεφάλαιο 2 και με την βοήθεια της παραγράφου 3.1 πως η βαθμίδα αποτυχίας για την κατανομή Maxwell-Boltzmann ισούται με τον τύπο

$$\lambda(x) = \frac{\left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^2 e^{-\frac{x^2}{2a^2}} \frac{1}{a^3}}{2\Phi\left(\frac{x}{a}\right) + \frac{1}{a} \sqrt{\frac{2}{\pi}} x e^{-\frac{x^2}{2a^2}} + 2}$$

Ως γνωστόν η γραφική παράσταση της βαθμίδας αποτυχίας είναι μια αύξουσα συνάρτηση και έχει παρουσιαστεί στην γενική της μορφή για την κατανομή στο διάγραμμα 3.1.8. Για τα δεδομένα μας τώρα χρησιμοποιώντας στην βοήθεια του πακέτου R η γραφική παράσταση της βαθμίδας αποτυχίας είναι η εξής:



Παρατηρούμε δηλαδή ότι είναι μια αύξουσα συνάρτηση η οποία αυξάνεται με σχετικά αργό ρυθμό. Επίσης παρακάτω παρουσιάζεται το σχήμα για την βαθμίδα αποτυχίας με την βοήθεια του προγράμματος Mathematica για  $\alpha=1$  και  $\alpha=6,49$ , όπου είναι και η παράμετρος των δεδομένων μας για την κατανομή Maxwell-Boltzmann



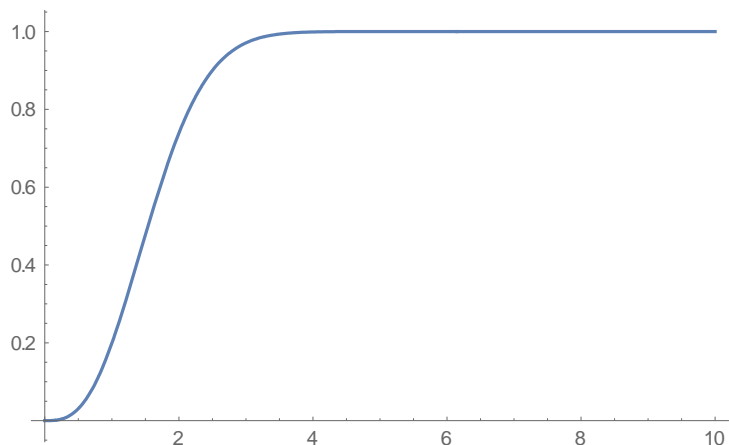
Διαπιστώνουμε πως όσο μεγαλύτερη παράμετρος έχει η κατανομή που αναλύσαμε τόσο πιο αργά κινείται και αυξάνεται, δηλαδή τόσο πιο αργά πεθαίνει κάποιος μιας και τα δεδομένα μας πρόκειται για υπολειπόμενους χρόνους ζωής μετρά τα εξήντα έτη ηλικίας.



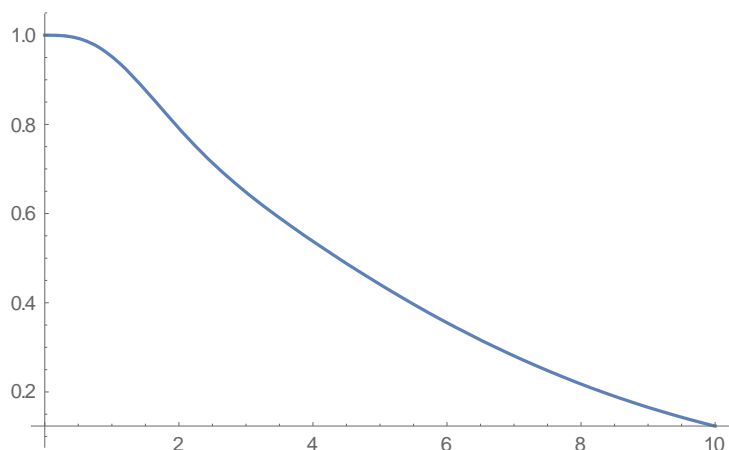
#### 4.7 Παρουσίαση Γραφικών παραστάσεων για $\alpha=6,49$ τιμή παραμέτρου

Στη υπό-ενότητα αυτή εφόσον έχουν αναλυθεί διεξοδικά οι γραφικές παραστάσεις την κατανομής Maxwell-Boltzmann γίνεται μια απλή αναφορά ώστε να διαπιστώσουμε πως απεικονίζονται οι γραφικές παραστάσεις για την τιμή της παραμέτρου της κατανομής που έχουμε.

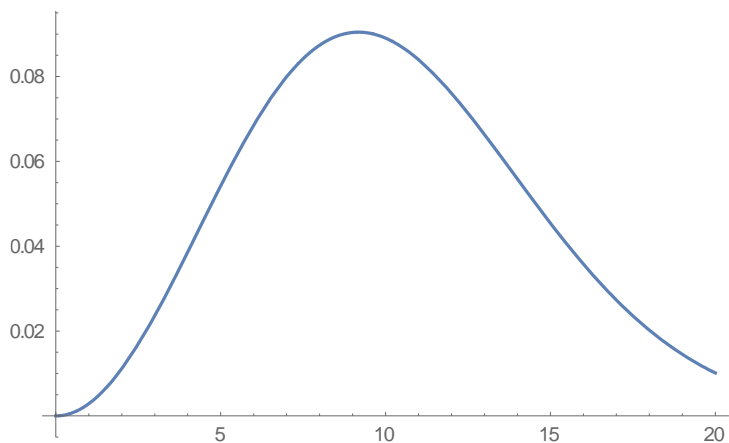
Πιο συγκεκριμένα για συνάρτηση κατανομής είναι η εξής:



Η συνάρτηση επιβίωσης :



Και η συνάρτηση πυκνότητας πιθανότητας παρίσταται ως εξής :



## 5. Παράρτημα :

Στην εν λόγω ενότητα παρουσιάζουμε το παράρτημα με κάποιους από τους κώδικες που βοήθησα στην εξαγωγή αποτελεσμάτων.

Ο Πίνακας αποτελεσμάτων από τον έλεγχο του Κολμογορον-Smirnov είναι :

x	maxw	e	diapora	abs(diapora)	max(abs(maxw-e))
1,48	0	0	0,00041725	0,00041725	0,063713488
1,55	9,73E-04	0	0,00060657	0,00060657	
1,85	7,78E-05	0	0,006519118	0,006519118	
2,25	2,63E-06	0	0,015442626	0,015442626	
2,31	6,23E-06	0	0,001108985	0,001108985	
2,91	1,22E-07	0	0,001914952	0,001914952	
3,06	2,10E-07	0	0,003043002	0,003043002	
3,26	3,34E-06	0	0,004541172	0,004541172	
3,60	4,98E-07	0	0,006447005	0,006447005	
3,98	7,09E-07	0	0,008872612	0,008872612	
4,24	9,73E-07	0	0,000432747	0,000432747	
4,29	1,29E-08	0	0,004166939	0,004166939	
4,44	1,68E-06	0	0,0052978	0,0052978	
4,91	2,14E-08	0	0,006616697	0,006616697	
5,05	2,67E-08	0	0,008138081	0,008138081	
5,06	3,28E-08	0	0,009876397	0,009876397	
5,13	3,98E-08	0	0,016035111	0,016035111	
5,19	4,78E-08	0	0,01902073	0,01902073	
5,47	5,67E-08	0	0,002237536	0,002237536	
5,74	6,67E-08	0	0,002698989	0,002698989	
5,95	7,78E-08	0	0,008216057	0,008216057	
5,99	9,01E-07	0	0,000128346	0,000128346	
6,02	1,04E+09	0	0,000145624	0,000145624	
6,14	1,18E+09	0	0,000163369	0,000163369	
6,16	1,34E+09	0	0,01023856	0,01023856	
6,18	1,52E+09	0	0,011516573	0,011516573	
6,26	1,71E+09	0	0,012896687	0,012896687	
6,30	1,91E+09	0	0,014382815	0,014382815	
6,34	2,13E+09	0	0,015978839	0,015978839	
6,50	2,37E+09	0	0,017688775	0,017688775	
6,58	2,63E+09	0	0,019516416	0,019516416	
6,73	2,90E+09	0	0,021465731	0,021465731	
6,84	3,19E+09	0	0,023540588	0,023540588	
6,87	3,49E+09	0	0,025738958	0,025738958	
6,91	3,82E+09	0	0,028997239	0,028997239	
6,91	4,17E+09	0	0,030557533	0,030557533	
7,02	4,54E+09	0	0,033173709	0,033173709	
7,09	4,92E+09	0	0,0359349	0,0359349	
7,10	5,33E+09	0	0,038845053	0,038845053	
7,10	5,77E+09	0	0,041908049	0,041908049	
7,12	6,22E+09	0	0,045127783	0,045127783	
7,17	6,70E+09	0	0,048508138	0,048508138	
7,21	7,20E+08	0	0,052053006	0,052053006	
7,53	7,73E+09	0	0,055790201	0,055790201	
7,76	8,28E+09	0	0,059630831	0,059630831	
7,80	8,85E+09	0	0,063713488	0,063713488	
7,92	9,46E+09	0	0,00010854	0,00010854	
7,92	0,00010855	0	0,000107423	0,000107423	
7,95	0,00010742	0	0,000114427	0,000114427	
7,99	0,00011427	0	0,000121401	0,000121401	
8,00	0,0001214	0	0,000128822	0,000128822	
8,03	0,00012882	0	0,00013654	0,00013654	
8,06	0,00013654	0	0,000144558	0,000144558	
8,21	0,00014456	0	0,000152885	0,000152885	
8,46	0,00015288	0	0,000161524	0,000161524	
8,54	0,00016152	0	0,000170482	0,000170482	
8,56	0,00017048	0	0,000179765	0,000179765	
8,56	0,00017976	0	0,000189377	0,000189377	
8,70	0,00018938	0	0,000199326	0,000199326	
8,75	0,00019933	0	0,000209616	0,000209616	
8,87	0,00020962	0	0,000220254	0,000220254	
8,87	0,00022025	0	0,000231244	0,000231244	
8,93	0,00023124	0	0,000242593	0,000242593	
9,05	0,00024259	0	0,000254307	0,000254307	
9,09	0,00025431	0	0,00026659	0,00026659	
9,25	0,00026659	0	0,000278849	0,000278849	
9,28	0,00027885	0	0,00029169	0,00029169	
9,41	0,00029169	0	0,000304917	0,000304917	
9,45	0,00030492	0	0,000318537	0,000318537	
9,61	0,00031854	0	0,000332555	0,000332555	
9,61	0,00033255	0	0,000346977	0,000346977	
9,77	0,00034698	0	0,000361809	0,000361809	
9,78	0,00036181	0	0,000377056	0,000377056	
9,84	0,00037706	0	0,000392723	0,000392723	
9,86	0,00039272	0	0,000408817	0,000408817	
9,92	0,00040882	0	0,000425343	0,000425343	
9,96	0,00042534	0	0,000442307	0,000442307	
10,01	0,00044231	0	0,000459714	0,000459714	
10,03	0,00045971	0	0,000477569	0,000477569	
10,29	0,00047757	0	0,000495879	0,000495879	
10,39	0,00049588	0	0,000514649	0,000514649	
10,40	0,00051465	0	0,000533885	0,000533885	
10,49	0,00053388	0	0,000553591	0,000553591	
10,66	0,00055359	0	0,000573774	0,000573774	
10,68	0,00057377	0	0,00059444	0,00059444	
10,78	0,00059444	0	0,000615593	0,000615593	
10,80	0,00061559	0	0,000637239	0,000637239	
10,94	0,00063724	0	0,000659384	0,000659384	
11,01	0,00065938	0	0,000682034	0,000682034	
11,01	0,00068203	0	0,000705193	0,000705193	
11,03	0,00070519	0	0,000728868	0,000728868	
11,09	0,00072887	0	0,000753063	0,000753063	
11,18	0,00075306	0	0,000777885	0,000777885	
11,32	0,00077789	0	0,000803039	0,000803039	
11,35	0,00080304	0	0,00082883	0,00082883	
11,48	0,00082883	0	0,000855164	0,000855164	
11,50	0,00085516	0	0,000882046	0,000882046	
11,50	0,00088205	0	0,000909482	0,000909482	
11,52	0,00090948	0	0,000937477	0,000937477	
11,60	0,00093748	0	0,000966037	0,000966037	
11,71	0,00096604	0	0,000995166	0,000995166	
11,76	0,00099517	0	0,010104871	0,010104871	
11,79	0,01010487	0	0,010515157	0,010515157	
11,83	0,01051516	0	0,010862029	0,010862029	
11,86	0,01086203	0	0,01117493	0,01117493	
11,87	0,01117493	0	0,01149554	0,01149554	
11,94	0,01149554	0	0,01182217	0,01182217	
11,94	0,01182217	0	0,01215489	0,01215489	
12,00	0,01215489	0	0,01249373	0,01249373	
12,00	0,01249373	0	0,01283876	0,01283876	
12,06	0,01283876	0	0,01319003	0,01319003	
12,06	0,01319003	0	0,01354759	0,01354759	
12,43	0,00135476	0	0,001391149	0,001391149	
12,47	0,00139115	0	0,00142818	0,00142818	
12,50	0,00142818	0	0,001465855	0,001465855	
12,52	0,00146586	0	0,001504181	0,001504181	
12,56	0,00150418	0	0,001543163	0,001543163	
12,63	0,00154316	0	0,001582806	0,001582806	
12,83	0,00158281	0	0,001623116	0,001623116	
12,94	0,00162312	0	0,001664097	0,001664097	
13,05	0,0016641	0	0,001705754	0,001705754	
13,10	0,00170575	0	0,001748094	0,001748094	
13,27	0,00174809	0	0,001791122	0,001791122	
13,43	0,00179112	0	0,001834841	0,001834841	
13,46	0,00183484	0	0,001879259	0,001879259	
13,58	0,00187926	0	0,00192438	0,00192438	
13,60	0,00192438	0	0,001970209	0,001970209	
13,78	0,00197021	0	0,002016751	0,002016751	
13,80	0,00201675	0	0,002064012	0,002064012	
13,85	0,00206401	0	0,002111996	0,002111996	
13,87	0,002112	0	0,00216071	0,00216071	
13,87	0,00216071	0	0,002210157	0,002210157	
13,96	0,00221016	0	0,002260344	0,002260344	
14,01	0,00226034	0	0,002311275	0,002311275	
14,11	0,00231128	0	0,002362955	0,002362955	
14,19	0,00236296	0	0,00241539	0,00241539	
14,20	0,00241539	0	0,002468585	0,002468585	
14,24	0,00246858	0	0,002522544	0,002522544	
14,26	0,00252254	0	0,002577273	0,002577273	
14,27	0,00257727	0	0,002632777	0,002632777	
14,46	0,00263278	0	0,002689061	0,002689061	
14,57	0,00268906	0	0,00274613	0,00274613	
14,60	0,00274613	0	0,002803988	0,002803988	
14,61	0,00280399	0	0,002862642	0,002862642	
14,78	0,00286264	0	0,002922096	0,002922096	
14,78	0,00292209	0	0,002982354	0,002982354	
14,94	0,00298235	0	0,003043423	0,003043423	
15,21	0,00304342	0	0,003105306	0,003105306	
15,31	0,00310531	0	-0,002784371	-0,002784371	
15,38	0,00316801	0,01	-0,002720843	-0,002720843	
15,39	0,00321354	0,01	-0,002656485	-0,002656485	
15,44	0,00325959	0,01	-0,002591292	-0,002591292	
15,59	0,00336109	0,01	-0,002525259	-0,002525259	
15,71	0,00342712	0,01	-0,002458382	-0,0024583	

```

w<-sort(w)
w
boxplot(w)
emp<-ecdf(w)
emp(w)
mean(w)
skewness(w,na.rm=FALSE)
library(moments)
skewness(w)
kurtosis(w)
#MLE#
a=6.49
#Maxwell Distribution#
x<-seq(0,by=0.01,length=5000)
maxw=2*pnorm(x/a) - (1/a)*sqrt(2/pi)*x*exp(-(x^2)/(2*(a^2)))-1
maxw
plot(x,maxw)
n=168
e=rep(0,5000)
for(i in 1:5000)
{e[i]=(length(w[w<x[i]])/n)}
xx=t(rbind(x,x))
empir=t(rbind(e,maxw))
matplot(xx,empir,type="l")
diafora=maxw-e
diafora
plot(x,diafora,type="l")
max(abs(emp(w)-maxw))

```

```

#which fuctions fit in our data#

library(MASS)

fitdistr(w,"weibull")

ks.test(w,pweibull,shape=2.7952485,scale=11.7581978,TRUE)

?Exponential(stats)

fitdistr(w,"exponential")

ks.test(w,pexp,rate=0.095430643,TRUE)

fitdistr(w,"gamma")

ks.test(w,pgamma,shape=5.23424723,rate=0.49950779 ,scale=1/0.49950779 ,TRUE)

g<-fitdist(w,"gamma",start=list(shape=5.23424723,scale=0.49950779))

fitdistr(w,"lognormal")

ks.test(w,"plnorm",meanlog=2.25079862,sdlog=0.48918958,TRUE)

#failure rate#

b<-(1/a^3)*sqrt(2/pi)*x^2*exp(-(x^2)/(2*(a^2)))

plot(b)

y<-(-2*pnorm(x/a))+((1/a)*sqrt(2/pi)*x*exp(-(x^2)/(2*(a^2))))+2)

y

plot(b/y,type="l")

plot(y)

#ypoleipomenos xronos zwhs#

library(ismev)

library(survival)

library(help=survival)

g<-mrl.plot(w,umin=min(w),umax=max(w)-0.1,conf=0.95,nint=100)

erfc <- function(x) 2 * pnorm(x * sqrt(2), lower = FALSE)

ll<-2*a*exp(-(x^2)/(2*(a^2)))*sqrt(2/pi)-x*erfc(x/(sqrt(2)*a))

ll

tt<-ll/y

```

```

tt
rr<-plot(tt,type="l")
lines(g,rr)
t<-seq(0,by=0.1,length=200)
lll<-2*a*exp(-(t^2)/(2*(a^2)))*sqrt(2/pi)-t*erfc(t/(sqrt(2)*a))
lll
yz<-(-2*pnorm(t/a))+((1/a)*sqrt(2/pi)*t*exp(-(t^2)/(2*(a^2)))+2)
tz<-plot(yz)
ty<-lll/yz
ty
rrr<-plot(ty)

```

### **B Τρόπος:**

```

test<-
read_excel("C:/Users/WZ131MR/Desktop/Διπλωματική/survival_data_greece2.xls")
test1<-test$Years
#define the CDF
pmaxbol <- function(data,theta){
  x=data
  a=theta
  2*pnorm(x/a) - (1/a)*sqrt(2/pi)*x*exp(-(x^2)/(2*(a^2)))-1 }
#define the Density function
dmaxbol <- function(data,theta){
  x=data
  a=theta
  1/(a^3)*sqrt(2/pi)*(x^2)*exp(-(x^2)/(2*(a^2))) }
# Neg-log-likelihood function
NLLike = function (theta, data)
{
  NlogL = -sum(log(dmaxbol(data = data, theta = theta)))

```

```

if (is.na(NlogL)) NlogL = .Machine$double.xmax
return(NlogL)
}
# Max likelihood estimation
ldist = function(data, L)
{
  theta = 6.49
  nlminbOut = optim(par = theta, fn = NLike, data = data)
  out = list(fit = nlminbOut)
}
#Fitting the distribution
lparam<-ldist(data=test1)
lparam$fit$par

#GoF tests
ADGofTest::ad.test(test1, pmaxbol, theta = lparam$fit$par)
stats::ks.test(test1,"pmaxbol", theta = lparam$fit$par)
#plot the histogram of x
x <- test1
h<-hist(x, breaks=100, col="red", xlab="Years")
xfit<-seq(min(x),max(x),length=40)
yfit<-dmaxbol(xfit,lparam$fit$par)
yfit <- yfit*diff(h$mids[1:2])*length(x)

```

Ο κώδικας που χρησιμοποιήθηκε για τον έλεγχο προσαρμογής των δεδομένων σε άλλες κατανομές είναι ο εξής:

```

library(MASS)
fitdistr(w,"weibull")
ks.test(w,pweibull,shape=2.7952485,scale=11.7581978,TRUE)
?Exponential(stats)

```

```

fitdistr(w,"exponential")
ks.test(w,pexp,rate=0.095430643,TRUE)
fitdistr(w,"gamma")
ks.test(w,pgamma,shape=5.23424723,rate=0.49950779 ,scale=1/0.49950779 ,TRUE)
g<-fitdist(w,"gamma",start=list(shape=5.23424723,scale=0.49950779))
fitdistr(w,"lognormal")
ks.test(w,"plnorm",meanlog=2.25079862,sdlog=0.48918958,TRUE)

```

## Κώδικας που χρησιμοποιήθηκε στο Mathematica

### Κεφάλαιο 3

#### **Γραφικές παραστάσεις**

$$f[x_]:= (x^2 * \text{Exp}[-\frac{x^2}{2a^2}] * (1/a^3) * ((2/\text{Pi})^{(1/2)}))$$

```
Integrate[f[x], {x, 0, Infinity}]
```

$$g[y_]:= \text{Integrate}[x^2 * \text{Exp}[-\frac{x^2}{2a^2}] * (1/a^3) * ((2/\text{Pi})^{(1/2)}), \{x, 0, y\}]$$

$$\Phi[x_]:= \text{CDF}[\text{NormalDistribution}[0,1], x]$$

```
Integrate[f[x], {x, 0, y}]
```

$$\text{Integrate}[x^2 * \text{Exp}[-\frac{x^2}{2a^2}] * (1/a^3) * ((2/\text{Pi})^{(1/2)}), \{x, 0, t\}]$$

```
Plot[f[x], {x, 0, 10}]
```

$$\text{Plot}[x^2 \frac{1}{a^3} (\sqrt{2} \sqrt{\frac{1}{\text{Pi}}}) \text{Exp}[-\frac{x^2}{2a^2}], \{x, 0, 10\}]$$

$$\text{Plot}[-2 * \Phi[x/a] + ((1/a) * x * \text{Exp}[-x^2/2] * \text{Sqrt}[2/\text{Pi}]) + 2, \{x, 0, 10\}]$$

#### **Βαθμίδα αποτυχίας**

$$r[x_]= (x^2 * \text{Exp}[-\frac{x^2}{2a^2}] * (1/a^3) * ((2/\text{Pi})^{(1/2)})) / (-2 * \Phi[x/a] + ((1/a) * x * \text{Exp}[-x^2/2] * \text{Sqrt}[2/\text{Pi}]) + 2)$$

$$e[x_]= (x^2 * \text{Exp}[-\frac{x^2}{2u^2}] * (1/u^3) * ((2/\text{Pi})^{(1/2)})) / (-2 * \Phi[x/u] + ((1/u) * x * \text{Exp}[-x^2/2] * \text{Sqrt}[2/\text{Pi}]) + 2)$$

```
Plot[{r[x], e[x]}, {x, 0, 8}]
```

#### **Υπολειπόμενος Χρόνος Ζωής**

$$f[x_]:= (x^2 * \text{Exp}[-\frac{x^2}{2a^2}] * (1/a^3) * ((2/\text{Pi})^{(1/2)}))$$

▮

$Fcum[x\_]:=Integrate[f[t],\{t,0,x\}]$

$Assuming[a > 0,FullSimplify[\%]]$

$Assuming[a > 0,Integrate[x * f[x],\{x,0,Infinity\}]$

$tailF[t\_]:=Assuming[a > 0,Integrate[f[x],\{x,t,Infinity\}]$

$$\frac{e^{-\frac{t^2}{2a^2}}\sqrt{\frac{2}{\pi}}t}{a} + Erfc\left[\frac{t}{\sqrt{2a}}\right]$$

$tail2F[t\_]:=Assuming[a > 0,Integrate[tailF[x],\{x,t,Infinity\}]$

$$2ae^{-\frac{t^2}{2a^2}}\sqrt{\frac{2}{\pi}} - tErfc\left[\frac{t}{\sqrt{2a}}\right]$$

$mrl[t\_]:=tail2F[t]/tailF[t]$

$$\frac{2ae^{-\frac{t^2}{2a^2}}\sqrt{\frac{2}{\pi}} - tErfc\left[\frac{t}{\sqrt{2a}}\right]}{\frac{e^{-\frac{t^2}{2a^2}}\sqrt{\frac{2}{\pi}}t}{a} + Erfc\left[\frac{t}{\sqrt{2a}}\right]}$$

$Plot[mrl[t],\{t,0,20\}]$



## 6. Βιβλιογραφία

### Ελληνική

**Μπούτσικας Μιχαήλ (2008).** Σημειώσεις Μαθήματος « Θεωρία Αξιοπιστίας». Τμ.Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς

**Μπούτσικας Μιχαήλ (2004).** Σημειώσεις Μαθήματος «Μέθοδοι Προσομοίωσης και Στατιστικές Υπολογιστικές Τεχνικές». Τμ.Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς

**Μπούτσικας Μιχαήλ (2004),** Σημειώσεις μαθήματος «Στατιστικά Προγράμματα») Τμ.Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς

**Πιτσέλης Γεώργιος (2015).** Σημειώσεις Μαθήματος «Ζημειοκατανομές». Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης ,ΠΜΣ Αναλογιστικής Επιστήμης και Διοικητικής Κινδύνου, Πανεπιστήμιο Πειραιώς

**Πολίτης Κωνσταντίνος (2016),** Σημειώσεις Μαθήματος «Θεωρία Χρεοκοπίας», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης ,ΠΜΣ Αναλογιστικής Επιστήμης και Διοικητικής Κινδύνου ,Πανεπιστήμιο Πειραιώς

**Τσίμπος Κλέων - Φώτης Γεωργιακώδης ,** Βιβλίο: Περιγραφική Στατιστική, Β έκδοση

**Φουσκάκης Δημήτρης (2011).** Σημειώσεις Μαθήματος «Περιγραφική Στατιστική». Τμήμα Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών ,Εθνικό Μετσόβιο Πολυτεχνείο

**Χαραλαμπίδης Χαράλαμπος Α. (2009),** Σημειώσεις Μαθήματος «Βασικές Συνεχείς κατανομές». Τμήμα Μαθηματικών, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

### Ηλεκτρονικές Διευθύνσεις

<http://www2.stat-athens.aueb.gr/~exek/NPar-Statistics/chapter4.pdf>

[www.Wikipedia.gr](http://www.Wikipedia.gr)

[http://www.unipi.gr/faculty/dantz/Introduction\\_to\\_R.pdf](http://www.unipi.gr/faculty/dantz/Introduction_to_R.pdf)

<http://www.iasj.net/iasj?func=fulltext&aid=33656>

<http://www.math.uah.edu/stat/special/Maxwell.html>

[http://www.unipi.gr/faculty/mbouts/prob\\_intro/prob\\_intro4.pdf](http://www.unipi.gr/faculty/mbouts/prob_intro/prob_intro4.pdf)

## Αγγλική

**Calabria, R., & Pulcini, G. (1990).** On the maximum likelihood and least-squares estimation in the Inverse Weibull distribution. *Statistica Applicata*, 2(1), 53–66

**Chin-Diew Lai**, Chapter 2, Generalized Weibull Distributions, Springer 2014

**Erto, P. (1989).** Genesis, properties and identification of the Inverse Weibull lifetime model. *Statistica Applicata*, 1(2), 117–128.

**Kusum Lata Singh and R.S. Srivastava (2014),** Inverse Maxwell Distribution as a Survival Model, Genesis and Parameter Estimation

**Marshall and Olkin (2007)** Book: Inequalities: Theory of Majorization and its Application, Second Edition

**Michael Bean (2001)** Book: Probability: The science of Uncertainty with applications to investments, insurance and engineering

**Norman L Johnson, Samuel Kotz, and N Balakrishnan,** [Continuous Multivariate Distributions: Models and Applications](#)

**R.C.Gupts and D.M. Bradley (2003),** Representing the Mean Residual life in Terms of the Failure Rate, Department of Mathematics and Statistics University of Maine, USA