



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

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

**ΙΔΙΟΤΗΤΕΣ ΚΑΙ ΠΡΟΣΕΓΓΙΣΕΙΣ ΓΙΑ ΜΕΙΞΕΙΣ ΚΑΙ ΣΥΝΕΛΙΞΕΙΣ ΓΑΜΜΑ
ΚΑΤΑΝΟΜΩΝ**

ΤΡΙΑΝΤΟΠΟΥΛΟΥ Α. ΕΛΕΝΗ

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

ΠΕΙΡΑΙΑΣ
Δεκέμβριος, 2020

Η παρούσα διπλωματική εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από την ΓΣΕΣ του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμόν συνεδρίαση του, σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Μεταπτυχιακού Προγράμματος στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου.

Τα μέλη της επιτροπής των καθηγητών ήταν:

- Αναπληρωτής Καθηγητής: Πολίτης Κωνσταντίνος (Επιβλέπων)
- Επίκουρος Καθηγητής: Πιτσέλης Γεώργιος
- Επίκουρος Καθηγητής: Τζαβελάς Γεώργιος

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



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

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

**PROPERTIES AND APPROXIMATIONS FOR MIXTURES AND CONVOLUTIONS
OF GAMMA DISTRIBUTIONS**

TRIANTOPOULOU A. ELENI

MSc Dissertation

Submitted to the Department of Statistics and Insurance Science
of the University of Piraeus as partial fulfillment of the requirements
for the Diploma of Master of Science in Actuarial Science
and Risk Management

Piraeus , Greece

December 2020

This thesis was approved unanimously by the three-member Commission of Inquiry which was appointed by the Department of Statistics and Insurance Science of the University of Piraeus in the No. meeting, according to the internal operational regulations of the postgraduate program in Actuarial Science and Risk Management.

The three committee members were:

- Associate Professor: Politis Konstantinos (Supervisor)
- Assistant Professor: Pitselis Georgios
- Assistant Professor: Tzavelas Georgios

The approval of the thesis by the Department of Statistics and Insurance Science of the University of Piraeus does not imply the acceptance of the opinions of the author.

Στους γονείς μου,
Αναστάσιο και Αγλαΐα

και στην αδερφή μου,
Βασιλική

ΕΥΧΑΡΙΣΤΙΕΣ

Ιδιαίτερες ευχαριστίες για τον επιβλέποντα καθηγητή μου, κύριο Κωνσταντίνο Πολίτη ο οποίος με στήριξε και με καθοδήγησε ορθά καθ' όλη την διάρκεια εκπόνησης της δοθείσας διπλωματικής εργασίας ενώ παράλληλα με βοήθησε να ασχοληθώ σε ερευνητικό επίπεδο με ένα ιδιαίτερα ενδιαφέρον κομμάτι της θεωρίας κινδύνων. Τον ευχαριστώ ολόψυχα για την υπομονή και την στήριξη του διότι χωρίς την πολύτιμη συμβολή του δεν θα ήταν εφικτή η ολοκλήρωση της εργασίας.

Καθώς επίσης θα ήθελα να ευχαριστήσω και τους καθηγητές της τριμελούς επιτροπής, κύριο Πιτσέλη και κύριο Τζαβελά για την σημαντική βοήθεια και την επίβλεψη τους.

Επιπλέον θα ήθελα να εκφράσω βαθύτατη ευγνωμοσύνη στην οικογένεια μου για την συμπαράσταση σε κάθε στάδιο της ζωής μου όλα αυτά τα χρόνια καθώς και την ενθάρρυνση ώστε να πετύχω κάθε στόχο σε επαγγελματικό αλλά και ακαδημαϊκό τομέα.

Τέλος οφείλω να αφιερώσω την δοθείσα διπλωματική εργασία, πέραν από τους γονείς και την αδερφή μου, και στους παππούδες μου οι οποίοι δεν βρίσκονται πλέον στη ζωή αλλά και στην πνευματική μου κόρη Γεωργία, η οποία αποτελεί ένα πολύ σημαντικό πρόσωπο της ζωής μου.

ΠΕΡΙΛΗΨΗ

Ο αναλογισμός αποτελεί μια επιστήμη και έναν κλάδο των εφαρμοσμένων και χρηματοοικονομικών μαθηματικών ο οποίος τα τελευταία χρόνια έχει εξελιχθεί ραγδαία στον τομέα της ασφάλισης. Συγκεκριμένα η θεωρία χρεοκοπίας, ως ένας από τους σημαντικότερους κλάδους της θεωρίας κινδύνου, μελετάει την ανέλιξη του πλεονάσματος δηλαδή τις μεταβολές στα έσοδα και τα έξοδα με την πάροδο του χρόνου για ένα ασφαλιστικό χαρτοφυλάκιο. Παράλληλα ιδιαίτερα σημαντική είναι και η μοντελοποίηση των μεγεθών των αποζημιώσεων στο συλλογικό πρότυπο της θεωρίας κινδύνων και η επιλογή των κατάλληλων κατανομών για την μοντελοποίηση τόσο των ατομικών μεγεθών αλλά και την περιγραφή των συνολικών ζημιών.

Εδώ και αρκετά χρόνια η κατανομή Γάμμα, η οποία είναι μια συνεχής κατανομή με δύο παραμέτρους (κλίμακας και σχήματος), χρησιμοποιείται εκτενώς όχι μόνο στην στατιστική ανάλυση αλλά και στον τομέα του αναλογισμού. Αποτελούν κατανομές οι οποίες πλέον έχουν μελετηθεί λεπτομερώς και για αυτό το λόγο εφαρμόζονται ευρέως στην μοντελοποίηση ασφαλιστικών χαρτοφυλακίων. Παρ' όλα αυτά ιδιαίτερο ενδιαφέρον μελέτης για την αναλογιστική επιστήμη αποτελούν και οι μείξεις αλλά και οι συνελίξεις αυτών των κατανομών. Για τον σκοπό της εργασίας, θα προσπαθήσουμε να εξετάσουμε μέσω παραδειγμάτων την ασυμμετρία και την κύρτωση για τις μείξεις και συνελίξεις Γάμμα κατανομών ώστε να έχουμε μια εικόνα σχετικά με την μορφή τους και να εντοπίσουμε ποιοι συνδυασμοί των παραμέτρων και των βαρών (για τις μείξεις) μας δίνουν συγκεκριμένα χαρακτηριστικά αντίστοιχα. Παράλληλα θα προσπαθήσουμε μέσω άλλων κατανομών να προσεγγίσουμε τις συνελίξεις Γάμμα κατανομών οι οποίες αρκετές φορές δεν έχουν αναλυτικό τύπο και να μελετήσουμε την εφαρμογή αυτών στην θεωρία κινδύνων και στο συλλογικό πρότυπο αποζημιώσεων των ασφαλιστικών χαρτοφυλακίων.

ABSTRACT

Actuarial Science is a branch of applied and financial mathematics which has nowadays developed rapidly in the field of insurance. Specifically, bankruptcy theory, as one of the most important branches of risk theory, studies the evolution of the surplus, id est the changes in the income and the expenses over time for an insurance portfolio. Moreover, it is particularly essential to model the amounts of compensation in the model of collective risk and select the appropriate distributions for the description of both individual sizes and total losses.

Over many years, Gamma distribution which is a continuous distribution with two parameters (scale and shape), is widely used not only in statistical analysis but also in the field of actuarial science. They are distributions that have now been studied in detail and for this reason are widely applied in the modeling of insurance portfolios. Nevertheless, the mixtures and the convolutions of these distributions are of particular interest for the actuarial science and risk management. For the purpose of this thesis, we will try to investigate through examples the asymmetry and the kurtosis for the mixtures and the convolutions of Gamma distributions in order to analyze their shape and identify what combinations of parameters and weights (for mixtures only) give us specific characteristics respectively. At the same time, we will attempt to approach the convolution of Gamma distributions through other distributions because it is widely known that convolutions of Gamma do not always have a closed form and we will also study their application in the risk theory and the collective model of portfolios.

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Πινάκων	1
Κατάλογος Διαγραμμάτων	2
Κατάλογος Σχημάτων	4
Κατάλογος Συντομογραφιών	5
0. Εισαγωγή	6
1. Εισαγωγικές έννοιες και χρησιμότητα των μείξεων-συνελίξεων Γάμμα κατανομών	
1.1 Εισαγωγή στην κατανομή Γάμμα.....	7
1.2 Η κατανομή Erlang.....	14
1.3 Χρησιμότητα των μείξεων και συνελίξεων Γάμμα κατανομών.....	18
2. Βαθμίδα αποτυχίας (Failure rate)	
2.1 Βαθμίδες αποτυχίας για διάφορες περιπτώσεις κατανομών.....	24
2.1.1 Βαθμίδα αποτυχίας για την απλή μορφή της εκθετικής κατανομής.....	24
2.1.2 Βαθμίδα αποτυχίας για την περίπτωση Γάμμα κατανομής.....	25
2.1.3 Βαθμίδα αποτυχίας στην περίπτωση διακριτής μείξης Γάμμα κατανομών.....	27
2.1.4 Βαθμίδα αποτυχίας για την περίπτωση συνεχούς μείξης εκθετικών κατανομών.....	29
2.2 Δύο βασικοί νόμοι που διέπουν την βαθμίδα αποτυχίας.....	30
2.2.1 Νόμος του Compertz.....	30
2.2.2 Νόμος του Makeham.....	31
2.3 Μονοτονία βαθμίδας αποτυχίας και κλάσεις κατανομών.....	32
2.3.1 Μονοτονία βαθμίδων αποτυχίας.....	33
2.3.2 Δύο βασικές κλάσεις κατανομών.....	33
3. Μέσος υπολειπόμενος χρόνος ζωής (Mean Residual Lifetime)	
3.1 Σύνδεση βαθμίδας αποτυχίας και μέσου υπολειπόμενου χρόνου ζωής.....	43
3.2 Επιπρόσθετες κλάσεις κατανομών.....	44
3.2.1 Οικογένειες κλάσης κατανομών NBU και NWU.....	45
3.2.2 Οικογένειες κλάσης κατανομών NBUE και NWUE.....	45
3.2.3 Οικογένειες κλάσης κατανομών UBA και UWA.....	47
3.2.4 Οικογένειες κλάσης κατανομών UBAE και UWAE.....	47
3.2.5 Οικογένειες κλάσης κατανομών HNBUE και HNWUE.....	48
3.2.6 Διάταξη μεταξύ των κλάσεων.....	48

4. Ασυμμετρία για μείξεις και συνελίξεις Γάμμα Κατανομών

4.1 Ασυμμετρία και μέτρα ασυμμετρίας.....	51
4.2 Κύρτωση και μέτρα κύρτωσης.....	55
4.3 Ασυμμετρία και κύρτωση για διάφορες κατανομές.....	57
4.3.1 Εκθετική κατανομή.....	57
4.3.2 Κατανομή Pareto.....	58
4.3.3 Κατανομή Γάμμα.....	60
4.3.4 Κατανομή Weibull.....	61
4.4 Ασυμμετρία και κύρτωση για μείξεις δύο εκθετικών κατανομών.....	64
4.5 Ασυμμετρία και κύρτωση για μείξεις δύο Γάμμα κατανομών.....	68
4.6 Ασυμμετρία και κύρτωση για συνελίξεις Γάμμα κατανομών.....	82

5. Προσεγγίσεις για συνελίξεις Γάμμα κατανομών και εφαρμογές στη θεωρία κινδύνου

5.1 Ακριβής πυκνότητα για την συνέλιξη Γάμμα κατανομών όταν οι παράμετροι σχήματος είναι ακέραιοι αριθμοί.....	88
5.2 Προσέγγιση για την συνέλιξη Γάμμα κατανομών όταν δεν υπάρχει κλειστός τύπος...91	
5.2.1 Η ακριβής συνάρτηση πυκνότητας για την συνέλιξη Γάμμα κατανομών.....92	
5.2.2 Ο προσεγγιστικός τύπος για την κατανομή των βαρών στη συνέλιξη.....94	
5.2.3 Εφαρμογές μέσω ενός παραδείγματος συνέλιξης 3 Γάμμα κατανομών.....99	
5.3 Εφαρμογές των συνελίξεων Γάμμα κατανομών στην Θεωρία Κινδύνου.....109	
5.3.1 Προσεγγίσεις για την κατανομή των συνολικών αποζημιώσεων..... 110	
5.3.2 Προσεγγίσεις για την κατανομή των συνολικών αποζημιώσεων με την χρήση συνελίξεων Γάμμα κατανομών..... 112	

Παράρτημα.....	121
-----------------------	------------

Βιβλιογραφία.....	133
--------------------------	------------

- 4.1** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για διάφορες μείξεις 2 εκθετικών κατανομών..... 66
- 4.2** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για διάφορες μείξεις 2 Γάμμα κατανομών με κοινή παράμετρο κλίμακας (b)..... 71
- 4.3** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για μείξεις 2 Γάμμα κατανομών με διαφορετική παράμετρο κλίμακας (b) και για διάφορες περιπτώσεις παραμέτρων σχήματος (a)..... 72
- 4.4** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για μείξεις 2 Γάμμα κατανομών με διαφορετική παράμετρο κλίμακας (b) όπου οι παράμετροι σχήματος και για τις δύο κατανομές είναι μικρότερες της μονάδας ($a_1 < 1$ και $a_2 < 1$)..... 78
- 4.5** Ζεύγη θετικών συντελεστών ασυμμετρίας και συντελεστών κύρτωσης > 3 , για μείξεις 2 Γάμμα κατανομών όταν για τις ίδιες μείξεις διατηρούμε σταθερές τις παραμέτρους κλίμακας και μεταβάλλουμε μόνο τις παραμέτρους σχήματος ($a_1 < 1$ και $a_2 < 1$) καθώς και τα βάρη αντίστοιχα..... 79
- 4.6** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για μείξεις 2 Γάμμα κατανομών όπου η παράμετρος σχήματος είναι ίση με 20 και για τις δύο κατανομές Γάμμα που συνθέτουν την μείξη..... 81
- 4.7** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για συνελίξεις 2 Γάμμα κατανομών με διαφορετική παράμετρο κλίμακας (b) και κοινή παράμετρο σχήματος (a)..... 85
- 4.8** Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για συνελίξεις 2 Γάμμα κατανομών με διαφορετική παράμετρο κλίμακας (b) και για διάφορες περιπτώσεις παραμέτρων σχήματος (a)..... 87
- 5.1** Τιμές ποσοστημορίων και ποσοστών απόκλισης για την ακριβή συνάρτηση κατανομής μιας συνέλιξης 3 Γάμμα κατανομών και την προσεγγιστική συνάρτηση κατανομής αντίστοιχα..... 108

1.1	Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της $Gamma(0.5,2)$ (κόκκινη καμπύλη) και της $Gamma(3.5,5.5)$ (μπλε καμπύλη).....	13
1.2	Διάγραμμα της συνάρτησης κατανομής της $Gamma(0.5,2)$ (κόκκινη καμπύλη) και της $Gamma(3.5,5.5)$ (μπλε καμπύλη).....	13
1.3	Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της $Erl(2,3)$ (κόκκινη καμπύλη) και της $Erl(6,1.5)$ (μπλε καμπύλη).....	17
1.4	Διάγραμμα της συνάρτησης κατανομής της $Erl(2,3)$ (κόκκινη καμπύλη) και της $Erl(6,1.5)$ (μπλε καμπύλη).....	17
2.1	Βαθμίδα αποτυχίας για την εκθετική κατανομή με παράμετρο $\lambda=0.25$ (κόκκινη γραμμή) και παράμετρο $\lambda=2.5$ (μπλε γραμμή).....	36
2.2	Βαθμίδα αποτυχίας για την Γάμμα κατανομή με παραμέτρους $(0.5,3)$, $(0.5,1.5)$ και $(4,2)$	36
2.3	Βαθμίδα αποτυχίας για την κατανομή Erlang με παραμέτρους $(3,0.5)$, $(2,0.5)$ και $(2,6)$	37
3.1	Μέσος υπολειπόμενος χρόνος ζωής για την εκθετική κατανομή με παράμετρο $\lambda=0.25$ (κόκκινη γραμμή) και παράμετρο $\lambda=2.5$ (μπλε γραμμή).....	49
3.2	Μέσος υπολειπόμενος χρόνος ζωής για την κατανομή Erlang με παραμέτρους $(3,0.25)$, $(2,0.25)$ και $(2,4)$	49
3.3	Μέσος υπολειπόμενος χρόνος ζωής για την κατανομή Weibull με παραμέτρους $(0.5,6)$, $(0.5,2.5)$ και $(4,2)$	50
4.1	Συνάρτηση πυκνότητας πιθανότητας της $p \cdot Gamma(a_1, b_1) + (1 - p) \cdot Gamma(a_2, b_2)$ (πράσινη καμπύλη) με $a_1 = 4, b_1 = 1, a_2 = 10, b_2 = 0.5$ και των συναρτήσεων $Gamma(4,1)$ (κόκκινη καμπύλη) και $Gamma(10,0.5)$ (μπλε καμπύλη). Έχουν επιλεγθεί βάρη $p = 0.15$ και $1 - p = 0.85$ αντίστοιχα.....	69
4.2	Συνάρτηση πυκνότητας πιθανότητας (αριστερά) και βαθμίδα αποτυχίας (δεξιά) για την μείξη των κατανομών $Gamma(20,5)$ και $Gamma(20,30)$ με βάρη $p = 0.8$ και $1 - p = 0.2$ αντίστοιχα, για την οποία διαπιστώσαμε αρνητική ασυμμετρία.....	74
4.3	Συνάρτηση πυκνότητας της βαθμίδα αποτυχίας του διαγράμματος (4.2) στο διάστημα $[15,17)$	74
4.4	Συνάρτηση πυκνότητας πιθανότητας (αριστερά) και συνάρτηση της βαθμίδα αποτυχίας (δεξιά) για την μείξη των κατανομών $Gamma(4,7)$ και $Gamma(30,20)$ με βάρη $p = 0.2$ και $1 - p = 0.8$ αντίστοιχα, για την οποία επίσης παρατηρήθηκε, μέσω του Πίνακα 4.3, αριστερή ή αρνητική ασυμμετρία.....	75

4.5 Συνάρτηση πυκνότητας πιθανότητας και συνάρτηση της βαθμίδας αποτυχίας για την μείξη των κατανομών $\text{Gamma}(1,4)$ (η οποία είναι μια εκθετική κατανομή καθώς $a_1 = 1$) και $\text{Gamma}(10,4)$ με βάρη $p = 0.1$ και $1 - p = 0.9$ αντίστοιχα.....	76
4.6 Συνάρτηση πυκνότητας πιθανότητας και συνάρτηση βαθμίδας αποτυχίας για την μείξη των κατανομών $\text{Gamma}(5,10)$ και $\text{Gamma}(40,10)$ με βάρη $p = 0.1$ και $1 - p = 0.9$ αντίστοιχα.....	76
4.7 Συνάρτηση πυκνότητας πιθανότητας για την μείξη των κατανομών $\text{Gamma}(4,0.3)$ και $\text{Gamma}(8,0.3)$ με βάρη $p = 0.5$ και $1 - p = 0.5$ αντίστοιχα.....	77
4.8 Συνάρτηση πυκνότητας πιθανότητας για την μείξη των κατανομών $\text{Gamma}(0.5,0.6)$ και $\text{Gamma}(0.8,0.6)$ με βάρη $p = 0.8$ και $1 - p = 0.2$ αντίστοιχα.....	77
4.9 Συνάρτηση πυκνότητας πιθανότητας για την μείξη των κατανομών $\text{Gamma}(8,6)$ και $\text{Gamma}(20,12)$ με βάρη $p = 0.75$ και $1 - p = 0.25$ αντίστοιχα.....	77
4.10 Συνάρτηση πυκνότητας πιθανότητας για την μείξη των κατανομών $\text{Gamma}(0.5,6)$ και $\text{Gamma}(10,30)$ με βάρη $p = 0.6$ και $1 - p = 0.4$ αντίστοιχα.....	77
4.11 Συνάρτηση πυκνότητας πιθανότητας για την μείξη των κατανομών $\text{Gamma}(4,7)$ και $\text{Gamma}(30,20)$ με βάρη $p = 0.75$ και $1 - p = 0.25$ αντίστοιχα.....	77
4.12 Συνάρτηση πυκνότητας πιθανότητας για την μείξη των κατανομών $\text{Gamma}(20,5)$ και $\text{Gamma}(20,30)$ με βάρη $p = 0.3$ και $1 - p = 0.7$ αντίστοιχα.....	77
5.1 Πυκνότητα συνέλιξης των $\text{Gamma}(3,6)$, $\text{Gamma}(10,8)$ και $\text{Gamma}(20,10)$ (αριστερά) και συνάρτηση κατανομής (δεξιά) με βάση τον τύπο του Moschopoulos.....	101
5.2 Προσέγγιση της πυκνότητας της συνέλιξης των $\text{Gamma}(3,6)$, $\text{Gamma}(10,8)$ και $\text{Gamma}(20,10)$ (αριστερά) και προσέγγιση της συνάρτησης κατανομής (δεξιά) με βάση τον τύπο του Barnabani.....	102
5.3 Γραφική απεικόνιση της απόλυτης διαφοράς μεταξύ της ακριβούς πυκνότητας της συνέλιξης των $\text{Gamma}(3,6)$, $\text{Gamma}(10,8)$ και $\text{Gamma}(20,10)$ και της προσέγγισης στο διάστημα $(0,7)$	103
5.4 Γραφική απεικόνιση της απόλυτης διαφοράς μεταξύ της ακριβούς συνάρτησης κατανομής της συνέλιξης των $\text{Gamma}(3,6)$, $\text{Gamma}(10,8)$ και $\text{Gamma}(20,10)$ και της προσέγγισης στο διάστημα $(0,7)$	104

3.1 Σχηματική απεικόνιση της τ.μ. $T(x)$ που εκφράζει τον μέσο υπολειπόμενο χρόνο ζωής.....	39
3.2 Σχηματική απεικόνιση για τις οικογένειες κλάσης κατανομών NBUE(NWUE), DMRL(IMRL), NBU(NWU) και HNBUE(HNWUE).....	48
4.1 Σχηματική απεικόνιση κατανομών με θετική ασυμμετρία.....	52
4.2 Σχηματική απεικόνιση κατανομών με αρνητική ασυμμετρία.....	52
4.3 Σχηματική απεικόνιση για μεσόκυρτες κατανομές.....	56
4.4 Σχηματική απεικόνιση για λεπτόκυρτες κατανομές.....	56
4.5 Σχηματική απεικόνιση για πλατύκυρτες κατανομές.....	56

ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

τ.μ.	τυχαία μεταβλητή
σ.π.π	συνάρτηση πυκνότητας πιθανότητας
σ.κ.	συνάρτηση κατανομής
IFR	Increasing Failure Rate
DFR	Decreasing Failure Rate
NBU	New Better than Used
NWU	New Worse than Used
NBUE	New Better than Used in Expectation
NWUE	New Worse than Used in Expectation
UBA	Used Better than Aged
UWA	Used Worse than Aged
UBAE	Used Better than Aged in Expectation
UWAE	Used Worse than Aged in Expectation
HNBU	Harmonic New Better than Used in Expectation
HNWUE	Harmonic New Worse than Used in Expectation
βλ.	βλέπε

ΕΙΣΑΓΩΓΗ

Στην δοθείσα διπλωματική εργασία θα γίνει μια διεξοδική μελέτη τόσο των μείξεων όσο και των συνελίξεων Γάμμα κατανομών, δίνοντας κυρίως έμφαση στο δεύτερο, οι οποίες χρησιμοποιούνται ευρέως στην αναλογιστική επιστήμη και ειδικότερα στην θεωρία συλλογικού κινδύνου ως μοντέλα ατομικών ή συνολικών ζημιών και σε διάφορες προσεγγιστικές μεθόδους.

Το κατάλληλο λογισμικό που θα χρησιμοποιηθεί είναι το πρόγραμμα Mathematica το οποίο θα βοηθήσει στην δημιουργία αντίστοιχων διαγραμμάτων αλλά και παραδειγμάτων για την μελέτη αυτής της ομάδας κατανομών καθώς επίσης θα γίνει χρήση και της γλώσσας προγραμματισμού R κυρίως στο πέμπτο κεφάλαιο της διατριβής. Θα μελετήσουμε τη χρήση των συνελίξεων Γάμμα σε προσεγγιστικές μεθόδους που αφορούν την κατανομή των συνολικών αποζημιώσεων στη θεωρία κινδύνου. Ενώ παράλληλα θα δοθεί έμφαση σε κάποιες βασικές ιδιότητες των μείξεων και συνελίξεων Γάμμα όπως η ασυμμετρία αλλά και η κύρτωση αντίστοιχα. Τέλος θα γίνει προσπάθεια προσέγγισης των πυκνοτήτων των συνελίξεων Γάμμα μέσω άλλων κατανομών.

Η παρούσα εργασία στοχεύει στην διεξοδική ανάλυση των παραπάνω εννοιών και την κατανόηση της βαθύτατης σημασίας που έχουν οι συγκεκριμένες κατανομές στο πεδίο της αναλογιστικής επιστήμης. Στο πρώτο κεφάλαιο θα έχουμε μια μικρή εισαγωγή στις Γάμμα κατανομές αλλά και τις κατανομές Erlang ως μια ειδική κατηγορία των Γάμμα, τις συναρτήσεις πυκνότητας πιθανότητάς τους καθώς επίσης και σε διάφορα χαρακτηριστικά στοιχεία αυτών των κατανομών όπως είναι η μέση τιμή και η διακύμανση. Επιπλέον αρκετές από τις μεθόδους προσέγγισης προβλημάτων όπως είναι η περιγραφή του χρόνου ζωής θα μας απασχολήσουν ιδιαίτερα στην παρούσα διπλωματική και θα μελετηθούν μέσω συγκεκριμένων συναρτήσεων όπως η βαθμίδα αποτυχίας και ο μέσος υπολειπόμενος χρόνος ζωής, θεωρώντας ότι ο χρόνος ακολουθεί διάφορες περιπτώσεις κατανομών αντίστοιχα, ώστε να γίνει μια λεπτομερή μελέτη αυτών των ιδιοτήτων. Οι έννοιες αυτές θα αναλυθούν στο δεύτερο και τρίτο κεφάλαιο της εργασίας αντίστοιχα. Στο τέταρτο κεφάλαιο θα μελετήσουμε διεξοδικά μια από τις σημαντικότερες ιδιότητες για την μορφή μιας κατανομής, την ασυμμετρία ενώ παράλληλα θα εξετάσουμε και μία συναφή έννοια με την ασυμμετρία η οποία είναι η κύρτωση υπό τις μείξεις και συνελίξεις Γάμμα κατανομών. Μέσω διαγραμμάτων και αντίστοιχων παραδειγμάτων θα προσπαθήσουμε από μείξεις και συνελίξεις 2 Γάμμα κατανομών να οδηγηθούμε σε ένα γενικότερο συμπέρασμα σχετικά με τις ιδιότητες αυτές. Στο τελευταίο κεφάλαιο θα παρουσιάσουμε προσεγγίσεις για συνελίξεις Γάμμα κατανομών καθώς αρκετές φορές δεν υπάρχει κλειστός τύπος για την συνάρτηση πυκνότητας πιθανότητας της συγκεκριμένης οικογένειας κατανομών. Ωστόσο μέσω συγκεκριμένων μεθόδων και συγκρίσεων, θα ερευνήσουμε κατά πόσο οι προσεγγίσεις αυτές πλησιάζουν την ακριβή κατανομή. Στο τελευταίο κεφάλαιο θα δούμε και μέσω κατάλληλης θεωρίας πώς αυτές οι κατανομές χρησιμεύουν στην θεωρία κινδύνου και το συλλογικό πρότυπο αποζημιώσεων μέσω διαφόρων προσεγγιστικών μεθόδων για την κατανομή των συνολικών απαιτήσεων.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΧΡΗΣΙΜΟΤΗΤΑ ΤΩΝ ΜΕΙΞΕΩΝ-ΣΥΝΕΛΙΞΕΩΝ ΓΑΜΜΑ ΚΑΤΑΝΟΜΩΝ

1.1 Εισαγωγή στην Κατανομή Γάμμα

Η κατανομή Γάμμα αποτελεί μια συνεχή κατανομή η οποία με την πάροδο των χρόνων έχει αρχίσει να χρησιμοποιείται ευρέως τόσο στην στατιστική επιστήμη όσο και στον αναλογισμό. Η χρήση της στον τομέα της αναλογιστικής επιστήμης και της θεωρίας κινδύνου είναι ιδιαίτερα ζωτική καθώς μπορεί να προσεγγίσει την κατανομή των συνολικών αποζημιώσεων στο συλλογικό πρότυπο ενώ παράλληλα να χρησιμοποιηθεί για την μοντελοποίηση των συνολικών αλλά και των ατομικών ζημιών ενός ασφαλιστικού χαρτοφυλακίου. Αρκετές από τις κατανομές των συνολικών αποζημιώσεων έχουν περίπου παρόμοια χαρακτηριστικά με τις κατανομές Γάμμα, δηλαδή δεξιά (ή θετική) ασυμμετρία, κάτι που ισχύει για την περίπτωση των κατανομών Γάμμα όπως θα εξετάσουμε αναλυτικά στην ενότητα 4.3 του κεφαλαίου 4, αλλά επίσης παίρνουν και μη αρνητικές τιμές. Παράλληλα αποτελούν κατανομές οι οποίες έχουν μελετηθεί εκτενώς από μια πληθώρα επιστημόνων και είναι ιδιαίτερα γνωστές οι ιδιότητές τους. Τέλος πολύ γνωστές κατανομές στον αναλογισμό όπως η εκθετική και η χ^2 αποτελούν δύο ειδικές περιπτώσεις των κατανομών Γάμμα.

Η κατανομή Γάμμα με θετικές παραμέτρους (a, b) έχει την εξής συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{b^a \cdot x^{a-1} \cdot e^{-b \cdot x}}{\Gamma(a)}, & x > 0 \end{cases} \quad (1.1.1)$$

όπου $b > 0$ η οποία ονομάζεται παράμετρος κλίμακας (scale parameter) ενώ $a > 0$ η αντίστοιχη παράμετρος σχήματος (shape parameter).

Μπορούμε εναλλακτικά εάν κάνουμε χρήση μιας διαφορετικής παραμετροποίησης να ορίσουμε την συνάρτηση πυκνότητας πιθανότητας ως εξής:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x^{a-1} \cdot e^{-\frac{x}{b}}}{b^a \cdot \Gamma(a)}, & x > 0 \end{cases} \quad (1.1.2)$$

Η πυκνότητα f είναι παντού συνεχής ως προς x , εκτός από το σημείο $x = 0$ στο οποίο παρουσιάζει μια ασυνέχεια όταν $0 < a < 1$.

Επιπλέον από τις παραπάνω σχέσεις για την συνάρτηση πυκνότητας πιθανότητας μπορούμε να διαπιστώσουμε ότι για παράμετρο σχήματος ίση με την μονάδα ($\alpha = 1$) έχουμε αντίστοιχα τον τύπο που ισχύει για την συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής γεγονός το οποίο δικαιολογεί το ότι αποτελεί μια ειδική περίπτωση της Γάμμα.

Στην συνέχεια θα κάνουμε χρήση του τύπου (1.1.2) για την παρουσίαση των παρακάτω αποτελεσμάτων. Το σύμβολο Γ στην σχέση (1.1.2) αναφέρεται στην συνάρτηση Γάμμα η οποία ορίζεται ως εξής για οποιοδήποτε x θετικό:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} \cdot e^{-t} dt, \quad 0 < x < \infty \quad (1.1.3)$$

Για να είναι η συνάρτηση f μια συνάρτηση πυκνότητας πιθανότητας θα πρέπει να ισχύει ότι: $\int_0^{\infty} f(x) dx = 1$ και συνεπώς έχουμε: $\int_0^{\infty} x^{a-1} \cdot e^{-x/b} = b^a \cdot \Gamma(a)$.

Η παραπάνω συνάρτηση Γάμμα εισήχθη για πρώτη φορά το 1730 από τον Ελβετό μαθηματικό Leonhard Euler (1707-1783) ως μια γενίκευση της παραγοντικής συνάρτησης $x!$, για μη ακέραιες τιμές του x (Bahnmann, 2015). Ο Euler πρότεινε τον τύπο:

$$\Gamma(x) = \int_0^1 \left[\log\left(\frac{1}{t}\right)\right]^{x-1} dt, \quad \text{ο οποίος είναι ισοδύναμος με την σχέση (1.1.3).}$$

Ωστόσο το σύμβολο Γ οφείλεται στον Γάλλο μαθηματικό Adrien-Marie Legendre (1752-1833).

Ας δούμε κάποιες από τις βασικότερες ιδιότητες της συνάρτησης Γάμμα ($\Gamma(x)$):

$$(\alpha) \Gamma(1) = 1$$

$$(\beta) \Gamma(x + 1) = x \cdot \Gamma(x), \quad x > 0$$

$$(\gamma) \Gamma(n + 1) = n!, \quad n = 1, 2, 3 \dots \dots$$

$$(\delta) \frac{\Gamma(x+n)}{\Gamma(x)} = \prod_{i=0}^{n-1} (x + i), \quad 0 < x < \infty \text{ και } n = 1, 2, 3 \dots$$

Η ιδιότητα (γ) δείχνει ότι η $\Gamma(x)$ είναι μια επέκταση της παραγοντικής συνάρτησης $n! = 1 \cdot 2 \cdot \dots \cdot (n - 1) \cdot n$ (όπου n είναι θετικός ακέραιος) σε όλους τους θετικούς πραγματικούς αριθμούς με αποτέλεσμα να έχουμε μια συνεχή παρεμβολή μεταξύ διαδοχικών ακέραιων παραγόντων.

Επιπλέον υπάρχει και η μη πλήρης συνάρτηση Γάμμα (incomplete Gamma Function) η οποία ορίζεται για θετικές πραγματικές τιμές του x μέσω του παρακάτω ολοκληρώματος:

$$\Gamma(x, a) = \int_0^x t^{a-1} \cdot e^{-t} dt, \quad \text{για } a > 0 \quad (1.1.4)$$

Είναι φανερό ότι ισχύει η εξής σχέση όταν το x πάρει πολύ μεγάλες τιμές κοντά στο άπειρο:

$$\lim_{x \rightarrow \infty} \Gamma(x, a) = \Gamma(a)$$

Έστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Γάμμα με συνάρτηση πυκνότητας πιθανότητας όπως ορίστηκε στην σχέση (1.1.2). Η συνάρτηση κατανομής (cumulative function) της τ.μ. X μπορεί να εκφραστεί μέσω μιας μη πλήρους συνάρτησης Γάμμα. Ας δούμε αρχικά το ολοκλήρωμα $I(x) = \int_0^x f(u)du$, για $0 < x < \infty$ το οποίο αποτελεί τον ορισμό για την συνάρτηση κατανομής όπως γνωρίζουμε από τις πιθανότητες. Αν αλλάξουμε την μεταβλητή u και την αντικαταστήσουμε με $u = b \cdot t$ τότε θα έχουμε αντίστοιχα (Bahnmann, 2015):

$$I(x) = \int_0^x \frac{u^{a-1} \cdot e^{-\frac{u}{b}}}{b^a \cdot \Gamma(a)} = \frac{1}{\Gamma(a)} \int_0^{x/b} t^{a-1} \cdot e^{-t} dt = \frac{\Gamma(\frac{x}{b}, a)}{\Gamma(a)}$$

όπου $\Gamma(x, a)$ είναι η μη πλήρης συνάρτηση Γάμμα όπως ορίστηκε προηγουμένως.

Συνεπώς για την συνάρτηση κατανομής της τ.μ. X η οποία ακολουθεί την $Gamma(a, b)$ θα έχουμε:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{\Gamma(\frac{x}{b}, a)}{\Gamma(a)}, & 0 < x < \infty \end{cases}$$

Μπορούμε ωστόσο μέσω της ίδιας διαδικασίας να παραγάγουμε και τις ροπές m τάξης αντίστοιχα. Αυτή την φορά χρησιμοποιούμε το ολοκλήρωμα:

$$I_m(x) = \int_0^x u^m \cdot f(u)du, \text{ για } 0 < x < \infty$$

Εάν όπως προηγουμένως θέσουμε όπου $u = b \cdot t$, για $m = 1, 2, 3 \dots$ τότε προκύπτει:

$$I_m(x) = \int_0^x u^m \cdot \frac{u^{a-1} \cdot e^{-\frac{u}{b}}}{b^a \cdot \Gamma(a)} du = \frac{b^m}{\Gamma(a)} \int_0^{x/b} t^{a+m-1} e^{-t} dt = \frac{b^m}{\Gamma(a)} \cdot \Gamma(\frac{x}{b}, a + m)$$

Ωστόσο για τις ροπές m -τάξης θα ισχύει βάσει ορισμού από τις πιθανότητες:

$$E[X^m] = \lim_{x \rightarrow \infty} I_m(x) = \frac{b^m}{\Gamma(a)} \cdot \Gamma(a + m) = a \cdot (a + 1) \cdot \dots \cdot (a + m - 1) \cdot b^m$$

Ενώ επιπλέον $E[X^m; x] = I_m(x) + x^m \cdot (1 - F(x)) = E(X^m) \cdot \frac{\Gamma(\frac{x}{b}, a+m)}{\Gamma(a+m)} + x^m \cdot (1 - \frac{\Gamma(\frac{x}{b}, a)}{\Gamma(a)})$

όπου ως $E[X^m; l] = \int_0^l x^m dF_X(x) + l^m \cdot (1 - F_X(l))$ ορίζεται η περιορισμένη ροπή τάξης m όταν η τ.μ. X στον αναλογισμό περιγράφει ατομικές ζημιές οι οποίες περιορίζονται από ένα

όριο l έτσι ώστε η ασφαλιστική εταιρεία να πληρώνει ολόκληρη την ζημιά όταν είναι μικρότερη από το παραπάνω όριο και να πληρώνει το μέγιστο όριο όταν η ζημιά το υπερβαίνει.

Η τ.μ. Y που περιγράφει τις παραπάνω ζημιές ορίζεται αντίστοιχα ως (Πιτσέλης, 2018):

$$Y = \min\{X, l\} = \begin{cases} X, & \text{για } 0 \leq X < l \\ l, & \text{για } l \leq X < \infty \end{cases}$$

Μέσω των παραπάνω σχέσεων προκύπτουν οι εξής τύποι για την μέση τιμή, την διακύμανση αλλά και τον συντελεστή ασυμμετρίας των κατανομών Γάμμα (τον τύπο για τον συντελεστή ασυμμετρίας θα τον μελετήσουμε διεξοδικά στο κεφάλαιο 4):

α) $E(X) = a \cdot b$

β) $Var(X) = a \cdot b^2$

γ) $\gamma_1 = \frac{2}{\sqrt{a}}$

Επιπλέον μια εξίσου σημαντική ιδιότητα των κατανομών Γάμμα, την οποία θα παρουσιάσουμε αναλυτικά στην ενότητα 4.6 του κεφαλαίου 4, είναι ότι το άθροισμα ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν την κατανομή Γάμμα με ίδια παράμετρο κλίμακας b ακολουθεί επίσης μια κατανομή Γάμμα. Το αποτέλεσμα αυτό είναι ιδιαίτερα σημαντικό για τις **συνελίξεις** Γάμμα κατανομών οι οποίες αποτελούν το κύριο αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας.

Ωστόσο το παραπάνω αποτέλεσμα προκύπτει άμεσα και μέσω της χρήσης των ροπογεννητριών.

Γνωρίζουμε ότι όταν εάν η τ.μ. X ακολουθεί την $Gamma(a, b)$ τότε ισχύει:

$$\begin{aligned} M_X(t) &= E[e^{tX}] = \frac{1}{b^a \cdot \Gamma(a)} \cdot \int_0^\infty x^{a-1} \cdot e^{-x \cdot (\frac{1}{b} - t)} dx \\ &= \frac{1}{b^a \cdot \Gamma(a)} \cdot \Gamma(a) \cdot (\frac{1}{b} - t)^{-a} \\ &= (1 - bt)^{-a}, \quad -\infty < t < \frac{1}{b} \end{aligned}$$

Το δεύτερο βήμα της παραπάνω ισότητας προκύπτει κάνοντας χρήση του $\int_0^\infty x^{a-1} \cdot e^{-x/b} = b^a \cdot \Gamma(a)$ εφόσον η f είναι συνάρτηση πυκνότητας πιθανότητας.

Ο περιορισμός $-\infty < t < \frac{1}{b}$ στην $M_X(t)$ ως συνάρτηση του t , μας εξασφαλίζει ότι το ολοκλήρωμα του πρώτου βήματος θα συγκλίνει και δεν θα απειρίζεται.

Ας θεωρήσουμε τις ανεξάρτητες και τυχαίες μεταβλητές $\{X_i\}$ η οποίες θα ακολουθούν Γάμμα κατανομή με διαφορετικές παραμέτρους σχήματος a_i και κοινή παράμετρο κλίμακας b δηλαδή ακολουθούν τις $Gamma(a_i, b)$.

Εφόσον έχουμε υποθέσει ανεξαρτησία μεταξύ των $\{X_i\}$ τότε για την ροπογεννήτρια της τ.μ. του αθροίσματος $Y = \sum_{i=1}^n X_i$ θα μπορούμε να την γράψουμε ως το γινόμενο των αντίστοιχων ροπογεννητριών ως εξής:

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (1 - b \cdot t)^{-a_i} = (1 - b \cdot t)^{-\sum_{i=1}^n a_i}, \quad -\infty < t < \frac{1}{b}$$

Η οποία ροπογεννήτρια είναι επίσης ροπογεννήτρια μιας κατανομής Γάμμα με παραμέτρους $(\sum_{i=1}^n a_i, b)$. Συνεπώς $Y \sim \text{Gamma}(\sum_{i=1}^n a_i, b)$.

Στην συνέχεια ας δούμε περιληπτικά και τις δύο βασικές περιπτώσεις των κατανομών Γάμμα:

1) Όπως αναφέραμε και προηγουμένως η πιο απλή περίπτωση για την κατανομή Γάμμα είναι όταν η παράμετρος σχήματος είναι σταθερή και ίση με την μονάδα ($a = 1$). Τότε η κατανομή θα ανήκει στην οικογένεια των εκθετικών κατανομών. Η εκθετική κατανομή (exponential distribution) είναι γνωστή και ως κατανομή Laplace (Laplace distribution) προς τιμήν του Γάλλου μαθηματικού και φυσικού Pierre-Simon Laplace (1749-1847), ο οποίος είχε μεγάλη συνεισφορά στον κλάδο των μαθηματικών και κυρίως την θεωρία πιθανοτήτων.

Εάν η τ.μ. $X \sim \text{Exp}(b)$ τότε θα έχει συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \left(\frac{1}{b}\right) \cdot e^{-\frac{x}{b}}, & 0 < x < \infty \end{cases}$$

και αθροιστική συνάρτηση κατανομής αντίστοιχα:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\frac{x}{b}}, & 0 \leq x < \infty \end{cases}$$

Ο υπολογισμός των ροπών προκύπτει με αντίστοιχο τρόπο όπως δείξαμε προηγουμένως για την γενική περίπτωση των κατανομών Γάμμα. Επομένως έχουμε για την μέση τιμή, την διακύμανση αλλά και τον συντελεστή ασυμμετρίας (θα τον υπολογίσουμε αναλυτικά στην ενότητα 4.3 του 4^{ου} κεφαλαίου):

α) $E(X) = b$

β) $\text{Var}(X) = b^2$

γ) $\gamma_1 = 2$

Επιπλέον για την περιορισμένη μέση τιμή όταν έχουμε μοντέλο ζημιών με ανώτερο όριο ισχύει:

$$E[X; x] = b \cdot \left(1 - e^{-\frac{x}{b}}\right) = E(X) \cdot \left(1 - e^{-\frac{x}{b}}\right)$$

Όπως και οι κατανομές Γάμμα στην γενική τους μορφή, έτσι και οι εκθετικές κατανομές έχουν μεγάλη εφαρμογή στο πεδίο της αναλογιστικής επιστήμης αλλά παρόλα αυτά έχουν περιορισμένη αξία ιδίως ως κατανομές αποζημιώσεων στην θεωρία κινδύνου. Αποτελούν

κατανομές οι οποίες έχοντας μόνο μια παράμετρο δεν είναι ιδιαίτερα ευέλικτες ώστε να προσαρμοστούν σε αναλογιστικά δεδομένα. Για τον σκοπό αυτό γίνεται συχνά χρήση των μείξεων εκθετικών κατανομών για την μοντελοποίηση των ατομικών ζημιών στην θεωρία κινδύνου καθώς με αυτό τον τρόπο επιτυγχάνεται η χρήση κατανομών με περισσότερες από μια παραμέτρους όπως έχει η απλή εκθετική. Στην συνέχεια, στο κεφάλαιο 2, θα ορίσουμε αναλυτικά την έννοια της μείξης κατανομών.

Ωστόσο αν αναφέρουμε ότι μια μείξη 2 εκθετικών κατανομών θα έχει αθροιστική συνάρτηση κατανομής (Πολίτης 2018):

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - p \cdot e^{-\frac{x}{b_1}} - (1-p) \cdot e^{-\frac{x}{b_2}}, & 0 \leq x < \infty \end{cases}$$

με 3 παραμέτρους αντίστοιχα, $b_1 > 0$, $b_2 > 0$ και $0 < p < 1$

2) Μία εξίσου σημαντική περίπτωση για την οικογένεια των Γάμμα κατανομών είναι η κατανομή χ^2 -τετράγωνο ($\chi^2(n)$) με n βαθμούς ελευθερίας. Η κατανομή αυτή προκύπτει όταν η παράμετρος σχήματος πάρει την τιμή $a = 1/2 \cdot n$ όπου n θετικός ακέραιος και η παράμετρος κλίμακας πάρει την τιμή $b = 2$. Η συνάρτηση πυκνότητας πιθανότητας για την $\chi^2(n)$ θα είναι η παρακάτω:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x^{n/2-1} \cdot e^{-\frac{x}{2}}}{2^{n/2} \cdot \Gamma\left(\frac{1}{2} \cdot n\right)}, & x > 0 \end{cases}$$

Επιπλέον όταν η τ.μ. $X \sim \chi^2(n)$ τότε θα έχει αντιστοίχως την παρακάτω μέση τιμή, διακύμανση και συντελεστή ασυμμετρίας:

α) $E[X] = n$

β) $Var[X] = 2 \cdot n$

γ) $\gamma_1 = 2 \cdot \sqrt{\frac{2}{n}}$

Η κατανομή χ^2 -τετράγωνο αποτελεί και την κατανομή του αθροίσματος των τετραγώνων ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν την κανονική κατανομή $N(\mu, \sigma^2)$ με μέση τιμή μ και τυπική απόκλιση ίση με σ . Ως κατανομή είναι ιδιαίτερα γνωστή κυρίως για την εφαρμογή της στους ελέγχους καλής προσαρμογής, τα λεγόμενα χ^2 τεστ, τα οποία εισήχθησαν για πρώτη φορά το 1900 από τον Βρετανό στατιστικό επιστήμονα Karl Pearson.

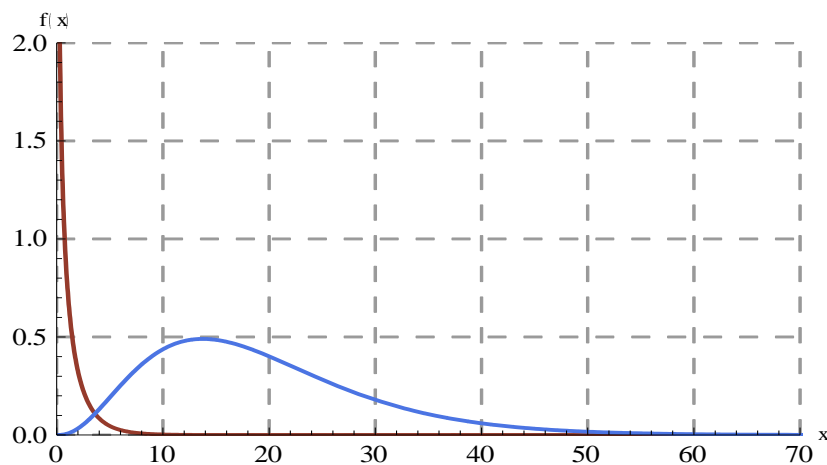
Τέλος σχετικά με τις συναρτήσεις Γάμμα ($\Gamma(x)$) τις οποίες αναφέραμε προηγουμένως και οι οποίες αποτελούν βασικό στοιχείο των κατανομών Γάμμα καθώς συμπεριλαμβάνονται μέσα στις συναρτήσεις πυκνότητας πιθανότητας τους, αξίζει να αναφερθεί ότι αρκετές φορές λόγω

δυσκολιών υπολογισμού γίνονται κάποιες προσεγγίσεις αυτών κυρίως μέσω της χρήσης διαφόρων πολυωνύμων.

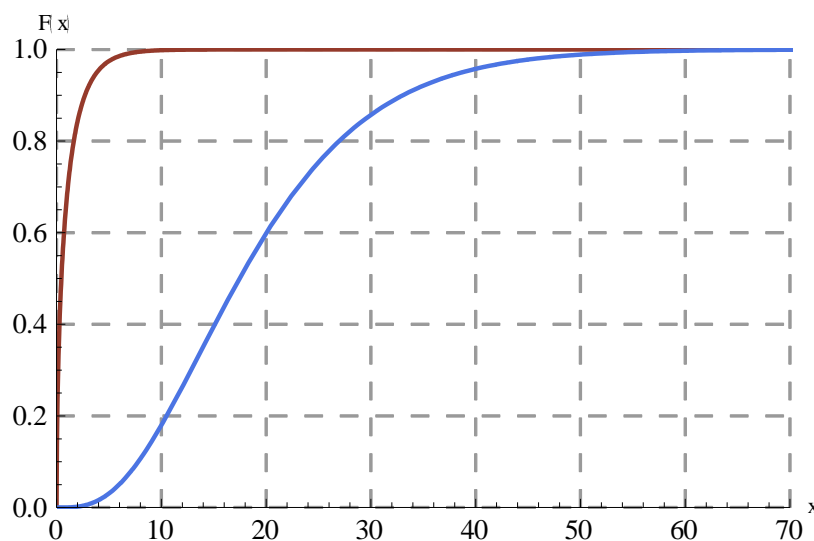
Διαγράμματα κατανομής Γάμμα

Στα παρακάτω Διαγράμματα 1.1 και 1.2 παρουσιάζονται οι συναρτήσεις πυκνότητας πιθανότητας και οι συναρτήσεις κατανομής για τις κατανομές Γάμμα με παραμέτρους $a = 0.5, b = 2$ και $a = 3.5, b = 5.5$.

Διάγραμμα 1.1 : Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της $Gamma(0.5,2)$ (κόκκινη καμπύλη) και της $Gamma(3.5,5.5)$ (μπλε καμπύλη)



Διάγραμμα 1.2 : Διάγραμμα της συνάρτησης κατανομής της $Gamma(0.5,2)$ (κόκκινη καμπύλη) και της $Gamma(3.5,5.5)$ (μπλε καμπύλη)



1.2 Η κατανομή Erlang

Μια εξίσου σημαντική κατανομή για την αναλογιστική επιστήμη και κυρίως την θεωρία κινδύνου είναι η κατανομή Erlang η οποία αποτελεί και αυτή μια ειδική περίπτωση των κατανομών Γάμμα. Προκύπτει όταν η παράμετρος σχήματος (shape parameter) της κατανομής Γάμμα είναι ένας θετικός ακέραιος αριθμός. Ταυτίζεται επίσης με την κατανομή του αθροίσματος n ανεξάρτητων και τυχαίων εκθετικών μεταβλητών όταν $n \in \mathbb{N}$. Δηλαδή αν οι τ.μ. $X_i \sim \text{Exp}(b)$ τότε η τ.μ. $Y = X_1 + X_2 + \dots + X_n$ θα ακολουθεί την $\text{Erl}(n, b)$ όπου n θετικός ακέραιος.

Η κατανομή Erlang οφείλει το όνομά της στον Δανό μαθηματικό και στατιστικό επιστήμονα Agner Krarup Erlang (1878-1929) ενώ παράλληλα έχει μεγάλη εφαρμογή στις στοχαστικές ανελίξεις, οι οποίες αποτελούν την βάση για την ανάπτυξη της Θεωρίας Ουρών Αναμονής, αλλά και σε διάφορους κλάδους των εφαρμοσμένων μαθηματικών.

Ορισμός 1.2: Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x^{n-1} \cdot e^{-x/b}}{b^n \cdot (n-1)!}, & x > 0 \end{cases}$$

όπου n θετικός ακέραιος και $0 < b < \infty$, τότε θα λέμε ότι η τ.μ. X ακολουθεί την κατανομή Erlang με παραμέτρους n, b .

Ο όρος $(n-1)!$ απαιτεί ακέραιες τιμές για την παράμετρο n , ωστόσο μπορούμε κάνοντας χρήση της συνάρτησης $\Gamma(x)$ (την οποία περιγράψαμε στην ενότητα 1.1) να συμπεριλάβουμε και τις τιμές εκείνες οι οποίες δεν είναι ακέραιοι αριθμοί αλλά σε αυτή τη περίπτωση θα έχουμε κατανομή Γάμμα και όχι Erlang.

Η αντίστοιχη συνάρτηση επιβίωσης (survival function): $S(x) = 1 - F(x)$ δίνεται από τον παρακάτω τύπο:

$$S(x) = \sum_{i=0}^{n-1} \frac{(x/b)^i}{i!} \cdot e^{-x/b}, \quad x \geq 0 \text{ και } n \in \mathbb{N}$$

ενώ η συνάρτηση κατανομής της τ.μ. X θα ισούται με:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{\Gamma(x/b, n)}{(n-1)!}, & 0 < x < \infty \end{cases}$$

όπου $\Gamma(x, a)$ είναι η μη πλήρης συνάρτηση Γάμμα όπως ορίστηκε στη σχέση (1.1.4)

ή εναλλακτικά: $F(x) = 1 - \sum_{i=0}^{n-1} \frac{(x/b)^i}{i!} \cdot e^{-x/b}, \quad x \geq 0 \text{ και } n \in \mathbb{N}$

Η συνάρτηση πυκνότητας πιθανότητας της τ.μ. X που ακολουθεί $Erl(n, b)$ είναι μη αρνητική ενώ παράλληλα ισχύει η εξής σχέση για το ολοκλήρωμα I_n :

$$I_n = \int_0^{\infty} y^{n-1} \cdot e^{-y} dy = (n-1)!, \quad n = 1, 2, \dots \quad (1.2.1)$$

Επιπλέον για να είναι η συνάρτηση f μια συνάρτηση πυκνότητας πιθανότητας θα πρέπει να ισχύει ότι: $\int_0^{\infty} f(x) dx = 1$ και συνεπώς έχουμε:

$$\frac{1}{b^n \cdot (n-1)!} \cdot \int_0^{\infty} x^{n-1} \cdot e^{-x/b} dx = \frac{1}{(n-1)!} \cdot \int_0^{\infty} y^{n-1} \cdot e^{-y} dy = 1 \quad (1.2.2)$$

(όπου θέτουμε $y = x/b$ και $dy = 1/b \cdot dx$)

Παρατηρούμε ότι η σχέση (1.2.1) ισχύει λόγω του παραπάνω ολοκληρώματος, άρα πρόκειται όντως για συνάρτηση πυκνότητας πιθανότητας.

Η μέση τιμή της τ.μ. $X \sim Erl(n, b)$ προκύπτει μέσω της σχέσης:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \frac{1}{b^n (n-1)!} \cdot \int_0^{\infty} x^n \cdot e^{-x/b} dx$$

θέτουμε $y = x/b$ και $dy = (1/b) \cdot dx$ και έχουμε:

$$E(X) = \frac{b}{(n-1)!} \cdot \int_0^{\infty} y^n \cdot e^{-y} dy$$

Με χρήση του ολοκληρώματος (1.2.1) ισχύει ότι:

$$E(X) = \frac{b \cdot n!}{(n-1)!} = n \cdot b$$

Με τον ίδιο τρόπο υπολογίζουμε την ροπή δεύτερης τάξης:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \frac{1}{b^n \cdot (n-1)!} \cdot \int_0^{\infty} x^{n+1} \cdot e^{-x/b} dx \\ &= \frac{b^2}{(n-1)!} \cdot \int_0^{\infty} y^{n+1} \cdot e^{-y} dy \end{aligned}$$

Άρα : $E(X^2) = \frac{b^2 \cdot (n+1)!}{(n-1)!} = b^2 \cdot n \cdot (n+1)$ και για επομένως για την διασπορά ισχύει:

$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = b^2 \cdot n \cdot (n+1) - n^2 \cdot b^2 = n \cdot b^2$ όπου n θετικός ακέραιος.

Στη συνέχεια ας θεωρήσουμε μια ανέλιξη Poisson X_t , $t \geq 0$ με μέση τιμή $E[X_t] = t/b$ η οποία εκφράζει τον αριθμό των εμφανίσεων ενός ενδεχομένου A στο διάστημα $[0,t]$ και συμβολίζουμε ως T_n το χρόνο αναμονής μέχρι να εμφανιστεί η n-οστή επιτυχία (πραγματοποίηση του ενδεχομένου A) (Χαραλαμπίδης, 2009).

Ισχύει ότι το ενδεχόμενο $\{T_n > t\}$ είναι ισοδύναμο με το ενδεχόμενο $\{X_t < n\}$ καθώς το ενδεχόμενο ο χρόνος αναμονής μέχρι να εμφανιστεί η ν-οστή επιτυχία να υπερβαίνει την χρονική στιγμή t ισούται με το ενδεχόμενο ο συνολικός αριθμός επιτυχιών στο $[0,t]$ να είναι μικρότερος από n.

Επομένως έχουμε:

$$P(T_n > t) = P(X_t < n) = \sum_{k=0}^{n-1} P(X_t = k) = \sum_{k=0}^{n-1} e^{-b/t} \cdot \frac{(t/b)^k}{k!}, \quad t \geq 0$$

Συνεπώς η συνάρτηση κατανομής για την τ.μ. T_n θα δίνεται από τον τύπο:

$$P(T_n \leq t) = F_{T_n}(t) = 1 - e^{-b/t} \cdot \sum_{k=0}^{n-1} \frac{(t/b)^k}{k!}, \quad t \geq 0$$

με $F_{T_n}(t) = 0$ για $t < 0$.

Αν στην συνέχεια παραγωγίσουμε ως προς t την συνάρτηση κατανομής της τ.μ. T_n θα προκύψει:

$$f_{T_n}(t) = \frac{d}{dt} F_{T_n}(t) = 1/b \cdot e^{-t/b} \sum_{k=0}^{n-1} \frac{(t/b)^k}{k!} - e^{-t/b} \sum_{k=0}^{n-1} \frac{1/b \cdot (t/b)^{k-1}}{(k-1)!}$$

Και επομένως η συνάρτηση πυκνότητας πιθανότητας της τ.μ. T_n είναι η:

$$f(t) = \frac{t^{n-1} \cdot e^{-t/b}}{b^n \cdot (n-1)!}, \quad 0 \leq t < \infty$$

Η κατανομή αυτή είναι ακριβώς ίδια με την κατανομή Erlang που μελετήθηκε για πρώτη φορά όπως αναφέραμε πριν από τον μαθηματικό A.K Erlang, ο οποίος ασχολήθηκε με διάφορα προβλήματα σχετιζόμενα με τηλεφωνικά δίκτυα όπως η διάρκεια των τηλεφωνικών κλήσεων αλλά και ο χρόνος αναμονής ώσπου να είναι ελεύθερη μια τηλεφωνική γραμμή. Συνεπώς παρατηρούμε ότι η χρήση της κατανομής αυτής είναι ιδιαίτερα σημαντική για τις στοχαστικές διαδικασίες και για την περιγραφή τέτοιων προβλημάτων.

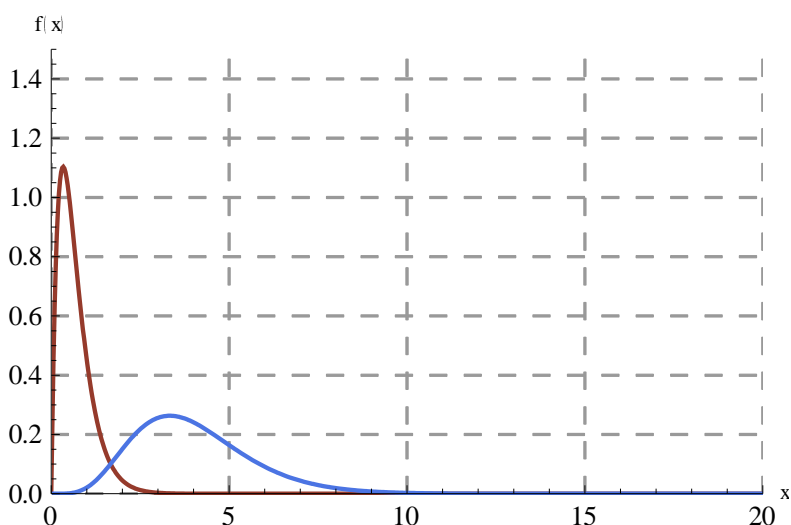
Τέλος μια σημαντική παρατήρηση είναι ότι αρκετές φορές η αθροιστική συνάρτηση της κατανομής Γάμμα είναι δύσκολο να υπολογιστεί με αποτέλεσμα η κατανομή Erlang να είναι

πιο εύκολη στη χρήση της καθώς η αθροιστική της συνάρτηση υπολογίζεται με διαδοχική ολοκλήρωση κατά παράγοντες.

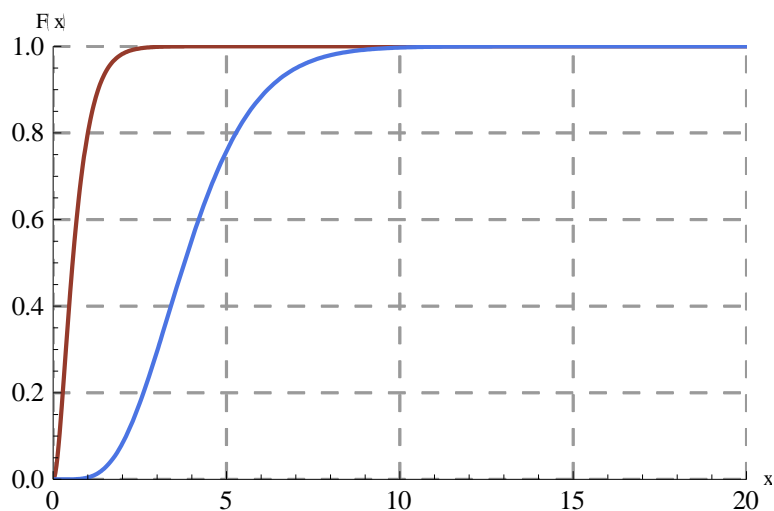
Διαγράμματα κατανομής Erlang

Στα παρακάτω Διαγράμματα 1.3 και 1.4 παρουσιάζονται οι συναρτήσεις πυκνότητας πιθανότητας και οι συναρτήσεις κατανομής για τις κατανομές Erlang με παραμέτρους $n = 2, b = 3$ και $n = 6, b = 1.5$.

Διάγραμμα 1.3: Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της $Erl(2,3)$ (κόκκινη καμπύλη) και της $Erl(6,1.5)$ (μπλε καμπύλη)



Διάγραμμα 1.4: Διάγραμμα της συνάρτησης κατανομής της $Erl(2,3)$ (κόκκινη καμπύλη) και της $Erl(6,1.5)$ (μπλε καμπύλη)



1.3 Χρησιμότητα των μείξεων και συνελίξεων Γάμμα στον αναλογισμό

Στις προηγούμενες δύο ενότητες κάναμε μια εισαγωγή στην κατανομή Γάμμα αλλά και στην κατανομή Erlang, η οποία αποτελεί μια ειδική υποπερίπτωση αυτών. Είναι ιδιαίτερα σημαντική η κατανόηση βασικών εννοιών σχετικά με την κατανομή Γάμμα διότι η κατανομή αυτή θα αποτελέσει τον πυρήνα της παρούσας διπλωματικής εργασίας καθώς θα μελετήσουμε ιδιότητες και προσεγγίσεις για μείξεις αλλά και συνελίξεις των συγκεκριμένων κατανομών. Ήδη αναφέραμε την ζωτική σημασία της κατανομής Γάμμα στον αναλογισμό και την ευρεία χρήση της στη θεωρία κινδύνου, συνεπώς ακόμη πιο αξιοσημείωτη είναι η χρήση των μείξεων και των συνελίξεων για αυτές τις κατανομές.

Στην συνέχεια της εργασίας θα ορίσουμε αναλυτικά με μαθηματικούς όρους τι σημαίνει μείξη αλλά και συνέλιξη κατανομών (Κεφάλαιο 2: Ορισμός 2.2 και Κεφάλαιο 4: Ορισμός 4.5.1).

Α) Σχετικά με την οικογένεια των μείξεων ή μεμιγμένων κατανομών όπως θα δούμε και στα επόμενα κεφάλαια προκύπτουν οι εξής δύο περιπτώσεις:

- Διακριτή περίπτωση: Στην περίπτωση αυτή η κατανομή της μείξης είναι ένα διακριτό σταθμισμένο άθροισμα κατανομών.
- Συνεχής περίπτωση: Στην περίπτωση αυτή η κατανομή της μείξης προκύπτει ως ένα αποτέλεσμα ολοκληρωμάτων.

Και τα δύο είδη μείξης κατανομών χρησιμοποιούνται στην αναλογιστική επιστήμη ενώ στην θεωρία χρεοκοπίας αρκετά συχνές είναι οι διακριτές μείξεις ως μοντέλο για τα μεγέθη των αποζημιώσεων ενός ασφαλιστικού χαρτοφυλακίου. Οι συνεχείς μείξεις αντίστοιχα εμφανίστηκαν προς τα τέλη του 1940 μέσω του μοντέλου του Ammeter και από τότε έχουν πολλές εφαρμογές σε διάφορους κλάδους της θεωρίας κινδύνου όπως είναι η θεωρία αξιοπιστίας. Το πιο ιδιαίτερο χαρακτηριστικό των μεμιγμένων κατανομών είναι κυρίως το γεγονός ότι αποτελούν μια ικανοποιητική προσέγγιση για χαρτοφυλάκια τα οποία παρουσιάζουν «ανομοιογένεια».

Ας θεωρήσουμε ένα χαρτοφυλάκιο ασφάλισης αυτοκινήτων για το οποίο μπορούμε εύκολα να διαπιστώσουμε ότι η πιθανότητα εμφάνισης μιας ζημιάς από κάποιο ατύχημα αλλά και η κατανομή του ύψους της απαίτησης που θα δημιουργηθεί για την ασφαλιστική εταιρεία, διαφέρει από οδηγό σε οδηγό. Μπορεί να εξαρτάται είτε από την ηλικία, είτε από τα συνολικά έτη που οδηγεί, είτε από το φύλλο ή ακόμη και από το εάν βρισκόταν υπό την επήρεια αλκοόλ. Επομένως θα πρέπει να θεωρήσουμε ότι το χαρτοφυλάκιο χωρίζεται σε ξεχωριστά τμήματα, καθένα από τα οποία θα παρουσιάζει ομοιογένεια ως τον αντίστοιχο κίνδυνο. Αν λοιπόν υποθέσουμε ότι η κατανομή του κινδύνου μέσα σε κάθε επιμέρους τμήμα του ασφαλιστικού χαρτοφυλακίου είναι της ίδιας μορφής (π.χ. κατανομή Γάμμα), οι παράμετροι της κατανομής αυτής θα είναι διαφορετικές από τμήμα σε τμήμα. Συνεπώς σε αυτή τη περίπτωση, ένας οδηγός τον οποίο επιλέγουμε με τυχαίο τρόπο από το χαρτοφυλάκιο θα έχει μια **μεμιγμένη** κατανομή κινδύνου (Πολίτης, 2012).

Επιπλέον ένα χαρτοφυλάκιο μιας ασφαλιστικής εταιρείας μπορεί να εμφανίζει ανομοιογένεια και ως προς την μονάδα του χρόνου. Υπάρχουν πολλές περιπτώσεις όπου μπορούμε να θεωρήσουμε ότι η συχνότητα εμφάνισης ζημιών (και συνεπώς απαιτήσεων) μεταβάλλεται από την μία ασφαλιστική περίοδο στην άλλη, καθώς ο χρόνος αποτελεί έναν σημαντικό παράγοντα. Αυτό έχει ως αποτέλεσμα να είναι εξίσου αναγκαία η χρήση μιας μεμιγμένης κατανομής, με την μείξη να γίνεται ως προς τον χρόνο αυτή τη φορά, για την μελέτη του αριθμού των απαιτήσεων στο συνολικό χρονικό διάστημα που μας ενδιαφέρει.

Μια επιπλέον αξιοσημείωτη εφαρμογή των μείξεων κατανομών προκύπτει μέσω της παρακάτω Πρότασης 1.3.1 (Πολίτης, 2012):

Πρόταση 1.3.1: Θεωρούμε τις ανεξάρτητες και τυχαίες μεταβλητές S_1, S_2, \dots, S_n , έτσι ώστε η κατανομή της S_i να είναι μια σύνθετη κατανομή Poisson με παράμετρο λ_i και οι αντίστοιχες ατομικές αποζημιώσεις να έχουν συνάρτηση κατανομής $F_i(x)$, για $i = 1, 2, 3 \dots$

Τότε η κατανομή του αθροίσματος $S = S_1 + S_2 + \dots + S_n$ θα ακολουθεί επίσης την σύνθετη κατανομή Poisson με παράμετρο

$$\lambda = \sum_{i=1}^n \lambda_i$$

Και συνάρτηση κατανομής:

$$F(x) = \sum_{i=1}^n \frac{\lambda_i}{\lambda} \cdot F_i(x)$$

Μιας και έγινε αναφορά στην σύνθετη Poisson κατανομή αξίζει να επισημάνουμε ότι γενικά μια τυχαία μεταβλητή ακολουθεί μια σύνθετη κατανομή όταν προκύπτει ως άθροισμα ανεξάρτητων και ισόνομων μεταβλητών των οποίων το πλήθος είναι τυχαίος αριθμός. Είναι γνωστόν ότι οι σύνθετες κατανομές έχουν κεντρικό ρόλο στο συλλογικό πρότυπο της θεωρίας κινδύνων αφού οι συνολικές αποζημιώσεις ακολουθούν τέτοιου είδους κατανομές.

Σχετικά με το αποτέλεσμα της παραπάνω Πρότασης (Πολίτης, 2012), ισχύει ότι εάν έχουμε n ανεξάρτητα χαρτοφυλάκια και για $1 \leq i \leq n$, ο συνολικός κίνδυνος στο χαρτοφυλάκιο i ακολουθεί μια σύνθετη κατανομή Poisson με συνάρτηση κατανομής των ατομικών κινδύνων ίση με $F_i(x)$, τότε ο αθροιστικός κίνδυνος από όλα τα επιμέρους χαρτοφυλάκια θα ακολουθεί επίσης μια σύνθετη κατανομή Poisson στην οποία η συνάρτηση κατανομής των ατομικών κινδύνων θα προκύπτει ως μια *διακριτή μείξη* των $F_i(x)$.

Γενικό συμπέρασμα: Όπως είδαμε και παραπάνω οι μείξεις κατανομών γενικότερα αποτελούν ένα αναπόσπαστο κομμάτι της αναλογιστικής επιστήμης. Ιδιαίτερα λοιπόν όταν πρόκειται για μείξεις των κατανομών Γάμμα οι οποίες κατανομές Γάμμα έχουν μελετηθεί εκτενώς, είναι μη αρνητικές και έχουν θετική ασυμμετρία τότε κατανοούμε ότι τέτοιου είδους μείξεις θα έχουν μια μεγάλη πληθώρα εφαρμογών στον αναλογισμό.

B) Σχετικά με τις συνελίξεις κατανομών, είναι ιδιαίτερα γνωστόν ότι η συνέλιξη αποτελεί μια πράξη ιδιαίτερα σημαντική γιατί σχετίζεται με την ανάλυση συστημάτων αλλά και με το γεγονός ότι μετατρέπεται σε γινόμενο όταν αλλάζουμε χώρους (από τον χρόνο στην συχνότητα και αντίστροφα). Η συνέλιξη λόγω του ότι εμπλέκει τον υπολογισμό ενός ολοκληρώματος (ή ενός αθροίσματος όταν έχουμε διακριτό χρόνο) έχει αρκετή δυσκολία ως προς την μελέτη της. Η δυσκολία αυτή έγκειται στο γεγονός ότι στην πράξη εμπεριέχεται το γινόμενο δύο συναρτήσεων, εκ των οποίων η μία έχει υποστεί ανάκλαση και μετατόπιση. Η αντίστοιχη ακολουθία μετατροπής (ανάκλασης και μετατόπισης) έχει ως εξής:

$$y(t) \rightarrow y(\tau) \rightarrow y(-\tau) \rightarrow y(-\tau + t) = y(t - \tau)$$

(πηγή: Wikipedia)

Η συνέλιξη επομένως είναι μια πράξη η οποία εφαρμόζεται σε δύο ή και περισσότερες συναρτήσεις και μπορεί να είναι (για τον ορισμό συνεχούς και διακριτής συνάρτησης βλ. Πολίτης, 2012) :

- Συνέλιξη συνεχών συναρτήσεων ή
- Συνέλιξη διακριτών συναρτήσεων

Επιπλέον η συνέλιξη έχει τις ίδιες ιδιότητες με τον πολλαπλασιασμό. Δηλαδή είναι αντιμεταθετική, προσεταιριστική, επιμεριστική ως προς την πρόσθεση και υπάρχει ένα ουδέτερο στοιχείο η γενικευμένη συνάρτηση $\delta(x)$ (συνάρτηση του Dirak), η οποία μηδενίζεται για κάθε x εκτός από το σημείο μηδέν όπου τείνει στο θετικό άπειρο.

Στα αναλογιστικά μαθηματικά είναι ιδιαίτερα συχνή η μελέτη προβλημάτων που αφορούν αθροίσματα τυχαίων μεταβλητών. Είδαμε και προηγουμένως ότι οι σύνθετες κατανομές αποτελούν μια βασική έννοια η οποία συνδέεται με το άθροισμα ανεξάρτητων και ισόνομων τυχαίων μεταβλητών όταν όμως το πλήθος των προσθετέων δεν είναι σταθερό αλλά μια τυχαία μεταβλητή. Όπως για παράδειγμα το πλήθος των απαιτήσεων που φτάνουν σε μια ασφαλιστική εταιρεία εντός ενός συγκεκριμένου χρονικού διαστήματος χωρίς να γνωρίζουμε τον ακριβή αριθμό του. Ωστόσο υπάρχουν πολλές περιπτώσεις όπου μπορεί να έχουμε την μελέτη αθροισμάτων της μορφής:

$$Y = Y_1 + Y_2 + \dots + Y_n$$

Όπου το n θα είναι ένας γνωστός ακέραιος αριθμός και οι Y_i ανεξάρτητες τ.μ.

Συνεπώς η έννοια της **συνέλιξης** σε αντίθεση με την έννοια της **σύνθετης** τυχαίας μεταβλητής χρησιμεύει για την μελέτη τέτοιου είδους αθροισμάτων στην αναλογιστική επιστήμη.

Έχοντας λοιπόν κατανοήσει την βαθύτατη σημασία των μείξεων αλλά και των συνελίξεων κυρίως των κατανομών Γάμμα θα προσπαθήσουμε να τις αναλύσουμε διεξοδικώς και να δούμε αποτελέσματα σχετικά με την μορφή τους. Στην συνέχεια της παρούσας διπλωματικής εργασίας θα μελετήσουμε έννοιες όπως η βαθμίδα αποτυχίας και ο μέσος υπολειπόμενος χρόνος ζωής, θα διεξάγουμε συμπεράσματα σχετικά με μία πολύ σημαντική ιδιότητα αυτών των κατανομών, την ασυμμετρία ενώ παράλληλα θα παρουσιάσουμε προσεγγίσεις για την συνάρτηση πυκνότητας των συνελίξεων και εφαρμογές στη θεωρία κινδύνου.

ΚΕΦΑΛΑΙΟ 2

ΒΑΘΜΙΔΑ ΑΠΟΤΥΧΙΑΣ (FAILURE RATE)

Η συνάρτηση της βαθμίδας αποτυχίας αποτελεί μια από τις βασικότερες ιδιότητες την οποία θα αναλύσουμε και θα μελετήσουμε με διεξοδικό τρόπο στην παρούσα διατριβή, δίνοντας έμφαση σε διάφορες κατανομές αντίστοιχα. Η συνάρτηση αυτή εμφανίζεται ιδιαίτερα στο κλάδο της αναλογιστικής επιστήμης και συγκεκριμένα στην ανάλυση επιβίωσης, όπου καλείται συνάρτηση κινδύνου καθώς επίσης και στα Συμβάντα Ζωής και Θανάτου (Ασφαλίσεις Ζωής), όπου καλείται ένταση θνησιμότητας ή *force of mortality*. Στα μοντέλα επιβίωσης (*survival models*) ένα πολύ σημαντικό ρόλο παίζει η τυχαία μεταβλητή X η οποία περιγράφει την διάρκεια ζωής ενός νεογέννητου ατόμου ($X \geq 0$) και είναι συνεχής διότι μετράει χρόνο ζωής. Επιπλέον παίρνει τιμές σε ένα κλειστό διάστημα $[0, \omega]$ όπου ω ορίζεται ως η μέγιστη ή οριακή ηλικία (ή χρονική στιγμή) μέχρι την οποία μπορεί να επιβιώσει ένα ανθρώπινο ον ή ένα μηχανήμα. Σε πολλές επιστήμες, οι συνεχείς αυτές τυχαίες μεταβλητές καλούνται χρόνοι ζωής (ή *lifetimes*) και χαρακτηρίζονται από την συνάρτηση κατανομής αλλά και από την συνάρτηση πυκνότητας πιθανότητάς τους.

Ωστόσο αρκετές φορές ο χρόνος ζωής μπορεί να διακοπεί με απροσδόκητο τρόπο λόγω μιας αποτυχίας, θανάτου ή οποιουδήποτε άλλου γεγονότος. Για αυτό το λόγο ιδιαίτερη σημασία έχει η συνάρτηση της βαθμίδας αποτυχίας η οποία σαν ποσότητα εκφράζει την πιθανότητα αποτυχίας ύστερα από ένα σημείο εκκίνησης (γέννηση ή έναρξη λειτουργίας στο κλάδο της μηχανικής) έως τη χρονική στιγμή που δύναται να επιβιώσει ένα ανθρώπινο ον (ή ένα σύστημα). Στα συμβάντα ζωής και θανάτου ή ασφαλίσεις ζωής, όπου θεωρείται ως ένταση θνησιμότητας εκφράζει τον στιγμιαίο ρυθμό θανάτου. Ωστόσο, παρ' όλο που αρκετές φορές θεωρείται ως η πιθανότητα μια αποτυχία να εμφανιστεί σε μια συγκεκριμένη χρονική στιγμή t , δεδομένου ότι προηγουμένως δεν έχει υπάρξει άλλη αποτυχία, στην ουσία δεν αποτελεί πιθανότητα καθώς μπορεί να πάρει τιμές μεγαλύτερες από την μονάδα. Γενικά, η ένταση θνησιμότητας ενός ανθρώπινου όντος ή η βαθμίδα αποτυχίας ενός μηχανικού εξαρτήματος δεν αυξάνεται πάντοτε με εκθετικό ρυθμό καθώς αυξάνεται ο χρόνος, γι' αυτό θα παρατηρήσουμε για διάφορες κατανομές ποια μπορεί να είναι η μορφή η οποία δύναται να πάρει η συγκεκριμένη συνάρτηση.

Συνεπώς είναι ιδιαίτερα φανερό ότι η κατανόηση αλλά και ο προσδιορισμός της συνάρτησης της βαθμίδας αποτυχίας αποτελεί βασικό στοιχείο για πολλούς επιστημονικούς κλάδους. Στο παρόν κεφάλαιο της δοθείσας εργασίας θα προσπαθήσουμε με απλό αλλά ταυτόχρονα τεκμηριωμένο τρόπο να ορίσουμε μαθηματικά την συγκεκριμένη ποσότητα αλλά παράλληλα και να ερμηνεύσουμε την σημασία της στον τομέα της Θεωρίας κινδύνων και της αναλογιστικής επιστήμης γενικότερα. Στη συνέχεια θα δούμε ορισμένα παραδείγματα για συνεχείς κατανομές και πώς ορίζεται η βαθμίδα αποτυχίας για αυτές αντίστοιχα.

Ορισμός 2.1: Έστω η θετική τυχαία μεταβλητή X η οποία εκφράζει την διάρκεια ζωής ενός ατόμου ή τον χρόνο μέχρι την αποτυχία ενός μηχανήματος, με αθροιστική συνάρτηση κατανομής:

$$F(x) = \begin{cases} \Pr(X \leq x), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

και συνάρτηση πυκνότητας πιθανότητας $f(x)$ αντίστοιχα. Επιπλέον ως δεξιά ουρά της συνάρτησης κατανομής ορίζεται η ποσότητα $\bar{F}(x) = S(x) = 1 - F(x)$ γνωστή ως συνάρτηση επιβίωσης της τ.μ. X (εκφράζει την πιθανότητα μη αποτυχίας μέχρι τη συγκεκριμένη χρονική στιγμή) και η οποία είναι ίση με:

$$\bar{F}(x) = \int_x^\infty f(t) dt = 1 - \int_0^x f(t) dt, \quad \text{ενώ παράλληλα ισχύει ότι } f(x) = \frac{dF(x)}{dx}.$$

Τότε ως βαθμίδα αποτυχίας ή failure rate ή ένταση θνησιμότητας ορίζεται:

$$\mu(x) = \frac{f(x)}{F(x)} = \frac{f(x)}{1-F(x)} = -\frac{d}{dx} \ln(\bar{F}(x)), \quad x \geq 0$$

Παρατηρούμε την εξής σχέση ανάμεσα στη βαθμίδα αποτυχίας (failure rate) και την συνάρτηση επιβίωσης:

Η ένταση θνησιμότητας (ασφαλίσεις ζωής) ή βαθμίδα αποτυχίας ορίζεται ως εξής:

$$\begin{aligned} \mu(x) &= \lim_{dx \rightarrow 0} \frac{\Pr(x < X < x+dx | X > x)}{dx} = \lim_{dx \rightarrow 0} \frac{F(x+dx) - F(x)}{dx \cdot \bar{F}(x)} = \lim_{dx \rightarrow 0} \frac{1}{dx} \cdot \left(1 - \frac{\bar{F}(x+dx)}{\bar{F}(x)}\right) = \\ &= \lim_{dx \rightarrow 0} \frac{1}{dx} \cdot (1 - \Pr(X > x + dx | X > x)) \end{aligned} \quad (2.0)$$

$$\text{ή} \quad \mu(x) = \lim_{dx \rightarrow 0} \frac{S(x) - S(x+dx)}{S(x) \cdot dx} = -\frac{1}{S(x)} \cdot \lim_{dx \rightarrow 0} \frac{S(x+dx) - S(x)}{dx} = -\frac{1}{S(x)} \cdot \lim_{dx \rightarrow 0} \frac{S(x+dx) - S(x)}{(x+dx) - x}$$

$$\text{Συνεπώς } \mu(x) = -\frac{S'(x)}{S(x)}$$

Από την τελευταία σχέση μπορούμε εύκολα να διακρίνουμε ότι η βαθμίδα αποτυχίας και η συνάρτηση επιβίωσης είναι δύο μεγέθη τα οποία συσχετίζονται με αρνητικό τρόπο, δηλαδή για αρκετά μεγάλες τιμές της $S(x)$, η $\mu(x)$ παίρνει μικρές τιμές και αντίστροφα για μικρές τιμές της συνάρτησης επιβίωσης, η βαθμίδα αποτυχίας μεγαλώνει. Αυτό είναι ιδιαίτερα προφανές καθώς όσο μειώνεται η πιθανότητα επιβίωσης τόσο πιο μεγάλος είναι ο ρυθμός αποτυχίας. Παράλληλα προκύπτει ότι μικρές τιμές της βαθμίδας αποτυχίας είναι ένα σημάδι για βαριά δεξιά ουρά ενώ αντίστοιχα μεγάλες τιμές της $\mu(x)$ συσχετίζονται με πιο ελαφριά δεξιά ουρά.

Παράλληλα μια πιο ισχυρή σύνδεση μεταξύ της βαθμίδας αποτυχίας ή έντασης θνησιμότητας και της κατανομής της τ.μ. X , η οποία σύνδεση υποδεικνύει ότι η $\mu(x)$ καθορίζει με έναν μοναδικό τρόπο την συνάρτηση κατανομής της τ.μ. X είναι:

$$\mu(x) = \frac{-S'(x)}{S(x)}$$

Ολοκληρώνοντας την παραπάνω σχέση από 0 έως x έχουμε ότι:

$$\int_0^x \mu(t)dt = - \int_0^x \frac{S'(t)}{S(t)} dt = - \int_0^x (\ln S(t))' dt = - \{\ln S(x) - \ln S(0)\}$$

$$\Rightarrow \ln S(x) = - \int_0^x \mu(t)dt .$$

Άρα $\bar{F}(x) = S(x) = e^{-\int_0^x \mu(t)dt}$.

Έστω τώρα ότι θεωρούμε το ενδεχόμενο :

$$A = \{\text{άτομο ηλικίας } x \text{ πεθαίνει μεταξύ των ηλικιών } x \text{ και } x + dx\}$$

Έχουμε ότι $A = \{x < X < x + dx | X > x\}$ και όταν $dx \rightarrow 0$ τότε το άτομο ηλικίας x πεθαίνει πλησίον της ηλικίας x .

Όμως $\Pr(A) = \frac{\Pr(x < X < x + dx)}{\Pr(X > x)}$, οπότε για $dx \rightarrow 0$ ισχύει ότι:

$$\Pr(A) \cong \frac{f(x)}{S(x)} dx \text{ και συνεπώς } \mu(x) \cdot dx \cong \Pr(A) .$$

Δηλαδή η ποσότητα $\mu(x) \cdot dx$ εκφράζει την πιθανότητα το άτομο ηλικίας x να πεθάνει μέσα στο διάστημα $(x, x + dx)$ ή γενικότερα την δεσμευμένη πιθανότητα αποτυχίας στο διάστημα αυτό, δοθέντος ότι έχει επιβιώσει μέχρι την ηλικία x .

Στη συνέχεια θα εξετάσουμε την σχέση της μέσης τιμής της τ.μ. X (η οποία αποτελεί μια βασική ποσότητα της στατιστικής ανάλυσης) με την βαθμίδα αποτυχίας.

Γνωρίζουμε ότι η μέση τιμή της τ.μ. X ορίζεται ως εξής:

$$E(X) = \int_0^\infty x \cdot f(x)dx = \int_0^\infty x \cdot dF(x)$$

θεωρώντας ότι αυτή η ποσότητα είναι πεπερασμένη.

Ισχύει ότι:

$$E(X) = \int_0^\infty -x \cdot d(1 - F(x)) = -[x \cdot (1 - F(x))] + \int_0^\infty (1 - F(x))dx$$

Όπου έχουμε $\lim_{x \rightarrow 0} x \cdot \bar{F}(x) = 0$ και $\lim_{x \rightarrow \infty} x \cdot \bar{F}(x) = 0$ καθώς η συνάρτηση επιβίωσης είναι φθίνουσα και όσο το x πλησιάζει στο άπειρο, τόσο η συνάρτηση επιβίωσης τείνει στο μηδέν.

Συνεπώς προκύπτει ότι:

$$E(X) = \int_0^\infty \bar{F}(x)dx = \int_0^\infty \frac{f(x)}{\mu(x)} dx$$

Το οποίο μπορεί εναλλακτικά να γραφθεί ως: $E(X) = E\left(\frac{1}{\mu(x)}\right)$

Σύμφωνα με την τελευταία σχέση που αποδείξαμε, προκύπτει ότι όταν η βαθμίδα αποτυχίας $\mu(x)$ παίρνει ιδιαίτερα μικρές τιμές τότε η μέση τιμή της τ.μ. X παίρνει αντίστοιχα μεγάλες τιμές και αντιστρόφως. Μέσω αυτής της σχέσης διαπιστώνουμε και την συσχέτιση μεταξύ δεξιάς ουράς και βαθμίδας αποτυχίας η οποία όπως αποδείξαμε προηγουμένως είναι αρνητική.

Σύμφωνα με τους Willmot & Lin (2001), η συνάρτηση κατανομής $F(x)$ έχει φθίνουσα βαθμίδα αποτυχίας ή DFR εάν ο λόγος $\frac{\bar{F}(x+y)}{\bar{F}(x)}$ είναι μια μη φθίνουσα συνάρτηση ως προς x και για σταθερό $y \geq 0$ και αν ο λογάριθμος $\log \bar{F}(x)$ είναι κυρτή συνάρτηση. Αντιστοίχως, η συνάρτηση κατανομής $F(x)$ έχει αύξουσα βαθμίδα αποτυχίας ή IFR εάν ο λόγος $\frac{\bar{F}(x+y)}{\bar{F}(x)}$ είναι μια μη αύξουσα συνάρτηση ως προς x και για σταθερό $y \geq 0$ και αν ο λογάριθμος $\log \bar{F}(x)$ είναι κοίλη συνάρτηση. Επιπλέον όπως παρατηρούμε από την σχέση (2.0), αν η συνάρτηση $F(x)$ είναι απόλυτα συνεχής τότε η φθίνουσα βαθμίδα αποτυχίας είναι ισοδύναμη με την $\mu(x)$ η οποία είναι φθίνουσα συνάρτηση ως προς x ενώ η αύξουσα βαθμίδα αποτυχίας είναι ισοδύναμη με την $\mu(x)$ η οποία είναι αύξουσα συνάρτηση ως προς x .

2.1 Βαθμίδες αποτυχίας για διάφορες περιπτώσεις κατανομών

2.1.1 Βαθμίδα αποτυχίας για την απλή μορφή της εκθετικής κατανομής

Έστω η θετική συνεχής τ.μ. X η οποία περιγράφει τον χρόνο ζωής ενός νεογέννητου ατόμου ή μιας μονάδας γενικότερα. Θεωρούμε ότι $X \sim \text{Exp}(\lambda)$, δηλαδή ότι ακολουθεί την εκθετική κατανομή με παράμετρο λ και έχει συνάρτηση κατανομής $F(x)$. Η συνάρτηση πυκνότητας πιθανότητας είναι ίση με $f(x) = \lambda \cdot e^{-\lambda \cdot x}$, $\lambda > 0$ και $x \geq 0$ και συνεπώς η συνάρτηση κατανομής αντίστοιχα θα είναι $F(x) = \int_0^x f(y) dy = 1 - e^{-\lambda \cdot x}$.

Αντίστοιχα για την συνάρτηση επιβίωσης ή δεξιά ουρά, έχουμε ότι:

$$S(x) = \bar{F}(x) = 1 - F(x) = e^{-\lambda \cdot x}, \text{ με } \lambda > 0 \text{ και } x \geq 0$$

Με βάση τον ορισμό 2.1 για την βαθμίδα αποτυχίας προκύπτει:

$$\mu(x) = \frac{f(x)}{S(x)} = \frac{\lambda \cdot e^{-\lambda \cdot x}}{e^{-\lambda \cdot x}} = \lambda \text{ όπου } \lambda \text{ μια σταθερά μεγαλύτερη του μηδενός.}$$

Μπορούμε μέσω της τελευταίας σχέσης να διαπιστώσουμε ότι η βαθμίδα αποτυχίας για μια εκθετική κατανομή είναι σταθερή και ίση με την παράμετρο λ και είναι η μοναδική κατανομή για την μοντελοποίηση των χρόνων ζωής η οποία εμφανίζει σταθερή βαθμίδα αποτυχίας. Παράλληλα αυτό σημαίνει όπως αναφέραμε και προηγουμένως ότι μόνο για την ειδική περίπτωση της εκθετικής, η συνάρτηση κατανομής $F(x)$ μπορεί να ανήκει ταυτόχρονα και στην κλάση DFR αλλά και στην κλάση IFR.

Αυτό οφείλεται σε μια σπουδαία ιδιότητα των εκθετικών κατανομών, την λεγόμενη “ιδιότητα έλλειψης μνήμης” σύμφωνα με την οποία ισχύει:

$$\Pr(X > t + x | X > x) = \frac{\Pr(X > t+x)}{\Pr(X > x)} = \frac{S(x+t)}{S(x)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda x}} = e^{-\lambda \cdot t} = \Pr(X > t)$$

Ωστόσο όταν αναφερόμαστε στην βαθμίδα αποτυχίας ως ένταση θνησιμότητας για τις Ασφαλίσεις Ζωής, η ιδιότητα αυτή υποδηλώνει ότι ο θάνατος είναι ανεξάρτητος από την ηλικία στην οποία βρίσκεται ένα άτομο, γεγονός το οποίο είναι μη ρεαλιστικό καθώς όσο πιο μεγάλο είναι ένα άτομο ηλικιακά τόσο μεγαλύτερη έκθεση στον θάνατο έχει.

2.1.2 Βαθμίδα αποτυχίας για την περίπτωση Γάμμα Κατανομής

Έστω τώρα ότι η τ.μ. X που περιγράφει τον χρόνο ζωής ενός ατόμου ή ενός συστήματος, ακολουθεί την κατανομή Γάμμα με παράμετρο μορφής a και παράμετρο κλίμακας b , για $a, b > 0$ ενώ παράλληλα έχει συνάρτηση πυκνότητας πιθανότητας ίση με:

$$f(x) = \frac{b^a \cdot x^{a-1} \cdot e^{-b \cdot x}}{\Gamma(a)}, \quad x \geq 0 \text{ και } a, b > 0$$

όπου $\Gamma(a)$ για $a > 0$ είναι η συνάρτηση Γάμμα η οποία ορίζεται μέσω του παρακάτω ολοκληρώματος: $\Gamma(a) = \int_0^\infty t^{a-1} \cdot e^{-t} dt$.

Γενικά ισχύει ότι η συνάρτηση κατανομής $F(x)$ και αντίστοιχα και η συνάρτηση επιβίωσης $S(x)=1-F(x)$ δεν μπορούν εύκολα να εκφραστούν μέσω μιας απλής αναλυτικής μορφής αλλά όταν η παράμετρος μορφής a είναι θετικός ακέραιος αριθμός ($a=1,2,3 \dots$) δηλαδή έχουμε την λεγόμενη Erlang κατανομή, με συνεχείς ολοκληρώσεις ανά μέρη και για $x > 0$ προκύπτει η συνάρτηση κατανομής:

$$F(x) = 1 - e^{-b \cdot x} \sum_{j=0}^{a-1} (b \cdot x)^j / j! \quad \text{Για } a = 1, 2, 3, \dots \quad (2.1)$$

Επομένως κάνοντας αντικατάσταση στο τύπο που ισχύει για την βαθμίδα αποτυχίας στην περίπτωση της Γάμμα Κατανομής με $b > 0$ και a να ανήκει στο σύνολο των ακεραίων, έχουμε:

$$\mu(x) = \frac{f(x)}{1 - F(x)} = \frac{b^a \cdot x^{a-1}}{\Gamma(a) \cdot \sum_{j=0}^{a-1} (b \cdot x)^j / j!}, \quad \text{με } b > 0 \text{ και } a = 1, 2, 3, \dots$$

Ωστόσο μπορούμε εύκολα να μελετήσουμε την ασυμπτωτική συμπεριφορά της βαθμίδας αποτυχίας καθώς το x τείνει στο άπειρο, εφαρμόζοντας κανόνα De L'Hospital:

$$\mu(\infty) = \lim_{x \rightarrow \infty} \mu(x) = \lim_{x \rightarrow \infty} \frac{f(x)}{\bar{F}(x)} = - \lim_{x \rightarrow \infty} \frac{f'(x)}{f(x)} \quad (2.2)$$

Επιπλέον για την πρώτη παράγωγο της συνάρτησης πυκνότητας προκύπτει ότι:

$$f'(x) = \frac{b^a}{\Gamma(a)} \cdot \{(a-1) \cdot x^{a-2} - b \cdot x^{a-1}\} \cdot e^{-b \cdot x} \quad (2.3)$$

Άρα αντικαθιστώντας την σχέση (2.3) στην (2.2) έχουμε:

$$\mu(\infty) = \lim_{x \rightarrow \infty} \frac{b \cdot x^{a-1} - (a-1) \cdot x^{a-2}}{x^{a-1}} = b - \lim_{x \rightarrow \infty} \frac{a-1}{x} = b$$

Συνεπώς η βαθμίδα αποτυχίας έχει όριο όταν το x τείνει στο άπειρο ($\lim_{x \rightarrow \infty} \mu(x) = b$) ίσο με την παράμετρο κλίμακας b .

Επιπλέον γνωρίζουμε γενικά ότι ισχύει η εξής σχέση:

$$\frac{1}{\mu(x)} = \frac{S(x)}{f(x)} = \int_x^{\infty} \frac{f(y)}{f(x)} dy = \int_0^{\infty} \frac{f(x+y)}{f(x)} dy$$

Άρα θα μπορούσαμε μελετώντας τον λόγο $\frac{f(x+y)}{f(x)}$, να αποκτήσουμε μια εικόνα σχετικά με την μονοτονία αλλά και την κλάση κατανομής (IFR ή DFR) της έντασης θνησιμότητας ή βαθμίδας αποτυχίας.

Προκύπτει:

$$\frac{f(x+y)}{f(x)} = \left(\frac{y+x}{x}\right)^{a-1} \cdot e^{-b \cdot y}$$

Συνεπώς για $a \geq 1$ ο λόγος είναι φθίνουσα συνάρτηση ως προς x και για σταθερό $y \geq 0$ αντίστοιχα ενώ για $0 < a \leq 1$ είναι αύξουσα ως προς x και για σταθερό $y \geq 0$.

Ωστόσο στην περίπτωση όπου η παράμετρος a πάρει την τιμή 1 τότε η βαθμίδα αποτυχίας θα είναι σταθερή και ίση με b και θα ανήκει τόσο στην κλάση κατανομών IFR όσο και στην κλάση DFR, γεγονός το οποίο είναι αναμενόμενο καθώς γνωρίζουμε από τις ζημιοκατανομές ότι εάν $a = 1$ τότε η συνάρτηση πυκνότητας πιθανότητας της Γάμμα Κατανομής συμπίπτει με την συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής την οποία μελετήσαμε στην υποενότητα 2.1.1.

Τέλος (από Willmot & Lin, 1994) έχουμε ότι εάν $a \leq 1$ τότε ο λογάριθμος της $f(x)$ είναι κυρτή συνάρτηση και η συνάρτηση κατανομής $F(x)$ ανήκει στην κλάση DFR ενώ εάν $a \geq 1$ τότε ο λογάριθμος της $f(x)$ είναι κοίλη συνάρτηση και η συνάρτηση κατανομής $F(x)$ ανήκει στην κλάση IFR.

Στην συνέχεια, στις δύο επόμενες υποενότητες θα μελετήσουμε την βαθμίδα αποτυχίας για την περίπτωση όπου έχουμε μια μείξη Γάμμα Κατανομών αλλά και μια μείξη Εκθετικών κατανομών αντίστοιχα. Για τον σκοπό αυτό είναι ιδιαίτερα σημαντικό να παρουσιάσουμε με τεκμηριωμένο τρόπο αυτή την κατηγορία κατανομών η οποία χρησιμοποιείται συχνά στην αναλογιστική επιστήμη και η οποία είναι η μείξη κατανομών (ή mixture distributions). Γενικά μπορούμε να παρατηρήσουμε ότι μια μείξη κατανομών δεν είναι τίποτε άλλο παρά μια εφαρμογή του νόμου ολικής πιθανότητας σε κάποιες γνωστές κατανομές.

Ορισμός 2.2: (Κ.Πολίτης, 2012) Η κατανομή F μιας τυχαίας μεταβλητής X θα λέμε ότι είναι μια μεμιγμένη κατανομή ή μια μείξη κατανομών όταν ικανοποιεί μια από τις παρακάτω συνθήκες:

A) Η κατανομή F γράφεται ως ένα διακριτό άθροισμα με την εξής μορφή:

$$F(x) = \sum_{i=1}^n a_i \cdot F_{x_i(x)}$$

Για κάποιες θετικές σταθερές a_1, a_2, \dots, a_n τέτοιες ώστε $\sum_{i=1}^n a_i = 1$ και για κάποιες τυχαίες μεταβλητές X_1, X_2, \dots, X_n έτσι ώστε η X_i να έχει συνάρτηση κατανομής $F_{x_i(x)}$ αντίστοιχα. Στην περίπτωση αυτή η κατανομή F ονομάζεται **διακριτή μείξη κατανομών**.

B) Η κατανομή F γράφεται ως ένα συνεχές ολοκλήρωμα με την εξής μορφή:

$$F(x) = \int_{-\infty}^{\infty} F_{X|Y}(x|y) \cdot f_Y(y) dy$$

Για κάποια συνεχή τυχαία μεταβλητή Y με συνάρτηση πυκνότητας $f_Y(y)$. Ενώ παράλληλα η κατανομή $F_{X|Y}(x|y)$ είναι η δεσμευμένη (αθροιστική) συνάρτηση κατανομής της τυχαίας μεταβλητής X δοθέντος ότι $Y = y$. Στην περίπτωση αυτή η κατανομή F ονομάζεται **συνεχή μείξη κατανομών**.

Γενικότερα στην αναλογιστική επιστήμη χρησιμοποιούνται και τα δύο είδη μείξης κατανομών, ενώ στο κομμάτι της θεωρίας χρεοκοπίας πολύ συχνά χρησιμοποιούνται οι διακριτές μείξεις για την μοντελοποίηση των μεγεθών των αποζημιώσεων ενός χαρτοφυλακίου. Ωστόσο και οι συνεχείς μείξεις ως μοντέλα για τα μεγέθη ζημιών έχουν βρει πολλές εφαρμογές σε διάφορους κλάδους της θεωρίας κινδύνων από το 1940 και ύστερα.

2.1.3 Βαθμίδα αποτυχίας στην περίπτωση διακριτής μείξης Γάμμα Κατανομών

Έστω η τ.μ. X η οποία περιγράφει τον χρόνο ζωής ενός ατόμου ή ενός συστήματος αντίστοιχα και ακολουθεί μια μείξη από συνολικά r , Γάμμα κατανομές με παραμέτρους $k = 1, 2, \dots, r$ και b σταθερό.

Τότε θεωρούμε την παρακάτω συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \sum_{k=1}^r q_k \cdot \frac{b^k \cdot x^{k-1} \cdot e^{-bx}}{(k-1)!}, \quad x > 0$$

όπου $\{q_1, q_2, \dots, q_r\}$ έχουν άθροισμα ίσο με την μονάδα.

Επομένως χρησιμοποιώντας την σχέση (2.1) για την συνάρτηση κατανομής έχουμε ότι:

$$\bar{F}(x) = \sum_{k=1}^r q_k \cdot \int_x^{\infty} \frac{b^k \cdot y^{k-1} \cdot e^{-by}}{(k-1)!} dy$$

$$= e^{-bx} \cdot \sum_{k=1}^r q_k \cdot \sum_{j=0}^{k-1} \frac{(b \cdot x)^j}{j!}$$

Στη συνέχεια αλλάζοντας τη σειρά των αθροισμάτων (βλέπε Willmot & Lin, 2001) προκύπτει:

$$\bar{F}(x) = e^{-b \cdot x} \cdot \sum_{j=0}^{r-1} \frac{(b \cdot x)^j}{j!} \cdot \sum_{k=j+1}^r q_k$$

ή ισοδύναμα:

$$\bar{F}(x) = e^{-b \cdot x} \cdot \sum_{j=0}^{r-1} \bar{P}_j \cdot \frac{(b \cdot x)^j}{j!},$$

όπου $\bar{P}_j = \sum_{k=j+1}^r q_k$ για $j \leq r-1$ και $\bar{P}_j = 0$ για $j \geq r$

Τότε η βαθμίδα αποτυχίας με βάση τον ορισμό της εκφράζεται ως εξής, θέτοντας στο άθροισμα που περιλαμβάνει η συνάρτηση πυκνότητας πιθανότητας όπου $j = k-1$:

$$\mu(x) = \frac{f(x)}{\bar{F}(x)} = \frac{b \cdot \sum_{j=0}^{r-1} q_{j+1} \cdot \frac{(b \cdot x)^j}{j!}}{\sum_{j=0}^{r-1} \bar{P}_j \cdot \frac{(b \cdot x)^j}{j!}},$$

Μπορούμε από την τελευταία σχέση να διαπιστώσουμε ότι για $x = 0$ (έναρξη λειτουργίας ενός συστήματος ή γέννηση ενός ατόμου) τότε $\mu(0) = b \cdot q_1$ και αντίστοιχα, $\mu(\infty) = b$ εφόσον ισχύει ότι $\bar{P}_{r-1} = q_r$.

Αυτό σημαίνει ότι η συνάρτηση κατανομής $F(x)$ είναι DFR δηλαδή έχει φθίνουσα βαθμίδα αποτυχίας μόνο εάν $q_1 = 1$. Σε αυτή τη περίπτωση η $F(x)$ είναι εκθετική και συνεπώς και DFR.

Επιπλέον εφόσον $q_{j+1} = \bar{P}_j - \bar{P}_{j+1}$, μπορούμε εύκολα να παρατηρήσουμε ότι:

$$\mu(x) = b \cdot \left\{ 1 - \frac{\sum_{j=0}^{r-2} \bar{P}_{j+1} \cdot \frac{(b \cdot x)^j}{j!}}{\sum_{j=0}^{r-1} \bar{P}_j \cdot \frac{(b \cdot x)^j}{j!}} \right\}$$

Από την οποία σχέση βλέπουμε ότι :

$$\mu(x) \leq b = \mu(\infty) \text{ για όλα τα } x.$$

Υπάρχουν ωστόσο συνθήκες κάτω από τις οποίες η συνάρτηση κατανομής $F(x)$ ανήκει στην κλάση IFR δηλαδή έχει αύξουσα βαθμίδα αποτυχίας, και έχουν μελετηθεί αναλυτικά από τους Esary, Marshall και Proschan (1973) αλλά δεν θα δώσουμε έμφαση σε αυτό το κομμάτι, στην παρούσα διπλωματική εργασία.

2.1.4 Βαθμίδα αποτυχίας για την περίπτωση συνεχούς μείξης εκθετικών κατανομών

Έστω ότι η συνάρτηση κατανομής $F(x)$ για μια συνεχή μείξη εκθετικών κατανομών μπορεί να εκφρασθεί ως εξής (Willmot & Lin, 1994):

$$F(x) = \int_0^{\infty} (1 - e^{-\theta \cdot x}) \cdot dG(\theta) , \quad x \geq 0$$

Όπου θεωρούμε ως $G(\theta)$ την συνάρτηση κατανομής μιας μη-αρνητικής τυχαίας μεταβλητής. Τότε η βαθμίδα αποτυχίας θα είναι ίση με:

$$\mu(x) = -\frac{d}{dx} \ln(\bar{F}(x)) = \frac{\int_0^{\infty} \theta \cdot e^{-\theta \cdot x} dG(\theta)}{\int_0^{\infty} e^{-\theta \cdot x} dG(\theta)} \quad (2.4)$$

Για να αναλύσουμε την $\mu(x)$, θεωρούμε την συνάρτηση κατανομής $G_x(\theta) = \Pr(\theta_x \leq \theta)$ μιας τυχαίας μεταβλητής θ_x , η οποία ικανοποιεί την σχέση:

$$dG_x(\theta) = \frac{e^{-\theta \cdot x} dG(\theta)}{\int_0^{\infty} e^{-\theta \cdot x} dG(\theta)}$$

Ενώ παράλληλα έχουμε:

$$E(\theta_x^k) = \frac{\int_0^{\infty} \theta^k \cdot e^{-\theta \cdot x} dG(\theta)}{\int_0^{\infty} e^{-\theta \cdot x} dG(\theta)}$$

Εάν πάρουμε την πρώτη παράγωγο της σχέσης (2.4) που δώσαμε προηγουμένως για την βαθμίδα αποτυχίας θα έχουμε ότι:

$$\begin{aligned} \mu'(x) &= -\frac{\int_0^{\infty} \theta^2 \cdot e^{-\theta \cdot x} dG(\theta)}{\int_0^{\infty} e^{-\theta \cdot x} dG(\theta)} + \left\{ \frac{\int_0^{\infty} \theta^1 \cdot e^{-\theta \cdot x} dG(\theta)}{\int_0^{\infty} e^{-\theta \cdot x} dG(\theta)} \right\}^2 \\ &= -E(\theta_x^2) + \{E(\theta_x)\}^2 = -Var(\theta_x) \end{aligned}$$

Συνεπώς πάντοτε θα ισχύει ότι η βαθμίδα αποτυχίας της συνεχούς μείξης εκθετικών κατανομών έχει αρνητική πρώτη παράγωγο ($\mu'(x) \leq 0$) και επομένως η μείξη εκθετικών θα ανήκει στην κλάση DFR. Αναλυτικότερα την μονοτονία αλλά και τις δύο βασικές κλάσεις κατανομών που έχουμε δει, θα τις παρουσιάσουμε σε επόμενη υποενότητα αντίστοιχα.

Ωστόσο γνωρίζουμε ότι η συνάρτηση κατανομής της εκθετικής $F(x|\theta) = 1 - e^{-\theta \cdot x}$ ανήκει στην κλάση της φθίνουσας βαθμίδα αποτυχίας. Με χρήση της ανισότητας Cauchy-Schwartz μπορεί ναδειχθεί ότι εάν:

$$F(x) = \int_{-\infty}^{+\infty} F(x|\theta) \cdot dG(\theta)$$

Όπου η $F(x|\theta)$ είναι μια συνάρτηση κατανομής η οποία ανήκει στην κλάση DFR και $G(\theta)$ είναι μια συνάρτηση κατανομής τότε και η $F(x)$ αντίστοιχα θα έχει φθίνουσα βαθμίδα αποτυχίας. Δηλαδή όπως θα δούμε αργότερα και ένα θεώρημα των Barlow & Proschan, η μείξη κατανομών διατηρεί την ιδιότητα της φθίνουσας βαθμίδα αποτυχίας και γενικά μπορούμε να θεωρήσουμε ότι η μείξη κατανομών είναι σαν μια διαδικασία η οποία «γεννάει» κατανομές με παχιά ουρά (Barlow & Proschan, 1975).

2.2 Δύο βασικοί νόμοι που διέπουν την βαθμίδα αποτυχίας

2.2.1 Νόμος του Gompertz

Το 1825 ο Gompertz ήταν ο πρώτος που πρότεινε μια εκθετική μορφή για την ένταση θνησιμότητας (βαθμίδα αποτυχίας) των ανθρώπινων όντων. Συμφωνά με τον νόμο αυτό η δύναμη του ατόμου ώστε να αντισταθεί στο θάνατο μειώνεται με ρυθμό ανάλογο της δύναμης αυτής όπου:

$$\frac{1}{\mu(x)} = \text{μέτρο αντίστασης στο θάνατο (measure of resistance to death)}$$

Και συνεπώς θα ισχύει:

$$\frac{d}{dx} \left(\frac{1}{\mu(x)} \right) = -h \cdot \frac{1}{\mu(x)}, \quad \text{όπου } h = \text{σταθερά αναλογίας, } h > 0$$

$$\text{Έστω: } g(x) = \frac{1}{\mu(x)} \quad \text{τότε } g'(x) = -h \cdot g(x) \Rightarrow \int \frac{g'(x)}{g(x)} = - \int h dx \Rightarrow$$

$$\int (\ln(g(x)))' dx = - \int h dx \Rightarrow \ln(g(x)) = -h \cdot x - A$$

$$\Rightarrow \ln\left(\frac{1}{\mu(x)}\right) = -h \cdot x - A \Rightarrow \ln(\mu(x)) = h \cdot x + A$$

$$\text{Άρα } \mu(x) = e^{A+h \cdot x} = e^A \cdot (e^h)^x \quad \text{όπου θέτουμε } B = e^A \quad \text{και } c = e^h$$

Η βαθμίδα αποτυχίας έχει την μορφή: $\mu(x) = B \cdot c^x$

Άρα η ένταση θνησιμότητας θα αυξάνεται με εκθετικό ρυθμό σε σχέση με τον χρόνο.

Η αντίστοιχη συνάρτηση επιβίωσης θα έχει την μορφή:

$$S(x) = \bar{F}(x) = e^{-\int_0^x \mu(t)dt} = \frac{1}{g} \cdot g^{c^x}, \quad \text{όπου έχουμε θέσει ως } \ln g = -\frac{B}{\ln c}$$

2.2.2 Νόμος του Makeham

Το 1867 ο Makeham παρουσίασε μια πιο γενικευμένη μορφή του παραπάνω τύπου για την βαθμίδα αποτυχίας, λαμβάνοντας υπόψιν την αρχική περίοδο της ζωής ενός ατόμου κατά την οποία η θνησιμότητα είναι σταθερή και οφείλεται σε εξωτερικές αιτίες (π.χ. ατυχήματα). Έτσι η βαθμίδα αποτυχίας θα έχει την μορφή:

$$\mu(x) = D + B \cdot c^x \quad \text{με } x \geq 0 \text{ και } D, B, C > 0$$

όπου θέτουμε $B = e^A$ και $c = e^h$ αντίστοιχα.

Σε αυτή τη περίπτωση η συνάρτηση επιβίωσης θα ισούται με:

$$\begin{aligned} \int_0^x \mu(t)dt &= \int_0^x (D + B \cdot c^t)dt = D \cdot x + \frac{B}{\ln c} \cdot [c^x - 1] \Rightarrow \\ \Rightarrow S(x) &= e^{-[D \cdot x + \frac{B}{\ln c} \cdot (c^x - 1)]} \end{aligned}$$

Αν θέσουμε όμως:

$$S = e^{-D} \quad \text{και} \quad \ln g = -\frac{B}{\ln c}$$

Τότε αντίστοιχα έχουμε :

$$S(x) = \frac{1}{g} \cdot S^x \cdot g^{c^x}$$

Και οι δύο παραπάνω μορφές για την βαθμίδα αποτυχίας (ή ένταση θνησιμότητας) χρησιμοποιούνται ευρέως στην αναλογιστική επιστήμη και κυρίως στον κλάδο της δημογραφίας ενώ επίσης έχουν πάρα πολλές εφαρμογές σε αρκετούς τομείς (Ε. Χατζηκωνσταντινίδης, 2018).

2.3 Μονοτονία βαθμίδας αποτυχίας και κλάσεις κατανομών

Στην ενότητα αυτή θα μελετήσουμε δύο βασικές κλάσεις μονοτονίας κατανομών όπως αυτές χρησιμοποιούνται στην Θεωρία Αξιοπιστίας. Η ταξινόμηση των κατανομών σε αυτές τις κλάσεις γίνεται με βάση την βαθμίδα αποτυχίας ή τον μέσο υπολειπόμενο χρόνο ζωής που αποτελεί μια στενά συνδεδεμένη έννοια με το failure rate και την οποία θα αναλύσουμε διεξοδικά στο επόμενο κεφάλαιο (3^ο κεφάλαιο) της παρούσας εργασίας.

Η μελέτη της μορφής αλλά και της μονοτονίας της συνάρτησης της βαθμίδας αποτυχίας αποτελεί βασικό στοιχείο ανάλυσης για διάφορους τομείς όπως η ανάλυση επιβίωσης αλλά και η θεωρία αξιοπιστίας.

Η δεσμευμένη πιθανότητα αποτυχίας $\mu(t) \cdot dt$ στο διάστημα $(t, dt]$ περιγράφει ουσιαστικά της ιδιότητες γήρανσης των αντίστοιχων κατανομών, οι οποίες είναι ιδιαίτερα σημαντικές για πάρα πολλές εφαρμογές σε διάφορους κλάδους. Για αυτό τον λόγο μια ποιοτική περιγραφή της μονοτονίας της βαθμίδας αποτυχίας $\mu(x)$ μπορεί να φανεί χρήσιμη για την στοχαστική ανάλυση αποτυχιών, θανάτων ή ακόμη και καταστροφών. Στη περίπτωση ωστόσο της εκθετικής κατανομής, όπου την μελετήσαμε αναλυτικά σε προηγούμενη υποενότητα, βλέπουμε ότι το failure rate είναι σταθερό και συνεπώς η πιθανότητα θανάτου ή αποτυχίας δεν αυξάνεται με τον χρόνο. Πρόκειται δηλαδή για τ.μ. που περιγράφουν τον χρόνο ζωής και οι οποίες χαρακτηρίζονται από ιδιότητες «μη γήρανσης».

Ωστόσο σε πάρα πολλά δεδομένα της ανάλυσης επιβίωσης γίνεται συχνά μοντελοποίηση με την χρήση μονότονων βαθμίδων αποτυχίας. Παρ' όλα αυτά η μονοτονία της συνάρτησης της βαθμίδας αποτυχίας δεν μπορεί να είναι μόνο αύξουσα ή μόνο φθίνουσα, με την αυστηρή έννοια, καθώς μπορεί για παράδειγμα η θνησιμότητα σε ένα συγκεκριμένο χρονικό διάστημα να αυξάνεται ραγδαία και ύστερα από μια χρονική στιγμή και μετά να αρχίσει να μειώνεται. Σε αυτή τη περίπτωση λέμε ότι η βαθμίδα αποτυχίας έχει μια ανάποδη λεκανοειδή μορφή (upside-down bathtub shape) και τα δεδομένα μας μπορούν να μοντελοποιηθούν με τη χρήση λογαριθμο-κανονικής ή αντίστροφης Γκαουσιανής κατανομής. Υπάρχουν και περιπτώσεις όπου λόγω της λεγόμενης βρεφικής θνησιμότητας στην αρχή της ζωής του ανθρώπου, η θνησιμότητα ξεκινάει από ένα αρκετά υψηλό επίπεδο, στην συνέχεια της ζωής ενός ανθρώπου μειώνεται έως ότου φτάσει σε ένα ελάχιστο σημείο και μετά αυξάνεται. Σε αυτή τη περίπτωση λέμε ότι η βαθμίδα αποτυχίας έχει μια λεκανοειδή μορφή (bathtub shape) και για την μοντελοποίηση των δεδομένων μπορούν να χρησιμοποιηθούν μείξεις κατανομών οι οποίες αποτελούν και βασικό κομμάτι της δοθείσας διπλωματικής εργασίας.

Ας δούμε στην συνέχεια κάποιους βασικούς κανόνες για την μονοτονία των βαθμίδων αποτυχίας:

2.3.1 Μονοτονία βαθμίδων αποτυχίας

Η βαθμίδα αποτυχίας $\mu(x)$ όπως έχουμε ήδη αναφέρει αποτελεί μια σημαντική ποσότητα στον αναλογισμό η οποία χρησιμοποιείται για την μοντελοποίηση των κατανομών θετικών συνεχών τ.μ. που περιγράφουν χρόνο ζωής. Παράλληλα μας δείχνει τον στιγμιαίο ρυθμό θνησιμότητας στην μονάδα του χρόνου αλλά και την γήρανση σε έναν πληθυσμό. Για αυτό είναι σημαντικό να αναλύσουμε και να διακρίνουμε τις μορφές τις οποίες μπορεί να πάρει αυτή η συνάρτηση ανάλογα με την μονοτονία της σε διάφορα διαστήματα. Εμείς θα εξετάσουμε τέσσερις βασικές μορφές της συνάρτησης της βαθμίδας αποτυχίας ως προς την μονοτονία της και οι οποίες είναι οι εξής:

- Η βαθμίδα αποτυχίας $\mu(x)$ είναι αύξουσα σε όλο το διάστημα $(0, \infty)$ και συνεπώς θα ισχύει για το διάστημα αυτό ότι $\mu'(x) > 0$. Τότε η συνάρτηση κατανομής $F(x)$ του χρόνου ζωής λέμε ότι έχει αύξουσα βαθμίδα αποτυχίας, ή ότι είναι IFR.

- Η βαθμίδα αποτυχίας $\mu(x)$ είναι φθίνουσα σε όλο το διάστημα $(0, \infty)$ και συνεπώς θα ισχύει για το διάστημα αυτό ότι $\mu'(x) < 0$. Τότε η συνάρτηση κατανομής $F(x)$ του χρόνου ζωής λέμε ότι έχει φθίνουσα βαθμίδα αποτυχίας, ή ότι είναι DFR.

- Η βαθμίδα αποτυχίας $\mu(x)$ είναι αύξουσα στο διάστημα $[0, t_0)$ και φθίνουσα στο διάστημα (t_0, ∞) . Δηλαδή ισχύει ότι:
$$\mu'(x) > 0 \text{ για } x \in [0, t_0), \quad \mu'(t_0) = 0 \text{ και } \mu'(x) < 0 \text{ για } x \in (t_0, \infty)$$

Τότε η συνάρτηση κατανομής $F(x)$ του χρόνου ζωής θα λέμε ότι έχει ανάποδη λεκανοειδή μορφή (upside-down bathtub failure rate).

- Η βαθμίδα αποτυχίας $\mu(x)$ είναι φθίνουσα στο διάστημα $[0, t_0)$ και αύξουσα στο διάστημα (t_0, ∞) . Δηλαδή ισχύει ότι:
$$\mu'(x) < 0 \text{ για } x \in [0, t_0), \quad \mu'(t_0) = 0 \text{ και } \mu'(x) > 0 \text{ για } x \in (t_0, \infty)$$

Τότε η συνάρτηση κατανομής $F(x)$ του χρόνου ζωής θα λέμε ότι έχει λεκανοειδή μορφή (bathtub failure rate).

2.3.2 Δύο βασικές κλάσεις κατανομών

Όπως ήδη αναφέραμε και προηγουμένως δύο από τις βασικότερες κλάσεις στις οποίες εντάσσονται οι κατανομές των χρόνων ζωής ανάλογα με την μονοτονία της βαθμίδας αποτυχίας τους είναι η κλάση IFR όταν η βαθμίδα αποτυχίας $\mu(x)$ είναι αύξουσα συνάρτηση

ως προς x για $x \in [0, \infty)$ και η κλάση DFR όταν η βαθμίδα αποτυχίας $\mu(x)$ είναι φθίνουσα συνάρτηση ως προς x για $x \in [0, \infty)$.

Στην προηγούμενη υποενότητα δείξαμε ότι η βαθμίδα αποτυχίας καθορίζει κατά έναν μοναδικό τρόπο την συνάρτηση επιβίωσης ή δεξιά ουρά της τ.μ. που περιγράφει τον χρόνο ζωής, άρα και η μονοτονία του failure rate θα εξαρτάται από την δεξιά ουρά.

Σύμφωνα με τους Willmot & Lin (2001) υπάρχει ένα βασικό κριτήριο σύμφωνα με το οποίο συνδέεται η μονοτονία της βαθμίδας αποτυχίας σε σχέση με την συνάρτηση επιβίωσης αντίστοιχα.

Ισχύει ο εξής κανόνας:

- Η συνάρτηση κατανομής $F(x)$ έχει φθίνουσα βαθμίδα αποτυχίας ή DFR εάν ο λόγος $\frac{\bar{F}(x+y)}{\bar{F}(x)}$ είναι μια αύξουσα(μη φθίνουσα) συνάρτηση ως προς x και για σταθερό $y \geq 0$ και αν η δεξιά ουρά $\bar{F}(x)$ είναι λογαριθμικά κυρτή συνάρτηση.
- Η συνάρτηση κατανομής $F(x)$ έχει αύξουσα βαθμίδα αποτυχίας ή IFR εάν ο λόγος $\frac{\bar{F}(x+y)}{\bar{F}(x)}$ είναι μια φθίνουσα(μη αύξουσα) συνάρτηση ως προς x και για σταθερό $y \geq 0$ και αν η δεξιά ουρά $\bar{F}(x)$ είναι λογαριθμικά κοίλη συνάρτηση.

Η εκθετική κατανομή εφόσον διαθέτει μια σταθερή στο χρόνο βαθμίδα αποτυχίας, μπορεί να θεωρηθεί είτε μια αύξουσα είτε μια φθίνουσα συνάρτηση και συνεπώς να ανήκει και στις δύο παραπάνω κλάσεις. Στην περίπτωση της κατανομής Γάμμα, η παράμετρος a είναι αυτή η οποία θα καθορίσει την μονοτονία της βαθμίδας αποτυχίας ανάλογα με το εάν είναι μεγαλύτερη ή μικρότερη από την μονάδα αντίστοιχα. Ενώ για την Erlang κατανομή η οποία αποτελεί γενίκευση της Γάμμα και η παράμετρος a ανήκει στο σύνολο των ακεραίων με $a=1,2,3,\dots$ δηλαδή εφόσον $a > 1$ πάντοτε θα έχει αύξουσα βαθμίδα αποτυχίας.

Στη συνέχεια θα δώσουμε έμφαση σε δυο σημαντικά θεωρήματα όπως αυτά αναφέρονται από τους Barlow & Proschan (1975).

Θεώρημα 2.3.1: Έστω δύο ανεξάρτητες και θετικές συνεχείς τυχαίες μεταβλητές X_1 και X_2 , με συναρτήσεις κατανομής F και G αντίστοιχα και έστω H η συνάρτηση κατανομής του αθροίσματος των δυο τ.μ. $X_1 + X_2$.

Τότε η H ονομάζεται συνέλιξη των δυο αυτών κατανομών και ικανοποιεί την ακόλουθη σχέση:

$$H(y) = \int_0^y G(y-x) \cdot dF(x)$$

Εάν οι συναρτήσεις κατανομών $F(x)$ και $G(x)$ ανήκουν και οι δύο στην κλάση IFR της αύξουσας βαθμίδας αποτυχίας τότε και η συνέλιξη αυτών $H(x)$ θα ανήκει επίσης στην ίδια κλάση IFR.

Θεώρημα 2.3.2: Έστω οι ανεξάρτητες και θετικές συνεχείς τυχαίες μεταβλητές X_1, X_2, \dots, X_v με συναρτήσεις κατανομών αντίστοιχα F_1, F_2, \dots, F_v .

Τότε η διακριτή μείξη αυτών των κατανομών ορίζεται μέσω της ακόλουθης σχέσης:

$$F(y) = \sum_{i=1}^n a_i \cdot F_i(x) \quad \text{για την οποία ισχύει } \sum_{i=1}^n a_i = 1 \text{ και } a_i > 0 \text{ για } \forall i = 1, 2, \dots, n$$

Εάν οι συναρτήσεις κατανομών F_1, F_2, \dots, F_v ανήκουν στην κλάση DFR της φθίνουσας βαθμίδας αποτυχίας τότε και η διακριτή μείξη $F(y)$ αυτών θα ανήκει επίσης στην ίδια κλάση DFR.

Αξίζει να σημειωθεί όμως ότι δεν ισχύει το παραπάνω θεώρημα για την κλάση της αύξουσας βαθμίδας αποτυχίας IFR καθώς όταν συνεχείς συναρτήσεις κατανομών ανήκουν στην IFR τότε δεν ισχύει απαραίτητα ότι και η μείξη αυτών θα ανήκει στην ίδια κλάση.

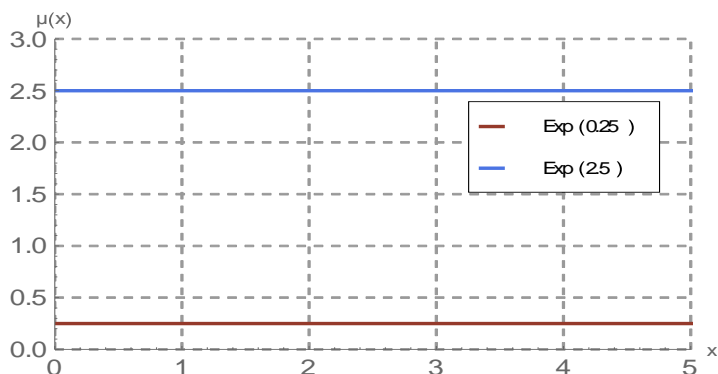
Συνεπώς αξίζει να παρατηρήσουμε ότι η μείξη εκθετικών κατανομών ανήκει στην κλάση DFR όπως αναφέραμε και προηγουμένως ενώ παράλληλα και η μείξη Γάμμα κατανομών με παράμετρο $a < 1$ θα είναι επίσης DFR.

Στον παρακάτω πίνακα έχουμε συγκεντρωτικά τα εξής αποτελέσματα:

ΚΑΤΑΝΟΜΗ ΧΡΟΝΟΥ ΖΩΗΣ	ΚΛΑΣΗ ΒΑΘΜΙΔΑΣ ΑΠΟΤΥΧΙΑΣ
ΕΚΘΕΤΙΚΗ ΚΑΤΑΝΟΜΗ	IFR & DFR
ΓΑΜΜΑ ΚΑΤΑΝΟΜΗ	Για $a > 1$ είναι IFR Για $a < 1$ είναι DFR Για $a = 1$ είναι DFR & IFR
ΜΕΙΞΗ ΕΚΘΕΤΙΚΩΝ ΚΑΤΑΝΟΜΩΝ	DFR
ΜΕΙΞΗ ΓΑΜΜΑ ΚΑΤΑΝΟΜΩΝ	Για $a_1, a_2, \dots, a_v < 1$ είναι DFR Για $a_1, a_2, \dots, a_v > 1$ ή $a_i > 1$ και $a_j < 1$ (με $i \neq j$) μπορεί να είναι IFR ή DFR ή bathtub ή upside-down bathtub

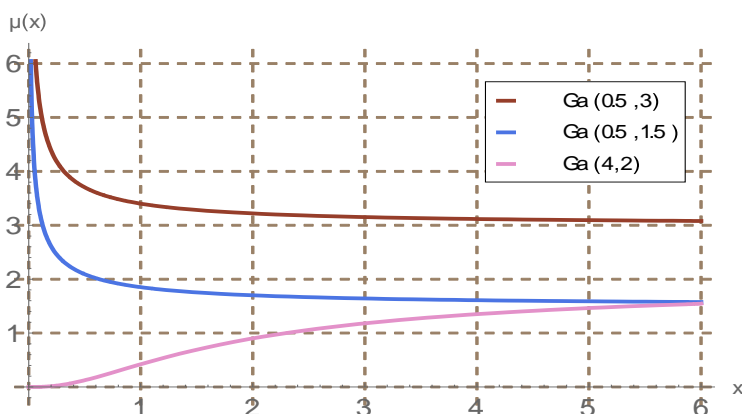
Ας δούμε ωστόσο με την βοήθεια του υπολογιστικού προγράμματος Mathematica και κάποια διαγράμματα όπου φαίνεται η μορφή της βαθμίδας αποτυχίας για τρεις βασικές κατανομές τις οποίες αναφέραμε προηγουμένως:

Διάγραμμα 2.1: Βαθμίδα αποτυχίας για την εκθετική κατανομή με παράμετρο $\lambda = 0.25$ (κόκκινη γραμμή) και παράμετρο $\lambda = 2.5$ (μπλε γραμμή)



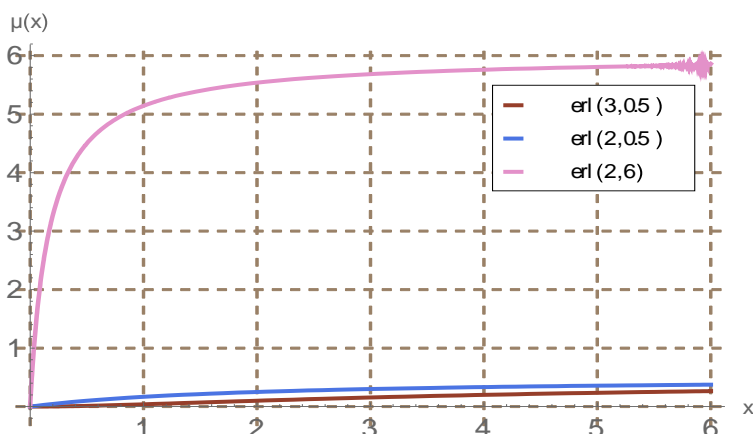
Όπως είχαμε ήδη αναφέρει στην υποενότητα 2.1.1, η εκθετική κατανομή είναι η μόνη κατανομή χρόνων ζωής για την οποία η βαθμίδα αποτυχίας είναι σταθερή και ίση με την παράμετρο λ , γεγονός το οποίο μπορούμε εύκολα να διαπιστώσουμε και από το παραπάνω διάγραμμα.

Διάγραμμα 2.2: Βαθμίδα αποτυχίας για την Γάμμα κατανομή με παραμέτρους $(0.5,3)$, $(0.5,1.5)$ και $(4,2)$



Για την περίπτωση της κατανομής Γάμμα από το παραπάνω διάγραμμα επιβεβαιώνουμε ότι όταν η παράμετρος a είναι μικρότερη από τη μονάδα (εδώ 0.5) τότε η βαθμίδα αποτυχίας είναι μια φθίνουσα συνάρτηση ως προς x . Ενώ όσο πιο μικρή είναι η παράμετρος b ($1.5 < 3$) τόσο η καμπύλη πλησιάζει τον άξονα των x . Αντίστοιχα όταν η παράμετρος a είναι μεγαλύτερη από τη μονάδα (εδώ 4) τότε η βαθμίδα αποτυχίας είναι μια αύξουσα συνάρτηση ως προς x .

Διάγραμμα 2.3: Βαθμίδα αποτυχίας για την κατανομή Erlang με παραμέτρους (3,0.5), (2,0.5) και (2,6)



Τέλος για την περίπτωση της κατανομής Erlang καθώς είναι μια κατανομή Γάμμα αλλά με παράμετρο a ακέραιο αριθμό (δηλαδή παράμετρο πάντα μεγαλύτερη της μονάδας) τότε η βαθμίδα αποτυχίας θα είναι αύξουσα. Βλέπουμε από το παραπάνω διάγραμμα ότι για παραμέτρους $a > 1$ (εδώ 3, 2 και 2) η συνάρτηση της βαθμίδας αποτυχίας είναι μια αύξουσα συνάρτηση ως προς x και για τις τρεις καμπύλες. Ενώ επιπλέον όσο η παράμετρος b μειώνεται ($0.5 < b$) τόσο η καμπύλη της βαθμίδας αποτυχίας πλησιάζει τον άξονα x .

ΚΕΦΑΛΑΙΟ 3

ΜΕΣΟΣ ΥΠΟΛΕΙΠΟΜΕΝΟΣ ΧΡΟΝΟΣ ΖΩΗΣ (MEAN RESIDUAL LIFETIME)

Στο κεφάλαιο αυτό θα δούμε αναλυτικά μια άλλη εξίσου σημαντική μεταβλητή για την ανάλυση επιβίωσης και τις ασφαλίσεις ζωής, η οποία είναι ο υπολειπόμενος ή μελλοντικός χρόνος ζωής. Ουσιαστικά η μεταβλητή αυτή μετράει την υπολειπόμενη διάρκεια ζωής που έχει ένα άτομο ή ένα σύστημα εφόσον έχει περάσει ένα συγκεκριμένο χρονικό διάστημα από την γέννηση ή την κατασκευή αντίστοιχα. Για παράδειγμα στον τομέα των συντάξεων μας ενδιαφέρει ιδιαίτερα να δούμε πόσος χρόνος ζωής απομένει σε ένα άτομο ηλικίας 65, που είναι η ηλικία συνταξιοδότησης σε πολλές χώρες σήμερα. Ωστόσο το βασικό αντικείμενο μελέτης μας σε αυτό το κεφάλαιο θα είναι η μέση τιμή της παραπάνω τυχαίας μεταβλητής η οποία ονομάζεται μέσος υπολειπόμενος χρόνος ζωής (mean residual lifetime-MRL).

Στη συνέχεια του κεφαλαίου θα μελετήσουμε και την ιδιαίτερα στενή σύνδεση μεταξύ του μέσου υπολειπόμενου χρόνου ζωής και της ποσότητας που αναλύσαμε προηγουμένως, της βαθμίδας αποτυχίας. Παράλληλα θα δούμε και κάποιες επιπλέον κλάσεις κατανομών που συσχετίζονται με τον μέσο υπολειπόμενο χρόνο ζωής αλλά και την βαθμίδα αποτυχίας ενώ επίσης θα μελετήσουμε και την μονοτονία του MRL.

Στην δημογραφία, την ανάλυση επιβίωσης αλλά και σε άλλους τομείς, ο ρόλος αυτής της ποσότητας είναι αρκετά σημαντικός. Ενώ η βαθμίδα αποτυχίας στον χρόνο t για μια τυχαία μεταβλητή T ή X μας δίνει πληροφορίες για ένα πολύ μικρό χρονικό διάστημα κοντά στο t , ο μέσος υπολειπόμενος χρόνος ζωής στον χρόνο t μπορεί να μας δώσει πληροφορίες για όλο το χρονικό διάστημα ζωής (t, ∞) που απομένει σε ένα άτομο ή σε ένα σύστημα (Guess and Proschan, 1988). Παρόλα αυτά οι δύο αυτές έννοιες συμπληρώνουν η μια την άλλη και στην θεωρία αξιοπιστίας που αφορά κυρίως την μηχανική, δίνεται κυρίως έμφαση και στις δύο ποσότητες αντίστοιχα.

Επιπρόσθετα θα δείξουμε και σε αυτό το κεφάλαιο ότι όπως η βαθμίδα αποτυχίας καθορίζει με έναν μοναδικό τρόπο την συνάρτηση κατανομής της τυχαίας μεταβλητής του χρόνου ζωής έτσι και ο μέσος υπολειπόμενος χρόνος ζωής καθορίζει μοναδικά την συνάρτηση κατανομής της τυχαίας μεταβλητής του υπολειπόμενου χρόνου ζωής.

Ας ορίσουμε λοιπόν αρχικά την τυχαία μεταβλητή T που αντιπροσωπεύει τον χρόνο ζωής που απομένει σε ένα άτομο ή σύστημα εφόσον έχει περάσει ένα χρονικό διάστημα.

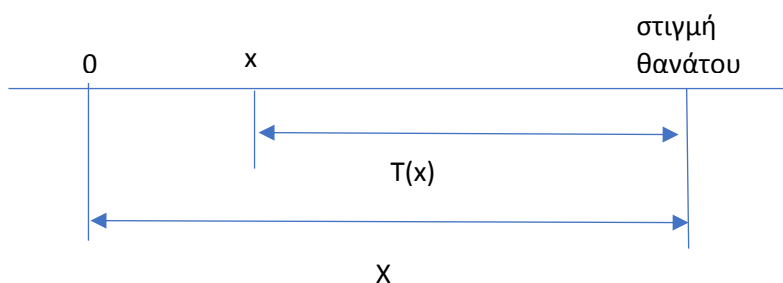
Ορισμός 3.1: Έστω η μη αρνητική και συνεχής τυχαία μεταβλητή X η οποία περιγράφει τον χρόνο ζωής ενός ατόμου ή ενός συστήματος. Τότε ορίζουμε μια νέα τυχαία μεταβλητή $T(x)$ η οποία θα εκφράζει την υπολειπόμενη διάρκεια ζωής ενός ατόμου ηλικίας (x) με $x > 0$.

Για παράδειγμα η $T(30)=60$ σημαίνει ότι για ένα άτομο ηλικίας 30 ετών, η υπολειπόμενη διάρκεια ζωής του θα είναι 60 έτη. Δηλαδή θα πεθάνει σε ηλικία 90 ετών.

Άρα η τυχαία μεταβλητή $x + T(x)$ θα εκφράζει αντίστοιχα την χρονική στιγμή θανάτου ενός νεογέννητου ατόμου. Επομένως θα ισχύει ότι: $T(x) = (X - x|X > x)$ και προφανώς

$$T(0) = X.$$

Σχήμα 3.1



Συνάρτηση κατανομής της $T(x)$

Η συνάρτηση κατανομής της $T(x)$ στο σημείο t συμβολίζεται με:

$$\begin{aligned} F_{T(x)}(t) &= t q_x = \Pr[T(x) \leq t] = \Pr(X - x \leq t | X > x) = \Pr(X \leq x + t | X > x) \\ &= \frac{\Pr[(X > x) \cap (X \leq x + t)]}{\Pr(X > x)} = \frac{\Pr[x < X \leq x + t]}{\Pr[X > x]} = \frac{S(x) - S(x + t)}{S(x)} \end{aligned}$$

Συνεπώς ισχύει ότι:

$$F_{T(x)}(t) = \frac{S(x) - S(x + t)}{S(x)} = 1 - \frac{S(x + t)}{S(x)} = 1 - \frac{\bar{F}(x + t)}{\bar{F}(x)}, \quad t \geq 0$$

Και επομένως για την συνάρτηση επιβίωσης της τυχαίας μεταβλητής $T(x)$ θα έχουμε αντιστοίχως ότι:

$$\bar{F}_{T(x)}(t) = 1 - F_{T(x)}(t) = t p_x = \frac{S(x + t)}{S(x)} = \frac{\bar{F}(x + t)}{\bar{F}(x)}, \quad t \geq 0$$

Αν θυμηθούμε ωστόσο την περίπτωση της εκθετικής κατανομής για την οποία ισχύει η ιδιότητα «έλλειψης μνήμης» σύμφωνα με την οποία ο θάνατος δεν εξαρτάται από την ηλικία στην οποία βρίσκεται ένα άτομο (μη ρεαλιστικό) τότε θα καταλάβουμε ότι η περιγραφή του υπολειπόμενου χρόνου ζωής δεν έχει ιδιαίτερη σημασία για την συγκεκριμένη κατανομή. Εξάλλου εφόσον όπως δείξαμε προηγουμένως ισχύει για την εκθετική κατανομή ότι: $Pr(X > t + x | X > x) = Pr(X > t) = S(t)$ μπορούμε εύκολα να αντιληφθούμε πως: $F_{T(x)}(t) = F_X(t)$.

Στην συνέχεια θα μελετήσουμε αναλυτικά την μέση τιμή της παραπάνω τυχαίας μεταβλητής η οποία περιγράφει τον υπολειπόμενο χρόνο ζωής. Η ποσότητα του μέσου υπολειπόμενου χρόνου ζωής σε διάφορους τομείς της αναλογιστικής επιστήμης είναι ιδιαίτερα χρήσιμη. Ας δούμε ωστόσο πως ορίζεται, ως μια απλή μέση τιμή.

Ορισμός 3.2: Με βάση τους Willmot & Lin (2001), ο μέσος υπολειπόμενος χρόνος ζωής δίνεται από την παρακάτω σχέση:

$$E(T(X)) = m(x) = \frac{\int_x^\infty (t-x) \cdot dF(t)}{\bar{F}(x)}, \quad x \geq 0 \quad (3.1)$$

Επιπλέον από το προηγούμενο κεφάλαιο δείξαμε ότι γενικά για μια τ.μ. X ισχύει η σχέση: $E(X) = \int_0^\infty \bar{F}(x) dx$

Συνεπώς έχουμε :

$$m(x) = \int_0^\infty Pr(T(x) > t) dt = \int_0^\infty \frac{\bar{F}(x+t)}{\bar{F}(x)} dt \quad (3.2)$$

Ας δούμε όμως και μια εξίσου σημαντική έννοια στην αναλογιστική επιστήμη η οποία είναι η *κατανομή ισορροπίας*.

Αν λοιπόν διαιρέσουμε την σχέση $E(X) = \int_0^\infty \bar{F}(x) dx$ με την μέση τιμή $E(X)$ μιας τ.μ. X τότε θα ισχύει ότι: $\int_0^\infty \frac{\bar{F}(x)}{E(X)} dx = 1$, γεγονός το οποίο συνεπάγεται πως η $f_1(x) = \frac{\bar{F}(x)}{E(X)}$ είναι μια συνάρτηση πυκνότητας πιθανότητας, καθώς έχει ολοκλήρωμα από μηδέν έως άπειρο ίσο με την μονάδα, ακόμη και αν η $F(x)$ δεν είναι απόλυτα συνεχής.

Η αντίστοιχη συνάρτηση κατανομής: $F_1(x) = 1 - \bar{F}_1(x) = \int_0^x \frac{\bar{F}(y)}{E(X)} dy$, $x \geq 0$ καλείται κατανομή ισορροπίας της $F(x)$. Άρα θα έχουμε και για την δεξιά ουρά της κατανομής ισορροπίας ότι:

$$\bar{F}_1(x) \cdot E(X) = \int_x^\infty \bar{F}(y) dy$$

Επιστρέφοντας λοιπόν στην σχέση (3.2) που είχαμε καταλήξει για τον μέσο υπολειπόμενο χρόνο ζωής εναλλακτικά μπορούμε να έχουμε:

$$\begin{aligned} m(x) &= \frac{1}{\bar{F}(x)} \cdot \int_0^{\infty} \bar{F}(x+t) dt = \frac{1}{\bar{F}(x)} \cdot \int_x^{\infty} \bar{F}(y) dy = \frac{\bar{F}_1(x) \cdot E(X)}{\bar{F}(x)} \\ &= \int_x^{\infty} \left(\frac{t \cdot f(t)}{\bar{F}(x)} - x \right) dt \end{aligned}$$

όπου $\bar{F}(x) > 0$ είναι η συνάρτηση δεξιάς ουράς της τυχαίας μεταβλητής X και $\bar{F}_1(x)$ είναι η συνάρτηση δεξιάς ουράς της κατανομής ισορροπίας.

Επιπλέον μπορούμε να δούμε ότι για $x = 0$ ισχύει:

$$\bar{F}(0) = 1 \text{ και } \bar{F}_1(0) \cdot E(X) = \int_0^{\infty} \bar{F}(y) dy = E(X) \text{ επομένως: } m(0) = E(X).$$

Ο μέσος υπολειπόμενος χρόνος ζωής συνδέεται αρκετά στενά με την βαθμίδα αποτυχίας όταν αυτή υπάρχει και χωρίς να είναι απαραίτητα απόλυτα συνεχής. Επιπλέον μεγάλες τιμές της $m(x)$ σχετίζονται κυρίως με παχιά ουρά μιας κατανομής. Παράλληλα ο μέσος υπολειπόμενος χρόνος ζωής μέσω της παρακάτω σχέσης που θα δείξουμε, παρατηρούμε ότι συνδέεται ιδιαίτερα και με την κατανομή ισορροπίας $F_1(x)$ καθώς:

$$-\frac{d}{dx} \ln(\bar{F}_1(x)) = \frac{\bar{F}(x)/E(x)}{\bar{F}_1(x)} = \frac{1}{m(x)}$$

Η τελευταία σχέση μας δηλώνει ότι ο αντίστροφος του μέσου υπολειπόμενου χρόνου ζωής ($1/m(x)$) είναι η βαθμίδα αποτυχίας για την συνάρτηση της κατανομής ισορροπίας $F_1(x)$. Άρα από την σχέση που δείξαμε ότι ισχύει σε προηγούμενο κεφάλαιο έχουμε:

$$\bar{F}_1(x) = e^{-\int_0^x \frac{1}{m(y)} dy} \quad (3.3)$$

Συνεπώς από την ισότητα: $F'_1(x) = \frac{\bar{F}(x)}{m(0)}$ και την σχέση (3.3) ταυτόχρονα παρατηρούμε πως η συνάρτηση κατανομής $F(x)$ είναι κατά έναν μοναδικό τρόπο καθορισμένη από τον μέσο υπολειπόμενο χρόνο ζωής $m(x)$.

Στην συνέχεια, με βάση τους Willmot & Lin (2001) θα δούμε κάποιους βασικούς κανόνες για την μονοτονία του MRL αντίστοιχα:

- Η συνάρτηση κατανομής $F(x)$ ανήκει στην κλάση του αύξοντα μέσου υπολειπόμενου χρόνου ζωής (IMRL) όταν το $m(x)$ είναι μη φθίνουσα (αύξουσα) συνάρτηση ως προς x . Θυμόμαστε από το προηγούμενο κεφάλαιο ότι η $F(x)$ ανήκει στην κλάση της φθίνουσας βαθμίδας αποτυχίας (DFR) όταν ο λόγος $\frac{\bar{F}(x+t)}{\bar{F}(x)}$ είναι μη φθίνουσα συνάρτηση ως προς x και για σταθερό $t > 0$. Συνεπώς μπορούμε να συμπεράνουμε ότι

η κλάση DFR συνεπάγεται κλάση IMRL. Άρα η κλάση DFR είναι ένα υποσύνολο της κλάσης IMRL.

- Αντίστοιχα μπορούμε να συμπεράνουμε ότι η κλάση IFR συνεπάγεται κλάση DMRL. Άρα η κλάση IFR είναι ένα υποσύνολο της κλάσης DMRL.

Η κλάση IMRL όπως αντίστοιχα και η κλάση DFR σύμφωνα με τον Bondesson (1983) διατηρούνται στις μείξεις κατανομών. Ωστόσο από την άλλη πλευρά όταν έχουμε μια κλάση IFR η οποία διατηρεί την ιδιότητα της στις συνελίξεις κατανομών, η κλάση DMRL αντίστοιχα δεν διατηρεί αυτή την ιδιότητα.

Παρ' όλα αυτά, δεν προκύπτει ότι κάθε μη αρνητική συνάρτηση $m(x) \geq 0$ θα είναι αντίστοιχα και συνάρτηση μέσου υπολειπόμενου χρόνου ζωής.

Θα παρουσιάσουμε λοιπόν την ακόλουθη πρόταση σύμφωνα με την οποία γίνονται γνωστές κάποιες από τις βασικές προϋποθέσεις που θα πρέπει να έχει μια συνάρτηση $m(x)$ ώστε να είναι συνάρτηση MRL όπως αυτές αναφέρονται στο βιβλίο των Shaked and Shanthikumar (Κεφάλαιο 2, 2007).

Πρόταση 3.1: Μια συνάρτηση $m(x)$ θα ονομάζεται συνάρτηση μέσου υπολειπόμενου χρόνου ζωής μιας μη αρνητικής και συνεχής τ.μ. $T(x)$ (η οποία περιγράφει τον υπολειπόμενο χρόνο ζωής ενός ατόμου ή ενός συστήματος εφόσον έχουν περάσει x έτη) εάν και μόνο αν ισχύουν τα κάτωθι:

- Η $m(x)$ να είναι συνεχής συνάρτηση
- $0 \leq m(x) < \infty$ για κάθε $x \geq 0$
- $m(0) > 0$
- $m(x) + x$ να είναι μια αύξουσα συνάρτηση ως προς $x \geq 0$
- αν υπάρχει ένα x_0 τέτοιο ώστε $m(x_0) = 0$, τότε $m(x) = 0$ για όλα τα $x \geq x_0$,

διαφορετικά όταν δεν υπάρχει x_0 τ. ω $m(x_0) = 0$, τότε $\int_0^{\infty} \frac{1}{m(x)} dx = \infty$

3.1. Σύνδεση βαθμίδας αποτυχίας και μέσου υπολειπόμενου χρόνου ζωής

Στην ενότητα αυτή θα δούμε αναλυτικά την σχέση μεταξύ των δύο ποσοτήτων (βαθμίδας αποτυχίας και μέσου υπολειπόμενου χρόνου ζωής) η οποία φαίνεται να είναι ιδιαίτερη στενή. Όπως ήδη γνωρίζουμε η βαθμίδα αποτυχίας ή ένταση θνησιμότητας εκφράζει τον στιγμιαίο κίνδυνο αποτυχίας, ο οποίος αυξάνεται με την πάροδο του χρόνου διότι λόγω της «γήρανσης» ενός οργανισμού ή ενός συστήματος, ο κίνδυνος αποτυχίας είναι μεγαλύτερος με το πέρασμα του χρόνου. Αντιθέτως όπως θα περιμένει κανείς, ο μέσος υπολειπόμενος χρόνος ζωής που απομένει σε ένα ανθρώπινο ον ή ένα σύστημα, μειώνεται καθ' όλη τη διάρκεια του χρόνου. Επομένως με μια πρώτη ματιά θα μπορούσαμε να πούμε ότι τα δυο αυτά μεγέθη είναι αντιστρόφως ανάλογα μεταξύ τους. Επομένως σχετικά με την μονοτονία των δύο συναρτήσεων θα ισχύει ότι όταν η βαθμίδα αποτυχίας είναι μια αύξουσα συνάρτηση του χρόνου τότε ο μέσος υπολειπόμενος χρόνος ζωής θα είναι μια φθίνουσα συνάρτηση του χρόνου και αντίστροφα.

Δείξαμε προηγουμένως ότι: $m(x) = \frac{1}{\bar{F}(x)} \cdot \int_x^\infty \bar{F}(y) dy = \frac{1}{S(x)} \cdot \int_x^\infty S(y) dy$

Επομένως παίρνοντας την πρώτη παράγωγο της παραπάνω σχέσης έχουμε:

$$m'(x) = \frac{-S(x)^2 + f(x) \cdot \int_x^\infty S(y) dy}{S(x)^2}$$

Η σχέση αυτή είναι ισοδύναμη με: $m'(x) = \mu(x) \cdot m(x) - 1$,

όπου $\mu(x)$ η βαθμίδα αποτυχίας αντίστοιχα.

Εναλλακτικά αν την εκφράσουμε ως προς τη βαθμίδα αποτυχίας έχουμε:

$$\mu(x) = \frac{m'(x)+1}{m(x)} \quad (3.4)$$

Τέλος γνωρίζοντας για την συνάρτηση επιβίωσης από προηγούμενο κεφάλαιο ότι ισχύει: $S(x) = e^{-\int_0^x \mu(y) dy}$ και σε συνδυασμό με την σχέση (3.4)

Μπορούμε εύκολα να διαπιστώσουμε ότι:

$$S(x) = \frac{m(0)}{m(x)} \cdot e^{-\int_0^x \frac{1}{m(y)} dy}$$

Συνεπώς η συνάρτηση επιβίωσης, η συνάρτηση της βαθμίδας αποτυχίας και η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής είναι έννοιες αλληλένδετες και αν γνωρίζουμε μια από αυτές μπορούμε να εύκολα να βρούμε τις άλλες δύο.

Στην συνέχεια, ας αναλύσουμε τον μέσο υπολειπόμενο χρόνο ζωής στην περίπτωση των Γάμμα κατανομών μέσω του παρακάτω παραδείγματος.

Γνωρίζουμε ότι η συνάρτηση πυκνότητας πιθανότητας της Γάμμα κατανομής με παραμέτρους a, b είναι η:

$$f(x) = \frac{b^a \cdot x^{a-1} \cdot e^{-b \cdot x}}{\Gamma(a)}, \quad x \geq 0$$

ενώ παράλληλα η μέση τιμή από τις πιθανότητες είναι ίση με $E(X) = \frac{a}{b}$. Γενικά ο μέσος υπολειπόμενος χρόνος ζωής $m(x)$ για την κατανομή Γάμμα είναι δύσκολο να εκφραστεί με μια απλή μορφή. Συνήθως μελετάμε για ευκολία την περίπτωση όπου η παράμετρος σχήματος είναι θετικός ακέραιος αριθμός (Erlang κατανομή).

Σε αυτή τη περίπτωση ο μέσος υπολειπόμενος χρόνος ζωής (Willmot & Lin, 2001) ισούται με:

$$m(x) = \frac{\sum_{j=0}^{a-1} (a-j) \cdot \frac{(b \cdot x)^j}{j!}}{b \cdot \sum_{j=0}^{a-1} \frac{(b \cdot x)^j}{j!}}, \quad x \geq 0$$

Γενικά για $a > 0$ δεν υπάρχει μια απλή μορφή για τον MRL. Ωστόσο όπως δείξαμε στο προηγούμενο κεφάλαιο για την βαθμίδα αποτυχίας της Γάμμα κατανομής, προκύπτει ότι εάν $a = 1$ τότε $m(x) = \frac{1}{b}$, $x \geq 0$ ενώ εάν $a < 1$ τότε η $F(x)$ ανήκει στην κλάση DFR και συνεπώς η $m(x)$ θα είναι αύξουσα συνάρτηση ως προς x . Επιπλέον εάν $a > 1$ τότε η $F(x)$ ανήκει στην κλάση IFR και συνεπώς η $m(x)$ θα είναι φθίνουσα συνάρτηση ως προς x .

$$\text{Επίσης } m(\infty) = \frac{1}{\mu(\infty)} = \frac{1}{b}.$$

Άρα εάν $a < 1$ (ή $a > 1$) τότε ο μέσος υπολειπόμενος χρόνος ζωής θα είναι αύξουσα συνάρτηση ως προς x (ή φθίνουσα συνάρτηση) από το $m(0) = \frac{a}{b}$ έως το όριο που παίρνει όταν το x τείνει στο άπειρο $m(\infty) = 1/b$.

3.2 Επιπρόσθετες Κλάσεις Κατανομών

Έχουμε ήδη εξετάσει τις δύο βασικές κλάσεις κατανομών IFR και DFR για αύξουσα και φθίνουσα βαθμίδα αποτυχίας και τις IMRL και DMRL για αύξοντα και φθίνοντα μέσο υπολειπόμενο χρόνο ζωής. Παρόλα αυτά έχει αρκετό ενδιαφέρον να δούμε και κάποιες επιπλέον κλάσεις οι οποίες είναι εξίσου σημαντικές.

Θα μελετήσουμε ωστόσο περιπτώσεις κλάσεων οι οποίες περιγράφουν κυρίως συνεχείς τυχαίες μεταβλητές. Ας τις δούμε λοιπόν αναλυτικά σε αυτή την ενότητα (Willmot & Lin, 2001) (Τσάκλα, 2014).

3.2.1 Οικογένειες κλάσης κατανομών NBU και NWU

Η πρώτη επιπλέον οικογένεια κλάσεων κατανομών που θα μελετήσουμε αντίστοιχα σε αυτή την υποενότητα είναι η NBU δηλαδή New Better than Used και η NWU δηλαδή New Worse than Used.

Ορισμός 3.3: Έστω η μη-αρνητική και συνεχής τυχαία μεταβλητή X , η οποία εκφράζει τον χρόνο ζωής ενός ατόμου ή ενός συστήματος, έχει συνάρτηση κατανομής $F(x)$ και συνάρτηση επιβίωσης $S(x)$ αντίστοιχα. Τότε η κατανομή $F(x)$ θα ανήκει στην κλάση NBU εάν ισχύει η ακόλουθη συνθήκη:

$$\Pr(X > x + y | X > x) \leq \Pr(X > y) \quad , \quad \text{για } x \geq 0 \text{ και } y \geq 0$$

Δηλαδή: $\Pr(T(x) > y) \leq \Pr(X > y)$ (η δεσμευμένη συνάρτηση επιβίωσης για ένα άτομο ή σύστημα ηλικίας x να είναι μικρότερη από την συνάρτηση επιβίωσης ενός νεογέννητου ατόμου ή νέου συστήματος).

Μια ισοδύναμη μορφή της παραπάνω σχέσης είναι:

$$S(x + y) \leq S(x) \cdot S(y), \quad \text{για κάθε } x \geq 0 \text{ και } y \geq 0$$

Αντίστοιχα η κατανομή $F(x)$ θα ανήκει στην κλάση NWU εάν:

$$\Pr(X > x + y | X > x) \geq \Pr(X > y), \quad \text{για } x \geq 0 \text{ και } y \geq 0$$

Δηλαδή: $\Pr(T(x) > y) \geq \Pr(X > y)$ (η δεσμευμένη συνάρτηση επιβίωσης για ένα άτομο ή σύστημα ηλικίας x να είναι μεγαλύτερη από την συνάρτηση επιβίωσης ενός νεογέννητου ατόμου ή νέου συστήματος)

Εναλλακτικά:

$$S(x + y) \geq S(x) \cdot S(y), \quad \text{για κάθε } x \geq 0 \text{ και } y \geq 0$$

Η ισότητα στις παραπάνω συνθήκες ισχύει μόνο όταν έχουμε εκθετική κατανομή λόγω της ιδιότητας έλλειψης μνήμης.

Επιπλέον προκύπτει ότι η κλάση IFR ανήκει στην κλάση κατανομών NBU ενώ η κλάση DFR ανήκει στην κλάση NWU.

3.2.2. Οικογένειες κλάσης κατανομών NBUE και NWUE

Μια δεύτερη οικογένεια κλάσης κατανομών, που θα εξετάσουμε στην συνέχεια ως επέκταση των προηγούμενων δύο που ήδη έχουμε δει, είναι η NBUE δηλαδή New Better than Used in Expectation και η NWUE δηλαδή New Worse than Used in Expectation.

Ορισμός 3.4: Έστω η μη-αρνητική και συνεχής τ.μ. X , η οποία εκφράζει τον χρόνο ζωής ενός ατόμου ή ενός μηχανήματος και έχει συνάρτηση κατανομής $F(x)$. Ο μέσος υπολειπόμενος χρόνος ζωής είναι ίσος με $m(x)$. Τότε θα ισχύει ότι η κατανομή $F(x)$ (εφόσον έχει πεπερασμένο μέσο μ) ανήκει στην κλάση NBUE εάν έχουμε την παρακάτω συνθήκη:

$$m(x) \leq m(0) \quad \text{δηλαδή εαν } E(T(X)) \leq E(X) \text{ για } \forall x \geq 0$$

Επιπλέον έχουμε ήδη δείξει ότι:

$$m(x) = \frac{E(X) \cdot \bar{F}_1(x)}{\bar{F}(x)}$$

Συνεπώς η παραπάνω σχέση μπορεί να εκφρασθεί ως $\bar{F}_1(x) \leq \bar{F}(x)$ όπου $\bar{F}_1(x)$ είναι η κατανομή ισορροπίας.

Μια ισοδύναμη ιδιότητα είναι (Willmot & Lin, 2001):

$$m(x) = \int_0^{\infty} \frac{S(x+t)}{S(x)} dt \leq \int_0^{\infty} S(t) dt = E(X) = m(0)$$

όπου $S(x) = 1 - F(x) = \bar{F}(x)$ είναι η συνάρτηση επιβίωσης ή δεξιά ουρά της τ.μ. X

Αντίστοιχα η κατανομή $F(x)$ (εφόσον έχει πεπερασμένο μέσο μ) ανήκει στην κλάση NWUE εάν έχουμε την παρακάτω συνθήκη:

$$m(x) \geq m(0) \quad \text{δηλαδή εαν } E(T(X)) \geq E(X)$$

ή ισοδύναμα:

$$m(x) = \int_0^{\infty} \frac{S(x+t)}{S(x)} dt \geq \int_0^{\infty} S(t) dt = E(X) = m(0) .$$

Οι παραπάνω σχέσεις εκφράζουν ότι ο μέσος υπολειπόμενος χρόνος ζωής για ένα άτομο ή ένα σύστημα ηλικίας x θα είναι μικρότερος (ή μεγαλύτερος αντίστοιχα) από τον μέσο χρόνο ζωής ενός νεογέννητου ατόμου ή ενός νέου συστήματος.

Επίσης προκύπτει ότι η κλάση NBU (η οποία προϋποθέτει ότι ο υπολειπόμενος χρόνος ζωής ατόμου ηλικίας x είναι μικρότερος από τον χρόνο ζωής ενός νεογέννητου ατόμου ή ενός νέου συστήματος) είναι υποσύνολο της κλάσης NBUE που αφορά τις μέσες τιμές των ποσοτήτων που αναφέραμε. Αντιστοίχως η κλάση NWU είναι υποσύνολο της κλάσης NWUE (Willmot & Lin, 2001).

3.2.3 Οικογένειες κλάσης κατανομών UBA και UWA

Στην προτελευταία αυτή υποενότητα θα εξετάσουμε μια επιπλέον οικογένεια κλάσης κατανομών η οποία είναι εξίσου ιδιαίτερη για την κατανόηση και την μελέτη των εννοιών που έχουμε εισαγάγει στα δύο αυτά κεφάλαια. Πρόκειται για την κλάση UBA που σημαίνει Used Better than aged και την κλάση UWA που σημαίνει Used Worse than aged.

Ορισμός 3.5: Έστω η μη-αρνητική και συνεχής τ.μ. X η οποία περιγράφει τον χρόνο ζωής ενός συστήματος ή ενός ατόμου και έχει συνάρτηση κατανομής $F(x)$. Τότε θα λέμε ότι η κατανομή $F(x)$ είναι UBA (ή UWA) εάν ισχύει ότι ο μέσος υπολειπόμενος χρόνος ζωής $m(x) = \int_0^{\infty} \frac{\bar{F}(x+t)}{\bar{F}(x)} dt$ ικανοποιεί τις συνθήκες:

$$0 < m(\infty) = \lim_{x \rightarrow \infty} m(x) < \infty \text{ και } \bar{F}(x+y) \geq (\leq) \bar{F}(x) \cdot e^{-y/m(\infty)} \text{ για } \forall x \geq 0 \text{ και } y \geq 0$$

Επιπλέον εάν η $F(x)$ είναι απόλυτα συνεχής με βαθμίδα αποτυχίας:

$$\mu(x) = -\frac{d}{dx} \ln \bar{F}(x),$$

τότε η κατανομή $F(x)$ είναι UBA (ή UWA) εάν $\mu(x) \leq (\geq) \mu(\infty)$ όπου $0 < \mu(\infty) < \infty$.

Αποδεικνύεται επίσης ότι αν μια κατανομή $F(x)$ ανήκει στην κλάση DMRL του φθίνοντα μέσου υπολειπόμενου χρόνου ζωής (ή στην κλάση IMRL του αύξοντα υπολειπόμενου χρόνου ζωής) με $m(\infty) > 0$ ($m(\infty) < \infty$) τότε θα ανήκει αντίστοιχα στην κλάση UBA (ή UWA).

3.2.4 Οικογένειες κλάσης κατανομών UBAE και UWAE

Επιπλέον αξίζει να δούμε και τις κλάσεις UBAE (used better than aged in expectation) και UWAE (used worse than aged in expectation) οι οποίες είναι μεγαλύτερες από τις κλάσεις UBA και UWA αντίστοιχα.

Ορισμός 3.6: Εάν η θετική και συνεχής τ.μ. X η οποία εκφράζει τον χρόνο ζωής ενός ατόμου ή ενός συστήματος έχει συνάρτηση κατανομής $F(x)$ τότε θα λέμε ότι η κατανομή αυτή ανήκει στην κλάση UBAE (ή UWAE) όταν ισχύει για τον μέσο υπολειπόμενο χρόνο ζωής ότι: $m(x) \geq (\leq) m(\infty)$ όπου $0 < m(\infty) < \infty$.

Επιπρόσθετα από την παρακάτω ανισότητα μπορεί εύκολα να δειχθεί ότι εάν η $F(x)$ είναι UBA (ή UWA) τότε η UBA (ή UWA) θα είναι υποσύνολο της κλάσης UBAE (ή UWAE) εφόσον έχουμε:

$$m(x) = \int_0^{\infty} \frac{\bar{F}(x+t)}{\bar{F}(x)} dt \geq (\leq) \int_0^{\infty} e^{-\frac{x}{m(\infty)}} dx = m(\infty)$$

3.2.5 Οικογένειες κλάσης κατανομών HNBUE και HNWUE

Τέλος υπάρχει και μια οικογένεια κλάσης η οποία περιλαμβάνει όλες τις κλάσεις που αναφέραμε προηγουμένως και θα την δούμε περιληπτικά (Τσάκλα, 2014). Αφορά τις κλάσεις Harmonic new better than used in expectation και Harmonic new worse than used in expectation σύμφωνα με τις οποίες ο χρόνος ζωής μια μονάδας θα είναι στοχαστικά μικρότερος(ή μεγαλύτερος) από τον χρόνο ζωής μιας μονάδας που ακολουθεί εκθετική κατανομή με παράμετρο $\mu = E(X)$.

Ορισμός 3.7: Έστω η τ.μ. X η οποία εκφράζει τον χρόνο ζωής μιας μονάδας με συνάρτηση επιβίωσης $S(x)$ και μια άλλη δεύτερη τ.μ. η οποία ακολουθεί την εκθετική κατανομή με παράμετρο $\mu = E(X)$ και έχει συνάρτηση επιβίωσης $S_\mu(x) = e^{-x/\mu}$. Τότε εάν $S(x) \leq S_\mu(x)$ θα λέμε ότι η κατανομή F ανήκει στην HNBUE (ή HNWUE) εφόσον:

$$\int_t^\infty S(x)dx \leq (\geq) \int_t^\infty S_\mu(x)dx \quad \text{ή} \quad \int_t^\infty S(x)dx \leq (\geq) \mu \cdot e^{-\frac{t}{\mu}}, \quad t \geq 0$$

3.2.6 Διάταξη μεταξύ των κλάσεων

Όλες οι παραπάνω κλάσεις ωστόσο συνδέονται μεταξύ τους μέσω διατάξεων καθώς κάποιες αποτελούν υποσύνολα άλλων μεγαλύτερων κλάσεων (Barlow and Proschan, 1975).

Ας δούμε λοιπόν μέσω του παρακάτω σχήματος την διάταξη των υποσυνόλων:

α) $IFR \subseteq NBU \subseteq NBUE \subseteq HNBUE$

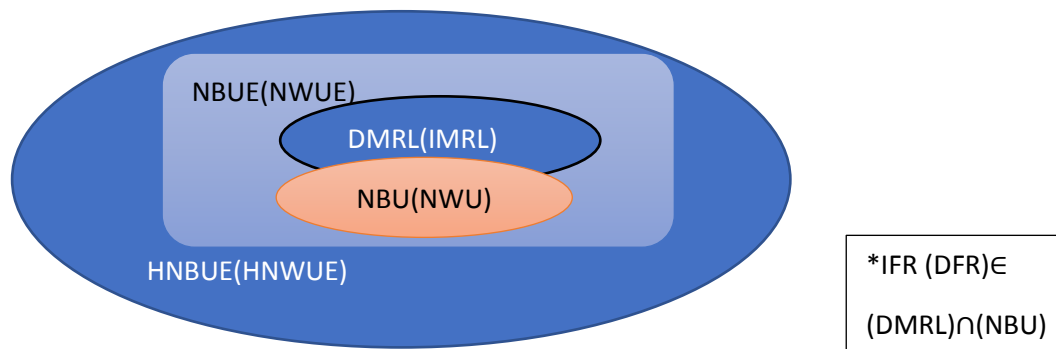
ενώ επίσης ισχύει ότι $IFR \subseteq DMRL$

β) $DFR \subseteq NWU \subseteq NWUE \subseteq HNWUE$

ενώ επίσης ισχύει ότι $DFR \subseteq IMRL$

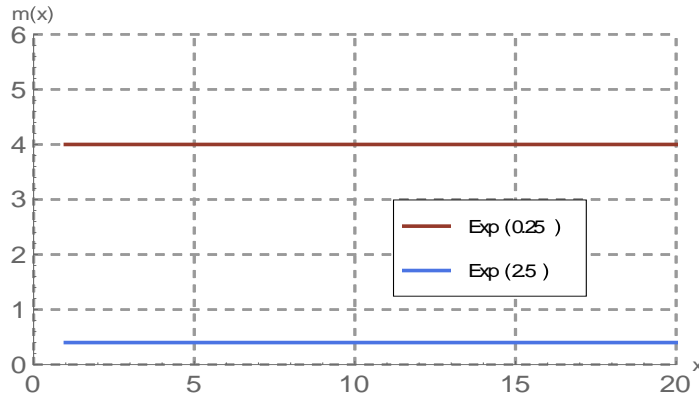
Επομένως εάν θέλουμε να δείξουμε ότι μια κατανομή F ανήκει σε μια από τις κλάσεις που ήδη αναφέραμε DMRL(IMRL) ή NBU(NWU) ή NBUE(NWUE) ή HNBUE(HNWUE) τότε αρκεί να αποδείξουμε ότι η F ανήκει στην μικρότερη κλάση αξιοπιστίας IFR(DFR) γεγονός το οποίο συνεπάγεται ότι θα εντάσσεται και σε όλες τις μεγαλύτερες κλάσεις κατανομών που ήδη έχουμε μελετήσει.

Σχήμα 3.2



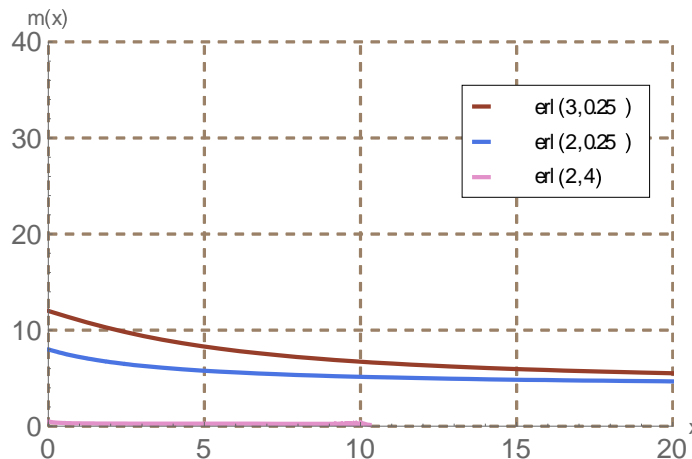
Στο τελευταίο κομμάτι αυτής της ενότητας θα δούμε τα παρακάτω διαγράμματα όπου απεικονίζεται η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής για τρεις κατανομές:

Διάγραμμα 3.1: Μέσος υπολειπόμενος χρόνος ζωής για την εκθετική κατανομή με παράμετρο $\lambda = 0.25$ (κόκκινη γραμμή) και παράμετρο $\lambda = 2.5$ (μπλε γραμμή)



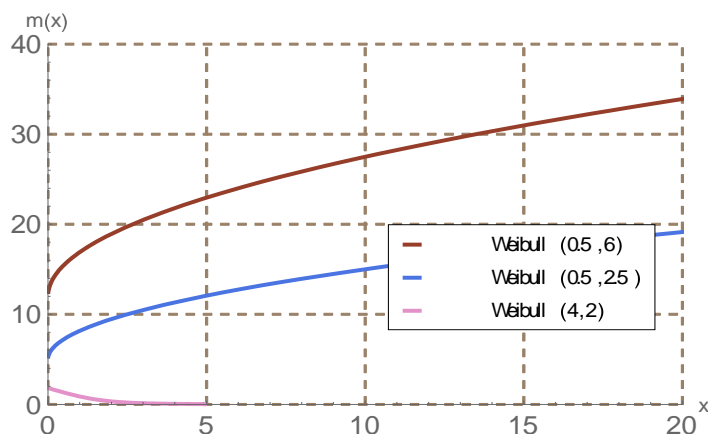
Όπως και η συνάρτηση της βαθμίδας αποτυχίας έτσι και η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής για την περίπτωση των εκθετικών χρόνων ζωής θα είναι μια σταθερή συνάρτηση ως προς x . Ωστόσο σε αντίθεση με την βαθμίδα αποτυχίας που μελετήσαμε στο προηγούμενο κεφάλαιο, εδώ όσο μεγαλύτερη είναι η παράμετρος λ ($0.25 < 2.5$) τόσο μικρότερος θα είναι ο μέσος υπολειπόμενος χρόνος ζωής.

Διάγραμμα 3.2: Μέσος υπολειπόμενος χρόνος ζωής για την κατανομή Erlang με παραμέτρους $(3, 0.25)$, $(2, 0.25)$ και $(2, 4)$



Για την κατανομή Erlang με παράμετρο a ακέραιο αριθμό, παρατηρήσαμε προηγουμένως ότι η συνάρτηση της βαθμίδας αποτυχίας είναι πάντοτε αύξουσα. Στο παραπάνω διάγραμμα ωστόσο επιβεβαιώνουμε την αρνητική σχέση μεταξύ μέσου υπολειπόμενου χρόνου ζωής και βαθμίδας αποτυχίας καθώς παρατηρούμε ότι για $a > 1$ (εδώ 3, 2 και 2), η συνάρτηση του MRL είναι πάντοτε φθίνουσα. Μάλιστα όσο μεγαλύτερη είναι η παράμετρος b ($0.25 < 4$) τόσο η καμπύλη πλησιάζει τον άξονα x .

Διάγραμμα 3.3: Μέσος υπολειπόμενος χρόνος ζωής για την κατανομή Weibull με παραμέτρους (0.5,6), (0.5,2.5) και (4,2)



Τέλος για την κατανομή Weibull όπως ακριβώς ισχύει και για την κατανομή Γάμμα για παράμετρο μορφής μεγαλύτερη της μονάδας η βαθμίδα αποτυχίας είναι αύξουσα ενώ για παράμετρο μορφής μικρότερη της μονάδας η βαθμίδα αποτυχίας είναι φθίνουσα (ανεξάρτητα από τις τιμές που παίρνει η παράμετρος κλίμακας). Έτσι λοιπόν από το παραπάνω διάγραμμα φαίνεται η αρνητική συσχέτιση μεταξύ των συναρτήσεων βαθμίδας αποτυχίας και μέσου υπολειπόμενου χρόνου ζωής. Συγκεκριμένα όταν έχουμε παράμετρο μορφής μεγαλύτερη της μονάδας (στη περίπτωση αυτή ίση με 4), ο μέσος υπολειπόμενος χρόνος ζωής είναι φθίνουσα συνάρτηση ενώ όταν έχουμε παράμετρο μορφής μικρότερη της μονάδας (στη περίπτωση αυτή ίση με 0.5), ο μέσος υπολειπόμενος χρόνος ζωής είναι αύξουσα συνάρτηση ως προς x . Επιπλέον για παράμετρο μορφής μικρότερη της μονάδας, όσο η παράμετρος κλίμακας μικραίνει ($2 < 2.5 < 6$) τόσο η καμπύλη του MRL θα πλησιάζει τον άξονα x .

ΚΕΦΑΛΑΙΟ 4

ΑΣΥΜΜΕΤΡΙΑ ΓΙΑ ΜΕΙΞΕΙΣ ΚΑΙ ΣΥΝΕΛΙΞΕΙΣ ΓΑΜΜΑ ΚΑΤΑΝΟΜΩΝ

Μια από τις βασικότερες ιδιότητες για την εικόνα μιας κατανομής στην αναλογιστική επιστήμη αλλά και την στατιστική ανάλυση γενικότερα είναι η ασυμμετρία ή λοξότητα (skewness). Η ιδιότητα αυτή μας δίνει σημαντικές πληροφορίες σχετικά με το σχήμα μιας κατανομής και είναι ιδιαίτερα χρήσιμη όσο η βαθμίδα αποτυχίας αλλά και ο μέσος υπολειπόμενος χρόνος ζωής. Ωστόσο θα εξετάσουμε πολύ συνοπτικά και μια επιπλέον έννοια η οποία είναι η κύρτωση (kurtosis).

Στο παρόν κεφάλαιο θα υπολογίσουμε ποσότητες όπως είναι ο συντελεστής ασυμμετρίας αλλά και ο συντελεστής κύρτωσης για διάφορες περιπτώσεις κατανομών με την βοήθεια του υπολογιστικού προγράμματος Mathematica. Παράλληλα θα δώσουμε έμφαση στην μελέτη της ασυμμετρίας για τις μειξείς και συνελίξεις Γάμμα κατανομών, οι οποίες αποτελούν το βασικό αντικείμενο μελέτης της παρούσας διατριβής και είναι οι κατανομές που χρησιμοποιούνται ευρέως στην θεωρία κινδύνων ως μοντέλα ατομικών ζημιών. Θα προσπαθήσουμε ξεκινώντας από την μείξη ή συνέλιξη 2 Γάμμα κατανομών, να οδηγηθούμε σε ένα γενικό συμπέρασμα για την ασυμμετρία αλλά και την κύρτωση n Γάμμα κατανομών.

Ωστόσο αρχικά έχει σημασία να δούμε το βασικό γνωστικό περιβάλλον για τις έννοιες της ασυμμετρίας και της κύρτωσης πριν προχωρήσουμε σε οποιαδήποτε περαιτέρω ανάλυση.

4.1 Ασυμμετρία και μέτρα ασυμμετρίας

Ένα βασικό μέτρο στατιστικής ανάλυσης είναι ο συντελεστής ασυμμετρίας ο οποίος μετράει κατά πόσο μια κατανομή αποκλίνει από μια απολύτως συμμετρική κατανομή (skewness=0) όπως για παράδειγμα είναι η κανονική κατανομή (normal distribution) ή η ομοιόμορφη (uniform distribution). Όσον αφορά την εικόνα μιας κατανομής είναι προφανές ότι οι ακραίες τιμές παίζουν σημαντικό ρόλο. Αν σε μια κατανομή δεν υπάρχουν ακραίες τιμές τότε η κατανομή αυτή συνήθως δεν παρουσιάζει μεγάλη ασυμμετρία.

Γενικά υπάρχουν δύο είδη ασυμμετρίας τα οποία θα δούμε αναλυτικά παρακάτω (Πανάρετος & Ξεκαλάκη, 2003):

A) Αν υπάρχουν ακραίες τιμές στα δεξιά του κύριου όγκου των τιμών ενός πληθυσμού ή ενός δείγματος τότε θα λέμε ότι η κατανομή των δεδομένων έχει *δεξιά ασυμμετρία* (ή λοξότητα) ή ότι έχει δεξιά κλίση (skewed to the right ή positively skewed). Ονομάζεται αλλιώς και θετική ασυμμετρία κατά την οποία παρατηρείται μεγάλη ουρά της κατανομής προς τα δεξιά. Όταν η κατανομή παρουσιάζει θετική ασυμμετρία τότε ισχύει $\mu > \delta > M_o$ όπου μ είναι η μέση τιμή ($E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$), δ είναι η διάμεσος και M_o η κορυφή της κατανομής αντίστοιχα.

Σχήμα 4.1

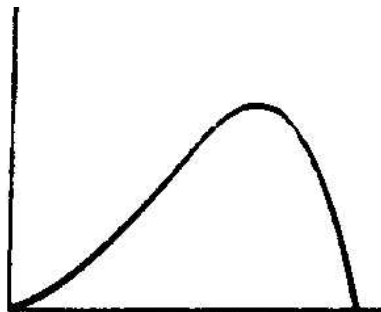


Θετική ασυμμετρία

(Πηγή: Wikipedia)

B) Αν υπάρχουν ακραίες τιμές στα αριστερά του κύριου όγκου των τιμών ενός πληθυσμού ή ενός δείγματος τότε θα λέμε ότι η κατανομή των δεδομένων έχει *αριστερή ασυμμετρία* (ή λοξότητα) ή ότι έχει αριστερή κλίση (skewed to the left ή negatively skewed). Ονομάζεται αλλιώς και αρνητική ασυμμετρία κατά την οποία παρατηρείται μεγάλη ουρά της κατανομής προς τα αριστερά. Όταν η κατανομή παρουσιάζει αριστερή ασυμμετρία τότε ισχύει $\mu < \delta < M_o$ όπου μ είναι η μέση τιμή ($E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$), δ είναι η διάμεσος και M_o η κορυφή της κατανομής αντίστοιχα.

Σχήμα 4.2



Αρνητική ασυμμετρία

(Πηγή: Wikipedia)

Γενικά έχουν προταθεί διάφορα μέτρα ασυμμετρίας (measures of skewness). Ωστόσο είναι γενικά αποδεκτό (Arnold and Groeneveld, 1995) ότι κάθε μέτρο ασυμμετρίας γ θα πρέπει να ικανοποιεί κάποιες συγκεκριμένες προϋποθέσεις. Για να αναφέρουμε τις παρακάτω προϋποθέσεις χρησιμοποιούμε τους συμβολισμούς $\gamma(X)$ και $\gamma(F)$ εναλλάξ, όπου X μια τυχαία μεταβλητή με συνάρτηση κατανομής $F(x)$.

Θα πρέπει να ισχύει (Marshall & Olkin, 2007) :

α) $\gamma(X) = \gamma(a \cdot X + b)$ για όλα τα $a > 0$ και όλα τα b .

β) $\gamma(X) = -\gamma(-X)$

γ) Εάν $F \leq_c G$, τότε $\gamma(F) \leq \gamma(G)$

Είναι προφανές ότι εάν $\gamma(F) = 0$ τότε η κατανομή F είναι συμμετρική.

Σχετικά με την τελευταία προϋπόθεση (γ) η οποία αφορά την διάταξη κυρτού μετασχηματισμού θα διατυπώσουμε τον παρακάτω ορισμό (Marshall & Olkin, 2007) :

Ορισμός 4.1 : Έστω η τυχαία μεταβλητή X με συνάρτηση κατανομής F και η τυχαία μεταβλητή Y με συνάρτηση κατανομής G αντίστοιχα. Τότε η X θα λέμε ότι είναι μικρότερη ως προς την διάταξη κυρτού μετασχηματισμού από την Y (smaller in the convex transform order) δηλαδή $X \leq_c Y$, εάν η $G^{-1}(F(x))$ είναι κυρτή ως προς x στο στήριγμα της X . Η έννοια $F \leq_c G$ είναι ισοδύναμη με την $X \leq_c Y$.

Όπως έχουμε ήδη αναφέρει ένας βασικός τρόπος ανάλυσης της ασυμμετρίας μιας κατανομής είναι η χρήση κάποιων άμεσων αριθμητικών μέτρων. Τέτοια μέτρα σχετίζονται κυρίως με την διασπορά των δεδομένων.

Ένα μέτρο το οποίο είχε αρχικά προταθεί από τον Sir Francis Galton για την μέτρηση της ασυμμετρίας ήταν το εξής:

$$\frac{Q_1 + Q_3 - 2 \cdot \delta}{IQR} \quad (4.1.1)$$

Όπου $Q_1 =$ κάτω τεταρτημόριο , $Q_3 =$ άνω τεταρτημόριο και $\delta =$ διάμεσος.

Ο παρονομαστής του παραπάνω κλάσματος ονομάζεται και εναλλακτικά ως ενδοτεταρτημοριακό εύρος ή $IQR =$ άνω τεταρτημόριο - κάτω τεταρτημόριο.

Ο Karl Pearson το 1895, πρότεινε ως μέτρο ασυμμετρίας την συνάρτηση:

$$\frac{\text{μέσος-επικρατούσα τιμή}}{\text{τυπική απόκλιση}} \quad (4.1.2)$$

που όμως έχει το μειονέκτημα ότι προϋποθέτει γνώση της κορυφής (mode) της κατανομής η οποία δεν είναι πάντα εύκολο να προσδιοριστεί. Έτσι λόγω αυτής της δυσκολίας και δεδομένου ότι για μια ελαφρά ασύμμετρη κατανομή ισχύει κατά προσέγγιση η σχέση:

$$\text{μέσος-επικρατούσα τιμή} = 3 \cdot (\text{μέσος-διάμεσος}) ,$$

συχνά χρησιμοποιείται ο συντελεστής:

$$\frac{3 \cdot (\text{μέσος-διάμεσος})}{\text{τυπική απόκλιση}}$$

Η σχέση (4.1.2) δίνει τιμή ίση με το μηδέν όταν πρόκειται για συμμετρικές κατανομές και δεν μεταβάλλεται όταν η παράμετρος κλίμακας σε διάφορες κατανομές αλλάζει.

Για μια τυχαία μεταβλητή X με συνάρτηση κατανομής F , ένα περισσότερο γνωστό μέτρο ασυμμετρίας του Pearson το οποίο είχε προταθεί και από τους Edgeworth (1904) και Charlier (1906) είναι:

$$\gamma_1 = \frac{E(X-\mu)^3}{\sigma^3} = \frac{E[X-E(X)]^3}{\{E[X-E(X)]^2\}^{3/2}} \quad (4.1.3)$$

για την οποία ποσότητα όμως απαιτείται η ύπαρξη της τρίτης ροπής. Η έννοια της ροπής προέρχεται από τον κλάδο της Μηχανικής και αναφέρεται στη μέτρηση της τάσης μιας δύναμης να παράγει περιστροφή. Η τάση αυτή εξαρτάται από το μέγεθος της δύναμης και από την απόσταση του σημείου στο οποίο ασκείται η δύναμη από κάποια αρχή (Πανάρετος & Ξεκαλάκη, 2003). Αυτό το μέτρο ασυμμετρίας είναι επίσης ίσο με μηδέν όταν η κατανομή είναι συμμετρική.

Η συνάρτηση (4.1.2) του Pearson ικανοποιεί τις προϋποθέσεις α - γ που πρέπει να πληροί ένα μέτρο ασυμμετρίας. Ωστόσο η συνάρτηση (4.1.1) του Francis Galton δεν ικανοποιεί την συνθήκη (γ) όπως αποδείχθηκε από τον Van Zwet (1964).

Το τελευταίο μέτρο (4.1.3) χρησιμοποιείται πλέον πιο συχνά και είναι γνωστό ως συντελεστής ασυμμετρίας (coefficient of skewness) ενώ παράλληλα ορίζεται μέσω της στατιστικής συνάρτησης μ_3/σ^3 . Όπου για τις συνεχείς κατανομές τις οποίες θα μελετήσουμε στο δοθέν κεφάλαιο έχουμε ότι: $\mu_3 = E[X - E(X)]^3 = \int_{-\infty}^{\infty} (x - \mu)^3 \cdot f(x) dx$,

με $\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$ να είναι ο μέσος της κατανομής ή η ροπή πρώτης τάξης γύρω από το μηδέν.

Γενικά ως ροπή r τάξης γύρω από το μηδέν ορίζεται η ποσότητα:

$$E(X^r) = \int_{-\infty}^{\infty} x^r \cdot f(x) dx$$

Είναι προφανές από τον ορισμό του συντελεστή μέσω της μ_3 ότι μεγάλες τιμές των αποκλίσεων των δεδομένων από την μέση τιμή της κατανομής είναι καθοριστικές για την τιμή του αριθμητή. Αν λοιπόν οι μεγάλες θετικές αποκλίσεις είναι επικρατέστερες τότε η ροπή μ_3 θα είναι θετική ενώ αντίστοιχα αν επικρατούν οι μεγάλες αρνητικές αποκλίσεις τότε η ροπή μ_3 θα είναι αρνητική. Επειδή οι μεγάλες αποκλίσεις συνδέονται με την μακριά ουρά μιας κατανομής βλέπουμε ότι η μ_3 και κατά συνέπεια ο συντελεστής ασυμμετρίας θα έχει θετική ή αρνητική τιμή ανάλογα με τον εάν η κατεύθυνση της ασυμμετρίας είναι δεξιά (θετική) ή αριστερή (αρνητική). Αν τα δεδομένα κατανέμονται συμμετρικά γύρω από την μέση τιμή τότε

η τρίτη κεντρική ροπή μ_3 και κατά συνέπεια ο συντελεστής ασυμμετρίας θα είναι μηδέν. Ας δούμε παρακάτω συνοπτικά τις τιμές που μπορεί να πάρει ο συντελεστής ασυμμετρίας:

- $\frac{\mu_3}{\sigma^3} > 0$ θετική ή δεξιά ασυμμετρία
- $\frac{\mu_3}{\sigma^3} = 0$ συμμετρική κατανομή
- $\frac{\mu_3}{\sigma^3} < 0$ αρνητική ή αριστερή ασυμμετρία

4.2 Κύρτωση και μέτρα κύρτωσης

Η κύρτωση αποτελεί και αυτή ένα εξίσου σημαντικό στοιχείο στατιστικής ανάλυσης της μορφής μια κατανομής. Ένα βασικό μέτρο για την κύρτωση είναι ο συντελεστής κύρτωσης (coefficient of kurtosis) ο οποίος μας δείχνει τον βαθμό συγκέντρωσης των δεδομένων γύρω από τον μέσο και τα άκρα της κατανομής και χρησιμοποιείται ως μέτρο της οξύτητας της κορυφής της κατανομής των δεδομένων.

Για μια τυχαία μεταβλητή X με συνάρτηση κατανομής F , ο συντελεστής κύρτωσης του Pearson ο οποίος εκφράζει τον βαθμό κύρτωσης της κατανομής ορίζεται από τον τύπο:

$$\gamma_2 = \frac{E(X-\mu)^4}{\sigma^4} = \frac{E[X-E(X)]^4}{\{E[X-E(X)]^2\}^{4/2}}$$

όπου απαιτείται ωστόσο η ύπαρξη της τέταρτης ροπής.

Παράλληλα μπορεί να γραφθεί ως το πηλίκο μ_4/σ^4 , όπου μ_4 είναι η τέταρτη κεντρική ροπή και σ η τυπική απόκλιση της κατανομής. Η τέταρτη κεντρική ροπή της κατανομής ορίζεται ως:

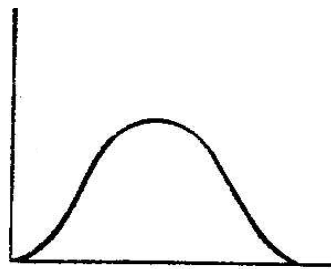
$$\mu_4 = E[X - E(X)]^4 = \int_{-\infty}^{\infty} (x - \mu)^4 \cdot f(x) dx$$

με $\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$ να είναι ο μέσος της κατανομής στην περίπτωση όπου έχουμε συνεχείς κατανομές οι οποίες αποτελούν το βασικό κομμάτι μελέτης για το κεφάλαιο αυτό.

Ας εξετάσουμε ωστόσο τις τιμές που μπορεί να πάρει ο συντελεστής κύρτωσης για μια κατανομή:

- Η τιμή του συντελεστή κύρτωσης είναι ίση με 3 (kurtosis=3) όταν μια κατανομή είναι κανονική. Στην περίπτωση αυτή οι κατανομές των δεδομένων λέγονται μεσόκυρτες (mesokurtic) και λέμε ότι έχουν μέτρια κύρτωση.

Σχήμα 4.3

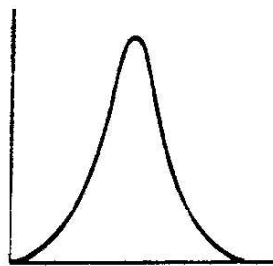


Μεσόκυρτη

(Πηγή: Wikipedia)

- Η τιμή του συντελεστή κύρτωσης είναι μεγαλύτερη του 3 ($kurtosis > 3$) όταν μια κατανομή εμφανίζει σχετικά μεγάλη μέγιστη συχνότητα (κορυφή) και επομένως μεγάλη συγκέντρωση τιμών γύρω από τον μέσο. Στην περίπτωση αυτή οι κατανομές των δεδομένων λέγονται λεπτόκυρτες (leptokurtic) και έχουν οξεία κύρτωση. Παράλληλα είναι οικογένεια κατανομών με παχιές ουρές.

Σχήμα 4.4

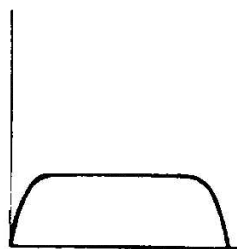


Λεπτόκυρτη

(Πηγή: Wikipedia)

- Η τιμή του συντελεστή κύρτωσης είναι μικρότερη του 3 ($kurtosis < 3$) όταν μια κατανομή εμφανίζει σχετικά μικρή μέγιστη συχνότητα (κορυφή) και επομένως μικρή συγκέντρωση τιμών γύρω από τον μέσο. Στην περίπτωση αυτή οι κατανομές των δεδομένων λέγονται πλατύκυρτες (platykurtic) και έχουν ελαφριά κύρτωση. Αποτελούν μια οικογένεια κατανομών με λεπτές ουρές.

Σχήμα 4.5



Πλατύκυρτη

(Πηγή: Wikipedia)

4.3 Ασυμμετρία και κύρτωση για διάφορες κατανομές

4.3.1 Εκθετική κατανομή

Έστω η συνεχής τυχαία μεταβλητή X η οποία ακολουθεί την εκθετική κατανομή με συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

και συνάρτηση κατανομής $F(x) = 1 - e^{-\lambda \cdot x}$. Από τον ορισμό προκύπτουν άμεσα ότι η μέση τιμή είναι ίση με $\mu = E(X) = \frac{1}{\lambda}$ και η διακύμανση ισούται με $\sigma^2 = Var(X) = \frac{1}{\lambda^2}$.

Για τον συντελεστή ασυμμετρίας του Pearson έχουμε ότι:

$$\begin{aligned} \gamma_1 &= \frac{E(X - \mu)^3}{\sigma^3} = \frac{E(X^3) - 3 \cdot \mu \cdot E(X^2) + 3 \cdot \mu^2 \cdot E(X) - \mu^3}{\sigma^3} = \\ &= \frac{E(X^3) - 3 \cdot \mu \cdot (E(X^2) - \mu \cdot E(X)) - \mu^3}{\sigma^3} = \\ &= \frac{E(X^3) - 3 \cdot \mu \cdot \sigma^2 - \mu^3}{\sigma^3} \end{aligned} \quad (4.3.1)$$

Με τη βοήθεια του Mathematica ορίζουμε την συνάρτηση πυκνότητας πιθανότητας ως: $f[x_]:= l * Exp[-l * x]$ και υπολογίζουμε την τρίτη ροπή γύρω από το μηδέν ως εξής:

$$m_3 = Integrate[(x^3) * f[x], {x, 0, Infinity}]$$

η οποία ισούται με $\frac{6}{\lambda^3}$.

Αντικαθιστώντας στην σχέση του συντελεστή ασυμμετρίας τις αντίστοιχες ποσότητες έχουμε ότι :

$$\gamma_1 = \frac{\frac{6}{\lambda^3} - 3 \cdot \frac{1}{\lambda} \cdot \frac{1}{\lambda^2} \cdot \frac{1}{\lambda^3}}{\frac{1}{\lambda^3}} = \frac{\frac{6}{\lambda^3} - \frac{4}{\lambda^3}}{\frac{1}{\lambda^3}} = 2$$

Παρατηρούμε ότι ο συντελεστής ασυμμετρίας είναι πάντοτε σταθερός και ίσος με 2, χωρίς να εξαρτάται από την παράμετρο λ της εκθετικής κατανομής.

Συνεπώς η εκθετική κατανομή εμφανίζει δεξιά (ή θετική) ασυμμετρία.

Ωστόσο ας μελετήσουμε και την τιμή που μπορεί να πάρει ο συντελεστής κύρτωσης αντίστοιχα χρησιμοποιώντας την τέταρτη κεντρική ροπή.

Για τον συντελεστή κύρτωσης ισχύει ότι ισούται με:

$$\gamma_2 = \frac{E(X-\mu)^4}{\sigma^4} = \frac{E[X-E(X)]^4}{\{E[X-E(X)]^2\}^{4/2}} \quad (4.3.2)$$

Γνωρίζουμε ότι $\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}$ και συνεπώς $\sigma^4 = \frac{1}{\lambda^4}$.

Επιπλέον για την τέταρτη κεντρική ροπή ισχύει: $E[X - E(X)]^4 = \int_0^{\infty} (x - \mu)^4 \cdot f(x) dx$

Με την βοήθεια του Mathematica για $\mu = \frac{1}{\lambda}$ υπολογίζουμε την τέταρτη κεντρική ροπή ως εξής:

$$\mu_4 = \text{Integrate}[(x - \mu)^4 * f[x], \{x, 0, Infinity\}]$$

για την οποία ισχύει ότι είναι ίση με $9/\lambda^4$.

Συνεπώς στον τύπο για τον συντελεστή κύρτωσης έχουμε ότι $\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{\frac{9}{\lambda^4}}{\frac{1}{\lambda^4}} = 9$.

Είναι μια ποσότητα σταθερή και ανεξάρτητη από την μεταβλητή λ της εκθετικής κατανομής. Επιπλέον ο συντελεστής κύρτωσης παίρνει τιμή μεγαλύτερη του 3, επομένως πρόκειται για μια λεπτόκυρτη κατανομή. Υπάρχει μεγάλη συγκέντρωση των δεδομένων γύρω από τον μέσο και έχει οξεία κύρτωση.

4.3.2 Κατανομή Pareto

Έστω η συνεχής τυχαία μεταβλητή X η οποία ακολουθεί την κατανομή Pareto με συνάρτηση πυκνότητας πιθανότητας ίση με:

$$f(x) = \frac{\alpha \cdot \lambda^\alpha}{(\lambda + x)^{\alpha+1}} \quad \text{με } \alpha, \lambda > 0 \text{ και } x > 0$$

όπου λ είναι η παράμετρος κλίμακας και α η παράμετρος σχήματος.

Επιπλέον η συνάρτηση κατανομής είναι ίση με: $F(x) = 1 - \left(\frac{\lambda}{\lambda+x}\right)^\alpha$

Για τις ροπές k -τάξης (Δημάκη, 2013) ισχύει η σχέση:

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

Ωστόσο για να ορίζεται η ροπή k -τάξης ($E(X^k) < \infty$) θα πρέπει $\alpha > k$.

Για $k = 1$ έχουμε την μέση τιμή ίση με: $E(X) = \frac{\lambda}{\alpha-1}$ με $\alpha > 1$

Για $k = 2$ έχουμε την δεύτερη ροπή ίση με: $E(X^2) = \frac{\lambda^2 \cdot 2!}{(\alpha-1) \cdot (\alpha-2)}$ με $\alpha > 2$

Για $k = 3$ έχουμε την τρίτη ροπή ίση με:

$$E(X^3) = \frac{\lambda^3 \cdot 3!}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} = \frac{6 \cdot \lambda^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} \text{ με } \alpha > 3$$

Επιπλέον η τυπική απόκλιση για την κατανομή Pareto θα είναι ίση με:

$$\begin{aligned} \sigma &= \sqrt{E(X^2) - (E(X))^2} = \sqrt{\frac{2 \cdot \lambda^2}{(\alpha - 1)(\alpha - 2)} - \frac{\lambda^2}{(\alpha - 1)^2}} = \sqrt{\frac{\lambda^2 \cdot (2 \cdot (\alpha - 1) - (\alpha - 2))}{(\alpha - 1)^2 \cdot (\alpha - 2)}} \\ &= \sqrt{\frac{\alpha \cdot \lambda^2}{(\alpha - 1)^2 \cdot (\alpha - 2)}} \end{aligned}$$

Για τον συντελεστή ασυμμετρίας αν αντικαταστήσουμε στην σχέση (4.3.1) τις ποσότητες που υπολογίσαμε έχουμε ότι:

$$\gamma_1 = \frac{\frac{6 \cdot \lambda^3}{(\alpha - 1)(\alpha - 2)(\alpha - 3)} - 3 \cdot \frac{\lambda}{\alpha - 1} \cdot \frac{\alpha \cdot \lambda^2}{(\alpha - 1)^2 \cdot (\alpha - 2)} - \frac{\lambda^3}{(\alpha - 1)^3}}{\sqrt{\frac{\alpha \cdot \lambda^2}{(\alpha - 1)^2 \cdot (\alpha - 2)} \cdot \frac{\alpha \cdot \lambda^2}{(\alpha - 1)^2 \cdot (\alpha - 2)}}$$

Υστερα από πράξεις με τη βοήθεια του Mathematica προκύπτει η εξής σχέση για τον συντελεστή ασυμμετρίας:

$$\gamma_1 = \frac{2 \cdot (\alpha + 1)}{\alpha - 3} \cdot \sqrt{\frac{\alpha - 2}{\alpha}}$$

Ο συντελεστής ορίζεται μόνο για τιμές του α μεγαλύτερες του 3 ($\alpha > 3$) ενώ παράλληλα δεν εξαρτάται από την παράμετρο κλίμακας λ . Στη περίπτωση όμως αυτή όπου $\alpha > 3$, η παραπάνω ποσότητα θα είναι πάντα θετική με αποτέλεσμα η κατανομή να έχει δεξιά (ή θετική) ασυμμετρία και ακραίες τιμές στα δεξιά του κύριου όγκου των δεδομένων.

Στη συνέχεια ας εξετάσουμε και την τιμή που δύναται να πάρει ο συντελεστής κύρτωσης με τον υπολογισμό της τέταρτης κεντρικής ροπής.

Για την τέταρτη κεντρική ροπή ισχύει: $E[X - E(X)]^4 = \int_0^{\infty} (x - \mu)^4 \cdot f(x) dx$

Με την βοήθεια του Mathematica για $\mu = \frac{\lambda}{\alpha - 1}$ με $\alpha > 1$ υπολογίζουμε την τέταρτη κεντρική ροπή ως εξής:

$$\mu_4 = \text{Integrate}[(x - \mu)^4 * f[x], \{x, 0, \text{Infinity}\}]$$

για την οποία ισχύει ότι είναι ίση με:

$$\frac{\lambda^4 \cdot 3 \cdot \alpha \cdot (2 + \alpha + 3 \cdot \alpha^2)}{(\alpha - 1)^4 \cdot (\alpha^3 - 9 \cdot \alpha^2 + 26 \cdot \alpha - 24)}$$

Έχοντας υπολογίσει παράλληλα την τυπική απόκλιση και υψώνοντας την εις την τέταρτη δύναμη προκύπτει η σ^4 .

Επομένως έχουμε το πηλίκο:

$$\frac{\mu_4}{\sigma^4} = \frac{6 \cdot (\alpha^3 + \alpha^2 - 6 \cdot \alpha - 2)}{\alpha(\alpha - 3)(\alpha - 4)} + 3,$$

το οποίο ορίζεται αντίστοιχα για $\alpha > 4$.

4.3.3 Κατανομή Γάμμα

Έστω η συνεχής τυχαία μεταβλητή X η οποία ακολουθεί την κατανομή Γάμμα με συνάρτηση πυκνότητας πιθανότητας ίση με:

$$f(x) = \begin{cases} \frac{b^a \cdot x^{a-1} \cdot e^{-b \cdot x}}{\Gamma(a)}, & \text{για } x \geq 0 \\ 0, & \text{για } x < 0 \end{cases}$$

όπου b είναι η παράμετρος κλίμακας και a η παράμετρος σχήματος με $a, b > 0$ και $a, b \in \mathcal{R}$. Παράλληλα ορίζουμε ως συνάρτηση Γάμμα ($\Gamma: (0, \infty) \mapsto \mathcal{R}$) το ολοκλήρωμα:

$$\Gamma(a) = \int_0^{\infty} t^{a-1} \cdot e^{-t} dt \quad (4.3.3)$$

Η μέση τιμή γνωρίζουμε ότι είναι ίση με $\mu = E(X) = \frac{a}{b}$ και η διακύμανση ισούται με $\sigma^2 = \text{Var}(X) = \frac{a}{b^2}$.

Στο Mathematica ορίζουμε την συνάρτηση πυκνότητας πιθανότητας ως:

$$f[x_] := (b^a) * (x^(a - 1)) * (Exp[-b * x])/Gamma[a]$$

και υπολογίζουμε την τρίτη ροπή γύρω από το μηδέν ως εξής:

$$m_3 = \text{Integrate}[(x^3) * f[x], \{x, 0, \text{Infinity}\}]$$

η οποία ισούται με: $m_3 = \frac{a(a+1)(a+2)}{b^3} = \frac{a^3 + 3 \cdot a^2 + 2 \cdot a}{b^3}$

Αντικαθιστώντας στην σχέση (4.3.1) του συντελεστή ασυμμετρίας τις αντίστοιχες ποσότητες έχουμε ότι :

$$\gamma_1 = \frac{\frac{a^3 + 3 \cdot a^2 + 2 \cdot a}{b^3} - 3 \cdot \frac{a}{b} \cdot \frac{a}{b^2} - \frac{a^3}{b^3}}{\frac{a}{b^2} \cdot \frac{\sqrt{a}}{b}} = \frac{\frac{2 \cdot a}{b^3}}{\frac{a \cdot \sqrt{a}}{b^3}} = \frac{2}{\sqrt{a}}$$

Συνεπώς για την κατανομή Γάμμα ο συντελεστής ασυμμετρίας ισούται με $2/\sqrt{a}$ και είναι μια ποσότητα η οποία εξαρτάται μόνο από την παράμετρο σχήματος a και όχι από την παράμετρο κλίμακας b . Επιπλέον παρατηρούμε μια αρνητική σχέση μεταξύ του συντελεστή ασυμμετρίας και της παραμέτρου σχήματος καθώς όσο αυξάνεται το a τόσο θα μειώνεται (φθίνει) η τιμή του συντελεστή. Εφόσον η παράμετρος a είναι θετικός αριθμός η παραπάνω ποσότητα θα είναι πάντα θετική με αποτέλεσμα η κατανομή να έχει δεξιά (ή θετική) ασυμμετρία και ενδεχομένως ακραίες τιμές στα δεξιά του κύριου όγκου των δεδομένων. Επιπλέον παρατηρούμε ότι όταν η παράμετρος $a \rightarrow \infty$ τότε ο συντελεστής $\gamma_1 \rightarrow 0$.

Στη συνέχεια ας εξετάσουμε και την τιμή που δύναται να πάρει ο συντελεστής κύρτωσης με τον υπολογισμό της τέταρτης κεντρικής ροπής.

Με την χρήση του Mathematica για $\mu = \frac{a}{b}$ με $a > 0, b > 0$ υπολογίζουμε την τέταρτη κεντρική ροπή ως εξής:

$$\mu_4 = \text{Integrate}[(x - \mu)^4 * f[x], \{x, 0, \text{Infinity}\}]$$

για την οποία ισχύει ότι είναι ίση με $\frac{3 \cdot a \cdot (a+2)}{b^4} = \frac{3 \cdot a^2 + 6 \cdot a}{b^4}$.

Παράλληλα αν υψώσουμε εις το τετράγωνο την διακύμανση, προκύπτει η ποσότητα: $\sigma^4 = \frac{a^2}{b^4}$

Επομένως έχουμε το πηλίκο: $\frac{\mu_4}{\sigma^4} = \frac{\frac{3 \cdot a^2 + 6 \cdot a}{b^4}}{\frac{a^2}{b^4}} = \frac{3 \cdot (a+2)}{a} = \frac{6}{a} + 3$

Όπως και ο συντελεστής ασυμμετρίας έτσι και ο συντελεστής κύρτωσης δεν εξαρτάται από την παράμετρο κλίμακας b ενώ επίσης έχει αρνητική σχέση με την παράμετρο σχήματος καθώς όσο αυξάνεται το a τόσο θα μειώνεται (φθίνει) η τιμή του συντελεστή. Τέλος εφόσον η παράμετρος a είναι θετικός αριθμός που ανήκει στο σύνολο των πραγματικών αριθμών, το πηλίκο $6/a + 3$ θα είναι πάντοτε θετικό με αποτέλεσμα ο συντελεστής κύρτωσης να παίρνει τιμή μεγαλύτερη του 3. Άρα πρόκειται για μια λεπτόκυρτη κατανομή με παχιά ουρά. Δεν υπάρχει απαραίτητα μεγάλη συγκέντρωση των δεδομένων γύρω από τον μέσο και οξεία κύρτωση πάντα διότι για πολύ μεγάλο a η κύρτωση είναι κοντά στο 3.

4.3.4 Κατανομή Weibull

Έστω η συνεχής τυχαία μεταβλητή X η οποία ακολουθεί την κατανομή Weibull (η οποία αποτελεί γενίκευση της εκθετικής) με συνάρτηση πυκνότητας πιθανότητας ίση με:

$$f(x) = \begin{cases} \frac{\kappa}{\lambda} \cdot \left(\frac{x}{\lambda}\right)^{\kappa-1} \cdot e^{-\left(\frac{x}{\lambda}\right)^\kappa}, & \text{για } x \geq 0 \\ 0, & \text{για } x < 0 \end{cases}$$

όπου λ είναι η παράμετρος κλίμακας και κ η παράμετρος σχήματος με $\kappa, \lambda > 0$.

Ας υπολογίσουμε αρχικά τον συντελεστή ασυμμετρίας γ_1 για την παραπάνω κατανομή:
Με τη βοήθεια του Mathematica ορίζουμε την συνάρτηση πυκνότητας πιθανότητας ως:

$$f[x_]:= (\kappa/\lambda) * \left(\frac{x}{\lambda}\right)^\kappa (\kappa - 1) * (\text{Exp}\left[-\left(\frac{x}{\lambda}\right)^\kappa\right])$$

και υπολογίζουμε την πρώτη ροπή γύρω από το μηδέν (μέση τιμή) ως εξής:

$$\mu = \text{Integrate}[x * f[x], \{x, 0, \text{Infinity}\}]$$

η οποία ισούται με:

$$\mu = \lambda \cdot \Gamma\left(1 + \frac{1}{\kappa}\right)$$

όπου Γ η συνάρτηση Γάμμα που ορίσαμε προηγουμένως μέσω της σχέσης (4.3.3).

Στην συνέχεια για $\mu = \lambda \cdot \Gamma\left(1 + \frac{1}{\kappa}\right)$ με $\alpha > 0, b > 0$ υπολογίζουμε την τρίτη κεντρική ροπή ως εξής:

$$\mu_3 = \text{Integrate}[(x - \mu)^3 * f[x], \{x, 0, \text{Infinity}\}]$$

για την οποία ισχύει ότι είναι ίση με:

$$\frac{\lambda^3 \cdot (2 \cdot \Gamma\left(\frac{1}{\kappa}\right))^3 - 6 \cdot \kappa \cdot \Gamma\left(\frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{2}{\kappa}\right) + \kappa^3 \cdot \Gamma\left(1 + \frac{3}{\kappa}\right)}{\kappa^3}$$

Στο Mathematica υπολογίζουμε την δεύτερη ροπή γύρω από το μηδέν ως εξής

$$m_2 = \text{Integrate}[(x^2) * f[x], \{x, 0, \text{Infinity}\}]$$

και έπειτα υπολογίζουμε την τυπική απόκλιση ως $\sigma = \text{Sqrt}[m_2 - (\mu^2)]$

Η διακύμανση της κατανομής είναι ίση με:

$$\sigma^2 = \lambda^2 \cdot \left[\Gamma\left(1 + \frac{2}{\kappa}\right) - \left(\Gamma\left(1 + \frac{1}{\kappa}\right) \right)^2 \right]$$

Επομένως προκύπτει το ακόλουθο πηλίκο για τον συντελεστή ασυμμετρίας:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\lambda^3 \cdot (2 \cdot \Gamma\left(\frac{1}{\kappa}\right))^3 - 6 \cdot \kappa \cdot \Gamma\left(\frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{2}{\kappa}\right) + \kappa^3 \cdot \Gamma\left(1 + \frac{3}{\kappa}\right)}{\kappa^3 \cdot \left(\lambda^2 \cdot \left[\Gamma\left(1 + \frac{2}{\kappa}\right) - \left(\Gamma\left(1 + \frac{1}{\kappa}\right) \right)^2 \right] \right)^{\frac{3}{2}}}$$

ή ισοδύναμα :

$$\gamma_1 = \frac{\lambda^3 \cdot \Gamma\left(1 + \frac{3}{\kappa}\right) - 3 \cdot \mu \cdot \sigma^2 - \mu^3}{\sigma^3}$$

Συνεπώς σε αντίθεση με την κατανομή Γάμμα, στην κατανομή Weibull ο συντελεστής ασυμμετρίας είναι μια ποσότητα η οποία εξαρτάται όχι μόνο από την παράμετρο σχήματος κ αλλά και από την παράμετρο κλίμακας λ. Άρα για διάφορες τιμές των παραμέτρων μπορεί να πάρει είτε θετική είτε αρνητική τιμή και η κατανομή να έχει αντίστοιχα είτε δεξιά είτε αριστερή ασυμμετρία.

Τέλος ας εξετάσουμε και την τιμή που μπορεί να πάρει ο συντελεστής κύρτωσης με τον υπολογισμό της τέταρτης κεντρικής ροπής.

Στο πρόγραμμα Mathematica για $\mu = \lambda \cdot \Gamma\left(1 + \frac{1}{\kappa}\right)$ με $a > 0, b > 0$ υπολογίζουμε την τέταρτη κεντρική ροπή ως εξής:

$$\mu_4 = \text{Integrate}[(x - \mu)^4 * f[x], \{x, 0, \text{Infinity}\}]$$

για την οποία ισχύει ότι είναι ίση με:

$$\frac{\lambda^4 \cdot (-3 \cdot \left(\Gamma\left(\frac{1}{\kappa}\right)\right)^4 + 12 \cdot \kappa \cdot \left(\Gamma\left(\frac{1}{\kappa}\right)\right)^2 \cdot \Gamma\left(\frac{2}{\kappa}\right) - 12 \cdot \kappa^2 \cdot \Gamma\left(\frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{3}{\kappa}\right) + \kappa^4 \cdot \Gamma\left(1 + \frac{4}{\kappa}\right)}{\kappa^4}$$

Επομένως για τον συντελεστή κύρτωσης αντίστοιχα θα έχουμε:

$$\begin{aligned} \frac{\mu_4}{\sigma^4} &= \frac{\lambda^4 \cdot (-3 \cdot \left(\Gamma\left(\frac{1}{\kappa}\right)\right)^4 + 12 \cdot \kappa \cdot \left(\Gamma\left(\frac{1}{\kappa}\right)\right)^2 \cdot \Gamma\left(\frac{2}{\kappa}\right) - 12 \cdot \kappa^2 \cdot \Gamma\left(\frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{3}{\kappa}\right) + \kappa^4 \cdot \Gamma\left(1 + \frac{4}{\kappa}\right)}{\kappa^4 \cdot \left(\lambda^2 \cdot \left[\Gamma\left(1 + \frac{2}{\kappa}\right) - \left(\Gamma\left(1 + \frac{1}{\kappa}\right)\right)^2\right]\right)^2} = \\ &= \frac{-3 \cdot \Gamma\left(\frac{1}{\kappa}\right) \cdot \left(\left(\Gamma\left(\frac{1}{\kappa}\right)\right)^3 - 4 \cdot \kappa \cdot \Gamma\left(\frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{2}{\kappa}\right) + 4 \cdot \kappa^2 \cdot \Gamma\left(\frac{3}{\kappa}\right)\right) + \kappa^4 \cdot \Gamma\left(1 + \frac{4}{\kappa}\right)}{\left(\left(\Gamma\left(\frac{1}{\kappa}\right)\right)^2 - 2 \cdot \kappa \cdot \Gamma\left(\frac{2}{\kappa}\right)\right)^2}. \end{aligned}$$

Ύστερα από πράξεις θα μπορούσε εναλλακτικά η παραπάνω ποσότητα να γραφθεί ως εξής:

$$\gamma_2 = \frac{\lambda^4 \cdot \Gamma\left(1 + \frac{4}{\kappa}\right) - 4 \cdot \gamma_1 \cdot \mu \cdot \sigma^3 - 6 \cdot \mu^2 \cdot \sigma^2 - \mu^4}{\sigma^4}$$

Από την τελευταία σχέση παρατηρούμε ότι ο συντελεστής κύρτωσης εξαρτάται και από την τιμή που θα πάρει ο συντελεστής ασυμμετρίας γ_1 ενώ παράλληλα εξαρτάται και από την παράμετρο σχήματος κ αλλά και από την παράμετρο κλίμακας λ. Συνεπώς η τιμή του συντελεστή γ_2 εξαρτάται από τις δυνατές τιμές των δύο αυτών παραμέτρων.

4.4 Ασυμμετρία και κύρτωση για μείξεις δύο Εκθετικών Κατανομών

Είναι ευρέως γνωστόν ότι μια κατηγορία κατανομών η οποία χρησιμοποιείται ιδιαίτερος στον αναλογισμό είναι οι μείξεις κατανομών. Στο προηγούμενο κεφάλαιο διατυπώσαμε τις δύο συνθήκες που πρέπει να ικανοποιεί μια κατανομή ώστε να ονομάζεται διακριτή ή συνεχής μείξη κατανομών. Στην ενότητα αυτή θα προσπαθήσουμε να μελετήσουμε βασικά χαρακτηριστικά όπως η κύρτωση αλλά και η ασυμμετρία για μια σημαντική οικογένεια κατανομών η οποία είναι η διακριτή μείξη εκθετικών. Οι εκθετικές κατανομές χρησιμοποιούνται κυρίως στην θεωρία χρεοκοπίας για την ποσοτικοποίηση των ατομικών αποζημιώσεων ενώ παράλληλα αποτελούν μια γενίκευση των κατανομών Γάμμα (a, b) (για $a = 1$).

Ορισμός 4.4.1 (Πολίτης, 2018):

Για μια τυχαία μεταβλητή X , με συνάρτηση πυκνότητας:

$$a_1 \cdot b_1 \cdot e^{-b_1 \cdot x} + a_2 \cdot b_2 \cdot e^{-b_2 \cdot x} + \dots + a_k \cdot b_k \cdot e^{-b_k \cdot x} = \sum_{i=1}^k a_i \cdot b_i \cdot e^{-b_i \cdot x}$$

με $a_i > 0$ για κάθε i , $b_i > 0$ για κάθε i και $\sum_{i=1}^k a_i = 1$, η κατανομή της θα ονομάζεται μείξη k εκθετικών κατανομών με παραμέτρους b_1, b_2, \dots, b_k και αντίστοιχα βάρη a_1, a_2, \dots, a_k .

Εάν $a_i < 0$ για κάποιο i , τότε έχουμε συνδυασμό εκθετικών κατανομών.

Ιδιότητες για τις μείξεις εκθετικών κατανομών:

α) για τις ροπές ισχύει ότι:

$$E(X^r) = a_1 \cdot E(X_1^r) + a_2 \cdot E(X_2^r) + \dots + a_k \cdot E(X_k^r), r = 1, 2, \dots$$

β) για την ροπογεννήτρια της μείξης εκθετικών ισχύει:

$$M_X(t) = a_1 \cdot M_{X_1}(t) + a_2 \cdot M_{X_2}(t) + \dots + a_k \cdot M_{X_k}(t), t < \min\{b_1, \dots, b_k\}$$

όπου $X_i \sim \text{Exp}(b_i)$ για $i = 1, 2, \dots, k$

Θα μελετήσουμε την διακριτή μείξη δύο Εκθετικών κατανομών με συναρτήσεις πυκνότητας πιθανότητας:

$$f_i(x) = \begin{cases} b_i \cdot e^{-b_i x}, & \text{για } x \geq 0 \\ 0, & \text{για } x < 0 \end{cases}$$

για $i = 1, 2$ και $b_1, b_2 > 0$

Η συνάρτηση πυκνότητας πιθανότητας για την διακριτή μείξη θα είναι:

$$f(x) = a \cdot f_1(x) + (1 - a) \cdot f_2(x)$$

Για την μείξη των δύο εκθετικών κατανομών με $X_1 \sim \text{Exp}(b_1)$ και $X_2 \sim \text{Exp}(b_2)$ και βάρη a και $1 - a$ αντίστοιχα, τα οποία αθροίζουν στη μονάδα, με την χρήση του προγράμματος Mathematica προκύπτουν οι εξής τύποι για την μέση τιμή αλλά και την διακύμανση που όπως παρατηρούμε πρόκειται για σταθμισμένες τιμές (με τα αντίστοιχα βάρη) των επιμέρους μέσων τιμών και διακυμάνσεων :

$$\begin{aligned} \bullet \quad E(X) &= \int_0^{\infty} x \cdot (a \cdot f_1(x) + (1 - a) \cdot f_2(x)) dx = a \cdot \frac{1}{b_1} + (1 - a) \cdot \frac{1}{b_2} \\ \bullet \quad \text{Var}(X) &= \int_0^{\infty} (x - \mu)^2 \cdot (a \cdot f_1(x) + (1 - a) \cdot f_2(x)) dx \\ &= \frac{1 - \frac{a(b_1 - b_2) \cdot (a \cdot (b_1 - b_2) + 2 \cdot b_2)}{b_1^2}}{b_2^2} \end{aligned}$$

Συνεπώς η τυπική απόκλιση σ για την κατανομή της μείξης θα είναι η τετραγωνική ρίζα της διακύμανσης $\text{Var}(X)$.

Στη συνέχεια για την εύρεση των δύο συντελεστών, θα υπολογίσουμε την τρίτη αλλά και την τέταρτη κεντρική ροπή αντίστοιχα.

Για την τρίτη κεντρική ροπή έχουμε:

$$\begin{aligned} \mu_3 &= \int_0^{\infty} (x - \mu)^3 \cdot (p \cdot f_1(x) + (1 - p) \cdot f_2(x)) dx \\ &= \frac{-2 \cdot (-1 + a^3) \cdot b_1^3 + 6 \cdot (-1 + a) \cdot a^2 \cdot b_1^2 \cdot b_2 - 6 \cdot (-1 + a)^2 \cdot a \cdot b_1 \cdot b_2^2 + 2 \cdot a \cdot (3 + (-3 + a) \cdot a) \cdot b_2^3}{b_1^3 \cdot b_2^3} \end{aligned}$$

Αν υπολογίσουμε το πηλίκο της τρίτης κεντρικής ροπής ως προς την τυπική απόκλιση σ υψωμένη εις την τρίτη δύναμη, τότε προκύπτει ο συντελεστής ασυμμετρίας για την μείξη δύο εκθετικών κατανομών. Αντιστοίχως αν υπολογίσουμε το πηλίκο της τέταρτης κεντρικής ροπής: $\mu_4 = \int_0^{\infty} (x - \mu)^4 \cdot (a \cdot f_1(x) + (1 - a) \cdot f_2(x)) dx$ ως προς την τυπική απόκλιση σ υψωμένη εις την τέταρτη δύναμη, τότε προκύπτει ο συντελεστής κύρτωσης. Και στις δύο περιπτώσεις οι τύποι οι οποίοι προκύπτουν είναι ιδιαίτερα μακροσκελείς και περίπλοκοι ενώ παράλληλα εξαρτώνται τόσο από τα βάρη a και $1 - a$ αλλά και από τις τιμές των παραμέτρων b_1 και b_2 των δύο εκθετικών κατανομών που συνθέτουν την μείξη.

Για τον λόγο αυτό είναι αρκετά δύσκολη η εξαγωγή συμπερασμάτων σχετικά με το εάν η ασυμμετρία για τη μείξη δύο εκθετικών κατανομών είναι θετική ή αρνητική και σχετικά με την κύρτωση αντίστοιχα.

Ωστόσο για περαιτέρω ανάλυση θα δούμε κάποια παραδείγματα όπου θα μελετήσουμε για συγκεκριμένες τιμές των παραμέτρων των δύο κατανομών αλλά και για συγκεκριμένες τιμές των βαρών της μείξης, τις τιμές των συντελεστών ασυμμετρίας και κύρτωσης.

Αρχικά αξίζει να σημειωθεί ότι στη περίπτωση όπου και οι δύο εκθετικές κατανομές έχουν την ίδια παράμετρο $b = b_1 = b_2$ και μεταβάλλουμε τα βάρη αντίστοιχα, τότε πρόκειται για εκθετική κατανομή και όχι για μείξη με αποτέλεσμα οι συντελεστές ασυμμετρίας και

κύρτωσης να παίρνουν πάντοτε τις τιμές 2 και 9 όπως ακριβώς είδαμε ότι ισχύει και για την εκθετική κατανομή.

Επομένως, θα εξετάσουμε την περίπτωση όπου οι παράμετροι των εκθετικών κατανομών που συνθέτουν τη μείξη διαφέρουν μεταξύ τους. Στον Πίνακα 4.1, θα δούμε τα αποτελέσματα για τα ζεύγη τιμών των συντελεστών όταν οι παράμετροι είναι και οι δύο μικρότεροι της μονάδας, όταν είναι και οι δύο μεγαλύτεροι της μονάδας και όταν για την μία κατανομή είναι μεγαλύτερη ενώ για την άλλη κατανομή είναι μικρότερη της μονάδας. Παράλληλα θα μεταβάλλουμε τις τιμές για τα βάρη της μείξης.

Πίνακας 4.1

(b_1, b_2)	$a = 0.1$	$a = 0.5$	$a = 0.8$
(0.01,0.5)	(6.59302, 61.1293)	(2.72343, 13.2074)	(2.12563, 9.58175)
(0.4,0.9)	(2.90547, 18.6419)	(2.55916, 13.0721)	(2.18679, 10.1667)
(0.05,0.09)	(2.40346, 13.0291)	(2.37486, 11.8366)	(2.14089, 9.9378)
(5,10)	(2.61291, 15.322)	(2.46691, 12.4711)	(2.16525, 10.0665)
(4,60)	(6.54624, 61.4788)	(2.77799, 13.6379)	(2.15857, 9.75873)
(70,80)	(2.01185, 9.09997)	(2.02626, 9.20936)	(2.01443, 9.11079)
(1,80)	(6.57952, 60.819)	(2.7129, 13.131)	(2.11965, 9.55147)
(1,2)	(2.61291, 15.322)	(2.46691, 12.4711)	(2.16525, 10.0665)
(0.04,2)	(6.59302, 61.1293)	(2.72343, 13.2074)	(2.12563, 9.58175)
(0.06,80)	(6.54693, 60.2025)	(2.69545, 13.008)	(2.10992, 9.50312)
(0.8,3)	(4.56078, 38.5996)	(2.78957, 14.2831)	(2.22045, 10.2244)
(0.5,90)	(6.56193, 60.4734)	(2.70273, 13.0588)	(2.11396, 9.52303)
(60,0.01)	(2.02843, 9.12158)	(2.69456, 13.0018)	(4.51893, 30.3383)
(50,3)	(2.05482, 9,24544)	(2.77138, 135815)	(4.60257, 31.6187)
(4,0.2)	(2.05088, 9.22592)	(2.76071, 13.4935)	(4.59773, 31.4786)
(0.5,0.01)	(2.03793, 9.16445)	(2.72343, 13.2074)	(4.56097, 30.8751)
(80,0.08)	(2.02885, 9.12341)	(2.69584, 13.0107)	(4.52097, 30.3628)
(0.9,0.01)	(2.03377, 9.14544)	(2.7109, 13.1167)	(4.54383, 30.6471)
(5,40)	(6.18809, 58.009)	(2.8185, 14.0467)	(2.18958, 9.94794)
(100,0.5)	(2.03082, 9.13218)	(2.7019, 13.053)	(4.53044, 30.4784)
(50,35)	(2.03767, 9.2625)	(2.16803, 10.3143)	(2.17027, 10.4951)
(0.02,0.04)	(2.61291, 15.322)	(2.46691, 12.4711)	(2.16525, 10.0665)
(10,3)	(2.09858, 9.53473)	(2.75923, 14.1716)	(3.76963, 25.5326)
(90,1)	(2.03377, 9.14544)	(2.7109, 13.1167)	(4.54383, 30.6471)
(0.8,25)	(6.60257, 61.4843)	(2.73928, 13.3256)	(2.13479, 9.62919)
(100,90)	(2.00505, 9.03873)	(2.01646, 9.13141)	(2.012, 9.09901)
(0.00006,0.00008)	(2.0685, 9.61423)	(2.11415, 9.89965)	(2.05382, 9.39342)
(65,0.00003)	(2.02835, 9.12121)	(2.6943, 13.0003)	(4.51852, 30.3333)

Γενικό συμπέρασμα σχετικά με την ασυμμετρία και κύρτωση της μείξης δύο εκθετικών κατανομών:

- Στην περίπτωση όπου οι παράμετροι για τις δύο εκθετικές κατανομές διαφέρουν μεταξύ τους, όπως παρατηρήσαμε μέσω των παραπάνω αποτελεσμάτων, ακόμη και όταν οι δύο παράμετροι είναι μικρότεροι της μονάδας ή και οι δύο μεγαλύτεροι της μονάδας ή η μια μεγαλύτερη και η άλλη μικρότερη της μονάδας, ανεξάρτητα από το εάν $b_1 < b_2$ ή $b_2 < b_1$ τότε ο συντελεστής ασυμμετρίας είναι πάντοτε θετικός ενώ ο συντελεστής κύρτωσης μεγαλύτερος του 3. Πρόκειται για λεπτόκυρτη κατανομή με θετική ασυμμετρία.

Η μείξη εκθετικών κατανομών μπορεί να προσεγγίσει αρκετά ικανοποιητικά μια κατανομή Pareto (διπλωματική εργασία: Μ. Τσίπα, 2018), επομένως είναι λογικό ότι θα έχει πάντοτε θετική ασυμμετρία όπως είδαμε και σε προηγούμενη υποενότητα ότι ισχύει για τις κατανομές Pareto.

Έχοντας λοιπόν μελετήσει ιδιότητες όπως η ασυμμετρία και η κύρτωση για την μείξη δύο εκθετικών κατανομών, στη συνέχεια του κεφαλαίου θα αναφερθούμε στην οικογένεια των κατανομών Γάμμα οι οποίες ως γνωστόν για παράμετρο $a = 1$ μας δίνουν εκθετικές κατανομές.

4.5 Ασυμμετρία και κύρτωση για μείξεις δύο Γάμμα Κατανομών

Όπως έχει ήδη αναφερθεί σε προηγούμενο κεφάλαιο η μείξη κατανομών έχει μεγάλη εφαρμογή στον αναλογισμό και κυρίως την θεωρία κινδύνων και μπορεί να είναι είτε διακριτή είτε συνεχής αντίστοιχα. Στην συνεχή περίπτωση, η κατανομή της μείξης προκύπτει από δεσμευμένες κατανομές και ολοκληρώματα ενώ στην διακριτή μείξη (την οποία θα μελετήσουμε στην παρούσα διπλωματική εργασία), η κατανομή προκύπτει από ένα σταθμισμένο άθροισμα κατανομών το οποίο μπορεί να είναι πεπερασμένο ή μη. Στη παρακάτω ενότητα θα προσπαθήσουμε να μελετήσουμε ιδιότητες όπως είναι η ασυμμετρία και η κύρτωση για την περίπτωση όπου έχουμε διακριτή μείξη κατανομών και όταν οι κατανομές αυτές ανήκουν στην οικογένεια των κατανομών Γάμμα.

Ας μελετήσουμε την διακριτή μείξη δύο κατανομών Γάμμα με συναρτήσεις πυκνότητας πιθανότητας:

$$f_i(x) = \begin{cases} \frac{b_i^{a_i} \cdot x^{a_i-1} \cdot e^{-b_i x}}{\Gamma(a_i)} & , \text{για } x \geq 0 \\ 0 & , \text{για } x < 0 \end{cases}$$

για $i = 1, 2$ και $a_1, a_2 > 0, b_1, b_2 > 0$ ενώ η συνάρτηση Γάμμα ορίζεται μέσω της σχέσης (4.3.3).

Η συνάρτηση πυκνότητας πιθανότητας για την διακριτή μείξη θα είναι:

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

Αρχικά για την μείξη των κατανομών με $X_1 \sim \text{Gamma}(a_1, b_1)$ και $X_2 \sim \text{Gamma}(a_2, b_2)$ και βάρη p και $1-p$ αντίστοιχα, τα οποία αθροίζουν στη μονάδα, με την χρήση του προγράμματος Mathematica προκύπτουν οι παρακάτω τύποι τόσο για την μέση τιμή όσο και την διακύμανση:

- $E(X) = \int_0^{\infty} x \cdot (p \cdot f_1(x) + (1 - p) \cdot f_2(x)) dx = p \cdot \frac{a_1}{b_1} + (1 - p) \cdot \frac{a_2}{b_2}$
- $Var(X) = \int_0^{\infty} (x - \mu)^2 \cdot (p \cdot f_1(x) + (1 - p) \cdot f_2(x)) dx =$
 $= \frac{-2 \cdot b_1 \cdot b_2 \cdot a_1 \cdot a_2 \cdot (1 - p) \cdot p + b_2^2 \cdot a_1 \cdot p \cdot (1 + a_1 \cdot (1 - p)) + b_1^2 \cdot a_2 \cdot (1 - p) \cdot (1 + a_2 \cdot p)}{(b_1 \cdot b_2)^2}$

Συνεπώς η τυπική απόκλιση σ για την κατανομή της μείξης θα είναι η τετραγωνική ρίζα της διακύμανσης $Var(X)$.

Έπειτα υπολογίζουμε την τρίτη κεντρική ροπή για την οποία ισχύει ο παρακάτω τύπος:

$$\begin{aligned} \mu_3 &= \int_0^{\infty} (x - \mu)^3 \cdot (p \cdot f_1(x) + (1 - p) \cdot f_2(x)) dx = \\ &= -\frac{3 \cdot a_2^2 \cdot b_1^2 \cdot (b_1 + a_1 \cdot b_2 \cdot (1 - 2 \cdot p)) \cdot (-1 + p) \cdot p + a_2^3 \cdot b_1^3 \cdot (-1 + p) \cdot p \cdot (-1 + 2 \cdot p)}{b_1^3 \cdot b_2^3} \\ &\quad + \frac{a_2 \cdot b_1 \cdot (-1 + p) \cdot (2 \cdot b_1^2 - 3 \cdot a_1 \cdot b_2 \cdot (b_1 + b_2 + a_1 \cdot b_2)) \cdot p + 6 \cdot a_1^2 \cdot b_2^2 \cdot p^2}{b_1^3 \cdot b_2^3} \end{aligned}$$

$$\frac{-a_1 \cdot b_2^3 \cdot p \cdot (2 + a_1 \cdot (-1 + p)) \cdot (-3 + a_1 \cdot (-1 + 2 \cdot p))}{b_1^3 \cdot b_2^3}$$

Διαιρώντας την τρίτη κεντρική ροπή με την τυπική απόκλιση υψωμένη εις την τρίτη δύναμη προκύπτει αντίστοιχα ένας ιδιαίτερα πολύπλοκος τύπος για τον συντελεστή ασυμμετρίας ο οποίος εξαρτάται από τα βάρη αλλά και από τις παραμέτρους σχήματος και κλίμακας των δύο Γάμμα κατανομών.

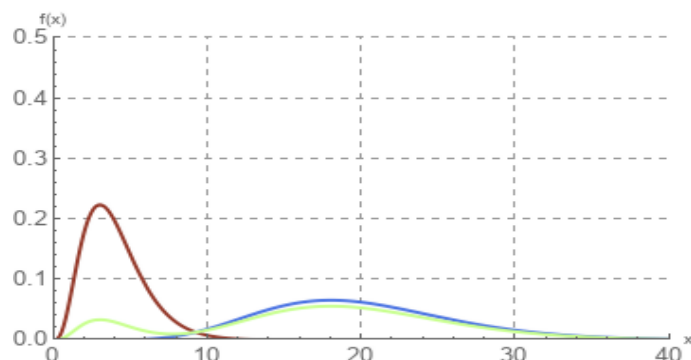
Αντίστοιχα διαιρώντας την τέταρτη κεντρική ροπή με την τυπική απόκλιση υψωμένη εις την τέταρτη δύναμη προκύπτει ένας εξίσου ιδιαίτερα πολύπλοκος τύπος για τον συντελεστή κύρτωσης ο οποίος επίσης εξαρτάται από τα βάρη αλλά και από τις παραμέτρους σχήματος και κλίμακας των δύο Γάμμα κατανομών.

Ωστόσο με τους τύπους της ασυμμετρίας και της κύρτωσης που προκύπτουν για την μείξη δύο Γάμμα κατανομών δεν μπορούμε να δούμε για ποιες τιμές των παραμέτρων των δύο κατανομών θα έχουμε θετική ή αρνητική ασυμμετρία και αντίστοιχη κύρτωση καθώς είναι δύσκολο να αναλυθούν. Παρόλα αυτά μπορούμε για συγκεκριμένες τιμές των παραμέτρων και βαρών των κατανομών να δούμε τι ισχύει σχετικά με την ασυμμετρία αλλά και τη κύρτωση.

Όπως παρατηρήσαμε προηγουμένως ο συντελεστής ασυμμετρίας για την απλή κατανομή Γάμμα ισούται με $2/\sqrt{a}$ και εξαρτάται μόνο από την παράμετρο σχήματος. Ας δούμε ωστόσο τι ισχύει για μία διακριτή μείξη 2 Γάμμα κατανομών.

Αρχικά μπορούμε να παρατηρήσουμε ότι όταν στις δύο κατανομές που σχηματίζουν τη μείξη, οι κορυφές τους βρίσκονται μακριά η μία από την άλλη (δηλαδή όταν ισχύει για τις παραμέτρους σχήματος $|a_1 - a_2| \gg 0$), από το παρακάτω διάγραμμα φαίνεται ότι η κατανομή με το μεγαλύτερο βάρος επιβάλλεται της άλλης και συνεπώς η κατανομή της μείξης τείνει προς την κατανομή με το μεγαλύτερο βάρος. Άρα τιμές όπως η διακύμανση και η μέση τιμή (και επομένως και τα αντίστοιχα στατιστικά της κατανομής όπως ο συντελεστής κύρτωσης και ο συντελεστής ασυμμετρίας που μελετάμε) τείνουν προς τις αντίστοιχες τιμές της κατανομής με το μεγαλύτερο βάρος. Δηλαδή τα βάρη των κατανομών που συνθέτουν τη μείξη επηρεάζουν τις τιμές των συντελεστών κύρτωσης και ασυμμετρίας.

Διάγραμμα 4.1



Το παραπάνω διάγραμμα αφορά την συνάρτηση πυκνότητας πιθανότητας της μείξης :

$p \cdot \text{Gamma}(a_1, b_1) + (1 - p) \cdot \text{Gamma}(a_2, b_2)$ (πράσινη καμπύλη) με $a_1 = 4, b_1 = 1, a_2 = 10$ και $b_2 = 0.5$ και των συναρτήσεων $\text{Gamma}(4,1)$ (κόκκινη καμπύλη) και $\text{Gamma}(10,0.5)$ (μπλε καμπύλη). Έχουν επιλεγθεί βάρη $p = 0.15$ και $1 - p = 0.85$ αντίστοιχα.

Ωστόσο στη συνέχεια για να μπορέσουμε να καταλήξουμε σε ένα συμπέρασμα σχετικά με την συμπεριφορά των συντελεστών κύρτωσης και ασυμμετρίας θα δούμε κάποια παραδείγματα στα οποία για σταθερές παραμέτρους σχήματος a_1, a_2 της μείξης, θα μεταβάλλουμε την παράμετρο κλίμακος, η οποία θα θεωρούμε ότι είναι κοινή και για τις δυο Γάμμα κατανομές που συνθέτουν την μείξη, καθώς και τα αντίστοιχα βάρη των κατανομών της μείξης. Θα δούμε και συγκεντρωτικά τα διάφορα αποτελέσματα μέσω ακόλουθου πίνακα. Επομένως όπως θα δούμε και αναλυτικά παρακάτω σε τέσσερις περιπτώσεις, για δεδομένες τιμές των παραμέτρων σχήματος a_1, a_2 και για τα ίδια βάρη, όταν αλλάζει η κοινή παράμετρος κλίμακος $b = b_1 = b_2$ τότε οι συντελεστές ασυμμετρίας και κύρτωσης δεν μεταβάλλονται.

πρώτη περίπτωση: Θεωρούμε δύο κατανομές Γάμμα με παραμέτρους σχήματος $a_1 = 4$ και $a_2 = 8$, παράμετρο κλίμακος ίδια και για τις δύο μείξεις (0.3,3,20) κάθε φορά, ενώ παράλληλα θεωρούμε βάρη 0.5 για καθεμία κατανομή την πρώτη φορά, βάρος $p = 0.10$ την δεύτερη φορά και βάρος $p = 0.8$ την τρίτη φορά.

Αν αντικαταστήσουμε τους τύπους για τον συντελεστή ασυμμετρίας και κύρτωσης στο υπολογιστικό πρόγραμμα Mathematica προκύπτουν τα εξής ζεύγη τιμών αντίστοιχα:

(0.758947,3.48) , (0.548635,3.54319) και (1.17313,4.76229)

δεύτερη περίπτωση: Θεωρούμε δύο κατανομές Γάμμα με παραμέτρους σχήματος $a_1 = 5$ και $a_2 = 40$, παράμετρο κλίμακος ίδια και για τις δύο μείξεις (5,10,30) κάθε φορά, ενώ παράλληλα θεωρούμε βάρη 0.5 για καθεμία κατανομή την πρώτη φορά, βάρος $p = 0.10$ την δεύτερη φορά και βάρος $p = 0.8$ την τρίτη φορά.

Αν αντικαταστήσουμε τους τύπους για τον συντελεστή ασυμμετρίας και κύρτωσης στο υπολογιστικό πρόγραμμα Mathematica προκύπτουν τα εξής ζεύγη τιμών για κάθε επιλογή βάρους:

(0.161684,1.29682) , (-1.50936,5.09122) και (1.5761,3.8443)

τρίτη περίπτωση: Θεωρούμε δύο κατανομές Γάμμα με παραμέτρους σχήματος $a_1 = 1$ και $a_2 = 10$, παράμετρο κλίμακος ίδια και για τις δύο μείξεις (2,4,10) κάθε φορά, ενώ παράλληλα θεωρούμε βάρη 0.5 για καθεμία κατανομή την πρώτη φορά, βάρος $p = 0.10$ την δεύτερη φορά και βάρος $p = 0.8$ την τρίτη φορά.

Αν αντικαταστήσουμε τους τύπους για τον συντελεστή ασυμμετρίας και κύρτωσης στο υπολογιστικό πρόγραμμα Mathematica προκύπτουν τα εξής ζεύγη τιμών για κάθε βάρος αντίστοιχα:

$$(0.549106, 2.14884), (-0.187147, 3.34057) \text{ και } (1.82951, 5.50125)$$

τέταρτη περίπτωση: Θεωρούμε δύο κατανομές Γάμμα με παραμέτρους σχήματος $a_1 = 0.5$ και $a_2 = 0.8$, παράμετρο κλίμακος ίδια και για τις δύο μείξεις (0.6, 2, 50) κάθε φορά, ενώ παράλληλα θεωρούμε βάρη 0.5 για καθεμία κατανομή την πρώτη φορά, βάρος $p = 0.1$ την δεύτερη φορά και βάρος $p = 0.8$ την τρίτη φορά.

Αν αντικαταστήσουμε τους τύπους για τον συντελεστή ασυμμετρίας και κύρτωσης στο υπολογιστικό πρόγραμμα Mathematica προκύπτουν τα εξής ζεύγη τιμών για κάθε περίπτωση βάρους αντίστοιχα:

$$(2.47964, 12.1684), (2.27629, 10.7593) \text{ και } (2.67793, 13.7112)$$

Επομένως όπως παρατηρούμε, η μεταβλητή του βάρους είναι αυτή η οποία θα δώσει διαφορετικές τιμές για τους συντελεστές αντίστοιχα κι αυτό διότι η κατανομή της μείξης τείνει προς την κατανομή με το μεγαλύτερο βάρος όπως είδαμε και διαγραμματικά προηγουμένως. Συγκεντρωτικά στον ακόλουθο πίνακα έχουμε τις εξής τιμές για τους συντελεστές (ασυμμετρίας, κύρτωσης):

Πίνακας 4.2

(a_1, a_2)	P=0.5	P=0.10	P=0.80
(4,8) b= (0.3,3,20)	(0.758947, 3.48)	(0.548635, 3.54319)	(1.17313, 4.76229)
(5,40) b= (5,10,30)	(0.161684, 1.29682)	(-1.50936, 5.09122)	(1.5761, 3.8443)
(1,10) b= (2,4,10)	(0.549106, 2.14884)	(-0.187147, 3.3405)	(1.82951, 5.5012)
(0.5,0.8) b= (0.6,2,50)	(2.47964, 12.1684)	(2.27629, 10.7593)	(2.67793, 13.7112)
(10,0.6) b= (0.4,7,80)	(0.536252, 2.0654)	(2.9658, 11.6427)	(-0.20078, 2.5474)

- Σε αντίθεση με την μείξη εκθετικών κατανομών, παρατηρούμε ότι στην μείξη Γάμμα κατανομών με κοινή παράμετρο κλίμακας b προκύπτουν και αρνητικές τιμές για τον συντελεστή γ_1 . Συγκεκριμένα, μέσω των παραπάνω αποτελεσμάτων του Πίνακα 4.2, βλέπουμε ότι όταν η τιμή του a_2 είναι αρκετά μεγαλύτερη από την τιμή του a_1 τότε η ασυμμετρία παίρνει τιμές κοντά στο μηδέν για βάρος $p = 0.5$ ενώ για βάρος $p = 0.1$ η ασυμμετρία γίνεται αρνητική. Αντίθετα όταν $a_1 \gg a_2$ τότε η ασυμμετρία παίρνει

επίσης τιμές κοντά στο μηδέν για βάρος $p = 0.5$ αλλά γίνεται αρνητική για βάρος $p = 0.8$.

Στη συνέχεια, θα παρουσιάσουμε σε έναν επιπλέον πίνακα τα διάφορα στατιστικά (μέση τιμή, τυπική απόκλιση, συντελεστής ασυμμετρίας και συντελεστής κύρτωσης) για μείξεις κατανομών όπου η κάθε κατανομή που συνθέτει τη μείξη θα έχει διαφορετική παράμετρο κλίμακας (όχι κοινή όπως προηγουμένως) και θα δούμε τις αντίστοιχες τιμές των στατιστικών όταν οι παράμετροι σχήματος των επιμέρους κατανομών είναι πολύ κοντά η μία στην άλλη ($|a_1 - a_2| \sim 0$) ή όταν οι παράμετροι των επιμέρους κατανομών είναι μακριά η μία από την άλλη.

Πίνακας 4.3

Μείξη κατανομών	Μέση τιμή	Τυπική απόκλιση	Συντελεστής Ασυμμετρίας	Συντελεστής Κύρτωσης
0.5*Gamma (3,7) + 0.5*Gamma (9,11)	0.623377	0.325194	0.562737	3.11586
0.75*Gamma(8,6) + 0.25*Gamma(20,12)	1.41667	0.471405	0.469563	3.30176
0.2*Gamma(20,5) + 0.8*Gamma (20,6)	3.46667	0.821922	0.574178	3.57523
0.75*Gamma(20,5) + 0.25*Gamma (20,6)	3.83333	0.906765	0.469693	3.29942
0.3*Gamma(20,5) + 0.7*Gamma(20,30)	1.66667	1.609	1.16227	2.84266
0.8*Gamma(20,5) + 0.2*Gamma(20,30)	3.33333	1.55635	-0.545069	2.47775
0.75*Gamma(4,7) + 0.25*Gamma(30,20)	0.803571	0.491574	0.765783	2.5914
0.2*Gamma(4,7) + 0.8*Gamma(30,20)	1.31429	0.46291	-0.697157	3.03417
0.5*Gamma (3,8) + 0.5*Gamma (25,10)	1.4375	1.1302	0.245621	1.51259
0.8*Gamma (3,8) + 0.2*Gamma (25,10)	0.8	0.9	1.57442	4.12363

0.7*Gamma(0.3,5)+ 0.3*Gamma(0.3,10)	0.051	0.0974115	4.01134	27.6859
0.1*Gamma(0.3,5)+ 0.9*Gamma(0.3,10)	0.033	0.0630952	4.35877	35.8591
0.85*Gamma(0.5,26)+ 0.15*Gamma(3,25)	0.0343462	0.514139	2.60318	11.698
0.6*Gamma(0.5,6)+ 0.4*Gamma(10,30)	0.183333	0.166667	0.718	3.0366

- Στη περίπτωση όπου έχουμε μείξεις με τις ίδιες κατανομές Γάμμα για όλες αλλά με διαφορετικά βάρη για την κάθε μείξη αντίστοιχα, παρατηρούμε ότι τα αποτελέσματα στους συντελεστές κύρτωσης και ασυμμετρίας αλλάζουν. Πολλές φορές μπορεί για μια μείξη δύο κατανομών με συγκεκριμένα βάρη να έχουμε θετική ασυμμετρία ενώ μεταβάλλοντας μόνο τις τιμές των βαρών να προκύπτει αρνητική ασυμμετρία. Συνεπώς η μεταβλητή του βάρους μπορεί να επηρεάσει σημαντικά την ασυμμετρία αλλά και την κύρτωση στις μείξεις Γάμμα κατανομών.

Παράλληλα είναι αρκετά ενδιαφέρον να μελετήσουμε και την συμπεριφορά της βαθμίδας αποτυχίας (failure rate) στην περίπτωση όπου έχουμε μείξεις Γάμμα κατανομών με *αρνητική* ασυμμετρία. Όπως διαπιστώσαμε στο 2^ο κεφάλαιο της διπλωματικής εργασίας, η βαθμίδα αποτυχίας συσχετίζεται με την δεξιά ουρά ενώ προκύπτει ότι μεγάλες τιμές της βαθμίδας αποτυχίας συνδέονται με παχιά δεξιά ουρά. Επιπλέον μια συνάρτηση κατανομής ανήκει στην κλάση IFR όταν η συνάρτηση της βαθμίδας αποτυχίας είναι αύξουσα (δηλαδή έχει θετική πρώτη παράγωγο, $\mu'(x) > 0$) ενώ ανήκει στην κλάση DFR όταν η συνάρτηση της βαθμίδας αποτυχίας είναι φθίνουσα (δηλαδή έχει αρνητική πρώτη παράγωγο, $\mu'(x) < 0$). Ωστόσο στη περίπτωση της μείξης κατανομών, όπου οι επιμέρους κατανομές θα έχουν βαθμίδες αποτυχίας $\mu_1(x)$ και $\mu_2(x)$ αντίστοιχα, θα πρέπει να ισχύει (Marshall & Olkin, 2007):

$$\frac{d\mu(x)}{dx} = \frac{d(p(x)\mu_1(x) + (1-p(x))\mu_2(x))}{dx} > 0 \text{ (ή } < 0 \text{)}$$

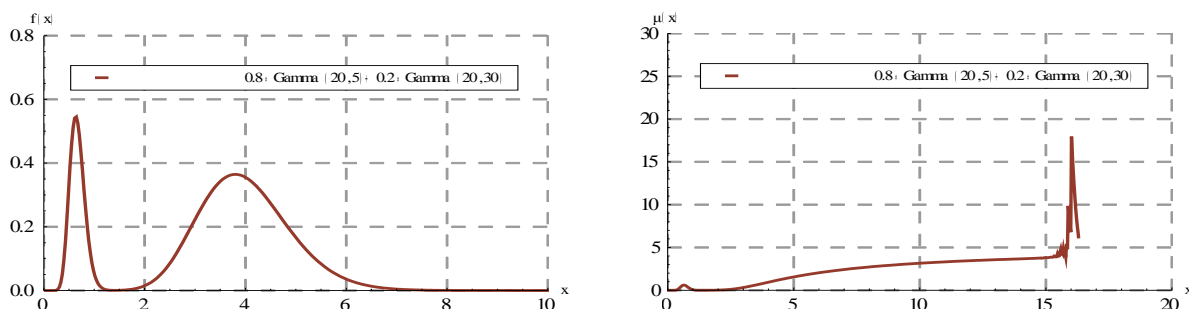
όπου

$$p(x) = \frac{p \cdot \bar{F}_1(x)}{p \cdot \bar{F}_1(x) + (1-p) \cdot \bar{F}_2(x)}$$

Επειδή η ανάλυση του παραπάνω τύπου με μαθηματικές μεθόδους είναι ιδιαίτερα περίπλοκη, μπορούμε για συγκεκριμένες τιμές των παραμέτρων των επιμέρους κατανομών να εξετάσουμε πότε έχουμε αύξουσα συνάρτηση της βαθμίδας αποτυχίας ή εάν προκύπτει μονοτονία ανά διαστήματα καθώς είναι πιθανό η μείξη να μην παρουσιάζει γνήσια μονοτονία.

Για την μείξη των κατανομών $\text{Gamma}(20,5)$ και $\text{Gamma}(20,30)$ με βάρη $p = 0.8$ και $1 - p = 0.2$ αντίστοιχα, για την οποία διαπιστώσαμε αρνητική ασυμμετρία, θα παρουσιάσουμε στο παρακάτω διάγραμμα την συνάρτηση πυκνότητας πιθανότητας αλλά και την μορφή της βαθμίδας αποτυχίας της.

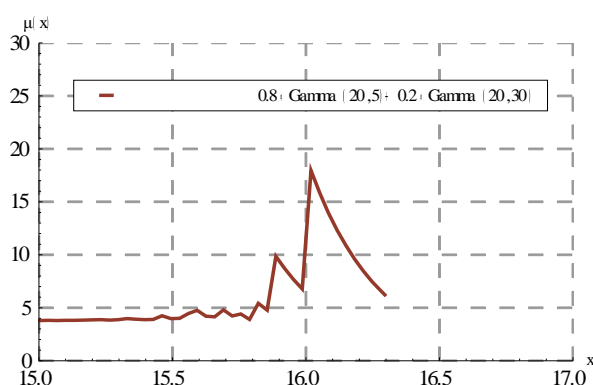
Διάγραμμα 4.2: Συνάρτηση πυκνότητας (αριστερά) και βαθμίδα αποτυχίας (δεξιά)



Πρόκειται για μια συνάρτηση πυκνότητας πιθανότητας η οποία δεν είναι μονοκόρυφη ενώ παράλληλα η συνάρτηση της βαθμίδας αποτυχίας θυμίζει μορφή roller-coaster, δηλαδή πρόκειται για μία μείξη των δύο βασικών μονοτονιών (IFR και DFR). Γενικά μια συνάρτηση με μορφή roller-coaster, εμφανίζει περισσότερα από ένα σημεία καμπής, τα οποία πιθανά σημεία καμπής προκύπτουν εάν μηδενίσουμε τη πρώτη παράγωγο της συνάρτησης. Στη συγκεκριμένη περίπτωση, για πολύ μικρές τιμές του x , η συνάρτηση $\mu(x)$ είναι αύξουσα, έπειτα φθίνουσα, στη συνέχεια πάλι αύξουσα ενώ τέλος στο διάστημα 15 έως 17 παρατηρούμε ότι η μονοτονία αλλάζει περισσότερες από 2 φορές.

Συνεπώς θα δούμε πιο αναλυτικά την μονοτονία της βαθμίδας αποτυχίας μέσω ενός επιπλέον διαγράμματος για το συγκεκριμένο διάστημα τιμών του x .

Διάγραμμα 4.3: Συνάρτηση πυκνότητας της βαθμίδας αποτυχίας του (4.2) στο διάστημα $[15,17)$

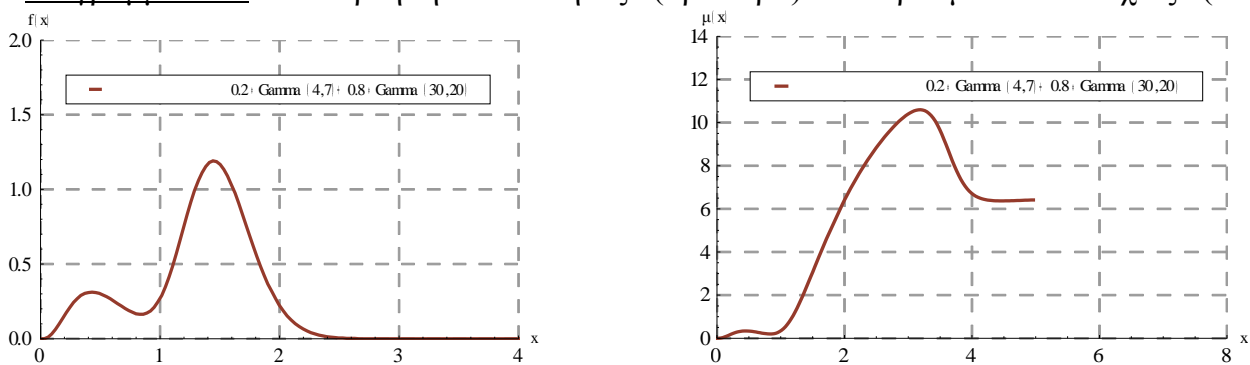


Είναι εμφανές, από το παραπάνω διάγραμμα, ότι αρκετές φορές μεταβάλλεται η μονοτονία της βαθμίδας αποτυχίας όταν το x βρίσκεται στο διάστημα $[15,17)$ καθώς αυξάνει και φθίνει συνεχόμενα. Ωστόσο θα χρησιμοποιήσουμε στο Mathematica την εντολή `NSolve[Derivative[1] [μ][x]==0,x]` ώστε να βρούμε τις ρίζες της εξίσωσης $\mu'(x) = 0$ καθώς τα σημεία στα οποία μηδενίζεται η πρώτη παράγωγος (σημεία καμπής) μας δείχνουν πόσες φορές αλλάζει η μονοτονία της αντίστοιχης συνάρτησης.

Συγκεκριμένα προκύπτουν 19 ρίζες για την εξίσωση $\mu'(x) = 0$, επομένως η μονοτονία της βαθμίδας αποτυχίας για την μείξη $\text{Gamma}(20,5)$ και $\text{Gamma}(20,30)$ με βάρος $p = 0.8$ αλλάζει συνολικά 19 φορές στο διάστημα $[0,20)$.

Επιπλέον θα παρουσιάσουμε διαγραμματικά την συνάρτηση πυκνότητας πιθανότητας αλλά και την συνάρτηση της βαθμίδας αποτυχίας για την μείξη των κατανομών $\text{Gamma}(4,7)$ και $\text{Gamma}(30,20)$ με βάρη $p = 0.2$ και $1 - p = 0.8$ αντίστοιχα, για την οποία επίσης παρατηρήθηκε, μέσω του πίνακα 4.3, αριστερή ή αρνητική ασυμμετρία.

Διάγραμμα 4.4: Συνάρτηση πυκνότητας (αριστερά) και βαθμίδα αποτυχίας (δεξιά)

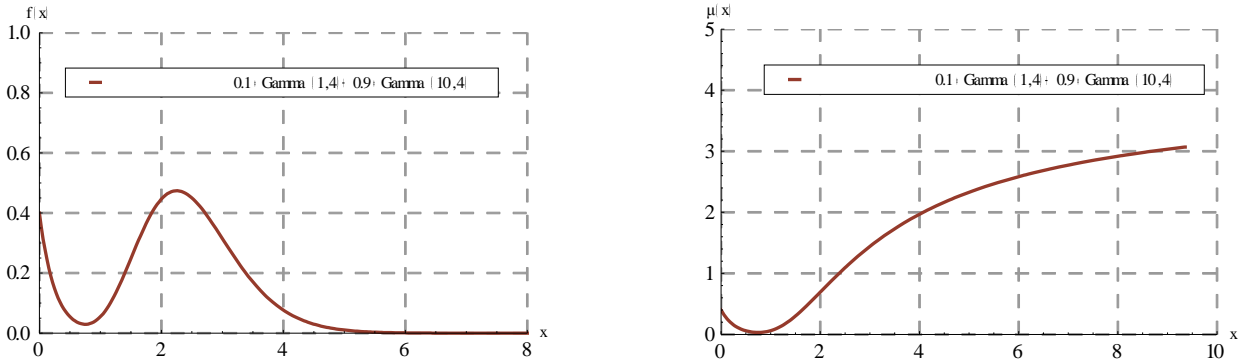


Όπως παρατηρούμε και σε αυτήν την περίπτωση μείξης Γάμμα κατανομών με αρνητική ασυμμετρία, η συνάρτηση πυκνότητας πιθανότητας δεν είναι μονοκόρυφη ενώ παράλληλα η μορφή της βαθμίδας αποτυχίας είναι roller-coaster όπως προηγουμένως. Η συνάρτηση της βαθμίδας αποτυχίας, ξεκινάει σταθερή, στη συνέχεια αυξάνεται απότομα, έπειτα φθίνει και στο τέλος παραμένει σταθερή.

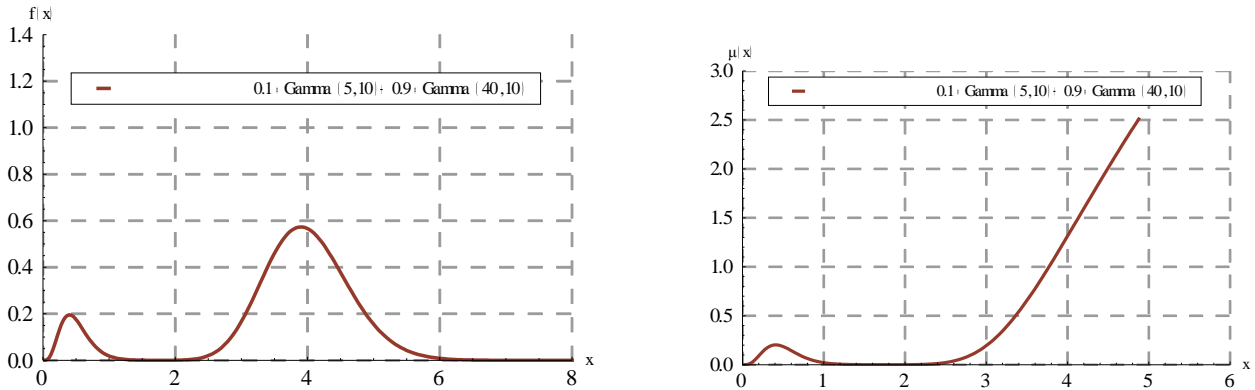
Τέλος θα εξετάσουμε μέσω διαγραμμάτων και δύο επιπλέον περιπτώσεις αριστερής ασυμμετρίας όπου οι κατανομές Γάμμα οι οποίες συνθέτουν τη μείξη θα έχουν κοινή παράμετρο κλίμακας (Πίνακας 4.2).

Επομένως θα παρουσιάσουμε την συνάρτηση πυκνότητας πιθανότητας και την συνάρτηση της βαθμίδας αποτυχίας για την μείξη των κατανομών $\text{Gamma}(1,4)$ η οποία είναι μια εκθετική κατανομή καθώς $a_1 = 1$ και $\text{Gamma}(10,4)$ με βάρη $p = 0.1$ και $1 - p = 0.9$ (διάγραμμα 4.5). Και την συνάρτηση πυκνότητας πιθανότητας μαζί με την συνάρτηση της βαθμίδας αποτυχίας για την μείξη των κατανομών $\text{Gamma}(5,10)$ και $\text{Gamma}(40,10)$ με βάρη $p = 0.1$ και $1 - p = 0.9$ αντίστοιχα (διάγραμμα 4.6).

Διάγραμμα 4.5: Συνάρτηση πυκνότητας (αριστερά) και βαθμίδα αποτυχίας (δεξιά)



Διάγραμμα 4.6: Συνάρτηση πυκνότητας (αριστερά) και βαθμίδα αποτυχίας (δεξιά)



Στο Διάγραμμα (4.6) παρατηρούμε ότι η συνάρτηση της βαθμίδας αποτυχίας είναι μορφής roller-coaster και συγκεκριμένα θα μπορούσαμε να διαπιστώσουμε ότι αρχικά έχει ανάποδη λεκανοειδή μορφή (upside-bathtub) ενώ στην συνέχεια αποκτά τη μορφή μιας αύξουσας (IFR) κατανομής. Και σε αυτή τη περίπτωση η συνάρτηση πυκνότητας πιθανότητας έχει δύο κορυφές αντίστοιχα. Στο διάγραμμα (4.5) όπου έχουμε μείξη η οποία σχηματίζεται από μια εκθετική κατανομή με παράμετρο b_1 ($a_1 = 1$) και μια Γάμμα με παραμέτρους a_2 και b_2 , η συνάρτηση της βαθμίδας αποτυχίας έχει λεκανοειδή μορφή (bathtub) γεγονός το οποίο επιβεβαιώνει την μία από τις δύο παρατηρήσεις των Block et al., (2012) οι οποίες είναι:

- Όταν $b_1 = b_2$ και $a_2 \leq 2$ τότε η βαθμίδα αποτυχίας της μείξης θα είναι IFR.
- Όταν $b_1 = b_2$ και $a_2 > 2$ τότε η βαθμίδα αποτυχίας της μείξης θα έχει λεκανοειδή μορφή (bathtub).

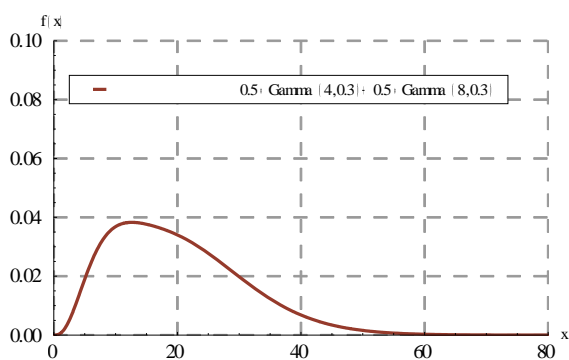
όπου $b_1, b_2 > 0$ και $a_2 \in \mathbb{N}$.

Γενικά όπως προκύπτει και μέσω των παραπάνω διαγραμμάτων, για τις μείξεις Γάμμα κατανομών με αρνητική ασυμμετρία, σε καμία περίπτωση δεν έχουμε βαθμίδα αποτυχίας με γνήσια μονοτονία ενώ παράλληλα τις περισσότερες φορές παρατηρούμε μορφή roller-coaster.

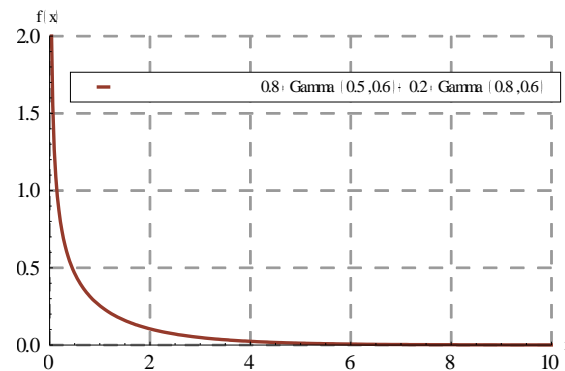
Επιπλέον διαπιστώνουμε ότι για όλες τις μείξεις Γάμμα κατανομών με αρνητική ασυμμετρία, οι συναρτήσεις πυκνότητας πιθανότητας δεν είναι μονοκόρυφες. Επομένως θα ήταν χρήσιμο να εξετάσουμε τι ισχύει και στην περίπτωση όπου έχουμε μείξεις Γάμμα κατανομών με θετική ασυμμετρία, για διάφορα παραδείγματα του Πίνακα 4.3.

Θα παρουσιάσουμε μέσω διαγραμμάτων τις αντίστοιχες συναρτήσεις:

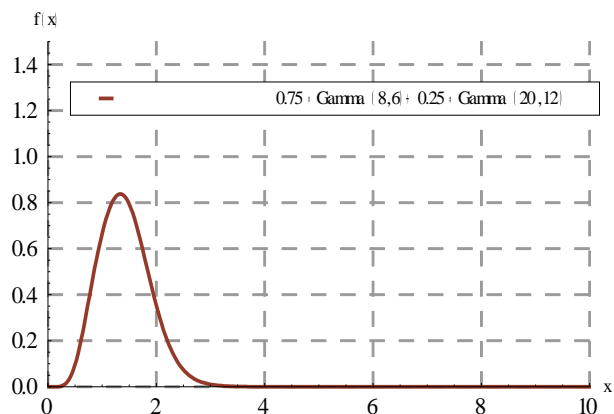
Διάγραμμα 4.7



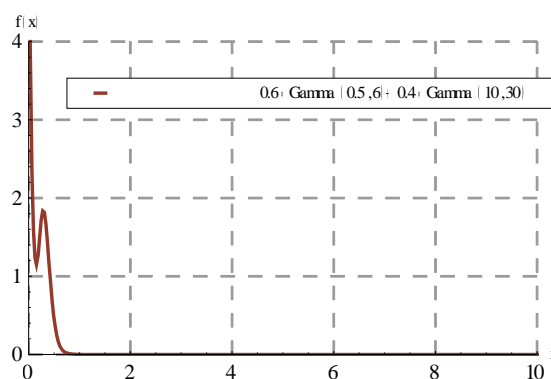
Διάγραμμα 4.8



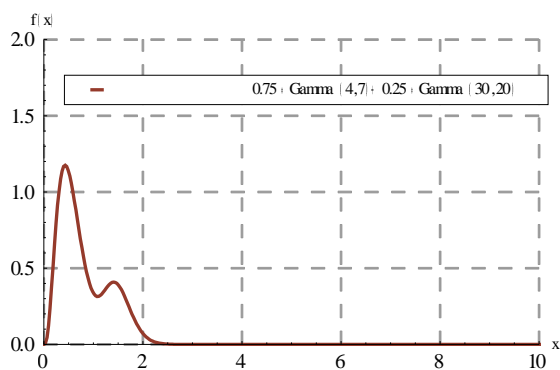
Διάγραμμα 4.9



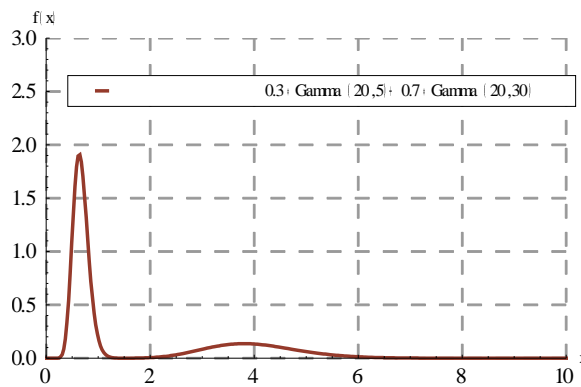
Διάγραμμα 4.10



Διάγραμμα 4.11



Διάγραμμα 4.12



Στις περισσότερες περιπτώσεις παραδειγμάτων για μείξεις Γάμμα κατανομών με θετική ασυμμετρία, διαπιστώνουμε ότι η συνάρτηση πυκνότητας πιθανότητας είναι μονοκόρυφη. Ωστόσο για τις μείξεις Γάμμα κατανομών στις οποίες διατηρώντας σταθερές τις παραμέτρους και μεταβάλλοντας τα βάρη, προκύπτει άλλοτε αρνητική και άλλοτε θετική ασυμμετρία παρατηρούμε ότι οι συναρτήσεις πυκνότητας πιθανότητας τους έχουν δύο κορυφές.

Επιπρόσθετα, παρατηρώντας τα παραπάνω αποτελέσματα του Πίνακα 4.3, θα ήταν ιδιαίτερα επικοδομητικό για την μελέτη μας να εξετάσουμε την περίπτωση όπου οι παράμετροι σχήματος και για τις δύο κατανομές θα είναι μικρότερες της μονάδας ($a_1 < 1$ και $a_2 < 1$).

Πίνακας 4.4

Μείξη κατανομών	Μέση τιμή	Τυπική απόκλιση	Συντελεστής Ασυμμετρίας	Συντελεστής Κύρτωσης
0.5*Gamma (0.3,7) + 0.5*Gamma (0.9,11)	0.0623377	0.0846151	2.62332	13.6279
0.75*Gamma(0.8,0.6) + 0.25*Gamma(0.4,0.6)	1.16667	1.424	2.3793	11.432
0.2*Gamma(0.1,0.6) + 0.8*Gamma (0.1,6)	0.0466667	0.247745	13.5505	283.037
0.75*Gamma(0.02,5) +0.25*Gamma (0.98,6)	0.0438333	0.110298	3.91687	23.8056
0.3*Gamma(0.79,8) + 0.7*Gamma(0.08,8)	0.036625	0.0789443	3.70639	22.4166
0.65*Gamma(0.3,30) +0.35*Gamma(0.3,30)	0.01	0.0182574	3.65148	23
0.8*Gamma(0.02,2) + 0.2*Gamma(0.09,40)	0.00845	0.0634103	15.7627	376.37
0.1*Gamma(0.7,0.5) + 0.9*Gamma(0.0008,0.9)	0.1408	0.676066	7.69086	84.4118

- Επομένως ένα γενικό συμπέρασμα (χωρίς να αποτελεί κανόνα) το οποίο προκύπτει μέσω των παραπάνω παραδειγμάτων σχετικά με την ασυμμετρία αλλά και την κύρτωση για την μείξη δύο Γάμμα Κατανομών είναι ότι για παραμέτρους σχήματος $a_1 < 1$ και $a_2 < 1$, ανεξάρτητα από τις τιμές των παραμέτρων κλίμακας και των βαρών, η ασυμμετρία είναι πάντοτε θετική ενώ ο συντελεστής κύρτωσης παίρνει τιμές μεγαλύτερες του 3 με αποτέλεσμα να έχουμε μία λεπτόκυρτη κατανομή. Όπως είδαμε και στην προηγούμενη υποενότητα, το ίδιο ισχύει και στην περίπτωση όπου $a_1 = 1$ και $a_2 = 1$ δηλαδή όταν έχουμε μείξη δύο εκθετικών κατανομών οι οποίες όπως αναφέραμε αποτελούν μια ειδική περίπτωση των κατανομών Γάμμα για παραμέτρους σχήματος ίσες με τη μονάδα.

Ωστόσο για να μπορέσουμε να επιβεβαιώσουμε το παραπάνω συμπέρασμα θα δούμε και κάποια επιπλέον παραδείγματα για τα οποία προκύπτει θετικός συντελεστής ασυμμετρίας και συντελεστής κύρτωσης μεγαλύτερος του 3, όταν για τις ίδιες μείξεις διατηρούμε σταθερές τις παραμέτρους κλίμακας και μεταβάλλουμε μόνο τις παραμέτρους σχήματος ($a_1 < 1$ και $a_2 < 1$) καθώς και τα βάρη αντίστοιχα.

Πίνακας 4.5

Μείξη κατανομών	Μέση τιμή	Τυπική απόκλιση	Συντελεστής Ασυμμετρίας	Συντελεστής Κύρτωσης
0.5*Gamma (0.3,7) + 0.5*Gamma (0.9,11)	0.0623377	0.0846151	2.62332	13.6279
0.5*Gamma(0.8,7) +0.5*Gamma(0.2,11)	0.0662338	0.106295	2.94777	15.6179
0.9*Gamma (0.3,7) + 0.1*Gamma (0.9,11)	0.0467532	0.0799414	3.3818	20.3227
0.8*Gamma(0.02,2) + 0.2*Gamma(0.09,40)	0.00845	0.0634103	15.7627	376.37
0.8*Gamma(0.6,2) + 0.2*Gamma(0.7,40)	0.2435	0.364495	2.85223	14.9552
0.1*Gamma(0.02,2) + 0.9*Gamma(0.09,40)	0.003025	0.0235803	39.1194	2496.32
0.01*Gamma(0.5,2) + 0.99*Gamma(0.01,40)	0.0027475	0.0432863	28.7107	1160.98

Τέλος, ως συνέχεια των προηγούμενων αναλύσεων, όπως παρατηρούμε από τα αποτελέσματα του Πίνακα 4.3, έχει ενδιαφέρον ότι για το ίδιο a ($a = 20$) και για τις δύο Γάμμα κατανομές έχουμε αρνητικό συντελεστή ασυμμετρίας ($\gamma_1 < 0$). Για τον σκοπό αυτό, μελετώντας μια μείξη Γάμμα κατανομών με την ίδια παράμετρο a (στην περίπτωση αυτή θα θεωρήσουμε $a = 20$), θα προσπαθήσουμε να βρούμε μια συνθήκη για να δούμε πότε ισχύει $\gamma_1 < 0$ συναρτήσει των b_1, b_2, p .

Στο υπολογιστικό πρόγραμμα Mathematica, αντικαθιστούμε όπου a ίσο με 20 στον τύπο για την ασυμμετρία που είχαμε βρει ότι ισχύει στην περίπτωση μείξης 2 Γάμμα κατανομών.

Προκύπτει ότι:

$$\gamma_1 = \frac{-(1200 \cdot b_1^2 \cdot (b_1 + 20 \cdot b_2 \cdot (1 - 2p)) \cdot (-1 + p) \cdot p + 8000 b_1^3 \cdot (-1 + p) \cdot p \cdot (-1 + 2p) + 20 \cdot b_1 \cdot (-1 + p) \cdot (2 \cdot b_1^2 - 60 b_2 \cdot (b_1 + 21 b_2) p + 2400 b_2^2 \cdot p^2))}{(b_1^3 b_2^3 \cdot (\frac{20 \cdot b_2^2 \cdot (21 - 20p) \cdot p - 400 \cdot b_1^2 \cdot (-1 + p) \cdot p - 20 b_1 \cdot (-1 + p) \cdot (b_1 - 40 p b_2)}{b_1^2 b_2^2})^{\frac{3}{2}})}$$

$$\frac{-20 \cdot b_2^3 \cdot p \cdot (2 + 20 \cdot (-1 + p) \cdot (-3 + 20 \cdot (-1 + 2p)))}{(b_1^3 b_2^3 \cdot (\frac{20 \cdot b_2^2 \cdot (21 - 20p) \cdot p - 400 \cdot b_1^2 \cdot (-1 + p) \cdot p - 20 b_1 \cdot (-1 + p) \cdot (b_1 - 40 p b_2)}{b_1^2 b_2^2})^{\frac{3}{2}})}$$

Από μαθηματικής άποψης, παρατηρούμε τελικά ότι η διεξαγωγή συμπερασμάτων μέσω του παραπάνω τύπου, ο οποίος ύστερα από την υπόθεση ότι $a_1 = a_2 = 20$ περιέχει μόνο 3 μεταβλητές αντί για 5, είναι αρκετά δύσκολη καθώς προκύπτει πρόσθεση αρνητικών αλλά και θετικών ποσοτήτων χωρίς να είναι εφικτή η εύρεση του προσήμου αντιστοίχως. Παράλληλα όπως θα εξετάσουμε και μέσω διαφόρων παραδειγμάτων για την μείξη Γάμμα κατανομών με κοινή παράμετρο a και ίση με 20, θα δούμε ότι όταν οι παράμετροι κλίμακας b_1 και b_2 είναι αρκετά κοντά ή αντίστοιχα απέχουν κατά πολύ η μία από την άλλη τότε προκύπτουν διαφορετικά αποτελέσματα σε κάθε περίπτωση, γεγονός το οποίο δεν μπορεί με μαθηματική έκφραση να προκύψει μέσω του παραπάνω τύπου. Ακόμη πιο δύσκολη είναι η μελέτη όταν η παράμετρος σχήματος είναι κοινή αλλά δεν παίρνει συγκεκριμένη τιμή.

Παρακάτω θα δούμε στον Πίνακα 4.6 κάποια ενδεικτικά παραδείγματα περίπτωσης όπου η παράμετρος σχήματος είναι ίση με 20 και για τις δύο κατανομές Γάμμα που συνθέτουν την μείξη ώστε να οδηγηθούμε σε κάποιες παρατηρήσεις.

Πίνακας 4.6

Μείξη κατανομών	Μέση τιμή	Τυπική απόκλιση	Συντελεστής Ασυμμετρίας	Συντελεστής Κύρτωσης
0.6*Gamma (20,7) + 0.4*Gamma (20,10)	2.51429	0.707972	0.519116	3.1114
0.8*Gamma(20,0.5) +0.2*Gamma(20,30)	32.1333	17.6506	-0.741224	2.55397
0.1*Gamma (20,6) + 0.9*Gamma (20,40)	0.783333	0.888429	2.95502	10.7504
0.3*Gamma(20,80) + 0.7*Gamma(20,0.8)	17.575	12.2684	-0.396201	1.77772
0.7*Gamma(20,6) + 0.3*Gamma(20,0.2)	32.3333	45.9644	1.11001	2.59007
0.1*Gamma(20,4) + 0.9*Gamma(20,2)	9.5	2.62202	0.0346714	3.08132
0.5*Gamma(20,4) + 0.5*Gamma(20,20)	3.0007	2.15639	0.390438	1.69556
0.5*Gamma(20,8) + 0.5*Gamma(20,0.1)	101.25	103.69	0.283629	1.45426

Για την μείξη κατανομών Γάμμα με κοινή παράμετρο σχήματος ίση με 20, παρατηρούμε ότι όταν η παράμετρος κλίμακας b_1 είναι πολύ μικρότερη από την παράμετρο κλίμακας b_2 , τότε προκύπτει αρνητική ασυμμετρία για $p > 1/2$ και ο συντελεστής ασυμμετρίας παίρνει τιμές κοντά στο μηδέν για $p = 0.5$ ενώ όταν η παράμετρος κλίμακας b_2 είναι πολύ μικρότερη από την παράμετρο κλίμακας b_1 , τότε προκύπτει αρνητική ασυμμετρία για $p < 1/2$ και ο συντελεστής ασυμμετρίας παίρνει τιμές κοντά στο μηδέν για $p = 0.5$. Επιπλέον όταν οι παράμετροι κλίμακας παίρνουν τιμές πολύ κοντά η μία στην άλλη, ανεξάρτητα από το εάν $b_1 < b_2$ ή $b_1 > b_2$, η ασυμμετρία της μείξης είναι πάντοτε θετική.

Γενικό συμπέρασμα: Η εύρεση προσήμου για τον συντελεστή ασυμμετρίας στην περίπτωση της μείξης Γάμμα κατανομών είναι ιδιαίτερα δύσκολη μέσω μαθηματικών τύπων. Για τον σκοπό αυτό, μέσω διάφορων παραδειγμάτων προσπαθήσαμε να οδηγηθούμε σε συμπεράσματα και παρατηρήσεις ύστερα από τον συνδυασμό ποικίλων παραμέτρων.

4.6 Ασυμμετρία και κύρτωση για συνελίξεις δύο Γάμμα Κατανομών

Στην ενότητα αυτή θα προσπαθήσουμε να μελετήσουμε την συμπεριφορά των συντελεστών ασυμμετρίας και κύρτωσης για συνελίξεις Γάμμα κατανομών, ξεκινώντας αρχικά από την συνέλιξη δύο κατανομών Γάμμα. Στην αναλογιστική επιστήμη αλλά και στην θεωρία πιθανοτήτων, είναι ευρέως γνωστή η χρήση των αθροισμάτων τυχαίων μεταβλητών για την επίλυση διάφορων προβλημάτων. Η πιο απλή περίπτωση αφορά την μελέτη αθροισμάτων της μορφής: $X_1 + X_2 + \dots + X_n$ όπου n είναι γνωστός ακέραιος αριθμός και X_i ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Επομένως, μια πολύ βασική έννοια για την μελέτη τέτοιων αθροισμάτων είναι η έννοια της συνέλιξης.

Ορισμός 4.5.1 (Κ.Πολίτης, 2012):

α) Έστω X, Y δύο *συνεχείς* ανεξάρτητες τυχαίες μεταβλητές με συναρτήσεις πυκνότητας πιθανότητας $f_X(x)$ και $f_Y(y)$ αντίστοιχα και Z μια τρίτη τυχαία μεταβλητή που ορίζεται από τη σχέση $Z = X + Y$. Τότε η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής Z δίνεται από την σχέση:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) \cdot f_Y(y) dy = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) dx$$

β) Έστω X, Y δύο *διακριτές* ανεξάρτητες τυχαίες μεταβλητές με συναρτήσεις πιθανότητας $f_X(x)$ και $f_Y(y)$ αντίστοιχα και Z μια τρίτη τυχαία μεταβλητή που ορίζεται από τη σχέση $Z = X + Y$. Τότε η συνάρτηση πιθανότητας της μεταβλητής Z δίνεται από την σχέση:

$$f_Z(z) = \sum_{y \in R_Y} f_X(z-y) \cdot f_Y(y) dy = \sum_{x \in R_X} f_X(x) \cdot f_Y(z-x) dx$$

όπου R_Y και R_X είναι τα πεδία ορισμού των Y, X αντίστοιχα.

Στην συνέχεια έχοντας ορίσει την έννοια της συνέλιξης, θα προσπαθήσουμε να μελετήσουμε την ασυμμετρία και κύρτωση για την περίπτωση της συνέλιξης δύο Γάμμα κατανομών.

Ωστόσο θα δούμε κάποιες περιπτώσεις αντίστοιχα, ξεκινώντας με την πιο απλή περίπτωση όπου οι δύο κατανομές Γάμμα θα έχουν την *ίδια παράμετρο κλίμακος*, b .

Α' περίπτωση: Έστω η τυχαία μεταβλητή $X \sim \text{Gamma}(a_1, b)$ και η τυχαία μεταβλητή $Y \sim \text{Gamma}(a_2, b)$ με αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας:

$$f_X(x) = \frac{b^{a_1} \cdot x^{a_1-1} \cdot e^{-b \cdot x}}{\Gamma(a_1)} \quad \text{και} \quad f_Y(y) = \frac{b^{a_2} \cdot y^{a_2-1} \cdot e^{-b \cdot y}}{\Gamma(a_2)}$$

Για $x \geq 0$ και $y \geq 0$, αντικαθιστώντας τις παραπάνω συναρτήσεις στην συνάρτηση της συνέλιξης έχουμε:

$$\begin{aligned}
f_Z(z) &= \int_0^z f_X(x) \cdot f_Y(z-x) dx = \int_0^z f_X(z-y) \cdot f_Y(y) dy \\
&= \int_0^z \frac{b^{a_1} \cdot x^{a_1-1} \cdot e^{-b \cdot x}}{\Gamma(a_1)} \cdot \frac{b^{a_2} \cdot (z-x)^{a_2-1} \cdot e^{-b \cdot (z-x)}}{\Gamma(a_2)} dx \\
&= \frac{b^{a_1+a_2}}{\Gamma(a_1) \cdot \Gamma(a_2)} \cdot e^{-b(x+y)} \cdot \int_0^z x^{a_1-1} \cdot (z-x)^{a_2-1} dx
\end{aligned}$$

Με αλλαγή μεταβλητής $t = \frac{x}{z}$ έτσι ώστε $0 \leq t \leq 1$, προκύπτει ότι:

$$f_Z(z) = \frac{b^{a_1+a_2}}{\Gamma(a_1) \cdot \Gamma(a_2)} \cdot e^{-b \cdot z} \cdot z^{a_1+a_2-1} \cdot \int_0^1 t^{a_1-1} \cdot (1-t)^{a_2-1} dt$$

Όμως ισχύει ότι: $\int_0^1 t^{a_1-1} \cdot (1-t)^{a_2-1} dt = B(a_1, a_2)$

όπου $B(a_1, a_2)$ είναι η συνάρτηση Βήτα για την οποία ισχύει:

$$B(a_1, a_2) = \frac{\Gamma(a_1) \cdot \Gamma(a_2)}{\Gamma(a_1+a_2)}$$

Επομένως η συνάρτηση πυκνότητας πιθανότητας του αθροίσματος $Z = X + Y$ θα δίνεται από την σχέση:

$$f_Z(z) = \frac{b^{a_1+a_2}}{\Gamma(a_1 + a_2)} \cdot e^{-bz} \cdot z^{a_1+a_2-1} \quad z \geq 0$$

Επομένως αποδείξαμε ότι η συνέλιξη δύο κατανομών Γάμμα με την ίδια παράμετρο κλίμακας b , θα είναι μια κατανομή Γάμμα με παραμέτρους $(a_1 + a_2, b)$. Δηλαδή εάν $X \sim \text{Gamma}(a_1, b)$ και $Y \sim \text{Gamma}(a_2, b)$ τότε $X + Y \sim \text{Gamma}(a_1 + a_2, b)$.

Σε αυτή τη περίπτωση εφόσον πρόκειται για κατανομή Γάμμα με συντελεστή σχήματος $a_1 + a_2$, ο συντελεστής ασυμμετρίας θα ισούται με $2/\sqrt{a_1 + a_2}$ και είναι μια ποσότητα η οποία εξαρτάται μόνο από τις παραμέτρους σχήματος $a_1 + a_2$ των δύο Γάμμα κατανομών και όχι από την κοινή παράμετρο κλίμακας b . Επιπλέον παρατηρούμε μια *αρνητική σχέση* μεταξύ του συντελεστή ασυμμετρίας και της παραμέτρου σχήματος $a_1 + a_2$ καθώς όσο αυξάνεται το $a_1 + a_2$ τόσο θα μειώνεται (φθίνει) η τιμή του συντελεστή. Εφόσον η παράμετρος $a_1 + a_2$ είναι θετικός αριθμός η παραπάνω ποσότητα θα είναι πάντα θετική με αποτέλεσμα η κατανομή να έχει δεξιά (ή θετική) ασυμμετρία.

Όσον αφορά τον συντελεστή κύρτωσης όπως δείξαμε προηγουμένως για τις Γάμμα κατανομές, ο συντελεστής θα είναι ίσος με: $\frac{6}{a_1+a_2} + 3$. Όπως και ο συντελεστής ασυμμετρίας

έτσι και ο συντελεστής κύρτωσης δεν εξαρτάται από την κοινή παράμετρο κλίμακας b ενώ επίσης έχει αρνητική σχέση με την παράμετρο σχήματος καθώς όσο αυξάνεται το $a_1 + a_2$ τόσο θα μειώνεται (φθίνει) η τιμή του συντελεστή. Τέλος εφόσον η παράμετρος $a_1 + a_2$ είναι θετικός αριθμός που ανήκει στο σύνολο των πραγματικών αριθμών, το πηλίκο $(6/a_1 + a_2) + 3$ θα είναι πάντοτε θετικό με αποτέλεσμα ο συντελεστής κύρτωσης να παίρνει τιμή μεγαλύτερη του 3. Άρα πρόκειται για μια λεπτόκυρτη κατανομή.

Β' περίπτωση: Στη συνέχεια θα εξετάσουμε την περίπτωση όπου οι δύο Γάμμα κατανομές θα έχουν κοινή παράμετρο σχήματος a αλλά διαφορετικές παραμέτρους κλίμακας.

Έστω η τυχαία μεταβλητή $X \sim \text{Gamma}(a, b_1)$ και η τυχαία μεταβλητή $Y \sim \text{Gamma}(a, b_2)$ με αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας:

$$f_X(x) = \frac{b_1^a \cdot x^{a-1} \cdot e^{-b_1 \cdot x}}{\Gamma(a)} \quad \text{και} \quad f_Y(y) = \frac{b_2^a \cdot y^{a-1} \cdot e^{-b_2 \cdot y}}{\Gamma(a)}$$

Με τη βοήθεια του Mathematica, αφού εισάγουμε τις συναρτήσεις πυκνότητας πιθανότητας των τυχαίων μεταβλητών X, Y ως εξής:

$$f_1[x_] := (b_1^a) * (x^{a-1}) * (Exp[-b_1 * x]) / Gamma[a]$$

$$f_2[y_] := (b_2^a) * (y^{a-1}) * (Exp[-b_2 * y]) / Gamma[a]$$

μπορούμε να υπολογίσουμε το ολοκλήρωμα:

$$\int_0^z f_X(x) \cdot f_Y(z-x) dx$$

με την εντολή: $f[z_] := \text{Integrate}[f_1[x] * f_2[z-x], \{x, 0, z\}]$

Στη περίπτωση αυτή προκύπτει ένας ιδιαίτερα περίπλοκος τύπος για την συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $Z=X+Y$ ο οποίος συμπεριλαμβάνει και συναρτήσεις Bessel.

$$f_z(z) = \frac{b_1^a \cdot (b_1 - b_2)^{\frac{1}{2}-a} \cdot b_2^a \cdot e^{-\frac{1}{2}(b_1+b_2) \cdot z} \sqrt{\pi} \cdot \sqrt{\frac{1}{z}} \cdot z^a \cdot I[-\frac{1}{2} + a, \frac{1}{2} \cdot (b_1 - b_2) \cdot z]}{\Gamma(a)}$$

όπου $I[n, z]$ είναι η τροποποιημένη συνάρτηση Bessel πρώτου είδους ($I_n(z)$) τάξης $n = -\frac{1}{2} + a$.

Έπειτα υπολογίζοντας την τρίτη και τέταρτη κεντρική ροπή καθώς επίσης και την τυπική απόκλιση, μέσω των τύπων για τους συντελεστές κύρτωσης και ασυμμετρίας προκύπτουν

αποτελέσματα τα οποία εξαρτώνται τόσο από τις παραμέτρους κλίμακας b_1 και b_2 των δύο κατανομών αλλά και από την κοινή παράμετρο σχήματος a .

Ωστόσο επειδή είναι δύσκολο να μελετήσουμε για ποιες τιμές των παραμέτρων θα έχουμε θετική ή αρνητική ασυμμετρία και αντίστοιχη κύρτωση (σε αντίθεση με την Α' περίπτωση), θα δούμε για συγκεκριμένες τιμές τι ισχύει σχετικά με τους συντελεστές. Θα παρουσιάσουμε στον παρακάτω πίνακα τα αποτελέσματα όπως προκύπτουν με την χρήση του προγράμματος Mathematica.

Πίνακας 4.7

Συνελίξεις Γάμμα κατανομών	Συντελεστής Ασυμμετρίας	Συντελεστής Κύρτωσης
$a = 0.1, b_1 = 6, b_2 = 8$	-2.23373	32.1665
$a = 0.3, b_1 = 5, b_2 = 50$	0.0443933	15.1914
$a = 8, b_1 = 8, b_2 = 10$	$\frac{-189}{41\sqrt{82}}$	$\frac{22815}{6724}$
$a = 4, b_1 = 12, b_2 = 14$	$\frac{-559}{85\sqrt{85}}$	$\frac{54441}{14450}$
$a = 4, b_1 = 6, b_2 = 60$	$\frac{-1001}{101\sqrt{101}}$	$\frac{91209}{20402}$
$a = 30, b_1 = 2, b_2 = 4$	$\frac{-3\sqrt{6}}{25}$	$\frac{392}{125}$
$a = 40, b_1 = 5, b_2 = 30$	$\frac{-217}{37\sqrt{370}}$	$\frac{86031}{27380}$
$a = 0.1, b_1 = 20, b_2 = 30$	-2.32105	33.9372
$a = 0.1, b_1 = 0.4, b_2 = 0.8$	-2.60711	39.5061
$a = 10, b_1 = 0.4, b_2 = 0.8$	-0.509117	3.408
$a = 80, b_1 = 0.6, b_2 = 80$	-0.223588	3.07499
$a = 3, b_1 = 0.4, b_2 = 20$	-1.15402	4.9984
$a = 0.5, b_1 = 0.4, b_2 = 40$	2.82801	14.9976
$a = 0.1, b_1 = 0.5, b_2 = 0.0001$	6.3245	63.004
$a = 0.3, b_1 = 5, b_2 = 6$	-0.156354	7.60722
$a = 0.4, b_1 = 7, b_2 = 10$	-0.395818	8.15969
$a = 0.8, b_1 = 0.5, b_2 = 70$	0.218388	5.11862
$a = 0.9, b_1 = 0.7, b_2 = 200$	0.565648	6.79506
$a = 8, b_1 = 700, b_2 = 4$	$\frac{1339844}{15313\sqrt{15313}}$	$\frac{7034547195}{1875903752}$
$a = 35, b_1 = 5, b_2 = 0.06$	0.337989	3.17138
$a = 10, b_1 = 5, b_2 = 4$	0.4553166	3.314455

- Μια παρατήρηση, η οποία προκύπτει άμεσα από τα παραδείγματα του πίνακα 4.6 είναι ότι για παραμέτρους σχήματος a μεγαλύτερες της μονάδας έχουμε αρνητική ασυμμετρία όταν $b_1 < b_2$ ενώ θετική ασυμμετρία όταν $b_1 > b_2$ είτε οι παράμετροι κλίμακας b_1 και b_2 είναι κοντά είτε είναι μακριά η μία από την άλλη. Ωστόσο αυτό δεν ισχύει για παραμέτρους σχήματος a μικρότερες της μονάδας καθώς μπορεί να προκύπτει αρνητική ασυμμετρία ή θετική ασυμμετρία είτε όταν $b_1 > b_2$ είτε όταν $b_1 < b_2$

ανάλογα με το εάν οι παράμετροι κλίμακας b_1 και b_2 είναι μακριά ή κοντά η μία από την άλλη.

Γ' περίπτωση: Τέλος θα εξετάσουμε την περίπτωση όπου οι δύο Γάμμα κατανομές θα έχουν διαφορετική παράμετρο σχήματος αλλά και διαφορετική παράμετρο κλίμακας ταυτόχρονα.

Έστω η τυχαία μεταβλητή $X \sim \text{Gamma}(a_1, b_1)$ και η τυχαία μεταβλητή $Y \sim \text{Gamma}(a_2, b_2)$ με αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας:

$$f_X(x) = \frac{b_1^{a_1} \cdot x^{a_1-1} \cdot e^{-b_1 \cdot x}}{\Gamma(a_1)} \quad \text{και} \quad f_Y(y) = \frac{b_2^{a_2} \cdot y^{a_2-1} \cdot e^{-b_2 \cdot y}}{\Gamma(a_2)}$$

Με τη χρήση του προγράμματος Mathematica, εισάγουμε τις συναρτήσεις πυκνότητας πιθανότητας των τυχαίων μεταβλητών X, Y με τις παρακάτω εντολές ως εξής:

$$f_1[x_] := (b_1^{a_1}) * (x^{a_1-1}) * (Exp[-b_1 * x]) / Gamma[a_1]$$

$$f_2[y_] := (b_2^{a_2}) * (y^{a_2-1}) * (Exp[-b_2 * y]) / Gamma[a_2]$$

Ενώ στη συνέχεια βρίσκουμε το ολοκλήρωμα:

$$\int_0^z f_X(x) \cdot f_Y(z-x) dx$$

με την εντολή: $f[z_] := \text{Integrate}[f_1[x] * f_2[z-x], \{x, 0, z\}]$

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $Z=X+Y$ δίνεται από τον τύπο:

$$f_Z(z) = b_1^{a_1} \cdot b_2^{a_2} \cdot e^{-b_2 \cdot z} \cdot \left(\frac{1}{z}\right)^{-a_1} \cdot z^{a_2-1} \cdot {}_1\bar{F}_1[a_1, a_1 + a_2, (-b_1 + b_2) \cdot z]$$

όπου ${}_1\bar{F}_1[a, b, z]$ είναι η κανονικοποιημένη υπεργεωμετρική συνάρτηση (Regularized confluent hypergeometric function ${}_1\bar{F}_1$).

Υπολογίζοντας την τρίτη και τέταρτη κεντρική ροπή καθώς επίσης και την τυπική απόκλιση, μέσω των τύπων για τους συντελεστές κύρτωσης και ασυμμετρίας προκύπτουν αποτελέσματα τα οποία εξαρτώνται τόσο από τις παραμέτρους κλίμακας b_1 και b_2 των δύο κατανομών αλλά και από τις παραμέτρους σχήματος a_1 και a_2 .

Όπως και στην δεύτερη περίπτωση, η ανάλυση των τύπων για τους συντελεστές είναι ιδιαίτερα δύσκολη. Μέσω του προγράμματος Mathematica, θα θέσουμε συγκεκριμένες τιμές για τις παραμέτρους κλίμακας και σχήματος ενώ παράλληλα θα δούμε τι τιμές παίρνουν οι συντελεστές όταν οι παράμετροι σχήματος των επιμέρους κατανομών είναι κοντά η μία από την άλλη ($|a_1 - a_2| \sim 0$), όταν είναι μακριά η μία από την άλλη ($|a_1 - a_2| \gg 0$) και όταν οι παράμετροι κλίμακας παίρνουν μικρές ή μεγάλες τιμές αντίστοιχα. Ωστόσο όπως παρατηρούμε για τις διάφορες περιπτώσεις παραδειγμάτων, προκύπτει θετικός συντελεστής ασυμμετρίας και συντελεστής κύρτωσης μεγαλύτερος του 3, δηλαδή έχουμε λεπτόκυρτες κατανομές με δεξιά ασυμμετρία.

Πίνακας 4.8

Συνελίξεις Γάμμα κατανομών	Συντελεστής ασυμμετρίας	Συντελεστής Κύρτωσης
$a_1 = 0.1, a_2 = 0.3, b_1 = 6, b_2 = 8$	3.25234	19.1925
$a_1 = 0.1, a_2 = 0.3, b_1 = 6, b_2 = 60$	6.06841	59.5727
$a_1 = 4, a_2 = 6, b_1 = 8, b_2 = 10$	0.6443148	3.6303623
$a_1 = 4, a_2 = 6, b_1 = 3, b_2 = 30$	0.9793801	4.456211
$a_1 = 0.3, a_2 = 30, b_1 = 14, b_2 = 16$	0.363455	3.1982
$a_1 = 0.3, a_2 = 30, b_1 = 4, b_2 = 40$	1.42009	8.0509
$a_1 = 8, a_2 = 40, b_1 = 6, b_2 = 8$	0.2953	3.1332501
$a_1 = 8, a_2 = 40, b_1 = 5, b_2 = 70$	0.6821277	3.7132434
$a_1 = 0.1, a_2 = 0.5, b_1 = 0.2, b_2 = 0.4$	3.04516	18.5556
$a_1 = 0.1, a_2 = 0.5, b_1 = 0.3, b_2 = 80$	6.32389	62.9916
$a_1 = 4, a_2 = 5, b_1 = 0.1, b_2 = 0.2$	0.768958	3.93878
$a_1 = 0.1, a_2 = 10, b_1 = 40, b_2 = 50$	0.629985	3.59588

ΚΕΦΑΛΑΙΟ 5

ΠΡΟΣΕΓΓΙΣΕΙΣ ΓΙΑ ΣΥΝΕΛΙΞΕΙΣ ΓΑΜΜΑ ΚΑΤΑΝΟΜΩΝ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΘΕΩΡΙΑ ΚΙΝΔΥΝΟΥ

Ένας από τους βασικούς σκοπούς του κεφαλαίου είναι να προσεγγίσουμε τις συνελίξεις Γάμμα κατανομών από κάποιες άλλες κατανομές καθώς είναι ιδιαίτερα γνωστόν ότι η ακριβής έκφραση για μία συνέλιξη Γάμμα κατανομών, όταν οι παράμετροι σχήματος a_1, a_2, \dots, a_n δεν είναι ακέραιοι αριθμοί, είναι αρκετά περίπλοκη. Επομένως θα δούμε παρακάτω αναλυτικά την κατανομή την οποία θα επιλέξουμε ώστε να προσεγγίσουμε ικανοποιητικά μια συνέλιξη Γάμμα κατανομών, για μεγάλο εύρος τιμών των παραμέτρων αλλά και για περισσότερες από 2 τυχαίες μεταβλητές στην συνέλιξη.

Ωστόσο προκειμένου να αξιολογήσουμε κατά πόσο καλά προσεγγίζεται η συνέλιξη από την κατάλληλη κατανομή θα χρησιμοποιήσουμε αντίστοιχα διάφορα κριτήρια. Θα συγκρίνουμε τις ροπές (τρίτη, τέταρτη, πέμπτη και έκτη) μεταξύ της ακριβούς κατανομής και της προσέγγισης, θα παρουσιάσουμε μέσω γραφημάτων και θα εξετάσουμε την απόλυτη διαφορά των συναρτήσεων πυκνότητας πιθανότητας αλλά και των συναρτήσεων κατανομών ενώ παράλληλα μια εξίσου σημαντική μέθοδος αξιολόγησης της προσέγγισης θα αποτελέσει και η μέθοδος των άνω ποσοστημορίων.

Τέλος θα μελετήσουμε και την εφαρμογή των συνελίξεων Γάμμα κατανομών στην θεωρία κινδύνου και το συλλογικό πρότυπο αποζημιώσεων.

5.1 Ακριβής πυκνότητα για την συνέλιξη Γάμμα κατανομών όταν οι παράμετροι σχήματος είναι ακέραιοι αριθμοί

Η συνάρτηση πυκνότητας πιθανότητας του αθροίσματος n ανεξάρτητων και τυχαίων μεταβλητών που ακολουθούν την κατανομή Γάμμα με παραμέτρους σχήματος a_i και παραμέτρους κλίμακας b_i , για $i = 1, 2, \dots, n$ αποτελεί εδώ και χρόνια αντικείμενο μελέτης πολλών επιστημόνων ενώ βρίσκει εφαρμογές σε διάφορα πεδία των εφαρμοσμένων μαθηματικών, της στατιστικής και του αναλογισμού. Γενικότερα όμως, η ακριβής κατανομή της συνέλιξης ανεξάρτητων και τυχαίων Γάμμα μεταβλητών δεν έχει πάντοτε κλειστό τύπο με αποτέλεσμα να μην είναι εφικτό να εκφραστεί μέσω μίας απλής μορφής (Diaconis and Perlman, 1990). Οι υπολογισμοί για ακριβείς εκφράσεις δεν είναι πάντοτε άμεσοι και βασίζονται κυρίως σε διάφορες προσεγγίσεις όπως θα μελετήσουμε αναλυτικότερα στη συνέχεια του πέμπτου κεφαλαίου. Παρόλα αυτά, στη δοθείσα ενότητα θα δούμε έναν από τους πρώτους ακριβείς τύπους για την συνέλιξη Γάμμα κατανομών που δόθηκε από τον Mathai (1982) στην πιο απλή περίπτωση όπου όλες οι παράμετροι σχήματος είναι *ακέραιοι αριθμοί*. Η έκφραση αυτή οδηγεί σε έναν εύκολο υπολογισμό υπό την συνθήκη που θέσαμε προηγουμένως ότι δηλαδή έχουμε ακέραιες παραμέτρους, ωστόσο όταν πρόκειται για

διάφορες τιμές των παραμέτρων κλίμακας και σχήματος ο υπολογισμός είναι πιο περίπλοκος και για πρώτη φορά μελετήθηκε από τον Moschoroulos (1985) μέσω μιας Γάμμα σειράς.

Ο Mathai (1982) έδωσε διάφορες εκφράσεις για την πυκνότητα της συνέλιξης: α) ως ένα πεπερασμένο άθροισμα Γάμμα πυκνοτήτων όταν οι παράμετροι a_i είναι ακέραιοι, β) ως μιας υπεργεωμετρικής συνάρτησης όταν οι παράμετροι σχήματος και κλίμακας είναι διαφορετικές για κάθε Γάμμα κατανομή και γ) ως μιας σειράς πολυωνύμων (series in terms of zonal polynomials) όταν οι παράμετροι σχήματος $a_i, i = 1, 2, \dots, n$ είναι ίδιες για όλες τις κατανομές Γάμμα. Στην περίπτωση όπου οι παράμετροι κλίμακας $b_i, i = 1, 2, \dots, n$ είναι ίδιες για όλες τις κατανομές Γάμμα τότε πολύ εύκολα προκύπτει ότι η συνάρτηση πυκνότητας της συνέλιξης είναι εξίσου μια κατανομή Γάμμα.

Θα δούμε την περίπτωση όπου όλες οι παράμετροι σχήματος είναι ακέραιοι, χωρίς απαραίτητα να ισχύει ότι και οι παράμετροι κλίμακας είναι επίσης ακέραιοι αριθμοί (Mathai, 1982).

Έστω οι ανεξάρτητες και τυχαίες μεταβλητές X_i , με $i = 1, 2, \dots, n$, οι οποίες έχουν συνάρτηση πυκνότητας πιθανότητας :

$$f(x_i) = \frac{x_i^{a_i-1} \cdot e^{-x_i/b_i}}{b_i^{a_i} \cdot \Gamma(a_i)}, \quad a_i > 0, b_i > 0, x_i > 0$$

και $f(x_i) = 0$, αλλού.

Τότε ορίζουμε την τυχαία μεταβλητή Y ως το άθροισμα $Y = X_1 + X_2 + \dots + X_n$. Επιπλέον θεωρούμε ότι a_1, a_2, \dots, a_n είναι ακέραιοι αριθμοί ενώ οι παράμετροι κλίμακας b_1, b_2, \dots, b_n μπορεί να μην είναι απαραίτητα ακέραιοι. Εφόσον οι ροπογεννήτριες συναρτήσεις των X_i και Y υπάρχουν τότε θα κάνουμε χρήση αυτών. Γνωρίζουμε ότι οι ανεξάρτητες και τυχαίες μεταβλητές X_1, X_2, \dots, X_n έχουν ροπογεννήτρια $M_{X_i}(t) = (1 - b_i \cdot t)^{-a_i}$ και συνεπώς λόγω της ανεξαρτησίας θα ισχύει για την τυχαία μεταβλητή του αθροίσματος ότι:

$$M_Y(t) = \prod_{i=1}^n (1 - b_i \cdot t)^{-a_i} \quad (5.1.1)$$

Η συνάρτηση πυκνότητας πιθανότητας της Y μπορεί να βρεθεί μέσω της παραπάνω ροπογεννήτριας. Θα εργαστούμε ωστόσο για την εύρεση της συνάρτησης πυκνότητας πιθανότητας μέσω μιας τεχνικής των μερικών κλασμάτων (general partial fraction technique). Όπως είναι ιδιαίτερα φανερό, αν μερικές από τις παραμέτρους κλίμακας b_i είναι ίδιες, τότε στο γινόμενο της σχέσης (5.1.1) κάποιοι από τους n παράγοντες θα συγχωνεύονται. Για τον σκοπό αυτό θεωρούμε στην περίπτωση μας ότι όλα τα b_i είναι μη αρνητικά και διαφορετικά μεταξύ τους. Επιπλέον καθώς τα a_i είναι ακέραιοι αριθμοί τότε μπορούμε την ροπογεννήτρια της Y να την εκφράσουμε ως ένα πεπερασμένο άθροισμα με την τεχνική των μερικών κλασμάτων, δηλαδή έχουμε:

$$M_Y(t) = \prod_{j=1}^n (-b_j)^{-a_j} \cdot \prod_{j=1}^n (t - 1/b_j)^{-a_j}$$

Επομένως θα ισχύει ότι:

$$M_Y(t) = \left[\prod_{j=1}^n (-b_j)^{-a_j} \right] \cdot \sum_{j=1}^n \sum_{r=1}^{a_j} b_{jr} \cdot (t - 1/b_j)^{-r}$$

όπου θα ορίσουμε στη συνέχεια τους συντελεστές b_{jr} . Ωστόσο η συνάρτηση πυκνότητας που αντιστοιχεί στη ροπογεννήτρια $(1 - b_j \cdot t)^{-r}$ είναι η $\frac{y^{r-1} \cdot e^{-y/b_j}}{b_j^r \cdot \Gamma(r)}$ για $y > 0$ και ίση με το μηδέν, αλλού. Επομένως για την πυκνότητα της τυχαίας μεταβλητής Y , την οποία ονομάζουμε $g(y)$ θα έχουμε ότι:

$$g(y) = \left[\prod_{j=1}^n (-b_j)^{-a_j} \right] \cdot \sum_{j=1}^n \sum_{r=1}^{a_j} b_{jr} \cdot (-1)^{-r} \cdot \frac{y^{r-1} \cdot e^{-\frac{y}{b_j}}}{(r-1)!}, \quad \text{για } y > 0$$

και $g(y) = 0$, αλλού. Συνεπώς παρατηρούμε ότι πρόκειται για ένα πεπερασμένο άθροισμα το οποίο περιλαμβάνει πυκνότητες Γάμμα κατανομών.

Στην συνέχεια όμως θα πρέπει να εκτιμήσουμε τους συντελεστές b_{jr} μέσω της ακόλουθης τεχνικής. Εφόσον θεωρούμε ότι οι παράμετροι a_j είναι ακέραιοι, χρησιμοποιώντας γνωστά αποτελέσματα τα οποία ισχύουν για τα παραγοντικά μπορούμε να γράψουμε ότι:

$$b_{jr} = \lim_{t \rightarrow 1/b_j} \left\{ \frac{1}{(a_j - r)!} \cdot \frac{\partial a_j - r}{\partial t} \left[(t - 1/b_j)^{-a_j} \cdot \prod_{i=1}^n (t - 1/b_i)^{-a_i} \right] \right\}$$

$$\text{Έστω } \Delta(t) = \prod_{\substack{i=1 \\ i \neq j}}^n (t - 1/b_i)^{-a_i} \quad \text{και} \quad A(t) = \frac{\partial}{\partial t} \log \Delta(t) = - \sum_{\substack{i=1 \\ i \neq j}}^n \left[\frac{a_i}{t - 1/b_i} \right]$$

τότε ισχύει:

$$\frac{\partial}{\partial t} \Delta(t) = A(t) \cdot \Delta(t)$$

και επομένως:

$$\frac{\partial a_j - r}{\partial t} \Delta(t) = \frac{\partial a_j - r}{\partial t} [A(t) \cdot \Delta(t)] = \sum_{j_1=0}^{a_j - r - 1} \binom{a_j - r - 1}{j_1} A^{(a_j - r - 1 - j_1)} \Delta^{(j_1)}$$

όπου $A^{(s)}$ και $\Delta^{(s)}$ θα είναι η μερική παράγωγος τάξης s των $A(t)$ και $\Delta(t)$ ως προς t . Έχουμε ότι $A^{(0)} = A(t)$ και $\Delta^{(0)} = \Delta(t)$ ενώ παράλληλα για το πλήθος των συνδυασμών στη παραπάνω σχέση ισχύει ο τύπος: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ και $0! = 1$.

Συνεπώς μπορούμε ύστερα από πράξεις (βλέπε Mathai, 1982), και εκφράζοντας όλες τις προηγούμενες ποσότητες στο χρόνο $t = 1/b_j$ να έχουμε πλέον τον εξής τύπο για τους συντελεστές b_{jr} :

$$b_{jr} = \left\{ \sum_{j_1=0}^{a_j-r-1} \binom{a_j-r-1}{j_1} A_j^{(a_j-r-1-j_1)} \cdot \sum_{j_2=0}^{j_1-1} \binom{j_1-1}{j_2} A_j^{(j_1-1-j_2)} \dots \dots \dots \right\} \cdot \Delta_j / (a_j - r)!$$

Όπου $\Delta_j = \prod_{i=1, i \neq j}^n (1/b_j - 1/b_i)^{-a_i}$ και $A_j^{(s)} = (-1)^{s+1} \cdot s! \cdot \sum_{i=1, i \neq j}^n a_i \cdot (1/b_j - 1/b_i)^{-(s+1)}$

Για παράδειγμα όταν έχουμε $a_j = 3$ τότε $b_{j3} = \Delta_j$, $b_{j2} = A_j \Delta_j$ και $b_{j1} = [A_j^{(1)} + (A_j)^2] \cdot \Delta_j / 2$.

Επομένως μπορούμε να έχουμε μια συνάρτηση πυκνότητας για την συνέλιξη n Γάμμα κατανομών στην περίπτωση όπου όλες οι παράμετροι σχήματος είναι ακέραιοι, αρκεί όμως να μπορούμε εύκολα να παράγουμε τους συντελεστές b_{jr} τους οποίους χρειαζόμαστε για το πεπερασμένο άθροισμα που μας δίνει την πυκνότητα της τ.μ. Y .

5.2 Προσέγγιση για την συνέλιξη Γάμμα κατανομών όταν δεν υπάρχει κλειστός τύπος

Όπως είναι ιδιαίτερα γνωστόν, η ακριβής συνάρτηση της πυκνότητας πιθανότητας για μια συνέλιξη Γάμμα κατανομών είναι αρκετές φορές περίπλοκη ενώ παράλληλα δεν επιδέχεται κλειστή μορφή. Από όλες τις συναρτήσεις πυκνότητας που έχουν προταθεί, η πιο εύκολη στην χρήση της είναι η σειρά Γάμμα κατανομών που πρότεινε το 1985 ο Moschopoulos. Ωστόσο για συγκεκριμένους συνδυασμούς παραμέτρων σχήματος και κλίμακας, ο υπολογισμός των βαρών που βρίσκονται μέσα στην σειρά αποδείχθηκε ότι μπορεί να είναι ιδιαίτερα χρονοβόρος και πολύπλοκος. Το 2009, οι Vellaisamy και Upadhye πρότειναν μια νέα προσέγγιση για την σειρά του Moschopoulos (1985). Η προσέγγιση αυτή έδωσε το έναυσμα ώστε να ερμηνευτούν τα βάρη της σειράς ως πιθανότητες μιας μη πεπερασμένης διακριτής κατανομής. Για τον σκοπό αυτό, λόγω της απλότητας της σειράς που παρουσίασε ο Moschopoulos και λόγω της έκφρασης των βαρών ως πιθανότητες, μελέτησε το 2017 ο Barnabani την προσέγγιση της συνέλιξης Γάμμα κατανομών μέσω της προσέγγισης των βαρών της σειράς από μια διακριτή κατανομή, με στόχο τον γρήγορο υπολογισμό της συνέλιξης και την εφαρμογή ενός μεγάλου εύρους παραμέτρων. Για δύο τυχαίες μεταβλητές στην συνέλιξη, τα βάρη μπορούν να εκφραστούν με μεγάλη ακρίβεια από μια αρνητική διωνυμική κατανομή με αποτέλεσμα να έχουμε έναν ακριβή τύπο για την συνέλιξη. Στη περίπτωση όμως της ύπαρξης περισσότερων από δύο τυχαίων μεταβλητών στη συνέλιξη τότε για την προσέγγιση των βαρών προτάθηκε η γενικευμένη αρνητική διωνυμική κατανομή των Jain και Consul (1971).

Στην ενότητα αυτή, αρχικά θα δούμε *θεωρητικά* την ακριβή κατανομή όπως δόθηκε από τον Moschopoulos (1985) αλλά και την προσέγγιση της ενώ στην συνέχεια για συγκεκριμένες τιμές των παραμέτρων θα προσπαθήσουμε να συγκρίνουμε με διάφορες μεθόδους τις δύο κατανομές και να αξιολογήσουμε τα αποτελέσματά μας αντίστοιχα.

5.2.1 Η ακριβής συνάρτηση πυκνότητας για την συνέλιξη Γάμμα κατανομών

Θα δούμε αρχικά σε αυτή την υποενότητα, τον πρώτο τύπο που δόθηκε από τον Mathai (1982) για την πυκνότητα της συνέλιξης Γάμμα κατανομών στην γενική περίπτωση όπου έχουμε διάφορες τιμές των παραμέτρων σχήματος και κλίμακας. Έστω λοιπόν X_1, X_2, \dots, X_n , n ανεξάρτητες τυχαίες μεταβλητές οι οποίες ακολουθούν Γάμμα κατανομή με παραμέτρους σχήματος $a_i > 0$ και παραμέτρους κλίμακας $b_i > 0$, $i = 1, 2, \dots, n$ και $Y = \sum_{i=1}^n X_i$, η συνέλιξη ανεξάρτητων Γάμμα μεταβλητών. Τότε σύμφωνα με τον Mathai (1982) η πυκνότητα της Y μπορεί να εκφραστεί ως μια σειρά με άπειρα αθροίσματα ως εξής:

$$f(x) = [\prod_{j=1}^n b_j^{a_j} \cdot \Gamma(\gamma)]^{-1} \cdot x^{\gamma-1} \cdot e^{-\frac{x}{b_1}} \cdot \varphi(a_2, \dots, a_n; \gamma; (\frac{1}{b_1} - \frac{1}{b_2}) \cdot x, \dots, (\frac{1}{b_1} - \frac{1}{b_n}) \cdot x)$$

όπου $b_1 = \min_j(b_j)$, $\gamma = \sum_{j=1}^n a_j$ και φ είναι μια υπεργεωμετρική συνάρτηση με $n-1$ μεταβλητές οι οποίες ορίζονται μέσω μιας πολλαπλής σειράς:

$$\begin{aligned} & \varphi(a_2, \dots, a_n; \gamma; (\frac{1}{b_1} - \frac{1}{b_2}) \cdot x, \dots, (\frac{1}{b_1} - \frac{1}{b_n}) \cdot x) = \\ & = \sum_{r_2=0}^{\infty} \dots \sum_{r_n=0}^{\infty} \{(a_2)_{r_2} \dots (a_n)_{r_n} \cdot [(\frac{1}{b_1} - \frac{1}{b_2}) \cdot x]^{r_2} \dots [(\frac{1}{b_1} - \frac{1}{b_n}) \cdot x]^{r_n} / [r_2! \dots r_n! (\gamma)_{r_2+\dots+r_n}]\} \end{aligned}$$

με $(a_n)_{r_n} = a_n \cdot (a_n + 1) \dots (a_n + r_n - 1)$ και $(\gamma)_r = \gamma \cdot (\gamma + 1) \dots (\gamma + r - 1)$ τα οποία ονομάζονται σύμβολα Pochhammer.

Ωστόσο στην περίπτωση όπου έχουμε συνέλιξη 2 Γάμμα τυχαίων μεταβλητών, δηλαδή $n = 2$ τότε η παραπάνω πυκνότητα μπορεί να εκφραστεί ως μια ποσότητα που περιέχει την υπεργεωμετρική συνάρτηση Kummer (${}_1F_1$) (Abramowitz και Stegun, 1972) :

$$\begin{aligned} f(x; (a_1, b_1), (a_2, b_2)) &= \frac{x^{\gamma-1} \cdot e^{-x/b_1}}{b_1^{a_1} \cdot b_2^{a_2} \cdot \Gamma(\gamma)} \cdot {}_1F_1\left(a_2; \gamma; (\frac{1}{b_1} - \frac{1}{b_2}) \cdot x\right) = \\ &= \left(\frac{b_1}{b_2}\right)^{a_2} \cdot \text{Gamma}(x; \gamma; b_1) \cdot {}_1F_1\left(a_2; \gamma; (\frac{1}{b_1} - \frac{1}{b_2}) \cdot x\right) \end{aligned}$$

η οποία υπεργεωμετρική συνάρτηση Kummel μπορεί να υπολογισθεί αρκετά εύκολα.

Στην παραπάνω έκφραση της πυκνότητας δεν είναι απαραίτητο να ισχύει πάντα $b_1 < b_2$ καθώς ακόμη και όταν $b_1 > b_2$ τότε έχουμε:

$$\begin{aligned} f(x; (a_1, b_1), (a_2, b_2)) &= f(x; (a_2, b_2), (a_1, b_1)) \\ &= \frac{x^{\gamma-1} \cdot e^{-\frac{x}{b_2}}}{b_1^{a_1} \cdot b_2^{a_2} \cdot \Gamma(\gamma)} \cdot {}_1F_1\left(a_1; \gamma; (\frac{1}{b_2} - \frac{1}{b_1}) \cdot x\right) \\ &= \frac{x^{\gamma-1} \cdot e^{-\frac{x}{b_2}}}{b_1^{a_1} \cdot b_2^{a_2} \cdot \Gamma(\gamma)} \cdot e^{-\left(\frac{1}{b_2} - \frac{1}{b_1}\right) \cdot x} \cdot {}_1F_1\left(a_2; \gamma; (\frac{1}{b_1} - \frac{1}{b_2}) \cdot x\right) \end{aligned}$$

$$= \frac{x^{\gamma-1} \cdot e^{-x/b_1}}{b_1^{a_1} \cdot b_2^{a_2} \cdot \Gamma(\gamma)} \cdot {}_1F_1\left(a_2; \gamma; \left(\frac{1}{b_1} - \frac{1}{b_2}\right) \cdot x\right),$$

όπου ισχύει η ιδιότητα ότι: ${}_1F_1(a; b; c) = e^c \cdot {}_1F_1(b - a; b; -c)$.

Ωστόσο στην συνέχεια θα χρησιμοποιήσουμε κυρίως τον ακριβή τύπο του *Moschopoulos* (1985), ο οποίος έδωσε μια πιο εύκολη έκφραση για την συνέλιξη μέσω μιας Γάμμα σειράς.

Σύμφωνα με τον *Moschopoulos* (1985), η ακριβής πυκνότητα για την συνέλιξη Γάμμα κατανομών μπορεί να εκφραστεί ως ένα άπειρο άθροισμα.

Έστω οι ανεξάρτητες και τυχαίες μεταβλητές X_i , με $i = 1, 2, \dots, n$, οι οποίες έχουν συνάρτηση πυκνότητας πιθανότητας :

$$f(x_i) = \frac{x_i^{a_i-1} \cdot e^{-x_i/b_i}}{b_i^{a_i} \cdot \Gamma(a_i)}, \quad a_i > 0, b_i > 0, x_i > 0$$

και $f(x_i) = 0$, αλλού.

Τότε ορίζουμε την τυχαία μεταβλητή Y ως το άθροισμα $Y = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$. Ισχύει ότι $Y \sim \text{Gamma}(\rho + K, b_{min})$ όπου $\rho = \sum_{i=1}^n a_i > 0$, $b_{min} = \min_i(b_i)$ και K είναι μια τυχαία μεταβλητή με $\Pr(K = k) = w_k = C \cdot d_k$, $k = 0, 1, 2, \dots$ και $d_0 = 1$.

Επιπλέον θεωρούμε ότι $C = \prod_{i=1}^n (b_{min}/b_i)^{a_i}$ και $d_k = (1/k) \sum_{i=1}^k i \cdot g_i \cdot d_{k-i}$ με

$$g_i = (1/i) \sum_{j=1}^n a_j \cdot (1 - b_{min}/b_j)^i$$

Δίχως βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι $b_{min} = b_1$. Τότε η (περιθώρια) συνάρτηση πυκνότητας της Y μπορεί να εκφραστεί ως εξής (βλέπε επίσης *Vellaisamy* και *Upadhye*, 2009) :

$$f(y) = \sum_{k=0}^{\infty} w_k \cdot \frac{y^{\rho+k-1} \cdot e^{-y/b_1}}{b_1^{\rho+k} \cdot \Gamma(\rho + k)} \quad (5.2.1)$$

Η παραπάνω συνάρτηση πιθανότητας ορίζεται όταν υπάρχει τουλάχιστον ένα b_i το οποίο να διαφέρει από τα υπόλοιπα. Θεωρούμε ότι έστω αυτή η παράμετρος είναι η b_1 . Επιπλέον με το σύμβολο n , αναφερόμαστε στον αριθμό των τυχαίων μεταβλητών της συνέλιξης που ακολουθούν Γάμμα κατανομή και έχουν διαφορετικές παραμέτρους κλίμακας και όχι στον συνολικό αριθμό των τυχαίων μεταβλητών που βρίσκονται στην συνέλιξη. Οι τυχαίες μεταβλητές οι οποίες ακολουθούν Γάμμα κατανομή με κοινή παράμετρο κλίμακας συγχωνεύονται στον παραπάνω τύπο.

Επιπλέον λόγω της ομοιόμορφης σύγκλισης της (5.2.1), μπορούν το ολοκλήρωμα και το άθροισμα να εναλλαχθούν με αποτέλεσμα η αθροιστική συνάρτηση κατανομής της Y να γραφθεί ως εξής (*Moschopoulos*, 1985):

$$F(t) = \sum_{k=0}^{\infty} w_k \cdot \int_0^t \frac{y^{\rho+k-1} \cdot e^{-y/b_1}}{b_1^{\rho+k} \cdot \Gamma(\rho+k)} dy \quad (5.2.2)$$

Ωστόσο για λόγους ευκολίας και για πρακτικούς σκοπούς θα πρέπει η διακριτή κατανομή (k, w_k) μετά από πχ. m όρους να περικόπτεται, όπου m να είναι τέτοιο ώστε να επιτυγχάνεται η καλύτερη δυνατή ακρίβεια για τον υπολογισμό της πυκνότητας της συνέλιξης.

Η εφαρμογή των σχέσεων (5.2.1) και (5.2.2) είναι ιδιαίτερα απλή και εύκολη αλλά ο υπολογισμός τους μπορεί να γίνει αρκετά χρονοβόρος ιδίως όταν έχουμε ένα πολύ μεγάλο εύρος τιμών για τις παραμέτρους κλίμακας και όταν οι παράμετροι σχήματος παίρνουν μικρές τιμές αντίστοιχα. Αυτό έχει ως αποτέλεσμα η χρήση των παραπάνω εκφράσεων να είναι δύσκολη όταν πρόκειται για τις παραπάνω περιπτώσεις.

Ύστερα από αρκετές προσομοιώσεις στην ακριβή κατανομή (Barnabani, 2017) διαπιστώθηκε ότι οι αργοί υπολογισμοί αλλά και οι δυσκολίες στην εφαρμογή οφείλονταν κυρίως στην διακριτή κατανομή (k, w_k) , $k = 0, 1, 2, \dots$. Το γεγονός αυτό δημιούργησε την ανάγκη ώστε ο Barnabani (2017) να ψάξει για μια θεωρητική προσέγγιση της διακριτής κατανομής των βαρών της σειράς που παρουσίασε ο Moschoroulos (1985) με σκοπό τον πιο εύκολο και γρήγορο υπολογισμό των συναρτήσεων πυκνότητας πιθανότητας για συνελίξεις Γάμμα κατανομών.

5.2.2 Ο προσεγγιστικός τύπος για την κατανομή των βαρών στη συνέλιξη

Σύμφωνα με τον Barnabani (2017), θεωρούμε αρχικά το άθροισμα δύο τυχαίων μεταβλητών που ακολουθούν την Γάμμα κατανομή.

Σε αυτή την περίπτωση θα έχουμε ότι: $C = (b_1/b_2)^{a_2}$, $g_i = (1/i) \cdot a_2 \cdot p^i$, $i = 1, 2, 3, \dots$ με $p = (1 - b_1/b_2)$ και επιπλέον $d_k = p^k \cdot (1/k!) \cdot (a_2)_k$, $(a_2)_0 = 1$ όπου ως $(a_2)_k$ θεωρούμε το σύμβολο Pochhammer,

$$(a_2)_k = a_2 \cdot (a_2 + 1) \cdot \dots \cdot (a_2 + k - 1) .$$

Για την κατανομή των βαρών w_k , εάν πολλαπλασιάσουμε το $C \cdot d_k$ τότε θα έχουμε ότι:

$$\Pr(K = k) = (1/k!) \cdot (a_2)_k \cdot (1 - p)^{a_2} \cdot p^k$$

η οποία είναι μια αρνητική διωνυμική κατανομή.

Επομένως για $n = 2$ στην συνέλιξη Γάμμα κατανομών θα ισχύει ότι $K \sim NB(a_2, p)$, $k = 0, 1, 2, \dots$ με $p = (1 - b_1/b_2)$. Συνεπώς για $n = 2$ για την πυκνότητα της Y έχουμε:

$$f(y) = \sum_{k=0}^{\infty} w_k \cdot \frac{y^{\rho+k-1} \cdot e^{-y/b_1}}{b_1^{\rho+k} \cdot \Gamma(\rho+k)}, \quad w_k \sim NB(a_2, p) \text{ και } p = (1 - b_1/b_2) \quad (5.2.3)$$

Ωστόσο με τη χρήση αλγεβρικών ιδιοτήτων μπορούμε εναλλακτικά την πυκνότητα της συνέλιξης δύο τυχαίων μεταβλητών που ακολουθούν Γάμμα κατανομή να την γράψουμε ως:

$$f(y) = (1 - p)^{a_2} \cdot \text{Gamma}(\rho, b_1; y) \cdot {}_1F_1(a_2, \rho; \frac{py}{b_1})$$

όπου ${}_1F_1(a_2, \rho; \frac{py}{b_1}) = \sum_{k=0}^{\infty} (t^k/k!)(a_2)_k/(\rho)_k$ με $t = \frac{py}{b_1}$. Η συνάρτηση αυτή ονομάζεται υπεργεωμετρική συνάρτηση (confluent hypergeometric function).

Θα θεωρήσουμε ωστόσο την πιο γενική περίπτωση όπου θα έχουμε το άθροισμα περισσοτέρων από 2 τυχαίων μεταβλητών στη συνέλιξη.

Στην περίπτωση αυτή, δεν μπορούμε να βρούμε μια διακριτή κατανομή για τα ζεύγη (k, w_k) , $k = 0, 1, 2, \dots$ και για τον σκοπό αυτό έχει προταθεί μια θεωρητική προσέγγιση για την κατανομή των βαρών (Barnabani, 2017).

Θεωρούμε την πρώτη ροπή για την τυχαία μεταβλητή K με $\Pr(K = k) = w_k = C \cdot d_k$ για την οποία εξ ορισμού ισχύει ότι

$$E(K) = \sum_{k=0}^{\infty} k \cdot \Pr(K = k) = C \cdot \sum_{k=0}^{\infty} k \cdot d_k$$

όπου τα d_k μπορούν να υπολογιστούν αναδρομικά όπως είδαμε προηγουμένως εφόσον πρώτα γίνει υπολογισμός των ποσοτήτων g_i . Ωστόσο η ποσότητα g_i όπως την ορίσαμε στην ακριβή πυκνότητα, εξαρτάται από την ποσότητα $1 - b_{min}/b_j = 1 - b_1/b_j$. Επομένως θεωρούμε μια παράμετρο p με $0 < p < 1 \forall j$ ώστε να προσεγγίσουμε την $1 - b_1/b_j$. Για τον σκοπό αυτό αντικαθιστούμε τα b_j με μια ψευδομεταβλητή \bar{b} ώστε η ποσότητα $1 - b_1/b_j$ να είναι μια σταθερά και ίση με p . Συνεπώς για την έκφραση των g_i έχουμε ότι:

$$g_i = (1/i) \sum_{j=1}^n a_j \cdot (1 - b_1/b_j)^i \cong \rho \cdot \frac{p^i}{i}$$

όπου $\rho = \sum_{i=1}^n a_i > 0$.

Παράλληλα προσεγγίζουμε τις ποσότητες d_k ως εξής:

$$d_k = (1/k) \sum_{i=1}^k i \cdot g_i \cdot d_{k-i} \cong (\rho/k) \sum_{i=1}^k p^i \cdot d_{k-i}, k = 1, 2, \dots$$

από την οποία σχέση προκύπτει άμεσα ότι: $d_k \cong p^k \cdot \frac{(\rho)_k}{k!}$

όπου $(\rho)_k = \rho \cdot (\rho + 1) \cdot \dots \cdot (\rho + k - 1) = \frac{\Gamma(\rho+k)}{\Gamma(\rho)}$.

Επομένως έχουμε για την πρώτη ροπή:

$$E(K) = C \cdot \sum_{k=0}^{\infty} k \cdot d_k = (1 - p)^\rho \cdot \sum_{k=0}^{\infty} k \cdot p^k \cdot \frac{\Gamma(\rho+k)}{k!} = \frac{p \cdot \rho}{1 - p}$$

με $C = \prod_{i=1}^n (b_1/b_i)^{a_i} = (1 - (1 - b_1/b_i)^{a_1}) \cdot (1 - (1 - b_1/b_i)^{a_2}) \dots (1 - (1 - b_1/b_i)^{a_n})$
 άρα $C = (1 - p)^{\sum_{i=1}^n a_i} = (1 - p)^\rho$.

Παρατηρούμε ότι σε αυτή τη περίπτωση η πρώτη ροπή της τ.μ. K είναι ίση με την αναμενόμενη τιμή μιας αρνητικής διωνυμικής κατανομής με παραμέτρους ρ και p .

Μέσω του παραπάνω αποτελέσματος δίνεται το έναυσμα ώστε να προσεγγιστεί η κατανομή $(k, w_k), k = 0, 1, 2, \dots$ από μια αρνητική διωνυμική κατανομή με παράμετρο p , η οποία παράμετρος θα πρέπει να είναι τέτοια ώστε να ισχύει η ισότητα μεταξύ των ροπών της τ.μ. $Y|K$ και της τ.μ. Y ότι δηλαδή $E_K E_Y(Y|K) = E(Y)$, άρα

$$\mu = \frac{p \cdot \rho}{1 - p} \Rightarrow p = \frac{\mu}{\mu + \rho} = 1 - \frac{b_1}{\bar{b}}, \quad (5.2.4)$$

όπου $\mu = \sum_{i=1}^n \frac{b_i}{b_1} \cdot (1 - \frac{b_1}{b_i}) \cdot a_i$. Επίσης μέσω διαφόρων αλγεβρικών αποτελεσμάτων μπορεί να προκύψει ότι ισοδύναμα το \bar{b} είναι ένας σταθμισμένος μέσος των b_i με βάρη a_i αντίστοιχα, δηλαδή $\bar{b} = \sum_{i=1}^n a_i \cdot b_i / \sum_{i=1}^n a_i$.

Η προσέγγιση αυτή είναι ιδιαίτερα ικανοποιητική όταν το \bar{b} παίρνει τιμές κοντά σε αυτές των b_i ή όταν γενικά οι παράμετροι b_i δεν παίρνουν πολύ διαφορετικές τιμές μεταξύ τους. Σύμφωνα με τον Barnabani (2017) όταν οι παράμετροι b_i παίρνουν ένα μεγάλο εύρος τιμών τότε η προσέγγιση για την κατανομή των w_k ($w_k \sim NB(\rho, p)$) δεν είναι αρκετά επικοδομητική για την συνέλιξη.

Συνεπώς, μια καλή προσέγγιση για την συνέλιξη θα πρέπει να λαμβάνει υπόψιν την μεταβλητότητα των παραμέτρων κλίμακας b_i . Για τον σκοπό αυτό θα χρησιμοποιήσουμε τις παρακάτω ισότητες που περιέχουν τις δύο πρώτες ροπές, με τη χρήση παραμέτρων r και p .

$$\begin{aligned} E_K[E_Y(Y|K)] &= E(Y) \\ E_K(Var_Y(Y|K)) + Var_K(E_Y(Y|K)) &= Var(Y) \end{aligned}$$

Οι δείκτες κάτω από την μέση τιμή και τη διακύμανση υποδεικνύουν για ποια μεταβλητή γίνεται ο υπολογισμός.

Σχετικά με τα r, p πιο συγκεκριμένα έχουμε:

$$\left\{ \begin{array}{l} \mu = \frac{p \cdot r}{1 - p} \\ \mu_2 = \frac{p \cdot \rho}{1 - p^2} \end{array} \right. \Rightarrow \left\{ \begin{array}{l} p = 1 - \frac{\mu}{\mu_2} \\ r = \frac{1 - p}{p} \cdot \mu \end{array} \right. \quad (5.2.5)$$

με $\mu_2 = \sum_{i=1}^n (\frac{b_i}{b_1})^2 \cdot (1 - \frac{b_1}{b_i}) \cdot a_i$

Για λόγους ευκολίας όταν η κατανομή των βαρών w_k προσεγγίζεται από μια αρνητική διωνυμική, τότε την πυκνότητα της συνέλιξης των Γάμμα κατανομών θα την συμβολίζουμε ως $g_{nb}(y)$ και την αθροιστική συνάρτηση κατανομής ως $F_{nb}(y)$ αντίστοιχα. Θα χρησιμοποιήσουμε αυτούς τους συμβολισμούς στη συνέχεια μέσω του παραδείγματος που θα κάνουμε ώστε να συγκρίνουμε την προσέγγιση της συνέλιξης με την ακριβή πυκνότητα.

Όπως ήδη αναφέραμε η αρνητική διωνυμική κατανομή για τα w_k είναι εξαιρετικά αποδοτική όταν οι παράμετροι κλίμακας δεν μεταβάλλονται αρκετά από την μία Γάμμα κατανομή στην άλλη και όταν οι παράμετροι σχήματος δεν παίρνουν τόσο μικρές τιμές. Σύμφωνα με τον Barnabani (2017), στην περίπτωση όπου υπάρχουν παράμετροι σχήματος με χαμηλές τιμές τότε η αριστερή ουρά της αρνητικής διωνυμικής κατανομής είναι πιο βαριά σε σχέση με την πραγματική κατανομή $(k, w_k), k = 0, 1, 2, \dots$ και η μεγαλύτερη διαφορά μεταξύ της προσέγγισης και της πραγματικής κατανομής των βαρών βρίσκεται κυρίως γύρω από την επικρατούσα τιμή.

Ωστόσο όταν η αρνητική διωνυμική κατανομή δεν είναι αρκετά αποτελεσματική ως προσέγγιση των κατανομών $(k, w_k), k = 0, 1, 2, \dots$, για να βελτιωθεί σημαντικά η προσεγγιστική κατανομή των βαρών w_k θα χρειαστεί να εκτιμηθεί και μια τρίτη παράμετρος εκτός των r, p . Για τον σκοπό αυτό προτάθηκε (Barnabani, 2017) η χρήση της γενικευμένης αρνητικής διωνυμικής κατανομής αντίστοιχα ως προσέγγιση της διακριτής κατανομής $(k, w_k), k = 0, 1, 2, \dots$, η οποία ορίστηκε για πρώτη φορά από τους Jain και Consul (1971).

Η συνάρτηση πιθανότητας για την γενικευμένη αρνητική διωνυμική κατανομή με 3 παραμέτρους δίνεται μέσω του ακόλουθου τύπου ως εξής:

$$GNB(k; r, p, b) = \begin{cases} \frac{r}{r + bk} \binom{r + bk}{k} p^k \cdot (1 - p)^{r + bk - k}, & k = 0, 1, 2, \dots \\ 0, & k \geq \mu, r + b\mu < 0 \end{cases}$$

για $0 < p < 1$ και $|p \cdot b| < 1$. Στην απλή περίπτωση όπου $b = 1$, τότε η γενικευμένη αρνητική διωνυμική κατανομή ανάγεται στην αρνητική διωνυμική κατανομή. Επίσης σύμφωνα με τους Consul και Gupta (1980), όταν $0 < b < 1$ τότε η γενικευμένη διωνυμική κατανομή δεν υπάρχει.

Ύστερα από πράξεις (βλέπε Barnabani, 2017) και με την χρήση των τριών πρώτων ροπών οι τρεις παράμετροι για την γενικευμένη αρνητική διωνυμική εκτιμώνται ως εξής:

$$\begin{aligned}
p &= 1 - \frac{A}{2} + \sqrt{\frac{A^2}{4} - 1} \\
b &= \frac{1}{p} \left[1 - \left(\frac{\mu \cdot (1-p)}{\mu_2} \right)^{\frac{1}{2}} \right] \\
r &= \mu \cdot \left(\frac{1-pb}{p} \right)
\end{aligned} \tag{5.2.6}$$

όπου $A = -2 + (\mu \cdot \mu_3 - 3 \cdot \mu_2^2)^2 / (\mu \cdot \mu_2^3)$. Θα πρέπει ωστόσο να ισχύει ότι $A \geq 2$, δηλαδή $(\mu \cdot \mu_3 - 3 \cdot \mu_2^2)^2 / (\mu \cdot \mu_2^3) \geq 4$ ώστε οι παραπάνω παράμετροι να ανήκουν στο σύνολο των πραγματικών αριθμών. Εάν $A < 2$ τότε θέτουμε $b = 1$ και συνεπώς θα έχουμε αρνητική διωνυμική κατανομή αντίστοιχα. Παράλληλα η τρίτη κεντρική ροπή προσεγγίζεται ως εξής:

$$\mu_3 = \sum_{i=1}^n \left(\frac{b_i}{b_1} \right)^3 \cdot \left(1 - \frac{b_1}{b_i} \right) \cdot \left(2 - \frac{b_1}{b_i} \right) \cdot a_i \tag{5.2.7}$$

Ωστόσο θα ήταν χρήσιμο να δούμε μέσω ενός αριθμητικού παραδείγματος, τι τιμές παίρνει η ποσότητα A για συγκεκριμένες κατανομές Γάμμα στην συνέλιξη. Όπως αναφέραμε προηγουμένως είναι απαραίτητο να ισχύει η ανισότητα $A \geq 2$ ώστε οι παράμετροι της γενικευμένης διωνυμικής κατανομής που χρησιμοποιούμε για την προσέγγιση των βαρών να παίρνουν τιμές στο σύνολο \mathcal{R} .

Έστω η τυχαία μεταβλητή Y για την οποία ισχύει ότι: $Y = X_1 + X_2 + X_3 + X_4$ με $X_1 \sim \text{Gamma}(2,3)$, $X_2 \sim \text{Gamma}(0.5,1)$, $X_3 \sim \text{Gamma}(20,10)$ και $X_4 \sim \text{Gamma}(50,30)$

Θα υπολογίσουμε με τη βοήθεια του Mathematica την ποσότητα: $\mu = \sum_{i=1}^n \frac{b_i}{b_1} \cdot \left(1 - \frac{b_1}{b_i} \right) \cdot a_i$

Έχουμε:

$$\frac{b_1}{b_1} \cdot \left(1 - \frac{b_1}{b_1} \right) \cdot a_1 + \frac{b_2}{b_1} \cdot \left(1 - \frac{b_1}{b_2} \right) \cdot a_2 + \frac{b_3}{b_1} \cdot \left(1 - \frac{b_1}{b_3} \right) \cdot a_3 + \frac{b_4}{b_1} \cdot \left(1 - \frac{b_1}{b_4} \right) \cdot a_4 = 496.33$$

Στην συνέχεια υπολογίζουμε μέσω Mathematica την ποσότητα: $\mu_2 = \sum_{i=1}^n \left(\frac{b_i}{b_1} \right)^2 \cdot \left(1 - \frac{b_1}{b_i} \right) \cdot a_i$

Ισχύει:

$$\begin{aligned}
&\left(\frac{b_1}{b_1} \right)^2 \cdot \left(1 - \frac{b_1}{b_1} \right) \cdot a_1 + \left(\frac{b_2}{b_1} \right)^2 \cdot \left(1 - \frac{b_1}{b_2} \right) \cdot a_2 + \left(\frac{b_3}{b_1} \right)^2 \cdot \left(1 - \frac{b_1}{b_3} \right) \cdot a_3 + \left(\frac{b_4}{b_1} \right)^2 \cdot \left(1 - \frac{b_1}{b_4} \right) \cdot a_4 \\
&= 4655.44
\end{aligned}$$

Και τέλος υπολογίζουμε την ποσότητα: $\mu_3 = \sum_{i=1}^n \left(\frac{b_i}{b_1}\right)^3 \cdot \left(1 - \frac{b_1}{b_i}\right) \cdot \left(2 - \frac{b_1}{b_i}\right) \cdot a_i$

Προκύπτει:

$$\left(\frac{b_1}{b_1}\right)^2 \cdot \left(1 - \frac{b_1}{b_1}\right) \cdot \left(2 - \frac{b_1}{b_1}\right) \cdot a_1 + \left(\frac{b_2}{b_1}\right)^2 \cdot \left(1 - \frac{b_1}{b_2}\right) \cdot \left(2 - \frac{b_1}{b_2}\right) \cdot a_2 + \left(\frac{b_3}{b_1}\right)^2 \cdot \left(1 - \frac{b_1}{b_3}\right) \cdot \left(2 - \frac{b_1}{b_3}\right) \cdot a_3 + \left(\frac{b_4}{b_1}\right)^2 \cdot \left(1 - \frac{b_1}{b_4}\right) \cdot \left(2 - \frac{b_1}{b_4}\right) \cdot a_4 = 86381.5$$

Η ποσότητα: $\frac{(\mu \cdot \mu_3 - 3 \cdot \mu_2^2)^2}{\mu \cdot \mu_2^3} = 9.79293$ είναι μεγαλύτερη από 4 και συνεπώς για την συνέλιξη αυτή, οι παράμετροι θα ανήκουν στο σύνολο των πραγματικών αριθμών. Δηλαδή μπορεί η κατανομή των βαρών w_k να προσεγγιστεί από μια γενικευμένη αρνητική διωνυμική κατανομή.

Για την συνέλιξη επιλέξαμε κατανομές με παραμέτρους κλίμακας b_i που παίρνουν ένα μεγάλο εύρος τιμών και παρατηρούμε ότι η προσέγγιση από μια γενικευμένη αρνητική διωνυμική είναι εποικοδομητική, γεγονός το οποίο δεν ισχύει για την αρνητική διωνυμική.

Επιπλέον όταν η κατανομή των βαρών w_k προσεγγίζεται από μια γενικευμένη αρνητική διωνυμική κατανομή, τότε την πυκνότητα της συνέλιξης των Γάμμα κατανομών θα την συμβολίζουμε ως $g_{gnb}(y)$ και την αθροιστική συνάρτηση κατανομής ως $F_{gnb}(y)$ αντίστοιχα.

Σύμφωνα με την βιβλιογραφία, ο Stewart (2007) πρότεινε την προσέγγιση των συνελιξεων Γάμμα κατανομών από μια άλλη Γάμμα κατανομή της οποίας οι παράμετροι σχήματος και κλίμακας μπορούν να προκύψουν εάν εξισωθούν οι δύο πρώτες ροπές της συνέλιξης με την προσεγγιστική Γάμμα κατανομή. Αν και ο Barnabani (2017) ασχολείται και με αυτήν την προσέγγιση που πρότεινε ο Stewart (2007), στην παρούσα διπλωματική εργασία δεν θα αναφερθούμε στην συγκεκριμένη προσέγγιση.

5.2.3 Εφαρμογές μέσω ενός παραδείγματος συνέλιξης 3 Γάμμα κατανομών

Στην υποενότητα αυτή θα προσπαθήσουμε για μια συνέλιξη 3 Γάμμα κατανομών με συγκεκριμένες παραμέτρους σχήματος και κλίμακας να υπολογίσουμε την ακριβή πυκνότητα της. Στην συνέχεια με βάση την θεωρία που αναπτύξαμε στις προηγούμενες υποενότητες θα υπολογίσουμε έναν προσεγγιστικό τύπο όπου τα βάρη της σειράς της ακριβούς κατανομής θα ακολουθούν γενικευμένη αρνητική διωνυμική κατανομή. Όπως είδαμε και προηγουμένως η κατανομή αυτή είναι η καταλληλότερη, όταν υπάρχουν περισσότερες από 2 τυχαίες μεταβλητές στην συνέλιξη, σύμφωνα με τον Barnabani (2017), και για διάφορες τιμές των παραμέτρων (χωρίς να έχουμε περιορισμούς). Σκοπός μας είναι να διαπιστώσουμε μέσω διαφόρων κριτηρίων αν η προσέγγιση αυτή επιφέρει εποικοδομητικά και γρήγορα

αποτελέσματα και παράλληλα να εξετάσουμε αν οι προσεγγίσεις για την κατανομή των βαρών από μια γενικευμένη αρνητική διωνυμική κατανομή, είναι αρκετά κοντά στην πραγματική πυκνότητα της συνέλιξης.

Έστω η τυχαία μεταβλητή Y για την οποία ισχύει ότι: $Y = X_1 + X_2 + X_3$ με $X_1 \sim \text{Gamma}(3,6)$, $X_2 \sim \text{Gamma}(10,8)$ και $X_3 \sim \text{Gamma}(20,10)$.

Στο στατιστικό πρόγραμμα R, υπάρχει το πακέτο *coga* (Chaoran Hu, 2018) μέσω του οποίου μπορεί να γίνει αξιολόγηση και υπολογισμός της πυκνότητας και της συνάρτησης κατανομής της συνέλιξης ανεξάρτητων Γάμμα τυχαίων μεταβλητών. Στο πακέτο αυτό υπάρχουν οι δύο ακριβείς μέθοδοι των Mathai (1982) και Moschopoulos (1985) καθώς επίσης και η προσεγγιστική μέθοδος του Barnabani (2017). Θα δούμε πως λειτουργούν αντίστοιχα οι συναρτήσεις αυτές μέσω εντολών στο R.

Εφόσον έχουμε μια συνέλιξη 3 τυχαίων Γάμμα μεταβλητών, δεν θα χρησιμοποιήσουμε τον τύπο του Mathai ο οποίος είναι περισσότερο γρήγορος και αποτελεσματικός κυρίως στην περίπτωση της συνέλιξης 2 Γάμμα τυχαίων μεταβλητών. Υπάρχουν ωστόσο οι εντολές της R: *dcoga2dim* και *pcoga2dim* μέσω των οποίων μπορούν να υπολογισθούν η πυκνότητα και η συνάρτηση κατανομής της συνέλιξης με τη χρήση του τύπου του Mathai.

Για τον υπολογισμό της πυκνότητας και της συνάρτησης κατανομής όταν $n = 3$, όπως ισχύει και στο συγκεκριμένο παράδειγμα μας, θα γίνει χρήση της Γάμμα σειράς που πρότεινε ο Moschopoulos (1985).

Μπορούμε αρχικά στα σημεία 1 έως 15, να δούμε τι τιμές παίρνει η πυκνότητα και η συνάρτηση κατανομής για τον τύπο του Moschopoulos μέσω του κώδικα:

```
> dcoga(1:15,c(3,10,20),c(6,8,10))
```

```
> pcoga(1:15,c(3,10,20),c(6,8,10))
```

Επίσης πριν δημιουργήσουμε τα αντίστοιχα διαγράμματα, θα παραγάγουμε 10.000 παρατηρήσεις από την κατανομή του Moschopoulos ώστε να καθορίσουμε κυρίως το εύρος τιμών των παρατηρήσεων και στη συνέχεια θα υπολογίσουμε την πυκνότητα και την συνάρτηση κατανομής της συνέλιξης σε 100 σημεία που θα βρίσκονται στο εύρος αυτό (Chaoran Hu, 2017). Παράλληλα βλέπουμε στο Διάγραμμα 5.1, την πυκνότητα της συνέλιξης (αριστερά) και την συνάρτηση κατανομής (δεξιά). Οι μαύρες γραμμές στα διαγράμματα αντιστοιχούν στο δείγμα παρατηρήσεων που πήραμε ενώ οι κόκκινες γραμμές θα αντιστοιχούν στα αποτελέσματα των συναρτήσεων *dcoga* και *pcoga*.

Έχουμε τον ακόλουθο κώδικα στην R:

```
> set.seed(123)
```

```
> y <- rcoga(100000,c(3,10,20),c(6,8,10))
```

```
> grid <- seq(0,5,length.out=100)
```

```
> pdf <- dcoga(grid, c(3,10,20),c(6,8,10))
```



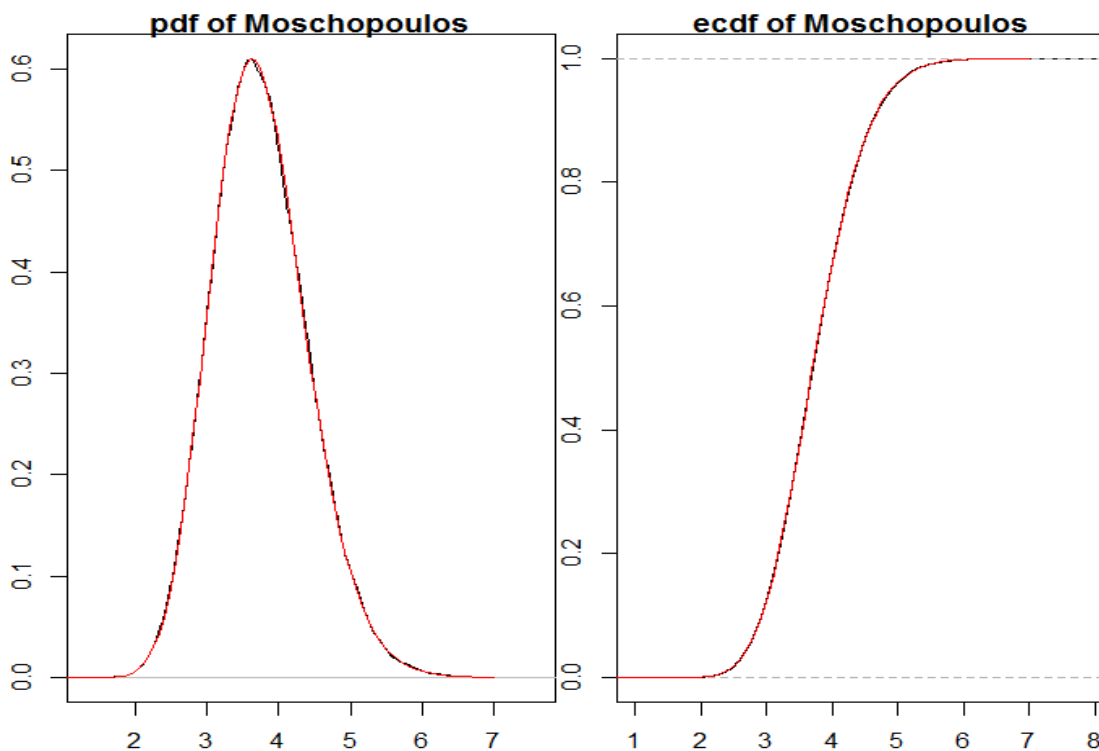
```

> cdf <- pcoga(grid, c(3,10,20),c(6,8,10))
> par (mfrow = c(1, 2), mar = c(2,2,1,0))
> plot (density(y), main="pdf of Moschopoulos", col="black")
> lines (grid, pdf, col="red")
> plot (ecdf (y), main="ecdf of Moschopoulos", col="black")
> lines (grid, cdf, col="red")

```

Και συνεπώς προκύπτει το παρακάτω διάγραμμα για το διάστημα $x \in (0,7)$:

Διάγραμμα 5.1: Πυκνότητα συνέλιξης των Gamma (3,6), Gamma (10,8) και Gamma (20,10) (αριστερά) και συνάρτηση κατανομής (δεξιά).



Ακολουθώντας την ίδια διαδικασία μπορούμε να υπολογίσουμε αντίστοιχα την πυκνότητα και την συνάρτηση κατανομής της συνέλιξης για τον προσεγγιστικό τύπο του Barnabani (2017) σύμφωνα με τον οποίο τα βάρη της σειράς Γάμμα του Moschopoulos (1985) ακολουθούν γενικευμένη αρνητική κατανομή. Οι εντολές στο πρόγραμμα R είναι: `dcoga_approx` και `pcoga_approx`.

Αρχικά θα δούμε στα σημεία 1 έως 20, τι τιμές παίρνει η πυκνότητα και η συνάρτηση κατανομής για τον τύπο του Barnabani μέσω του κώδικα:

```

> dcoga_approx (1:20,c(3,10,20),c(6,8,10))

```

```
> pcoga_approx (1:20,c(3,10,20),c(6,8,10))
```

Επιπρόσθετα, θα παρουσιάσουμε στο Διάγραμμα 5.2, την προσέγγιση της πυκνότητας της συνέλιξης (αριστερά) και την προσέγγιση της συνάρτησης κατανομής (δεξιά). Οι μπλε γραμμές στα διαγράμματα αντιστοιχούν στο δείγμα παρατηρήσεων που πήραμε με τον ίδιο ακριβώς τρόπο όπως προηγουμένως ενώ οι κόκκινες γραμμές θα αντιστοιχούν στα αποτελέσματα των συναρτήσεων `dcoga_approx` και `pcoga_approx`.

Θα κάνουμε χρήση των παρακάτω εντολών στο πρόγραμμα R:

```
> set.seed (123)
```

```
> y1<- rcoga (100000,c(3,10,20),c(6,8,10))
```

```
> grid<-seq(0,7,length.out=100)
```

```
> pdf1<-dcoga_approx(grid, c(3,10,20),c(6,8,10))
```

```
> cdf1<-pcoga_approx(grid, c(3,10,20),c(6,8,10))
```

```
> par(mfrow = c(1, 2), mar = c(2,2,1,0))
```

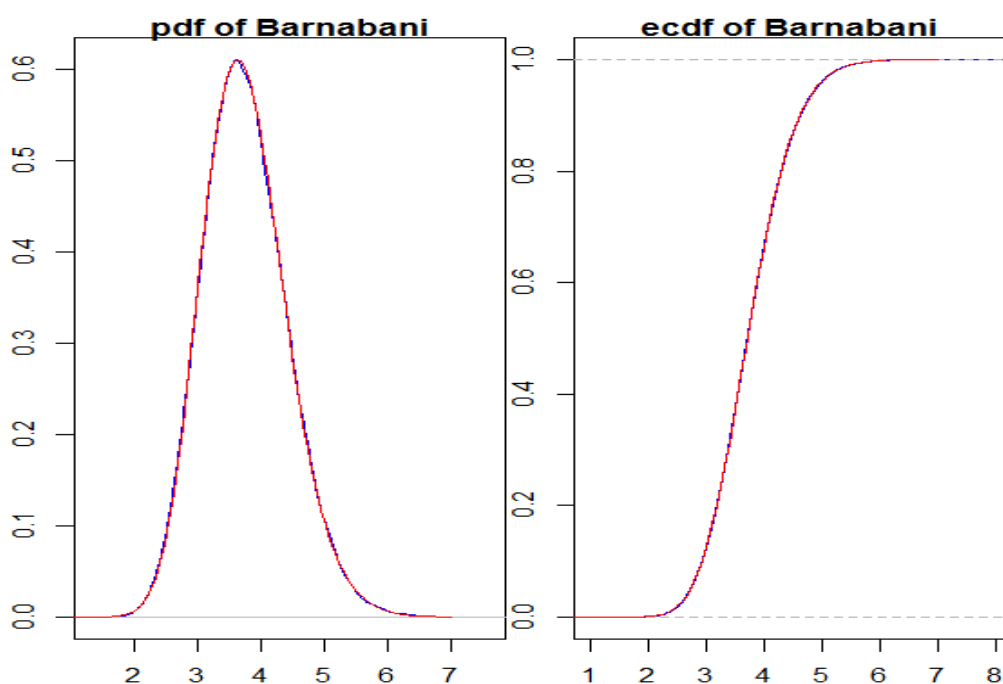
```
> plot(density(y1), main="pdf of Barnabani", col="blue")
```

```
> lines(grid, pdf1, col="red")
```

```
> plot(ecdf(y1), main="ecdf of Barnabani", col="blue")
```

```
> lines(grid, cdf1, col="red")
```

Διάγραμμα 5.2: Προσέγγιση της πυκνότητας της συνέλιξης των Gamma (3,6), Gamma (10,8) και Gamma (20,10) (αριστερά) και προσέγγιση της συνάρτησης κατανομής (δεξιά).



Παρατηρούμε ότι για $n = 3$ στην συνέλιξη Γάμμα κατανομών, η προσέγγιση που πρότεινε ο Barnabani φαίνεται να είναι ιδιαίτερα αποτελεσματική ενώ παράλληλα είναι μια πολύ γρήγορη μέθοδος υπολογισμού των συναρτήσεων πυκνότητας και κατανομής.

Ας προσπαθήσουμε στη συνέχεια να αξιολογήσουμε με διάφορα κριτήρια, την αντίστοιχη προσέγγιση ώστε να διαπιστώσουμε εάν όντως αποτελεί έναν εναλλακτικό τρόπο υπολογισμού της πυκνότητας του Moschoroulos (η οποία είναι αρκετά χρονοβόρα όταν οι παράμετροι κλίμακας παίρνουν μεγάλο εύρος τιμών π.χ. $(b_1, b_2) = (0.3, 10)$).

Μέγιστη διαφορά συναρτήσεων πυκνότητας

Ένα αρχικό κριτήριο αξιολόγησης της προσέγγισης είναι η μέγιστη διαφορά των συναρτήσεων πυκνότητας, η οποία προκύπτει από τον παρακάτω τύπο ως εξής:

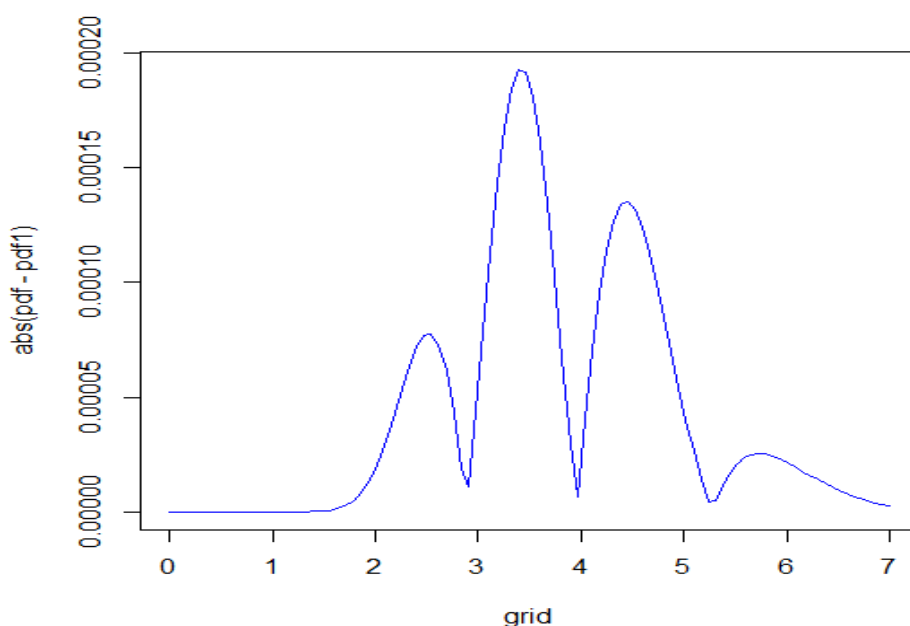
$$\sup_{x \geq 0} |g(y) - g_{gnb}(y)|$$

ενώ με την εντολή στην R: `max(abs(pdf-pdf1))`, βρίσκουμε ότι η μέγιστη τιμή που μπορεί να πάρει η απόλυτη διαφορά των δύο πυκνοτήτων (Moschoroulos και Barnabani) είναι 0.0001928652.

Στο διάγραμμα παρακάτω θα απεικονίσουμε γραφικά την απόλυτη διαφορά

$|g(y) - g_{gnb}(y)|$ στο διάστημα 0 έως 7.

Διάγραμμα 5.3: Γραφική απεικόνιση της απόλυτης διαφοράς μεταξύ της ακριβής πυκνότητας της συνέλιξης και της προσέγγισης στο διάστημα (0,7)



Επιβεβαιώνουμε και από το γράφημα ότι η μέγιστη τιμή της απόλυτης διαφοράς των πυκνοτήτων είναι κοντά στο 0.00020 ενώ παράλληλα παρατηρούμε και άλλα τοπικά μέγιστα στο διάστημα $x \in (0,7)$. Ωστόσο οι πολύ μικρές διαφορές ανάμεσα στις τιμές των πυκνοτήτων της προσέγγισης και της ακριβής κατανομής φανερώνουν μια ικανοποιητική προσέγγιση του τύπου του Barnabani.

Μέγιστη διαφορά συναρτήσεων κατανομής

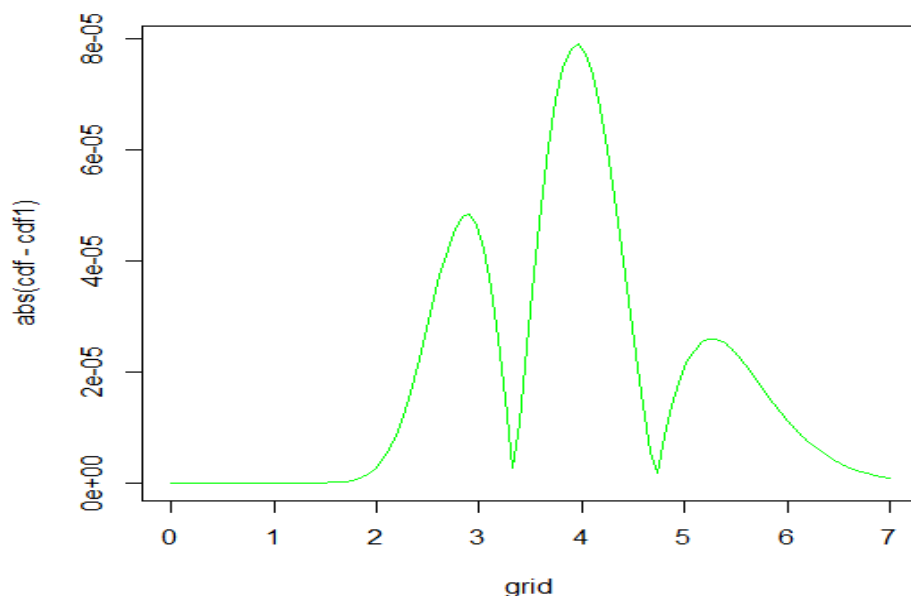
Ένας εξίσου παρόμοιος τρόπος αξιολόγησης της προσέγγισης του Barnabani, εκτός από την μέγιστη διαφορά των πυκνοτήτων όπου εξετάσαμε προηγουμένως, είναι η μέγιστη διαφορά των συναρτήσεων κατανομής. Ισχύει ο εξής τύπος:

$$\sup_{x \geq 0} |F(y) - F_{gnb}(y)|$$

Επιπλέον, όπως προηγουμένως με την εντολή στην R: `max(abs(cdf-cdf1))`, υπολογίζουμε τη μέγιστη τιμή που μπορεί να πάρει η απόλυτη διαφορά των δύο συναρτήσεων κατανομής. Η τιμή αυτή όπως προκύπτει είναι ένας πολύ μικρός αριθμός: 7.896021e-05.

Στην συνέχεια θα δούμε την γραφική αναπαράσταση για την απόλυτη διαφορά των συναρτήσεων κατανομής ($|F(y) - F_{gnb}(y)|$) στο διάστημα 0 έως 7.

Διάγραμμα 5.4: Γραφική απεικόνιση της απόλυτης διαφοράς μεταξύ της ακριβής συνάρτησης κατανομής της συνέλιξης και της προσέγγισης της στο διάστημα (0,7).



Στο παραπάνω διάγραμμα διακρίνουμε ότι η απόλυτη διαφορά μεταξύ των δύο συναρτήσεων κατανομής παίρνει πολύ μικρές έως αμελητέες τιμές ενώ παράλληλα η διαφορά άλλοτε

ακολουθεί ανοδική και άλλοτε καθοδική πορεία. Στο διάστημα $x \in (0,7)$ παρατηρούμε ότι υπάρχουν συνολικά 3 τοπικά μέγιστα και το ολικό μέγιστο όπως υπολογίσαμε και προηγουμένως φαίνεται και από την γραφική απεικόνιση ότι είναι η πολύ μικρή τιμή $7.896021e-05$ στο σημείο y περίπου κοντά στο 4. Ως συμπέρασμα καταλήγουμε στο γεγονός ότι οι διαφορές μεταξύ τόσο των συναρτήσεων πυκνότητας όσο και των συναρτήσεων κατανομής είναι ελάχιστες με αποτέλεσμα να προκύπτουν ικανοποιητικές προσεγγίσεις αντίστοιχα.

Σύγκριση των ροπών

Μια επίσης καλή μέθοδος ώστε να αξιολογήσουμε την προσέγγιση του Barnabani για την συνέλιξη Γάμμα κατανομών είναι εξετάζοντας τις ροπές για τις πυκνότητες του Moschoroulos (1985) και του Barnabani (2017) αντίστοιχα. Επειδή οι 2 πρώτες ροπές είναι περισσότερο κοντά η μια στην άλλη, έχει μεγαλύτερο νόημα να συγκρίνουμε την τρίτη, τέταρτη, πέμπτη και έκτη ροπή.

Τρίτη ροπή

Για την κατανομή του Moschoroulos ισχύει: $E(Y^3) = 57.78614$

Για την κατανομή του Barnabani ισχύει: $E(Y^3) = 57.78653$

Τέταρτη ροπή

Για την κατανομή του Moschoroulos ισχύει: $E(Y^4) = 237.0602$

Για την κατανομή του Barnabani ισχύει: $E(Y^4) = 237.0666$

Πέμπτη ροπή

Για την κατανομή του Moschoroulos ισχύει: $E(Y^5) = 1000.465$

Για την κατανομή του Barnabani ισχύει: $E(Y^5) = 1000.534$

Έκτη ροπή

Για την κατανομή του Moschoroulos ισχύει: $E(Y^6) = 4340.487$

Για την κατανομή του Barnabani ισχύει: $E(Y^6) = 4341.087$

Παρατηρούμε αντίστοιχα ότι τα αποτελέσματα των ροπών και για τις δύο πυκνότητες είναι παρόμοια μεταξύ χωρίς να εμφανίζονται σημαντικές ή μεγάλες διαφορές καθώς η τάξη της ροπής αυξάνει. Το γεγονός αυτό μας δείχνει ότι η ακριβής πυκνότητα του Moschopoulos φαίνεται να προσεγγίζεται αρκετά καλά από την προσεγγιστική πυκνότητα του Barnabani μέχρι και τον υπολογισμό της έκτης ροπής.

Σύγκριση άνω ποσοστημορίων

Τα ποσοστημόρια αποτελούν στην στατιστική επιστήμη ένα εξίσου βασικό χαρακτηριστικό για μια κατανομή. Επομένως η μέθοδος των ποσοστημορίων είναι αρκετά σημαντική καθώς μας βοηθάει ώστε να αποκτήσουμε μια πλήρη εικόνα σχετικά με τα άκρα της κατανομής αντίστοιχα. Μάλιστα για κατανομές οι οποίες έχουν βαριά ουρά είναι ιδιαίτερα σημαντικό το δεξιό άκρο τους.

Έστω ότι συμβολίζουμε το άνω ποσοστημόριο μιας κατανομής ως p_a .

Ορισμός 5.1: Το άνω ποσοστημόριο p_a μιας κατανομής αποτελεί το σημείο αυτό για το οποίο το $a\%$ των τιμών είναι μεγαλύτερες από αυτό ενώ το υπόλοιπο $(1 - a)\%$ είναι μικρότερες ή ίσες από αυτό το σημείο.

Επομένως στη συνέχεια θα κάνουμε χρήση αυτής της μεθόδου ώστε να υπολογίσουμε τα άνω ποσοστημόρια των συναρτήσεων κατανομής τόσο του Moschopoulos όσο και του Barnabani και να συγκρίνουμε εάν είναι κοντά οι τιμές που δίνει η προσέγγιση με τις πραγματικές τιμές.

Με βάση τον ορισμό θα πρέπει να ισχύει:

$$P(Y > p_a) = a$$

Ωστόσο ας δούμε αντίστοιχα και κάποιες ειδικές περιπτώσεις ποσοστημορίων:

- Πρώτο τεταρτημόριο (quartile) Q_1 : αφήνει δεξιά του το 75% των παρατηρήσεων
- Τρίτο τεταρτημόριο (quartile) Q_3 : αφήνει δεξιά του το 25% των παρατηρήσεων
- Διάμεσος Q_2 : αφήνει δεξιά της το 50% των παρατηρήσεων
- Δεκατημόρια D_k
- Εκατοστημόρια P_k

Για τον υπολογισμό των άνω ποσοστημορίων στην συνάρτηση κατανομής του Moschopoulos θα γίνει χρήση της κάτωθι εξίσωσης:

$$P(Y > p_{Y,a}) = 1 - F_{p_{Y,a}}$$

Ενώ αντιστοίχως για τον υπολογισμό των άνω ποσοστημορίων στην συνάρτηση κατανομής του Barnabani (2017) θα γίνει η χρήση της κάτωθι εξίσωσης:

$$P(Y_1 > p_{Y_1, \alpha}) = 1 - F_{gnb, p_{Y_1, \alpha}}$$

Για διάφορες τιμές του α , θα υπολογίσουμε τα άνω ποσοστημόρια $p_{Y_1, \alpha}$ και $p_{Y, \alpha}$ και για τις δύο κατανομές.

Για $\alpha=0.005$

Κώδικας στην R:

```
> library (GoFKernel)
> f<-function(x) pcoga(x, c (3,10,20),c(6,8,10))
> f.inv<-inverse(f, lower=0,upper=8)
> f.inv(0.995)
> f1<-function(x) pcoga_approx(x, c(3,10,20),c(6,8,10))
> f.inv1<-inverse(f1,lower=0,upper=8)
> f.inv1(0.995)
> ((f.inv(0.995)-f.inv1(0.995))/f.inv(0.995))*100
```

Προκύπτει ως λύση της εξίσωσης για κάθε κατανομή αντίστοιχα: $p_{Y, \alpha} = 5.686413$ και $p_{Y_1, \alpha} = 5.685258$.

Αντίστοιχος είναι ο κώδικας και για τις υπόλοιπες περιπτώσεις.

Για $\alpha=0.03$

Προκύπτει ως λύση της εξίσωσης για κάθε κατανομή αντίστοιχα: $p_{Y, \alpha} = 5.095424$ και $p_{Y_1, \alpha} = 5.095136$.

Για $\alpha=0.05$

Προκύπτει ως λύση της εξίσωσης για κάθε κατανομή αντίστοιχα: $p_{Y, \alpha} = 4.904807$ και $p_{Y_1, \alpha} = 4.904687$.

Για $\alpha=0.075$

Προκύπτει ως λύση της εξίσωσης για κάθε κατανομή αντίστοιχα: $p_{Y, \alpha} = 4.742957$ και $p_{Y_1, \alpha} = 4.742944$.

Για $\alpha=0.1$

Προκύπτει ως λύση της εξίσωσης για κάθε κατανομή αντίστοιχα: $p_{Y,a} = 4.620949$ και $p_{Y_1,a} = 4.620997$.

Αντίστοιχα θα υπολογίσουμε και τις ποσότητες: $\frac{p_{Y,a}-p_{Y_1,a}}{p_{Y,a}} \%$, οι οποίες μας δείχνουν το ποσοστό απόκλισης μεταξύ των τιμών των ποσοστημορίων για τις δύο συναρτήσεις κατανομών (ακριβής και προσέγγιση).

Παρουσιάζουμε συγκεντρωτικά τα αποτελέσματα μας στον παρακάτω πίνακα:

Για ευκολία τις λύσεις τις συμβολίζουμε αντί για $p_{Y,a}$ και $p_{Y_1,a}$ ως y και y_1 .

Πίνακας 5.1:

α	$F_Y(y)$	$F_{Y_1}(y_1)$	y	y_1	Ποσοστό απόκλισης
0.005	0.995	0.995	5.686413	5.685258	0.02032%
0.03	0.97	0.97	5.095424	5.095136	0.005652%
0.05	0.95	0.95	4.904807	4.904687	0.002447%
0.075	0.925	0.925	4.742957	4.742944	0.000274%
0.1	0.9	0.9	4.620949	4.620997	-0.001041%

Όπως διαπιστώνουμε και στον πίνακα παραπάνω, οι τιμές των ποσοστημορίων για την συνάρτηση κατανομής του Moschoroulos (y) υπερβαίνουν τις αντίστοιχες τιμές των ποσοστημορίων για την συνάρτηση κατανομής του Barnabani (y_1) για $\alpha = \{0.005, 0.03, 0.05, 0.075\}$. Παρόλα αυτά για $\alpha = 0.1$ παρατηρούμε ότι η τιμή του y_1 βρίσκεται ελάχιστα πιο πάνω από την τιμή του y . Ωστόσο εξίσου σημαντικός είναι και ο υπολογισμός των ποσοστών απόκλισης για τις δύο κατανομές καθώς όσο μικρότερο είναι το αντίστοιχο ποσοστό τόσο καλύτερα προσεγγίζεται η πραγματική κατανομή. Το ποσοστό απόκλισης μπορεί να πάρει είτε θετικές είτε αρνητικές τιμές ενώ όσο πιο κοντά βρίσκεται στο μηδέν τόσο το καλύτερο. Στην περίπτωση μας, οι τιμές των ποσοστών είναι όντως πολύ μικρές και αγγίζουν το μηδέν γεγονός το οποίο επιβεβαιώνει ότι η κατανομή του Barnabani πρόκειται για μια αξιόλογη προσέγγιση.

Γενικό συμπέρασμα: Ύστερα από την διεξαγωγή των παραπάνω κριτηρίων μέσω των οποίων αποκτήσαμε μια εικόνα για διάφορα χαρακτηριστικά των πυκνοτήτων και των συναρτήσεων κατανομής τόσο του Moschoroulos όσο και του Barnabani για την περίπτωση ενός απλού παραδείγματος συνέλιξης 3 Γάμμα κατανομών, διαπιστώσαμε με κάθε μέθοδο ότι όντως η μία κατανομή προσεγγίζει ικανοποιητικά την άλλη. Αυτό μας δίνει τη δυνατότητα να χρησιμοποιούμε την κατανομή που πρότεινε ο Barnabani η οποία είναι ιδιαίτερα γρήγορη στον υπολογισμό της για διάφορες τιμές των παραμέτρων σχήματος και κλίμακας και για την συνέλιξη περισσότερων από 2 Γάμμα τυχαίων μεταβλητών, έχοντας ταυτόχρονα ένα πολύ μικρό ποσοστό σφάλματος και απόκλισης από την ακριβή κατανομή αντίστοιχα.

5.3 Εφαρμογές των συνελίξεων Γάμμα κατανομών στην Θεωρία Κινδύνου

Το συλλογικό πρότυπο της θεωρίας κινδύνου αποτελεί ένα από τα σημαντικότερα μοντέλα στην αναλογιστική επιστήμη καθώς μελετάει την κατανομή των συνολικών απαιτήσεων ενός συγκεκριμένου χαρτοφυλακίου που φτάνουν σε μια ασφαλιστική εταιρεία, μέσα σε ένα καθορισμένο χρονικό διάστημα.

Κύριο αντικείμενο μελέτης για την ασφαλιστική εταιρεία αποτελούν οι συνολικές απαιτήσεις S , τις οποίες θα πρέπει να πληρώσει για ένα χαρτοφυλάκιο. Ωστόσο για το συνολικό ποσό των αποζημιώσεων θα πρέπει να λάβει υπόψιν τους εξής παράγοντες:

- τα μεγέθη των απαιτήσεων και
- το πλήθος των απαιτήσεων που φτάνουν στην εταιρεία

Οι τιμές των παραπάνω ποσοτήτων εφόσον δεν είναι γνωστές παριστάνονται από τυχαίες μεταβλητές. Συγκεκριμένα συμβολίζουμε ως N την διακριτή τυχαία μεταβλητή η οποία παίρνει τιμές στο σύνολο $\{0,1,2 \dots\}$ και παριστάνει το συνολικό πλήθος των απαιτήσεων για το χρονικό διάστημα που μελετάμε. Ενώ τα μεγέθη των αποζημιώσεων συμβολίζονται ως X_1, X_2, \dots, X_N και ακολουθούν όλα την ίδια κατανομή F .

Ισχύει για το μέγεθος S των συνολικών αποζημιώσεων (Πολίτης, 2012):

$$S = \begin{cases} X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i, & \text{αν } N = 1, 2, \dots \\ 0, & \text{αν } N = 0 \end{cases}$$

Η S είναι μια *σύνθετη* τυχαία μεταβλητή όπως προκύπτει και από την παραπάνω σχέση. Επιπλέον θεωρούμε ως F την κοινή συνάρτηση κατανομής που ακολουθούν οι ατομικές απαιτήσεις X_i ενώ παράλληλα: $p_n = P(N = n)$, $n = 0, 1, 2, \dots$ είναι η συνάρτηση πιθανότητας για το πλήθος N των αποζημιώσεων στο αντίστοιχο χρονικό διάστημα.

Έστω:

$$G(x) = P(S \leq x), \quad x \geq 0$$

η συνάρτηση κατανομής της τυχαίας μεταβλητής S , η οποία παριστάνει το μέγεθος των συνολικών απαιτήσεων στο συλλογικό πρότυπο. Η $G(x)$ μπορεί να παρασταθεί ως ένα άπειρο άθροισμα δυναμοσυνελίξεων της F , η οποία δυναμοσυνέλιξη τάξης n θα είναι πολλαπλασιασμένη με την πιθανότητα $p_n = P(N = n)$.

Πρόταση 5.3.1 (Πολίτης, 2012): Ισχύει η εξής σχέση για την συνάρτηση κατανομής $G(x)$ των συνολικών αποζημιώσεων S

$$G(x) = \sum_{n=0}^{\infty} P(N = n) \cdot F^{*n}(x) = \sum_{n=0}^{\infty} p_n \cdot F^{*n}(x), \quad x \geq 0$$

Για $n = 0$, η συνέλιξη $F^{*0}(x)$ ορίζεται ως $F^{*0}(x) = 0$ για $x < 0$ και ως $F^{*0}(x) = 1$ για $x \geq 0$.

Παράλληλα για την δυναμοσυνέλιξη τάξης n ($F^{*n}(x)$) έχουμε ότι:

$$F^{*n}(x) = (F^{*n-1} \cdot F)(x) = \int_{-\infty}^{\infty} F^{*n-1}(x-u) dF(u), \quad n = 1, 2, 3..$$

5.3.1 Προσεγγίσεις για την κατανομή των συνολικών αποζημιώσεων

Κάθε ασφαλιστική εταιρεία δίνει ιδιαίτερη σημασία στο σύνολο των κινδύνων στους οποίους είναι εκτεθειμένη παρά σε κάθε ατομικό κίνδυνο ξεχωριστά. Συνεπώς η κατανομή των συνολικών απαιτήσεων έχει ζωτικό ρόλο στο τομέα του αναλογισμού.

Σε αρκετές περιπτώσεις είναι εφικτός ο ακριβής υπολογισμός της κατανομής της τ.μ. S , η οποία περιγράφει τις συνολικές αποζημιώσεις που πληρώνει η ασφαλιστική εταιρεία για ένα συγκεκριμένο χαρτοφυλάκιο. Ωστόσο, υπάρχουν πολλές περιπτώσεις για τις οποίες δεν μπορεί να υπολογιστεί αναλυτικός τύπος για την κατανομή $G(x)$, συνεπώς είναι σημαντική η εύρεση μιας ικανοποιητικής προσέγγισης. Υπάρχουν παρ' όλα αυτά δύο από τις βασικότερες προσεγγίσεις οι οποίες βασίζονται στην κανονική κατανομή αλλά και στην μετατοπισμένη κατανομή Γάμμα. Αυτές οι μέθοδοι προσέγγισης της κατανομής του αθροίσματος των απαιτήσεων βασίζονται κυρίως στην ισότητα μεταξύ των ροπών των αντίστοιχων κατανομών (moment matching approximations) ενώ κάνουν χρήση μόνο των τριών πρώτων ροπών. Παράλληλα μια πιο σύνθετη μέθοδος προσέγγισης είναι μέσω της μείξης Erlang κατανομών. Αλλά, επειδή τα μεγέθη των ατομικών ζημιών μπορεί να ακολουθούν είτε κατανομές με βαριά είτε κατανομές με ελαφριά ουρά, στην πρώτη περίπτωση (με βαριά ουρά) οι ροπές μεγαλύτερης τάξης μπορεί να μην υπάρχουν και συνεπώς η προσέγγιση με την χρήση των μείξεων Erlang κατανομών να μην είναι αρκετά αποδοτική.

Ας δούμε ωστόσο περιληπτικά τι ισχύει στην περίπτωση της κανονικής προσέγγισης:

(Πολίτης, 2012): Έστω η τυχαία μεταβλητή W η οποία ακολουθεί κανονική κατανομή με την ίδια μέση τιμή και διακύμανση με την τ.μ. S που περιγράφει το σύνολο των απαιτήσεων σε ένα ασφαλιστικό χαρτοφυλάκιο δηλαδή $W \sim N(E(S), Var(S))$, τότε θα ισχύει ο προσεγγιστικός τύπος:

$$P(S \leq x) = P(W \leq x) \quad (5.3.1)$$

Για την τ.μ. W , η αντίστοιχη τυποποιημένη τ.μ. Z η οποία ακολουθεί την τυποποιημένη κανονική κατανομή ($Z \sim N(0,1)$) είναι η:

$$Z = \frac{W - E(W)}{\sqrt{Var(W)}}$$

Εφόσον από υπόθεση ισχύει ότι $E(W) = E(S)$ και $Var(W) = Var(S)$ τότε ο προσεγγιστικός τύπος στην σχέση (5.3.1) θα μπορεί να γραφθεί αντίστοιχα ως:

$$P(S \leq x) \cong P\left(\frac{W - E(S)}{\sqrt{Var(S)}} \leq \frac{x - E(S)}{\sqrt{Var(S)}}\right)$$

$$P(S \leq x) \cong \Phi\left(\frac{x - E(S)}{\sqrt{Var(S)}}\right), \quad x \geq 0$$

Ωστόσο η προσέγγιση μέσω της κανονικής κατανομής δεν είναι εποικοδομητική όταν έχουμε μικρό πλήθος ατομικών ζημιών. Αν το πλήθος των X_i ήταν σταθερό (ίσο με n) τότε το Κεντρικό Οριακό Θεώρημα θα μας εξασφάλιζε ότι για μεγάλες τιμές του n , η κατανομή του άθροισματος θα ήταν προσεγγιστικά κανονική γεγονός το οποίο δεν μπορεί να ισχύει πάντοτε καθώς το πλήθος N των ατομικών κινδύνων είναι μια τυχαία μεταβλητή.

Επιπρόσθετα ένα βασικό μειονέκτημα της κανονικής προσέγγισης είναι ότι χρησιμοποιεί μόνο τις δύο πρώτες ροπές της κατανομής των συνολικών αποζημιώσεων χωρίς να δίνει έμφαση στην ασυμμετρία. Όταν για παράδειγμα η ακριβής κατανομή $G(x)$ είναι σύνθετη Poisson (γεγονός το οποίο είναι αρκετά συχνό στο συλλογικό πρότυπο) η οποία εμφανίζει πάντοτε θετική ασυμμετρία τότε δεν θα ήταν σωστή η προσέγγιση από μια συμμετρική κατανομή όπως η κανονική. Για τον σκοπό αυτό και για την βελτίωση των αποτελεσμάτων μελετήθηκε η προσέγγιση της $G(x)$ από μια μετατοπισμένη κατανομή Γάμμα.

(Πολίτης, 2012): Αν μια μεταβλητή Y ακολουθεί την κατανομή Γάμμα με παραμέτρους a, b τότε η τ.μ. $W = Y + k$ θα ακολουθεί την μετατοπισμένη κατανομή Γάμμα με παραμέτρους a, b και k , όπου k μια πραγματική σταθερά. Θα πρέπει να γίνεται επιλογή των κατάλληλων παραμέτρων έτσι ώστε οι δύο κατανομές (ακριβής και προσέγγιση) να έχουν κοινή μέση τιμή, διακύμανση και συντελεστή ασυμμετρίας. Αν η $Y \sim \text{Gamma}(a, b)$ γνωρίζουμε από τις πιθανότητες ότι έχει μέση τιμή ίση με a/b , διακύμανση ίση με a/b^2 ενώ ο συντελεστής ασυμμετρίας ισούται με $2/\sqrt{a}$. Επομένως η μεταβλητή $W = Y + k$ θα έχει την ίδια διακύμανση και συνεπώς και τον ίδιο συντελεστή ασυμμετρίας. Άρα ο υπολογισμός των παραμέτρων της μετατοπισμένης Γάμμα για την προσέγγιση της $G(x)$ προκύπτει από την λύση των κάτωθι εξισώσεων:

$$E(S) = \frac{a}{b} + k$$

$$Var(S) = a/b^2$$

$$\gamma_{1,s} = 2/\sqrt{a}$$

Για θετικές τιμές της παραμέτρου k , η προσέγγιση είναι αρκετά ικανοποιητική και η παραπάνω μέθοδος αποτελεί μια αξιόπιστη επιλογή. Αν όμως η παράμετρος k πάρει αρνητικές τιμές από την λύση του παραπάνω συστήματος τότε η μετατοπισμένη Γάμμα μας δίνει θετική πιθανότητα να έχουμε αρνητικό άθροισμα απαιτήσεων γεγονός το οποίο δεν μπορεί να ισχύει.

5.3.2 Προσεγγίσεις για την κατανομή των συνολικών αποζημιώσεων με την χρήση συνελίξεων Γάμμα κατανομών

Στην υποενότητα αυτή θα παρουσιάσουμε μια ιδιαίτερα ικανοποιητική μέθοδο προσέγγισης της κατανομής $G(x)$ (Furman, 2018) η οποία θα μπορεί να εφαρμοστεί σε οποιαδήποτε περίπτωση κατανομών των συνολικών απαιτήσεων (είτε με ελαφριά είτε με βαριά ουρά) και για μεγάλο ή μικρό πλήθος αποζημιώσεων αντίστοιχα.

Ωστόσο για την μέθοδο προσέγγισης που θα παρουσιάσουμε θα γίνει χρήση του μοντέλου με το σταθερό πλήθος απαιτήσεων (Individual Risk Model ή IRM) όπου η τ.μ. που θα περιγράψει τις συνολικές απαιτήσεις θα είναι της μορφής $S_n = X_1 + X_2 + \dots + X_n$ όπου n σταθερός αριθμός και οι ατομικές ζημιές $\{X_1, X_2, \dots, X_n\}$ θα είναι ανεξάρτητες τ.μ. αλλά χωρίς να ακολουθούν απαραίτητα την ίδια κατανομή. Θα δούμε ότι η μέθοδος λειτουργεί με αντίστοιχο τρόπο και στην περίπτωση όπου έχουμε το μοντέλο με το μεταβλητό πλήθος απαιτήσεων (Collective Risk Model ή CRM) της μορφής $S_N = X_1 + X_2 + \dots + X_N$ με το πλήθος των απαιτήσεων N να είναι μια τυχαία μεταβλητή.

Έστω οι θετικές και ανεξάρτητες τ.μ. $X_1 + X_2 + \dots + X_n$ που περιγράφουν τις ατομικές ζημιές στο συλλογικό πρότυπο με αντίστοιχες αθροιστικές συναρτήσεις κατανομών $F_1(x), F_2(x), \dots, F_n(x)$. Θεωρούμε επίσης ως μετασχηματισμούς Laplace των $X_i, i = 1, 2, \dots, n$ τις ποσότητες $\varphi_i(z) := E[\exp(-z \cdot X_i)]$.

Όπως έχει ήδη αναφερθεί ο βασικός σκοπός της μεθόδου είναι η προσέγγιση της συνάρτησης κατανομής της τ.μ. των συνολικών απαιτήσεων $S = X_1 + X_2 + \dots + X_n$. Αν συμβολίσουμε τον μετασχηματισμό Laplace της τ.μ. S ως $\varphi(z) := E[\exp(-z \cdot S)]$ ή αλλιώς $\mathcal{L}(z)$ με βάση την ανεξαρτησία των X_i θα ισχύει:

$$\varphi(z) := \prod_{i=1}^n \varphi_i(z) \quad (5.3.2)$$

Και επίσης
$$F(x) := \mathcal{L}^{-1} \left\{ \frac{\varphi(z)}{z} \right\} (x), x \geq 0 \quad (5.3.3)$$

Συνεπώς με τον κατάλληλο συνδυασμό των σχέσεων 5.3.2 και 5.3.3 θα μπορούσαμε να έχουμε μια προσέγγιση για την αθροιστική συνάρτηση κατανομής $F(x)$ αρκεί όπως θα εξετάσουμε και στην συνέχεια να καταφέρουμε με ικανοποιητικές μεθόδους να προσεγγίσουμε τους μετασχηματισμούς Laplace $\varphi_i(z)$ των τ.μ. X_i και να αντιστρέψουμε στη συνέχεια τον μετασχηματισμό Laplace της τ.μ. S (Furman, 2018).

Στην συνέχεια λοιπόν θα προσεγγίσουμε τους μετασχηματισμούς Laplace $\varphi_i(z)$ με την χρήση των μετασχηματισμών Laplace κάποιων συγκεκριμένων *συνελίξεων Γάμμα κατανομών* παρουσιάζοντας με αυτόν τον τρόπο την ιδιαίτερη σημασία στη θεωρία κινδύνων αυτής της οικογένειας κατανομών που αποτελούν το αντικείμενο μελέτης της παρούσας εργασίας. Ωστόσο για την επίτευξη αυτή θα χρησιμοποιήσουμε τις προσεγγίσεις Pade ώστε να καθορίσουμε τις αντίστοιχες παραμέτρους σχήματος και κλίμακας των κατανομών που θα επιλέξουμε (Furman, 2018).

Θα χρησιμοποιήσουμε τις *συνελίξεις* m ανεξάρτητων και τυχαίων μεταβλητών που θα ακολουθούν την κατανομή Γάμμα, $\Gamma_i \sim \text{Gamma}(a_i, b_i)$, με $i = 1, 2, \dots, m$ και θα αναζητήσουμε τις παραμέτρους αυτές (a_i, b_i) ώστε η αθροιστική συνάρτηση κατανομής της τ.μ. που χρησιμοποιούμε ως προσέγγιση ($\tilde{X}_{i,(m)} := \sum_{i=1}^m \Gamma_i$) να βρίσκεται αρκετά κοντά στην αθροιστική συνάρτηση κατανομής της τ.μ. X_i , $i = 1, 2, \dots, n$.

Για ευκολία στην συνέχεια θα γράφουμε X αντί για X_i και $\varphi(z)$ αντί για $\varphi_i(z)$.

Με όρους που αφορούν τους μετασχηματισμούς Laplace, ο σκοπός μας είναι να βρούμε για ποιες παραμέτρους (a_i, b_i) ο μετασχηματισμός Laplace της προσεγγιστικής τ.μ.

$$\tilde{\varphi}(z) := E \left[\exp \left(-z \cdot \sum_{i=1}^m \Gamma_i \right) \right] = \prod_{i=1}^m \left(1 + \frac{z}{b_i} \right)^{-a_i}$$

θα πλησιάζει την ποσότητα $\varphi(z) = E[\exp(-z \cdot X)]$. Η μέθοδος των ροπών είναι αυτή η οποία θα βοηθήσει στην κατάλληλη επιλογή παραμέτρων σχήματος και κλίμακας για τις κατανομές Γάμμα που περιέχονται στην συνέλιξη (moment matching approximations). Οι ροπές της κάθε τ.μ. X μπορούν να υπολογισθούν ως $E[X^k] = (-1)^k \varphi^{(k)}(0)$ και εάν θέσουμε τις πρώτες m ροπές της τ.μ. \tilde{X} να είναι ίσες με τις αντίστοιχες m ροπές της τ.μ. X τότε ουσιαστικά θα είναι ισοδύναμα σαν να θέτουμε ίσες τις παραγώγους τάξης $k = 1, 2, \dots, m$ των μετασχηματισμών Laplace $\varphi(z)$ και $\tilde{\varphi}(z)$ στο σημείο $z = 0$.

Παρόλα αυτά αν έχουμε στο συλλογικό πρότυπο ατομικές ζημιές που να ακολουθούν κατανομές με βαριές ουρές για τις οποίες μπορεί να μην υπάρχουν κάποιες από τις ροπές (και συνεπώς και οι παράγωγοι τάξης k των μετασχηματισμών Laplace στο σημείο μηδέν : $\varphi^{(k)}(0)$) τότε θα ήταν εύλογο να θέσουμε ίσες τις παραγώγους τάξης $k = 1, 2, \dots, m$ των μετασχηματισμών Laplace $\varphi(z)$ και $\tilde{\varphi}(z)$ όχι στο σημείο $z = 0$, αλλά σε κάποιο σημείο $z = z^* > 0$. Εάν οι τ.μ. X των ατομικών ζημιών ακολουθούν κατανομές με ελαφριές ουρές θα μπορούσαμε σε αυτή την περίπτωση να επιλέξουμε το σημείο $z^* = 0$ αλλά γενικά το σημείο z^* θα πρέπει να είναι αυστηρώς θετικό.

Ένα εξίσου σημαντικό χαρακτηριστικό της μεθόδου που εφαρμόζουμε είναι το γεγονός ότι οι μετασχηματισμοί Laplace $\tilde{\varphi}(z)$ για τις τ.μ. που χρησιμοποιούμε ως προσεγγιστικές συγκλίνουν στους μετασχηματισμούς Laplace $\varphi(z)$ των τ.μ. X με ομοιόμορφο και ταυτόχρονα εκθετικά γρήγορο τρόπο με αποτέλεσμα η κατανομή της τ.μ. \tilde{X} να συγκλίνει πολύ ικανοποιητικά στην κατανομή της τ.μ. X (Furman, 2018).

Επιπλέον εφόσον έχουμε την συνέλιξη m Γάμμα κατανομών με 2 παραμέτρους για την κάθε κατανομή Γάμμα (a, b) , συνολικά θα απαιτείται η επιλογή $2 \cdot m$ παραμέτρων. Άρα θα θέτουμε ίσες τις πρώτες $2 \cdot m$ ροπές των τ.μ. \tilde{X} και X . Η μέθοδος που χρησιμοποιούμε είναι η ενδεδειγμένη για την περίπτωση αυτή.

Προσέγγιση με την χρήση συνελίξεων 2 Γάμμα κατανομών

Θα παρουσιάσουμε αρχικά πως λειτουργεί η παραπάνω τεχνική για την πολύ απλή περίπτωση όπου $m = 2$ και $z^* = 1 > 0$. Συνεπώς θα πρέπει να βρούμε συνολικά $2 \cdot m = 4$ θετικούς αριθμούς (a_1, b_1, a_2, b_2) τέτοιους ώστε οι παράγωγοι του μετασχηματισμού Laplace $\varphi(z)$ στο σημείο $z = 1$ να συμπίπτουν στο ίδιο σημείο με τις παραγώγους του μετασχηματισμού Laplace της προσεγγιστικής τ.μ. δηλαδή του:

$$\tilde{\varphi}_{(2)}(z) = \left(1 + \frac{z}{b_1}\right)^{-a_1} \cdot \left(1 + \frac{z}{b_2}\right)^{-a_2}$$

Για να μπορέσουμε να υπολογίσουμε τις παραμέτρους (a_1, b_1, a_2, b_2) των κατανομών Γάμμα της συνέλιξης, θα πρέπει να έχουμε ένα σύστημα με τουλάχιστον 4 εξισώσεις αντίστοιχα. Επομένως θα πρέπει να επιλυθούν οι παρακάτω εξισώσεις για $k = 1, 2, 3, 4$:

$$\left. \frac{d^k}{dz^k} \left(1 + \frac{z}{b_1}\right)^{-a_1} \cdot \left(1 + \frac{z}{b_2}\right)^{-a_2} \right|_{z=1} = \varphi^{(k)}(1), \quad k \in \{1, 2, 3, 4\} \quad (5.3.4)$$

Όπου $\varphi^{(k)}(1)$ είναι η παράγωγος κ τάξης της τ.μ. X (την οποία προσεγγίζουμε) στο σημείο $z=1$. Αν μπορούσαμε να υπολογίσουμε τις παραγώγους στην αριστερή πλευρά της σχέσης 5.3.4 τότε θα είχαμε ως αποτέλεσμα ένα αρκετά περίπλοκο μη γραμμικό σύστημα των $\{a_1, b_1, a_2, b_2\}$ και δεν θα ήταν εφικτή η ανάλυση και ο υπολογισμός των εξισώσεων που θα προέκυπταν (Furman, 2018).

Για τον σκοπό αυτό και για να αποφύγουμε την πολυπλοκότητα και την δυσκολία υπολογισμού που υπάρχει στις εξισώσεις της σχέσης 5.3.4 θα ήταν πιο εποικοδομητικό να παραγωγίσουμε τους λογαρίθμους των μετασχηματισμών Laplace. Δηλαδή αντί να θέσουμε ίσες τις παραγώγους κ-τάξης $(\varphi^{(k)}$ και $\tilde{\varphi}_{(2)}^{(k)})$ στο σημείο $z=1$ για $k = 1, 2, 3, 4$, θα ήταν πιο εύκολο να θέσουμε ίσες τις παραγώγους των λογαρίθμων $\ln[\varphi(z)]$ και $\ln[\tilde{\varphi}_{(2)}(z)]$ στο σημείο $z=1$ για $k = 1, 2, 3, 4$. Σε αυτή την περίπτωση μπορεί το τελικό αποτέλεσμα να ήταν ίδιο αλλά σαφέστατα θα ήταν πολύ πιο γρήγορος ο υπολογισμός για τις τέσσερις παραμέτρους. Το σύστημα εξισώσεων που θα χρησιμοποιούσαμε θα ήταν το εξής:

$$\left. \frac{d^k}{dz^k} \ln \left[\left(1 + \frac{z}{b_1}\right)^{-a_1} \cdot \left(1 + \frac{z}{b_2}\right)^{-a_2} \right] \right|_{z=1} = \left. \frac{d^k}{dz^k} \ln[\varphi(z)] \right|_{z=1}, \quad k \in \{1, 2, 3, 4\} \quad (5.3.5)$$

Στην συνέχεια θα ορίσουμε την παρακάτω ποσότητα:

$$s_k := -\left(\frac{1}{k!}\right) \cdot \left. \frac{d^{k+1}}{dz^{k+1}} \ln[\varphi(z)] \right|_{z=1}$$

Για τους σκοπούς των παρακάτω υπολογισμών θα θεωρήσουμε τις ποσότητες s_k ως γνωστές αν και μπορεί ναδειχθεί ότι ο υπολογισμός τους είναι ιδιαίτερα απλός (Furman, 2018).

Υπολογίζοντας τις παραγώγους στην αριστερή πλευρά της σχέσης 5.3.5 θα προκύψει το εξής σύστημα εξισώσεων για $k = 1, 2, 3, 4$ (βλ. Furman, 2018):

$$\frac{a_1}{(1+b_1)^k} + \frac{a_2}{(1+b_2)^k} + \frac{a_2}{(1+b_2)^k} + \frac{a_2}{(1+b_2)^k} = (-1)^k \cdot s_{k-1}, \text{ για } k = 1, 2, 3, 4 \quad (5.3.6)$$

Για την εύρεση των παραμέτρων (a_1, b_1, a_2, b_2) που μας ενδιαφέρει και για τον υπολογισμό των εξισώσεων της σχέσης 5.3.6 θα γίνει χρήση των προσεγγίσεων Pade.

Εισάγουμε μια νέα συνάρτηση ως: $\psi(z) := -\varphi'(z)/\varphi(z)$ όπου $\varphi(z)$ και $\varphi'(z)$ είναι ο μετασχηματισμός Laplace για την τ.μ. X και η πρώτη παράγωγος του μετασχηματισμού αυτού αντίστοιχα. Συνεπώς το σύστημα εξισώσεων της σχέσης 5.3.5 μπορεί να εκφρασθεί και ως:

$$\frac{d^k}{dz^k} \left[\frac{a_1}{z+b_1} + \frac{a_2}{z+b_2} \right] \Big|_{z=1} = \psi^{(k)}(1), k \in \{1, 2, 3, 4\} \quad (5.3.7)$$

Για κάποιες σταθερές A, B, C, D μπορούμε να έχουμε την εξής ισότητα :

$$\frac{a_1}{z+b_1} + \frac{a_2}{z+b_2} = \frac{A + B \cdot (z-1)}{1 + C \cdot (z-1) + D \cdot (z-1)^2} \quad (5.3.8)$$

Συνεπώς μπορούμε ισοδύναμα για την σχέση (5.3.7) να πούμε ότι θα πρέπει οι τέσσερις πρώτοι όροι του αναπτύγματος Maclaurin για την συνάρτηση:

$$\frac{P(w)}{Q(w)} := \frac{A + B \cdot w}{1 + C \cdot w + D \cdot w^2}$$

να είναι ίσοι με τους αντίστοιχους τέσσερις πρώτους όρους του αναπτύγματος Maclaurin για την:

$$\psi(1+w) = s_0 + s_1 \cdot w + s_2 \cdot w^2 + \dots$$

με s_k να είναι η ποσότητα που ορίσαμε προηγουμένως.

Έχουμε δηλαδή ορίσει την νέα μεταβλητή $w = z - 1$ και για $w \rightarrow 0$ ισχύει:

$$\frac{A + B \cdot w}{1 + C \cdot w + D \cdot w^2} = s_0 + s_1 \cdot w + s_2 \cdot w^2 + s_3 \cdot w^3 + O(w^4) \quad (5.3.9)$$

όπου $O(\cdot)$ (ή Big-O notation) αποτελεί το ασυμπτωτικό άνω όριο του ρυθμού ανάπτυξης μιας συνάρτησης f . Στην περίπτωση αυτή θα ισχύει ότι:

$$f(w) = O(w^4) \text{ καθώς } w \rightarrow 0$$

εάν
$$\limsup_{w \rightarrow 0} \frac{|f(w)|}{w^4} < \infty$$

Ορισμός 5.3.1 : Γενικά ονομάζουμε την λογική συνάρτηση $P(w)/Q(w)$ ως μια $[p/q]$ προσέγγιση Pade μιας συνάρτησης $f(w)$, η οποία προσέγγιση θα έχει ίδιους τους πρώτους $p + q + 1$ όρους με την συνάρτηση $f(w)$, στο ανάπτυγμα του Maclaurin. Να αναφέρουμε ότι τα $P(w), Q(w)$ είναι πολυώνυμα με βαθμό: $\deg(P) = p$ και $\deg(Q) = q$ αντίστοιχα. Συνεπώς μέσω της σχέσης (5.3.9) προκύπτει ότι η συνάρτηση $\frac{P(w)}{Q(w)}$ είναι μια $[1/2]$ προσέγγιση Pade της συνάρτησης $\psi(1+w)$.

Σκοπός μας λοιπόν είναι η εύρεση των σταθερών A, B, C και D . Στην ισότητα 5.3.9 πολλαπλασιάζουμε και τις δύο πλευρές με το πολυώνυμο $Q(w) = 1 + C \cdot w + D \cdot w^2$ και έχουμε αντιστοίχως, για $w \rightarrow 0$:

$$A + B \cdot w = (1 + C \cdot w + D \cdot w^2) \cdot (s_0 + s_1 \cdot w + s_2 \cdot w^2 + s_3 \cdot w^3 + O(w^4)) \quad (5.3.10)$$

Κάνοντας τις πράξεις, στο δεύτερο σκέλος της ισότητας θα έχουμε:

$$\begin{aligned} & s_0 + s_1 \cdot w + s_2 \cdot w^2 + s_3 \cdot w^3 + O(w^4) + C \cdot w \cdot s_0 + C \cdot s_1 \cdot w^2 + C \cdot s_2 \cdot w^3 + C \cdot s_3 \cdot w^4 \\ & + C \cdot O(w^4) + D \cdot s_0 \cdot w^2 + D \cdot s_1 \cdot w^3 + D \cdot s_2 \cdot w^4 + D \cdot s_3 \cdot w^5 + D \\ & \cdot O(w^4) = \\ & = s_0 + (s_1 + C \cdot s_0) \cdot w + (s_2 + C \cdot s_1 + D \cdot s_0) \cdot w^2 + (s_3 + C \cdot s_2 + D \cdot s_1) \cdot w^3 \\ & + (C \cdot s_3 + D \cdot s_2) w^4 + D \cdot s_3 \cdot w^5 + (1 + C + D) \cdot O(w^4) \end{aligned}$$

Επομένως θα πρέπει να ισχύει, για $w \rightarrow 0$:

$$\begin{aligned} A + B \cdot w = & s_0 + (s_1 + C \cdot s_0) \cdot w + (s_2 + C \cdot s_1 + D \cdot s_0) \cdot w^2 + (s_3 + C \cdot s_2 + D \cdot s_1) \\ & \cdot w^3 + (C \cdot s_3 + D \cdot s_2) w^4 + D \cdot s_3 \cdot w^5 + (1 + C + D) \cdot O(w^4) \end{aligned}$$

Άρα για τους συντελεστές μπροστά από τις δυνάμεις του w , από ισότητα πολυωνύμων προκύπτει το παρακάτω σύστημα 4 γραμμικών εξισώσεων:

$$\left\{ \begin{array}{l} A = s_0 \\ B = s_1 + C \cdot s_0 \\ 0 = s_2 + C \cdot s_1 + D \cdot s_0 \\ 0 = s_3 + C \cdot s_2 + D \cdot s_1 \end{array} \right. \quad (5.3.11)$$

Το παραπάνω σύστημα εξισώσεων μπορεί να επιλυθεί αρκετά εύκολα για γνωστές ποσότητες s_k , ξεκινώντας αρχικά από την εύρεση των C, D μέσω της τρίτης και τέταρτης εξίσωσης και στην συνέχεια με αντικατάσταση στις δύο πρώτες εξισώσεις να γίνει εύρεση των A, B .

Εάν έχουμε βρει τις σταθερές A, B, C, D του παραπάνω συστήματος τότε θα μπορούμε να βρούμε αντίστοιχα και τις παραμέτρους κλίμακας των κατανομών Γάμμα της συνέλιξης μέσω της σχέσης 5.3.8 όπου για $w = z - 1$, θα έχουμε:

$$(1 + b_1 + w) \cdot (1 + b_2 + w) = 1 + C \cdot w + D \cdot w^2$$

και στην συνέχεια οι παράμετροι σχήματος μπορούν να υπολογισθούν μέσω των σχέσεων

$$a_1 = P'(-1 - b_1) \quad \text{και} \quad a_2 = P'(-1 - b_2)$$

Όπου $P'(w)$ είναι η πρώτη παράγωγος του πολυωνύμου $P(w)$.

Μέσω της παραπάνω διαδικασίας είναι εφικτή η εύρεση των επιθυμητών παραμέτρων των κατανομών Γάμμα στη συνέλιξη που χρησιμοποιούμε ώστε να προσεγγίσουμε την κατανομή

των ατομικών αποζημιώσεων στο συλλογικό πρότυπο. Συνεπώς θα μπορούμε πλέον να υπολογίσουμε και τον μετασχηματισμό Laplace της προσεγγιστικής τ.μ. \tilde{X} .

Προσέγγιση με την χρήση συνελίξεων m Γάμμα κατανομών

Θα εξετάσουμε σε αυτή την περίπτωση πως μπορεί να γίνει η προσέγγιση από μια τ.μ. η οποία θα ακολουθεί μια συνέλιξη m Γάμμα κατανομών με τον αριθμό m όπως επίσης και το σημείο z^* στο οποίο υπολογίζουμε τις παραγώγους των μετασχηματισμών Laplace των τ.μ. X και \tilde{X} , να είναι αυθαίρετοι.

Η διαδικασία είναι ακριβώς ίδια με την περίπτωση της συνελίξης 2 Γάμμα κατανομών μόνο που αυτή τη φορά θα πρέπει να βρεθεί μια συνάρτηση $\frac{P(w)}{Q(w)}$ που να είναι $[(m-1)/m]$ προσέγγιση Pade της συνάρτησης $\psi(z^* + w)$. Προηγουμένως είχαμε θεωρήσει ότι $z^* = 1$.

Επομένως αντί για τις σχέσεις 5.3.8 και 5.3.9 που είχαμε στην περίπτωση όπου $m = 2$ τώρα θα ισχύει η σχέση:

$$\begin{aligned} \sum_{i=1}^m \frac{a_i}{z^* + b_i + w} &= \frac{a_0 + a_1 \cdot w + \dots + a_{m-1} \cdot w^{m-1}}{1 + b_1 \cdot w + b_2 \cdot w^2 + \dots + b_m \cdot w^m} \\ &= s_0 + s_1 \cdot w + s_2 \cdot w^2 + \dots + s_{2m-1} \cdot w^{2m-1} + O(w^{2m}) \end{aligned} \quad (5.3.12)$$

αντίστοιχα έχουμε την ποσότητα:

$$s_k := -\left(\frac{1}{k!}\right) \cdot \psi^{(k)}(z^*)$$

με $\psi^{(k)}(z^*)$ να είναι η παράγωγος k -τάξης της συνάρτησης ψ , όπως την ορίσαμε προηγουμένως για τους σκοπούς των προσεγγίσεων Pade, στο σημείο z^* .

Μέσω της σχέσης 5.3.12 θα μπορούσαμε να δημιουργήσουμε ένα σύστημα εξισώσεων όπως το 5.3.11 ώστε να υπολογίσουμε τις παραμέτρους κλίμακας και σχήματος $\{a_i, b_i\}$ για $i = 1, \dots, m$. Ωστόσο στην γενική περίπτωση των συνελίξεων m Γάμμα κατανομών, οι ποσότητες s_k δεν έχουν κλειστή μορφή αλλά είναι εφικτός ο υπολογισμός τους αριθμητικά.

Ας δούμε αναλυτικά πως βρίσκουμε τις ποσότητες αυτές, οι οποίες αποτελούν το βασικό κομμάτι της μεθόδου που περιγράψαμε.

Αρχικά υπολογίζουμε τις παραγώγους k -τάξης των μετασχηματισμών Laplace της τ.μ. X ως:

$$g_k := \varphi^{(k)}(z^*) = (-1)^k \int_0^\infty x^k \cdot e^{-z^* \cdot x} dF(x), \quad k \geq 0 \quad (5.3.13)$$

Για τους υπολογισμούς αυτούς έχει δειχθεί ότι είναι κατάλληλη η εφαρμογή του διπλού εκθετικού τετραγωνισμού των Takahasi και Mori (1974) (double-exponential quadrature). Πρόκειται για φόρμουλες οι οποίες εφαρμόζουν τον λεγόμενο τραπεζοειδή κανόνα σε άπειρα ολοκληρώματα, τα οποία προκύπτουν από δεδομένα ολοκληρώματα $\int_a^b f(x) dx$ με

κατάλληλους μετασχηματισμούς μεταβλητών της μορφής $x = \varphi(u)$. Οι φόρμουλες αυτές έχουν χαρακτηριστεί έτσι καθώς οι συναρτήσεις που ολοκληρώνονται με $u \rightarrow \infty$ έχουν διπλή εκθετική ασυμπτωτική συμπεριφορά.

Η τραπεζοειδής μέθοδος αριθμητικής ολοκλήρωσης χρησιμοποιείται για την προσέγγιση ολοκληρωμάτων της μορφής $\int_a^b f(x)dx$ και στην πιο γενική περίπτωση όπου έχουμε ένα διάστημα $[a, b]$ χωρίς υποδιαίρεσεις θα ισχύει ότι:

$$\int_a^b f(x)dx \approx (b - a) \cdot \frac{f(a) + f(b)}{2}$$

Ύστερα από τον υπολογισμό των ποσοτήτων στη σχέση (5.3.13), να αναφέρουμε πως από τον ορισμό της νέας συνάρτησης $\psi(z)$ ισχύει ότι:

$$\frac{d}{dz} \varphi(z) = -\psi(z) \cdot \varphi(z) \quad (5.3.14)$$

Η παραπάνω ισότητα 5.3.14 σε όρους σειρών Taylor κεντραρισμένες στο σημείο z^* μπορεί να γραφθεί ισοδύναμα ως:

$$\sum_{k \geq 0} g_{k+1} (z - z^*)^k / k! = - \left(\sum_{n \geq 0} s_n (z - z^*)^n \right) \cdot \left(\sum_{k \geq 0} \frac{g_k (z - z^*)^k}{k!} \right)$$

Συγκρίνοντας και τα δύο μέλη της νέας πλέον ισότητας προκύπτει ότι έχουμε σταθερό όρο ίσο με $s_0 = -g_1/g_0$ και συγκρίνοντας αντίστοιχα και τις σταθερές μπροστά από τις ποσότητες $(z - z^*)^k$ προκύπτει ο εξής τύπος για τις ποσότητες s_k :

$$s_k = - \left(\frac{1}{g_0} \right) \cdot \left(\frac{g_{k+1}}{k!} + \sum_{i=0}^{k-1} s_i \cdot \frac{g_{k-i}}{(k-i)!} \right) \quad (5.3.15)$$

Η ισότητα 5.3.15 μας δίνει την δυνατότητα υπολογισμού των s_k για $k \geq 1$, μέσω αναδρομικών τύπων. Επομένως γνωρίζοντας τις ποσότητες αυτές είναι εφικτή και η εύρεση των παραμέτρων σχήματος και κλίμακας των κατανομών Γάμμα που εμπεριέχονται στην συνέλιξη που χρησιμοποιούμε ως προσέγγιση των κατανομών των ατομικών ζημιών.

Πρόταση 5.3.1 (Furman, 2018): Η παραπάνω μέθοδος προσέγγισης της κατανομής των συνολικών ζημιών στο συλλογικό πρότυπο της Θεωρίας κινδύνου με την χρήση συνελίξεων Γάμμα κατανομών είναι περισσότερο εποικοδομητική και μας δίνει επιθυμητά αποτελέσματα όταν η τ.μ. X την οποία θέλουμε να προσεγγίσουμε ανήκει στην οικογένεια των γενικευμένων Γάμμα συνελίξεων (Generalized Gamma Convolutions) (βλέπε Furman, 2018). Σε αυτή την περίπτωση ο αλγόριθμος που περιγράψαμε προηγουμένως θα μας δίνει θετικούς αριθμούς a_i, b_i που θα καθορίζουν την κατανομή της προσεγγιστικής τ.μ. $\tilde{X} = \sum_{i=1}^m \Gamma_i$, όπου $\Gamma_i \sim \text{Gamma}(a_i, b_i)$.

Η κλάση των γενικευμένων Γάμμα συνελίξεων είναι ιδιαίτερα σημαντική καθώς περιέχει πολλές κατανομές οι οποίες χρησιμοποιούνται ευρέως στην αναλογιστική επιστήμη, όπως η Γάμμα, η αντίστροφη Γάμμα, η Pareto, η λογαριθμοκανονική και η Weibull για παραμέτρους σχήματος μικρότερες της μονάδας.

Αλγόριθμος Gaver-Stehfest

Ο αλγόριθμος Gaver-Stehfest αποτελεί μια μέθοδο αντιστροφής των μετασχηματισμών Laplace, η οποία χρησιμοποιεί μόνο τις τιμές του μετασχηματισμού στον θετικό ημίαξονα και δεν απαιτεί πολύπλοκους υπολογισμούς. Η μέθοδος αυτή εφευρέθηκε από τον Stehfest το 1970 ως μια προσπάθεια βελτίωσης της μεθόδου του Gaver που είχε ήδη χρησιμοποιηθεί λίγο καιρό νωρίτερα. Από τότε ο αλγόριθμος αυτός, ο οποίος οφείλει το όνομα του και στους δύο προαναφερθέντες επιστήμονες, λόγω της απλότητας του και της καλής του επίδοσης έχει χρησιμοποιηθεί σε διάφορους τομείς των πιθανοτήτων και της στατιστικής, των χρηματοοικονομικών μαθηματικών, της αναλογιστικής επιστήμης αλλά και της υπολογιστικής φυσικής. Όπως είδαμε προηγουμένως οι προσεγγίσεις Pade βοήθησαν ώστε να υπολογίσουμε τις παραμέτρους σχήματος και κλίμακας των κατανομών Γάμμα στην συνέλιξη που θα προσεγγίσει την κατανομή της κάθε τ.μ. X_i . Έτσι θα μπορέσουμε να προσεγγίσουμε τον μετασχηματισμό Laplace $\varphi_i(z)$ της κάθε τ.μ. X_i και συνεπώς και τον μετασχηματισμό Laplace $\varphi(z) = \prod_{i=1}^n \varphi_i(z)$ του συνολικού κινδύνου S . Στην συνέχεια λοιπόν για να ολοκληρωθεί η παραπάνω διαδικασία προσέγγισης, ο αλγόριθμος αυτός θα αποτελέσει την βασική μέθοδο ώστε να αντιστρέψουμε τον μετασχηματισμό Laplace του συνολικού κινδύνου και να προσεγγίσουμε την συνάρτηση κατανομής της τ.μ. S (που είναι και το ζητούμενο).

Ας δούμε με λίγα λόγια τι ισχύει σχετικά με τον παραπάνω αλγόριθμο (Furman, 2018):

Θεωρούμε μια συνάρτηση $f(x)$ με αντίστοιχο μετασχηματισμό Laplace ίσο με

$$\varphi(z) := \int_0^{\infty} e^{-zx} f(x) dx$$

όπου θεωρούμε ότι η $\varphi(z)$ συγκλίνει για όλα τα $z > 0$ και ορίζουμε για όλους τους ακέραιους αριθμούς $m \geq 1$ την παρακάτω συνάρτηση:

$$f_m(x) := \ln(2) \cdot x^{-1} \sum_{k=1}^{2m} a_k(n) \varphi(k \cdot \ln(2) \cdot x^{-1}), \quad x > 0$$

με τους συντελεστές $a_k(n)$ να προκύπτουν μέσω του παρακάτω τύπου:

$$a_k(m) := \frac{(-1)^{m+k}}{m!} \sum_{j=\frac{k+1}{2}}^{\min(m,k)} j^{m+1} \binom{n}{j} \binom{2j}{j} \binom{j}{k-j}, \quad \text{για } m \geq 1 \text{ και } 1 \leq k \leq 2m$$

Οι προσεγγίσεις $f_m(x)$ θα συγκλίνουν στην $f(x)$ όταν η $f(x)$ είναι συνεχής ως προς x ενώ παράλληλα έχει αποδειχθεί ότι η σύγκλιση γίνεται με ιδιαίτερα γρήγορο τρόπο.

Το μόνο πρόβλημα του αλγορίθμου είναι ότι μπορεί στο άθροισμα για τον υπολογισμό των $f_m(x)$ να χαθούν σημαντικοί όροι λόγω του ότι οι συντελεστές $a_k(n)$ παίρνουν πολύ μεγάλες τιμές (για μεγάλες τιμές των k, m αντίστοιχα) και εναλλάσσονται εύκολα.

Μέσω λοιπόν μιας σειράς συναρτήσεων $f_m(x)$, όπως ορίζονται από τον αλγόριθμο Gaver-Stehfest, μπορούμε να προσεγγίσουμε την συνάρτηση $f(x)$ για την οποία θα γνωρίζουμε τον μετασχηματισμό Laplace της. Συνεπώς η μέθοδος αυτή θα βοηθήσει ώστε έχοντας προσεγγίσει τον μετασχηματισμό Laplace: $\varphi(z) = \prod_{i=1}^n \varphi_i(z)$ του συνολικού κινδύνου S να μπορέσουμε να υπολογίσουμε προσεγγιστικά την συνάρτηση πυκνότητας πιθανότητας και επομένως και την συνάρτηση κατανομής για τον συνολικό κίνδυνο $S = X_1 + X_2 + \dots + X_n$ που θα έχει τον συγκεκριμένο μετασχηματισμό τον οποίο γνωρίζουμε.

Γενικό συμπέρασμα: Είναι ιδιαίτερα φανερό ότι οι συνελίξεις κατανομών Γάμμα κατέχουν μια εξέχουσα θέση την αναλογιστική επιστήμη και συγκεκριμένα στον κλάδο της θεωρίας κινδύνου ενώ παράλληλα μπορούμε πλέον να τις προσεγγίσουμε με έναν αρκετά επικοδομητικό και αποτελεσματικό τρόπο. Επιπλέον μέσω της παραπάνω διαδικασίας όπου περιγράψαμε, παρατηρούμε ότι χωρίς την συμβολή των συνελίξεων Γάμμα δεν θα ήταν εφικτή η προσέγγιση των κατανομών των συνολικών απαιτήσεων στο συλλογικό πρότυπο αποζημιώσεων. Είναι πάρα πολύ χρήσιμες ως κατανομές και ο βασικός σκοπός της παρούσας διατριβής ήταν να αναδείξουμε την σπουδαιότητα τους. Τέλος και οι μείξεις όπως και οι συνελίξεις Γάμμα κατανομών έχουν σημαντική χρησιμότητα στον αναλογισμό και για τον λόγο αυτό η μελέτη βασικών ιδιοτήτων τους, όπως η ασυμμετρία, είναι απαραίτητη.

Σκοπός αυτής της ενότητας είναι να παρουσιάσουμε τους κώδικες που χρησιμοποιήθηκαν για την κατασκευή των διαγραμμάτων αλλά και τον υπολογισμό σημαντικών ποσοτήτων όπως αυτές παρατέθηκαν στο κύριο μέρος της παρούσας διατριβής. Για την διεξαγωγή των αποτελεσμάτων έχει γίνει χρήση της γλώσσας προγραμματισμού R και του μαθηματικού προγράμματος Wolfram Mathematica 8.

Όλα τα διαγράμματα έχουν πραγματοποιηθεί στο Mathematica εκτός από τα διαγράμματα στο κεφάλαιο 5 για τα οποία χρησιμοποιήθηκε η γλώσσα R. Οι κώδικες στην R έχουν παρατεθεί στο κύριο μέρος της διπλωματικής εργασίας καθώς δεν είχαν μεγάλη έκταση.

Διάγραμμα 1.1 : Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της $Gamma(0.5,2)$ (κόκκινη καμπύλη) και της $Gamma(3.5,5.5)$ (μπλε καμπύλη)

```
f[x_] := (x^(a1-1)) * (Exp[-x/b1]) / (b1^a1) * Gamma[a1]
f[x]

a1=0.5
b1=2

f2[x_] := (x^(a2-1)) * (Exp[-x/b2]) / (b2^a2) * Gamma[a2]
f2[x]

a2=3.5
b2=5.5

Plot[{f[x], f2[x]}, {x, 0, 100},
AxesLabel -> {"x", "f(x)"},
PlotRange -> {{0, 70}, {0, 2}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
Directive[RGBColor[0.78888, 0.999, 0.56], Thick]},
AxesLabel -> {Style[x, 16],
Style[f (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed]]
```

Διάγραμμα 1.2 : Διάγραμμα της συνάρτησης κατανομής της $Gamma(0.5,2)$ (κόκκινη καμπύλη) και της $Gamma(3.5,5.5)$ (μπλε καμπύλη)

```
F[x_] := Gamma[a1, 0, x/b1] / Gamma[a1]
F[x]

a1=0.5
b1=2
F2[x_] := Gamma[a2, 0, x/b2] / Gamma[a2]
F2[x]

Plot[{F[x], F2[x]}, {x, 0, 100},
AxesLabel -> {"x", "F(x)"},
PlotRange -> {{0, 70}, {0, 1}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
Directive[RGBColor[0.78888, 0.999, 0.56], Thick]},
AxesLabel -> {Style[x, 16],
Style[F (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
```

GridLines -> Automatic,

GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed]]

Διάγραμμα 1.3 : Διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της $Erl(2,3)$ (κόκκινη καμπύλη) και της $Erl(6,1.5)$ (μπλε καμπύλη)

```
f3[x_]:= (b3^a3)*(x^(a3-1))*Exp[-b3*x]/(a3-1)!
```

```
F3[x_]:= 1-Sum[(1/n!)*Exp[-b3*x]*(b3*x)^n,{n,0,a3-1}]
```

```
a3=2
```

```
b3=3
```

```
f3[x]
```

```
F3[x]
```

```
f4[x_]:= (b4^a4)*(x^(a4-1))*Exp[-b4*x]/(a4-1)!
```

```
F4[x_]:= 1-Sum[(1/n!)*Exp[-b4*x]*(b4*x)^n,{n,0,a4-1}]
```

```
a4=6
```

```
b4=1.5
```

```
f4[x]
```

```
F4[x]
```

```
Plot[{f3[x], f4[x]}, {x, 0, 100},
```

```
AxesLabel -> {"x", "f(x)"},
```

```
PlotRange -> {{0, 20}, {0, 1.5}},
```

```
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
```

```
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
```

```
Directive[RGBColor[0.78888, 0.999, 0.56], Thick}},
```

```
AxesLabel -> {Style[x, 16],
```

```
Style[f (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
```

```
GridLines -> Automatic,
```

```
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed]]
```

Διάγραμμα 1.4 : Διάγραμμα της συνάρτησης κατανομής της $Erl(2,3)$ (κόκκινη καμπύλη) και της $Erl(6,1.5)$ (μπλε καμπύλη)

```
Plot[{F3[x], F4[x]}, {x, 0, 100},
```

```
AxesLabel -> {"x", "F(x)"},
```

```
PlotRange -> {{0, 20}, {0, 1}},
```

```
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
```

```
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
```

```
Directive[RGBColor[0.78888, 0.999, 0.56], Thick}},
```

```

AxesLabel -> {Style[x, 16],
Style[F(x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed]]

```

Διάγραμμα 2.1: Βαθμίδα αποτυχίας για την εκθετική κατανομή με παράμετρο $\lambda = 0.25$ (κόκκινη γραμμή) και παράμετρο $\lambda = 2.5$ (μπλε γραμμή)

```

f[l_, x_] := 1*Exp[-l*x]
F[l_, x_] := 1 - Exp[-l*x]
<< "PlotLegends`"
Plot[{f[0.25, x]/(1 - F[0.25, x]), f[2.5, x]/(1 - F[2.5, x])}, {x, 0, 6},
AxesLabel -> {"x", "μ(x)"},
PlotRange -> {{0, 5}, {0, 3}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick]},
AxesLabel -> {Style[x, 16],
Style[μ(x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed],
PlotLegend -> {"Exp(0.25)", "Exp(2.5)"}, LegendSize -> {0.5, 0.3},
LegendPosition -> {0.2, 0.008}, LegendShadow -> None]

```

Διάγραμμα 2.2: Βαθμίδα αποτυχίας για την Γάμμα κατανομή με παραμέτρους (0.5,3), (0.5,1.5) και (4,2)

```

f[a_, b_, x_] := (b^a/Gamma[a])*(x^(a - 1))* Exp[-x*b];
F[a_, b_, x_] := Gamma[a, 0, b*x]/Gamma[a];
<< "PlotLegends`"
Plot[{f[0.5, 3, x]/(1 - F[0.5, 3, x]),
f[0.5, 1.5, x]/(1 - F[0.5, 1.5, x]),
f[4, 2, x]/(1 - F[4, 2, x])}, {x, 0, 6},
AxesLabel -> {"x", "μ(x)"},
PlotRange -> {{0.5}, {0.3}},
PlotStyle -> {Directive[RGBColor[0.587877, 0.24553, 0.16788], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],

```



```

Directive[RGBColor[0.890, 0.56788, 0.788899], Thick}},
AxesLabel -> {Style[x, 16],
Style[μ (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.5, 0.4], Dashed],
PlotLegend -> {"Ga(0.5,3)", "Ga(0.5,1.5)", "Ga(4,2)"},
LegendSize -> {0.5, 0.3},
LegendPosition -> {0.2, 0.1}, LegendShadow -> None]

```

Διάγραμμα 2.3: Βαθμίδα αποτυχίας για την κατανομή Erlang με παραμέτρους (3,0.5), (2,0.5) και (2,6)

```

f[b_, a_, x_] := (b^a)*(x^(a - 1))*Exp[-b*x]/(a - 1)!
F[b_, a_, x_] := 1 - Sum[(1/n!)*Exp[-b*x]*(b*x)^n, {n, 0, a - 1}];
<< "PlotLegends`"
Plot[{f[0.5, 3, x]/(1 - F[0.5, 3, x]),
f[0.5, 2, x]/(1 - F[0.5, 2, x]), f[6, 2, x]/(1 - F[6, 2, x])}, {x, 0, 6},
AxesLabel -> {"x", "μ(x)"},
PlotRange -> {{0.5}, {0.3}},
PlotStyle -> {Directive[RGBColor[0.587877, 0.24553, 0.16788], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
Directive[RGBColor[0.890, 0.56788, 0.788899], Thick]},
AxesLabel -> {Style[x, 16],
Style[μ (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.5, 0.4], Dashed],
PlotLegend -> {"erl(3,0.5)", "erl(2,0.5)", "erl(2,6)"},
LegendSize -> {0.5, 0.3},
LegendPosition -> {0.2, 0.1}, LegendShadow -> None]

```

Διάγραμμα 3.1: Μέσος υπολειπόμενος χρόνος ζωής για την εκθετική κατανομή με παράμετρο $\lambda = 0.25$ (κόκκινη γραμμή) και παράμετρο $\lambda = 2.5$ (μπλε γραμμή)

```

F[l_, x_] := 1 - Exp[-l*x]; x = Range[-1, 26];
m[l_] = Integrate[(1 - F[l, x + t])/(1 - F[l, x]), {t, 0, Infinity}];
<< "PlotLegends`"

```

```
ListLinePlot[{m[0.25], m[2.5]},
PlotRange -> {{0, 20}, {0, 6}},
AxesLabel -> {"x", "m(x)"},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick]},
AxesLabel -> {Style[x, 16],
Style[m (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed],
PlotLegend -> {"Exp(0.25)", "Exp(2.5)"}, LegendSize -> {0.5, 0.3},
LegendPosition -> {-0.01, -0.3}, LegendShadow -> None]
```

Διάγραμμα 3.2: Μέσος υπολειπόμενος χρόνος ζωής για την κατανομή Erlang με παραμέτρους (3,0.25), (2,0.25) και (2,4)

```
F[l_, a_, x_] := 1 - Sum[(1/n!)*Exp[-l*x]*(l*x)^n, {n, 0, a - 1}];
x = Range[1, 30];
m2[l_, a_, x_] :=
Integrate[(1 - F[l, a, x + t])/(1 - F[l, a, x]), {t, 0, Infinity}];
<< "PlotLegends`"
Plot[{m2[0.25, 3, x], m2[0.25, 2, x], m2[4, 2, x]}, {x, 0.01, 30},
PlotRange -> {{0, 20}, {0, 40}},
AxesLabel -> {"x", "m(x)"},
PlotStyle -> {Directive[RGBColor[0.587877, 0.24553, 0.16788], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
Directive[RGBColor[0.890, 0.56788, 0.788899], Thick]},
AxesLabel -> {Style[x, 16],
Style[m (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.5, 0.4], Dashed],
PlotLegend -> {"erl(3,0.25)", "erl(2,0.25)", "erl(2,4)"},
LegendSize -> {0.5, 0.3},
LegendPosition -> {0.2, 0.1}, LegendShadow -> None]
```

Διάγραμμα 3.3: Μέσος υπολειπόμενος χρόνος ζωής για την κατανομή Weibull με παραμέτρους (0.5,6), (0.5,2.5) και (4,2)

```
F[l_, a_, x_] := 1 - Exp[-(x/l)^a];
me[l_, a_, x_] :=
Integrate[(1 - F[l, a, x + t])/(1 - F[l, a, x]), {t, 0, 1000000}];
<< "PlotLegends`"
Plot[{me[6, 0.5, x], m2[2.5, 0.5, x], m2[2, 4, x]}, {x, 0.01, 30},
PlotRange -> {{0, 20}, {0, 40}},
AxesLabel -> {"x", "m(x)"},
PlotStyle -> {Directive[RGBColor[0.587877, 0.24553, 0.16788], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
Directive[RGBColor[0.890, 0.56788, 0.788899], Thick]},
AxesLabel -> {Style[x, 16],
Style[m (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.5, 0.4], Dashed],
PlotLegend -> {"Weibull(0.5,6)", "Weibull(0.5,2.5)", "Weibull(4,2)"},
LegendSize -> {0.7, 0.3},
LegendPosition -> {-0.01, -0.3}, LegendShadow -> None]
```

Πίνακας 4.1: Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για διάφορες μείξεις 2 εκθετικών κατανομών

```
f1[x_] := (b1)*(Exp[-b1*x])
f2[x_] := (b2)*(Exp[-b2*x])
f[x_] = a*f1[x] + (1 - a)*f2[x]
μ=Integrate[x*f[x], {x,0,Infinity}]
m2 = Integrate[((x - μ)^2)*f[x], {x, 0, Infinity}]
FullSimplify[m2]
sigma = Sqrt[m2]
m3 = Integrate[((x - μ)^3)*f[x], {x, 0, Infinity}]
γ1 = (m3)/(sigma^3)
m4 = Integrate[((x - μ)^4)*f[x], {x, 0, Infinity}]
γ2 = (m4)/(sigma^4)
```

Διάγραμμα 4.1: Συνάρτηση πυκνότητας πιθανότητας της μείξης $p \cdot \text{Gamma}(a_1, b_1) + (1 - p) \cdot \text{Gamma}(a_2, b_2)$ (πράσινη καμπύλη) με $a_1 = 4, b_1 = 1, a_2 = 10$ και $b_2 = 0.5$ και των συναρτήσεων $\text{Gamma}(4,1)$ (κόκκινη καμπύλη) και $\text{Gamma}(10,0.5)$ (μπλε καμπύλη). Με βάρη $p = 0.15$ και $1 - p = 0.85$ αντίστοιχα.

```
f1[x_] := (b1^a1)*(x^(a1 - 1))*(Exp[-b1*x])/Gamma[a1]
f2[x_] := (b2^a2)*(x^(a2 - 1))*(Exp[-b2*x])/Gamma[a2]
f[x_] = p*f1[x] + (1 - p)*f2[x]
a1 = 4
b1 = 1
a2 = 10
b2 = 0.5
p = 0.15
f[x]
<< "PlotLegends`"
Plot[{f1[x], f2[x], f[x]}, {x, 0, 100},
AxesLabel -> {"x", "f(x)"},
PlotRange -> {{0, 40}, {0, 0.5}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick],
Directive[RGBColor[0.294567, 0.4566777, 0.8897678], Thick],
Directive[RGBColor[0.78888, 0.999, 0.56], Thick]},
AxesLabel -> {Style[x, 16],
Style[f (x), 16]}, TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed]]
```

Για τον υπολογισμό των συντελεστών ασυμμετρίας και κύρτωσης για μείξεις 2 Γάμμα κατανομών με κοινή ή διαφορετική παράμετρο κλίμακας (b), για διάφορους συνδυασμούς τιμών των παραμέτρων σχήματος (a) και για διάφορες περιπτώσεις βαρών (όπως παραθέτουμε στους πίνακες 4.2, 4.3, 4.4, 4.5 και 4.6) το κύριο μέρος του κώδικα είναι το εξής:

```
f1[x_] := (b1^a1)*(x^(a1 - 1))*(Exp[-b1*x])/Gamma[a1]
f2[x_] := (b2^a2)*(x^(a2 - 1))*(Exp[-b2*x])/Gamma[a2]
f[x_] = p*f1[x] + (1 - p)*f2[x]
μ = Integrate[x*f[x], {x, 0, Infinity}]
m2 = Integrate[((x - μ)^2)*f[x], {x, 0, Infinity}]
```

```

FullSimplify[m2]
sigma = Sqrt[m2]
m3 = Integrate[((x - μ)^3)*f[x], {x, 0, Infinity}]
γ1 = (m3)/(sigma^3)
m4 = Integrate[((x - μ)^4)*f[x], {x, 0, Infinity}]
γ2 = (m4)/(sigma^4)

```

Διάγραμμα 4.2: Συνάρτηση πυκνότητας πιθανότητας (αριστερά) και βαθμίδας αποτυχίας (δεξιά) για την μείξη των κατανομών $\text{Gamma}(20,5)$ και $\text{Gamma}(20,30)$ με βάρη $p = 0.8$ και $1 - p = 0.2$ αντίστοιχα, για την οποία διαπιστώσαμε αρνητική ασυμμετρία

```

f1[x_] := (b1^a1)*(x^(a1 - 1))*(Exp[-b1*x])/Gamma[a1]
f2[x_] := (b2^a2)*(x^(a2 - 1))*(Exp[-b2*x])/Gamma[a2]
f[x_] = p*f1[x] + (1 - p)*f2[x]
a1 = 20
b1 = 5
a2 = 20
b2 = 30
p = 0.8
f[x]
F1[x_] := Integrate[f1[y], {y, 0, x}]
TailF1[x_] := 1 - F1[x]
F2[x_] := Integrate[f2[y], {y, 0, x}]
TailF2[x_] := 1 - F2[x]
t[x_] := (p*TailF1[x])/(p*TailF1[x] + (1 - p)*TailF2[x])
μ1[x_] := f1[x]/TailF1[x]
μ2[x_] := f2[x]/TailF2[x]
μ[x_] := t[x]*μ1[x] + (1 - t[x])*μ2[x]
<< "PlotLegends`"
Plot[{μ[x]}, {x, 0, 100},
AxesLabel -> {"x", "μ(x)"},
PlotRange -> {{0, 20}, {0, 30}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick]},
AxesLabel -> {Style[x, 16],

```

```

Style[μ (x), 16]], TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed],
PlotLegend -> {"0.8*Gamma(20,5)+0.2*Gamma(20,30)"},
LegendSize -> {1.49, 0.1},
LegendPosition -> {-0.8, 0.3}, LegendShadow -> False]
<< "PlotLegends`"

```

```

Plot[{f[x]}, {x, 0, 100},
AxesLabel -> {"x", "f(x)"},
PlotRange -> {{0, 20}, {0, 30}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick]},
AxesLabel -> {Style[x, 16],
Style[f (x), 16]], TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed],
PlotLegend -> {"0.8*Gamma(20,5)+0.2*Gamma(20,30)"},
LegendSize -> {1.49, 0.1},
LegendPosition -> {-0.8, 0.3}, LegendShadow -> False]

```

Διάγραμμα 4.3: Συνάρτηση πυκνότητας της βαθμίδας αποτυχίας του διαγράμματος (4.2) στο διάστημα [15,17)

```

<< "PlotLegends`"
Plot[{μ[x]}, {x, 0, 100},
AxesLabel -> {"x", "μ(x)"},
PlotRange -> {{15, 17}, {0, 30}},
PlotStyle -> {Directive[RGBColor[0.5856, 0.22876, 0.17089], Thick]},
AxesLabel -> {Style[x, 16],
Style[μ (x), 16]], TicksStyle -> Directive[FontSize -> 13],
GridLines -> Automatic,
GridLinesStyle -> Directive[RGBColor[0.6, 0.6, 0.6], Dashed],
PlotLegend -> {"0.8*Gamma(20,5)+0.2*Gamma(20,30)"},
LegendSize -> {1.49, 0.1},

```

LegendPosition -> {-0.8, 0.3}, LegendShadow -> False]

Αντίστοιχοι είναι οι κώδικες των διαγραμμάτων 4.4 έως 4.12

Πίνακας 4.7: Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για συνελίξεις 2 Γάμμα κατανομών με διαφορετική παράμετρο κλίμακας (b) και κοινή παράμετρο σχήματος (a)

```
f1[x_] := (b1^a)*(x^(a - 1))*(Exp[-b1*x])/Gamma[a]
f2[y_] := (b2^a)*(y^(a - 1))*(Exp[-b2*y])/Gamma[a]
f[z_] = Integrate[f1[x]*f2[z - x], {x, 0, z}, Assumptions -> {Re[a] > 0}]
μ = Integrate[z*f[z], {z, 0, Infinity}, Assumptions -> {Re[a] > 0}]
FullSimplify[μ]
m2 = Integrate[((x - μ)^2)*f[x], {x, 0, Infinity}, Assumptions -> {Re[a] > 0}]
FullSimplify[m2]
sigma = Sqrt[m2]
m3 = Integrate[((x - μ)^3)*f[x], {x, 0, Infinity}, Assumptions -> {Re[a] > 0}]
γ1 = (m3)/(sigma^3)
m4 = Integrate[((x - μ)^4)*f[x], {x, 0, Infinity}, Assumptions -> {Re[a] > 0}]
γ2 = (m4)/(sigma^4)
μ3 = Integrate[(x^3)*f[x], {x, 0, Infinity}, Assumptions -> {Re[a] > 0}]
γ11 = (μ3 - 3*μ*(sigma^2) - (μ^3))/(sigma^3)
a = 0.1
b1 = 0.4
b2 = 0.8
γ1
γ11
γ2
```

Πίνακας 4.8: Ζεύγη συντελεστών ασυμμετρίας και κύρτωσης για συνελίξεις 2 Γάμμα κατανομών με διαφορετική παράμετρο κλίμακας (b) και για διάφορες περιπτώσεις παραμέτρων σχήματος (a)

```
f1[x_] := (b1^a1)*(x^(a1 - 1))*(Exp[-b1*x])/Gamma[a1]
f2[y_] := (b2^a2)*(y^(a2 - 1))*(Exp[-b2*y])/Gamma[a2]
f[z_] = Integrate[f1[x]*f2[z - x], {x, 0, z}]
μ = Integrate[z*f[z], {z, 0, Infinity}]
```

```

μ2 = Integrate[(z^2)*f[z], {z, 0, Infinity}]
sigma = Sqrt[μ2 - (μ^2)]
m3 = Integrate[((z - μ)^3)*f[z], {z, 0, Infinity}]
FullSimplify[m3]
γ1 = (m3)/(sigma^3)
FullSimplify[γ1]
m4 = Integrate[((x - μ)^4)*f[x], {x, 0, Infinity}]
γ2 = (m4)/(sigma^4)
a1 = 0.1
a2 = 10
b1 = 40
b2 = 50
γ1
γ2

```

Για τον υπολογισμό των ροπών της πυκνότητας του Moschopoulos και της πυκνότητας του Barnabani στο 5^ο κεφάλαιο χρησιμοποιήθηκαν οι παρακάτω εντολές στην R:

```

> integrate(function(x) (x^3)*dcoga_approx(x,c(3,10,20),c(6,8,10)),0,Inf)
> integrate(function(x) (x^3)*dcoga(x,c(3,10,20),c(6,8,10)),0,Inf)

> integrate(function(x) (x^4)*dcoga_approx(x,c(3,10,20),c(6,8,10)),0,Inf)
> integrate(function(x) (x^4)*dcoga(x,c(3,10,20),c(6,8,10)),0,Inf)

> integrate(function(x) (x^5)*dcoga_approx(x,c(3,10,20),c(6,8,10)),0,Inf)
> integrate(function(x) (x^5)*dcoga(x,c(3,10,20),c(6,8,10)),0,Inf)

> integrate(function(x) (x^6)*dcoga_approx(x,c(3,10,20),c(6,8,10)),0,Inf)
> integrate(function(x) (x^6)*dcoga(x,c(3,10,20),c(6,8,10)),0,Inf)

```


ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Abramowitz M. and Stegun I. (1972), Handbook of Mathematical Functions. 10th edition. Dover, New York.
- Arnold, Barry C. and Richard A. Groeneveld (1995) Measuring Skewness with respect to the mode. The American Statistician, Vol 49(1), 34-38.
- Bahnemann D. (2015) Distributions for actuaries. Casualty actuarial society, Virginia.
- Barlow R. and Proschan F. (1975) Statistical Theory of Reliability and Life Testing Probability Models. Holt, Rinehart and Winston.
- Barnabani M. (2017) An approximation to the convolution of Gamma distributions. Communications in Statistics-Simulation and Computation, Vol 46(1), 331-345.
- Block H., Langberg N., Savits T. (2012) A Mixture of Exponential and IFR Gamma distributions having an Upsidedown Bathtub-Shaped Failure Rate. Probability in the Engineering and Informational Sciences, Vol 26, 573-850.
- Bondesson L. (1983) On preservation of classes of life distributions under reliability operations: some complementary results. Naval Research Logistics Quarterly, Vol 30 (3), 443-447.
- Chaoran Hu, Vladimir Pozdnyakov and Jun Yan (2019) Density and Distribution Evaluation for Convolution of Independent Gamma Variables. Computational Statistics, Vol 35, 327-342.
- Charlier C.V.L. (1906) Uber das Fehlergesetz. Meddelanden fran Lunds Astronomiska Observatorium, Vol. 25, Gleerups, Lund.
- Consul P.C. and Gupta H.C (1980) The generalized negative binomial distribution and its characterization by zero regression. SIAM Journal on Applied Mathematics, Vol 39 (2) , 231-237.
- Diaconis P. and Perlman M.D. (1990) Bounds for tail probabilities of weighted sums of independent gamma random variables. Topics in Statistical Dependence, Vol 16, 147-166.
- Edgeworth F.Y (1904) The law of error. Transactions of the Cambridge Philosophical Society , Vol 20, Part I , 36-65 and 113-114.
- Furman Edward, Justin Miles and Alexey Kuznetsov (2018) Risk aggregation: A general approach via the class of Generalized Gamma Convolutions. Research Gate.
- Guess F. and Proschan F. (1988) Mean residual life: theory and applications. Handbook of Statistics, Vol 7, 215-224.
- Jain G.C and Consul P.C (1971) A Generalized negative binomial distribution. SIAM Journal on Applied Mathematics, Vol 21(4), 501-513.

- Marshall W. and Olkin I. (2007) Life Distributions: Structure of Nonparametric Semiparametric and Parametric Families. Springer Science and Business Media, New York, USA.
- Marshall W. , Esary J and Proschan F. (1973) Shock models and wear processes. The Annals of Probability, Vol 1(4), 627-649.
- Mathai A.M (1982) Storage capacity of a dam with Gamma type inputs. Ann. Inst. Statist. Math. , Vol 34, Part I, 591-597.
- Moschopoulos P.G. (1985) The distribution of the sum of independent Gamma random variables. Ann. Inst. Statist. Math., Vol 35, Part I, 541-544.
- Shaked M. and Shanthikumar J. (2007) Stochastic orders. Springer Science and Business Media, New York, USA.
- Stewart T., Strijbosch L., Moors H, Van Batenburg P. (2007) A simple approximation to the convolution of Gamma distributions. Center Discussion Paper, No 2006-27, ISSN 0924-7815, Revised Version Sept. 2007.
- Takahasi H. and Mori M. (1974) Double exponential formulas for numerical integration. Publ. RIMS, Kyoto Univ, Vol 9, 721-741.
- Van Zwet (1964) Convex transformations of random variables. Mathematisch Centrum
- Vellaisamy P. and Upadhye N.S. (2009) On the sums of compound negative binomial and gamma random variables. Journal of Applied probability, Vol 46(1) , 272-283.
- Willmot G.E. and Lin X.S. (2001) Lundberg approximations for compound distributions with Insurance applications. Springer Science and Business media, New York, USA.
- Willmot G.E. and Lin X.S. (1994) Lundberg bounds on the tails of compound distributions. Journal of Applied probability, Vol 31(3), 743-756.

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Δημάκη Α. (2013) Πανεπιστημιακές σημειώσεις στο μάθημα «Θεωρία κατανομών». Τμήμα Στατιστικής, Οικονομικό Πανεπιστήμιο Αθηνών.
- Πανάρετος Ι. και Ξεκαλάκη Ε. (2000) Εισαγωγή στην Στατιστική Σκέψη. Τόμος Ι Εκδόσεις Μπένου.
- Πιτσέλης Γ. (2018) Κατανομές Μ.Π.Λ.Ο.Υ.Ζ, Εκδόσεις Παπαζήση.
- Πολίτης Κ. (2012) Εισαγωγή στην θεωρία συλλογικού κινδύνου. Εκδόσεις Σταμούλη
- Πολίτης Κ. (2018) Πανεπιστημιακές σημειώσεις στο μάθημα «Θεωρία Κινδύνου 2» του ΠΜΣ Αναλογιστικής επιστήμης και Διοικητικής Κινδύνου.
- Τσάκλα Χ. (2014) Διπλωματική εργασία: Η συσχέτιση του πλεονάσματος πριν και μετά τη χρεοκοπία στη θεωρία κινδύνων.
- Τσίπα Μ. (2018) Διπλωματική εργασία: Προσεγγίσεις κατανομών Pareto από συνδυασμούς εκθετικών κατανομών με εφαρμογές στη θεωρία κινδύνων.
- Χαραλαμπίδης Χ. (2009) Πανεπιστημιακές σημειώσεις στο μάθημα «Βασικές συνεχείς κατανομές». Τμήμα Μαθηματικών, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.
- Χατζηκωνσταντινίδης Ε. (2018) Πανεπιστημιακές σημειώσεις στο μάθημα «Συμβάντα Ζωής και Θανάτου Ι» του ΠΜΣ Αναλογιστικής επιστήμης και Διοικητικής Κινδύνου.

Για την χρήση των υπολογιστικών προγραμμάτων χρησιμοποιήθηκε η εξής βιβλιογραφία:

MATHEMATICA

- Δερμιτζάκης Β. (2018) Σημειώσεις για το εργαστήριο του μαθήματος «Θεωρία Κινδύνου 2» του ΠΜΣ Αναλογιστικής επιστήμης και Διοικητικής Κινδύνου.
- Peasly E. (2013) An Introduction to Mathematica. University of Oxford, Oxford.

ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R

- Ατζουλάκος Δ. (2018) Πανεπιστημιακές σημειώσεις στο μάθημα «Γλώσσα προγραμματισμού R με εφαρμογές στον αναλογισμό» του ΠΜΣ Αναλογιστικής επιστήμης και Διοικητικής Κινδύνου.
- Ντζούφρας Ι. και Καρλής Δ. (2015) Εισαγωγή στον προγραμματισμό και την στατιστική ανάλυση με R. Εκδόσεις Κάλλιπος.

