

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ**  
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



**ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ**  
ΤΟΜΕΑΣ ΠΙΘΑΝΟΤΗΤΩΝ, ΣΤΑΤΙΣΤΙΚΗΣ  
& ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

**ΒΑΣΙΛΕΙΟΥ Α. ΠΑΠΠΑ**



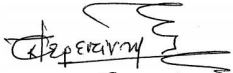



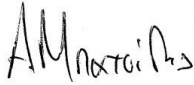
**ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΑΣ  
ΣΤΗΝ ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ**

**ΙΩΑΝΝΙΝΑ, 2013**



Η παρούσα Διδακτορική Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του διδακτορικού Διπλώματος που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 29/03/2013 από την επταμελή εξεταστική επιτροπή:


<b>ΟΝΟΜΑΤΕΠΩΝΥΜΟ</b>	<b>ΒΑΘΜΙΑ</b>	<b>ΥΠΟΓΡΑΦΗ</b>
Σωτήριος Λουκάς (Επιβλέπων)	Καθηγητής	
Κωνσταντίνος Αδαμίδης	Αναπλ. Καθηγητής	
Κοσμάς Φερεντίνος	Ομότιμος Καθηγητής	
Κωνσταντίνος Ζωγράφος	Καθηγητής	
Σταύρος Κουρούκλης	Καθηγητής	
Κωνσταντίνα Σκούρη	Επικ. Καθηγήτρια	
Απόστολος Μπατσίδης	Λέκτορας	

Η έγκριση της Διδακτορικής Διατριβής από το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων δεν υποδηλοί αποδοχή των γνώμών του συγγραφέα (Άρθρο 202 2, Ν. 5343/1932 και Άρθρο 50 8, Ν.1268/1982)



## ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

“Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή.”



Βασίλειος Α. Παπάς

Μαθηματικός, ΜΔΕ στη Στατιστική και Επιχειρησιακή Έρευνα



Στους γονείς μου,  
Αλέξανδρο και Χρυσούλα.





# Ευχαριστίες

Η συγγραφή μιας διδακτορικής διατριβής αποτελεί για τον υποψήφιο διδάκτορα ένα πολυετές πόνημα υπομονετικής προσήλωσης στην έρευνα. Αισθάνομαι ευγνώμων που σε αυτό μου το εγχείρημα έτυχα της αρωγής διαφόρων ατόμων τους οποίους και ευχαριστώ από καρδιάς.

Ειδικότερα, ευχαριστώ το Ίδρυμα Κρατικών Υποτροφιών για την υποτροφία που μου χορήγησε κατά τα δύο πρώτα έτη εκπόνησης της διατριβής, βοηθώντας με να ξεπεράσω τις όποιες οικονομικές δυσκολίες.

Πρωτίστως, ευχαριστώ ιδιαίτερα τον κ. Σωτήριο Λουκά, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, ο οποίος πίστεψε στις δυνατότητές μου, ανέλαβε με προθυμία το ρόλο του επιβλέποντα και στάθηκε ουσιαστικότερος αρωγός σε όλη τη διάρκεια εκπόνησης της διατριβής.

Παράλληλα, ευχαριστώ τον κ. Κωνσταντίνο Αδαμίδα, Αναπληρωτή Καθηγητή του Τμήματος Διοίκησης Επιχειρήσεων Αγροτικών Προϊόντων και Τροφίμων του Πανεπιστημίου Δυτικής Ελλάδας, η καθοριστικότερη βοήθεια του οποίου συνέβαλλε τα μέγιστα στην ολοκλήρωση της διατριβής.

Επιπρόσθετα, ευχαριστώ τον κ. Κοσμά Φερεντίνο, Ομότιμο Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για τη συμμετοχή του στην τριμελή επιτροπή, τις πολύτιμες συμβουλές και τη σημαντική συμβολή του στην υλοποίηση της διατριβής.

Ακόμη, θα ήθελα να ευχαριστήσω τον κ. Κωνσταντίνο Ζωγράφο, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για την εποικοδομητικές παρατηρήσεις και επισημάνσεις, που συνέβαλλαν στην βελτιστοποίηση της διατριβής.

Θέλω επίσης να ευχαριστήσω τους υπόλοιπους διδάσκοντες, διδάκτορες και μεταπτυχιακούς φοιτητές, καθώς και τη γραμματεία του Τομέα

Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας για την άριστη συνεργασία που είχαμε καθ' όλη τη διάρκεια εκπόνησης της διατριβής.

Εκφράζω τέλος ένα μεγάλο ευχαριστώ στην οικογένειά μου, τους γονείς μου Αλέξανδρο και Χρυσούλα και τις αδερφές μου Κατερίνα και Μαρία, που με στηρίζουν υλικά και πνευματικά εξ απαλών ονύχων.

Βασίλειος Α. Παπάς

Φεβρουάριος, 2013

# Περιεχόμενα

	σελ
<b>1. Εισαγωγή</b> .....	<b>1</b>
1.1 Ιστορική αναδρομή .....	4
1.2 Ορολογία και βασικές έννοιες .....	5
1.3 Κατανομές ανάλυσης επιβίωσης .....	15
1.3.1 Τρόποι παραγωγής .....	15
1.3.2 Αναλυτική μεθοδολογία μελέτης .....	16
1.3.3 Βασικές κατανομές ανάλυσης επιβίωσης .....	21
<b>2. Η Επεκτεταμένη Κατανομή του Chen</b> .....	<b>31</b>
2.1 Το μοντέλο .....	32
2.2 Συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής .....	35
2.3 Στατιστική συμπερασματολογία .....	37
2.4 Εφαρμογές .....	40
2.5 Συμπεράσματα .....	43
<b>3. Η Weibull-Διωνυμική Κατανομή</b> .....	<b>45</b>
3.1 Η κατανομή .....	46
3.2 Ιδιότητες .....	49
3.3 Ε.Μ.Π. και ασυμπτωτικά διαστήματα εμπιστοσύνης .....	53
3.4 Εφαρμογές .....	55
3.5 Συμπεράσματα .....	60
<b>4. Η Γενικευμένη Κατανομή EXTG</b> .....	<b>61</b>
4.1 Η γενικευμένη οικογένεια κατανομών .....	61
4.2 Μια επεκτεταμένη κατανομή Weibull .....	63
4.2.1 Συναρτήσεις κινδύνου και μέσου υπολειπόμενου χρόνου ζωής .....	66
4.2.2 Στατιστική συμπερασματολογία .....	68
4.2.3 Εφαρμογές .....	75
4.3 Συμπεράσματα .....	76
<b>5. Η Γενικευμένη Εκθετική-Λογαριθμική Κατανομή</b> .....	<b>79</b>
5.1 Το μοντέλο .....	79

5.2 Συνάρτηση πυκνότητας πιθανότητας, συνάρτηση κινδύνου και διατεταγμένα στατιστικά .....	81
5.3 Ιδιότητες .....	88
5.4 Εντροπίες .....	90
5.5 Στατιστική συμπερασματολογία .....	91
5.6 Εφαρμογές .....	95
5.7 Συμπεράσματα .....	96
<b>6. Βήτα Γενικευμένες Κατανομές .....</b>	<b>99</b>
6.1 Η γενικευμένη οικογένεια .....	99
6.2 Ειδικές περιπτώσεις κατανομών .....	107
6.3 Προτάσεις για μελλοντική έρευνα .....	126
<b>Περίληψη .....</b>	<b>127</b>
<b>Abstract .....</b>	<b>131</b>
<b>Βιβλιογραφία .....</b>	<b>135</b>

---

---

# Κεφάλαιο 1

## Εισαγωγή

Η ανάλυση επιβίωσης αποτελεί ταυτόσημη έννοια της ανάλυσης του χρονικού διαστήματος που προηγείται της καταγραφής ενός συμβάντος. Ως όρος χρησιμοποιείται κατά κύριο λόγο σε βιοϊατρικές επιστήμες, όπου το ενδιαφέρον επικεντρώνεται στην παρατήρηση του χρόνου ζωής είτε των ασθενών είτε των πειραματόζωων. Η ανάλυση του χρονικού διαστήματος που μεσολαβεί μέχρι την καταγραφή ενός συμβάντος έχει επίσης χρησιμοποιηθεί ευρέως στις κοινωνικές επιστήμες, όπου το ενδιαφέρον εστιάζεται σε γεγονότα όπως η αλλαγή θέσεως εργασίας, ο γάμος, η γέννηση παιδιών και ούτω καθεξής. Επιπρόσθετα, οι επιστήμες της μηχανικής έχουν συμβάλει στην ανάπτυξη της ανάλυσης επιβίωσης. Στον επιστημονικό αυτό κλάδο συνηθίζεται να χρησιμοποιείται ο όρος «ανάλυση αξιοπιστίας», μιας και η κύρια εφαρμογή του είναι η μοντελοποίηση του χρόνου που απαιτείται για την αποτυχία μηχανημάτων ή ηλεκτρονικών εξαρτημάτων. Η μελέτη όλων αυτών των επιστημονικών εξειδικεύσεων έχει ως επί το πλείστον ενοποιηθεί στο πεδίο της ανάλυσης επιβίωσης.

Γίνεται εύκολα αντιληπτό ότι η περιγραφή και η μοντελοποίηση του χρόνου ζωής, αποτελεί το κύριο ζητούμενο σε πολλές εκφάνσεις της ανθρώπινης ζωής και δραστηριότητας. Ως εκ τούτου, οι μεθοδολογίες που αναπτύσσονται στην ανάλυση επιβίωσης αποτελούν πολύτιμο εργαλείο για ερευνητές διαφόρων επιστημονικών κλάδων, όπως της μηχανικής, της οικονομίας, της ιατρικής και των υπόλοιπων βιολογικών επιστημών.

Η παρούσα διδακτορική διατριβή πραγματεύεται θέματα που αφορούν σε συνεχείς κατανομές πιθανότητας που έχουν άμεση εφαρμογή στην ανάλυση επιβίωσης. Στο **πρώτο κεφάλαιο**, παρουσιάζεται μια σύντομη ιστορική αναδρομή του ερευνητικού πεδίου της ανάλυσης επιβίωσης, καταγράφεται η ορολογία και οι βασικές έννοιες συνεχών κατανομών

χρόνων ζωής και περιγράφονται οι διάφορες χαρακτηριστικές συναρτήσεις και ιδιότητες που αυτές έχουν. Ειδική βαρύτητα δίνεται στη συνάρτηση κινδύνου και τη σημασία της στην περιγραφή της ικανότητας επιβίωσης ενός πληθυσμού. Επιπρόσθετα, παρουσιάζονται τρόποι παραγωγής κατανομών χρόνων ζωής, η αναλυτική μεθοδολογία μελέτης τους, καθώς και οι πιο χαρακτηριστικές κατανομές με σταθερές, μονοκόρυφες ή κοίλες συναρτήσεις κινδύνου.

Στο **δεύτερο κεφάλαιο**, παρουσιάζεται μια κατανομή ανάλυσης επιβίωσης τριών παραμέτρων με αύξουσα, φθίνουσα ή κοίλη συνάρτηση κινδύνου. Μελετώνται διάφορες ιδιότητές της, όπως οι τρόποι παραγωγής της, οι σχέσεις της με άλλες κατανομές, η μονοτονία της συνάρτησης πυκνότητας πιθανότητας (σ.π.π.) της, οι ροπές της και οι αντίστοιχες συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας και τους πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας, ενώ για τον υπολογισμό των εκτιμητών γίνεται χρήση της γλώσσας προγραμματισμού στατιστικών εφαρμογών R. Το κεφάλαιο ολοκληρώνεται με συγκριτικά σε σχέση με άλλες κατανομές, αποτελέσματα από την προσαρμογή της σε πραγματικά σύνολα δεδομένων. Η κατανομή και τα παραγόμενα από τη μελέτη της αποτελέσματα, παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2011).

Στο **τρίτο κεφάλαιο**, μελετάται ένα μοντέλο ανάλυσης επιβίωσης τριών παραμέτρων με μονότονες, μονοκόρυφες και τροποποιημένες κοίλες συναρτήσεις κινδύνου. Ερευνώνται διάφορες μαθηματικές και στατιστικές ιδιότητές του και παρουσιάζεται η φυσική διεργασία παραγωγής του, μέσα από τη θεωρία ανταγωνιστικών κινδύνων. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας. Τέλος, παρουσιάζονται εφαρμογές του μοντέλου σε δεδομένα που απορρέουν από πραγματικές καταστάσεις και η προσαρμογή του σε αυτά συγκρίνεται με άλλη εναλλακτική

κατανομή. Τα αποτελέσματα αυτού του κεφαλαίου παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2010).

Στο **τέταρτο κεφάλαιο**, εισάγεται μια οικογένεια κατανομών μέσω της οποίας δύναται να γενικευτεί μια κατανομή ανάλυσης επιβίωσης, με την προσθήκη μιας επιπλέον παραμέτρου, σε αναλογία με τη μελέτη των Marshall and Olkin (1997). Στη συνέχεια, η οικογένεια κατανομών χρησιμοποιείται για την ανάπτυξη μιας τετρα-παραμετρικής τροποποιημένης επέκτασης της κατανομής Weibull, με αύξουσα, φθίνουσα ή κοίλη συνάρτηση κινδύνου. Μελετώνται διάφορες ιδιότητές της, όπως οι σχέσεις της με άλλες κατανομές, η μονοτονία της σ.π.π. της, οι ροπές της και οι αντίστοιχες συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας και τους πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας, ενώ για τον υπολογισμό των εκτιμητών γίνεται χρήση της γλώσσας προγραμματισμού στατιστικών εφαρμογών R. Το κεφάλαιο ολοκληρώνεται με συγκριτικά σε σχέση με άλλες κατανομές αποτελέσματα από την προσαρμογή της σε πραγματικά σύνολα δεδομένων. Η οικογένεια κατανομών καθώς και η αναλυτικά μελετώμενη κατανομή και τα παραγόμενα αυτής αποτελέσματα, παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2012).

Στο **πέμπτο κεφάλαιο**, παρουσιάζεται μια κατανομή ανάλυσης επιβίωσης τριών παραμέτρων με αύξουσα, φθίνουσα ή μονοκόρυφη συνάρτηση κινδύνου. Μελετώνται διάφορες ιδιότητές της, όπως οι τρόποι παραγωγής της, οι σχέσεις της με άλλες κατανομές, η μονοτονία της σ.π.π. της, οι ροπές της, η εντροπία και οι αντίστοιχες συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας και τους πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας, ενώ για τον υπολογισμό των εκτιμητών γίνεται χρήση της γλώσσας προγραμματισμού στατιστικών εφαρμογών

R. Το κεφάλαιο ολοκληρώνεται με εφαρμογές της κατανομής σε πραγματικά σύνολα δεδομένων. Η κατανομή και τα παραγόμενα από τη μελέτη της αποτελέσματα, παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2013).

Στο **έκτο κεφάλαιο**, παρουσιάζονται οι βήτα γενικευμένες κατανομές πιθανότητας. Αναφέρεται ο τρόπος παραγωγής τους και περιγράφονται οι βασικές μαθηματικές ιδιότητές τους. Επιπρόσθετα, γίνεται αναφορά σε διάφορες κατανομές που ανήκουν στην οικογένεια και έχουν προταθεί μέχρι σήμερα στη βιβλιογραφία, ενώ δίνονται και προτάσεις για μελλοντική έρευνα στο συγκεκριμένο πεδίο.

Τέλος, παρατίθεται περίληψη της διδακτορικής διατριβής στην ελληνική και αγγλική γλώσσα, και η σχετική βιβλιογραφία που χρησιμοποιήθηκε κατά τη διάρκεια εκπόνησης της έρευνας.

## 1.1 Ιστορική αναδρομή

Το επίπεδο γνώσης που έχει αποκτηθεί έως σήμερα για την ανάλυση επιβίωσης προκύπτει από μία διαδικασία εξέλιξης που διήρκεσε αρκετά χρόνια και η οποία εμφανίζει ιδιαίτερη ανάπτυξη τα τελευταία χρόνια. Διάφοροι επιστημονικοί τομείς, όπως η Βιολογία, η Ιατρική, η Φαρμακευτική, ακόμη και κλάδοι της βιομηχανίας κ.α., συνέβαλλαν σε αυτήν την ανάπτυξη, επιχειρώντας να βρουν λύση και να αντιμετωπίσουν τα διάφορα προβλήματα που αφορούν την ερευνά τους.

Η δημιουργία του ερευνητικού κλάδου της ανάλυσης επιβίωσης χρονολογείται ήδη από το 17<sup>ο</sup> αιώνα (1662), οπότε και δημοσιεύθηκε στη Μ. Βρετανία το πρώτο βιβλίο, το οποίο περιείχε καταγεγραμμένους καταλόγους με στοιχεία γεννήσεων και θανάτων, που αναφέρονταν στις προηγούμενες δεκαετίες. Έτσι, ήταν η πρώτη φορά που οι «θάνατοι» αντιμετωπίστηκαν ως «γεγονότα», για τα οποία έγιναν αναλυτικές μελέτες.

Οι παγκόσμιοι πόλεμοι που ακολούθησαν, έδωσαν το έναυσμα για περαιτέρω ανάπτυξη της έρευνας στον τομέα της αξιοπιστίας και στη



μελέτη της διάρκειας ζωής των στρατευμάτων, αλλά και αργότερα η έρευνα επικεντρώθηκε στη μελέτη κάποιων ιδιαίτερων πιθανοθεωρητικών προβλημάτων σχετιζόμενων με την «παύση λειτουργίας» και την «αντικατάσταση» εξαρτημάτων μηχανικών ή ηλεκτρονικών κυκλωμάτων, όπως κάποιας βαλβίδας, ή ενός θερμοστάτη σε ένα μηχανικό κύκλωμα, μιας λυχνίας, ή μίας αντίστασης σε ένα ηλεκτρικό κύκλωμα. Υπήρξε δηλαδή μεγάλη πρόοδος στη βιομηχανία των ηλεκτρονικών συσκευών.

Οι μέθοδοι της ανάλυσης επιβίωσης είχαν τεράστιες εφαρμογές σε κλινικά δεδομένα και σκοπός ήταν να απαντηθούν ερωτήματα όπως ποια είναι η πιθανότητα ένας ασθενής να ζήσει μέχρι μια συγκεκριμένη χρονική στιγμή, ή με ποιο ρυθμό θα πεθάνουν κάποιοι ασθενείς, οι οποίοι έχουν ήδη επιβιώσει μέχρι ένα συγκεκριμένο χρονικό σημείο (κ.α.).

Η ανάπτυξη των επιχειρησιακών ερευνών κατέδειξε ότι υπάρχουν και πολλά άλλα προβλήματα και μοντέλα διαφορετικής υφής, στα οποία η ανάλυση επιβίωσης μπορεί να εφαρμοστεί και να προσφέρει λύσεις. Ακόμα περισσότερο με την ταχύτατη εξέλιξη των ηλεκτρονικών υπολογιστών, η εφαρμογή και η ανάπτυξη της ανάλυσης επιβίωσης γίνεται ολοένα και μεγαλύτερη, καθώς παράγεται μεγάλος αριθμός λογισμικών πακέτων για το σκοπό αυτό.

## 1.2 Ορολογία και βασικές έννοιες

Στην παράγραφο αυτή, γίνεται αναφορά στην ορολογία και τις βασικές έννοιες που χρησιμοποιούνται στην ανάλυση επιβίωσης. Κάθε μία από αυτές αποτελεί ισοδύναμο τρόπο ορισμού της κατανομής μιας συνεχούς, μη αρνητικής τυχαίας μεταβλητής  $T$  η οποία παριστάνει χρόνο ζωής, κάτω από την υπόθεση ενός ομογενούς πληθυσμού.

### 1.2.1 Αθροιστική συνάρτηση κατανομής

Ως **αθροιστική συνάρτηση κατανομής** (α.σ.κ. ή cumulative distribution function), ορίζεται η πιθανότητα του ενδεχομένου η τυχαία μεταβλητή  $T$  να πάρει τιμές μικρότερες ή ίσες του  $t$ :

$$F(t) = P(T \leq t), \quad t \geq 0.$$

Κατά συνέπεια, εκφράζει την πιθανότητα αποτυχίας μέχρι τη χρονική στιγμή  $t$  ή διαφορετικά, το ποσοστό των πειραματικών μονάδων που αναμένεται να έχουν αποβιώσει μέχρι το χρόνο  $t$ . Η αθροιστική συνάρτηση κατανομής είναι αύξουσα και συνεχής συνάρτηση του  $t$ , με  $F(0) = 0$  και  $\lim_{t \rightarrow +\infty} F(t) = 1$ .

### 1.2.2 Συνάρτηση πυκνότητας πιθανότητας

Η **συνάρτηση πυκνότητας πιθανότητας** (σ.π.π. ή probability density function) ορίζεται από τη σχέση:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = F'(t), \quad t \geq 0. \quad (1.1)$$

Για μικρό  $\Delta t$ , η ποσότητα  $f(t)\Delta t$ , εκφράζει κατά προσέγγιση την πιθανότητα «θανάτου-αποτυχίας» στο διάστημα  $[t, t + \Delta t)$ . Η σ.π.π. κατανομών χρόνων ζωής, έχει τις παρακάτω ιδιότητες:

$$f(t) \geq 0, \quad \text{για κάθε } t \geq 0$$

και

$$\int_0^{+\infty} f(t) dt = 1.$$

### 1.2.3 Συνάρτηση επιβίωσης

Η **συνάρτηση επιβίωσης** (survivor ή reliability function), αποτελεί τη δυική συνάρτηση της αθροιστικής συνάρτησης κατανομής και ορίζεται ως η πιθανότητα του ενδεχομένου η τυχαία μεταβλητή  $T$  να πάρει τιμές μεγαλύτερες του  $t$ :

$$S(t) = P(T > t) = 1 - F(t), \quad t \geq 0.$$

Κατά συνέπεια, εκφράζει την πιθανότητα επιβίωσης μέχρι τη χρονική στιγμή  $t$  ή διαφορετικά, το ποσοστό των πειραματικών μονάδων που αναμένεται να έχουν επιβιώσει στο χρόνο  $t$ . Από τις ιδιότητες της αθροιστικής συνάρτησης κατανομής, έπεται ότι η  $S$  είναι φθίνουσα και συνεχής, με  $S(0) = 1$  και  $\lim_{t \rightarrow +\infty} S(t) = 0$ . Η σημασία της στην ανάλυση επιβίωσης είναι μεγάλη αφού μεταξύ άλλων, αφενός αποτελεί μέτρο σύγκρισης της βιωσιμότητας διαφορετικών πληθυσμών, αφετέρου διευκολύνει το στατιστικό χειρισμό δεξιά λογοκριμένων δεδομένων (right censored data).

### 1.2.4 Συνάρτηση κινδύνου

Η **συνάρτηση κινδύνου** (hazard function ή failure rate ή force mortality), εκφράζει το στιγμιαίο ρυθμό «θανάτου» ή «αποτυχίας» τη χρονική στιγμή  $t$ , δοθείσης της επιβίωσης μέχρι το χρόνο  $t$  και δίνεται από τη σχέση:

$$h(t) = \lim_{\Delta t \rightarrow +\infty} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad t \geq 0. \quad (1.2)$$

Για μικρό  $\Delta t$ , η ποσότητα  $h(t)\Delta t$ , εκφράζει κατά προσέγγιση την πιθανότητα θανάτου στο διάστημα  $[t, t + \Delta t)$ , δοθείσης της επιβίωσης μέχρι το χρόνο  $t$ . Η συνάρτηση κινδύνου για συνεχή κατανομή χρόνου ζωής, έχει τις παρακάτω ιδιότητες:

$$h(t) \geq 0, \quad \text{για κάθε } t \geq 0$$

και

$$\int_0^{+\infty} h(t) dt \rightarrow +\infty.$$

Αν και κάθε μία από τις αναφερόμενες συναρτήσεις αποτελεί μια ξεχωριστή ερμηνεία της επιβίωσης, αδιαμφισβήτητα, η συνάρτηση κινδύνου είναι το πιο αντιπροσωπευτικό μέτρο περιγραφής της ικανότητας επιβίωσης ενός πληθυσμού. Δεδομένου ότι περιγράφει τον τρόπο με τον οποίο μεταβάλλεται ο στιγμιαίος ρυθμός «θανάτου-αποτυχίας» στο χρόνο, η εκ των προτέρων πληροφορία για τη μονοτονία της, βοηθά στην επιλογή κατάλληλου μοντέλου για την περιγραφή χρόνων ζωής.

Με εξαίρεση τις περιπτώσεις που η υπόθεση της σταθερής συνάρτησης κινδύνου είναι ρεαλιστική, κατά κανόνα, η επιβίωση περιγράφεται με κατανομές των οποίων η αντίστοιχη συνάρτηση κινδύνου είναι αύξουσα, φθίνουσα, κοίλη ή μονοκόρυφη. Με βάση τα τέσσερα τελευταία είδη μονοτονίας, μια κατανομή ανάλυσης επιβίωσης χαρακτηρίζεται αντίστοιχα ως IFR (increasing failure rate), DFR (decreasing failure rate), BTFR (bathtub shaped failure rate) και UBTFR (upside down bathtub shaped failure rate). Οι μαθηματικοί ορισμοί των παραπάνω ιδιοτήτων, δίνονται στη συνέχεια (Glaser, 1980).

### Ορισμός

Μια κατανομή ανάλυσης επιβίωσης, χαρακτηρίζεται ως:

- IFR (DFR)  $\Leftrightarrow h'(t) > 0$  ( $h'(t) < 0$ ) για κάθε  $t > 0$ .
- BTFR (UBTFR)  $\Leftrightarrow$  υπάρχει  $t_0 > 0$ , τέτοιο ώστε:
  - $h'(t) < 0$  ( $h'(t) > 0$ ) για κάθε  $0 < t < t_0$ ,
  - $h'(t) > 0$  ( $h'(t) < 0$ ) για κάθε  $t > t_0$  και
  - $h'(t_0) = 0$ .

Συχνά, το ενδιαφέρον επικεντρώνεται σε μια περίοδο της συνολικής διάρκειας ζωής, στην οποία επέρχεται γήρανση ή φθορά χρήσης. Το

γεγονός αυτό, δικαιολογεί εν μέρει, τη μεγάλη δημοτικότητα των κατανομών χρόνων ζωής με αύξουσα συνάρτηση κινδύνου.

Αντίστοιχα, κατανομές με φθίνουσα συνάρτηση κινδύνου χρησιμοποιούνται ως επί το πλείστον για την περιγραφή της διάρκειας ζωής ηλεκτρονικών ή μηχανικών εξαρτημάτων, τα οποία παρουσιάζουν βελτίωση της λειτουργίας τους με την πάροδο του χρόνου. Παράλληλα έχουν ευρεία εφαρμογή σε κλάδους όπως αυτός της Πληροφορικής. Σε πληροφοριακά συστήματα και εφαρμογές λογισμικού είναι σύνηθες το φαινόμενο να υπάρξει κατάρρευση (crash) κατά τις αρχικές εκτελέσεις του προγράμματος λόγω πιθανών προγραμματιστικών σφαλμάτων. Ωστόσο, αφού αυτά διορθωθούν και με την πάροδο του χρόνου, κρίνεται ως μηδενική η πιθανότητα μελλοντικής δυσλειτουργίας του προγράμματος. Εφαρμογές, με φθίνουσα συνάρτηση κινδύνου εντοπίζονται ακόμη σε ιατρικά δεδομένα ίασης συγκεκριμένων ασθενειών.

Επιπρόσθετα, όταν μελετάται η συνολική διάρκεια επιβίωσης, τα πλέον ρεαλιστικά μοντέλα για τη περιγραφή χρόνων ζωής, είναι εκείνα με συνάρτηση κινδύνου κυπελλοειδούς μορφής. Για παράδειγμα, σε ότι αφορά στη διάρκεια ζωής έμψυχων οντοτήτων, ο ρυθμός θανάτου τους αρχικά είναι σχετικά υψηλός (περίοδος νεογνικής θνησιμότητας), φθίνει προοδευτικά μέχρι μια συγκεκριμένη τιμή όπου προσωρινά διατηρείται σχεδόν σταθερός (περίοδος νεότητας) και στη συνέχεια αυξάνει και πάλι λόγω γήρατος (περίοδος γηρατειών). Ανάλογη συμπεριφορά παρουσιάζει και η συνάρτηση κινδύνου πολλών εργοστασιακών προϊόντων (π.χ. ηλεκτρονικές συσκευές), με μεγάλη συχνότητα δυσλειτουργικών προβλημάτων με την έναρξη λειτουργίας τους (λόγω αστοχίας υλικού) ή μετά από χρήση (λόγω φθοράς) και σχετικά ομαλή λειτουργία στο μεσοδιάστημα. Βεβαίως, όταν διενεργείται από τον κατασκευαστή έλεγχος της ποιότητας της λειτουργίας τους, με σκοπό την απομάκρυνση των ελαττωματικών πριν την προώθησή τους στην αγορά, ο χρόνος ζωής των υπολοίπων, όπως καταγράφεται από τον καταναλωτή, αναμένεται να περιγράφεται από αύξουσα συνάρτηση κινδύνου.

Τα μοντέλα με μονοκόρυφη συνάρτηση κινδύνου χρησιμοποιούνται κυρίως στην περίπτωση που παρατηρείται μεγάλη συχνότητα αποτυχιών κοντά στη χρονική αφετηρία της διάρκειας ζωής του υπό μελέτη αντικειμένου. Εφαρμογές της συγκεκριμένης μορφής συνάρτησης κινδύνου εντοπίζονται τόσο σε ιατρικά δεδομένα, όσο και σε δεδομένα τηλεπικοινωνιών και φυσικών καταστροφών.

Τέλος, η εκθετική κατανομή με σταθερή συνάρτηση κινδύνου, παρουσιάζει μεγάλο εύρος εφαρμογών στη μοντελοποίηση χρόνων ζωής, λόγω της ευκολίας μελέτης της και των ελκυστικών της ιδιοτήτων.

Περιπτώσεις πληθυσμών με διαφορετικό, σε σχέση με τα προαναφερόμενα, είδος μονοτονίας του στιγμιαίου ρυθμού «αποτυχίας», όπως για παράδειγμα τροποποιημένη κυρτή ή κοίλη συνάρτηση κινδύνου (roller-coaster hazard function), αποτελούν μειοψηφία και αντιμετωπίζονται μεμονωμένα με συγκεκριμένες κατανομές χρόνων ζωής (βλέπε Wong, 1988, 1989 και 1991 καθώς και αναφορές εκεί).

### 1.2.6 Αντίστροφη συνάρτηση κινδύνου

Παράλληλα με τη συνάρτηση κινδύνου και η αντίστροφη συνάρτηση αυτής, βρίσκει τα τελευταία χρόνια πολλές εφαρμογές. Η **αντίστροφη συνάρτηση κινδύνου** (reverse hazard function), εκφράζει τη δεσμευμένη πιθανότητα «θανάτου» ή «αποτυχίας» μέσα στο χρονικό διάστημα  $(t - \Delta t, t]$ , δοθείσης της αποτυχίας μέχρι το χρόνο  $t$  και δίνεται από τη σχέση:

$$r(t) = \lim_{\Delta t \rightarrow +\infty} \frac{P(t - \Delta t \leq T \leq t | T \leq t)}{\Delta t} = \frac{f(t)}{F(t)}, \quad t \geq 0. \quad (1.3)$$

Οι Keilson and Sumita (1982) ήταν ανάμεσα στους πρώτους ερευνητές που όρισαν την αντίστροφη συνάρτηση κινδύνου. Επιπλέον μελέτη των ιδιοτήτων της επετεύχθη από τους Andersen et al. (1993), Shaked and Shanthikumar (1994), Ruiz and Navarro (1994), Block et al. (1998), Chandra and Roy (2001), Finkelstein (2002) και Gupta et al. (2004). Όλοι οι παραπάνω μελετητές απέδειξαν ότι η αντίστροφη συνάρτηση

κινδύνου αποτελεί ένα σημαντικό «εργαλείο» των μοντέλων ανάλυσης επιβίωσης. Μπορεί να χρησιμοποιηθεί για την εκτίμηση της συνάρτησης κινδύνου σε περιπτώσεις αριστερά λογοκριμένων δεδομένων, δύναται να έχει άμεση εφαρμογή σε  $k$  από  $n$  συστήματα και στον υπολογισμό του χρόνου αναμονής (χρόνος που μεσολαβεί από την αποτυχία μέχρι την καταγραφή της), ενώ συνδέεται άμεσα και με το μέσο χρόνο αναμονής (mean waiting time - χρόνος που καταγράφεται μέχρι την αποτυχία), ο οποίος μπορεί να χρησιμοποιηθεί ως εναλλακτικός τρόπος ορισμού συναρτήσεων επιβίωσης.

### 1.2.6 Αθροιστική συνάρτηση κινδύνου

Από τις σχέσεις (1.1) και (1.2), προκύπτει ότι:

$$S(t) = e^{-H(t)},$$

όπου

$$H(t) = \int_0^t h(x) dx,$$

είναι η **αθροιστική συνάρτηση κινδύνου** (cumulative hazard ή integrated hazard function), η οποία είναι μη φθίνουσα συνάρτηση του  $t$ , με  $H(0) = 0$  και  $\lim_{t \rightarrow +\infty} H(t) = +\infty$ .

### 1.2.7 Μέσος υπολειπόμενος χρόνος ζωής

Ένα μέτρο βιωσιμότητας (ή αξιοπιστίας) στη μελέτη της διάρκειας ζωής βιολογικών οργανισμών (ή εργοστασιακών προϊόντων), είναι ο **μέσος υπολειπόμενος χρόνος ζωής** (mean residual lifetime). Η τιμή του τη χρονική στιγμή  $t$ , δίνει την τιμή της αναμενόμενης διάρκειας ζωής από τη στιγμή  $t$  μέχρι την «αποτυχία», δοθείσης της επιβίωσης μέχρι το χρόνο  $t$  και δίνεται από τη σχέση:

$$m(t) = E(T - t | T \geq t) = \frac{\int_t^{+\infty} S(t) dt}{S(t)}, t \geq 0.$$

Για κατανομές ανάλυσης επιβίωσης με πεπερασμένη μέση τιμή, η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής, έχει τις ακόλουθες ιδιότητες:

$$m(t) \geq 0, m'(t) \geq -1, \text{ για κάθε } t \geq 0$$

και

$$\int_0^{+\infty} \frac{1}{m(t)} dt = +\infty.$$

Στον Πίνακα 1.1, παρουσιάζεται ο τρόπος προσδιορισμού κάθε μιας από τις συναρτήσεις που προαναφέρθηκαν, αν μια από τις υπόλοιπες είναι γνωστή.



Πίνακας 1.1 Συσχετίσεις βασικών συναρτήσεων

$f(t)$	$F(t)$	$S(t)$	$h(t)$	$r(t)$	$H(t)$	$m(t)$
$f(t)$	$\int_0^t f(t) dt$	$\int_t^{+\infty} f(t) dt$	$\frac{f(t)}{\int_t^{+\infty} f(t) dt}$	$\frac{f(t)}{\int_0^t f(t) dt}$	$-\log \left\{ \int_t^{+\infty} f(t) dt \right\}$	$\frac{\int_t^{+\infty} (t-t)f(t) dt}{\int_t^{+\infty} f(t) dt}$
$F(t)$	-	$I - F(t)$	$-\left[ \log \{ I - F(t) \} \right]'$	$-\left[ \log \{ F(t) \} \right]'$	$-\log \{ I - F(t) \}$	$\frac{\int_t^{+\infty} \{ I - F(t) \} dt}{I - F(t)}$
$S(t)$	$I - S(t)$	-	$-\left[ \log \{ S(t) \} \right]'$	$-\left[ \log \{ I - S(t) \} \right]'$	$-\log \{ S(t) \}$	$\frac{\int_t^{+\infty} S(t) dt}{S(t)}$
$h(t)$	$h(t)e^{-\int_0^t h(t) dt}$	$e^{-\int_0^t h(t) dt}$	-	$\frac{h(t)e^{-\int_0^t h(t) dt}}{I - e^{-\int_0^t h(t) dt}}$	$\int_0^t h(t) dt$	$\frac{\int_t^{+\infty} e^{-\int_0^t h(t) dt} dt}{e^{-\int_0^t h(t) dt}}$
$r(t)$	$r(t)e^{-\int_0^t r(t) dt}$	$I - e^{-\int_0^t r(t) dt}$	$\frac{r(t)e^{-\int_0^t r(t) dt}}{I - e^{-\int_0^t r(t) dt}}$	-	$\int_0^t \frac{r(t)e^{-\int_0^t r(t) dt}}{I - e^{-\int_0^t r(t) dt}} dt$	$\frac{\int_t^{+\infty} \left\{ I - e^{-\int_0^t r(t) dt} \right\} dt}{I - e^{-\int_0^t r(t) dt}}$
$H(t)$	$H'(t)e^{-H(t)}$	$e^{-H(t)}$	$H'(t)$	$\frac{H'(t)e^{-H(t)}}{I - e^{-H(t)}}$	-	$e^{H(t)} \int_t^{+\infty} e^{-H(t)} dt$
$m(t)$	$\frac{I + m'(t)}{m(t)} e^{-\int_0^t \frac{I + m(t)}{m(t)} dt}$	$e^{-\int_0^t \frac{I + m(t)}{m(t)} dt}$	$\frac{I + m'(t)}{m(t)}$	$\frac{I + m'(t)}{m(t)} \times \frac{e^{-\int_0^t \frac{I + m(t)}{m(t)} dt}}{I - e^{-\int_0^t \frac{I + m(t)}{m(t)} dt}}$	$\int_0^t \frac{I + m'(t)}{m(t)} dt$	-

### 1.2.8 Λοιπές συναρτήσεις

Αξίζει να σημειωθεί ότι υπάρχουν και άλλοι ισοδύναμοι τρόποι ορισμού της κατανομής  $T$ , όπως για παράδειγμα, αυτός του αναμενόμενου χρόνου αδράνειας (expected inactivity time, Ghitany et al., 2005) που εμφανίζεται σχετικά πρόσφατα στη βιβλιογραφία ή οι προγενέστεροί του, της ροπογεννήτριας συνάρτησης  $E(e^{sT})$  (moment generating function) και της χαρακτηριστικής συνάρτησης  $E(e^{isT})$  (characteristic function). Η επιλογή αυτών που παρουσιάστηκαν, έγινε στη βάση της δημοτικότητας τους στην ανάλυση επιβίωσης, λόγω της απτής φυσικής τους ερμηνείας.

Οι ροπές και τα εκατοστιαία σημεία δεν προσδιορίζουν μονοσήμαντα την κατανομή του χρόνου ζωής, αλλά συχνά αποτελούν ένα χρήσιμο τρόπο συνοπτικής περιγραφής της. Οι ροπές  $r$  τάξης περί το μηδέν, προσδιορίζονται από τη σχέση:

$$E(T^r) = \int_0^{+\infty} t^r f(t) dt.$$

Η μέση τιμή  $E(T)$ , αποτελεί ένα μέτρο της κεντρικής τάσης των τιμών της  $T$ , ενώ η διακύμανση

$$Var(T) = E(T^2) - \{E(T)\}^2,$$

ένα μέτρο της διασποράς των τιμών της, γύρω από τη μέση τιμή.

Τα  $p$  εκατοστιαία σημεία, δίνονται από τη σχέση:

$$F(t_p) = p \quad \text{ή} \quad t_p = F^{-1}(p)$$

και προσδιορίζουν το χρόνο μέχρι τον οποίο δυνητικά επιβιώνει ένα προκαθορισμένο ποσοστό  $(1-p)*100\%$  επί του συνόλου των πειραματικών μονάδων, ή διαφορετικά το χρόνο μέχρι τον οποίο η πιθανότητα επιβίωσης ισούται με  $1-p$ .

Τέλος, η κατά Rényi εντροπία (Rényi, 1961), που αποτελεί ένα μέτρο της αβέβαιης μεταβλητότητας μιας τυχαίας μεταβλητής  $T$ , μπορεί να υπολογιστεί από τη σχέση:

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left\{ \int_{-\infty}^{+\infty} f^\gamma(t) dt \right\},$$

όπου  $\gamma > 0$  και  $\gamma \neq 1$ . Παράλληλα, η κατά Shannon εντροπία (Shannon, 1948), η οποία αποτελεί ειδική περίπτωση της προηγούμενης, υπολογίζεται από τη σχέση:

$$\lim_{\gamma \rightarrow 1} I_R(\gamma) = E[-\log\{f(T)\}].$$

## 1.3 Κατανομές ανάλυσης επιβίωσης

Όπως έχει ήδη αναφερθεί, οι κατανομές πιθανότητας βρίσκουν άμεση εφαρμογή στην ανάλυση επιβίωσης εξυπηρετώντας τις ανάγκες του ερευνητικού αυτού πεδίου, χάριν της καλής προσαρμογής τους σε δεδομένα χρόνων ζωής. Για το λόγο αυτό, καθ' όλη τη διάρκεια εξέλιξης του κλάδου της ανάλυσης επιβίωσης, πολλοί ερευνητές ανέπτυξαν διάφορα μοντέλα με σκοπό τη βέλτιστη περιγραφή πραγματικών καταστάσεων. Στην παράγραφο αυτή, γίνεται αναφορά στους τρόπους με τους οποίους δύναται να παραχθούν μοντέλα χρόνων ζωής, αναφέρεται αναλυτικά η μεθοδολογία μελέτης της οποίας αυτά τυγχάνουν και καταγράφονται ορισμένες από τις βασικότερες κατανομές ανάλυσης επιβίωσης.

### 1.3.1 Τρόποι παραγωγής

Γενικότερα στη βιβλιογραφία έχουν προταθεί τρεις διαφορετικοί τρόποι παραγωγής κατανομών χρόνου ζωής. Ο πρώτος, αφορά τον ορισμό μιας κατανομής από μηδενική βάση, έχοντας ως γνώμονα την ικανοποίηση των επιθυμητών προϋποθέσεων και ιδιοτήτων που οφείλει αυτή να πληροί. Εναλλακτικά, μια νέα κατανομή δύναται να προκύψει με την κατάλληλη προσθήκη αριθμού παραμέτρων σε ήδη υπάρχουσα

κατανομή. Ως επί το πλείστον, ο τρόπος αυτός υστερεί έναντι του πρώτου όσον αφορά το εύρος των δυνατοτήτων που δυνητικά έχει η νέα κατανομή, μιας και εφαρμόζεται σε δεδομένη αρχική κατανομή, υπερτερεί ωστόσο, σε ευκολία υλοποίησης. Τέλος, ένας από τους συνηθέστερα χρησιμοποιούμενους τρόπους παραγωγής νέων μοντέλων ανάλυσης επιβίωσης είναι η μίξη υπαρχουσών κατανομών. Δεδομένου του πλήθους των κατανομών που έχουν ήδη προταθεί στη βιβλιογραφία, η μεθοδολογία αυτή αποτελεί ένα τρόπο διεύρυνσης των δυνατοτήτων προσαρμογής των μοντέλων σε μετρήσεις που απορρέουν από σύνθετες πραγματικές καταστάσεις.

Βέβαια, όλοι οι προαναφερθέντες τρόποι διέπονται από τον ακόλουθο κανόνα:

«Ο αριθμός παραμέτρων ενός μοντέλου είναι ανάλογος με την ευελιξία του και αντιστρόφως ανάλογος με την ευκολία μελέτης του»

### *1.3.2 Αναλυτική μεθοδολογία μελέτης*

Η μεθοδολογία μελέτης ενός μοντέλου ανάλυσης επιβίωσης θα μπορούσε να ταξινομηθεί σε πέντε στάδια.

Το **πρώτο στάδιο** περιλαμβάνει τη συλλογή των δεδομένων επιβίωσης. Τα δεδομένα μπορούν να είναι είτε εργαστηριακά, είτε πραγματικά. Τα εργαστηριακά δεδομένα λαμβάνονται συχνά κάτω από ελεγχόμενο περιβάλλον και βασίζονται σε ένα σωστά σχεδιασμένο πείραμα. Σε αντίθεση, τα πραγματικά δεδομένα πάσχουν από μεταβλητότητα στο περιβάλλον λειτουργίας, καθώς και από άλλους ανεξέλεγκτους παράγοντες.

Η μορφή των δεδομένων μπορεί να ποικίλει. Στην περίπτωση των δεδομένων επιβίωσης, θα μπορούσε να είναι συνεχής (π.χ., η διάρκεια ζωής ενός μεμονωμένου αντικειμένου) ή διακριτή (π.χ., ο αριθμός των αντικειμένων που αποτυγχάνουν σε ένα συγκεκριμένο χρονικό διάστημα). Επιπλέον, τα δεδομένα επιβίωσης δύναται να ταξινομηθούν ως πλήρη και λογοκριμένα, με τη δεύτερη κατηγορία να επιδέχεται

υποδιαίρεσης σε περισσότερες υποκατηγορίες. Παρακάτω, αναλύονται οι δύο κατηγορίες δεδομένων.

Έστω  $T_1, T_2, \dots, T_n$  ένα ανεξάρτητο τυχαίο δείγμα  $n$  τυχαίων μεταβλητών από την κατανομή  $f(t, \theta)$ , όπου  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$  είναι μια  $k$ -διάστατη παράμετρος και  $t_1, t_2, \dots, t_n$  οι τιμές των  $T_i$ .

### **Πλήρη δεδομένα**

Το διαθέσιμο προς εκτίμηση σύνολο δεδομένων είναι το  $\{t_1, t_2, \dots, t_n\}$ . Με άλλα λόγια, οι ακριβείς χρόνοι επιβίωσης για κάθε παρατήρηση είναι γνωστοί.

### **Λογοκριμένα δεδομένα**

Σε αυτήν την περίπτωση, οι ακριβείς χρόνοι επιβίωσης για ορισμένες ή και όλες τις παρατηρήσεις δεν είναι γνωστοί. Οι διάφοροι τύποι λογοκρισίας που εμφανίζονται στην ανάλυση επιβίωσης είναι οι εξής:

#### *Δεξιά λογοκρισία*

Μια παρατήρηση  $t_i, i = 1, 2, \dots, n$  λέγεται δεξιά λογοκριμένη όταν το γεγονός συμβαίνει μετά το χρονικό διάστημα συλλογής του δείγματος. Για παράδειγμα, ένα από τα υπό μελέτη ηλεκτρονικά εξαρτήματα δεν αποτυγχάνει μέσα στο προκαθορισμένο χρονικό διάστημα ενός πειράματος, αλλά συνεχίζει να λειτουργεί και μετά τη λήξη. Οι περιπτώσεις δεξιάς λογοκρισίας είναι πολύ συχνά εμφανιζόμενες στις μελέτες ανάλυσης επιβίωσης.

#### *Αριστερή λογοκρισία*

Μια παρατήρηση  $t_i$  λέγεται αριστερά λογοκριμένη όταν το γεγονός συμβαίνει πριν το χρονικό διάστημα συλλογής του δείγματος. Για παράδειγμα, σε έναν ασθενή γνωρίζουμε τη χρονική στιγμή της εξέτασης που πιστοποίησε ότι νοσεί, ωστόσο δε γνωρίζουμε τη χρονική στιγμή κατά την οποία ο ασθενής μολύνθηκε.

*Λογοκρισία τύπου I (II)*

Η λογοκρισία τύπου I (II) πολλές φορές συνδυάζεται με τη δεξιά ή αριστερή λογοκρισία. Εμφανίζεται για μια παρατήρηση  $t_i$ , όταν αυτή βρίσκεται εκτός ενός προκαθορισμένου χρονικού διαστήματος (αριθμού αποτυχιών), όπου έχει θέσει ο ερευνητής για τη συλλογή του δείγματος.

*Διαστηματική λογοκρισία*

Μια παρατήρηση  $t_i$  λέγεται ότι εμφανίζει διαστηματική λογοκρισία όταν ανήκει εντός ενός διαστήματος, ωστόσο δε γνωρίζουμε την ακριβή στιγμή της αποτυχίας. Για παράδειγμα, σε έναν ασθενή γνωρίζουμε τη χρονική στιγμή κάθε επαναλαμβανόμενης εξέτασης μέχρι και εκείνη που πιστοποιεί ότι νοσεί, ωστόσο δε γνωρίζουμε την ακριβή χρονική στιγμή κατά την οποία ο ασθενής μολύνθηκε, παρά μόνον ότι αυτή βρίσκεται στο μεσοδιάστημα μεταξύ των δύο τελευταίων εξετάσεων.

*Τυχαία λογοκρισία*

Έστω τυχαία μεταβλητή  $S_i$  ανεξάρτητη της  $T_i$ . Η παρατηρούμενη τιμή του δείγματος θα δίνεται από την ποσότητα  $\min\{t_i, s_i\}$ , έτσι ώστε η παρατήρηση να είναι λογοκριμένη εάν  $t_i > s_i$ .

Περισσότερες πληροφορίες σχετικά με τις διάφορες μορφές λογοκρισίας, μπορούν να αναζητηθούν, μεταξύ άλλων στους Lawless (1982), Nelson (1982) και Leemis (1995).

Το **δεύτερο στάδιο** περιλαμβάνει την προκαταρκτική ανάλυση των συλλεγόμενων δεδομένων. Αυτή περιλαμβάνει τον υπολογισμό διαφόρων στατιστικών μέτρων όπως η μέγιστη και η ελάχιστη τιμή, η μέση τιμή, η διακύμανση, η διάμεσος και τα τεταρτημόρια. Εάν το εύρος είναι μικρό σε σχέση με τη μέση τιμή του δείγματος, θα μπορούσε κανείς να αγνοήσει τη μεταβλητότητα στα δεδομένα και να μοντελοποιήσει τα δεδομένα με χρήση της μέσης τιμής. Ωστόσο, όταν αυτό δεν είναι εφικτό, το επιλεγόμενο μοντέλο πρέπει να ικανοποιεί αυτή τη μεταβλητότητα των δεδομένων. Επιπρόσθετα, στην περίπτωση χρονικά

διατεταγμένων δεδομένων, η προκαταρκτική ανάλυση χρησιμοποιείται για να προσδιοριστούν οι ιδιότητες, όπως η αυξητική (ή φθίνουσα) τάση, η αυτοσυσχέτιση, κ.ά..

Ο κύριος σκοπός της ανάλυσης αυτής είναι να βοηθήσει στον καθορισμό εάν ένα συγκεκριμένο μοντέλο είναι ή όχι κατάλληλο να περιγράψει τα συγκεκριμένα δεδομένα. Διάφορα γραφήματα έχουν αναπτυχθεί για να βοηθήσουν σε αυτό. Μερικά από αυτά τα γραφήματα (π.χ. ιστόγραμμα) είναι γενικά, ενώ άλλα (π.χ., γράφημα πιθανοτήτων Weibull) αναπτύχθηκαν αρχικά για συγκεκριμένα μοντέλα, αλλά έκτοτε έχουν χρησιμοποιηθεί σε μία ευρύτερη κατηγορία μοντέλων.

Το **τρίτο στάδιο** περιλαμβάνει την επιλογή κατάλληλου μοντέλου (κατανομής πιθανότητας) για την περιγραφή του συλλεγόμενου συνόλου δεδομένων. Για να επιτευχθεί αυτό το βήμα, ο ερευνητής οφείλει να έχει καλή γνώση των ιδιοτήτων των διαφορετικών μοντέλων που κρίνονται ως κατάλληλα για προσαρμογή. Η θεωρία πιθανοτήτων πραγματεύεται με τέτοιου είδους μελέτη για μια ποικιλία προσαρμοζόμενων μοντέλων. Ένα σημαντικό χαρακτηριστικό της προσαρμογής δεδομένων είναι ότι συχνά υπάρχουν περισσότερες από μία κατανομές που περιγράφουν επαρκώς τα δεδομένα. Κατά αυτόν τον τρόπο, ο ερευνητής έχει τη δυνατότητα της επιλογής εκείνου που ικανοποιεί περισσότερο τα δικά του κριτήρια.

Η πηγή των δεδομένων παρέχει συχνά μια ένδειξη για την επιλογή του κατάλληλου μοντέλου. Για παράδειγμα, η λογαριθμική-κανονική κατανομή, ή η κατανομή Weibull, χρησιμοποιούνται συχνά σε δεδομένα αποτυχίας λόγω κόπωσης (αύξουσα συνάρτηση κινδύνου), ενώ η εκθετική κατανομή βρίσκει αρκετές φορές εφαρμογή σε δεδομένα αποτυχίας ηλεκτρονικών εξαρτημάτων. Για να χρησιμοποιηθεί ωστόσο αυτή η γνώση, οφείλει ο ερευνητής να είναι εξοικειωμένος με πλήθος κατανομών ανάλυσης επιβίωσης που έχουν προταθεί στη βιβλιογραφία για τη μελέτη διαφορετικών καταστάσεων.

Εάν τα δεδομένα δεν είναι ανεξάρτητα, θα πρέπει να γίνει χρήση μοντέλων που περιλαμβάνουν πολυπαραγοντική συνάρτηση κατανομής. Εάν ο χρόνος είναι ένας παράγοντας που χρειάζεται να συμπεριληφθεί στο μοντέλο, τότε αυτό γίνεται ακόμη πιο σύνθετο. Η κατασκευή τέτοιων μοντέλων απαιτεί χρήση στοχαστικών διαδικασιών και δε θα αναλυθεί περαιτέρω στην παρούσα διατριβή.

Το **τέταρτο στάδιο** περιλαμβάνει την εκτίμηση των παραμέτρων. Εφόσον ένα μοντέλο επιλεγεί, χρειάζεται να εκτιμηθούν οι παράμετροί του χρησιμοποιώντας τα διαθέσιμα δεδομένα. Για αυτήν τη διαδικασία έχουν αναπτυχθεί μία ποικιλία από τεχνικές, οι οποίες μπορούν να ταξινομηθούν σε δύο βασικές κατηγορίες - γραφικές και αναλυτικές. Η ακρίβεια της εκτίμησης εξαρτάται από το μέγεθος των δεδομένων και τη μέθοδο που χρησιμοποιείται. Οι γραφικές μέθοδοι αποδίδουν προσεγγιστικές εκτιμήσεις, ενώ οι αναλυτικές μέθοδοι αποδίδουν καλύτερες εκτιμήσεις και όρια εμπιστοσύνης για τους εκτιμητές. Συχνά για τεκμηρίωση των αποτελεσμάτων εφαρμόζονται συνδυαστικά και οι δύο μέθοδοι εκτίμησης. Γνωστές γραφικές μέθοδοι αποτελούν τα γραφήματα hazard plot και WPP plot, ενώ από τις στατιστικές ξεχωρίζουν οι μέθοδοι μέγιστης πιθανοφάνειας, ροπών, Bayes, εκατοστιαίων σημείων και διαστηματική. Επιπρόσθετα, η εκτίμηση των παραμέτρων του μοντέλου μπορεί να επιτευχθεί σε σημείο ή σε διάστημα (εμπιστοσύνης), ενώ οι εκτιμητές δύναται να ελεγχθούν για την ύπαρξη διαφορών ιδιοτήτων, όπως αμεροληψία, συνέπεια και αποτελεσματικότητα.

Το **πέμπτο στάδιο** περιλαμβάνει την προσαρμογή του μοντέλου στα συλλεγόμενα δεδομένα. Οποιοδήποτε μοντέλο μπορεί να προσαρμοστεί σε ένα συγκεκριμένο σύνολο δεδομένων. Ωστόσο, το μοντέλο μπορεί να μην είναι κατάλληλο ή επαρκές. Ένα ακατάλληλο μοντέλο, σε γενικές γραμμές, δε θα δώσει την επιθυμητή λύση στο πρόβλημα. Ως εκ τούτου, είναι απαραίτητο να ελέγχεται η εγκυρότητα του μοντέλου που επιλέχθηκε. Υπάρχουν αρκετές μέθοδοι για να γίνει αυτό, οι οποίες μπορούν να ταξινομηθούν σε δύο βασικές κατηγορίες - γραφικές και



στατιστικές. Οι συνηθέστερα χρησιμοποιούμενες γραφικές μέθοδοι είναι το γράφημα εμπειρικής συνάρτησης κατανομής (EDF plot), το γράφημα εμπειρικής συνάρτησης κινδύνου (hazard plot), το ιστόγραμμα και διάφορα μετασχηματισμένα γραφήματα (probability plot, Weibull plot, P-P plot, Q-Q plot). Αντίστοιχα, στατιστικές μέθοδοι που έχουν μεγάλη εφαρμογή είναι τα τεστ των Kolmogorov-Smirnov, των Anderson-Darling, των Cramer-von Mises, του Pearson ( $\chi^2$  τεστ) και του πηλίκου πιθανοφανειών (για σύγκριση ένθετων κατανομών).

### *1.3.3 Βασικές κατανομές ανάλυσης επιβίωσης*

Στην παράγραφο αυτή, παρουσιάζονται ορισμένες βασικές κατανομές πιθανότητας που βρίσκουν εφαρμογή στην ανάλυση επιβίωσης. Εκτενής μελέτη της εκθετικής κατανομής και των κατανομών Weibull, γάμμα και log-normal, υπάρχει στα κλασικά βιβλία θεωρίας κατανομών όπως των Johnson et al. (1994) και στα περισσότερα εγχειρίδια ανάλυσης επιβίωσης (π.χ. Nelson, 1982, Lee, 1992, Leemis, 1995 και Marshall and Olkin, 2007). Επιπρόσθετα, παρουσιάζονται οι κατανομές: γενικευμένη γάμμα (Stacy, 1962), Gompertz-Makeham (Jordan, 1967), γενικευμένη F (Kalbfleisch and Prentice, 1980), log-logistic (Tadikamalla, 1980), Lomax (Johnson et al., 1994), modified extreme value (Johnson et al., 1995) και εκθετική γεωμετρική (Adamidis and Loukas, 1998). Μεγάλο πλήθος κατανομών, εκτός από αυτές που προαναφέρθηκαν, περιγράφονται στους Johnson et al. (1995), Leemis (1995) και Marshall and Olkin (2007), ενώ οι Rajarshi and Rajarshi (1988) και οι Lai et al. (2001), παρουσιάζουν μια ανασκόπηση στα μοντέλα με κυπελλοειδούς μορφής συνάρτηση κινδύνου. Οι Kleiber and Kotz (2003), μελετούν κατανομές επιβίωσης που έχουν άμεση εφαρμογή στις οικονομικές και αναλογιστικές επιστήμες, οι Pham and Lai (2007), περιγράφουν ένα σύνολο κατανομών που αποτελούν γενικεύσεις της Weibull κατανομής, ενώ οι Chahkandi and Ganjali (2009) μελετούν και συγκρίνουν κατανομές φθίνουσας συνάρτησης κινδύνου.

### Η εκθετική κατανομή

Η εκθετική κατανομή με σ.π.π.

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

έχει συνάρτηση επιβίωσης  $S(t) = e^{-\lambda t}$  και σταθερή συνάρτηση κινδύνου  $h(t) = \lambda$ . Αποδεικνύεται ότι  $E(T) = 1/\lambda$  και  $Var(T) = 1/\lambda^2$ , άρα ο συντελεστής μεταβλητότητας ισούται με τη μονάδα.

Η εκθετική κατανομή σχετίζεται με την κατανομή extreme value. Ειδικότερα, η τ.μ.  $T$  ακολουθεί εκθετική κατανομή με παράμετρο  $\lambda$ ,  $T \sim E(\lambda)$ , εάν και μόνον εάν

$$Y = \log T = \alpha + W,$$

όπου  $\alpha = -\log \lambda$  και  $W$  η τυχαία μεταβλητή που ακολουθεί την τυπική extreme-value κατανομή, με σ.π.π.

$$f_w(w) = e^{w-e^w}, w > 0.$$

Η παραπάνω αποτελεί μονοκόρυφη συνάρτηση με  $E(W) = -\gamma$ , όπου  $\gamma = 0.5722$  είναι η σταθερά του Euler και  $Var(W) = \pi^2/6$ . Η λοξότητα είναι  $-1.14$ . Η απόδειξη της παραπάνω πρότασης προκύπτει άμεσα με αλλαγή μεταβλητών.

Ιστορικά, η εκθετική κατανομή αποτελεί το πρώτο ευρέως χρησιμοποιούμενο μοντέλο ανάλυσης επιβίωσης, αφενός λόγω της απλής συναρτησιακής της μορφής και της ευκολίας στο στατιστικό χειρισμό της, αφετέρου λόγω της καταλληλότητάς της στην περιγραφή χρόνων ζωής σε πληθώρα περιπτώσεων. Οι εφαρμογές της κυμαίνονται από μελέτες χρόνων ζωής εργοστασιακών προϊόντων, μέχρι έρευνες στους χρόνους ανάνηψης ή υποτροπής από χρόνιες ασθένειες. Βεβαίως, η

υπόθεση της σταθερής συνάρτησης κινδύνου δεν παύει να είναι αρκετά περιοριστική.

### Η κατανομή Weibull

Η κατανομή Weibull είναι ίσως η πιο δημοφιλής κατανομή ανάλυσης επιβίωσης, με ένα εξαιρετικά ευρύ φάσμα εφαρμογών τόσο στο χώρο της βιομηχανίας όσο και σε κλάδους της ιατρικής και των άλλων βιολογικών επιστημών. Η καταλληλότητά της στην περιγραφή χρόνων ζωής καθώς και οι απλές εκφράσεις των αντίστοιχων συναρτήσεων πυκνότητας πιθανότητας, επιβίωσης και κινδύνου, εξηγούν τη δημοτικότητά της.

Η τ.μ.  $T$  ακολουθεί την κατανομή Weibull με παραμέτρους  $\lambda$  και  $p$ ,  $T \sim W(\lambda, p)$ , εάν  $T^p \sim E(\lambda)$ . Η αθροιστική συνάρτηση κινδύνου της Weibull κατανομής είναι  $H(t) = (\lambda t)^p$ , η συνάρτηση επιβίωσης  $S(t) = e^{-(\lambda t)^p}$  και η συνάρτηση κινδύνου  $h(t) = \lambda p (\lambda t)^{p-1}$ . Ο λογάριθμος της συνάρτησης κινδύνου της Weibull κατανομής είναι γραμμική συνάρτηση του λογαρίθμου του χρόνου με σταθερά  $p \log \lambda + \log p$  και κλίση  $p-1$ . Επομένως, η συνάρτηση κινδύνου είναι αύξουσα εάν  $p > 1$ , σταθερή εάν  $p = 1$  και φθίνουσα εάν  $p < 1$ .

Η Weibull κατανομή συνδέεται με την κατανομή extreme value, ως εξής: η τ.μ.  $T$  ακολουθεί κατανομή Weibull με παραμέτρους  $\lambda, p$ ,  $T \sim W(\lambda, p)$ , εάν και μόνον εάν

$$Y = \log T = \alpha + \sigma W,$$

όπου  $\alpha = -\log \lambda$ ,  $\sigma = 1/p$  και  $W$  η τυχαία μεταβλητή που ακολουθεί την τυπική extreme value. Η απόδειξη προκύπτει με αλλαγή μεταβλητών, θεωρώντας την τ.μ.  $W$ , αλλάζοντας σε  $Y = \alpha + \sigma W$  και στη συνέχεια σε  $T = e^Y$ .

### Η κατανομή Gompertz-Makeham

Η κατανομή Gompertz-Makeham χαρακτηρίζεται από το γεγονός ότι ο λογάριθμος της συνάρτησης κινδύνου είναι γραμμική συνάρτηση του χρόνου  $t$ , δηλαδή  $h(t) = e^{\alpha + \beta t}$ . Επομένως, η συνάρτηση Gompertz είναι στενά συνδεδεμένη με την Weibull κατανομή και για την ακρίβεια η κατανομή Gompertz-Makeham είναι η log-Weibull κατανομή. Η συγκεκριμένη κατανομή παρέχει αξιοσημείωτα καλές προσαρμογές σε δεδομένα θνησιμότητας ενηλίκων σε σύγχρονα αναπτυγμένες χώρες.

### Η κατανομή γάμμα

Η κατανομή γάμμα με παραμέτρους  $\lambda$  και  $k$ ,  $G(\lambda, k)$ , έχει σ.π.π.

$$f(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)},$$

και συνάρτηση επιβίωσης

$$S(t) = 1 - \frac{\Gamma_{\lambda t}(k)}{\Gamma(k)},$$

όπου  $\Gamma_t(k)$  είναι η μη πλήρης συνάρτηση γάμμα, που ορίζεται ως:

$$\Gamma_t(k) = \int_0^t x^{k-1} e^{-x} dx.$$

Δεν υπάρχει κλειστή μορφή προσδιορισμού της συνάρτησης επιβίωσης, έχουν ωστόσο αναπτυχθεί υπολογιστικοί αλγόριθμοι για τον υπολογισμό αυτής (π.χ. στη γλώσσα R υπάρχει η συνάρτηση `rgamma` που υπολογίζει την α.σ.κ. και την συνάρτηση επιβίωσης της κατανομής γάμμα με παραμέτρους  $1/\lambda$  και  $k$ ).

Αντίστοιχα, και η συνάρτηση κινδύνου στερείται κλειστής μορφής, υπολογίζεται ωστόσο από τον τύπο ορισμού της  $h(t) = f(t)/S(t)$ . Όσον αφορά τη μονοτονία της:

- είναι γνησίως αύξουσα στο διάστημα  $[0, \lambda]$  για  $k > 1$ ,
- είναι σταθερή για  $k = 1$  και
- είναι γνησίως φθίνουσα στο διάστημα  $[+\infty, \lambda)$  για  $k < 1$ .

Εάν  $k = 1$  η κατανομή γάμμα ταυτίζεται με την εκθετική κατανομή, η οποία μπορεί να περιγράψει το χρόνο αναμονής μέχρι το πρώτο συμβάν σε μια διαδικασία Poisson. Σε περίπτωση που  $k$  ακέραιος με  $k > 1$ , τότε η κατανομή γάμμα καλείται και κατανομή Erlang και περιγράφει το χρόνο αναμονής μέχρι το  $k$ -οστό συμβάν σε μια διαδικασία Poisson.

Σύνδεση με την κατανομή extreme value υφίσταται και για την γάμμα κατανομή. Με αλλαγή μεταβλητών, μπορεί να αποδειχθεί ότι  $T \sim G(\lambda, k)$  εάν και μόνον εάν

$$\log T = Y = \alpha + W,$$

όπου η  $W$  ακολουθεί τη γενικευμένη extreme value κατανομή με σ.π.π.

$$f_w(w) = \frac{e^{kw - e^w}}{\Gamma(k)}.$$

### Η γενικευμένη κατανομή γάμμα

Η γενικευμένη κατανομή γάμμα, αποτελεί μια γενίκευση της γάμμα κατανομής και σχετίζεται με παρόμοιο τρόπο με την κατανομή extreme value με την προσθήκη μιας παραμέτρου κλίμακας. Δηλαδή, μια τ.μ.  $T$  ακολουθεί τη γενικευμένη γάμμα κατανομή,  $T \sim GG(\lambda, p, k)$ , εάν και μόνο εάν

$$Y = \log T = \alpha + \sigma W,$$

όπου η  $W$  ακολουθεί τη γενικευμένη extreme value κατανομή με παράμετρο  $k$ . Η σ.π.π. της γενικευμένης κατανομής γάμμα δίνεται από τη σχέση:

$$f(t) = \frac{\lambda p (\lambda t)^{pk-1} e^{-(\lambda t)^p}}{\Gamma(k)},$$

όπου  $p = 1/\sigma$ .

Η γενικευμένη κατανομή γάμμα αποτελεί επέκταση διαφόρων γνωστών κατανομών. Έτσι:

- για  $p = 1$ , ταυτίζεται με την κατανομή γάμμα,
- για  $k = 1$ , μετατρέπεται στη Weibull κατανομή,
- για  $p = k = 1$ , υποβιβάζεται στην εκθετική κατανομή, ενώ
- όταν  $k \rightarrow +\infty$ , συγκλίνει στην κατανομή log-normal.

### Η κατανομή log-normal

Η τ.μ.  $T$  ακολουθεί την κατανομή log-normal, εάν και μόνο εάν

$$Y = \log T = \alpha + \sigma W,$$

όπου η  $W$  ακολουθεί την τυπική κανονική κατανομή.

Η συνάρτηση κινδύνου της κατανομής log-normal είναι μονοκόρυφη, δηλαδή, αυξάνει από το 0 μέχρι να λάβει τη μέγιστη τιμή της και στη συνέχεια φθίνει μονότονα, προσεγγίζοντας το 0 για  $t \rightarrow +\infty$ .

Όπως προαναφέρθηκε, καθώς  $k \rightarrow +\infty$  η γενικευμένη extreme value κατανομή προσεγγίζει την τυπική κανονική και επομένως η γενικευμένη γάμμα προσεγγίζει την κατανομή log-normal.

### Η κατανομή log-logistic

Η τ.μ.  $T$  ακολουθεί την κατανομή log-logistic εάν και μόνον εάν

$$Y = \log T = \alpha + \sigma W,$$

όπου η  $W$  ακολουθεί την τυπική λογιστική (standard logistic) κατανομή με σ.π.π.

$$f_W(w) = \frac{e^w}{(1+e^w)^2}, w \in \mathbb{R},$$

α.σ.κ.

$$F_W(w) = \frac{e^w}{1+e^w},$$

και συνάρτηση επιβίωσης

$$S_W(w) = \frac{1}{1+e^w}.$$

Με αλλαγή μεταβλητών στην  $T$  έχουμε ότι η συνάρτηση επιβίωσης της κατανομής log-logistic δίνεται από τη σχέση:

$$S_T(t) = \frac{1}{1+(\lambda t)^p},$$

όπου  $\alpha = -\log \lambda$  και  $\sigma = 1/p$ . Λογαριθμίζοντας παίρνουμε την (αρνητική) αθροιστική συνάρτηση κινδύνου, ενώ παραγωγίζοντας ως προς  $t$ , λαμβάνουμε τον τύπο υπολογισμού της συνάρτησης κινδύνου

$$h(t) = \frac{\lambda p (\lambda t)^{p-1}}{1+(\lambda t)^p}.$$

Παρατηρείται ότι το logit της συνάρτησης επιβίωσης  $S(t)$ , δηλαδή η ποσότητα  $\log\left(\frac{S(t)}{F(t)}\right)$ , είναι γραμμική συνάρτηση του  $\log t$ . Το γνώρισμα αυτό παρέχει τη δυνατότητα ενός γραφικού διαγνωστικού ελέγχου: δεδομένης μιας μη-παραμετρικής εκτίμησης της συνάρτησης επιβίωσης, εάν το γράφημα του logit αυτής ως προς το λογάριθμο του χρόνου προσεγγίζει την ευθεία, τότε η συνάρτηση επιβίωσης ταυτίζεται με αυτή της κατανομής log-logistic.

Η συνάρτηση κινδύνου μπορεί να λάβει διάφορες μορφές ανάλογα με την τιμή των παραμέτρων της, και ειδικότερα:

- είναι γνησίως φθίνουσα από το  $+\infty$ , για  $p < 1$ ,
- είναι γνησίως φθίνουσα από το  $\lambda$ , για  $p = 1$ ,
- είναι αντίστοιχης μορφής με τη log-normal για  $p > 1$ .

### Η κατανομή Lomax

Η κατανομή Pareto τύπου II (γνωστή και ως κατανομή Lomax), βρίσκει κυρίως εφαρμογές στη μελέτη της βιωσιμότητας επιχειρήσεων.

Η σ.π.π. της, δίνεται από τη σχέση:

$$f(t) = \frac{a\lambda}{(1+\lambda t)^{a+1}},$$

όπου  $t \geq 0$ ,  $a, \lambda \in \mathbb{R}$ , με  $a$  και  $\lambda$  παραμέτρους μορφής και κλίμακας αντίστοιχα.

Επιπρόσθετα, οι συναρτήσεις επιβίωσης και κινδύνου της κατανομής, δίνονται από τις σχέσεις:

$$S(t) = \frac{1}{(1+\lambda t)^a}$$

και



$$h(t) = \frac{a\lambda}{1 + \lambda t}.$$

Αποδεικνύεται ότι, η συνάρτηση κινδύνου είναι φθίνουσα συνάρτηση του χρόνου με αρχική τιμή  $h(0) = a\lambda$  και  $\lim_{t \rightarrow +\infty} h(t) = 0$ .

### Η κατανομή modified extreme value

Η κατανομή modified extreme value, χρησιμοποιείται για την περιγραφή της επιβίωσης πληθυσμών που εμφανίζουν ραγδαίο ρυθμό φθοράς ως προς το χρόνο.

Η σ.π.π. της κατανομής, δίνεται από τη σχέση:

$$f(t) = a\lambda e^{\lambda t + a(1 - e^{\lambda t})},$$

όπου  $t \geq 0$ ,  $a, \lambda \in \mathbb{R}$ , με  $a$  και  $\lambda$  παραμέτρους μορφής και κλίμακας αντίστοιχα.

Επιπρόσθετα, οι συναρτήσεις επιβίωσης και κινδύνου, δίνονται από τις σχέσεις:

$$S(t) = e^{a(1 - e^{\lambda t})}$$

και

$$h(t) = a\lambda e^{\lambda t}.$$

Αποδεικνύεται ότι, η συνάρτηση κινδύνου είναι αύξουσα συνάρτηση του χρόνου με αρχική τιμή  $h(0) = a\lambda$  και  $\lim_{t \rightarrow +\infty} h(t) = +\infty$ .

### Η γενικευμένη F κατανομή

Μια ενδιαφέρουσα περίπτωση κατανομής αποτελεί η γενικευμένη F κατανομή και αυτό γιατί περιλαμβάνει όλες τις προαναφερθείσες κατανομές ως ειδικές ή οριακές περιπτώσεις και επομένως είναι χρήσιμη για τον έλεγχο διαφόρων παραμετρικών μορφών.

Η τ.μ.  $T$  ακολουθεί τη γενικευμένη F κατανομή εάν και μόνον εάν

$$Y = \log T = \alpha + \sigma W,$$

όπου  $W$  παριστάνει το λογάριθμο μιας τ.μ.  $X \sim F(\beta, \gamma)$ .

### Η εκθετική γεωμετρική κατανομή

Η εκθετική γεωμετρική κατανομή αποτελεί μια από τις πιο συχνά χρησιμοποιούμενες κατανομές φθίνουσας συνάρτησης κινδύνου, λόγω της καλής της προσαρμογής και της ευκολίας μελέτης της. Μεταξύ άλλων, έχει χρησιμοποιηθεί για τη μοντελοποίηση χρόνων μεταξύ διαδοχικών ατυχημάτων και σφαλμάτων, καθώς και χρόνων ζωής ηλεκτρικών και ηλεκτρονικών συσκευών.

Η σ.π.π. της κατανομής, δίνεται από τη σχέση:

$$f(t) = a\lambda e^{\lambda t + a(1 - e^{\lambda t})},$$

όπου  $t \geq 0$ ,  $a, \lambda \in \mathbb{R}$ , με  $a$  και  $\lambda$  παραμέτρους μορφής και κλίμακας αντίστοιχα.

Επιπρόσθετα, οι συναρτήσεις επιβίωσης και κινδύνου, δίνονται από τις σχέσεις:

$$S(t) = e^{a(1 - e^{\lambda t})}$$

και

$$h(t) = a\lambda e^{\lambda t}.$$

Αποδεικνύεται ότι, η συνάρτηση κινδύνου είναι αύξουσα συνάρτηση του χρόνου με αρχική τιμή  $h(0) = a\lambda$  και  $\lim_{t \rightarrow +\infty} h(t) = +\infty$ .

# Κεφάλαιο 2

## Η Ελεκτηταμένη Κατανομή του Chen

Οι κατανομές πιθανότητας αποτελούν το βασικό ερευνητικό εργαλείο στην ανάλυση επιβίωσης, λόγω της καλής τους προσαρμογής σε αντίστοιχα σύνολα δεδομένων, των πολλαπλών ιδιοτήτων τους και κυρίως της συνάρτησης κινδύνου τους. Η επιλογή του κατάλληλου παραμετρικού μοντέλου για την περιγραφή του χρόνου ζωής ενός πληθυσμού, βασίζεται τόσο στις φυσικές διεργασίες που διέπουν τη διάρκεια ζωής των οντοτήτων του πληθυσμού μέχρι τη στιγμή της αποτυχίας, όσο και στην προσαρμογή του επιλεγόμενου μοντέλου στα παρατηρούμενα δεδομένα. Έτσι, η μελέτη των φυσικών καταστάσεων, οδήγησε στη δημιουργία κατανομών με διάφορες μορφές συνάρτησης κινδύνου, προκειμένου να καταστεί δυνατή η περιγραφή των πραγματικών καταστάσεων στο σύνολό τους, είτε αυτές παρουσιάζουν φθορά (ή βελτίωση) στη λειτουργία τους με την πάροδο του χρόνου, είτε εμφανίζουν μια πιο σύνθετη συνδυαστική λειτουργία (π.χ. αρχικά φθείρονται και εν συνεχεία ενδυναμώνουν με την πάροδο του χρόνου ή το αντίθετο).

Στο συγκεκριμένο κεφάλαιο μελετάται μια νέα παραμετρική κατανομή που αποτελεί επέκταση της κατανομής του Chen (Chen, 2000). Περιγράφεται η διαδικασία με την οποία αυτή έχει παραχθεί και εν συνεχεία υπολογίζονται οι σχετικές με αυτή συναρτήσεις πυκνότητας πιθανότητας, επιβίωσης και κινδύνου, ελέγχοντας τις συνθήκες κάτω από τις οποίες οι συναρτήσεις αυτές μπορούν να αλλάξουν σχηματική μορφή. Επιπρόσθετα, εξετάζονται διάφορες ιδιότητες της γενικευμένης κατανομής. Τέλος, γίνεται εκτίμηση των παραμέτρων της κατανομής, ενώ ελέγχεται και η ικανότητα προσαρμογής της σε πραγματικά σύνολα δεδομένων.

## 2.1 Το μοντέλο

Ο Chen (2000) μελέτησε μια δι-παραμετρική κατανομή, η οποία έχει αύξουσα ή κοίλη συνάρτηση κινδύνου και συνάρτηση επιβίωσης που δίνεται από τη σχέση:

$$S_{Chen}(t) = e^{-\lambda(1-e^{t^\beta})}, \quad t, \beta, \lambda > 0. \quad (2.1)$$

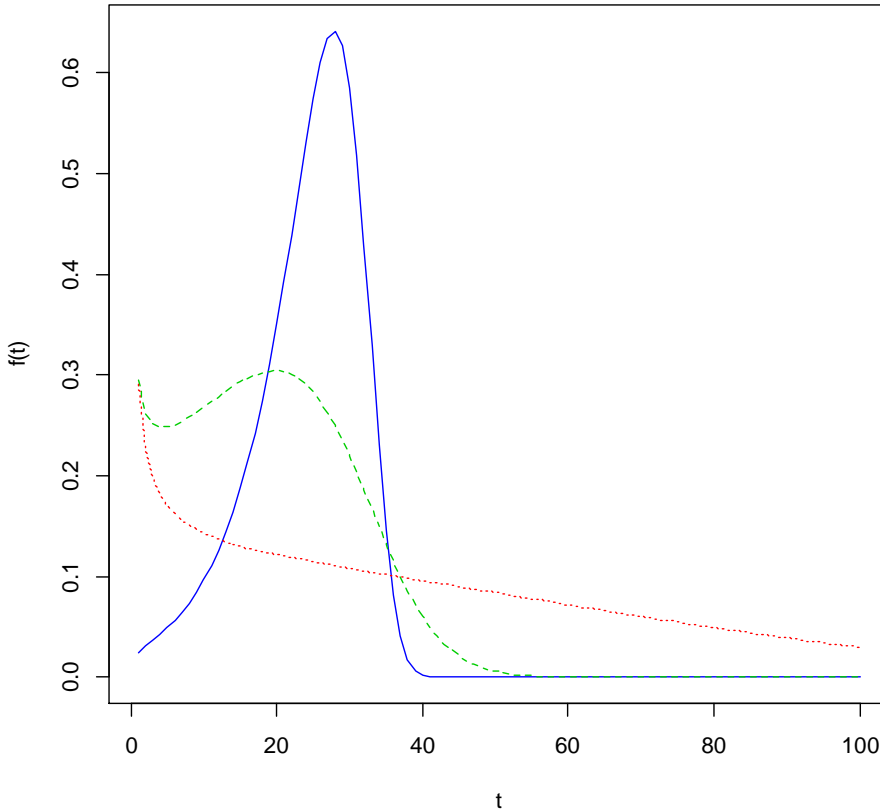
Όμως, ο μικρός αριθμός παραμέτρων, ταυτόχρονα με την ευκολία μελέτης της κατανομής, περιορίσε την ευελιξία της στην περιγραφή πραγματικών καταστάσεων. Σε αυτό το κεφάλαιο εισάγεται μια γενίκευση της παραπάνω κατανομής, η οποία διευρύνει το φάσμα εφαρμογών της αρχικής κατανομής του Chen (2000), δίχως παράλληλα να περιορίζει σε σημαντικό βαθμό τη δυνατότητα μελέτης των ιδιοτήτων της.

Υποθέτοντας την ύπαρξη μιας κατανομής με συνάρτηση επιβίωσης που δίνεται από την (2.1), μια τυχαία μεταβλητή  $T$  θα λέγεται ότι ακολουθεί την επεκτεταμένη κατανομή του Chen (EC) όταν η συνάρτηση πυκνότητας πιθανότητάς της δίνεται από τη σχέση:

$$f(t; \theta) = \frac{(\alpha - 1) \beta \lambda t^{\beta-1} e^{-\lambda(e^{t^\beta} - 1)}}{\left\{ 1 - (1 - \alpha) e^{-\lambda(e^{t^\beta} - 1)} \right\} \ln \alpha}, \quad t, \alpha, \beta, \lambda > 0. \quad (2.2)$$

Ανάλογα με τις τιμές των τριών παραμέτρων της EC κατανομής, η σ.π.π. δύναται να είναι φθίνουσα, μονοκόρυφη ή τροποποιημένη κοίλη (roller-coaster). Οι διάφορες μορφές της σ.π.π. παρουσιάζονται στο γράφημα 2.1 για επιλεγμένες τιμές των παραμέτρων.

**Γράφημα 2.1** Συναρτήσεις πυκνότητας πιθανότητας της EC κατανομής για  $(\alpha, \beta, \lambda) = (2, 0.2, 0.5)$  (μπλε συνεχής γραμμή),  $(\alpha, \beta, \lambda) = (8, 0.8, 0.01)$  (πράσινη διακεκομμένη γραμμή) και  $(\alpha, \beta, \lambda) = (50, 0.4, 0.4)$  (κόκκινη γραμμή με τελείες).



Προφανώς, για  $\alpha \rightarrow 1$ , η νέο-προταθείσα κατανομή ταυτίζεται με αυτή του Chen, ενώ για  $\alpha \in (0,1)$ , αποδεικνύεται ότι η επεκτεταμένη κατανομή αποτελεί μίξη της κατανομής του Chen με τη λογαριθμική κατανομή. Πράγματι, κάνοντας χρήση των αποτελεσμάτων των Barlow and Proschan (1981) και Arnold et al. (1992), θεωρούμε τους χρόνους ζωής  $T = \min(T_1, T_2, \dots, T_Z)$  ενός συστήματος  $Z$  όμοιων εξαρτημάτων σε

σειρά, όπου η αποτυχία συμβαίνει εάν τουλάχιστον ένα από τα εν σειρά εξαρτήματα πάψει να λειτουργεί. Υποθέτοντας ότι οι χρόνοι ζωής των εξαρτημάτων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με σ.π.π. που δίνεται από τη σχέση

$$f(t; \beta, \lambda) = \beta \lambda t^{\beta-1} e^{t^\beta - \lambda(e^{t^\beta} - 1)},$$

για  $t, \beta, \lambda \in \mathbb{R}_+$  και η κατανομή του αριθμού τους  $Z$  είναι η λογαριθμική, ανεξάρτητη των  $T$ , με συνάρτηση πιθανότητας

$$p(z; \alpha) = -\frac{(1-\alpha)^z}{z \ln \alpha},$$

για  $z \in \mathbb{N} - \{0\}$ ,  $\alpha \in (0,1)$ , τότε η κατανομή του  $T | Z$  έχει σ.π.π.

$$f(t | z; \beta, \lambda) = \beta \lambda z t^{\beta-1} e^{t^\beta - \alpha z (e^{t^\beta} - 1)},$$

για  $t, \beta, \lambda \in \mathbb{R}_+$  και η κατανομή του  $T$  είναι η EC κατανομή με σ.π.π. που δίνεται από τη σχέση (2.2).

υπολογισμός των  $r$ -τάξεως ροπών απαιτεί χρήση τυπικών διαδικασιών αριθμητικής ολοκλήρωσης, που είναι διαθέσιμες στην πλειοψηφία των μαθηματικών υπολογιστικών πακέτων. Για  $\alpha \in (0,1)$ , ωστόσο, μπορούν να εκφραστούν από τη σχέση:

$$E(T^r) = -\frac{1}{\ln \alpha} \sum_{k=1}^{+\infty} \frac{(1-\alpha)^k}{k} \int_0^{+\infty} \left\{ \ln \left( \frac{t}{\kappa \lambda} + 1 \right) \right\}^{r/\beta} e^{-t} dt.$$

Επιπλέον με αντιστροφή της α.σ.κ., η οποία μπορεί εύκολα να βρεθεί από τη σχέση (2.2), προκύπτει η συνάρτηση υπολογισμού των εκατοστιαίων σημείων, η οποία δίνεται από τη σχέση:

$$F^{-1}(p) = \left[ \ln \left\{ 1 - \frac{1}{\lambda} \ln \left( \frac{1 - \alpha^{1-p}}{1 - \alpha} \right) \right\} \right]^{1/\beta},$$

για  $p \in (0,1)$ . Επομένως, εύκολα προκύπτει ότι η διάμεσος θα είναι  $M = \left[ \ln \left\{ 1 + \frac{1}{\lambda} \ln(1 + \sqrt{\alpha}) \right\} \right]^{1/\beta}$ .

## 2.2 Συναρτήσεις κινδύνου και μέσου υπολειπόμενου χρόνου ζωής

Από τη σχέση (2.2) και χρησιμοποιώντας τη σχέση

$$S(t; \theta) = 1 - F(t; \theta),$$

προκύπτει ότι η συνάρτηση επιβίωσης της EC κατανομής δίνεται από τον τύπο

$$S(t; \theta) = \frac{\ln \left\{ 1 - (1 - \alpha) e^{-\lambda(e^{t^\beta} - 1)} \right\}}{\ln \alpha}.$$

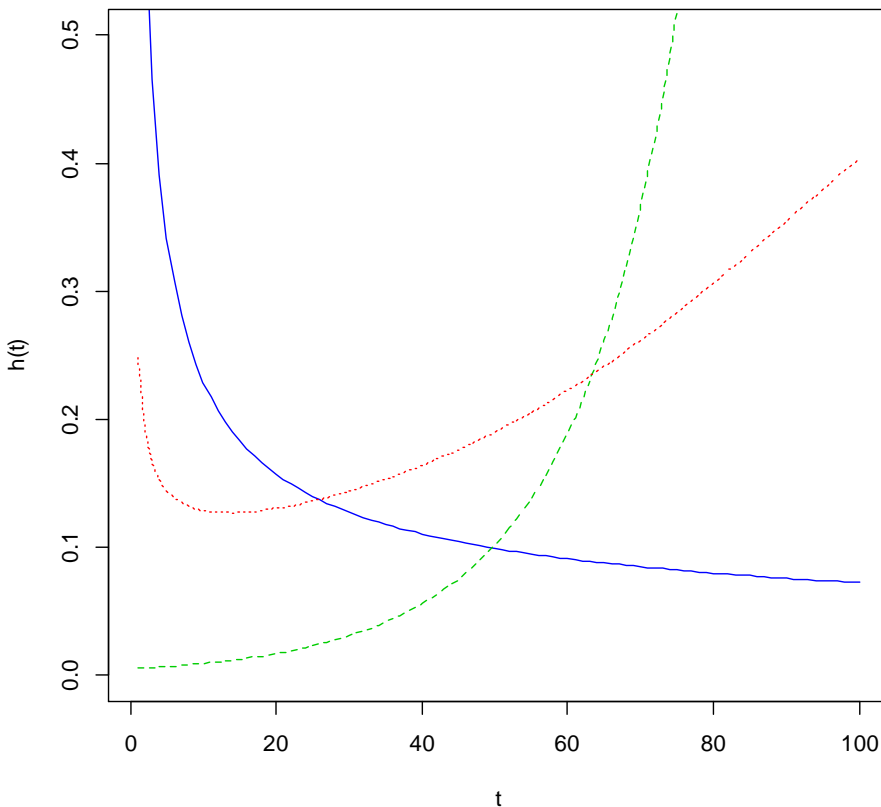
Επομένως και η συνάρτηση κινδύνου της προτεινόμενης κατανομής θα δίνεται από τη σχέση:

$$h(t; \theta) = \frac{(\alpha - 1)\beta\lambda t^{\beta-1} e^{t^\beta - \lambda(e^{t^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t^\beta} - 1)}\} \ln\{1 - (1 - \alpha)e^{-\lambda(e^{t^\beta} - 1)}\}}.$$

Με αντίστοιχη μεθοδολογία με αυτή που εφαρμόζεται αναλυτικά στο πέμπτο κεφάλαιο, μπορεί να αποδειχθεί ότι για  $\beta \geq 1$ , η συνάρτηση κινδύνου της κατανομής EC είναι γνησίως αύξουσα (IFR) (Glaser,

1980). Σε διαφορετική περίπτωση, δηλαδή για  $\beta < 1$ , η συνάρτηση κινδύνου μπορεί να είναι γνησίως αύξουσα, φθίνουσα, ή ακόμη να λάβει και κυπελλοειδή μορφή. Οι διάφορες μορφές της συνάρτησης κινδύνου της EXTG κατανομής φαίνονται στο Γράφημα 2.2, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \beta, \lambda)$ .

**Γράφημα 2.2** Συναρτήσεις κινδύνου της EC κατανομής για  $(\alpha, \beta, \lambda) = (2, 0.2, 0.5)$  (μπλε συνεχής γραμμή),  $(\alpha, \beta, \lambda) = (8, 0.8, 0.01)$  (πράσινη διακεκομμένη γραμμή) και  $(\alpha, \beta, \lambda) = (50, 0.4, 0.4)$  (κόκκινη γραμμή με τελείες).





Δεδομένου ότι δεν υπήρξε αποτυχία προγενέστερη του  $t_0$ , ο υπολοιπόμενος χρόνος ζωής ορίζεται ως το χρονικό διάστημα από το  $t_0$  μέχρι τη στιγμή της αποτυχίας. Η συνάρτηση του μέσου υπολοιπόμενου χρόνου ζωής της κατανομής EC, για  $\alpha \in (0,1)$  δίνεται από τη σχέση:

$$\begin{aligned} m(t_0; \theta) &= E(T - t_0 | T \geq t_0) = \int_{t_0}^{+\infty} \frac{(t - t_0) f(t; \theta)}{1 - F(t_0; \theta)} dt \\ &= \frac{1}{\ln \left\{ 1 - (1 - \alpha) e^{-\lambda (e^{t_0^\beta} - 1)} \right\}} \sum_{k=1}^{+\infty} \frac{(1 - \alpha)^k}{k} \\ &\quad \times \int_{k\lambda (e^{t_0^\beta} - 1)}^{+\infty} \left[ \left\{ \ln \left( \frac{t}{k\lambda} + 1 \right) \right\}^{1/\beta} - t_0 \right] e^{-t} dt. \end{aligned}$$

### 2.3 Στατιστική συμπερασματολογία

Θεωρώντας ένα τυχαίο δείγμα  $n$  παρατηρήσεων  $y_{obs} = (t_i; i = 1, \dots, n)$  από την (2.2), ο λογάριθμος της συνάρτησης πιθανοφάνειας  $L(\theta; y_{obs})$  δίνεται από τη σχέση:

$$\begin{aligned} \ln L(\theta; y_{obs}) = l(\theta; y_{obs}) &= n[\ln\{(\alpha - 1)\beta\lambda\} - \ln(\ln \alpha) + \lambda] \\ &\quad + (\beta - 1) \sum_{i=1}^n \ln t_i + \sum_{i=1}^n t_i^\beta - \lambda \sum_{i=1}^n e^{t_i^\beta} \\ &\quad - \sum_{i=1}^n \ln\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}. \end{aligned}$$

Παραγωγίζοντας ως προς τις παραμέτρους  $\theta = (\alpha, \beta, \lambda)$  προκύπτουν οι εξής σχέσεις:

$$\frac{\partial l}{\partial \alpha} = n\{(\alpha - 1)^{-1} + (\alpha \ln \alpha)^{-1}\} - \sum_{i=1}^n \frac{e^{-\lambda(e^{t_i^\beta} - 1)}}{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}}, \quad (2.3)$$

$$\begin{aligned} \frac{\partial l}{\partial \beta} = & n\beta^{-1} + \sum_{i=1}^n \{1 + t_i^\beta - \lambda t_i^\beta e^{t_i^\beta}\} \ln t_i \\ & - (1 - \alpha)\lambda \sum_{i=1}^n \frac{t_i^\beta \ln t_i e^{t_i^\beta - \lambda(e^{t_i^\beta} - 1)}}{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}}, \end{aligned} \quad (2.4)$$

$$\frac{\partial l}{\partial \lambda} = n(\lambda^{-1} + 1) - \sum_{i=1}^n e^{t_i^\beta} - (1 - \alpha) \sum_{i=1}^n \frac{(e^{t_i^\beta} - 1)e^{-\lambda(e^{t_i^\beta} - 1)}}{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}}. \quad (2.5)$$

Οι εκτιμητές μέγιστης πιθανοφάνειας  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$  των παραμέτρων  $\theta = (\alpha, \beta, \lambda)$  προκύπτουν από την επίλυση του συστήματος των τριών μη-γραμμικών κανονικών εξισώσεων:  $\frac{\partial l}{\partial \alpha} = 0$ ,  $\frac{\partial l}{\partial \beta} = 0$  και  $\frac{\partial l}{\partial \lambda} = 0$ .

Είναι προφανές ότι για την επίλυση του εν λόγω συστήματος, θα πρέπει να χρησιμοποιηθεί κάποια από τις μεθόδους αριθμητικής ανάλυσης όπως για παράδειγμα η μέθοδος Newton-Raphson, ή η μέθοδος scoring.

Επιπλέον, σύμφωνα με την ασυμπτωτική θεωρία κανονικότητας, η κατανομή των εκτιμητών μέγιστης πιθανοφάνειας  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ , είναι κατά προσέγγιση η τρισδιάστατη κανονική με μέση τιμή  $\theta = (\alpha, \beta, \lambda)$  και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα πληροφορίας του Fisher  $I(\theta)$ , ο οποίος υπολογίζεται από τη σχέση  $I(\theta) = E(J(\theta))$ , όπου,  $J(\theta)$  είναι ο συμμετρικός, τρίτης τάξης παρατηρούμενος πίνακας πληροφορίας, ενώ η αναμενόμενη τιμή λαμβάνεται ως προς την κατανομή της τυχαίας μεταβλητής  $T$ .

Τα στοιχεία του πίνακα  $J(\theta)$ :  $J_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}$ , όπου  $i, j = 1, 2, 3$ , δίνονται

αναλυτικά από τις σχέσεις:

$$J_{11} = n\{(\alpha - 1)^{-2} - (1 + \ln \alpha)(\alpha \ln \alpha)^{-2}\} - \sum_{i=1}^n \frac{e^{-2\lambda(e^{t_i^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}^2},$$

$$J_{12} = -\lambda \sum_{i=1}^n \frac{t_i^\beta \ln t_i e^{t_i^\beta - \lambda(e^{t_i^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}^2},$$

$$J_{13} = -\sum_{i=1}^n \frac{(e^{t_i^\beta} - 1)e^{-\lambda(e^{t_i^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}^2},$$

$$J_{22} = n\beta^{-2} - \sum_{i=1}^n t_i^\beta \{1 - \lambda(1 + t_i^\beta)e^{t_i^\beta}\} \ln^2 t_i \\ + (1 - \alpha)\lambda \sum_{i=1}^n \frac{t_i^\beta \ln^2 t_i e^{t_i^\beta - \lambda(e^{t_i^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}^2} \\ \times [1 + t_i^\beta - \lambda t_i^\beta e^{t_i^\beta} - (1 - \alpha)\{1 + t_i^\beta\}e^{-\lambda(e^{t_i^\beta} - 1)}],$$

$$J_{23} = \sum_{i=1}^n t_i^\beta e^{t_i^\beta} \ln t_i + (1 - \alpha) \sum_{i=1}^n \frac{t_i^\beta \ln t_i e^{t_i^\beta - \lambda(e^{t_i^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}^2} \\ \times \{1 - \lambda(e^{t_i^\beta} - 1) - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\},$$

$$J_{33} = n\lambda^{-2} - (1 - \alpha) \sum_{i=1}^n \frac{(e^{t_i^\beta} - 1)^2 e^{-\lambda(e^{t_i^\beta} - 1)}}{\{1 - (1 - \alpha)e^{-\lambda(e^{t_i^\beta} - 1)}\}^2}.$$

Ο παραπάνω πίνακας  $J(\theta)$  αποτελεί συνεπή εκτιμητή του  $I(\theta)$  και μπορεί να χρησιμοποιηθεί για την κατασκευή ασυμπτωτικών

διαστημάτων εμπιστοσύνης για τις άγνωστες παραμέτρους. Ωστόσο, αν οποιαδήποτε από τις πραγματικές τιμές των παραμέτρων είναι κοντά στο μηδέν, τότε η ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας αποτελεί μίξη κατανομών, ως επί το πλείστον Χι-τετράγωνο (Self and Liang, 1987) και η δημιουργία ασυμπτωτικών διαστημάτων εμπιστοσύνης δυσχεραίνεται αρκετά.

## 2.4 Εφαρμογές

Η προσαρμογή της προτεινόμενης EC κατανομής ελέγχεται σε δύο σύνολα πραγματικών δεδομένων. Το πρώτο σύνολο δεδομένων περιλαμβάνει αποτελέσματα από μια δοκιμή σχετικά με τη διάρκεια ζωής 50 ηλεκτρικών εξαρτημάτων (Aarset, 1987). Το δεύτερο σύνολο δεδομένων αποτελείται από 60 παρατηρήσεις που αφορούν το χρόνο μέχρι την αποτυχία μιας συστοιχίας ηλεκτρικών συσκευών (Lawless, 2003). Και στις δύο περιπτώσεις τα δεδομένα ομαδοποιήθηκαν και εκτιμήθηκε η εμπειρική συνάρτηση κινδύνου, επιδεικνύοντας καθαρά κυπελλοειδή μορφή. Εκτός από την προτεινόμενη EC κατανομή, άλλες δύο κατανομές προσαρμόζονται στα δύο σύνολα δεδομένων ούτως ώστε να υπάρξει και μια άμεση αντιπαράθεση των κατανομών μεταξύ τους. Στο πρώτο σύνολο, η προτεινόμενη κατανομή συγκρίνεται με αυτή του Chen (2000), ενώ στο δεύτερο σύνολο ως εναλλακτική χρησιμοποιείται μια τρι-παραμετρική κατανομή που μελετήθηκε από τους Dimitrakoulou et al. (2007), έχει σ.π.π. που δίνεται από τη σχέση

$$g(t; \theta) = \alpha \beta \lambda t^{\beta-1} (1 + \lambda t^\beta)^{\alpha-1} e^{-\lambda t^\beta}, t, \alpha, \beta, \lambda > 0,$$

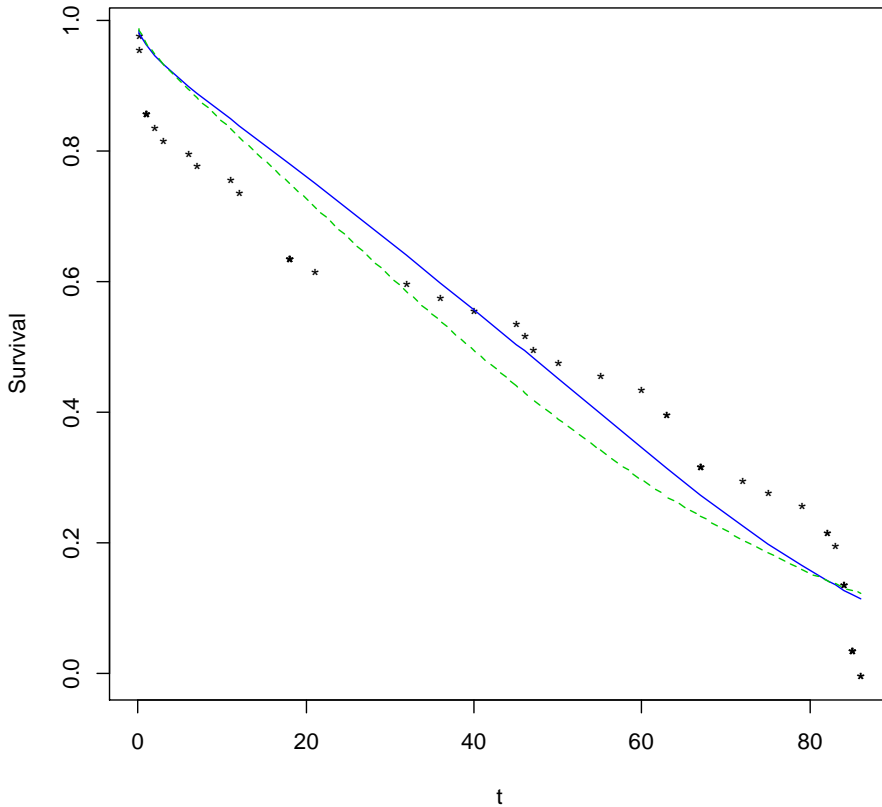
όπου  $\theta = (\alpha, \beta, \lambda)$  και για χάριν συντομίας θα συμβολίζεται ως DAL. Υπολογίζοντας τους εκτιμητές μέγιστης πιθανοφάνειας, χρησιμοποιήθηκε το τεστ καλής προσαρμογής των Kolmogorov-Smirnov (K-S), για τον έλεγχο της προσαρμοστικότητας κάθε μιας από τις δύο κατανομές στα δύο σύνολα δεδομένων. Οι εκτιμητές, ο λογάριθμος της συνάρτησης πιθανοφάνειας (LL), το κριτήριο AIC, η τιμή του στατιστικού του K-S τεστ καθώς και η αντίστοιχη κρίσιμη τιμή (p-value)

δίνονται στον Πίνακα 2.1. Επιπρόσθετα, η τιμή του στατιστικού τεστ πηλίκου πιθανοφάνειας για τον έλεγχο της μηδενικής υπόθεσης  $H_0: \alpha = 1$  υπολογίστηκε να είναι 8.9909 ( $p=0.0027$ ), υποδεικνύοντας ότι η EC περιγράφει το πρώτο σύνολο δεδομένων καλύτερα από την εναλλακτική κατανομή του Chen. Το συμπέρασμα αυτό ενισχύεται και από τη γραφική παράσταση των εμπειρικών και προσαρμοζόμενων συναρτήσεων επιβίωσης του Γραφήματος 2.3.

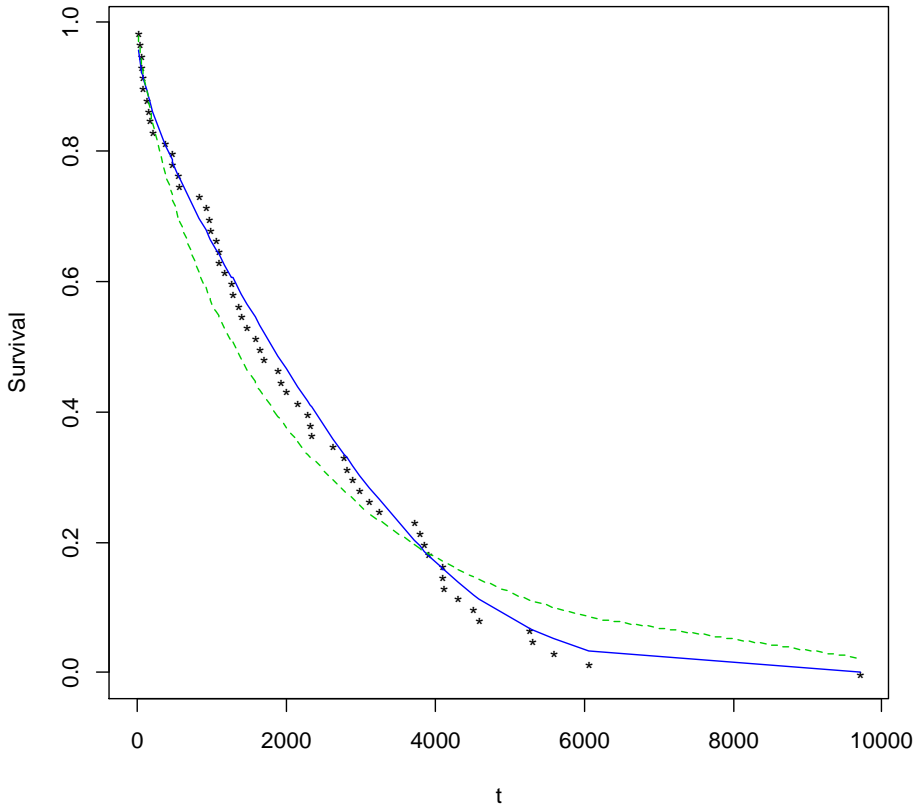
**Πίνακας 2.1** Εκτιμητές παραμέτρων, τιμές λογαρίθμου συνάρτησης πιθανοφάνειας, κριτηρίου AIC, στατιστικού K-S και στατιστικού πηλίκου πιθανοφανειών, καθώς και κρίσιμες τιμές από την προσαρμογή των συγκρινόμενων κατανομών πιθανότητας στα δύο διαθέσιμα σύνολα δεδομένων.

Σύνολο Δεδομένων	Κατανομή	$\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$	LL	AIC	K-S (p-value)	LL ratio test
1) $n = 50$	EC	(154.249, 0.305, 0.11)	-228.6728	463.3456	0.1403 (0.2789)	$H_0: \alpha = 1$ 8.9909 (0.0027)
	Chen	(-----, 0.3444, 0.0205)	-233.1682	470.3364	0.1676 (0.1204)	
2) $n = 60$	EC	(150.3987, 0.1804, 0.0548)	-520.5755	1047.151	0.0522 (0.9939)	
	DAL	(1.52, 0.73, 0.0022)	-522.7890	1051.578	0.1416 (0.1635)	

**Γράφημα 2.3** Γραφικές παραστάσεις της συνάρτησης επιβίωσης της εμπειρικής κατανομής (σημεία \*), της EC κατανομής (μπλε συνεχής γραμμή) και των κατανομών Chen και DAL (πράσινη διακεκομμένη γραμμή) για τα δύο σύνολα δεδομένων αντίστοιχα.



(α) Ηλεκτρικά Εξαρτήματα



(β) Ηλεκτρικές συσκευές

## 2.5 Συμπεράσματα

Η προτεινόμενη EC κατανομή αποτελεί επέκταση της κατανομής του Chen (2000). Σκοπός της επέκτασης είναι η βελτιστοποίηση της ευελιξίας της αρχικής κατανομής και η καλύτερη προσαρμογή της σε πραγματικές καταστάσεις, δίχως αυτό να δυσχεραίνει υπερβολικά τη δυνατότητα μελέτης της. Με τη μελέτη της και την συγκρινόμενη εφαρμογή της σε δεδομένα έναντι της αρχικής, η νέα κατανομή αποδεικνύεται ότι ικανοποιεί το λόγο ύπαρξής της.





# Κεφάλαιο 3

## Η Weibull-Διωνυμική Κατανομή

Τα τελευταία χρόνια, πολλές νέες κατανομές πιθανότητας αναπτύχθηκαν ως τροποποιήσεις της Εκθετικής και της Weibull κατανομής. Ενδεικτικά αναφέρονται: η Weibull-Geometric (WG) που προτάθηκε από τους Marshall and Olkin (1997) και μελετήθηκε επίσης και από τους Souza et al. (2008), η Exponential-Geometric (EG) που μελετήθηκε από τους Adamidis and Loukas (1998), η Exponential-Poisson (EP) που προτάθηκε από τον Kus (2007), η Exponential-Logarithmic (EL) που προτάθηκε από τους Tahmasbi and Rezaei (2008), η Exponential-Binomial (EB) που προτάθηκε από τους Chahkandi and Ganjali (2009) και τέλος η Weibull-Logarithmic (WL) που προτάθηκε από τους Ciumara and Preda (2009).

Στο κεφάλαιο αυτό μελετάται μια κατανομή με τρεις παραμέτρους, η οποία αποτελεί γενίκευση της EB κατανομής που προτάθηκε από τους Chahkandi and Ganjali (2009) και έχει καλύτερη εφαρμογή σε πραγματικά δεδομένα έναντι τόσο της EB όσο και άλλων τρι-παραμετρικών κατανομών. Αναφέρεται η διαδικασία με την οποία παρήχθη η προτεινόμενη κατανομή και δίνονται χαρακτηριστικά στοιχεία της όπως η συνάρτηση πυκνότητας πιθανότητάς της καθώς και σχέσεις της με άλλες κατανομές. Η περαιτέρω μελέτη των ιδιοτήτων της κατανομής κάνει φανερή την ύπαρξη συνάρτησης κινδύνου πολλαπλών μορφών (αύξουσα, φθίνουσα, κυρτή και τροποποιημένη κοίλη). Επιπρόσθετα εκτιμώνται οι παράμετροι της κατανομής με τη μέθοδο Μέγιστης Πιθανοφάνειας και υπολογίζεται ο παρατηρούμενος πίνακας πληροφορίας. Τέλος η ποιότητα προσαρμογής της προτεινόμενης κατανομής ελέγχεται σε πραγματικά δεδομένα.

### 3.1 Η κατανομή

Έστω  $Y_1, Y_2, \dots, Y_Z$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την κατανομή Weibull με παραμέτρους  $\beta$  και  $\lambda$ , δηλαδή έχουν συνάρτηση πυκνότητας πιθανότητας που δίνεται από τη σχέση:

$$g(y; \beta, \lambda) = \beta \lambda (\lambda y)^{\beta-1} e^{-(\lambda y)^\beta}, \quad y, \beta, \lambda > 0.$$

Θεωρούμε επίσης ως  $Z$  μια τυχαία μεταβλητή που ακολουθεί Διωνυμική κατανομή με παραμέτρους  $n$  και  $\alpha$ , δηλαδή  $Z \sim B(n, \alpha)$ , με συνάρτηση πιθανότητας:

$$p(z) = \binom{n}{z} \alpha^z (1-\alpha)^{n-z}, \quad n \in \mathbb{N}, \alpha \in (0, 1).$$

Υποθέτοντας ότι οι τυχαίες μεταβλητές  $Y$  και  $Z$  είναι ανεξάρτητες, ορίζουμε την τυχαία μεταβλητή  $T = \min(Y_1, Y_2, \dots, Y_Z)$ . Σε αυτή την περίπτωση η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας της  $T/Z=z$  είναι

$$f(t | z; \beta, \lambda) = \beta \lambda z (\lambda t)^{\beta-1} e^{-(\lambda t)^\beta}, \quad t, \beta, \lambda > 0,$$

ενώ, η συνάρτηση πυκνότητας πιθανότητας της  $T$  δίνεται από τη σχέση:

$$f(t; \theta) = \frac{n \alpha \beta \lambda (\lambda t)^{\beta-1} e^{-(\lambda t)^\beta} \left\{ 1 + \alpha e^{-(\lambda t)^\beta} \right\}^{n-1}}{(1+\alpha)^n - 1}, \quad (3.1)$$

όπου  $t > 0$  και  $\theta = (\alpha, \beta, \lambda)$ , με  $\beta > 0$  παράμετρο μορφής και  $\alpha, \lambda > 0$  παραμέτρους κλίμακας.

Στη συνέχεια θα λέμε ότι μια κατανομή ακολουθεί την Weibull διωνυμική (Weibull Binomial ή WB) κατανομή, όταν η συνάρτηση πυκνότητας πιθανότητάς της θα δίνεται από τη σχέση (3.1).

**Θεώρημα:**

Η σ.π.π. (3.1) είναι γνησίως φθίνουσα όταν  $\beta \leq 1$ , ενώ έχει τουλάχιστον μια κορυφή σε οποιαδήποτε άλλη περίπτωση.

**Απόδειξη:**

Έστω

$$\frac{d \ln f(t; \theta)}{dt} = 0 \Leftrightarrow$$

$$(\beta - 1) - \beta(\lambda t)^\beta - (n - 1)\alpha\beta(\lambda t)^\beta e^{-(\lambda t)^\beta} \left\{ 1 + \alpha e^{-(\lambda t)^\beta} \right\}^{-1} = 0.$$

Περιπτώσεις:

1)  $0 < \beta \leq 1$ , τότε είναι:

$$\frac{d \ln f(t; \theta)}{dt} < 0 \Rightarrow f(t; \theta) \downarrow \forall t > 0 \text{ φθίνουσα σ.π.π.}$$

2)  $\beta > 1$ , τότε είναι:

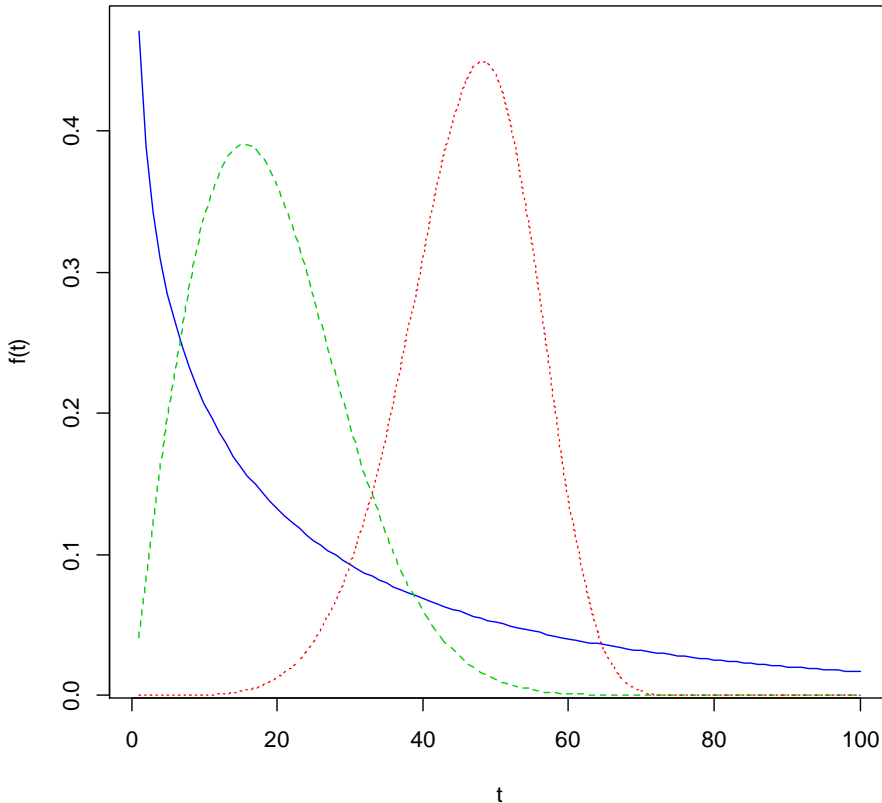
$$\frac{d \ln f(t; \theta)}{dt} = 0 \Leftrightarrow 1 + \alpha e^{-(\lambda t)^\beta} \left[ \frac{n - 1}{1 + \alpha e^{-(\lambda t)^\beta}} \right] = \frac{1 - \beta^{-1}}{(\lambda t)^\beta} \Leftrightarrow$$

$$g_1(t; \theta) = g_2(t; \theta).$$

Όμως οι  $g_1$  και  $g_2$  είναι συνεχείς συναρτήσεις με κοινό πεδίο ορισμού  $(0, +\infty)$  και σύνολα τιμών  $\left( 1, \frac{1 + n\alpha}{1 + \alpha} \right)$  και  $(0, +\infty)$  αντίστοιχα. Επομένως η παραπάνω εξίσωση έχει τουλάχιστον μια ρίζα στο  $(0, +\infty)$ , δηλαδή η σ.π.π. (3.1) έχει τουλάχιστον μια κορυφή στο  $(0, +\infty)$ .

Οι διάφορες μορφές της σ.π.π. φαίνονται και στο παρακάτω γράφημα, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \beta, \lambda)$ .

**Γράφημα 3.1** Συναρτήσεις πυκνότητας πιθανότητας της WB κατανομής για  $(\alpha, \beta, \lambda) = (0.1, 0.8, 0.02)$  (μπλε συνεχής γραμμή),  $(\alpha, \beta, \lambda) = (5, 2, 0.05)$  (πράσινη διακεκομμένη γραμμή) και  $(\alpha, \beta, \lambda) = (2, 6, 0.1)$  (κόκκινη γραμμή με τελείες).



Είναι εύκολο να δει κανείς ότι για  $\beta = 1$ , η προτεινόμενη κατανομή δίνει την EB κατανομή που προτάθηκε από τους Chahkandi and Ganjali (2009). Επιπρόσθετα για  $\beta = 2$ , προκύπτει η κατανομή Reyleigh-Binomial (RB). Τέλος για  $\alpha = n = 1$  προκύπτει η κατανομή Weibull, ενώ όταν  $\alpha = \beta = n = 1$  τότε η προτεινόμενη κατανομή ισοδυναμεί με την εκθετική κατανομή.

### 3.2 Ιδιότητες

Ολοκληρώνοντας τη σχέση (3.1) στο διάστημα  $(0, t)$  προκύπτει η αθροιστική συνάρτηση κατανομή της WB κατανομής, η οποία δίνεται από τη σχέση:

$$F(t; \theta) = \frac{\left\{1 + \alpha e^{-(\lambda t)^\beta}\right\}^n - (1 + \alpha)^n}{1 - (1 + \alpha)^n}, \quad t, \alpha, \beta, \lambda > 0. \quad (3.2)$$

Επομένως και η συνάρτηση επιβίωσης της προτεινόμενης κατανομής θα δίνεται από τη σχέση:

$$S(t; \theta) = 1 - F(t; \theta) = \frac{1 - \left\{1 + \alpha e^{-(\lambda t)^\beta}\right\}^n}{1 - (1 + \alpha)^n}, \quad t, \alpha, \beta, \lambda > 0. \quad (3.3)$$

Επιπρόσθετα, με αντιστροφή της (3.2), προκύπτει η συνάρτηση υπολογισμού των εκατοστιαίων σημείων, η οποία θα δίνεται από τη σχέση:

$$F^{-1}(p) = \lambda^{-1} \left[ \ln \left[ \alpha^{-1} \left[ \left\{ p + (1-p)(1 + \alpha)^n \right\}^{n^{-1}} - 1 \right] \right] \right]^{\beta^{-1}}, \quad (3.4)$$

όπου  $p \in (0, 1)$ .

Με απλή εφαρμογή για  $p = 1/2$  στην (3.4) έχουμε ότι η διάμεσος της WB κατανομής είναι

$$M = \lambda^{-1} \left[ \ln \left[ \alpha \left[ \left[ \frac{1}{2} \{1 + (1 + \alpha)^n\} \right]^{n^{-1}} - 1 \right] \right] \right]^{\beta^{-1}}.$$

Οι ροπές  $r$ -τάξης της κατανομής δίνονται από τη σχέση:

$$E(T^r) = r \int_0^{\infty} t^{r-1} S(t; \theta) dt = r \int_0^{\infty} t^{r-1} \frac{1 - \left\{1 + \alpha e^{-(\lambda t)^\beta}\right\}^n}{1 - (1 + \alpha)^n} dt.$$

Περαιτέρω υπολογισμοί απαιτούν χρήση τυπικών διαδικασιών αριθμητικής ολοκλήρωσης, που είναι διαθέσιμοι στην πλειοψηφία των μαθηματικών πακέτων.

Χρησιμοποιώντας τις σχέσεις (3.1) και (3.3) παίρνουμε τη συνάρτηση κινδύνου της WB κατανομής, η οποία δίνεται από τη σχέση:

$$h(t; \theta) = \frac{f(t; \theta)}{S(t; \theta)} = \frac{n\alpha\beta\lambda(\lambda t)^{\beta-1} e^{-(\lambda t)^\beta} \left\{1 + \alpha e^{-(\lambda t)^\beta}\right\}^{n-1}}{\left\{1 + \alpha e^{-(\lambda t)^\beta}\right\}^n - 1}, \quad (3.5)$$

όπου  $t, \alpha, \beta, \lambda > 0$ .

#### Θεώρημα:

Η συνάρτηση κινδύνου (3.5) είναι γνησίως φθίνουσα όταν  $\beta \leq 1$ .

#### Απόδειξη:

Εφαρμόζοντας το θεώρημα του Glaser (1980) προκύπτει η συνάρτηση:

$$\begin{aligned} n(t; \theta) &= -\frac{f'(t; \theta)}{f(t; \theta)} = \\ &= (1 - \beta)t^{-1} + \beta\lambda(\lambda t)^{\beta-1} + n(n-1)\alpha\beta\lambda(\lambda t)^{\beta-1} e^{-(\lambda t)^\beta} \left\{1 + \alpha e^{-(\lambda t)^\beta}\right\}^{n-2}. \end{aligned}$$

Παραγωγίζοντας, την  $n(t; \theta)$  έχουμε:

$$n'(t; \theta) = (\beta - 1)t^{-2} \left[ 1 + \beta(\lambda t)^\beta \left[ 1 + n(n-1)\alpha e^{-(\lambda t)^\beta} \left\{ 1 + a e^{-(\lambda t)^\beta} \right\}^{n-2} \right] \right] \\ - n(n-1)(\alpha\beta\lambda)^2 (\lambda t)^{2(\beta-1)} \left\{ 1 + a e^{-(\lambda t)^\beta} \right\}^{n-3} \left\{ 1 + (n-1)a e^{-(\lambda t)^\beta} \right\}.$$

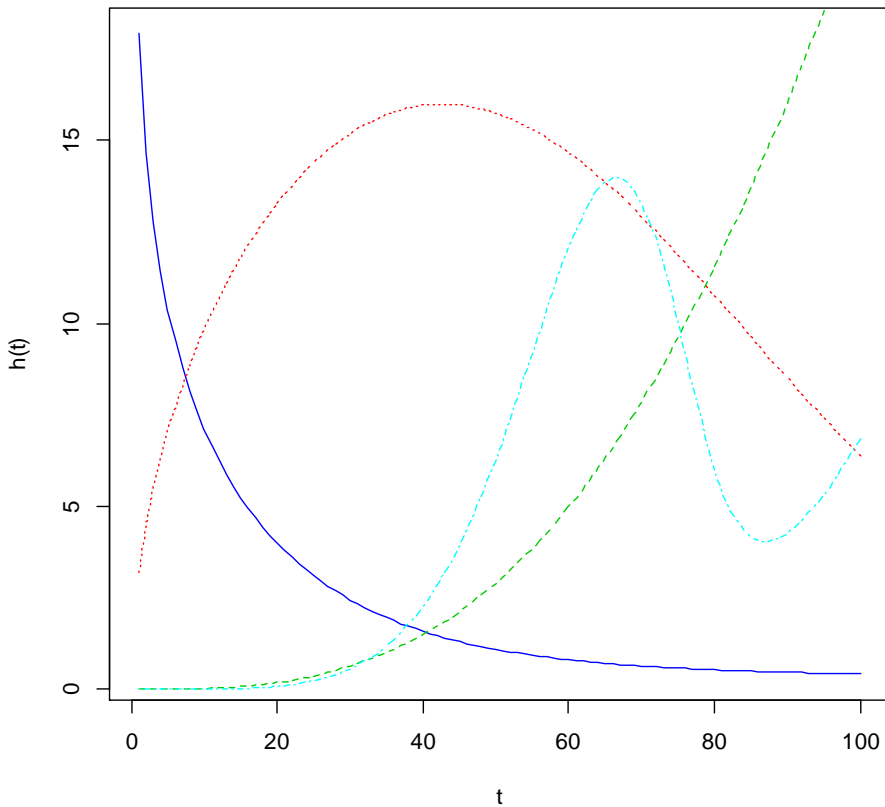
Επομένως, όταν  $\beta \leq 1$  είναι:

$$n'(t; \theta) < 0 \Rightarrow h(t; \theta) \downarrow \forall t > 0 \text{ φθίνουσα συνάρτηση κινδύνου.}$$

Παράλληλα για  $\beta > 1$  η συνάρτηση κινδύνου (3.5) δύναται να αποκτήσει άλλες τρεις διαφορετικές μορφές (αύξουσα, κυρτή και τροποποιημένη κοίλη). Αυτό δίνει τη δυνατότητα στον εκάστοτε ερευνητή να τη χρησιμοποιεί τόσο για θέματα μηχανικής αξιοπιστίας, όσο και για χρόνους ζωής μετάλλων ή ηλεκτρικών συσκευών, καθώς και για τη μελέτη χρονικών διαστημάτων μεταξύ καιρικών φαινομένων.

Οι διάφορες μορφές της συνάρτησης κινδύνου φαίνονται και στο παρακάτω γράφημα, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \beta, \lambda)$ .

**Γράφημα 3.2** Συναρτήσεις κινδύνου της WB κατανομής για  $(\alpha, \beta, \lambda) = (0.3, 0.8, 0.6)$  (μπλε συνεχής γραμμή),  $(\alpha, \beta, \lambda) = (8, 4, 0.09)$  (πράσινη διακεκομμένη γραμμή),  $(\alpha, \beta, \lambda) = (3, 1.5, 0.2)$  (κόκκινη γραμμή με τελείες) και  $(\alpha, \beta, \lambda) = (0.5, 6, 0.15)$  (γαλάζια αραιά διακεκομμένη γραμμή).





### 3.3 Ε.Μ.Π. και ασυμπτωτικά διαστήματα εμπιστοσύνης

Θεωρώντας ένα τυχαίο δείγμα  $m$  παρατηρήσεων  $y_{obs} = (t_i; i = 1, \dots, m)$  από την (3.1), ο λογάριθμος της συνάρτησης πιθανοφάνειας δίνεται από τη σχέση:

$$l(\theta; y_{obs}) = m \log(n\alpha\beta\lambda) + (\beta - 1) \sum_{i=1}^m \log(\lambda t_i) - \sum_{i=1}^m (\lambda t_i)^\beta + (n-1) \sum_{i=1}^m \log\left\{1 + \alpha e^{-(\lambda t_i)^\beta}\right\} - m \log\left\{(1-\alpha)^n - 1\right\}.$$

Παραγωγίζοντας ως προς  $\theta = (\alpha, \beta, \lambda)$  προκύπτουν οι σχέσεις:

$$\frac{\partial l}{\partial \alpha} = -m\alpha^{-1} - \frac{mn}{(1+\alpha) - (1+\alpha)^{1-n}} + (n-1) \sum_{i=1}^m \left\{ \frac{1}{\alpha + e^{(\lambda t_i)^\beta}} \right\}, \quad (3.6)$$

$$\begin{aligned} \frac{\partial l}{\partial \beta} = & -m\beta^{-1} + \sum_{i=1}^m \left\{ 1 - (\lambda t_i)^\beta \right\} \log(\lambda t_i) \\ & - (n-1) \sum_{i=1}^m \left\{ \frac{\alpha (\lambda t_i)^\beta \log(\lambda t_i)}{\alpha + e^{(\lambda t_i)^\beta}} \right\}, \end{aligned} \quad (3.7)$$

$$\frac{\partial l}{\partial \lambda} = m(\beta - 2)\lambda^{-1} - \beta\lambda^{-1} \sum_{i=1}^m (\lambda t_i)^\beta \left\{ 1 + \frac{(n-1)\alpha}{\alpha + e^{(\lambda t_i)^\beta}} \right\}. \quad (3.8)$$

Οι εκτιμητές μέγιστης πιθανοφάνειας  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$  των παραμέτρων  $\theta = (\alpha, \beta, \lambda)$  προκύπτουν από την επίλυση του συστήματος των τριών μη-γραμμικών κανονικών εξισώσεων:  $\frac{\partial l}{\partial \alpha} = 0$ ,  $\frac{\partial l}{\partial \beta} = 0$  και  $\frac{\partial l}{\partial \lambda} = 0$ , η οποία δύναται να επιτευχθεί με χρήση μιας επαναληπτικής αριθμητικής μεθόδου.

Εφόσον οι εκτιμητές μέγιστης πιθανοφάνειας των αγνώστων παραμέτρων  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$  δεν μπορούν να προκύψουν σε κλειστή μορφή, δεν είναι εύκολο να προσδιοριστεί η ακριβής κατανομή των εκτιμητών αυτών. Ωστόσο, εφαρμόζοντας προσέγγιση μεγάλου μεγέθους δείγματος, μπορεί να υποθεθεί ότι οι εκτιμητές των παραμέτρων ακολουθούν προσεγγιστικά τρισδιάστατη κανονική κατανομή με μέση τιμή  $\theta = (\alpha, \beta, \lambda)$  και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα πληροφορίας του Fisher  $I(\theta) = E(J(\theta))$  (βλέπε Lawless, 2003), όπου  $J(\theta)$  είναι ο παρατηρούμενος πίνακας πληροφορίας με στοιχεία  $J_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}$ ,  $i, j = 1, 2, 3$ , ο οποίος κάτω από ασθενείς συνθήκες αποτελεί συνεπή εκτιμητή του  $I(\theta)$ .

Παραγωγίζοντας επιμέρους τις σχέσεις (3.6)-(3.8) μπορούν εύκολα να προκύψουν τα στοιχεία του, δευτέρας τάξεως, συμμετρικού, παρατηρούμενου πίνακα πληροφορίας  $J(\theta)$ , τα οποία δίνονται από τις σχέσεις:

$$J_{11} = m\alpha^{-2} + \frac{mn\{1+(n-1)(1+\alpha)^{-n}\}}{\{(1+\alpha)-(1+\alpha)^{1-n}\}^2} - (n-1)\sum_{i=1}^m \frac{1}{\{\alpha + e^{(\lambda t_i)^\beta}\}^2},$$

$$J_{12} = -(n-1)\sum_{i=1}^m \frac{(\lambda t_i)^\beta e^{(\lambda t_i)^\beta} \log(\lambda t_i)}{\{\alpha + e^{(\lambda t_i)^\beta}\}^2},$$

$$J_{13} = -(n-1)\beta\lambda^{-1}\sum_{i=1}^m \frac{(\lambda t_i)^\beta e^{(\lambda t_i)^\beta}}{\{\alpha + e^{(\lambda t_i)^\beta}\}^2},$$

$$J_{22} = m\beta^{-2} - \beta\lambda^{-1} \sum_{i=1}^m (\lambda t_i)^\beta \log(\lambda t_i) \\ - (n-1)\alpha\beta\lambda^{-1} \sum_{i=1}^m \frac{(\lambda t_i)^\beta \left\{ \alpha + e^{(\lambda t_i)^\beta} - (\lambda t_i)^\beta e^{(\lambda t_i)^\beta} \right\} \log(\lambda t_i)}{\left\{ \alpha + e^{(\lambda t_i)^\beta} \right\}^2},$$

$$J_{23} = m\lambda^{-1} - \lambda^{-1} \sum_{i=1}^m (\lambda t_i)^\beta \{1 + \beta \log(\lambda t_i)\} \\ - (n-1)\alpha\lambda^{-1} \sum_{i=1}^m \frac{(\lambda t_i)^\beta \{1 + \beta \log(\lambda t_i)\}}{\alpha + e^{(\lambda t_i)^\beta}} \\ + (n-1)\alpha\beta\lambda^{-1} \sum_{i=1}^m \frac{(\lambda t_i)^{2\beta} e^{(\lambda t_i)^\beta} \log(\lambda t_i)}{\left\{ \alpha + e^{(\lambda t_i)^\beta} \right\}^2},$$

$$J_{33} = m(2-\beta)\lambda^{-2} - \beta(\beta-1)\lambda^{-2} \sum_{i=1}^m (\lambda x_i)^\beta \left\{ 1 + \frac{(n-1)\alpha}{\alpha + e^{(\lambda x_i)^\beta}} \right\} \\ - (n-1)\alpha\beta^2\lambda^{-2} \sum_{i=1}^m \frac{(\lambda x_i)^{2\beta} e^{(\lambda x_i)^\beta}}{\left\{ \alpha + e^{(\lambda x_i)^\beta} \right\}^2}.$$

### 3.4 Εφαρμογές

Η προσαρμογή της προτεινόμενης WB κατανομής ελέγχεται σε τρία σύνολα πραγματικών δεδομένων. Το πρώτο σύνολο δεδομένων περιλαμβάνει αποτελέσματα από μια δοκιμή σχετικά με τη διάρκεια ζωής 18 ηλεκτρονικών συσκευών (βλέπε Wang, 2000). Τα δεδομένα χωρίστηκαν και η εκτίμηση με διάφορες μεθόδους της συνάρτησης κινδύνου έδειξε σαφώς κυρτή μορφή. Το δεύτερο σύνολο δεδομένων αποτελείται από 24 παρατηρήσεις που αφορούν το χρονικό διάστημα μεταξύ διαδοχικών σεισμών που έλαβαν χώρα τον περασμένο αιώνα στο σεισμικό ρήγμα της Βόρειας Ανατολίας (βλέπε Kus, 2007, Tahmasbi and Rezaei, 2008 και Chahkandi and Ganjali, 2009). Τέλος, στο τελευταίο

παράδειγμα, συμπεριλαμβάνονται δεδομένα που αντιπροσωπεύουν το χρόνο ζωής 20 ηλεκτρικών εξαρτημάτων (βλέπε Zimmer et al., 1998). Εκτός από την προτεινόμενη WB κατανομή, άλλη μια κατανομή με τρεις παραμέτρους προσαρμόζεται στα τρία σύνολα δεδομένων ούτως ώστε να υπάρξει και μια άμεση αντιπαράθεση των δύο κατανομών μεταξύ τους. Αυτή είναι η εναλλακτική κατανομή που προτάθηκε από τους Dimitrakoulou et al. (2007), η οποία έχει σ.π.π. που δίνεται από τη σχέση

$$g(t; \theta) = \alpha \beta \lambda t^{\beta-1} (1 + \lambda t^\beta)^{\alpha-1} e^{-1-(1+\lambda t^\beta)^\alpha}, t, \alpha, \beta, \lambda > 0,$$

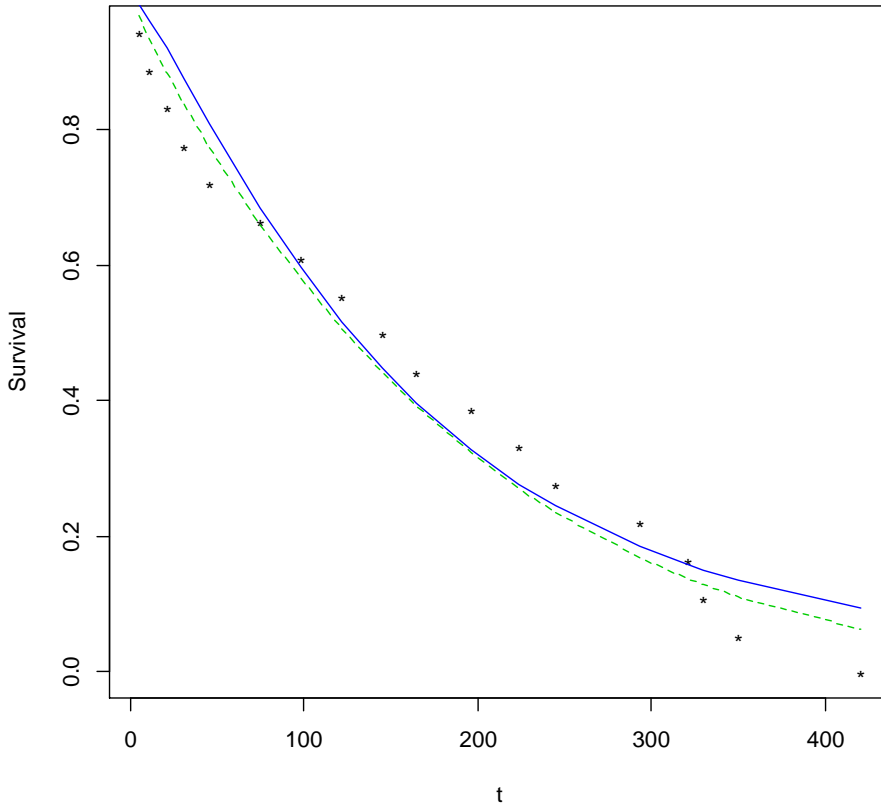
όπου  $\theta = (\alpha, \beta, \lambda)$  και για χάριν συντομίας θα συμβολίζεται ως DAL. Υπολογίζοντας τους εκτιμητές μέγιστης πιθανοφάνειας, χρησιμοποιήθηκε το τεστ καλής προσαρμογής των Kolmogorov-Smirnov (K-S), για τον έλεγχο της προσαρμοστικότητας κάθε μιας από τις δύο κατανομές στα τρία σύνολα δεδομένων. Οι εκτιμητές, ο λογάριθμος της συνάρτησης πιθανοφάνειας (LL), το κριτήριο AIC, η τιμή του στατιστικού του K-S τεστ καθώς και η αντίστοιχη κρίσιμη τιμή (p-value) περιέχονται στον Πίνακα 3.1.

**Πίνακας 3.1** Εκτιμητές παραμέτρων, τιμές λογαρίθμου συνάρτησης πιθανοφάνειας, κριτηρίου AIC και στατιστικού K-S, καθώς και κρίσιμες τιμές από την προσαρμογή και των δύο κατανομών πιθανότητας στα τρία διαθέσιμα σύνολα δεδομένων.

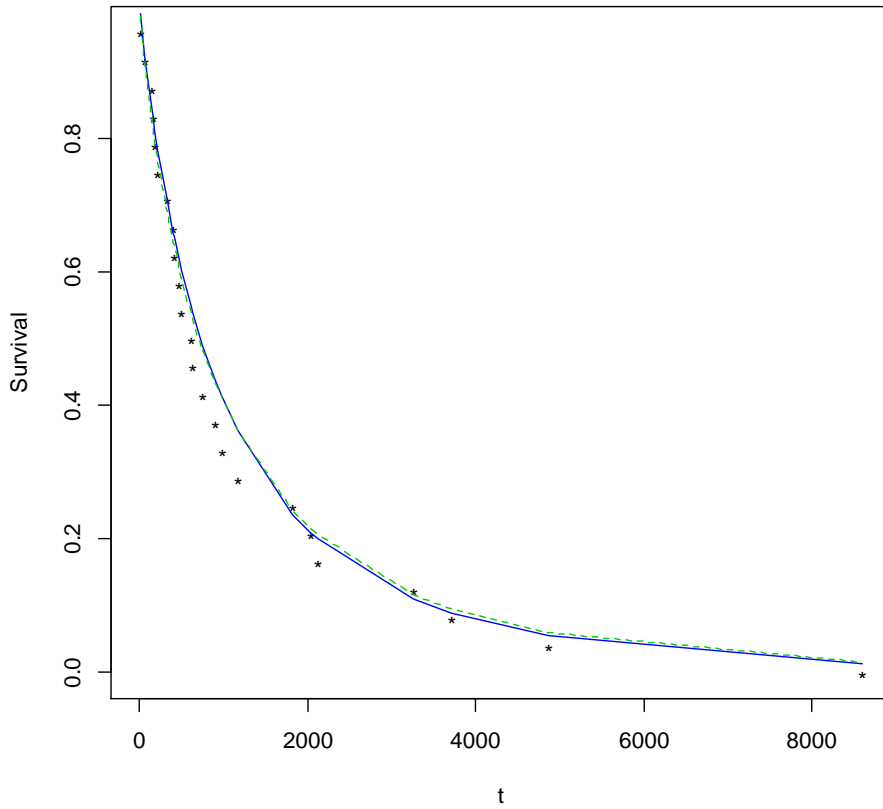
Σύνολο Δεδομένων	Κατανομή	$\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$	LL	AIC	K-S (p-value)
1) $n = 18$	WB	(0.2361, 1.2145, 0.0022)	-111.2438	228.4876	0.1165 (0.9441)
	DAL	(2.4923, 0.8977, 0.0031)	-109.934	225.868	0.121 (0.9266)
2) $n = 24$	WB	(0.0961, 0.9131, 0.0004)	-196.6479	399.2958	0.0804 (0.9942)
	DAL	(0.6428, 0.8899, 0.0037)	-196.7304	399.4608	0.0876 (0.985)
3) $n = 20$	WB	(0.0535, 1.2106, 0.0512)	-73.4886	152.9772	0.0548 (1)
	DAL	(22.25, 0.807, 0.0032)	-74.0409	154.0818	0.0873 (0.9945)

Γίνεται φανερό ότι, για όλα τα σύνολα δεδομένων, η προτεινόμενη WB κατανομή λειτουργεί επαρκώς ανταγωνιστικά απέναντι στην εναλλακτική DAL κατανομή, κάτι το οποίο ενισχύεται και από τη γραφική παράσταση των εμπειρικών και προσαρμοζόμενων συναρτήσεων επιβίωσης (Γράφημα 3.3).

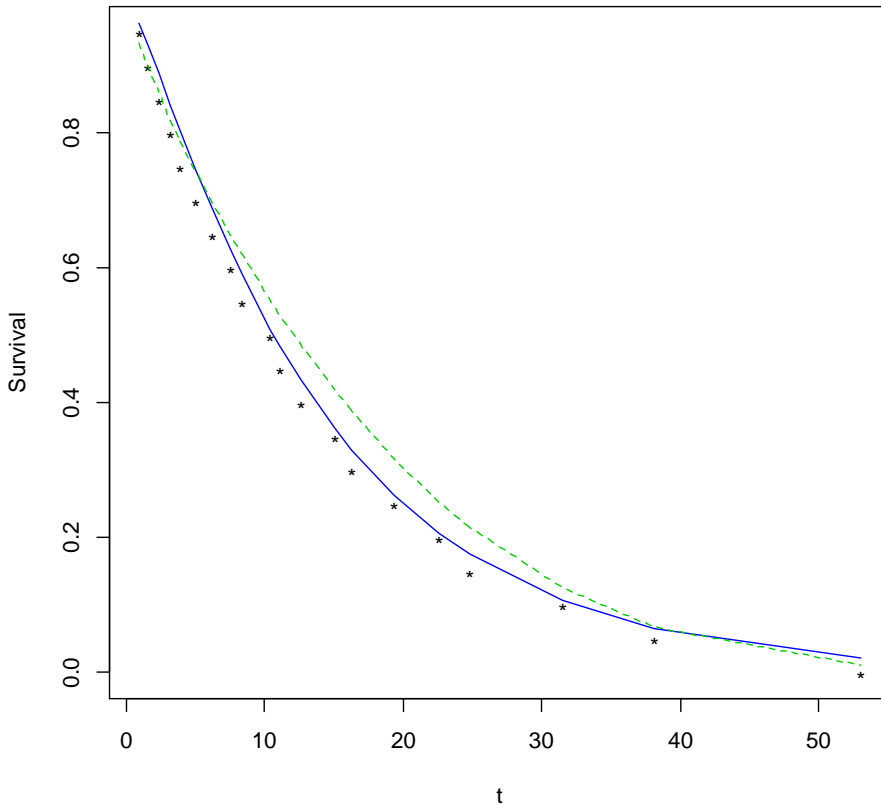
**Γράφημα 3.3** Γραφικές παραστάσεις της συνάρτησης επιβίωσης της εμπειρικής κατανομής (σημεία \*), της WB κατανομής (μπλε συνεχής γραμμή) και της DAL (πράσινη διακεκομμένη γραμμή) για τα τρία σύνολα δεδομένων.



(α) Ηλεκτρονικές Συσκευές



(β) Μεσοδιαστήματα Σεισμών



(γ) Ηλεκτρικά Εξαρτήματα

### 3.5 Συμπεράσματα

Η προτεινόμενη WB κατανομή αποτελεί γενίκευση της EB κατανομής που προτάθηκε από τους Chahkandi and Ganjali (2009). Έχει διευρυμένη δυνατότητα λήψης διαφόρων μορφών συνάρτησης κινδύνου, έναντι της αρχικής EB κατανομής, και αυτό την καθιστά περισσότερο ευέλικτη στην περιγραφή πραγματικών καταστάσεων. Επιπλέον, λειτουργεί αν όχι καλύτερα, τουλάχιστον ανταγωνιστικά απέναντι σε άλλες τρι-παραμετρικές κατανομές, όσον αφορά την προσαρμογή της σε πραγματικά δεδομένα.



# Κεφάλαιο 4

## Η Γενικευμένη Κατανομή EXTG

Κατά τη διάρκεια των τελευταίων δεκαετιών, ένα σημαντικό μέρος της έρευνας είναι αφιερωμένο στη δημιουργία κατανομών πιθανότητας με διαφόρων μορφών συνάρτηση κινδύνου. Προφανώς, το κίνητρο για την τάση αυτή, είναι να δοθεί μεγαλύτερη ελευθερία επιλογής και βελτιστοποίηση της περιγραφής δεδομένων που συχνά απορρέουν από σύνθετες πραγματικές καταστάσεις.

Σε αυτό το κεφάλαιο εισάγεται μια γενικευμένη οικογένεια κατανομών, με την προσθήκη μιας επιπλέον παραμέτρου σε μια γενική κατανομή πιθανότητας, σε αντιστοιχία με τους Marshall and Olkin (1997). Στη συνέχεια, η οικογένεια κατανομών χρησιμοποιείται για την ανάπτυξη μιας τετρα-παραμετρικής τροποποιημένης επέκτασης της κατανομής Weibull, με συνάρτηση κινδύνου διαφόρων μορφών, που συναγωνίζεται επαρκώς άλλες κατανομές στην προσαρμογή σε πραγματικά δεδομένα.

Το κεφάλαιο είναι οργανωμένο ως εξής: Η ενότητα 4.1 περιλαμβάνει τη γενικευμένη οικογένεια κατανομών και ορισμένες ιδιότητες αυτής. Η προτεινόμενη τετρα-παραμετρική κατανομή εισάγεται στην ενότητα 4.2, μαζί με τη μελέτη των ιδιοτήτων της και θεμάτων αξιοπιστίας. Οι παράμετροι της κατανομής εκτιμώνται με τη μέθοδο μέγιστης πιθανοφάνειας και υπολογίζεται ο παρατηρούμενος πίνακας πληροφορίας. Τέλος, ελέγχεται η προσαρμογή της προτεινόμενης κατανομής έναντι άλλης εναλλακτικής κατανομής, σε δύο πραγματικά σύνολα δεδομένων.

### 4.1 Η γενικευμένη οικογένεια κατανομών

Η γενίκευση μιας κατανομής πιθανότητας μπορεί να επιτευχθεί με την προσθήκη μιας παραμέτρου μορφής, σε αντιστοιχία με τους Marshall and Olkin (1997). Έτσι, έχοντας μια κατανομή με συνάρτηση επιβίωσης

$S_0$ , η συνάρτηση επιβίωσης της προτεινόμενης οικογένειας κατανομών με την επιπρόσθετη παράμετρο  $p$  δίνεται από τον τύπο:

$$S(t) = \frac{\ln \{1 - (1-p)S_0(t)\}}{\ln p}, t, p \in \mathbb{R}_+, \quad (4.1)$$

όπου, όταν  $p \rightarrow 1$ , τότε  $S \rightarrow S_0$ . Οι συναρτήσεις πυκνότητας πιθανότητας και κινδύνου υπολογίζονται άμεσα και δίνονται από τις σχέσεις:

$$f(t) = \frac{(p-1)f_0(t)}{\{1 - (1-p)S_0(t)\} \ln p}$$

και

$$h(t) = \frac{(p-1)S_0(t)h_0(t)}{\{1 - (1-p)S_0(t)\} \ln \{1 - (1-p)S_0(t)\}}, \quad (4.2)$$

αντίστοιχα, όπου  $f_0$  και  $h_0$  είναι οι συναρτήσεις πυκνότητας πιθανότητας και κινδύνου που αντιστοιχούν στην αρχική κατανομή με συνάρτηση επιβίωσης  $S_0$ , και  $t, p \in \mathbb{R}_+$ . Εφόσον,

$$\lim_{t \rightarrow 0^+} h(t) = \frac{p-1}{p \ln p} \lim_{t \rightarrow 0^+} h_0(t) \quad \text{και} \quad \lim_{t \rightarrow +\infty} h(t) = \lim_{t \rightarrow +\infty} h_0(t),$$

από τη (4.2) έπεται ότι,

$$\frac{p-1}{p \ln p} h_0(t) \leq h(t) \leq h_0(t), t \in \mathbb{R}_+, p \geq 1,$$

και

$$h_0(t) \leq h(t) \leq \frac{p-1}{p \ln p} h_0(t), t \in \mathbb{R}_+, p \in (0, 1].$$

Επομένως, η ποσότητα  $h(t)/h_0(t)$  με  $t \in \mathbb{R}_+$  είναι αύξουσα για  $p \geq 1$  και φθίνουσα για  $p \in (0,1]$ . Όταν  $S_0(0)=1$ , η συνάρτηση κινδύνου στην αρχική της τιμή,  $h(0)$ , συμπεριφέρεται διαφορετικά από τις αντίστοιχες συναρτήσεις των Weibull και γάμμα κατανομών. Για τις δύο αυτές οικογένειες, η κατανομή μπορεί να είναι εκθετική με  $h(0)=0$ , ή  $h(0)=+\infty$ , έτσι ώστε η  $h(0)$  να είναι μη συνεχής στην παράμετρο μορφής. Αυτό δεν ισχύει στην περίπτωση συναρτήσεων κινδύνου της μορφής (4.2), και επομένως η προτεινόμενη οικογένεια κατανομών μπορεί να φανεί χρήσιμη στη βελτιστοποίηση της προσαρμογής της κατανομής με συνάρτηση επιβίωσης  $S_0$ .

## 4.2 Μια επεκτεταμένη κατανομή Weibull

Η συνάρτηση επιβίωσης της τροποποιημένης επέκτασης της κατανομής Weibull, που προτάθηκε από τους Xie *et al.* (2002) και μελετήθηκε επιπρόσθετα από τους Tang *et al.* (2003), δίνεται από τη σχέση:

$$S_0(t) = e^{-\alpha \lambda \left( e^{(t/\alpha)^\beta} - 1 \right)}, \quad (4.3)$$

για  $t, \alpha, \beta, \lambda \in \mathbb{R}_+$ . Στη συνέχεια θα αναφέρεται ως XTG κατανομή, η κατανομή με συνάρτηση επιβίωσης της μορφής (4.3). Εισάγωντας την (4.3) στη σχέση (4.1), προκύπτει η συνάρτηση επιβίωσης της προτεινόμενης τετρα-παραμετρικής κατανομής:

$$S(t; \theta) = \frac{\ln \{ 1 - (1-p)e^{-\alpha \lambda \left( e^{(t/\alpha)^\beta} - 1 \right)} \}}{\ln p}, \quad (4.4)$$

όπου  $\theta = (\alpha, \beta, \lambda, p)$  και  $t, \alpha, \beta, \lambda, p \in \mathbb{R}_+$ . Από την (4.2) ή την (4.4), μπορεί άμεσα να υπολογιστεί η σ.π.π., η οποία δίνεται από τον τύπο:

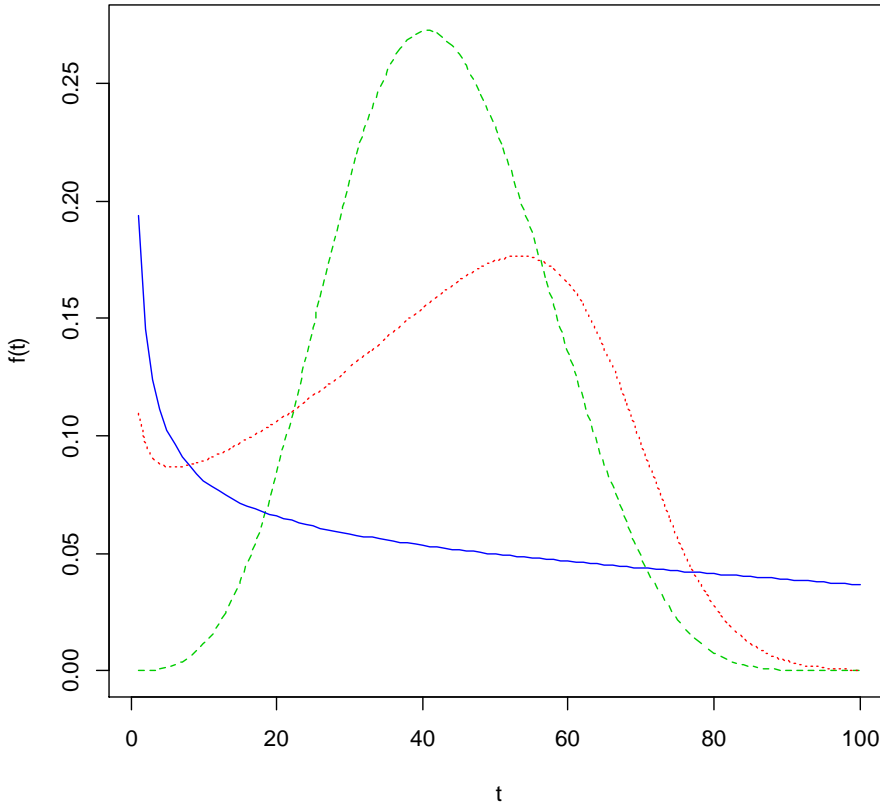
$$f(t; \theta) = \frac{\beta \lambda (p-1) \left(\frac{t}{\alpha}\right)^{\beta-1} e^{\left(\frac{t}{\alpha}\right)^\beta - \alpha \lambda \left(e^{(t/\alpha)^\beta} - 1\right)}}{\left\{1 - (1-p)e^{-\alpha \lambda \left(e^{(t/\alpha)^\beta} - 1\right)}\right\} \ln p}, \quad (4.5)$$

όπου  $\alpha$  είναι παράμετρος κλίμακας και  $\beta, \lambda, p$  είναι παράμετροι μορφής. Ανάλογα με τις τιμές των τεσσάρων παραμέτρων, η σ.π.π. δύναται να λάβει μονότονα φθίνουσα, μονοκόρυφη ή ακόμη και τροποποιημένη κυρτή μορφή (roller-coster type). Οι διάφορες μορφές της σ.π.π. φαίνονται στο Γράφημα 4.1, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \beta, \lambda, p)$ .

Γίνεται εύκολα αντιληπτό ότι για  $p \rightarrow 1$  η προτεινόμενη κατανομή δίνει την XTG κατανομή, επομένως η προτεινόμενη κατανομή μπορεί να θεωρηθεί ως μια επέκταση της XTG (η οποία με τη σειρά της αποτελεί ασυμπτωτική επέκταση της συνήθους δι-παραμετρικής Weibull κατανομής) και εάν, επιπλέον,  $\alpha = 1$  τότε η (4.5) ορίζει την κατανομή του Chen (2000). Στη συνέχεια η προτεινόμενη επεκτεταμένη XTG κατανομή θα αναφέρεται ως EXTG.

Επιπρόσθετα, μπορεί να αποδειχθεί ότι για  $p \in (0, 1)$  η (4.5) αποτελεί μίξη λογαριθμικής και XTG κατανομής. Θεωρούμε τους χρόνους ζωής  $T = \min(T_1, T_2, \dots, T_Z)$  ενός συστήματος  $Z$  όμοιων εξαρτημάτων σε σειρά, όπου η αποτυχία συμβαίνει εάν τουλάχιστον ένα από τα εν σειρά εξαρτήματα πάψει να λειτουργεί.

**Γράφημα 4.1** Συναρτήσεις πυκνότητας πιθανότητας της EXTG κατανομής για  $(\alpha, \beta, \lambda, p) = (2, 0.5, 0.03, 0.2)$  (μπλε συνεχής γραμμή),  $(\alpha, \beta, \lambda, p) = (12, 4, 2, 0.2)$  (πράσινη διακεκομμένη γραμμή) και  $(\alpha, \beta, \lambda, p) = (2, 0.7, 0.3, 0.2)$  (κόκκινη γραμμή με τελείες).



Υποθέτοντας ότι οι χρόνοι ζωής των εξαρτημάτων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση επιβίωσης που δίνεται από τη σχέση (4.3) και η κατανομή του αριθμού τους  $Z$  είναι η λογαριθμική, ανεξάρτητη των  $T_i$ , με συνάρτηση πιθανότητας:

$$p(z; p) = -\frac{(1-p)^z}{z \ln p},$$

για  $z \in \mathbb{N} - \{0\}$ ,  $p \in (0,1)$ , τότε η κατανομή του  $T | Z = z$  έχει σ.π.π.:

$$f(t | z; \beta, \lambda) = \beta \lambda z \left( \frac{t}{\alpha} \right)^{\beta-1} e^{\left( \frac{t}{\alpha} \right)^\beta - \alpha \lambda \left( e^{(t/\alpha)^\beta} - 1 \right)},$$

για  $t, \alpha, \beta, \lambda \in \mathbb{R}_+$  και προκύπτει εύκολα ότι η κατανομή του  $T$  είναι η EXTG κατανομή με σ.π.π. που δίνεται από τη σχέση (4.5).

Ο υπολογισμός των  $r$ -τάξεως ροπών απαιτεί χρήση τυπικών διαδικασιών αριθμητικής ολοκλήρωσης, που είναι διαθέσιμες στην πλειοψηφία των μαθηματικών υπολογιστικών πακέτων. Για  $p \in (0,1)$  ωστόσο, μπορούν να εκφραστούν από τη σχέση:

$$E(T^r) = -\frac{1}{\ln p} \sum_{k=1}^{+\infty} \frac{(1-p)^k}{k} \int_0^{\infty} \alpha^r \left\{ \ln \left( \frac{t}{\alpha \kappa \lambda} + 1 \right) \right\}^{r/\beta} e^{-t} dt.$$

Επιπλέον, γνωρίζοντας ότι  $F(x; \theta) = 1 - S(x; \theta)$ , με αντιστροφή της σχέσης (4.4) προκύπτει η συνάρτηση υπολογισμού των εκατοστιαίων σημείων, η οποία δίνεται από τη σχέση:

$$F^{-1}(q) = \alpha \left[ \ln \left\{ 1 - \frac{1}{\alpha \lambda} \ln \left( \frac{1 - p^{1-q}}{1 - p} \right) \right\} \right]^{1/\beta},$$

για  $p \in (0,1)$ . Επομένως, η διάμεσος θα είναι

$$M = \alpha \left[ \ln \left\{ 1 + \frac{1}{\alpha \lambda} \ln(1 + \sqrt{p}) \right\} \right]^{1/\beta}.$$

#### 4.2.1 Συναρτήσεις κινδύνου και μέσου υπολειπόμενου χρόνου ζωής

Από τις σχέσεις (4.4) και (4.5) προκύπτει η συνάρτηση κινδύνου της EXTG κατανομής:

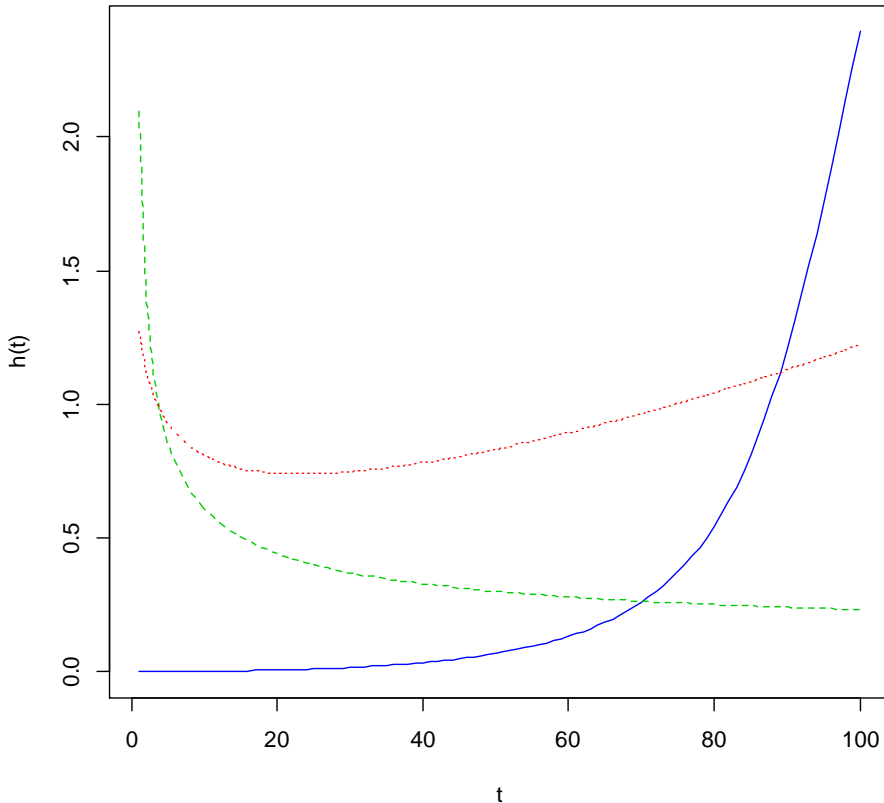
$$h(t; \theta) = \frac{\beta\lambda(p-1)\left(\frac{t}{\alpha}\right)^{\beta-1} e^{\left(\frac{t}{\alpha}\right)^\beta - \alpha\lambda\left(e^{\left(\frac{t}{\alpha}\right)^\beta} - 1\right)}}{\left\{1 - (1-p)e^{-\alpha\lambda\left(e^{\left(\frac{t}{\alpha}\right)^\beta} - 1\right)}\right\} \ln\left\{1 - (1-p)e^{-\alpha\lambda\left(e^{\left(\frac{t}{\alpha}\right)^\beta} - 1\right)}\right\}}.$$

Με αντίστοιχη μεθοδολογία με αυτή που εφαρμόζεται αναλυτικά στο πέμπτο κεφάλαιο, μπορεί να αποδειχθεί ότι για  $\beta \geq 1$ , η συνάρτηση κινδύνου της κατανομής EXTG είναι γνησίως αύξουσα (IFR) (Glaser, 1980). Σε διαφορετική περίπτωση, δηλαδή για  $\beta < 1$ , η συνάρτηση κινδύνου μπορεί να είναι γνησίως αύξουσα, φθίνουσα, ή ακόμη να λάβει και κυπελλοειδή μορφή. Οι διάφορες μορφές της συνάρτησης κινδύνου της EXTG κατανομής φαίνονται στο Γράφημα 4.2, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \beta, \lambda, p)$ .

Δεδομένου ότι δεν υπήρξε αποτυχία προγενέστερη του  $t_0$ , ο υπολειπόμενος χρόνος ζωής ορίζεται ως το χρονικό διάστημα από το  $t_0$  μέχρι τη στιγμή της αποτυχίας. Η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής της κατανομής EXTG, για  $p \in (0,1)$ , δίνεται από τη σχέση:

$$\begin{aligned} m(t_0; \theta) &= E(T - t_0 | T \geq t_0) = \int_{t_0}^{+\infty} \frac{(t - t_0) f(t; \theta)}{1 - F(t_0; \theta)} dt \\ &= \frac{1}{\ln\left\{1 - (1-p)e^{-\alpha\lambda\left(e^{\left(\frac{t_0}{\alpha}\right)^\beta} - 1\right)}\right\}} \sum_{k=1}^{+\infty} \frac{(1-p)^k}{k} \\ &\quad \times \int_{\alpha k \lambda \left(e^{\left(\frac{t_0}{\alpha}\right)^\beta} - 1\right)}^{+\infty} \left[ \alpha \left\{ \ln\left(\frac{t}{\alpha k \lambda} + 1\right) \right\}^{1/\beta} - t_0 \right] e^{-t} dt. \end{aligned}$$

**Γράφημα 4.2** Συναρτήσεις κινδύνου της EXTG κατανομής για  $(\alpha, \beta, \lambda, p) = (0.05, 0.45, 0.002, 20)$  (μπλε συνεχής γραμμή),  $(\alpha, \beta, \lambda, p) = (0.2, 0.2, 3, 2)$  (πράσινη διακεκομμένη γραμμή) και  $(\alpha, \beta, \lambda, p) = (10, 0.9, 0.5, 0.3)$  (κόκκινη γραμμή με τελείες).



#### 4.2.2 Στατιστική συμπερασματολογία

Θεωρώντας ένα τυχαίο δείγμα  $n$  παρατηρήσεων  $y_{obs} = (t_i; i=1, \dots, n)$  από την (4.5), ο λογάριθμος της συνάρτησης πιθανοφάνειας θα δίνεται από τη σχέση:



$$l(\theta; y_{obs}) = n[\ln\{(p-1)\beta\lambda\} - \ln(\ln p) + \alpha\lambda + (1-\beta)\ln\alpha] + (\beta-1)\sum_{i=1}^n \ln t_i \\ + \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta - \alpha\lambda \sum_{i=1}^n e^{\left(\frac{t_i}{\alpha}\right)^\beta} - \sum_{i=1}^n \ln \left\{ 1 - (1-p)e^{-\alpha\lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)} \right\}.$$

Παραγωγίζοντας ως προς τις παραμέτρους  $\theta = (\alpha, \beta, \lambda, p)$  προκύπτουν οι σχέσεις:

$$\frac{\partial l}{\partial \alpha} = n\lambda - n\alpha^{-1}(\beta-1) - \alpha^{-1}\beta \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \\ - \lambda \sum_{i=1}^n e^{\left(\frac{t_i}{\alpha}\right)^\beta} \left\{ 1 + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} \tag{4.6} \\ - \lambda(1-p) \sum_{i=1}^n \frac{\left[ e^{\left(\frac{t_i}{\alpha}\right)^\beta} \left\{ 1 + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} - 1 \right] e^{-\alpha\lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}{1 - (1-p)e^{-\alpha\lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}},$$

$$\frac{\partial l}{\partial \beta} = n\beta^{-1} + \sum_{i=1}^n \left\{ 1 + \left(\frac{t_i}{\alpha}\right)^\beta - \alpha\lambda \left(\frac{t_i}{\alpha}\right)^\beta e^{\left(\frac{t_i}{\alpha}\right)^\beta} \right\} \ln \left(\frac{t_i}{\alpha}\right) \\ - \alpha\lambda(1-p) \sum_{i=1}^n \frac{\left(\frac{t_i}{\alpha}\right)^\beta \ln \left(\frac{t_i}{\alpha}\right) e^{\left(\frac{t_i}{\alpha}\right)^\beta} e^{-\alpha\lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}{1 - (1-p)e^{-\alpha\lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}, \tag{4.7}$$

$$\frac{\partial l}{\partial \lambda} = n(\alpha + \lambda^{-1}) - \alpha \sum_{i=1}^n e^{\left(\frac{t_i}{\alpha}\right)^\beta} - \alpha(1-p) \sum_{i=1}^n \frac{\left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right) e^{-\alpha \lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}{1 - (1-p)e^{-\alpha \lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}, \quad (4.8)$$

$$\frac{\partial l}{\partial p} = n\left\{(p-1)^{-1} + (p \ln p)^{-1}\right\} - \sum_{i=1}^n \frac{e^{-\alpha \lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}{1 - (1-p)e^{-\alpha \lambda \left(e^{\left(\frac{t_i}{\alpha}\right)^\beta} - 1\right)}}. \quad (4.9)$$

Οι εκτιμητές μέγιστης πιθανοφάνειας  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{p})$  των παραμέτρων  $\theta = (\alpha, \beta, \lambda, p)$  προκύπτουν από την επίλυση του συστήματος των τεσσάρων μη-γραμμικών κανονικών εξισώσεων:  $\frac{\partial l}{\partial \alpha} = 0$ ,  $\frac{\partial l}{\partial \beta} = 0$ ,  $\frac{\partial l}{\partial \lambda} = 0$  και  $\frac{\partial l}{\partial p} = 0$ , η οποία προφανώς δε μπορεί να γίνει με αναλυτικές μεθόδους. Κατά συνέπεια, για την απόκτηση των εκτιμητών μέγιστης πιθανοφάνειας θα πρέπει να χρησιμοποιηθεί κάποια από τις μεθόδους αριθμητικής ανάλυσης.

Σύμφωνα με γνωστές ιδιότητες των εκτιμητών μέγιστης πιθανοφάνειας, η ασυμπτωτική κατανομή του  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{p})$  είναι κατά προσέγγιση η τετραδιάστατη κανονική με μέση τιμή  $\theta = (\alpha, \beta, \lambda, p)$  και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του αναμενόμενου πίνακα πληροφορίας του Fisher  $I(\theta)$ , με στοιχεία:

$$I_{ij} = E\left(-\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}\right), i, j = 1, 2, 3, 4.$$

Η ασυμπτωτική κανονικότητα εξακολουθεί να ισχύει αν για λόγους υπολογιστικής διευκόλυνσης ο  $I(\theta)$  αντικατασταθεί από τον παρατηρούμενο πίνακα πληροφορίας  $J(\theta)$ , με στοιχεία:

$$J_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}, i, j = 1, 2, 3, 4,$$

εκτιμώμενα στο  $\theta = \hat{\theta}$ , αφού ο τελευταίος πίνακας κάτω από ασθενείς συνθήκες αποτελεί συνεπή εκτιμητή του πρώτου (βλέπε Lawless, 2003). Τα άνω τριγωνικά στοιχεία του συμμετρικού παρατηρούμενου πίνακα πληροφορίας, δίνονται αναλυτικά παραγωγίζοντας τις σχέσεις (4.6)-(4.9):

$$J_{11} = -n\alpha^{-2}(\beta-1) - \alpha^{-1}\beta \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \left[ \alpha^{-1}(\beta+1) - \lambda e^{(t_i/\alpha)^\beta} \right. \\ \times \left. \left\{ \beta \left(\frac{t_i}{\alpha}\right)^\beta + \beta + 1 \right\} \right] + \lambda(1-p) \sum_{i=1}^n \frac{e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)}}{1 - (1-p)e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)}} \\ \times \left[ e^{(t_i/\alpha)^\beta} \left[ \lambda \left[ e^{(t_i/\alpha)^\beta} \left\{ \beta^2 \left(\frac{t_i}{\alpha}\right)^{2\beta} - 1 \right\} + 2 \right] - \alpha^{-1}\beta \left(\frac{t_i}{\alpha}\right)^\beta \right. \right. \right. \\ \times \left. \left. \left\{ \beta \left(\frac{t_i}{\alpha}\right)^\beta + \beta + 1 \right\} \right] \right. \\ \left. \left. - \frac{\lambda(1-p)e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)} \left[ e^{(t_i/\alpha)^\beta} \left\{ 1 + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} - 1 \right]^2}{1 - (1-p)e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)}} \right] \right],$$

$$\begin{aligned}
J_{12} = & n\alpha^{-1} + \alpha^{-1} \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \left[ \left\{ 1 + \beta \ln\left(\frac{t_i}{\alpha}\right) \right\} + \lambda e^{\left(\frac{t_i}{\alpha}\right)^\beta} \right. \\
& \times \left. \left\{ 1 + (1 + \beta) \ln\left(\frac{t_i}{\alpha}\right) + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} \right] + \lambda(1-p) \\
& \times \sum_{i=1}^n \frac{\left(\frac{t_i}{\alpha}\right)^\beta e^{\left(\frac{t_i}{\alpha}\right)^\beta - \alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)}}{1 - (1-p)e^{-\alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)}} \left\{ 1 + (1 + \beta) \ln\left(\frac{t_i}{\alpha}\right) + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right. \\
& \left. - \alpha\lambda \ln\left(\frac{t_i}{\alpha}\right) \left[ e^{\left(\frac{t_i}{\alpha}\right)^\beta} \left\{ 1 + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} - 1 \right] \right. \\
& \left. \times \left( 1 + \frac{(1-p)e^{-\alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)}}{1 - (1-p)e^{-\alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)}} \right) \right\},
\end{aligned}$$

$$\begin{aligned}
J_{13} = & -n + \sum_{i=1}^n e^{\left(t_i/\alpha\right)^\beta} \left\{ 1 + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} \\
& + (1-p) \sum_{i=1}^n \frac{e^{-\alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)} \left[ e^{\left(t_i/\alpha\right)^\beta} \left\{ 1 + \beta \left(\frac{t_i}{\alpha}\right)^\beta \right\} - 1 \right]}{\left\{ 1 - (1-p)e^{-\alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)} \right\}^2} \\
& \times \left\{ 1 - (1-p)e^{-\alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right)} - \alpha\lambda \left(e^{\left(t_i/\alpha\right)^\beta} - 1\right) \right\},
\end{aligned}$$

$$J_{14} = - \sum_{i=1}^n \frac{e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)}}{\left\{ 1 - (1-p)e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)} \right\}^2},$$

$$J_{22} = n\beta^{-2} - \sum_{i=1}^n \left( \frac{t_i}{\alpha} \right)^\beta \left[ 1 - \alpha \lambda \left\{ 1 + \left( \frac{t_i}{\alpha} \right)^\beta \right\} e^{\left( \frac{t_i}{\alpha} \right)^\beta} \right] \ln^2 \left( \frac{t_i}{\alpha} \right) \\ + \alpha \lambda (1-p) \sum_{i=1}^n \frac{\left( \frac{t_i}{\alpha} \right)^\beta \ln^2 \left( \frac{t_i}{\alpha} \right) e^{\left( \frac{t_i}{\alpha} \right)^\beta - \alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)}}{\left\{ 1 - (1-p)e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)} \right\}^2} \\ \times \left[ 1 + \left( \frac{t_i}{\alpha} \right)^\beta - \alpha \lambda \left( \frac{t_i}{\alpha} \right)^\beta e^{\left( \frac{t_i}{\alpha} \right)^\beta} \right. \\ \left. - (1-p) \left\{ 1 + \left( \frac{t_i}{\alpha} \right)^\beta \right\} e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)} \right],$$

$$\begin{aligned}
J_{23} &= \alpha \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta e^{\left(\frac{t_i}{\alpha}\right)^\beta} \ln\left(\frac{t_i}{\alpha}\right) \\
&\quad + \alpha(1-p) \sum_{i=1}^n \frac{\left(\frac{t_i}{\alpha}\right)^\beta \ln\left(\frac{t_i}{\alpha}\right) e^{\left(\frac{t_i}{\alpha}\right)^\beta - \alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}}{\left\{1 - (1-p)e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}\right\}^2} \\
&\quad \times \left\{1 - \alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right) - (1-p)e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}\right\}, \\
J_{24} &= -\alpha \lambda \sum_{i=1}^n \frac{\left(\frac{t_i}{\alpha}\right)^\beta \ln\left(\frac{t_i}{\alpha}\right)^\beta e^{\left(\frac{t_i}{\alpha}\right)^\beta - \alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}}{\left\{1 - (1-p)e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}\right\}^2}, \\
J_{33} &= n\lambda^{-2} - \alpha^2(1-p) \sum_{i=1}^n \frac{\left(e^{(t_i/\alpha)^\beta} - 1\right)^2 e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}}{\left\{1 - (1-p)e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}\right\}^2}, \\
J_{34} &= -\alpha \sum_{i=1}^n \frac{\left(e^{(t_i/\alpha)^\beta} - 1\right) e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}}{\left\{1 - (1-p)e^{-\alpha \lambda \left(e^{(t_i/\alpha)^\beta} - 1\right)}\right\}^2},
\end{aligned}$$

$$J_{44} = n \left\{ (p-1)^{-2} - (1 + \ln p)(p \ln p)^{-2} \right\} - \sum_{i=1}^n \frac{e^{-2\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)}}{\left\{ 1 - (1-p)e^{-\alpha \lambda \left( e^{(t_i/\alpha)^\beta} - 1 \right)} \right\}^2}.$$

### 4.2.3 Εφαρμογές

Η προσαρμογή της προτεινόμενης EXTG κατανομής ελέγχεται σε δύο σύνολα πραγματικών δεδομένων. Το πρώτο σύνολο δεδομένων περιλαμβάνει αποτελέσματα από μια δοκιμή σχετικά με τη διάρκεια ζωής 50 ηλεκτρικών εξαρτημάτων (Aarset, 1987). Το δεύτερο σύνολο δεδομένων αποτελείται από 44 παρατηρήσεις που αφορούν το χρόνο ζωής (σε ώρες) όλων των υποσυστημάτων ενός αυτοκινήτου, δηλ. κινητήρα, υδραυλικό υποσύστημα και σύστημα κλιματισμού, φρένα, σύστημα μετάδοσης, λάστιχα, τροχούς και αμάξωμα (Kumar et al., 1989 και Pulcini, 2001). Τα δεδομένα των δύο περιπτώσεων ομαδοποιήθηκαν και εκτιμήθηκε η εμπειρική συνάρτηση κινδύνου, υποδεικνύοντας καθαρά κυπελλοειδή μορφή. Εκτός από την προτεινόμενη EXTG κατανομή, και η XTG κατανομή προσαρμόζεται στα δύο σύνολα δεδομένων ούτως ώστε να υπάρξει και μια άμεση αντιπαράθεση των δύο κατανομών μεταξύ τους. Υπολογίζοντας τους εκτιμητές μέγιστης πιθανοφάνειας, χρησιμοποιήθηκε το τεστ καλής προσαρμογής των Kolmogorov-Smirnov (K-S), για τον έλεγχο της προσαρμοστικότητας κάθε μιας από τις δύο κατανομές στα δύο σύνολα δεδομένων. Οι εκτιμητές, ο λογάριθμος της συνάρτησης πιθανοφάνειας (LL), το κριτήριο AIC, η τιμή του στατιστικού του K-S τεστ καθώς και η αντίστοιχη κρίσιμη τιμή (p-value) περιέχονται στον παρακάτω Πίνακα 4.1. Επιπρόσθετα, οι τιμές του στατιστικού τεστ πηλίκου πιθανοφάνειας για τον έλεγχο της μηδενικής υπόθεσης  $H_0: p = 1$ , υπολογίστηκαν να είναι 8.7939 (p=0.003) και 6.8366 (p=0.0089) αντίστοιχα,

επιδεικνύοντας ότι η EXTG περιγράφει τα δύο σύνολα δεδομένων καλύτερα από την εναλλακτική XTG κατανομή. Το συμπέρασμα αυτό ενισχύεται και από τη γραφική παράσταση των εμπειρικών και προσαρμοζόμενων συναρτήσεων επιβίωσης του Γραφήματος 4.3.

**Πίνακας 4.1** Εκτιμητές παραμέτρων, τιμές λογαρίθμου συνάρτησης πιθανοφάνειας, κριτηρίου AIC, στατιστικού K-S και στατιστικού πηλίκου πιθανοφανειών, καθώς και κρίσιμες τιμές από την προσαρμογή των συγκρινόμενων κατανομών πιθανότητας στα δύο διαθέσιμα σύνολα δεδομένων.

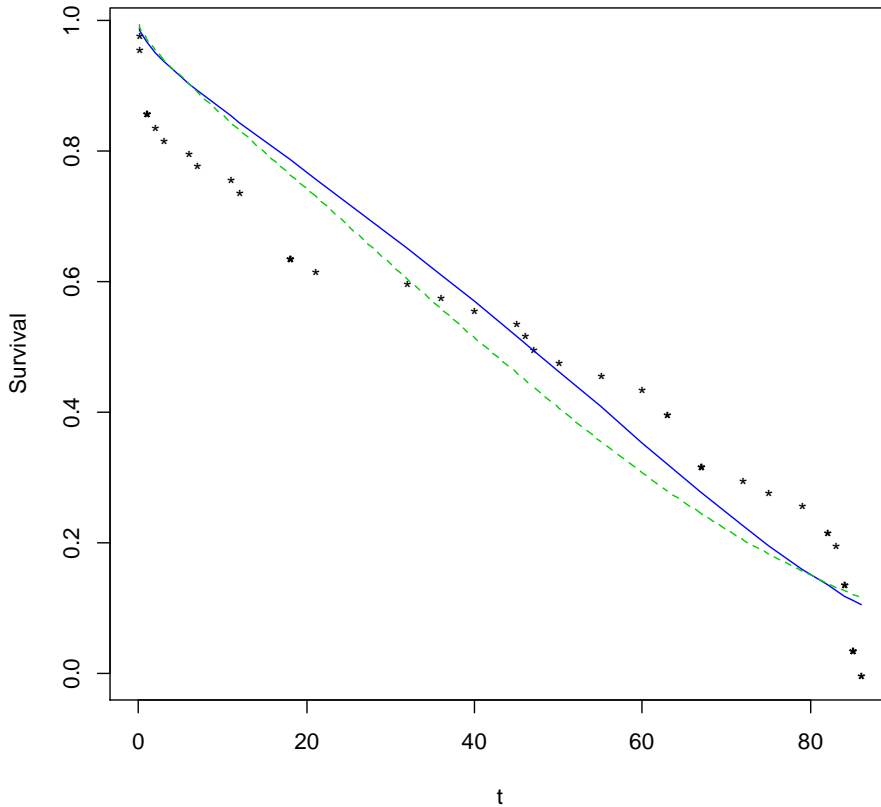
Σύνολο Δεδομένων	Κατανομή	$\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\rho})$	LL	AIC	K-S (p-value)	LL ratio test
1) $n = 50$	EXTG	(14.06, 0.4838, 0.042, 313.57)	-227.2496	462.4992	0.1470 (0.23)	$H_0: p = 1$
	XTG	(13.747, 0.588, 0.00876, -----)	-231.6466	469.2932	0.1597 (0.1562)	8.7939 (0.003)
2) $n = 44$	EXTG	(0.0831, 0.2007, 0.0211, 20.4087)	-696.0237	1400.0474	0.1397 (0.3262)	$H_0: p = 1$
	XTG	(0.0806, 0.2086, 0.0057, -----)	-702.8603	1411.7206	0.1667 (0.1544)	6.8366 (0.0089)

### 4.3 Συμπεράσματα

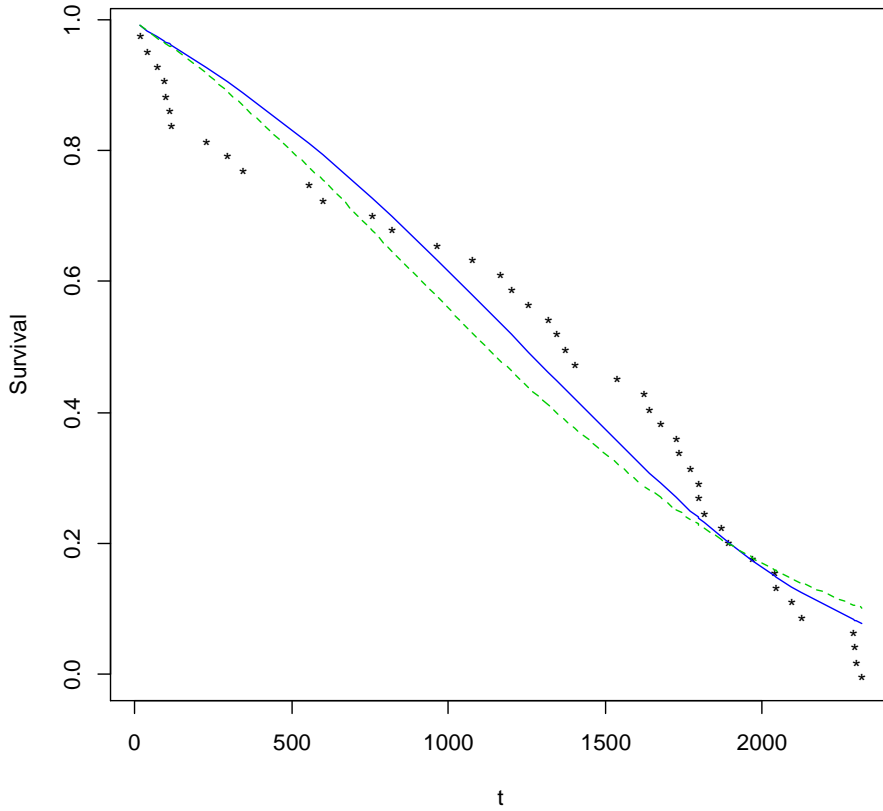
Η ανάγκη εύρεσης κατανομών που αποδίδουν βέλτιστη περιγραφή σε σύνθετες πραγματικές καταστάσεις, οδήγησε τους Marshall and Olkin (1997) στη μελέτη μιας γενικευμένη οικογένειας κατανομών που προκύπτει ως μίξη μιας κατανομής με τη γεωμετρική. Στο ίδιο πλαίσιο και έχοντας ως γνώμονα τις ιδιότητες της λογαριθμικής κατανομής, ορίζεται ένα γενικευμένο μοντέλο το οποίο μπορεί να παράγει εύχρηστες κατανομές, ικανές να βελτιστοποιήσουν την προσαρμογή δυσεφάρμοστων δεδομένων. Η ανάπτυξη και μελέτη μιας ειδικής τέτοιας κατανομής, αποδεικνύει τις δυνατότητες της γενικευμένης αυτής οικογένειας.



**Γράφημα 4.3** Γραφικές παραστάσεις της συνάρτησης επιβίωσης της εμπειρικής κατανομής (σημεία \*), της EXTG κατανομής (μπλε συνεχής γραμμή) και της XTG (πράσινη διακεκομμένη γραμμή) για τα δύο σύνολα δεδομένων.



(α) Ηλεκτρικά Εξαρτήματα



(β) Υποσυστήματα αυτοκινήτου

# Κεφάλαιο 5

## Η Γενικευμένη Εκθετική-Λογαριθμική Κατανομή

Οι Adamidis and Loukas (1998) εισήγαγαν μια κατανομή με φθίνουσα συνάρτηση κινδύνου. Η εν λόγω κατανομή αποτελεί μίξη εκθετικής με γεωμετρική κατανομή και είναι γνωστή ως εκθετική γεωμετρική (EG) κατανομή. Σε αντιστοιχία με τους Adamidis and Loukas (1998), οι Tahmasbi and Rezaei (2008) μελέτησαν μια δι-παραμετρική κατανομή, την εκθετική λογαριθμική (EL), η οποία αποτελεί μίξη εκθετικής και λογαριθμικής κατανομής, και έχει φθίνουσα συνάρτηση κινδύνου. Ήδη, ο Proschan (1963), είχε αποδείξει ότι η ιδιότητα της φθίνουσας συνάρτησης κινδύνου διατηρείται και σε μίξεις κατανομών με σταθερές συναρτήσεις κινδύνου. Η αθροιστική συνάρτηση της EL κατανομής δίνεται από τη σχέση:

$$F_{EL}(t) = 1 - \frac{\ln\{1 - (1-p)e^{-\lambda t}\}}{\ln p}, t, \lambda > 0 \text{ και } 0 < p < 1. \quad (5.1)$$

Σε αυτό το κεφάλαιο εισάγεται μια κατανομή τριών παραμέτρων ως γενίκευση της εκθετικής λογαριθμικής κατανομής. Η μελέτη της νέας κατανομής αποδεικνύει ότι δύναται να έχει φθίνουσα, αύξουσα ακόμη και μονοκόρυφη συνάρτηση κινδύνου αναλόγως με τις τιμές των παραμέτρων της. Επιπρόσθετα ελέγχονται στατιστικές ιδιότητες αυτής, καθώς και η χρησιμότητά της στη μοντελοποίηση ρεαλιστικών καταστάσεων με χρήση των εκτιμητών μέγιστης πιθανοφάνειας σε πραγματικά δεδομένα. Υπολογίζονται επίσης οι εκφράσεις της εντροπίας τόσο του Rényi (1961) όσο και του Shannon (1948).

### 5.1 Το μοντέλο

Υποθέτοντας την ύπαρξη μιας τυχαίας μεταβλητής με αθροιστική συνάρτηση κατανομής που δίνεται από την (5.1), μια τυχαία μεταβλητή

$T$  θα λέγεται ότι ακολουθεί γενικευμένη εκθετική λογαριθμική (GEL) κατανομή όταν η συνάρτηση επιβίωσής της  $S$  δίνεται από τη σχέση:

$$S(t) = 1 - \left[ 1 - \frac{\ln \{1 - (1-p)e^{-\lambda t}\}}{\ln p} \right]^\alpha, t, \alpha, \lambda > 0 \text{ και } 0 < p < 1. \quad (5.2)$$

Ακολουθώς, δίνεται ένας χαρακτηρισμός για την προτεινόμενη κατανομή. Έστω  $\alpha > 0$  ακέραιος και  $Y_1, Y_2, \dots, Y_\alpha$  τυχαίο δείγμα από την EL κατανομή. Τότε ισχύει ότι η τυχαία μεταβλητή  $T = \max \{Y_i\}_{i=1}^\alpha$  ακολουθεί την GEL κατανομή. Επομένως, η γενικευμένη κατανομή δύναται να χρησιμοποιηθεί για τη μοντελοποίηση της μέγιστης διάρκειας ζωής τυχαίων δειγμάτων από την EL. Αξίζει επιπλέον να σημειωθεί ότι η GEL κατανομή έχει επιθυμητή φυσική ερμηνεία. Αν υπάρχουν  $\alpha$  ανεξάρτητα εξαρτήματα σε ένα παράλληλο σύστημα και ο χρόνος ζωής αυτών ακολουθεί την EL κατανομή, τότε η διάρκεια ζωής του συστήματος μπορεί να περιγραφεί από την GEL κατανομή.

Στη συνέχεια, στην ενότητα 5.2 δίνονται οι συναρτήσεις πυκνότητας πιθανότητας και κινδύνου της προτεινόμενης κατανομής. Επιπλέον, αποδεικνύεται ότι η συνάρτηση κινδύνου δύναται να λάβει τρεις μορφές, δηλαδή φθίνουσα, αύξουσα και μονοκόρυφη και δίνεται τύπος υπολογισμού του  $i$ -οστού διατεταγμένου στατιστικού. Οι εκφράσεις των εκατοστιαίων σημείων, των ροπών  $r$ -τάξεως και του μέσου υπολειπόμενου χρόνου ζωής, δίνονται στην ενότητα 5.3 Στην ενότητα 5.4, μελετώνται τύποι υπολογισμού της εντροπίας της γενικευμένης κατανομής, ενώ στην ενότητα 5.5 εκτιμώνται οι παράμετροί της με τη μέθοδο μέγιστης πιθανοφάνειας και υπολογίζεται ο παρατηρούμενος πίνακας πληροφορίας. Στην τελευταία ενότητα 5.6, ελέγχεται η προσαρμολή της προτεινόμενης κατανομής έναντι άλλων δύο εναλλακτικών κατανομών, σε δύο πραγματικά σύνολα δεδομένων.

## 5.2 Συναρτήσεις πυκνότητας πιθανότητας, κινδύνου και διατεταγμένα στατιστικά

Έστω τυχαία μεταβλητή  $T \sim GEL(\alpha, \lambda, p)$ , με συνάρτηση επιβίωσης που δίνεται από την (5.2). Τότε, η αντίστοιχη σ.π.π. της  $T$  δίνεται από τη σχέση:

$$f(t; \theta) = \frac{\alpha \lambda (p-1) e^{-\lambda t}}{\{1 - (1-p)e^{-\lambda t}\} \ln p} \left[ 1 - \frac{\ln \{1 - (1-p)e^{-\lambda t}\}}{\ln p} \right]^{\alpha-1}, \quad (5.3)$$

όπου  $t, \alpha, \lambda > 0$  και  $0 < p < 1$ .

Γίνεται εύκολα αντιληπτό ότι για  $\alpha = 1$ , η προτεινόμενη κατανομή δίνει την EL κατανομή. Εάν επιπρόσθετα,  $p \rightarrow 1$  τότε η (5.3) ταυτίζεται με τη σ.π.π. της εκθετικής κατανομής.

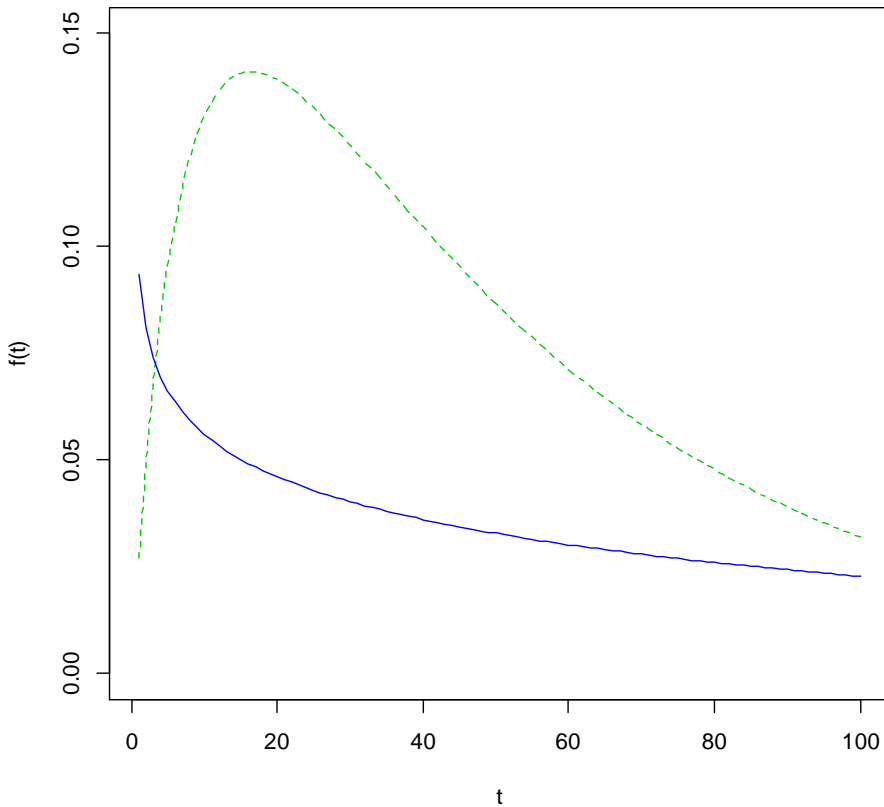
Για  $\alpha > 0$  ακέραιο, αποδεικνύεται ότι η (5.3), δύναται να γραφεί ως γραμμικός συνδυασμός:

$$f(t; \theta) = \Gamma(\alpha + 1) f_{EL}(t) \sum_{j=0}^{\alpha-1} \frac{(-1)^j}{\Gamma(\alpha - j) j!} S_{EL}^j(t),$$

όπου  $S_{EL}(t) = 1 - G_{EL}(t)$  είναι η συνάρτηση επιβίωσης τυχαίας μεταβλητής που ακολουθεί EL κατανομή με παραμέτρους  $\lambda$  και  $p$ .

Οι διάφορες μορφές της σ.π.π. (5.3) φαίνονται στο παρακάτω γράφημα 5.1, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \lambda, p)$ .

**Γράφημα 5.1** Συναρτήσεις πυκνότητας πιθανότητας της GEL κατανομής για  $(\alpha, \lambda, p) = (0.8, 0.02, 0.3)$  (μπλε συνεχής γραμμή) και  $(\alpha, \lambda, p) = (2, 0.2, 0.3)$  (πράσινη διακεκομμένη γραμμή).



Επιπρόσθετα, η συνάρτηση κινδύνου της GEL δίνεται από τη σχέση:

$$h(t; \theta) = \frac{\frac{\alpha \lambda}{\ln p} \frac{(1-p)e^{-\lambda t}}{\{1-(1-p)e^{-\lambda t}\}} \left[ 1 - \frac{\ln \{1-(1-p)e^{-\lambda t}\}}{\ln p} \right]^{\alpha-1}}{1 - \left[ 1 - \frac{\ln \{1-(1-p)e^{-\lambda t}\}}{\ln p} \right]^{\alpha}}, \quad (5.4)$$

όπου  $t, \alpha, \lambda > 0$  και  $0 < p < 1$ .

**Θεώρημα:**

Η συνάρτηση κινδύνου (5.4) είναι γνησίως φθίνουσα όταν  $0 < \alpha \leq 1$ , ενώ σε διαφορετική περίπτωση μπορεί να αποκτήσει και άλλες δύο διαφορετικές μορφές (αύξουσα και μονοκόρυφη).

**Απόδειξη:**

Παραγωγίζοντας την (5.3) προκύπτει ότι:

$$\begin{aligned}
 f'(t; \theta) &= \frac{\alpha \lambda (1-p)}{\ln p} \left[ 1 - \frac{\ln \{1 - (1-p)e^{-\lambda t}\}}{\ln p} \right]^{\alpha-1} \frac{\lambda e^{-\lambda t}}{\{1 - (1-p)e^{-\lambda t}\}^2} \\
 &+ \frac{\alpha(\alpha-1)\lambda^2(1-p)^2 e^{-2\lambda t}}{[\ln p \{1 - (1-p)e^{-\lambda t}\}]^2} \left[ 1 - \frac{\ln \{1 - (1-p)e^{-\lambda t}\}}{\ln p} \right]^{\alpha-2} \Leftrightarrow \\
 f'(t; \theta) &= \frac{\alpha \lambda^2 (1-p) e^{-\lambda t} \left[ \ln \left\{ \frac{p}{1 - (1-p)e^{-\lambda t}} \right\} \right]^{\alpha-2}}{(\ln p)^\alpha \{1 - (1-p)e^{-\lambda t}\}^2} \\
 &\times \left[ (\alpha-1)(1-p)e^{-\lambda t} + \ln \left\{ \frac{p}{1 - (1-p)e^{-\lambda t}} \right\} \right].
 \end{aligned}$$

Εφαρμόζοντας το θεώρημα του Glaser (1980) προκύπτει η συνάρτηση:

$$\begin{aligned}
 n(t; \theta) &= -\frac{f'(t; \theta)}{f(t; \theta)} = \frac{\alpha \lambda^2 (1-p) e^{-\lambda t} \left[ \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^{\alpha-2}}{(\ln p)^\alpha \{1-(1-p)e^{-\lambda t}\}^2} \\
 &= \frac{\alpha \lambda (1-p) e^{-\lambda t} \left[ \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^{\alpha-1}}{(\ln p)^\alpha \{1-(1-p)e^{-\lambda t}\}} \\
 &\quad \times \left[ (\alpha-1)(1-p)e^{-\lambda t} + \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right] \\
 &= \frac{\lambda \left[ (\alpha-1)(1-p)e^{-\lambda t} + \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]}{\{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\}}.
 \end{aligned}$$

Παραγωγίζοντας, την  $n(t; \theta)$  έχουμε:

$$\begin{aligned}
 n'(t; \theta) &= \frac{-\lambda^2 (\alpha-1)(1-p) e^{-\lambda t} \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\}}{\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2} \\
 &\quad - \frac{\lambda^2 (\alpha-1)(1-p)^2 e^{-2\lambda t} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\}}{\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2} \\
 &\quad + \frac{\lambda^2 (\alpha-1)(1-p)^2 e^{-2\lambda t}}{\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2} + \frac{\lambda^2 (1-p) e^{-\lambda t}}{\{1-(1-p)e^{-\lambda t}\}^2} \Leftrightarrow
 \end{aligned}$$



$$\begin{aligned}
 n'(t; \theta) &= \frac{\lambda^2 (1-p) e^{-\lambda t} \left[ \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]}{\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2} \\
 &\quad \times \left[ (\alpha-1) \{1-(1-p)e^{-\lambda t}\} + \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right] \\
 &\quad + \frac{(\alpha-1)(1-p)e^{-\lambda t} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\}}{\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2} \Leftrightarrow \\
 n'(t; \theta) &= \frac{\lambda^2 (1-p) e^{-\lambda t} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \left[ \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} + \alpha - 1 \right]}{\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2}.
 \end{aligned}$$

Δεδομένου ότι  $t, \alpha, \lambda > 0$  και  $0 < p < 1$  ισχύουν τα ακόλουθα:

- $\lambda^2 (1-p) e^{-\lambda t} > 0, \forall t > 0,$
- $\ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} < 0, \forall t > 0$  και
- $\left[ \{1-(1-p)e^{-\lambda t}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} \right]^2 > 0, \forall t > 0.$

Άρα, το πρόσημο της συνάρτησης  $n'(t; \theta)$  καθορίζεται από την

$$\text{ποσότητα } A(t) = \left[ \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t}} \right\} + \alpha - 1 \right].$$

Περιπτώσεις:

1)  $0 < \alpha \leq 1$ , τότε είναι:

$A(t) < 0 \Leftrightarrow n'(t; \theta) < 0 \Rightarrow h(t; \theta) \downarrow \forall t > 0$  φθίνουσα συνάρτηση κινδύνου.

2)  $1 < \alpha < 1 - \ln p$ , τότε είναι:

$$A(t_0) = 0 \Leftrightarrow \ln \left\{ \frac{p}{1 - (1-p)e^{-\lambda t_0}} \right\} + \alpha - 1 = 0 \Leftrightarrow$$

$$\frac{p}{1 - (1-p)e^{-\lambda t_0}} = e^{1-\alpha} \Leftrightarrow 1 - (1-p)e^{-\lambda t_0} = pe^{\alpha-1} \Leftrightarrow$$

$$(1-p)e^{-\lambda t_0} = 1 - pe^{\alpha-1} \Leftrightarrow e^{-\lambda t_0} = \frac{1 - pe^{\alpha-1}}{1-p} \Leftrightarrow$$

$$-\lambda t_0 = \ln \left( \frac{1 - pe^{\alpha-1}}{1-p} \right) \Leftrightarrow t_0 = -\frac{1}{\lambda} \ln \left( \frac{1 - pe^{\alpha-1}}{1-p} \right) = \ln \left( \frac{1-p}{1 - pe^{\alpha-1}} \right)^{\lambda^{-1}}.$$

Οπότε  $\exists t_0 > 0$ , τέτοιο ώστε  $A(t_0) = 0 \Leftrightarrow n'(t_0; \theta) = 0$ , ενώ το πρόσημο της ποσότητας  $n''(t_0; \theta)$  θα καθορίσει το είδος της καμπύλης (κυρτή ή κοίλη). Όμως, είναι:

$$n''(t; \theta) = \frac{\lambda^3 y (1-y) \ln \left( \frac{p}{y} \right)}{\left[ y \ln \left( \frac{p}{y} \right) \right]^3} \times \left[ \left\{ \frac{1-y}{y} - \ln \left( \frac{p}{y} \right) \right\} \left\{ \ln \left( \frac{p}{y} \right) + \alpha - 1 \right\} + \ln \left( \frac{p}{y} \right) \frac{1-y}{y} \right] - \frac{2\lambda^3 (1-y)^2 \ln \left( \frac{p}{y} \right) \left\{ \ln \left( \frac{p}{y} \right) + \alpha - 1 \right\} \left\{ \ln \left( \frac{p}{y} \right) + 1 \right\}}{\left[ y \ln \left( \frac{p}{y} \right) \right]^3},$$

όπου  $y = 1 - (1-p)e^{-\lambda t}$  και

$$n''(t_0; \theta) = -\frac{\lambda^3 (1-y_0)^2}{y_0^3 \ln\left(\frac{p}{y_0}\right)} = \frac{\lambda^3 (1-pe^{\alpha-1})^2}{(1-a)(pe^{\alpha-1})^3} < 0, \text{ όταν } \alpha > 1.$$

Επομένως, η χρονική στιγμή  $t_0$  αποτελεί μέγιστο για τη συνάρτηση  $n(t; \theta)$ , γεγονός που σημαίνει ότι:

- $n'(t; \theta) > 0 \Leftrightarrow n(t; \theta) \uparrow$  για κάθε  $0 < t < t_0$ ,
- $n'(t_0) = 0$  για  $t = t_0$  και
- $n'(t; \theta) < 0 \Leftrightarrow n(t; \theta) \downarrow$  για κάθε  $t > t_0$ .

και άρα η  $h(t; \theta)$  είναι μονοκόρυφη συνάρτηση (UBTFR).

3)  $1 < \alpha < 1 - \ln p$ , τότε είναι:

$$A(t) \geq -\ln\{1 - (1-p)e^{-\lambda t}\} > 0 \Leftrightarrow n'(t; \theta) > 0 \Rightarrow h(t; \theta) \uparrow \forall t > 0$$

αύξουσα συνάρτηση κινδύνου.

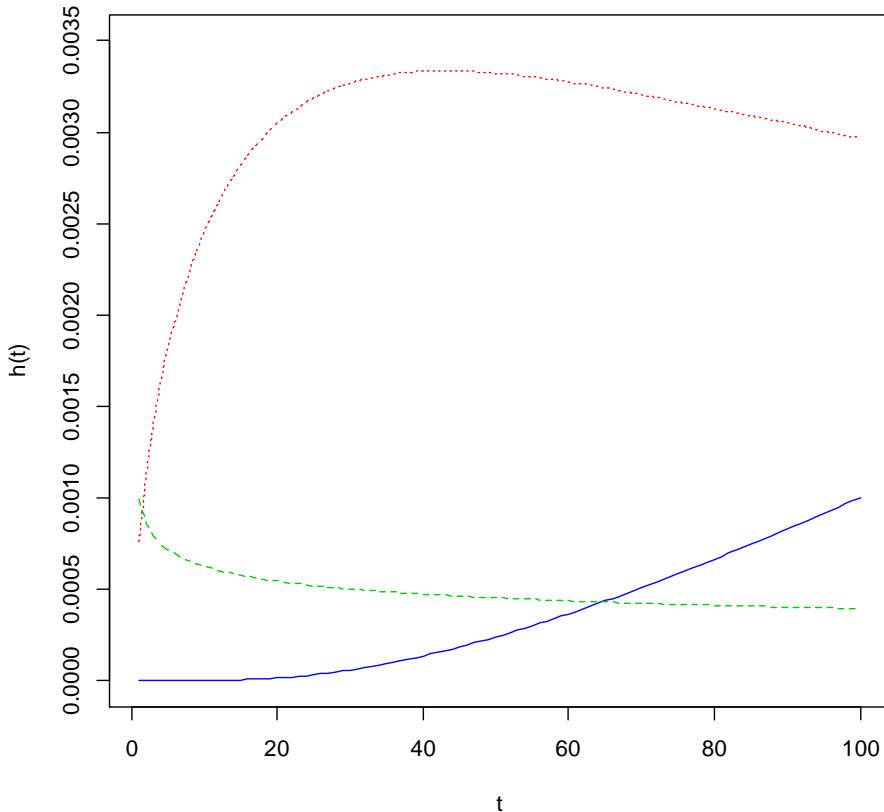
Επίσης, προκύπτει ότι η σ.π.π. του  $i$ -οστού διατεταγμένου στατιστικού ενός τυχαίου δείγματος  $T_1, T_2, \dots, T_n$  από την GEL κατανομή μπορεί να δοθεί από τη σχέση:

$$f_{i:n}(t; \theta) = \frac{n!}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} \frac{(-1)^k}{i+k} \binom{n-i}{k} f_{GEL}(t; \alpha(i+k), \lambda, p),$$

όπου  $f_{GEL}(\cdot; \alpha(i+k), \lambda, p)$ , η σ.π.π. τυχαίας μεταβλητής που ακολουθεί την κατανομή GEL με παραμέτρους  $\theta = (\alpha(i+k), \lambda, p)$ .

Οι διάφορες μορφές της συνάρτησης κινδύνου φαίνονται και στο παρακάτω γράφημα 5.2, για συγκεκριμένες τιμές των παραμέτρων  $\theta = (\alpha, \lambda, p)$ .

**Γράφημα 5.2** Συναρτήσεις κινδύνου της GEL κατανομής για  $(\alpha, \lambda, p) = (10, 0.01, 0.005)$  (μπλε συνεχής γραμμή),  $(\alpha, \lambda, p) = (0.8, 0.0001, 0.6)$  (πράσινη διακεκομμένη γραμμή) και  $(\alpha, \lambda, p) = (1.6, 0.00001, 0.00005)$  (κόκκινη γραμμή με τελείες).



### 5.3 Ιδιότητες

Από τη (5.2) προκύπτει ότι η α.σ.κ. της GEL δίνεται από τη σχέση:

$$F(t) = \left[ 1 - \frac{\ln \{1 - (1-p)e^{-\lambda t}\}}{\ln p} \right]^\alpha, t, \alpha, \lambda > 0 \text{ και } 0 < p < 1. \quad (5.5)$$

Επιπλέον με αντιστροφή της (5.5) προκύπτει η σχέση υπολογισμού των εκατοστιαίων σημείων

$$t_q = F^{-1}(q) = \frac{1}{\lambda} \ln \left\{ \frac{1-p}{1-p^{(\sqrt[q]{q-1})}} \right\}, p \in (0,1).$$

Επομένως, η διάμεσος είναι  $M = \frac{1}{\lambda} \ln \left\{ \frac{1-p}{1-p^{(\frac{1}{\sqrt{2}}-1)}} \right\}$ .

Ο υπολογισμός των r-τάξεως ροπών απαιτεί γενικά χρήση τυπικών διαδικασιών αριθμητικής ολοκλήρωσης, που είναι διαθέσιμοι στην πλειοψηφία των μαθηματικών πακέτων. Για τη προτεινόμενη ωστόσο κατανομή, οι ροπές μπορούν να εκφραστούν από τη σχέση:

$$E(T^r) = \frac{\alpha}{\lambda^r} \int_0^1 (1-t)^{\alpha-1} \ln^r \left( \frac{1-p}{1-p^t} \right) dt.$$

Επιπρόσθετα, δεδομένου ότι δεν υπήρξε αποτυχία προγενέστερη του  $t_0$ , ο υπολοιπόμενος χρόνος ζωής ορίζεται ως το χρονικό διάστημα από το  $t_0$  μέχρι τη στιγμή της αποτυχίας. Η συνάρτηση του μέσου υπολοιπόμενου χρόνου ζωής της κατανομής GEL δίνεται από τη σχέση:

$$\begin{aligned} m(t_0; \theta) &= E(T - t_0 | T \geq t_0) = \int_{t_0}^{+\infty} \frac{(t-t_0) f(t; \theta)}{1-F(t_0; \theta)} dt \\ &= \frac{\alpha}{1-(1-A)^\alpha} \int_0^A \left\{ \frac{1}{\lambda} \ln \left( \frac{1-p}{1-p^t} \right) - t_0 \right\} (1-t)^{\alpha-1} dt, \end{aligned}$$

όπου  $A = \frac{\log \{1-(1-p)e^{-\lambda t_0}\}}{\log p}$ .

## 5.4 Εντροπίες

Η εντροπία αποτελεί έννοια ιδιάζουσας σημασίας σε διάφορους επιστημονικούς κλάδους, όπως στη φυσική, στις πιθανότητες και τη στατιστική, στη θεωρία τηλεπικοινωνιών και στις οικονομικές επιστήμες. Προπομπός της μελέτης της εντοπίας αποτέλεσε η συνάρτηση που όρισε ο Hartley (1928) ως ένα μέτρο αβεβαιότητας. Ωστόσο, η μελέτη του Shannon (1948) στη μαθηματική θεωρία τηλεπικοινωνιών ήταν αυτή που καθιέρωσε την καθολική χρήση της εντροπίας ως ένα βασικό «εργαλείο» σχεδόν κάθε επιστημονικού και μηχανολογικού κλάδου. Έπειτα, ακολούθησε η μελέτη του Rényi (1961), όπου γενίκευσε τον τύπο υπολογισμού της εντροπίας κατά Shannon. Στην ανάλυση επιβίωσης, η εντροπία μιας τυχαίας μεταβλητής  $T$  αποτελεί ένα μέτρο της αβέβαιης μεταβλητότητας αυτής (Rényi, 1961), ενώ μπορεί να χρησιμοποιηθεί ακόμη και για την ταυτοποίηση ενός πιθανοθεωρητικού μοντέλου, μέσω της αρχής της μέγιστης εντροπίας.

Η εντροπία του Rényi (1961) ορίζεται ως

$$I_R(\gamma) = \frac{1}{1-\gamma} \ln \left\{ \int_{-\infty}^{+\infty} f^\gamma(t) dt \right\}, \gamma > 0 \text{ και } \gamma \neq 1,$$

ενώ ειδική περίπτωση αυτής αποτελεί η εντροπία του Shannon (1948), η οποία δίνεται από τον τύπο

$$\lim_{\gamma \rightarrow 1} I_R(\gamma) = H_{sh}(f) \equiv E_T[-\ln f(T)].$$

Στην περίπτωση της GEL κατανομής, και για  $(\alpha-1)\gamma+1 > 0$ , η κατά Rényi εντροπία δίνεται από τη σχέση:

$$I_R(\gamma) = \frac{1}{1-\gamma} \ln \left\{ \left[ -\frac{\alpha\lambda(1-p)}{\ln p} \right]^\gamma \sum_{j=0}^{+\infty} \frac{(-1)^j \Gamma((\alpha-1)\gamma+1)}{\Gamma((\alpha-1)\gamma+1-j) j!} \times \int_0^{+\infty} \left[ \frac{e^{-\lambda t}}{1-(1-p)e^{-\lambda t}} \right]^\gamma \left[ \frac{\ln \{1-(1-p)e^{-\lambda t}\}}{\ln p} \right]^j dt \right\},$$

ενώ η εντροπία του Shannon μπορεί να υπολογιστεί από τη σχέση:

$$H_{Sh}(f) \equiv E[-\ln f(T)] = \ln(-\ln p) - (\alpha - 1)\alpha^{-1} + (\alpha + 1)^{-1} \ln p - \ln\{\alpha\lambda(1-p)\} + \alpha \int_0^1 (1-t)^{\alpha-1} \ln\left(\frac{1-p}{1-p^t}\right) dt.$$

## 5.5 Στατιστική συμπερασματολογία

Θεωρώντας ένα τυχαίο δείγμα  $n$  παρατηρήσεων  $y_{obs} = (t_i; i=1, \dots, n)$  από την (5.5), ο λογάριθμος της συνάρτησης πιθανοφάνειας θα δίνεται από τη σχέση:

$$l(\theta; y_{obs}) = -n \ln(-\ln p) + n \ln \alpha + n \ln \lambda + n \ln(1-p) - \lambda \sum_{i=1}^n t_i - \sum_{i=1}^n \left[ \ln\{1 - (1-p)e^{-\lambda t_i}\} \right] + (\alpha + 1) \sum_{i=1}^n \left[ \ln \left[ 1 - \frac{\ln\{1 - (1-p)e^{-\lambda t_i}\}}{\ln p} \right] \right].$$

Παραγωγίζοντας ως προς τις παραμέτρους  $\theta = (\alpha, \lambda, p)$  προκύπτουν οι μερικές παράγωγοι  $\frac{\partial l}{\partial \alpha}$ ,  $\frac{\partial l}{\partial \lambda}$  και  $\frac{\partial l}{\partial p}$ , οι τύποι υπολογισμού των οποίων δίνονται από τις ακόλουθες σχέσεις:

$$\frac{\partial l}{\partial \alpha} = n\alpha^{-1} + \sum_{i=1}^n \left[ \ln \left[ 1 - \frac{\ln\{1 - (1-p)e^{-\lambda t_i}\}}{\ln p} \right] \right], \quad (5.6)$$

$$\frac{\partial l}{\partial \lambda} = n\lambda^{-1} - \lambda \sum_{i=1}^n \left\{ \frac{t_i}{1 - (1-p)e^{-\lambda t_i}} \right\} - (\alpha - 1) \sum_{i=1}^n \left[ \frac{(1-p)t_i e^{-\lambda t_i}}{\{1 - (1-p)e^{-\lambda t_i}\} \ln \left\{ \frac{p}{1 - (1-p)e^{-\lambda t_i}} \right\}} \right], \quad (5.7)$$

$$\frac{\partial l}{\partial p} = -n \left\{ (p \ln p)^{-1} + (1-p)^{-1} \right\} - \sum_{i=1}^n \left\{ \frac{e^{-\lambda t_i}}{1 - (1-p)e^{-\lambda t_i}} \right\} - (\alpha - 1) \times \sum_{i=1}^n \left[ \frac{p \ln p e^{-\lambda t_i} - \{1 - (1-p)e^{-\lambda t_i}\} \ln \{1 - (1-p)e^{-\lambda t_i}\}}{p \ln p \{1 - (1-p)e^{-\lambda t_i}\} \ln \left\{ \frac{p}{1 - (1-p)e^{-\lambda t_i}} \right\}} \right]. \quad (5.8)$$

Οι εκτιμητές μέγιστης πιθανοφάνειας  $\hat{\theta} = (\hat{\alpha}, \hat{\lambda}, \hat{p})$  των παραμέτρων  $\theta = (\alpha, \lambda, p)$  προκύπτουν από την επίλυση του συστήματος των τριών μη-γραμμικών κανονικών εξισώσεων:  $\frac{\partial l}{\partial \alpha} = 0$ ,  $\frac{\partial l}{\partial \lambda} = 0$  και  $\frac{\partial l}{\partial p} = 0$ . Είναι προφανές ότι η επίλυση του εν λόγω συστήματος, δύναται να επιτευχθεί μόνο με χρήση μιας επαναληπτικής αριθμητικής μεθόδου.

Επιπρόσθετα, σύμφωνα με την ασυμπτωτική θεωρία κανονικότητας, η κατανομή των εκτιμητών μέγιστης πιθανοφάνειας  $\hat{\theta} = (\hat{\alpha}, \hat{\lambda}, \hat{p})$ , είναι κατά προσέγγιση η τρισδιάστατη κανονική με μέση τιμή  $\theta = (\alpha, \lambda, p)$  και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα πληροφορίας του Fisher  $I(\theta)$ , ο οποίος υπολογίζεται από τη σχέση  $I(\theta) = E(J(\theta))$ , όπου,  $J(\theta)$  είναι ο συμμετρικός, τρίτης τάξης



παρατηρούμενος πίνακας πληροφορίας, ο οποίος αποτελεί συνεπή εκτιμητή του  $I(\theta)$ . Τα στοιχεία του πίνακα  $J(\theta)$ :  $J_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}$ , όπου

$i, j = 1, 2, 3$ , δίνονται αναλυτικά από τις σχέσεις:

Παραγωγίζοντας επιμέρους τις σχέσεις (5.6)-(5.8) των μερικών παραγώγων μπορούν εύκολα να προκύψουν τα άνω τριγωνικά στοιχεία του, δευτέρας τάξεως, συμμετρικού, παρατηρούμενου πίνακα πληροφορίας  $I$ , τα οποία δίνονται από τις σχέσεις:

$$J_{11} = \frac{n}{a^2},$$

$$J_{12} = \sum_{i=1}^n \left[ \frac{(1-p)t_i e^{-\lambda t_i}}{\{1-(1-p)e^{-\lambda t_i}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t_i}} \right\}} \right],$$

$$J_{13} = \sum_{i=1}^n \left[ \frac{p \ln p e^{-\lambda t_i} - \{1-(1-p)e^{-\lambda t_i}\} \ln \{1-(1-p)e^{-\lambda t_i}\}}{p \ln p \{1-(1-p)e^{-\lambda t_i}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t_i}} \right\}} \right],$$

$$J_{22} = \frac{n}{\lambda^2} - \sum_{i=1}^n \frac{(1-p)t_i^2 e^{-\lambda t_i}}{\{1-(1-p)e^{-\lambda t_i}\}^2} - (a-1) \sum_{i=1}^n \frac{(1-p)t_i^2 e^{-\lambda t_i} \left[ \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t_i}} \right\} - (1-p)e^{-\lambda t_i} \right]}{\{1-(1-p)e^{-\lambda t_i}\}^2 \ln^2 \left\{ \frac{p}{1-(1-p)e^{-\lambda t_i}} \right\}},$$

$$J_{23} = -\sum_{i=1}^n \frac{t_i e^{-\lambda t_i}}{\{1-(1-p)e^{-\lambda t_i}\}^2} - (a-1) \sum_{i=1}^n \frac{t_i e^{-\lambda t_i} \left[ \frac{(1-p)(1-e^{-\lambda t_i})}{p} + \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda t_i}} \right\} \right]}{\{1-(1-p)e^{-\lambda t_i}\}^2 \ln^2 \left\{ \frac{p}{1-(1-p)e^{-\lambda t_i}} \right\}},$$

$$J_{33} = -\frac{n(1+\ln p)}{p^2 \ln^2 p} + \frac{n}{(1-p)^2} - \sum_{i=1}^n \frac{e^{-2\lambda x_i}}{\{1-(1-p)e^{-\lambda x_i}\}^2} - (a-1) \sum_{i=1}^n \frac{1}{p^2 \ln^2 p \{1-(1-p)e^{-\lambda x_i}\}^2 \ln^2 \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\}} \times \left[ p^2 e^{-2\lambda x_i} \ln p \ln^2 \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\} + p e^{-\lambda x_i} \{1-(1-p)e^{-\lambda x_i}\} \ln^2 \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\} - p e^{-\lambda x_i} (1-e^{-\lambda x_i}) \ln p \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\} - (1-e^{-\lambda x_i}) \{1-(1-p)e^{-\lambda x_i}\} \ln p \ln \{1-(1-p)e^{-\lambda x_i}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\} - (1-e^{-\lambda x_i}) p e^{-\lambda x_i} \ln p \ln \{1-(1-p)e^{-\lambda x_i}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\} - (1-e^{-\lambda x_i}) \{1-(1-p)e^{-\lambda x_i}\} \ln \{1-(1-p)e^{-\lambda x_i}\} \ln \left\{ \frac{p}{1-(1-p)e^{-\lambda x_i}} \right\} - (1-e^{-\lambda x_i})^2 \ln p \{1-(1-p)e^{-\lambda x_i}\} \right].$$

## 5.6 Εφαρμογές

Η προσαρμογή της προτεινόμενης GEL κατανομής ελέγχεται σε δύο σύνολα πραγματικών δεδομένων. Το πρώτο σύνολο δεδομένων περιλαμβάνει 213 αποτελέσματα από χρόνους ζωής συστημάτων κλιματισμού σε αεροσκάφη τύπου Boeing 720 (Proschan, 1963, Dahiya and Gurland, 1972, Gleser, 1989 και Adamidis and Loukas, 1998). Το δεύτερο σύνολο δεδομένων αποτελείται από 51 παρατηρήσεις που αφορούν το χρονικό διάστημα (σε ημέρες) από την ύφεση μέχρι την υποτροπή ασθενών με οξεία μη-λεμφοβλαστική λευχαιμία (Glucksberg et al., 1981 και Ebrahimi, 1991). Παράλληλα με την προτεινόμενη GEL κατανομή, η EL και η DAL κατανομή (Dimitrakopoulou et al., 2007) προσαρμόζονται στα δύο σύνολα δεδομένων, ούτως ώστε να υπάρξει μια άμεση αντιπαράθεση των τριών κατανομών μεταξύ τους. Υπολογίζοντας τους εκτιμητές μέγιστης πιθανοφάνειας, χρησιμοποιήθηκε το τεστ καλής προσαρμογής των Kolmogorov-Smirnov (K-S), για τον έλεγχο της προσαρμοστικότητας κάθε μιας από τις τρεις κατανομές στα δύο σύνολα δεδομένων. Οι εκτιμητές, ο λογάριθμος της συνάρτησης πιθανοφάνειας (LL), το κριτήριο AIC, η τιμή του στατιστικού του K-S τεστ, καθώς και η αντίστοιχη κρίσιμη τιμή (p-value) περιέχονται στον παρακάτω Πίνακα 5.1. Επιπρόσθετα, οι τιμές του στατιστικού τεστ πηλίκου πιθανοφανειών για τον έλεγχο της μηδενικής υπόθεσης  $H_0: \alpha = 1$ , υπολογίστηκαν να είναι 4.694 (p=0.0303) και 10.485 (p=0.0012) αντίστοιχα, επιδεικνύοντας ότι η GEL περιγράφει τα δύο σύνολα δεδομένων σαφώς καλύτερα από τη EL κατανομή. Το συμπέρασμα αυτό ενισχύεται και από τη γραφική παράσταση των εμπειρικών και προσαρμοζόμενων συναρτήσεων επιβίωσης του Γραφήματος 5.3.

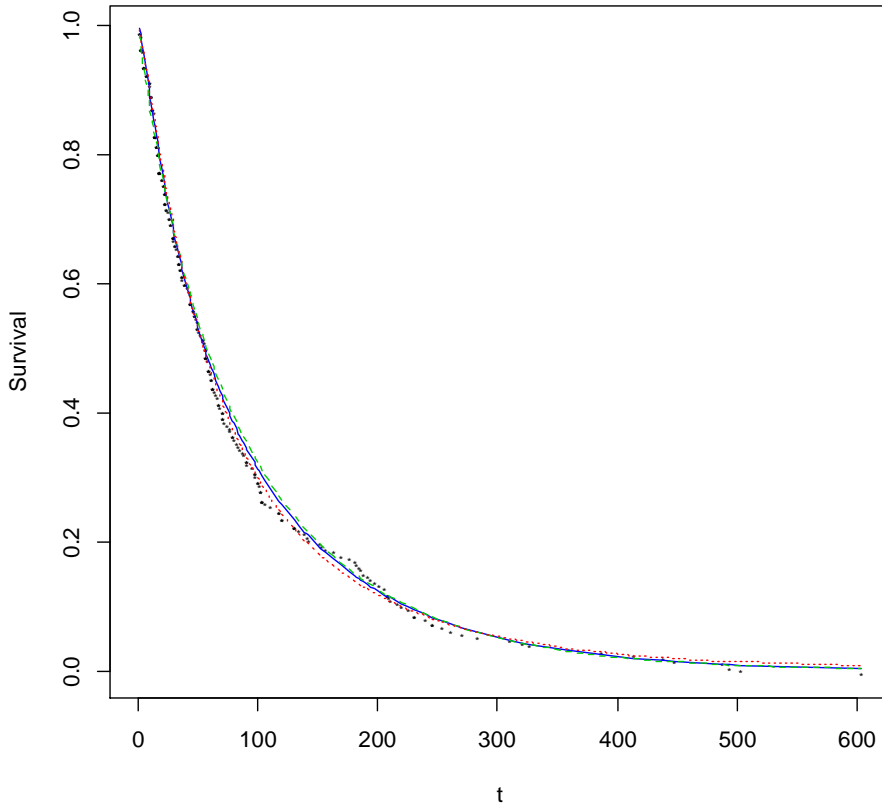
**Πίνακας 5.1** Εκτιμητές παραμέτρων, τιμές λογαρίθμου συνάρτησης πιθανοφάνειας, κριτηρίου AIC, στατιστικού K-S και στατιστικού πηλίκου πιθανοφανειών, καθώς και κρίσιμες τιμές από την προσαρμογή των συγκρινόμενων κατανομών πιθανότητας στα δύο διαθέσιμα σύνολα δεδομένων.

Σύνολο Δεδομένων	Κατανομή	$\hat{\theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\rho})$	LL	AIC	K-S (p-value)	LL ratio test
1) $n = 213$	GEL	(1.7441, 0.0081, 0.0623)	-1173.7865	2353.573	0.0345 (0.9618)	$H_0: \alpha = 1$
	EL	(-----, 0.0084, 0.3874)	-1176.1335	2356.267	0.0486 (0.6965)	4.694 (0.0303)
	DAL	(0.4638, 1.256, 0.0138)	-1174.762	2355.524	0.0389 (0.904)	
2) $n = 51$	GEL	(3.0377, 0.0047, 0.2168)	-335.4452	676.8903	0.0643 (0.9842)	$H_0: \alpha = 1$
	EL	(-----, 0.0034, 0.9432)	-340.6876	685.3753	0.1413 (0.2607)	10.485 (0.0012)
	DAL	(0.9978, 1.0692, 0.0024)	-339.2599	684.5198	0.1279 (0.3746)	

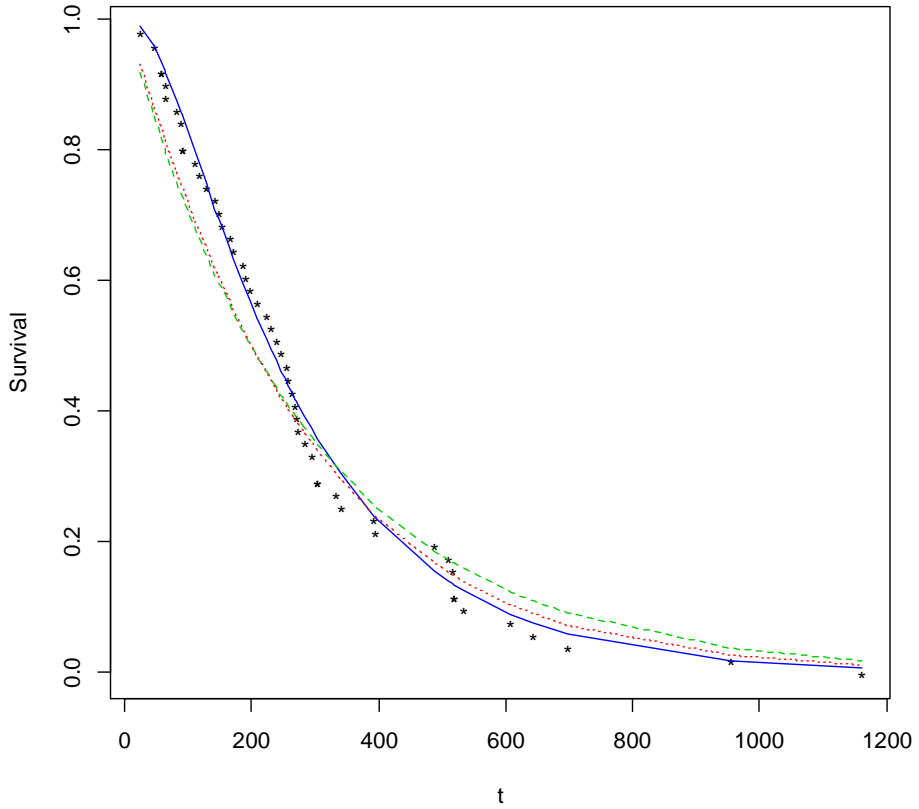
## 5.7 Συμπεράσματα

Η προτεινόμενη GEL κατανομή αποτελεί γενίκευση της EL κατανομής που εισήγαγαν οι Tahmasbi and Rezaei (2008). Η μελέτη της αποδεικνύει ότι η γενικευμένη κατανομή δύναται να έχει φθίνουσα, αύξουσα ακόμη και μονοκόρυφη (UBFTR) συνάρτηση κινδύνου, ανάλογα με τις τιμές των παραμέτρων της, σε αντίθεση με την EL κατανομή που έχει μόνο φθίνουσα συνάρτηση κινδύνου. Αυτό σημαίνει ότι η προτεινόμενη κατανομή έχει μεγαλύτερη ευελιξία. Επιπλέον, η προσαρμογή της σε συγκεκριμένα σύνολα δεδομένων, πιστοποιεί την υπεροχή της γενικευμένης κατανομής στην περιγραφή σύνθετων πραγματικών καταστάσεων έναντι τόσο της EL, όσο και άλλης αντίστοιχης τρι-παραμετρικής κατανομής.

**Γράφημα 5.3** Γραφικές παραστάσεις της συνάρτησης επιβίωσης της εμπειρικής κατανομής (σημεία \*), της GEL (μπλε συνεχής γραμμή) της EL (πράσινη διακεκομμένη γραμμή) και της DAL κατανομής (κόκκινη γραμμή με τελείες) για τα δύο σύνολα δεδομένων.



(α) Συστήματα κλιματισμού αεροσκαφών



(β) Χρόνοι ζωής ασθενών

# Κεφάλαιο 6

## Βήτα Γενικευμένες Κατανομές

Η ανάγκη για βέλτιστη προσομοίωση πραγματικών καταστάσεων, ώθησε τους ερευνητές να στραφούν στις γενικευμένες κατανομές. Τον πρόδρομο αυτής της κατηγορίας κατανομών αποτέλεσε ο Amoroso (1925), ο οποίος μελέτησε τη γενικευμένη γάμμα κατανομή με σκοπό την προσαρμογή της παρατηρούμενης κατανομής εισοδήματος. Έκτοτε, πολυάριθμες έρευνες έχουν εκπονηθεί αναπτύσσοντας ποικίλες κατηγορίες γενικευμένων κατανομών. Στο κεφάλαιο αυτό, μελετάται μια ειδική κατηγορία γενικευμένων κατανομών, οι βήτα γενικευμένες κατανομές, η ευελιξία των οποίων, τείνει να καλύπτει τη δυσκολία μελέτης των ιδιοτήτων τους. Αρχικά, δίνεται μια αναλυτική περιγραφή της γενικευμένης οικογένειας, στη συνέχεια παρουσιάζονται ειδικές περιπτώσεις κατανομών που ανήκουν σε αυτήν και έχουν προταθεί τα τελευταία χρόνια στη βιβλιογραφία, ενώ το κεφάλαιο ολοκληρώνεται με προτάσεις για μελλοντική έρευνα στο συγκεκριμένο πεδίο.

### 6.1 Η γενικευμένη οικογένεια

Η συγκεκριμένη μέθοδος γενίκευσης κατανομών εισήχθη από τους Eugene et al. (2002), οι οποίοι όρισαν τη βήτα κανονική κατανομή. Βασικό πλεονέκτημα αυτής, αποτέλεσε το γεγονός ότι σε αντίθεση με την κανονική κατανομή, δύναται να είναι μονοκόρυφη ή δικόρυφη. Στα χρόνια που ακολούθησαν, πολλοί ερευνητές μελέτησαν συγκεκριμένες περιπτώσεις κατανομών που ανήκουν στην οικογένεια των βήτα γενικευμένων. Οι Famoye et al. (2003), Gupta and Nadarajah (2004) και Razzaghi (2009), επέκτειναν με τη σειρά τους την έρευνα των Eugene et al. (2002), μελετώντας αντίστοιχα τη δικόρυφη περιοχή της βήτα κανονικής κατανομής και διάφορες εκφράσεις των ροπών της. Παράλληλα, οι Brown et al. (2002) και Olapade (2004) μελέτησαν τη βήτα λογιστική κατανομή, ενώ η μελέτη των Nadarajah and Kotz (2004), βασίστηκε στην βήτα Gumbel κατανομή, η οποία έχει μεγαλύτερη

ευελιξία ουρών από ότι η Gumbel κατανομή (γνωστή και ως extreme value τύπου I κατανομή). Την ίδια χρονική περίοδο, ο Jones (2004) μελέτησε εκτενώς τις βήτα γενικευμένες κατανομές βασιζόμενος στην κατανομή των διατεταγμένων στατιστικών. Οι Nadarajah and Gupta (2004) ήταν αυτοί που μελέτησαν και την βήτα Fréchet κατανομή, για την οποία οι Barreto-Souza et al. (2008) παρουσίασαν ορισμένες επιπρόσθετες μαθηματικές ιδιότητες. Οι Famoye et al. (2005) πρότειναν τη βήτα-Weibull κατανομή, οι Lee et al. (2007) παρουσίασαν εφαρμογές της σε λογοκριμένα δεδομένα και ο Zografos (2008) μελέτησε αναλυτικότερα. Οι Nadarajah and Kotz (2006) όρισαν τη βήτα εκθετική κατανομή, της οποίας η συνάρτηση κινδύνου δίνεται να λάβει αύξουσα ή φθίνουσα μορφή. Οι Kong et al. (2007) εισήγαγαν τη βήτα γάμμα κατανομή, οι Akinsete et al. (2008) όρισαν τη βήτα Pareto κατανομή και εν συνεχεία οι Lemos de Morais et al. (2009) μελέτησαν τη βήτα γενικευμένη λογιστική τύπου IV. Ακολούθησαν οι Cordeiro and Barreto-Souza (2009) οι οποίοι εξήγαγαν ορισμένα γενικά αποτελέσματα για την εν λόγω γενικευμένη οικογένεια κατανομών, ενώ στην εκτενέστατη μελέτη των Zografos and Balakrishnan (2009), εκφράζονται συν τοις άλλοις τύποι υπολογισμού της κατά Shannon εντροπίας για τις οικογένειες των βήτα και γάμμα γενικευμένων κατανομών, καθώς και μια εναλλακτική μέθοδος των ροπών για την εκτίμηση των παραμέτρων των γενικευμένων αυτών κατανομών. Έκτοτε, πολλαπλασιάστηκαν οι έρευνες που αφορούν κατανομές που ανήκουν στη συγκεκριμένη οικογένεια, με τους Fischer and Vaughan (2010) να μελετούν τη βήτα υπερβολική τέμνουσα, τους Barreto-Souza et al. (2010) τη βήτα γενικευμένη εκθετική και τους Pescim et al. (2010) τη βήτα γενικευμένη ημι-κανονική κατανομή. Μεταγενέστερα, ακολούθησαν οι Paranaíba et al. (2010) με τη μελέτη της βήτα Burr τύπου XII, ο Mahmoudi (2011) με τη μελέτη της βήτα γενικευμένης Pareto κατανομής και πιο πρόσφατα οι Singla et al. (2012) και Bidram (2012), ερευνώντας τη βήτα γενικευμένη Weibull και βήτα εκθετική γεωμετρική κατανομή, αντίστοιχα.



### 6.1.1 Ορισμός

Η οικογένεια των βήτα γενικευμένων (B-Γ) κατανομών ορίσθηκε από τους Eugene et al. (2002), μέσω της αθροιστικής συνάρτησης κατανομής της. Η α.σ.κ. της εν λόγω οικογένειας γενικευμένων κατανομών δίνεται από τη σχέση:

$$F(t) = \frac{1}{B(a,b)} \int_0^{G(t)} x^{a-1} (1-x)^{b-1} dx, \quad a, b > 0,$$

όπου  $G(t)$  είναι η α.σ.κ. μιας αρχικής τυχαίας μεταβλητής και

$$B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx.$$

Η οικογένεια αυτή των B-Γ κατανομών, αποτελεί γενίκευση της αρχικής κατανομής  $G$  μιας τυχαίας μεταβλητής με α.σ.κ.  $G(t)$ . Μάλιστα, στην περίπτωση όπου  $a = b = 1$ , η α.σ.κ. της B-Γ κατανομής ταυτίζεται με τη  $G(t)$ . Ο κύριος λόγος ύπαρξης των επιπρόσθετων παραμέτρων είναι η προσθήκη ευελιξίας στη μορφή και τις ουρές της κατανομής με άμεσο αντίκτυπο στην προσαρμογή της σε πραγματικές καταστάσεις.

Η α.σ.κ. της B-Γ οικογένειας κατανομών μπορεί να δοθεί από τη σχέση:

$$F(t) = I_{G(t)}(a,b) = \frac{B_{G(t)}(a,b)}{B(a,b)},$$

όπου  $B_{G(t)}(a,b)$  παριστάνει τη μη πλήρη συνάρτηση βήτα που δίνεται από τη σχέση:

$$B_{G(t)}(a,b) = \int_0^{G(t)} x^{a-1} (1-x)^{b-1} dx, \quad a, b > 0 \quad (6.1)$$

και  $I_{G(t)}(a,b)$  είναι η μη πλήρης βήτα συνάρτηση ηλίκου.

### 6.1.2 Συνάρτηση πυκνότητας πιθανότητας

Η οικογένεια των Β-Γ κατανομών δημιουργήθηκε με σκοπό τη γενίκευση συνεχών κατανομών, αν και μπορεί να χρησιμοποιηθεί και για τη γενίκευση διακριτών. Η Β-Γ κατανομή θα έχει σ.π.π. ή σ.π. εάν η αρχική κατανομή  $G$  είναι απόλυτα συνεχής ή διακριτή, αντίστοιχα. Στο παρόν κεφάλαιο εστιάζουμε στις γενικευμένες συνεχείς κατανομές, επομένως ο ορισμός της σ.π.π. της Β-Γ κατανομής θα δίνεται από τη σχέση:

$$f(t) \equiv F'(t) = \frac{g(t)}{B(a,b)} G(t)^{a-1} [1-G(t)]^{b-1}, \quad a, b > 0, \quad (6.2)$$

όπου  $g(t)$  είναι η σ.π.π. της αρχικής κατανομής.

Οι Cordeiro and Barreto-Souza (2009) έδωσαν μια χρήσιμη επέκταση της σ.π.π. της Β-Γ κατανομής για να εξάγουν ορισμένες γενικές ιδιότητες της συγκεκριμένης οικογένειας κατανομών. Έστω  $b$  ένας μη-ακέραιος, πραγματικός αριθμός και  $|t| < 1$ . Θεωρώντας την επέκταση της δυναμοσειράς  $(1-t)^{b-1}$

$$(1-t)^{b-1} = \sum_{j=0}^{+\infty} \frac{(-1)^j \Gamma(b)}{\Gamma(b-j) j!} t^j, \quad (6.3)$$

η μη πλήρης συνάρτηση βήτα δύναται να εκφραστεί ως εξής:

$$B_i(a,b) = t^a \sum_{j=0}^{+\infty} \frac{(-1)^j \Gamma(b) t^j}{\Gamma(b-j) j! (j+a)}. \quad (6.4)$$

Σε περίπτωση που ο  $b$  είναι ακέραιος, τότε το άθροισμα γίνεται πεπερασμένο και ο δείκτης  $j$  σταματά στο  $b-1$ .

Έτσι με χρήση της σχέσης (6.4), η α.σ.κ. της Β-Γ κατανομής μπορεί να γραφεί ως εξής:

$$F(t) = \frac{1}{B(a,b)} \sum_{r=0}^{+\infty} w_r(a,b) G(t)^{a+r}, \quad (6.5)$$

όπου  $w_j(a,b) = \frac{(-1)^j \Gamma(b)}{\Gamma(b-j) j! (a+j)}$  σταθερά.

Επομένως, με χρήση της (6.5), η σ.π.π. μπορεί πλέον να γραφεί και ως:

$$f(t) = \frac{1}{B(a,b)} g(t) \sum_{r=0}^{+\infty} (a+r) w_r(a,b) G(t)^{a+r-1}. \quad (6.6)$$

Επιπρόσθετα, κάνοντας επανειλημμένη χρήση της σχέσης (6.3) για  $\alpha$  μη-ακέραιο πραγματικό αριθμό, είναι:

$$\begin{aligned} G(t)^\alpha &= \{1 - [1 - G(t)]\}^\alpha = \sum_{j=0}^{+\infty} \frac{(-1)^j \Gamma(\alpha+1)}{\Gamma(\alpha-j+1) j!} [1 - G(t)]^j \\ &= \sum_{j=0}^{+\infty} \sum_{r=0}^{+\infty} \frac{(-1)^{j+r} \Gamma(\alpha+1)}{\Gamma(\alpha-j+1) (j-r)! r!} G(t)^r \\ &= \sum_{r=0}^{+\infty} \sum_{j=r}^{+\infty} \frac{(-1)^{j+r} \Gamma(\alpha+1)}{\Gamma(\alpha-j+1) (j-r)! r!} G(t)^r \\ &= \sum_{r=0}^{+\infty} s_r(\alpha) G(t)^r, \end{aligned} \quad (6.7)$$

όπου  $s_r(\alpha) = \sum_{j=r}^{+\infty} \frac{(-1)^{j+r} \Gamma(\alpha+1)}{\Gamma(\alpha-j+1) (j-r)! r!}$ .

Εαν  $a$  μη-ακέραιος πραγματικός αριθμός, τότε με χρήση της (6.7) η α.σ.κ. της Β-Γ κατανομής (6.5) δίνεται να γραφεί ως:

$$F(t) = \frac{1}{B(a,b)} \sum_{r=0}^{+\infty} t_r(a,b) G(t)^r,$$

όπου  $t_r(a,b) = \sum_{l=0}^{+\infty} w_l(a,b) s_r(a+l)$ .

Με απλή διαφοροποίηση της παραπάνω σχέσης, προκύπτει μια άλλη μορφή της σ.π.π. να είναι:

$$f(t) = \frac{1}{B(a,b)} g(t) \sum_{r=0}^{+\infty} (r+1) t_{r+1}(a,b) G(t)^r. \quad (6.8)$$

Ο σκοπός αυτής της εναλλακτικής μορφής υπολογισμού της σ.π.π. όταν  $a$  μη-ακέραιος, είναι να ορισθεί η δυναμοσειρά με τη  $G(t)$  υψωμένη μόνο σε ακέραιες δυνάμεις. Εάν η  $G(t)$  δεν έχει κλειστή μορφή, υποτίθεται ότι επιδέχεται της επέκτασης:

$$G(t) = \sum_{k=0}^{+\infty} a_k t^{k+c},$$

όπου  $\{a_k\}_{k \geq 0}$  είναι μια ακολουθία πραγματικών αριθμών. Έτσι, για κάθε  $\alpha$  θετικό ακέραιο, θα είναι:

$$\left( \sum_{k=0}^{+\infty} a_k t^k \right)^\alpha = \sum_{k=0}^{+\infty} c_{\alpha,k} t^k, \quad (6.9)$$

όπου οι συντελεστές  $c_{\alpha,k}$  για  $k=1,2,\dots$  λαμβάνονται από την επαναληπτική εξίσωση:

$$c_{\alpha,k} = (ka_0)^{-1} \sum_{m=1}^i (\alpha m - k + m) a_m c_{\alpha,k-m}$$

και  $c_{\alpha,0} = a_0^\alpha$  (Gradshteyn and Ryzhik, 2007).

Έαν  $\alpha$  ακέραιος, από τη σχέση (6.9) προκύπτει ότι:

$$G(t)^\alpha = \sum_{k=0}^{+\infty} c_{\alpha,k} t^{k+c\alpha}. \quad (6.10)$$

Χρησιμοποιώντας την τελευταία σχέση (6.10), οι σχέσεις (6.6) και (6.8) δύνανται να ξαναγραφούν ως εξής:

$$f(t) = \frac{1}{B(a,b)} g(t) \sum_{r,k=0}^{+\infty} (a+r) w_r(a,b) c_{a+r-1,k} t^{k+c(a+r-1)} \quad (6.11)$$

και

$$f(t) = \frac{1}{B(a,b)} g(t) \sum_{r,k=0}^{+\infty} (r+1) t_{r+1}(a,b) c_{r,k} t^{k+cr}. \quad (6.12)$$

Οι σχέσεις υπολογισμού της σ.π.π. της Β-Γ κατανομής (6.11) και (6.12), παρουσιάστηκαν αρχικά από τους Cordeiro and Barreto-Souza (2009) και αποτελούν χρήσιμο εργαλείο για τη μελέτη μαθηματικών ιδιοτήτων της οικογένειας των Β-Γ κατανομών. Οι σχέσεις αυτές ορίζουν τη σ.π.π. της Β-Γ κατανομής, με χρήση της σ.π.π.  $g(t)$  της αρχικής κατανομής  $G$ , πολλαπλασιαζόμενη με μια άπειρη δυναμοσειρά του  $t$ .

### 6.1.3 Συναρτήσεις επιβίωσης και κινδύνου

Οι συναρτήσεις επιβίωσης και κινδύνου αποτελούν χρήσιμα εργαλεία σε μελέτες επιβίωσης και αξιοπιστίας, περιπτώσεις κατά τις οποίες το πεδίο ορισμού επικεντρώνεται στο  $\mathbb{R}^+$ .

Μια σημαντική ιδιότητα της μη πλήρους συνάρτησης βήτα είναι η εξής:

$$B_t(a,b) = B(a,b) - B_{1-t}(b,a).$$

Με χρήση αυτής της ιδιότητας, η συνάρτηση επιβίωσης μιας κατανομής που ανήκει στην οικογένεια των Β-Γ κατανομών θα δίνεται από τη σχέση:

$$S(t) = 1 - \frac{B_{G(t)}(a,b)}{B(a,b)} = \frac{B_{1-G(t)}(b,a)}{B(a,b)} = \frac{B_{S^*(t)}(b,a)}{B(a,b)},$$

όπου  $S^*(t) = 1 - G(t)$  είναι η συνάρτηση επιβίωσης της  $G$  αρχικής κατανομής. Παρατηρείται ότι η συνάρτηση επιβίωσης δίνεται από μια κανονικοποιημένη μη πλήρη συνάρτηση βήτα και εξαρτάται από το χρόνο  $t$  μόνο μέσω της  $S^*(t)$ .

Επιπρόσθετα, η συνάρτηση κινδύνου για οποιαδήποτε κατανομή της οικογένειας των Β-Γ δίνεται από τη σχέση:

$$h(t) = \frac{[1 - G(t)]^{b-1} G(t)^{a-1}}{B_{1-G(t)}(b, a)} g(t).$$

#### 6.1.4 Ροπές

Οι Cordeiro and Barreto-Souza (2009), εξήγαγαν μια μορφή για τον υπολογισμό της  $\nu$ -οστής τάξης ροπής της Β-Γ κατανομής, χρησιμοποιώντας τις σχέσεις (6.11) και (6.12) της σ.π.π..

Έστω  $X$  μια Β-Γ τυχαία μεταβλητή και  $Y$  μια τυχαία μεταβλητή με α.σ.κ.  $G(t)$ . Για  $a$  ακέραιο, από τη σχέση (6.11) έπεται ότι η  $\nu$ -οστής τάξης ροπή της Β-Γ κατανομής μπορεί να εκφραστεί από άπειρα αθροίσματα των ροπών της  $G$  κατανομής ως εξής:

$$E(X^\nu) = \frac{1}{B(a, b)} \sum_{j,k=0}^{+\infty} (a+j) w_j(a, b) c_{a+j-1, k} E(Y^{\nu+k+c(a+j-1)})$$

και από τη σχέση (6.12) για  $a$  ακέραιο, προκύπτει ότι:

$$E(X^\nu) = \frac{1}{B(a, b)} \sum_{r,k=0}^{+\infty} (r+1) t_{r+1}(a, b) c_{r, k} E(Y^{\nu+k+cr}).$$

#### 6.1.5 Εντροπίες

Στη μελέτη των Zografos and Balakrishnan (2009), η οποία αποτελεί την εκτενέστατη και πιο επίκαιρη αναφορά στις οικογένειες των βήτα και γάμμα γενικευμένων κατανομών, εκφράζεται τύπος υπολογισμού της

εντροπίας κατά Shannon. Έτσι η κατά Shannon εντροπία για μια τ.μ.  $T$  που ανήκει σε οποιαδήποτε κατανομή της οικογένειας των Β-Γ δύναται να υπολογιστεί από τη σχέση:

$$H_{Sh}(f) \equiv E[-\log f(T)] = \log B(a, b) - (a-1)[\Psi(a) - \Psi(a+b)] - (b-1)[\Psi(b) - \Psi(a+b)] - E_Y[\log f(F^{-1}(Y))],$$

όπου  $Y \sim Beta(a, b)$  και  $\Psi$  συμβολίζει τη συνάρτηση Δίγαμμα με

$$\Psi(x) = \frac{d}{dx} [\log \Gamma(x)] = \frac{\Gamma'(x)}{\Gamma(x)}.$$

## 6.2 Ειδικές περιπτώσεις κατανομών

Σε αυτήν την παράγραφο παρουσιάζονται διάφορες ειδικές κατανομές που ανήκουν στην οικογένεια των Β-Γ κατανομών και έχουν προταθεί τα τελευταία χρόνια στη βιβλιογραφία. Καταγράφονται συνολικά στον ακόλουθο Πίνακα 6.1, ενώ παρουσιάζονται για κάθε περίπτωση αναλυτικά οι συναρτήσεις πυκνότητας πιθανότητας, επιβίωσης και κινδύνου. Αξίζει να σημειωθεί ότι το δεύτερο συνθετικό της ονομασίας κάθε γενικευμένης κατανομής αντιστοιχεί στην χρησιμοποιούμενη κάθε φορά αρχική κατανομή  $G$ .

**Πίνακας 6.1** Ειδικές βήτα γενικευμένες κατανομές.

Κατανομή	Ερευνητές (Έτος)	Γεννήτορας $G(t)$
Βήτα - Κανονική	Eugene et al. (2002) και Razzaghi (2009)	$G(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$
Βήτα - Λογιστική	Brown et al. (2002) και Olapade (2004)	$G(t) = \frac{1}{1+e^{-\lambda t}}$
Βήτα - Gumbel	Nadarajah and Kotz (2004)	$G(t) = e^{-e^{-\frac{t-\mu}{\sigma}}}$
Βήτα - Fréchet	Nadarajah and Gupta (2004) και Barreto-	$G(t) = e^{-\left(\frac{t}{\sigma}\right)^{-\lambda}}$

	Souza et al. (2008)	
Βήτα - Weibull	Famoye et al. (2005), Lee et al. (2007), Zografos (2008) και Cordeiro et al. (2008)	$G(t) = 1 - e^{-(\lambda t)^c}$
Βήτα - Εκθετική	Nadarajah and Kotz (2006)	$G(t) = 1 - e^{-\lambda t}$
Βήτα - Γάμμα	Kong et al. (2007)	$G(t) = \frac{\Gamma_{t/\lambda}(\rho)}{\Gamma(\rho)}$
Βήτα - Pareto	Akinsete et al. (2008)	$G(t) = 1 - \left(\frac{t}{\theta}\right)^{-k}$
Βήτα - Γενικευμένη Λογιστική τύπου IV	Lemos de Morais et al. (2009)	$G(t) = \frac{B_{\frac{1}{1+e^{-t}}}(p, q)}{B(p, q)}$
Βήτα - Υπερβολική Τέμνουσα	Fischer and Vaughan (2010)	$G(t) = \frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)$
Βήτα - Γενικευμένη Ημι-κανονική	Pescim et al. (2010)	$G(t) = 2\Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] - 1$
Βήτα - Γενικευμένη Εκθετική	Barreto-Souza et al. (2010)	$G(t) = (1 - e^{-\lambda t})^\alpha$
Βήτα - Burr τύπου XII	Paranaiba et al. (2010)	$G(t) = 1 - \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}$
Βήτα - Γενικευμένη Pareto	Mahmoudi (2011)	$G(t) = \begin{cases} 1 - \left[1 - \frac{\xi(t-\mu)}{\sigma}\right]^{\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{(t-\mu)}{\sigma}}, & \xi = 0 \end{cases}$
Βήτα - Γενικευμένη Weibull	Singla et al. (2012)	$G(t) = \left[1 - e^{-(\lambda t)^\beta}\right]^\alpha$
Βήτα - Εκθετική Γεωμετρική	Bidram (2012)	$G(t) = \frac{1 - e^{-\beta t}}{1 - p e^{-\beta t}}$



### 6.2.1 Βήτα Κανονική (BN)

Η βήτα κανονική κατανομή ορίστηκε από τους Eugene et al. (2002), και η σημαντικότητά της δεν περιορίζεται απλώς στη γενίκευση της κανονικής κατανομής. Από την BN κατανομή, οι Eugene et al. (2002) έθεσαν το πλαίσιο δημιουργίας της οικογένειας των Β-Γ κατανομών. Η κατανομή και οι ιδιότητές της μελετήθηκαν αργότερα και από τον Razzaghi (2009).

Έστω  $\Phi(t)$  και  $\varphi(t)$  η α.σ.κ. και η σ.π.π. αντίστοιχα, της κανονικής κατανομής. Τότε από τη σχέση (6.2) προκύπτει η σ.π.π. της BN κατανομής να δίνεται από τη σχέση:

$$f(t) = \frac{1}{\sigma B(a,b)} \left[ \Phi\left(\frac{t-\mu}{\sigma}\right) \right]^{a-1} \left[ 1 - \Phi\left(\frac{t-\mu}{\sigma}\right) \right]^{b-1} \varphi\left(\frac{t-\mu}{\sigma}\right),$$

όπου  $t, a, b, \sigma > 0$  και  $\mu \in \mathbb{R}$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BN κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}(b,a)}{B(a,b)}$$

και

$$h(t) = \frac{\left[ 1 - \Phi\left(\frac{t-\mu}{\sigma}\right) \right]^{b-1} \left[ \Phi\left(\frac{t-\mu}{\sigma}\right) \right]^{a-1}}{B_{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}(b,a)} \varphi\left(\frac{t-\mu}{\sigma}\right).$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BN κατανομής μπορεί να λάβει μονοκόρυφη ή δικόρυφη μορφή, ανάλογα με τις τιμές των παραμέτρων της.

### 6.2.2 Βήτα Λογιστική (BL)

Η βήτα λογιστική κατανομή (ή log-F κατανομή) προτάθηκε από τους Brown et al. (2002) οι οποίοι μελέτησαν διάφορες μαθηματικές ιδιότητες της κατανομής, ενώ επιπρόσθετη έρευνα διεξήχθη από τον Olapade (2004), ο οποίος μελέτησε μια γενίκευση καθώς και τη σχέση της με άλλες κατανομές.

Έστω  $G(t) = \frac{1}{1+e^{-\lambda t}}$ , με  $t, \lambda > 0$  η α.σ.κ. της λογιστικής κατανομής.

Τότε από τη σχέση (6.2) προκύπτει η σ.π.π. της BL κατανομής να δίνεται από τη σχέση:

$$f(t) = \frac{\lambda}{B(a,b)} \frac{e^{-\lambda bt}}{(1+e^{-\lambda t})^{a+b}},$$

όπου  $t, a, b, \lambda > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BL κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{\frac{e^{-\lambda t}}{1+e^{-\lambda t}}}(b,a)}{B(a,b)}$$

και

$$h(t) = \frac{\lambda}{\frac{B_{\frac{e^{-\lambda t}}{1+e^{-\lambda t}}}(b,a)}{1+e^{-\lambda t}}} \frac{e^{-\lambda bt}}{(1+e^{-\lambda t})^{a+b}}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BL κατανομής μπορεί να λάβει γνησίως φθίνουσα μορφή.

### 6.2.3 Βήτα Gumbel (BGU)

Η βήτα Gumbel κατανομή μελετήθηκε από τους Nadarajah and Kotz (2004) και αποτελεί γενίκευση της κατανομής Gumbel που χρησιμοποιείται ευρέως στη μηχανολογία. Οι ερευνητές διατύπωσαν

εκφράσεις υπολογισμού των ροπών ν-οστής τάξης, διερεύνησαν ορισμένες ειδικές περιπτώσεις και μελέτησαν τη σ.π.π. της κατανομής.

Έστω  $G(t) = e^{-e^{-\frac{t-\mu}{\sigma}}}$ , με  $t, \mu \in \mathbb{R}$  και  $\sigma > 0$  η α.σ.κ. της Gumbel κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BGU κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{e^{-\frac{t-\mu}{\sigma}}}{\sigma B(a,b)} e^{-ae^{-\frac{t-\mu}{\sigma}}} \left\{ 1 - e^{-e^{-\frac{t-\mu}{\sigma}}} \right\}^{b-1},$$

όπου  $t, \mu \in \mathbb{R}$  και  $a, b, \sigma > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BGU κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1-e^{-e^{-\frac{t-\mu}{\sigma}}}}(b,a)}{B(a,b)}$$

και

$$h(t) = \frac{e^{-\frac{t-\mu}{\sigma}}}{\sigma B_{1-e^{-e^{-\frac{t-\mu}{\sigma}}}}(b,a)} e^{-ae^{-\frac{t-\mu}{\sigma}}} \left\{ 1 - e^{-e^{-\frac{t-\mu}{\sigma}}} \right\}^{b-1}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BGU κατανομής δύναται να λάβει διάφορες μορφές για συγκεκριμένες τιμές των παραμέτρων της.

#### 6.2.4 Βήτα Fréchet (BF)

Η βήτα Fréchet κατανομή εισήχθη από τους Nadarajah and Gupta (2004), οι οποίοι μελέτησαν ορισμένες ιδιότητες της σ.π.π. και της συνάρτησης κινδύνου και διατύπωσαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης. Στη συνέχεια, οι Barreto-Souza et al. (2008) μελέτησαν διαφορετικές εκφράσεις υπολογισμού των ροπών και ορισμένες επιπρόσθετες μαθηματικές ιδιότητες της κατανομής.

Έστω  $G(t) = e^{-\left(\frac{t}{\sigma}\right)^{-\lambda}}$ , με  $t, \sigma, \lambda > 0$  η α.σ.κ. της Fréchet κατανομής. Τότε από τη σχέση (6.2) προκύπτει η σ.π.π. της BF κατανομής να δίνεται από τη σχέση:

$$f(t) = \frac{\lambda \sigma^\lambda e^{-a\left(\frac{t}{\sigma}\right)^{-\lambda}} \left\{ 1 - e^{-\left(\frac{t}{\sigma}\right)^{-\lambda}} \right\}^{b-1}}{t^{1+\lambda} B(a, b)},$$

όπου  $t, a, b, \sigma, \lambda > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BF κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1-e^{-\left(\frac{t}{\sigma}\right)^{-\lambda}}}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{\lambda \sigma^\lambda e^{-a\left(\frac{t}{\sigma}\right)^{-\lambda}} \left\{ 1 - e^{-\left(\frac{t}{\sigma}\right)^{-\lambda}} \right\}^{b-1}}{t^{1+\lambda} B_{1-e^{-\left(\frac{t}{\sigma}\right)^{-\lambda}}}(b, a)}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BF κατανομής μπορεί να λάβει γνησίως φθίνουσα ή μονοκόρυφη μορφή, ανάλογα με τις τιμές των παραμέτρων της.

### 6.2.5 Βήτα Weibull (BW)

Η βήτα Weibull κατανομή προτάθηκε από τους Famoye et al. (2005). Επιπρόσθετη μελέτη της κατανομής διεξήχθη από τους Lee et al. (2007), οι οποίοι ερεύνησαν ορισμένες ιδιότητες της συνάρτησης κινδύνου, τις εντροπίες και την προσαρμογή της κατανομής σε ένα σύνολο λογοκριμένων δεδομένων. Επιπλέον μαθηματικές ιδιότητες της

κατανομής, όπως εκφράσεις των ροπών της, διατυπώθηκαν από τους Cordeiro et al. (2008).

Έστω  $G(t) = 1 - e^{-(\lambda t)^c}$ , με  $t, \lambda, c > 0$  η α.σ.κ. της Weibull κατανομής. Τότε από τη σχέση (6.2) προκύπτει η σ.π.π. της BW κατανομής να δίνεται από τη σχέση:

$$f(t) = \frac{c\lambda}{B(a,b)} (\lambda t)^{c-1} \left(1 - e^{-(\lambda t)^c}\right)^{a-1} e^{-b(\lambda t)^c},$$

όπου  $t, a, b, \lambda, c > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BE κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{e^{-(\lambda t)^c}}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{c\lambda^c t^{c-1} e^{-b(\lambda t)^c} \left(1 - e^{-(\lambda t)^c}\right)^{a-1}}{B_{e^{-(\lambda t)^c}}(b, a)}.$$

Οι Lee et al. (2007) απέδειξαν ότι η συνάρτηση κινδύνου της BW κατανομής είναι:

- σταθερή και ίση με  $b\lambda$ , όταν  $a = c = 1$ ,
- φθίνουσα όταν  $ac \leq 1$  και  $c \leq 1$ ,
- αύξουσα όταν  $ac \geq 1$  και  $c \geq 1$ ,
- κοίλη (bathtub) όταν  $ac < 1$  και  $c > 1$  και
- μονοκόρυφη όταν  $ac > 1$  και  $c < 1$ .

### 6.2.6 Βήτα Εκθετική (BE)

Η βήτα εκθετική κατανομή προτάθηκε από τους Nadarajah and Kotz (2006). Στην εργασία τους δόθηκαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης, ιδιότητες της συνάρτησης κινδύνου, διάφορα αποτελέσματα για την κατανομή του αθροίσματος BE τυχαίων

μεταβλητών, οι εκτιμητές μέγιστης πιθανοφάνειας της ΒΕ καθώς και ορισμένα ασυμπτωτικά αποτελέσματα.

Έστω  $G(t) = 1 - e^{-\lambda t}$ , με  $t, \lambda > 0$  η α.σ.κ. της εκθετικής κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της ΒΕ κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{\lambda}{B(a, b)} e^{-b\lambda t} (1 - e^{-\lambda t})^{a-1},$$

όπου  $t, a, b, \lambda > 0$ .

Παρατηρείται ότι για  $b = 1$ , προκύπτει η ειδική περίπτωση της υψωμένης σε δύναμη εκθετικής κατανομής (exponentiated exponential distribution) που μελετήθηκε από τους Gupta and Kundu (2001), ενώ για  $a = 1$  η ΒΕ κατανομή ταυτίζεται με την εκθετική με παράμετρο  $b\lambda$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της ΒΕ κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{e^{-\lambda t}}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{\lambda e^{-b\lambda t} (1 - e^{-\lambda t})^{a-1}}{B_{e^{-\lambda t}}(b, a)},$$

όπου  $B_{e^{-\lambda t}}(b, a)$  είναι η μη πλήρης συνάρτηση βήτα που ορίζεται από τη σχέση (6.1).

Αποδεικνύεται ότι η μορφή της συνάρτησης κινδύνου της ΒΝ κατανομής εξαρτάται μόνον από την παράμετρο  $a$ . Έτσι, όταν  $a < 1$ , η  $h(t)$  είναι γνησίως φθίνουσα, ενώ όταν  $a > 1$  η  $h(t)$  είναι γνησίως αύξουσα. Σε περίπτωση όπου  $a = 1$ , τότε η συνάρτηση κινδύνου είναι σταθερή.

### 6.2.7 Βήτα Γάμμα (BG)

Η βήτα γάμμα κατανομή προτάθηκε από τους Kong et al. (2007), οι οποίοι μελέτησαν διάφορες ιδιότητες του ορίου της σ.π.π. και της συνάρτησης κινδύνου, διατύπωσαν εκφράσεις υπολογισμού των ροπών όταν η παράμετρος μορφής  $a$  είναι ακέραιος αριθμός και εφάρμοσαν και μια προσαρμογή της κατανομής.

Έστω  $G(t) = \frac{\Gamma_{t/\lambda}(\rho)}{\Gamma(\rho)}$ , με  $t, \rho, \lambda > 0$  η α.σ.κ. της γάμμα κατανομής,

όπου  $\Gamma_t(\rho) = \int_0^t x^{\rho-1} e^{-x} dx$  είναι η μη πλήρης συνάρτηση γάμμα. Τότε

από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BG κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{t^{\rho-1} e^{-t/\lambda}}{B(a,b) \{\Gamma(\rho)\}^a \lambda^\rho} \{\Gamma_{t/\lambda}(\rho)\}^{a-1} \left\{ 1 - \frac{\Gamma_{t/\lambda}(\rho)}{\Gamma(\rho)} \right\}^{b-1},$$

όπου  $t, a, b, \rho, \lambda > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BG κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1-\frac{\Gamma_{t/\lambda}(\rho)}{\Gamma(\rho)}}(b,a)}{B(a,b)}$$

και

$$h(t) = \frac{t^{\rho-1} e^{-t/\lambda}}{B_{1-\frac{\Gamma_{t/\lambda}(\rho)}{\Gamma(\rho)}}(b,a) \{\Gamma(\rho)\}^a \lambda^\rho} \{\Gamma_{t/\lambda}(\rho)\}^{a-1} \left\{ 1 - \frac{\Gamma_{t/\lambda}(\rho)}{\Gamma(\rho)} \right\}^{b-1}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BG κατανομής δύναται να λάβει διάφορες μορφές (γνησίως αύξουσα, γνησίως φθίνουσα, κοίλη, μονοκόρυφη) για συγκεκριμένες τιμές των παραμέτρων της.

### 6.2.8 Βήτα Pareto (BP)

Η βήτα Pareto κατανομή εισήχθη από τους Akinsete et al. (2008), οι οποίοι μελέτησαν διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσαν εκφράσεις υπολογισμού της εντροπίας και των ροπών  $\nu$ -οστής τάξης, υπολόγισαν τις τυπικές αποκλίσεις για τη μέση τιμή και τη διάμεσο και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας μελέτησαν την προσαρμογή της κατανομής σε σύνολα δεδομένων.

Έστω  $G(t) = 1 - \left(\frac{t}{\theta}\right)^{-k}$ , με  $t \geq \theta$  και  $\theta, k > 0$  η α.σ.κ. της Pareto κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π. της BP κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{k}{\theta B(a, b)} \left\{ 1 - \left(\frac{t}{\theta}\right)^{-k} \right\}^{a-1} \left(\frac{t}{\theta}\right)^{-kb-1},$$

όπου  $t \geq \theta$  και  $a, b, \theta, k > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BP κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B\left(\frac{t}{\theta}\right)^{-k}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{k}{\theta B\left(\frac{t}{\theta}\right)^{-k}(b, a)} \left\{ 1 - \left(\frac{t}{\theta}\right)^{-k} \right\}^{a-1} \left(\frac{t}{\theta}\right)^{-kb-1}.$$

Αποδεικνύεται ότι η μορφή της συνάρτησης κινδύνου της BP κατανομής εξαρτάται μόνον από την παράμετρο  $a$ . Έτσι, όταν  $0 < a \leq 1$ , η  $h(t)$  είναι γνησίως φθίνουσα, ενώ όταν  $a > 1$  η  $h(t)$  είναι μονοκόρυφη.



### 6.2.9 Βήτα Γενικευμένη Λογιστική τύπου IV (BGL)

Η βήτα γενικευμένη λογιστική τύπου IV κατανομή εισήχθη από τους Lemos de Morais et al. (2009), οι οποίοι μελέτησαν διάφορες ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές και διατύπωσαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης.

Έστω  $G(t) = \frac{B_{\frac{1}{1+e^{-t}}}(p, q)}{B(p, q)}$ , με  $t, p, q > 0$  η α.σ.κ. της γενικευμένης λογιστικής τύπου IV κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BGL κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{\{B(p, q)\}^{1-a-b}}{B(a, b)} \frac{e^{-qt}}{(1+e^{-t})^{p+q}} \left\{ B_{\frac{1}{1+e^{-t}}}(p, q) \right\}^{a-1} \left\{ B_{\frac{e^{-t}}{1+e^{-t}}}(q, p) \right\}^{b-1},$$

όπου  $t, a, b, p, q > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BGL κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{\frac{e^{-t}}{1+e^{-t}}}(q, p)(b, a)}{B(p, q)} \frac{1}{B(a, b)}$$

και

$$h(t) = \frac{\{B(p, q)\}^{1-a-b}}{B_{\frac{e^{-t}}{1+e^{-t}}}(q, p)(b, a)} \frac{e^{-qt}}{(1+e^{-t})^{p+q}} \left\{ B_{\frac{1}{1+e^{-t}}}(p, q) \right\}^{a-1} \left\{ B_{\frac{e^{-t}}{1+e^{-t}}}(q, p) \right\}^{b-1}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BGL κατανομής μπορεί να λάβει γνησίως φθίνουσα ή μονοκόρυφη μορφή, ανάλογα με τις τιμές των παραμέτρων της.

### 6.2.10 Βήτα Υπερβολική Τέμνουσα (BHS)

Η βήτα υπερβολική τέμνουσα (beta hyperbolic secant) κατανομή προτάθηκε από τους Fischer and Vaughan (2010), οι οποίοι μελέτησαν διάφορες ιδιότητες της κατανομής, όπως οι ροπές αυτής και ειδικές και οριακές περιπτώσεις της, ενώ συνέκριναν την προσαρμογή της σε πραγματικά δεδομένα έναντι άλλων κατανομών.

Έστω  $G(t) = \frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)$ , με  $t, \sigma > 0$  και  $\mu \in \mathbb{R}$  η α.σ.κ. της υπερβολικής τέμνουσας κατανομής. Τότε από τη σχέση (6.2) προκύπτει η σ.π.π. της BHS κατανομής να δίνεται από τη σχέση:

$$f(t) = \frac{\left\{\frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)\right\}^{a-1} \left\{1 - \frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)\right\}^{b-1}}{\pi \sigma B(a, b) \cosh\left(\frac{t-\mu}{\sigma}\right)},$$

όπου  $t, a, b, \sigma > 0$  και  $\mu \in \mathbb{R}$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BHS κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1 - \frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{\left\{\frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)\right\}^{a-1} \left\{1 - \frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)\right\}^{b-1}}{\pi \sigma B_{1 - \frac{2}{\pi} \arctan\left(e^{\frac{t-\mu}{\sigma}}\right)}(b, a) \cosh\left(\frac{t-\mu}{\sigma}\right)}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BHS κατανομής είναι μονοκόρυφη για κάθε  $a, b > 0$ .

### 6.2.11 Βήτα Γενικευμένη Ημι-κανονική (BGHN)

Η βήτα γενικευμένη ημι-κανονική κατανομή εισήχθη από τους Pescim et al. (2010), οι οποίοι μελέτησαν διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας συνέκριναν την προσαρμογή της κατανομής σε πραγματικά δεδομένα έναντι άλλων κατανομών.

Έστω  $G(t) = 2\Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] - 1$ , με  $t, \alpha, \theta > 0$  η α.σ.κ. της γενικευμένης

ημι-κανονικής κατανομή, όπου  $\Phi(t) = \frac{1}{2} \left\{ 1 + \frac{2}{\sqrt{\pi}} \int_0^{t/\sqrt{2}} e^{-x^2} dx \right\}$ . Τότε από

τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BGHN κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{t}\right) \left(\frac{t}{\theta}\right)^\alpha e^{-\frac{1}{2}\left(\frac{t}{\theta}\right)^{2\alpha}} 2^{b-1} \left\{ 2\Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] - 1 \right\}^{a-1} \left\{ 1 - \Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] \right\}^{b-1}}{B(a, b)},$$

όπου  $t, a, b, \alpha, \theta > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BGHN κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{1}{B(a, b)} \int_{2\Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] - 1}^{+\infty} x^{a-1} (1-x)^{b-1} dx$$

και

$$h(t) = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{t}\right) \left(\frac{t}{\theta}\right)^\alpha e^{-\frac{1}{2}\left(\frac{t}{\theta}\right)^{2\alpha}} 2^{b-1} \left\{ 2\Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] - 1 \right\}^{a-1} \left\{ 1 - \Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] \right\}^{b-1}}{B(a, b) \left\{ 1 - I_{2\Phi\left[\left(\frac{t}{\theta}\right)^\alpha\right] - 1}(a, b) \right\}}$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BGHN κατανομής μπορεί να λάβει γνησίως φθίνουσα ή αύξουσα μορφή, ανάλογα με τις τιμές των παραμέτρων της.

### 6.2.12 Βήτα Γενικευμένη Εκθετική (BGE)

Η βήτα γενικευμένη εκθετική κατανομή εισήχθη από τους Barreto-Souza et al. (2010), οι οποίοι μελέτησαν διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας συνέκριναν την προσαρμογή της κατανομής σε πραγματικά δεδομένα έναντι άλλων κατανομών, που προκύπτουν ως ειδικές περιπτώσεις αυτής.

Έστω  $G(t) = (1 - e^{-\lambda t})^\alpha$ , με  $t, \alpha, \lambda > 0$  η α.σ.κ. της γενικευμένης εκθετικής κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BGE κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{\alpha\lambda}{B(a,b)} e^{-\lambda t} (1 - e^{-\lambda t})^{\alpha(a-1)} \left\{ 1 - (1 - e^{-\lambda t})^\alpha \right\}^{b-1},$$

όπου  $t, a, b, \alpha, \lambda > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BGE κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1-(1-e^{-\lambda t})^\alpha}(b,a)}{B(a,b)}$$

και

$$h(t) = \frac{\alpha\lambda e^{-\lambda t} (1 - e^{-\lambda t})^{\alpha(a-1)} \left\{ 1 - (1 - e^{-\lambda t})^\alpha \right\}^{b-1}}{B_{1-(1-e^{-\lambda t})^\alpha}(b,a)}.$$

Αποδεικνύεται ότι η μορφή της συνάρτησης κινδύνου της BGE κατανομής εξαρτάται μόνον από την παράμετρο  $\alpha$ . Έτσι, όταν  $\alpha < 1$ , η

$h(t)$  είναι φθίνουσα, ενώ όταν  $\alpha > 1$  η  $h(t)$  είναι αύξουσα. Σε περίπτωση όπου  $\alpha = 1$ , τότε η συνάρτηση κινδύνου είναι σταθερή.

### 6.2.13 Βήτα Burr τύπου XII (BB)

Η βήτα Burr τύπου XII κατανομή εισήχθη από τους Paranaíba et al. (2010), οι οποίοι μελέτησαν διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας συνέκριναν την προσαρμογή της κατανομής σε πραγματικά δεδομένα έναντι άλλων κατανομών.

Έστω  $G(t) = 1 - \left[ 1 + \left( \frac{t}{s} \right)^c \right]^{-k}$ , με  $t, k, c, s > 0$  η α.σ.κ. της Burr τύπου XII κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BB κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{ckt^{c-1}}{s^c B(a, b)} \left\{ 1 + \left( \frac{t}{s} \right)^c \right\}^{-k(b-1)} \left\{ 1 - \left[ 1 + \left( \frac{t}{s} \right)^c \right]^{-k} \right\}^{a-1},$$

όπου  $t, a, b, k, c, s > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BB κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{\left[ 1 + (t/s)^c \right]^{-k}}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{ckt^{c-1} \left[ 1 + \left( \frac{t}{s} \right)^c \right]^{-k(b-1)} \left\{ 1 - \left[ 1 + \left( \frac{t}{s} \right)^c \right]^{-k} \right\}^{a-1}}{s^c B_{\left[ 1 + (t/s)^c \right]^{-k}}(b, a)}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BB κατανομής μπορεί να λάβει γνησίως φθίνουσα, γνησίως αύξουσα ή μονοκόρυφη μορφή ανάλογα με τις τιμές των παραμέτρων της.

#### 6.2.14 Βήτα Γενικευμένη Pareto (BGP)

Η βήτα γενικευμένη Pareto κατανομή εισήχθη από τον Mahmoudi (2011), ο οποίος μελέτησε διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσε εκφράσεις υπολογισμού των ροπών ν-οστής τάξης και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας συνέκρινε την προσαρμογή της κατανομής σε πραγματικά δεδομένα έναντι άλλων κατανομών. Η συγκεκριμένη κατανομή αποτελεί γενίκευση της βήτα Pareto κατανομής που μελετήθηκε από τους Akinsete et al. (2008).

$$\text{Έστω } G(t) = \begin{cases} 1 - \left[ 1 - \frac{\xi(t-\mu)}{\sigma} \right]^{\frac{1}{\xi}} & , \xi \neq 0 \\ 1 - e^{-\frac{(t-\mu)}{\sigma}} & , \xi = 0 \end{cases}, \text{ με } t \geq \xi \text{ για } \xi \leq 0, \text{ ή}$$

$\mu < t < \mu + \frac{\sigma}{\xi}$  για  $\xi > 0$  και  $\xi, \mu \in \mathbb{R}, \sigma > 0$ , η α.σ.κ. της γενικευμένης

Pareto κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BGP κατανομής δίνεται από τη σχέση:

$$f(t) = \begin{cases} \frac{1}{\sigma B(a,b)} \left\{ 1 - \frac{\xi(t-\mu)}{\sigma} \right\}^{\frac{b}{\xi}-1} \left\{ 1 - \left[ 1 - \frac{\xi(t-\mu)}{\sigma} \right]^{\frac{1}{\xi}} \right\}^{a-1} & , \xi \neq 0 \\ \frac{1}{\sigma B(a,b)} e^{-\frac{b(t-\mu)}{\sigma}} \left\{ 1 - e^{-\frac{(t-\mu)}{\sigma}} \right\}^{a-1} & , \xi = 0 \end{cases},$$

όπου  $t \geq \xi$  για  $\xi \leq 0$  ή  $\mu < t < \mu + \frac{\sigma}{\xi}$  για  $\xi > 0$  και  $\xi, \mu \in \mathbb{R}, a, b, \sigma > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BGP κατανομής δίνονται από τις σχέσεις:

$$S(t) = \begin{cases} 1 - I_{1 - \left[1 - \frac{\xi(t-\mu)}{\sigma}\right]^{\frac{1}{\xi}}} (a, b) & , \xi \neq 0 \\ 1 - I_{1 - e^{-\frac{(t-\mu)}{\sigma}}} (a, b) & , \xi = 0 \end{cases}$$

και

$$h(t) = \begin{cases} \frac{\left\{1 - \frac{\xi(t-\mu)}{\sigma}\right\}^{\frac{b}{\xi}-1} \left\{1 - \left[1 - \frac{\xi(t-\mu)}{\sigma}\right]^{\frac{1}{\xi}}\right\}^{a-1}}{\sigma B_{\left[1 - \frac{\xi(t-\mu)}{\sigma}\right]^{\frac{1}{\xi}}} (b, a)} & , \xi \neq 0 \\ \frac{e^{-\frac{b(t-\mu)}{\sigma}} \left\{1 - e^{-\frac{(t-\mu)}{\sigma}}\right\}^{a-1}}{\sigma B_{e^{-\frac{(t-\mu)}{\sigma}}} (b, a)} & , \xi = 0 \end{cases} .$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BGP κατανομής μπορεί να λάβει γνησίως φθίνουσα, γνησίως αύξουσα, μονοκόρυφη ή κοίλη μορφή, ανάλογα με τις τιμές των παραμέτρων της.

### 6.2.15 Βήτα Γενικευμένη Weibull (BGW)

Η βήτα γενικευμένη Weibull κατανομή εισήχθη από τους Singla et al. (2012), οι οποίοι μελέτησαν διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσαν εκφράσεις υπολογισμού των ροπών ν-οστής τάξης και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας συνέκριναν την προσαρμογή της κατανομής σε πραγματικά δεδομένα έναντι άλλων κατανομών.

Έστω  $G(t) = \left[1 - e^{-(\lambda t)^\beta}\right]^\alpha$ , με  $t, \alpha, \beta, \lambda > 0$  η α.σ.κ. της γενικευμένης Weibull κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BGW κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{\alpha\beta\lambda^\beta t^{\beta-1} e^{-(\lambda t)^\beta}}{B(a, b)} \left\{1 - e^{-(\lambda t)^\beta}\right\}^{\alpha(a-1)} \left\{1 - \left[1 - e^{-(\lambda t)^\beta}\right]^\alpha\right\}^{b-1},$$

όπου  $t, a, b, p, q > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BGW κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{1 - \left[1 - e^{-(\lambda t)^\beta}\right]^\alpha}(b, a)}{B(a, b)}$$

και

$$h(t) = \frac{\alpha\beta\lambda^\beta t^{\beta-1} e^{-(\lambda t)^\beta}}{B_{1 - \left[1 - e^{-(\lambda t)^\beta}\right]^\alpha}(b, a)} \left\{1 - e^{-(\lambda t)^\beta}\right\}^{\alpha(a-1)} \left\{1 - \left[1 - e^{-(\lambda t)^\beta}\right]^\alpha\right\}^{b-1}.$$

Αποδεικνύεται ότι η συνάρτηση κινδύνου της BGW κατανομής μπορεί να λάβει γνησίως φθίνουσα, γνησίως αύξουσα, μονοκόρυφη ή κοίλη μορφή, ανάλογα με τις τιμές των παραμέτρων της.

### 6.2.16 Βήτα Εκθετική Γεωμετρική (BEG)

Η βήτα εκθετική γεωμετρική κατανομή εισήχθη από τον Bidram (2012), ο οποίος μελέτησε διάφορες μαθηματικές ιδιότητες της κατανομής, τη σχέση της με άλλες κατανομές, διατύπωσε εκφράσεις υπολογισμού των ροπών  $n$ -οστής τάξης και με χρήση των εκτιμητών μέγιστης πιθανοφάνειας συνέκρινε την προσαρμογή της κατανομής σε πραγματικά δεδομένα έναντι άλλων κατανομών.



Έστω  $G(t) = \frac{1 - e^{-\beta t}}{1 - pe^{-\beta t}}$ , με  $t, \beta, p > 0$  η α.σ.κ. της εκθετικής γεωμετρικής κατανομής. Τότε από τη σχέση (6.2) προκύπτει ότι η σ.π.π. της BEG κατανομής δίνεται από τη σχέση:

$$f(t) = \frac{(1-p)^b}{B(a,b)} \beta e^{-b\beta t} (1 - e^{-\beta t})^{a-1} (1 - pe^{-\beta t})^{-(a+b)},$$

όπου  $t, a, b, \beta, p > 0$ .

Επιπλέον οι συναρτήσεις επιβίωσης και κινδύνου της BEG κατανομής δίνονται από τις σχέσεις:

$$S(t) = \frac{B_{\frac{(1-p)e^{-\beta t}}{1-pe^{-\beta t}}}(b,a)}{B(a,b)}$$

και

$$h(t) = \frac{(1-p)^b \beta e^{-b\beta t} (1 - e^{-\beta t})^{a-1} (1 - pe^{-\beta t})^{-(a+b)}}{B_{\frac{(1-p)e^{-\beta t}}{1-pe^{-\beta t}}}(b,a)}.$$

Επιπλέον, ο Bidram απέδειξε ότι η συνάρτηση κινδύνου της BEG κατανομής είναι:

- φθίνουσα, όταν  $a \leq 1$ ,
- αύξουσα, όταν  $a > 1$  και  $p \leq \frac{a-1}{a+b}$  και
- μονοκόρυφη, όταν  $a > 1$  και  $p > \frac{a-1}{a+b}$ .

### 6.3 Προτάσεις για μελλοντική έρευνα

Η οικογένεια των γενικευμένων κατανομών που παρουσιάστηκε σε αυτό το κεφάλαιο, αποτελεί τα τελευταία χρόνια το επίκεντρο της μελέτης των ερευνητών στο χώρο της ανάλυσης επιβίωσης. Αυτό δικαιολογείται από το γεγονός ότι οι συγκεκριμένες κατανομές με τις ποικίλες ιδιότητες, διευρύνουν το πεδίο εφαρμογής των αρχικών υπάρχουσών κατανομών και βελτιστοποιούν το ρόλο τους στην περιγραφή σύνθετων πραγματικών καταστάσεων. Οι δυνατότητες επέκτασης των συγκεκριμένων κατανομών είναι ακόμη πολλές και για το λόγο αυτό προτείνεται η στροφή του ενδιαφέροντος μελλοντικών ερευνητών προς αυτήν την κατεύθυνση. Αρκετές γενικευμένες βήτα και γάμμα κατανομές παραμένουν δίχως πλήρη μελέτη, ενώ το ερευνητικό πεδίο διαστέλλεται σε σημαντικό βαθμό αν αναλογιστεί κανείς τη χρήση διδιάστατων γενικευμένων κατανομών και την πληθώρα των πρόσθετων εφαρμογών που θα μπορούσαν αυτές να έχουν. Εξάλλου, ένα από τα βασικά προτερήματα της ανάλυσης επιβίωσης, αποτελεί το γεγονός ότι πραγματεύεται με έννοιες άμεσα εφαρμόσιμες σε πραγματικές καταστάσεις, η τάση των οποίων κρατά τους ερευνητές συνεχώς σε εγρήγορση, παροτρύνοντάς τους να μελετούν μεθοδολογίες περιγραφής, ερμηνείας και αντιμετώπισής τους.

# Περίληψη

## Κατανομές Πιθανότητας στην Ανάλυση Επιβίωσης

Η ανάλυση επιβίωσης αποτελεί ταυτόσημη έννοια της ανάλυσης του χρονικού διαστήματος που προηγείται της καταγραφής ενός συμβάντος. Ως όρος χρησιμοποιείται κατά κύριο λόγο σε βιοϊατρικές επιστήμες, όπου το ενδιαφέρον επικεντρώνεται στην παρατήρηση του χρόνου ζωής είτε των ασθενών είτε των πειραματόζωων. Η ανάλυση του χρονικού διαστήματος που μεσολαβεί μέχρι την καταγραφή ενός συμβάντος έχει επίσης χρησιμοποιηθεί ευρέως στις κοινωνικές επιστήμες, όπου το ενδιαφέρον εστιάζεται σε γεγονότα όπως η αλλαγή θέσεως εργασίας, ο γάμος, η γέννηση παιδιών και ούτω καθεξής. Επιπρόσθετα, οι επιστήμες της μηχανικής έχουν συμβάλει στην ανάπτυξη της ανάλυσης επιβίωσης. Στον επιστημονικό αυτό κλάδο συνηθίζεται να χρησιμοποιείται ο όρος «ανάλυση αξιοπιστίας», μιας και η κύρια εφαρμογή του είναι η μοντελοποίηση του χρόνου που απαιτείται για την αποτυχία μηχανημάτων ή ηλεκτρονικών εξαρτημάτων. Η μελέτη όλων αυτών των επιστημονικών εξειδικεύσεων έχει ως επί το πλείστον ενοποιηθεί στο πεδίο της ανάλυσης επιβίωσης.

Γίνεται εύκολα αντιληπτό ότι η περιγραφή και η μοντελοποίηση του χρόνου ζωής, αποτελεί το κύριο ζητούμενο σε πολλές εκφάνσεις της ανθρώπινης ζωής και δραστηριότητας. Ως εκ τούτου, οι μεθοδολογίες που αναπτύσσονται στην ανάλυση επιβίωσης αποτελούν πολύτιμο εργαλείο για ερευνητές διαφόρων επιστημονικών κλάδων, όπως της μηχανικής, της οικονομίας, της ιατρικής και των υπόλοιπων βιολογικών επιστημών.

Η παρούσα διδακτορική διατριβή πραγματεύεται θέματα που αφορούν σε συνεχείς κατανομές πιθανότητας που έχουν άμεση εφαρμογή στην ανάλυση επιβίωσης. Στο **πρώτο κεφάλαιο**, παρουσιάζεται μια σύντομη

ιστορική αναδρομή του ερευνητικού πεδίου της ανάλυσης επιβίωσης, καταγράφεται η ορολογία και οι βασικές έννοιες συνεχών κατανομών χρόνων ζωής και περιγράφονται οι διάφορες χαρακτηριστικές συναρτήσεις και ιδιότητες που αυτές έχουν. Ειδική βαρύτητα δίνεται στη συνάρτηση κινδύνου και τη σημασία της στην περιγραφή της ικανότητας επιβίωσης ενός πληθυσμού. Επιπρόσθετα, παρουσιάζονται τρόποι παραγωγής κατανομών χρόνων ζωής, η αναλυτική μεθοδολογία μελέτης τους, καθώς και οι πιο χαρακτηριστικές κατανομές με σταθερές, μονοκόρυφες ή κοίλες συναρτήσεις κινδύνου.

Στο **δεύτερο κεφάλαιο**, παρουσιάζεται μια κατανομή ανάλυσης επιβίωσης τριών παραμέτρων με αύξουσα, φθίνουσα ή κοίλη συνάρτηση κινδύνου. Μελετώνται διάφορες ιδιότητές της, όπως οι τρόποι παραγωγής της, οι σχέσεις της με άλλες κατανομές, η μονοτονία της συνάρτησης πυκνότητας πιθανότητας (σ.π.π.) της, οι ροπές της και οι αντίστοιχες συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας και τους πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας, ενώ για τον υπολογισμό των εκτιμητών γίνεται χρήση της γλώσσας προγραμματισμού στατιστικών εφαρμογών R. Το κεφάλαιο ολοκληρώνεται με συγκριτικά σε σχέση με άλλες κατανομές, αποτελέσματα από την προσαρμογή της σε πραγματικά σύνολα δεδομένων. Η κατανομή και τα παραγόμενα από τη μελέτη της αποτελέσματα, παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2011).

Στο **τρίτο κεφάλαιο**, μελετάται ένα μοντέλο ανάλυσης επιβίωσης τριών παραμέτρων με μονότονες, μονοκόρυφες και τροποποιημένες κοίλες συναρτήσεις κινδύνου. Ερευνώνται διάφορες μαθηματικές και στατιστικές ιδιότητές του και παρουσιάζεται η φυσική διεργασία παραγωγής του, μέσα από τη θεωρία ανταγωνιστικών κινδύνων. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας. Τέλος, παρουσιάζονται εφαρμογές

του μοντέλου σε δεδομένα που απορρέουν από πραγματικές καταστάσεις και η προσαρμογή του σε αυτά συγκρίνεται με άλλη εναλλακτική κατανομή. Τα αποτελέσματα αυτού του κεφαλαίου παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2010).

Στο **τέταρτο κεφάλαιο**, εισάγεται μια οικογένεια κατανομών μέσω της οποίας δύναται να γενικευτεί μια κατανομή ανάλυσης επιβίωσης, με την προσθήκη μιας επιπλέον παραμέτρου, σε αναλογία με τη μελέτη των Marshall and Olkin (1997). Στη συνέχεια, η οικογένεια κατανομών χρησιμοποιείται για την ανάπτυξη μιας τετρα-παραμετρικής τροποποιημένης επέκτασης της κατανομής Weibull, με αύξουσα, φθίνουσα ή κοίλη συνάρτηση κινδύνου. Μελετώνται διάφορες ιδιότητες της, όπως οι σχέσεις της με άλλες κατανομές, η μονοτονία της σ.π.π. της, οι ροπές της και οι αντίστοιχες συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας και τους πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας, ενώ για τον υπολογισμό των εκτιμητών γίνεται χρήση της γλώσσας προγραμματισμού στατιστικών εφαρμογών R. Το κεφάλαιο ολοκληρώνεται με συγκριτικά σε σχέση με άλλες κατανομές αποτελέσματα από την προσαρμογή της σε πραγματικά σύνολα δεδομένων. Η οικογένεια κατανομών καθώς και η αναλυτικά μελετώμενη κατανομή και τα παραγόμενα αυτής αποτελέσματα, παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2012).

Στο **πέμπτο κεφάλαιο**, παρουσιάζεται μια κατανομή ανάλυσης επιβίωσης τριών παραμέτρων με αύξουσα, φθίνουσα ή μονοκόρυφη συνάρτηση κινδύνου. Μελετώνται διάφορες ιδιότητες της, όπως οι τρόποι παραγωγής της, οι σχέσεις της με άλλες κατανομές, η μονοτονία της σ.π.π. της, οι ροπές της, η εντροπία και οι αντίστοιχες συναρτήσεις επιβίωσης, κινδύνου και μέσου υπολειπόμενου χρόνου ζωής. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας και τους πίνακες παρατηρούμενης

και αναμενόμενης πληροφορίας, ενώ για τον υπολογισμό των εκτιμητών γίνεται χρήση της γλώσσας προγραμματισμού στατιστικών εφαρμογών R. Το κεφάλαιο ολοκληρώνεται με εφαρμογές της κατανομής σε πραγματικά σύνολα δεδομένων. Η κατανομή και τα παραγόμενα από τη μελέτη της αποτελέσματα, παρουσιάζονται συνοπτικά στην εργασία των Pappas, Adamidis and Loukas (2013).

Στο **έκτο κεφάλαιο**, παρουσιάζονται οι βήτα γενικευμένες κατανομές πιθανότητας. Αναφέρεται ο τρόπος παραγωγής τους και περιγράφονται οι βασικές μαθηματικές ιδιότητές τους. Επιπρόσθετα, γίνεται αναφορά σε διάφορες κατανομές που ανήκουν στην οικογένεια και έχουν προταθεί μέχρι σήμερα στη βιβλιογραφία, ενώ δίνονται και προτάσεις για μελλοντική έρευνα στο συγκεκριμένο πεδίο.

Τέλος, παρατίθεται περίληψη της διδακτορικής διατριβής στην ελληνική και αγγλική γλώσσα, και η σχετική βιβλιογραφία που χρησιμοποιήθηκε κατά τη διάρκεια εκπόνησης της έρευνας.

# Abstract

## Probability Distributions in Survival Analysis

Survival analysis is identical to meaning analysis of the period preceding the recording of an event. As a term, it is primarily used in biomedical sciences, where the focus is on the observation of life of patients or laboratory animals. The analysis of the time leading up to an event record has also been widely used in the social sciences, where the focus is on events such as job change, marriage, birth of children and so on. Additionally, the engineering sciences have contributed to the development of survival analysis. In this scientific field is customary to use the term "reliability analysis", since the main application is the modeling of the time required for the failure of machinery or electronics. The study of all these scientific specialization is mostly consolidated in the area of survival analysis.

It is easily understood that the description and modeling of life is the main issue in many aspects of human life and activity. Therefore, the methodologies developed in survival analysis are a valuable tool for researchers in various disciplines such as engineering, economics, medicine and other life sciences.

This thesis addresses issues related to continuous probability distributions that have direct application in survival analysis. In this context, the **first chapter** provides a brief historical overview of the research field of survival analysis, record the terminology and basic concepts of continuous distributions of lifetimes and describes the various characteristic functions and properties that they have. Special attention is given to the hazard function and its importance in the description of the survivability of a population. Additionally, ways to produce lifetime distributions, the analytical methodology for their study, and the most characteristic distributions with fixed, unimodal or concave hazard functions, are presented in the same chapter.

In the **second chapter**, a three-parameter lifetime distribution with ascending, descending or concave hazard function is introduced. Various statistical properties and reliability aspects, such as ways of production, relations with other distributions, the monotony of the probability density function (pdf), its moments and the corresponding survival, hazard and mean residual life functions are studied. The parameter estimation problem is treated by the maximum likelihood method and observed and expected information tables. The free programming language for statistical applications R is used for calculating the estimators. Finally, the model is fitted to real datasets and compared with alternative models. The distribution and obtained from the study results, are included in Pappas, Adamidis and Loukas (2011).

In the **third chapter**, we introduce and study a three-parameter survival analysis model, with monotonous, unimodal and modified-concave hazard functions. Various mathematical and statistical properties are investigated and the physical production process, through the theory of competing risks, is presented. The parameter estimation problem is treated with the maximum likelihood method. Applications of the model to real data are derived and comparisons with an alternative distribution are given. The results of this chapter are included in Pappas, Adamidis and Loukas (2010).

In the **fourth chapter** a family of lifetime distributions, which may generalize a survival distribution with the addition of an extra parameter, in parallel with the study of Mashall and Olkin (1997), is introduced. Subsequently, the family of distributions is used to develop a modified four-parameter expansion of Weibull distribution, with ascending, descending or concave hazard function. Various properties, such as relationships with other distributions, the monotony of the pdf, its moments and the corresponding survival, hazard and mean residual life functions are studied. The parameter estimation problem is treated by the maximum likelihood method, the observed and expected information tables are derived, while the programming language for statistical



applications R is used for calculating the estimators. The chapter concludes with results obtained from fitting the proposed model, as well as other previously studied models, to real data. This family of distributions and the particular model studied analytically are presented in Pappas, Adamidis and Loukas (2012).

In the **fifth chapter**, a three-parameter lifetime distribution with ascending, descending or unimodal hazard function is introduced. Various properties, such as ways of production, relations with other distributions, the monotony of the pdf, its moments, its entropy and the corresponding survival, hazard and mean residual life functions are studied. The parameter estimation problem is treated by the maximum likelihood method, the observed and expected information tables are given, while the programming language for statistical applications R is used for calculating the estimators. The chapter concludes with applications of fitting the model to real datasets. The distribution and the results obtained are included in Pappas, Adamidis and Loukas (2012).

The **sixth chapter** presents a review on beta generalized probability distributions. Their production method is introduced and their basic mathematical properties are described. Additionally, various lifetime distributions that belong to the beta generalized family that have been proposed so far in the literature are presented, together with suggestions for future research in this field.

By completing, a summary of the thesis in Greek and English language and the literature that was used throughout the research development are given.



# Βιβλιογραφία

- Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36 (1), 106-108.
- Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters*, 39, 35-42.
- Adamidis, K., Dimitrakopoulou, T. and Loukas S. (2005). On an extension of the exponential-geometric distribution. *Statistics and Probability Letters*, 73, 259-269.
- Akinsete, A., Famoye, F. and Lee, C. (2008). The beta-Pareto distribution. *Statistics*, 42 (6), 547-563.
- Amoroso, L. (1925). Ricerche intorno alla curva dei redditi. *Annali di matematica pura e applicata*, 2 (4), 123-157.
- Andersen, P. K., Borgan, O., Gill, R. G. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Arnold, B. C., Balakrishnan, N. and Natarajah, H. N. (1992). *First Course in Order Statistics*. John Wiley & Sons, New York.
- Barlow, R. E., Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models, To Begin With*. Silver Spring, MD.
- Barreto-Souza, W. and Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. *Statistics and Probability Letters*, 79, 2493-2500.
- Barreto-Souza, W., Lemos de Moraes, A. and Cordeiro G., M. (2008). The Weibull-Geometric Distribution. *Preprint: arXiv:0809.2703v1*.
- Barreto-Souza, Cordeiro, G. M. and Simas, A. B. (2008). Some results for Beta-Fréchet distribution. *Preprint: arXiv:0809.1873v1*.

- Barreto-Souza, W., Santos, A. H. S. and Cordeiro, G. M. (2010). The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 80 (2), 159-172.
- Bidram, H. (2012). The Beta Exponential-Geometric Distribution. *Communications in Statistics – Simulation and Computation*, 41 (9), 1606-1622.
- Block, H. W., Savits, T. H. and Singh, H. (1998). The Reversed Hazard Rate Function. *Probability in the Engineering and Informational Sciences*, 12 (1), 69-90.
- Chahkandi, M. and Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, 53, 4433-4440.
- Chandra, N. K. and Roy, D. (2001). Some results on reversed hazard rate. *Probability in the Engineering and Informational Sciences*, 15 (1), 95-102.
- Chen Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, 49, 155-161.
- Ciumara, R. and Preda, V. (2009). The Weibull-Logarithmic distribution in lifetime analysis and its properties. *The XIII International Conference: "Applied Stochastic Models and Data Analysis"*, Lithuania.
- Cordeiro, G. M., Simas, A. B. and Stosic, B. D. (2008). Explicit expressions for moments of the beta Weibull distribution. *Preprint: arXiv:0809.1860v1*.
- Dahiya, R. C. and Gurland, J. (1972). Goodness of t-tests for gamma and exponential distributions. *Technometrics*, 14, 791-801.

- Dimitrakopoulou, T., Adamidis, K. and Loukas S. (2007). A lifetime distribution with an upside-down bathtub-shaped hazard function. *IEEE Transactions on Reliability*, 56 (2), 308-311.
- Ebrahimi, N. (1991). On estimating change point in a mean residual life function. *Sankhyā A*, 53 (2), 206-219.
- Eugene, N., Lee, C. and Famoye, F. (2002). Beta-Normal distribution and its applications. *Communications in Statistics – Theory and Methods*, 31 (4), 497-512.
- Famoye, F., Lee, C. and Eugene, N. (2003). Beta-Normal Distribution: Bimodality Properties and Application. *Journal of Modern Applied Statistical Methods*, 2, 314-326.
- Famoye, F., Lee, C. and Olumolade, O. (2005). The Beta-Weibull Distribution. *Journal of Statistical Theory and Applications*, 4, 121-136.
- Finkelstein, M. S. (2002). On the reversed hazard rate. *Reliability Engineering and System Safety*, 78 (1), 71-75.
- Fischer, M. J. and Vaughan, D. (2010). The Beta-Hyperbolic Secant Distribution. *Austrian Journal of Statistics*, 39, 245-258.
- Ghitany, M. E., Kotz, S. and Xie, M. (2005). On some reliability measures and their stochastic orderings for the Topp-Leone distribution. *Journal of Applied Statistics*, 32 (7), 715-722.
- Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75, 667-672.
- Gleser, L. J. (1989). The gamma distribution as a mixture of exponential distributions. *The American Statistician*, 43 (2), 115-117.

- Glucksberg, H., Cheever, M. A., Farewell, V. T., Fefer, A., Sale, A. and Thomas, E. D. (1981). High dose combination chemotherapy for acute nonlymphoblastic leukaemia adults. *Cancer*, 48, 1073-1081.
- Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series and Products* (7<sup>th</sup> edition Edited by Jeffrey, A. and Zwillinger, D.). Elsevier, San Diego.
- Gupta, A. and Nadarajah, S. (2004). On the moments of the beta normal distribution. *Communication in Statistics – Theory and Methods*, 31, 1-13.
- Gupta, R. D. and Kundu, D. (2001). Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, 43 (1), 117-130.
- Gupta, R. D., Gupta, R. C. and Sankaran, P. G. (2004). Some Characterization Results Based on Factorization of the (Reversed) Hazard Rate Function. *Communication in Statistics – Theory and Methods*, 33, 3009-3031.
- Hartley, R. V. L. (1928). Transmission of Information. *Bell System Technical Journal*, 7 (3), 535-563.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions* (2<sup>nd</sup> edition, Vol. 1). John Wiley & Sons, New York.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions* (2<sup>nd</sup> edition, Vol. 2). John Wiley & Sons, New York.
- Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test*, 13, 1-43.
- Jordan, C. W. (1967). *Life Contingencies*. Society of Actuaries, Chicago.

- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.
- Keilson, J. and Sumita, U. (1982). Uniform stochastic ordering and related inequalities. *Canadian Journal of Statistics*, 10 (3), 181-189.
- Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, New York.
- Kong, L., Lee, C. and Sepanski, J. H. (2007). On the Properties of Beta-Gamma Distribution. *Journal of Modern Applied Statistical Methods*, 6 (1), 187-211.
- Kumar, U., Klefsjo, B. and Granholm, S. (1989). Reliability investigation for a fleet of load haul dump machines in Swedish mine. *Reliability Engineering and System Safety*, 26 (4), 341-361.
- Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*, 51, 4497-4509.
- Lai, C. D., Xie, M. and Murthy, D. N. P. (2001). Bathtub Shaped Failure Rate Distributions. *Handbook in Reliability*, N. Balakrishnan and C.R. Rao, eds., 20, 69-104.
- Lawless, J. F. (2003). *Statistical Models and Methods for Life Time Data*, 2<sup>nd</sup> Edition. John Wiley & Sons, New York.
- Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis*, 2<sup>nd</sup> Edition. John Wiley & Sons, New York.
- Lee, C., Famoye, F. and Olumolade, O. (2007). Beta-Weibull Distribution: Some Properties and Applications to Censored Data. *Journal of Modern Applied Statistical Methods*, 6, 173-186.
- Leemis, L. M. (1995). *Reliability, Probabilistic Models and Statistical Methods*. Prentice-Hall, New Jersey.

- Lemos de Morais, A., Cordeiro, G. M. and Cysneiros, A. H. M. A. (2011). The Beta Generalized Logistic Distribution. *Brazilian Journal of Probability and Statistics*,
- Mahmoudi, E. (2011). The beta generalized Pareto distribution with application to lifetime data. *Mathematics and Computers in Simulation*, 81 (11), 2414-2430.
- Marshall, A. W. and Olkin, I., (1997). A new method for adding a parameter to a family of distributions with application to the Exponential and Weibull families. *Biometrika*, 84, 641-652.
- Marshall, A. W. and Olkin, I. (2007). *Life Distributions*. Springer, New York.
- Murthy, R. D. N., Xie, M. and Jiang R. (2004). *Weibull Models*. John Wiley & Sons, New York.
- Nadarajah, S. and Gupta, A. K. (2004). The beta Fréchet distribution. *Far East Journal of Theoretical Statistics*, 14, 15-24.
- Nadarajah, S. and Kotz, S. (2004). The beta Gumbel distribution. *Mathematical Problems in Engineering*, 10, 323-332.
- Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution. *Reliability Engineering and System Safety*, 91 (6), 689-697.
- Nelson, W. (1982). *Applied Life Data Analysis*. John Wiley & Sons, New York.
- Pappas, V., Adamidis, K. and Loukas, S. (2010). A probability distribution with applications in Survival Analysis. *23<sup>rd</sup> Hellenic Statistical Institute Conference*, Veroia, Greece.
- Pappas, V., Adamidis, K. and Loukas, S. (2011). A three-parameter lifetime distribution. *Advances and Applications in Statistics*, 39, 35-42.



- Pappas, V., Adamidis, K. and Loukas, S. (2012). A family of lifetime distributions. *International Journal of Quality, Statistics and Reliability*, doi:10.1155/2012/760687.
- Pappas, V., Adamidis, K. and Loukas, S. (2012). A generalization of the exponential-logarithmic distribution, submitted.
- Pescim, R. R., Demetrio, C. G. B., Cordeiro, G. M., Ortega, E. M. M. and Urbano, M. R. (2010). The beta generalized half-normal distribution. *Computational Statistics and Data Analysis*, 54, 945-957.
- Pham, H. and Lai, C. D. (2007). On Recent Generalizations of the Weibull Distribution. *IEEE Transactions on Reliability*, 56 (3), 454-458.
- Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics*, 5, 375-383.
- Pulcini, G. (2001). Modeling the failure data of a repairable equipment with bathtub type failure intensity. *Reliability Engineering and System Safety*, 71, 209-218.
- Rajarshi, S. and Rajarshi, M. B. (1988). Bathtub Distributions: A Review. *Communications in Statistics – Theory and Methods*, 17 (8), 2597-2621.
- Razzaghi, M. (2009). Beta-normal distribution in dose-response modeling and risk assessment for quantitative responses. *Environmental and Ecological Science*, 16, 21-36.
- Rényi, A. (1961). On measure of entropy and information. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 547–561.
- Ruiz, J. M. and Navarro, J. (1994). Characterization of distributions by relationships between failure rate and mean residual life. *IEEE Transactions on Reliability*, 43, 640-644.

- Self, S. G. and Liang, K.-L. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605-610.
- Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic orders and their applications*. New York: Academic Press.
- Shannon C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-432.
- Singla, N., Jain, K. and Sharma, S. K. (2012). The Beta Generalized Weibull distribution: Properties and applications. *Reliability Engineering and System Safety*, 102, 5-15.
- Stacy, E. W. (1962). A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*, 33 (3), 1187-1192.
- Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review*, 48, 337-344.
- Tahmasbi, R. and Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, 52, 3889-3901.
- Tang, Y., Xie, M. and Goh, T. N. (2003). Statistical analysis of a Weibull extension model. *Communications in Statistics - Theory and Methods*, 32 (5), 913-928.
- Wang, F. K. (2000). A new model with bathtub-shaped failure rate using an additive Burr XII distribution. *Reliability Engineering and System Safety*, 70, 305-312.
- Wong, K. L. (1988). The bathtub does not hold water any more. *Quality and Reliability Engineering International*, 4 (3), 279-282.

- Wong, K. L. (1989). The roller-coaster curve is in. *Quality and Reliability Engineering International*, 5 (1), 29-36.
- Wong, K. L. (1991). The physical basis for the roller-coaster hazard rate curve for electronics. *Quality and Reliability Engineering International*, 7 (6), 489-495.
- Xie, M., Tang, Y. and Goh, T. N. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and System Safety*, 76, 279-285.
- Zimmer, W., Keats, J. B. and Wang, F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology*, 30 (4), 386-394.
- Zografos, K. (2008). On some beta generated distributions and their maximum entropy characterization: The beta-Weibull distribution, in N.S. Barnett, S.S. Dragomir (Eds.). *Advances in Inequalities from Probability Theory and Statistics*. Nova Science Publishers, New Jersey, 237-260.
- Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, 6, 344-362.