



ΑΙΚΑΤΕΡΙΝΗ ΒΑΣΙΛΕΙΑΔΟΥ

**ΜΙΞΕΙΣ ΕΚΘΕΤΙΚΗΣ ΚΑΤΑΝΟΜΗΣ
ΜΕ ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ
ΣΕ ΜΟΝΤΕΛΑ ΕΠΙΒΙΩΣΗΣ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ, 2013

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη

ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ

που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 26/03/2013 από την εξεταστική επιτροπή:

ΟΝΟΜΑΤΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ	ΥΠΟΓΡΑΦΗ
Σωτήριος Λουκάς (Επιβλέπων)	Καθηγητής
Κωνσταντίνος Ζωγράφος	Καθηγητής
Απόστολος Μπατσίδης	Λέκτορας

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.”

..... (Υπογραφή υποψηφίου)

Αικατερίνη Βασιλειάδου

*Στους γονείς μου Γιάννη και Μαρία
και τον αδερφό μου Αλέξη.*

Ευχαριστίες

Ιωάννινα, Μάϊος 2013

Η διατριβή αυτή, στα πλαίσια του Μεταπτυχιακού Προγράμματος Ειδίκευσης, εκπονήθηκε στο Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων στον τομέα “Στατιστικής και Επιχειρησιακής Έρευνας”. Θα ήθελα να ευχαριστήσω όλους αυτούς που συνέβαλαν στην πραγματοποίηση της παρούσας διατριβής και ιδιαίτερα τους Καθηγητές, που ήταν πρόθυμοι για την επίβλεψη της εργασίας μου και για την επιλογή του θέματος της διατριβής μου, που με έκανε ιδιαίτερα χαρούμενη.

Θέλω να ευχαριστήσω τον κ. Σωτήριο Λουκά, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για την επίβλεψη της Μεταπτυχιακής μου Διατριβής και για την βοήθεια και συμπαράστασή του.

Τον κ. Κωνσταντίνο Ζωγράφο, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων για τις πολύτιμες συμβουλές του κατά τη διάρκεια ολοκλήρωσης της Μεταπτυχιακής μου Διατριβής.

Τον κ. Απόστολο Μπατσίδη, Λέκτορα του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για τη συμμετοχή του στην εξεταστική επιτροπή, όπως και για τις σημαντικές παρατηρήσεις και επισημάνσεις του.

Επίσης, ευχαριστώ τον κ. Βασίλειο Παππά, Υποψήφιο Διδάκτορα του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για τις πολύτιμες συμβουλές και παρεμβάσεις του στη διαμόρφωση της διατριβής και στη συγγραφή της.

Ακόμη, θέλω να ευχαριστήσω την κ. Ευαγγελία Γεωργάκη για την σημαντική στήριξη σε όλη τη διάρκεια των Μεταπτυχιακών μου Σπουδών.

Περισσότερο από όλους, ευχαριστώ τους γονείς μου Γιάννη και Μαρία και τον αδερφό μου Αλέξη που με στήριξαν υλικά και πνευματικά για αυτή μου την προσπάθεια. Τέλος, θα ήθελα να ευχαριστήσω τον ξάδερφο μου Βαλάντη και τους φίλους μου Δημήτρη, Στέλλα, Μαρία, Σωτηρία, Ευαγγελία, Νίκο και Αντρέα για την ψυχολογική στήριξή τους.

Αικατερίνη Βασιλειάδου

ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ.
ΠΡΟΛΟΓΟΣ	1
1. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ	5
1.1 Ιστορική αναδρομή.....	6
1.2 Εισαγωγή	8
1.2.1 Συνάρτηση πυκνότητας πιθανότητας.....	8
1.2.2 Συνάρτηση επιβίωσης	9
1.2.3 Συνάρτηση κινδύνου	10
1.2.4 Σχέσεις μεταξύ των συναρτήσεων επιβίωσης.....	14
1.2.5 Συνάρτηση μέσου υπολειπόμενου χρόνου ζωής.....	15
1.3 Οι συναρτήσεις συνοπτικά σε πίνακα	16
1.4 Η σημασία της συνάρτησης κινδύνου	17
1.5 Βασικά μοντέλα κατανομών στην ανάλυση επιβίωσης.....	18
1.6 Συνάρτηση πιθανοφάνειας.....	29
1.7 EM αλγόριθμος.....	30
2. ΚΑΤΑΝΟΜΕΣ ΦΘΙΝΟΥΣΑΣ ΣΥΝΑΡΤΗΣΗΣ ΚΙΝΔΥΝΟΥ	37
2.1 Η Εκθετική-Γεωμετρική κατανομή	38
2.1.1 Ορισμοί και φυσική ερμηνεία	38
2.1.2 Ιδιότητες της κατανομής	40
2.1.3 Χαρακτηριστικά της κατανομής	44
2.1.4 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας	51
2.1.5 Η μέθοδος EM.....	54

2.1.6	Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας	55
2.2	Η Εκθετική-Poisson κατανομή	57
2.2.1	Ορισμοί και φυσική ερμηνεία	57
2.2.2	Χαρακτηριστικά της κατανομής	59
2.2.3	Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας.....	63
2.2.4	Η μέθοδος EM.....	66
2.2.5	Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας.....	67
2.3	Η Εκθετική-Λογαριθμική κατανομή.....	68
2.3.1	Ορισμοί και φυσική ερμηνεία	68
2.3.2	Χαρακτηριστικά της κατανομής	69
2.3.3	Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας.....	74
2.3.4	Η μέθοδος EM.....	75
2.3.5	Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας	77
2.4	Η Εκθετική-Δυναμοσειρά κατανομή	78
2.4.1	Φυσική ερμηνεία	78
2.4.2	Χαρακτηριστικά της κατανομής	80
2.4.3	Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας.....	82
2.4.4	Η μέθοδος EM.....	88
2.4.5	Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας	89
3.	ΕΦΑΡΜΟΓΕΣ	91
4.	ΑΛΛΕΣ ΜΟΡΦΕΣ ΣΥΝΑΡΤΗΣΗΣ ΚΙΝΔΥΝΟΥ	99

4.1 Κατανομές με κυπελλοειδούς μορφής συνάρτησης κινδύνου, B.T.F.R. (Bathtub Shaped Failure Rate)	100
4.2 Η Γενικευμένη Εκθετική-Poisson κατανομή	102
4.2.1 Φυσική ερμηνεία της κατανομής	102
4.2.2 Ορισμοί και χαρακτηριστικά της κατανομής	102
4.2.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας ...	103
4.3 Η Γενικευμένη Εκθετική-Γεωμετρική κατανομή	105
4.3.1 Φυσική ερμηνεία της κατανομής	105
4.3.2 Ορισμοί και χαρακτηριστικά της κατανομής	106
4.3.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας ...	107
4.4 Η Γενικευμένη Εκθετική-Λογαριθμική κατανομή	108
4.4.1 Φυσική ερμηνεία της κατανομής	108
4.4.2 Ορισμοί και χαρακτηριστικά της κατανομής	109
4.4.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας ...	111
4.4.4 Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας	113
ΠΑΡΑΡΤΗΜΑ	115
ΠΕΡΙΛΗΨΗ	123
ABSTRACT OF THE DISSERTATION	125
BIBΛΙΟΓΡΑΦΙΑ	127

ΠΡΟΛΟΓΟΣ

Η μελέτη του μήκους ζωής των βιολογικών οργανισμών, αλλά και των συσκευών (για παράδειγμα, ηλεκτρικές συσκευές, κ.τ.λ.) είναι μέγιστης σημασίας στις βιολογικές και μηχανικές επιστήμες και αποτελούν το επίκεντρο της αξιοπιστίας και της ανάλυσης επιβίωσης.

Με τον όρο *ανάλυση επιβίωσης* εννοούμε το σύνολο των στατιστικών μεθοδολογιών που έχουν ως στόχο την περιγραφή και τη μοντελοποίηση της διάρκειας ενός φαινομένου ή συμβάντος στο χρόνο. Οι μεθοδολογίες της ανάλυσης επιβίωσης αποτελούν χρήσιμο εργαλείο για ερευνητές από διάφορους χώρους, όπως της οικονομίας, της ιατρικής και των άλλων βιολογικών επιστημών.

Ο κλάδος της οικονομίας είναι μια περιοχή έρευνας, όπου το εισόδημα, το μέγεθος της εταιρίας, οι τιμές και οι αναλογιστικές ζημιές είναι από τη φύση τους μη αρνητικές. Μη αρνητικές τυχαίες μεταβλητές προκύπτουν ως μεγέθη που σχετίζονται με φυσικά φαινόμενα. Τέτοια παραδείγματα είναι: η διάμετρος των δέντρων και το ύψος, η ταχύτητα των ανέμων, η δύναμη των υλικών, η ροή του ρεύματος, η βροχόπτωση, η φθορά των ελαστικών και η χημική σύνθεση κ.α.

Η μαθηματική περιγραφή της διάρκειας ζωής (εξαρτημάτων, συσκευών, κ.τ.λ.) από τις κατανομές με φθίνουσα συνάρτηση κινδύνου (DFR) αποτελεί ένα σημαντικό μέρος της μελέτης. Όπως επίσης, η έρευνα μίξης κατανομών αποτελεί ένα σημαντικό εργαλείο για την περιγραφή της κατανομής του χρόνου λειτουργίας εξαρτημάτων ή συσκευών κ.τ.λ. Ο Proschan (1963) χρησιμοποιώντας τη μίξη εκθετικών κατανομών, θέλησε να εξηγήσει το χρόνο μεταξύ διαδοχικών αποτυχιών του συστήματος κλιματισμού σε αεροπλάνα τύπου Boeing 720. Επίσης, ο Proschan (1963), είχε αποδείξει ότι η ιδιότητα της φθίνουσας συνάρτησης κινδύνου διατηρείται και σε μίξεις κατανομών με σταθερές συναρτήσεις κινδύνου.

Το αντικείμενο της διατριβής είναι η μελέτη των κατανομών που προκύπτουν από τη Μίξη Εκθετικής με Διακριτές Κατανομές (Διωνυμική, Γεωμετρική, Poisson, Λογαριθμική) σε μοντέλα επιβίωσης.

Στο Κεφάλαιο 1, γίνεται αναφορά κάποιων βασικών εννοιών πάνω στα μοντέλα επιβίωσης, που αποτελούν τη βάση για την υποστήριξη της διατριβής. Στη συνέχεια, παρουσιάζεται η ιστορική αναδρομή της ανάλυσης επιβίωσης. Αναλύονται, χάριν πληρότητας, κάποιες βασικές συναρτήσεις, όπως είναι: η συνάρτηση πυκνότητας πιθανότητας, η συνάρτηση επιβίωσης, η συνάρτηση κινδύνου και η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής. Επίσης, εξηγούμε τη σημασία της συνάρτησης κινδύνου, αποδεικνύουμε τη σχέση της με τη συνάρτηση επιβίωσης και αναφέρουμε τα σημαντικότερα μοντέλα χρόνων ζωής. Τέλος, γίνεται μια σύντομη αναφορά στη συνάρτηση πιθανοφάνειας και αναλύεται ο EM αλγόριθμος.

Στο Κεφάλαιο 2, εξετάζονται οι τέσσερις σημαντικότερες κατανομές με φθίνουσα συνάρτηση κινδύνου (DFR). Εξετάζονται κατά σειρά η Εκθετική-Γεωμετρική κατανομή (EG), η Εκθετική-Poisson κατανομή (EP), η Εκθετική-Λογαριθμική κατανομή (EL) και η Εκθετική-Δυναμοσειρά κατανομή (EPS). Οι κατανομές αυτές εισήχθησαν και μελετήθηκαν από τους Adamidis and Loukas (1998), Kus (2007), Tahmasbi and Rezaei (2008) και Chahkandi and Ganjali (2009), αντίστοιχα. Αρχικά, θα δοθεί η φυσική ερμηνεία αυτών των κατανομών. Όσον αφορά τα χαρακτηριστικά των κατανομών γίνεται μελέτη της αθροιστικής συνάρτησης κατανομής, της διαμέσου, των ροπών, της μέσης τιμής, της διακύμανσης, της συνάρτησης επιβίωσης, της συνάρτησης κινδύνου και της συνάρτησης του μέσου υπολειπόμενου χρόνου ζωής. Στη συνέχεια, το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας (EMΠ). Προσδιορίζονται οι πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας (Fisher information matrix) και παρουσιάζεται ένας EM αλγόριθμος.

Στο Κεφάλαιο 3, γίνεται μελέτη προσαρμογής των προαναφερθέντων κατανομών χρησιμοποιώντας τρία σύνολα δεδομένων από πραγματικά δεδομένα. Τέλος, παρουσιάζονται οι γραφικές παραστάσεις της συνάρτησης επιβίωσης κάθε κατανομής για κάθε σύνολο δεδομένων ξεχωριστά.

Στο Κεφάλαιο 4, δίνονται κάποιοι ορισμοί της κυπελλοειδούς ιδιότητας (bathtub shaped failure rate ή B.T.F.R.). Στη συνέχεια, εξετάζονται κατά σειρά η Γενικευμένη Εκθετική-Poisson κατανομή, η Γενικευμένη Εκθετική-Γεωμετρική κατανομή και η Γενικευμένη Εκθετική-Λογαριθμική κατανομή. Δίνεται η φυσική ερμηνεία αυτών των κατανομών και μελετώνται τα χαρακτηριστικά τους. Τέλος, το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας (EMΠ) και τους πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας.

Στο Παράρτημα, δίνονται τα τρία σύνολα δεδομένων που χρησιμοποιήθηκαν στο Κεφάλαιο 3, οι συναρτήσεις στην R γλώσσα προγραμματισμού για το πρώτο σύνολο δεδομένων, από τις οποίες προέκυψαν ο πίνακας και οι καμπύλες επιβίωσης των κατανομών EL, EG, EB και EP του Κεφαλαίου 3 και οι συναρτήσεις στη Mathematica, από τις οποίες προέκυψαν τα γραφήματα του Κεφαλαίου 2. Η διατριβή ολοκληρώνεται με την περίληψη του περιεχομένου της στην Αγγλική και Ελληνική γλώσσα και την βιβλιογραφία.

1. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Η *ανάλυση επιβίωσης (survival analysis)* είναι μια περιοχή έρευνας στη στατιστική, η οποία δημιουργήθηκε για την ανάλυση δεδομένων, τα οποία δε μπορούν να υποστούν επεξεργασία από τις συνηθισμένες στατιστικές μεθόδους. Τα δεδομένα αυτά δίνουν τη χρονική διάρκεια μέχρι να γίνει ένα συγκεκριμένο γεγονός. Η ανάλυση επιβίωσης μπορεί να εφαρμοστεί σε αρκετές περιπτώσεις, όπως για παράδειγμα, στη μηχανολογία για την ανάλυση του χρόνου μέχρι την καταστροφή ενός μηχανήματος ή στη γεωργία για την ανάλυση του χρόνου μέχρι τη στιγμή να βγάλει καρπό ένα δέντρο. Στην περίπτωση της μηχανολογίας, η ανάλυση αναφέρεται και ως θεωρία αξιοπιστίας (reliability theory).

Τα *δεδομένα επιβίωσης (survival data)* περιλαμβάνουν το χρόνο επιβίωσης, την αντίδραση σε δεδομένη θεραπεία, τα χαρακτηριστικά του ασθενούς που σχετίζονται με την απόκριση στη θεραπεία, την επιβίωση και την εξέλιξη μιας πάθησης ή ενός συμπτώματος.

Με τον όρο *χρόνος επιβίωσης ή χρόνος αποτυχίας (survival time ή failure time)* εννοούμε το χρονικό διάστημα από τη στιγμή της παρατήρησης κάποιου ενδεχόμενου μέχρι το θάνατο ή την αποτυχία (όπως αντίδραση σε θεραπεία, θάνατος, εμφάνιση πάθησης). Ο χρόνος επιβίωσης χρίζει ειδικής μεταχείρισης για το λόγο ότι είναι περιορισμένος στο να είναι θετικός. Ο χρόνος επιβίωσης ονομάζεται και χρόνος ως το “γεγονός” ή την “αποτυχία”.

Επίσης, ο χρόνος επιβίωσης αποτελεί ένα από τα βασικότερα σημεία ενδιαφέροντος. Ενδεικτικά,

- Ο χρόνος μέχρι να αντιδράσει ο οργανισμός του ασθενούς σε κάποιο φάρμακο, εφαρμόζεται σε βιοχημικές εφαρμογές.
- Ο χρόνος μέχρι να χαλάσει μια συσκευή, εφαρμόζεται στη μηχανική.
- Ο χρόνος μέχρι ένας δείκτης να φτάσει το απαιτούμενο όριο, εφαρμόζεται στις οικονομικές επιστήμες.

Παράδειγμα

Έστω ότι θέλουμε να αξιολογήσουμε ένα πειραματικό φάρμακο σε σχέση με το χρόνο επιβίωσης πειραματόζωου σε μια εργαστηριακή μελέτη. Για το σκοπό αυτό εφαρμόζεται το νέο φάρμακο σε έναν αριθμό μονάδων. Έστω T μια μη-αρνητική τυχαία μεταβλητή που χαρακτηρίζει το χρόνο επιβίωσης από την αρχή της δόσης μέχρι το θάνατο για κάθε πειραματική μονάδα. Η τυχαία μεταβλητή T μπορεί να είναι συνεχής ή διακριτή.

Παρατηρήσεις

- Πολλές από τις συνηθισμένες στατιστικές μεθόδους προϋποθέτουν ότι η κατανομή της μεταβλητής που εξετάζουμε είναι κανονική. Όμως, οι χρόνοι επιβίωσης έχουν μια κατανομή που διαφέρει πολύ από την κανονική. Αυτό συμβαίνει, γιατί οι χρόνοι επιβίωσης ορισμένων παρατηρήσεων είναι λογοκριμένοι, δηλαδή κάποια άτομα μπορεί να εισέρχονται στη μελέτη σε διαφορετικούς χρόνους, έχοντας ως συνέπεια ο χρόνος παρακολούθησης των ατόμων αυτών να μην είναι επαρκής, ώστε να καταγραφεί ο χρόνος μέχρι την πραγματοποίηση του υπό μελέτη γεγονότος.
- Οι μέθοδοι για την ανάλυση επιβίωσης προϋποθέτουν ότι οι παρατηρήσεις είναι ανεξάρτητες.

1.1 Ιστορική αναδρομή

Για την ανάλυση επιβίωσης αυτό που γνωρίζουμε μέχρι σήμερα οφείλεται στους διάφορους τομείς έρευνας που η ανάλυση επιβίωσης έχει συμβάλει, όπως είναι η Ιατρική, η Φαρμακευτική, η Βιολογία. Στόχος τους ήταν και είναι η εύρεση λύσης για την αντιμετώπιση των διάφορων προβλημάτων τους.

Το 1662 στη Μ. Βρετανία δημοσιεύθηκε το πρώτο βιβλίο, το οποίο είχε καταγεγραμμένους καταλόγους με γεννήσεις και θανάτους που αναφέρονταν στις προηγούμενες δεκαετίες. Ήταν η πρώτη φορά που οι θάνατοι αντιμετωπίστηκαν ως γεγονότα, για τα οποία έγιναν αναλυτικές μελέτες (Δημάκη 2006, σελ 2).

Οι παγκόσμιοι πόλεμοι που ακολούθησαν, έδωσαν το έναυσμα για ανάπτυξη της έρευνας στον τομέα της αξιοπιστίας και στην μελέτη της διάρκειας ζωής των στρατιωτικών στρατευμάτων, αλλά και αργότερα η έρευνα επικεντρώθηκε στην μελέτη κάποιων ιδιαίτερων πιθανοθεωρητικών προβλημάτων σχετιζομένων με την πάυση λειτουργίας και την αντικατάσταση εξαρτημάτων μηχανικών ή ηλεκτρικών κυκλωμάτων, όπως κάποιας βαλβίδας ή ενός θερμοστάτη σε ένα μηχανικό κύκλωμα, μιας λυχνίας ή μιας αντίστασης σε ένα ηλεκτρικό. Υπήρξε δηλαδή μεγάλη πρόοδος στη Βιομηχανία των ηλεκτρικών συσκευών.

Έπειτα, οι μέθοδοι της ανάλυσης επιβίωσης είχαν τεράστιες εφαρμογές σε κλινικά δεδομένα και σκοπός ήταν να απαντηθούν ερωτήματα, όπως ποιά είναι η πιθανότητα ένας ασθενής να ζήσει μέχρι μια συγκεκριμένη χρονική περίοδο ή με ποιό ρυθμό θα πεθάνουν κάποιοι ασθενείς, οι οποίοι έχουν ήδη επιβιώσει μέχρι ένα συγκεκριμένο χρονικό σημείο, κ.α. (Δημάκη, 2006).

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 εξέλιξη των ηλεκτρονικών υπολογιστών, η εφαρμογή και η ανάπτυξή της γίνεται ολοένα και μεγαλύτερη, καθώς παράγονται ολοένα και περισσότερα λογισμικά πακέτα για το σκοπό αυτό.

1.2 Εισαγωγή

Η κατανομή του χρόνου επιβίωσης, χαρακτηρίζεται από τέσσερις συναρτήσεις, τη συνάρτηση πυκνότητας πιθανότητας (probability density function), τη συνάρτηση επιβίωσης (survivor function), τη συνάρτηση κινδύνου (hazard function) και τη συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής (mean residual lifetime).

1.2.1 Συνάρτηση πυκνότητας πιθανότητας

Στην ενότητα αυτή θα δοθεί ο ορισμός της συνάρτησης πυκνότητας πιθανότητας ή της συνάρτησης πυκνότητας του χρόνου επιβίωσης T .

Ορισμός 1.1

Η συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση πυκνότητας (probability density function ή density function) του χρόνου επιβίωσης T , ορίζεται ως το όριο της πιθανότητας ότι ένα άτομο αποτυγχάνει σε ένα μικρό διάστημα $(t, t + \Delta t)$ ανά μονάδα πλάτους Δt . Η έκφραση του είναι η εξής:

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

όπου με $S(t)$ συμβολίζεται η συνάρτηση επιβίωσης, ο ορισμός της οποίας θα δοθεί στη συνέχεια.

Για μικρό Δt , η ποσότητα $f(t)\Delta t$ εκφράζει κατά προσέγγιση την πιθανότητα “θανάτου-αποτυχίας” στο διάστημα $[t, t + \Delta t]$. Η συνάρτηση πυκνότητας πιθανότητας κατανομών χρόνων ζωής, έχει τις παρακάτω ιδιότητες:

$$f(t) \geq 0, \text{ για κάθε } t \geq 0 \text{ και } \int_0^{\infty} f(t)dt = 1.$$

Η αθροιστική συνάρτηση κατανομής της $f(t)$ (*cumulative distribution function*) είναι η:

$$F(t) = P(T \leq t), \text{ για } t \geq 0.$$

1.2.2 Συνάρτηση επιβίωσης

Στην ενότητα αυτή θα δοθεί ο ορισμός της συνάρτησης επιβίωσης και έπειτα κάποιες ιδιότητες αυτής.

Ορισμός 1.2

Συμβολίζοντας το χρόνο επιβίωσης μιας *συνεχούς, μη αρνητικής τοχαίας μεταβλητής* (τ.μ.) με T , η *συνάρτηση επιβίωσης (survival function)* $S(t)$ ορίζεται ως η πιθανότητα επιβίωσης ενός ατόμου πέραν της χρονικής στιγμής t και δίνεται από τη σχέση:

$$S(t) = P(T > t) = \int_t^{+\infty} f(x)dx = 1 - F(t), \quad t \geq 0,$$

όπου $f(t)$ και $F(t)$ η συνάρτηση πυκνότητας πιθανότητας και η αθροιστική συνάρτηση κατανομής, αντίστοιχα.

Ιδιότητες

Η συνάρτηση επιβίωσης $S(t)$ είναι μη αρνητική, φθίνουσα και συνεχής με

$$S(0) = 1 \text{ και } \lim_{t \rightarrow \infty} S(t) = 0.$$

Η γραφική παράσταση της $S(t)$ συναρτήσει του t ονομάζεται *καμπύλη επιβίωσης (survival curve)* και προτάθηκε από τον Berkson (1942).

Αντίθετα, όταν έχουμε ομαδοποιημένα δεδομένα ή οι χρόνοι ζωής αναφέρονται σε ακέραιο αριθμό κύκλων κάποιου τύπου, τότε είναι απαραίτητη η εισαγωγή *διακριτής τυχαίας μεταβλητής* για τη μέτρηση των χρόνων ζωής.

Ορισμός 1.3

Έστω T *διακριτή τυχαία μεταβλητή*, με τιμές t_1, t_2, t_3, \dots , τέτοιες ώστε $0 \leq t_1 \leq t_2 \leq \dots$ και συνάρτηση πιθανότητας:

$$p(t_j) = P(T = t_j), j = 1, 2, \dots$$

Τότε, η συνάρτηση επιβίωσης δίνεται από τη σχέση:

$$S(t) = P(T \geq t) = \sum_{j: t_j \geq t} p(t_j). \quad (1.1)$$

Όπως και στη συνεχή περίπτωση, η $S(t)$ είναι μονότονα φθίνουσα και δεξιά συνεχής συνάρτηση του χρόνου, με $S(0) = 1$ και $S(\infty) = 0$.

1.2.3 Συνάρτηση κινδύνου

Στην ενότητα αυτή θα δοθούν οι ορισμοί της συνάρτησης κινδύνου και της αθροιστικής συνάρτησης κινδύνου και κάποιες ιδιότητες αυτών.

Ορισμός 1.4

Η συνάρτηση κινδύνου (hazard function ή failure rate) $h(t)$, η οποία καθορίζει το στιγμιαίο ποσοστό θανάτου ή αποτυχίας στο χρόνο t , δεδομένου ότι η πειραματική μονάδα επιβιώνει πέραν του χρόνου t , είναι μια δεσμευμένη πιθανότητα και δίνεται από τη σχέση:

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

Προσεγγιστικά, $h(t)\Delta t$ είναι η πιθανότητα θανάτου στο διάστημα $[t, t+\Delta t)$, δεδομένου ότι το άτομο επιζεί πέραν του χρόνου t . Για **συνεχή κατανομή χρόνου ζωής**, η συνάρτηση κινδύνου έχει τις παρακάτω ιδιότητες:

- $h(t) \geq 0$, για όλα τα $t \geq 0$,
- $\int_0^{\infty} h(t)dt = \infty$.

Παρατηρήσεις

Εάν η συνάρτηση κινδύνου $h(t)$ λαμβάνει ελάχιστο όταν $t = t_0$, τότε η συνάρτηση πυκνότητας πιθανότητας $f(t)$, επίσης λαμβάνει ελάχιστο στο t_0 (από σχέση 1.2).

Η αθροιστική συνάρτηση κινδύνου (cumulative hazard function) δίνεται από τη σχέση:

$$H(t) = \int_0^t h(x)dx = -\log S(t).$$

Η τελευταία είναι μη φθίνουσα συνάρτηση του t , με

$$H(0) = 0 \text{ και } \lim_{t \rightarrow \infty} H(t) = \infty.$$

Οι συναρτήσεις $f(t)$, $S(t)$ και $h(t)$ συνδέονται μεταξύ τους με τους εξής τύπους (Lawless, 1982):

$$f(t) = h(t) \exp\left(-\int_0^t h(x)dx\right),$$

και

$$S(t) = \exp\left(-\int_0^t h(x)dx\right).$$

Αντίστοιχα, στη διακριτή περίπτωση η συνάρτηση κινδύνου δίνεται από τη σχέση:

$$h(t_j) = P(T = t_j | T \geq t_j) = \frac{p(t_j)}{S(t_j)}, j = 1, 2, \dots$$

Τότε, έχουμε:

$$h(t_j) = \frac{P(T \geq t_j) - P(T \geq t_j + 1)}{P(T \geq t_j)},$$

δηλαδή

$$h(t_j)P(T \geq t_j) = P(T \geq t_j) - P(T \geq t_j + 1),$$

και

$$P(T \geq t_j + 1) - [1 - h(t_j)]P(T \geq t_j) = 0,$$

άρα

$$1 - h(t_j) = \frac{S(t_{j+1})}{S(t_j)}. \quad (1.3)$$

Επίσης,
$$S(t_j) = P(T \geq t_j | T \geq t_{j-1})P(T \geq t_{j-1}),$$

και

$$S(t_j) = \frac{S(t_j)}{S(t_{j-1})} S(t_{j-1}).$$

Από τη σχέση (1.3):

$$S(t_j) = (1 - h(t_{j-1}))S(t_{j-1}).$$

Επομένως,

$$P(T \geq t_j) = P(T \geq 0) \prod_{t_j=0}^{t-1} (1 - h(t_j)).$$

Όμως, $P(T \geq 0) = S(0) = 1$, άρα

$$P(T \geq t_j) = \prod_{t_j=0}^{t-1} (1 - h(t_j)),$$

και το ανάλογο της αθροιστικής συνάρτησης κινδύνου *στη διακριτή περίπτωση*, ορίζεται ως εξής:

$$H(t) = -\log S(t),$$

όπου $S(t)$ δίνεται από τη σχέση (1.1).

Οι $p(t)$, $h(t)$ και $S(t)$ συνδέονται μεταξύ τους με τις εξής σχέσεις:

$$S(t) = \prod_{j:t_j < t} [1 - h(t_j)],$$

και

$$p(t_j) = S(t_j) - S(t_{j+1}).$$

Το σημαντικότερο ίσως αποτέλεσμα που συνοδεύει τον ορισμό της συνάρτησης κινδύνου είναι το θεώρημα του μονοσήμαντου της συνάρτησης κινδύνου, το οποίο διατυπώνεται στο θεώρημα που ακολουθεί. Δίνεται στη συνέχεια χωρίς απόδειξη.

Θεώρημα 1.1

Έστω X τυχαία μεταβλητή με μη αρνητικές τιμές και συνάρτηση επιβίωσης $S(t)$. Η κατανομή της X ορίζεται μονοσήμαντα από τη συνάρτηση κινδύνου $h(t)$.

1.2.4 Σχέσεις μεταξύ των συναρτήσεων επιβίωσης

Γνωρίζουμε ότι, η συνάρτηση επιβίωσης συνδέεται με τη συνάρτηση κινδύνου με την ακόλουθη σχέση (βλέπε ορισμό 1.2):

$$h(t) = \frac{f(t)}{S(t)}. \quad (1.4)$$

Η συνάρτηση πυκνότητας πιθανότητας οποιασδήποτε κατανομής είναι ίση με την παράγωγο της συνάρτησης κατανομής, επομένως:

$$f(t) = \frac{d(F(t))}{dt} = \frac{d[1-S(t)]}{dt} = -S'(t). \quad (1.5)$$

Από τις (1.3) και (1.4) προκύπτει:

$$h(t) = \frac{-S'(t)}{S(t)} = -\frac{d}{dt} \log S(t). \quad (1.6)$$

Στη συνέχεια, ολοκληρώνοντας την (1.5) και λαμβάνοντας υπόψη ότι $S(0) = 1$, έχουμε:

$$\begin{aligned} \int_0^t h(x) dx &= \int_0^t -\frac{d}{dx} \log S(x) \\ &= -\log S(t). \end{aligned}$$

Επομένως, $S(t) = \exp[-H(t)]$ και καταλήγουμε στο συμπέρασμα ότι $f(t) = h(t) \exp[-H(t)]$.

1.2.5 Συνάρτηση μέσου υπολειπόμενου χρόνου ζωής

Στην ενότητα αυτή θα δοθεί ο ορισμός της συνάρτησης του μέσου υπολειπόμενου χρόνου ζωής και κάποιες ιδιότητες αυτής.

Ορισμός 1.5

Η αναμενόμενη διάρκεια ζωής από τη στιγμή t μέχρι την “αποτυχία”, δοθείσης της επιβίωσης μέχρι το χρόνο t είναι μια συνάρτηση του t και ονομάζεται **μέσος υπολειπόμενος χρόνος ζωής**.

Πιο συγκεκριμένα, αν η συνεχής τυχαία μεταβλητή T αντιπροσωπεύει τη διάρκεια ζωής, τότε ο μέσος υπολειπόμενος χρόνος ζωής δίνεται από τη σχέση:

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

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

$$m(t) \geq 0, \quad m'(t) \geq -1, \quad \text{για κάθε } t \geq 0 \text{ και } \int_0^{\infty} \frac{1}{m(t)} dt = \infty.$$

Παρατηρήσεις

- Ο ορισμός της αναμενόμενης διάρκειας ζωής εφαρμόζεται στις περιπτώσεις δοκιμών ζωής (για παράδειγμα, η διάρκεια ζωής βιολογικών οργανισμών ή εργοστασιακών προϊόντων).
- Είναι δείκτης γήρανσης ενός ατόμου ή ενός εξαρτήματος.

Ορισμός 1.6

Έστω T μια διακριτή τυχαία μεταβλητή, τότε η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής είναι:

$$m(t) = \frac{\sum_{s=t}^{\infty} S(s)}{S(t)}, \text{ για } t = 0, 1, 2, \dots$$

Οι ροπές δεν προσδιορίζουν μονοσήμαντα την κατανομή του χρόνου ζωής, αλλά συχνά αποτελούν ένα χρήσιμο τρόπο συνοπτικής περιγραφής της. **Οι ροπές r τάξης περί το μηδέν**, μιας συνεχούς τ.μ. T που περιγράφει τη διάρκεια ζωής, προσδιορίζονται από το ολοκλήρωμα

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

Η μέση τιμή $E(T)$, αποτελεί ένα μέτρο της κεντρικής τάσης των τιμών της τ.μ. T , ενώ η διακύμανση $Var(T) = E(T^2) - \{E(T)\}^2$ ένα μέτρο της διασποράς των τιμών της, γύρω από τη μέση τιμή.

1.3 Οι συναρτήσεις συνοπτικά σε πίνακα

Έστω T μια συνεχής τ.μ. που περιγράφει τη διάρκεια ζωής, με σ.π.π. $f(\cdot)$ και α.σ.κ. $F(\cdot)$. Στον πίνακα που ακολουθεί δίνεται στην 1^η στήλη η συναρτησιακή σχέση ορισμού διάφορων ποσοτήτων που απαντώνται στην υπόλοιπη διατριβή, ενώ στη 2^η στήλη ο τρόπος προσδιορισμού της συνάρτησης επιβίωσης, αν είναι γνωστές οι συναρτήσεις της 1^{ης} στήλης.

Συναρτήσεις	Συνάρτηση Επιβίωσης
<p>Συνάρτηση πυκνότητας πιθανότητας</p> $f(x) = F'(x)$	$S(x) = \int_x^{\infty} f(z) dz$
<p>Συνάρτηση κινδύνου</p> $H(x) = -\log S(x)$	$S(x) = \exp\{-H(x)\}$
<p>Ποσοστό κινδύνου</p> $h(x) = H'(x) = \frac{f(x)}{S(x)}$	$S(x) = \exp\left\{-\int_0^x h(z) dz\right\}$
<p>Αντίστροφη συνάρτηση κινδύνου</p> $T(x) = \log F(x)$	$S(x) = 1 - \exp\{T(x)\}$
<p>Αντίστροφο ποσοστό κινδύνου</p> $t(x) = T'(x) = \frac{f(x)}{F(x)}$	$S(x) = 1 - \exp\left\{\int_0^x t(z) dz\right\}$
<p>Κατανομή υπολειπόμενου χρόνου ζωής</p> $S_t(x) = \frac{S(x+t)}{S(t)}$	$S(x) = S_0(x)$
<p>Κατανομή μέσου υπολειπόμενου χρόνου ζωής</p> $m(t) = \int_0^{\infty} \frac{S(x+t)}{S(t)} dx$	$S(t) = \frac{\mu}{m(t)} \exp\left\{-\int_0^t \frac{dz}{m(t)}\right\}$

1.4 Η σημασία της συνάρτησης κινδύνου

Η συνάρτηση κινδύνου (*hazard function*) είναι ιδιαίτερα χρήσιμη στη θεωρία αξιοπιστίας και ανάλυσης επιβίωσης, γιατί περιγράφει τον τρόπο με τον οποίο μεταβάλλεται ο στιγμιαίος ρυθμός “θανάτου-αποτυχίας” στο χρόνο και έτσι βοηθά στην επιλογή του κατάλληλου

μοντέλου για την περιγραφή χρόνων ζωής. Συνήθως, η επιβίωση περιγράφεται με κατανομές των οποίων η αντίστοιχη συνάρτηση κινδύνου είναι αύξουσα, φθίνουσα, κυπελλοειδούς μορφής ή μονοκόρυφη. Η περίπτωση της εκθετικής κατανομής με σταθερή συνάρτηση κινδύνου, παρουσιάζει μεγάλο εύρος εφαρμογών στη μοντελοποίηση χρόνων ζωής. Ωστόσο, το μεγαλύτερο ενδιαφέρον επικεντρώνεται σε μια περίοδο συνολικής διάρκειας ζωής στην οποία επέρχεται γήρανση ή φθορά χρήσης. Αυτός είναι και ο λόγος που εξηγεί τη μεγάλη δημοτικότητα των κατανομών χρόνου ζωής με αύξουσα συνάρτηση κινδύνου (increasing rate). Τα μοντέλα με φθίνουσα συνάρτηση κινδύνου είναι αυτά που χρησιμοποιούνται για την περιγραφή της διάρκειας ζωής ηλεκτρονικών ή μηχανικών εξαρτημάτων, τα οποία παρουσιάζουν βελτίωση στη λειτουργία τους με την πάροδο του χρόνου. Τα μοντέλα με μονοκόρυφη συνάρτηση κινδύνου χρησιμοποιούνται κυρίως στην περίπτωση που παρατηρείται μεγάλη συχνότητα “αποτυχιών” κοντά στη χρονική αφετηρία του πειράματος που παράγει τα δεδομένα. Ακόμη, τα μοντέλα με μη μονότονη συνάρτηση κινδύνου είναι τα λιγότερο χρησιμοποιούμενα στην πράξη, με εξαίρεση αυτά με συνάρτηση κινδύνου κυπελλοειδούς μορφής (bathtub models), τα οποία χρησιμοποιούνται για να περιγράψουν φαινόμενα που παρακολουθούνται από τη γέννηση μέχρι την κατάληξή τους. Οι ιδιότητες αυτών των μοντέλων θα αποτελέσουν αντικείμενο μελέτης του τέταρτου κεφαλαίου.

1.5 Βασικά μοντέλα κατανομών στην ανάλυση επιβίωσης

Οι παρακάτω κατανομές αναφέρονται σε βιβλία θεωρίας κατανομών (π.χ. Johnson et al., 1994) και σε εγχειρίδια ανάλυσης επιβίωσης (π.χ. Leemis, 1995).

Η Εκθετική Κατανομή

Η αξία της εκθετικής κατανομής στην ανάλυση επιβίωσης είναι μεγάλη λόγω της απλής συναρτησιακής της μορφής και της χρησιμότητάς της στην ερμηνεία χρόνου ζωής. Αυτό έχει ως συνέπεια τις πολλές εφαρμογές της σε μελέτες διάρκειας ζωής μηχανών, ειδικότερα εξαρτημάτων και σε έρευνες χρόνων επιβίωσης σε περιπτώσεις χρόνιων ασθενειών. Το βασικό χαρακτηριστικό της εκθετικής κατανομής είναι η σταθερή συνάρτηση κινδύνου που οφείλεται στην ιδιότητα της αμνησίας που χαρακτηρίζει τη συγκεκριμένη κατανομή. Η εκθετική κατανομή έχει συνάρτηση πυκνότητας πιθανότητας και συνάρτηση επιβίωσης, αντίστοιχα

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

και

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

Συμβολίζουμε ότι $T \sim \text{Exp}(\lambda)$. Προκύπτει ότι, έχει σταθερή συνάρτηση κινδύνου

$$h(t; \lambda) = \lambda > 0 \quad (\lambda \text{ σταθερά}),$$

όπως και σταθερό μέσο υπολειπόμενο χρόνο ζωής που έχει τη μορφή

$$m(t; \lambda) = \frac{1}{\lambda}, \quad t \geq 0.$$

Οι ροπές r τάξης περί το μηδέν της εκθετικής κατανομής, δίνονται από τη σχέση

$$E(T^r; \lambda) = \frac{\Gamma(r+1)}{\lambda^r},$$

όπου

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt,$$

είναι η συνάρτηση Γάμμα.

Η μέση τιμή και η διακύμανση της, είναι αντίστοιχα

$$E(T; \lambda) = \frac{1}{\lambda} \text{ και } Var(T; \lambda) = \frac{1}{\lambda^2}.$$

Η Κατανομή Weibull

Η κατανομή Weibull είναι γενίκευση της εκθετικής κατανομής με τη διαφορά ότι δεν έχει σταθερή συνάρτηση κινδύνου, όπως η εκθετική, γι' αυτό και έχει ευρύτερες εφαρμογές στο χώρο της βιομηχανίας, της ιατρικής και των βιολογικών επιστημών. Λόγω της καταλληλότητας στην περιγραφή χρόνων ζωής, αποτελεί μια από τις σημαντικότερες κατανομές στην ανάλυση επιβίωσης.

Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull, δίνεται από τη σχέση

$$f(t; \theta) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha), \quad t > 0,$$

με $\theta = (\alpha, \lambda) \in R^+ \times R^+$, όπου α και λ είναι η παράμετρος μορφής (shape) και κλίμακας (scale), αντίστοιχα. Τότε, συμβολίζουμε $T \sim Weib(\alpha, \lambda)$. Η εκθετική κατανομή αποτελεί ειδική περίπτωση της Weibull για $\alpha = 1$.

Η συνάρτηση επιβίωσης και κινδύνου της, είναι αντίστοιχα

$$S(t; \theta) = \exp(-\lambda t^\alpha) \text{ και } h(t; \theta) = \lambda \alpha t^{\alpha-1} \text{ με } t > 0,$$

όπου για την τελευταία έχουμε

- Για $\alpha > 1$, $\lim_{t \rightarrow 0^+} h(t; \theta) = 0$ και $\lim_{t \rightarrow \infty} h(t; \theta) = \infty$,
- Για $\alpha < 1$, $\lim_{t \rightarrow 0^+} h(t; \theta) = \infty$ και $\lim_{t \rightarrow \infty} h(t; \theta) = 0$,
- Για $\alpha = 1$, σταθερή (Εκθετική).

Η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής της, είναι της μορφής

$$m(t; \theta) = \frac{\exp(\lambda t^\alpha)}{\alpha \lambda^{\frac{1}{\alpha}}} \Gamma\left(\frac{1}{\alpha}\right) \left[1 - I\left(\frac{1}{\alpha}, \lambda t^\alpha\right) \right], t > 0,$$

όπου $I(\alpha, t) = \frac{1}{\Gamma(\alpha)} \int_0^t u^{\alpha-1} e^{-u} du$ είναι η μη πλήρης συνάρτηση Γάμμα.

Οι ροπές r τάξης περί το μηδέν της κατανομής Weibull, δίνονται από τη σχέση

$$E(T^r; \theta) = \frac{r}{\alpha \lambda^{\frac{r}{\alpha}}} \Gamma\left(\frac{r}{\alpha}\right),$$

όπου $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

Κατά συνέπεια, η μέση τιμή και η διακύμανση της, δίνονται αντίστοιχα από τις σχέσεις

$$E(T; \theta) = \frac{1}{\alpha \lambda^{\frac{1}{\alpha}}} \Gamma\left(\frac{1}{\alpha}\right),$$

και

$$\text{Var}(T; \theta) = \frac{1}{\lambda^{\frac{2}{\alpha}}} \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right) \right]^2 \right].$$

Η Κατανομή Γάμμα

Η κατανομή Γάμμα χρησιμοποιείται ευρύτατα στην ανάλυση επιβίωσης. Όμως, οι συναρτήσεις επιβίωσης και κινδύνου της, δεν μπορούν να γραφούν σε απλή κλειστή μορφή και έτσι μειώνεται ο βαθμός χρησιμότητάς της σε σχέση με άλλες κατανομές, όπως για παράδειγμα, η κατανομή Weibull.

Η συνάρτηση πυκνότητας πιθανότητας της Γάμμα κατανομής, δίνεται από τη σχέση

$$f(t; \theta) = \frac{1}{\Gamma(\alpha)} \lambda (\lambda t)^{\alpha-1} \exp(-\lambda t), \quad t > 0,$$

με $\theta = (\alpha, \lambda) \in R^+ \times R^+$, όπου α είναι παράμετρος μορφής και λ η παράμετρος κλίμακας. Τότε, συμβολίζουμε $T \sim G(\alpha, \lambda)$. Η εκθετική κατανομή αποτελεί ειδική περίπτωση της Γάμμα για $\alpha = 1$.

Οι συναρτήσεις επιβίωσης και κινδύνου της κατανομής Γάμμα, γράφονται αντίστοιχα συναρτήσεις της μη πλήρους συνάρτησης Γάμμα, στη μορφή

$$S(t; \theta) = 1 - I(\alpha, \lambda t), \quad t > 0,$$

και

$$h(t; \theta) = \frac{f(t; \theta)}{1 - I(\alpha, \lambda t)}, \quad t > 0.$$

Η συνάρτηση κινδύνου της, είναι

- αύξουσα για $\alpha > 1$, με $\lim_{t \rightarrow 0^+} h(t; \theta) = 0$ και $\lim_{t \rightarrow \infty} h(t; \theta) = \lambda$,
- φθίνουσα για $0 < \alpha < 1$, με $\lim_{t \rightarrow 0^+} h(t; \theta) = \infty$ και $\lim_{t \rightarrow \infty} h(t; \theta) = \lambda$,
- σταθερή για $\alpha = 1$ (Εκθετική).

Οι ροπές r τάξης περί το μηδέν της κατανομής Γάμμα, δίνονται από τη σχέση

$$E(T^r; \theta) = \frac{\alpha(\alpha+1)\dots(\alpha+r-1)}{\lambda^r}.$$

Κατά συνέπεια, η μέση τιμή και διακύμανση της, είναι αντίστοιχα

$$E(T; \theta) = \frac{\alpha}{\lambda} \text{ και } Var(T; \theta) = \frac{\alpha}{\lambda^2}.$$

Στην περίπτωση που η παράμετρος μορφής α παίρνει θετικές ακέραιες τιμές, $\alpha = n$, οι συναρτήσεις πυκνότητας πιθανότητας και επιβίωσης, είναι

$$f(t; \theta) = \frac{1}{(n-1)!} \lambda (\lambda t)^{n-1} \exp(-\lambda t), \quad t > 0,$$

και

$$S(t; \theta) = \sum_{\kappa=0}^{n-1} e^{-\lambda t} (\lambda t)^\kappa / \kappa!, \quad t > 0,$$

αντίστοιχα.

Η Κατανομή Log-normal

Η σημασία της λογαριθμοκανονικής (Log-normal) κατανομής στην ανάλυση επιβίωσης είναι ιδιαίτερα σημαντική. Παράδειγμα εφαρμογής

της είναι στην περιγραφή του χρόνου μέχρι την εμφάνιση καρκίνου των πνευμόνων σε καπνιστές. Έστω T τυχαία μεταβλητή που μετρά το χρόνο ζωής, τότε η T ακολουθεί λογαριθμοκανονική κατανομή με παραμέτρους μ και σ , αν η τυχαία μεταβλητή $Y = \log T$ ακολουθεί κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 .

Η συνάρτηση πυκνότητας πιθανότητας της $Y = \log T$, δίνεται από την σχέση

$$f(y) = \frac{1}{(2\pi)^{1/2} \sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad -\infty < y < \infty.$$

Επομένως, η συνάρτηση πυκνότητας πιθανότητας της λογαριθμοκανονικής κατανομής $T = \exp(y)$, είναι

$$f(t; \theta) = \frac{1}{(2\pi)^{1/2} \sigma t} \exp\left[-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right], \quad t > 0,$$

όπου $\theta = (\mu, \sigma^2) \in R \times R^+$ και συμβολίζουμε $T \sim \Lambda(\mu, \sigma^2)$.

Οι συναρτήσεις επιβίωσης και κινδύνου της, είναι αντίστοιχα

$$S(t; \theta) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right), \quad t > 0 \quad \text{και} \quad h(t; \theta) = \frac{f(t; \theta)}{S(t; \theta)}, \quad t > 0,$$

όπου με $\Phi(\cdot)$ συμβολίζεται η α.σ.κ. της $N(0, 1)$.

Η μέση τιμή και η διακύμανση της, είναι αντίστοιχα

$$E(T; \theta) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad \text{και} \quad \text{Var}(T; \theta) = [\exp(\sigma^2) - 1][\exp(2\mu + \sigma^2)].$$

Η Κατανομή Pareto II

Η κατανομή Pareto II (γνωστή και ως κατανομή του Lomax), έχει χρησιμοποιηθεί στη μελέτη βιωσιμότητας των επιχειρήσεων.

Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Pareto II, δίνεται από τη σχέση

$$f(t; \theta) = \frac{\alpha \lambda}{(1 + \lambda t)^{\alpha+1}}, \quad t \geq 0,$$

με $\theta = (\alpha, \lambda) \in R^+ \times R^+$, όπου α και λ είναι οι παράμετροι μορφής και κλίμακας, αντίστοιχα. Τότε, συμβολίζουμε $T \sim \text{Pareto}(\alpha, \lambda)$.

Οι συναρτήσεις επιβίωσης και κινδύνου της, είναι αντίστοιχα

$$S(t; \theta) = \frac{1}{(1 + \lambda t)^\alpha}, \quad t \geq 0,$$

και

$$h(t; \theta) = \frac{\alpha \lambda}{1 + \lambda t}, \quad t \geq 0,$$

όπου η τελευταία είναι φθίνουσα με $h(0; \theta) = \alpha \lambda$ και $\lim_{t \rightarrow \infty} h(t; \theta) = 0$.

Οι ροπές r τάξης περί το μηδέν της κατανομής Pareto II, είναι

$$E(T^r; \theta) = \frac{\Gamma(1+r)\Gamma(\alpha-r)}{\Gamma(\alpha)}, \quad -1 < r < \alpha.$$

Η μέση τιμή και η διακύμανση της, είναι αντίστοιχα

$$E(T; \theta) = \frac{1}{\lambda(\alpha - 1)}, \quad \alpha > 1,$$

και

$$\text{Var}(T; \theta) = \frac{\alpha}{(\alpha - 1)^2 (\alpha - 2) \lambda^2}, \quad \alpha > 2.$$

Η Γεωμετρική Κατανομή

Έστω T η τ.μ. που περιγράφει τον αριθμό των δοκιμών μέχρι την πρώτη επιτυχία ενός διωνυμικού τυχαίου πειράματος με πιθανότητα επιτυχίας p .

Η συνάρτηση πιθανότητας της τ.μ. T (γεωμετρικής κατανομής), δίνεται από τη σχέση

$$P(T = t) = pq^t, \text{ με } t = 0, 1, 2, \dots,$$

όπου $p + q = 1$, $0 \leq p \leq 1$, $0 \leq q \leq 1$. Συμβολίζουμε $T \sim \text{Geo}(p)$.

Η αθροιστική συνάρτηση κατανομής της, είναι

$$F(t) = \sum_{y=0}^t pq^y = p(1 + q + q^2 + \dots + q^t) = p \frac{1 - q^{t+1}}{1 - q} = 1 - q^{t+1}, \quad t \geq 0.$$

Η συνάρτηση επιβίωσης και η συνάρτηση κινδύνου της, είναι αντίστοιχα

$$S(t) = q^{t+1},$$

και

$$h(t) = \frac{P(T=t)}{P(T \geq t)} = \frac{pq^t}{\sum_{x=t}^{\infty} P(T=x)} = \frac{pq^t}{\sum_{x=t}^{\infty} pq^x} = \frac{pq^t}{1 - \sum_{x=0}^{t-1} pq^x} = \frac{pq^t}{1 - p \frac{1-q^t}{1-q}} = p.$$

Άρα, $h(t) = p$ δηλαδή είναι σταθερή συνάρτηση.

Η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής της, είναι

$$m(t) = \frac{1}{S(t)} \sum_{x=t}^{\infty} S(x) = \frac{1}{q^{t+1}} \sum_{x=t}^{\infty} q^{x+1}.$$

Όμως,

$$\sum_{x=0}^{\infty} q^{x+1} = \frac{q}{1-q} = \sum_{x=0}^{t-1} q^{x+1} + \sum_{x=t}^{\infty} q^{x+1}.$$

Επομένως,

$$\sum_{x=t}^{\infty} q^{x+1} = \frac{q}{1-q} - \sum_{x=0}^{t-1} q^{x+1} = \frac{q}{p} - q \frac{1-q^t}{1-q} = \frac{q}{p} (1 - 1 + q^t) = \frac{q^{t+1}}{p}.$$

Άρα, τελικά $m(t) = \frac{1}{p}$.

Η μέση τιμή και η διακύμανση της, είναι αντίστοιχα

$$E(T) = \frac{q}{p} \text{ και } Var(T) = \frac{q}{p^2}, \text{ με } t=0,1,2,\dots \text{ και } 0 < p < 1, 0 < q < 1.$$

Στον Πίνακα που ακολουθεί συνοψίζονται όσα δόθηκαν στην ενότητα αυτή και αφορούν τη σ.π.π. ή τη σ.π., τη συνάρτηση επιβίωσης, τη συνάρτηση κινδύνου, τη μέση τιμή και τη διακύμανση των κατανομών που πρωτότερα αναφέρθηκαν.

Κατανομή	Συνάρτηση Πιθανότητας $f(t)$	Συνάρτηση Επιβίωσης $S(t)$	Συνάρτηση Κινδύνου $h(t)$	Μέση Τιμή $E(T)$	Διακύμανση $Var(T)$
Εκθετική $Exp(\lambda), \lambda > 0$	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Weibull $Weib(\alpha, \lambda)$ $\alpha, \lambda > 0$	$\lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)$	$\exp(-\lambda t^\alpha)$	$\lambda \alpha t^{\alpha-1}$	$\frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right)$ $\alpha \lambda^\alpha$	$\frac{1}{\alpha^2} \left[\Gamma\left(\frac{2}{\alpha}\right) - \left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2 \right]$
Γάμμα $G(\alpha, \lambda)$ $\alpha, \lambda > 0$	$\frac{1}{\Gamma(\alpha)} \lambda (\lambda t)^{\alpha-1} \exp(-\lambda t)$	$1 - I(\alpha, \lambda t)$	$\frac{f(t; \theta)}{1 - I(\alpha, \lambda t)}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Log-normal $\Lambda(\mu, \sigma^2)$ $\mu \geq 0, \sigma > 0$	$\frac{1}{(2\pi)^{1/2} \sigma t} \exp\left[-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma}\right)^2\right]$	$1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$	$\frac{f(t)}{S(t)}$	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$	$\frac{[\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)}{[\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)}$
Pareto $Pareto(\alpha, \lambda)$ $\alpha, \lambda > 0$	$\frac{\alpha \lambda}{(1 + \lambda t)^{\alpha+1}}$	$\frac{1}{(1 + \lambda t)^\alpha}$	$\frac{\alpha \lambda}{1 + \lambda t}$	$\frac{1}{\lambda(\alpha - 1)}$	$\frac{\alpha}{(\alpha - 1)^2 (\alpha - 2) \lambda^2}$
Γεωμετρική $Geo(p)$	$p q^i, p + q = 1$	q^{i+1}	p	$\frac{q}{p}, p \neq 0$	$\frac{q}{p^2}, p \neq 0$

1.6 Συνάρτηση πιθανοφάνειας

Η μέθοδος της μέγιστης πιθανοφάνειας προτάθηκε για πρώτη φορά από τον Gauss το 1821, πιστώνεται όμως στον Fisher, γιατί πρώτος αυτός ερεύνησε το 1922 τις ιδιότητες της μεθόδου.

Οι παρακάτω ορισμοί αναφέρονται στο βιβλίο Μαθηματικής Στατιστικής (Παπαϊωάννου και Φερεντίνος, 2002).

Ορισμός 1.7

Έστω $x = (x_1, x_2, \dots, x_n)$ τυχαίο δείγμα από ανεξάρτητες και ισόνομες παρατηρήσεις με συνάρτηση πιθανότητας ή συνάρτηση πυκνότητας πιθανότητας $p(x; \theta)$, όπου θ μια μονοδιάστατη ή πολυδιάστατη παράμετρος. Τότε, η συνάρτηση $L(\theta, x) = \prod_{i=1}^n p(x_i; \theta)$ ονομάζεται **συνάρτηση πιθανοφάνειας**.

Ορισμός 1.8

Ο **εκτιμητής μέγιστης πιθανοφάνειας (EMΠ)**, $\hat{\theta}$, της παραμέτρου θ του μοντέλου, είναι η τιμή που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας. Δηλαδή $L(\hat{\theta}; x) = \sup_{\theta \in \Theta} L(\theta; x)$, όπου Θ ο παραμετρικός χώρος.

Συχνά, στη μέθοδο της μέγιστης πιθανοφάνειας χρησιμοποιείται ο λογάριθμος της συνάρτησης πιθανοφάνειας $l(\theta, x) = \log L(\theta, x)$, καθώς είναι ευκολότερη η χρήση του και η μεγιστοποίηση των δυο συναρτήσεων επιτυγχάνεται στην ίδια τιμή του θ .

Εκτός από ορισμένες περιπτώσεις (ομοιόμορφη κατανομή, κ.α.), για την εύρεση του εκτιμητή μέγιστης πιθανοφάνειας, χρησιμοποιούμε τις εξισώσεις πιθανοφάνειας

$$\frac{\partial}{\partial \theta} l(\theta, x) = 0.$$

1.7 EM αλγόριθμος

Πολλές φορές η μέθοδος της μέγιστης πιθανοφάνειας είναι πολύπλοκη και δεν φέρνει τα αναμενόμενα αποτελέσματα όσον αφορά τον τρόπο εύρεσης εκτιμητή. Οι λόγοι που καταφεύγουμε σε αριθμητικές επαναληπτικές μεθόδους είναι:

- Η συνάρτηση πιθανοφάνειας δεν είναι παραγωγίσιμη ή έχει πολλαπλά μέγιστα.
- Στις διακριτές κατανομές, ελάχιστες είναι οι περιπτώσεις στις οποίες είναι εφικτή η επίλυση του συστήματος των εξισώσεων πιθανοφάνειας και η εύρεση λύσης του σε κλειστή μορφή.

Οι πιο δημοφιλείς μέθοδοι βελτιστοποίησης είναι η μέθοδος Newton-Raphson και η μέθοδος scoring.

- Η διαδικασία των **Newton-Raphson** περιγράφεται ως εξής: Έστω θ_0 μια τυχούσα αρχική εκτίμηση της παραμέτρου θ . Τότε, η ακολουθία $\{\theta_i\}$, όπου

$$\theta_{i+1} = \theta_i - \left[\frac{\partial^2 \log L(\theta / x)}{\partial \theta^2} \right]^{-1} \left[\frac{\partial \log L(\theta / x)}{\partial \theta} \right]_{\theta=\theta_i}, \quad i = 0, 1, 2, \dots$$

συγκλίνει προς τη ζητούμενη λύση.

- Η μέθοδος *scoring* είναι μια παραλλαγή της Newton-Raphson, όπου στη θέση του πίνακα $I(\theta; x) = -\frac{\partial^2}{\partial \theta^2} \log p(\theta; x)$, εμφανίζεται η αναμενόμενη τιμή του ως προς X :

$$E_x [I(\theta; X)] = E_x \left[-\frac{\partial^2}{\partial \theta^2} \log p(\theta; X) \right],$$

που είναι το μέτρο πληροφορίας του Fisher $I_X^F(\theta)$.

Όμως, κατά τη διάρκεια ενός τυχαίου πειράματος μπορεί να έχουμε απώλεια πληροφορίας για διάφορους λόγους. Τότε, μπορούμε να εφαρμόσουμε τον EM αλγόριθμο για την εύρεση εκτιμητή μέγιστης πιθανοφάνειας των παραμέτρων.

Ο αλγόριθμος EM (Expectation Maximization) είναι μια επαναληπτική διαδικασία εύρεσης εκτιμητών μέγιστης πιθανοφάνειας των παραμέτρων μιας δοθείσας κατανομής. Εφαρμόζεται σε προβλήματα, όπου κάποιες μεταβλητές δεν έχουν παρατηρηθεί (μη παρατηρήσιμες ή κρυμμένες μεταβλητές). Συνεπώς, η εφαρμογή του EM συνιστάται για την επίλυση δυο βασικών προβλημάτων.

- Το πρώτο αφορά περιπτώσεις που έχουμε ελλιπή δεδομένα, δηλαδή δεδομένα από τα οποία ορισμένες τιμές λείπουν, εξαιτίας λαθών ή περιορισμών που υπάρχουν κατά τη διάρκεια της διαδικασίας παρατήρησης. Για παράδειγμα, στη μελέτη του χρόνου ζωής κάποιων ασθενών που πάσχουν από μια ασθένεια, το πείραμα το πιθανότερο είναι να σταματήσει πριν πεθάνουν όλοι οι ασθενείς. Κατά συνέπεια, δε γνωρίζουμε τον ακριβή χρόνο ζωής για όλους τους ασθενείς παρά μόνον αυτών που πέθαναν.
- Το δεύτερο αφορά κυρίως εφαρμογές μικτών μοντέλων, στα οποία η μεγιστοποίηση της πιθανοφάνειας είναι αναλυτικά αδύνατη. Γι' αυτό υποθέτουμε την ύπαρξη κάποιων επιπρόσθετων κρυμμένων μεταβλητών, που προσδιορίζουν την ομάδα στην οποία ανήκει το κάθε πρότυπο.

Η γενική φιλοσοφία του EM αλγορίθμου διατυπώνεται ακολούθως. Ξεκινάμε με μια αρχική εκτίμηση $\Theta^{(0)}$ των παραμέτρων του μικτού μοντέλου, που πρέπει να εκτιμηθούν. Κάθε επανάληψη αποτελείται από δυο βήματα.

- Το πρώτο είναι το βήμα αναμενόμενης τιμής (E-step ή expectation step), στο οποίο προσπαθούμε να υπολογίσουμε ένα τοπικό κάτω φράγμα του λογαρίθμου της συνάρτησης πιθανοφάνειας και να το μεγιστοποιήσουμε ως προς την κατανομή των κρυμμένων μεταβλητών.
- Το δεύτερο βήμα είναι το βήμα μεγιστοποίησης (M-step ή maximization step), στο οποίο μεγιστοποιείται το κάτω φράγμα ως προς τις παραμέτρους $\Theta = \left[(\pi_j, \theta_j), j = 1, \dots, M \right]$ της μικτής κατανομής, υποθέτοντας πάντα ότι η κατανομή των κρυμμένων μεταβλητών που βρέθηκε στο E-step, είναι σωστή.

Αυτά τα δυο βήματα επαναλαμβάνονται μέχρι να υπάρξει σύγκλιση στην ακολουθία των παραμέτρων, δηλαδή όταν φτάσουμε σε κάποιο τοπικό μέγιστο.

Βελτιστοποίηση μέσω του αλγορίθμου EM

Η μέθοδος του EM αλγορίθμου γίνεται πιο εύκολα αντιληπτή στην περίπτωση τυχαίου δείγματος, το οποίο αποτελείται από παρατηρηθείσες και μη παρατηρηθείσες ή ελλιπούσες τιμές.

Ένα παράδειγμα δείγματος με ελλιπούσες τιμές προκύπτει στην περίπτωση ελέγχου του χρόνου επιβίωσης. Ενδεικτικά, αναφέρουμε το πείραμα, όπου ένας αριθμός ηλεκτρικών λαμπτήρων ανάβει συνεχώς και καταμετράται ο χρόνος που χρειάζεται μέχρι να πάψουν να λειτουργούν. Σε ένα τέτοιο παράδειγμα, είναι συνήθες φαινόμενο το πείραμα να διακοπεί πριν πάψουν να λειτουργούν όλοι οι λαμπτήρες. Ο χρόνος επιβίωσης των λαμπτήρων, οι οποίοι συνεχίζουν να δουλεύουν δεν έχει

παρατηρηθεί. Όμως, προφανώς ο αριθμός των λογοκριμένων παρατηρήσεων και ο χρόνος της λογοκρισίας περιέχουν πληροφορία για τη κατανομή του χρόνου επιβίωσης.

Ακόμη, ένα γνωστό παράδειγμα, στο οποίο μπορεί να χρησιμοποιηθεί ο EM αλγόριθμος είναι το πεπερασμένο μοντέλο μίξης κατανομών. Κάθε παρατήρηση προέρχεται από μια άγνωστη παρατήρηση ενός υποτιθέμενου συνόλου κατανομών. Οι ελλιπούσες τιμές προσδιορίζουν την κατανομή. Οι παράμετροι των κατανομών πρόκειται να εκτιμηθούν. Ένα παράπλευρο κέρδος της μεθόδου είναι ότι εκτιμάται σε ποιά κατηγορία ανήκουν τα δεδομένα.

Τα ελλιπή δεδομένα μπορούν να είναι ελλιπούσες παρατηρήσεις της ίδιας τυχαίας μεταβλητής, η οποία παράγει το δείγμα που παρατηρήθηκε, όπως στην περίπτωση του παραδείγματος λογοκρισίας ή μπορούν να προέρχονται από μια διαφορετική τυχαία μεταβλητή, η οποία σχετίζεται με κάποιο τρόπο με την τυχαία μεταβλητή που έχει παρατηρηθεί.

Ο Sundberg (1974, 1976) μελέτησε τις ιδιότητες της εξίσωσης πιθανοφάνειας στο γενικό πλαίσιο των προβλημάτων ελλιπών δεδομένων από εκθετικές οικογένειες και έφτασε σε ειδικές σχέσεις για την εξίσωση πιθανοφάνειας και τον πίνακα πληροφορίας, οι οποίες έχουν μείνει γνωστές ως σχέσεις του Sundberg. Όλα αυτά έγιναν πριν την γενική μορφή της μεθόδου EM, που εμφανίζεται στο άρθρο των Dempster, Laird and Rubin (1977) και McLachlan and Krishnan (2008). Όπως επισημάνθηκε από τους Little and Rubin (1983), ο EM αλγόριθμος θα συγκλίνει αξιόπιστα, αλλά με αργούς ρυθμούς, όταν ο όγκος των πληροφοριών στα στοιχεία που λείπουν είναι σχετικά μεγάλος. Πρόσφατα, ο EM αλγόριθμος χρησιμοποιήθηκε από πολλούς συγγραφείς, όπως οι Adamidis and Loukas (1998), Adamidis (1999), Ng et al. (2002), Karlis (2002) και Adamidis et al. (2005). Ο EM αλγόριθμος περιέχει δυο βήματα: Στο E-βήμα (expectation step) υπολογίζουμε την υπό συνθήκη αναμενόμενη τιμή του λογαρίθμου της συνάρτησης πιθανοφάνειας των πλήρων δεδομένων $l(\theta; x)$, δοθέντων των ελλιπών

δεδομένων και μιας τιμής της παραμέτρου. Στο M-βήμα (maximization step) γίνεται η μεγιστοποίηση της εξίσωσης.

Δύο σημαντικές ερμηνείες της μεθόδου είναι:

- i. Η μέθοδος προκύπτει λόγω ειδικών μορφών που παίρνουμε από τις παραγωγίσεις του λογαρίθμου της συνάρτησης πιθανοφάνειας. Πολλοί συγγραφείς στην προσπάθεια τους να χειριστούν τις εξισώσεις πιθανοφάνειας, έχουν φτάσει στον EM, ώστε να τις λύσουν με πιο κομψό τρόπο.
- ii. Η εύρεση εκτιμητή μέγιστης πιθανοφάνειας σε πολλά προβλήματα είναι πολύπλοκη, όμως με κατάλληλη μορφοποίηση, θεωρείται ως πρόβλημα ελλειπών δεδομένων και μπορεί να λυθεί με τη μέθοδο EM με απλούστερο τρόπο.

Ο EM αλγόριθμος σε μικτές κατανομές

Οι μίξεις κατανομών είναι μια μεγάλη κατηγορία μοντέλων που χρησιμοποιούνται ευρύτατα για να περιγράψουν ανομοιογενείς πληθυσμούς. Οι εφαρμογές τους εντοπίζονται στη στατιστική, στη βιολογία, κ.τ.λ. Ειδικά στους βιολογικούς πληθυσμούς, τα μοντέλα μίξεων αποτελούν βασικό εργαλείο για την περιγραφή τους. Στο χώρο της αλιείας, η παρουσία ψαριών με διαφορετική ηλικία στο δείγμα, όπου το βάρος του ψαριού είναι συνάρτηση της ηλικίας του, μας οδηγεί στην απόρριψη του απλού ομογενούς μοντέλου. Επομένως, τα μοντέλα μίξεων είναι καταλληλότερα στις περιπτώσεις αυτές.

Μια μικτή κατανομή ορίζεται ως ειδική περίπτωση γραμμικού συνδυασμού ενός πεπερασμένου αριθμού συναρτήσεων πυκνότητας πιθανότητας. Πιο συγκεκριμένα, η πιθανότητα μιας τυχαίας μεταβλητής X , η οποία ακολουθεί μικτή κατανομή, υπολογίζεται από το άθροισμα M συναρτήσεων πυκνότητας πιθανότητας σταθμισμένων με βάρη. Η διατύπωση ενός τέτοιου μοντέλου αποτυπώνεται στην παρακάτω σχέση:

$$p(x; \Theta) = \sum_{j=1}^M \pi_j f(x; \theta_j),$$

όπου $f(x; \theta_j)$ αναπαριστά την j -οστή συνιστώσα κατανομή του μικτού μοντέλου και θ_j είναι το αντίστοιχο διάνυσμα των παραμέτρων της. Τα π_j αποκαλούνται παράμετροι μίξης, οι οποίες θα πρέπει να ικανοποιούν τους παρακάτω περιορισμούς:

$$0 \leq \pi_j \leq 1, \text{ για κάθε } j,$$

και

$$\sum_{j=1}^M \pi_j = 1.$$

Οι παραπάνω περιορισμοί είναι ταυτόσημοι με αυτούς που πρέπει να πληρούνται, έτσι ώστε μια ποσότητα να θεωρείται πιθανότητα. Άρα, τα βάρη π_j μπορούν να ερμηνευτούν ως η εκ των προτέρων πιθανότητα, σύμφωνα με την οποία ένα διάνυσμα μπορεί να παραχθεί από την j -οστή συνιστώσα κατανομής του ολικού μοντέλου.

Έστω τυχαίο δείγμα $x = (x_1, x_2, \dots, x_n)$. Τότε, η συνάρτηση πιθανοφάνειας είναι:

$$L = \prod_{i=1}^n p(x_i; \Theta) = \prod_{i=1}^n \left(\sum_{j=1}^M \pi_j f(x_i; \theta_j) \right),$$

και

$$l = \log L = \sum_{i=1}^n \log \left(\sum_{j=1}^M \pi_j f(x_i; \theta_j) \right).$$

Οι εξισώσεις, από τις οποίες προκύπτουν οι εκτιμητές μέγιστης πιθανοφάνειας $\hat{\theta}_j$ και $\hat{\pi}_j$, λαμβάνοντας υπόψη ότι $\sum_{j=1}^M \pi_j = 1$, είναι οι εξής:

$$\frac{\partial l}{\partial \theta_j} = \sum_{i=1}^n \frac{\pi_j}{p(x_i; \Theta)} \frac{\partial f(x_i; \theta_j)}{\partial \theta_j} = 0, \text{ με } j = 1, \dots, M,$$

και

$$\frac{\partial l}{\partial \pi_j} = \sum_{i=1}^n \frac{f(x_i; \theta_j) - f(x_i; \theta_k)}{p(x_i; \Theta)} = 0, \text{ με } j = 1, \dots, M-1.$$

2. ΚΑΤΑΝΟΜΕΣ ΦΘΙΝΟΥΣΑΣ ΣΥΝΑΡΤΗΣΗΣ ΚΙΝΔΥΝΟΥ

Εισαγωγή

Στη Στατιστική, η σημασία των μοντέλων κατανομών μίξεων είναι μεγάλη και το πεδίο εφαρμογών τους συνεχώς αυξάνεται. Θα εξηγήσουμε τη σημαντικότητα των μίξεων κατανομής Poisson. Στις περιπτώσεις που η απλή κατανομή Poisson αποτυγχάνει να περιγράψει τα δεδομένα, οι μίξεις της κατανομής Poisson χρησιμοποιούνται ως εναλλακτικά μοντέλα. Μια χαρακτηριστική ιδιότητα της κατανομής Poisson είναι ότι η διακύμανσή της είναι ίση με την αναμενόμενη τιμή της, το οποίο αρκετές φορές δε συμβαίνει στην πράξη.

Οι μίξεις της κατανομής Poisson μπορούν να περιγράψουν την ανομοιογένεια του πληθυσμού και έτσι αποκαλούνται ευέλικτα εναλλακτικά μοντέλα. Τα άτομα που συνιστούν τον πληθυσμό δεν έχουν την ίδια συχνότητα εμφάνισης του υπό εξέταση γεγονότος, εξαιτίας της ανομοιογένειας του πληθυσμού. Η συχνότητα αυτή περιγράφεται από την παράμετρο της κατανομής Poisson και συνεπώς, ο καθένας μπορεί να έχει μια διαφορετική τιμή για τη συχνότητα αυτή. Επομένως, η συχνότητα εμφάνισης του γεγονότος που περιγράφεται από την παράμετρο της κατανομής Poisson είναι μια τυχαία μεταβλητή που ακολουθεί κάποια κατανομή, η οποία αποκαλείται κατανομή μίξης. Τότε, από το θεώρημα ολικής πιθανότητας προκύπτουν οι μίξεις της κατανομής Poisson και εξετάζονται οι ιδιότητές τους. Λόγω της πολυπλοκότητάς τους, μόνο λίγες τέτοιες κατανομές έχουν μελετηθεί.

Οι πεπερασμένες μίξεις αποτελούν μια ενδιαφέρουσα κατηγορία μεταξύ των μίξεων της κατανομής Poisson. Είναι ιδιαίτερα διαδεδομένα λόγω της απλής φυσικής ερμηνείας τους ως μοντέλα που περιγράφουν έναν πληθυσμό με πεπερασμένο αριθμό υποπληθυσμών.

Ακόμη και στις περιπτώσεις, στις οποίες οι κατανομές μίξης είναι συνεχείς, η εκτίμησή τους συνίσταται στην εκτίμηση μιας κατανομής με πεπερασμένο αριθμό σημείων με μη αρνητική πιθανότητα.

Όμως, η επιλογή του κατάλληλου παραμετρικού μοντέλου για την περιγραφή του χρόνου ζωής του υπό εξέταση πληθυσμού αποτελεί ένα σημαντικό πρόβλημα στην ανάλυση της επιβίωσης. Μια τέτοια επιλογή γίνεται, είτε με γνώμονα τις φυσικές διεργασίες που οδηγούν στην “αποτυχία”, είτε με κριτήριο την ποιότητα προσαρμογής του εκάστοτε μοντέλου στα δεδομένα χρόνων ζωής.

Για να προσεγγίσουμε μια ερμηνεία των εν λόγω μοντέλων μίξεων (EG, EP, EL, EPS), ας σκεφτούμε μια κατάσταση, όπου η αποτυχία (για παράδειγμα, μιας συσκευής ή ενός συστήματος) παρουσιάζεται εξαιτίας της παρουσίας ενός άγνωστου αριθμού Z αρχικών ελαττωμάτων του ίδιου τύπου. Τα Y αντιπροσωπεύουν τη διάρκεια ζωής των εν λόγω εξαρτημάτων ή εν γένει συστατικών του συστήματος και για το λόγο αυτό ακολουθούν εκθετική κατανομή. Κάθε ελάττωμα μπορεί να ανιχνευθεί μόνο αφού προκαλέσει αποτυχία και σε αυτήν την περίπτωση επισκευάζεται πλήρως.

2.1 Η Εκθετική - Γεωμετρική κατανομή

Στην ενότητα αυτή παρουσιάζεται η Εκθετική-Γεωμετρική κατανομή, η οποία μελετήθηκε αναλυτικά από τους Adamidis and Loukas (1998).

2.1.1 Ορισμοί και φυσική ερμηνεία

Θεώρημα 2.1 (Adamidis and Loukas, 1998)

Έστω $\{Y_i\}_{i=1}^Z$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με πυκνότητα

$$f(y; \beta) = \beta e^{-\beta y}, \quad y \geq 0,$$

όπου $\beta \in R^+$ και Z γεωμετρική τυχαία μεταβλητή με συνάρτηση πιθανότητας $P(z; p) = (1-p)p^{z-1}$, $z \in N - \{0\}$ με $p \in (0,1)$. Τότε, η τυχαία μεταβλητή $X = \min\left(\{Y_i\}_{i=1}^Z\right)$, ακολουθεί την κατανομή Exponential-Geometric (EG), που θα αναφέρεται στο εξής ως Εκθετική-Γεωμετρική με παραμέτρους $\theta = (\beta, \gamma) \in R^+ \times R^+$, όπου $\gamma = 1-p$.

Απόδειξη (Adamidis and Loukas, 1998)

Από τις υποθέσεις του θεωρήματος προκύπτει εύκολα ότι η δεσμευμένη κατανομή της τυχαίας μεταβλητής X/Z είναι εκθετική με παράμετρο $z\beta$. Κατά συνέπεια, η από κοινού κατανομή των X και Z δίνεται από τη σχέση

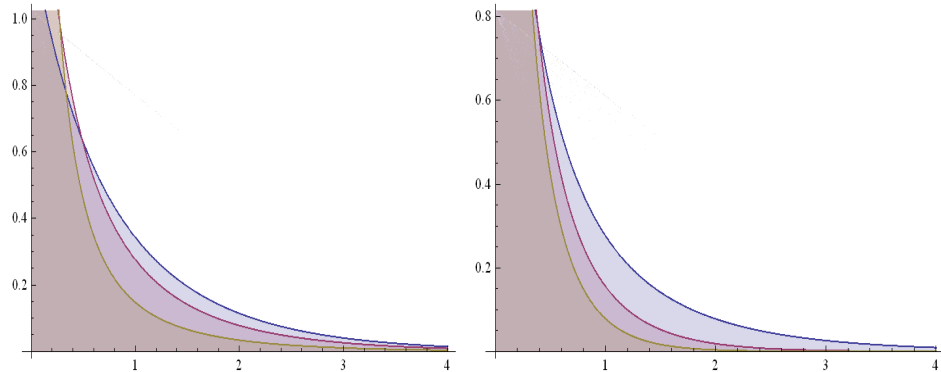
$$f(x, z; \beta, p) = z\beta(1-p)p^{z-1}e^{-z\beta x}.$$

Άρα,

$$\begin{aligned} f(x; \beta, p) &= \beta(1-p)e^{-\beta x} \sum_{n=1}^{\infty} z \left(p e^{-\beta x} \right)^{z-1} \\ &= \beta(1-p)e^{-\beta x} \left(\sum_0^{\infty} t^z \right)' \Big|_{t=pe^{-\beta x}} \\ &= \frac{\beta(1-p)e^{-\beta x}}{\left(1-pe^{-\beta x}\right)^2}. \end{aligned}$$

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X ορίζει την κατανομή που καλείται Εκθετική-Γεωμετρική.

Στο σχήμα που ακολουθεί δίνεται η γραφική παράσταση αυτής της σ.π.π. για ειδικές τιμές των παραμέτρων και για $x \in (0, 4)$.



Εικόνα 1. Η συνάρτηση πυκνότητας πιθανότητας της EG α) για $\beta=1$ και $p=0.2, 0.5, 0.8$ και β) για $p=0.5$ και $\beta=1, 2, 3$.

2.1.2 Ιδιότητες της κατανομής

Στη συνέχεια παραθέτουμε κάποια χρήσιμα πορίσματα που απορρέουν από τον ορισμό της Εκθετικής-Γεωμετρικής κατανομής και ουσιαστικά συνδέουν αυτήν την κατανομή με υπάρχουσες. Υποθέτουμε λοιπόν ότι η X ακολουθεί την EG κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f(x; \Theta) = \beta(1-p)e^{-\beta x}(1-pe^{-\beta x})^{-2}, \quad x \geq 0, \quad (2.1)$$

με $\theta = (\beta, \gamma) \in R^+ \times R^+$, όπου οι παράμετροι β και $\gamma = 1-p$, είναι παράμετροι κλίμακας (scale) και μορφής (shape), αντίστοιχα.

Πόρισμα 2.1 (Adamidis and Loukas, 1998)

Αν η X ακολουθεί την EG κατανομή με συνάρτηση πυκνότητας πιθανότητας που δίνεται από την (2.1), η τυχαία μεταβλητή

$Y = p^{-1}(e^{\beta X} - 1)$ ακολουθεί την κατανομή Pareto II με παραμέτρους σχήματος και κλίμακας 1 και $p/(1-p)$, αντίστοιχα.

Απόδειξη (Adamidis and Loukas, 1998)

Έστω $F_Y(y)$ η αθροιστική συνάρτηση κατανομής της μεταβλητής Y , για $y \in R$. Τότε, έχουμε:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(p^{-1}(e^{\beta X} - 1) \leq y) = P(e^{\beta X} \leq py + 1) = P(\beta X \leq \log(py + 1)) \\ &= P(X \leq 1/\beta \cdot \log(py + 1)) = F_X(1/\beta \cdot \log(py + 1)), \end{aligned}$$

όπου F_X είναι η αντίστοιχη συνάρτηση κατανομής της τυχαίας μεταβλητής X . Παραγωγίζοντας προκύπτει:

$$f_Y(y) = \left(F_X(1/\beta \cdot \log(py + 1)) \right)' = \frac{p}{\beta(py + 1)} \cdot f_X(1/\beta \cdot \log(py + 1)).$$

Με αντικατάσταση στην (2.1), όπου x την τιμή $1/\beta \cdot \log(py + 1)$ έχουμε:

$$\begin{aligned} f_Y(y) &= \frac{p}{\beta(py + 1)} \cdot \beta(1-p) e^{-\beta(1/\beta \cdot \log(py + 1))} \left(1 - p e^{-\beta(1/\beta \cdot \log(py + 1))} \right)^{-2} \\ &= \frac{p}{py + 1} \cdot (1-p) e^{\log(py + 1)^{-1}} \left(1 - p e^{\log(py + 1)^{-1}} \right)^{-2}. \end{aligned}$$

Επομένως,

$$\begin{aligned} f_Y(y) &= \frac{p}{py + 1} (1-p) \frac{1}{py + 1} \left(1 - \frac{p}{py + 1} \right)^{-2} \\ &= \frac{p(1-p)}{(py - p + 1)^2}. \end{aligned}$$

Άρα, είναι της μορφής $f(y; \alpha, \lambda) = \frac{\alpha \lambda}{(1 + \lambda y)^{\alpha+1}}$, όπου α και λ είναι

οι παράμετροι σχήματος και κλίμακας, αντίστοιχα, της συνάρτησης πυκνότητας πιθανότητας της κατανομής Pareto II, που ορίστηκε στο προηγούμενο κεφάλαιο με $\alpha = 1$ και $\lambda = p / (1 - p)$.

Πόρισμα 2.2 (Adamidis and Loukas, 1998)

Αν η X ακολουθεί την EG κατανομή με συνάρτηση πυκνότητας πιθανότητας που δίνεται από την (2.1), η τυχαία μεταβλητή $Y = -\beta^{-1} \log(1 - e^{-\beta X})$ ακολουθεί την αποκομμένη λογιστική κατανομή στο R^+ με παραμέτρους θέσης και κλίμακας $\beta^{-1} \log\{p(1-p)^{-1}\}$ και β , αντίστοιχα.

Απόδειξη (Adamidis and Loukas, 1998)

Έστω $F_Y(y)$ η αθροιστική συνάρτηση κατανομής της μεταβλητής Y , για $y \in R$. Τότε, έχουμε:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(-\beta^{-1} \log(1 - e^{-\beta X}) \leq y) = P(\log(1 - e^{-\beta X})^{-1} \leq \beta y) \\ &= P\left(\left(1 - e^{-\beta X}\right)^{-1} \leq e^{\beta y}\right) = P(1 - e^{-\beta X} \geq e^{-\beta y}) = P(-e^{-\beta X} \geq e^{-\beta y} - 1) \\ &= P(e^{-\beta X} \leq 1 - e^{-\beta y}) = P(X \geq -\beta^{-1} \log(1 - e^{-\beta y})). \end{aligned}$$

Επομένως, $F_Y(y) = 1 - F_X(-\beta^{-1} \log(1 - e^{-\beta y}))$, $y \in R^+$,

όπου F_X είναι η αντίστοιχη συνάρτηση κατανομής για την τυχαία μεταβλητή X . Παραγωγίζοντας προκύπτει:

$$\begin{aligned} f_Y(y) &= F'_Y(y) = -\left(F_X\left(-\beta^{-1} \cdot \log(1-e^{-\beta y})\right)\right)' \\ &= \frac{\beta e^{-\beta y}}{\beta(1-e^{-\beta y})} f_X\left(-\beta^{-1} \log(1-e^{-\beta y})\right). \end{aligned}$$

Με αντικατάσταση στην (2.1), όπου x την τιμή $-\beta^{-1} \log(1-e^{-\beta y})$ έχουμε:

$$\begin{aligned} f_Y(y) &= \frac{e^{-\beta y}}{(1-e^{-\beta y})} \cdot \beta(1-p) e^{-\beta(-\beta^{-1} \cdot \log(1-e^{-\beta y}))} \left(1 - p e^{-\beta(-\beta^{-1} \cdot \log(1-e^{-\beta y}))}\right)^{-2} \\ &= \frac{e^{-\beta y}}{(1-e^{-\beta y})} \cdot \beta(1-p)(1-e^{-\beta y}) (1-p(1-e^{-\beta y}))^{-2} \\ &= \beta(1-p) \frac{e^{-\beta y}}{(1-p + p e^{-\beta y})^2}, y \in R^+. \end{aligned}$$

Είναι γνωστό ότι, αν X ακολουθεί λογιστική κατανομή, τότε η σ.π.π. της δίνεται από τη σχέση

$$f(x; \mu, s) = \frac{e^{-\left(\frac{x-\mu}{s}\right)}}{s \left(1 + e^{-\left(\frac{x-\mu}{s}\right)}\right)^2}, x \in R.$$

Επομένως, η Y ακολουθεί τη λογιστική κατανομή με παραμέτρους

$$\mu = \beta^{-1} \log\{p(1-p)^{-1}\} \text{ και } \beta = \frac{1}{s},$$

και περιορισμό του πεδίου ορισμού της στο $(0, \infty)$. Γι' αυτό το λόγο λέμε ότι ακολουθεί την αποκομμένη (truncated) λογιστική κατανομή.

2.1.3 Χαρακτηριστικά της κατανομής

Δίνονται κατά σειρά η *αθροιστική συνάρτηση κατανομής*, η *διάμεσος*, οι *ροπές*, η *μέση τιμή*, η *διακύμανση*, η *συνάρτηση επιβίωσης*, η *συνάρτηση κινδύνου* και ο *μέσος υπολειπόμενος χρόνος*.

Η *αθροιστική συνάρτηση* κατανομής της τ.μ. X που ακολουθεί την Εκθετική-Γεωμετρική κατανομή με σ.π.π. αυτής της σχέσης (2.1), δίνεται από τη σχέση:

$$F(x; \theta) = \int_0^x \beta(1-p)e^{-\beta x} (1-pe^{-\beta x})^{-2} dx, \quad x \geq 0,$$

όπου με αντικατάσταση:

$$y = 1 - pe^{-\beta x} \Rightarrow dy = p\beta e^{-\beta x} dx,$$

το αρχικό ολοκλήρωμα γράφεται:

$$\begin{aligned} F(x; \theta) &= \int_{1-p}^y \frac{(1-p)}{p} y^{-2} dy \\ &= \left[\frac{1-p}{p} \left(-\frac{1}{y} \right) \right]_{1-p}^y \\ &= -\frac{1-p}{p} (1-pe^{-\beta x})^{-1} + \frac{1}{p} \\ &= (1-e^{-\beta x})(1-pe^{-\beta x})^{-1}. \end{aligned}$$

Η *διάμεσος* είναι: $x = \beta^{-1} \log(2-p)$.

Για την απόδειξη της σχέσης αυτής, αρκεί να δείξουμε ότι:

$$F(\beta^{-1} \log(2-p); \theta) = \frac{1}{2}.$$

Όντως είναι:

$$\begin{aligned} F(\beta^{-1} \log(2-p); \theta) &= \left(1 - e^{\log(2-p)^{-1}}\right) \left(1 - p e^{\log(2-p)^{-1}}\right)^{-1} \\ &= \left(1 - \frac{1}{2-p}\right) \left(1 - \frac{p}{2-p}\right)^{-1} = \frac{1}{2}. \end{aligned}$$

Οι r -ροπές δίνονται από τη σχέση:

$$E(X^r; \theta) = (1-p)r!(\beta^r p)^{-1} L(p; r),$$

όπου $L(p; r) = \sum_{j=1}^{\infty} p^j j^{-r}$ είναι η γενίκευση της διλογαριθμικής συνάρτησης του Euler.

Θέτοντας $\gamma = 1-p$,

$$E(X^r; \theta) = \int_0^{\infty} \frac{x^r \beta \gamma e^{-\beta x}}{\{1 - (1-\gamma)e^{-\beta x}\}^2} dx,$$

όπου με αντικατάσταση $\beta x = t$ προκύπτει:

$$\begin{aligned}
E(X^r; \theta) &= \frac{\gamma}{\beta^r} \int_0^{\infty} \frac{t^r e^{-t}}{\{1 - (1 - \gamma)e^{-t}\}^2} dt \\
&= \frac{-\gamma}{(1 - \gamma)\beta^r} \int_0^{\infty} t^r d \left\{ \frac{1}{1 - (1 - \gamma)e^{-t}} \right\} \\
&= \frac{-\gamma}{(1 - \gamma)\beta^r} \frac{t^r}{1 - (1 - \gamma)e^{-t}} \Big|_0^{\infty} + \frac{\gamma}{(1 - \gamma)\beta^r} \int_0^{\infty} \frac{rt^{r-1}}{1 - (1 - \gamma)e^{-t}} dt \\
&= \frac{-\gamma}{(1 - \gamma)\beta^r} \frac{t^r}{1 - (1 - \gamma)e^{-t}} \Big|_0^{\infty} \\
&\quad + \frac{r\gamma}{(1 - \gamma)\beta^r} \int_0^{\infty} t^{r-1} dt + \frac{r\gamma}{\beta^r} \int_0^{\infty} \frac{t^{r-1} e^{-t}}{1 - (1 - \gamma)e^{-t}} dt.
\end{aligned}$$

Άρα,

$$\begin{aligned}
E(X^r; \theta) &= \frac{-\gamma}{(1 - \gamma)\beta^r} t^r \left\{ \frac{1}{1 - (1 - \gamma)e^{-t}} - 1 \right\} \Big|_0^{\infty} + \frac{r\gamma}{\beta^r} \int_0^{\infty} \frac{t^{r-1} e^{-t}}{1 - (1 - \gamma)e^{-t}} dt \\
&= \frac{-\gamma}{\beta^r} \frac{t^r e^{-t}}{1 - (1 - \gamma)e^{-t}} \Big|_0^{\infty} + \frac{r\gamma}{\beta^r} \int_0^{\infty} \frac{t^{r-1} e^{-t}}{1 - (1 - \gamma)e^{-t}} dt.
\end{aligned}$$

Στη συνέχεια εφαρμόζοντας τον κανόνα L'Hospital, προκύπτει ότι:

$$\lim_{t \rightarrow \infty} \frac{t^r e^{-t}}{1 - (1 - \gamma)e^{-t}} = \lim_{t \rightarrow \infty} \frac{t^r}{e^t - (1 - \gamma)} \left(\text{μορφή } \frac{\infty}{\infty} \right) = \lim_{t \rightarrow \infty} \frac{rt^{r-1}}{e^t} = \dots = \lim_{t \rightarrow \infty} \frac{r!}{e^t} = 0.$$

Άρα,

$$E(X^r; \theta) = \frac{r! \gamma}{\Gamma(r) \beta^r} \int_0^{\infty} \frac{t^{r-1} e^{-t}}{1 - (1 - \gamma)e^{-t}} dt = \frac{r! \gamma}{\beta^r} \Phi(1 - \gamma, r, 1),$$

όπου

$\Phi(z, s, \alpha) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{x^{s-1} e^{-\alpha x}}{1 - ze^{-x}} dx$ με $\alpha > 0, z < 1$ και $s > 0$ είναι η συνάρτηση του Lerch (Lerch's transcendent function, Magnus et al., 1966, σελ. 32).

Για $r = 1$ παίρνουμε ότι η **μέση τιμή** δίνεται από τη σχέση:

$$\begin{aligned} E(X; \theta) &= \frac{\gamma}{\beta} \Phi(1-\gamma, 1, 1) = \frac{\gamma}{\beta} \int_0^{\infty} \frac{e^{-x}}{1 - (1-\gamma)e^{-x}} dx = \frac{\gamma}{\beta(1-\gamma)} \int_0^{\infty} d[\log\{1 - (1-\gamma)e^{-x}\}] \\ &= -\frac{\gamma \log \gamma}{\beta(1-\gamma)} = \frac{(p-1) \log(1-p)}{\beta p}. \end{aligned}$$

Η **διακύμανση** προκύπτει με εφαρμογή της σχέσης $Var(X) = EX^2 - (EX)^2$, απ'όπου εξάγεται ότι:

$$Var(X; \theta) = \frac{1-p}{p\beta^2} \left\{ 2L(p; 2) - \left(\frac{1-p}{p} \right) \log^2(1-p) \right\}.$$

Η **συνάρτηση επιβίωσης** δίνεται από τη σχέση $S(x; \theta) = 1 - F(x; \theta)$, απ'όπου προκύπτει:

$$\begin{aligned} S(x; \theta) &= \frac{1 - pe^{-\beta x}}{1 - pe^{-\beta x}} - \frac{1 - e^{-\beta x}}{1 - pe^{-\beta x}} \\ &= (1-p)e^{-\beta x} (1 - pe^{-\beta x})^{-1}. \end{aligned}$$

Η **συνάρτηση κινδύνου** δίνεται από τη σχέση $h(x; \theta) = \frac{f(x; \theta)}{S(x; \theta)}$, οπότε

έχουμε αντίστοιχα:

$$h(x; \theta) = \frac{\beta(1-p)e^{-\beta x}(1-pe^{-\beta x})^{-2}}{(1-p)e^{-\beta x}(1-pe^{-\beta x})^{-1}} = \beta(1-pe^{-\beta x})^{-1}.$$

Άρα,

$$\lim_{x \rightarrow 0} h(x; \theta) = \beta(1-p)^{-1},$$

και

$$\lim_{x \rightarrow \infty} h(x; \theta) = \beta,$$

και με τις δυο τιμές πεπερασμένες σε αντίθεση με τις αντίστοιχες της κατανομής Weibull, όπου $h(0) = \infty$ και $h(\infty) = 0$.

Η παράγωγος της συνάρτησης κινδύνου ως προς x είναι:

$$h'(x; \theta) = \frac{\beta^2(\gamma-1)e^{-\beta x}}{\{1-(1-\gamma)e^{-\beta x}\}^2}.$$

Επομένως, για $0 < \gamma < 1$ η συνάρτηση κινδύνου είναι γνησίως φθίνουσα, για $\gamma > 1$ είναι γνησίως αύξουσα και σταθερή για $\gamma = 1$.

Στη συνέχεια, η τιμή του **μέσου υπολειπόμενου χρόνου ζωής**, δίνεται από τη σχέση:

$$\begin{aligned} m(x_0; \theta) &= \frac{\int_{x_0}^{\infty} S(y; \theta) dy}{S(x_0; \theta)} \\ &= e^{\beta x_0} (1 - pe^{-\beta x_0}) \int_{x_0}^{\infty} \frac{e^{-\beta y}}{(1 - pe^{-\beta y})} dy. \end{aligned}$$

Άρα,

$$\begin{aligned}
 m(x_0; \theta) &= e^{\beta x_0} (1 - pe^{-\beta x_0}) \int_{x_0}^{\infty} d \left(\frac{\log(1 - pe^{-\beta y})}{\beta p} \right) \\
 &= e^{\beta x_0} (1 - pe^{-\beta x_0}) (\beta p)^{-1} \log(1 - pe^{-\beta y}) \Big|_{x_0}^{\infty} \\
 &= e^{\beta x_0} (1 - pe^{-\beta x_0}) (\beta p)^{-1} [\log(1) - \log(1 - pe^{-\beta x_0})] \\
 &= e^{\beta x_0} (1 - pe^{-\beta x_0}) (\beta p)^{-1} (0 - \log(1 - pe^{-\beta x_0})) \\
 &= -e^{\beta x_0} (1 - pe^{-\beta x_0}) (\beta p)^{-1} \log(1 - pe^{-\beta x_0}).
 \end{aligned}$$

Επομένως,

$$m(x_0; \theta) = -(\beta p)^{-1} e^{\beta x_0} (1 - pe^{-\beta x_0}) \log(1 - pe^{-\beta x_0}).$$

Η παραπάνω συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής είναι:

- γνησίως αύξουσα για $p > 0$,
- γνησίως φθίνουσα για $p < 0$,
- σταθερή για $p = 0$ (Εκθετική Κατανομή).

Η τιμή της $m(x_0; \theta)$ στο μηδέν είναι:

$$\lim_{x_0 \rightarrow 0} m(x_0; \theta) = -(\beta p)^{-1} (1 - p) \log(1 - p) = E(x; \theta),$$

δηλαδή ταυτίζεται με εκείνη του μέσου χρόνου ζωής, ενώ η οριακή τιμή της στο άπειρο (εφαρμόζοντας τον κανόνα του L'Hospital) θα είναι:

$$\begin{aligned}
\lim_{x_0 \rightarrow \infty} m(x_0; \theta) &= \lim_{x_0 \rightarrow \infty} \frac{1 - pe^{-\beta x_0}}{-\beta p} \lim_{x_0 \rightarrow \infty} e^{\beta x_0} \log\{1 - pe^{-\beta x_0}\} \\
&= \lim_{x_0 \rightarrow \infty} \frac{1 - pe^{-\beta x_0}}{-\beta p} \lim_{x_0 \rightarrow \infty} \frac{\log\{1 - pe^{-\beta x_0}\}}{e^{-\beta x_0}} \\
&= \lim_{x_0 \rightarrow \infty} \frac{1 - pe^{-\beta x_0}}{-\beta p} \lim_{x_0 \rightarrow \infty} \frac{-p}{1 - pe^{-\beta x_0}} \\
&= \frac{1}{\beta}.
\end{aligned}$$

Συνεπώς,

$$\begin{aligned}
\lim_{x_0 \rightarrow \infty} \frac{m(x_0; \theta)}{E(x_0; \theta)} &= \lim_{x_0 \rightarrow \infty} \frac{-(\beta p^{-1})e^{\beta x_0} (1 - pe^{-\beta x_0}) \log(1 - pe^{-\beta x_0})}{(\beta p)^{-1} (p-1) \log(1-p)} \\
&= - \lim_{x_0 \rightarrow \infty} \frac{(1 - pe^{-\beta x_0}) \log(1 - pe^{-\beta x_0})}{e^{-\beta x_0} (p-1) \log(1-p)},
\end{aligned}$$

απ'όπου με εφαρμογές του κανόνα L'Hospital προκύπτει:

$$\begin{aligned}
\lim_{x_0 \rightarrow \infty} m(x_0; \theta) \{E(x_0; \theta)\}^{-1} &= \lim_{x_0 \rightarrow \infty} \frac{p\beta e^{-\beta x_0} (\log(1 - pe^{-\beta x_0}) + 1)}{\beta e^{-\beta x_0} (p-1) \log(1-p)} \\
&= \lim_{x_0 \rightarrow \infty} \frac{p (\log(1 - pe^{-\beta x_0}) + 1)}{(p-1) \log(1-p)} \\
&= -p \{(1-p) \log(1-p)\}^{-1} > 1,
\end{aligned}$$

αφού $p < -\log(1-p) < p(1-p)^{-1}$, για $0 < p < 1$. Οπότε, η τιμή του μέσου υπολειπόμενου χρόνου ζωής υπερβαίνει τη μέση τιμή για κάθε x , δηλαδή ισχύει $m(x_0; \theta) > E(x; \theta)$. Για $p < 0$, η τιμή του μέσου υπολειπόμενου χρόνου ζωής δεν υπερβαίνει τη μέση τιμή για κάθε x και για $p = 0$ ταυτίζονται.

2.1.4 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

Θα κληθούμε να χρησιμοποιήσουμε τη Μέθοδο της Μέγιστης Πιθανοφάνειας (ΜΜΠ) για την εκτίμηση των παραμέτρων αυτών. Ας θεωρήσουμε ένα τυχαίο δείγμα μεγέθους n , $y_{obs} = (x_i; i = 1, \dots, n)$. Η συνάρτηση πιθανοφάνειας και ο λογάριθμός της (2.1) είναι αντίστοιχα:

$$L = L(\theta; y_{obs}) = \prod_{i=1}^n f_i(x_i; \theta) = \prod_{i=1}^n \beta(1-p)e^{-\beta x_i} (1 - pe^{-\beta x_i})^{-2},$$

και

$$\begin{aligned} l(\theta; y_{obs}) &= \sum_{i=1}^n \left(\log(\beta(1-p)) + \log e^{-\beta x_i} + \log(1 - pe^{-\beta x_i})^{-2} \right) \\ &= n \log(\beta(1-p)) - \beta \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \log(1 - pe^{-\beta x_i}). \end{aligned}$$

Οπότε, θέτοντας τις μερικές παραγώγους ίσες με το μηδέν $\frac{dl}{d\theta_i} = 0$,

$i = 1, 2$, προκύπτουν οι εξισώσεις μέγιστης πιθανοφάνειας:

$$\frac{\partial l(\theta; y_{obs})}{\partial \beta} = n\beta^{-1} + \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \left\{ x_i (1 - pe^{-\beta x_i})^{-1} \right\},$$

και

(2.2)

$$\frac{\partial l(\theta; y_{obs})}{\partial p} = p^{-1} \left\{ n(p-2)(1-p)^{-1} + 2 \sum_{i=1}^n (1 - pe^{-\beta x_i})^{-1} \right\}.$$

Για την επίλυση του συστήματος των παραπάνω μη γραμμικών ως προς τις παραμέτρους εξισώσεων, θα πρέπει να χρησιμοποιηθεί κάποια από τις μεθόδους αριθμητικής ανάλυσης, όπως για παράδειγμα, η μέθοδος Newton-Raphson ή η μέθοδος scoring.

Θεώρημα 2.2 (Adamidis and Loukas, 1998)

Ορίζουμε $g(\beta; p, y_{obs}) = n\beta^{-1} + \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \left\{ x_i (1 - pe^{-\beta x_i})^{-1} \right\}$, δηλαδή τη συνάρτηση του δεξιού μέλους της πρώτης έκφρασης της (2.2) και $\bar{x} = n^{-1} \sum_{i=1}^n x_i$. Τότε, για $p \in (0, 1)$, η $g(\beta; p, y_{obs}) = 0$ έχει μία και μοναδική ρίζα, η οποία βρίσκεται εντός του διαστήματος $(\bar{x}^{-1}, (1-p)\{\bar{x}(1+p)\}^{-1})$.

Απόδειξη (Adamidis and Loukas, 1998)

$$\text{Έστω } w(\beta; p, y_{obs}) = 2 \sum_{i=1}^n \left\{ x_i (1 - pe^{-\beta x_i})^{-1} \right\}.$$

Είναι προφανές ότι η w είναι γνησίως φθίνουσα ως προς β με

$$\lim_{\beta \rightarrow \infty} w(\beta; p, y_{obs}) = 2 \sum_{i=1}^n x_i.$$

Επίσης, προκύπτει ότι:

$$g(\beta; p, y_{obs}) < n\beta^{-1} + \sum_{i=1}^n x_i - \lim_{\beta \rightarrow \infty} w(\beta; p, y_{obs}) = n\beta^{-1} - \sum_{i=1}^n x_i < 0, \quad (2.3)$$

όταν $\beta > \bar{x}^{-1}$.

Επίσης,

$$\lim_{\beta \rightarrow 0} w(\beta; p, y_{obs}) = 2(1-p)^{-1} \sum_{i=1}^n x_i,$$

οπότε:

$$\begin{aligned} g(\beta; p, y_{obs}) &> n\beta^{-1} + \sum_{i=1}^n x_i - \lim_{\beta \rightarrow 0} w(\beta; p, y_{obs}) \\ &= n\beta^{-1} - (1+p)(1-p)^{-1} \sum_{i=1}^n x_i. \end{aligned} \quad (2.4)$$

Είναι $g(\beta; p, y_{obs}) > 0$, όταν $\beta < (1-p)\{\bar{x}(1+p)\}^{-1}$. Συνδυάζοντας τις (2.3) και (2.4) προκύπτει ότι υπάρχει μια τουλάχιστον ρίζα στο εν λόγω διάστημα. Για να αποδείξουμε τη μοναδικότητα της ρίζας θεωρούμε την πρώτη παράγωγο $g'(\beta; p, y_{obs})$, η οποία δίνεται από τη σχέση:

$$g'(\beta; p, y_{obs}) = -n\beta^{-2} + 2p \sum_{i=1}^n \left\{ x_i^2 e^{-\beta x_i} (1 - pe^{-\beta x_i})^{-2} \right\}.$$

Συμβολίζοντας με β^* τη λύση της $g'(\beta; p, y_{obs}) = 0$, προκύπτει ότι:

$$n\beta^{*-1} = 2p\beta^* \sum_{i=1}^n \left\{ x_i^2 e^{-\beta^* x_i} (1 - pe^{-\beta^* x_i})^{-2} \right\},$$

και

$$g(\beta^*; p, y_{obs}) = \sum_{i=1}^n \frac{x_i \left(p^2 e^{-2\beta^* x_i} + 2p\beta^* x_i e^{-\beta^* x_i} - 1 \right)}{\left(1 - pe^{-\beta^* x_i} \right)^2}.$$

Παρατηρούμε ότι για την έκφραση μέσα στην παρένθεση ισχύει:

$$\begin{aligned}
 p^2 e^{-2\beta^* x_i} + 2p\beta^* x_i e^{-\beta^* x_i} - 1 &= 2e^{-\beta^* x_i} \left(p^2 \frac{e^{-\beta^* x_i}}{2} + p\beta^* x_i - \frac{1}{2e^{-\beta^* x_i}} \right) \\
 &= 2e^{-\beta^* x_i} \left(p^2 \frac{1}{2e^{\beta^* x_i}} + p\beta^* x_i - \frac{e^{\beta^* x_i}}{2} \right) \\
 &\leq 2e^{-\beta^* x_i} \left(\frac{1}{2e^{\beta^* x_i}} + \beta^* x_i - \frac{e^{\beta^* x_i}}{2} \right) \\
 &= 2e^{-\beta^* x_i} \left(\beta^* x_i - \sinh(\beta^* x_i) \right) < 0,
 \end{aligned}$$

για κάθε β^* , $x_i > 0$, $i = 1, \dots, n$ με $\sinh(\beta^* x_i) = \frac{1 - e^{-2\beta^* x_i}}{2e^{\beta^* x_i}}$.

Αυτό αποδεικνύει ότι η g είναι αρνητική στα στάσιμα σημεία της. Επιπλέον, $\lim_{\beta \rightarrow \infty} g(\beta; p, \mathcal{Y}_{obs}) = -\sum_{i=1}^n x_i$, γεγονός που αποδεικνύει ότι η g δε μπορεί να έχει παραπάνω από μία ρίζα.

2.1.5 Η μέθοδος EM

Για την εφαρμογή του αλγορίθμου θεωρούμε ότι το δείγμα είναι “πλήρες”, δηλαδή αποτελείται από τις παρατηρούμενες τιμές της τυχαίας μεταβλητής X και τις τιμές μιας μη παρατηρούμενης τυχαίας μεταβλητής Z με από κοινού κατανομή:

$$f(x, z; \theta) = \beta(1-p)z p^{z-1} e^{-\beta z x},$$

για $x, \beta \in R^+$, $z \in N$, $p \in (0, 1)$ και $\theta = (\beta, p)$.

Έστω $Y_{obs} = (x_i, z_i; i = 1, \dots, n)$ τα υποθετικά δεδομένα. Στο E-βήμα υπολογίζουμε την υπο συνθήκη αναμενόμενη τιμή, δοθέντων των παρατηρούμενων τιμών $X_i, i = 1, \dots, n$, δηλαδή την αναμενόμενη τιμή της $(Z | X; \theta^{(t)})$, όπου $\theta^{(t)} = (\beta^{(t)}, p^{(t)})$ είναι η τιμή της παραμέτρου θ . Αποδεικνύεται ότι η δεσμευμένη κατανομή της τυχαίας μεταβλητής $Z | X$ έχει πυκνότητα:

$$p(z | x; \theta) = z \left(p e^{-\beta x} \right)^{z-1} \left(1 - p e^{-\beta x} \right)^2, \quad z \in N,$$

όπου $x, \beta, p \in R^+$. Άρα, η αναμενόμενη τιμή της είναι:

$$E(Z | X = x; \theta) = 2 \left(1 - p e^{-\beta x} \right)^{-1} - 1.$$

Επομένως, αντιμετωπίζοντας τις δεσμευμένες αναμενόμενες τιμές ως δεδομένα, μια επανάληψη της μεθόδου EM ορίζεται από τις σχέσεις (M-βήμα μεγιστοποίησης):

$$\beta^{(t+1)} = n \left[\sum_{i=1}^n \frac{x_i \left(1 + p^{(t)} e^{-\beta^{(t)} x_i} \right)}{1 - p^{(t)} e^{-\beta^{(t)} x_i}} \right]^{-1},$$

και

$$p^{(t+1)} = 1 - n \left[\sum_{i=1}^n \frac{1 + p^{(t)} e^{-\beta^{(t)} x_i}}{1 - p^{(t)} e^{-\beta^{(t)} x_i}} \right]^{-1}.$$

2.1.6 Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας

Η κατανομή των εκτιμητών μέγιστης πιθανοφάνειας, $\hat{\theta}$, σύμφωνα με τις ασυμπτωτικές ιδιότητες των εκτιμητών αυτών, είναι κατά προσέγγιση

η διδιάστατη κανονική με μέση τιμή θ και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα αναμενόμενης πληροφορίας $J(\theta) = E(I; \theta)$, όπου $I = I(\theta; y_{obs})$ είναι ο συμμετρικός, δεύτερης τάξης πίνακας παρατηρούμενης πληροφορίας, ενώ η αναμενόμενη τιμή λαμβάνεται ως προς την κατανομή της τυχαίας μεταβλητής X . Τα στοιχεία του πίνακα

$$I(\theta; y_{obs}), I_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}, i, j = 1, 2,$$

δίνονται αναλυτικά από τις σχέσεις:

$$I_{11} = -g'(\beta; p, Y_{obs}),$$

όπου

$$g'(\beta; p, y_{obs}) = -n\beta^{-2} + 2p \sum_{i=1}^n \left\{ x_i^2 e^{-\beta x_i} (1 - pe^{-\beta x_i})^{-2} \right\},$$

$$I_{12} = 2 \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - pe^{-\beta x_i})^2},$$

και

$$I_{22} = \frac{n}{(1-p)^2} - 2 \sum_{i=1}^n \frac{e^{-2\beta x_i}}{(1 - pe^{-\beta x_i})^2}.$$

Επιπλέον, τα στοιχεία του συμμετρικού δεύτερης τάξης πίνακα αναμενόμενης πληροφορίας $J(\theta)$ είναι:

$$J_{11} = \frac{n}{3\beta^2 p} \left[3p - 2 \{ p - (1-p)L(p; 2) \} \right],$$

όπου

$$L(p; 2) = \sum_{j=1}^{\infty} p^j j^{-2},$$

$$J_{12} = \frac{n \{4p^2 - p + (1-p)^2 \log(1-p)\}}{3\beta p^2 (1-p)},$$

και

$$J_{22} = n(1-p)^{-2} / 3.$$

2.2 Η Εκθετική - Poisson κατανομή

Στην ενότητα αυτή παρουσιάζεται η Εκθετική-Poisson κατανομή, η οποία μελετήθηκε αναλυτικά από τον Kus (2007).

2.2.1 Ορισμοί και φυσική ερμηνεία

Έστω $\{Y_i\}_{i=1}^Z$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση πυκνότητας πιθανότητας

$$f(y; \beta) = \beta e^{-\beta y},$$

με $y, \beta \in R^+$ και Z είναι μια Poisson αποκομμένη στο 0, με Z ανεξάρτητη των Y_i , με συνάρτηση πιθανότητας

$$P(z; \lambda) = e^{-\lambda} \lambda^z \Gamma^{-1}(z+1) (1 - e^{-\lambda})^{-1}, z \in N, \lambda \in R^+,$$

όπου $\Gamma(\cdot)$ είναι η Γάμμα συνάρτηση.

Ορίζοντας $X = \min\left(\{Y_i\}_{i=1}^Z\right)$, προκύπτει $f(x|z; \beta) = \beta z e^{-\beta z x}$ και επομένως, η συνάρτηση πυκνότητας πιθανότητας του X , δίνεται από τη σχέση

$$f(x; \theta) = \frac{\lambda \beta}{(1 - e^{-\lambda})} e^{-\lambda - \beta x + \lambda \exp(-\beta x)}, \quad (2.5)$$

για $x, \beta, \lambda \in R^+$, όπου $\theta = (\lambda, \beta)$.

Παρατηρήσεις

- Η τελευταία θα αναφέρεται ως Εκθετική-Poisson (EP) κατανομή.
- $\lim_{x \rightarrow 0} f(x; \theta) = \lim_{x \rightarrow 0} \frac{\lambda \beta}{(1 - e^{-\lambda})} e^{-\lambda - \beta x + \lambda \exp(-\beta x)} = \lambda \beta (1 - e^{-\lambda})^{-1}$.
- Για όλες τις τιμές των παραμέτρων, η EP είναι φθίνουσα ως προς x και τείνει στο 0 όταν $x \rightarrow \infty$, δηλαδή

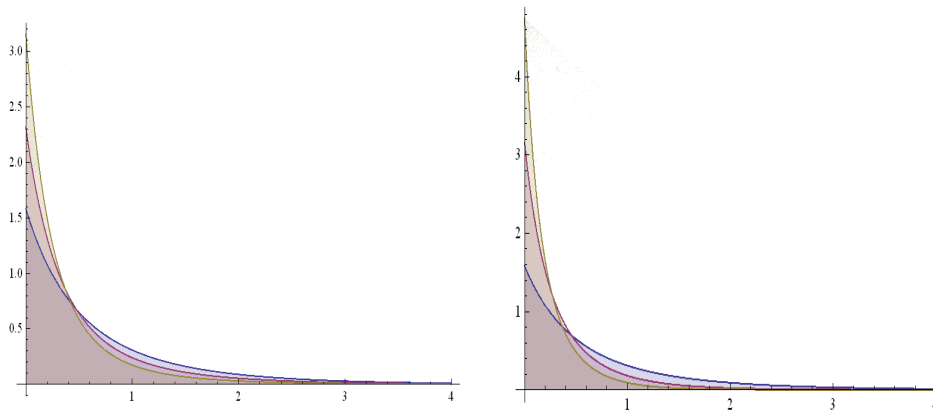
$$\lim_{x \rightarrow \infty} f(x; \theta) = \lim_{x \rightarrow \infty} \frac{\lambda \beta}{(1 - e^{-\lambda})} e^{-\lambda - \beta x + \lambda \exp(-\beta x)} = 0.$$

Η EP ακολουθεί εκθετική κατανομή με παράμετρο β , όταν $\lambda \rightarrow 0$ και συγκεκριμένα με εφαρμογή του κανόνα L'Hospital παίρνει τη μορφή

$$f(x; \theta) = \beta e^{-\beta x}.$$

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X ορίζει την κατανομή που καλείται Εκθετική-Poisson.

Στο σχήμα που ακολουθεί δίνεται η γραφική παράσταση αυτής της σ.π.π. για ειδικές τιμές των παραμέτρων και για $x \in (0, 4)$.



Εικόνα 2. Η συνάρτηση πυκνότητας πιθανότητας της EP α) για $\beta=1$ και $\lambda=1, 2, 3$ και β) για $\lambda=1$ και $\beta=1, 2, 3$.

2.2.2 Χαρακτηριστικά της κατανομής

Δίνονται κατά σειρά η *αθροιστική συνάρτηση κατανομής*, η *διάμεσος*, οι *ροπές*, η *μέση τιμή*, η *διακύμανση*, η *συνάρτηση επιβίωσης*, η *συνάρτηση κινδύνου* και ο *μέσος υπολειπόμενος χρόνος*.

Η *αθροιστική συνάρτηση* κατανομής της τ.μ. X που ακολουθεί την Εκθετική-Poisson κατανομή με σ.π.π. αυτής της σχέσης (2.5), δίνεται από τη σχέση:

$$F(x; \theta) = (e^{\lambda \exp(-\beta x)} - e^{-\lambda}) (1 - e^{-\lambda})^{-1}.$$

Έχουμε:

$$\begin{aligned} F(x; \theta) &= \int_0^x \frac{\lambda \beta}{(1 - e^{-\lambda})} e^{-\lambda - \beta x + \lambda \exp(-\beta x)} dx \\ &= \frac{\lambda \beta}{(1 - e^{-\lambda})} e^{-\lambda} \int_0^x e^{-\beta x + \lambda \exp(-\beta x)} dx. \end{aligned}$$

Θέτω $y = e^{-\beta x}$, $dy = -\beta e^{-\beta x} dx$,

$$\begin{aligned} F(x; \theta) &= -\frac{\lambda e^{-\lambda}}{(1 - e^{-\lambda})} \int_1^y e^{\lambda y} dy \\ &= (-e^{-\lambda + \lambda y} + 1)(1 - e^{-\lambda})^{-1} \\ &= -(e^{\lambda} - e^{\lambda y})(1 - e^{\lambda})^{-1} \\ &= (e^{\lambda \exp(-\beta x)} - e^{\lambda})(1 - e^{\lambda})^{-1}. \end{aligned}$$

Η *διάμεσος* είναι:

$$\log \left\{ \log \left[2^{-1} (e^{\lambda} + 1) \right]^{-1} \lambda \right\} \beta^{-1}.$$

Οι *r-ροπές* δίνονται από τη σχέση:

$$E(X^r; \theta) = \frac{\lambda \Gamma(r+1)}{(e^{\lambda} - 1) \beta^r} F_{r+1, r+1}([1, 1, \dots, 1], [2, 2, \dots, 2], \lambda), \quad r \in N,$$

όπου $F_{p,q}(n, d, \lambda)$ είναι η γενικευμένη υπεργεωμετρική κατανομή και ορίζεται ως εξής:

$$F_{p,q}(n, d, \lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k \prod_{i=1}^p \Gamma(n_i + k) \Gamma^{-1}(n_i)}{\Gamma(k+1) \prod_{i=1}^q \Gamma(d_i + k) \Gamma^{-1}(d_i)},$$

όπου $n = [n_1, n_2, \dots, n_p]$, p είναι ο αριθμός των τελεστών του n και $d = [d_1, d_2, \dots, d_q]$, q είναι ο αριθμός των τελεστών του d .

Η μέση τιμή και η διακύμανση είναι αντίστοιχα:

$$E(X; \theta) = \frac{\lambda}{(e^\lambda - 1)\beta} F_{2,2}([1,1], [2,2], \lambda),$$

και

$$\text{Var}(X; \theta) = \frac{\lambda}{(e^\lambda - 1)\beta^2} \left\{ 2F_{3,3}([1,1,1], [2,2,2], \lambda) - \frac{\lambda}{e^\lambda - 1} F_{2,2}([1,1], [2,2], \lambda) \right\}.$$

Οι αντίστοιχες συναρτήσεις επιβίωσης και κινδύνου, είναι:

$$\begin{aligned} S(x; \theta) &= 1 - (e^{\lambda \exp(-\beta x)} - e^\lambda)(1 - e^\lambda)^{-1} \\ &= 1 - \left(\frac{e^{\lambda \exp(-\beta x)}}{1 - e^\lambda} - \frac{e^\lambda}{1 - e^\lambda} \right) \\ &= \frac{1 - e^\lambda - e^{\lambda \exp(-\beta x)} + e^\lambda}{1 - e^\lambda} \\ &= (1 - e^{\lambda \exp(-\beta x)})(1 - e^\lambda)^{-1}, x \geq 0, \end{aligned}$$

και

$$h(x; \theta) = \frac{f(x; \theta)}{S(x; \theta)} = \frac{\beta \lambda e^{-\lambda - \beta x + \lambda \exp(-\beta x)} (1 - e^\lambda)}{(1 - e^\lambda)(1 - e^{\lambda \exp(-\beta x)})}, x \geq 0.$$

Η συνάρτηση κινδύνου είναι φθίνουσα με:

$$\lim_{x \rightarrow 0} h(x; \theta) = \lim_{x \rightarrow 0} \frac{\beta \lambda e^{-\lambda - \beta x + \lambda \exp(-\beta x)} (1 - e^\lambda)}{(1 - e^{-\lambda})(1 - e^{\lambda \exp(-\beta x)})} = \lambda \beta e^\lambda (e^\lambda - 1)^{-1},$$

και

$$\lim_{x \rightarrow \infty} h(x; \theta) = \lim_{x \rightarrow \infty} \frac{\beta \lambda e^{-\lambda - \beta x + \lambda \exp(-\beta x)} (1 - e^\lambda)}{(1 - e^{-\lambda})(1 - e^{\lambda \exp(-\beta x)})} = \beta,$$

και με τις δυο τιμές πεπερασμένες σε αντίθεση με τις αντίστοιχες της κατανομής Weibull, όπου $h(0) = \infty$ και $h(\infty) = 0$.

Η συνάρτηση του *μέσου υπολειπόμενου χρόνου ζωής* είναι:

$$\begin{aligned} m(x_0; \theta) &= E(X - x_0 | X \geq x_0; \theta) \\ &= \frac{\lambda e^{-\beta x_0}}{\beta (e^{\lambda \exp(-\beta x_0)} - 1)} F_{2,2}([1, 1], [2, 2], \lambda e^{-\beta x_0}), \end{aligned}$$

όπου $F_{p,q}(n, d, \lambda)$ είναι η γενικευμένη υπεργεωμετρική κατανομή.

Επιπλέον,

$$\begin{aligned} \lim_{x_0 \rightarrow 0} m(x_0; \theta) &= \lim_{x_0 \rightarrow 0} \frac{\lambda e^{-\beta x_0}}{\beta (e^{\lambda \exp(-\beta x_0)} - 1)} F_{2,2}([1, 1], [2, 2], \lambda e^{-\beta x_0}) \\ &= \frac{\lambda}{\beta (e^\lambda - 1)} F_{2,2}([1, 1], [2, 2], \lambda) = E(X; \theta), \end{aligned}$$

και

$$\lim_{x_0 \rightarrow \infty} m(x_0; \theta) = \beta^{-1}.$$

2.2.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

Στην ενότητα αυτή θα παρουσιαστούν τα αποτελέσματα του Kus (2007) για την εκτίμηση των παραμέτρων με τη μέθοδο της μέγιστης πιθανοφάνειας.

Ο λογάριθμος της συνάρτησης πιθανοφάνειας που βασίζεται σε τυχαίο δείγμα μεγέθους n , $y_{obs} = (x_i; i = 1, \dots, n)$ που ακολουθεί κατανομή EP είναι της μορφής:

$$\ell(\theta; y_{obs}) = n \log(\lambda\beta) - \beta \sum_{i=1}^n x_i - n\lambda + \lambda \sum_{i=1}^n e^{-\beta x_i} - n \log(1 - e^{-\lambda}),$$

απ' όπου προκύπτουν οι εξισώσεις μέγιστης πιθανοφάνειας:

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \lambda} = \sum_{i=1}^n e^{-\beta x_i} + n \left[\lambda^{-1} - (1 - e^{-\lambda})^{-1} \right], \quad (2.6)$$

και

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \beta} = n\beta^{-1} - \sum_{i=1}^n x_i - \lambda \sum_{i=1}^n x_i e^{-\beta x_i}. \quad (2.7)$$

Παρατηρούμε ότι για τη δεύτερη εκ των δύο εξισώσεων,

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \beta} = 0,$$

μπορεί να υπάρξει αναλυτική λύση ως προς λ , με τιμή:

$$\hat{\lambda} = (\hat{\beta}^{-1} - \bar{X}) \left\{ \frac{1}{n} \sum_{i=1}^n X_i \exp(-\hat{\beta} X_i) \right\}^{-1},$$

όπου $\hat{\lambda}$ και $\hat{\beta}$ είναι οι ΕΜΠ.

Θεώρημα 2.3 (Kus, 2007)

Ορίζουμε $h(\lambda; \beta, y_{obs}) = \sum_{i=1}^n e^{-\beta x_i} + n \left[\lambda^{-1} - (1 - e^{-\lambda})^{-1} \right]$, δηλαδή τη συνάρτηση στο δεξιό μέλος της (2.6), όπου β είναι η πραγματική τιμή της παραμέτρου. Τότε, η ρίζα $\hat{\lambda}$ της $h(\lambda; \beta, y_{obs}) = 0$ είναι μοναδική για $n/2 < \sum_{i=1}^n e^{-\beta x_i}$.

Απόδειξη (Kus, 2007)

Η $h(\lambda; \beta, y_{obs})$ είναι αυστηρά φθίνουσα.

Εάν

$$\begin{aligned} \lim_{\lambda \rightarrow 0} h(\lambda; \beta, Y_{obs}) &= \sum_{i=1}^n e^{-\beta x_i} + \lim_{\lambda \rightarrow 0} n \left[\lambda^{-1} - (1 - e^{-\lambda})^{-1} \right] \\ &= \sum_{i=1}^n e^{-\beta x_i} + n \lim_{\lambda \rightarrow 0} \frac{1 - e^{-\lambda} - \lambda}{\lambda(1 - e^{-\lambda})}, \end{aligned}$$

εφαρμόζω τον κανόνα L'Hospital και έχω:

$$\lim_{\lambda \rightarrow 0} h(\lambda; \beta, Y_{obs}) = \sum_{i=1}^n e^{-\beta x_i} - \frac{n}{2}.$$

Επομένως, $\lim_{\lambda \rightarrow 0} h(\lambda; \beta, Y_{obs}) > 0$, όταν $\sum_{i=1}^n e^{-\beta x_i} > \frac{n}{2}$.

Εάν $\sum_{i=1}^n e^{-\beta x_i} > \frac{n}{2}$, τότε καθώς

$$\lim_{\lambda \rightarrow \infty} h(\lambda; \beta, Y_{obs}) = \sum_{i=1}^n e^{-\beta x_i} - n,$$

προκύπτει ότι $\lim_{\lambda \rightarrow \infty} h(\lambda; \beta, Y_{obs}) < 0$, όταν $\sum_{i=1}^n e^{-\beta x_i} < n$.

Έτσι, υπάρχει τουλάχιστον μια ρίζα της $h(\lambda; \beta, Y_{obs}) = 0$.

Για να αποδείξουμε ότι η ρίζα είναι μοναδική θα υπολογίσουμε την:

$$h'(\lambda; \beta, Y_{obs}) = -n \frac{\exp(-2\lambda) - 2 \exp(-\lambda) - \lambda^2 \exp(-\lambda) + 1}{\lambda^2 (1 - \exp(-\lambda))^2}.$$

Θέτω

$$\begin{aligned} z(\lambda) &= \exp(-2\lambda) - 2 \exp(-\lambda) - \lambda^2 \exp(-\lambda) + 1 \\ &= 4 \exp(-\lambda) \left[\sin h \left(\frac{\lambda}{2} \right) + \frac{\lambda}{2} \right] \left[\sin h \left(\frac{\lambda}{2} \right) - \frac{\lambda}{2} \right]. \end{aligned}$$

Επειδή $\sin h \left(\frac{\lambda}{2} \right) - \frac{\lambda}{2} > 0$, τότε $z(\lambda) > 0$. Επομένως, $h'(\lambda; \beta, Y_{obs}) < 0$

και η ρίζα είναι μοναδική.

Θεώρημα 2.4 (Kus, 2007)

Ορίζουμε $g(\beta; \lambda, y_{obs}) = n\beta^{-1} - \sum_{i=1}^n x_i - \lambda \sum_{i=1}^n x_i e^{-\beta x_i}$, δηλαδή τη

συνάρτηση στο δεξιό μέλος της (2.7) και $\bar{x} = n^{-1} \sum_{i=1}^n x_i$, όπου λ είναι η

πραγματική τιμή της παραμέτρου. Τότε, δεδομένου ότι $\lambda \in (0, e^2)$, η ρίζα $\hat{\beta}$ της $g(\beta; \lambda, y_{obs}) = 0$ βρίσκεται στο εσωτερικό του $([\bar{x}(\lambda+1)]^{-1}, \bar{x}^{-1})$ και είναι μοναδική.

2.2.4 Η μέθοδος EM

Για την εφαρμογή του αλγορίθμου θεωρούμε ότι το δείγμα είναι “πλήρες”, δηλαδή αποτελείται από τις παρατηρούμενες τιμές της τυχαίας μεταβλητής X και τις τιμές μιας μη παρατηρούμενης τυχαίας μεταβλητής Z με από κοινού κατανομή:

$$f(x, z; \theta) = \beta z e^{-\beta z x} e^{-\lambda} \lambda^z \Gamma^{-1}(z+1) (1 - e^{-\lambda})^{-1},$$

με $x > 0$, $z = 1, 2, \dots$ και $\beta, \lambda > 0$.

Έστω $Y_{obs} = (x_i, z_i; i = 1, \dots, n)$ τα υποθετικά δεδομένα. Στο E-βήμα υπολογίζουμε την υπό συνθήκη αναμενόμενη τιμή, δοθέντων των παρατηρούμενων τιμών X_i , $i = 1, \dots, n$, δηλαδή η αναμενόμενη τιμή της $(Z | X; \theta^{(t)})$, όπου $\theta^{(t)} = (\beta^{(t)}, \lambda^{(t)})$ είναι η τιμή της παραμέτρου θ . Αποδεικνύεται ότι η δεσμευμένη κατανομή της τυχαίας μεταβλητής $Z | X$ έχει πυκνότητα:

$$P(z / x; \theta) = e^{-\beta z x - \lambda \exp(-\beta x) + \beta x} \lambda^{z-1} \Gamma^{-1}(z), \quad z \in N,$$

όπου $x, \beta, \lambda \in R^+$. Άρα, η αναμενόμενη τιμή της είναι:

$$E(Z | X; \theta) = 1 + \lambda e^{-\beta x}.$$

Επομένως, αντιμετωπίζοντας τις δεσμευμένες αναμενόμενες τιμές ως δεδομένα, μια επανάληψη της μεθόδου EM ορίζεται από τις σχέσεις (Μ-βήμα μεγιστοποίησης):

$$\beta^{(h+1)} = n \left[\sum_{i=1}^n x_i \left(1 + \lambda^{(h)} e^{-\beta^{(h)} x_i} \right) \right]^{-1},$$

και

$$\lambda^{(h+1)} = n^{-1} \left(1 - e^{-\lambda^{(h+1)}} \right) \sum_{i=1}^n \left(1 + \lambda^{(h)} e^{-\beta^{(h)} x_i} \right).$$

2.2.5 Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας

Η κατανομή των εκτιμητών μέγιστης πιθανοφάνειας, $\hat{\theta}$, σύμφωνα με τις ασυμπτωτικές ιδιότητες των εκτιμητών αυτών, είναι κατά προσέγγιση η διδιάστατη κανονική με μέση τιμή θ και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα αναμενόμενης πληροφορίας $J(\theta) = E(I; \theta)$, όπου $I = I(\theta; y_{obs})$ είναι ο συμμετρικός, δεύτερης τάξης πίνακας παρατηρούμενης πληροφορίας, ενώ η αναμενόμενη τιμή λαμβάνεται ως προς την κατανομή της τυχαίας μεταβλητής X . Τα στοιχεία του πίνακα

$$I(\theta; y_{obs}), I_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, 2,$$

δίνονται αναλυτικά από τις σχέσεις:

$$I_{11} = n \left(1 + e^{2\lambda} - \lambda^2 e^\lambda - 2e^\lambda \right) \left[\lambda \left(1 - e^{-\lambda} \right) \right]^{-2},$$

$$I_{12} = \sum_{i=1}^n x_i e^{-\beta x_i},$$

και

$$I_{22} = n\beta^{-2} - \lambda \sum_{i=1}^n x_i^2 e^{-\beta x_i}.$$

Επιπλέον, τα στοιχεία του συμμετρικού δεύτερης τάξης πίνακα *αναμενόμενης πληροφορίας* $J(\theta)$ είναι:

$$J_{11} = n(1 + e^{2\lambda} - \lambda^2 e^\lambda - 2e^\lambda) [\lambda(1 - e^{-\lambda})]^{-2},$$

$$J_{12} = \frac{n\lambda e^{-\lambda}}{4\beta(1 - e^{-\lambda})} F_{2,2}([2, 2], [3, 3], \lambda),$$

και

$$J_{22} = n\beta^{-2} - \frac{n\lambda^2 e^{-\lambda}}{4\beta^2(1 - e^{-\lambda})} F_{3,3}([2, 2, 2], [3, 3, 3], \lambda).$$

2.3 Η Εκθετική - Λογαριθμική κατανομή (EL)

Η Εκθετική-Λογαριθμική κατανομή (EL) μελετάται αναλυτικά από τους Tahmasbi and Rezaei (2008).

2.3.1 Ορισμοί και φυσική ερμηνεία

Έστω $\{Y_i\}_{i=1}^Z$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση πυκνότητας πιθανότητας

$$f_Y(y; \beta) = \beta e^{-\beta y}, \quad y \geq 0,$$

όπου $\beta \in R^+$ και Z τυχαία μεταβλητή με συνάρτηση πιθανότητας

$$P_Z(z; p) = \frac{(1-p)^z}{-z \ln p}, \quad z \in N, \text{ με } p \in (0,1).$$

Ορίζοντας με $X = \min\left(\{Y_i\}_{i=1}^Z\right)$ και υποθέτοντας την ανεξαρτησία των Y_1, \dots, Y_Z, Z έχουμε

$$f_{X|Z}(x|z; \beta) = \beta z e^{-\beta x},$$

και η συνάρτηση πυκνότητας πιθανότητας της X , δίνεται από τη σχέση

$$f_x(x; p, \beta) = \frac{1}{-\ln p} \frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}}, \quad x > 0, \quad (2.8)$$

με $0 < p < 1$ και $\beta > 0$.

2.3.2 Χαρακτηριστικά της κατανομής

Δίνονται κατά σειρά η *αθροιστική συνάρτηση κατανομής*, η *διάμεσος*, η *συνάρτηση επιβίωσης*, η *συνάρτηση κινδύνου*, ο *μέσος υπολειπόμενος χρόνος*, οι *ροπές*, η *μέση τιμή* και η *διακύμανση*.

Η *αθροιστική συνάρτηση* κατανομής της τ.μ. X που ακολουθεί την Εκθετική-Λογαριθμική κατανομή με σ.π.π. αυτής της σχέσης (2.8), δίνεται από τη σχέση:

$$F_X(x; p, \beta) = 1 - \frac{\ln(1-(1-p)e^{-\beta x})}{\ln p}.$$

Έχουμε:

$$\begin{aligned}
 F_X(x; p, \beta) &= \int_0^x f(x; p, \beta) dx \\
 &= \int_0^x \frac{1}{-\ln p} \frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}} dx \\
 &= \frac{\beta(1-p)}{-\ln p} \int_0^x \frac{e^{-\beta x}}{1-(1-p)e^{-\beta x}} dx.
 \end{aligned}$$

Θέτοντας $y = (1-p)e^{-\beta x}$, θα είναι:

$$dy = -\beta(1-p)e^{-\beta x} dx,$$

και επομένως,

$$\begin{aligned}
 F_X(x; p, \beta) &= \frac{1}{\ln p} \int_{1-p}^y \frac{1}{1-y} dy \\
 &= \frac{1}{\ln p} [-\ln(1-y) + \ln p] \\
 &= 1 - \frac{\ln(1-(1-p)e^{-\beta x})}{\ln p}.
 \end{aligned}$$

Η *διάμεσος* είναι:

$$x = \beta^{-1} \ln(1 + \sqrt{p}).$$

Έχουμε:

$$F_X(x; p, \beta) = \frac{1}{2} \Leftrightarrow 1 - \frac{\ln(1 - (1-p)e^{-\beta x})}{\ln p} = \frac{1}{2}$$

$$\Leftrightarrow \ln p = \ln(1 - (1-p)e^{-\beta x})^2$$

$$\Leftrightarrow (1-p)e^{-\beta x} = 1 - \sqrt{p}$$

$$\Leftrightarrow -\beta x = \ln\left(\frac{1 - \sqrt{p}}{1-p}\right)$$

$$\Leftrightarrow x = \beta^{-1} \ln(1 + \sqrt{p}).$$

Οι συναρτήσεις *επιβίωσης* και *κινδύνου* δίνονται αντίστοιχα από τις σχέσεις:

$$S(x; \theta) = 1 - F(x; \theta) = \frac{\ln(1 - (1-p)e^{-\beta x})}{\ln p}, \quad x \geq 0,$$

και

$$h(x; \theta) = \frac{f(x; \theta)}{S(x; \theta)} = \frac{-\beta(1-p)e^{-\beta x}}{(1 - (1-p)e^{-\beta x}) \ln(1 - (1-p)e^{-\beta x})}, \quad x \geq 0.$$

Η συνάρτηση κινδύνου είναι φθίνουσα με:

$$\lim_{x \rightarrow 0} h(x; \theta) = \lim_{x \rightarrow 0} \frac{-\beta(1-p)e^{-\beta x}}{(1 - (1-p)e^{-\beta x}) \ln(1 - (1-p)e^{-\beta x})} = \frac{-\beta(1-p)}{p \ln p},$$

και

$$\begin{aligned}\lim_{x \rightarrow \infty} h(x; \theta) &= \lim_{x \rightarrow \infty} \frac{-\beta(1-p)e^{-\beta x}}{(1-(1-p)e^{-\beta x}) \ln(1-(1-p)e^{-\beta x})} \\ &= -\beta(1-p) \lim_{x \rightarrow \infty} \frac{e^{-\beta x}}{(1-(1-p)e^{-\beta x}) \ln(1-(1-p)e^{-\beta x})}.\end{aligned}$$

Είναι απροσδιόριστη μορφή $\left(\frac{0}{0}\right)$, επομένως θα εφαρμόσω τον κανόνα του L'Hospital και έχουμε:

$$\lim_{x \rightarrow \infty} h(x; \theta) = \lim_{x \rightarrow \infty} \frac{\beta^2(1-p)e^{-\beta x}}{\beta(1-p)e^{-\beta x} [\ln(1-(1-p)e^{-\beta x}) + 1]} = \beta,$$

και με τις δυο τιμές πεπερασμένες σε αντίθεση με τις αντίστοιχες της κατανομής Weibull, όπου $h(0) = \infty$ και $h(\infty) = 0$.

Η συνάρτηση του **μέσου υπολειπόμενου χρόνου ζωής** είναι:

$$m(x_0; p, \beta) = E(X - x_0 | X \geq x_0; \beta, p) = -\frac{di \log(1 - (1-p)e^{-\beta x_0})}{\beta \ln(1 - (1-p)e^{-\beta x_0})},$$

όπου $di \log(\cdot)$ είναι η διλογαριθμική κατανομή και ορίζεται ως:

$$di \log(a) = \int_1^a \frac{\ln(x)}{1-x} dx.$$

Επιπλέον,

$$\lim_{x_0 \rightarrow 0} m(x_0; \beta, p) = E(X; \beta, p),$$

και

$$\lim_{x_0 \rightarrow \infty} m(x_0; \beta, p) = \beta^{-1}.$$

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

$$E(X^r; p, \beta) = -\frac{r! \text{poly log}(r+1, 1-p)}{\beta^r \ln p}, \quad r \in N,$$

όπου $\text{poly log}(\cdot)$ είναι η πολυλογαριθμική συνάρτηση που ορίζεται ως εξής:

$$\text{poly log}(a, z) = \sum_{k=1}^{\infty} \frac{z^k}{k^a}.$$

Επομένως, η μέση τιμή και η διακύμανση, δίνονται από τις σχέσεις:

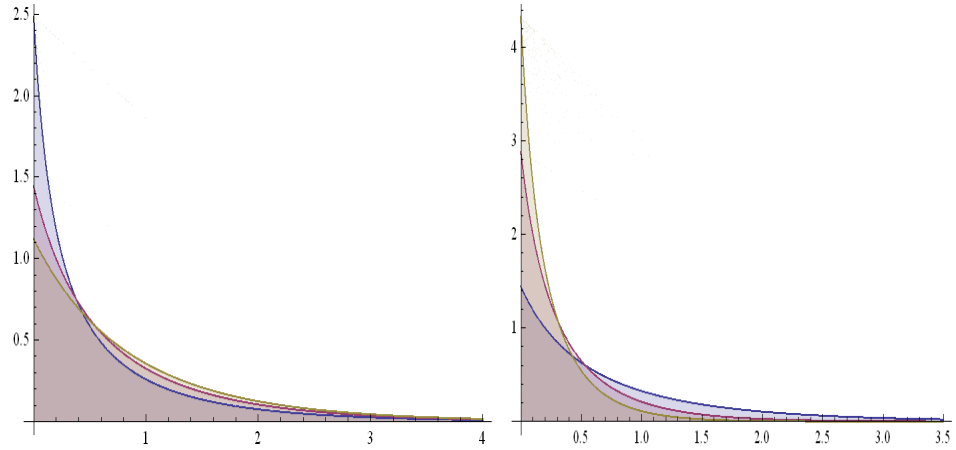
$$E(X; \theta) = -\frac{\text{poly log}(2, 1-p)}{\beta \ln p},$$

και

$$\text{Var}(X; \theta) = -\frac{2 \text{poly log}(3, 1-p)}{\beta^2 \ln p} - \frac{\text{poly log}^2(2, 1-p)}{\beta^2 \ln^2 p}.$$

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X ορίζει την κατανομή που καλείται Εκθετική-Λογαριθμική.

Στο σχήμα που ακολουθεί δίνεται η γραφική παράσταση αυτής της σ.π.π. για ειδικές τιμές των παραμέτρων και για $x \in (0, 4)$.



Εικόνα 3. Η συνάρτηση πυκνότητας πιθανότητας της EL α) για $\beta=1$ και $p=0.2, 0.5, 0.8$ και β) για $p=0.5$ και $\beta=1, 2, 3$.

2.3.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

Η συνάρτηση του λογαρίθμου της πιθανοφάνειας με παρατηρούμενο δείγμα μεγέθους n , $y_{obs} = (x_i; i=1, \dots, n)$ με κατανομή EL δίνεται από τη σχέση:

$$\ell(p, \beta; y_{obs}) = -n \ln(-\ln(p)) + n \ln \beta + n \ln(1-p) - \beta \sum_{i=1}^n x_i - \sum_{i=1}^n \ln[1 - (1-p)e^{-\beta x_i}]$$

και συνακόλουθα οι μερικές παράγωγοι για τις παραμέτρους p και β δίνονται από τις σχέσεις:

$$\frac{\partial \ell(p, \beta; y_{obs})}{\partial p} = \frac{-n}{p \ln p} - \frac{n}{1-p} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}}, \quad (2.9)$$

και

$$\frac{\partial \ell(p, \beta; y_{obs})}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}}. \quad (2.10)$$

Λόγω της δυσκολίας εξεύρεσης αριθμητικών λύσεων του συστήματος των εξισώσεων $\frac{\partial \ell(p, \beta; y_{obs})}{\partial p} = 0$, $\frac{\partial \ell(p, \beta; y_{obs})}{\partial \beta} = 0$, χρησιμοποιούμε δύο θεωρήματα που αποδεικνύονται με μεθόδους παρόμοιες με αυτές της εργασίας των Adamidis and Loukas (1998).

Θεώρημα 2.5 (Tahmasbi and Rezaei (2008))

Ορίζουμε $h(p; \beta, y_{obs}) = \frac{-n}{p \ln p} - \frac{n}{1-p} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}}$, δηλαδή τη συνάρτηση του δεξιού μέλους της σχέσης (2.9), όπου β είναι η πραγματική τιμή της παραμέτρου. Τότε, για $n/2 < \sum_{i=1}^n e^{-\beta x_i}$, η εξίσωση $h(p; \beta, y_{obs}) = 0$ έχει τουλάχιστον μια ρίζα.

Θεώρημα 2.6 (Tahmasbi and Rezaei (2008))

Ορίζουμε $g(\beta; p, y_{obs}) = \frac{n}{\beta} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}}$, δηλαδή τη συνάρτηση του δεξιού μέλους της σχέσης (2.10) και $\bar{x} = n^{-1} \sum_{i=1}^n x_i$, όπου p είναι η πραγματική τιμή της παραμέτρου. Τότε, για δοθέν $p \in (0,1)$, η ρίζα της $g(\beta; p, y_{obs}) = 0$ βρίσκεται στο εσωτερικό του διαστήματος $[p\bar{x}^{-1}, \bar{x}^{-1}]$ και είναι μοναδική.

2.3.4 Η μέθοδος EM

Για την εφαρμογή του αλγορίθμου θεωρούμε ότι το δείγμα είναι “πλήρες”, δηλαδή αποτελείται από τις παρατηρούμενες τιμές της τυχαίας μεταβλητής X και τις τιμές μιας μη παρατηρούμενης τυχαίας μεταβλητής Z με από κοινού κατανομή:

$$f(x, z; p, \beta) = -(\ln p)^{-1} \beta e^{-\beta x} (1-p)^z, \quad z=1, \dots, x, p, \beta > 0.$$

Έστω $Y_{obs} = (x_i, z_i; i=1, \dots, n)$ τα υποθετικά δεδομένα. Στο Ε-βήμα υπολογίζουμε την υπό συνθήκη αναμενόμενη τιμή, δοθέντων των παρατηρούμενων τιμών $X_i, i=1, \dots, n$, δηλαδή την αναμενόμενη τιμή της $(Z | X; \theta^{(t)})$, όπου $\theta^{(t)} = (\beta^{(t)}, p^{(t)})$ είναι η τιμή της παραμέτρου θ . Αποδεικνύεται ότι η δεσμευμένη κατανομή της τυχαίας μεταβλητής $Z | X$ έχει πυκνότητα:

$$P(z | x; p, \beta) = (1 - (1-p)e^{-\beta x}) [(1-p)e^{-\beta x}]^{z-1}, \quad z \in N,$$

όπου $x, \beta, p \in R^+$. Άρα, η αναμενόμενη τιμή της είναι:

$$E(Z | X = x; p^{(h)}, \beta^{(h)}) = \left[1 - (1-p^{(h)})e^{-\beta^{(h)}x} \right]^{-1}.$$

Επομένως, αντιμετωπίζοντας τις δεσμευμένες αναμενόμενες τιμές ως δεδομένα, μια επανάληψη της μεθόδου EM ορίζεται από τις σχέσεις (Μ-βήμα μεγιστοποίησης):

$$\beta^{(h+1)} = n \left[\sum_{i=1}^n \frac{x_i}{1 - (1-p^{(h)})e^{-\beta^{(h)}x_i}} \right]^{-1},$$

και

$$p^{(h+1)} = \frac{-n(1-p^{(h+1)})}{\log(p^{(h+1)}) \sum_{i=1}^n \left[1 - (1-p^{(h)})e^{-\beta^{(h)}x_i} \right]^{-1}}.$$

2.3.5 Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας

Η κατανομή των εκτιμητών μέγιστης πιθανοφάνειας, $\hat{\theta}$, σύμφωνα με τις ασυμπτωτικές ιδιότητες των εκτιμητών αυτών, είναι κατά προσέγγιση η διδιάστατη κανονική με μέση τιμή θ και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα αναμενόμενης πληροφορίας $J(\theta) = E(I; \theta)$, όπου $I = I(\theta; y_{obs})$ είναι ο συμμετρικός, δεύτερης τάξης πίνακας παρατηρούμενης πληροφορίας, ενώ η αναμενόμενη τιμή λαμβάνεται ως προς τη κατανομή της τυχαίας μεταβλητής X . Τα στοιχεία του πίνακα

$$I(\theta; y_{obs}), I_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}, i, j = 1, 2,$$

δίνονται αναλυτικά από τις σχέσεις:

$$I_{11} = \frac{-n(\log p + 1)}{(p \log p)^2} + \frac{n}{(1-p)^2} - \sum_{i=1}^n \frac{e^{-2\beta x_i}}{(1 - (1-p)e^{-\beta x_i})^2},$$

$$I_{12} = I_{21} = -\sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - (1-p)e^{-\beta x_i})^2},$$

και

$$I_{22} = \frac{n}{\beta^2} - (1-p) \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - (1-p)e^{-\beta x_i})^2}.$$

Επιπλέον, τα στοιχεία του συμμετρικού δεύτερης τάξης πίνακα αναμενόμενης πληροφορίας $J(\theta)$ είναι:

$$J_{11} = \frac{-n(\log p + 1)}{(p \log p)^2} + \frac{n}{(1-p)^2} + n \frac{(1-4p)/p^2 - 2 \log p + 3}{2(1-p)^2 \log p},$$

$$J_{12} = J_{21} = n \frac{1-p+p \log p}{2\beta p(1-p) \log p},$$

και

$$J_{22} = -n \frac{d \log(p)}{\beta^2 \log p},$$

όπου $d \log(\cdot)$ η διλογαριθμική συνάρτηση.

2.4 Η Εκθετική - Δυναμοσειρά κατανομή (EPS)

Η Εκθετική-Δυναμοσειρά κατανομή (EPS) μελετάται αναλυτικά από τους Chahkandi and Ganjali (2009).

2.4.1 Φυσική ερμηνεία

Η φυσική διεργασία παραγωγής της κατανομής EPS περιγράφεται από το ακόλουθο θεώρημα.

Θεώρημα 2.7 (Chahkandi and Ganjali (2009))

Έστω $\{Y_i\}_{i=1}^z$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με πυκνότητα

$$f(y, \beta) = \beta e^{-\beta y}, \quad y \geq 0,$$

όπου $\beta \in R^+$ και Z είναι μια κατανομή δυναμοσειράς αποκομμένη στο 0, με συνάρτηση πιθανότητας

$$P(z; \theta) = \frac{\alpha_z \theta^z}{A(\theta)(1-\alpha)}, \quad z \in N, \quad \theta > 0,$$

όπου $(1-\alpha)A(\theta) = \sum_{z \in N} \alpha_z \theta^z$ και $\alpha = \frac{\alpha_0}{A(\theta)}$,

με

$$\alpha_z > 0 \text{ και } A(\theta) = \sum_z \alpha_z \theta^z,$$

η οποία είναι μια οικογένεια κατανομών που περιλαμβάνει τη Διωνυμική, την Poisson, την Αρνητική Διωνυμική και τη Λογαριθμική κατανομή (Johnson et al., 1993).

Τότε, η τυχαία μεταβλητή $X = \min(\{Y_i\}_{i=1}^z)$ ακολουθεί τη κατανομή EPS.

Απόδειξη (Chahkandi and Ganjali (2009))

Από τις υποθέσεις του θεωρήματος προκύπτει ότι η δεσμευμένη κατανομή της τυχαίας μεταβλητής $X|Z$ είναι εκθετική με παράμετρο βz . Κατά συνέπεια, η συνάρτηση πυκνότητας πιθανότητας του X όταν $z \in N$ είναι:

$$\begin{aligned} f(x; \xi) &= \sum_{z \in N} f(x|z; \beta) p(z; \theta) \\ &= \sum_{z \in N} \beta z e^{-\beta z x} \frac{\alpha_z \theta^z}{A(\theta) - \alpha_0} \\ &= \sum_{z \in N} \beta z \alpha_z \frac{(\theta e^{-\beta x})^z}{A(\theta) - \alpha_0}. \end{aligned}$$

Αλλά,

$$\begin{aligned} \sum_{z \in \mathbb{N}} \beta z a_z (\theta e^{-\beta x})^z &= -\frac{\partial}{\partial x} \sum_{z \in \mathbb{N}} a_z (\theta e^{-\beta x})^z \\ &= -\frac{\partial}{\partial x} (A(\theta e^{-\beta x}) - a_0) \\ &= -\frac{\partial}{\partial x} A(\theta e^{-\beta x}). \end{aligned}$$

Οπότε προκύπτει:

$$f(x; \xi) = \frac{-\frac{\partial}{\partial x} A(\theta e^{-\beta x})}{A(\theta) - a_0}, \quad (2.11)$$

με $\xi = (\theta, \beta)$, όπου οι παράμετροι θ και β είναι παράμετροι κλίμακας (scale) και μορφής (shape), αντίστοιχα. Αυτή είναι η σ.π.π. της EPS.

2.4.2 Χαρακτηριστικά της κατανομής

Δίνονται κατά σειρά η *αθροιστική συνάρτηση κατανομής*, η *ροπογεννήτρια*, η *συνάρτηση επιβίωσης*, η *συνάρτηση κινδύνου* και ο *μέσος υπολειπόμενος χρόνος*.

Η *αθροιστική συνάρτηση* κατανομής της τ.μ. X που ακολουθεί την Εκθετική-Δυναμοσειρά κατανομή με σ.π.π. αυτής της σχέσης (2.11), δίνεται από τη σχέση:

$$F(x; \xi) = 1 - \frac{A(\theta e^{-\beta x}) - a_0}{A(\theta) - a_0}, \quad x \geq 0.$$

Η *ροπογεννήτρια* συνάρτηση δίνεται από τις σχέσεις:

$$\begin{aligned}
M_X(t) &= \int_0^{\infty} e^{tx} f(x; \xi) dx \\
&= \int_0^{\infty} \sum_{z \in N} \frac{\beta z a_z \theta^z e^{(t-\beta z)x}}{\Lambda(\theta) - \alpha_0} dx \\
&= \sum_{z \in N} \frac{\beta z a_z \theta^z}{\Lambda(\theta) - \alpha_0} \frac{1}{t - \beta z} e^{(t-\beta z)x} \Big|_0^{\infty} \\
&= \left\{ \begin{array}{l} -\sum_{z \in N} \frac{\beta z a_z \theta^z}{(t - \beta z)(\Lambda(\theta) - \alpha_0)}, t < \beta z \\ \infty, t > \beta z \end{array} \right\}.
\end{aligned}$$

Η πιθανότητα επιβίωσης πέραν του χρόνου x , γνωστή και ως *συνάρτηση επιβίωσης* και η *συνάρτηση κινδύνου* έχουν αντίστοιχα τη μορφή:

$$S(x; \xi) = 1 - F(x; \xi) = \frac{\Lambda(\theta e^{-\beta x}) - a_0}{\Lambda(\theta) - a_0},$$

και

$$h(x; \xi) = \frac{-\frac{\partial}{\partial x} \Lambda(\theta e^{-\beta x})}{\Lambda(\theta e^{-\beta x}) - a_0}.$$

Ο μέσος υπολειπόμενος χρόνος ζωής δίνεται από τις σχέσεις:

$$\begin{aligned}
m(x_0; \xi) &= E(X - x_0 | X > x_0; \xi) \\
&= \int_{\{x: x > x_0\}} \frac{(x - x_0) f_x(x)}{P(X > x_0)} dx \\
&= [p(X > x_0)]^{-1} \int_0^{\infty} y f_x(x_0 + y) dy \\
&= [p(X > x_0)]^{-1} \int_0^{\infty} y \sum_{z \in N} \{\beta z e^{-\beta z(y+x_0)} \frac{a_z \theta^z}{A(\theta) - \alpha_0}\} dy \\
&= [p(X > x_0)]^{-1} \sum_{z \in N} \frac{a_z \theta^z e^{-\beta z x_0}}{A(\theta) - \alpha_0} \frac{1}{\beta z} \\
&= [p(X > x_0) \beta]^{-1} \frac{A(\theta e^{-\beta x_0}) - \alpha_0}{A(\theta) - \alpha_0} E_{\theta e^{-\beta x_0}} \left(\frac{1}{z} \right),
\end{aligned}$$

όπου $E_{\theta e^{-\beta x_0}} \left(\frac{1}{z} \right)$ είναι η αναμενόμενη τιμή στο $\frac{1}{z}$, περικόπτοντας το 0 από τη δυναμοσειρά κατανομή με παράμετρο $\theta e^{-\beta x_0}$.

2.4.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

Έστω ένα τυχαίο δείγμα μεγέθους n , $y_{obs} = (x_i; i=1, \dots, n)$ από τη (2.11), η συνάρτηση πιθανοφάνειας και ο λογάριθμος της δίνονται αντίστοιχα από τις σχέσεις:

$$L = L(\xi; y_{obs}) = \left(\frac{1}{A(\theta) - \alpha_0} \right)^n (\theta \beta)^n e^{-\beta \sum_{i=1}^n x_i} \sum_{i=1}^n A'(\theta e^{-\beta x_i}),$$

και

$$l(\xi; y_{obs}) = -n \log[A(\theta) - \alpha_0] + n \log(\theta\beta) - \sum_{i=1}^n \beta x_i + \sum_{i=1}^n \log A'(\theta e^{-\beta x_i}),$$

όπου $A'(\cdot)$ είναι η πρώτη παράγωγος ως προς θ .

Οι εξισώσεις μέγιστης πιθανοφάνειας βρέθηκαν να είναι:

$$\frac{\partial l(\xi; y_{obs})}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n x_i + \sum_{i=1}^n \frac{A''(\theta e^{-\beta x_i})(-\theta x_i e^{-\beta x_i})}{A'(\theta e^{-\beta x_i})}, \quad (2.12)$$

και

$$\frac{\partial l(\xi; y_{obs})}{\partial \theta} = \frac{n}{\theta} - \frac{nA'(\theta)}{A(\theta) - \alpha_0} + \sum_{i=1}^n \frac{A''(\theta e^{-\beta x_i})e^{-\beta x_i}}{A'(\theta e^{-\beta x_i})}, \quad (2.13)$$

όπου $A''(\cdot)$ είναι η δεύτερη παράγωγος ως προς θ .

Θέτοντας τις μερικές παραγώγους $\frac{\partial l}{\partial \theta_i}$, $i=1,2$ ίσες με το μηδέν, είναι

προφανής η δυσκολία επίλυσης του συστήματος των παραπάνω μη γραμμικών ως προς τις παραμέτρους εξισώσεων. Γι'αυτό, για την επίλυσή του χρησιμοποιείται κάποια από τις μεθόδους αριθμητικής ανάλυσης, όπως για παράδειγμα, η μέθοδος Newton-Raphson ή η μέθοδος scoring.

Τα ακόλουθα Θεωρήματα 2.8 και 2.9 εκφράζουν τις συνθήκες για την ύπαρξη και μοναδικότητα των εκτιμητών μέγιστης πιθανοφάνειας.

Θεώρημα 2.8 (Chahkandi and Ganjali (2009))

Ορίζουμε
$$g(\beta, \theta, y_{obs}) = \frac{n}{\beta} - \sum_{i=1}^n x_i + \sum_{i=1}^n \frac{A''(\theta e^{-\beta x_i})(-\theta x_i e^{-\beta x_i})}{A'(\theta e^{-\beta x_i})},$$

δηλαδή τη συνάρτηση στο δεξιό μέλος της (2.12) και $\bar{x} = n^{-1} \sum_{i=1}^n x_i$. Τότε,

δοθείσης της τιμής του θ στο διάστημα $(0, \infty)$, η εξίσωση $g(\beta, \theta, y_{obs})=0$ έχει λύση, η οποία βρίσκεται στο διάστημα $(\bar{X}^{-1}, A'(\theta) \{ \bar{X}(\theta A''(\theta) + A'(\theta)) \}^{-1})$ και είναι μοναδική.

Απόδειξη (Chahkandi and Ganjali (2009))

Έστω

$$w(\beta, \theta, y_{obs}) = \sum_{i=1}^n x_i \theta e^{-\beta x_i} \frac{A''(\theta e^{-\beta x_i})}{A'(\theta e^{-\beta x_i})},$$

η οποία είναι φθίνουσα ως προς β .

Έχουμε,

$$\lim_{\beta \rightarrow \infty} w(\beta, \theta, y_{obs}) = 0,$$

$$g(\beta, \theta, y_{obs}) < n\beta^{-1} - \sum_{i=1}^n x_i - \lim_{\beta \rightarrow \infty} w(\beta, \theta, y_{obs}) = n\beta^{-1} - \sum_{i=1}^n x_i.$$

Ως εκ τούτου $g(\beta, \theta, y_{obs}) < 0$, όταν $\beta > \bar{x}^{-1}$.

Από την άλλη πλευρά,

$$\lim_{\beta \rightarrow 0} w(\beta, \theta, y_{obs}) = \frac{\theta A''(\theta)}{A'(\theta)} \sum_{i=1}^n x_i,$$

έτσι ώστε

$$g(\beta, \theta, y_{obs}) > n\beta^{-1} - \sum_{i=1}^n x_i - \lim_{\beta \rightarrow 0} w(\beta, \theta, y_{obs})$$

$$= n\beta^{-1} - \sum_{i=1}^n x_i - \frac{\theta A''(\theta)}{A'(\theta)} \sum_{i=1}^n x_i,$$

και

$$\lim_{\beta \rightarrow 0} g(\beta, \theta, y_{obs}) = \sum_{i=1}^n x_i \left(-1 - \frac{\theta A''(\theta)}{A'(\theta)} \right) > 0.$$

Ως εκ τούτου $g(\beta, \theta, y_{obs}) > 0$, όταν $\beta < A'(\theta) \{ \bar{X} (\theta A''(\theta) + A'(\theta)) \}^{-1}$.

Άρα, η εξίσωση $g(\beta, \theta, y_{obs})$ έχει μια τουλάχιστον λύση στο διάστημα $(A'(\theta) \{ \bar{X} (\theta A''(\theta) + A'(\theta)) \}^{-1}, \bar{X}^{-1})$. Για να αποδείξουμε τη μοναδικότητα της ρίζας θεωρούμε την πρώτη παράγωγο $g'(\beta, \theta, y_{obs})$, η οποία δίνεται από τη σχέση:

$$g'(\beta, \theta, y_{obs}) = -n\beta^{-2} + \frac{\partial^2}{\partial \beta^2} \sum_{i=1}^n \ln(A'(\theta e^{-\beta x_i})).$$

Συμβολίζοντας με β^* τη λύση της $g'(\beta, \theta, y_{obs}) = 0$, προκύπτει ότι:

$$n\beta^{*-1} = \beta^* \frac{\partial^2}{\partial \beta^{*2}} \sum_{i=1}^n \ln(A'(\theta e^{-\beta^* x_i})),$$

και

$$g(\beta^*, \theta, y_{obs}) = \beta^* \frac{\partial^2}{\partial \beta^{*2}} \sum_{i=1}^n \ln(A'(\theta e^{-\beta^* x_i})) - \sum_{i=1}^n X_i \\ + \sum_{i=1}^n \frac{\partial}{\partial \beta^*} \sum_{i=1}^n \ln(A'(\theta e^{-\beta^* x_i})),$$

για κάθε β^* , $x_i > 0$, $i = 1, \dots, n$. Αυτό αποδεικνύει ότι η g είναι αρνητική στα στάσιμα σημεία της. Επιπλέον, $\lim_{\beta \rightarrow \infty} g(\beta, \theta, y_{obs}) = -\sum_{i=1}^n x_i$, γεγονός που αποδεικνύει ότι η g δε μπορεί να έχει παραπάνω από μία ρίζα.

Θεώρημα 2.9 (Chahkandi and Ganjali (2009))

Ορίζουμε $h(\theta; \beta, y_{obs}) = \frac{n}{\theta} - \frac{nA'(\theta)}{A(\theta) - \alpha_0} + \sum_{i=1}^n \frac{A''(\theta e^{-\beta x_i}) e^{-\beta x_i}}{A'(\theta e^{-\beta x_i})}$, δηλαδή

τη συνάρτηση στο δεξιό μέλος της (2.13), όπου β η αληθινή τιμή της παραμέτρου. Τότε,

i) Η εξίσωση $h(\theta; \beta, y_{obs}) = 0$ έχει τουλάχιστον μια ρίζα για τις κατανομές Exponential-Geometric (EG), Exponential-Poisson (EP), Exponential-Logarithmic (EL), Exponential-Negative binomial (EN), εάν $\frac{n}{2} < \sum_{i=1}^n e^{-\beta x_i}$.

ii) Εάν $h(p; \beta, y_{obs}) = \frac{\partial l(p, \beta, y_{obs})}{\partial p}$, όπου $p = \frac{\theta}{1+\theta}$ και $p \in (0, 1)$, τότε η εξίσωση $h(\theta; \beta, y_{obs}) = 0$ έχει τουλάχιστον μια ρίζα για την κατανομή Exponential-Binomial (EB), εάν $\frac{n}{2} < \sum_{i=1}^n e^{-\beta x_i}$ και $\sum_{i=1}^n e^{\beta x_i} > \frac{n^2}{n-1}$.

Απόδειξη (Chahkandi and Ganjali (2009))

i) Αποδείχτηκε για τις EG, EP και EL κατανομές από τους Adamidis and Loukas (1998), Kus (2007) και Tahmasbi and Rezaei (2008), αντίστοιχα. Εμείς θα αποδείξουμε το αντίστοιχο συμπέρασμα για την κατανομή Exponential-Negative binomial (EN).

Σε αυτήν την ειδική περίπτωση της EPS είναι:

$$A(\theta) = (1 - \theta)^{-r}.$$

Έτσι,

$$h(\theta; \beta, y_{obs}) = \frac{-nr}{(1 - \theta) - (1 - \theta)^{r+1}} + \frac{n}{\theta} + \sum_{i=1}^n \frac{(r+1)e^{-\beta x_i}}{1 - \theta e^{-\beta x_i}},$$

και προκύπτει ότι:

$$\lim_{\theta \rightarrow 0} h(\theta; \beta, y_{obs}) = -\frac{n}{2}(r+1) + (r+1) \sum_{i=1}^n e^{-\beta x_i},$$

και

$$\lim_{\theta \rightarrow 1} h(\theta; \beta, y_{obs}) = -\infty.$$

Οπότε, η εξίσωση $h(\theta; \beta, y_{obs}) = 0$ έχει τουλάχιστον μια ρίζα για $\theta > 0$,

εάν

$$-\frac{n}{2}(r+1) + (r+1) \sum_{i=1}^n e^{-\beta x_i} > 0 \quad \text{ή} \quad \frac{n}{2} < \sum_{i=1}^n e^{-\beta x_i}.$$

ii) Στην EB κατανομή έχουμε

$$A(p) = \left(\frac{1}{1-p} \right)^n.$$

Έτσι,

$$h(p, \beta, y_{obs}) = \frac{-n^2}{(1-p)(1-(1-p)^n)} + \frac{n}{p(1-p)} + \frac{n-1}{1-p} \sum_{i=1}^n \frac{e^{-\beta x_i}}{1-p + pe^{-\beta x_i}},$$

και προκύπτει ότι:

$$\lim_{p \rightarrow 0} h(p; \beta, y_{obs}) = -\frac{n}{2}(n-1) + (n-1) \sum_{i=1}^n e^{-\beta x_i},$$

και

$$\lim_{p \rightarrow 1} h(p; \beta, y_{obs}) = \sum_{i=1}^n \frac{-n+1+ne^{-\beta x_i}}{e^{-\beta x_i}}.$$

Οπότε, η εξίσωση $h(p; \beta, y_{obs}) = 0$ έχει τουλάχιστον μια ρίζα για $p \in (0, 1)$, εάν

$$-\frac{n}{2}(n-1) + (n-1) \sum_{i=1}^n e^{-\beta x_i} > 0 \quad \text{και} \quad \sum_{i=1}^n \frac{-n+1+ne^{-\beta x_i}}{e^{-\beta x_i}} < 0,$$

ή

$$\frac{n}{2} < \sum_{i=1}^n e^{-\beta x_i} \quad \text{και} \quad \sum_{i=1}^n e^{\beta x_i} > \frac{n^2}{n-1}.$$

2.4.4 Η μέθοδος EM

Για την εφαρμογή του αλγορίθμου θεωρούμε ότι το δείγμα είναι “πλήρες”, δηλαδή αποτελείται από τις παρατηρούμενες τιμές της τυχαίας μεταβλητής X και τις τιμές μιας μη παρατηρούμενης τυχαίας μεταβλητής Z , με από κοινού κατανομή:

$$f(x, z; \xi) = \beta z \exp(-\beta zx) a_z \theta^z (A(\theta) - \alpha_0)^{-1}, \quad \text{με } x > 0, z \in Z, \beta, \theta > 0.$$

Έστω $Y_{obs} = (x_i, z_i; i=1, \dots, n)$ τα υποθετικά δεδομένα. Στο E-βήμα υπολογίζουμε την υπο συνθήκη αναμενόμενη τιμή, δοθέντων των παρατηρούμενων τιμών X_i , $i=1, \dots, n$, δηλαδή η αναμενόμενη τιμή της $(Z | X; \xi^{(t)})$, όπου $\xi^{(t)} = (\theta^{(t)}, \beta^{(t)})$ είναι η τιμή της παραμέτρου ξ .

Αποδεικνύεται ότι η δεσμευμένη κατανομή της τυχαίας μεταβλητής $Z | X$ έχει πυκνότητα:

$$P(z | x; \xi) = \beta z (\theta e^{-\beta x})^z a_z \left(-\frac{\partial}{\partial x} A(\theta e^{-\beta x}) \right)^{-1}, \quad z \in N,$$

όπου $x, \beta, \theta \in R^+$. Άρα, η αναμενόμενη τιμή της είναι:

$$E(Z | X = x; \xi) = 1 + \theta e^{-\beta x} \frac{A''(\theta e^{-\beta x})}{A'(\theta e^{-\beta x})}.$$

Επομένως, αντιμετωπίζοντας τις δεσμευμένες αναμενόμενες τιμές ως δεδομένα, μια επανάληψη της μεθόδου EM ορίζεται από τις σχέσεις (Μ-βήμα μεγιστοποίησης):

$$\beta^{(t+1)} = n \left[\sum_{i=1}^n x_i \left(1 + \frac{\theta^{(t)} e^{-\beta^{(t)} x_i} A''(\theta^{(t)} e^{-\beta^{(t)} x_i})}{A'(\theta^{(t)} e^{-\beta^{(t)} x_i})} \right) \right]^{-1},$$

και το $\theta^{(t+1)}$ είναι η λύση της ακόλουθης εξίσωσης:

$$\frac{\theta^{(t+1)} A'(\theta^{(t+1)})}{A(\theta^{(t+1)}) - \alpha_0} = 1 + \frac{1}{n} \sum_{i=1}^n \frac{\theta^{(t)} e^{-\beta^{(t)} x_i} A''(\theta^{(t)} e^{-\beta^{(t)} x_i})}{A'(\theta^{(t)} e^{-\beta^{(t)} x_i})}.$$

2.4.5 Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας

Εφαρμόζοντας τη συνήθη προσέγγιση μεγάλου δείγματος, η κατανομή των εκτιμητών μέγιστης πιθανοφάνειας, $\hat{\xi}$, είναι κατά προσέγγιση η διδιάστατη κατανομή με μέση τιμή ξ και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του πίνακα αναμενόμενης πληροφορίας $J(\xi) = E(I; \xi)$, όπου $I = I(\xi; y_{obs})$ είναι ο συμμετρικός δεύτερης τάξης πίνακας παρατηρούμενης πληροφορίας και

η αναμενόμενη τιμή λαμβάνεται ως προς την κατανομή της τυχαίας μεταβλητής X .

Τα στοιχεία του πίνακα $I(\xi; y_{obs})$ δίνονται από τις εξής σχέσεις:

$$I_{11} = \frac{\partial^2 n \log(A(\theta) - \alpha_0)}{\partial \theta^2} + \frac{n}{\theta^2} - \frac{\partial^2 \sum_{i=1}^n \log(A'(\theta e^{-\beta x_i}))}{\partial \theta^2},$$

$$I_{12} = - \frac{\partial^2 \sum_{i=1}^n \log(A'(\theta e^{-\beta x_i}))}{\partial \beta \partial \theta},$$

και

$$I_{22} = n\beta^{-2} - \frac{\partial^2 \sum_{i=1}^n \log(A'(\theta e^{-\beta x_i}))}{\partial \beta^2}.$$

3. ΕΦΑΡΜΟΓΕΣ

Οι εφαρμογές αφορούν σε τρία σύνολα πραγματικών δεδομένων (παρατίθενται στο Παράρτημα). Το πρώτο σύνολο δεδομένων, αποτελείται από 213 παρατηρήσεις που αφορούν στον αριθμό των διαδοχικών αποτυχιών του συστήματος κλιματισμού του κάθε μέλους της από ένα στόλο από 13 αεροσκάφη. Τα δεδομένα αναλύθηκαν από τους Adamidis and Loukas (1998), Kus (2007) και Tahmasbi and Rezaei (2008). Το δεύτερο σύνολο δεδομένων, αποτελείται από 24 παρατηρήσεις που αναφέρονται στην περίοδο ανάμεσα σε διαδοχικούς σεισμούς στη ζώνη Βόρειας Ανατολίας. Αυτά τα δεδομένα αναλύθηκαν από τον Kus (2007) και τους Tahmasbi and Rezaei (2008). Το τρίτο σύνολο δεδομένων, αποτελείται από 13 παρατηρήσεις που αφορούν τις ώρες διαστημάτων (σε μέρες) των διαδοχικών σεισμών. Αυτά τα δεδομένα αναλύθηκαν στην εργασία των Tahmasbi and Rezaei (2008). Στα δεδομένα αυτά προσαρμόστηκαν τα μοντέλα **EL**, **EG**, **EB** και **EP**. Οι σ.π.π. αυτών των κατανομών είναι:

$$\mathbf{EP}: f_1(x; \theta_1) = \frac{\lambda_1 \beta_1}{(1 - e^{-\lambda_1})} e^{-\lambda_1 - \beta_1 x + \lambda_1 e^{-\beta_1 x}}, \quad x, \beta_1, \lambda_1 \in R^+,$$

$$\mathbf{EG}: f_2(x; \theta_2) = \lambda_2 (1 - \beta_2) e^{-\lambda_2 x} (1 - \beta_2 e^{-\lambda_2 x})^{-2}, \quad x > 0, \lambda_2 > 0, \beta_2 \in (0, 1),$$

$$\mathbf{EB}: f_3(x; \theta_3) = \frac{n \lambda_3 \beta_3 e^{-\lambda_3 x} (1 + \beta_3 e^{-\lambda_3 x})^{n-1}}{(1 + \beta_3)^n - 1}, \quad \lambda_3 > 0, \beta_3 \in (0, 1),$$

$$\mathbf{EL}: f_4(x; \theta_4) = -\frac{\lambda_4 \beta_4 e^{-\lambda_4 x}}{\log(1 - \beta_4)(1 - \beta_4 e^{-\lambda_4 x})}, \quad \lambda_4 > 0, \beta_4 \in (0, 1),$$

όπου $\theta_j = (\beta_j, \lambda_j)$, $j=1, 2, \dots, 4$.

Το EG μοντέλο αποδίδει τις αποτυχίες του συστήματος κλιματισμού με την παρουσία των ελαττωματικών εξαρτημάτων ή σφαλμάτων στα οποία υπέπεσε κατά τη διάρκεια της παραγωγής. Με όμοιο τρόπο αποδίδει τις καταστροφές των ανθρακωρυχείων σε κατασκευαστικά σφάλματα ή σε ανθρώπινα λάθη που διαπράχθηκαν από τους άπειρους ανθρακωρύχους.

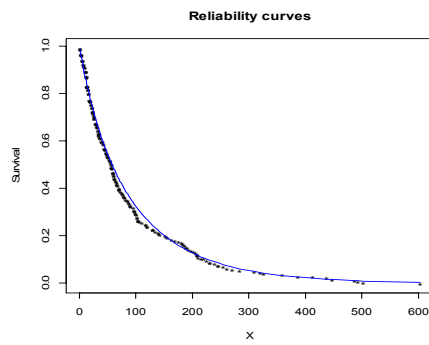
Η προσαρμογή τους έγινε με τη μέθοδο της μέγιστης πιθανοφάνειας και η καταλληλότητά τους για την περιγραφή των δεδομένων εξετάστηκε τόσο με στατιστικές όσο και με γραφικές μεθόδους. Στον Πίνακα 1, παρουσιάζονται οι τιμές των εκτιμητών μέγιστης πιθανοφάνειας, η τιμή του λογαρίθμου της συνάρτησης πιθανοφάνειας (LL), το κριτήριο (AIC) και η τιμή του στατιστικού των Kolmogorov-Smirnov (K-S) τεστ, καθώς και η αντίστοιχη κρίσιμη τιμή (p-value).

Πίνακας 1. Εκτιμητές παραμέτρων, τιμές λογαρίθμου συνάρτησης πιθανοφάνειας, κριτηρίου AIC και στατιστικού K-S, καθώς και κρίσιμες τιμές από την προσαρμογή των συγκρινόμενων κατανομών πιθανότητας στα τρία διαθέσιμα σύνολα δεδομένων.

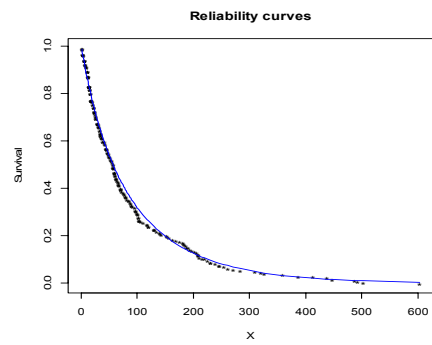
Σύνολο Δεδομένων	Κατανομή	$\hat{\theta} = (\hat{\beta}, \hat{\lambda})$	LL	AIC (-2logL)	K-S (p-value)
Plane (n=213)	EL	(0.3874, 0.0084)	-1176.1335	2352.267	0.0486 (0.6965)
	EG	(0.4276, 0.008)	-1175.925	2351.849	0.0493 (0.6797)
	EB	(0.00628, 0.00751)	-1175.793	2351.585	0.0467 (0.7425)
	EP	(0.00751, 1.3324)	-1175.793	2351.585	0.047 (0.7348)
Earth1 (n=24)	EL	(0.12585, 0.000414)	-196.6475	393.295	0.0885 (0.9833)
	EG	(0.74, 0.000333)	-196.5148	393.0296	0.0734 (0.9983)
	EB	(0.10457, 0.000426)	-196.8592	393.7184	0.0802 (0.9944)
	EP	(0.00036, 2.64)	-196.7696	393.5392	0.0936 (0.9715)
Earth2 (n=13)	EL	(0.20852, 0.00221)	-91.79099	183.582	0.1308 (0.9792)
	EG	(0.10753, 0.00221)	-91.73976	183.4795	0.1316 (0.9561)
	EB	(0.01740, 0.00222)	-91.74046	183.4809	0.1327 (0.9531)
	EP	(0.00222, 0.213)	-91.74033	183.4807	0.132 (0.9552)

Για το πρώτο σύνολο δεδομένων, το οποίο αφορά τον αριθμό των διαδοχικών αποτυχιών του συστήματος κλιματισμού του κάθε μέλους της από ένα στόλο από 13 αεροσκάφη, οι καμπύλες επιβίωσης των κατανομών: (α) EL, (β) EG, (γ) EB και (δ) EP, αντίστοιχα, δίνονται παρακάτω:

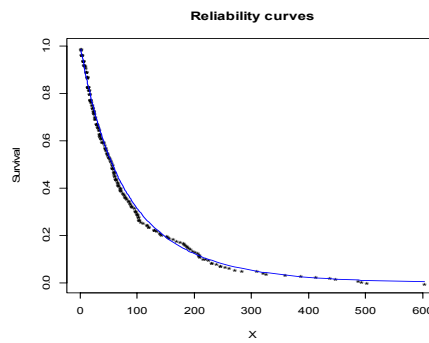
(α)



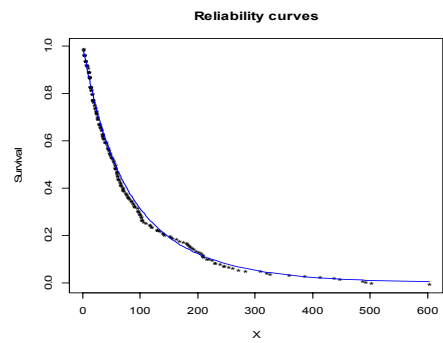
(β)



(γ)

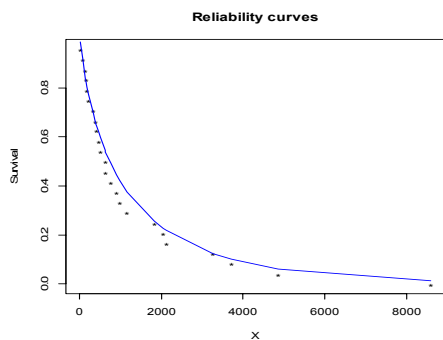


(δ)

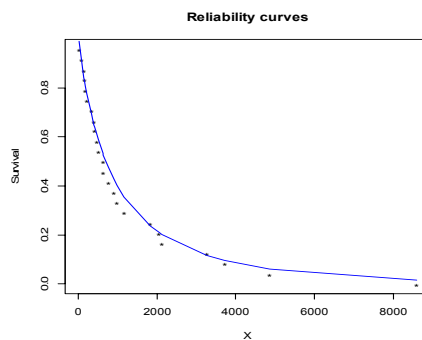


Για το δεύτερο σύνολο δεδομένων, το οποίο αφορά την περίοδο ανάμεσα σε διαδοχικούς σεισμούς στη ζώνη της Βόρειας Ανατολίας, οι καμπύλες επιβίωσης των κατανομών: (α) EL, (β) EG, (γ) EB και (δ) EP, αντίστοιχα, δίνονται παρακάτω:

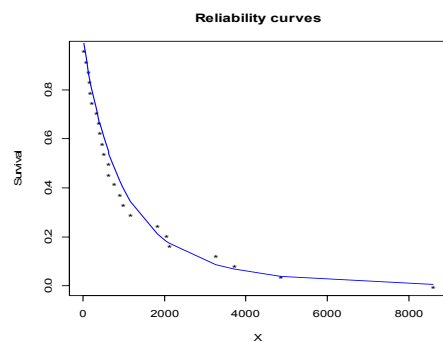
(α)



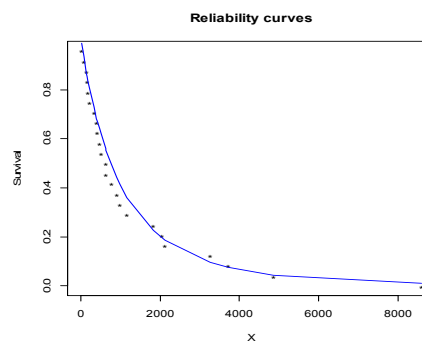
(β)



(γ)

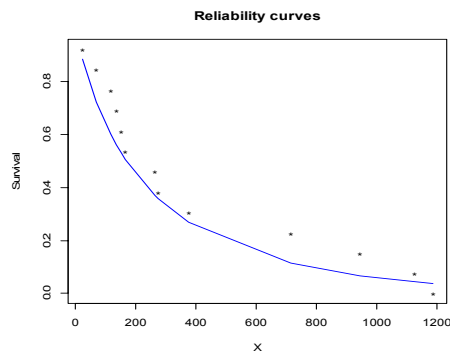


(δ)

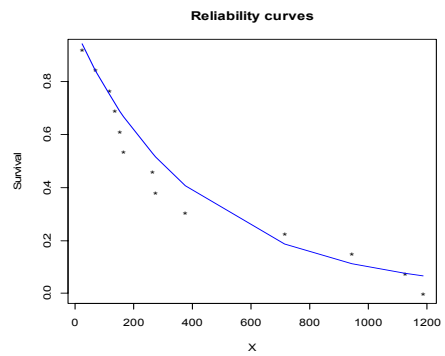


Για το τρίτο σύνολο δεδομένων, το οποίο αφορά τις ώρες διαστημάτων (σε μέρες) των διαδοχικών σεισμών, οι καμπύλες επιβίωσης των κατανομών: (α) EL, (β) EG, (γ) EB και (δ) EP, αντίστοιχα, δίνονται παρακάτω:

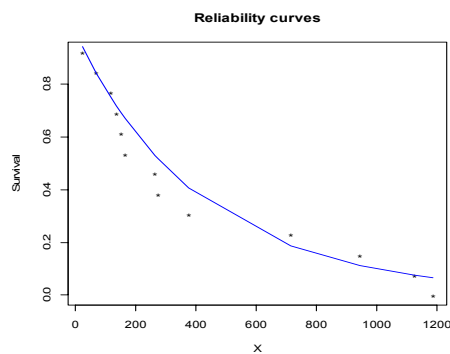
(α)



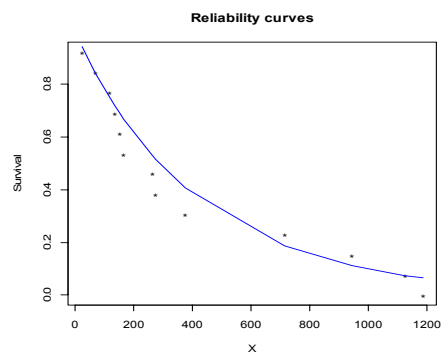
(β)



(γ)



(δ)



Τα αποτελέσματα για το πρώτο σύνολο δεδομένων αποδεικνύουν ότι οι κατανομές EB και EP δίνουν εξίσου καλή προσαρμογή. Για το δεύτερο σύνολο δεδομένων, η καλύτερη προσαρμογή είναι αυτή της κατανομής EG. Για το τρίτο σύνολο δεδομένων, η καλύτερη προσαρμογή επιτυγχάνεται από το μοντέλο των Tahmasbi and Rezaei (2008).

Οι στατιστικές όσο και οι γραφικές μέθοδοι που χρησιμοποιήθηκαν αποδεικνύουν ότι οι κατανομές EL, EG, EB και EP έχουν εξίσου το ίδιο ικανοποιητική προσαρμογή και για τα τρία σύνολα δεδομένων.

4. ΑΛΛΕΣ ΜΟΡΦΕΣ ΣΥΝΑΡΤΗΣΗΣ ΚΙΝΔΥΝΟΥ

Εισαγωγή

Οι πιο δημοφιλείς κατανομές ανάλυσης επιβίωσης είναι εκείνες με μονότονη συνάρτηση κινδύνου (π.χ. Weibull, Γάμμα). Όμως, η φύση του μηχανισμού “θανάτου-αποτυχίας” τις καθιστά ανεπαρκείς (ή και ακατάλληλες) για την περιγραφή της ικανότητας επιβίωσης του υπό εξέταση πληθυσμού.

Κατά συνέπεια, τα πλέον ρεαλιστικά μοντέλα για την περιγραφή χρόνων ζωής βιολογικών οργανισμών, αλλά και εργοστασιακών προϊόντων είναι τα μοντέλα με συνάρτηση κινδύνου κυπελλοειδούς μορφής, όταν το φαινόμενο “χρόνος ζωής” παρακολουθείται από τη γέννηση μέχρι την κατάληξή του. Πράγματι, στα πρώτα στάδια ζωής βιολογικών οργανισμών, η πιθανότητα θανάτου τους είναι αρκετά υψηλή λόγω νεογνικής θνησιμότητας, που οφείλεται κυρίως σε κληρονομικές ανεπάρκειες. Καθώς όμως, η επιρροή των τελευταίων εξασθενεί με την πάροδο του χρόνου, η πιθανότητα θανάτου φθίνει σταδιακά, έως ότου φθάσει σε μια κατώτατη τιμή, όπου και διατηρείται σχεδόν σταθερή για κάποιο χρονικό διάστημα. Αυτό το διάστημα αντιστοιχεί στη νεαρή ηλικία, στην οποία συμβαίνουν θάνατοι κυρίως λόγω ξαφνικών ατυχημάτων. Τέλος, από κάποια χρονική στιγμή και μετά η πιθανότητα θανάτου αυξάνει σταδιακά λόγω γήρανσης (Rajarshi and Rajarshi, 1988 και Lai et al, 2001).

Σε άλλες περιπτώσεις, όπως για παράδειγμα, όταν παρατηρείται μεγάλη συχνότητα πρόωρων “θανάτων-αποτυχιών”, καταλληλότερες για την περιγραφή της βιωσιμότητας ή ανθεκτικότητας του υπό μελέτη πληθυσμού, είναι οι κατανομές με μονοκόρυφη συνάρτηση κινδύνου, όπως η λογαριθμική και αντίστροφη καν/κή (Johnson et al., 1994).

Επιπρόσθετα, οι γενικευμένες εκθετικές κατανομές δύναται να χρησιμοποιηθούν για τη μοντελοποίηση της μέγιστης διάρκειας ζωής

των τυχαίων δειγμάτων από τις EP, EG, EL κατανομές, αντίστοιχα. Αξίζει να σημειωθεί ότι οι κατανομές αυτές έχουν επιθυμητή φυσική ερμηνεία, η οποία θα αναφερθεί παρακάτω.

4.1 Κατανομές με Κυπελλοειδούς μορφής συνάρτησης κινδύνου, B.T.F.R. (Bathtub Shaped Failure Rate)

Ορισμός 4.1

Έστω T μια θετική, συνεχής τυχαία μεταβλητή, με κατανομή $f(t)$ και συνάρτηση κινδύνου $h(t)$, συνεχή και δυο φορές παραγωγίσιμη στο $(0, \infty)$. Η $f(t)$ είναι B.T.F.R., αν υπάρχει $t_0 \in (0, \infty)$, τέτοιο ώστε $h'(t) < 0$ για κάθε $t < t_0$ και $h'(t) > 0$ για κάθε $t > t_0$, με $h'(t_0) = 0$.

Ορισμός 4.2

Μια κατανομή ανάλυσης επιβίωσης έχει τη γενικευμένη ή τροποποιημένη B.T.F.R. ιδιότητα, αν υπάρχει ένα t^* με $0 \leq t^* \leq \infty$, τέτοιο ώστε η $h(t)$ να είναι κυπελλοειδούς μορφής για κάθε $t \geq t^*$ (Block et al, 2003).

Η συνάρτηση κινδύνου περιγράφει τον τρόπο με τον οποίο μεταβάλλεται η στιγμιαία πιθανότητα “θανάτου” με το χρόνο. Έτσι, η συνάρτηση κινδύνου παίζει πολύ σημαντικό ρόλο στην ανάλυση επιβίωσης. Για την προσαρμογή κατάλληλου μοντέλου στα δεδομένα χρόνων ζωής, είναι ιδιαίτερα χρήσιμη η εκ των προτέρων πληροφορία για τη συνάρτηση κινδύνου ενός πληθυσμού.

Στις εφαρμογές χρησιμοποιούνται τα μοντέλα με αύξουσα συνάρτηση κινδύνου (I.F.R.), φθίνουσα συνάρτηση κινδύνου (D.F.R.), κυπελλοειδούς μορφής συνάρτησης κινδύνου (B.T.F.R.) ή ανάποδης της κυπελλοειδούς μορφής συνάρτησης κινδύνου (U.B.T.F.R.) (Lawless, 1982).

Ορισμός 4.3

Μια κατανομή ανάλυσης επιβίωσης ορίζεται ως:

α) I.F.R. (D.F.R.) αν και μόνο αν $h'(t) > 0$ για κάθε $t > 0$ ($h'(t) < 0$ για κάθε $t > 0$),

β) B.T.F.R. (U.B.T.F.R.) αν και μόνο αν υπάρχει $t_0 > 0$, τέτοιο ώστε:

- $h'(t) < 0$ ($h'(t) > 0$) για κάθε $t \in (0, t_0)$,
- $h'(t_0) = 0$,
- $h'(t) > 0$ ($h'(t) < 0$) για κάθε $t \in (t_0, \infty)$.

Ορισμός 4.4

Έστω T διακριτή τυχαία μεταβλητή, με τιμές t_1, t_2, \dots , τέτοιες ώστε $0 \leq t_1 < t_2 < \dots$. Τότε, η κατανομή της τυχαίας μεταβλητής T είναι B.T.F.R., αν για την ακολουθία $\{h(t_j), j \geq 1\}$ υπάρχουν t_{n_1} και t_{n_2} με $1 \leq t_{n_1} \leq t_{n_2} \leq \infty$, τέτοια ώστε να ισχύουν οι ανισότητες:

$$h(t_1) > h(t_2) > \dots > h(t_{n_1-1}) > h(t_{n_1}) = \dots = h(t_{n_2}) < h(t_{n_2+1}) < \dots$$

Τα t_{n_1} και t_{n_2} λέγονται κρίσιμα σημεία (change points) της ακολουθίας $\{h(t_j), j \geq 1\}$.

4.2 Η Γενικευμένη Εκθετική - Poisson κατανομή

Η Γενικευμένη Εκθετική-Poisson κατανομή (GEP) μελετάται αναλυτικά από τους Barreto-Souza and Cordeiro-Neto (2009).

4.2.1 Φυσική ερμηνεία της κατανομής

Έστω X μια τυχαία μεταβλητή, η οποία ακολουθεί τη Γενικευμένη Εκθετική-Poisson κατανομή, δηλαδή $X \sim GEP(\lambda, \beta, \alpha)$.

Η γενικευμένη εκθετική (GE) κατανομή έχει αθροιστική συνάρτηση κατανομής της μορφής:

$$G_{GE}(x) = (1 - e^{-\lambda x})^\alpha, \quad x, \lambda, \alpha > 0 \quad (\text{Gupta and Kundu, (1999)}).$$

Επομένως, η αθροιστική συνάρτηση κατανομής της Γενικευμένης Εκθετικής-Poisson (GEP) κατανομής, είναι της μορφής:

$$F(x) = G_{EP}(x)^\alpha = \left(\frac{1 - e^{-\lambda + \lambda \exp(-\beta x)}}{1 - e^{-\lambda}} \right)^\alpha, \quad x, \lambda, \alpha > 0.$$

Η σημασία της είναι αξιοσημείωτη. Για παράδειγμα, αν υπάρχουν n ανεξάρτητα εξαρτήματα σε ένα παράλληλο σύστημα και ο χρόνος ζωής αυτών ακολουθεί την κατανομή EP, τότε η διάρκεια ζωής του συστήματος μπορεί να περιγραφεί από τη κατανομή GEP.

4.2.2 Ορισμοί και χαρακτηριστικά της κατανομής

Η **αθροιστική συνάρτηση** κατανομής της X , δίνεται από τη σχέση:

$$F(x) = G_{EP}(x)^\alpha = \left(\frac{1 - e^{-\lambda + \lambda \exp(-\beta x)}}{1 - e^{-\lambda}} \right)^\alpha, \quad x, \lambda, \alpha > 0.$$

Επιπλέον, η *συνάρτηση πυκνότητας πιθανότητας* της, δίνεται από τη σχέση:

$$f(x) = \frac{\alpha\lambda\beta}{(1-e^{-\lambda})^\alpha} \left\{1 - e^{-\lambda+\lambda\exp(-\beta x)}\right\}^{\alpha-1} e^{-\lambda-\beta x+\lambda\exp(-\beta x)}, x, \lambda, \beta, \alpha > 0.$$

Για $\alpha = 1$ προκύπτει η Εκθετική-Poisson κατανομή.

Οι *συναρτήσεις επιβίωσης και κινδύνου* της, είναι αντίστοιχα:

$$S(x) = 1 - \left(\frac{1 - e^{-\lambda+\lambda\exp(-\beta x)}}{1 - e^{-\lambda}} \right)^\alpha, x > 0,$$

και

$$h(x) = \alpha\lambda\beta \frac{\left\{1 - e^{-\lambda+\lambda\exp(-\beta x)}\right\}^{\alpha-1} e^{-\lambda-\beta x+\lambda\exp(-\beta x)}}{(1 - e^{-\lambda})^\alpha - \left\{1 - e^{-\lambda+\lambda\exp(-\beta x)}\right\}}, x > 0.$$

Στην περίπτωση, όπου $X \sim GEP(\lambda, \beta, \alpha)$ και α μη ακέραιος θετικός, οι *ροπές r τάξης* δίνονται από τη σχέση:

$$E(X^r) = \frac{r!\Gamma(\alpha+1)\lambda}{(1-e^{-\lambda})^\alpha \beta^r} \sum_{j=0}^{\infty} \frac{(-1)^j e^{-\lambda(j+1)}}{\Gamma(\alpha-j)j!} F_{r+1,r+1}([1,1,\dots,1],[2,2,\dots,2],\lambda(j+1))$$

4.2.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

Ο λογάριθμος της συνάρτησης πιθανοφάνειας που βασίζεται σε τυχαίο δείγμα μεγέθους n , $y_{obs} = (x_i; i = 1, \dots, n)$ που ακολουθεί τη κατανομή GEP, είναι της μορφής:

$$\begin{aligned} \ell(\theta; y_{obs}) &= -\lambda + \log(\alpha\lambda\beta) - \alpha \log(1 - e^{-\lambda}) \\ &\quad + (\alpha - 1) \log\left\{1 - e^{-\lambda + \lambda \exp(-\beta x)}\right\} - \beta x + \lambda e^{-\beta x}, \end{aligned}$$

απ' όπου προκύπτουν οι εξισώσεις μέγιστης πιθανοφάνειας:

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \lambda} = \frac{1}{\lambda} - \frac{\alpha}{e^{\lambda} - 1} + (e^{-\beta x} - 1) \left[1 - (\alpha - 1) \left\{ e^{\lambda - \lambda \exp(-\beta x)} - 1 \right\}^{-1} \right],$$

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \beta} = \frac{1}{\beta} - x + \lambda x e^{-\beta x} \left[(\alpha - 1) \left\{ e^{\lambda - \lambda \exp(-\beta x)} - 1 \right\}^{-1} - 1 \right],$$

και

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \alpha} = \frac{1}{\alpha} - \log(1 - e^{-\lambda}) + \log\left\{1 - e^{-\lambda + \lambda \exp(-\beta x)}\right\}.$$

Οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων $\theta = (\lambda, \beta, \alpha)$ προκύπτουν από την επίλυση του συστήματος των τριών μη-γραμμικών εξισώσεων πιθανοφάνειας

$$\frac{\partial l}{\partial \lambda} = 0, \quad \frac{\partial l}{\partial \beta} = 0 \quad \text{και} \quad \frac{\partial l}{\partial \alpha} = 0.$$

Είναι προφανές ότι η επίλυση του εν λόγω συστήματος, δύναται να επιτευχθεί μόνο με τη χρήση μιας επαναληπτικής αριθμητικής μεθόδου, όπως είναι οι μέθοδοι Newton-Raphson, quasi-Newton και Nelder-Mead.

4.3 Η Γενικευμένη Εκθετική - Γεωμετρική κατανομή

Η Γενικευμένη Εκθετική-Γεωμετρική κατανομή (GEG) μελετάται αναλυτικά από τους Silva and Barreto Souza (2010).

4.3.1 Φυσική ερμηνεία της κατανομής

Έστω X μια τυχαία μεταβλητή, η οποία ακολουθεί τη Γενικευμένη Εκθετική-Γεωμετρική κατανομή, δηλαδή $X \sim GEG(\beta, p, \alpha)$.

Η γενικευμένη εκθετική (GE) κατανομή έχει αθροιστική συνάρτηση κατανομής της μορφής:

$$G_{GE}(x) = (1 - e^{-\lambda x})^\alpha, \quad x, \lambda, \alpha > 0 \quad (\text{Gupta and Kundu, (1999)}).$$

Επομένως, η αθροιστική συνάρτηση κατανομής της Γενικευμένης Εκθετικής-Γεωμετρικής (GEG) κατανομής, είναι της μορφής:

$$F(x; \beta, p, \alpha) = G_{EG}(x; \beta, p, \alpha)^\alpha = \left(\frac{1 - e^{-\beta x}}{1 - pe^{-\beta x}} \right)^\alpha, \quad x > 0,$$

όπου $\beta > 0$, $p \in (0, 1)$ και $\alpha > 0$.

Η σημασία της είναι αξιοσημείωτη, γιατί η κατανομή GEG περιγραφεί τη διάρκεια ζωής του συστήματος, όταν έχουμε n ανεξάρτητα εξαρτήματα σε ένα παράλληλο σύστημα και ο χρόνος ζωής αυτών ακολουθεί την κατανομή EG.

4.3.2 Ορισμοί και χαρακτηριστικά της κατανομής

Η **αθροιστική συνάρτηση** κατανομής της X , δίνεται από τη σχέση:

$$F(x; \beta, p, \alpha) = \left(\frac{1 - e^{-\beta x}}{1 - pe^{-\beta x}} \right)^\alpha, \quad x > 0,$$

όπου $\beta > 0$, $p \in (0, 1)$ και $\alpha > 0$.

Επιπλέον, η **συνάρτηση πυκνότητας πιθανότητας** της, δίνεται από τη σχέση:

$$f(x; \beta, p, \alpha) = \frac{\alpha \beta (1-p) e^{-\beta x} (1 - e^{-\beta x})^{\alpha-1}}{(1 - pe^{-\beta x})^{\alpha+1}}, \quad x > 0,$$

όπου $\beta > 0$, $p \in (0, 1)$ και $\alpha > 0$. Οι παράμετροι β , α είναι παράμετροι κλίμακας και σχήματος, αντίστοιχα.

Για $\alpha = 1$ προκύπτει η Εκθετική-Γεωμετρική κατανομή.

Οι **συναρτήσεις επιβίωσης** και **κινδύνου** της, είναι αντίστοιχα:

$$S(x; \beta, p, \alpha) = 1 - \left(\frac{1 - e^{-\beta x}}{1 - pe^{-\beta x}} \right)^\alpha, \quad x > 0,$$

και

$$h(x; \beta, p, \alpha) = \frac{\alpha \beta (1-p) e^{-\beta x} (1 - e^{-\beta x})^{\alpha-1} \{1 - pe^{-\beta x}\}^{-1}}{(1 - pe^{-\beta x})^\alpha - \{1 - e^{-\beta x}\}^\alpha}, \quad x > 0.$$

Στην περίπτωση, όπου $X \sim GEP(\beta, p, \alpha)$, οι **ροπές r τάξης** δίνονται από τη σχέση:

$$E(X^r; \beta, p, \alpha) = \frac{(1-p)(-1)^r \lambda}{\beta^r \Gamma(\alpha)} \sum_{j=0}^{\infty} \frac{p^j}{j!} \Gamma(\alpha + j + 1) \frac{\partial^r}{\partial t^r} B(\alpha, 1+t-\alpha) /_{t=\alpha+j}$$

4.3.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

Ο λογάριθμος της συνάρτησης πιθανοφάνειας που βασίζεται σε τυχαίο δείγμα μεγέθους n , $y_{obs} = (x_i; i=1, \dots, n)$ που ακολουθεί τη κατανομή GEG, είναι της μορφής:

$$\begin{aligned} \ell(\theta; y_{obs}) = & n \log \{ \alpha \beta (1-p) \} - \beta \sum_{i=1}^n x_i + (\alpha - 1) \sum_{i=1}^n \log(1 - e^{-\beta x_i}) \\ & - (\alpha + 1) \sum_{i=1}^n \log(1 - p e^{-\beta x_i}), \end{aligned}$$

απ' όπου προκύπτουν οι εξισώσεις μέγιστης πιθανοφάνειας:

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n x_i + (\alpha - 1) \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{1 - e^{-\beta x_i}} - p(\alpha + 1) \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{1 - p e^{-\beta x_i}},$$

$$\frac{\partial \ell(\theta; y_{obs})}{\partial p} = -\frac{n}{1-p} + (\alpha + 1) \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - p e^{-\beta x_i}},$$

και

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log \left(\frac{1 - e^{-\beta x_i}}{1 - p e^{-\beta x_i}} \right).$$

Οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων $\theta = (\beta, p, \alpha)$ προκύπτουν από την επίλυση του συστήματος των τριών μη-γραμμικών εξισώσεων πιθανοφάνειας

$$\frac{\partial l}{\partial \beta} = 0, \quad \frac{\partial l}{\partial p} = 0 \quad \text{και} \quad \frac{\partial l}{\partial \alpha} = 0.$$

Παρατηρούμε ότι για τη τελευταία εξίσωση,

$$\frac{\partial \ell(\theta; y_{obs})}{\partial \alpha} = 0,$$

μπορεί να υπάρξει ακριβής λύση η:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \log \left(\frac{1 - \hat{p}e^{-\hat{\beta}x_i}}{1 - e^{-\hat{\beta}x_i}} \right)},$$

όπου $\hat{\alpha}$, $\hat{\beta}$ και \hat{p} είναι οι ΕΜΠ.

4.4 Η Γενικευμένη Εκθετική - Λογαριθμική κατανομή

Η Γενικευμένη Εκθετική-Λογαριθμική κατανομή (GEL) μελετάται αναλυτικά από τους Παππάς, Αδαμίδης και Λουκάς (2012).

4.4.1 Φυσική ερμηνεία της κατανομής

Έστω X μια τυχαία μεταβλητή, η οποία ακολουθεί τη Γενικευμένη Εκθετική-Λογαριθμική κατανομή, δηλαδή $X \sim GEL(\alpha, \lambda, p)$.

Η γενικευμένη εκθετική (GE) κατανομή έχει αθροιστική συνάρτηση κατανομής της μορφής:

$$G_{GE}(x) = (1 - e^{-\lambda x})^\alpha, \quad x, \lambda, \alpha > 0 \quad (\text{Gupta and Kundu, (1999)}).$$

Επομένως, η αθροιστική συνάρτηση κατανομής της Γενικευμένης Εκθετικής-Λογαριθμικής (GEL) κατανομής, είναι της μορφής:

$$F(x; \alpha, \lambda, p) = G_{EL}(x; \alpha, \lambda, p)^\alpha = \left(1 - \frac{\log\{1 - (1-p)e^{-\lambda x}\}}{\log p} \right)^\alpha, \quad x > 0,$$

όπου $\lambda > 0$, $p \in (0,1)$ και $\alpha > 0$.

Αν υπάρχουν n ανεξάρτητα εξαρτήματα σε ένα παράλληλο σύστημα και ο χρόνος ζωής αυτών ακολουθεί την κατανομή EL, τότε η διάρκεια ζωής του συστήματος μπορεί να περιγραφεί από τη κατανομή GEL.

4.4.2 Ορισμοί και χαρακτηριστικά της κατανομής

Η **αθροιστική συνάρτηση** κατανομής της X , δίνεται από τη σχέση:

$$F(x; \alpha, \lambda, p) = \left(1 - \frac{\log\{1 - (1-p)e^{-\lambda x}\}}{\log p} \right)^\alpha, \quad x > 0,$$

όπου $\lambda > 0$, $p \in (0,1)$ και $\alpha > 0$.

Επιπλέον, η **συνάρτηση πυκνότητας πιθανότητας** της, δίνεται από τη σχέση:

$$f(x; \alpha, \lambda, p) = \frac{\alpha \lambda (p-1) e^{-\lambda x}}{\{1 - (1-p)e^{-\lambda x}\} \log p} \left[1 - \frac{\log\{1 - (1-p)e^{-\lambda x}\}}{\log p} \right]^{\alpha-1}, \quad (4.1)$$

όπου $\lambda > 0$, $p \in (0,1)$ και $\alpha > 0$. Οι παράμετροι λ , α είναι παράμετροι κλίμακας και σχήματος, αντίστοιχα.

Για $\alpha = 1$ προκύπτει η Εκθετική-Λογαριθμική κατανομή και αν το p τείνει στο 1, τότε η (4.1) ταυτίζεται με τη συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής.

Οι *συναρτήσεις επιβίωσης* και *κινδύνου* της, είναι αντίστοιχα:

$$S(x; \alpha, \lambda, p) = 1 - \left(1 - \frac{\log \{1 - (1-p)e^{-\lambda x}\}}{\log p} \right)^\alpha, \quad x > 0,$$

και

$$h(x; \alpha, \lambda, p) = \frac{\frac{-\alpha\lambda}{\log p} \frac{(1-p)e^{-\lambda x}}{1 - (1-p)e^{-\lambda x}} \left[1 - \frac{\log \{1 - (1-p)e^{-\lambda x}\}}{\log p} \right]^{\alpha-1}}{1 - \left[1 - \frac{\log \{1 - (1-p)e^{-\lambda x}\}}{\log p} \right]^\alpha}, \quad x > 0.$$

Η συνάρτηση κινδύνου είναι γνησίως φθίνουσα, όταν $0 < \alpha \leq 1$ (Glaser, 1980), ενώ σε διαφορετική περίπτωση μπορεί να αποκτήσει και άλλες δυο διαφορετικές μορφές, μονοκόρυφη, εάν $\log \left(\frac{1-p}{1-pe^{\alpha-1}} \right)^{\frac{1}{\lambda}} > 0$, ή αύξουσα διαφορετικά.

Στην περίπτωση, όπου $X \sim GEP(\beta, p, \alpha)$, οι *ροπές r τάξης* δίνονται από τη σχέση:

$$E(X^r; \alpha, \lambda, p) = \frac{\alpha}{\lambda^r} \int_0^1 (1-x)^{\alpha-1} \log^r \left(\frac{1-p}{1-p^x} \right) dx.$$

Η διάμεσος υπολογίζεται να είναι:

$$\frac{1}{\lambda} \log \left\{ \frac{1-p}{1-p^{\left(\frac{1}{\alpha\sqrt{x}}\right)}} \right\}.$$

Η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής είναι:

$$\begin{aligned} m(x_0; \alpha, \lambda, p) &= E(X - x_0 | X \geq x_0; \alpha, \lambda, p) \\ &= \int_{x_0}^{+\infty} \frac{(x - x_0) f(x; \alpha, \lambda, p)}{1 - F(x_0; \alpha, \lambda, p)} dx \\ &= \frac{\alpha}{1 - (1 - A)^\alpha} \int_0^A \left\{ \frac{1}{\lambda} \log \left(\frac{1-p}{1-p^x} \right) - x_0 \right\} (1-x)^{\alpha-1} dx, \end{aligned}$$

$$\text{όπου } A = \frac{\log \{1 - (1-p)e^{-\lambda x_0}\}}{\log p}.$$

4.4.3 Εκτιμητική με τη μέθοδο της μέγιστης πιθανοφάνειας

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

$$\begin{aligned}
l(\theta; y_{obs}) &= -n \log(-\log p) + n \log \alpha + n \log \lambda + n \log(1-p) - \lambda \sum_{i=1}^n x_i \\
&\quad - \sum_{i=1}^n \left[\log \left\{ 1 - (1-p)e^{-\lambda x_i} \right\} \right] \\
&\quad + (\alpha + 1) \sum_{i=1}^n \left[\log \left[1 - \frac{\log \left\{ 1 - (1-p)e^{-\lambda x_i} \right\}}{\log p} \right] \right].
\end{aligned}$$

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:

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

$$\frac{\partial l}{\partial \lambda} = n\lambda^{-1} - \lambda \sum_{i=1}^n \left\{ \frac{x_i}{1 - (1-p)e^{-\lambda x_i}} \right\}$$

$$-(\alpha - 1) \sum_{i=1}^n \left[\frac{(1-p)x_i e^{-\lambda x_i}}{\left\{ 1 - (1-p)e^{-\lambda x_i} \right\} \log \left\{ \frac{p}{1 - (1-p)e^{-\lambda x_i}} \right\}} \right],$$

και

$$\frac{\partial l}{\partial p} = -n \left\{ (p \log p)^{-1} + (1-p)^{-1} \right\} - \sum_{i=1}^n \left\{ \frac{e^{-\lambda x_i}}{1 - (1-p)e^{-\lambda x_i}} \right\}$$

$$-(\alpha - 1) \sum_{i=1}^n \left[\frac{p \log p e^{-\lambda x_i} - \{1 - (1-p)e^{-\lambda x_i}\} \log \{1 - (1-p)e^{-\lambda x_i}\}}{p \log p \{1 - (1-p)e^{-\lambda x_i}\} \log \left\{ \frac{p}{1 - (1-p)e^{-\lambda x_i}} \right\}} \right].$$

Οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων $\theta = (\alpha, \lambda, p)$ προκύπτουν από την επίλυση του συστήματος των τριών μη-γραμμικών εξισώσεων πιθανοφάνειας

$$\frac{\partial l}{\partial \alpha} = 0, \quad \frac{\partial l}{\partial \lambda} = 0 \quad \text{και} \quad \frac{\partial l}{\partial p} = 0.$$

Είναι προφανές ότι η επίλυση του εν λόγω συστήματος, δύναται να επιτευχθεί μόνο με τη χρήση μιας επαναληπτικής αριθμητικής μεθόδου, όπως είναι οι μέθοδοι Newton-Raphson, quasi-Newton και Nelder-Mead.

4.4.4 Ασυμπτωτική κατανομή των εκτιμητών μέγιστης πιθανοφάνειας

Εφόσον οι εκτιμητές μέγιστης πιθανοφάνειας των άγνωστων παραμέτρων, $\hat{\theta}$, 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 θ και πίνακα διακυμάνσεων-συνδιακυμάνσεων τον αντίστροφο του αναμενόμενου πίνακα πληροφορίας $J(\theta) = E(I; \theta)$ (βλ. Lawless 2003),

όπου $I(\theta)$ είναι ο παρατηρούμενος πίνακας πληροφορίας με στοιχεία

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

Παραγωγίζοντας επιμέρους τις σχέσεις των μερικών παραγώγων μπορούν εύκολα να προκύψουν τα άνω τριγωνικά στοιχεία του, δευτέρας τάξεως, συμμετρικού, παρατηρούμενου πίνακα πληροφορίας $I(\theta)$.

Ο παραπάνω πίνακας $I(\theta)$ αποτελεί συνεπή εκτιμητή του $J(\theta)$ και μπορεί να χρησιμοποιηθεί για τη κατασκευή ασυμπτωτικών διαστημάτων εμπιστοσύνης για τις άγνωστες παραμέτρους.

ΠΑΡΑΡΤΗΜΑ

1) Δεδομένα που χρησιμοποιήθηκαν στο τρίτο κεφάλαιο

Πίνακας 1. Παρατηρήσεις που αφορούν στον αριθμό των διαδοχικών αποτυχιών του συστήματος κλιματισμού από ένα στόλο 13 αεροσκαφών (Adamidis and Loukas (1998), Kus (2007) και Tahmasbi and Rezaei (2008)).

194, 413, 90, 74, 55, 23, 97, 50, 359, 50, 130, 487, 102, 15, 14, 10, 57, 320, 261, 51, 44, 9, 254, 493, 18, 209, 41, 58, 60, 48, 56, 87, 11, 102, 12, 5, 100, 14, 29, 37, 186, 29, 104, 7, 4, 72, 270, 283, 7, 57, 33, 100, 61, 502, 220, 120, 141, 22, 603, 35, 98, 54, 181, 65, 49, 12, 239, 14, 18, 39, 3, 12, 5, 32, 9, 14, 70, 47, 62, 142, 3, 104, 85, 67, 169, 24, 21, 246, 47, 68, 15, 2, 91, 59, 447, 56, 29, 176, 225, 77, 197, 438, 43, 134, 184, 20, 386, 182, 71, 80, 188, 230, 152, 36, 79, 59, 33, 246, 1, 79, 3, 27, 201, 84, 27, 21, 16, 88, 130, 14, 118, 44, 15, 42, 106, 46, 230, 59, 153, 104, 20, 206, 5, 66, 34, 29, 26, 35, 5, 82, 5, 61, 31, 118, 326, 12, 54, 36, 34, 18, 25, 120, 31, 22, 18, 156, 11, 216, 139, 67, 310, 3, 46, 210, 57, 76, 14, 111, 97, 62, 26, 71, 39, 30, 7, 44, 11, 63, 23, 22, 23, 14, 18, 13, 34, 62, 11, 191, 14, 16, 18, 130, 90, 163, 208, 1, 24, 70, 16, 101, 52, 208, 95.

Πίνακας 2. Παρατηρήσεις που αναφέρονται στην περίοδο ανάμεσα σε διαδοχικούς σεισμούς στη ζώνη Βόρειας Ανατολίας (Kus (2007) και Tahmasbi and Rezaei (2008)).

1163, 501, 2039, 4863, 3258, 616, 217, 143, 323, 398, 9, 182, 159, 67, 633, 2117, 756, 896, 461, 3709, 409, 8592, 1821, 979.

Πίνακας 3. Ώρες διαστημάτων (σε μέρες) των διαδοχικών σεισμών (Tahmasbi and Rezaei (2008)).

136, 1187, 117, 944, 24, 70, 716, 1126, 378, 166, 152, 264, 275.

2) *Συναρτήσεις στην R γλώσσα προγραμματισμού* (Για το πρώτο σύνολο δεδομένων, το οποίο αφορά τον αριθμό των διαδοχικών αποτυχιών του συστήματος κλιματισμού του κάθε μέλους της από ένα στόλο από 13 αεροσκάφη)

```
plane=c(194, 413, 90, 74, 55, 23, 97, 50, 359, 50, 130, 487, 102, 15, 14,
10, 57, 320, 261, 51, 44, 9, 254, 493, 18, 209, 41, 58, 60, 48, 56, 87, 11,
102, 12, 5, 100, 14, 29, 37, 186, 29, 104, 7, 4, 72, 270, 283, 7, 57, 33,
100, 61, 502, 220, 120, 141, 22, 603, 35, 98, 54, 181, 65, 49, 12, 239, 14,
18, 39, 3, 12, 5, 32, 9, 14, 70, 47, 62, 142, 3, 104, 85, 67, 169, 24, 21,
246, 47, 68, 15, 2, 91, 59, 447, 56, 29, 176, 225, 77, 197, 438, 43, 134,
184, 20, 386, 182, 71, 80, 188, 230, 152, 36, 79, 59, 33, 246, 1, 79, 3, 27,
201, 84, 27, 21, 16, 88, 130, 14, 118, 44, 15, 42, 106, 46, 230, 59, 153,
104, 20, 206, 5, 66, 34, 29, 26, 35, 5, 82, 5, 61, 31, 118, 326, 12, 54, 36,
34, 18, 25, 120, 31, 22, 18, 156, 11, 216, 139, 67, 310, 3, 46, 210, 57, 76,
14, 111, 97, 62, 26, 71, 39, 30, 7, 44, 11, 63, 23, 22, 23, 14, 18, 13, 34,
62, 11, 191, 14, 16, 18, 130, 90, 163, 208, 1, 24, 70, 16, 101, 52, 208, 95)
```

```
plane=sort(plane)
library(stats4)
library(gregmisc)
library(maxLik)
```

```
##### EL #####
cdfel <- function(x, b, l){
  cdf = (1-(log(1-(1-b)*exp(-l*x))/log(b)))
  return(cdf)
}
pdfel <- function(x, b, l){
  n = length(x)
  pdf = ((-l*(1-b)*exp(-l*x))/(log(b)*(1-(1-b)*exp(-l*x))))
  return(pdf)
}
```



```
hrfel <- function(x, b, l){
  n = length(x)
  hrf = pdfel(x, b, l) / (1-cdfel(x, b, l))
  return(hrf)
}

llef <- function(b, l){
  x = plane
  n=length(x)
  n*log(-log(b))-n*log(l)-n*log(1-b)+l*sum(x)+sum(log(1-(1-b)*exp(-
l*x)))
}
#####

##### EG #####
cdfeg <- function(x, b, l){
  cdf = (1-exp(-l*x))/(1-b*exp(-l*x))
  return(cdf)
}

pdfeg <- function(x, b, l){
  pdf = (l*(1-b)*exp(-l*x))/((1-b*exp(-l*x))^2)
  return(pdf)
}

hrfeg <- function(x, b, l){
  hrf = pdfeg(x, b, l) / (1-cdfeg(x, b, l))
  return(hrf)
}

lleg <- function(b, l){
  x = plane
  n=length(x)
  -n*log(l*(1-b))+l*sum(x)+2*sum(log(1-b*exp(-l*x)))
```

```

}
#####

##### EB #####
cdfeb <- function(x, b, l){
  n = length(x)
  cdf = ((1+b*exp(-l*x))^n-(1+b)^n) / (1-(1+b)^n)
  return(cdf)
}

pdfeb <- function(x, b, l){
  n = length(x)
  pdf = (n*l*b*exp(-l*x)*((1+b*exp(-l*x))^(n-1))) / ((1+b)^n-1)
  return(pdf)
}

hrfeb <- function(x, b, l){
  hrf = pdfeb(x, b, l) / (1-cdfeb(x, b, l))
  return(hrf)
}

lleb <- function(b, l){
  x = plane
  n = length(x)
  m = n
  -n*log(m*b*l)+l*sum(x)+n*log((1+b)^n-1)-(m-1)*sum(log(1+b*exp(-
l*x)))
}
#####

##### EP #####

cdfep <- function(x, b, l){
  cdf = (1-exp(-l+exp(-b*x))) / (1-exp(-l))

```

```
return(cdf)
}

pdfep <- function(x, b, l){
pdf = l*b*exp(-l-b*x+l*exp(-b*x))/(1-exp(-l))
return(pdf)
}

hrfep <- function(x, b, l){
hrf = pdfep(x, b, l) / (1-cdfep(x, b, l))
return(hrf)
}

llep <- function(b, l){
x = plane
n=length(x)
-n*log(b*l)+n*l+b*sum(x)-l*sum(exp(-b*x))+n*log(1-exp(-l))
}
#####

##### FITTING #####
est <- mle(minuslog = llel, start = list(b=0.3874, l=0.0084))
summary(est)
ks.test(plane, cdfel, b=0.3874, l=0.0084)

est <- mle(minuslog = lleg, start = list(b=0.4276, l=0.008))
summary(est)
ks.test(plane, cdfeg, b=0.4276, l=0.008)

est <- mle(minuslog = lleb, start = list(b=0.00628, l=0.00751))
summary(est)
ks.test(plane, cdfeb, b=0.00628, l=0.00751)

est <- mle(minuslog = llep, start = list(b=0.00751, l=1.3324))
```

```

summary(est)
ks.test(plane, cdfep, b=0.00751, l=1.3324)
#####

## SURVIVAL PLOTS ###
esdf = function(x){
  x <- sort(x)
  n <- length(x)
  if (n < 1)stop("'x' must have 1 or more non-missing values")
  t=tabulate(match(x, x))
  if(length(t)<length(x))t=c(t, rep(0, (length(x)-length(t))))
  y = cumsum(t)/n
  plot(x, 1-y, xlab="X", ylab="Survival", main="Reliability curves",
  type="p", pch="*")
}

# Σύνολο δεδομένων από το αρχείο DATA: plane (Dhrf)
esdf(plane)
points(plane, 1-cdfel(plane, b=0.3874, l=0.0084), type='l', lty=1, col=4,
cex=0.5)
points(plane, 1-cdfeg(plane, b=0.4276, l=0.008), type='l', lty=1, col=3,
cex=0.5)
points(plane, 1-cdfeb(plane, b=0.00628, l=0.00751), type='l', lty=1,
col=2, cex=0.5)
points(plane, 1-cdfep(plane, b=0.00751, l=1.3324), type='l', lty=1, col=5,
cex=0.5)
#####

```

3) Συναρτήσεις στη MATHEMATICA**Η κατανομή EL**

```
dist=ProbabilityDistribution[(-Log[0.5]^(-
1)) (β(0.5)Exp[-β x])/(1-0.5Exp[-β x]), {x, 0,
Infinity}]
Plot[Evaluate@Table[PDF[dist, x], {β, {1, 2,
3}}], {x, 0, 3.5}, Filling→Axis, PlotRange→All]
```

```
dist=ProbabilityDistribution[(-Log[p]^(-1)) (1-
p)Exp[-x]/(1-(1-p)Exp[-x]), {x, 0, Infinity}]
Plot[Evaluate@Table[PDF[dist, x], {p, {0.2, 0.5,
0.8}}], {x, 0, 4}, Filling→Axis, PlotRange→All]
```

Η κατανομή EG

```
dist = ProbabilityDistribution[(1-p)Exp[-x] (1-p
Exp[-x])^(-2), {x, 0, Infinity}]
Plot[Evaluate@Table[PDF[dist, x], {p, {0.2, 0.5,
0.8}}], {x, 0, 4}, Filling→Axis]
```

```
dist=ProbabilityDistribution[β(0.5)Exp[-β x] (1-
0.5Exp[-β x])^(-2), {x, 0, +Infinity}]
Plot[Evaluate@Table[PDF [dist, x], {β, {1, 2,
3}}], {x, 0, 4}, Filling→Axis]
```

Η κατανομή EP

```
dist=ProbabilityDistribution[β Exp[-1-β x+Exp[-β
x]]/(1-Exp[-1]), {x, 0, Infinity}]
Plot[Evaluate@Table[PDF[dist, x], {β, {1, 2,
3}}], {x, 0, 4}, Filling->Axis, PlotRange ->All]
```

```
dist=ProbabilityDistribution[λ Exp[-λ-x+λ Exp[-
x]]/(1-Exp[-λ]), {x, 0, Infinity}]
Plot[Evaluate@Table[PDF[dist, x], {λ, {1, 2,
3}}], {x, 0, 4}, Filling->Axis, PlotRange ->All]
```


ΠΕΡΙΛΗΨΗ

Τίτλος: Μίξεις Εκθετικής Κατανομής με Διακριτές Κατανομές σε μοντέλα επιβίωσης.

Συγγραφέας: Αικατερίνη Βασιλειάδου

Επιβλέπων: Σωτήριος Λουκάς

Με τον όρο ανάλυση επιβίωσης εννοούμε το σύνολο των στατιστικών μεθοδολογιών που έχουν ως στόχο τη περιγραφή και τη μοντελοποίηση της διάρκειας ενός φαινομένου ή συμβάντος στο χρόνο. Ο κύριος στόχος της διατριβής είναι η μελέτη των μοντέλων που προκύπτουν ως Μίξεις Εκθετικής Κατανομής με Διακριτές Κατανομές (Διωνυμική, Poisson, Γεωμετρική και Λογαριθμική) σε μοντέλα επιβίωσης.

Στο Κεφάλαιο 1, παρουσιάζονται κάποιες βασικές συναρτήσεις των μοντέλων επιβίωσης και η σημασία της συνάρτησης κινδύνου. Επίσης, παρουσιάζονται οι πιο δημοφιλείς κατανομές χρόνων ζωής. Επιπλέον, παρουσιάζονται η μέθοδος μέγιστης πιθανοφάνειας και ο EM αλγόριθμος.

Στο Κεφάλαιο 2, ελέγχεται η φυσική ερμηνεία και τα χαρακτηριστικά τεσσάρων κατανομών (Εκθετική-Γεωμετρική, Εκθετική-Poisson, Εκθετική-Λογαριθμική και Εκθετική-Δυναμοσειρά) με φθίνουσα συνάρτηση κινδύνου. Οι κατανομές αυτές μελετήθηκαν από τους Adamidis and Loukas (1998), Kus (2007), Tahmasbi and Rezaei (2008) and Chahkandi and Ganjali (2009), αντίστοιχα. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας. Παρουσιάζεται ο EM αλγόριθμος και οι πίνακες παρατηρούμενης και αναμενόμενης πληροφορίας.

Στο Κεφάλαιο 3, ελέγχεται η προσαρμογή αυτών των κατανομών σε πραγματικά δεδομένα.

Στο Κεφάλαιο 4, ελέγχεται η φυσική ερμηνεία και τα χαρακτηριστικά τριών κατανομών (Γενικευμένη Εκθετική-Γεωμετρική, Γενικευμένη Εκθετική-Poisson, Γενικευμένη Εκθετική- Λογαριθμική) με αύξουσα, φθίνουσα, κυπελλοειδούς και ανάποδη της κυπελλοειδούς συνάρτησης κινδύνου. Αυτές οι κατανομές μελετήθηκαν από τους Barreto-Souza and Cribari-Neto (2009), Silva and Barreto-Souza (2010) και Παππάς, Αδαμίδης και Λουκάς (2012), αντίστοιχα. Το πρόβλημα της εκτίμησης των παραμέτρων, αντιμετωπίζεται με τη μέθοδο της μέγιστης πιθανοφάνειας.

ABSTRACT OF THE DISSERTATION

Title: Mixtures of Exponential Distribution with Discrete Distributions in survival models.

Author: Aikaterini Vasiliadou

Supervisor: Sotirios Loukas

Survival analysis is a collection of statistical procedures for the analysis of data, where the response of interest is the time until an event occurs. The main goal of this dissertation is to study the models arising from Mixing Exponential with Discrete Distributions (Binomial, Poisson, Geometric and Logarithmic) in survival models.

Chapter 1 presents the basic distributions that are used in survival analysis and the importance of the hazard function. In the sequel, several popular lifetime distributions are presented. Furthermore, the method of maximum likelihood and EM algorithm are derived.

Chapter 2 examines the natural interpretation and the properties of four distributions (EG, EP, EL, EPS) with decreasing failure rate (D.F.R.). The above distributions with D.F.R. were studied by Adamidis and Loukas (1998), Kus (2007), Tahmasbi and Rezaei (2008) and Chahkandi and Ganjali (2009), respectively. The estimation of parameters is studied by the method of maximum likelihood. An EM algorithm and expressions for their asymptotic variances and covariances are derived.

Chapter 3 examines the adjustment of these distributions to real datasets.

Chapter 4 examines the natural interpretation and the properties of three distributions (GEP, GEG, GEL) with increasing, decreasing, bathtub and upside down bathtub shaped failure rates. These distributions were studied by Barreto-Souza and Cribari-Neto (2009), Silva and Barreto-Souza (2010) and Pappas, Adamidis and Loukas (2012), respectively. The maximum likelihood method is used for the parameter estimation problem.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Adamidis, K. (1999). An EM algorithm for estimating negative binomial parameters. *Austral.New Zealand J. Statist.* 41(2), 213-221.
- Adamidis, K., Loukas, S. (1998). A life time distribution with decreasing failure rate. *Statist. Probab. Lett.* 39, 35-42.
- Adamidis, K., Dimitrakopoulou, T., Loukas, S. (2005). On the extension of the exponential-geometric distribution. *Statist. Probab. Lett.* 73, 259-269.
- Alkarni, S and Oraby, A. (2012). A compound class of Poisson and lifetime distributions. *J Stat. Appl. Pro.* 1 (1), 45-51.
- Barlow, R.E., Proschan, F., Marshall, A.W. (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.* 34, 375-389.
- Barlow, R.E. and Proschan, F. (1965). *Mathematical theory of reliability and life testing*. Holt, Rinehart and Winston: New York.
- Barlow, R.E. and Proschan, F. (1965). *Mathematical theory of reliability*. John Wiley and Sons, New York.
- Barlow, R.E., Marshall, A.W. (1964). Bounds for distributions with monotone hazard rate I and II. *Ann. Math. Statist.* 35, 1234-1274.
- Barreto-Souza, W., Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. *Statistics and Probability Letters*, 79, 2493-2500.
- Barreto-Souza, W., B. Silva, R. and M. Cordeiro, G. (2010). A new distribution with decreasing, increasing and upside-down bathtub failure rate, *Computational Statistics and Data Analysis*, 54, 935-944.

- Block, H.W., Y.Li, and Savits T.H. (2003). Initial and final behavior of failure rate functions for mixtures and systems. *Journal of Applied Probability*, 40, 721-740.
- Block, H.W., Savits, T.H., Wondmagegnehu, E.T. (2003). Mixtures of distributions with increasing linear failure rates. *J. Appl.Prob.*, 40, 485-504.
- Chahkandi, M. and Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, 53, 4433-4440.
- Cozzolino, J.M. (1968). Probabilistic models of decreasing failure rate processes. *Naval Research Logistics Quarterly*, 15, 361-374.
- Dahiya, R. C. and Gurland, J. (1972). Goodness of t-tests for gamma and exponential distributions. *Technometrics*, 14, 791-801.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.
- Gleser, L.J. (1989). The Gamma distribution as a mixture of exponential distributions. *The American Statistician*, 43, 115-117.
- Gupta, R.D., Kundu, D., (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics* 41, 173-188.
- Gurland, J., Sethuraman, J. (1994). Reversal of increasing failure rates when pooling failure rate data. *Technometrics*, 36, 416-418.
- Johnson, N. L., Kotz, S., Balakrishnan, N. (1993). *Univariate Discrete Distributions*. John Wiley and Sons. INC, New York.
- Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics*, 30(1), 63-77.

- Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*, 51, 4497-4509.
- Lawless, J.F. (1982). *Statistical models and methods for lifetime data*. John Wiley and Sons, New York.
- Leemis, L.N. (1995). *Reliability: Probabilistic models and statistical methods*. Prentice-Hall, Inc.
- Lomax, K.S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49, 847-852.
- Marshall, A.W., Proschan, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *The Annals of Mathematical Statistics*, 36, 69-77.
- McLachlan, G.J., Krishnan, T. (1997). *The EM Algorithm and Extension*. Wiley, New York.
- McNolty, F., Doyle, J., Hansen, E. (1980). Properties of the mixed exponential failure process. *Technometrics* 22, 555-565.
- Ng, H.K.T., Chan, P.S., Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using EM algorithm. *Computational Statistics and Data Analysis*, 371-376.
- Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* 5, 375-383.
- Rajarshi, S. and Rajarshi, M.B. (1988). Bathtub distribution: A review. *Commun. Statist.-Theory Meth.*, 17(8), 2597-2621.
- Saunders, S.C., Myhre, J.M. (1983). Maximum likelihood estimation for two-parameter decreasing hazard rate distributions using censored data. *Journal of the American Statistical Association*. 78, 664-673.

- Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scand J Statist*, 1, 49-58.
- Sundberg, R. (1976). An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Commun. Statist.-Simula. Computa.*, B 5(1), 55-64.
- Tahmasbi, R., Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, 52, 3889-3901.
- Δημάκη, Κ. (2006). Ανάλυση Επιβίωσης. *Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών*.
- Κολυβά, Μ. (1998). Μαθηματική Στατιστική. *Τόμος 1, Εκδόσεις Ζήτη*.
- Παπαϊωάννου, Τ. και Φερεντίνου Κ. (2000). Μαθηματική Στατιστική. *Β' Έκδοση, Εκδόσεις Αθ.Σταμούλης*.
- Παπάς, Β.Α., Αδαμίδης, Κ. και Λουκάς, Σ. (2012). Μία γενίκευση της εκθετικής-λογαριθμικής κατανομής, 25^ο Συνέδριο Ε.Σ.Ι., Βόλος, 1-10.