

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

Η Lindley κατανομή και οι Εφαρμογές της στον
Αναλογισμό

ΚΩΝΣΤΑΝΤΙΝΟΣ-ΠΑΝΑΓΙΩΤΗΣ
ΤΑΧΛΙΑΜΠΟΥΡΗΣ

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

Πειραιάς
Ιούλιος 2018

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

Τα μέλη της Επιτροπής ήταν:

- Αντζουλάκος Δημήτριος (Επιβλέπων)
- Πολίτης Κωνσταντίνος
- Τζαβέλας Γεώργιος

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND INSURANCE
SCIENCE

POSTGRADUATE PROGRAM IN ACTUARIAL
SCIENCE AND RISK MANAGEMENT

Lindley distribution and its Application in Actuarial
Science

KONSTANTINOS-PANAGIOTIS
TACHLIAMPOURIS

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

Piraeus
July 2018

Αφιερωμένο στην μητέρα μου Σοφία.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον αναπληρωτή καθηγητή Αντζουλάκο Δημήτριο, ειδικότερα όσον αφορά την πολύτιμη βοήθειά του στην ολοκλήρωση της διπλωματικής μου εργασίας, όπως και τους επιβλέποντες Μπούτσικα Μιχαήλ και Τζαβέλα Γεώργιο. Τέλος, θα ήθελα να ευχαριστήσω όλους τους διδάσκοντες είτε μέλη ΔΕΠ είτε εξωτερικούς, για την πολύτιμη γνώση που μας μετέδωσαν.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΕΧΟΜΕΝΑ	7
ΕΙΣΑΓΩΓΗ.....	10
ABSTRACT	11
ΚΕΦΑΛΑΙΟ 1.....	1
Η κατανομή Lindley.....	1
1.1 Εισαγωγή.....	1
1.2 Γέννηση της κατανομής Lindley	1
1.3 Συνάρτηση πυκνότητας	2
Σχήμα 1-1	3
Σχήμα 1-2	3
1.4 Επικρατούσα τιμή, διάμεσος και μέση τιμή	4
Θεώρημα 1.1	4
1.5 Ροπές, λοξότητα και κύρτωση.....	5
Σχήμα 1-3	8
1.6 Χαρακτηριστική συνάρτηση	8
1.7 Ρυθμός αποτυχίας και συνάρτηση μέσης υπολειπόμενης ζωής.....	9
Σχήμα 1-4	10
Σχήμα 1-5	11
1.8 Μέσες αποκλίσεις	12
1.9 Εκτιμητής μέγιστης πιθανοφάνειας.....	13
Θεώρημα 1.2	14
Θεώρημα 1.3	15
1.10 Μελέτη προσομοίωσης.....	16
Πίνακας 1-1	17
Πίνακας 1-2.....	17
Πίνακας 1-3.....	18
1.11 Προσαρμογή σε δεδομένα	18
Πίνακας 1-4.....	19
Σχήμα 1-6	19
Σχήμα 1-7	20
ΚΕΦΑΛΑΙΟ 2.....	21
Κατανομή Lindley και ανταγωνιστικοί κίνδυνοι	21
2.1 Εισαγωγή.....	21
2.2 Μοντέλο ανταγωνιστικών κινδύνων για την κατανομή Lindley	22

Πίνακας 2-1	25
Πίνακας 2-2	26
Πίνακας 2-3	26
ΚΕΦΑΛΑΙΟ 3.....	29
Η γενικευμένη κατανομή Lindley	29
3.1 Εισαγωγή.....	29
3.2 Ορισμός και βασικές ιδιότητες της γενικευμένης κατανομής Lindley	29
3.2.1 Η γενίκευση της κατανομής	29
3.2.2 Το σχήμα της κατανομής	30
Σχήμα 3-1	31
Σχήμα 3-2	32
Σχήμα 3-3	32
Πρόταση 3.1	33
3.2.3 Κατανομή αθροίσματος	34
Πρόταση 3.2	34
3.2.4 Παραγωγή τυχαίου δείγματος	35
3.3 Εκτίμηση.....	36
3.3.1 Εκτιμητές μέγιστης πιθανοφάνειας	36
3.3.2 Προσαρμογή σε δεδομένα	38
Πίνακας 3-1	38
Πίνακας 3-2	38
Πίνακας 3-3	39
ΚΕΦΑΛΑΙΟ 4.....	41
Η γενικευμένη Poisson-Lindley κατανομή	41
4.1 Εισαγωγή.....	41
4.2 Η γενικευμένη Poisson-Lindley κατανομή	42
Θεώρημα 4.1	42
Σχήμα 4-1	45
4.3 Ιδιότητες της γενικευμένης Poisson-Lindley κατανομής.....	46
4.3.1 Σχήμα.....	46
4.3.2 Γεννήτριες συναρτήσεις.....	46
4.3.3 Ροπές, λοξότητα και κύρτωση.....	47
Πίνακας 4-1	48
Πίνακας 4-2	48
Σχήμα 4-2	49
Σχήμα 4-3	49
4.4 Αλγόριθμος παραγωγής τυχαίου δείγματος	50

4.5 Εκτίμηση	51
4.5.1 Μέθοδος των ροπών	51
Πρόταση 4.2	51
Πρόταση 4.3	52
4.5.2 Μέθοδος μέγιστης πιθανοφάνειας.....	53
4.6 Το συλλογικό μοντέλο κινδύνου	54
Πρόταση 4.4	54
Πρόταση 4.5	55
4.7 Εφαρμογή σε πραγματικά δεδομένα	56
Σύνολο δεδομένων 1	56
Πίνακας 4-3	57
Σύνολο δεδομένων 2	57
Πίνακας 4-4	58
ΚΕΦΑΛΑΙΟ 5.....	59
Συμπεράσματα	59
ΠΑΡΑΡΤΗΜΑ.....	62
ΚΕΦΑΛΑΙΟ 1.....	62
ΚΕΦΑΛΑΙΟ 3.....	65
ΚΕΦΑΛΑΙΟ 4.....	68
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	71
ΕΛΛΗΝΙΚΗ.....	71
ΞΕΝΗ	71

ΕΙΣΑΓΩΓΗ

Στην διπλωματική εργασία που θα ακολουθήσει, εξετάζουμε την κατανομή Lindley και κάποιες παραλλαγές της. Η κατανομή Lindley εισήχθη από τον Lindley (1958, 1965) και πρόκειται για κατανομή που προέρχεται από την οικογένεια των Εκθετικών κατανομών. Στο Κεφάλαιο 1, γίνεται η γνωριμία με την απλή μορφή της κατανομής (συνάρτηση πυκνότητας, κλπ.) ενώ γίνεται σύγκριση με την εκθετική κατανομή ως προς την εκτίμηση της παραμέτρου θ αλλά και της προσαρμογής των κατανομών σε τυχαία δεδομένα. Τα αποτελέσματα δείχνουν ξεκάθαρα την καλύτερη προσαρμογή της Lindley κατανομής έναντι της εκθετικής. Στο Κεφάλαιο 2, εξετάζεται το μοντέλο των πολλαπλών αιτιών εξόδου που εφαρμόζεται στην Αναλογιστική επιστήμη. Δίχως εξαίρεση της γενικότητας, θα δούμε τη συμπεριφορά της Lindley κατανομής στο μοντέλο με 2 αίτια εξόδου και θα επιχειρήσουμε να εκτιμήσουμε τις παραμέτρους αυτής. Στο Κεφάλαιο 3, θα εξετάσουμε τη μορφή και τα χαρακτηριστικά στοιχεία της Γενικευμένης Lindley κατανομής με τρεις παραμέτρους (GL), η οποία παρουσιάζει μεγαλύτερη ευελιξία σε σχέση με την $Lindley(\theta)$ (Zakerzadeh (2009)). Ειδικές περιπτώσεις της GL αποτελεί η εκθετική και η γάμμα κατανομή. Μέσω ενός αλγορίθμου παραγωγής τυχαίων δειγμάτων, θα δούμε δύο παραδείγματα με διαφορετικά δεδομένα στα οποία η GL έχει καλύτερη προσαρμογή από την γάμμα, την Weibull και την Lognormal κατανομή. Τέλος, στο Κεφάλαιο 4 γίνεται η παρουσίαση μίας νέας γενικευμένης Poisson-Lindley κατανομής, η οποία πηγάζει από την Poisson κατανομή όταν η παράμετρος λ ακολουθεί τη Lindley κατανομή με δύο παραμέτρους την οποία συμβολίζουμε $TPLD(\theta, a)$ (Shanker (2013)). Η ανάγκη για τη δημιουργία ενός τέτοιου μοντέλου προκύπτει από την ασφαλιστική αγορά και για τον υπολογισμό των ασφαλιστρών, όπου υιοθετείται μία κατανομή για τη ζημία η οποία μπορεί να επαληθευθεί ή όχι. Συγκεκριμένα, πέρα από την ελαχιστοποίηση της απόκλισης από τα πραγματικά δεδομένα, η νέα αυτή κατανομή μπορεί να περιγράψει τις ακραίες τιμές («βαριά ουρά») (Shanker (2013)). Γίνεται αναφορά στην συνάρτηση κατανομής και σε χαρακτηριστικά στοιχεία της όπως το σχήμα, η λοξότητα και η κύρτωση ενώ γίνεται σύγκριση με άλλες κατανομές ως προς την εφαρμογή τους. Τα αποτελέσματα δείχνουν την καλύτερη προσαρμογή της $NGPL(\theta, a)$ συγκριτικά με άλλες ευρέως χρησιμοποιούμενες κατανομές στην Αναλογιστική επιστήμη.

ABSTRACT

The Diploma Thesis that will follow, we will examine the Lindley Distribution and variations of it. The Lindley Distribution was introduced by Lindley (1958, 1965) and belongs to the Exponential Distributions family. In Chapter 1, we will get acquainted with the simple form of the Distribution (Density function, etc.) while compared to the Exponential Distribution with regard to the assessment of the θ parameter but also adapting the Distribution to data. The results clearly indicate Lindley's better adaptation against the Exponential Distribution. In Chapter 2, the competing risks life time data model is examined applied in the Actuarial science. Without excluding generality, we shall observe the Lindley Distribution behavior with 2 causes of competing failures and we shall attempt to evaluate said parameters. In Chapter 3, we will examine the form and parameters of the Generalized Lindley Distribution with three parameters (*GL*), which presents more flexibility than *Lindley*(θ) (Zakerzadeh (2009)). Special cases of the *GL* is the Exponential and the Gamma Distributions. Through an algorithm of random samples production, we shall examine two examples with different Data in which *GL* has better application than Gamma, Weibull and Lognormal Distributions. In conclusion Chapter 4 presents a new Generalized Poisson-Lindley Distribution, which derives from the Poisson Distribution when the λ parameter follows the Lindley Distribution with two parameters which we symbolize as *TPLD*(θ, a) (Shanker (2013)). The need for the creation of such model derives from the Insurance Market and the Premium rates calculation, where a Distribution referencing potential Liability is adopted which can be verified or not. Specifically, except from minimizing the deviation from the actual Data, this new Distribution can describe the extreme values («heavy tail») (Shanker (2013)). Reference is made to the allocation Distribution and its general characteristics such as shape, skewness and kurtosis while a comparison is made with other Distributions as per their application. The results show the better adaptation of the *NGPL*(θ, a) in comparison to other widely used Distributions in the Actuarial Science.

ΚΕΦΑΛΑΙΟ 1

Η κατανομή Lindley

1.1 Εισαγωγή

Η κατανομή Lindley (συμβ. $Lindley(\theta)$), που εισήχθη από τον Lindley (Lindley (1958, 1965)), έχει συνάρτηση πυκνότητας (σ.π.) που δίνεται από τον τύπο

$$f(x) = \frac{\theta^2}{\theta + 1} (1 + x)e^{-\theta x}, \quad x > 0, \theta > 0,$$

και συνάρτηση κατανομής

$$F(x) = 1 - \frac{\theta + 1 + \theta x}{\theta + 1} e^{-\theta x}, \quad x > 0, \theta > 0.$$

Η εκθετική κατανομή (συμβ. $Exp(\theta)$) με συνάρτηση πυκνότητας που δίνεται από τον τύπο

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

είναι μια πυκνότητα που μοιάζει με αυτή της κατανομής Lindley. Λόγω του γεγονότος πως η εκθετική κατανομή είναι δημοφιλέστερη, η Lindley κατανομή και οι ιδιότητες της δεν έχουν γίνει ιδιαίτερα γνωστές. Στο παρόν κεφάλαιο θα παρουσιαστούν οι βασικές ιδιότητες της κατανομής Lindley, και θα διαπιστώσουμε την καλύτερη προσαρμογή και την ακριβέστερη εκτίμηση που παρέχει σε πραγματικά δεδομένα έναντι της εκθετικής κατανομής. Η μελέτη αυτών των αποτελεσμάτων βρίσκονται στην εργασία των Ghitany, Atieh & Nadarajah (2007).

1.2 Γέννηση της κατανομής Lindley

Έστω X_1 μια συνεχής τυχαία μεταβλητή που έχει την εκθετική κατανομή με παράμετρο θ ($X_1 \sim Exp(\theta)$) και X_2 μια άλλη συνεχής τυχαία μεταβλητή που έχει την κατανομή γάμμα με

παράμετρο μορφής $a = 2$ και παράμετρο κλίμακας θ ($X_2 \sim \text{Gamma}(2, \theta)$). Τότε η τυχαία μεταβλητή X με συνάρτηση πυκνότητας

$$\begin{aligned} f_X(x) &= \frac{\theta}{\theta + 1} f_{X_1}(x) + \left(1 - \frac{\theta}{\theta + 1}\right) f_{X_2}(x) \\ &= \frac{\theta}{\theta + 1} \theta e^{-\theta x} + \left(1 - \frac{\theta}{\theta + 1}\right) \theta^2 x e^{-\theta x} = \frac{\theta^2}{\theta + 1} (1 + x) e^{-\theta x} \end{aligned}$$

έχει την κατανομή Lindley με παράμετρο θ ($X \sim \text{Lindley}(\theta)$). Επομένως η κατανομή Lindley αποτελεί μίξη μιας εκθετικής και μιας γάμμα κατανομής. Από την παραπάνω γέννηση της κατανομής Lindley προκύπτει ότι η συνάρτηση κατανομής της δίνεται από τον τύπο

$$\begin{aligned} F_X(x) &= \frac{\theta}{\theta + 1} F_{X_1}(x) + \left(1 - \frac{\theta}{\theta + 1}\right) F_{X_2}(x) \\ &= \frac{\theta}{\theta + 1} (1 - e^{-\theta x}) + \left(1 - \frac{\theta}{\theta + 1}\right) (1 - e^{-\theta x} (1 + \theta x)) \\ &= 1 - \frac{\theta + 1 + \theta x}{\theta + 1} e^{-\theta x}, \quad x > 0. \end{aligned}$$

1.3 Συνάρτηση πυκνότητας

Η πρώτη παράγωγος της σ.π. της κατανομής Lindley είναι

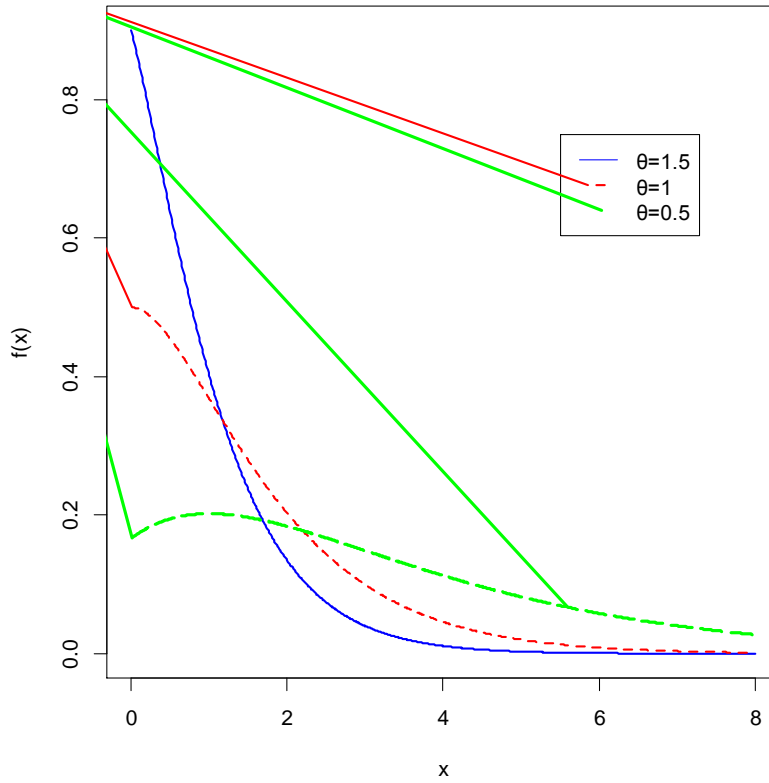
$$\frac{d}{dx} f(x) = f'(x) = \frac{\theta^2}{\theta + 1} (1 - \theta - \theta x) e^{-\theta x}.$$

Συνεπώς

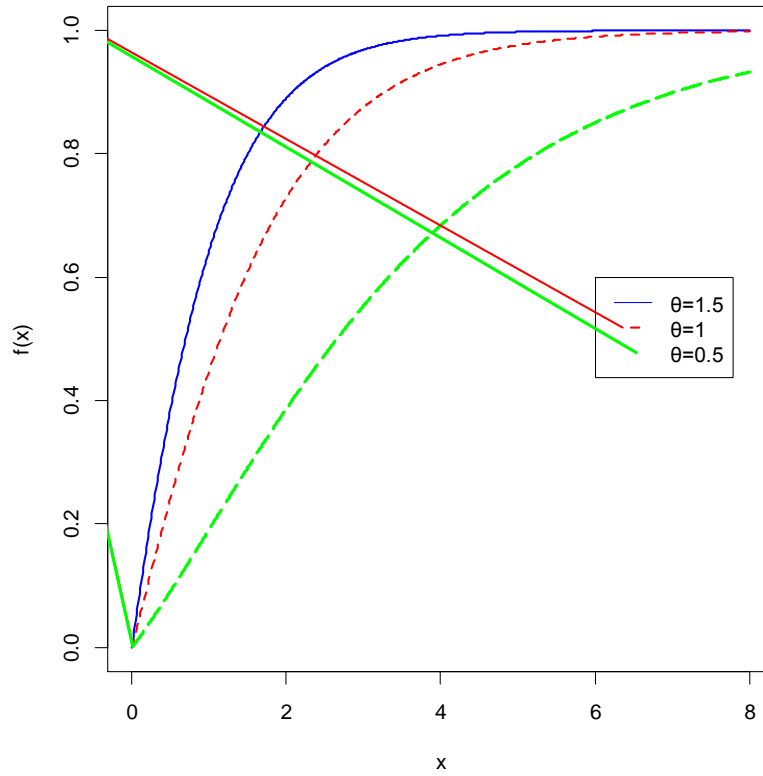
- για $\theta < 1$ έχουμε ότι η λύση της εξίσωσης $f'(x) = 0$ είναι η $x_0 = \frac{1-\theta}{\theta}$ όπου και μεγιστοποιείται η $f(x)$, και
- για $\theta \geq 1$ έχουμε ότι $f'(x) \leq 0$ που σημαίνει ότι η $f(x)$ είναι φθίνουσα.

Στα ακόλουθα σχήματα δίνονται διάφορες γραφικές παραστάσεις της συνάρτησης πυκνότητας και της συνάρτησης κατανομής της κατανομής Lindley για διάφορες τιμές της παραμέτρου θ .

Σχήμα 1-1
Συνάρτηση πυκνότητας της κατανομής Lindley



Σχήμα 1-2
Συνάρτηση κατανομής της κατανομής Lindley



Παρατηρούμε πως η επικρατούσα τιμή (ή κορυφή) της κατανομής Lindley μπορεί να πάρει διάφορες τιμές ανάλογα με τις τιμές της παραμέτρου θ , κάτι που δεν συμβαίνει με την εκθετική κατανομή όπου η επικρατούσα τιμή είναι πάντα στο 0.

1.4 Επικρατούσα τιμή, διάμεσος και μέση τιμή

Έστω X μια τ.μ. που έχει την κατανομή Lindley με παράμετρο θ . Τότε για την επικρατούσα τιμή της, $M = mode(X)$, τη διάμεσό της, $m = median(X)$, και τη μέση τιμή της, $\mu = E(X)$, ισχύει το εξής θεώρημα (δείτε Abadir (2005)).

Θεώρημα 1.1

Έστω X μια τ.μ. που έχει την κατανομή Lindley με παράμετρο θ Τότε

$$M < m < \mu.$$

Απόδειξη

Για τη μέση τιμή της X έχουμε

$$\begin{aligned} E(X) &= \int_0^{+\infty} xf(x)dx = \int_0^{+\infty} x \frac{\theta^2}{\theta+1} (1+x)e^{-\theta x} dx \\ &= \frac{\theta^2}{\theta+1} \left(\int_0^{+\infty} xe^{-\theta x} dx + \int_0^{+\infty} x^2 e^{-\theta x} dx \right) = \frac{\theta^2}{\theta+1} (I_1 + I_2) \end{aligned}$$

όπου

$$I_1 = \int_0^{+\infty} x e^{-\theta x} dx = \left[x \frac{e^{-\theta x}}{-\theta} \right]_0^{+\infty} + \int_0^{+\infty} \frac{e^{-\theta x}}{\theta} dx = \frac{1}{\theta} \left[\frac{e^{-\theta x}}{-\theta} \right]_0^{+\infty} = \frac{1}{\theta^2},$$

$$I_2 = \int_0^{+\infty} x^2 e^{-\theta x} dx = \left[x^2 \frac{e^{-\theta x}}{-\theta} \right]_0^{+\infty} + \frac{2}{\theta} \int_0^{+\infty} x e^{-\theta x} dx = \frac{2}{\theta} I_1 = \frac{2}{\theta^3}.$$

Επομένως

$$\mu = E(X) = \frac{\theta^2}{\theta+1} (I_1 + I_2) = \frac{\theta^2}{\theta+1} \left(\frac{1}{\theta^2} + \frac{2}{\theta^3} \right) = \frac{\theta+2}{\theta(\theta+1)}.$$

Για την επικρατούσα τιμή έχουμε ήδη αναφέρει ότι εμφανίζεται στο σημείο $x_0 = \frac{1-\theta}{\theta}$ όταν $\theta < 1$, και στο σημείο 0 όταν $\theta \geq 1$.

Συνεπώς ισχύουν οι σχέσεις

- $F(M) = \begin{cases} 1 - \frac{2}{\theta + 1} e^{-(1-\theta)}, & \theta < 1 \\ 0, & \theta \geq 1 \end{cases}$
- $F(m) = \frac{1}{2}$
- $F(\mu) = 1 - \frac{\theta^2 + 3\theta + 3}{(\theta + 1)^2} e^{-(\theta+2)/(\theta+1)}$

Παρατηρούμε πως η $F(M)$ είναι φθίνουσα συνάρτηση για $\theta \in (0,1)$, ενώ για κάθε $\theta > 0$, ισχύει ότι

$$0 \leq F(M) < 1 - 2e^{-1} < \frac{1}{2}.$$

Επίσης, η $F(\mu)$ είναι αύξουσα συνάρτηση για $\theta > 0$ και

$$\frac{1}{2} < 1 - 3e^{-2} < F(\mu) < 1.$$

Τέλος, λόγω του ότι η $F(x)$ είναι αύξουσα συνάρτηση για κάθε $x > 0$ και για κάθε $\theta > 0$, παίρνουμε ότι

$$M < m < \mu. \quad \blacksquare$$

Αξίζει να σημειώσουμε ότι το παραπάνω θεώρημα ισχύει και για την εκθετική κατανομή.

1.5 Ροπές, λοξότητα και κύρτωση

Για τη ροπή r -τάξης της κατανομής Lindley έχουμε

$$\begin{aligned} E(X^r) &= \int_0^{+\infty} x^r f(x) dx = \int_0^{+\infty} x^r \frac{\theta^2}{\theta + 1} (1+x) e^{-\theta x} dx \\ &= \frac{\theta^2}{\theta + 1} \left(\int_0^{+\infty} x^r e^{-\theta x} dx + \int_0^{+\infty} x^{r+1} e^{-\theta x} dx \right) = \frac{\theta^2}{\theta + 1} (I_r + I_{r+1}) \end{aligned}$$

όπου

$$I_r = \int_0^{+\infty} x^r e^{-\theta x} dx.$$

Όμως

$$I_r = \int_0^{+\infty} x^r e^{-\theta x} dx = \frac{1}{\theta^{r+1}} \int_0^{+\infty} (\theta x)^r e^{-\theta x} d(\theta x) = \frac{1}{\theta^{r+1}} \int_0^{+\infty} y^r e^{-y} dy = \frac{\Gamma(r+1)}{\theta^{r+1}} = \frac{r!}{\theta^{r+1}}.$$

Έτσι η ροπή r -τάξης της κατανομής Lindley δίνεται από τη σχέση

$$\mu'_r = E(X^r) = \frac{r! (\theta + r + 1)}{\theta^r (\theta + 1)}, \quad r = 1, 2, 3, \dots$$

Για $r = 1, 2, 3, 4$ προκύπτει ότι

$$\mu'_1 = \frac{\theta + 2}{\theta(\theta + 1)} = \mu,$$

$$\mu'_2 = \frac{2(\theta + 3)}{\theta^2(\theta + 1)},$$

$$\mu'_3 = \frac{6(\theta + 4)}{\theta^3(\theta + 1)},$$

$$\mu'_4 = \frac{24(\theta + 5)}{\theta^4(\theta + 1)}.$$

Σημειώνουμε πως η ροπή r -τάξης της εκθετικής κατανομής είναι $\mu'_r = \frac{r!}{\theta^r}$.

Η κεντρική ροπή k -τάξης της κατανομής Lindley προκύπτει από τον τύπο

$$\mu_k = E[(X - \mu)^k] = \sum_{r=0}^k \binom{k}{r} \mu'_r (-\mu)^{k-r}.$$

Για $r = 2, 3, 4$ προκύπτει ότι

$$\mu_2 = \frac{\theta^2 + 4\theta + 2}{\theta^2(\theta + 1)^2} = \sigma^2,$$

$$\mu_3 = \frac{2(\theta^3 + 6\theta^2 + 6\theta + 2)}{\theta^3(\theta + 1)^3},$$

$$\mu_4 = \frac{3(3\theta^4 + 24\theta^3 + 44\theta^2 + 32\theta + 8)}{\theta^4(\theta + 1)^4}.$$

Ο συντελεστής μεταβλητότητας (γ), ο συντελεστής λοξότητα ($\sqrt{\beta_1}$) και ο συντελεστής κύρτωσης (β_2) είναι

$$\gamma = \frac{\sigma}{\mu} = \frac{\sqrt{\theta^2 + 4\theta + 2}}{\theta + 2},$$

$$\sqrt{\beta_1} = \frac{\mu_3}{(\sigma^2)^{3/2}} = \frac{2(\theta^3 + 6\theta^2 + 6\theta + 2)}{(\theta^2 + 4\theta + 2)^{3/2}},$$

$$\beta_2 = \frac{3(3\theta^4 + 24\theta^3 + 44\theta^2 + 32\theta + 8)}{(\theta^2 + 4\theta + 2)^2}.$$

Από τους παραπάνω τύπους προκύπτει ότι (δείτε ΣΧΗΜΑ 1-3):

- (i) Ο συντελεστής μεταβλητότητας γ είναι αύξουσα συνάρτηση ως προς θ και

$$\frac{1}{\sqrt{2}} < \gamma < 1$$

- (ii) Ο συντελεστής λοξότητας $\sqrt{\beta_1}$ είναι αύξουσα συνάρτηση ως προς θ και

$$\sqrt{2} < \sqrt{\beta_1} < 2$$

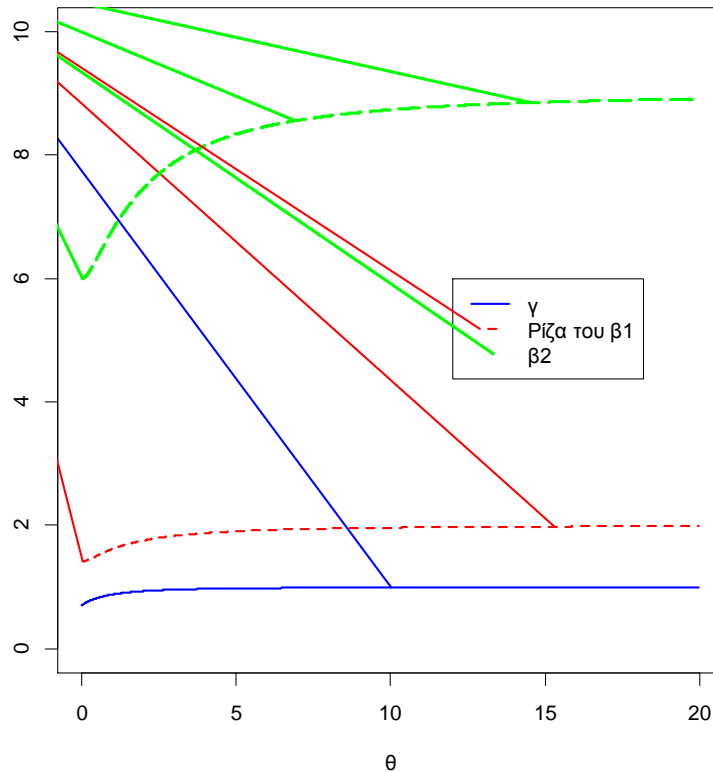
- (iii) Ο συντελεστής κύρτωσης β_2 είναι αύξουσα συνάρτηση ως προς θ και

$$6 < \beta_2 < 9$$

Γραφική παράσταση του συντελεστή μεταβλητότητας, λοξότητας και κύρτωσης της κατανομής Lindley δίνεται στο ακόλουθο σχήμα.

Σχήμα 1-3

Συντελεστής μεταβλητότητας, λοξότητας και κύρτωσης της κατανομής Lindley



Σημειώνουμε ότι στην εκθετική κατανομή συντελεστής μεταβλητότητας (γ), ο συντελεστής λοξότητας ($\sqrt{\beta_1}$) και ο συντελεστής κύρτωσης (β_2) είναι 1, 2 και 6 αντίστοιχα. Αυτό δείχνει για ακόμα μία φορά την ευελιξία της κατανομής Lindley έναντι της εκθετικής κατανομής.

1.6 Χαρακτηριστική συνάρτηση

Η χαρακτηριστική συνάρτηση της κατανομής Lindley δίνεται από τον τύπο

$$\varphi(t) = \frac{\theta^2(\theta - it + 1)}{(\theta + 1)(\theta - it)^2} .$$

Πράγματι

$$\begin{aligned} \varphi(t) = E[e^{itX}] &= \int_0^{+\infty} e^{itx} \frac{\theta^2}{\theta + 1} (1 + x)e^{-\theta x} dx \\ &= \left(\frac{\theta^2}{\theta + 1} \right) \frac{(\theta - it + 1)}{(\theta - it)^2} \int_0^{+\infty} \frac{(\theta - it)^2}{(\theta - it + 1)} (1 + x)e^{-(\theta - it)x} dx \end{aligned}$$

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

όπου $i = \sqrt{-1}$ είναι η φανταστική μονάδα. Έτσι η γεννήτρια συνάρτηση ημιαναλλοίωτων δίνεται από τη σχέση

$$H(t) = \log(E[e^{itX}]) = \log\left(1 - \frac{it}{\theta + 1}\right) - 2 \log\left(1 - \frac{it}{\theta}\right) = \sum_{r=0}^{\infty} \kappa_r \frac{(it)^r}{r}.$$

Κάνοντας χρήση του αναπτύγματος

$$\log(1 - x) = - \sum_{r=0}^{\infty} \frac{x^r}{r}, \quad -1 \leq x < 1$$

προκύπτει ότι η ημιαναλλοίωτη r -τάξης δίνεται από τον τύπο

$$\kappa_r = \frac{2(r-1)!}{\theta^r} - \frac{(r-1)!}{(\theta+1)^r}$$

Σημειώνεται πως η ημιαναλλοίωτη r -τάξης της εκθετικής κατανομής είναι $(r-1)!/\theta^r$.

1.7 Ρυθμός αποτυχίας και συνάρτηση μέσης υπολειπόμενης ζωής

Ο ρυθμός αποτυχίας (ένταση θνησιμότητας / ρυθμός κινδύνου) μιας θετικής συνεχούς τυχαίας μεταβλητή X δίνεται από τον τύπο

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(X < x + \Delta x \mid X > x)}{\Delta x} = \frac{f(x)}{1 - F(x)}.$$

Στην περίπτωση της κατανομής Lindley έχουμε ότι (Barlow & Proschan (1981))

$$h(x) = \frac{\theta^2(1+x)}{\theta+1+\theta x}.$$

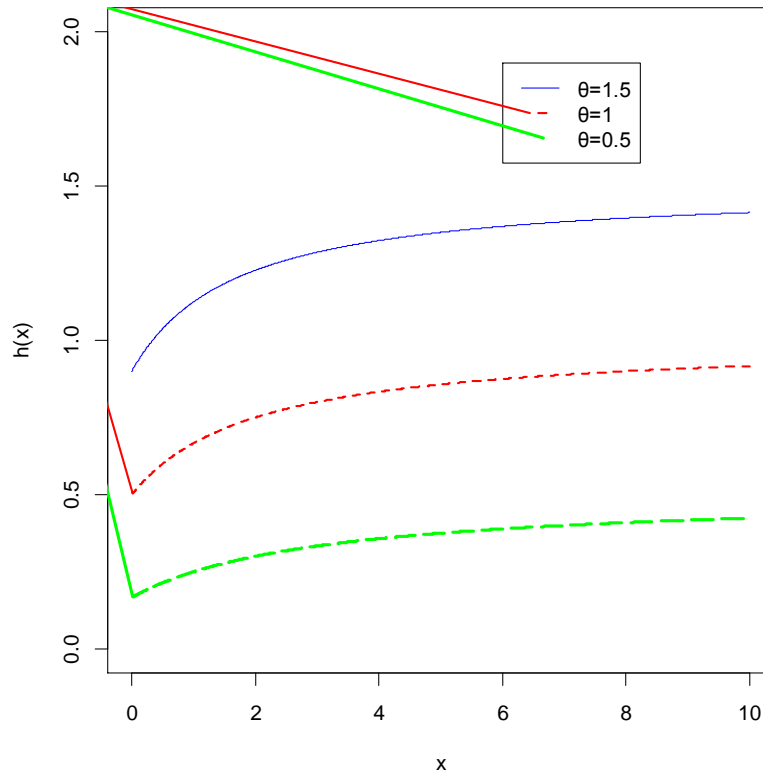
Παρατηρούμε ότι

1. $h(0) = f(0) = \theta^2/(\theta+1)$.
2. Η $h(x)$ είναι αύξουσα συνάρτηση του x και του θ και

$$\frac{\theta^2}{1+\theta} < h(x) < \theta.$$

3. Η ένταση θνησιμότητας για την εκθετική κατανομή $Exp(\theta)$ ως γνωστόν είναι $h(x) = \theta$, και έτσι είναι πάλι εμφανής η ευελιξία της Lindley κατανομής σε σχέση με την εκθετική.

Σχήμα 1-4
Ρυθμός αποτυχίας της κατανομής Lindley



Γενικά, σε μια συνεχή κατανομή με σ.π. $f(x)$ και σ.κ. $F(x)$, η συνάρτηση μέσης υπολειπόμενης ζωής δίνεται από τον τύπο

$$E(X - x | X > x) = m(x) = \frac{1}{1 - F(x)} \int_x^{+\infty} (1 - F(t)) dt.$$

Στην περίπτωση της κατανομής Lindley έχουμε ότι

$$m(x) = \frac{1}{(\theta + 1 + \theta x)e^{-\theta x}} \int_x^{+\infty} (\theta + 1 + \theta t)e^{-\theta t} dt$$

Όμως

$$I_1 = \int_x^{+\infty} \theta e^{-\theta t} dt = \left[\theta \frac{e^{-\theta t}}{-\theta} \right]_x^{+\infty} = e^{-\theta x},$$

$$I_2 = \int_x^{+\infty} e^{-\theta t} dt = \left[\frac{e^{-\theta t}}{-\theta} \right]_x^{+\infty} = \frac{e^{-\theta x}}{\theta},$$

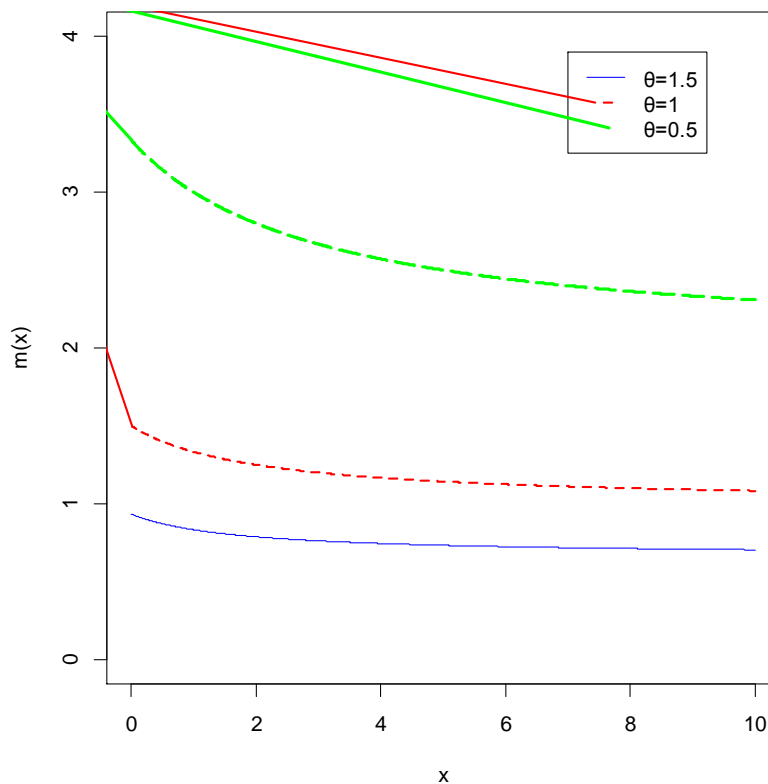
$$I_3 = \int_x^{+\infty} \theta t e^{-\theta t} dt = \left[\theta t \frac{e^{-\theta t}}{\theta} \right]_x^{+\infty} + \int_x^{+\infty} e^{-\theta t} dt = x e^{-\theta x} + \frac{e^{-\theta x}}{\theta}.$$

Έτσι

$$m(x) = \frac{1}{(\theta + 1 + \theta x)e^{-\theta x}} (I_1 + I_2 + I_3) = \frac{\theta + 2 + \theta x}{\theta(\theta + 1 + \theta x)}.$$

Σχήμα 1-5

Συνάρτηση μέσης υπολειπόμενης ζωής της κατανομής Lindley



Παρατηρούμε ότι

1. $m(0) = \mu$.
2. Η $m(x)$ είναι φθίνουσα συνάρτηση του x και του θ και

$$\frac{1}{\theta} < m(x) < \frac{\theta + 2}{\theta(\theta + 1)} = \mu.$$

3. Η συνάρτηση μέσης υπολειπόμενης ζωής για την εκθετική κατανομή $Exp(\theta)$ ως γνωστόν είναι $m(x) = 1/\theta$ που αποδεικνύει για άλλη μια φορά την ευελιξία της Lindley κατανομής σε σχέση με την εκθετική..

1.8 Μέσες αποκλίσεις

Η διασπορά ενός πληθυσμού X πολλές φορές αποτιμάται με τη μέση απόκλιση από το μέσο και τη διάμεσο αυτού. Για μία θετική συνεχή κατανομή η μέση απόκλιση από το μέσο και τη διάμεσο ορίζονται από τις παρακάτω σχέσεις αντιστοίχως

$$\delta_1(x) = E(|X - \mu|) = \int_0^{+\infty} |x - \mu|f(x)dx ,$$

και

$$\delta_2(x) = E(|X - M|) = \int_0^{+\infty} |x - M|f(x)dx ,$$

όπου $\mu = E(X)$ και $M = \text{Median}(X)$.

Παρατηρούμε ότι

$$E(|X - m|) = \int_0^m (m - x)f(x)dx + \int_m^{+\infty} (x - m)f(x)dx = 2 \left\{ mF(m) - \int_0^m xf(x)dx \right\}.$$

Συνεπώς

$$\begin{aligned} E(|X - \mu|) &= 2 \left\{ \mu F(\mu) - \int_0^\mu xf(x)dx \right\} \\ &= 2 \left\{ \mu - \left(\frac{\mu\theta + \mu + \mu^2\theta}{\theta + 1} e^{-\mu\theta} \right) - \int_0^\mu x \frac{\theta^2}{\theta + 1} e^{-\theta x} dx - \int_0^\mu x^2 \frac{\theta^2}{\theta + 1} e^{-\theta x} dx \right\} \\ &= 2 \left\{ \mu - \left(\frac{\mu\theta + \mu + \mu^2\theta}{\theta + 1} e^{-\mu\theta} \right) - I_1 - I_2 \right\}. \end{aligned}$$

Όμως

$$\begin{aligned} I_1 &= \int_0^\mu x \frac{\theta^2}{\theta + 1} e^{-\theta x} dx = \left[x \frac{\theta^2 e^{-\theta x}}{-\theta(\theta + 1)} \right]_0^\mu + \int_0^\mu \frac{e^{-\theta x} \theta^2}{\theta(\theta + 1)} dx \\ &= \left(\frac{-\mu\theta^2 e^{-\theta\mu}}{\theta(\theta + 1)} \right) + \frac{\theta}{\theta + 1} \left[\frac{e^{-\theta x}}{-\theta} \right]_0^\mu = \frac{1 - \mu\theta e^{-\mu\theta} - e^{-\mu\theta}}{\theta + 1} \end{aligned}$$

και

$$I_2 = \int_0^\mu x^2 \frac{\theta^2}{\theta + 1} e^{-\theta x} dx = \left[x^2 \frac{\theta^2 e^{-\theta x}}{-\theta(\theta + 1)} \right]_0^\mu + \frac{2}{\theta} I_1$$

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

Επομένως

$$= E(|X - \mu|) = 2\left\{\mu - \frac{\theta + 2}{\theta(\theta + 1)} + \frac{e^{-\mu\theta}}{\theta + 1}\left(\mu + 1 + \frac{2}{\theta}\right)\right\}$$

Ομοίως προκύπτει η μέση απόκλιση από τη διάμεσο αρκεί να αντικαταστήσουμε το μ με M .

Ο τελικός τύπος είναι

$$\delta_2(x) = 2\left\{M - \frac{\theta + 2}{\theta(\theta + 1)} + \frac{e^{-M\theta}}{\theta + 1}\left(M + 1 + \frac{2}{\theta}\right)\right\}$$

Σημειώνουμε ότι οι αντίστοιχες σχέσεις για την εκθετική κατανομή είναι

$$\delta_1(x) = 2\left\{\mu - \frac{1}{\theta} + \frac{e^{-\mu\theta}}{\theta}\right\}$$

και

$$\delta_2(x) = 2\left\{M - \frac{1}{\theta} + \frac{e^{-M\theta}}{\theta}\right\}.$$

1.9 Εκτιμητής μέγιστης πιθανοφάνειας

Έστω ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από την κατανομή $Lindley(\theta)$. Η συνάρτηση πιθανοφάνειας είναι

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \frac{\theta^2}{\theta + 1} (1 + x_i) e^{-\theta x_i} = \frac{\theta^{2n}}{(\theta + 1)^n} e^{-\theta x} \prod_{i=1}^n (1 + x_i)$$

όπου $x = \sum_{i=1}^n x_i$. Έστω $\ell(\theta) = \log L(\theta)$. Τότε

$$\begin{aligned} \ell(\theta) &= \log \frac{\theta^{2n}}{(\theta + 1)^n} + \sum_{i=1}^n \log(1 + x_i) - \theta x \\ &= 2n \log \theta - n \log(\theta + 1) + \sum_{i=1}^n \log(1 + x_i) - \theta x \end{aligned}$$

Παραγωγίζοντας ως προς θ έχουμε

$$\frac{d\ell(\theta)}{d\theta} = \frac{2n}{\theta} - \frac{n}{\theta+1} - x$$

Συνεπώς ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου θ προκύπτει από τη λύση της εξίσωσης

$$\frac{d\ell(\theta)}{d\theta} = \frac{2n}{\theta} - \frac{n}{\theta+1} - x = 0$$

ή ισοδύναμα

$$x\theta^2 + \theta(x - n) - 2n = 0$$

ή ισοδύναμα

$$\bar{x}\theta^2 + \theta(\bar{x} - 1) - 2 = 0.$$

Επομένως ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου θ είναι

$$\hat{\theta} = \frac{-(\bar{x} - 1) + \sqrt{(\bar{x} - 1)^2 + 8\bar{x}}}{2\bar{x}}.$$

Σημειώνεται πως η ίδια εκτίμηση προκύπτει και με τη μέθοδο των ροπών.

Στη συνέχεια θα δείξουμε ότι ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου θ είναι θετικά μεροληπτικός (δείτε Song (2001)).

Θεώρημα 1.2

Η εκτιμήτρια $\hat{\theta}$ της παραμέτρου θ είναι θετικά μεροληπτική, δηλαδή ισχύει ότι

$$E(\hat{\theta}) - \theta > 0.$$

Απόδειξη

Έστω $\hat{\theta} = g(\bar{X})$ όπου

$$g(t) = \frac{-(t-1) + \sqrt{(t-1)^2 + 8t}}{2t}, \quad t > 0.$$

Επειδή

$$g(t)'' = \frac{1}{t^3} \left[1 + \frac{3t^3 + 15t^2 + 9t + 1}{[(t-1)^2 + 8t]^{3/2}} \right] > 0$$

η $g(t)$ είναι γνήσια κυρτή. Τότε, από την ανισότητα του Jensen, θα έχουμε

$$E(g(\bar{X})) > g(E(\bar{X})).$$

Όμως

$$g(E(\bar{X})) = g(\mu) = g\left(\frac{\theta + 2}{\theta(\theta + 1)}\right) = \theta$$

οπότε

$$E(\hat{\theta}) > \theta. \quad \blacksquare$$

Στη συνέχεια δίνουμε την οριακή κατανομή του εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου θ .

Θεώρημα 1.3

Η εκτιμήτρια $\hat{\theta}$ της παραμέτρου θ ικανοποιεί την ακόλουθη σχέση

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{1}{\sigma^2}\right)$$

Για μεγάλο αριθμό τυχαίων δειγμάτων, το $100(1 - \alpha)\%$ διάστημα εμπιστοσύνης της παραμέτρου θ το

$$\hat{\theta} \pm z_{\alpha/2} \frac{1}{\sqrt{n\hat{\sigma}^2}}$$

Απόδειξη

Επειδή η μέση τιμή μ είναι πεπερασμένη, ισχύει ότι $\bar{X} \xrightarrow{P} \mu$. Αφού η συνάρτηση $g(t)$ είναι συνεχής συνάρτηση για $t = \mu$, τότε $g(\bar{X}) \xrightarrow{P} g(\mu)$, δηλαδή $\hat{\theta} \xrightarrow{P} \theta$. Καθώς $\sigma^2 < \infty$, από το Κεντρικό Οριακό Θεώρημα, θα έχουμε

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Εφόσον η $g(\mu)$ είναι διαφορίσιμη και $g'(\mu) \neq 0$, από τη μέθοδο Δέλτα θα έχουμε

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{D} N(0, [g'(\mu)]^2 \sigma^2).$$

Τέλος, επειδή $g(\bar{X}) = \hat{\theta}$, $g(\mu) = \theta$ και

$$g'(\mu) = -\left(\frac{1}{2\mu^2}\right) \left[1 + \frac{1 + 3\mu}{\sqrt{(\mu - 1)^2 + 8\mu}}\right] = -\frac{1}{\sigma^2},$$

παίρνουμε

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{1}{\sigma^2}\right).$$

1.10 Μελέτη προσομοίωσης

Αναφέραμε στην Παράγραφο 1.2 ότι η κατανομή Lindley αποτελεί μίξη μιας εκθετικής και μιας γάμμα κατανομής. Πιο συγκεκριμένα η συνάρτηση πυκνότητας $f(x)$ της κατανομής Lindley με παράμετρο θ ικανοποιεί τη σχέση

$$f_X = \frac{\theta}{\theta + 1} f_{X_1}(x) + \left(1 - \frac{\theta}{\theta + 1}\right) f_{X_2}(x)$$

όπου $X_1 \sim \text{Exp}(\theta)$ και $X_2 \sim \text{Gamma}(2, \theta)$.

Για την παραγωγή ενός τυχαίου δείγματος $X_i, i = 1, 2, \dots, n$, από την κατανομή Lindley(θ) εργαζόμαστε ως εξής:

1. Παράγουμε τυχαίο δείγμα $U_i \sim \text{Uniform}(0,1)$, $i = 1, 2, \dots, n$.
2. Παράγουμε τυχαίο δείγμα $V_i \sim \text{Exp}(\theta)$, $i = 1, 2, \dots, n$.
3. Παράγουμε τυχαίο δείγμα $W_i \sim \text{Gamma}(2, \theta)$, $i = 1, 2, \dots, n$.
4. Εάν $U_i \leq p = \frac{\theta}{\theta+1}$, θέτουμε $X_i = V_i$, διαφορετικά θέτουμε $X_i = W_i$, $i = 1, 2, \dots, n$.

Η μελέτη της προσομοίωσης έγινε για $N = 10000$ δείγματα και για επιλεγμένες τιμές των n , θ . Δίνουμε τα ακόλουθα αποτελέσματα:

- (i) Μέση μεροληψία (Average bias) για τους προσομοιωμένους εκτιμητές $\hat{\theta}_i$ ($i = 1, 2, \dots, N$)

$$\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta).$$

- (ii) Το μέσο τετραγωνικό σφάλμα (MSE) για τους προσομοιωμένους εκτιμητές $\hat{\theta}_i$ ($i = 1, 2, \dots, N$)

$$\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)^2.$$

- (iii) Πιθανότητα κάλυψης (coverage probability) που προκύπτει από τα N ασυμπτωτικά 90% και 95%. δ.ε.

Πίνακας 1-1

Μέση μεροληψία, μέσο τετραγωνικό σφάλμα και πιθανότητες κάλυψης
για την κατανομή *Lindley*(0.1)

<i>n</i>	Μέση μεροληψία	Μέσο τετραγωνικό σφάλμα	Πιθανότητα κάλυψης 90% δ.ε	Πιθανότητα κάλυψης 95% δ.ε
10	0.00482	0.00063	0.9128	0.9485
20	0.00271	0.00029	0.9069	0.9510
30	0.00175	0.00018	0.9052	0.9499
40	0.00105	0.00013	0.9033	0.9521
50	0.00084	0.00010	0.8998	0.9510
60	0.00091	0.00009	0.9032	0.9515
70	0.00074	0.00008	0.8995	0.9504
80	0.00063	0.00006	0.9028	0.9502
90	0.00060	0.00006	0.8972	0.9505
100	0.00037	0.00005	0.8982	0.9515
200	0.00021	0.00003	0.8973	0.9479
500	0.00006	0.00001	0.8985	0.9476
1000	0.00004	0.00000	0.8994	0.9518

Πίνακας 1-2

Μέση μεροληψία, μέσο τετραγωνικό σφάλμα και πιθανότητες κάλυψης
για την κατανομή *Lindley*(1)

<i>n</i>	Μέση μεροληψία	Μέσο τετραγωνικό σφάλμα	Πιθανότητα κάλυψης 90% δ.ε	Πιθανότητα κάλυψης 95% δ.ε
10	0.06061	0.07875	0.9151	0.9525
20	0.03164	0.03323	0.9153	0.9544
30	0.01976	0.02153	0.9082	0.9530
40	0.01468	0.01565	0.9020	0.9490
50	0.01038	0.01202	0.9024	0.9501
60	0.01110	0.01008	0.9024	0.9489
70	0.00858	0.00850	0.9032	0.9500
80	0.00736	0.00767	0.9001	0.9497
90	0.00725	0.00671	0.9006	0.9490
100	0.00523	0.00583	0.9036	0.9507
200	0.00233	0.00291	0.8990	0.9508
500	0.00107	0.00114	0.9026	0.9499
1000	0.00047	0.00057	0.9015	0.9523

Πίνακας 1-3
Μέση μεροληψία, μέσο τετραγωνικό σφάλμα και πιθανότητες κάλυψης
για την κατανομή *Lindley*(9)

n	Μέση μεροληψία	Μέσο τετραγωνικό σφάλμα	Πιθανότητα κάλυψης 90% δ.ε	Πιθανότητα κάλυψης 95% δ.ε
10	0.89862	11.07727	0.9211	0.9576
20	0.46029	4.41642	0.9126	0.9538
30	0.27466	2.67985	0.9087	0.9510
40	0.19864	1.90620	0.9058	0.9493
50	0.15386	1.48009	0.9062	0.9550
60	0.14394	1.20734	0.9074	0.9567
70	0.10971	1.04137	0.9040	0.9493
80	0.10520	0.90611	0.9044	0.9526
90	0.08923	0.81229	0.9006	0.9498
100	0.08580	0.72226	0.9030	0.9476
200	0.03529	0.35303	0.9013	0.9486
500	0.01376	0.13632	0.9034	0.9554
1000	0.00688	0.06836	0.8995	0.9505

Ο κώδικας στη γλώσσα προγραμματισμού *R* για την εξαγωγή των αριθμητικών αποτελεσμάτων των παραπάνω πινάκων δίνεται στο Παράρτημα.

Αξίζει να σημειώσουμε ότι από τους παραπάνω πίνακες προκύπτει ότι έχουμε θετική μεροληψία σύμφωνα και με το Θεώρημα 1.2. Παρατηρούμε πως η μεροληψία μειώνεται (αυξάνεται) καθώς το n (θ) αυξάνεται. Επίσης το μέσο τετραγωνικό σφάλμα μειώνεται (αυξάνεται) καθώς το n (θ) αυξάνεται, και η πιθανότητα κάλυψης της σε κάθε επίπεδο σημαντικότητας, είναι πολύ κοντά σε αυτό για όλα τα n , θ .

1.11 Προσαρμογή σε δεδομένα

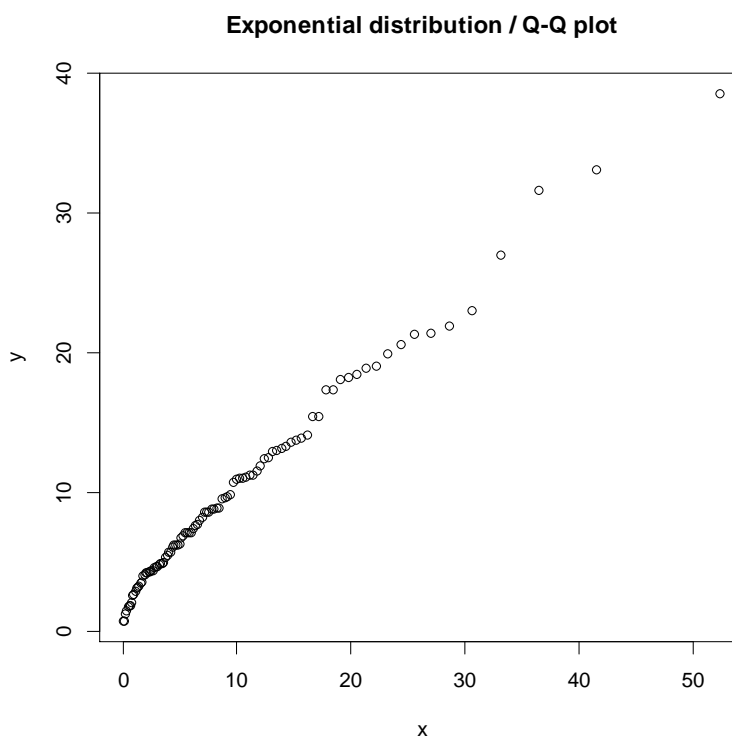
Στο σημείο αυτό, θα κάνουμε χρήση εμπειρικών δεδομένων αποδεικνύοντας πως ένα μοντέλο βασισμένο στην *Lindley* κατανομή έχει καλύτερα αποτελέσματα σε σχέση με ένα μοντέλο βασισμένο στην εκθετική κατανομή. Τα δεδομένα του παρακάτω πίνακα αποτελούν τους χρόνους αναμονής 100 πελατών μίας τράπεζας μέχρι να εξυπηρετηθούν.

Πίνακας 1-4
Χρόνοι αναμονής 100 πελατών μίας τράπεζας (σε λεπτά)

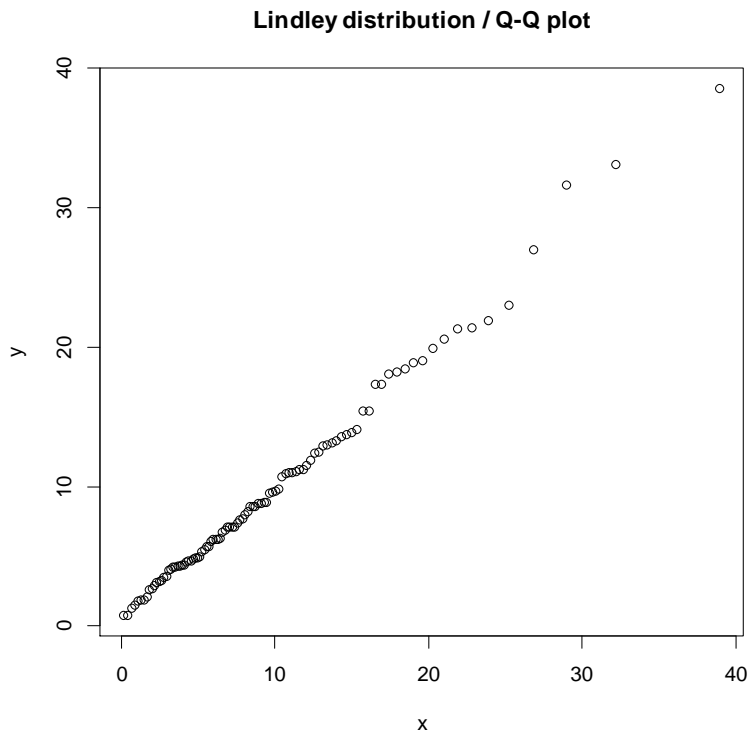
0.8	0.8	1.3	1.5	1.8	1.9	1.9	2.1	2.6	2.7
2.9	3.1	3.2	3.3	3.5	3.6	4.0	4.1	4.2	4.2
4.3	4.3	4.4	4.4	4.6	4.7	4.7	4.8	4.9	4.9
5.0	5.3	5.5	5.7	5.7	6.1	6.2	6.2	6.2	6.3
6.7	6.9	7.1	7.1	7.1	7.1	7.4	7.6	7.7	8.0
8.2	8.6	8.6	8.6	8.8	8.8	8.9	8.9	9.5	9.6
9.7	9.8	10.7	10.9	11.0	11.0	11.1	11.2	11.2	11.5
11.9	12.4	12.5	12.9	13.0	13.1	13.3	13.6	13.7	13.9
14.1	15.4	15.4	17.3	17.3	18.1	18.2	18.4	18.9	19.0
19.9	20.6	21.3	21.4	21.9	23.0	27.0	31.6	33.1	38.5

Στα παραπάνω δεδομένα, προσαρμόζουμε την Lindley και την εκθετική κατανομή. Οι εκτιμητές μέγιστης πιθανοφάνειας είναι $\hat{\theta} = 0.187$ για την Lindley κατανομή με $S.E.(\hat{\theta}) = 0.013$ και $\hat{\theta} = 0.101$ με $S.E.(\hat{\theta}) = 0.010$ για την εκθετική κατανομή. Η μέγιστη τιμή της συνάρτησης πιθανοφάνειας είναι -319 για την Lindley κατανομή και -329 για την εκθετική κατανομή από όπου προκύπτει η υπεροχή της Lindley κατανομής. Τέλος δίνουμε τα δύο ακόλουθα Q-Q διαγράμματα τα οποία επιβεβαιώνουν την υπεροχή της κατανομής Lindley. Ο κώδικας στη γλώσσα προγραμματισμού R για την εξαγωγή των Q-Q διαγραμμάτων δίνεται στο Παράρτημα.

Σχήμα 1-6
Διάγραμμα Q-Q για την εκθετική κατανομή



Σχήμα 1-7
Διάγραμμα Q-Q για την κατανομή Lindley



ΚΕΦΑΛΑΙΟ 2

Κατανομή Lindley και ανταγωνιστικοί κίνδυνοι

2.1 Εισαγωγή

Στη μελέτη των συμβάντων ζωής και θανάτου ενός πληθυσμού, παρατηρείται πως ο θάνατος ή γενικότερα το αίτιο εξόδου ενός ατόμου από ένα πληθυσμό μπορεί να είναι περισσότερα του ενός. Στην αναλογιστική επιστήμη και στην μελέτη των παραπάνω συμβάντων γίνεται ευρεία χρήση των κατανομών Weibull και Εκθετικής για τη μοντελοποίηση του χρόνου εξόδου του ατόμου. Εδώ, προτείνεται η χρήση της Lindley κατανομής. Τα κυριότερα αποτελέσματα αυτού του Κεφαλαίου βρίσκονται στην εργασία Mazucheli & Achcar (2011). Στο μοντέλο που θα χρησιμοποιήσουμε, βασική υπόθεση είναι πως τα $k (\geq 2)$ αίτια εξόδου είναι ανεξάρτητα και πως μία έξοδος μπορεί να συμβεί από ένα και μόνο αίτιο κάθε φορά.

Ας υποθέσουμε πως ένα άτομο (αντικείμενο) έχει $k (\geq 2)$ πιθανά αίτια εξόδου (ή αποτυχίας) από τον πληθυσμό. Έστω $Y_j, j = 1, 2, \dots, k$, ο χρόνος εξόδου λόγω του αιτίου εξόδου j . Τότε η τυχαία μεταβλητή T που δηλώνει το χρόνο εξόδου είναι η $T = \min(Y_1, Y_2, \dots, Y_k)$. Έχοντας υποθέσει πως τα αίτια εξόδου είναι ανεξάρτητα μεταξύ τους, η (συνολική) ένταση εξόδου (ή ένταση θνησιμότητας) $h_T(t)$ και η συνάρτηση επιβίωσης $S_T(t)$ της τ.μ. T είναι

$$h_T(t) = \sum_{j=1}^k h_j(t)$$

και

$$S_T(t) = \prod_{j=1}^k S_j(t)$$

αντίστοιχα, όπου $h_j(t)$ και $S_j(t)$, η ένταση θνησιμότητας και η συνάρτηση επιβίωσης για το αίτιο εξόδου j (δείτε Αντζουλάκος (2016)). Κάνοντας χρήση της σχέσης

$$h_T(t) = \frac{f_T(t)}{S_T(t)}$$

προκύπτει ότι

$$f_T(t) = \left(\sum_{j=1}^k h_j(t) \right) \left(\prod_{j=1}^k S_j(t) \right) = \left(\sum_{j=1}^k \frac{f_{T_j}(t)}{S_{T_j}(t)} \right) \left(\prod_{j=1}^k S_j(t) \right) \sum_{j=1}^k f_j(t) \prod_{\substack{l=1 \\ l \neq j}}^k S_l(t)$$

όπου $f_j(t)$ είναι η σ.π. της τ.μ. Y_j , $j = 1, 2, \dots, k$.

2.2 Μοντέλο ανταγωνιστικών κινδύνων για την κατανομή Lindley

Έστω πως ένα άτομο (ή αντικείμενο) ενός πληθυσμού έχει k (≥ 2) πιθανά αίτια εξόδου από τον πληθυσμό. Ο χρόνος για να συμβεί αυτή η έξοδος είναι μια τ.μ. $T = \min(Y_1, Y_2, \dots, Y_k)$, όπου Y_j ($j = 1, 2, \dots, k$) είναι ο χρόνος εξόδου λόγω του αιτίου εξόδου j . Υποθέτουμε πως οι τ.μ. Y_j ($j = 1, 2, \dots, k$) είναι ανεξάρτητες τυχαίες μεταβλητές που κάθε μια ακολουθεί την Lindley κατανομή $Lindley(\theta_j)$. Τότε

$$h_j(y|\theta_j) = \frac{\theta_j^2(1+j)}{(\theta_j + 1 + y\theta_j)}$$

και

$$S_j(y|\theta_j) = \frac{(\theta_j + 1 + y\theta_j)}{(\theta_j + 1)} e^{-y\theta_j}$$

($y \geq 0$ και $\theta_j > 0$ για $j = 1, 2, \dots, k$).

Για διαφορετικές τιμές της παραμέτρου θ_j , παρατηρούνται διαφορετικές τιμές της έντασης θνησιμότητας. Το γεγονός αυτό μας δίνει μεγαλύτερη ευελιξία στην προσαρμογή του μοντέλου μας σε πραγματικούς χρόνους ζωής.

Στο Κεφάλαιο 1, είδαμε πως η ροπή r -τάξης της κατανομής Lindley δίνεται από τη σχέση

$$\mu'_r = E(Y^r) = \frac{r!(\theta + r + 1)}{\theta^r(\theta + 1)}, \quad r = 1, 2, 3, \dots$$

οπότε η μέση τιμή και η διασπορά της είναι, αντίστοιχα,

$$E(Y) = \frac{\theta + 2}{\theta(\theta + 1)}$$

και

$$Var(Y) = \frac{2(\theta + 3)}{\theta^2(\theta + 1)} - \frac{(\theta + 2)^2}{\theta^2(\theta + 1)^2}.$$

Για την τυχαία μεταβλητή $T = \min(Y_1, Y_2, \dots, Y_k)$ έχουμε ότι η σ.π. της δίνεται από τον τύπο

$$f(t|\boldsymbol{\theta}) = \left(\sum_{j=1}^k \frac{\theta_j^2(1+t)}{(\theta_j+1+t\theta_j)} \right) \left(\prod_{j=1}^k \frac{(\theta_j+1+t\theta_j)}{(\theta_j+1)} e^{-t\theta_j} \right)$$

όπου $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. Επίσης

$$\begin{aligned} f(t|\boldsymbol{\theta}) &= \sum_{j=1}^k f_j(t) \prod_{\substack{l=1 \\ l \neq j}}^k S_l(t) = \sum_{j=1}^k h_j(t) S_j(t) \prod_{\substack{l=1 \\ l \neq j}}^k S_l(t) \\ &= \sum_{j=1}^k \frac{\theta_j^2(1+t)}{(\theta_j+1)} e^{-t\theta_j} \prod_{\substack{l=1 \\ l \neq j}}^k \frac{(\theta_l+1+t\theta_l)}{(\theta_l+1)} e^{-t\theta_l}. \end{aligned}$$

Για $k = 2$ ας υποθέσουμε πως έχουν καταγραφεί οι χρόνοι εξόδου και τα αίτια εξόδου n παρατηρήσεων. Δηλαδή τα δεδομένα μας είναι της μορφής

$$\mathbf{t} = \{(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)\}$$

όπου t_i είναι ο παρατηρούμενος χρόνος εξόδου του i ατόμου ($i = 1, 2, \dots, n$) και δ_i είναι μια δίτιμη τ.μ. που παίρνει τις τιμές 1 και 2 ανάλογα με το αίτιο που προξένησε την έξοδο του i ατόμου. Η συνάρτηση πιθανοφάνειας που αντιστοιχεί στα παραπάνω δεδομένα είναι (δείτε για παράδειγμα David & Moeschberger (1978), Pintilie (2006), Sarhan (2007))

$$\begin{aligned} L(\theta_1, \theta_2 | \mathbf{t}) &= \prod_{i=1}^n \left(\frac{\theta_1^2(1+t_i)}{(\theta_1+1)} e^{-t_i\theta_1} \frac{(\theta_2+1+t_i\theta_2)}{(\theta_2+1)} e^{-t_i\theta_2} \right)^{I[\delta_i=1]} \times \\ &\times \left(\frac{\theta_2^2(1+t_i)}{(\theta_2+1)} e^{-t_i\theta_2} \frac{(\theta_1+1+t_i\theta_1)}{(\theta_1+1)} e^{-t_i\theta_1} \right)^{I[\delta_i=2]} \end{aligned}$$

Χωρίς βλάβη της γενικότητας ας υποθέσουμε ότι οι πρώτοι n_1 χρόνοι εξόδου οφείλονται στο αίτιο εξόδου 1 και οι υπόλοιποι στο αίτιο εξόδου 2. Τότε

$$\begin{aligned} L(\theta_1, \theta_2 | \mathbf{t}) &= \\ &= \prod_{i=1}^{n_1} \left(\frac{\theta_1^2(1+t_i)}{(\theta_1+1)} e^{-t_i\theta_1} \frac{(\theta_2+1+t_i\theta_2)}{(\theta_2+1)} e^{-t_i\theta_2} \right) \prod_{i=n_1+1}^n \left(\frac{\theta_2^2(1+t_i)}{(\theta_2+1)} e^{-t_i\theta_2} \frac{(\theta_1+1+t_i\theta_1)}{(\theta_1+1)} e^{-t_i\theta_1} \right) \\ &= \frac{\theta_1^{2n_1} e^{-t\theta_1}}{(\theta_1+1)^n} \cdot \frac{\theta_2^{2(n-n_1)} e^{-t\theta_2}}{(\theta_2+1)^n} \left(\prod_{i=1}^{n_1} (1+t_i) \right) \left(\prod_{i=1}^{n_1} (\theta_2+1+t_i\theta_2) \right) \left(\prod_{i=n_1+1}^n (\theta_1+1+t_i\theta_1) \right) \end{aligned}$$

όπου $t = \sum_{i=1}^n t_i$. Έτσι

$$\begin{aligned}\log L(\theta_1, \theta_2 | \mathbf{t}) &= \ell(\theta_1, \theta_2 | \mathbf{t}) = \\ &\propto 2 \log \left(\theta_1^{n_1} \theta_2^{(n-n_1)} \right) - n \log [(\theta_1 + 1)(\theta_2 + 1)] - t(\theta_1 + \theta_2) \\ &+ \sum_{i=1}^{n_1} \log(\theta_2 + 1 + t_i \theta_2) + \sum_{i=n_1+1}^n \log(\theta_1 + 1 + t_i \theta_1)\end{aligned}$$

Παραγωγίζοντας ως προς θ_1 και θ_2 παίρνουμε

$$\begin{aligned}\frac{\partial \ell(\theta_1, \theta_2 | \mathbf{t})}{\partial \theta_1} &= \sum_{i=1}^n \frac{1 + t_i}{\theta_1(1 + t_i) + 1} - t - \frac{n}{\theta_1 + 1} + 2 \frac{n_1}{\theta_1}, \\ \frac{\partial \ell(\theta_1, \theta_2 | \mathbf{t})}{\partial \theta_2} &= \sum_{i=1}^n \frac{1 + t_i}{\theta_2(1 + t_i) + 1} - t - \frac{n}{\theta_2 + 1} + 2 \frac{n_2}{\theta_2}.\end{aligned}$$

Από τη λύση του συστήματος

$$\frac{\partial \ell(\theta_1, \theta_2 | \mathbf{t})}{\partial \theta_1} = 0, \quad \frac{\partial \ell(\theta_1, \theta_2 | \mathbf{t})}{\partial \theta_2} = 0$$

μπορούμε να πάρουμε τους Ε.Μ.Π. $\hat{\theta}_1$ και $\hat{\theta}_2$ των παραμέτρων θ_1 και θ_2 αντίστοιχα. Το $100 \times (1 - \alpha)\%$ διάστημα εμπιστοσύνης για τις παραμέτρους θ_1 και θ_2 μπορεί να προκύψει εφαρμόζοντας τη θεωρία πιθανοφάνειας και πιο συγκεκριμένα με χρήση των $\text{Var}(\hat{\theta}_1)$ και $\text{Var}(\hat{\theta}_2)$ που βρίσκονται στη διαγώνιο του αντίστροφου παρατηρούμενου πίνακα πληροφορίας του Fischer, ο οποίος είναι ο αντίστροφος πίνακας των δευτέρων μερικών παραγώγων του λογαρίθμου της συνάρτησης πιθανοφάνειας υπολογισμένος στο Ε.Μ.Π. των παραμέτρων θ_1 και θ_2 . Από τις προηγούμενες σχέσεις προκύπτει άμεσα πως ο παρατηρούμενος πίνακας πληροφορίας του Fischer είναι ένας διαγώνιος πίνακας. Οι παραπάνω εξισώσεις δεν έχουν ακριβή λύση ως προς θ_1 και θ_2 . Οι Ε.Μ.Π. των θ_1 και θ_2 μπορούν να βρεθούν με προσεγγιστικές μεθόδους.

Για να μελετηθεί η συμπεριφορά των Ε.Μ.Π., οι Mazucheli & Achcar (2011) παρουσίασαν αποτελέσματα μιας μελέτη προσομοίωσης. Προσομοίωσαν παρατηρήσεις από ένα πληθυσμό με 2 αίτια εξόδου, έχοντας υποθέσει την Lindley κατανομή για κάθε αίτιο. Την προσομοίωση παρατηρήσεων από την Lindley κατανομή την παρουσιάσαμε στο προηγούμενο κεφάλαιο. Εδώ προσομοιώνονται δεδομένα από δύο κατανομές Lindley όπου καταγράφεται η ελάχιστη τιμή μεταξύ των δύο παρατηρήσεων (χρόνος εξόδου) και ποια από τις δύο ήταν η ελάχιστη τιμή (αίτιο εξόδου). Η μελέτη της προσομοίωσης έγινε για $N = 10000$ δείγματα και για επιλεγμένες τιμές των n , θ_1 και θ_2 . Δίνονται αποτελέσματα για τις κάτωθι ποσότητες:

- (i) Μέση μεροληψία (Average bias) για τους προσομοιωμένους εκτιμητές $\hat{\theta}_{ji}$ ($i = 1, 2, \dots, N, j = 1, 2$)

$$\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_{ji} - \theta_j).$$

- (ii) Το μέσο τετραγωνικό σφάλμα (MSE) για τους προσομοιωμένους εκτιμητές $\hat{\theta}_i \hat{\theta}_{ji}$ ($i = 1, 2, \dots, N, j = 1, 2$)

$$\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_{ji} - \theta_j)^2.$$

- (iii) Πιθανότητα κάλυψης (coverage probability) που προκύπτει από τα N ασυμπτωτικά 95%. δ.ε.

Τα αποτελέσματα της προσομοίωσης παρουσιάζονται στους παρακάτω Πίνακες 2-1, 2-2 και 2-3.

Πίνακας 2-1

Μέση μεροληψία, μέσο τετραγωνικό σφάλμα, κάλυψη πιθανότητας και ποσοστημόρια (5%, 95%) για $\theta_1=0.9$ και $\theta_2=0.2$

n	Παράμετρος	BIAS	MSE	CP	5%	95%
20	θ_1	0.0266	0.0282	0.9538	0.6898	1.2241
	θ_2	0.0095	0.0109	0.8588	0.0000	0.3550
30	θ_1	0.0177	0.0177	0.9545	0.7224	1.1512
	θ_2	0.0047	0.0064	0.9270	0.0731	0.3245
50	θ_1	0.0102	0.0105	0.9494	0.7541	1.0900
	θ_2	0.0017	0.0034	0.9329	0.1055	0.2941
100	θ_1	0.0051	0.0050	0.9510	0.7951	1.0265
	θ_2	0.0011	0.0016	0.9431	0.1347	0.2645

Πίνακας 2-2

Μέση μεροληψία, μέσο τετραγωνικό σφάλμα, κάλυψη πιθανότητας και ποσοστημόρια (5%, 95%) για $\theta_1=0.9$ και $\theta_2=0.7$

n	Παράμετρος	BIAS	MSE	CP	5%	95%
20	θ_1	0.0257	0.0406	0.9497	0.6320	1.2774
	θ_2	0.0214	0.0333	0.9463	0.4536	1.0429
30	θ_1	0.0170	0.0257	0.9505	0.6771	1.2019
	θ_2	0.0140	0.0209	0.9483	0.4972	0.9664
50	θ_1	0.0096	0.0153	0.9469	0.7202	1.1240
	θ_2	0.0081	0.0123	0.9481	0.5382	0.8981
100	θ_1	0.0043	0.0074	0.9514	0.7705	1.0518
	θ_2	0.0041	0.0059	0.9485	0.5840	0.8375

Πίνακας 2-3

Μέση μεροληψία, μέσο τετραγωνικό σφάλμα, κάλυψη πιθανότητας και ποσοστημόρια (5%, 95%) για $\theta_1=0.5$ και $\theta_2=0.2$

n	Παράμετρος	BIAS	MSE	CP	5%	95%
20	θ_1	0.0139	0.0090	0.9519	0.3765	0.6792
	θ_2	0.0020	0.0046	0.9289	0.0967	0.3152
30	θ_1	0.0096	0.0058	0.9533	0.3978	0.6423
	θ_2	0.0015	0.0028	0.9399	0.1181	0.2910
50	θ_1	0.0054	0.0034	0.9480	0.4162	0.6052
	θ_2	0.0010	0.0016	0.9449	0.1367	0.2687
100	θ_1	0.0027	0.0017	0.9499	0.4397	0.5724
	θ_2	0.0004	0.0008	0.9485	0.1562	0.2478

Από τους παραπάνω πίνακες, συμπεραίνουμε πως για τις παραμέτρους θ_1 και θ_2 υπάρχει θετική και φθίνουσα μεροληψία καθώς το μέγεθος του πληθυσμού αυξάνεται. Το μέγεθος της μεροληψίας και του μέσου τετραγωνικού σφάλματος (MSE) δεν φαίνεται να επηρεάζεται από την απόσταση των θ_1 , θ_2 , ενώ η πιθανότητα κάλυψης είναι πολύ κοντά στα φυσιολογικά επίπεδα εκτός από την περίπτωση όπου $\theta_1 = 0.9$ και $\theta_2 = 0.2$ στην περίπτωση $n = 20$.

Οι Mazucheli & Achcar (2011) επίσης προσάρμοσαν την κατανομή Lindley σε ένα πραγματικό σεν δεδομένων του Lagakos (1978) όπου τα αποτελέσματα έδειξαν καλύτερη προσαρμογή σε αυτά έναντι της εκθετικής κατανομής και της κατανομής Weibull.

ΚΕΦΑΛΑΙΟ 3

Η γενικευμένη κατανομή Lindley

3.1 Εισαγωγή

Η Lindley κατανομή με μία παράμετρο δεν παρέχει αρκετή ευελιξία στην ανάλυση διαφόρων εμπειρικών δεδομένων. Προς ενίσχυση της ευελιξίας της έχουν προταθεί διάφορες γενικεύσεις της. Μια τέτοια γενίκευση είναι η κατανομή Lindley με τρεις παραμέτρους που θα εξετάσουμε παρακάτω και η οποία γενικεύει την Lindley κατανομή, περιέχοντας ως ειδικές περιπτώσεις την εκθετική και την γάμμα κατανομή. Η διαδικασία που ακολουθείται βασίζεται στη μίξη γάμμα κατανομών. Θα μελετήσουμε διάφορες ιδιότητες του γενικευμένου μοντέλου και συγκεκριμένα στην Παράγραφο 3.2 θα κάνουμε μια εισαγωγή στην γενικευμένη Lindley κατανομή και τα βασικά χαρακτηριστικά της όπως η συμπεριφορά της σ.π.π., της συνάρτησης κινδύνου, κατανομή αθροίσματος, αλγόριθμο παραγωγής τυχαίων παρατηρήσεων από τη γενικευμένη κατανομή Lindley, εκτίμηση παραμέτρων, κλπ. Στις υπόλοιπες παραγράφους θα εξεταστούν άλλα χαρακτηριστικά της γενικευμένης κατανομής Lindley. Τα συμπεράσματα του Κεφαλαίου 3 βασίζονται στην δημοσίευση της εργασίας των Zakerzadeh & Dolati (2009).

3.2 Ορισμός και βασικές ιδιότητες της γενικευμένης κατανομής Lindley

3.2.1 Η γενίκευση της κατανομής

Η σ.π. της κατανομής γάμμα με παράμετρο μορφής a και παράμετρο κλίμακας θ ($Gamma(a, \theta)$) δίνεται από τον τύπο

$$f_g(x; a, \theta) = \frac{\theta(\theta x)^{a-1} e^{-\theta x}}{\Gamma(a)}, \quad a, \theta, x > 0.$$

Έστω X_1 μια συνεχής τυχαία μεταβλητή που έχει την κατανομή γάμμα με παράμετρο μορφής a και παράμετρο κλίμακας θ ($X_1 \sim Gamma(a, \theta)$) και X_2 μια άλλη συνεχής τυχαία μεταβλητή, ανεξάρτητη από τη X_1 , που έχει την κατανομή γάμμα με παράμετρο μορφής $a + 1$ και

παράμετρο κλίμακας θ ($X_2 \sim \text{Gamma}(a + 1, \theta)$). Τότε η τυχαία μεταβλητή X με συνάρτηση πυκνότητας

$$\begin{aligned} f_X(x; a, \theta, \gamma) &= \frac{\theta}{\theta + \gamma} f_{X_1}(x) + \left(1 - \frac{\theta}{\theta + \gamma}\right) f_{X_2}(x) \\ &= \frac{\theta}{\theta + \gamma} f_g(x; a, \theta) + \left(\frac{\gamma}{\theta + \gamma}\right) f_g(x; a + 1, \theta) \\ &= \frac{\theta^2 (\theta x)^{a-1} (a + \gamma x) e^{-\theta x}}{(\theta + \gamma) \Gamma(a + 1)}, \quad a, \theta, \gamma, x > 0 \end{aligned}$$

θα λέμε ότι έχει τη γενικευμένη κατανομή Lindley με παραμέτρους a, θ, γ ($X \sim GL(a, \theta, \gamma)$). Επομένως η γενικευμένη κατανομή Lindley αποτελεί μίξη δύο κατανομών γάμμα.

Από την παραπάνω γέννηση της κατανομής Lindley προκύπτει ότι η συνάρτηση κατανομής της δίνεται από το γενικό τύπο

$$F_X(x) = \frac{\theta}{\theta + \gamma} F_{X_1}(x) + \left(1 - \frac{\theta}{\theta + \gamma}\right) F_{X_2}(x), \quad x > 0.$$

Η γενικευμένη κατανομή Lindley περιέχει ως ειδικές περιπτώσεις την κατανομή Lindley, την κατανομή γάμμα και την εκθετική κατανομή. Πράγματι, για $a = \gamma = 1$ προκύπτει η κατανομή $Lindley(\theta)$, δηλαδή

$$GL(1, \theta, 1) \equiv Lindley(\theta).$$

Για $\gamma = 0$ προκύπτει η κατανομή $\text{Gamma}(a, \theta)$, δηλαδή

$$GL(a, \theta, 0) \equiv \text{Gamma}(a, \theta).$$

Για $a = 1$ και $\gamma = 0$ προκύπτει η κατανομή $\text{Exp}(\theta)$, δηλαδή

$$GL(1, \theta, 0) \equiv \text{Exp}(\theta).$$

3.2.2 Το σχήμα της κατανομής

Η πρώτη και η δεύτερη παράγωγος του λογάριθμου της σ.π. της κατανομής $GL(a, \theta, \gamma)$ είναι

$$\begin{aligned} \frac{d}{dx} \log(f(x)) &= \frac{(a - 1)(a + \gamma x) + (\gamma - \theta(a + \gamma + x))x}{x(a + \gamma x)}, \\ \frac{d^2}{dx^2} \log(f(x)) &= \frac{(1 - a)(a + \gamma x)^2 - (\gamma x)^2}{x^2(a + \gamma x)^2}. \end{aligned}$$

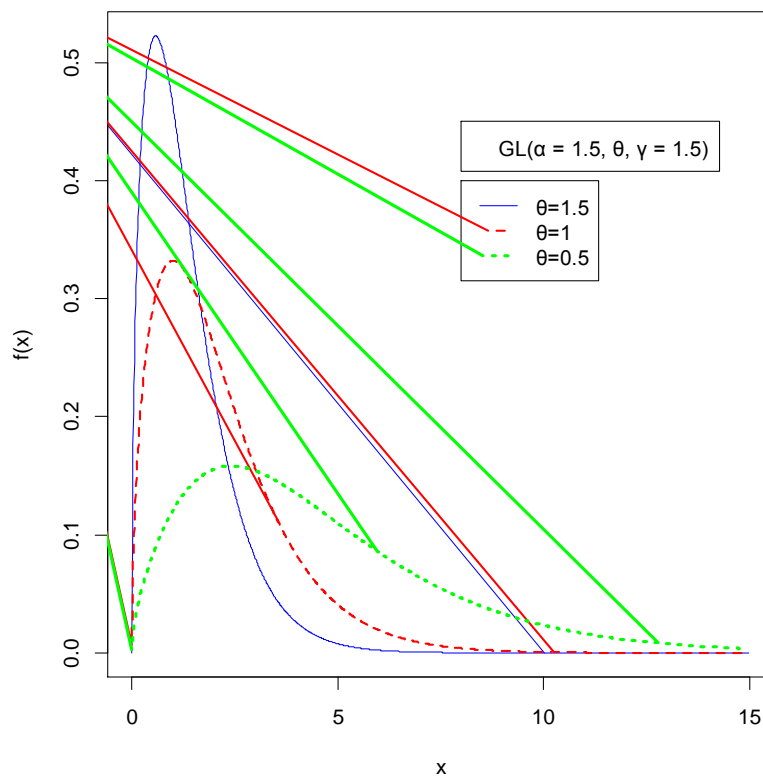
Για $a \geq 1$ έχουμε ότι $\frac{d^2}{dx^2} \log(f(x)) \leq 0$ και επομένως η σ.π. $f(x)$ είναι λογαριθμικά κοίλη (log-concave). Επειδή $\log(f'(0)) = \infty$ και $\log(f'(\infty)) = -\theta < 0$, προκύπτει ότι για $a \geq 1$ η $f(x)$ έχει μοναδική κορυφή στο σημείο

$$x_0 = \frac{a(\gamma - \theta) + \sqrt{(a(\gamma + \theta))^2 - 4a\theta\gamma}}{2\gamma\theta}$$

που είναι η μοναδική λύση της εξίσωσης $\frac{d}{dx} \log(f(x)) = 0$.

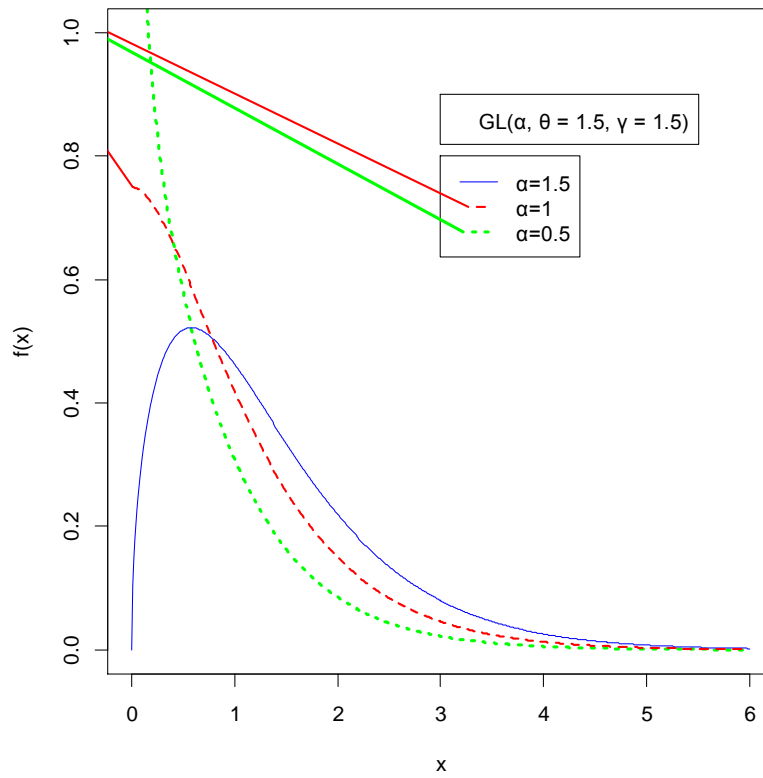
Για $a < 1$ έχουμε ότι $\frac{d}{dx} \log(f(x)) \leq 0$, δηλαδή η $f(x)$ είναι φθίνουσα (Glaser (1980)).

Σχήμα 3-1
Συνάρτησης πυκνότητας της κατανομής $GL(1.5, \theta, 1.5)$



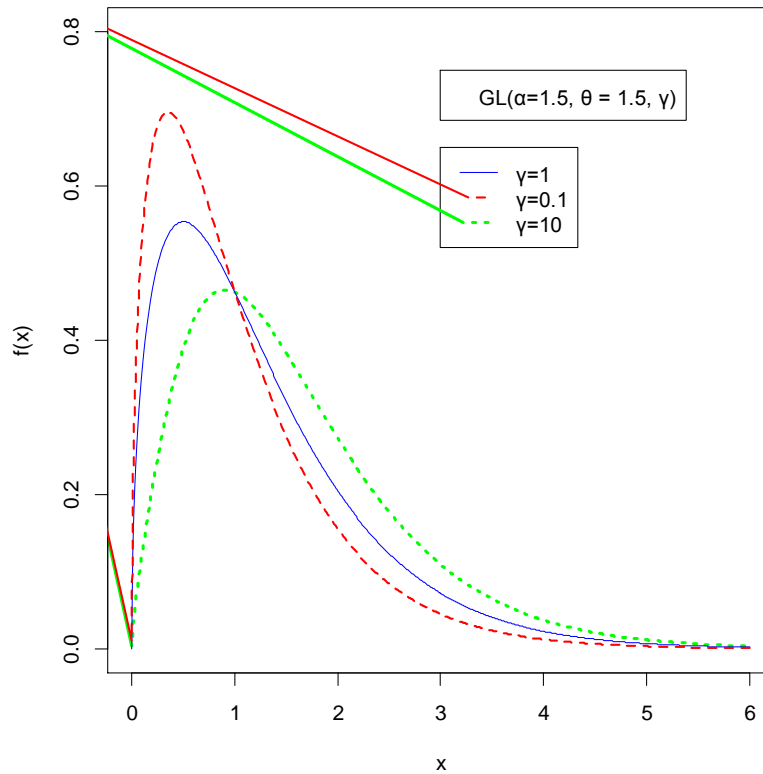
Σχήμα 3-2

Συνάρτησης πυκνότητας της κατανομής $GL(a, 1.5, 1.5)$



Σχήμα 3-3

Συνάρτησης πυκνότητας της κατανομής $GL(1.5, 1.5, \gamma)$



Έστω τώρα

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{S(x)}, \quad x > 0$$

η συνάρτηση κινδύνου της κατανομής $GL(a, \theta, \gamma)$. Επειδή η συνάρτηση επιβίωσης της κατανομής $GL(a, \theta, \gamma)$ μπορεί να δοθεί συναρτήσει της μη πλήρους γάμμα συνάρτησης όταν το a δεν είναι ακέραιος, η συνάρτηση κινδύνου δεν μπορεί να δοθεί σε κλειστή μορφή. Ωστόσο, οι ιδιότητές της μπορούν να προσδιοριστούν σύμφωνα με την ακόλουθη πρόταση.

Πρόταση 3.1

Έστω X μια τ.μ. που έχει την γενικευμένη κατανομή Lindley με παραμέτρους a, θ, γ και έστω $h(t)$ η συνάρτηση κινδύνου της. Τότε

- (i) Η $h(t)$ είναι αύξουσα για $a \geq 1$.
- (ii) Η $h(t)$ είναι λεκανοειδούς μορφής (bathtub) για $a < 1$ και $\gamma > 0$.
- (iii) Η $h(t)$ είναι φθίνουσα για $a \leq 1$ και $\gamma = 0$.

Απόδειξη

Παρατηρούμε ότι

$$\rho(t) = -\frac{f'(t)}{f(t)} = \frac{1 - \alpha}{t} - \frac{\gamma}{\alpha + \gamma t} + \theta$$

και

$$\rho'(t) = \frac{\alpha - 1}{t^2} + \frac{\gamma^2}{(\alpha + \gamma t)^2}.$$

Έτσι $\rho'(t) \geq 0$ όταν $a \geq 1$, και $\rho'(t) < 0$ όταν $a \leq 1$ και $\gamma = 0$. Όταν $a < 1$ και $\gamma > 0$ έχουμε

- $\rho'(t) < 0$ για $t \in \left(-\infty, \frac{\sqrt{1-a}+1-a}{\gamma}\right)$
- $\rho'(t) > 0$ για $t \in \left(\frac{\sqrt{1-a}+1-a}{\gamma}, +\infty\right)$
- $\rho'(t) = 0$ για $t = \frac{\sqrt{1-a}+1-a}{\gamma}$.

Τα παραπάνω αποτελέσματα αποδεικνύουν την πρόταση (δείτε Glaser (1980)). ■

3.2.3 Κατανομή αθροίσματος

Είναι γνωστό πως το άθροισμα ανεξάρτητων τ.μ. που ακολουθούν όλες γάμμα κατανομή με την ίδια παράμετρο κλίμακας θ αλλά διαφορετικές παραμέτρους μορφής, ακολουθεί γάμμα κατανομή με παράμετρο μορφής το άθροισμα των παραμέτρων μορφής των επιμέρους γάμμα κατανομών και την ίδια παράμετρο κλίμακας θ . Η παρακάτω πρόταση μας δείχνει πως η κατανομή του αθροίσματος ανεξάρτητων τ.μ. που ακολουθούν την γενικευμένη Lindley κατανομή μπορεί να γραφτεί ως μίξη γάμμα κατανομών.

Πρόταση 3.2

Έστω X_1, X_2, \dots, X_n ανεξάρτητες τ.μ. με $X_i \sim GL(a_i, \theta, \gamma)$ για $i = 1, 2, \dots, n$. Τότε η σ.π. της τ.μ. $S = \sum_{i=1}^n X_i$, δίνεται από τη σχέση

$$f_S(x) = \sum_{k=0}^n p_k f_g(x; a^* + k, \theta)$$

όπου f_g η σ.π. της γάμμα κατανομής παράμετρο κλίμακας θ , με παράμετρο μορφής $a^* + k$

όπου $a^* = \sum_{i=1}^n a_i$, και

$$p_k = \frac{\binom{n}{k} \theta^{n-k} \gamma^k}{(\gamma + \theta)^n}, \quad k = 0, 1, \dots, n$$

όπου $\sum_{k=0}^n p_k = 1$.

Απόδειξη

Η ροπογεννήτρια συνάρτηση της κατανομής $GL(a, \theta, \gamma)$ δίνεται από τον τύπο (δείτε προηγούμενη παράγραφο)

$$\begin{aligned} M(t) &= E(e^{tX}) = \frac{\theta}{\theta + \gamma} M_{X_1}(t) + \frac{\gamma}{\theta + \gamma} M_{X_2}(t) \\ &= \frac{\theta}{\theta + \gamma} \left(\frac{\theta}{\theta - t}\right)^a + \frac{\gamma}{\theta + \gamma} \left(\frac{\theta}{\theta - t}\right)^{a+1} = \left(\frac{\theta}{\theta - t}\right)^{a+1} \left(\frac{\theta - t + \gamma}{\theta + \gamma}\right). \end{aligned}$$

Έτσι η ροπογεννήτρια συνάρτηση της τ.μ. $S = \sum_{i=1}^n X_i$, όπου $X_i \sim GL(a, \theta, \gamma)$, δίνεται από τον τύπο

$$M_S(t) = \left(\frac{\theta}{\theta - t}\right)^{a^*+n} \left(\frac{\theta - t + \gamma}{\theta + \gamma}\right)^n.$$

Έστω τώρα

$$M_g(t; a^* + k, \theta) = \left(\frac{\theta}{\theta - t}\right)^{a^*+k}, \quad k = 0, 1, \dots, n$$

η ροπογεννήτρια συνάρτηση της γάμμα κατανομής $Gamma(a^* + k, \theta)$. Ας συμβολίσουμε με $M_f(t)$ τη ροπογεννήτρια συνάρτηση που αντιστοιχεί στη σ.π.

$$f_S(x) = \sum_{k=0}^n p_k f_g(x; a^* + k, \theta).$$

Τότε

$$\begin{aligned} M_f(t) &= \sum_{k=0}^n p_k M_g(x; a^* + k, \theta) = \frac{\theta^{n+a^*}}{(\gamma + \theta)^n} \left(\frac{1}{\theta - t}\right)^{a^*} \sum_{k=0}^n \binom{n}{k} \left(\frac{\gamma}{\theta - t}\right)^k \\ &= \frac{\theta^{n+a^*}}{(\gamma + \theta)^n} \left(\frac{1}{\theta - t}\right)^{a^*} \left(1 + \frac{\gamma}{\theta - t}\right)^n = \left(\frac{\theta}{\theta - t}\right)^{a^*+n} \left(\frac{\theta - t + \gamma}{\theta + \gamma}\right)^n. \end{aligned}$$

από όπου προκύπτει το ζητούμενο αφού $M_S(t) = M_f(t)$. ■

3.2.4 Παραγωγή τυχαίου δείγματος

Αναφέραμε στην Παράγραφο 3.1 ότι η γενικευμένη κατανομή Lindley αποτελεί μίξη δύο κατανομών γάμμα. Πιο συγκεκριμένα η συνάρτηση πυκνότητας $f(x)$ της γενικευμένης κατανομής Lindley με παραμέτρους a, θ, γ ικανοποιεί τη σχέση

$$f_X(x; a, \theta, \gamma) = \frac{\theta}{\theta + \gamma} f_{X_1}(x) + \left(1 - \frac{\theta}{\theta + \gamma}\right) f_{X_2}(x)$$

όπου $X_1 \sim \text{Gamma}(a, \theta)$ και $X_2 \sim \text{Gamma}(a + 1, \theta)$ (δείτε Lawless (2003)).

Για την παραγωγή ενός τυχαίου δείγματος $X_i, i = 1, 2, \dots, n$, από τη γενικευμένη κατανομή Lindley με παραμέτρους εργαζόμαστε ως εξής:

1. Παράγουμε τυχαίο δείγμα $U_i \sim \text{Uniform}(0,1), i = 1, 2, \dots, n$.
2. Παράγουμε τυχαίο δείγμα $V_i \sim \text{Gamma}(a, \theta), i = 1, 2, \dots, n$.
3. Παράγουμε τυχαίο δείγμα $W_i \sim \text{Gamma}(a + 1, \theta), i = 1, 2, \dots, n$.
4. Εάν $U_i \leq p = \frac{\theta}{\theta + \gamma}$, θέτουμε $X_i = V_i$, διαφορετικά θέτουμε $X_i = W_i$ για $i = 1, 2, \dots, n$.

3.3 Εκτίμηση

3.3.1 Εκτιμητές μέγιστης πιθανοφάνειας

Έστω ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από την κατανομή $GL(a, \theta, \gamma)$. Η συνάρτηση πιθανοφάνειας είναι

$$\begin{aligned} L(a, \theta, \gamma) &= \prod_{i=1}^n f(x_i; a, \theta, \gamma) = \prod_{i=1}^n \frac{\theta^2 (\theta x_i)^{a-1} (a + \gamma x_i) e^{-\theta x_i}}{(\theta + \gamma) \Gamma(a + 1)} \\ &= \frac{\theta^{n(a+1)} \exp(-\theta \sum x_i) \prod (a + \gamma x_i) x_i^{a-1}}{((\theta + \gamma) \Gamma(a + 1))^n}. \end{aligned}$$

Συνεπώς

$$\begin{aligned} \ell(a, \theta, \gamma) &= \log L(a, \theta, \gamma) = n(a + 1) \log(\theta) - n \log(\gamma + \theta) - n \log \Gamma(a + 1) \\ &\quad + (a - 1) \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \log(a + \gamma x_i) - \theta \sum_{i=1}^n x_i. \end{aligned}$$

Οι παράγωγοι της $\ell(a, \theta, \gamma)$ ως προς a, θ, γ είναι αντίστοιχα:

$$\begin{aligned} \frac{\partial \ell(a, \theta, \gamma)}{\partial a} &= n \log(\theta) - n \Psi(a + 1) + \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \frac{1}{a + \gamma x_i}, \\ \frac{\partial \ell(a, \theta, \gamma)}{\partial \theta} &= \frac{n(a + 1)}{\theta} - \frac{n}{\gamma + \theta} - \sum_{i=1}^n x_i, \end{aligned}$$

$$\frac{\partial \ell(\alpha, \theta, \gamma)}{\partial \gamma} = \sum_{i=1}^n \frac{x_i}{\alpha + \gamma x_i} - \frac{n}{\gamma + \theta}.$$

όπου $\Psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}$, και $\Gamma(t)$ είναι η συνάρτηση δίζαμμα. Από τις παραπάνω σχέσεις μπορούν να προκύψουν Ε.Μ.Π. των παραμέτρων α, θ, γ .

Η κατανομή $GL(\alpha, \theta, \gamma)$ ικανοποιεί όλες τις απαραίτητες συνθήκες κανονικότητας (δείτε, π.χ., Bain (1978), σελ. 86-87] προκειμένου να εφαρμοστεί η μεθοδολογία των μεγάλων δειγμάτων. Συνεπώς το διάνυσμα $(\hat{\alpha}, \hat{\theta}, \hat{\gamma})$ ακολουθεί ασυμπτωτικά τριδιάστατη κανονική κατανομή με μέση τιμή (α, θ, γ) με πίνακα διακυμάνσεων-συνδιακυμάνσεων I^{-1} , όπου I ο πίνακας της πληροφορίας του Fischer, του οποίου τα στοιχεία προκύπτουν από τις παρακάτω σχέσεις:

$$-E\left(\frac{\partial^2 l}{\partial \alpha^2}\right) = n\Psi'(1 + \alpha) + nJ_0(\alpha, \theta, \gamma),$$

$$-E\left(\frac{\partial^2 l}{\partial \alpha \partial \theta}\right) = \frac{-n}{\theta},$$

$$-E\left(\frac{\partial^2 l}{\partial \alpha \partial \gamma}\right) = nJ_1(\alpha, \theta, \gamma),$$

$$-E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = \frac{n(\alpha + 1)}{\theta^2} - \frac{n}{(\gamma + \theta)^2},$$

$$-E\left(\frac{\partial^2 l}{\partial \theta \partial \gamma}\right) = -\frac{n}{(\gamma + \theta)^2},$$

$$-E\left(\frac{\partial^2 l}{\partial \gamma^2}\right) = nJ_2(\alpha, \theta, \gamma) - \frac{n}{(\gamma + \theta)^2},$$

όπου

$$J_i(\alpha, \theta, \gamma) = E\left(\frac{X^i}{(\alpha + \gamma X)^2}\right) = \frac{\theta^{1-i}}{a\Gamma(a+1)(\gamma + \theta)} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left(\frac{\gamma}{a\theta}\right)^k \Gamma(a + k + i).$$

3.3.2 Προσαρμογή σε δεδομένα

Για να μελετήσουμε την προσαρμογή της κατανομή $GL(a, \theta, \gamma)$ θα χρησιμοποιήσουμε δύο σετ δεδομένων χρόνων ζωής του Lawless (δείτε Lawless (2003, pp. 204, 263)) που δίνονται στους δύο παρακάτω πίνακες.

Πίνακας 3-1

Σετ δεδομένων 1: Χρόνοι αποτυχίας (σε min) 15 ηλεκτρονικών εξαρτημάτων

1.4	5.1	6.3	10.8	12.1
18.5	19.7	22.2	23	30.6
37.3	46.3	53.9	59.8	66.2

Πίνακας 3-2

Σετ δεδομένων 2: Αριθμός κύκλων έως την αποτυχία 25 νημάτων ενός μέτρου σε συγκεκριμένο επίπεδο πίεσης

15	20	38	42	61
76	86	98	121	146
149	157	175	176	180
180	198	220	224	251
264	282	321	325	653

Στα παραπάνω δεδομένα, πέρα από τη γενικευμένη κατανομή Lindley με παραμέτρους a, θ, γ ($GL(a, \theta, \gamma)$) θα προσαρμοστούν και οι κατανομές Γάμμα, Weibull και lognormal με συναρτήσεις πυκνότητας

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

$$f_w(x; \alpha, \theta) = \left(\frac{\alpha}{\theta}\right) \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\theta}\right)^\alpha\right),$$

$$f_{ln}(x; \alpha, \theta) = \frac{1}{\sqrt{2\pi\alpha x}} e^{-\frac{1}{2}\left(\frac{\log x - \theta}{\alpha}\right)^2},$$

αντίστοιχα.

Στον παρακάτω πίνακα βλέπουμε τις εκτιμήσεις των παραμέτρων, την τιμή της Log-Likelihood (LL) συνάρτησης και την τιμή της στατιστικής συνάρτησης ελέγχου του τεστ Kolmogorov-Smirnov (K-S). Παρατηρούμε πως η γενικευμένη κατανομή Lindley συναγωνίζεται σε ικανοποιητικό βαθμό τις κατανομές Γάμμα, Weibull και lognormal.

Πίνακας 3-3

EMPI, Log-Likelihood και K-S έλεγχος για τα δεδομένα των Πινάκων 3-1 και 3-2

Σετ Δεδομένων	Κατανομή	a	θ	γ	LL	$K - S$
1	GL	1.200	0.064	0.083	-64.080	0.095
	Gamma	1.442	0.052	-	-64.186	0.103
	Weibull	1.306	29.8	-	-64.020	0.099
	Lognormal	1.026	2.931	-	-65.617	0.162
2	GL	1.531	0.012	0.018	-152.371	0.129
	Gamma	1.794	0.010	-	-152.438	0.135
	Weibull	1.414	196.155	-	-152.443	0.122
	Lognormal	0.874	4.880	-	-154.086	0.187

Ο κώδικας στη γλώσσα προγραμματισμού *R* για την εξαγωγή των αριθμητικών αποτελεσμάτων του παραπάνω πίνακα δίνεται στο Παράρτημα.

ΚΕΦΑΛΑΙΟ 4

Η γενικευμένη Poisson-Lindley κατανομή

4.1 Εισαγωγή

Στην ασφαλιστική αγορά, η αναλογιστική επιστήμη για τον υπολογισμό των ασφαλίσεων, υιοθετεί μία κατανομή για τη ζημία η οποία μπορεί να επαληθευθεί ή όχι. Κατά συνέπεια, η αναλογιστική επιστήμη αντιμετωπίζει την πρόκληση να βρει την κατάλληλη κατανομή που προσαρμόζεται καλύτερα στα πραγματικά δεδομένα. Εκτός από την καλή προσαρμογή στα δεδομένα, η προτεινόμενη κατανομή θα πρέπει να έχει κάποια συγκεκριμένα χαρακτηριστικά όπως το σχήμα της και η συμπεριφορά της ουράς της. Η κατανομή μπορεί να παρουσιάζει βαριά ή ελαφριά ουρά. Όταν η ουρά πλησιάζει αργά στο μηδέν τότε οι ακραίες τιμές μπορούν να δικαιολογηθούν. Άρα εκτός από την καλή προσαρμογή στα εμπειρικά δεδομένα, στο σχήμα της κατανομής θα πρέπει να αποτυπώνεται η συμπεριφορά των ακραίων τιμών. Το παραπάνω κρίσιμο ζήτημα αποτέλεσε το κίνητρο για την έρευνα μιας διακριτής κατανομής η οποία να τείνει στο μηδέν πιο αργά σε σχέση με την κλασική σύνθετη Poisson-Lindley κατανομή (δείτε Sankaran (1970)).

Σε αυτή την ενότητα, θα εξετάσουμε μια γενικευμένη Poisson-Lindley κατανομή, η οποία πηγάζει από την Poisson κατανομή όταν η παράμετρος της λ ακολουθεί τη διπαραμετρική Lindley κατανομή την οποία συμβολίζουμε με $TPLD(\theta, a)$ (δείτε Shanker et al. (2013)). Η σ.π. της κατανομής $TPLD(\theta, a)$ δίνεται από τη σχέση

$$f(x; a, \theta) = \frac{\theta^2}{(\theta + a)} (1 + ax)e^{-\theta x}, \quad x, a, \theta > 0.$$

Για $a = 1$ προκύπτει η κατανομή $Lindley(\theta)$, δηλαδή

$$TPLD(\theta, 1) \equiv Lindley(\theta).$$

Επίσης αν θέσουμε $a = 1$ στη γενικευμένη κατανομή Lindley με παραμέτρους a, θ, γ (δείτε Κεφάλαιο 3) τότε προκύπτει η κατανομή $TPLD(\theta, a)$, δηλαδή

$$GL(1, \theta, \gamma) \equiv TPLD(\theta, \gamma).$$

Στην Ενότητα 4.2, θα εισάγουμε την νέα κατανομή και στην Ενότητα 4.3 θα παρουσιαστούν οι βασικές ιδιότητές της. Στην Ενότητα 4.4 θα γίνει εκτίμηση των παραμέτρων της κατανομής με τη μέθοδο των ροπών και την μέθοδο Ε.Μ.Π. Στην Ενότητα 4.4 θα μελετηθεί το συλλογικό μοντέλο κινδύνου και στην Ενότητα 4.7 θα δούμε την αναλογιστική εφαρμογή της νέας κατανομής σε πραγματικά δεδομένα ένα εκ των οποίων αναφέρεται στον κλάδο της ασφάλισης αυτοκινήτου.

Τα κυριότερα αποτελέσματα αυτού του Κεφαλαίου βρίσκονται στην εργασία των Bhati et al. (2015).

4.2 Η γενικευμένη Poisson-Lindley κατανομή

Θα λέμε ότι μια τ.μ. X έχει τη γενικευμένη Poisson-Lindley κατανομή με παραμέτρους θ και a ($\theta, a > 0$), και θα τη συμβολίζουμε με $NGPL(\theta, a)$, εάν η συνάρτηση πιθανότητάς της $f(x; a, \theta)$ δίνεται από τον τύπο

$$f(x; a, \theta) = \frac{\theta^2}{(\theta + a)(1 + \theta)^{x+1}} \left(1 + \frac{a(x + 1)}{1 + \theta} \right), \quad x = 0, 1, 2, \dots$$

Για τη γέννηση της κατανομής $NGPL(\theta, a)$ σχετικό είναι το ακόλουθο θεώρημα.

Θεώρημα 4.1

Αν

$$X|\lambda \sim P(\lambda)$$

$$\lambda|\theta, a \sim TPLD(\theta, a)$$

τότε

$$X \sim NGPL(\theta, a).$$

Απόδειξη

Η συνάρτηση πιθανότητας της διακριτής τ.μ. X δίνεται από τον τύπο

$$\Pr(X = x) = \int_0^{\infty} \Pr(X = x | \lambda) f(\lambda; \theta, a) d\lambda = \int_0^{\infty} \frac{\theta^2 e^{-\lambda} \lambda^x (1 + a\lambda) e^{-\theta\lambda}}{(\theta + a)x!} d\lambda$$

$$\begin{aligned}
&= \frac{\theta^2}{(\theta + a)x!} \int_0^{\infty} e^{-\lambda(1+\theta)} \lambda^x d\lambda + \frac{a\theta^2}{(\theta + a)x!} \int_0^{\infty} e^{-\lambda(1+\theta)} \lambda^{x+1} d\lambda \\
&= \frac{\theta^2}{(\theta + a)x! (1 + \theta)^{x+1}} \int_0^{\infty} e^{-\lambda(1+\theta)} [\lambda(1 + \theta)]^x d[\lambda(1 + \theta)] \\
&\quad + \frac{a\theta^2}{(\theta + a)x! (1 + \theta)^{x+2}} \int_0^{\infty} e^{-\lambda(1+\theta)} [\lambda(1 + \theta)]^{x+1} d[\lambda(1 + \theta)] \\
&= \frac{\theta^2}{(\theta + a)x!} \frac{\Gamma(x + 1)}{(1 + \theta)^{x+1}} + \frac{a\theta^2}{(\theta + a)x!} \frac{\Gamma(x + 2)}{(1 + \theta)^{x+2}} \\
&= \frac{\theta^2}{(\theta + a)(1 + \theta)^{x+1}} \left(\frac{\Gamma(x + 1)}{x!} + \frac{a\Gamma(x + 2)}{x!(1 + \theta)} \right) \\
&= \frac{\theta^2}{(\theta + a)(1 + \theta)^{x+1}} \left(1 + \frac{a(x + 1)}{1 + \theta} \right)
\end{aligned}$$

■

Αξίζει να σημειώσουμε ότι:

- i. Για $a \rightarrow 0$, η συνάρτηση πιθανότητας της κατανομής $NGPL(\theta, a)$ ανάγεται στην γεωμετρική κατανομή με πιθανότητα επιτυχίας $\frac{\theta}{1+\theta}$, δηλαδή

$$NGPL(\theta, 0) \equiv G\left(\frac{\theta}{1 + \theta}\right).$$

- ii. Για $a = 1$, η συνάρτηση πιθανότητας της κατανομής $NGPL(\theta, a)$ ανάγεται στη συνάρτηση πιθανότητας της κατανομής Poisson-Lindley που προτάθηκε από τον Shankaran (1970). Η κατανομή Poisson-Lindley με παράμετρο θ (συμβολισμός $PL(\theta)$) έχει συνάρτηση πιθανότητας

$$f(x; \theta) = \frac{\theta^2(\theta + x + 2)}{(1 + \theta)^{x+3}}, \quad x = 0, 1, 2, \dots$$

Συνεπώς

$$NGPL(\theta, 1) \equiv PL(\theta).$$

- iii. Για $a \rightarrow \infty$ η συνάρτηση πιθανότητας της κατανομής $NGPL(\theta, a)$ ανάγεται στην αρνητική διωνυμική κατανομή με παραμέτρους $r = 2$ και πιθανότητα επιτυχίας $\frac{\theta}{1+\theta}$, δηλαδή

$$NGPL(\theta, \infty) \equiv NB\left(2, \frac{\theta}{1+\theta}\right).$$

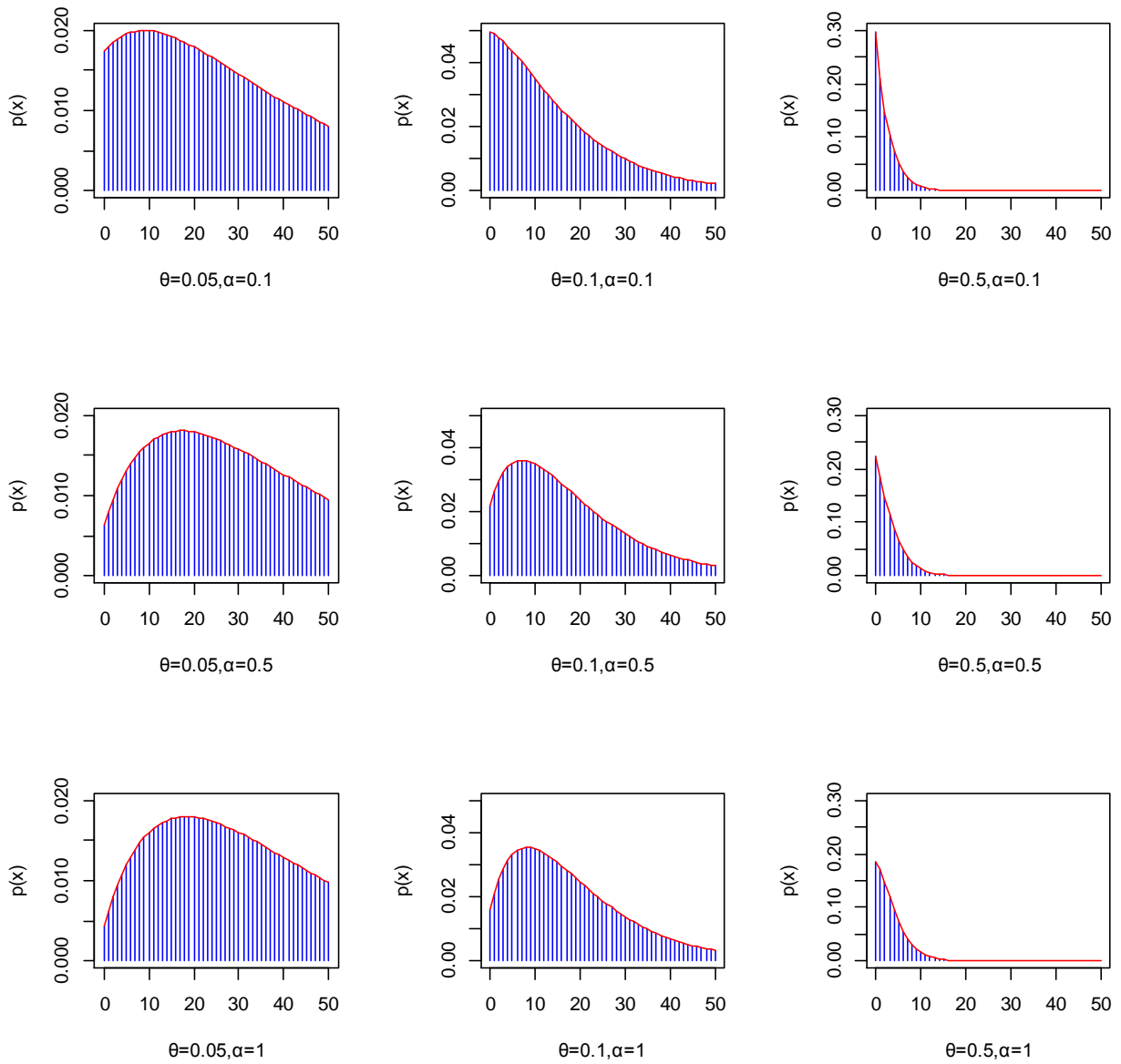
- iv. Η κατανομή $NGPL(\theta, a)$ αποτελεί μίξη της κατανομής $G\left(\frac{\theta}{1+\theta}\right)$ και $NB\left(2, \frac{\theta}{1+\theta}\right)$ με αναλογία $\frac{\theta}{\theta+a}$. Πράγματι

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

Στο Σχήμα 4-1 δίνονται διάφορες γραφικές παραστάσεις της συνάρτησης πιθανότητας της κατανομής $NGPL(\theta, a)$ για διάφορες τιμές των παραμέτρων θ και a . Παρατηρούμε πως όταν η παράμετρος θ παραμένει σταθερή και η παράμετρος a αυξάνει, η ουρά της κατανομής γίνεται πιο «βαριά». Επίσης όταν η παράμετρος a είναι σταθερή και η παράμετρος θ αυξάνει, η ουρά γίνεται πιο «ελαφριά».

Σχήμα 4-1

Η συνάρτηση μάζας πιθανότητας για διάφορες τιμές των παραμέτρων (θ, a)



Η συνάρτηση κατανομής της της κατανομής $NGPL(\theta, a)$ δίνεται από τον τύπο

$$F_x(x) = \sum_{n=0}^x \frac{\theta^2}{(\theta + a)} \frac{(1 + \theta + an + a)}{(1 + \theta)^{n+2}}$$

$$= \frac{(a + \theta)(\theta + 1)^{x+2} - (2a\theta + a + \theta^2 + \theta + a\theta x)}{(\theta + 1)^{x+2}(a + \theta)}.$$

4.3 Ιδιότητες της γενικευμένης Poisson-Lindley κατανομής

4.3.1 Σχήμα

Για $x = 0$, έχουμε ότι

$$p(0) = \frac{\theta^2(1+a+\theta)}{(a+\theta)(1+\theta)^2}$$

και για $x = 1, 2, 3, \dots$ έχουμε ότι

$$\frac{p(x+1)}{p(x)} = \frac{1+(2+x)a+\theta}{(1+\theta)(1+(1+x)a+\theta)}$$

Σημειώνουμε ότι:

- i. Για $\theta > 1$, η κατανομή $NGPL(\theta, a)$ είναι μονοκόρυφη με κορυφή στο 0.
- ii. Για $0 < \theta \leq 1$ και $a \geq \frac{\theta(\theta+1)}{1-\theta}$, η κατανομή $NGPL(\theta, a)$ είναι μονοκόρυφη με κορυφή στο σημείο $[x^*] + 1$, όπου $[x^*]$ είναι το ακέραιο μέρος του

$$x^* = \frac{a - \theta - a\theta - \theta^2}{a\theta}$$

- iii. Για $0 < \theta \leq 1$ και $a < \frac{\theta(\theta+1)}{1-\theta}$, η κατανομή $NGPL(\theta, a)$ είναι μονοκόρυφη με κορυφή στο 0.

Επιπλέον, εάν ορίσουμε τον δείκτη

$$R = \frac{\Pr(N_1 = n; \theta, a)}{\Pr(N_2 = n; \theta)} = \frac{1 + \frac{1+ax}{a+\theta}}{1 + \frac{1+x}{1+\theta}}$$

όπου $N_1 \sim NGPL(\theta, a)$ και $N_2 \sim PL(\theta)$, τότε η $NGPL(\theta, a)$ θα έχει πιο «βαριά» («ελαφριά») δεξιά ουρά σε σύγκριση με την $PL(\theta)$ εάν $R > (<) 1$ για $a > (<) 1$ και $x > (<) x_0 = 1/\theta$.

4.3.2 Γεννήτριες συναρτήσεις

Η πιθανογεννήτρια συνάρτηση $P(t)$ της κατανομής $NGPL(\theta, a)$ δίνεται από τον τύπο

$$P(t) = \sum_{x=0}^{\infty} t^x \frac{\theta^2}{(\theta+a)(1+\theta)^{x+1}} \left(1 + \frac{a(x+1)}{1+\theta}\right)$$

$$\begin{aligned}
&= \left(\frac{\theta^2}{\theta + a} \right) \left(\frac{a\theta + a + \theta^2 + 2\theta - \theta t - t + 1}{(\theta + 1)(\theta - t + 1)^2} \right) \\
&= \frac{\theta^2(a + \theta - t + 1)}{(a + \theta)(\theta - t + 1)^2}
\end{aligned}$$

Η ροπογεννήτρια συνάρτηση $M(t)$ της κατανομής $NGPL(\theta, a)$ δίνεται προφανώς από τον τύπο

$$M(t) = \frac{\theta^2(a + \theta - e^t + 1)}{(a + \theta)(\theta - e^t + 1)^2}.$$

4.3.3 Ροπές, λοξότητα και κύρτωση

Η ροπή πρώτης, δεύτερης τρίτης και τέταρτης τάξης της κατανομής $NGPL(\theta, a)$ είναι αντίστοιχα

$$\begin{aligned}
\mu'_1 &= \frac{2a + \theta}{\theta(a + \theta)} \\
\mu'_2 &= \frac{2a(\theta + 3) + \theta(\theta + 2)}{\theta^2(a + \theta)} \\
\mu'_3 &= \frac{\theta(\theta^2 + 6\theta + 6) + 2a(\theta^2 + 9\theta + 12)}{\theta^3(a + \theta)} \\
\mu'_4 &= \frac{2a(\theta^3 + 21\theta^2 + 72\theta + 60) + \theta(\theta^3 + 14\theta^2 + 36\theta + 24)}{\theta^4(a + \theta)}
\end{aligned}$$

Οι κεντρικές ροπές δεύτερης και τρίτης τάξης της κατανομής της κατανομής $NGPL(\theta, a)$ είναι αντίστοιχα

$$\begin{aligned}
\mu_2 &= \frac{2a^2(1 + \theta) + \theta^2(1 + \theta) + a\theta(4 + 3\theta)}{\theta^2(a + \theta)^2} \\
\mu_3 &= \frac{2a^3(\theta^2 + 3\theta + 2) + a^2\theta(5\theta^2 + 18\theta + 12) + a\theta^2(4\theta^2 + 15\theta + 12) + \theta^3(\theta^2 + 3\theta + 2)}{\theta^3(a + \theta)^3}
\end{aligned}$$

Σύμφωνα με τα παραπάνω ο συντελεστής λοξότητας a_3 και ο συντελεστής κύρτωσης a_4 που δίνονται, αντίστοιχα, από τους τύπους

$$a_3 = \frac{\mu_3}{\mu_2^{3/2}}, \quad a_4 = \frac{\mu_4}{\mu_2^2}$$

μπορούν να υπολογιστούν σχετικά εύκολα. Τιμές αυτών των συντελεστών για διάφορες τιμές των παραμέτρων της κατανομής $NGPL(\theta, a)$ δίνονται στους Πίνακες 4-1 και 4-2 (δείτε Παράρτημα). Γραφικές παραστάσεις των τιμών των συντελεστών a_3 και a_4 δίνονται στα Σχήματα 4.2 και 4.3 (δείτε Παράρτημα).

Πίνακας 4-1

Συντελεστής λοξότητας της κατανομής $NGPL(\theta, a)$ για διάφορες τιμές των παραμέτρων θ, a

a_3	$a = 0.5$	$a = 1$	$a = 1.5$	$a = 2$	$a = 2.5$	$a = 3$
$\theta = 0.25$	1.556	1.488	1.464	1.453	1.446	1.442
$\theta = 0.5$	1.708	1.598	1.550	1.525	1.509	1.498
$\theta = 1.0$	1.916	1.792	1.723	1.679	1.650	1.628
$\theta = 1.5$	2.063	1.949	1.875	1.824	1.787	1.759
$\theta = 2.0$	2.186	2.083	2.010	1.956	1.915	1.884
$\theta = 2.5$	2.297	2.203	2.132	2.078	2.035	2.001
$\theta = 3.0$	2.401	2.314	2.246	2.191	2.148	2.112

Πίνακας 4-2

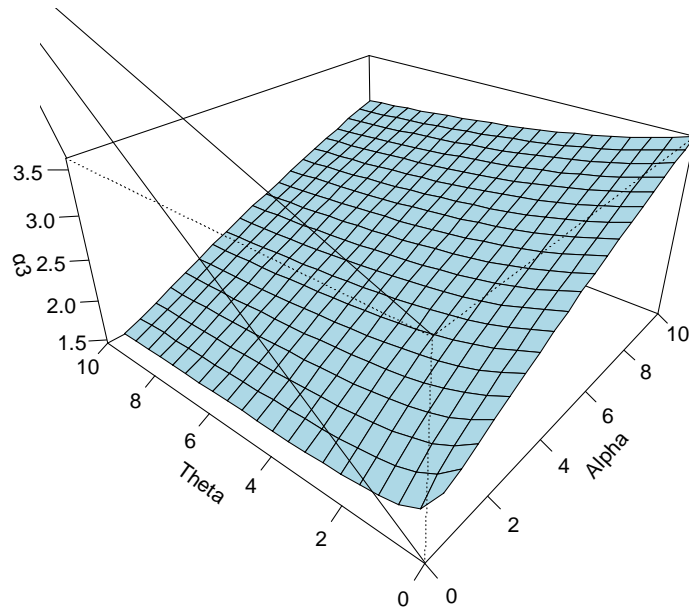
Συντελεστής κύρτωσης της κατανομής $NGPL(\theta, a)$ για διάφορες τιμές των παραμέτρων θ, a

a_4	$a = 0.5$	$a = 1$	$a = 1.5$	$a = 2$	$a = 2.5$	$a = 3$
$\theta = 0.25$	6.511	6.245	6.161	6.122	6.100	6.085
$\theta = 0.5$	7.180	6.672	6.475	6.373	6.312	6.272
$\theta = 1.0$	8.210	7.532	7.187	6.984	6.851	6.758
$\theta = 1.5$	8.957	8.282	7.875	7.609	7.425	7.289
$\theta = 2.0$	9.573	8.941	8.513	8.213	7.994	7.827
$\theta = 2.5$	10.129	9.542	9.111	8.792	8.549	8.360
$\theta = 3.0$	10.655	10.106	9.679	9.349	9.091	8.884

Σχήμα 4-2

Συντελεστής λοξότητας της κατανομής $NGPL(\theta, a)$ για διάφορες τιμές των παραμέτρων θ, a

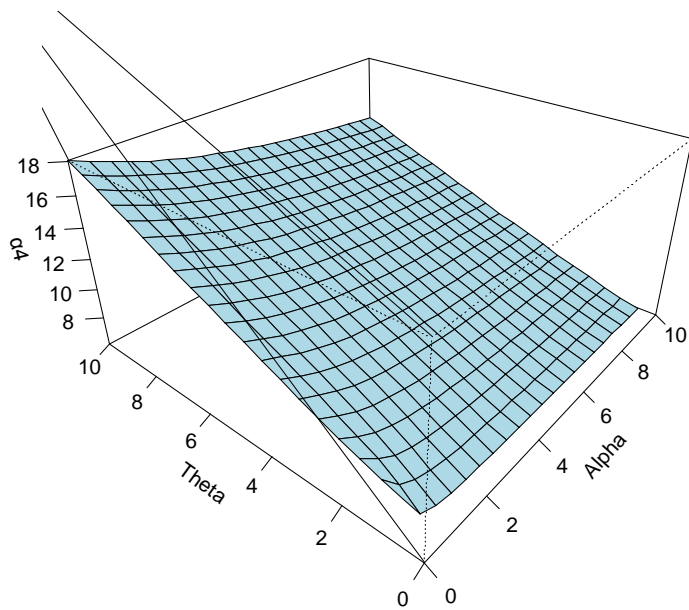
ΣΥΝΤΕΛΕΣΤΗΣ ΛΟΞΟΤΗΤΑΣ



Σχήμα 4-3

Συντελεστής κύρτωσης της κατανομής $NGPL(\theta, a)$ για διάφορες τιμές των παραμέτρων θ, a

ΣΥΝΤΕΛΕΣΤΗΣ ΚΥΡΤΩΣΗΣ



Ο συντελεστής μεταβλητότητας (C.V.) δίνεται από τον τύπο

$$C.V. = \frac{\sigma}{\mu}$$

και είναι ίσος με

$$\begin{aligned} C.V. &= \left(\frac{\theta(a + \theta)}{2a + \theta} \right) \sqrt{\frac{2a^2(1 + \theta) + \theta^2(1 + \theta) + a\theta(4 + 3\theta)}{\theta^2(a + \theta)^2}} \\ &= \frac{\sqrt{2a^2(1 + \theta) + \theta^2(1 + \theta) + a\theta(4 + 3\theta)}}{(2a + \theta)}. \end{aligned}$$

Για το συντελεστή $r = \frac{\sigma^2}{\mu}$ έχουμε

$$\begin{aligned} r &= \left(\frac{2a^2(1 + \theta) + \theta^2 + a\theta(4 + 3\theta)}{\theta^2(a + \theta)^2} \right) \left(\frac{\theta(a + \theta)}{2a + \theta} \right) \\ &= \frac{2a^2(\theta + 1) + a\theta(3\theta + 4) + \theta^2(\theta + 1)}{\theta(a + \theta)(2a + \theta)}. \end{aligned}$$

Μπορεί ναδειχθεί ότι $r > 1$ και έτσι η κατανομή $NGPL(\theta, a)$ παρουσιάζει υπερσκέδαση (over dispersed).

4.4 Αλγόριθμος παραγωγής τυχαίου δείγματος

Αν $X \sim NGPL(\theta, a)$, τότε η X προκύπτει από την κατανομή $P(\lambda)$ όπου η λ ακολουθεί την κατανομή $TPLD(\theta, a)$ η οποία μπορεί να αναπαρασταθεί ως μίξη δύο ανεξάρτητων τ.μ. με $V \sim Exp(\theta)$ και $W \sim Gamma(2, \theta)$ με παράμετρο μίξης την $p = \frac{\theta}{\theta + a}$.

Για την παραγωγή ενός τυχαίου δείγματος $X_i, i = 1, 2, \dots, n$, από την κατανομή $NGPL(\theta, a)$ εργαζόμαστε ως εξής:

1. Παράγουμε τυχαίο δείγμα $U_i \sim Uniform(0,1), i = 1, 2, \dots, n$.
2. Παράγουμε τυχαίο δείγμα $V_i \sim Exp(\theta), i = 1, 2, \dots, n$.
3. Παράγουμε τυχαίο δείγμα $W_i \sim Gamma(2, \theta), i = 1, 2, \dots, n$.
4. Εάν $U_i \leq p = \frac{\theta}{\theta + a}$, θέτουμε $\lambda_i = V_i$, διαφορετικά θέτουμε $\lambda_i = W_i, i = 1, 2, \dots, n$.
5. Παράγουμε τυχαίο δείγμα $X_i, i = 1, 2, \dots, n$, από την $P(\lambda_i)$.

4.5 Εκτίμηση

4.5.1 Μέθοδος των ροπών

Έστω ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από την κατανομή $NGPL(\theta, a)$. Οι εκτιμήτριες των παραμέτρων με τη μέθοδο των ροπών προκύπτουν ως εξής:

$$m_1 = \mu'_1 = \frac{2a + \theta}{\theta(a + \theta)}$$
$$m_2 = \mu'_2 = \frac{2a(\theta + 3) + \theta(\theta + 2)}{\theta^2(a + \theta)}$$

όπου m_1 και m_2 η δειγματική ροπή πρώτης και δεύτερης τάξης. Επιλύοντας τις παραπάνω εξισώσεις παίρνουμε τις εκτιμήτριες

$$\tilde{a} = \frac{\tilde{\theta} - m_1 \tilde{\theta}^2}{m_1 \tilde{\theta} - 2}$$

και

$$\tilde{\theta} = \frac{2m_1 + \sqrt{4m_1^2 + 2m_1 - 2m_2}}{m_2 - m_1}$$

Πρόταση 4.2

Για σταθερό a , η εκτιμήτρια ροπών $\tilde{\theta}$ της θ έχει θετική μεροληψία δηλαδή ισχύει $E(\tilde{\theta}) > \theta$.

Απόδειξη

Παρατηρούμε ότι $\tilde{\theta} = g(\bar{X})$ όπου

$$g(t) = \frac{1 - at + \sqrt{1 + 6at + a^2 t^2}}{2t}, \quad t > 0.$$

Αφού

$$g''(t) = \frac{1}{t^3} + \frac{9at + 3a^3 t^3 + 15a^2 t^2 + 1}{t^3(1 + 6at + a^2 t^2)^{3/2}} > 0$$

η $g(t)$ είναι αυστηρώς κυρτή. Έτσι, από την ανισότητα του Jensen έχουμε ότι $E\{g(\bar{X})\} > g\{E(\bar{X})\}$. Τέλος, επειδή

$$g\{E(\bar{X})\} = g\left(\frac{2a + \theta}{\theta(a + \theta)}\right) = \theta$$

προκύπτει άμεσα το ζητούμενο. ■

Πρόταση 4.3

Για σταθερές τιμές της παραμέτρου a , για την εκτιμήτρια με τη μέθοδο των ροπών $\tilde{\theta}$ της παραμέτρου θ ισχύει ότι είναι συνεπής και

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N(0, v^2(\theta))$$

όπου

$$v^2(\theta) = \frac{\theta^2(a + \theta)^2(2a^2\theta + 2a^2 + 3a\theta^2 + 4a\theta + \theta^3 + \theta^2)}{(2a^2 + 4a\theta + \theta^2)^2}.$$

Απόδειξη

Συνέπεια: Αφού $\mu < \infty$, $\bar{X} \xrightarrow{P} \mu$. Επειδή η

$$g(t) = \frac{1 - at + \sqrt{1 + 6at + a^2t^2}}{2t}$$

είναι συνεχής συνάρτηση για $t = \mu$ τότε $g(\bar{X}) \xrightarrow{P} g(\mu)$, δηλαδή $\tilde{\theta} \xrightarrow{P} \theta$.

Ασυμπτωτική κατανομή: Εφόσον $\sigma^2 < \infty$, τότε από το Κ.Ο.Θ. θα έχουμε

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Επίσης, καθώς $g(\mu)$ διαφορίσιμη και $g'(\mu) \neq 0$ από τη μέθοδο Δέλτα έχουμε

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} N(0, v^2(\theta))$$

όπου

$$v^2(\theta) = \left[g' \left(\frac{2a + \theta}{\theta(a + \theta)} \right) \right]^2 \sigma^2$$

δηλαδή

$$v^2(\theta) = \frac{\theta^2(a + \theta)^2(2a^2\theta + 2a^2 + 3a\theta^2 + 4a\theta + \theta^3 + \theta^2)}{(2a^2 + 4a\theta + \theta^2)^2}$$

Η απόδειξη ολοκληρώθηκε. ■

Ως αποτέλεσμα της ασυμπτωτικής κανονικότητας το διάστημα εμπιστοσύνης για την παράμετρο θ δίνεται από τη σχέση

$$\hat{\theta} \pm z_{\alpha/2} \frac{v(\hat{\theta})}{\sqrt{n}}.$$

4.5.2 Μέθοδος μέγιστης πιθανοφάνειας

Έστω ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από την κατανομή $NGPL(\theta, a)$. Ο λογάριθμος της συνάρτησης πιθανοφάνειας δίνεται από τον τύπο

$$\ell_n(a, \theta) = 2n \log(\theta) - n \log(\theta + a) - \sum_{i=1}^n (x_i + 1) \log(1 + \theta) + \sum_{i=1}^n \log\left(1 + \frac{a(x_i + 1)}{1 + \theta}\right)$$

Οι μερικές παράγωγοι ως προς θ και a είναι, αντίστοιχα,

$$\frac{\partial \ell_n(a, \theta)}{\partial \theta} = \frac{2}{n} - \frac{n}{\theta + a} - \frac{(\sum_{i=1}^n x_i) + n}{1 + \theta} - \sum_{i=1}^n \frac{a(x_i + 1)}{(1 + \theta + a(x_i + 1))(1 + \theta)}$$

$$\frac{\partial \ell_n(a, \theta)}{\partial a} = \frac{-n}{\theta + a} + \sum_{i=1}^n \frac{(x_i + 1)}{(1 + \theta + a(x_i + 1))}$$

Από το σύστημα

$$\frac{\partial \ell_n(a, \theta)}{\partial \theta} = 0, \quad \frac{\partial \ell_n(a, \theta)}{\partial a} = 0$$

λαμβάνουμε τους εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων θ, a που είναι

$$\hat{a} = \frac{\hat{\theta}^2 \bar{x} - \hat{\theta}}{2 - \hat{\theta} \bar{x}}, \quad \frac{-n}{\hat{\theta} + \hat{a}} + \sum_{i=1}^n \left(\frac{x_i + 1}{1 + \hat{\theta} + \hat{a}(x_i + 1)} \right) = 0$$

Οι μερικές παράγωγοι δευτέρας τάξης θα μας δώσουν

$$\frac{\partial^2 \ell_n(a, \theta)}{\partial \theta^2} = \sum_{i=1}^n \frac{a(x_i + 1)(a + 2\theta + ax_i + 2)}{(\theta + 1)^2(a + \theta + ax_i + 1)^2}$$

$$+ \sum_{i=1}^n \frac{x_i + 1}{(\theta + 1)^2} + \frac{n}{(a + \theta)^2} - \frac{2n}{\theta^2},$$

$$\frac{\partial^2 \ell_n(a, \theta)}{\partial \theta \partial a} = \frac{n}{(a + \theta)^2} - \sum_{i=1}^n \frac{x_i + 1}{(a + \theta + ax_i + 1)^2},$$

$$\frac{\partial^2 \ell_n(a, \theta)}{\partial a^2} = - \sum_{i=1}^n \frac{(x_i + 1)^2}{(\theta + 1)^2 \left(\frac{a(x_i + 1)}{\theta + 1} + 1 \right)^2} + \frac{n}{(a + \theta)^2}.$$

4.6 Το συλλογικό μοντέλο κινδύνου

Στις ασφαλίσεις, εκτός του κλάδου ζωής, όπως είναι οι ασφαλίσεις του κλάδου αυτοκινήτου, το συνολικό ποσό των ζημιών σε συγκεκριμένη χρονική περίοδο είναι μια τυχαία μεταβλητή που αποτελείται από το άθροισμα των ατομικών ζημιών σε αυτή τη χρονική περίοδο. Έστω N ο αριθμός των ζημιών σε μια συγκεκριμένη χρονική περίοδο ο οποίος είναι μια τυχαία μεταβλητή και έστω X_i ($i = 1, 2, \dots$) η τυχαία μεταβλητή που δηλώνει το μέγεθος της i ζημιάς. Τότε το συνολικό ποσό των ζημιών S δίνεται από τη σχέση

$$S = \begin{cases} X_1 + X_2 + \dots + X_N, & N = 1, 2, \dots \\ 0, & N = 0. \end{cases}$$

Στη συνέχεια υποθέτουμε ότι οι τυχαίες μεταβλητές X_1, X_2, \dots είναι ισόνομες και ανεξάρτητες από την τυχαία μεταβλητή N .

Η συνάρτηση πυκνότητας της S δίνεται από τον τύπο

$$f_S(x) = \sum_{n=0}^{\infty} p_n f^{*n}(x), \quad x \geq 0,$$

όπου $p_n = P(N = n)$, $n = 0, 1, \dots$, η συνάρτηση πιθανότητας της N και f^{*n} η n -οστή συνέλιξη που αντιστοιχεί στη συνάρτηση πυκνότητας $f(x)$ του ύψους μιας ζημιάς.

Πρόταση 4.4

Υποθέτουμε πως ο αριθμός των ζημιών N έχει την κατανομή $NGPL(\theta, a)$ και ότι το ύψος μιας ζημιάς X έχει την κατανομή $Exp(\lambda)$. Τότε η συνάρτηση πυκνότητας των συνολικών ζημιών S δίνεται από τον τύπο

$$f_S(x) = \frac{\theta^2 \lambda e^{-\theta \lambda x / (1+\theta)} (2a\theta + 2a + \theta^2 + 2\theta + a\lambda x + 1)}{(1+\theta)^4 (\theta + a)}, \quad x > 0$$

και

$$f_S(0) = \frac{\theta^2 (1 + a + \theta)}{(\theta + a)(1 + \theta)^2}.$$

Απόδειξη.

Έστω $N \sim NGPL(\theta, a)$ και $X_i \sim Exp(\lambda)$, $i = 1, 2, \dots$. Η n -οστή συνέλιξη $f^{*n}(x)$ της εκθετικής κατανομής δίνεται από τον τύπο

$$f^{*n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad n = 1, 2, \dots$$

που είναι η συνάρτηση πυκνότητας της κατανομής $Gamma(n, \lambda)$.

Έτσι για τη συνάρτηση πυκνότητας της τ.μ. S για $x > 0$ έχουμε

$$\begin{aligned} f_S(x) &= \sum_{n=1}^{\infty} \frac{\theta^2}{(\theta + a)(1 + \theta)^{n+1}} \left(1 + \frac{a(n+1)}{1 + \theta}\right) \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \\ &= \left(\frac{\theta^2 e^{-\lambda x}}{\theta + a}\right) \sum_{n=1}^{\infty} \frac{\lambda^n}{(1 + \theta)^{n+1} (n-1)!} x^{n-1} \left(1 + \frac{a(n+1)}{1 + \theta}\right) \\ &= \left(\frac{\theta^2 e^{-\lambda x}}{\theta + a}\right) \frac{e^{\lambda x/(1+\theta)} \lambda (1 + 2a + xa\lambda + 2\theta + 2a\theta + \theta^2)}{(1 + \theta)^4} \\ &= \frac{\theta^2 \lambda e^{-\theta \lambda x/(1+\theta)} (1 + 2a + xa\lambda + 2\theta + 2a\theta + \theta^2)}{(\theta + a)(1 + \theta)^4}. \end{aligned}$$

Η πιθανότητα για μηδενικό ύψος αποζημιώσεων είναι

$$f_S(0) = P(N = 0) = \frac{\theta^2 (1 + a + \theta)}{(\theta + a)(1 + \theta)^2}.$$

Η απόδειξη ολοκληρώθηκε. ■

Η μέση τιμή της S (δείτε Χατζηκωνσταντινίδης (2015)) ικανοποιεί τη σχέση $E(S) = E(X)E(N)$ οπότε

$$E(S) = \frac{2a + \theta}{\lambda \theta (a + \theta)}.$$

Πρόταση 4.5

Υποθέτουμε πως ο αριθμός των ζημιών N έχει την κατανομή $NGPL(\theta, a)$ και ότι το ύψος μιας ζημιάς X έχει την κατανομή $Gamma(2, \lambda)$. Τότε η συνάρτηση πυκνότητας των συνολικών ζημιών S δίνεται από τον τύπο

$$f_S(x) = \frac{\theta^2 e^{-\lambda x} \left[(\theta + 1)(3a + 2\theta + 2) \sinh\left(\frac{\lambda x}{\sqrt{\theta + 1}}\right) + a\sqrt{\theta + 1} \lambda x \cosh\left(\frac{\lambda x}{\sqrt{\theta + 1}}\right) \right]}{2(\theta + a)(\theta + 1)^{7/2}}$$

για $x > 0$ και

$$f_S(0) = \frac{\theta^2 (1 + a + \theta)}{(\theta + a)(1 + \theta)^2}$$

Απόδειξη

Έστω $N \sim NGPL(\theta, a)$ και $X_i \sim Gamma(2, \lambda)$, $i = 1, 2, \dots$. Η n -οστή συνέλιξη $f^{n*}(x)$ της κατανομής $Gamma(2, \lambda)$ δίνεται από τον τύπο

$$f^{n*}(x) = \frac{\lambda^{2n}}{(2n-1)!} x^{2n-1} e^{-\lambda x}, \quad n = 1, 2, \dots$$

που είναι η συνάρτηση πυκνότητας της κατανομής $Gamma(2n, \lambda)$.

Έτσι για τη συνάρτηση πυκνότητας της τ.μ. S για $x > 0$ έχουμε

$$\begin{aligned} f_S(x) &= \sum_{n=1}^{\infty} \frac{\theta^2}{(\theta+a)(1+\theta)^{n+1}} \left(1 + \frac{a(n+1)}{1+\theta}\right) \frac{\lambda^{2n}}{(2n-1)!} x^{2n-1} e^{-\lambda x} \\ &= \left(\frac{\theta^2 e^{-\lambda x}}{\theta+a}\right) \sum_{n=1}^{\infty} \frac{\lambda^{2n}}{(1+\theta)^{n+1}(2n-1)!} x^{2n-1} \left(1 + \frac{a(n+1)}{1+\theta}\right) \\ &= \left(\frac{\theta^2 e^{-\lambda x}}{\theta+a}\right) \frac{\lambda \left[(\theta+1)(3a+2\theta+2) \sinh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right) + a\sqrt{\theta+1} \lambda x \cosh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right) \right]}{2(\theta+1)^{7/2}} \\ &= \frac{\theta^2 e^{-\lambda x} \left[(\theta+1)(3a+2\theta+2) \sinh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right) + a\sqrt{\theta+1} \lambda x \cosh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right) \right]}{2(\theta+a)(\theta+1)^{7/2}}. \end{aligned}$$

Η πιθανότητα για μηδενικό ύψος αποζημιώσεων είναι

$$f_S(0) = \frac{\theta^2(1+a+\theta)}{(\theta+a)(1+\theta)^2}.$$

Η απόδειξη ολοκληρώθηκε. ■

4.7 Εφαρμογή σε πραγματικά δεδομένα

Στην ενότητα αυτή, θα προσαρμόσουμε την προτεινόμενη κατανομή σε 2 σύνολα δεδομένων ζημιών και θα συγκρίνουμε την προσαρμογή τους με άλλες κατανομές. Το πρώτο σύνολο δεδομένων έχει «βαριά» ουρά και πλησιάζει αργά το 0 ενώ το άλλο σύνολο δεδομένων πλησιάζει το 0 με γρηγορότερο ρυθμό.

Σύνολο δεδομένων 1

Τα συγκεκριμένα δεδομένα αφορούν στις μετρήσεις επιληπτικών κρίσεων ενός πληθυσμού (δείτε Chakraborty (2010)). Η σύγκριση με άλλες κατανομές όπως η Poisson (P), η Αρνητική Διωνυμική (NB), η Generalized Poisson-Lindley (GPL) των Mahmoudi & Zakerzadeh (2010)

μας δίνει τα αποτελέσματα του Πίνακα 4.3 . Σε κάθε από αυτές τις κατανομές η εκτίμηση των παραμέτρων γίνεται με τη μέθοδο των Ε.Μ.Π. Επιπλέον δίνεται η τιμή της στατιστικής χ^2 και της Log-Likelihood. Διαπιστώνουμε πως η κατανομή $NGPL(\theta, a)$ παρουσιάζει καλύτερη προσαρμογή σε σχέση με τις υπόλοιπες κατανομές. Συγκεκριμένα, η τιμή της στατιστικής χ^2 και της Log-Likelihood είναι η χαμηλότερη κάτι που σημαίνει την καλύτερη προσαρμογή στα δεδομένα.

Πίνακας 4-3
Προσαρμογή σε δεδομένα επιληπτικών κρίσεων

Τιμή	Συχνότητα	<i>P</i>	<i>NB</i>	<i>GPL</i>	<i>NGPL</i>
0	126	74.94	91.00	121.51	122.00
1	80	115.71	86.60	92.00	91.00
2	59	89.34	63.37	59.00	58.74
3	42	46.00	42.57	35.10	35.22
4	24	17.75	27.60	20.10	20.52
5	8	5.48	17.60	11.18	11.22
6	5	1.41	10.50	6.10	6.39
7	4	0.31	6.52	3.30	3.25
8	3	0.06	5.00	2.71	2.50
ΣΥΝΟΛΟ	351	351	351	351	351
	Παράμετροι	$\hat{\lambda} = 1.544$	$\hat{r} = 1.757$ $\hat{p} = .463$	$\hat{\theta} = 1.139$ $\hat{a} = 1.292$	$\hat{\theta} = 1.116$ $\hat{a} = 2.9061$
	Log Likelihood	-636.05	-595.22	-594.61	-594.48
	Chi square statistic	256.54	22.53	5.94	5.75

Σύνολο δεδομένων 2

Το σύνολο αυτό προέρχεται από δεδομένα της ασφάλισης του κλάδου αυτοκινήτου (δείτε Klugman (2012) σελ. 664)). Το συγκεκριμένο σύνολο σετ δεδομένων εμφανίζει «βαριά ουρά» και μεγάλη διασπορά (η διασπορά είναι μεγαλύτερη της μέσης τιμής). Διαπιστώνουμε πως το μοντέλο μας μπορεί να εφαρμοστεί σε δεδομένα με μεγάλη διασπορά. Τα αποτελέσματα παρουσιάζονται στον επόμενο πίνακα.

Πίνακας 4-4
 Προσαρμογή σε δεδομένα ασφάλισης αυτοκινήτων

Τιμή	Συχνότητα	<i>P</i>	<i>NB</i>	<i>GPL</i>	<i>NGPL</i>
0	1563	1544.153	1566.40	1566.407	1564.504
1	271	299.772	261.500	261.384	264.251
2	32	29.1	40.146	40.202	39.686
3	7	1.883	5.990	5.985	5.589
4	2	0.0914	0.880	0.874	0.870
ΣΥΝΟΛΟ	1875	1875	1875	1875	1875
	Παράμετροι	$\hat{\lambda} = 0.19413$	$\hat{\tau} = 6,135$ $\hat{\alpha} = 1.191$	$\hat{\theta} = 6.6715$ $\hat{\alpha} = 1.1648$	$\hat{\theta} = 7.8747$ $\hat{\alpha} = 8.835$
	Chi square statistic	57.04	3.610	3.655	3.489

Από τα παραπάνω είναι εμφανής η καλύτερη προσαρμογή του μοντέλου της κατανομής $NGPL(\theta, \alpha)$ σε σχέση με τις υπόλοιπες κατανομές.

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα

Η Lindley κατανομή είναι μία συνεχής κατανομή, η οποία προέρχεται από την εκθετική οικογένεια κατανομών. Γενικότερα, οι κατανομές από αυτή την οικογένεια όπως η εκθετική, Weibull κ.α., χρησιμοποιούνται ευρέως στην Αναλογιστική επιστήμη με σκοπό την προσαρμογή σε εμπειρικά δεδομένα και την δυνατότητα για καλύτερη πρόβλεψη αποτελεσμάτων.

Όπως είδαμε στο Κεφάλαιο 1, το μοντέλο προσαρμογής σε εμπειρικά δεδομένα με βάση την Lindley κατανομή ($Lindley(\theta)$) παρουσιάζει καλύτερα αποτελέσματα σε σχέση με αυτά για την εκθετική κατανομή, μία ιδιαίτερα δημοφιλή κατανομή. Συγκεκριμένα, η υπόθεση πως τα εμπειρικά δεδομένα προέρχονται από τη ($Lindley(\theta)$), για μεγάλα δείγματα, έδειξε μικρή θετική μεροληψία ενώ στο δεύτερο παράδειγμα τα διαγράμματα Q-Q και η μέγιστη τιμή της συνάρτησης πιθανοφάνειας, εμφάνισαν την καλύτερη προσαρμογή της ($Lindley(\theta)$) σε σχέση με την εκθετική

Ένα σημαντικό στοιχείο που παρατηρήθηκε είναι η ευελιξία της Lindley κατανομής. Στο Κεφάλαιο 2, είδαμε την περίπτωση που το αίτιο εξόδου από έναν πληθυσμό είναι μεγαλύτερο του ενός, όπως για παράδειγμα σε ένα χαρτοφυλάκιο ασφαλίσεων ζωής, την μορφή που μπορεί να λάβει η κατανομή θεωρώντας την παράμετρο θ ως διάνυσμα το οποίο παίρνει συγκεκριμένες τιμές ανάλογα με το αίτιο εξόδου. Στην περίπτωση αυτή, είδαμε μικρές τιμές της θετικής μεροληψίας για την εκτίμηση των παραμέτρων και για τα αντίστοιχα τετραγωνικά σφάλματα καθώς και τα πολύ κοντινά, στην αρχική υπόθεση, διαστήματα εμπιστοσύνης.

Στο Κεφάλαιο 3, είδαμε μια άλλη εκδοχή της κατανομής η οποία προέρχεται από τη μίξη δύο κατανομών γάμμα ($Gamma(a, \theta)$ και $Gamma(a + 1, \theta)$) και μας δίνει ως αποτέλεσμα τη Γενικευμένη Lindley κατανομή ($GL(a, \theta, \gamma)$). Συγκρίναμε την κατανομή με άλλες δημοφιλείς κατανομές όπως η εκθετική, η Weibull και τη Lognormal κατανομή ως προς την προσαρμογή σε εμπειρικά δεδομένα. Ειδικότερα, διαπιστώσαμε την καλύτερη προσαρμογή της Γενικευμένης Lindley κατανομής σε όρους Log-Likelihood (L-L) και της στατιστικής συνάρτησης ελέγχου Kolmogorov-Smirnov (K-S) σε σχέση με τις παραπάνω δημοφιλείς κατανομές.

Τέλος, στο Κεφάλαιο 4 είδαμε πως η Lindley κατανομή μπορεί να χρησιμοποιηθεί για διακριτές περιπτώσεις ενός συλλογικού μοντέλου όπως η συχνότητα ζημιών. Πιο συγκεκριμένα, είδαμε την κατανομή $NPGL(\theta, a)$ η οποία προκύπτει όταν ο αριθμός ζημιών ακολουθεί την Poisson κατανομή με παράμετρο λ που ακολουθεί με τη σειρά της την Lindley κατανομή με 2 παραμέτρους. Η εκτίμηση - προσαρμογή της συχνότητας ζημιών ενός χαρτοφυλακίου ασφαλιστηρίων για τον κλάδο αυτοκινήτου ήταν καλύτερη και στα 2 σετ δεδομένων σε σχέση με άλλες πολύ γνωστές διακριτές κατανομές όπως η Poisson και η Αρνητική Διωνυμική.

Με την κατανομή Lindley και τις γενικεύσεις της έχουν ασχοληθεί πολλοί ερευνητές. Ενδεικτικά αναφέρουμε τις παρακάτω εργασίες με γενικεύσεις της κατανομής Lindley, οι οποίες μπορούν να αποτελέσουν σημαντικό εργαλείο στην μελέτη και πρόβλεψη διάφορων παραμέτρων και μεγεθών στην Αναλογιστική επιστήμη.

Μία από αυτές είναι η Pareto - Poisson Lindley κατανομή (συμβ. $PPL(\beta, \theta, \lambda)$), η οποία βρίσκει εφαρμογή σε δεδομένα χρόνων ζωής και τις αντίστοιχες ασφαλίσεις (δείτε Asgharzadeh, Bakouch & Esmaeili (2013)). Η κατανομή αυτή είναι η μίξη της $Pareto(\lambda, \beta)$ και της αποκομμένης στο μηδέν Poisson-Lindley με παραμέτρους $\eta\beta$ και λ .

Επίσης, η αντίστροφη Power Lindley ($IPL(\alpha, \beta)$) κατανομή η οποία είναι το αποτέλεσμα της μίξης κατανομής $Weibull(\alpha, \beta)$ και $Generalized\ Gamma(2, \alpha, \beta)$ (δείτε Barco, Mazucheli & Janeiro (2016)).

Μία γενικότερη παραλλαγή της Lindley κατανομής η οποία συμπεριλαμβάνει ως ειδικές περιπτώσεις τη $Lindley(\theta)$ και τη $Weibull(\alpha, \beta)$, είναι η Weibull-Lindley κατανομή (συμβ. $WL(\alpha, \beta, \theta)$) και είναι η συνέλιξη των δύο παραπάνω κατανομών (δείτε A.Asgharzadeh et. Al (2016)). Συγκεκριμένα, εάν για 2 ανεξάρτητες μεταβλητές σε περιπτώσεις χαρτοφυλακίων με χρόνους ζωής ισχύει $Y \sim Lindley(\theta)$ και $Z \sim Weibull(\alpha, \beta)$, για την τ.μ. X με $X = \min(Y, Z)$, λέμε πως $X \sim WL(\alpha, \beta, \theta)$ με συνάρτηση επιβίωσης $\bar{F}(x) = \bar{G}(x)\bar{Q}(x)$, όπου $\bar{G}(x)$ και $\bar{Q}(x)$, οι συναρτήσεις επιβίωσης της $Lindley(\theta)$ και $Weibull(\alpha, \beta)$ αντίστοιχα. Εύκολα μπορεί να βρεθεί η σ.κ. από τη σχέση $F(x) = 1 - \bar{F}(x)$ και με παραγωγή ως προς x , η σ.π. $f(x)$.

Δύο άλλες μορφές της Γενικευμένης Lindley κατανομής που είδαμε στο Κεφάλαιο 3, είναι η $GLD4(\alpha, \theta)$ και $GLD5(\alpha, \theta)$ (δείτε Abouammoh et al (2015)). Η κατανομή $GLD4(\alpha, \theta)$ η οποία προκύπτει από τη μίξη 2 γάμμα κατανομών, $Gamma(\eta, \theta)$ και $Gamma(\tau, \theta)$ με βάρη μίξης $\theta^\varphi / (\gamma + \theta^\varphi)$ και $\gamma / (\gamma + \theta^\varphi)$ αντίστοιχα. Η κατανομή $GLD5(\alpha, \theta)$ το αποτέλεσμα της μίξης $Gamma(\alpha, \theta)$ και $Gamma(\alpha-1, \theta)$ και αντίστοιχα βάρη $\rho_1 = 1 - \rho_2 = 1 / (\theta + 1)$. Εύκολα αντιλαμβάνεται κανείς πως για $\alpha=2$ στην $GLD5(\alpha, \theta)$, ως ειδική περίπτωση, λαμβάνουμε την $Lindley(\theta)$ που είδαμε στο Κεφάλαιο 1.

Μία διακριτή μορφή μίξης της Lindley κατανομής είναι η Αρνητική Διωνυμική-Lindley κατανομή, η οποία συμβ. $NB\sim L(r, \theta)$ και προκύπτει όταν ο αριθμός ζημιών ενός χαρτοφυλακίου ακολουθεί τη $NB(r, p) = e^{-\lambda}$ και $\lambda \sim Lindley(\theta)$ (δείτε H. Zamani and N. Ismail (2010)). Τέλος, μια επίσης πολύ χρήσιμη διακριτή μορφή της Lindley κατανομής είναι η τροποποιημένη διακριτή Lindley κατανομή (συμβ. $MDL(\theta, m)$). Η κατανομή αυτή αποτελεί την μίξη της γεωμετρικής κατανομής και αρνητικής διωνυμικής κατανομής με βάρη $\theta/(1+\theta)$ και $1/(1+\theta)$ και παρουσιάζει πολύ καλύτερη προσαρμογή σε δεδομένα υπολειπόμενων χρόνων ζωής σε σχέση με τη γεωμετρική κατανομή η οποία έχει σταθερό ρυθμό θνησιμότητας (δείτε E. Gomez Déniz και E. Calderin Ojeda (2011) & Bakouch et al. (2014)).

Καταλήγοντας, μπορεί κανείς να συμπεράνει πως η Lindley κατανομή εμφανίζει πολλές εναλλακτικές μορφές, κάτι το οποίο μπορεί να προσφέρει μεγάλη ευελιξία στην Αναλογιστική Επιστήμη. Δίνεται η δυνατότητα κανείς να επιλέξει, ανάλογα με το αντικείμενο μελέτης, την κατάλληλη μορφή της κατανομής και να μπορέσει να προσαρμόσει και να εξάγει αποτελέσματα και συμπεράσματα όσο πιο κοντινά γίνεται σε εμπειρικά δεδομένα.

ΠΑΡΑΡΤΗΜΑ

ΚΕΦΑΛΑΙΟ 1

Κώδικας για την εξαγωγή των σχημάτων 1.1 έως 1.5

```
> x<-seq(0,10,0.5)
> u<-c(0.5,1,1.5)
> y1<-((0.5^2)/(0.5+1))*(1+x)*exp(-0.5*x)
> plot(x,y1,type="c",col="green",ylim=c(0,0.8),lwd=2,xlab="x",ylab="f(x)")
> y2<-((1^2)/(1+1))*(1+x)*exp(-1*x)
> lines(x,y2,lty=3,col="red",lwd=2)
> y3<-((1.5^2)/(1.5+1))*(1.5+x)*exp(-1.5*x)
> lines(x,y3,lty=1,col="blue",lwd=2)
>
legend(6,1,c("θ=0.5","θ=1","θ=1.5"),lty=c(1,1,1),col=c("green","red","blue"),
), lwd=c(2,2,2))
>
> x<-seq(0,10,0.5)
> Y1<-1-exp(-u[1]*x)*(u[1]+1+u[1]*x)/(u[1]+1)
> plot(x,Y1,type="c",col="green",ylim=c(0,0.8),lwd=2,xlab="x",ylab="F(x)")
> Y2<-1-exp(-u[2]*x)*(u[2]+1+u[2]*x)/(u[2]+1)
> lines(x,Y2,lty=3,col="red",lwd=2)
> Y3<-1-exp(-u[3]*x)*(u[3]+1+u[3]*x)/(u[3]+1)
> lines(x,Y3,lty=3,col="blue",lwd=2)
>
legend(6,1,c("θ=0.5","θ=1","θ=1.5"),lty=c(1,1,1),col=c("green","red","blue"),
), lwd=c(2,2,2))
>
> x<-seq(0,20,1)
> y1<-sqrt(x^2+4*x+2)/(x+2)
> y2<-2*(x^3+6*x^2+6*x+2)/(x^2+4*x+2)^(3/2)
> y3<-3*(3*x^4+24*x^3+44*x^2+32*x+8)/(x^2+4*x+2)^2
> plot(x,y1,type="l",col="blue",ylim=c(0,10),lwd=2,xlab="θ",ylab="")
> lines(x,y2,lty=3,col="red",lwd=2)
> lines(x,y3,lty=3,col="green",lwd=2)
> legend(10,6,c("γ","Πίζα του
β1","β2"),lty=c(1,3,3),col=c("blue","red","green"),lwd=c(2,2,2))
>
> x<-seq(0,10,0.5)
> u<-c(0.5,1,2)
> y1<-u[1]^2*(1+x)/(u[1]+u[1]*x+1)
> y2<-u[2]^2*(1+x)/(u[2]+u[2]*x+1)
> y3<-u[3]^2*(1+x)/(u[3]+u[3]*x+1)
> plot(x,y1,type="c",ylim=c(0,2.5),lwd=2,xlab="x",ylab="h(x)",col="green")
> lines(x,y2,lty=3,col="red",lwd=2)
> lines(x,y3,lty=1,col="blue",lwd=2)
>
legend(6,2.4,c("θ=1.5","θ=1","θ=0.5"),lty=c(1,3,3),col=c("blue","red","green"),
), lwd=c(2,2,2))
>
> x<-seq(0,10,0.5)
> u<-c(0.5,1,1.5)
> y1<-(u[1]+2+u[1]*x)/(u[1]*(u[1]+1+u[1]*x))
> y2<-(u[2]+2+u[2]*x)/(u[2]*(u[2]+1+u[2]*x))
> y3<-(u[3]+2+u[3]*x)/(u[3]*(u[3]+1+u[3]*x))
```

```

> plot(x,y1,type="c",col="green",ylim=c(0,4),lwd=2,xlab="x",ylab="m(x)")
> lines(x,y2,lty=3,col="red",lwd=2)
> lines(x,y3,lty=1,col="blue",lwd=2)
>
legend(7,3.5,c("θ=1.5","θ=1","θ=0.5"),lty=c(1,3,3),col=c("blue","red","green"),lwd=c(2,2,2))

```

Κώδικας για την εξαγωγή του Πίνακα 1.3

```

> set.seed(10000)
> theta <- 9
> p <- theta/(theta+1)
> N <- 10000
> n <- c(10,20,30,40,50,60,70,80,90,100,200,500,1000)
> uni <- c() #uniform
> v <- c() #exponential
> w <- c() #gamma
> lin <- c() #lindley
> emp <- c() #MLE
> m <- c() #mean of Lindley samples
> bias <- c()
> MSE <- c()
> cover95 <- c()
> cover90 <- c()
> y95 <- c() #
> y90 <- c()
> #Confidence limit
> cl95=0.95
> cl90=0.9
> # z percentile
> z95 <- qnorm((1+cl95)/2)
> z90 <- qnorm((1+cl90)/2)
>
> length <- length(n)
>
> for (r in 1:length ) {
+   for (i in 1:N) {
+     uni <- runif(n[r],0,1)
+     v <- rexp(n[r],rate=theta)
+     w <- rgamma(n[r], shape=2, rate=theta)
+     for (j in 1:n[r]) {
+       ifelse(uni[j]>p,lin[j]<-w[j],lin[j]<-v[j])
+     }
+     m[i] <- mean(lin)
+     emp[i] <- -(m[i]-1)+sqrt((m[i]-1)^2+8*m[i]))/(2*m[i])
+     lb95 <- emp[i]-z95/(sd(lin)*sqrt(n[r]))
+     ub95 <- emp[i]+z95/(sd(lin)*sqrt(n[r]))
+     y95[i] <- as.numeric((lb95<theta) & (theta<ub95))
+     lb90 <- emp[i]-z90/(sd(lin)*sqrt(n[r]))
+     ub90 <- emp[i]+z90/(sd(lin)*sqrt(n[r]))
+     y90[i] <- as.numeric((lb90<theta) & (theta<ub90))
+   }
+   bias[r] <- (1/N)*sum(emp-theta)
+   MSE[r] <- (1/N)*sum((emp-theta)^2)
+   cover95[r] <- mean(y95)
+   cover90[r] <- mean(y90)
+ }
>
> mat <- cbind(n,bias,MSE,cover90,cover95)

```

```

> rownames(mat) <- rep("", nrow(mat))
> colnames(mat) <- c("n", "Average bias", "Average MSE", "Coverage 90%",
"Coverage 95%")
> round(mat,digits=5)
  n Average bias Average MSE Coverage 90% Coverage 95%
10  0.89862    11.07727    0.9211    0.9576
20  0.46029     4.41642    0.9126    0.9538
30  0.27466     2.67985    0.9087    0.9510
40  0.19864     1.90620    0.9058    0.9493
50  0.15386     1.48009    0.9062    0.9550
60  0.14394     1.20734    0.9074    0.9567
70  0.10971     1.04137    0.9040    0.9493
80  0.10520     0.90611    0.9044    0.9526
90  0.08923     0.81229    0.9006    0.9498
100 0.08580     0.72226    0.9030    0.9476
200 0.03529     0.35303    0.9013    0.9486
500 0.01376     0.13632    0.9034    0.9554
1000 0.00688     0.06836    0.8995    0.9505

```

Κώδικας για την εξαγωγή των Σχημάτων 1.6 και 1.7

```

> a <-read.table("Tab4.txt")
> attach(a)
> names(a)
[1] "V1"
> n <- length(V1)
> m <- mean(V1)
> (thetaexp <- 1/m)
[1] 0.1012453
> (thetaLin <- (- (m-1)+sqrt((m-1)^2+8*m))/(2*m))
[1] 0.1865713
> x1 <-qexp(ppoints(n),rate=thetaexp)
> qqplot(x1, V1, main="Exponential distribution / Q-Q plot", xlab="x",
ylab="y")
> library(LindleyR)
> x2 <-qlindley(ppoints(n),theta=thetaLin)
> qqplot(x2, V1, main="Lindley distribution / Q-Q plot", xlab="x",
ylab="y")

```

ΚΕΦΑΛΑΙΟ 3

Κώδικας για την εξαγωγή των σχημάτων 3.1 έως 3.3

```
> a<-c(0.5,1,1.5)
> u<-c(0.5,1,1.5)
> g<-c(0.5,1,1.5)
>
plot(x,gl(x,a[3],u[1],g[3]),type="c",col="green",ylim=c(0,0.5),lwd=2,xlab="
x",ylab="f(x)")
> lines(x,gl(x,a[3],u[2],g[3]),lty=3,col="red",lwd=2)
> lines(x,gl(x,a[3],u[3],g[3]),lty=1,col="blue",lwd=2)
>
legend(7,3.5,c("GL( $\alpha=1.5,\theta,\gamma=1.5$ )"," $\theta=1.5$ "," $\theta=1$ "," $\theta=0.5$ "),lty=c(1,3,3),col=
c("blue","red","green"),lwd=c(2,2,2))

>
plot(x,gl(x,a[1],u[3],g[3]),type="c",col="green",ylim=c(0,0.5),lwd=2,xlab="
x",ylab="f(x)")
> lines(x,gl(x,a[2],u[3],g[3]),lty=3,col="red",lwd=2)
> lines(x,gl(x,a[3],u[3],g[3]),lty=1,col="blue",lwd=2)
>
legend(7,3.5,c("GL( $\alpha,\theta=1.5,\gamma=1.5$ )"," $\alpha=1.5$ "," $\alpha=1$ "," $\alpha=0.5$ "),lty=c(1,3,3),col=
c("blue","red","green"),lwd=c(2,2,2))

>
plot(x,gl(x,a[3],u[3],g[1]),type="c",col="green",ylim=c(0,0.5),lwd=2,xlab="
x",ylab="f(x)")
> lines(x,gl(x,a[3],u[3],g[2]),lty=3,col="red",lwd=2)
> lines(x,gl(x,a[3],u[3],g[3]),lty=1,col="blue",lwd=2)
>
legend(7,3.5,c("GL( $\alpha=1.5,\theta=1.5,\gamma$ )"," $\gamma=1.5$ "," $\gamma=1$ "," $\gamma=0.5$ "),lty=c(1,3,3),col=
c("blue","red","green"),lwd=c(2,2,2))
```

Κώδικας για την εξαγωγή του Πίνακα 3.3

ΣΕΤ ΔΕΔΟΜΕΝΩΝ 1

```
> y <- read.table("D1.txt", header=TRUE)
> attach(y)
> names(y)
[1] "x"
> library(fitdistrplus)

ΚΑΤΑΝΟΜΗ GAMMA

> gamMLE <- fitdist(x, "gamma")
> gamMLE$estimate
      shape      rate
1.44193975 0.05235443
> gamMLE$loglik
[1] -64.18593
> ks.test(x,"pgamma", shape=gamMLE$estimate[1], rate=gamMLE$estimate[2])

      One-sample Kolmogorov-Smirnov test

data:  x
```

```
D = 0.10313, p-value = 0.9921
alternative hypothesis: two-sided
```

KATANOMH WEIBULL

```
> weiMLE <- fitdist(x, "weibull")
> weiMLE$estimate
  shape    scale
1.305605 29.764543
> weiMLE$loglik
[1] -64.0202
> ks.test(x,"pweibull", shape=weiMLE$estimate[1], scale=weiMLE$estimate[2])
```

One-sample Kolmogorov-Smirnov test

```
data: x
D = 0.098099, p-value = 0.9957
alternative hypothesis: two-sided
```

KATANOMH LOGNORMAL

```
> lnMLE <- fitdist(x, "lnorm")
> lnMLE$estimate
 meanlog  sdlog
2.930596 1.025274
> lnMLE$loglik
[1] -65.61742
> ks.test(x,"plnorm", meanlog=lnMLE$estimate[1], sdlog=lnMLE$estimate[2])
```

One-sample Kolmogorov-Smirnov test

```
data: x
D = 0.16168, p-value = 0.7713
alternative hypothesis: two-sided
```

KATANOMH GL

```
> library(LindleyR)
> linMLE <- fitdist(x, "genlindley", fix.arg=list(beta=0.083, theta=0.064),
start = list( alpha =1.2))
> linMLE$estimate
  alpha
1.200131
> linMLE$loglik
[1] -64.08046
> ks.test(x,"pgenlindley", beta=0.083, theta=0.064,
alpha=linMLE$estimate[1])
```

One-sample Kolmogorov-Smirnov test

```
data: x
D = 0.095393, p-value = 0.997
alternative hypothesis: two-sided
```

ΣΕΤ ΔΕΔΟΜΕΝΩΝ 2

```
> y <- read.table("D2.txt", header=TRUE)
> names(y)
[1] "x"
> attach(y)
> library(fitdistrplus)
```

KATANOMH GAMMA

```
> gamMLE <- fitdist(x, "gamma")
```

```

> gamMLE$estimate
  shape      rate
1.79394334 0.01006049
> gamMLE$loglik
[1] -152.438
> ks.test(x,"pgamma", shape=gamMLE$estimate[1], rate=gamMLE$estimate[2])

      One-sample Kolmogorov-Smirnov test

data:  x
D = 0.13832, p-value = 0.7252
alternative hypothesis: two-sided

KATANOMH WEIBULL
> weiMLE <- fitdist(x, "weibull")
> weiMLE$estimate
  shape      scale
1.414046 196.154872
> weiMLE$loglik
[1] -152.4432
> ks.test(x,"pweibull", shape=weiMLE$estimate[1], scale=weiMLE$estimate[2])

      One-sample Kolmogorov-Smirnov test

data:  x
D = 0.12245, p-value = 0.8477
alternative hypothesis: two-sided

KATANOMH LOGNORMAL
> lnMLE <- fitdist(x, "lnorm")
> lnMLE$estimate
  meanlog      sdlog
4.8797771 0.8734907
> lnMLE$loglik
[1] -154.0864
> ks.test(x,"plnorm", meanlog=lnMLE$estimate[1], sdlog=lnMLE$estimate[2])

      One-sample Kolmogorov-Smirnov test

data:  x
D = 0.18731, p-value = 0.3443
alternative hypothesis: two-sided

KATANOMH GL
> library(LindleyR)
> linMLE <- fitdist(x, "genlindley", fix.arg=list(beta=0.018, theta=0.012),
start = list( alpha =1))
> linMLE$estimate
  alpha
1.530829
> linMLE$loglik
[1] -152.3707
> ks.test(x,"pgenlindley", beta=0.018, theta=0.012,
alpha=linMLE$estimate[1])

      One-sample Kolmogorov-Smirnov test

data:  x
D = 0.12943, p-value = 0.7964
alternative hypothesis: two-sided

```

ΚΕΦΑΛΑΙΟ 4

Κώδικας για την εξαγωγή του Σχήματος 4.1

```
> par(mfcol=c(3,3))
>
> u<-0.05
> #####
> a<-0.1
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
>
plot(x,fpl,type="h",ylim=c(0,0.030),col="blue",ylab="p(x)",xlab="θ=0.05,α=0.1")
> lines(x,fpl,type="l",col="red")
>
> a<-0.5
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
>
plot(x,fpl,type="h",ylim=c(0,0.020),col="blue",ylab="p(x)",xlab="θ=0.05,α=0.5")
> lines(x,fpl,type="l",col="red")
>
> a<-1
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
> plot(x,fpl,type="h",ylim=c(0,0.020),col="blue",ylab="p(x)",xlab="θ=0.05,α=1")
> lines(x,fpl,type="l",col="red")
>
> u<-0.1
> #####
> a<-0.1
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
>
plot(x,fpl,type="h",ylim=c(0,0.050),col="blue",ylab="p(x)",xlab="θ=0.1,α=0.1")
> lines(x,fpl,type="l",col="red")
>
> a<-0.5
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
> plot(x,fpl,type="h",ylim=c(0,0.05),col="blue",ylab="p(x)",xlab="θ=0.1,α=0.5")
> lines(x,fpl,type="l",col="red")
>
> a<-1
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
>
plot(x,fpl,type="h",ylim=c(0,0.05),col="blue",ylab="p(x)",xlab="θ=0.1,α=1")
> lines(x,fpl,type="l",col="red")
>
> u<-0.5
> #####
> a<-0.1
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
> plot(x,fpl,type="h",ylim=c(0,0.30),col="blue",ylab="p(x)",xlab="θ=0.5,α=0.1")
> lines(x,fpl,type="l",col="red")
>
> a<-0.5
> x<-0:50
```

```

> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
> plot(x,fpl,type="h",ylim=c(0,0.30),col="blue",ylab="p(x)",xlab="θ=0.5,α=0.5")
> lines(x,fpl,type="l",col="red")
>
> a<-1
> x<-0:50
> fpl<-((u^2)*(1+a*(x+1)/(1+u)))/((u+a)*(1+u)^(x+1))
> plot(x,fpl,type="h",ylim=c(0,0.30),col="blue",ylab="p(x)",xlab="θ=0.5,α=1")
> lines(x,fpl,type="l",col="red")

```

Κώδικας για την εξαγωγή των Σχημάτων 4.2-4.3

```

> sk<-function(u,a) (
+
+
+ ((2*(a^3)*((u^2)+3*u+2)+(a^2)*u*(5*(u^2)+18*u+12)+a*(u^2)*(4*(u^2)+15*u+12)
+ (u^3)*((u^2)+3*u+2))/((u^3)*((a+u)^3)))/((2*(a^2)*(1+u)+(u^2)*(1+u)+a*u*(4
+ 3*u))/((u^2)*((a+u)^2)))^(3/2)
+ )
>
> u<-seq(0,10,0.5)
> a<-seq(0,10,0.5)
>
persp(u,a,outer(u,a,sk),ticktype="detailed",main="ΛΟΞΟΤΗΤΑ",xlab="Alpha",yl
ab="Theta",xlim=range(u),ylim=range(a),theta=-
50,phi=30,expand=0.5,col="lightblue")

> dm4<-function(a,u) (
+ (2*a)*((u^3)+21*(u^2)+72*u+60)+u*((u^3)+14*(u^2)+36*u+24))/((u^4)*(a+u))
> dm3<-function(a,u) (
+ (u*((u^2)+6*u+6)+2*a*((u^2)+9*u+12))/((u^3)*(a+u)))
> dm2<-function(a,u) (
+ (2*a*(u+3)+u*(u+2))/((u^2)*(a+u)))
> m<-function(a,u) (
+ (2*a+u)/(u*(a+u)))
>
> m4<-function(a,u) (
+ dm4(a,u)-3*(m(a,u)^4)+6*dm2(a,u)*(m(a,u)^2)-4*dm3(a,u)*m(a,u))
> m2<-function(a,u) (
+ (2*(a^2)*(1+u)+(u^2)*(1+u)+a*u*(4+3*u))/((u^2)*(a+u)^2))
>
> kyrptosis<-function(a,u) (
+ m4(a,u)/(m2(a,u)^2))
>
>
persp(u,a,outer(u,a,kyrptosis),ticktype="detailed",main="ΚΥΡΠΩΣΗ",xlab="Alph
a",ylab="Theta",xlim=range(u),ylim=range(a),theta=-
50,phi=30,expand=0.5,col="lightblue")

```

Κώδικας για την εξαγωγή των Πινάκων 4.1-4.2

```

> matrix41<-matrix(0,nrow=7,ncol=6)
> theta<-c(0.25,0.5,1,1.5,2,2.5,3)
> alpha<-c(0.5,1,1.5,2,2.5,3)
> for(i in 1:7){
+ for(j in 1:6)(matrix41[i,j]<-sk(theta[i],alpha[j]))
+

```



```

+ }
> rownames(matrix41)<-
c("θ=0.25","θ=0.50","θ=1.00","θ=1.50","θ=2.00","θ=2.50","θ=3.00")
> colnames(matrix41)<-
c("α=0.50","α=1.00","α=1.50","α=2.00","α=2.50","α=3.00")
>
> matrix41
      α=0.50  α=1.00  α=1.50  α=2.00  α=2.50  α=3.00
θ=0.25 1.555875 1.487841 1.464454 1.453072 1.446453 1.442161
θ=0.50 1.707630 1.597690 1.550340 1.524594 1.508586 1.497733
θ=1.00 1.915904 1.792108 1.722817 1.679379 1.649812 1.628465
θ=1.50 2.063000 1.949172 1.875000 1.824036 1.787140 1.759289
θ=2.00 2.185735 2.083265 2.009840 1.956071 1.915347 1.883555
θ=2.50 2.296714 2.203384 2.132292 2.077800 2.035098 2.000875
θ=3.00 2.400718 2.314307 2.245771 2.191455 2.147759 2.112000

> matrix42<-matrix(0,nrow=7,ncol=6)
> theta<-c(0.25,0.5,1,1.5,2,2.5,3)
> alpha<-c(0.5,1,1.5,2,2.5,3)
> for(i in 1:7){
+ for(j in 1:6)(matrix42[i,j]<-kyrtosis(alpha[j],theta[i]))
+
+ }
> rownames(matrix42)<-
c("θ=0.25","θ=0.50","θ=1.00","θ=1.50","θ=2.00","θ=2.50","θ=3.00")
> colnames(matrix42)<-
c("α=0.50","α=1.00","α=1.50","α=2.00","α=2.50","α=3.00")
>
> matrix42
      α=0.50  α=1.00  α=1.50  α=2.00  α=2.50  α=3.00
θ=0.25 6.511794 6.245816 6.161779 6.122477 6.100155 6.085913
θ=0.50 7.180000 6.672845 6.475556 6.373930 6.312870 6.272447
θ=1.00 8.210059 7.532544 7.187939 6.984375 6.851345 6.758115
θ=1.50 8.957280 8.282493 7.875000 7.609720 7.425124 7.289938
θ=2.00 9.573563 8.941828 8.513270 8.213296 7.994041 7.827664
θ=2.50 10.129155 9.542415 9.111072 8.792066 8.549587 8.360143
θ=3.00 10.655161 10.106111 9.679200 9.349878 9.091555 8.884800

```

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΕΛΛΗΝΙΚΗ

1. Αντζουλάκος Δ., Εισαγωγή στην Γλώσσα Προγραμματισμού R, ΜΑΕ, ΠΑ.ΠΕΙ (2015)
2. Χατζηκωνσταντινίδης Ευ., Θεωρία Κινδύνου I, ΜΑΕ, ΠΑ.ΠΕΙ (2015)

ΞΕΝΗ

1. Abadir K.M., The mean–median–mode inequality: counterexamples, *Econometric Theory* 21, 477–482 (2005)
2. Abouammoh A.M. , Alshangiti A. M. & I.E. Ragab I.E., Journal of Statistical Computation and Simulation, Reliability estimation of the generalized Lindley distributions. *Journal of Statistical Computation and Simulation*, 85 (17), 3662–3678 (2015)
3. Asgharzadeh A. , Bakouch H. S. & Esmaeili L. , Pareto Poisson–Lindley distribution with application (2013)
4. Asgharzadeh A., Nadarajah S. and Sharafi F., Weibull Lindley Distribution, *Primary* 62E99 (2016)
5. Barlow R.E., Proschan F., *Statistical Analysis of Reliability and Life Testing*, Silver Spring (1981).
6. Bhatti D., Sastry D.V.S., Maha Qadri P.Z., *A New Generalized Poisson-Lindley Distribution: Applications and Properties* (2015)
7. Chakraborty S., On Some Distributional Properties of the Family of Weighted Generalized Poisson Distribution. *Communications in Statistics - Theory and Methods*, 39 (15), 2767–2788 (2010)
8. Ghitany M.E., Atieh B., Nadarajah S., Lindley distribution and its application (2007)
9. Glaser R. E., Bathtub and related failure rate characterizations. *J. Amer. Statist. Assoc.*, 75, 667-672 (1980)
10. Gomez Deniz E. and Calderin Ojeda E., The discrete Lindley distribution: properties and applications. *J Stat Comput. Simul.* 81(11), 1405–1416 (2011)
11. Klugman S., Panjer H., Willmot G., *Loss Models, From Data to Decisions*, John Wiley and Sons, New York (2012)
12. Lagakos S.K., A covariate model for partially censored data subject to competing causes of failure, *Journal of the Royal Statistical Society* 235-241 (1978)
13. Lawless J. F., *Statistical models and methods for lifetime data*. Wiley, New York (2003)
14. Lindley D.V., Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B* 20, 102–107 (1958)

15. Lindley D.V., Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II: Inference, Cambridge University Press, New York (1965)
16. Mazucheli L., Achcar J. A., The Lindley distribution applied to competing risks lifetime data (2011)
17. Moeschberger Melvin D.H., Theory Competing Risks, Lubrecht & Cramer Ltd (1978)
18. Nadarajah S., Bakouch H.S, Jazi M.A, A new discrete distribution. Statistics 48(1), 200–240 (2014)
19. Parede Barco K. V., Mazucheli J. & Janeiro V. ,The inverse power Lindley distribution (2016)
20. Pintilie M. ,Competing Risks: A Practical Perspective by Melania Pintilie, Wiley (2006)
21. Sarhan A.M., Analysis of incomplete, censored data in competing risks models with generalized exponential distributions (2007)
22. Song K.-S., Re'nyi information, loglikelihood and an intrinsic distribution measure, Journal of Statistical Planning and Inference 93, 51–69 (2001)
23. Zamani H. and Ismail N., Negative Binomial-Lindley Distribution and Its Application Actuarial Science Program, School of Mathematical Sciences, University Kebangsaan Malaysia (2010)