

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**Η ΛΟΓΙΣΤΙΚΗ ΚΑΤΑΝΟΜΗ,
ΓΕΝΙΚΕΥΣΕΙΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ**

Γεώργιος Ι. Μασούρας

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Σεπτέμβριος 2014

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 Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Κούτρας Μάρκος (Επιβλέπων)
- Πολίτης Κωσταντίνος
- Τσίμπος Κλέων

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
APPLIED STATISTICS**

**LOGISTIC DISTRIBUTION,
GENERALIZATIONS AND ITS
APPLICATIONS**

By

George J. Masouras

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 Science in
Applied Statistics

Piraeus, Greece
September 2014

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

*Στους γονείς μου
Ιωάννη και Μαρία*

Πανεπιστήμιο Πειραιώς

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Κύριο Μάρκο Κούτρα για την πολύτιμη συμπαράστασή του και υπομονή που επέδειξε σε όλη τη διάρκεια της διπλωματικής μου. Προσπερνώ τις παιδαγωγικές του αρετές που κάθε καθηγητής οφείλει να έχει και στέκομαι στην ανθρώπινή του ιδιότητα που είχα την χαρά να γνωρίσω και να γευτώ. Ιδιότητα που μόνο συναισθήματα θαυμασμού και ζήλειας, με την έννοια της έμπνευσης και παραδειγματισμού, μπορούν να προκαλέσουν. Επίσης θέλω να ευχαριστήσω τη συμφοιτήτρια και φίλη μου Μαρία Αμπατζιάδου για τη ψυχολογική στήριξη.

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Περίληψη

Στο πρώτο κεφάλαιο της παρούσας διπλωματικής εργασίας εισάγουμε το μοντέλο της λογιστικής εξίσωσης κάνοντας μια ιστορική αναδρομή για τη θέση που κατείχε στη βιβλιογραφία. Παρουσιάζουμε τις διάφορες εκφράσεις που έχουν εμφανισθεί στη βιβλιογραφία για αυτή και καταλήγουμε στη σύνδεσή της με τη λογιστική κατανομή.

Στο Κεφάλαιο 2 παραθέτουμε τους τύπους της λογιστικής κατανομής και παρουσιάζουμε αναλυτικά τις ιδιότητες και τα χαρακτηριστικά της.

Στο Κεφάλαιο 3 γίνεται αναφορά σε διάφορες μεθόδους εκτίμησης των παραμέτρων της λογιστικής κατανομής. Παράγουμε ψευδοτυχαίους αριθμούς από τη λογιστική κατανομή με συγκεκριμένες παραμέτρους και γίνεται πρακτική σύγκριση των μεθόδων εκτίμησης.

Στο Κεφάλαιο 4 παρουσιάζονται οι τέσσερις κύριοι τύποι γενικευμένων λογιστικών μοντέλων καθώς και ένα πρόσφατο γενικευμένο λογιστικό μοντέλο με έξι παραμέτρους. Συζητάμε μερικές από τις ιδιότητες και τα χαρακτηριστικά τους, και παραθέτουμε διάφορα αποτελέσματα που αφορούν τις σχέσεις τους με άλλες γνωστές κατανομές.

Στο τελευταίο κεφάλαιο εφαρμόζουμε μεθόδους εκτίμησης των παραμέτρων σε πραγματικά δεδομένα που ακολουθούν τη λογιστική.

Abstract

The logistic growth function is one of the oldest growth functions in the literature and is used to describe both population growth as well as organismic growth. In the first chapter of this thesis we introduce the above function as originally proposed by Verhulst(1838) and we discuss all these common characteristics that led to the born of the logistic distribution. In Chapter 2 we present several results for the logistic distribution which can be utilized in model building.

In Chapter 3, we discuss several estimation methods as applied to the logistic population. An illustrative example is worked out by generating random numbers from the logistic distribution with specific parameters.

In Chapter 4, the four major types of generalized logistic distributions are presented. Furthermore, a reference to the most recent six-parameter generalized logistic model is made. We discuss their properties and characteristics as well as their relationships to some other common statistical distributions.

Finally, in Chapter 5 an example with real data is presented to illustrate how all these methods of estimation can be applied.

Περιεχόμενα

Περίληψη.....	ix
Abstract.....	x
Περιεχόμενα.....	xi
Κατάλογος πινάκων.....	xiii
Κατάλογος σχημάτων.....	xv
Κατάλογος συντομογραφιών.....	xviii

ΚΕΦΑΛΑΙΟ 1: Η Λογιστική Συνάρτηση

1.1 Ιστορική Αναδρομή.....	1
1.1.1 «Εναλλακτική έκφραση» του τύπου της λογιστικής καμπύλης.....	4
1.1.2 Η Logit συνάρτηση.....	5
1.2 Τύπος Λογιστικής Συνάρτησης.....	5
1.3 Η Λογιστική Κατανομή.....	8

ΚΕΦΑΛΑΙΟ 2: Λογιστική Κατανομή και Ιδιότητες

2.1 Τύποι Λογιστικής Κατανομής.....	9
2.2 Η Τυποποιημένη Λογιστική Κατανομή.....	10
2.3 Χαρακτηριστικά Λογιστικής Κατανομής.....	12
2.4 Ιδιότητες Λογιστικής Κατανομής.....	15
2.5 Παραγωγή Τυχαίων Αριθμών από τη Λογιστική Κατανομή.....	18
2.6 Ομοιότητες Λογιστικής με Κανονική Κατανομή.....	19
2.7 Χαρακτηριστικά της Λογιστικής Κατανομής.....	21
2.8 Μέτρα θέσης και διασποράς.....	24
2.9 Διάμεσος και Ποσοστιαία σημεία.....	29

ΚΕΦΑΛΑΙΟ 3: Στατιστική Συμπερασματολογία

3.1	Εκτιμήτριες Μέγιστης Πιθανοφάνειας για πλήρη δεδομένα.....	30
3.2	Διαστήματα εμπιστοσύνης για εκτιμήτριες μέγιστης πιθανοφάνειας.....	33
3.3	Εκτιμήτριες Μέγιστης Πιθανοφάνειας για λογοκριμένα δεδομένα.....	35
3.4	Εκτιμήτριες ροπών για τις παραμέτρους α και β	35
3.5	Γραφικές μέθοδοι	40
ΚΕΦΑΛΑΙΟ 4: Γενικεύσεις/Τροποποιήσεις της Λογιστικής Κατανομής		
4.1	Εισαγωγή.....	48
4.2	Χαρακτηριστικά.....	54
4.3	Σχέση της γενικευμένης λογιστικής κατανομής με άλλες κατανομές	58
4.4	Εισαγωγή του γενικευμένου λογιστικού μοντέλου με έξι παραμέτρους.....	73
4.4.1	Γενικευμένο λογιστικό μοντέλο με έξι παραμέτρους.....	73
4.4.2	Σχέση της γενικευμένης λογιστικής κατανομής με έξι παραμέτρους με άλλες κατανομές.....	75
ΚΕΦΑΛΑΙΟ 5: Εφαρμογές της Λογιστικής Κατανομής		
Παράρτημα		
	Εντολές στο Mathematica.....	93
	Βιβλιογραφία	99

Κατάλογος Πινάκων

1.1	Προσαρμογή λογιστικής καμπύλης στην απογραφή πληθυσμού των ΗΠΑ για το χρονικό διάστημα 1790-1960.....	2
3.1	Προσομοιωμένες τιμές εκτίμησης παραμέτρων λογιστικής κατανομής.....	44
4.1	Οι τέσσερις κύριοι τύποι των γενικευμένων λογιστικών κατανομών.....	54
4.2	Οι Ροπογεννήτριες των τεσσάρων κυρίων τύπων των γενικευμένων λογιστικών κατανομών.....	67
5.1	Χρόνοι ζωής σε ώρες εκατό λαμπτήρων πυρακτώσεως.....	81

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Κατάλογος Σχημάτων

1.1	Γράφημα λογιστικής καμπύλης.....	4
1.2	Γράφημα λογιστικής συνάρτησης για διάφορες τιμές του k	6
2.1	Γράφημα της συνάρτησης πυκνότητας της τυποποιημένης λογιστικής κατανομής.....	8
2.2	Γράφημα της συνάρτησης κατανομής της τυποποιημένης λογιστικής κατανομής.....	12
2.3	Γράφημα των συναρτήσεων πυκνότητας της τυποποιημένης λογιστικής και τυποποιημένης κανονικής κατανομής.....	13
2.4	Γράφημα της συνάρτησης πυκνότητας της λογιστικής κατανομής για σταθερό σ και διαφορετικό μ	13
2.5	Γράφημα της συνάρτησης πυκνότητας της λογιστικής κατανομής για διαφορετικό σ και σταθερό μ	14
2.6	Γράφημα της βαθμίδας αποτυχίας $h(t)$	17
2.7	Γράφημα της διαφοράς $G_2(x) - G_1(x)$ των συναρτήσεων κατανομής της λογιστικής και κανονικής κατανομής.....	20
4.1	Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I για διάφορες τιμές της παραμέτρου α	49
4.2	Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου II για διάφορες τιμές της παραμέτρου α	50
4.3	Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου III για διάφορες τιμές της παραμέτρου α	52
4.4	Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I όταν η τιμή της παραμέτρου α παίρνει τιμές στο διάστημα $(0, 1)$	57
4.5	Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I όταν η τιμή της παραμέτρου α παίρνει τιμές στο διάστημα $(1, \infty)$	57
4.6	Γράφημα της συνάρτησης πυκνότητας της τυποποιημένης εκθετικής κατανομής.....	59

4.7	Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I για τιμές της παραμέτρου α κοντά στο 0.....	59
4.8	Γράφημα της συνάρτησης πυκνότητας της κατανομής Gumbel.....	61
4.9	Γράφημα της γενικευμένης λογιστικής συνάρτησης Τύπου I για μεγάλες τιμές της παραμέτρου α	61
4.10	Γράφημα της συνάρτησης πυκνότητας της τυποποιημένης κανονικής κατανομής και της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής Τύπου III για συγκεκριμένη τιμή της παραμέτρου α	65
4.11	Γράφημα της συνάρτησης πυκνότητας της κατανομής Student(15) και της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής Τύπου III για συγκεκριμένη τιμή της παραμέτρου α	66
5.1	Γράφημα διασποράς και της προσαρμοσμένης ευθείας παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων που εφαρμόστηκε στα δεδομένα του Πίνακα 5.1.....	86
5.2	Γράφημα της εμπειρικής συνάρτησης κατανομής για τα δεδομένα του Πίνακα 5.1.....	87
5.3	Γράφημα της συνάρτησης κατανομής από τη λογιστική κατανομή με παραμέτρους όπως εκτιμήθηκαν (EMΠ) από τον Πίνακα 5.1.....	88
5.4	Γράφημα της εμπειρικής συνάρτησης και της συνάρτησης κατανομής από τη λογιστική κατανομή με παραμέτρους όπως εκτιμήθηκαν (EMΠ) από τον Πίνακα 5.1.....	88

Πανεπιστήμιο Πειραιώς

Κατάλογος Συντομογραφιών

$P(t)$	Μέγεθος πληθυσμού
$f(t)$	Συνάρτηση πυκνότητας πιθανότητας
$F(t)$	Αθροιστική συνάρτηση κατανομής
$E[X]$	Μέση τιμή της τυχαίας μεταβλητής X
$\text{var}(X)$	Διακύμανση της τυχαίας μεταβλητής X
$X \sim L(\mu, \sigma^2)$	Η τυχαία μεταβλητή X ακολουθεί τη λογιστική κατανομή με μέσο μ και διακύμανση σ^2
$Y \sim L(0, 1)$	Η τυχαία μεταβλητή Y ακολουθεί την <i>τυποποιημένη</i> λογιστική κατανομή
$Z \sim L\left(0, \frac{\pi^2}{3}\right)$	Η τυχαία μεταβλητή Z ακολουθεί την <i>τυπική</i> λογιστική κατανομή
$X \sim N(0, 1)$	Η τυχαία μεταβλητή X ακολουθεί την <i>τυποποιημένη</i> κανονική κατανομή
$R(t)$	Συνάρτηση αξιοπιστίας
$S(t)$	Συνάρτηση επιβίωσης
$h(t)$	Συνάρτηση κινδύνου
$r(x)$	Αντίστροφη συνάρτηση κινδύνου
IFR	Αύξουσα βαθμίδα αποτυχίας
$U \sim U(0, 1)$	Η τυχαία μεταβλητή U ακολουθεί την ομοιόμορφη κατανομή στο διάστημα $(0, 1)$
$M_Z(t)$	Ροπογεννήτρια συνάρτηση της τυχαίας μεταβλητής Z
$\phi_Z(t)$	Χαρακτηριστική συνάρτηση της τυχαίας μεταβλητής Z
$\zeta(s)$	Συνάρτηση ζήτα του Riemann
EMΠ	Εκτιμητήρια Μέγιστης Πιθανοφάνειας
ΕΣΚ	Εμπειρική Συνάρτηση Κατανομής
LPP	Logistic Probability Plot ή Χάρτης Πιθανότητας για τη λογιστική

	κατανομή
MP	Μέθοδος των ροπών
ΓM	Γραφική μέθοδος
MSV	<i>Mean sum variation</i>
$G_a^I(\mathbf{y})$	Αθροιστική συνάρτηση κατανομής του λογιστικού μοντέλου τύπου I
$g_a^I(\mathbf{y})$	Συνάρτηση πυκνότητας πιθανότητας του λογιστικού μοντέλου τύπου I
$G_a^{II}(\mathbf{y})$	Αθροιστική συνάρτηση κατανομής του λογιστικού μοντέλου τύπου II
$g_a^{II}(\mathbf{y})$	Συνάρτηση πυκνότητας πιθανότητας του λογιστικού μοντέλου τύπου II
$G_a^{III}(\mathbf{y})$	Αθροιστική συνάρτηση κατανομής του λογιστικού μοντέλου τύπου III
$g_a^{III}(\mathbf{y})$	Συνάρτηση πυκνότητας πιθανότητας του λογιστικού μοντέλου τύπου III
$G_{pq}^{IV}(\mathbf{y})$	Αθροιστική συνάρτηση κατανομής του λογιστικού μοντέλου τύπου IV
$g_{pq}^{IV}(\mathbf{y})$	Συνάρτηση πυκνότητας πιθανότητας του λογιστικού μοντέλου τύπου IV
$T \sim \exp(\lambda)$	Η τυχαία μεταβλητή T ακολουθεί την εκθετική κατανομή με παράμετρο λ
$T \sim \Gamma(\alpha, \beta)$	Η τυχαία μεταβλητή T ακολουθεί την κατανομή Γάμμα με παραμέτρους α και β
$T \sim B(\alpha, \beta)$	Η τυχαία μεταβλητή T ακολουθεί την κατανομή Βήτα με παραμέτρους α και β
τ.μ	Τυχαία μεταβλητή
δ.ε	Διάστημα εμπιστοσύνης
ε.σ	Επίπεδο σημαντικότητας

Κεφάλαιο 1^ο

Η Λογιστική Συνάρτηση

1.1 Ιστορική Αναδρομή

Η λογιστική συνάρτηση (*logistic function*) αποτελεί μία από τις πρώτες συναρτήσεις η οποία χρησιμοποιήθηκε για να παραστήσει μοντέλα ανάπτυξης. Απεικονίζεται με μια καμπύλη σχήματος «S», γι' αυτό συνηθίζεται να συναντάται και με το όνομα «σιγμοειδής» (*sigmoid*). Ο όρος «σιγμοειδής» προέκυψε από το ελληνικό γράμμα «ς». Ο πρώτος που χρησιμοποίησε τον όρο *λογιστική* ήταν ο Pierre Francois Verhulst (1844-1845). Παρόλο που στο σύγγραμμά του «*Mathematical Researches into the Law of Population Growth Increase*» δεν εξηγεί το λόγο για τον οποίο την ονόμασε λογιστική, η λογαριθμική βάση της συνάρτησης φαίνεται να αποτέλεσε την βασική αιτία. Ο όρος καθιερώθηκε αργότερα από τους Reed and Berkson (1929) οι οποίοι ήταν δύο από τους πρώιμους μελετητές και υποστηρικτές της συνάρτησης.

Η λογιστική συνάρτηση χρησιμοποιήθηκε από τον Verhulst σε δημογραφικές μελέτες που έκανε, επεκτείνοντας το έργο του προγενέστερου του Thomas Malthus (1798). Το έργο του αφορούσε την ανάπτυξη ενός βιολογικού πληθυσμού. Εισηγήθηκε τη λογιστική εξίσωση υποθέτοντας ότι ο ρυθμός με τον οποίο αναπτύσσεται ένας πληθυσμός εξαρτάται από δυο κύριους παράγοντες:

- Τους διαθέσιμους πόρους που υπάρχουν στο φυσικό περιβάλλον
- Την κατανάλωση – «ζήτηση» που υπάρχει από τον εν λόγω πληθυσμό

Η ισορροπία που υπάρχει μεταξύ των δυο παραπάνω καταστάσεων, επηρεάζει το ρυθμό. Μικροί πληθυσμοί κατά την εμφάνισή τους αυξάνονται γεωμετρικά λόγω της αφθονίας των πόρων. Με την πάροδο του χρόνου όμως, ο πληθυσμός αυξάνεται και η κάλυψη των πόρων εξασθενεί. Η εξάντληση των πόρων σε συνδυασμό με την ολοένα και αυξανόμενη «ζήτηση» ωθεί το σύστημα προς μια συντηρητική κατάσταση, οδηγώντας το σε μια ισορροπία. Ο Verhulst υπέθεσε ότι η αύξηση του φυσικού λογάριθμου του μεγέθους ενός πληθυσμού συναρτήσει του χρόνου για μια

συγκεκριμένη γεωγραφική περιοχή ισούται με τη διαφορά μιας σταθερής ποσότητας και μιας συνάρτησης που αυξάνει με τον πληθυσμό.

Ας συμβολίσουμε το μέγεθος του πληθυσμού τη χρονική στιγμή t με $P(t)$. Το μοντέλο όπως παρουσιάστηκε από τον Verhulst περιγράφεται από την εξής διαφορική εξίσωση

$$\frac{dP(t)}{dt} = rP(t)\left(1 - \frac{P(t)}{K}\right) \quad (1.1)$$

όπου r ένας σταθερός όρος (παράμετρος του Malthus) και K η μέγιστη τιμή που μπορεί να πάρει η συνάρτηση $P(t)$. Η τιμή K είναι σημαντικό να θεωρηθεί ανεξάρτητη από τυχόν τεχνικές και κοινωνικές περιβαλλοντικές αλλαγές. Για να κατανοήσουμε τη σημαντικότητα αυτής της υπόθεσης, αναφέρουμε το παράδειγμα της περιγραφής της αύξησης του πληθυσμού των Η.Π.Α από τη λογιστική συνάρτηση για το διάστημα 1790-1960 (Πίνακας 1.1). Οι εκτιμήσεις μέσω της λογιστικής καμπύλης αρχίζουν και αποκλίνουν από τις πραγματικές τιμές όσο απομακρυνόμαστε από το 1950 και μετά. Αυτό οφείλεται στο γεγονός ότι η συνάρτηση αδυνατεί να ερμηνεύσει και να αποτυπώσει την τεράστια αύξηση γεννήσεων μετά το τέλος του δευτέρου παγκόσμιου πολέμου, καθώς και τις βελτιωμένες συνθήκες στον τομέα της υγείας.

Πίνακας 1.1 Προσαρμογή λογιστικής καμπύλης στην απογραφή πληθυσμού των ΗΠΑ (εξαιρούνται οι περιοχές Χαβάη και Αλάσκα) για το χρονικό διάστημα 1790-1960

Πληθυσμός (σε εκατομμύρια)			
Έτος	Δεδομένα Απογραφής	Εκτίμηση Λογιστικής καμπύλης	Σφάλμα Εκτίμησης
1790	3.9	3.7	0.2
1800	5.3	5.1	0.2
1810	7.2	7.0	0.2
1820	9.6	9.5	0.1
1830	12.9	12.8	0.1
1840	17.1	17.3	-0.2
1850	23.2	23.0	0.2
1860	31.4	30.3	1.1
1870	38.6	39.3	-0.7
1880	50.2	50.2	0.0
1890	62.9	62.8	0.1
1900	76.0	76.7	0.7

1910	92.0	91.4	0.6
1920	105.7	106.1	-0.4
1930	122.8	120.1	2.7
1940	131.4	132.8	-1.4
1950	150.7	143.8	6.9
1960	178.5	153.0	25.5

Η εξίσωση (1.1) δείχνει το ρυθμό με τον οποίο μεταβάλλεται η συνάρτηση $P(t)$. Η τιμή της σταθεράς r καθορίζει την αναλογική αύξηση της $P(t)$ στη μονάδα του χρόνου. Ο συνεχής αυξανόμενος ρυθμός φαίνεται από την ποσότητα $+rP(t)$. Η ποσότητα $-rP^2(t)/k$ έχει το ρόλο του αντισταθμιστικού όρου μειώνοντας το ρυθμό μεταβολής όσο το μέγεθος του πληθυσμού πλησιάζει προς την τιμή K . Όταν η $P(t)$ πάρει την τιμή K τότε ο ρυθμός μηδενίζεται και η $P(t)$ σταθεροποιείται.

Διαιρώντας και τα δύο μέλη της (1.1) με K έχουμε

$$\frac{d}{dt} \frac{P(t)}{K} = r \frac{P(t)}{K} \left(1 - \frac{P(t)}{K} \right)$$

και θέτοντας $x(t) = \frac{P(t)}{K}$ παίρνουμε

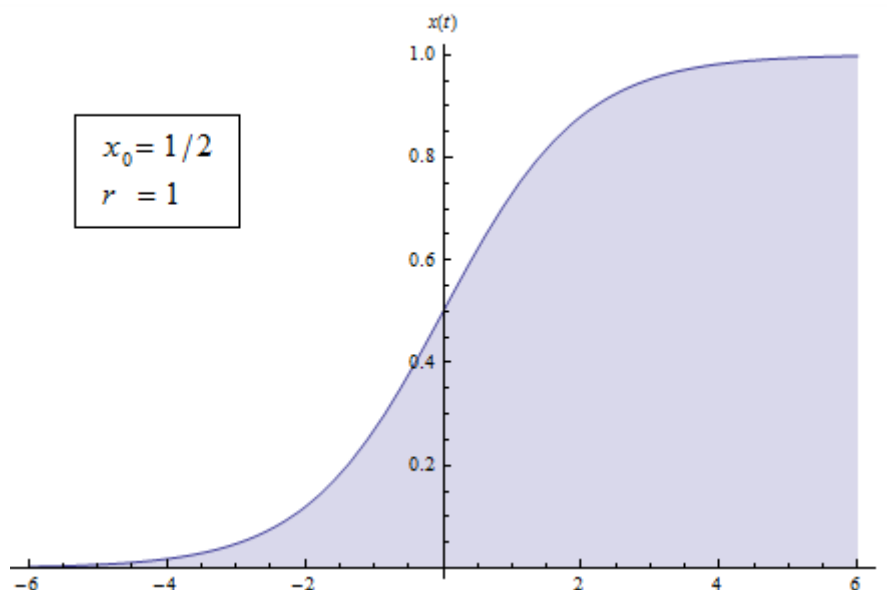
$$\frac{dx(t)}{dt} = r \cdot x(t) [1 - x(t)]. \quad (1.2)$$

Η λύση της πρώτης τάξεως μη γραμμικής διαφορικής εξίσωσης (1.2) γνωστή και ως λογιστική εξίσωση

$$x(t) = \frac{1}{1 + \left(\frac{1}{x_0} - 1 \right) \exp(-rt)} \quad (1.3)$$

αποτελεί τη λογιστική συνάρτηση ή λογιστική καμπύλη. Με x_0 θα συμβολίζουμε την αρχική συνθήκη $x_0 = x(0)$. Παρακάτω δίνεται το γράφημα της λογιστικής καμπύλης συναρτήσει του t .

Σχήμα 1.1 Γράφημα λογιστικής καμπύλης



Η λογιστική συνάρτηση έχει καλή προσαρμογή ως μοντέλο στην ανάπτυξη πληθυσμών διαφόρων ειδών. Πληθυσμοί όπως αυτού του ανθρώπινου είδους (Pearl και Reed (1920)), των ψαριών (Jensen (1975)), των ζώων (Miller και Botkin (1974)), των βακτηρίων και των κυττάρων (Tan (1983)), και των καρκινικών κυττάρων (Eisen (1974)). Επίσης προσαρμόζεται ως μοντέλο ανάπτυξης στην εξάπλωση μιας φήμης (Oliver (1969)) ή στη διάδοση μιας πληροφορίας μέσα σε ένα κοινωνικό σύνολο καθώς και στην εισχώρηση νέων προϊόντων στην αγορά. Σήμερα χρησιμοποιείται σε ένα ευρύ πεδίο εφαρμογών όπως τεχνητά νευρωνικά δίκτυα, βιολογία, βιομαθηματικά, δημογραφία, οικονομικά, χημεία, πολιτικές επιστήμες και στατιστική.

1.1.1 «Εναλλακτική έκφραση» του τύπου της (απλής) λογιστικής καμπύλης

Οι βασικές υπερβολικές συναρτήσεις και οι τύποι τους είναι:

- Υπερβολικό ημίτονο $\sinh(x) = \frac{e^x - e^{-x}}{2}$
- Υπερβολικό συνημίτονο $\cosh(x) = \frac{e^x + e^{-x}}{2}$
- Υπερβολική εφαπτομένη $\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

Παίρνοντας τον τύπο της υπερβολικής εφαπτομένης έχουμε

$$1 + \tanh(x) = \frac{2}{1 + e^{-2x}} \Leftrightarrow \frac{1}{2} [\tanh(x)] = \frac{1}{1 + e^{-2x}}$$

και τελικά

$$\frac{1}{2} \left[\tanh \left(\frac{x}{2} \right) \right] = \frac{1}{1 + e^{-x}} \quad (1.4)$$

Επομένως σύμφωνα με την (1.4), η (1.3) για $x_0 = \frac{1}{2}$ διαμορφώνεται ως εξής

$$x(t) = \frac{1}{1 + \exp(-rt)} = \frac{1}{2} \tanh \left(\frac{rt}{2} \right).$$

1.1.2 Η συνάρτηση Logit

Η συνάρτηση logit για έναν αριθμό p που παίρνει τιμές στο υποσύνολο των πραγματικών αριθμών $[0,1)$ ορίζεται από τον τύπο

$$\text{logit}(p) = \log \left(\frac{p}{1-p} \right)$$

Αντιστρέφοντας την παραπάνω συνάρτηση προκύπτει

$$y = \log \left(\frac{p}{1-p} \right) \Leftrightarrow e^y = \frac{p}{1-p} \Leftrightarrow p = \frac{e^y}{1+e^y} = \frac{1}{1+e^{-y}}$$

Άρα η σχέση που συνδέει τις δυο συναρτήσεις είναι

$$\text{logit}^{-1}(y) = \frac{1}{1+e^{-y}}$$

Η αντίστροφη της logit συνάρτησης για έναν αριθμό y που παίρνει τιμές στο υποσύνολο των πραγματικών αριθμών $[0,1]$ είναι η λογιστική συνάρτηση.

1.2 Ο γενικός τύπος της Λογιστικής Συνάρτησης

Ο γενικός τύπος της λογιστικής συνάρτησης περιέχει τρεις θετικές παραμέτρους L, C, k και δίνεται από τον τύπο

$$f(t) = \frac{L}{1 + C \exp(-kt)} \quad (1.5)$$

Όσο η μεταβλητή αυξάνεται ($t \rightarrow \infty$) θα ισχύει $\exp(-t) \rightarrow 0$ και επομένως, η συνάρτηση $f(t)$ θα συγκλίνει προς τη σταθερά L , δηλαδή $f(t) \rightarrow L$. Η παράμετρος L αποτελεί την οριακή μέγιστη τιμή για τη συνάρτηση. Για να κατανοήσουμε την παράμετρο C δίνουμε την τιμή 0 στην μεταβλητή t οπότε παίρνουμε

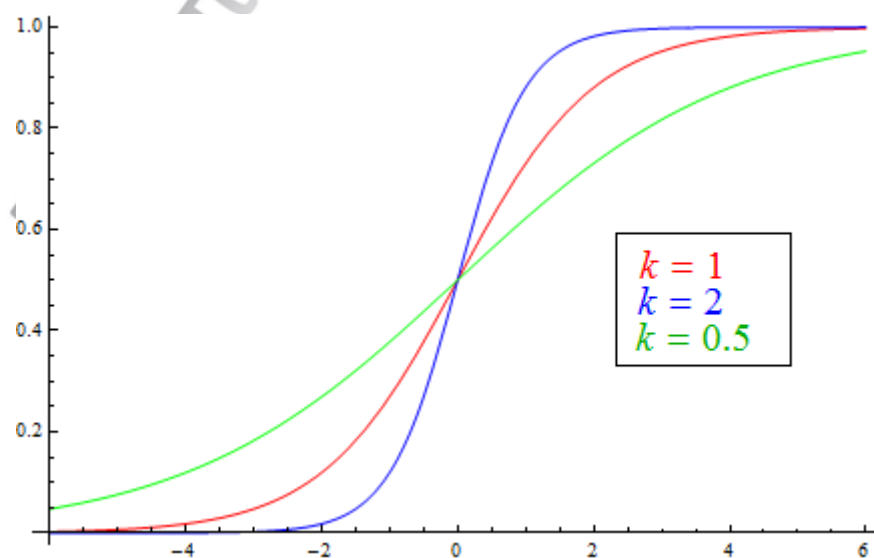
$$f(0) = \frac{L}{1+C}.$$

Αν λύσουμε ως προς C έχουμε

$$C = \frac{L}{f(0)} - 1 = \frac{L - f(0)}{f(0)}.$$

Η παράμετρος C μας δείχνει το λόγο της διαφοράς της οριακής μέγιστης τιμής L από την αρχική τιμή $f(0)$ προς την αρχική τιμή $f(0)$. Τέλος, η παράμετρος k μας δείχνει πόσο μεγάλος είναι ο ρυθμός αύξησης της συνάρτησης, δηλαδή όσο πιο μεγάλο k έχουμε τόσο πιο γρήγορα η καμπύλη θα προσεγγίσει τον οριζόντιο άξονα $y = L$. Παρακάτω δίνεται το γράφημα της λογιστικής συνάρτησης για διάφορες τιμές του k και $L = C = 1$.

Σχήμα 1.2 Γράφημα λογιστικής συνάρτησης για διάφορες τιμές του k



Ένα σημαντικό χαρακτηριστικό της λογιστικής συνάρτησης είναι ότι περιέχει ένα μοναδικό σημείο καμπής που χωρίζει τη συνάρτηση σε δύο όμοιες καμπύλες με αντίθετα κοίλα. Υπολογιστικά (παραγωγίζοντας δυο φορές την (1.5) και εξισώνοντας με το μηδέν) έχει βρεθεί ότι η λογιστική καμπύλη αλλάζει κοίλα στο σημείο t για το οποίο ισχύει

$$f(t) = \frac{L}{2}.$$

Εφόσον γνωρίζουμε την τεταγμένη του σημείου καμπής, εύκολα υπολογίζεται και η τεταγμένη του θέτοντας

$$\frac{L}{2} = \frac{L}{1 + C \exp(-kt)}$$

και λύνοντας ως προς t προκύπτει

$$t = \frac{\ln C}{k}$$

Έτσι, οι συντεταγμένες του σημείου καμπής είναι οι $\left[\frac{\ln C}{k}, \frac{L}{2} \right]$.

Η λογιστική συνάρτηση (1.3) προκύπτει αν στον τύπο (1.5) θέσουμε $L=1, C = \frac{1}{x_0} - 1$ και $k=r$. Οι συντεταγμένες του σημείου καμπής της θα είναι

$$\left[\frac{\ln \frac{1-x_0}{x_0}}{r}, \frac{1}{2} \right].$$

Ο τύπος (1.5) αφορά διαδικασίες ανάπτυξης για αυτό θεωρήσαμε ότι οι παράμετροι C και k είναι θετικές. Ένας πιο γενικός τύπος της λογιστικής συνάρτησης είναι ο εξής

$$f(x, \theta) = \frac{a}{1 + \beta \cdot \exp(\gamma \cdot x)}. \quad (1.6)$$

με $\theta = (\alpha, \beta, \gamma)$ το διάνυσμα των παραμέτρων. Ένας ακόμη πιο γενικός τύπος ο οποίος δόθηκε από τον Count (1943) είναι ο επόμενος

$$f(x, \theta) = \left(\frac{a}{1 + \beta e^{\gamma x}} \right)^{-\frac{1}{a}}. \quad (1.7)$$

Στόχος του ήταν να έχει μια πιο ευέλικτη μορφή της συνάρτησης όσον αφορά το σημείο καμπής.

Η λογιστική καμπύλη συναντάται στη βιβλιογραφία και με πιο περίπλοκες μορφές όπως

$$f_T(x, \theta_T) = \alpha_T \{1 + \tanh[\gamma_T(x - \beta_T)]\} \quad (1.8)$$

που δεν αποτελούν παρά μόνο απλές αναπαραμετροποιήσεις της. Αν στην σχέση

(1.6) αντί θ πάρουμε το παραμετρικό διάνυσμα $\theta = (\alpha_T, \beta_T, \gamma_T)$ με $\alpha_T = \frac{a}{2}$,

$\beta_T = -\frac{1}{\gamma} \ln \beta$ και $\gamma_T = -\frac{\gamma}{2}$ τότε προκύπτει η (1.7).

1.3 Η Λογιστική Κατανομή

Με βάση τη μορφή της λογιστικής συνάρτησης έχει προκύψει η λογιστική κατανομή η οποία έχει συνάρτηση κατανομής της μορφής

$$F(x) = P(X \leq x) = \frac{1}{1 + \exp\left(-\frac{x-a}{b}\right)} \quad -\infty < x < \infty. \quad (1.9)$$

όπου a, b οι παράμετροι θέσης και κλίμακας αντίστοιχα. Μια τυχαία μεταβλητή X που έχει συνάρτηση κατανομής την (1.9) θα λέμε ότι ακολουθεί τη λογιστική κατανομή με παραμέτρους a και b .

Αν παρατηρήσουμε τη συνάρτηση (1.5), ο (1.9) τύπος οδηγεί στη λογιστική συνάρτηση χρησιμοποιώντας το γραμμικό μετασχηματισμό $x = a + bt$.

Λύνοντας ως προς t προκύπτει $t = \frac{x-a}{b}$ και αν αντικαταστήσουμε στην (1.5) έχουμε

$$f(x) = \frac{L}{1 + C \exp\left[-k \frac{(x-a)}{b}\right]}.$$

Θέτοντας τιμές στις παραμέτρους $L, C, k = 1$ καταλήγουμε στη συνάρτηση κατανομής της λογιστικής κατανομής.

Αξίζει να σημειώσουμε ότι μέθοδοι που χρησιμοποιήθηκαν για την προσαρμογή της λογιστικής συνάρτησης ή λογιστικής καμπύλης ως μοντέλο ανάπτυξης, μπορούν να εφαρμοσθούν και στην περίπτωση της λογιστικής κατανομής. (Erkelens (1968), Balakrishnan (1992)).

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 2^ο

Λογιστική Κατανομή και Ιδιότητες

2.1 Τύποι Λογιστικής Κατανομής

Θα λέμε ότι μια τυχαία μεταβλητή X ακολουθεί τη λογιστική κατανομή με μέσο μ και διακύμανση σ^2 ($X \sim L(\mu, \sigma^2)$) αν η συνάρτηση κατανομής της X δίνεται από τον τύπο

$$F(x; \mu, \sigma) = \frac{1}{1 + \exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right]} \quad -\infty < x < \infty. \quad (2.1)$$

Η συνάρτηση πυκνότητας της X θα δίνεται από τον τύπο

$$f(x; \mu, \sigma) = \frac{\pi}{\sigma\sqrt{3}} \frac{\exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right]}{\left\{1 + \exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right]\right\}^2} \quad -\infty < x < \infty.$$

Για να αποφύγουμε στις παραπάνω σχέσεις το λόγο $\pi/\sqrt{3}$, κάνουμε αναπαραμετροποίηση θέτοντας $a = \mu$ και $\beta = \frac{\sigma\sqrt{3}}{\pi}$ οπότε προκύπτει

$$F(x; \alpha, \beta) = \frac{1}{1 + \exp\left(-\frac{x - \alpha}{\beta}\right)} \quad -\infty < x < \infty. \quad (2.2)$$

δηλαδή παίρνουμε την έκφραση (1.9). Η αντίστοιχη συνάρτηση πυκνότητας θα δίνεται από τον τύπο

$$f(x; \alpha, \beta) = \frac{1}{\beta} \frac{\exp\left(-\frac{x - \alpha}{\beta}\right)}{\left\{1 + \exp\left(-\frac{x - \alpha}{\beta}\right)\right\}^2} \quad -\infty < x < \infty. \quad (2.3)$$

Όταν η τυχαία μεταβλητή X ακολουθεί την τελευταία κατανομή προκύπτει ότι

$$E[X] = \alpha \text{ και } \text{var}(X) = \beta^2 \frac{\pi^2}{3} \text{ με } \beta > 0.$$

Όποιο από τα παραπάνω μοντέλα και αν χρησιμοποιηθεί θα πάρουμε τα ίδια αποτελέσματα.

Στο προηγούμενο κεφάλαιο αναφέραμε τη σχέση που συνδέει τη λογιστική συνάρτηση με την υπερβολική εφαπτομένη. Εφαρμόζοντας τη σχέση (1.4) στην (2.1), έχουμε

$$F(x; \mu, \sigma) = \frac{1}{2} \left\{ 1 + \tanh \left[\frac{\pi(x - \mu)}{2\sigma\sqrt{3}} \right] \right\} \quad -\infty < x < \infty.$$

Η υπερβολική τέμνουσα ($\text{sech}(x)$) ορίζεται ως ο λόγος $1/\cosh(x)$ και ισούται με $2/(e^x + e^{-x})$. Η συνάρτηση πυκνότητας της X σύμφωνα με την (2.3) και με χρήση της υπερβολικής τέμνουσας γράφεται

$$f(x; \mu, \sigma) = \frac{\pi}{4\sigma\sqrt{3}} \text{sech}^2 \left[\frac{\pi(x - \mu)}{\sigma\sqrt{3}} \right] \quad -\infty < x < \infty.$$

Η παραπάνω έκφραση της συνάρτησης πυκνότητας είναι και ο λόγος για τον οποίο η λογιστική κατανομή αναφέρεται πολλές φορές και με την ονομασία *sech-squared* κατανομή.

2.2 Η Τυποποιημένη Λογιστική Κατανομή

Θεωρούμε τη λογιστική κατανομή της σχέσης (2.1) και θέτουμε $E[Y] = \mu = 0$ και $\text{var}(Y) = \sigma^2 = 1$. Έτσι όταν η τυχαία μεταβλητή $X \square L(\mu, \sigma^2)$ τότε η τυχαία μεταβλητή $Y = (X - \mu)/\sigma \square L(0, 1)$. Η συνάρτηση κατανομής της θα δίνεται από τον τύπο

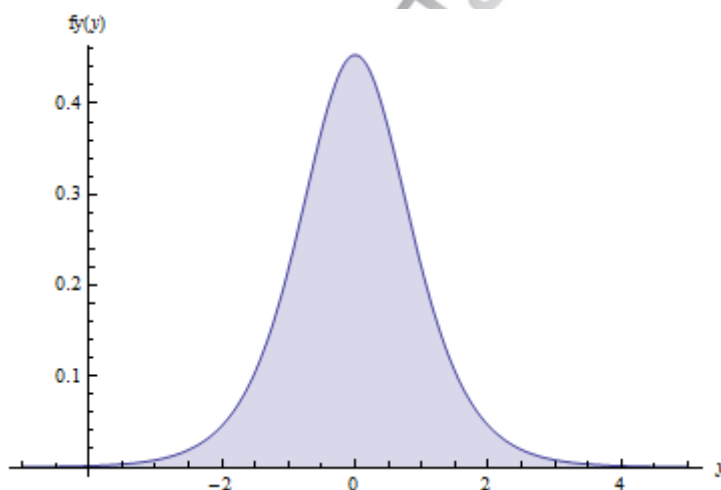
$$F(y; 0, 1) = \frac{1}{1 + \exp\left(-\frac{\pi y}{\sqrt{3}}\right)} \quad -\infty < y < \infty. \quad (2.4)$$

ενώ η συνάρτηση πυκνότητας από τον τύπο

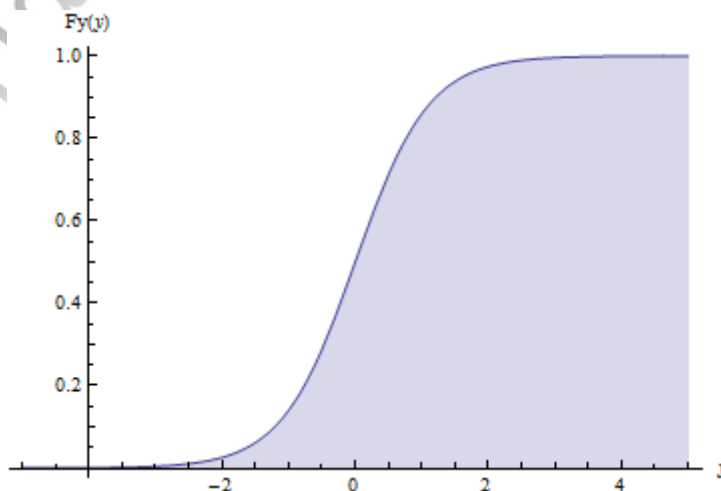
$$f(y;0,1) = \frac{\pi}{\sqrt{3}} \frac{\exp\left(-\frac{\pi y}{\sqrt{3}}\right)}{\left[1 + \exp\left(-\frac{\pi y}{\sqrt{3}}\right)\right]^2} \quad -\infty < y < \infty.$$

Η Y θα λέμε ότι ακολουθεί την τυποποιημένη λογιστική κατανομή και θα τη συμβολίζουμε με $Y \square L(0,1)$. Δίνονται παρακάτω γραφήματα της συνάρτησης πυκνότητας και της συνάρτησης κατανομής για την τυχαία μεταβλητή $Y \square L(0,1)$.

Σχήμα 2.1 Γράφημα της συνάρτησης πυκνότητας της τυποποιημένης λογιστικής κατανομής



Σχήμα 2.2 Γράφημα της συνάρτησης κατανομής της τυποποιημένης λογιστικής κατανομής



Αν στη σχέση (2.2) θέσουμε τιμές στις παραμέτρους $\alpha=0$ και $\beta=1$, η τυχαία μεταβλητή $Z = \pi(X - \mu) / \sigma\sqrt{3} = \pi Y / \sqrt{3}$ θα λέμε ότι ακολουθεί την τυπική λογιστική κατανομή. Οι συναρτήσεις κατανομής και πυκνότητας θα δίνονται από τους τύπους

$$F(z;0,1) = \frac{1}{1 + \exp(-z)} \quad -\infty < z < \infty. \quad (2.5)$$

και

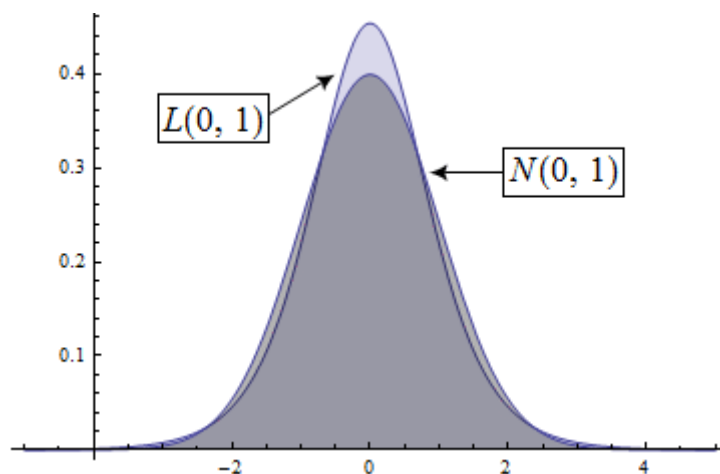
$$f(z;0,1) = \frac{\exp(-z)}{[1 + \exp(-z)]^2} \quad -\infty < z < \infty. \quad (2.6)$$

Η μέση τιμή $E[Z]$ για την τυχαία μεταβλητή Z θα είναι ίση με μηδέν και η διακύμανση $\text{var}(Z)$ ίση με $\pi^2/3$. Θα συμβολίζουμε με $Z \square L\left(0, \frac{\pi^2}{3}\right)$ και θα χρησιμοποιείται για ευκολία στις μαθηματικές πράξεις.

2.3 Χαρακτηριστικά Λογιστικής Κατανομής

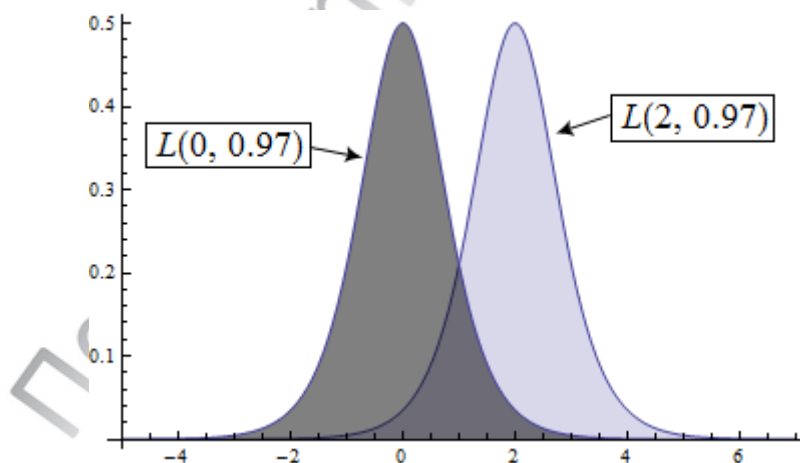
Η λογιστική κατανομή έχει δύο παραμέτρους. Την παράμετρο θέσης $\mu \in \mathbf{R}$ (*position parameter*) και την παράμετρο κλίμακας $\sigma > 0$ (*scale parameter*), ενώ δεν περιέχει παράμετρο μορφής (*shape parameter*). Η μορφή της λογιστικής κατανομής είναι παρόμοια με αυτή της κανονικής, μόνο που η λογιστική είναι πιο λεπτόκυρτη καταλήγοντας σε πιο ψηλή κορυφή στον μέσο. Μία από τις σημαντικότερες διαφορές τους είναι ότι η λογιστική κατανομή έχει πιο μακριές ουρές και ότι συνάρτηση κινδύνου (η οποία ορίζεται παρακάτω) στην άνω ουρά της κατανομής γίνεται σχεδόν παράλληλη με τον άξονα της μεταβλητής για μεγάλες τιμές της, προσεγγίζοντας την τιμή $1/\sigma$. Παρακάτω απεικονίζονται γραφικά οι συναρτήσεις πυκνότητας της τυποποιημένης κανονικής ($N(0,1)$) και λογιστικής ($L(0,1)$) κατανομής.

Σχήμα 2.2 Γράφημα των συναρτήσεων πυκνότητας της τυποποιημένης λογιστικής και τυποποιημένης κανονικής κατανομής



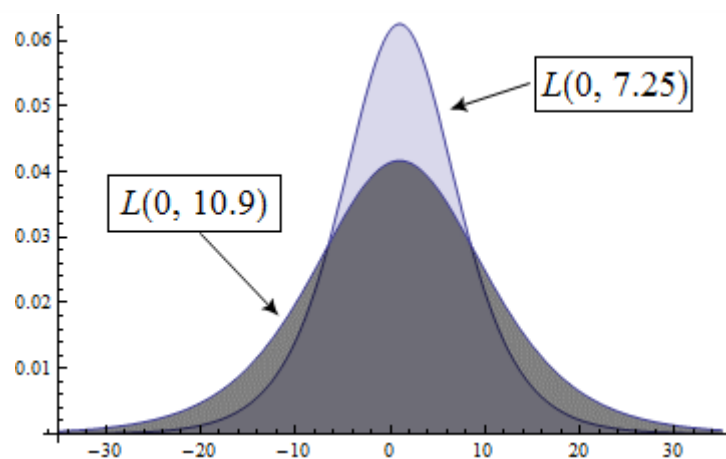
Όπως και η κανονική κατανομή, η λογιστική είναι συμμετρική γύρω από το μηδέν. Αυτό εύκολα προκύπτει από τον τύπο της συνάρτησης πυκνότητας για την οποία ισχύει $f_X(-x) = f_X(x) \quad \forall x \in \mathbf{R}$. Όσο αυξάνει η παράμετρος μ , η συνάρτηση πυκνότητας της κατανομής μετακινείται προς τα δεξιά ενώ όσο μειώνεται, μετακινείται προς την αντίθετη κατεύθυνση. Παρακάτω δίνεται η γραφική παράσταση της συνάρτησης πυκνότητας για σταθερό σ και διαφορετικό μ .

Σχήμα 2.4 Γράφημα της συνάρτησης πυκνότητας της λογιστικής κατανομής για σταθερό σ και διαφορετικό μ



Όσο η παράμετρος κλίμακας σ μειώνεται, η συνάρτηση πυκνότητας πιέζεται προς το μέσο δηλαδή γίνεται στενότερη και ψηλότερη ενώ όσο αυξάνεται, απλώνεται μακριά από το μέσο δηλαδή γίνεται φαρδύτερη και πιο κοντή. Παρακάτω δίνεται η γραφική παράσταση της συνάρτησης πυκνότητας για σταθερό μ και διαφορετικό σ .

Σχήμα 2.5 Γράφημα της συνάρτησης πυκνότητας της λογιστικής κατανομής για διαφορετικό σ και σταθερό μ



Για τιμές της τυχαίας μεταβλητής X στο διάστημα $(-\infty, \mu)$ η συνάρτηση πυκνότητας είναι αύξουσα (αφού αποδεικνύεται εύκολα ότι ισχύει $\frac{df(x; \mu, \sigma)}{dx} > 0$) και για τιμές στο (μ, ∞) είναι φθίνουσα (ισχύει $\frac{df(x; \mu, \sigma)}{dx} < 0$).

Συνεπώς σημείο μεγίστου παρουσιάζει για $x = \mu$ και ισούται με $\frac{\pi}{4\sigma\sqrt{3}}$. Για παράδειγμα, η συνάρτηση πυκνότητας της τυχαίας μεταβλητής $Y \sim L(0,1)$ έχει σημείο μεγίστου στο 0 και τιμή $f_Y(0) = \frac{\pi}{4\sqrt{3}}$.

Παίρνοντας τη δεύτερη παράγωγο της συνάρτησης πυκνότητας και εξισώνοντας με το 0 βρίσκουμε ότι τα σημεία στα οποία αλλάζει η καμπυλότητα και είναι τα $\mu + \sigma \log[2 \pm \sqrt{3}]$. Για την τυποποιημένη λογιστική κατανομή $L(0,1)$ τα σημεία καμπής είναι τα $\frac{\sqrt{3}}{\pi} \log[2 \pm \sqrt{3}]$.

2.4 Ιδιότητες Λογιστικής Κατανομής

Επειδή στην παρούσα ενότητα θα χρησιμοποιηθούν έννοιες όπως, συνάρτηση κινδύνου ή βαθμίδα αποτυχίας (*hazard rate* ή *failure rate*) και συνάρτηση αξιοπιστίας ή επιβίωσης (*Reliability function* ή *Survival function*) παρακάτω δίνονται οι ορισμοί τους.

Έστω T ο χρόνος ζωής μιας μονάδας (τυχαία μεταβλητή) και ας συμβολίσουμε με $R(t) = P(T > t)$ την πιθανότητα η μονάδα να επιβιώσει μέχρι τη χρονική στιγμή t . Η συνάρτηση $R(t)$ ονομάζεται συνάρτηση αξιοπιστίας (ή επιβίωσης $S(t)$). Αν υποθέσουμε ότι η υπό παρατήρηση μονάδα έχει επιζήσει στο χρονικό διάστημα $[0, t]$ τότε η δεσμευμένη πιθανότητα να επιβιώσει στο χρονικό διάστημα $(t, t+dt]$ θα ισούται με

$$P(T > t+dt | T > t) = \frac{P(T > t+dt)}{P(T > t)} = \frac{R(t+dt)}{R(t)}$$

ενώ η δεσμευμένη πιθανότητα να μην επιβιώσει στο ίδιο χρονικό διάστημα θα ισούται με

$$P(T \leq t+dt | T > t) = \frac{P(t < T \leq t+dt)}{P(T > t)} = \frac{R(t) - R(t+dt)}{R(t)}.$$

Αν υπάρχει η πρώτη παράγωγος της συνάρτησης αξιοπιστίας $R(t)$, τότε θα δίνεται από τον τύπο

$$R'(t) = \frac{dR(t)}{dt} = \lim_{dt \rightarrow 0} \frac{R(t+dt) - R(t)}{dt}$$

και παίρνοντας το λόγο

$$\frac{R'(t)}{R(t)} = \frac{1}{R(t)} \lim_{dt \rightarrow 0} \frac{R(t+dt) - R(t)}{dt} = \lim_{dt \rightarrow 0} \frac{1}{dt} \frac{R(t+dt) - R(t)}{R(t)}.$$

προκύπτει

$$-\frac{R'(t)}{R(t)} = \lim_{dt \rightarrow 0} \frac{P(T \leq t+dt | T > t)}{dt}.$$

Ο λόγος

$$h(t) = -\frac{R'(t)}{R(t)} = \lim_{dt \rightarrow 0} \frac{P(T \leq t+dt | T > t)}{dt}$$

ονομάζεται *βαθμίδα αποτυχίας* ή αλλιώς *συνάρτηση κινδύνου*. Η $h(t)$ ικανοποιεί τις σχέσεις

$$h(t) \geq 0 \quad \text{και} \quad \int_0^{\infty} h(t) dt = \infty.$$

Χρησιμοποιώντας την τυχαία μεταβλητή $Z \square L\left(0, \frac{\pi^2}{3}\right)$, ο Gumbel (1961) επισήμανε τις εξής ιδιότητες που ισχύουν για τη λογιστική κατανομή:

$$f_z(z) = \frac{\exp(-z)}{[1 + \exp(-z)]^2} = \frac{1}{1 + \exp(-z)} \frac{1}{1 + \exp(z)} = F_z(z)[1 - F_z(z)] \quad (2.7)$$

και

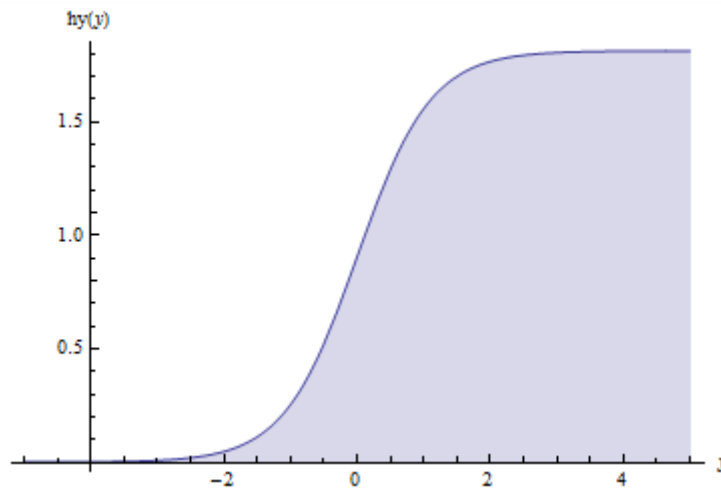
$$z = \log\left(\frac{F_z(z)}{1 - F_z(z)}\right). \quad (2.8)$$

Από τη σχέση (2.7) προκύπτει άμεσα ότι

$$h_z(z) = \frac{f_z(z)}{1 - F_z(z)} = F_z(z). \quad (2.9)$$

Σύμφωνα με την (2.9), καταλήγουμε στο συμπέρασμα ότι για μια τυχαία μεταβλητή $Z \square L\left(0, \frac{\pi^2}{3}\right)$, η βαθμίδα αποτυχίας (ή συνάρτηση κινδύνου) είναι ίση με τη συνάρτηση κατανομής. Αν χρησιμοποιούσαμε την τυχαία μεταβλητή $Y \sim L(0,1)$ αντί της $Z \square L\left(0, \frac{\pi^2}{3}\right)$ τότε η βαθμίδα αποτυχίας $h_Y(y)$ θα ισούταν με $\frac{\pi}{\sqrt{3}} F_Y(y)$, δηλαδή η $h_Y(y)$ είναι ανάλογη της $F_Y(y)$. Παρακάτω δίνεται το γράφημα της $h_Y(y)$ για την τυχαία μεταβλητή Y .

Σχήμα 2.6 Γράφημα της βαθμίδας αποτυχίας $h(t)$



Έχοντας σαν οδηγό τη σχέση (2.5), ας δούμε κάποιες άλλες σχέσεις που προκύπτουν μεταξύ των συναρτήσεων που περιγράφουν την τυχαία μεταβλητή $Z \sim L\left(0, \frac{\pi^2}{3}\right)$.

Η συνάρτηση κινδύνου στο σημείο $-z$ είναι ίση με

$$h_z(-z) = \frac{f_z(-z)}{1 - F_z(-z)} = \frac{f_z(-z)}{F_z(z)}$$

και λόγω της συμμετρικότητας της συνάρτησης πυκνότητας της λογιστικής κατανομής, θα έχουμε

$$h_z(-z) = \frac{f_z(z)}{F_z(z)} = r_z(z) \quad (2.10)$$

όπου με $r(x)$ θα συμβολίζουμε την αντίστροφη συνάρτηση κινδύνου (*reversed hazard rate*). Χρησιμοποιώντας τη σχέση (2.9) έχουμε

$$r_z(z) = h_z(-z) = F_z(-z) = 1 - F_z(z) = R_z(z)$$

δηλαδή η αντίστροφη συνάρτηση κινδύνου και η συνάρτηση αξιοπιστίας είναι ίσες. Επίσης, για όλα τα $-\infty < z < \infty$ είδαμε ότι ισχύει

$$h_z(z) = F_z(z)$$

Πολλαπλασιάζοντας και τα δυο μέλη της εξίσωσης με $r_z(z)$, προκύπτει:

$$\begin{aligned}h_z(z)r_z(z) &= F_z(z)r_z(z) \Leftrightarrow \\h_z(z)r_z(z) &= f_z(z)\end{aligned}$$

Γνωρίζοντας ότι $h_z(z) = F_z(z)$ και $r_z(z) = R_z(z)$, έχουμε

$$h_z(z) + r_z(z) = 1$$

και

$$\frac{h_z(z)}{r_z(z)} = \frac{F_z(z)}{R_z(z)} = \frac{1}{\frac{1+e^{-z}}{1+e^z}} = (e^{-z})^{-1} = e^z.$$

Μια συνεχής τυχαία μεταβλητή X (ή η αντίστοιχη κατανομή) θα λέγεται ότι έχει την ιδιότητα *IFR* (*increasing failure rate*) ή απλά ότι είναι *IFR*, αν η βαθμίδα αποτυχίας $\lambda(x)$ είναι αύξουσα συνάρτηση του x .

Η *IFR* ιδιότητα ικανοποιείται από τη λογιστική κατανομή λόγω του γεγονότος ότι η βαθμίδα αποτυχίας είναι ανάλογη με τη συνάρτηση κατανομής της.

2.5 Παραγωγή Τυχαίων Αριθμών από τη Λογιστική κατανομή

Η ιδιότητα (2.7) χρησιμεύει στο να παράγουμε εύκολα τυχαίους αριθμούς από τη λογιστική κατανομή με παραμέτρους α και $\beta > 0$.

Ας συμβολίσουμε με γ τη συνάρτηση κατανομής για τη συνεχή τυχαία μεταβλητή

$$Z \sim L\left(0, \frac{\pi^2}{3}\right).$$

Τότε σύμφωνα με την (2.8) έχουμε

$$z = \log\left(\frac{\gamma}{1-\gamma}\right) \quad (2.11)$$

Αν συμβολίσουμε με γ τη συνάρτηση αξιοπιστίας τότε έχουμε αντίστοιχα

$$z = \log\left(\frac{1-\gamma}{\gamma}\right) \quad (2.12)$$

Οι παραπάνω σχέσεις εκφράζουν τις αντίστροφες συναρτήσεις της συνάρτησης κατανομής και αξιοπιστίας αντίστοιχα, της συνεχούς τυχαίας μεταβλητής $Z \sim L\left(0, \frac{\pi^2}{3}\right)$.

Υπενθυμίζεται ότι ισχύει το επόμενο:

Πρόταση. Αν μια συνεχής τυχαία μεταβλητή U ακολουθεί την ομοιόμορφη κατανομή στο διάστημα $(0,1)$ (συμβολικά $U \sim U(0,1)$) και F είναι μία οποιαδήποτε συνάρτηση κατανομής, τότε η τυχαία μεταβλητή $X = F^{-1}(U)$ έχει συνάρτηση κατανομής F .

Σύμφωνα με την πρόταση, αν αντικαταστήσουμε όπου z με U , δηλαδή

$$z = \log\left(\frac{U}{1-U}\right) \quad \text{ή} \quad z = \log\left(\frac{1-U}{U}\right)$$

στις σχέσεις (2.9) και (2.10), θα παράγουμε τυχαίους αριθμούς από την $L\left(0, \frac{\pi^2}{3}\right)$.

Κάνοντας το γραμμικό μετασχηματισμό

$$\alpha + \beta \log\left(\frac{U}{1-U}\right) \quad \text{ή} \quad \alpha + \beta \log\left(\frac{1-U}{U}\right)$$

παράγουμε τυχαίους αριθμούς από την λογιστική κατανομή με παραμέτρους a και $\beta > 0$.

2.6 Ομοιότητες Λογιστικής με Κανονική Κατανομή

Η ομοιότητα της λογιστικής με την κανονική αποτέλεσε ένα μεγάλο πλεονέκτημα για τους ερευνητές. Σε ορισμένες περιπτώσεις είναι πιο εύκολο να χρησιμοποιηθεί η λογιστική κατανομή στη θέση της κανονικής. Η αντικατάσταση αυτή δε θα αποτελέσει πρόβλημα, απλουστεύοντας το έργο του αναλυτή χωρίς να υπάρχουν μεγάλες αποκλίσεις στα αποτελέσματα. Βέβαια αυτή η αντικατάσταση πρέπει να γίνεται με προσοχή λαμβάνοντας υπόψη τις ομοιότητες και τις διαφορές μεταξύ των δυο κατανομών.

Έστω μια συνεχής τυχαία μεταβλητή X η οποία ακολουθεί την τυποποιημένη κανονική κατανομή ($X \sim N(0,1)$). Η συνάρτηση κατανομής της θα δίνεται από τον τύπο

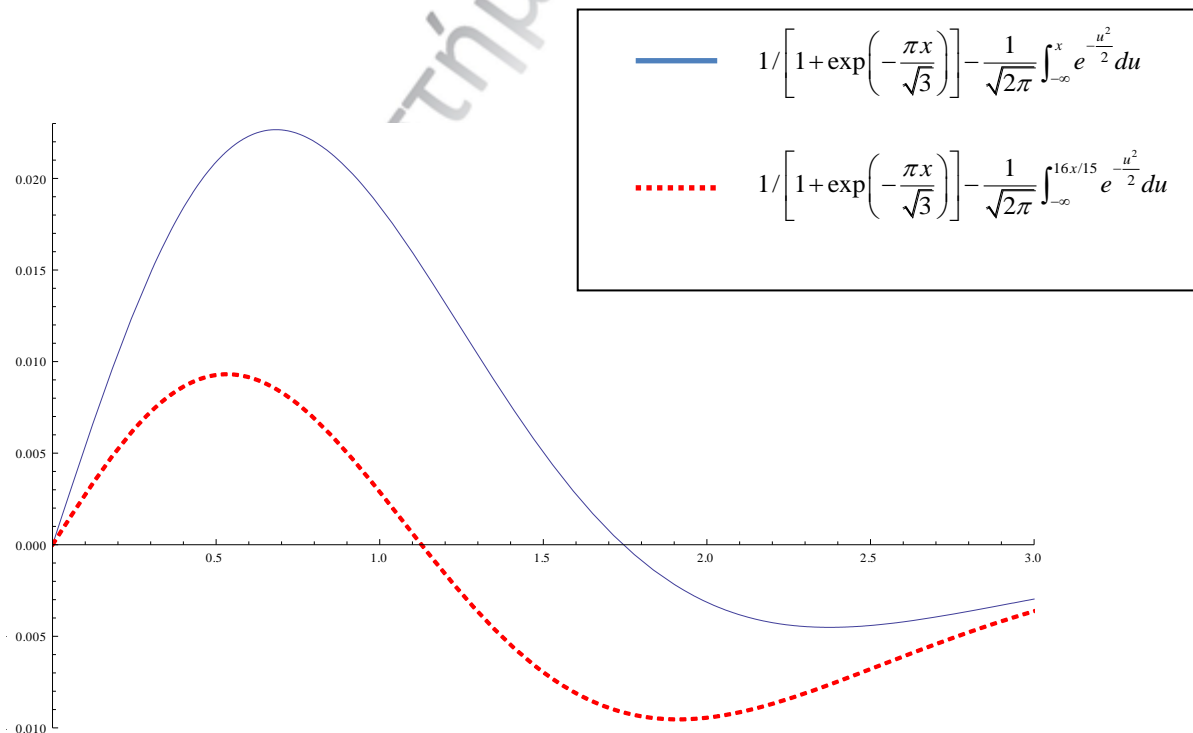
$$G_1(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du, \quad -\infty < x < \infty.$$

Αν η συνεχής τυχαία μεταβλητή X ακολουθούσε την τυποποιημένη λογιστική κατανομή ($X \sim L(0,1)$), τότε η συνάρτηση κατανομής της θα δινόταν από τον τύπο

$$G_2(x) = \frac{1}{1 + \exp\left(-\frac{\pi x}{\sqrt{3}}\right)}, \quad -\infty < x < \infty.$$

Στο παρακάτω γράφημα απεικονίζεται η διαφορά $G_2(x) - G_1(x)$ των δύο συναρτήσεων κατανομής

Σχήμα 2.7 Γράφημα της διαφοράς $G_2(x) - G_1(x)$ των συναρτήσεων κατανομής της λογιστικής και κανονικής κατανομής



Εφόσον και οι δύο κατανομές είναι συμμετρικές ως προς το 0 απεικονίζεται η διαφορά τους για $x \geq 0$. Από το σχήμα αυτό (μπλε ενιαία γραμμή) διακρίνουμε ότι η μέγιστη διαφορά είναι λίγο μεγαλύτερη παραπάνω από 0.02 (για την ακρίβεια είναι 0.0228 στο σημείο $x = 0.7$). Αυτή η μέγιστη διαφορά μπορεί να πάρει τιμή μικρότερη του 0.01 (κόκκινη διακεκομμένη γραμμή) αν πάρουμε τη συνάρτηση κατανομής της $N(0,1)$ στη θέση $\frac{16x}{15}$ ως προσέγγιση της τυποποιημένης λογιστικής δηλαδή αν θεωρήσουμε το $G_1\left(\frac{16x}{15}\right)$ αντί του $G_1(x)$ (Johnson και Kotz (1970)).

$$\left| \left\{ 1 + \exp\left(\frac{-\pi \cdot x}{\sqrt{3}}\right) \right\}^{-1} - \Phi\left(\frac{16x}{15}\right) \right| < 0.01 \quad -\infty < x < \infty.$$

2.7 Χαρακτηριστικά για τη Λογιστική Κατανομή

Για να κατανοήσουμε πλήρως μια συγκεκριμένη κατανομή αρκεί να αναγνωρίσουμε τις ιδιότητες που την χαρακτηρίζουν. Αρκετές φορές οι ιδιότητες μιας συγκεκριμένης κατανομής με κατάλληλους μονότονους μετασχηματισμούς μπορούν να εφαρμοσθούν και να χαρακτηρίζουν και μια άλλη κατανομή. Έτσι και στην περίπτωση μας θα δούμε πως ιδιότητες της εκθετικής κατανομής με κατάλληλους μετασχηματισμούς ισχύουν και για τη λογιστική κατανομή.

Θεωρούμε μια τυχαία μεταβλητή U με συνεχή συνάρτηση κατανομής $G(x)$. Τότε η

$$V = G(U) \quad \text{και} \quad W = -\frac{1}{a} \log V,$$

με $a > 0$ (σταθερά), κατανέμονται ομοιόμορφα στο $(0,1)$ και εκθετικά με παράμετρο a αντίστοιχα. Για την τυχαία μεταβλητή V είναι εύκολο να το διαπιστώσει κάποιος εφόσον για όλες τις συνεχείς κατανομές μετασχηματίζοντας την τυχαία μεταβλητή με εφαρμογή της συνάρτησης κατανομής καταλήγουμε στην ομοιόμορφη κατανομή στο $(0,1)$. Για την W αρκεί να παρατηρήσει κανείς ότι ο τύπος της είναι η αντίστροφη συνάρτηση της εκθετικής συνάρτησης αξιοπιστίας $R(x;a) = e^{-ax}$ με $x \geq 0$ και $a > 0$. Πιο αναλυτικά, η R είναι αντιστρέψιμη συνάρτηση στο $(0, \infty)$ και ειδικότερα,

$$R(x, a) = u \Leftrightarrow e^{-ax} = u \Leftrightarrow \log u = -ax \Leftrightarrow x = -\frac{1}{a} \log u$$

με $u \geq 0$ και $a > 0$.

Η πιο απλή ιδιότητα που χαρακτηρίζει την εκθετική κατανομή είναι η έλλειψη μνήμης (βλέπε Galambos και Kotz 1978 για περισσότερες πληροφορίες), δηλαδή

$$P(W > x + y | W > x) = P(W > y), \quad x, y \geq 0$$

ή ισοδύναμα

$$\begin{aligned} P\left(-\frac{1}{a} \log V > x + y \mid -\frac{1}{a} \log V > x\right) &= P\left(-\frac{1}{a} \log V > y\right) \Leftrightarrow \\ P(V \leq e^{-a(x+y)} \mid V \leq e^{-ax}) &= P(V \leq e^{-ay}) \end{aligned}$$

Θέτοντας $V = G(U)$ και λύνοντας ως προς U παίρνουμε

$$\begin{aligned} P(G(U) \leq e^{-a(x+y)} \mid G(U) \leq e^{-ax}) &= P(G(U) \leq e^{-ay}) \Leftrightarrow \\ P\left[U \leq G^{-1}(e^{-a(x+y)}) \mid U \leq G^{-1}(e^{-ax})\right] &= P\left[U \leq G^{-1}(e^{-ay})\right] \end{aligned} \quad (2.13)$$

όπου $G^{-1}(t)$ η αντίστροφη συνάρτηση της G .

Αν $G(z)$ είναι η συνάρτηση κατανομής της λογιστικής $L\left(0, \frac{\pi^2}{3}\right)$, η αντίστροφη συνάρτηση της G θα δίνεται από τον τύπο

$$G^{-1}(t) = \log \frac{t}{1-t}$$

και έτσι η (2.13) γίνεται

$$\begin{aligned} P\left[U \leq \log \frac{e^{-a(x+y)}}{1-e^{-a(x+y)}} \mid U \leq \log \frac{e^{-ax}}{1-e^{-ax}}\right] &= P\left[U \leq \log \frac{e^{-ay}}{1-e^{-ay}}\right] \Leftrightarrow \\ P\left[e^U \leq \frac{e^{-a(x+y)}}{1-e^{-a(x+y)}} \mid e^U \leq \frac{e^{-ax}}{1-e^{-ax}}\right] &= P\left[e^U \leq \frac{e^{-ay}}{1-e^{-ay}}\right] \Leftrightarrow \\ P\left[e^U \leq \frac{1}{1-e^{a(x+y)}} \mid e^U \leq \frac{1}{1-e^{ax}}\right] &= P\left[e^U \leq \frac{1}{1-e^{ay}}\right] \Leftrightarrow \\ P\left[e^{-U} + 1 \geq e^{a(x+y)} \mid e^{-U} + 1 \geq e^{ax}\right] &= P\left[e^{-U} + 1 \geq e^{ay}\right] \end{aligned}$$

Θέτοντας $T = 1 + e^{-U}$ και $u = e^x$, $v = e^y$ η παραπάνω ισότητα παίρνει τη μορφή

$$P(T \geq (uv)^a \mid T \geq u^a) = P(T \geq v^a) \quad (2.14)$$

Έτσι από την ιδιότητα της έλλειψης μνήμης προκύπτει ο επόμενος χαρακτηρισμός για τη λογιστική κατανομή

Θεώρημα. Έστω U μια τυχαία μεταβλητή και $T = 1 + e^{-U}$. Τότε η συνάρτηση κατανομής της τυχαίας μεταβλητής U είναι η λογιστική κατανομή με τύπο

$$F(z) = \frac{1}{1 + \exp(-z)} \text{ αν και μόνο αν για όλα τα } u \geq 1, v \geq 1 \text{ ισχύει η σχέση (2.14).}$$

Θεωρούμε πάλι την τυχαία μεταβλητή U και $T = 1 + e^{-U}$. Ο μέσος υπολειπόμενος χρόνος ζωής του λογαρίθμου της T δίνεται από τον τύπο

$$R(x) = E[\log T - x \mid \log T > x] \quad (2.15)$$

Αν $G(x)$ είναι η συνεχής συνάρτηση κατανομής της U τότε θα έχουμε

$$P(\log T > x) = P(T > e^x) = P(1 + e^{-U} > e^x) = P(U \leq -\log(e^x - 1)) = G(-\log(e^x - 1)). \quad (2.16)$$

Η (2.15) σύμφωνα με τη (2.16) γίνεται

$$R(x) = \frac{\int_x^\infty P(\log T > x) dt}{P(\log T > x)} = \frac{\int_x^\infty G(-\log(e^t - 1)) dt}{G(-\log(e^x - 1))}.$$

Κάνοντας αλλαγή μεταβλητής $y = -\log(e^t - 1)$ έχουμε

$$\int_{-\infty}^z \frac{G(y)}{1 + e^y} dy = G(z) \cdot R[\ln(1 + e^z)],$$

με $z = -\log(e^x - 1)$ όπου $z \in \mathbb{R}$. Στην περίπτωση που η συνάρτηση κατανομής

$G(x)$ είναι η λογιστική $F(x) = \frac{1}{1 + e^{-x}}$ δείχνεται ότι

$$R(x) = 1$$

και αντίστροφα.

Έτσι η τελευταία σχέση δίνει

$$\int_{-\infty}^z \frac{G(y)}{1+e^y} dy = G(z) \quad (2.17)$$

Οπότε ένας πληθυσμός με συνεχή συνάρτηση κατανομής G έχει τη λογιστική κατανομή αν και μόνο αν ο μέσος υπολειπόμενος χρόνος του $\log T$ είναι μονάδα, ή ισοδύναμα αν η (2.17) αληθεύει.

2.8 Μέτρα θέσης και διασποράς

Για χάρη μαθηματικής ευκολίας θεωρούμε τη τυχαία μεταβλητή $Z \sim L\left(0, \frac{\pi^2}{3}\right)$. Η συνάρτηση κατανομής της, καθώς και η συνάρτηση πυκνότητας δίνονται από τους τύπους (2.5) και (2.6) αντίστοιχα. Η ροπογεννήτρια συνάρτηση της τυχαίας μεταβλητής Z δίνεται από τον τύπο

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} f_Z(z) dz = \int_{-\infty}^{\infty} e^{tz} \frac{e^{-z}}{(1+e^{-z})^2} dz = \int_{-\infty}^{\infty} \frac{e^{(t-1)z}}{(1+e^{-z})^2} dz \quad (2.18)$$

Κάνουμε αλλαγή μεταβλητής θέτοντας $u = \frac{1}{1+e^{-z}}$ και η (2.18) γίνεται

$$M_Z(t) = \int_0^1 \left(\frac{u}{1-u}\right)^t du = \int_0^1 u^t (1-u)^{-t} du \quad (2.19)$$

Λαμβάνοντας υπόψη ότι η συνάρτηση Βήτα δίνεται από τον τύπο

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad \forall x, y \in \mathbf{R}^+ - \{0\}$$

η (2.19) μπορεί να γραφεί στη μορφή

$$M_Z(t) = B(1+t, 1-t). \quad (2.20)$$

Χρησιμοποιώντας τον τύπο που συνδέει τη συνάρτηση Βήτα με τη Γάμμα προκύπτει ότι

$$M_Z(t) = B(1+t, 1-t) = \frac{\Gamma(1+t)\Gamma(1-t)}{\Gamma(1+1)} = \Gamma(1+t)\Gamma(1-t). \quad (2.21)$$

Παίρνοντας πρώτη και δεύτερη παράγωγο της ροπογεννήτριας συνάρτησης και θέτοντας $t=0$, προκύπτουν η πρώτη και η δεύτερη ροπή αντίστοιχα της λογιστικής συνάρτησης $L\left(0, \frac{\pi^2}{3}\right)$. Πράγματι έχουμε

$$\frac{dM_Z(t)}{dt} = \frac{dE(e^{tZ})}{dt} = E\left(\frac{de^{tZ}}{dt}\right) = E(Ze^{tZ}) \quad (2.22)$$

και θέτοντας $t=0$ στην (2.22) προκύπτει η $E(Z)$. Όμοια παραγωγίζοντας δυο φορές έχουμε

$$\frac{d^2M_Z(t)}{dt^2} = \frac{d^2E(e^{tZ})}{dt^2} = E\left(\frac{d^2e^{tZ}}{dt^2}\right) = E\left(\frac{d(Ze^{tZ})}{dt}\right) = E(Z^2e^{tZ}) \quad (2.23)$$

και θέτοντας $t=0$ στην (2.23) προκύπτει η $E(Z^2)$. Γενικεύοντας τα παραπάνω καταλήγουμε ότι η ροπή k -τάξεως μιας τυχαίας μεταβλητής Z είναι ίση με την τιμή της k -οστής παραγώγου της $M_Z(t)$ στο σημείο $t=0$.

Σύμφωνα με τη σχέση (2.21) έχουμε

$$\mu_1 = E(Z) = \left[\Gamma'(1+t)\Gamma(1-t) - \Gamma(1+t)\Gamma'(1-t) \right]_{t=0} = \Gamma'(1) - \Gamma'(1) = 0$$

και

$$\begin{aligned} \mu_2 = E(Z^2) &= \left[\Gamma'(1+t)\Gamma(1-t) - \Gamma(1+t)\Gamma'(1-t) \right]' = \\ &= \left[(\Gamma''(1+t)\Gamma(1-t) - \Gamma'(1+t)\Gamma'(1-t)) - (\Gamma'(1+t)\Gamma'(1-t) - \Gamma(1+t)\Gamma''(1-t)) \right]_{t=0} = \\ &= 2\Gamma''(1)\Gamma(1) - 2\Gamma'(1)\Gamma'(1) = 2\left\{ \Gamma''(1) - [\Gamma'(1)]^2 \right\} \end{aligned}$$

όπου $\Gamma'(\cdot)$ και $\Gamma''(\cdot)$ η πρώτη και η δεύτερη παράγωγος της συνάρτησης Γάμμα. Ομοίως, ροπές μεγαλύτερης τάξεως για τη λογιστική συνάρτηση θα εκφράζονται μέσω μεγαλύτερης τάξεως παραγώγους της συνάρτησης Γάμμα. (Davis (1935), Abramowitz και Stegun (1965)). Η ροπογεννήτρια που δίνεται από την εξίσωση (2.21) μπορεί να γραφεί και στην εξής μορφή

$$M_Z(t) = \Gamma(1+t)\Gamma(1-t) = \pi t \csc \pi t \quad (2.22)$$

όπου \csc ή cosec είναι η συντέμνουσα $\left(\csc \theta = \frac{1}{\sin \theta}\right)$

Από την (2.22) έχουμε

$$E(e^{iZ}) = \pi t \csc \pi t$$

Έτσι,

$$E(e^{i^2 Z}) = \pi t i \csc i \pi t \quad (2.23)$$

όπου i η φανταστική μονάδα ($i^2 = -1$)

Ο τύπος της υπερβολικής συντέμνουσας είναι $\operatorname{cosech} x = i \csc ix$. Η (2.23) σύμφωνα με χρήση του τύπου της υπερβολικής συντέμνουσας μπορεί να εκφραστεί ως εξής

$$E(e^{i^2 Z}) = \pi t \operatorname{cosech} \pi t. \quad (2.24)$$

Το πρώτο μέρος της εξίσωσης (2.24) αποτελεί τη **χαρακτηριστική συνάρτηση**

$\varphi_Z(t) = E(e^{i^2 Z})$ της Z . Όταν η $Z \sim L\left(0, \frac{\pi^2}{3}\right)$, η χαρακτηριστική συνάρτηση της Z

δίνεται από τον παραπάνω τύπο.

Οι ροπές της λογιστικής κατανομής $L\left(0, \frac{\pi^2}{3}\right)$ μπορούν επίσης να υπολογιστούν αν ολοκληρώσουμε κατευθείαν ως προς τη συνάρτηση πυκνότητας της τυχαίας μεταβλητής Z . Οι ροπές μονής-τάξεως είναι όλες μηδέν εφόσον η κατανομή της τυχαίας μεταβλητής Z είναι συμμετρική γύρω από το μηδέν. Οι ροπές ζυγής-τάξεως θα δίνονται από τον τύπο

$$E(Z^{2r}) = \int_{-\infty}^{\infty} z^{2r} f_Z(z) dz = \int_{-\infty}^{\infty} z^{2r} \frac{e^{-z}}{(1+e^{-z})^2} dz.$$

Αν θέσουμε

$$g(z) = z^{2r} \frac{e^{-z}}{(1+e^{-z})^2}$$

θα έχουμε $\forall z \in \mathbf{R}$

$$g(-z) = (-z)^{2r} \frac{e^z}{(1+e^z)^2} = (-z)^{2r} \frac{e^z}{1+2e^z+e^{2z}}$$

και διαιρώντας αριθμητή και παρονομαστή με e^{2z} προκύπτει

$$g(-z) = (-z)^{2r} \frac{e^{-z}}{e^{-2z} + 2e^{-z} + 1} = z^{2r} \frac{e^{-z}}{(1+e^{-z})^2} = g(z).$$

Παρατηρούμε δηλαδή ότι η συνάρτηση g είναι άρτια, οπότε το ολοκλήρωμα μπορεί να γραφεί

$$E(Z^{2r}) = \int_{-\infty}^{\infty} z^{2r} \frac{e^{-z}}{(1+e^{-z})^2} dz = 2 \int_0^{\infty} z^{2r} \frac{e^{-z}}{(1+e^{-z})^2} dz.$$

Αναπτύσσοντας την παράσταση $(1+e^{-z})^{-2}$ σε σειρά βρίσκουμε

$$\begin{aligned} E(Z^{2r}) &= 2 \int_0^{\infty} z^{2r} \frac{e^{-z}}{(1+e^{-z})^2} dz = 2 \int_0^{\infty} z^{2r} \left\{ \sum_{j=1}^{\infty} (-1)^{j-1} j e^{-jz} \right\} dz \\ &= 2\Gamma(2r+1) \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j^{2r}} = 2\Gamma(2r+1) \left(1 - \frac{1}{2^{2r-1}} \right) \zeta(2r) \end{aligned}$$

όπου με $\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}$ συμβολίσαμε τη ζήτα του Riemann.

Η ροπή 2^{ης} τάξης για $r=1$ σύμφωνα με την τελευταία σχέση είναι ίση με

$$E(Z^2) = 2\zeta(2) = 2 \left(\frac{\pi^2}{6} \right) = \frac{\pi^2}{3}.$$

Το παραπάνω αποτέλεσμα ήταν αναμενόμενο αφού η κατανομή $L\left(0, \frac{\pi^2}{3}\right)$ έχει

διακύμανση $\pi^2/3$. Από τον ορισμό της διακύμανσης έχουμε

$$\text{Var}(Z) = E(Z^2) - E^2(Z) = \frac{\pi^2}{3} - 0 = \frac{\pi^2}{3}$$

Για $r=2$ η ροπή 4^{ης} τάξης είναι

$$E(Z^4) = 42\zeta(4) = 42\left(\frac{\pi^4}{90}\right) = \frac{7\pi^4}{15}$$

Η ροπή 3^{ης} τάξης είναι μηδενική οπότε και η ασυμμετρία της λογιστικής κατανομής $L\left(0, \frac{\pi^2}{3}\right)$ που εκφράζεται ως $\gamma_1 = \frac{\mu_3}{\sigma^3}$ θα είναι ίση με μηδέν.

Η κύρτωση της λογιστικής κατανομής $L\left(0, \frac{\pi^2}{3}\right)$ είναι ίση με

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{\frac{7\pi^4}{15}}{\left(\frac{\pi^2}{3}\right)^2} = \frac{7\pi^4}{\frac{\pi^4}{9}} = \frac{63}{15} = 4.2$$

Η τιμή του συντελεστή κύρτωσης της λογιστικής κατανομής είναι μεγαλύτερη από αυτή της κανονικής (η τιμή για την κανονική κατανομή είναι 3) και αυτό οφείλεται στις «μακρύτερες» ουρές της.

Η τυχαία μεταβλητή $Z \sim L\left(0, \frac{\pi^2}{3}\right)$ που θεωρήσαμε ισούται με $Z = \frac{\pi(X - \mu)}{\sigma\sqrt{3}}$.

Η τυχαία μεταβλητή $X = \mu + \sigma \frac{\sqrt{3}}{\pi} Z$ της οποίας η συνάρτηση κατανομής δίνεται από την εξίσωση (2.1) έχει μέση τιμή

$$E(X) = \mu + \sigma \frac{\sqrt{3}}{\pi} E(Z) = \mu$$

και διακύμανση

$$Var(X) = Var\left(\mu + \sigma \frac{\sqrt{3}}{\pi} Z\right) = \sigma^2 \frac{3}{\pi^2} Var(Z) = \sigma^2$$

2.9 Διάμεσος και Ποσοστιαία σημεία

Για όλες τις συμμετρικές κατανομές η διάμεσος συμπίπτει με το μέσο. Έτσι και για την τυχαία μεταβλητή $Z \sim L\left(0, \frac{\pi^2}{3}\right)$ η διάμεσος θα βρίσκεται στο σημείο 0. Ας το

δούμε και στην πράξη για την τυχαία μεταβλητή $X \sim L(\mu, \sigma^2)$. Έχουμε

$$F(x; \mu, \sigma) = \frac{1}{2} \Leftrightarrow 1 + \exp\left[\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right] = 2 \Leftrightarrow \frac{\pi(x - \mu)}{\sigma\sqrt{3}} = \ln 1 \Leftrightarrow x - \mu = 0 \Leftrightarrow x = \mu.$$

Δηλαδή όντως η διάμεσος συμπίπτει με το μέσο.

Το p -ποσοστημόριο υπολογίζεται από τον τύπο

$$P(X \leq x_p) = p \Leftrightarrow F(x_p) = p \Leftrightarrow \frac{\pi(x_p - \mu)}{\sigma\sqrt{3}} = \ln\left(\frac{p}{1-p}\right) \Leftrightarrow x_p = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{p}{1-p}\right)$$

Τα ποσοστημόρια $p = 0.25$ και $p = 0.75$ καλούνται πρώτο και τρίτο αντίστοιχα, ενώ το 0.5 καλείται διάμεσος. Για τις ποσότητες αυτές έχουμε

$$Q_1 = x_{0.25} = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{0.25}{0.75}\right) = \mu - \frac{\sigma\sqrt{3}}{\pi} \ln 3 = \mu - \left(\ln 3 \frac{\sqrt{3}}{\pi}\right) \sigma.$$

$$\delta = x_{0.5} = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{0.5}{0.5}\right) = \mu - \frac{\sigma\sqrt{3}}{\pi} \ln 1 = \mu.$$

$$Q_3 = x_{0.75} = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln\left(\frac{0.75}{0.25}\right) = \mu + \frac{\sigma\sqrt{3}}{\pi} \ln 3 = \mu + \left(\ln 3 \frac{\sqrt{3}}{\pi}\right) \sigma.$$

(ο δεύτερος τύπος συμφωνεί με αυτό που είχαμε βρεί λίγο παραπάνω)

Κεφάλαιο 3^ο

Στατιστική Συμπερασματολογία

Σ' αυτό το κεφάλαιο θα αναφερθούμε στη μέθοδο της μέγιστης πιθανοφάνειας, στη μέθοδο των ροπών καθώς και σε γραφικές μεθόδους για να εκτιμήσουμε τις παραμέτρους της λογιστικής κατανομής.

3.1 Εκτιμήτριες Μέγιστης Πιθανοφάνειας για πλήρη δεδομένα

Με τη μέθοδο αυτή θα εκτιμήσουμε τις παραμέτρους της λογιστικής κατανομής με τη χρήση πλήρων ή λογοκριμένων δεδομένων.

Έστω ότι διαθέτουμε ένα πλήρες δείγμα X_1, \dots, X_n από n ανεξάρτητες τυχαίες μεταβλητές οι οποίες ακολουθούν τη λογιστική κατανομή με άγνωστες παραμέτρους μ και σ ($X \sim L(\mu, \sigma^2)$). Η συνάρτηση πυκνότητας πιθανότητας της λογιστικής κατανομής με παραμέτρους μ, σ θα δίνεται από τον τύπο

$$f(x) = \frac{\pi}{\sigma\sqrt{3}} \frac{\exp\left(-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right)}{\left\{1 + \exp\left[-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right]\right\}^2}, \quad x \in \mathbf{R}$$

οπότε η συνάρτηση πιθανοφάνειας θα είναι ίση με

$$L(\mu, \sigma; x_1, \dots, x_n) = \left(\frac{\pi}{\sigma\sqrt{3}}\right)^n \prod_{i=1}^n \frac{\exp\left(-\frac{\pi(x_i-\mu)}{\sigma\sqrt{3}}\right)}{\left\{1 + \exp\left[-\frac{\pi(x_i-\mu)}{\sigma\sqrt{3}}\right]\right\}^2}$$

με αντίστοιχο πιθανολογάρημο

$$\begin{aligned}
\ell(\mu, \sigma; x_1, \dots, x_n) &= \ln \left(\left(\frac{\pi}{\sigma\sqrt{3}} \right)^n \prod_{i=1}^n \frac{\exp\left(-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)}{\left\{1 + \exp\left[-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right]\right\}^2} \right) = \\
&= n \ln \frac{\pi}{\sigma\sqrt{3}} + \ln \prod_{i=1}^n \exp\left(-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right) - 2 \ln \prod_{i=1}^n \left[1 + \exp\left(-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)\right] \\
&= n \ln \frac{\pi}{\sigma\sqrt{3}} - \frac{\pi}{\sigma\sqrt{3}} \left[\sum_{i=1}^n (x_i - \mu) \right] - 2 \left\{ \sum_{i=1}^n \ln \left[1 + \exp\left(-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)\right] \right\}
\end{aligned}$$

Για την εύρεση των εκτιμητριών μέγιστης πιθανοφάνειας των παραμέτρων μ , σ θα πρέπει να λύσουμε το σύστημα

$$\frac{\partial \ell(\mu, \sigma; x_1, \dots, x_n)}{\partial \mu} = 0, \quad \frac{\partial \ell(\mu, \sigma; x_1, \dots, x_n)}{\partial \sigma} = 0.$$

Η παράσταση που εμφανίζεται στην πρώτη εξίσωση του συστήματος είναι ίση με

$$\begin{aligned}
\frac{\partial \ell(\mu, \sigma; x_1, \dots, x_n)}{\partial \mu} &= n \frac{\pi}{\sigma\sqrt{3}} - 2 \frac{\pi}{\sigma\sqrt{3}} \sum_{i=1}^n \frac{\exp\left(-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)}{1 + \exp\left(-\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)} = \\
&= n \frac{\pi}{\sigma\sqrt{3}} - 2 \frac{\pi}{\sigma\sqrt{3}} \sum_{i=1}^n \frac{1}{1 + \exp\left(\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)} \quad (3.10)
\end{aligned}$$

και εξισώνοντας με το μηδέν προκύπτει η ακόλουθη εξίσωση

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \exp\left(\frac{\pi(x_i - \mu)}{\sigma\sqrt{3}}\right)} = \frac{1}{2} \quad (3.11)$$

Η παράσταση που εμφανίζεται στη δεύτερη εξίσωση του συστήματος είναι ίση με

$$\begin{aligned}
\frac{\partial l(\mu, \sigma; x_1, \dots, x_n)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{\pi}{\sigma^2 \sqrt{3}} \sum_{i=1}^n (x_i - \mu) - 2 \frac{\pi}{\sigma^2 \sqrt{3}} \sum_{i=1}^n \frac{(x_i - \mu) \exp \left[-\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]}{1 + \exp \left[-\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]} \\
&= -\frac{n}{\sigma} + \frac{\pi}{\sigma^2 \sqrt{3}} \sum_{i=1}^n (x_i - \mu) \left(1 - 2 \frac{\exp \left[-\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]}{1 + \exp \left[-\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]} \right) \\
&= -\frac{n}{\sigma} + \frac{\pi}{\sigma^2 \sqrt{3}} \sum_{i=1}^n (x_i - \mu) \left(1 - \frac{2}{1 + \exp \left[\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]} \right) \\
&= -\frac{n}{\sigma} - \frac{\pi}{\sigma^2 \sqrt{3}} \sum_{i=1}^n (x_i - \mu) \left(\frac{1 - \exp \left[\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]}{1 + \exp \left[\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]} \right) \tag{3.12}
\end{aligned}$$

και εξισώνοντας με το μηδέν προκύπτει η ακόλουθη εξίσωση

$$-\frac{\sqrt{3}}{\pi} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) \left(\frac{1 - \exp \left[\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]}{1 + \exp \left[\frac{\pi(x_i - \mu)}{\sigma \sqrt{3}} \right]} \right) \tag{3.13}$$

Παρατηρούμε ότι η λύση του συστήματος των εξισώσεων (3.11), (3.13) δεν οδηγεί σε τύπους κλειστής μορφής για την εκτίμηση των παραμέτρων μ , σ . Μπορεί όμως να λυθεί αριθμητικά με κάποια επαναληπτική μέθοδο (π.χ μέθοδος Newton-Raphson). Ως αρχικές τιμές των παραμέτρων μπορούμε να χρησιμοποιήσουμε τις αντίστοιχες εκτιμήτριες με τη μέθοδο των ροπών

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{και} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

3.2 Διαστήματα εμπιστοσύνης για τις εκτιμήτριες μέγιστης πιθανοφάνειας

Ο ασυμπτωτικός πίνακας διακύμανσης-συνδιακύμανσης για τις παραμέτρους μ και σ , σύμφωνα με τον Nelson (1982) υπολογίζεται από τον αντίστροφο του πίνακα πληροφορίας του Fisher

$$I_{ij} = E \left[-\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right], \quad i, j = 1, \dots, n$$

με $\theta_1, \theta_2 = \mu$ ή σ . Δηλαδή,

$$\begin{bmatrix} \text{Var}(\hat{\mu}) & \text{Cov}(\hat{\mu}, \hat{\sigma}) \\ \text{Cov}(\hat{\mu}, \hat{\sigma}) & \text{Var}(\hat{\sigma}) \end{bmatrix} = \begin{bmatrix} E \left(-\frac{\partial^2 \ell}{\partial \mu^2} \Big|_{\hat{\mu}, \hat{\sigma}} \right) & E \left(-\frac{\partial^2 \ell}{\partial \mu \partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} \right) \\ E \left(-\frac{\partial^2 \ell}{\partial \mu \partial \sigma} \Big|_{\hat{\mu}, \hat{\sigma}} \right) & E \left(-\frac{\partial^2 \ell}{\partial \sigma^2} \Big|_{\hat{\mu}, \hat{\sigma}} \right) \end{bmatrix}^{-1}$$

Τα προσεγγιστικά διαστήματα εμπιστοσύνης, $[\mu_L, \mu_U]$ και $[\sigma_L, \sigma_U]$ για τις παραμέτρους μ και σ με συντελεστή εμπιστοσύνης $1 - \gamma$ είναι τα

$$[\mu_L, \mu_U] = \hat{\mu} \pm z_{\gamma/2} \sqrt{V(\hat{\mu})} \quad \text{και} \quad [\sigma_L, \sigma_U] = \hat{\sigma} \pm z_{\gamma/2} \sqrt{V(\hat{\sigma})}$$

αντίστοιχα, με $z_{\gamma/2}$ να συμβολίζει το $\gamma/2$ -σημείο της τυποποιημένης κανονικής κατανομής ($P(Z > z_\gamma) = \gamma \Leftrightarrow Z_\gamma = \Phi^{-1}(1 - \gamma)$).

Ας υπολογίσουμε τις μερικές παραγώγους δεύτερης τάξης της πιθανοφάνειας που δίνεται από τον πιθανολογάρημο $\ell(\mu, \sigma; x_1, \dots, x_n)$. Για να διευκολυνθούμε στους υπολογισμούς θα θεωρήσουμε τις τυχαίες μεταβλητές $Z_i = \frac{\pi(X_i - \mu)}{\sigma\sqrt{3}}$ οι οποίες θα

ακολουθούν την κατανομή $L\left(0, \frac{\pi^2}{3}\right)$. Η (3.10) διαμορφώνεται ως εξής

$$\frac{\partial \ell(\mu, \sigma; z_1, \dots, z_n)}{\partial \mu} = n \frac{\pi}{\sigma \sqrt{3}} - 2 \frac{\pi}{\sigma \sqrt{3}} \sum_{i=1}^n \frac{f(z_i)}{F(z_i)} \quad (3.14)$$

όπου $f(z)$, $F(z)$ είναι οι συνάρτηση πυκνότητας πιθανότητας και αθροιστική συνάρτηση κατανομής της $L\left(0, \frac{\pi^2}{3}\right)$ αντίστοιχα, δηλαδή

$$f(z) = \frac{e^{-z}}{(1+e^{-z})^2} \quad \text{και} \quad F(z) = \frac{1}{1+e^{-z}}.$$

Είναι φανερό ότι

$$\frac{f(z)}{F(z)} = \frac{1}{1+e^z}.$$

Παραγωγίζοντας την (3.14) ως προς μ βρίσκουμε

$$\begin{aligned} \frac{\partial^2 \ell(\mu, \sigma; z_1, \dots, z_n)}{\partial \mu^2} &= -2 \frac{\pi}{\sigma \sqrt{3}} \sum_{i=1}^n \left\{ \frac{f'(z_i) F(z_i) \left(-\frac{\pi}{\sigma \sqrt{3}}\right) - f(z_i) f'(z_i) \left(-\frac{\pi}{\sigma \sqrt{3}}\right)}{F^2(z_i)} \right\} \\ &= \frac{\pi^2}{3\sigma^2} \left\{ 2 \sum_{i=1}^n \frac{f'(z_i)}{F(z_i)} - 2 \sum_{i=1}^n \left[\frac{f(z_i)}{F(z_i)} \right]^2 \right\} \end{aligned} \quad (3.15)$$

ενώ παραγωγίζοντας ως προς σ προκύπτει

$$\begin{aligned} \frac{\partial^2 \ell(\mu, \sigma; z_1, \dots, z_n)}{\partial \mu \partial \sigma} &= \frac{\partial \left(n \frac{\pi}{\sigma \sqrt{3}} - 2 \frac{\pi}{\sigma \sqrt{3}} \sum_{i=1}^n \frac{f(z_i)}{F(z_i)} \right)}{\partial \sigma} = -\frac{n\pi}{\sigma^2 \sqrt{3}} + \frac{2\pi}{\sigma^2 \sqrt{3}} \sum_{i=1}^n \frac{f(z_i)}{F(z_i)} \\ &\quad - \frac{2\pi}{\sigma \sqrt{3}} \left\{ \sum_{i=1}^n \frac{f'(z_i)}{F(z_i)} \left(-\frac{(x_i - \mu)\pi}{\sigma^2 \sqrt{3}} \right) - \sum_{i=1}^n \left[\frac{f(z_i)}{F(z_i)} \right]^2 \left(-\frac{(x_i - \mu)\pi}{\sigma^2 \sqrt{3}} \right) \right\} \\ &= \frac{\pi}{\sigma^2 \sqrt{3}} \left\{ -n + 2 \sum_{i=1}^n \frac{f(z_i)}{F(z_i)} + 2 \sum_{i=1}^n z_i \frac{f'(z_i)}{F(z_i)} - 2 \sum_{i=1}^n z_i \left[\frac{f(z_i)}{F(z_i)} \right]^2 \right\}. \end{aligned}$$

Η (3.12) σύμφωνα με το γραμμικό μετασχηματισμό $z_i = \frac{\pi(x_i - \mu)}{\sigma \sqrt{3}}$ γίνεται

$$\frac{\partial \ell(\mu, \sigma; z_1, \dots, z_n)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n z_i - 2 \frac{1}{\sigma} \sum_{i=1}^n z_i \frac{f(z_i)}{F(z_i)}$$

και παραγωγίζοντας ως προς σ προκύπτει

$$\begin{aligned} \frac{\partial^2 \ell(\mu, \sigma; z_1, \dots, z_n)}{\partial \sigma^2} &= \frac{n}{\sigma^2} + \sum_{i=1}^n \left(-\frac{2\pi(x_i - \mu)}{\sigma^3 \sqrt{3}} \right) - 2 \sum_{i=1}^n \left(-\frac{2\pi(x_i - \mu)}{\sigma^3 \sqrt{3}} \frac{f(z_i)}{F(z_i)} \right) \\ &\quad - 2 \sum_{i=1}^n \left(\frac{z_i}{\sigma} \right) \left\{ \frac{f'(z_i) F(z_i) \left(-\frac{\pi(x_i - \mu)}{\sigma^2 \sqrt{3}} \right) - f(z_i) f'(z_i) \left(-\frac{\pi(x_i - \mu)}{\sigma^2 \sqrt{3}} \right)}{F^2(z_i)} \right\} \\ &= \frac{n}{\sigma^2} - 2 \frac{1}{\sigma^2} \sum_{i=1}^n z_i + 4 \frac{1}{\sigma^2} \sum_{i=1}^n \left(z_i \frac{f(z_i)}{F(z_i)} \right) - 2 \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \frac{f'(z_i) F(z_i) (-z_i^2) - f(z_i) f'(z_i) (-z_i^2)}{F^2(z_i)} \right\} \\ &= \frac{n}{\sigma^2} - 2 \frac{1}{\sigma^2} \sum_{i=1}^n z_i + 4 \frac{1}{\sigma^2} \sum_{i=1}^n \left(z_i \frac{f(z_i)}{F(z_i)} \right) + 2 \frac{1}{\sigma^2} \sum_{i=1}^n \left[\frac{z_i^2 f'(z_i)}{F(z_i)} \right] - 2 \frac{1}{\sigma^2} \sum_{i=1}^n \left(z_i \frac{f(z_i)}{F(z_i)} \right)^2 \\ &= \frac{1}{\sigma^2} \left\{ n - 2 \sum_{i=1}^n z_i + 4 \sum_{i=1}^n \left(z_i \frac{f(z_i)}{F(z_i)} \right) + 2 \sum_{i=1}^n \left[\frac{z_i^2 f'(z_i)}{F(z_i)} \right] - 2 \sum_{i=1}^n \left(z_i \frac{f(z_i)}{F(z_i)} \right)^2 \right\}, \end{aligned}$$

όπου

$$f'(z) = \frac{df(z)}{dz}.$$

3.3 Εκτιμήτριες μέγιστης πιθανοφάνειας για λογοκριμένα δεδομένα

Έστω ότι ο χρόνος ζωής των υπό παρακολούθηση μονάδων ενός πληθυσμού περιγράφεται από μια τυχαία μεταβλητή X και έστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα χρόνων ζωής μεγέθους n από τον πληθυσμό. Στην περίπτωση που οι παρατηρούμενοι χρόνοι ζωής x_1, x_2, \dots, x_n είναι γνωστοί, μπορούμε να χρησιμοποιήσουμε γνωστές (παραμετρικές και μη παραμετρικές) στατιστικές μεθόδους για τη μελέτη του πληθυσμού. Υπάρχουν περιπτώσεις στις οποίες για διάφορους λόγους, οι χρόνοι ζωής ορισμένων παρατηρήσεων του δείγματος είναι λογοκριμένοι (*censored*). Μια λογοκριμένη παρατήρηση παρέχει μόνο μερική πληροφόρηση για το χρόνο ζωής της

αντίστοιχης μονάδας και μπορεί να είναι είτε λογοκριμένη από δεξιά (*right censoring*) όταν γνωρίζουμε ότι ο χρόνος ζωής του ατόμου είναι μεγαλύτερος από κάποιο χρόνο U , είτε λογοκριμένος από αριστερά (*left censoring*) όταν γνωρίζουμε ότι ο χρόνος ζωής του ατόμου είναι μικρότερος από κάποιο χρόνο U , είτε λογοκριμένος σε διάστημα (*interval censoring*) όταν γνωρίζουμε ότι ο χρόνος ζωής του ατόμου βρίσκεται σε διάστημα της μορφής (L, R) με $L < R$. Σε όλες τις περιπτώσεις λογοκρισίας για τη συνάρτηση πιθανοφάνειας ισχύει ότι:

$$L(\theta) \propto \left(\prod_{i \in D} f(x_i | \theta) \right) \left(\prod_{i \in C_l} F(x_i | \theta) \right) \left(\prod_{i \in C_r} [1 - F(x_i | \theta)] \right) \left(\prod_{i \in I} [F(R_i | \theta) - F(L_i | \theta)] \right) \quad (3.16)$$

όπου τα σύνολα D , C_r , C_l , και I αποτελούν διαμέριση του συνόλου των δεικτών $\{1, 2, \dots, n\}$ και ορίζονται ως εξής:

- το σύνολο D περιέχει τους δείκτες που δηλώνουν τα πλήρη δεδομένα
- το σύνολο C_r περιέχει τους δείκτες που δηλώνουν τα λογοκριμένα δεδομένα από δεξιά
- το σύνολο C_l περιέχει τους δείκτες που δηλώνουν τα λογοκριμένα δεδομένα από αριστερά και,
- το σύνολο I περιέχει τους δείκτες που δηλώνουν τα λογοκριμένα δεδομένα σε διάστημα.

Στην παρούσα εργασία θα ασχοληθούμε κυρίως με διπλά λογοκριμένους χρόνους τύπου II (*doubly type II censored*).

Στην περίπτωση της δεξιάς λογοκρισίας τύπου II, αποφασίζεται από την αρχή της έρευνας ότι αυτή θα τερματιστεί όταν αποτύχουν συνολικά r μονάδες. Επομένως, τα δεδομένα αποτελούνται από τους πλήρεις χρόνους των πρώτων r μονάδων που απέτυχαν, ενώ για τις υπόλοιπες $n-r$ μονάδες γνωρίζουμε ότι ο χρόνος ζωής τους είναι μεγαλύτερος από τον μέγιστο χρόνο ζωής των r μονάδων. Ένας άλλος τύπος αο πραγματικός χρόνος επιβίωσης είναι μικρότερος από αυτόν που παρατηρήθηκε. Υπάρχουν ειδικές περιπτώσεις που η έρευνα έχει προκαθορισμένο χρόνο λήξης, και τότε μπορεί να έχουμε ταυτόχρονα λογοκριμένα δεδομένα και από δεξιά και από αριστερά, οπότε μιλάμε για διπλή λογοκρισία (*double censored*).

Θεωρούμε ένα διπλά λογοκριμένο δείγμα τύπου II. Στην περίπτωση της διπλής λογοκρισίας τύπου II οι παρατηρήσεις είναι της μορφής $(T_{(i)}, \Delta_i)$ με $i=1, 2, \dots, n$ όπου

$$T_{(i)} = \begin{cases} X_{(i)}, & r+1 \leq i \leq n-s \\ X_{(c)}, & i \leq r \text{ κ και } \geq n-s+1 \end{cases}, \quad \Delta_i = \begin{cases} 1, & r+1 \leq i \leq n-s \\ 0, & i \leq r \text{ κ και } \geq n-s+1 \end{cases}$$

με r οι μικρότερες και s οι μεγαλύτερες λογοκριμένες παρατηρήσεις σε ένα διατεταγμένο δείγμα.

Θεωρούμε επίσης το διπλά λογοκριμένο τύπου II δείγμα

$$X_{r+1:n} \leq X_{r+2:n} \leq \dots \leq X_{n-s-1:n} \leq X_{n-s:n} \quad (3.17)$$

από τη λογιστική κατανομή $L(\mu, \sigma^2)$ με συνάρτηση κατανομής $F(x; \mu, \sigma)$ όπως δίνεται από την εξίσωση (2.1).

Ο γενικός τύπος της συνάρτησης πιθανοφάνειας (3.16) προσαρμόζεται σύμφωνα με τα παραπάνω, ως εξής:

$$L(\theta) \propto \left(\prod_{i \in D} f(x_i | \theta) \right) \left(\prod_{i \in C_\ell} F(x_i | \theta) \right) \left(\prod_{i \in C_r} [1 - F(x_i | \theta)] \right)$$

όπου

- $D = \{r+1, \dots, n-s\}$
- $C_\ell = \{1, \dots, r\}$
- $C_r = \{n-s+1, \dots, n\}$

Από τη θεωρία των διατεταγμένων δειγμάτων (*order statistics*) προκύπτει ότι για ένα τύπου II διπλά λογοκριμένο δείγμα η συνάρτηση πιθανοφάνειας θα δίνεται από τον τύπο

$$L(\mu, \sigma; x_{r+1}, \dots, x_{n-s}) = \frac{n!}{r!s!} \{F(x_{r+1:n}; \mu, \sigma)\}^r \{1 - F(x_{n-s:n}; \mu, \sigma)\}^s \prod_{i=r+1}^{n-s} f(x_{i:n}; \mu, \sigma)$$

Εκτελούμε το γραμμικό μετασχηματισμό $Z_i = \frac{\pi(X_i - \mu)}{\sigma\sqrt{3}}$ όπου $Z \sim L\left(0, \frac{\pi^2}{3}\right)$ οπότε

το δείγμα (3.16) γίνεται

$$Z_{r+1:n} \leq Z_{r+2:n} \leq \dots \leq Z_{n-s-1:n} \leq Z_{n-s:n}$$

με αντίστοιχη συνάρτηση πιθανοφάνειας

$$L(\mu, \sigma; z_{r+1}, \dots, z_{n-s}) = \frac{n!}{r!s!} \{F(z_{r+1:n}; \mu, \sigma)\}^r \{1 - F(z_{n-s:n}; \mu, \sigma)\}^s \prod_{i=r+1}^{n-s} \frac{\pi}{\sigma\sqrt{3}} f(z_{i:n}; \mu, \sigma)$$

διότι $f(x_{i:n}; \mu, \sigma) = \frac{\pi}{\sigma\sqrt{3}} f(z_{i:n}; \mu, \sigma)$.

Ο αντίστοιχος πιθανολογάρημος είναι ίσος με

$$\begin{aligned} \ell(\mu, \sigma; z_{r+1}, \dots, z_{n-s}) &= \ln n! - \ln r! - \ln s! + r \ln [F(z_{r+1:n}; \mu, \sigma)] + s \ln [1 - F(z_{n-s:n}; \mu, \sigma)] \\ &\quad + \ln \left\{ \prod_{i=r+1}^{n-s} \frac{\pi}{\sigma\sqrt{3}} f(z_{i:n}; \mu, \sigma) \right\} \\ &= \ln n! - \ln r! - \ln s! + r \ln [F(z_{r+1:n}; \mu, \sigma)] + s \ln [1 - F(z_{n-s:n}; \mu, \sigma)] \\ &\quad + \ln \left\{ \left(\frac{\pi}{\sigma\sqrt{3}} \right)^{n-s-r} \prod_{i=r+1}^{n-s} f(z_{i:n}; \mu, \sigma) \right\} \\ &= \ln n! - \ln r! - \ln s! + (n-r-s) \ln \left(\frac{\pi}{\sigma\sqrt{3}} \right) \\ &\quad + r \ln [F(z_{r+1:n}; \mu, \sigma)] + s \ln [1 - F(z_{n-s:n}; \mu, \sigma)] + \sum_{i=r+1}^{n-s} \ln f(z_{i:n}; \mu, \sigma) \end{aligned}$$

Χρησιμοποιώντας τη σχέση (2.5), ο πιθανολογάρημος τελικά παίρνει τη μορφή

$$\begin{aligned} \ell(\mu, \sigma; z_{r+1}, \dots, z_{n-s}) &= \ln n! - \ln r! - \ln s! + (n-r-s) \ln \left(\frac{\pi}{\sigma\sqrt{3}} \right) + r \ln [F(z_{r+1:n}; \mu, \sigma)] \\ &\quad + s \ln [1 - F(z_{n-s:n}; \mu, \sigma)] - \sum_{i=r+1}^{n-s} Z_{i:n} + 2 \sum_{i=r+1}^{n-s} \ln F(z_{i:n}; \mu, \sigma) \end{aligned}$$

Στην τελευταία σχέση λαμβάνοντας υπόψη ότι

$$\sum_i \ln f_Z(z_i) = \sum_i \ln [F_Z(z_i)(1-F_Z(z_i))] = \sum_i \ln \left[\frac{F_Z^2(z_i)}{F_Z(z_i)(1-F_Z(z_i))} \right] = \sum_i \ln \left[\frac{F_Z^2(z_i)}{e^{z_i}} \right].$$

οι πρώτες μερικοί παράγωγοι του πιθανολογάριθμου είναι ίσες με

$$\begin{aligned} \frac{\partial \ell(\mu, \sigma; z_{r+1}, \dots, z_{n-s})}{\partial \mu} &= r \frac{f(z_{r+1:n}; \mu, \sigma)}{F(z_{r+1:n}; \mu, \sigma)} \left(-\frac{\pi}{\sigma\sqrt{3}} \right) - s \frac{f(z_{n-s:n}; \mu, \sigma)}{1-F(z_{n-s:n}; \mu, \sigma)} \left(-\frac{\pi}{\sigma\sqrt{3}} \right) \\ &\quad - \sum_{i=r+1}^{n-s} \left(-\frac{\pi}{\sigma\sqrt{3}} \right) + 2 \sum_{i=r+1}^{n-s} \frac{f(z_{i:n}; \mu, \sigma)}{F(z_{i:n}; \mu, \sigma)} \left(-\frac{\pi}{\sigma\sqrt{3}} \right) \\ &= \frac{\pi}{\sigma\sqrt{3}} \left\{ -r \frac{f(z_{r+1:n}; \mu, \sigma)}{F(z_{r+1:n}; \mu, \sigma)} + s \frac{f(z_{n-s:n}; \mu, \sigma)}{1-F(z_{n-s:n}; \mu, \sigma)} + (n-s-r) - 2 \sum_{i=r+1}^{n-s} \frac{f(z_{i:n}; \mu, \sigma)}{F(z_{i:n}; \mu, \sigma)} \right\} \end{aligned}$$

και

$$\begin{aligned} \frac{\partial \ell(\mu, \sigma; z_{r+1}, \dots, z_{n-s})}{\partial \sigma} &= r \frac{f(z_{r+1:n}; \mu, \sigma)}{F(z_{r+1:n}; \mu, \sigma)} \left(-\frac{\pi(x_{r+1} - \mu)}{\sigma^2\sqrt{3}} \right) - s \frac{f(z_{n-s:n}; \mu, \sigma)}{1-F(z_{n-s:n}; \mu, \sigma)} \left(-\frac{\pi(x_{n-s} - \mu)}{\sigma^2\sqrt{3}} \right) \\ &\quad - \sum_{i=r+1}^{n-s} \left(-\frac{\pi(x_i - \mu)}{\sigma^2\sqrt{3}} \right) + 2 \sum_{i=r+1}^{n-s} \frac{f(z_{i:n}; \mu, \sigma)}{F(z_{i:n}; \mu, \sigma)} \left(-\frac{\pi(x_i - \mu)}{\sigma^2\sqrt{3}} \right) - \frac{(n-r-s)}{\sigma} \\ &= r \frac{f(z_{r+1:n}; \mu, \sigma)}{F(z_{r+1:n}; \mu, \sigma)} \left(-\frac{z_{r+1:n}}{\sigma} \right) - s \frac{f(z_{n-s:n}; \mu, \sigma)}{1-F(z_{n-s:n}; \mu, \sigma)} \left(-\frac{z_{n-s:n}}{\sigma} \right) - \sum_{i=r+1}^{n-s} \left(-\frac{z_{i:n}}{\sigma} \right) \\ &\quad + 2 \sum_{i=r+1}^{n-s} \frac{f(z_{i:n}; \mu, \sigma)}{F(z_{i:n}; \mu, \sigma)} \left(-\frac{z_{i:n}}{\sigma} \right) - \frac{(n-r-s)}{\sigma} \\ &= \frac{1}{\sigma} \left\{ -(n-r-s) - rz_{r+1:n} \frac{f(z_{r+1:n}; \mu, \sigma)}{F(z_{r+1:n}; \mu, \sigma)} - sz_{n-s:n} \frac{f(z_{n-s:n}; \mu, \sigma)}{1-F(z_{n-s:n}; \mu, \sigma)} + \sum_{i=r+1}^{n-s} z_{i:n} \right\} \\ &\quad - \frac{2}{\sigma} \sum_{i=r+1}^{n-s} z_{i:n} \frac{f(z_{i:n}; \mu, \sigma)}{F(z_{i:n}; \mu, \sigma)} \end{aligned}$$

Για $r=s=0$, δηλαδή για την περίπτωση όπου δεν έχουμε λογοκριμένα δεδομένα ούτε από αριστερά (μικρότερες παρατηρήσεις) ούτε από δεξιά (μεγαλύτερες

παρατηρήσεις), οι παραπάνω πρώτες μερικές παράγωγοι είναι όμοιες με εκείνες των πλήρων δεδομένων. Όπως και στην περίπτωση των πλήρων δεδομένων, η λύση του συστήματος μπορεί να λυθεί αριθμητικά με κάποια επαναληπτική μέθοδο. (Για περισσότερες πληροφορίες ο αναγνώστης παραπέμπεται στο άρθρο των Harter και Moore (1967))

3.4 Εκτιμήτριες ροπών για τις παραμέτρους α και β

Έστω ότι έχουμε ένα μοντέλο που ακολουθεί τη λογιστική κατανομή με παράμετρο θέσης α και παράμετρο κλίμακας β . Οι εκτιμήσεις των παραμέτρων αυτών μπορούν να υπολογισθούν χρησιμοποιώντας τη μέση τιμή και τη δειγματική διασπορά.

Η μέση τιμή ενός δείγματος $\{x_1, \dots, x_n\}$ δίνεται από τον τύπο

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n},$$

ενώ η δειγματική διασπορά

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Γνωρίζοντας τις ροπές πρώτης και δεύτερης τάξης της λογιστικής κατανομής και εξισώνοντας, έχουμε το παρακάτω σύστημα

$$\left\{ \begin{array}{l} \bar{x} = a \\ s^2 = \frac{\pi^2 \beta^2}{3} + \alpha^2 - \bar{x}^2 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \bar{x} = a \\ s^2 = \frac{\pi^2 \beta^2}{3} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \bar{x} = a \\ \sqrt{\frac{3s^2}{\pi^2}} = \beta \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \bar{x} = a \\ \sqrt{3\left(\frac{s}{\pi}\right)^2} = \beta \end{array} \right\}$$

Τελικά έχουμε,

$$\hat{\alpha} = \bar{x} \text{ και } \hat{\beta} = \sqrt{3\left(\frac{s}{\pi}\right)^2}$$

3.5 Γραφικές μέθοδοι

Στην παράγραφο αυτή θα γίνει αναφορά σε δύο γραφικές τεχνικές που βοηθούν στο να ελεγχθούν κατά πόσο τα δεδομένα προέρχονται από τη λογιστική κατανομή χωρίς όμως να γίνεται κάποιος τυπικός έλεγχος υποθέσεων.

α. Γράφημα εμπειρικής συνάρτησης κατανομής

Ας θεωρήσουμε ένα τυχαίο δείγμα μεγέθους n που προέρχεται από τη λογιστική κατανομή με αθροιστική συνάρτηση κατανομής

$$F(x; \mu, \sigma) = \frac{1}{1 + \exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right]} \quad -\infty < x < \infty, \quad (3.18)$$

με μέσο μ και τυπική απόκλιση σ .

Η εμπειρική συνάρτηση κατανομής του δείγματος θα δίνεται από τον τύπο

$$F_n(x) = \frac{\#(X_j \leq x)}{n},$$

όπου $j=1, \dots, n$. Για τις διατεταγμένες παρατηρήσεις του δείγματος $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$, η εμπειρική συνάρτηση κατανομής γράφεται

$$F_n(x) = \frac{i}{n}, \quad (3.19)$$

με $i = \#(X_{j:n} \leq x)$. Για να δημιουργήσουμε το γράφημα της εμπειρικής συνάρτησης κατανομής κατασκευάζουμε το γράφημα των $y_i = \frac{i}{n}$ ως προς $x_i = X_{i:n}$ για $i=1, \dots, n$. Όσο πιο μεγάλο το δείγμα (n) τόσο καλύτερη «εικόνα» μας δίνει το γράφημα της εμπειρικής συνάρτησης κατανομής για τον εν λόγω πληθυσμό ή για την αθροιστική συνάρτηση κατανομής που τον χαρακτηρίζει. Το ελάττωμα αυτής της γραφικής μεθόδου είναι ότι αν προσαρμόσουμε την καμπύλη της συνάρτησης κατανομής με συγκεκριμένες παραμέτρους, παρόλο που έχουμε μια εικόνα για το αν προσαρμόζεται το μοντέλο ή όχι, δεν μπορούμε να μετρήσουμε τις αποστάσεις των δεδομένων από την καμπύλη. Γι' αυτό στην πράξη προτιμάται η ακόλουθη γραφική μέθοδος.

β. LPP γράφημα (*Logistic probability plot*)

Μετασχηματίζοντας την (3.18) προκύπτουν διαδοχικά τα εξής:

$$\begin{aligned} F(x; \mu, \sigma) = y &\Leftrightarrow \frac{1}{1 + \exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right]} = y \Leftrightarrow 1 + \exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right] = \frac{1}{y} \\ &\Leftrightarrow \exp\left[-\frac{\pi(x - \mu)}{\sigma\sqrt{3}}\right] = \frac{1 - y}{y} \Leftrightarrow \frac{\pi(x - \mu)}{\sigma\sqrt{3}} = \ln \frac{y}{1 - y} \\ &\Leftrightarrow \frac{x - \mu}{\sigma\sqrt{3}} = \frac{1}{\pi} \ln \frac{y}{1 - y} \end{aligned}$$

και τελικά έχουμε

$$\frac{1}{\pi} \ln \frac{y}{1 - y} = \frac{1}{\sigma\sqrt{3}} x - \frac{\mu}{\sigma\sqrt{3}}. \quad (3.20)$$

Θέτοντας $a = \frac{1}{\sigma\sqrt{3}}$ και $\beta = -\frac{\mu}{\sigma\sqrt{3}}$ η (3.20) γίνεται

$$\frac{1}{\pi} \ln \frac{F(x; \mu, \sigma)}{1 - F(x; \mu, \sigma)} = ax + \beta.$$

Ο χάρτης πιθανότητας για τη λογιστική κατανομή (*logistic probability plot*)

δημιουργείται αν κατασκευάσουμε το γράφημα των

$$x \text{ έναντι των } y = \frac{1}{\pi} \ln \frac{F(x; \mu, \sigma)}{1 - F(x; \mu, \sigma)}.$$

Σε περίπτωση που θέλουμε να εξετάσουμε αν τα δεδομένα μας, d_1, \dots, d_n ,

προέρχονται από τη λογιστική κατανομή τότε πράττουμε ως εξής:

1. Διατάσσουμε τα δεδομένα μας κατά αύξουσα τιμή, έτσι ώστε $d_{(1)} \leq d_{(2)} \leq \dots \leq d_{(n)}$
2. Υπολογίζουμε την εμπειρική συνάρτηση κατανομής με τύπο $\hat{F}(d_{(i)}) = \frac{i - 0.5}{n}$, με

$1 \leq i \leq n$. Ο τελευταίος τύπος της $F_n(x)$ διαφέρει από τον τύπο της σχέσης (3.19)

έτσι ώστε να αποφύγουμε τη δημιουργία λογαρίθμων του απείρου. Για περισσότερες πληροφορίες ο αναγνώστης παραπέμπεται στον Barnett (1975).

- Υπολογίζουμε τις τιμές των $y_i = \frac{1}{\pi} \ln \frac{\hat{F}(d_{(i)})}{1 - \hat{F}(d_{(i)})}$ για $1 \leq i \leq n$.
- Κατασκευάζουμε το γράφημα της x_i ως προς y_i για $1 \leq i \leq n$.

Αν τα σημεία του γραφήματος βρίσκονται γύρω από μια ευθεία γραμμή τότε δεν μπορούμε να απορρίψουμε ότι τα δεδομένα μας προέρχονται από τη λογιστική κατανομή. Αν τα δεδομένα μας δε βρίσκονται γύρω από μια ευθεία γραμμή τότε καταφεύγουμε σε άλλου είδους μοντέλα.

Θα συγκρίνουμε στη συνέχεια τις μεθόδους εκτίμησης που προαναφέρθηκαν: μέθοδος μέγιστης πιθανοφάνειας (EMΠ), μέθοδος των ροπών (MP) και γραφική μέθοδος (ΓΜ), για να εξετάσουμε ποιά είναι η καταλληλότερη στο να εκτιμήσει τις παραμέτρους της λογιστικής κατανομής, ανάλογα με το μέγεθος του δείγματος.

Για τη σύγκριση των μεθόδων θα προσομοιωθούν 1000 δείγματα των 10, 50 και 120 παρατηρήσεων που προέρχονται από τη λογιστική κατανομή με προκαθορισμένες τιμές - παραμέτρους.

Το κριτήριο που θα κρίνει την καταλληλότερη μέθοδο είναι το μέσο άθροισμα αποκλίσεων (*mean sum variation, MSV*)

$$MSV = \frac{\sum_{k=1}^{1000} \left\{ \frac{|\mu - \hat{\mu}_k|}{\mu} + \frac{|\sigma - \hat{\sigma}_k|}{\sigma} \right\}}{1000}.$$

Καταλληλότερη κρίνεται η μέθοδος με τη μικρότερη *MSV* τιμή. Παρακάτω δίνεται ο σχετικός πίνακας με τις προσομοιωμένες τιμές.

Πίνακας 3.1 Προσομοιωμένες τιμές εκτίμησης παραμέτρων λογιστικής κατανομής

μ	σ	n	EMΠ	$\hat{\mu}$	$\hat{\sigma}$	ΓΜ	$\hat{\mu}$	$\hat{\sigma}$	MP	$\hat{\mu}$	$\hat{\sigma}$	Καταλληλότερη
2	3	10	1,2533	2,0115	0,03178	0,6169	2,0139	2,1464	0,6198	2,0139	2,142	ΓΜ
		50	0,69226	1,9924	1,24077	0,3694	2,0055	2,2900	0,3639	2,0055	2,312	MP
		120	0,23914	1,9962	2,49936	0,3153	1,9995	2,3008	0,3086	1,9995	2,321	EMΠ
5	2	10	1,09983	5,0633	0,04573	0,3961	4,9994	1,4876	0,3984	4,9994	1,487	ΓΜ
		50	0,60047	4,9948	0,85634	0,2726	4,9996	1,5299	0,2674	4,9996	1,545	MP
		120	0,17946	4,9965	1,68337	0,2482	4,9930	1,5494	0,2417	4,9930	1,563	EMΠ
10	5	10	1,12731	0,7746	0,02309	0,4076	9,9329	3,7032	0,4118	9,9329	3,704	ΓΜ
		50	0,60623	9,9948	2,14335	0,2816	9,9975	3,8249	0,2768	9,9975	3,857	MP
		120	0,19396	9,9854	4,15419	0,2621	9,9843	3,8260	0,2551	9,9843	3,863	EMΠ

Από τον παραπάνω πίνακα συμπεραίνουμε ότι η καταλληλότερη μέθοδος εκτίμησης των παραμέτρων της λογιστικής κατανομής εξαρτάται από το μέγεθος του δείγματος. Για μικρά δείγματα οι εκτιμήσεις των παραμέτρων με τη μέθοδο των ροπών (MP) και η γραφική μέθοδος (ΓΜ) φαίνονται να είναι κατάλληλες, με αυτή της γραφικής να είναι ελάχιστα καλύτερη για πολύ μικρά δείγματα ($n=10$). Για μεγάλα δείγματα μεγέθους $n \geq 100$ η μέθοδος μέγιστης πιθανοφάνειας φαίνεται με αρκετή διαφορά να είναι η καταλληλότερη. Το πλεονέκτημα αυτό γίνεται μειονέκτημα όταν εφαρμόζεται σε πολύ μικρά δείγματα. Ο αλγόριθμος με τις επαναληπτικές μεθόδους (100) αδυνατεί να συγκλίνει και να δώσει λύση για το διπλό σύστημα δύο εξισώσεων με δυο αγνώστους.

Παρακάτω δίνεται μέρος των αποτελεσμάτων όπως προσομοιώθηκαν μέσω του Mathematica

```

Πραγματικές παράμετροι  $\mu=2$  και  $\sigma=3$ 
Με τη μέθοδο EMΠ ο μέσος για εκτιμώμενο  $\mu$  είναι 1.99355
Με τη μέθοδο EMΠ ο μέσος για εκτιμώμενο  $\sigma$  είναι 2.50746
Με τη γραφική μέθοδο ο μέσος για εκτιμώμενο  $\mu$  είναι 1.99527
Με τη γραφική μέθοδο ο μέσος για εκτιμώμενο  $\sigma$  είναι 2.30927
Με τη μέθοδο ροπών ο μέσος για εκτιμώμενο  $\mu$  είναι 1.99527
Με τη μέθοδο ροπών ο μέσος για εκτιμώμενο  $\sigma$  είναι 2.33033
Με τη μέθοδο EMΠ, MSV= 0.238622
Με τη γραφική μέθοδο, MSV= 0.316411
Με τη μέθοδο των ροπών, MSV= 0.309474
    
```



```

Πραγματικές παράμετροι  $\mu=10$  και  $\sigma=5$ 
Με τη μέθοδο ΕΜΠ ο μέσος για εκτιμώμενο  $\mu$  είναι 10.0062
Με τη μέθοδο ΕΜΠ ο μέσος για εκτιμώμενο  $\sigma$  είναι 4.17128
Με τη γραφική μέθοδο ο μέσος για εκτιμώμενο  $\mu$  είναι 10.0018
Με τη γραφική μέθοδο ο μέσος για εκτιμώμενο  $\sigma$  είναι 3.84246
Με τη μέθοδο ροπών ο μέσος για εκτιμώμενο  $\mu$  είναι 10.0018
Με τη μέθοδο ροπών ο μέσος για εκτιμώμενο  $\sigma$  είναι 3.87928
Με τη μέθοδο ΕΜΠ, MSV= 0.192653
Με τη γραφική μέθοδο, MSV= 0.261298
Με τη μέθοδο των ροπών, MSV= 0.254019

```

Οι εντολές που χρησιμοποιήθηκαν είναι οι εξής:

```

(*MSV*)
n = 120;  $\mu$  = 10;  $\sigma$  = 5; myEktimitesRopwn = {}; myEktimitesElaxistwnTetragwnwn = {};
myEktimitesPithanofaneias = {}; monotimesGiam = {}; monotimesGias = {};
myEktimitesPithanofaneiasArithmoi = {}; maxfores = 1000;
Do[
  (*n τυχαίοι αριθμοί από Λογιστική κατανομή
  με παραμέτρους  $\mu$  και  $\sigma$ *)
  t = Table[ $\mu$  + (( $\sigma$  * Sqrt[3]) / Pi) * Log[Random[] / (1 - Random[])], {n}];
  (* Εκτιμητές Μέγιστης Πιθανοφάνειας *)
  emp = FindRoot[{ Sum[1 / (1 + Exp[(Pi * (t[[i]] -  $\mu$ )) / (s * Sqrt[3])]),
    {i, 1, Length[t]}] == n/2, Sum[((1 - Exp[(Pi * (t[[i]] -  $\mu$ )) / (s * Sqrt[3])]) /
    (1 + Exp[(Pi * (t[[i]] -  $\mu$ )) / (s * Sqrt[3])])) * (t[[i]] -  $\mu$ ) / s,
    {i, 1, Length[t]}] == -(100 * Sqrt[3]) / Pi}, {{m, 1}, {s, 2}}];
  AppendTo[myEktimitesPithanofaneias, emp];

  (*Διαδικαστικά βήματα για να απομονώσουμε τις
  τιμές των αριθμών από τα αποτελέσματα*)
  AppendTo[monotimesGiam, myEktimitesPithanofaneias[[fores, 1]]];
  AppendTo[monotimesGias, myEktimitesPithanofaneias[[fores, 2]]];
  AppendTo[myEktimitesPithanofaneiasArithmoi, {monotimesGiam[[fores, 2]],
    monotimesGias[[fores, 2]]}];

  (*Εκτιμητές με μέθοδο Ελαχίστων Τετραγώνων*)
  (*Διατάσσουμε τα δεδομένα της λίστας*)
  orderList = Sort[t]; eskList = {};
  (*Υπολογίζουμε την εμπειρική συνάρτηση κατανομής για τα δεδομένα*)
  For[i = 1, i ≤ Length[orderList], i++, AppendTo[eskList, (Sqrt[3] / Pi) *
    Log[((i - 0.5) / Length[orderList]) / (1 - ((i - 0.5) / Length[orderList]))]];
  (*Υπολογισμός συντελεστή κλίσης της ευθείας ( $\sigma$ )*)
  mp2 = Sum[(eskList[[i]] - Mean[eskList]) * orderList[[i]], {i, 1, Length[t]}] /
    (Sum[(eskList[[i]] - Mean[eskList])^2, {i, 1, Length[orderList]}]);

```

```

(*Υπολογισμός σταθεράς της ευθείας (μ)*)
mp1 = Mean[orderList] - mp2 * Mean[eskList];
(*Τοποθετούνται σε μια λίστα οι δύο εκτιμητές*)
AppendTo[myEktimitesElaxistwnTetragwnwn, {mp1, mp2}];
(*Εκτιμητής για μ με μέθοδο ροπών*)
mr1 = Mean[t];
(*Εκτιμητής για σ με μέθοδο ροπών*)
mr2 = StandardDeviation[t];
(*Τοποθετούνται σε μια λίστα οι δύο εκτιμητές*)
AppendTo[myEktimitesRopwn, {mr1, mr2}]

, {fores, 1, maxfores}
]

Print["Πραγματικές παράμετροι μ=", μ, " και σ=", σ]
(*Υπολόγισμα μέσο για -μ- και μέσο για -σ- για κάθε μέθοδο *)
empNewlistGiam = {}; empNewlistGias = {}; gmNewlistGiam = {}; gmNewlistGias = {};
mpNewlistGiam = {}; mpNewlistGias = {};
Do[
  AppendTo[empNewlistGiam, myEktimitesPithanofaneiasArithmoi[[i, 1]]];
  , {i, 1, maxfores}
]
Do[
  AppendTo[empNewlistGias, myEktimitesPithanofaneiasArithmoi[[i, 2]]];
  , {i, 1, maxfores}
]
Print["Με τη μέθοδο ΕΜΠ ο μέσος για εκτιμώμενο μ είναι ",
  Mean[empNewlistGiam]]
Print["Με τη μέθοδο ΕΜΠ ο μέσος για εκτιμώμενο σ είναι ",
  Mean[empNewlistGias]]
Do[
  AppendTo[gmNewlistGiam, myEktimitesElaxistwnTetragwnwn[[i, 1]]];
  , {i, 1, maxfores}
]
Do[
  AppendTo[gmNewlistGias, myEktimitesElaxistwnTetragwnwn[[i, 2]]];
  , {i, 1, maxfores}
]
Print["Με τη γραφική μέθοδο ο μέσος για εκτιμώμενο μ είναι ",
  Mean[gmNewlistGiam]]

```

```

Print["Με τη γραφική μέθοδο ο μέσος για εκτιμώμενο σ είναι ",
  Mean[gmNewlistGias]]
Do[
  AppendTo[mpNewlistGiam, myEktimitesRopwn[[i, 1]]];
  , {i, 1, maxfores}
]
Do[
  AppendTo[mpNewlistGias, myEktimitesRopwn[[i, 2]]];
  , {i, 1, maxfores}
]
Print["Με τη μέθοδο ροπών ο μέσος για εκτιμώμενο μ είναι ",
  Mean[mpNewlistGiam]]
Print["Με τη μέθοδο ροπών ο μέσος για εκτιμώμενο σ είναι ",
  Mean[mpNewlistGias]]

Print["Με τη μέθοδο ΕΜΠ, MSV= ",
  Sum[(Abs[μ - myEktimitesPithanofaneiasArithmoi[[i, 1]]] / μ)
    + (Abs[σ - myEktimitesPithanofaneiasArithmoi[[i, 2]]] / σ),
  {i, 1, maxfores}] / maxfores]
Print["Με τη γραφική μέθοδο, MSV= ",
  Sum[(Abs[μ - myEktimitesElaxistwnTetragwnwn[[i, 1]]] / μ)
    + (Abs[σ - myEktimitesElaxistwnTetragwnwn[[i, 2]]] / σ),
  {i, 1, maxfores}] / maxfores]
Print["Με τη μέθοδο των ροπών, MSV= ",
  Sum[(Abs[μ - myEktimitesRopwn[[i, 1]]] / μ)
    + (Abs[σ - myEktimitesRopwn[[i, 2]]] / σ), {i, 1, maxfores}] / maxfores]

```

Κεφάλαιο 4^ο

Γενικεύσεις/Τροποποιήσεις της Λογιστικής Κατανομής

4.1 Εισαγωγή

Στο παρόν κεφάλαιο θα εισάγουμε μια οικογένεια στατιστικών μοντέλων που αποτελούν φυσική γενίκευση του κλασικού μοντέλου της λογιστικής κατανομής. Λόγω της προσαρμοστικότητάς τους σε πραγματικά δεδομένα, τα τελευταία χρόνια δόθηκε ιδιαίτερη έμφαση στη μελέτη των συγκεκριμένων μοντέλων (Balakrishnan και Leung (1988a,b)). Τα περισσότερα από αυτά τα μοντέλα έχουν θεωρητικό ενδιαφέρον. Μερικά όμως, όπως το μοντέλο της γενίκευσης της λογιστικής κατανομής **Τύπου I** έχουν παράλληλα και μεγάλο πρακτικό ενδιαφέρον. Άλλα αναπτύχθηκαν και χρησιμοποιήθηκαν ως εναλλακτικά μοντέλα αντί του μοντέλου της λογιστικής παλινδρόμησης. Ο Prentice (1976) πρότεινε ένα τέτοιο μοντέλο με το όνομα **Τύπου IV**, το οποίο θα παρουσιαστεί παρακάτω.

A. Γενικευμένη λογιστική κατανομή Τύπου I

Η συνάρτηση κατανομής του λογιστικού μοντέλου **Τύπου I** δίνεται από την έκφραση

$$G_a^{-1}(y) = (1 + \exp(-y))^{-a} \quad \alpha > 0, -\infty < y < \infty. \quad (4.1)$$

Παρατηρούμε ότι αν πάρουμε τη συνάρτηση κατανομής της τυχαίας μεταβλητής

$Z \sim L\left(0, \frac{\pi^2}{3}\right)$ και την υψώσουμε στη δύναμη $\alpha > 0$ προκύπτει η σχέση (4.1).

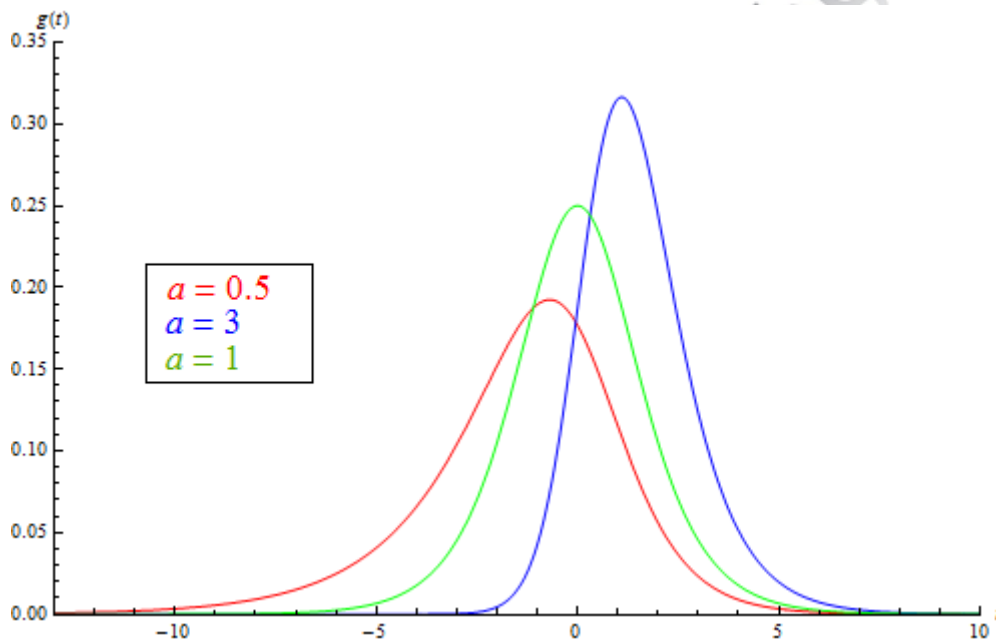
Συνεπώς, για $\alpha = 1$ καταλήγουμε στη συνάρτηση κατανομής της λογιστικής κατανομής όπως δίνεται από τη σχέση (2.5).

Η **συνάρτηση πυκνότητας** θα δίνεται από τον τύπο

$$g_a^{-I}(y) = \frac{\alpha \exp(-y)}{(1 + \exp(-y))^{\alpha+1}} \quad \alpha > 0, -\infty < y < \infty. \quad (4.2)$$

Παρακάτω δίνεται το γράφημα της συνάρτησης πυκνότητας (4.2) για διάφορες τιμές του $\alpha > 0$.

Σχήμα 4.1 Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I για διάφορες τιμές της παραμέτρου α .



B. Γενικευμένη λογιστική κατανομή Τύπου II

Η **συνάρτηση κατανομής** του λογιστικού μοντέλου **Τύπου II** δίνεται από την έκφραση

$$G_a^{-II}(y) = 1 - \frac{\exp(-\alpha \cdot y)}{(1 + \exp(-y))^\alpha} \quad \alpha > 0, -\infty < y < \infty. \quad (4.3)$$

Αν μια τυχαία μεταβλητή Y έχει τη **συνάρτηση κατανομής** του λογιστικού μοντέλου Τύπου I τότε η τυχαία μεταβλητή $-Y$ έχει **συνάρτηση κατανομής** τη σχέση (4.3). Αυτός είναι και ο λόγος που η γενικευμένη λογιστική κατανομή Τύπου II συναντάται με το όνομα **Αρνητική Τύπου I** (*Negative Type I*). Πράγματι, αν η τυχαία

μεταβλητή Y ακολουθεί τη γενικευμένη λογιστική κατανομή Τύπου I, για την τυχαία μεταβλητή $-Y$ θα έχουμε

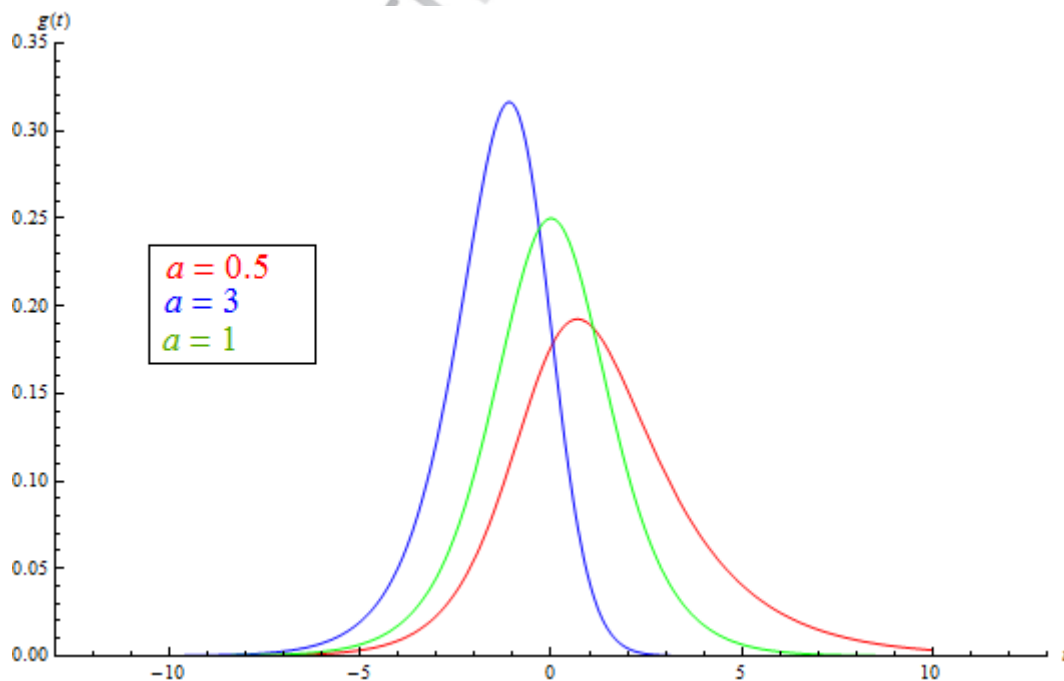
$$\begin{aligned} F_{-Y}(y) &= P(-Y \leq y) = P(Y > -y) = 1 - P(Y \leq -y) = 1 - G_a^I(-y) \\ &= 1 - (1 + \exp(y))^{-a} = 1 - \left(\frac{1}{1 + \exp(y)} \right)^a = 1 - \left(\frac{\exp(-y)}{1 + \exp(-y)} \right)^a \\ &= 1 - \frac{\exp(-\alpha \cdot y)}{(1 + \exp(-y))^a}. \end{aligned}$$

Η συνάρτηση πυκνότητας του λογιστικού μοντέλου Τύπου II δίνεται από τον τύπο

$$g_a^{II}(y) = \frac{\alpha \exp(-\alpha \cdot y)}{(1 + \exp(-y))^{a+1}} \quad \alpha > 0, -\infty < y < \infty. \quad (4.4)$$

Παρακάτω δίνεται το γράφημα της συνάρτησης πυκνότητας (4.4) για διάφορες τιμές του $\alpha > 0$.

Σχήμα 4.2 Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου II για διάφορες τιμές της παραμέτρου α .



Γ. Γενικευμένη λογιστική κατανομή Τύπου III

Η **συνάρτηση πυκνότητας** για το λογιστικό μοντέλο **Τύπου III** δίνεται από τον τύπο

$$g_a^{\text{III}}(y) = \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot y)}{(1 + \exp(-y))^{2\alpha}} \quad \alpha > 0, -\infty < y < \infty. \quad (4.5)$$

Με $\Gamma : (0, \infty) \rightarrow \mathbf{R}$ συμβολίζεται η συνάρτηση Γάμμα η οποία ορίζεται από τον τύπο

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} \exp(-t) dt \quad \alpha > 0.$$

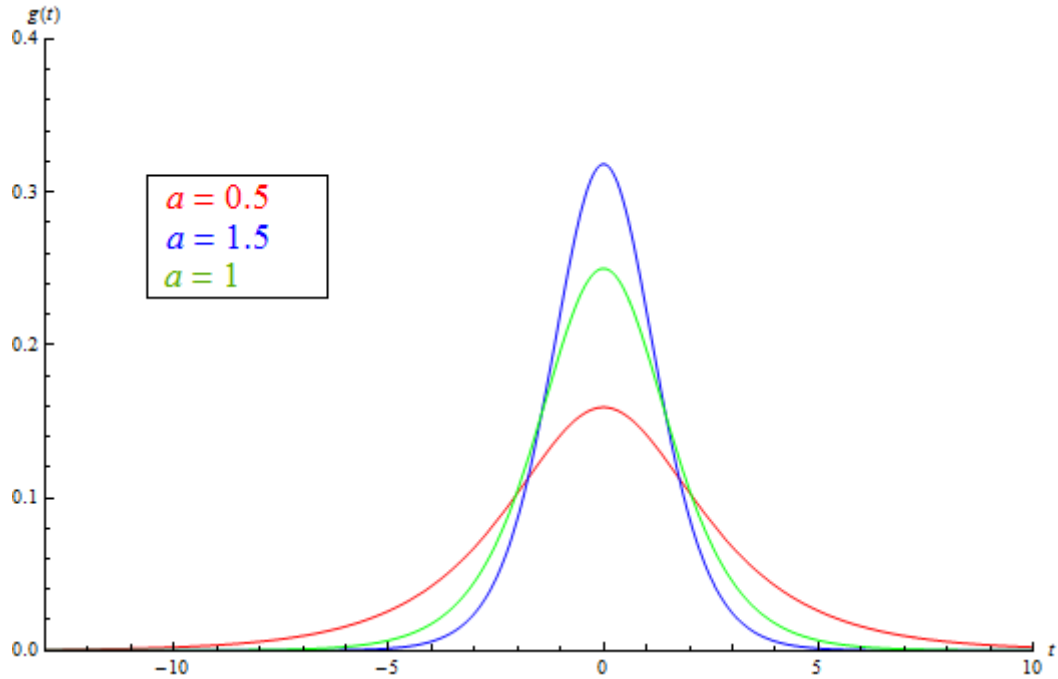
Θέτοντας $\alpha = 1$ στη σχέση (4.5) οδηγούμαστε στη συνάρτηση πυκνότητας της λογιστικής κατανομής όπως δίνεται από τη σχέση (2.6).

Το γενικευμένο λογιστικό μοντέλο Τύπου III είναι το μόνο γενικευμένο λογιστικό μοντέλο που έχει **συνάρτηση πυκνότητας** συμμετρική ως προς το μηδέν. Αναλυτικά έχουμε

$$\begin{aligned} g_a^{\text{III}}(-y) &= \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot (-y))}{(1 + \exp(-(-y)))^{2\alpha}} = \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(\alpha \cdot y)}{(1 + \exp(y))^{2\alpha}} \\ &= \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \left(\frac{\exp(y)}{(1 + \exp(y))^2} \right)^{\alpha} = \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \left(\frac{\exp\left(\frac{y}{2}\right)}{1 + \exp(y)} \right)^{2\alpha} \\ &= \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \left(\frac{\exp\left(-\frac{y}{2}\right)}{1 + \exp(-y)} \right)^{2\alpha} = \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot y)}{(1 + \exp(-y))^{2\alpha}} = g_a^{\text{III}}(y) \end{aligned}$$

Παρακάτω δίνεται το γράφημα της συνάρτησης πυκνότητας (4.5) για διάφορες τιμές του $\alpha > 0$.

Σχήμα 4.3 Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου III για διάφορες τιμές της παραμέτρου a .



Δ. Γενικευμένη λογιστική κατανομή Τύπου IV

Το γενικευμένο λογιστικό μοντέλο που μελετήθηκε από τους Prentice (1976) , (Prentice και Kalbfleisch (1980) θα το ονομάζουμε γενικευμένο λογιστικό μοντέλο **Τύπου IV**. Η **συνάρτηση πυκνότητας** για το λογιστικό μοντέλο Τύπου IV θα δίνεται από τον τύπο

$$g_{pq}^{IV}(y) = \Gamma(p+q)[\Gamma(p)\Gamma(q)]^{-1} \left[\frac{1}{1+\exp(-y)} \right]^p \left[\frac{\exp(-y)}{1+\exp(-y)} \right]^q \quad (4.6)$$

με $p, q > 0$, $-\infty < y < \infty$.

Το μοντέλο Τύπου IV είναι το πιο γενικό μοντέλο από όσα προαναφέρθηκαν. Θέτοντας τιμές στις παραμέτρους $p = \alpha$ και $q = 1$ στη σχέση (4.6) έχουμε

$$\begin{aligned} g_{pq}^{IV}(y) &= \Gamma(\alpha+1)[\Gamma(\alpha)\Gamma(1)]^{-1} [F^*(y)]^\alpha [1-F^*(y)] \\ &= \alpha \Gamma(\alpha)\Gamma(\alpha)^{-1} [1+\exp(-y)]^{-\alpha} [\exp(-y)][1+\exp(-y)]^{-1} \\ &= \frac{\alpha \exp(-y)}{(1+\exp(-y))^{\alpha+1}} \end{aligned}$$

η οποία αποτελεί έκφραση της συνάρτησης πυκνότητας του μοντέλου Τύπου I.

Με $F^*(\cdot)$ συμβολίζουμε τη συνάρτηση κατανομής της σχέσης (2.5).

Θέτοντας $p = 1$ και $q = \alpha$ στη σχέση (4.6) έχουμε

$$\begin{aligned} g_{pq}^{IV}(y) &= \Gamma(1+\alpha) [\Gamma(1)\Gamma(\alpha)]^{-1} [F^*(y)] [1-F^*(y)]^\alpha \\ &= \alpha \Gamma(\alpha) \Gamma(\alpha)^{-1} [1+\exp(-y)]^{-1} [\exp(-y)]^\alpha [1+\exp(-y)]^{-\alpha} \\ &= \frac{\alpha [\exp(-y)]^\alpha}{(1+\exp(-y))^{a+1}} = \frac{\alpha \exp(-\alpha \cdot y)}{(1+\exp(-y))^{a+1}} \end{aligned}$$

η οποία αποτελεί έκφραση της συνάρτησης πυκνότητας του μοντέλου Τύπου II.

Επίσης αν θέσουμε $p = \alpha$ και $q = \alpha$ στη σχέση (4.6) έχουμε

$$\begin{aligned} g_{pq}^{IV}(y) &= \Gamma(\alpha+\alpha) [\Gamma(\alpha)\Gamma(\alpha)]^{-1} [F^*(y)]^\alpha [1-F^*(y)]^\alpha \\ &= \Gamma(2\alpha) [\Gamma(\alpha)]^{-2} [F^*(y)(1-F^*(y))]^\alpha \\ &= \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \left[\frac{\exp(-y)}{[1+\exp(-y)]^2} \right]^\alpha \\ &= \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot y)}{(1+\exp(-y))^{2\alpha}} \end{aligned}$$

η οποία αποτελεί έκφραση της συνάρτησης πυκνότητας του μοντέλου Τύπου III.

Τέλος για $p = q = 1$ εύκολα καταλήγουμε στη συνάρτηση πυκνότητας της λογιστικής της σχέσης (2.6).

Παρακάτω παρουσιάζεται συνοπτικός πίνακας για τους τέσσερις κύριους τύπους της γενικευμένης λογιστικής κατανομής.

Πίνακας 4.1 Οι Τέσσερις Κύριοι Τύποι των Γενικευμένων Λογιστικών Κατανομών

Τύπος	Παράμετροι	Συνάρτηση Πυκνότητας $g(y)$
I	$\alpha > 0$	$\frac{\alpha \exp(-y)}{(1+\exp(-y))^{a+1}}$
II	$\alpha > 0$	$\frac{\alpha \exp(-\alpha \cdot y)}{(1+\exp(-y))^{a+1}}$
III	$\alpha > 0$	$\frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot y)}{(1+\exp(-y))^{2\alpha}}$
IV	$p > 0$ $q > 0$	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{\exp(-y)}{1+\exp(-y)} \right)^q \left(\frac{1}{1+\exp(-y)} \right)^p$

4.2 Ιδιότητες

Η ροπογεννήτρια συνάρτηση για τη γενικευμένη λογιστική συνάρτηση Τύπου I είναι ίση με

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \int_{-\infty}^{\infty} g_a^{-1}(y) e^{ty} dy = \int_{-\infty}^{\infty} \frac{\alpha \exp(-y)}{(1 + \exp(-y))^{a+1}} \exp(ty) dy \\ &= \alpha \int_{-\infty}^{\infty} \frac{\exp[(t-1)y]}{(1 + \exp(-y))^{a+1}} dy \end{aligned} \quad (4.7)$$

Κάνοντας την αλλαγή μεταβλητής $u = \frac{1}{1 + \exp(-y)}$, η (4.7) γίνεται

$$M_Y(t) = \alpha \int_0^1 \left(\frac{u}{1-u} \right)^{t-1} \frac{u^{a+1}}{u(1-u)} du = \alpha \int_0^1 u^{t+a-1} (1-u)^{-t} du \quad (4.8)$$

Λαμβάνοντας υπόψη ότι η συνάρτηση Βήτα δίνεται από τον τύπο

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad \forall x, y \in \mathbf{R}^+ - \{0\}$$

η (4.8) μπορεί να γραφεί στη μορφή

$$M_Y(t) = \alpha \cdot B(t + \alpha, 1 - t) \quad (4.9)$$

Χρησιμοποιώντας την ιδιότητα που συνδέει τη συνάρτηση Βήτα με τη Γάμμα προκύπτει ότι

$$M_Y(t) = \alpha \cdot \frac{\Gamma(t + \alpha) \Gamma(1 - t)}{\Gamma(\alpha + 1)} = \frac{\Gamma(t + \alpha) \Gamma(1 - t)}{\Gamma(\alpha)} \quad (4.10)$$

Για να υπολογίσουμε τη μέση τιμή, υπολογίζουμε τη παράγωγο

$$\begin{aligned} M'(t) &= \frac{1}{\Gamma(\alpha)} \left[\Gamma'(1-t)(1-t)' \Gamma(\alpha+t) + \Gamma(1-t) \Gamma'(\alpha+t)(\alpha+t)' \right] \\ &= \frac{1}{\Gamma(\alpha)} \left[-\Gamma'(1-t) \Gamma(\alpha+t) + \Gamma(1-t) \Gamma'(\alpha+t) \right] \end{aligned}$$

και για $t > 0$ προκύπτει

$$M'(0) = \frac{1}{\Gamma(\alpha)} \left[\Gamma'(\alpha) \Gamma(1) - \Gamma'(1) \Gamma(\alpha) \right] = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma'(1)}{\Gamma(1)} = \psi(\alpha) - \psi(1)$$

όπου με $\psi(\cdot)$ θα συμβολίζουμε την πρώτη παράγωγο της συνάρτησης $\log \Gamma(\cdot)$. Με $\psi'(\cdot), \psi''(\cdot) \dots$ θα συμβολίζουμε τις διαδοχικές παραγώγους της αντίστοιχα. Επομένως

$$E(Y) = M(0) = \psi(\alpha) - \psi(1).$$

Για να υπολογίσουμε τη διακύμανση, υπολογίζουμε τη δεύτερη παράγωγο

$$\begin{aligned} M''(t) &= \left[\frac{1}{\Gamma(\alpha)} [-\Gamma'(1-t)\Gamma(\alpha+t) + \Gamma(1-t)\Gamma'(\alpha+t)] \right]' \\ &= \frac{1}{\Gamma(\alpha)} [\Gamma''(\alpha+t)\Gamma(1-t) - \Gamma'(\alpha+t)\Gamma'(1-t) + \Gamma''(1-t)\Gamma(\alpha+t) - \Gamma'(\alpha+t)\Gamma'(1-t)] \\ &= \frac{1}{\Gamma(\alpha)} [\Gamma''(\alpha+t)\Gamma(1-t) + \Gamma''(1-t)\Gamma(\alpha+t) - 2\Gamma'(\alpha+t)\Gamma'(1-t)] \end{aligned}$$

Θέτοντας όπου $t = 0$ προκύπτει

$$M''(0) = \frac{1}{\Gamma(\alpha)} [\Gamma''(\alpha)\Gamma(1) + \Gamma''(1)\Gamma(\alpha) - 2\Gamma'(1)\Gamma'(\alpha)] = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} + \Gamma''(1) - \frac{2\Gamma'(1)\Gamma'(\alpha)}{\Gamma(\alpha)}$$

Έτσι, η διακύμανση ισούται

$$\begin{aligned} M''(0) - [M'(0)]^2 &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} + \Gamma''(1) - \frac{2\Gamma'(1)\Gamma'(\alpha)}{\Gamma(\alpha)} - [\psi(\alpha) - \psi(1)]^2 \\ &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma''(1)}{\Gamma(1)} - [\psi^2(\alpha) - 2\psi(\alpha)\psi(1) + \psi^2(1)] = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} + \Gamma''(1) \\ &\quad - \frac{2\Gamma'(1)\Gamma'(\alpha)}{\Gamma(\alpha)} - \left[\frac{[\Gamma'(\alpha)]^2}{\Gamma^2(\alpha)} - 2\frac{\Gamma'(\alpha)\Gamma'(1)}{\Gamma(\alpha)\Gamma(1)} + \frac{[\Gamma'(1)]^2}{\Gamma^2(1)} \right] \\ &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} + \Gamma''(1) - \frac{[\Gamma'(\alpha)]^2}{\Gamma^2(\alpha)} - [\Gamma'(1)]^2 \end{aligned} \quad (4.11)$$

Παρατηρώντας ότι

$$\begin{aligned} \psi'(\alpha) + \psi'(1) &= \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right]' + \left[\frac{\Gamma'(1)}{\Gamma(1)} \right]' = \left[\frac{\Gamma''(\alpha)\Gamma(\alpha) - [\Gamma'(\alpha)]^2}{\Gamma^2(\alpha)} \right] + \Gamma''(1) - [\Gamma'(1)]^2 \\ &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \frac{[\Gamma'(\alpha)]^2}{\Gamma^2(\alpha)} + \Gamma''(1) - [\Gamma'(1)]^2 \end{aligned} \quad (4.12)$$

συμπεραίνουμε ότι

$$M''(0) - [M'(0)]^2 = \psi'(\alpha) + \psi'(1)$$

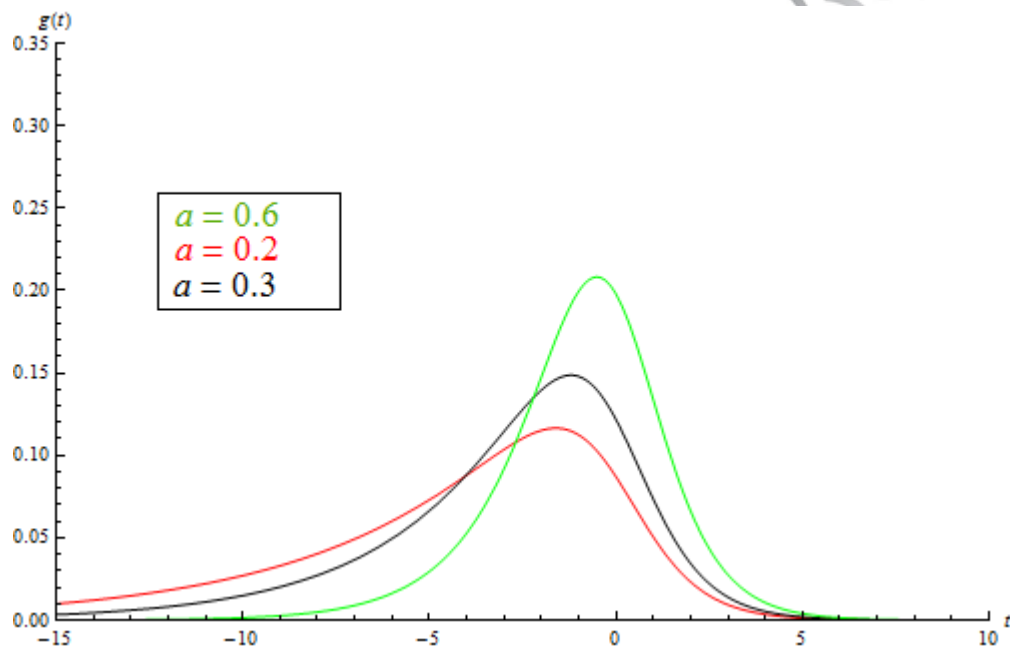
Οι Balakrishnan & Leung (1988a), υπολόγισαν τη ροπογεννήτρια της γενικευμένης λογιστικής κατανομής Τύπου I όπως δίνεται από τη σχέση (4.10) και βρήκαν την

πρώτη και δεύτερη ροπή όπως υπολογίστηκαν παραπάνω. Επίσης υπολόγισαν την τρίτη και τέταρτη ροπή οι οποίες δίνονται από τους παρακάτω τύπους

- $\mu_3 = \psi''(\alpha) - \psi''(1)$
- $\mu_4 = \psi'''(\alpha) + \psi'''(1) + 3[V(Y)]^2$

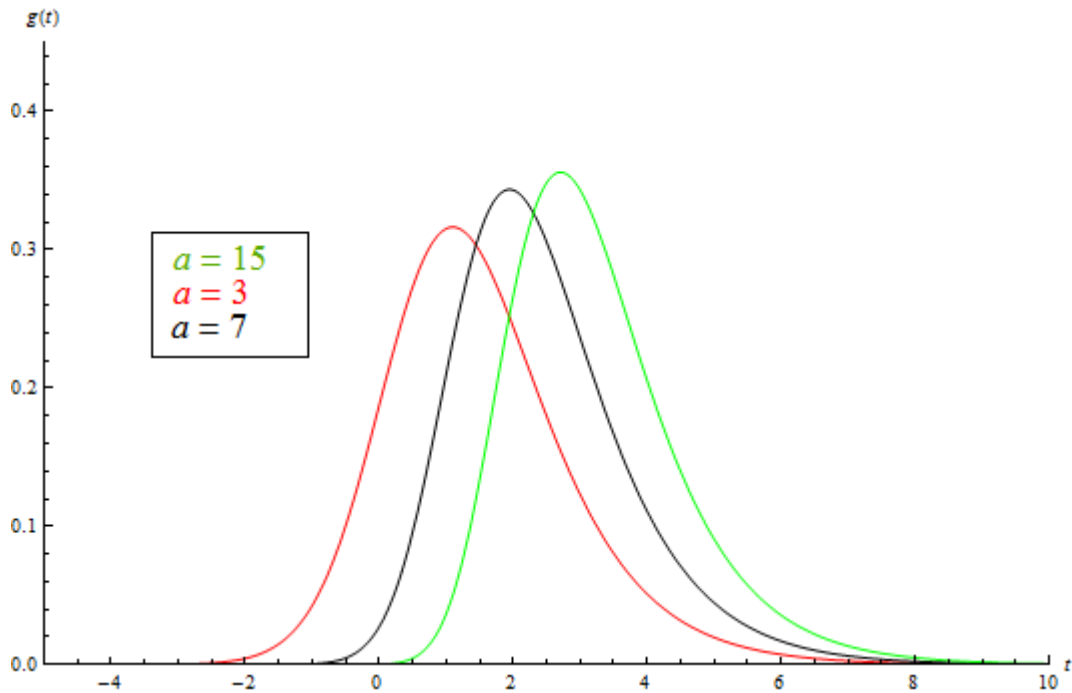
Η γενικευμένη λογιστική Τύπου I έχει αρνητική καμπυλότητα όταν η τιμή της παραμέτρου α βρίσκεται στο διάστημα $(0, 1)$

Σχήμα 4.4 Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I όταν η τιμή της παραμέτρου α παίρνει τιμές στο διάστημα $(0, 1)$



ενώ γίνεται θετική η καμπυλότητα της όταν η τιμή της παραμέτρου α παίρνει τιμές στο διάστημα $(1, \infty)$.

Σχήμα 4.5 Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I όταν η τιμή της παραμέτρου a παίρνει τιμές στο διάστημα $(1, \infty)$.



4.3 Σχέση της γενικευμένης λογιστικής κατανομής με άλλες κατανομές

Οι Ahuja και Nash (1967) έδειξαν ότι όταν μια τυχαία μεταβλητή Y ακολουθεί τη γενικευμένη λογιστική κατανομή Τύπου I τότε η τυχαία μεταβλητή $-\alpha Y$ για τιμές της παραμέτρου α κοντά στο μηδέν, ακολουθεί την τυποποιημένη εκθετική κατανομή ($\lambda = 1$). Υπενθυμίζεται ότι μια συνεχής τυχαία μεταβλητή X με συνάρτηση πυκνότητας

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

όπου $\lambda > 0$ λέμε ότι ακολουθεί την **εκθετική κατανομή** με παράμετρο λ .

Ας θεωρήσουμε αρχικά το γραμμικό μετασχηματισμό $Z = -\alpha Y$. Η **συνάρτηση κατανομής** του νέου μοντέλου βρίσκεται από την έκφραση

$$\begin{aligned} F_Z(t) &= P(Z \leq t) = P(-\alpha Y \leq t) = P\left(Y > -\frac{t}{\alpha}\right) = 1 - P\left(Y \leq -\frac{t}{\alpha}\right) \\ &= 1 - G_a^{-1}\left(-\frac{t}{\alpha}\right) \end{aligned}$$

Με $G_a^{-1}(\cdot)$ συμβολίζουμε τη **συνάρτηση κατανομής** της γενικευμένης λογιστικής Τύπου I.

Η **συνάρτηση πυκνότητας** θα δίνεται από τον τύπο

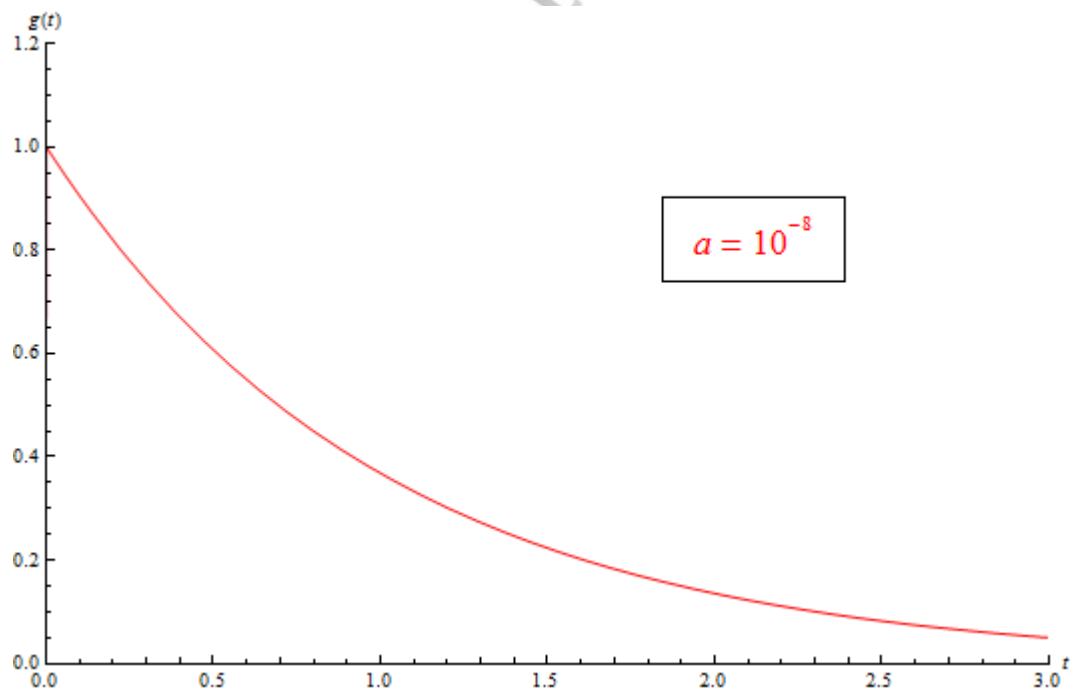
$$f_Z(t) = F'_Z(t) = \left[1 - G_a^{-1}\left(-\frac{t}{\alpha}\right) \right]' = -G_a^{-1}\left(-\frac{t}{\alpha}\right) \left(-\frac{t}{\alpha}\right)' = \frac{1}{\alpha} \cdot g_a^{-1}\left(-\frac{t}{\alpha}\right)$$

$$= \frac{1}{\alpha} \frac{\alpha \exp\left(-\frac{t}{\alpha}\right)}{\left(1 + \exp\left(-\frac{t}{\alpha}\right)\right)^{a+1}} = \frac{\exp\left(-\frac{t}{\alpha}\right)}{\left(1 + \exp\left(-\frac{t}{\alpha}\right)\right)^{a+1}}$$

με $\alpha > 0$ και $-\infty < t < \infty$.

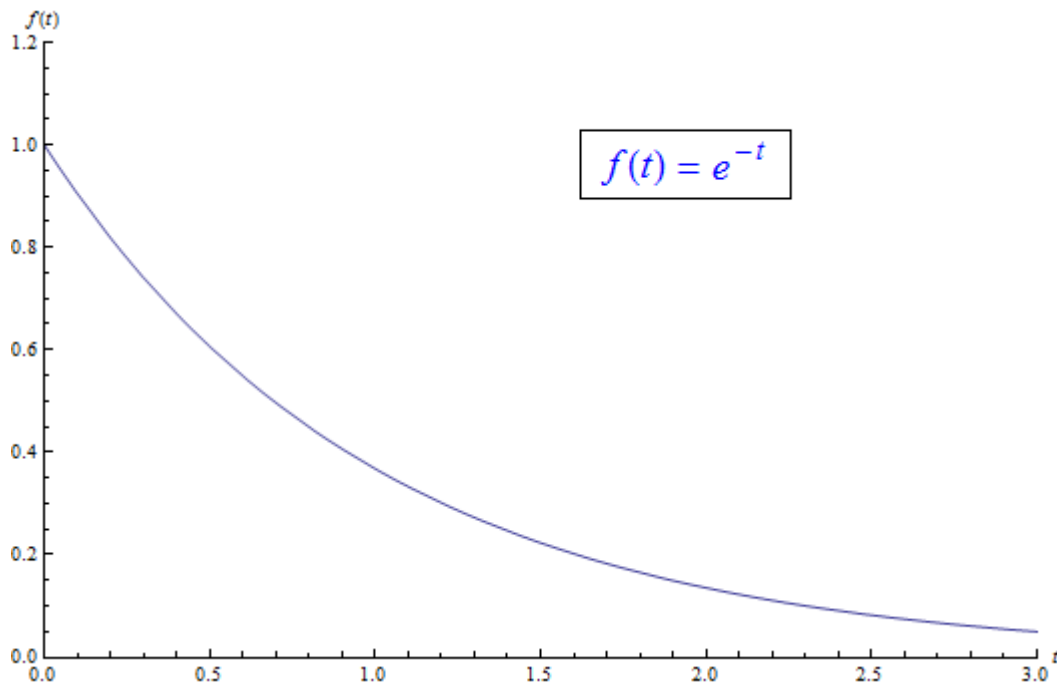
Παρακάτω δίνεται το γράφημα της τυποποιημένης εκθετικής συνάρτησης και της γενικευμένης λογιστικής συνάρτησης Τύπου I για τιμές της παραμέτρου α κοντά στο 0.

Σχήμα 4.6 Γράφημα της συνάρτησης πυκνότητας της τυποποιημένης εκθετικής κατανομής



και

Σχήμα 4.7 Γράφημα της συνάρτησης πυκνότητας του λογιστικού μοντέλου Τύπου I για τιμές της παραμέτρου α κοντά στο 0.



Επίσης, όταν μια τυχαία μεταβλητή Y ακολουθεί τη γενικευμένη λογιστική κατανομή Τύπου I τότε η τυχαία μεταβλητή $Y - \log(\alpha)$ για μεγάλες τιμές της παραμέτρου α , ακολουθεί τη κατανομή Gumbel. Υπενθυμίζεται ότι μια συνεχής τυχαία μεταβλητή X που ακολουθεί τη κατανομή Gumbel έχει **συνάρτηση κατανομής** που δίνεται από τη σχέση

$$F_X(x) = \exp(-\exp(-x)) \quad -\infty < x < \infty.$$

ενώ η **συνάρτηση πυκνότητας** είναι

$$f_X(x) = \exp(-x + \exp(-x)) \quad -\infty < x < \infty.$$

Ας θεωρήσουμε το γραμμικό μετασχηματισμό $Z = Y - \log(\alpha)$. Η **συνάρτηση κατανομής** του νέου μοντέλου βρίσκεται ως εξής

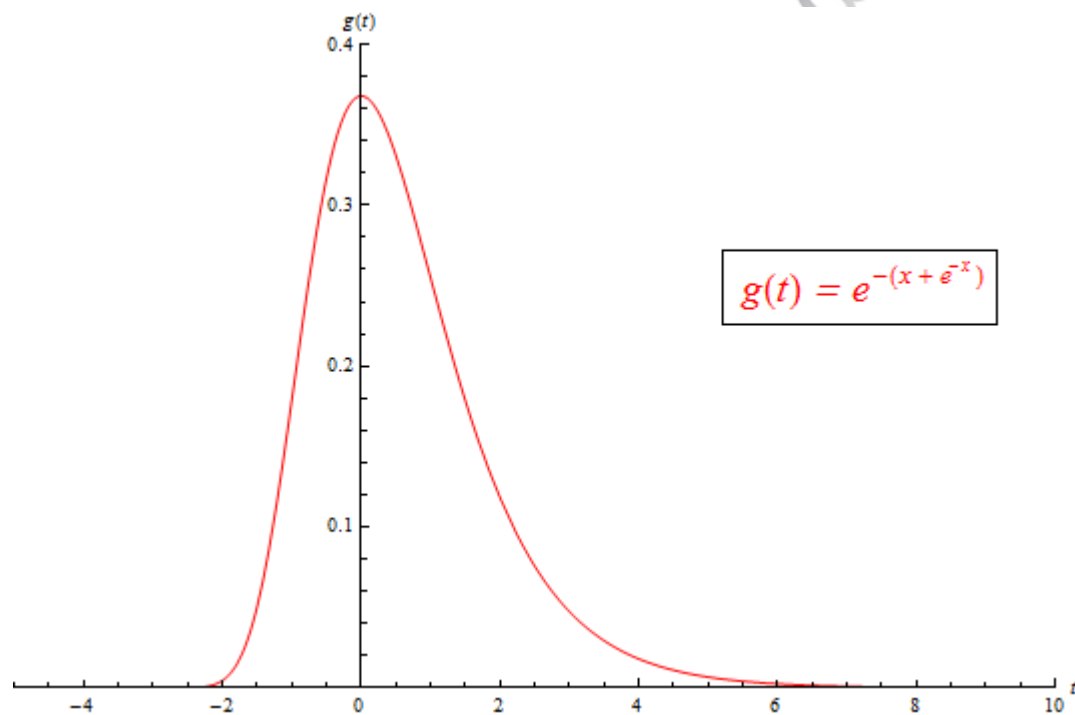
$$\begin{aligned} F_Z(t) &= P(Z \leq t) = P(Y - \log(\alpha) \leq t) = P(Y \leq t + \log(\alpha)) \\ &= G_a^{-1}(t + \log(\alpha)) \end{aligned}$$

Παραγωγίζοντας βρίσκουμε τη **συνάρτηση πυκνότητας** η οποία θα δίνεται από τον τύπο

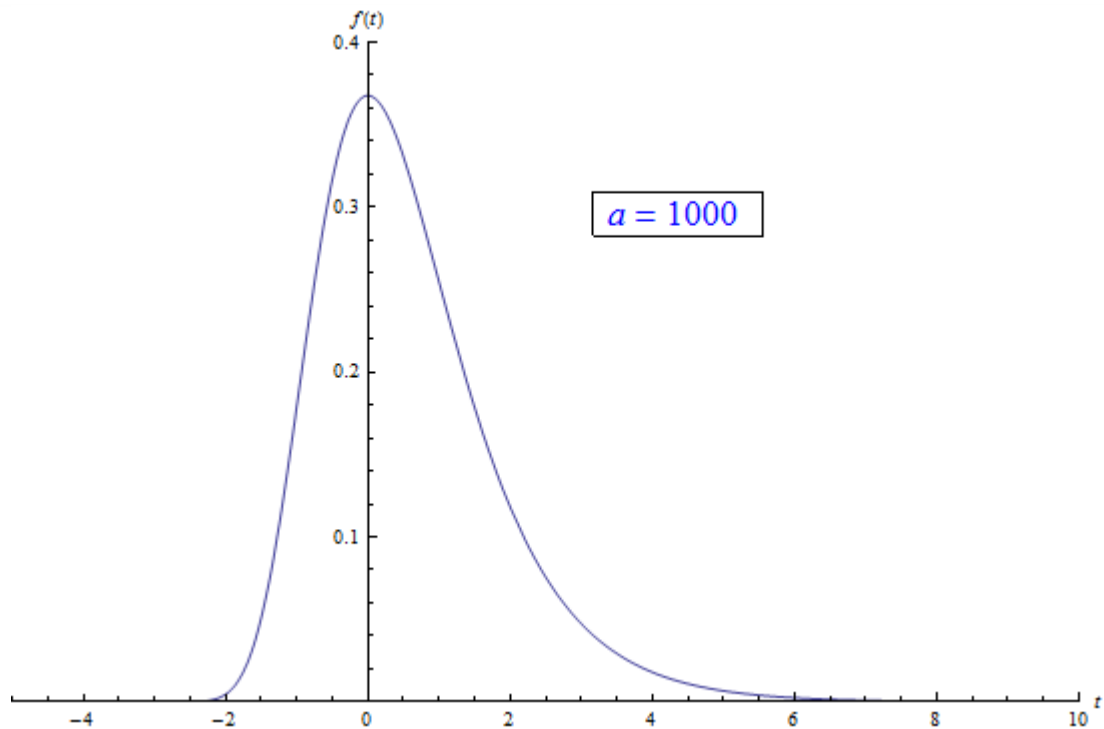
$$\begin{aligned}
 f_z(t) = F'_z(t) &= \left[G_a^{-1}(t + \log(\alpha)) \right]' = -G_a^{-1}(t + \log(\alpha))(t + \log(\alpha))' = g_a^{-1}(t + \log(\alpha)) \\
 &= \exp \left[-[t + \log(\alpha)] + \exp(-[t + \log(\alpha)]) \right] \\
 &= \exp \left[-\frac{1}{\alpha} \exp(-t) \right]
 \end{aligned}$$

Παρακάτω δίνεται το γράφημα της κατανομής Gumbel και της γενικευμένης λογιστικής συνάρτησης Τύπου I για μεγάλη τιμή του α .

Σχήμα 4.8 Γράφημα της συνάρτησης πυκνότητας της κατανομής Gumbel.



Σχήμα 4.9 Γράφημα της γενικευμένης λογιστικής συνάρτησης Τύπου I για μεγάλες τιμές της παραμέτρου α



Η ροπογεννήτρια συνάρτηση για τη γενικευμένη λογιστική συνάρτηση Τύπου II είναι

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \int_{-\infty}^{\infty} g_a^{\text{II}}(y) e^{ty} dy = \int_{-\infty}^{\infty} \frac{\alpha \exp(-\alpha \cdot y)}{(1 + \exp(-y))^{\alpha+1}} \exp(ty) dy \\ &= \alpha \int_{-\infty}^{\infty} \frac{\exp[(t - \alpha)y]}{(1 + \exp(-y))^{\alpha+1}} dy \end{aligned} \quad (4.13)$$

Κάνουμε αλλαγή μεταβλητής θέτοντας $u = \frac{1}{1 + \exp(-y)}$ και η (4.13) γίνεται

$$M_Y(t) = \alpha \int_0^1 \left(\frac{u}{1-u} \right)^{t-\alpha} \frac{u^{a+1}}{u(1-u)} du = \alpha \int_0^1 u^t (1-u)^{\alpha-t-1} du \quad (4.14)$$

Η (4.14) μπορεί να γραφεί στη μορφή

$$M_Y(t) = \alpha \cdot B(t+1, \alpha-t).$$

Χρησιμοποιώντας την ιδιότητα που συνδέει τη συνάρτηση Βήτα με τη Γάμμα προκύπτει ότι

$$M_Y(t) = \alpha \cdot \frac{\Gamma(t+1)\Gamma(\alpha-t)}{\Gamma(\alpha+1)} = \frac{\Gamma(t+1)\Gamma(\alpha-t)}{\Gamma(\alpha)}.$$

Θεωρούμε μια τυχαία μεταβλητή X που ακολουθεί την εκθετική κατανομή με παράμετρο λ . Τότε η τυχαία μεταβλητή $Y = \log[\exp(X) - 1]$ ακολουθεί τη γενικευμένη λογιστική κατανομή Τύπου II.

Η συνάρτηση κατανομής της Y είναι

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\log(\exp(X) - 1) \leq y) = P(\exp(X) - 1 \leq \exp(y)) \\ &= P(X \leq \log(\exp(y) + 1)) = F(\log(\exp(y) + 1)) \end{aligned}$$

δηλαδή

$$F_Y(y) = F(\log(\exp(y) + 1))$$

και παραγωγίζοντας ως προς y βρίσκουμε

$$\begin{aligned} f_Y(y) &= F'_Y(y) = (F_Y(\log(\exp(y) + 1)))' = F'(\log(\exp(y) + 1))(\log(\exp(y) + 1))' \\ &= f(\log(\exp(y) + 1)) \frac{\exp(y)}{1 + \exp(y)} \end{aligned} \quad (4.15)$$

Αντικαθιστώντας τη συνάρτηση πυκνότητας $f(t)$ της εκθετικής κατανομής, η (4.15) γίνεται

$$\begin{aligned} f_Y(y) &= \lambda \cdot \exp(-\lambda \cdot \log(\exp(y) + 1)) \frac{\exp(y)}{1 + \exp(y)} = \lambda (\exp(y) + 1)^{-\lambda} \frac{\exp(y)}{1 + \exp(y)} \\ &= \lambda \cdot \frac{[\exp(-y)]^\lambda}{[1 + \exp(-y)]^\lambda} \cdot \frac{1}{1 + \exp(-y)} = \lambda \frac{\exp(-\lambda \cdot y)}{(1 + \exp(-y))^{\lambda+1}} \end{aligned}$$

με $\lambda > 0$ και $-\infty < y < \infty$. Παρατηρούμε ότι η συνάρτηση πυκνότητας $f_Y(y)$ οδηγεί στον τύπο της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής Τύπου II.

Η ροπογεννήτρια συνάρτηση για τη γενικευμένη λογιστική συνάρτηση Τύπου III είναι

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \int_{-\infty}^{\infty} g_a^{\text{III}}(y) e^{tY} dy = \int_{-\infty}^{\infty} \frac{1}{B(\alpha, \alpha)} \frac{\exp(-\alpha y)}{(1 + \exp(-y))^{2-\alpha}} \exp(ty) dy \\ &= \frac{1}{B(\alpha, \alpha)} \int_{-\infty}^{\infty} \frac{\exp((t - \alpha)y)}{(1 + \exp(-y))^{2-\alpha}} dy \end{aligned} \quad (4.16)$$

Κάνουμε αλλαγή μεταβλητής θέτοντας $u = \frac{1}{1 + \exp(-y)}$ και η (4.16) γίνεται

$$\begin{aligned}
M_Y(t) &= \frac{1}{B(\alpha, \alpha)} \int_0^1 \left(\frac{u}{1-u} \right)^{t-\alpha} \frac{u^{2\alpha}}{u(1-u)} du \\
&= \frac{1}{B(\alpha, \alpha)} \int_0^1 u^{t+\alpha-1} (1-u)^{\alpha-t-1} du
\end{aligned} \tag{4.17}$$

Η (4.17) μπορεί να γραφεί στη μορφή

$$M_Y(t) = \frac{B(t+\alpha, \alpha-t)}{B(\alpha, \alpha)}$$

Χρησιμοποιώντας την ιδιότητα που συνδέει τη συνάρτηση Βήτα με τη Γάμμα προκύπτει ότι

$$M_Y(t) = \frac{\frac{\Gamma(t+\alpha)\Gamma(\alpha-t)}{\Gamma(2\alpha)}}{\frac{[\Gamma(\alpha)]^2}{\Gamma(2\alpha)}} = \frac{\Gamma(t+\alpha)\Gamma(\alpha-t)}{[\Gamma(\alpha)]^2}$$

Ο Davidson (1980) υπολόγισε τη ροπογεννήτρια της γενικευμένης λογιστικής Τύπου III όπως δίνεται από την παραπάνω σχέση, και βρήκε τις ροπές

- $E(Y) = \mu_1 = 0$
- $V(Y) = \mu_2 = 2\psi'(\alpha)$
- $\mu_3 = 0$
- $\mu_4 = 2\psi'''(\alpha) + 12[\psi'(\alpha)]^2$

Οι ροπές περιττής τάξης είναι ίσες με το μηδέν εφόσον η συνάρτηση πυκνότητας είναι συμμετρική ως προς το μηδέν.

Επίσης παρατήρησε ότι όταν μια τυχαία μεταβλητή Y ακολουθεί τη γενικευμένη λογιστική Τύπου III κατανομή με παράμετρο $\alpha > 0$, τότε η $(2/\alpha)^{1/2}Y$ για μεγάλες τιμές της παραμέτρου $\alpha > 0$, συμπεριφέρεται σχεδόν σαν τυποποιημένη κανονική κατανομή. Θα δείξουμε στη συνέχεια πως προέκυψε στη συνέχεια η παρατήρηση του Davidson (1980).

Έχουμε

$$F_Y(y) = P(Y \leq y) = P\left((2/\alpha)^{1/2}Y \leq y\right) = P\left(Y \leq (2/\alpha)^{-1/2}y\right) = F\left((2/\alpha)^{-1/2}y\right)$$

δηλαδή

$$F_Y(y) = F\left((2/\alpha)^{-1/2} y\right)$$

και παραγωγίζοντας ως προς y βρίσκουμε

$$\begin{aligned} f_Y(y) = F'_Y(y) &= \left(F_Y\left((2/\alpha)^{-1/2} y\right)\right)' = F'\left((2/\alpha)^{-1/2} y\right) \left((2/\alpha)^{-1/2} y\right)' \\ &= f\left((2/\alpha)^{-1/2} y\right) \frac{1}{(2/\alpha)^{1/2}} \end{aligned} \quad (4.18)$$

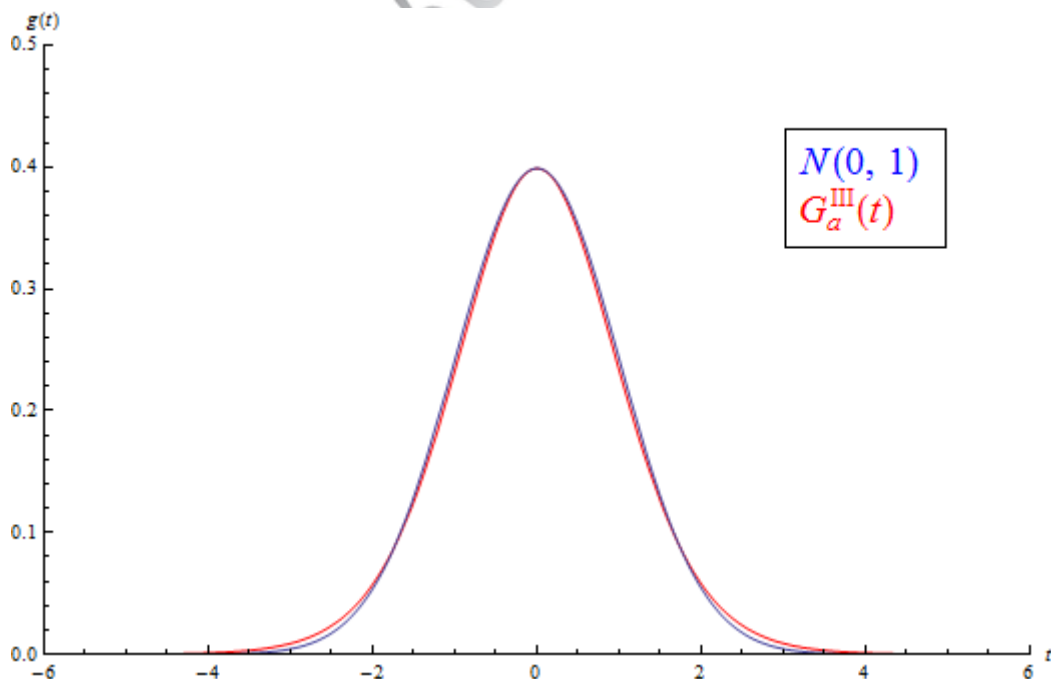
Σύμφωνα με τη **συνάρτηση πυκνότητας** της γενικευμένης λογιστικής Τύπου III, η (4.18) γίνεται

$$f_Y(y) = (2/\alpha)^{-1/2} \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot (2/\alpha)^{-1/2} y)}{\left(1 + \exp(-(2/\alpha)^{-1/2} y)\right)^{2\alpha}}$$

με $\alpha > 0$ και $-\infty < y < \infty$.

Παρακάτω δίνεται το γράφημα των δύο κατανομών ($N(0,1)$ και Γενικευμένη λογιστική Τύπου III με $\alpha = 2.12$)

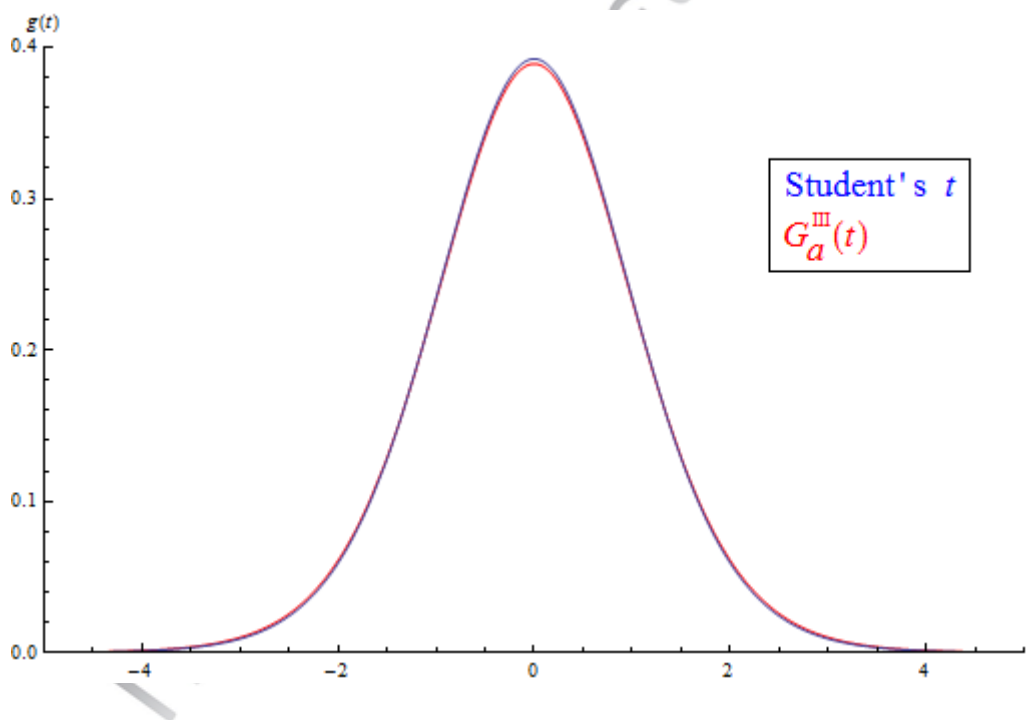
Σχήμα 4.10 Γράφημα της συνάρτησης πυκνότητας της τυποποιημένης κανονικής κατανομής και της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής Τύπου III για συγκεκριμένη τιμή της παραμέτρου α .



Οι Mudholkar και George (1978) παρατήρησαν τη μεγάλη ομοιότητα της γενικευμένης λογιστικής Τύπου III συνάρτησης με τη Student με εννιά βαθμούς ελευθερίας. Βασισμένοι σε αυτή την παρατήρηση οι George και Ojo (1980) και George, El-Saidi, και Singh (1986) πρότειναν ότι όταν θέλουμε να προσεγγίσουμε τη Student κατανομή με $\nu > 9$ βαθμούς ελευθερίας μέσω της γενικευμένης λογιστικής Τύπου III κατανομής να παίρνουμε ως τιμή της παραμέτρου $\alpha > 0$ θα την $\alpha = \frac{(\nu - 3.25)}{5.5}$, όπου ν οι βαθμοί ελευθερίας της αντίστοιχης κατανομής Student.

Παρακάτω δίνεται το γράφημα των δύο κατανομών (Student(15) και Γενικευμένη λογιστική Τύπου III με $\alpha = 2.13636$)

Σχήμα 4.11 Γράφημα της συνάρτησης πυκνότητας της κατανομής Student(15) και της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής Τύπου III για συγκεκριμένη τιμή της παραμέτρου α .



Η ροπογεννήτρια συνάρτηση για τη γενικευμένη λογιστική συνάρτηση Τύπου IV είναι

$$\begin{aligned}
 M_Y(t) = E(e^{tY}) &= \int_{-\infty}^{\infty} g_a^{IV}(y) e^{ty} dy = \int_{-\infty}^{\infty} \frac{1}{B(p, q)} \frac{\exp(-qy)}{(1 + \exp(-y))^{p+q}} \exp(ty) dy \\
 &= \frac{1}{B(p, q)} \int_{-\infty}^{\infty} \frac{\exp((t - q)y)}{(1 + \exp(-y))^{p+q}} dy \quad (4.19)
 \end{aligned}$$

Κάνουμε αλλαγή μεταβλητής θέτοντας $u = \frac{1}{1 + \exp(-y)}$ και η (4.19) γίνεται

$$\begin{aligned} M_Y(t) &= \frac{1}{B(p, q)} \int_0^1 \left(\frac{u}{1-u} \right)^{t-q} \frac{u^{p+q}}{u(1-u)} du \\ &= \frac{1}{B(p, q)} \int_0^1 u^{p+t-1} (1-u)^{q-t-1} du \end{aligned} \quad (4.20)$$

Η (4.20) μπορεί να γραφεί στη μορφή

$$M_Y(t) = \frac{B(p+t, q-t)}{B(p, q)}$$

Χρησιμοποιώντας την ιδιότητα που συνδέει τη συνάρτηση Βήτα με τη Γάμμα προκύπτει ότι

$$M_Y(t) = \frac{\frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p+q)}}{\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}} = \frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p)\Gamma(q)}.$$

Παρακάτω δίνεται συνοπτικός πίνακας των τεσσάρων κύριων τύπων γενικευμένων λογιστικών κατανομών με τις ροπογεννήτριες συναρτήσεις τους.

Πίνακας 4.2 Οι Ροπογεννήτριες των Τεσσάρων Κύριων Τύπων των Γενικευμένων Λογιστικών Κατανομών

Τύπος	Παράμετροι	Συνάρτηση Πυκνότητας $g(y)$	Ροπογεννήτριες
I	$\alpha > 0$	$\frac{\alpha \exp(-y)}{(1 + \exp(-y))^{\alpha+1}}$	$\frac{\Gamma(t + \alpha)\Gamma(1-t)}{\Gamma(\alpha)}$
II	$\alpha > 0$	$\frac{\alpha \exp(-\alpha \cdot y)}{(1 + \exp(-y))^{\alpha+1}}$	$\frac{\Gamma(t+1)\Gamma(\alpha-t)}{\Gamma(\alpha)}$
III	$\alpha > 0$	$\frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} \frac{\exp(-\alpha \cdot y)}{(1 + \exp(-y))^{2\alpha}}$	$\frac{\Gamma(t + \alpha)\Gamma(\alpha - t)}{[\Gamma(\alpha)]^2}$
IV	$p > 0$ $q > 0$	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{\exp(-y)}{1 + \exp(-y)} \right)^q \left(\frac{1}{1 + \exp(-y)} \right)^p$	$\frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p)\Gamma(q)}$

Οι τέσσερις πρώτες ροπές της γενικευμένης λογιστικής συνάρτησης Τύπου IV είναι

- $E(Y) = \mu_1 = \psi(p) - \psi(q)$
- $V(Y) = \mu_2 - [\mu_1]^2 = \psi'(p) + \psi'(q)$
- $\mu_3 = \psi''(p) - \psi''(q)$
- $\mu_4 = \psi'''(p) + \psi'''(q) + 3[V(Y)]^2$

Η γενικευμένη λογιστική κατανομή Τύπου IV μπορεί να προκύψει ως ο λογάριθμος των odds (logit) μιας τυχαίας μεταβλητής που ακολουθεί την κατανομή Βήτα. Συγκεκριμένα, αν μια τυχαία μεταβλητή T ακολουθεί την κατανομή Βήτα με παραμέτρους $p > 0$ και $q > 0$, τότε η τυχαία μεταβλητή $S = \log\left(\frac{T}{1-T}\right)$ έχει τη γενικευμένη λογιστική κατανομή Τύπου IV με παραμέτρους $p > 0$ και $q > 0$.

Αναλυτικά

$$\begin{aligned} F_S(s) &= P(S \leq s) = P\left(\log\left(\frac{T}{1-T}\right) \leq s\right) = P\left(T \leq \frac{\exp(s)}{1 + \exp(s)}\right) \\ &= F_T\left(\frac{\exp(s)}{1 + \exp(s)}\right) \end{aligned}$$

δηλαδή

$$F_S(s) = F_T\left(\frac{\exp(s)}{1 + \exp(s)}\right)$$

και παραγωγίζοντας ως προς s βρίσκουμε

$$\begin{aligned} f_S(s) &= F'_S(s) = \left(F_T\left(\frac{\exp(s)}{1 + \exp(s)}\right)\right)' = F'_T\left(\frac{\exp(s)}{1 + \exp(s)}\right) \left(\frac{\exp(s)}{1 + \exp(s)}\right)' \\ &= F'_T\left(\frac{\exp(s)}{1 + \exp(s)}\right) \left(\frac{\exp(s)(1 + \exp(s)) - \exp(s)\exp(s)}{(1 + \exp(s))^2}\right) \\ &= F'_T\left(\frac{\exp(s)}{1 + \exp(s)}\right) \left(\frac{\exp(s)}{(1 + \exp(s))^2}\right) \end{aligned} \quad (4.21)$$

Υπενθυμίζεται ότι μια συνεχής τυχαία μεταβλητή X ακολουθεί την **κατανομή Βήτα** με παραμέτρους α και β αν έχει συνάρτηση πυκνότητας

$$f(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 \leq x \leq 1 \\ 0, & \text{αλλού,} \end{cases}$$

όπου $\alpha > 0$, $\beta > 0$ είναι δύο πραγματικοί αριθμοί και

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

Χρησιμοποιώντας τη συνάρτηση πυκνότητας της **κατανομής Βήτα**, η (4.21) γίνεται

$$\begin{aligned} f_s(s) &= \left[\frac{\exp(s)}{(1+\exp(s))^2} \right] \frac{1}{B(p, q)} \left[\frac{\exp(s)}{1+\exp(s)} \right]^{p-1} \left[1 - \frac{\exp(s)}{1+\exp(s)} \right]^{q-1} \\ &= \left[\frac{\exp(s)}{(1+\exp(s))^2} \right] \frac{1}{B(p, q)} \left[\frac{\exp(s)}{1+\exp(s)} \right]^p \left[\frac{1}{1+\exp(s)} \right]^q \left[\frac{(1+\exp(s))^2}{\exp(s)} \right] \\ &= \frac{1}{B(p, q)} \left[\frac{\exp(s)}{1+\exp(s)} \right]^p \left[\frac{1}{1+\exp(s)} \right]^q = \frac{1}{\Gamma(p)\Gamma(q)} \left[\frac{\exp(s)}{1+\exp(s)} \right]^p \left[\frac{1}{1+\exp(s)} \right]^q \\ &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left[\frac{1}{1+\exp(-s)} \right]^p \left[\frac{\exp(-s)}{1+\exp(-s)} \right]^q = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} [F^*(s)]^p [1-F^*(s)]^q \end{aligned}$$

Η συνάρτηση πυκνότητας $f_s(s)$ μας οδηγεί στον τύπο της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής Τύπου IV.

Επίσης, αν μια τυχαία μεταβλητή Y ακολουθεί την κατανομή F (Fisher-Snedecor) με παραμέτρους $2p > 0$ και $2q > 0$, τότε η τυχαία μεταβλητή $-\log Y$ έχει τη γενικευμένη λογιστική κατανομή Τύπου IV.

Έχουμε

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(-\log Y \leq y) = P(\log Y > -y) \\ &= P(Y > \exp(-y)) = 1 - P(Y \leq \exp(-y)) \end{aligned}$$

δηλαδή

$$F_Y(y) = 1 - F(\exp(-y))$$

και παραγωγίζοντας ως προς y βρίσκουμε

$$\begin{aligned} f_Y(y) &= F'_Y(y) = (1 - F_Y(\exp(-y)))' = -F'(\exp(-y))(\exp(-y))' \\ &= f(\exp(-y))\exp(-y) \end{aligned} \quad (4.22)$$

Υπενθυμίζεται ότι

Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας

$$f(x; \alpha, \beta) = \frac{1}{B\left(\frac{\alpha}{2}, \frac{\beta}{2}\right)} \left(\frac{\alpha}{\beta}\right)^{\frac{\alpha}{2}} x^{\frac{\alpha}{2}-1} \left(1 + \frac{\alpha}{\beta} x\right)^{-\frac{\alpha+\beta}{2}} \quad -\infty < x < \infty,$$

όπου $\alpha > 0$, $\beta > 0$ είναι δύο πραγματικοί αριθμοί.

Η κατανομή της τυχαίας μεταβλητής X καλείται **κατανομή F** με παραμέτρους α και β .

Σύμφωνα με τη συνάρτηση πυκνότητας της **κατανομής F** , η (4.22) γίνεται

$$\begin{aligned} f_Y(y) &= \frac{1}{B\left(\frac{2p}{2}, \frac{2q}{2}\right)} \left(\frac{2p}{2q}\right)^{\frac{2p}{2}} (\exp(-y))^{\frac{2p}{2}-1} \left(1 + \frac{2p}{2q} \exp(-y)\right)^{-\frac{2p+2q}{2}} \exp(-y) \\ &= \frac{1}{B(p, q)} \left(\frac{p}{q}\right)^p (\exp(-y))^p \left(1 + \frac{p}{q} \exp(-y)\right)^{-(p+q)} (\exp(-y))^{-1} \exp(-y) \\ &= \frac{1}{B(p, q)} \left(\frac{p}{q}\right)^p \left(\frac{1}{\exp(y)}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^q \\ &= \frac{1}{B(p, q)} \left(\frac{p}{q \cdot \exp(y)}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^q \\ &= \frac{1}{B(p, q)} \left(\frac{p \exp(-y)}{q}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^q \\ &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{p \exp(-y)}{q}\right)^p \left(\frac{1}{1 + \frac{p}{q} \exp(-y)}\right)^q \end{aligned} \quad (4.23)$$

Για $p = q$ η (4.23) γίνεται

$$f_Y(y) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{\exp(-y)}{1+\exp(-y)} \right)^q \left(\frac{1}{1+\exp(-y)} \right)^p$$

Η τελευταία μορφή είναι ο τύπος (4.6) της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής **Τύπου IV**.

Τέλος, έστω μια συνεχής τυχαία μεταβλητή X που ακολουθεί την κατανομή t -Student με k βαθμούς ελευθερίας. Η τυχαία μεταβλητή $Y = \log\left(\frac{X^2}{k}\right)$ θα έχει τη γενικευμένη λογιστική κατανομή **Τύπου IV** με παραμέτρους $p = \frac{1}{2}$ και $q = \frac{k}{2}$.

Υπενθυμίζεται ότι μια τυχαία μεταβλητή Z η οποία ακολουθεί την τυποποιημένη κανονική κατανομή και S_ν μια τυχαία μεταβλητή ανεξάρτητη από τη Z η οποία ακολουθεί την κατανομή χ_ν^2 . Η τυχαία μεταβλητή T_ν που ορίζεται από τον τύπο

$$T_\nu = \frac{Z}{\sqrt{\frac{S_\nu}{\nu}}}$$

θα λέμε ότι ακολουθεί την κατανομή t ή την Student με ν βαθμούς ελευθερίας (συμβολικά: t_ν). Με χ_ν^2 συμβολίσαμε την κατανομή χι τετράγωνο. Η συνάρτηση πυκνότητας για την T_ν θα δίνεται από τον τύπο

$$f_{T_\nu}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}}, \quad -\infty < x < \infty.$$

Έχουμε

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(\log\left(\frac{X^2}{k}\right) \leq y\right) = P\left(\frac{X^2}{k} \leq \exp(y)\right) \\ &= P\left(X \leq \pm\sqrt{k} \cdot \exp\left(\frac{y}{2}\right)\right) \end{aligned}$$

δηλαδή

$$F_Y(y) = F_X\left(\pm\sqrt{k} \cdot \exp\left(\frac{y}{2}\right)\right) \quad (4.24)$$

Θεωρούμε $g_1(y) = \sqrt{k} \cdot \exp\left(\frac{y}{2}\right)$ και $g_2(y) = -\sqrt{k} \cdot \exp\left(\frac{y}{2}\right)$. Η πρώτη παράγωγος ισούται

$$\frac{d[g_i(y)]}{dy} = \pm \frac{1}{2} \sqrt{k} \cdot \exp\left(\frac{y}{2}\right)$$

Οπότε παραγωγίζοντας την (4.24) ως προς y έχουμε

$$\begin{aligned} f_Y(y) = F'_Y(y) &= \sum_{i=1}^2 \left| \frac{d[g_i(y)]}{dy} \right| \cdot f_X\left(\pm\sqrt{k} \cdot \exp\left(\frac{y}{2}\right)\right) \\ &= 2 \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k} \cdot \Gamma\left(\frac{k}{2}\right)} \cdot \frac{1}{\left(1 + \frac{\left(\pm\sqrt{k} \cdot \exp\left(\frac{y}{2}\right)\right)^2}{k}\right)^{\frac{k+1}{2}}} \cdot \frac{1}{2} \sqrt{k} \cdot \exp\left(\frac{y}{2}\right) \\ &= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{k}{2}\right)} \cdot \frac{1}{\left(1 + \frac{k \cdot \exp(y)}{k}\right)^{\frac{k+1}{2}}} \exp\left(\frac{y}{2}\right) \\ &= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{k}{2}\right)} \frac{1}{(1 + \exp(y))^{\frac{k}{2}}} \frac{1}{(1 + \exp(y))^{\frac{1}{2}}} [\exp(y)]^{\frac{1}{2}} \\ &= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{1 + \exp(y)}\right)^{\frac{k}{2}} \left(\frac{\exp(y)}{1 + \exp(y)}\right)^{\frac{1}{2}} \end{aligned} \quad (4.25)$$

Η σχέση (4.25) εύκολα έρχεται στη μορφή

$$f_Y(y) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{k}{2}\right)} \left(\frac{\exp(-y)}{1+\exp(-y)}\right)^{\frac{k}{2}} \left(\frac{1}{1+\exp(-y)}\right)^{\frac{1}{2}} = \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} [1-F^*(y)]^q [F^*(y)]^p$$

όπου $F^*(\cdot)$ η συνάρτηση κατανομής της λογιστικής $L\left(0, \frac{\pi^2}{3}\right)$. Η τελευταία μορφή είναι ο τύπος (4.6) της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής **Τύπου IV** με παραμέτρους $p = \frac{1}{2}$ και $q = \frac{k}{2}$.

4.4 Εισαγωγή του γενικευμένου λογιστικού μοντέλου με έξι παραμέτρους (Ροπές, χαρακτηριστικά και ιδιότητες)

Η αυξημένη χρήση γενικευμένων μοντέλων λόγω της ελαστικότητας και ευκολότερης προσαρμογής τους σε ανάλυση δεδομένων, οδήγησε πολλούς ερευνητές στο να μελετήσουν τις ιδιότητες και τις εφαρμογές τους. Μεγάλη αναφορά έχει γίνει σε βιβλιογραφίες στη μελέτη των χαρακτηριστικών και εφαρμογών των τεσσάρων γενικευμένων μοντέλων που προαναφέρθηκαν.

Ο Wu Jong-Wuu et al (2000) προχώρησε κάνοντας μια επέκταση του πιο γενικού μοντέλου από τα τέσσερα, δηλαδή του γενικευμένου λογιστικού μοντέλου Τύπου IV με τέσσερις παραμέτρους. Ανέπτυξε μια μέθοδο και εξήγαγε γενικευμένο λογιστικό μοντέλο με πέντε παραμέτρους.

Παρακάτω παρουσιάζεται η ανάπτυξη ενός πιο σύνθετου μοντέλου με έξι παραμέτρους, του γενικευμένου λογιστικού μοντέλου με έξι παραμέτρους.

4.4.1 Γενικευμένο λογιστικό μοντέλο με έξι παραμέτρους

Χωρίς βλάβη της γενικότητας, όπως και στα προηγούμενα τέσσερα γενικευμένα μοντέλα έτσι και εδώ θα υποθέσουμε τις τυποποιημένες μορφές για όλες τις κατανομές. Δηλαδή, θα υποθέσουμε ότι η παράμετρο θέσης και κλίμακας παίρνουν την τιμή 0 και 1 αντίστοιχα.

Ας θεωρήσουμε μια τυχαία μεταβλητή X που ακολουθεί τη γενικευμένη Βήτα Τύπου II κατανομή με παραμέτρους λ, p και q . Η συνάρτηση κατανομής της δίνεται από τον τύπο

$$f(x; 0, 1, \lambda, p, q) = \frac{\lambda^q}{B(p, q)} \frac{x^{p-1}}{(\lambda + x)^{p+q}}, \quad 0 < x < \infty, \quad \lambda > 0, \quad p > 0, \quad q > 0, \quad (4.26)$$

Τότε, η τυχαία μεταβλητή $Y = \log\left(\frac{X}{\beta}\right)$ θα λέμε ότι ακολουθεί τη γενικευμένη λογιστική κατανομή με έξι παραμέτρους $(0, 1, \lambda, \beta, p, q)$.

Αναλυτικά έχουμε

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(\log\left(\frac{X}{\beta}\right) \leq y\right) = P\left(\frac{X}{\beta} \leq \exp(y)\right) \\ &= P(X \leq \beta \exp(y)) = F(\beta \exp(y)) \end{aligned}$$

Παραγωγίζοντας τη συνάρτηση κατανομής ως προς y προκύπτει

$$\begin{aligned} f_Y(y) &= F'_Y(y) = (F_Y(\beta \exp(y)))' = F'(\beta \exp(y))(\beta \exp(y))' \\ &= f(\beta \exp(y))\beta \exp(y) \end{aligned} \quad (4.27)$$

Χρησιμοποιώντας τη συνάρτηση πυκνότητας που δίνεται στο τύπου (4.26), η (4.27) γίνεται

$$\begin{aligned} f_Y(y) &= \frac{\lambda^q}{B(p, q)} \frac{[\beta \exp(y)]^{p-1}}{(\lambda + \beta \exp(y))^{p+q}} \cdot \beta \exp(y) \\ &= \frac{\lambda^q \beta^p}{B(p, q)} \frac{[\exp(y)]^p}{(\lambda + \beta \exp(y))^{p+q}} \beta^{-1} \beta \exp(y) [\exp(y)]^{-1} \\ &= \frac{\lambda^q \beta^p}{B(p, q)} \frac{\exp(py)}{(\lambda + \beta \exp(y))^{p+q}} \end{aligned} \quad (4.28)$$

Αν θέσουμε στη σχέση (4.28) την παράμετρο $\beta = 1$, τότε προκύπτει η συνάρτηση πυκνότητας για το γενικευμένο λογιστικό μοντέλο του Wu Jong-Wuu et al (2000) με πέντε παραμέτρους. Ακόμη, αν θέσουμε $\beta = \lambda = 1$ στην (4.28) εύκολα καταλήγουμε στη συνάρτηση πυκνότητας για το γενικευμένο λογιστικό μοντέλο Τύπου IV

$$\begin{aligned}
f_Y(y) &= \frac{1}{B(p, q)} \frac{\exp(py)}{(1 + \exp(y))^{p+q}} = \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} \left(\frac{\exp(y)}{1 + \exp(y)} \right)^p \left(\frac{1}{1 + \exp(y)} \right)^q \\
&= \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} \left(\frac{1}{1 + \exp(-y)} \right)^p \left(\frac{\exp(-y)}{1 + \exp(-y)} \right)^q \\
&= \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} [F^*(y)]^p [1 - F^*(y)]^q
\end{aligned}$$

όπου $F^*(\cdot)$ η συνάρτηση κατανομής της λογιστικής $L\left(0, \frac{\pi^2}{3}\right)$.

Με τον ίδιο τρόπο που θέταμε τιμές στις παραμέτρους p και q για το γενικευμένο λογιστικό μοντέλο Τύπου IV, έτσι ώστε να οδηγηθούμε στα μικρότερα μοντέλα με λιγότερες παραμέτρους (Τύπου III, Τύπου II και Τύπου I), έτσι και εδώ θέτουμε ίδιες τιμές. Τέλος, θέτοντας $(\lambda, \beta, p, q) = (1, 1, 1, 1)$ οδηγούμαστε στο κλασικό μοντέλο

$$L\left(0, \frac{\pi^2}{3}\right).$$

4.4.2 Σχέση της γενικευμένης λογιστικής κατανομής με έξι παραμέτρους με άλλες κατανομές

Η ροπογεννήτρια συνάρτηση για τη γενικευμένη λογιστική συνάρτηση με έξι παραμέτρους είναι

$$M_Y(t) = E(e^{ty}) = \lambda' \beta^{-t} \frac{\Gamma(p+t)}{\Gamma(p)} \frac{\Gamma(q-t)}{\Gamma(q)}. \quad (4.29)$$

Έστω X μια τυχαία μεταβλητή που ακολουθεί την κατανομή Βήτα με παραμέτρους p και q . Ορίζουμε την τυχαία μεταβλητή $Y = \log\left(\frac{\lambda X}{\beta(1-X)}\right)$. Αποδεικνύεται ότι η Y ακολουθεί τη γενικευμένη λογιστική κατανομή με έξι παραμέτρους $(0, 1, \lambda, \beta, p, q)$.

Αναλυτικά έχουμε

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P\left(\log\left(\frac{\lambda X}{\beta(1-X)}\right) \leq y\right) = P\left(\frac{\lambda X}{\beta(1-X)} \leq \exp(y)\right) \\
&= P\left(X \leq \frac{\beta \exp(y)}{\lambda + \beta \exp(y)}\right) = F\left(\frac{\beta \exp(y)}{\lambda + \beta \exp(y)}\right)
\end{aligned}$$

Παραγωγίζοντας τη συνάρτηση κατανομής ως προς y προκύπτει

$$\begin{aligned} f_Y(y) &= F'_Y(y) = \left(F_Y \left(\frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right) \right)' = F' \left(\frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right) \left(\frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right)' \\ &= f \left(\frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right) \frac{\lambda \beta \exp(y)}{(\lambda + \beta \exp(y))^2} \end{aligned} \quad (4.30)$$

Σύμφωνα με τη συνάρτηση πυκνότητας της κατανομής Βήτα η (4.30) γίνεται

$$\begin{aligned} f_Y(y) &= \frac{1}{B(p, q)} \left[\frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right]^{p-1} \left[1 - \frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right]^{q-1} \frac{\lambda \beta \exp(y)}{(\lambda + \beta \exp(y))^2} \\ &= \frac{1}{B(p, q)} \left[\frac{\beta \exp(y)}{\lambda + \beta \exp(y)} \right]^{p-1} \left[\frac{\lambda}{\lambda + \beta \exp(y)} \right]^{q-1} \frac{\lambda \beta \exp(y)}{(\lambda + \beta \exp(y))^2} \\ &= \frac{1}{B(p, q)} [\beta \exp(y)]^{p-1} \lambda^{q-1} [\lambda + \beta \exp(y)]^{-q-p+2} [\lambda + \beta \exp(y)]^{-2} \lambda \beta \exp(y) \\ &= \frac{\lambda^q}{B(p, q)} \frac{[\beta \exp(y)]^p}{[\lambda + \beta \exp(y)]^{p+q}} = \frac{\lambda^q \beta^p}{B(p, q)} \frac{\exp(py)}{[\lambda + \beta \exp(y)]^{p+q}} \end{aligned} \quad (4.31)$$

Ο τύπος (4.31) αποτελεί τη συνάρτηση πυκνότητας της γενικευμένης λογιστικής κατανομής με έξι παραμέτρους $(0, 1, \lambda, \beta, p, q)$.

Έστω X_1 και X_2 δυο τυχαίες μεταβλητές που ακολουθούν την γενικευμένη Gumbel κατανομή με συναρτήσεις πυκνότητας

$$g_1(x_1) = \frac{\lambda^{-p}}{\Gamma(p)} \exp(px_1) \exp\left(-\frac{\exp(x_1)}{\lambda}\right)$$

και

$$g_2(x_2) = \frac{\beta^{-q}}{\Gamma(q)} \exp(qx_2) \exp\left(-\frac{\exp(x_2)}{\beta}\right)$$

αντίστοιχα.

Αν οι X_1 και X_2 είναι ανεξάρτητες τότε η τυχαία μεταβλητή $Y = X_1 - X_2$ ακολουθεί τη γενικευμένη λογιστική κατανομή με έξι παραμέτρους $(0, 1, \lambda, \beta, p, q)$.

Η ροπογεννήτρια συνάρτηση για τη X_1 δίνεται από τον τύπο

$$\begin{aligned} M_{X_1}(t) &= \int_{-\infty}^{\infty} \exp(t x_1) g_1(x_1) dx = \int_{-\infty}^{\infty} \exp(t x_1) g_1(x_1) dx_1 \\ &= \int_{-\infty}^{\infty} \exp(t x_1) \frac{\lambda^{-p}}{\Gamma(p)} \exp(p x_1) \exp\left(-\frac{\exp(x_1)}{\lambda}\right) dx_1 \\ &= \frac{\lambda^{-p}}{\Gamma(p)} \int_{-\infty}^{\infty} \exp((p+t)x_1) \exp\left(-\frac{\exp(x_1)}{\lambda}\right) dx_1 \end{aligned} \quad (4.32)$$

Κάνουμε αλλαγή μεταβλητής θέτοντας $u = \frac{\exp(x_1)}{\lambda}$ και η (4.32) γίνεται

$$\begin{aligned} M_{X_1}(t) &= \frac{\lambda^{-p}}{\Gamma(p)} \int_0^{\infty} (\lambda u)^{p+t} \exp(-u) \frac{1}{u} du \\ &= \frac{\lambda^{-p} \lambda^{p+t}}{\Gamma(p)} \int_0^{\infty} u^{p+t-1} \exp(-u) du \\ &= \frac{\lambda^t \Gamma(p+t)}{\Gamma(p)} \end{aligned}$$

Ομοίως για τη $-X_2$ έχουμε

$$M_{-X_2}(t) = \frac{\beta^{-t} \Gamma(q-t)}{\Gamma(q)}$$

Έστω X_1, X_2, \dots, X_ν ανεξάρτητες τυχαίες μεταβλητές με αντίστοιχες ροπογεννήτριες

$M_{X_1}(t), M_{X_2}(t), \dots, M_{X_\nu}(t)$. Τότε η ροπογεννήτρια του αθροίσματος

$$S_\nu = X_1 + X_2 + \dots + X_\nu$$

δίνεται από τον τύπο

$$M_{S_\nu}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \dots \cdot M_{X_\nu}(t) = \prod_{i=1}^{\nu} M_{X_i}(t).$$

Σύμφωνα με την παραπάνω πρόταση έχουμε για τη $Y = X_1 - X_2$

$$M_Y(t) = M_{X_1}(t) \cdot M_{-X_2}(t) = \frac{\lambda^t \Gamma(p+t)}{\Gamma(p)} \cdot \frac{\beta^{-t} \Gamma(q-t)}{\Gamma(q)} \quad (4.33)$$

Η σχέση (4.33) είναι η (4.29), δηλαδή η ροπογεννήτρια συνάρτηση για τη γενικευμένη λογιστική κατανομή με έξι παραμέτρους.

Τέλος, έστω X μια τυχαία μεταβλητή που ακολουθεί την **κατανομή F** με $(2p, 2q)$

βαθμούς ελευθερίας. Θεωρούμε την τυχαία μεταβλητή $Y = \log\left(\frac{\lambda \cdot p}{\beta \cdot q} X\right)$,

με $\lambda, \beta > 0$. Η τυχαία μεταβλητή Y θα έχει τη γενικευμένη λογιστική κατανομή με έξι παραμέτρους. Αναλυτικά έχουμε

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(\log\left(\frac{\lambda \cdot p}{\beta \cdot q} X\right) \leq y\right) = P\left(\frac{\lambda \cdot p}{\beta \cdot q} X \leq \exp(y)\right) \\ &= P\left(X \leq \frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right) \end{aligned}$$

δηλαδή

$$F_Y(y) = F\left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)$$

και παραγωγίζοντας ως προς y βρίσκουμε

$$\begin{aligned} f_Y(y) &= F'_Y(y) = \left(F\left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)\right)' = F'\left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right) \left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)' \\ &= f\left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right) \frac{\beta \cdot q}{\lambda \cdot p} \exp(y) \end{aligned} \quad (4.34)$$

Σύμφωνα με τη συνάρτηση πυκνότητας της **κατανομής F** , η (4.34) γίνεται

$$\begin{aligned} f_Y(y) &= \frac{1}{B\left(\frac{2p}{2}, \frac{2q}{2}\right)} \left(\frac{2p}{2q}\right)^{\frac{2p}{2}} \left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)^{\frac{2p}{2}-1} \left(1 + \frac{2p}{2q} \frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)^{-\frac{2p+2q}{2}} \frac{\beta \cdot q}{\lambda \cdot p} \exp(y) \\ &= \frac{1}{B(p, q)} \left(\frac{p}{q}\right)^p \left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)^{p-1} \left(1 + \frac{2p}{2q} \frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)^{-(p+q)} \\ &= \frac{1}{B(p, q)} \left(\frac{p}{q}\right)^p \left(\frac{\beta \cdot q}{\lambda \cdot p} \exp(y)\right)^{p-1} \left(1 + \frac{\beta}{\lambda} \exp(y)\right)^{-(p+q)} \\ &= \frac{1}{B(p, q)} \left(\frac{p}{q}\right)^p \beta^p \left(\frac{q}{p}\right)^p \left(\frac{1}{\lambda}\right)^p \exp(py) \left(1 + \frac{\beta}{\lambda} \exp(y)\right)^{-(p+q)} \\ &= \frac{\beta^p}{B(p, q)} \left(\frac{1}{\lambda}\right)^p \exp(py) \left(1 + \frac{\beta}{\lambda} \exp(y)\right)^{-(p+q)} \end{aligned} \quad (4.35)$$

Εφόσον $\lambda > 0$ πολλαπλασιάζουμε τη σχέση (4.35) με $\lambda^q \cdot \lambda^{-q}$ και έχουμε

$$\begin{aligned}
 f_Y(y) &= \frac{\beta^p}{B(p,q)} \left(\frac{1}{\lambda}\right)^p \exp(py) \left(1 + \frac{\beta}{\lambda} \exp(y)\right)^{-(p+q)} \lambda^q \cdot \lambda^{-q} \\
 &= \frac{\lambda^q \beta^p}{B(p,q)} \exp(py) \left(\frac{\frac{1}{\lambda}}{1 + \frac{\beta}{\lambda} \exp(y)}\right)^{p+q} \\
 &= \frac{\lambda^q \beta^p}{B(p,q)} \frac{\exp(py)}{[\lambda + \beta \exp(y)]^{p+q}}
 \end{aligned} \tag{4.36}$$

Η σχέση (4.36) αποτελεί έκφραση της συνάρτησης πυκνότητας της γενικευμένης λογιστικής κατανομής με έξι παραμέτρους.

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 5^ο

Χρήση της Λογιστικής Κατανομής για την προσαρμογή πραγματικών δεδομένων

5.1 Εκτίμηση παραμέτρων

Στο κεφάλαιο αυτό θα εφαρμόσουμε τις μεθόδους που αναλύθηκαν στο 3^ο Κεφάλαιο με στόχο να εκτιμήσουμε τις παραμέτρους της διπαραμετρικής λογιστικής κατανομής από την οποία προέρχονται τα δεδομένα μας. Τα δεδομένα που θα χρησιμοποιήσουμε προέρχονται από μετρήσεις που έγιναν όσον αφορά το χρόνο ζωής σε ώρες, σε λαμπτήρες πυρακτώσεως των 40 Watt (Davis (1952)). Όπως θα διαπιστώσουμε στη συνέχεια η λογιστική κατανομή είναι η κατανομή που προσαρμόζεται καλύτερα από κάθε άλλη κατανομή στις συγκεκριμένες μετρήσεις.

Στον πίνακα 5.1 δίνονται τα δεδομένα από εκατό μετρήσεις.

Πίνακας 5.1 Χρόνοι ζωής σε ώρες εκατό λαμπτήρων πυρακτώσεως

Χρονολογικά Δεδομένα							
	Χρόνος Ζωής Λαμπτήρων (ώρες)		Χρόνος Ζωής Λαμπτήρων (ώρες)		Χρόνος Ζωής Λαμπτήρων (ώρες)		Χρόνος Ζωής Λαμπτήρων (ώρες)
Παρατήρηση		Παρατήρηση		Παρατήρηση		Παρατήρηση	
1	1067	26	938	51	830	76	1250
2	919	27	970	52	1063	77	1203
3	1196	28	1237	53	930	78	1078
4	785	29	956	54	807	79	890
5	1126	30	1102	55	954	80	1303
6	936	31	1022	56	1063	81	1011
7	918	32	978	57	1002	82	1102
8	1156	33	832	58	909	83	996
9	920	34	1009	59	1077	84	780
10	948	35	1157	60	1021	85	900
11	855	36	1151	61	1062	86	1106
12	1092	37	1009	62	1157	87	704
13	1162	38	765	63	999	88	621
14	1170	39	958	64	932	89	854
15	929	40	1311	65	1035	90	1178
16	950	41	1037	66	944	91	1138
17	905	42	702	67	1049	92	951
18	972	43	521	68	940	93	902
19	1035	44	933	69	1122	94	923
20	1045	45	928	70	1115	95	1333
21	1157	46	1153	71	833	96	811
22	1195	47	946	72	1320	97	1217
23	1195	48	858	73	901	98	1085
24	1340	49	1071	74	1324	99	896
25	1122	50	1069	75	818	100	958

α. Μέθοδος Μέγιστης Πιθανοφάνειας

Εισάγουμε τα εκατό δεδομένα του Πίνακα 5.1 σε μορφή λίστας στο Mathematica. Ονομάζουμε LifeDataSet τη λίστα που περιέχει τα δεδομένα μας. Δημιουργούμε τις εξισώσεις (3.11) και (3.13) και θέτουμε τις αρχικές τιμές των παραμέτρων. Μέσω της

επαναληπτικής μεθόδου υπολογίζεται ότι οι εκτιμητές μέγιστης πιθανοφάνειας για τις παραμέτρους μ και σ είναι:

$$\hat{\mu} = 1010.79 \quad \text{και} \quad \hat{\sigma} = 160.712$$

Οι εντολές που χρησιμοποιήθηκαν είναι οι εξής:

```
(*Λίστα δεδομένων*)
LifeDataSet = {1067, 919, 1196, 785, 1126, 936, 918, 1156, 920, 948, 855,
  1092, 1162, 1170, 929, 950, 905, 972, 1035, 1045, 1157, 1195, 1195, 1340,
  1122, 938, 970, 1237, 956, 1102, 1022, 978, 832, 1009, 1157, 1151, 1009,
  765, 958, 1311, 1037, 702, 521, 933, 928, 1153, 946, 858, 1071, 1069, 830,
  1063, 930, 807, 954, 1063, 1002, 909, 1077, 1021, 1062, 1157, 999, 932, 1035,
  944, 1049, 940, 1122, 1115, 833, 1320, 901, 1324, 818, 1250, 1203, 1078, 890,
  1303, 1011, 1102, 996, 780, 900, 1106, 704, 621, 854, 1178, 1138, 951, 902,
  923, 1333, 811, 1217, 1085, 896, 958};

(*Εκτιμητές Μέγιστης Πιθανοφάνειας*)
EMPI = FindRoot[{Sum[1 / (1 + Exp[(Pi * (LifeDataSet[[i]] - μ) / (σ * Sqrt[3]))],
  {i, 1, Length[LifeDataSet]}] == 50,
  Sum[((1 - Exp[(Pi * (LifeDataSet[[i]] - μ) /
    (σ * Sqrt[3]))] / (1 + Exp[(Pi * (LifeDataSet[[i]] - μ) /
    (σ * Sqrt[3]))])) * (LifeDataSet[[i]] - μ) / σ,
  {i, 1, Length[LifeDataSet]}] == -(100 * Sqrt[3]) / Pi},
  {{μ, 100}, {σ, 100}}];
(*Εκτύπωση αποτελεσμάτων*)
Print[EMPI];
```

β. Μέθοδος των ροπών

Εισάγουμε τα δεδομένα με μορφή λίστας στο Mathematica όπως προηγουμένως. Υπολογίζουμε τον δειγματικό μέσο

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{100} x_i}{100} = 1011.05$$

και τη δειγματική διασπορά

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{100} (x_i - \bar{x})^2}{99} = 24682.1$$

Επομένως

$$\hat{\mu} = 1011.05 \text{ και } \hat{\sigma} = 157.106$$

Οι εντολές που χρησιμοποιήθηκαν είναι οι εξής:

```
(* Υπολογισμός δειγματικού μέσου *)
ΔειγματικόςΜέσος = Sum[LifeDataSet[[i]],
  {i, 1, Length[LifeDataSet]}] / Length[LifeDataSet];
(* Υπολογισμός δειγματικής διασποράς *)
ΔειγματικήΔιασπορά = Sum[(LifeDataSet[[i]] - ΔειγματικόςΜέσος)^2,
  {i, 1, Length[LifeDataSet]}] / (Length[LifeDataSet] - 1);
(* Εκτύπωση αποτελεσμάτων για εκτιμητές *)
Print["Η εκτιμήτρια ροπών της παραμέτρου α είναι ίση με "
, ΔειγματικόςΜέσος];
Print["Η εκτιμήτρια ροπών της παραμέτρου β είναι ίση με "
, Sqrt[3 * (ΔειγματικήΔιασπορά / (Pi^2))]];

```

γ. Γραφική μέθοδος

Εφόσον κατασκευάουμε το LPP γράφημα κάνουμε τις εξής ενέργειες

- Προσαρμόζουμε στο γράφημα την ευθεία παλινδρόμησης χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων (*ordinary least squares*)
- Η κλίση της ευθείας παλινδρόμησης δίνει την εκτίμηση $\hat{\sigma}$ για την παράμετρο κλίμακας σ .
- Αφού εκτιμήσουμε την παράμετρο κλίμακας, υπολογίζουμε και την εκτίμηση της παραμέτρου θέσης $\hat{\mu}$.

Η γενική λύση για την εξίσωση $x = \mu + \sigma y$ είναι

$$\hat{\sigma} = \frac{\sum_{i=1}^n (y_i - \bar{y})x_i}{\sum_{i=1}^n (y_i - \bar{y})^2} \text{ και } \hat{\mu} = \bar{x} - \hat{\sigma} \bar{y}.$$

Αφού έχουν εισαχθεί τα δεδομένα του Πίνακα 5.1 σε μορφή λίστας, το επόμενο βήμα είναι να τα διατάξουμε

{521, 621, 702, 704, 765, 780, 785, 807, 811, 818, 830, 832, 833, 854, 855, 858, 890, 896, 900, 901, 902, 905, 909, 918, 919, 920, 923, 928, 929, 930, 932, 933, 936, 938, 940, 944, 946, 948, 950, 951, 954, 956, 958, 958, 970, 972, 978, 996, 999, 1002, 1009, 1009, 1011, 1021, 1022, 1035, 1035, 1037, 1045, 1049, 1062, 1063, 1063, 1067, 1069, 1071, 1077, 1078, 1085, 1092, 1102, 1102, 1106, 1115, 1122, 1122, 1126, 1138, 1151, 1153, 1156, 1157, 1157, 1157, 1162, 1170, 1178, 1195, 1195, 1196, 1203, 1217, 1237, 1250, 1303, 1311, 1320, 1324, 1333, 1340}

Έπειτα υπολογίζουμε την εμπειρική συνάρτηση κατανομής σύμφωνα με τον τύπο

$$\hat{F}(d_{(i)}) = \frac{i - 0.5}{n}, \quad i = 1, \dots, 100 \text{ και } n = 100.$$

Στη συνέχεια υπολογίζουμε τις τιμές των ποσοτήτων

$$y_i = \frac{\sqrt{3}}{\pi} \ln \frac{\hat{F}(d_{(i)})}{1 - \hat{F}(d_{(i)})} \text{ για } 1 \leq i \leq 100.$$

Οι εκτιμήσεις των παραμέτρων θα γίνουν μέσω του SPSS με τη μέθοδο των ελαχίστων τετραγώνων (*OLS*).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.992 ^a	.984	.984	20.092

a. Predictors: (Constant), EskList

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2403969.846	1	2403969.846	5954.797	.000 ^a
	Residual	39562.904	98	403.703		
	Total	2443532.750	99			

a. Predictors: (Constant), EskList

b. Dependent Variable: OrderList

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1011.050	2.009		503.201	.000
	EskList	157.317	2.039	.992	77.167	.000

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1011.050	2.009		503.201	.000
	EskList	157.317	2.039	.992	77.167	.000

a. Dependent Variable: OrderList

Από τα αποτελέσματα του παραπάνω πίνακα **Model Summary** παρατηρούμε ότι ο συντελεστής προσδιορισμού (*coefficient of determination*) $R^2 = 0.984$ είναι αρκετά κοντά στη μονάδα (η μεταβλητότητα λόγω της παλινδρόμησης είναι αυξημένη σε σχέση με την ανεξάρτητη μεταβλητότητα), δηλαδή το μοντέλο της λογιστικής κατανομής προσαρμόζεται αρκετά καλά στα δεδομένα μας. Βλέποντας τα αποτελέσματα του πίνακα **Coefficients**, τα p-values (τελευταία στήλη Sig.) είναι ίσα με μηδέν. Άρα η μηδενική υπόθεση που ισχυρίζεται ότι οι συντελεστές παλινδρόμησης είναι μηδενικοί, απορρίπτεται.

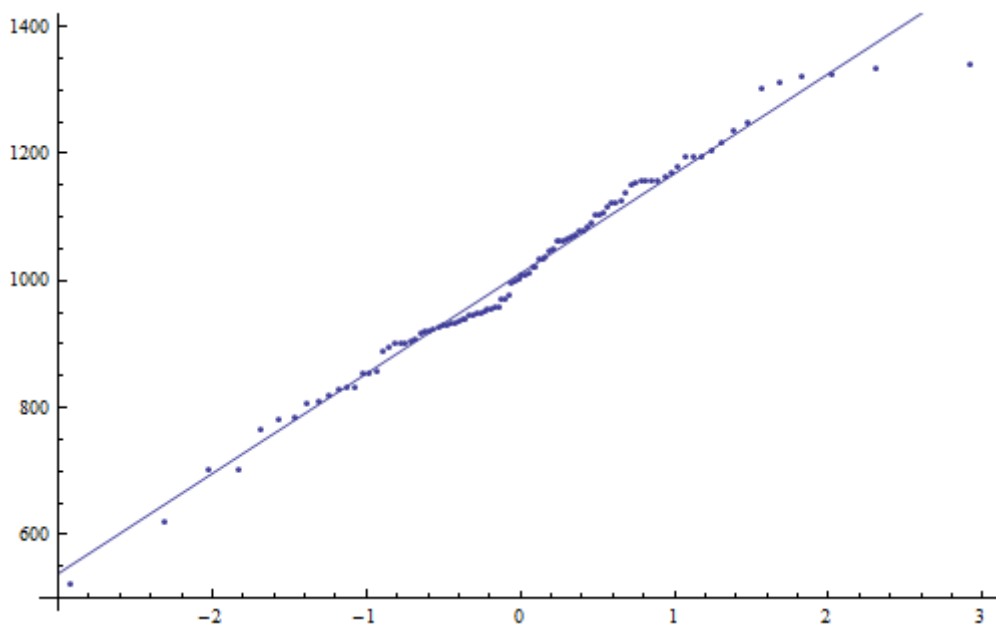
Η ευθεία της παλινδρόμησης έχει την παρακάτω μορφή

$$f(x) = 1011.05 + 157.317 \cdot x.$$

Η κλίση της ευθείας παλινδρόμησης είναι ίση με 157.317 ενώ ο σταθερός όρος είναι ίσος με 1011.05. Άρα η εκτίμηση της παραμέτρου κλίμακας ισούται με $\hat{\sigma} = 157.317$ και η εκτίμηση της παραμέτρου θέσης με $\hat{\mu} = 1011.05$.

Παρακάτω δίνεται το γράφημα διασποράς και της προσαρμοσμένης ευθείας παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων.

Σχήμα 5.1 Γράφημα διασποράς και της προσαρμοσμένης ευθείας παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων που εφαρμόστηκε στα δεδομένα του Πίνακα 5.1



Οι εντολές που χρησιμοποιήθηκαν είναι οι εξής:

```
(*Διατάσσουμε τα δεδομένα της λίστας*)
orderList = Sort[LifeDataSet]; xlist = {};
(*Υπολογίζουμε την εμπειρική συνάρτηση κατανομής για τα δεδομένα*)
For[i = 1, i ≤ Length[orderList], i++, AppendTo[xlist, (Sqrt[3] / Pi) *
  Log[((i - 0.5) / Length[LifeDataSet]) / (1 - ((i - 0.5) / Length[LifeDataSet]))]]]
(*Τοποθετούμε σε πίνακα τις λίστες σε μορφή (x,y) δεδομένα*)
data = Table[{xlist[[i]], orderList[[i]]}, {i, Length[orderList]};
(*Εφαρμόζουμε Μέθοδο Παλινδρόμησης*)
Needs["LinearRegression`"]
line = Fit[data, {1, x}, x]
Regress[data, {1, x}, x]
(*Δημιουργούμε διάγραμμα διασποράς*)
Show[ListPlot[data], Plot[{line}, {x, -3, 3}], PlotRange → {{-3, 3},
  {500, 1400}}, AxesOrigin → {-3, 500}]
```

5.2 Έλεγχος της ποιότητας προσαρμογής

Προκειμένου να ελέγξουμε αν τα δεδομένα μας προέρχονται από τη συγκεκριμένη κατανομή θα χρησιμοποιήσουμε μια γραφική και μια στατιστική μέθοδο. Η γραφική

μέθοδος αφορά τη σύγκριση των γραφημάτων της ΕΣΚ και της συνάρτησης κατανομής από την κατανομή που μας ενδιαφέρει.

α. Γράφημα Εμπειρικής Συνάρτησης Κατανομής

Για να κατασκευάσουμε το γράφημα της εμπειρικής συνάρτησης κατανομής ακολουθούμε τα παρακάτω βήματα:

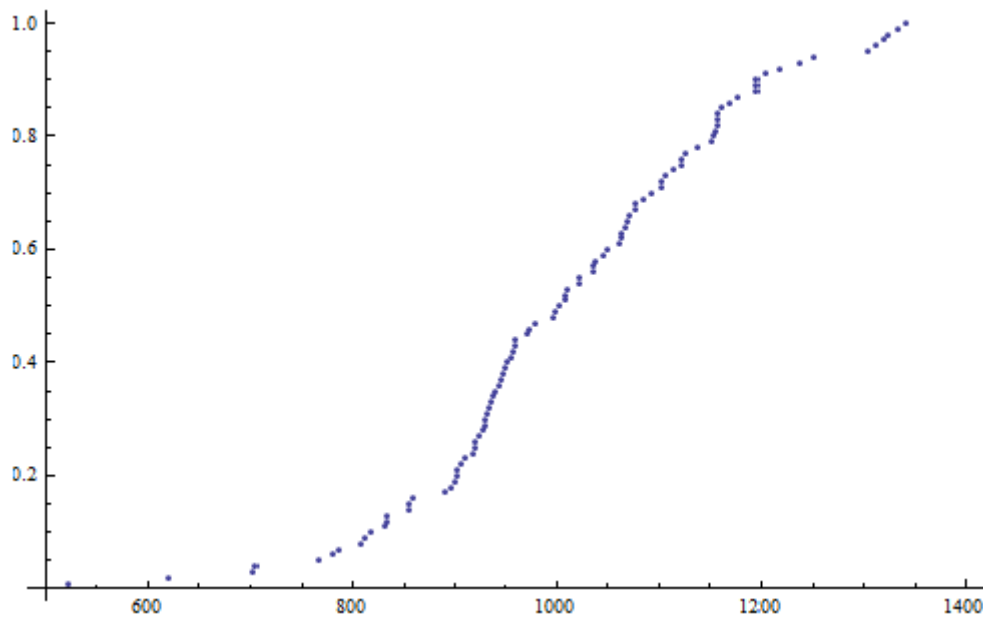
1. Υπολογίζουμε την εμπειρική συνάρτηση κατανομής σύμφωνα με τον τύπο της σχέσης (3.19), δηλαδή

$$\hat{F}(d_{(i)}) = \frac{i}{n}, \quad i = 1, \dots, 100 \text{ και } n = 100.$$

2. Εισάγουμε τα δεδομένα του πίνακα 3.2 σε μορφή λίστας και τα διατάσσουμε.
3. Τέλος κατασκευάζουμε το διάγραμμα των $\hat{F}(d_{(i)})$ ως προς X_{in} για $1 \leq i \leq 100$. (Υπενθυμίζεται ότι με X_{in} συμβολίζουμε τα διατεταγμένα δεδομένα μας)

Παρακάτω δίνεται το τελευταίο γράφημα

Σχήμα 5.2 Γράφημα της εμπειρικής συνάρτησης κατανομής για τα δεδομένα του Πίνακα 5.1



Έπειτα κατασκευάζουμε το γράφημα της συνάρτησης κατανομής από τη λογιστική κατανομή με παραμέτρους τις τιμές που υπολογίστηκαν σύμφωνα με τους εκτιμητές μέγιστης πιθανοφάνειας, δηλαδή

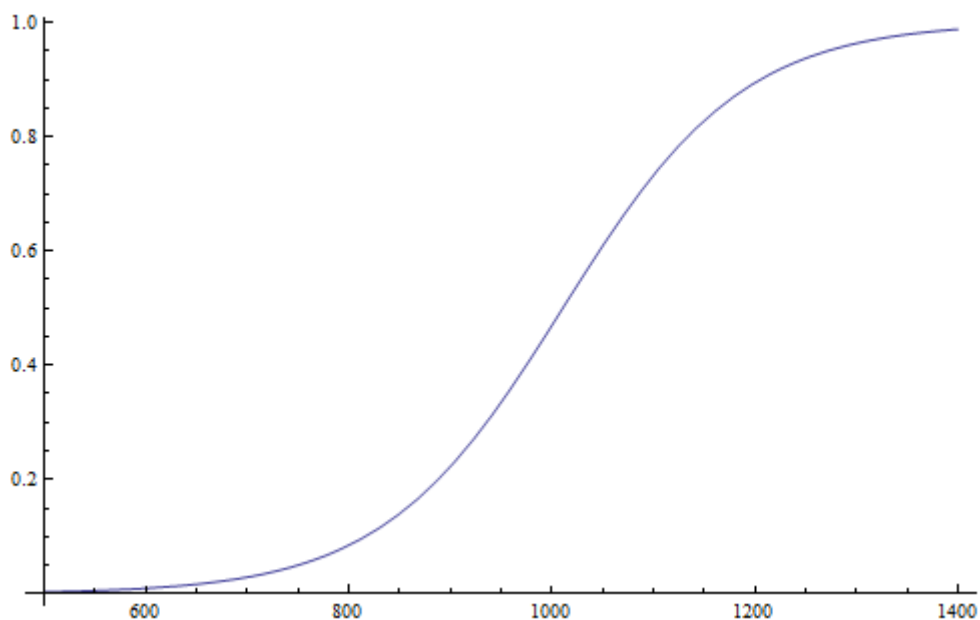
$$\hat{\mu} = 1010.79 \text{ και } \hat{\sigma} = 160.712$$

Η τύπος της συνάρτησης κατανομής θα είναι

$$F(x; \hat{\mu}, \hat{\sigma}) = \frac{1}{1 + \exp\left[-\frac{\pi(x - \hat{\mu})}{\hat{\sigma}\sqrt{3}}\right]} = \frac{1}{1 + \exp\left[-\frac{\pi(x - 1010.79)}{160.172 \cdot \sqrt{3}}\right]} \quad -\infty < x < \infty,$$

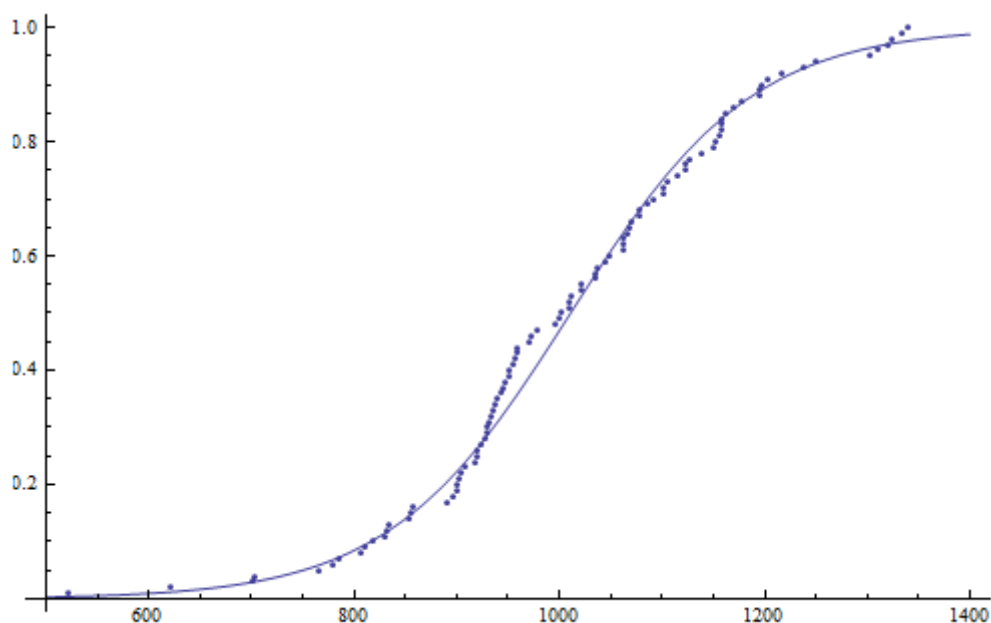
ενώ το γράφημα της δίνεται στο Σχήμα 5.3.

Σχήμα 5.3 Γράφημα της συνάρτησης κατανομής από τη λογιστική κατανομή με παραμέτρους όπως εκτιμήθηκαν (EMΠ) από τον Πίνακα 5.1



Στο Σχήμα 5.4 συγκρίνουμε τις τιμές από την εμπειρική συνάρτηση κατανομής (Σχήμα 5.2) με τις τιμές της συνάρτησης κατανομής (Σχήμα 5.3)

Σχήμα 5.4 Γράφημα της εμπειρικής συνάρτησης και της συνάρτησης κατανομής από τη λογιστική κατανομή με παραμέτρους όπως εκτιμήθηκαν (EMΠ) από τον Πίνακα 5.1



Οι εντολές που χρησιμοποιήθηκαν είναι οι εξής:

```
(*Γραφική μέθοδος για ΕΕΚ*)
z = Sort[myLamps]; edfList = {};
(*Υπολογίζουμε την Εμπειρική Συνάρτηση Κατανομής
 και τοποθετούμε τις τιμές σε μια λίστα*)
For[i = 1, i ≤ Length[z], i++, AppendTo[edfList, i / Length[z]]]
(*Βάζουμε σε πίνακα τις διατεταγμένες τιμές και τις τιμές της ΕΕΚ*)
newdata = Table[{z[[i]], edfList[[i]]}, {i, Length[z]}];
(*Γράφημα τιμών της ΕΕΚ*)
z = Show[ListPlot[newdata], PlotRange → {{500, 1400}, {0, 1}},
  AxesOrigin → {500, 0}]
(*Γράφημα Λογιστικής Κατανομής με παραμέτρους τους EMΠ*)
q = Plot[CDF[LogisticDistribution[1010.79, (160.712 * Sqrt[3]) / Pi], x]
, {x, 500, 1400}]
(*Προσαρμογή και των δύο*)
Show[z, q]
```

β. Έλεγχος καλής προσαρμογής στη λογιστική κατανομή (Κριτήριο Kolmogorov-Smirnov ή K-S)

Το κριτήριο K-S χρησιμοποιείται για τον έλεγχο καλής προσαρμογής ενός τυχαίου δείγματος σε μια συνεχή κατανομή ($H_0 : X_i \sim F_0$). Βασίζεται στην διαφορά της

εμπειρικής συνάρτησης κατανομής (που προέρχεται από το δείγμα) και της αναμενόμενης F_0 (υπό την H_0).

Έστω λοιπόν ένα τ.δ. X_1, X_2, \dots, X_m και έστω ότι επιθυμούμε να ελέγξουμε την υπόθεση $H_0 : X_i \sim F_0$. Η εμπειρική συνάρτηση κατανομής (ΕΣΚ) από το δείγμα υπολογίζεται σύμφωνα με με τον τύπο της σχέσης (3.19). Υπό την μηδενική υπόθεση θα πρέπει η ΕΣΚ να είναι «κοντά» στην F_0 . Αντίθετα, αν δεν ισχύει η H_0 αναμένουμε σημαντική απόκλιση της ΕΣΚ από την F_0 . Για να κατασκευαστεί ένας έλεγχος με βάση αυτό τον συλλογισμό, πρέπει να οριστεί μια «απόσταση» μεταξύ των δύο κατανομών (της ΕΣΚ και της F_0) και να απορρίπτουμε την H_0 όταν αυτή η απόσταση γίνεται «μεγάλη».

Ορισμός. Αν F, G είναι δύο συναρτήσεις κατανομής στον \mathbf{R} , τότε η ποσότητα

$$d_K(F, G) = \sup_{x \in \mathbf{R}} \{ |F(x) - G(x)| \} = \|F - G\|_\infty$$

καλείται απόσταση Kolmogorov μεταξύ της F και της G .

Σύμφωνα με τα παραπάνω, θα απορρίπτουμε την $H_0 : X_i \sim F_0$ όταν η στατιστική συνάρτηση

$$D = d_K(\hat{F}, F_0) = \sup_{x \in \mathbf{R}} \{ |\hat{F}(x) - F_0(x)| \},$$

λαμβάνει ασυνήθιστα μεγάλες τιμές, δηλαδή όταν $D > c$. Με \hat{F} συμβολίζουμε την ΕΣΚ. Προκειμένου να χρησιμοποιήσουμε το συγκεκριμένο κριτήριο στην πράξη, θα πρέπει πρώτα από όλα να προσδιοριστεί η κατανομή της τ.μ. D κάτω από την H_0 έτσι ώστε να υπολογίσουμε το c (για δεδομένο επίπεδο σημαντικότητας α) ή το p -value ενός δείγματος. Εφόσον η F_0 είναι αύξουσα συνάρτηση και η \hat{F} είναι σταθερή στα διαστήματα $[X_{i-1}, X_i)$ ενώ παρουσιάζει άλματα ύψους $1/m$ στα σημεία X_1, X_2, \dots, X_m η ποσότητα supremum της $\hat{F}(x) - F_0(x)$ ισούται

$$\sup_{x \in \mathbf{R}} \{ \hat{F}(x) - F_0(x) \} = \max_{i=1,2,\dots,m} \{ \hat{F}(X_{(i)}) - F_0(X_{(i)}) \} = \max_{i=1,2,\dots,m} \left\{ \frac{i}{m} - F_0(X_{(i)}) \right\} \geq 0.$$

Όμοια, το supremum της $F_0(x) - \hat{F}(x)$ ισούται

$$\sup_{x \in R} \{F_0(x) - \hat{F}(x)\} = \max_{i=1,2,\dots,m} \left\{ F_0(X_{(i)}) - \frac{i-1}{m} \right\} \geq 0.$$

Άρα η στατιστική συνάρτηση D γίνεται τελικά

$$\begin{aligned} D &= \sup_{x \in R} \left\{ \left| \hat{F}(x) - F_0(x) \right| \right\} = \max \left\{ \sup_{x \in R} \{ \hat{F}(x) - F_0(x) \}, \sup_{x \in R} \{ F_0(x) - \hat{F}(x) \} \right\} = \\ &= \max \left\{ \frac{i}{m} - F_0(X_{(i)}), F_0(X_{(i)}) - \frac{i-1}{m}, i = 1, 2, \dots, m \right\}. \end{aligned}$$

Η κατανομή της D , υπό την H_0 δεν εξαρτάται από την F_0 . Οι τ.μ $U_i = F_0(X_i), i = 1, 2, \dots, m$ είναι ανεξάρτητες και ακολουθούν την ομοιόμορφη κατανομή στο $(0,1)$. Συνεπώς, οποιαδήποτε και αν είναι η F_0 , η D έχει ίδια κατανομή με την τ.μ

$$D' = \max \left\{ \frac{i}{m} - U_i, U_i - \frac{i-1}{m}, i = 1, 2, \dots, m \right\}.$$

Επομένως η μηδενική υπόθεση θα απορρίπτεται όταν

$$D > D_m(\alpha),$$

όπου $D_m(\alpha)$ είναι το άνω α -σημείο της κατανομής της τ.μ D' που προέρχεται από δείγμα μεγέθους m . Η ακριβής κατανομή της τ.μ D' είναι δύσκολο να δοθεί και για αυτό έχουν κατασκευαστεί πίνακες με τα άνω α -σημεία της. Το p -value ενός δείγματος που έδωσε $D = d$ θα είναι

$$\Pr(D \geq d | H_0) = \Pr(D' \geq d) = \Pr \left(\max \left\{ \frac{i}{m} - U_i, U_i - \frac{i-1}{m}, i = 1, 2, \dots, m \right\} \geq d \right),$$

το οποίο είναι επίσης δύσκολο να υπολογιστεί.

Μπορούμε όμως με τη μέθοδο της προσομοίωσης να το εκτιμήσουμε. Η διαδικασία έχει ως εξής

- Παράγουμε όσες τιμές αντιστοιχούν στο δείγμα μας (100 στην περίπτωση μας) από την ομοιόμορφη κατανομή $(0,1)$.
- Τις τοποθετούμε σε μια λίστα και τις διατάσσουμε σε αύξουσα σειρά.
- Βρίσκουμε τη στατιστική συνάρτηση D και εξετάζουμε αν ξεπερνάει την τιμή d .

- Επαναλαμβάνουμε τη διαδικασία n φορές (10.000 στην περίπτωσή μας) και καταγράφουμε σε μια νέα λίστα πόσες φορές το D ξεπέρασε το d .
- Το ποσοστό των φορών που η στατιστική συνάρτηση D ξεπέρασε το d αποτελεί την εκτίμηση του p -value.

Επιθυμούμε να ελέγξουμε αν τα δεδομένα του Πίνακα 5.1 προέρχονται από τη λογιστική κατανομή με παραμέτρους έτσι όπως εκτιμήθηκαν με τη μέθοδο της μέγιστης πιθανοφάνειας, δηλαδή

$$\hat{\mu} = 1010.79 \text{ και } \hat{\sigma} = 160.712$$

Ακολουθούμε τα παράνω βήματα στο Mathematica.

Οι εντολές που χρησιμοποιήθηκαν είναι οι εξής:

```
(*Διατάσσουμε τα δεδομένα μας*)
mySortedLamps = Sort[myLamps];
(*Υπολογίζουμε το πλήθος της λίστας*)
arithmosLamps = Length[myLamps];
(*Ορίζουμε τη λογιστική κατανομή με συγκεκριμένες παραμέτρους*)
myLogisticDistribution = LogisticDistribution[1010.79, (160.712 * Sqrt[3]) / Pi];
(*Υπολογίζουμε την απόσταση d *)
distance1 = Max[Table[i / arithmosLamps - CDF[myLogisticDistribution,
    mySortedLamps[[i]]], {i, 1, arithmosLamps}]];
distance2 = Max[Table[CDF[myLogisticDistribution,
    mySortedLamps[[i]]] - (i - 1) / arithmosLamps, {i, 1, arithmosLamps}]];
maxDistance = Max[distance1, distance2];
Print["Η απόσταση d είναι: ", maxDistance]
```

Η απόσταση d είναι: 0.0846923

```

(*Θέτουμε την απόσταση d και το πλήθος των επαναλήψεων*)
s = 0; d = 0.084692344; n = 10 000; m = 100;
Do[
  (*Παράγουμε 100 τυχαίους αριθμούς από την U(0,1)*)
  U = Table[Random[], {m}]; S = Sort[U];
  (*Υπολογίζουμε την απόσταση D *)
  d1 = Max[Table[i/m - S[[i]], {i, 1, m}]];
  d2 = Max[Table[S[[i]] - (i - 1)/m, {i, 1, m}]];
  If[Max[d1, d2] ≥ d, s = s + 1]
  , {j, 1, n}];
(*Υπολογίζουμε το ποσοστό των επαναλήψεων που
 παρατηρήθηκε D ≥ d *)
ektimwmenoPvalue = N[s/n];
Print["Το εκτιμώμενο p-value είναι: ", ektimwmenoPvalue,
 " και το αντίστοιχο διάστημα εμπιστοσύνης"]
Print[" σε επίπεδο σημαντικότητας 5%: ",
 {ektimwmenoPvalue - (ektimwmenoPvalue (1 - ektimwmenoPvalue) / n)^0.5*1.96,
 ektimwmenoPvalue + (ektimwmenoPvalue (1 - ektimwmenoPvalue) / n)^0.5*1.96}]
Το εκτιμώμενο p-value είναι: 0.4464 και το αντίστοιχο διάστημα εμπιστοσύνης
σε επίπεδο σημαντικότητας 5%: {0.436656, 0.456144}

```

Επομένως $d = 0.0846923$ και $p\text{-value} = 0.4464$.

Συνεπώς δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση σε ε.σ 5%.

Παράρτημα

Εντολές στο *Mathematica*

Στο παράρτημα αυτό δίνονται οι εντολές που χρησιμοποιήθηκαν στο μαθηματικό πακέτο *Mathematica* για την κατασκευή των σχημάτων στη διπλωματική εργασία.

- **Κατασκευή σχήματος 1.1**

```
Plot[(1+Exp[-x])^-1, {x, -6, 6}, AxesLabel->{t, x[t]}, Filling->Axis];
```

- **Κατασκευή σχήματος 1.2**

```
h=Plot[1/(1+Exp[-x]), {x, -6, 6}, AxesOrigin->{-6, 0}, PlotStyle->Red];  
i=Plot[1/(1+Exp[-2x]), {x, -6, 6}, AxesOrigin->{-6, 0}, PlotStyle->Blue];  
z=Plot[1/(1+Exp[-0.5x]), {x, -6, 6}, AxesOrigin->{-6, 0}, PlotStyle->Green];  
Show[h, i, z]
```

- **Κατασκευή σχήματος 2.1**

```
Plot[PDF[LogisticDistribution[0, Sqrt[3]/Pi], x], {x, -5, 5},  
AxesOrigin->{-4, 0}, AxesLabel-> {y, fy[y]}, Filling->Axis]
```

- **Κατασκευή σχήματος 2.2**

```
Plot[CDF[LogisticDistribution[0, Sqrt[3]/Pi], x], {x, -5, 5},  
AxesOrigin->{-4, 0}, AxesLabel-> {y, Fy[y]}, Filling->Axis]
```

- **Κατασκευή σχήματος 2.3**

```
h=Plot[PDF[LogisticDistribution[0, Sqrt[3]/Pi], x],  
{x, -5, 5}, AxesOrigin->{-4, 0}, Filling->Axis];  
i=Plot[PDF[NormalDistribution[0, 1], x], {x, -5, 5}, AxesOrigin->  
{-4, 0}, Filling->Axis, FillingStyle->Opacity[0.3]];  
Show[h, i]
```

- **Κατασκευή σχήματος 2.4**

```
h=Plot[PDF[LogisticDistribution[2,0.5],x],{x,-5,7},AxesOrigin→{-5,0},Filling→Axis];  
i=Plot[PDF[LogisticDistribution[0,0.5],x],{x,-5,7},AxesOrigin→{-5,0},Filling→Axis,FillingStyle→Opacity[0.5]];  
Show[h,i]
```

- **Κατασκευή σχήματος 2.5**

```
h=Plot[PDF[LogisticDistribution[1,4],x],{x,-35,35},AxesOrigin→{-35,0},Filling→Axis];  
i=Plot[PDF[LogisticDistribution[1,6],x],{x,-35,35},AxesOrigin→{-35,0},Filling→Axis,FillingStyle→Opacity[0.5]];  
Show[h,i]
```

- **Κατασκευή σχήματος 2.6**

```
Plot[(Pi/(Sqrt[3]))*CDF[LogisticDistribution[0,Sqrt[3]/Pi],x],  
{x,-5,5},AxesOrigin→{-4,0},AxesLabel→{y,hy[y]},Filling→Axis]
```

- **Κατασκευή σχήματος 4.1**

```
f[y_,α_]=α*Exp[-y]*(1+Exp[-y])^-(α+1);  
Plot[{f[y,0.5],f[y,3],f[y,1]},{y,-15,10},PlotStyle→  
{Red,Blue,Green},PlotRange→{{-13,10},{0,0.35}},AxesLabel→  
{t,g[t]},AxesOrigin→{-13,0}]
```

- **Κατασκευή σχήματος 4.2**

```
f[y_,α_]=α*Exp[-α*y]*(1+Exp[-y])^-(α+1);  
Plot[{f[y,0.5],f[y,3],f[y,1]},{y,-15,10},PlotStyle→  
{Red,Blue,Green},PlotRange→{{-13,13},{0,0.35}},AxesLabel→  
{t,g[t]},AxesOrigin→{-13,0}]
```

- **Κατασκευή σχήματος 4.3**

```
f[y_,α_]=(Gamma[2α])/(Gamma[α])^2*(Exp[-α*y]*(1+Exp[-y])^-(2*α));  
Plot[{f[y,0.5],f[y,1.5],f[y,1]},{y,-15,10},PlotStyle→  
{Red,Blue,Green},PlotRange→{{-13,10},{0,0.4}},AxesLabel→  
{t,g[t]},AxesOrigin→{-13,0}]
```

- **Κατασκευή σχήματος 4.4**

```
f[y_,α_]=α*Exp[-y]*(1+Exp[-y])^-(α+1);  
Plot[{f[y,0.2],f[y,0.3],f[y,0.6]},{y,-15,10},PlotStyle→  
{Red,Black,Green},PlotRange→{{-15,10},{0,0.35}},AxesLabel→  
{t,g[t]},AxesOrigin→{-15,0}]
```

- **Κατασκευή σχήματος 4.5**

```
f[y_,α_]=α*Exp[-y]*(1+Exp[-y])^-(α+1);  
Plot[{f[y,3],f[y,7],f[y,15]},{y,-15,10},PlotStyle→  
{Red,Black,Green},PlotRange→{{-5,10},{0,0.45}},AxesLabel→  
{t,g[t]},AxesOrigin→{-5,0}]
```

- **Κατασκευή σχήματος 4.6**

```
Plot[Exp[-x]},{x,-10,10},PlotRange→{{0,3},{0,1.2}},AxesLabel→  
{t,f[t]},AxesOrigin→{0,0}]
```

- **Κατασκευή σχήματος 4.7**

```
s[y_,α_]=Exp[y/α]*(1+Exp[y/α])^-(α+1);  
Plot[s[y,0.00000001]},{y,-1,3},PlotStyle→Red,AxesLabel→  
{t,g[t]},AxesOrigin→{0,0},PlotRange→{{0,3},{0,1.2}}]
```

- **Κατασκευή σχήματος 4.8**

```
Plot[Exp[-Exp[-x]], {x, -10, 30}, PlotRange -> {{-10, 10}, {0, 1}}, AxesLabel ->
{t, g[t]}, AxesOrigin -> {0, 0}]
```

- **Κατασκευή σχήματος 4.9**

```
s[y_, α_] = (1 + Exp[-y]/α) ^ -α;
Plot[s[y, 100], {y, -10, 10}, PlotRange -> {{-10, 10}, {0, 1}}, AxesLabel ->
{σ, σ[t]}, AxesOrigin -> {0, 0}]
```

- **Κατασκευή σχήματος 4.10**

```
f[y_, α_] = (1/Sqrt[2/α]) * (Gamma[2α]) / (Gamma[α]) ^ 2 *
(Exp[-α * ((1/Sqrt[2/α]) * y)] * (1 + Exp[-((1/Sqrt[2/α]) * y)]) ^ - (2 * α));
z = Plot[f[y, 2.12], {y, -15, 15}, PlotStyle -> {Red}, PlotRange ->
{{-6, 6}, {0, 0.5}}, AxesLabel -> {t, g[t]}, AxesOrigin -> {-6, 0}];
k = Plot[PDF[NormalDistribution[0, 1], x], {x, -6, 6}];
Show[z, k]
```

- **Κατασκευή σχήματος 4.11**

```
f[y_, α_] = (Gamma[2α]) / (Gamma[α]) ^ 2 * (Exp[-α * y] * (1 + Exp[-y]) ^ - (2 * α));
z = Plot[f[y, (15 - 3.25) / 5.5], {y, -15, 15}, PlotStyle -> {Red}, PlotRange ->
{{-5, 5}, {0, 0.4}}, AxesLabel -> {t, g[t]}, AxesOrigin -> {-5, 0}];
l = Plot[PDF[StudentTDistribution[15], x], {x, -5, 5}];
```

- **Κατασκευή σχήματος 5.1**

```
orderList = Sort[myList]; xlist = {};
For[i = 1, i ≤ Length[orderList], i++, AppendTo[xlist, (Sqrt[3]/Pi) *
Log[((i - 0.5) / Length[myList]) / (1 - ((i - 0.5) / Length[myList]))]];
data = Table[{xlist[[i]], orderList[[i]]}, {i, Length[orderList]}];
Needs["LinearRegression`"]
line = Fit[data, {1, x}, x]
```

```
Regress[data, {1, x}, x]
Show[ListPlot[data], Plot[{line}, {x, -3, 3}], PlotRange→
{{-3, 3}, {50, 180}}, AxesOrigin→ {-3, 50}]
```

- **Κατασκευή σχήματος 5.2**

```
z=Sort[myList];edfList={};
For[i=1,i<= Length[z],i++,AppendTo[edfList,i/Length[z] ]]
newdata = Table[{z[[i]],edfList[[i]]},{i,Length[z]}];
z=Show[ListPlot[newdata],PlotRange→ {{50,160},{0,1}},
AxesOrigin→ {50,0}]
```

- **Κατασκευή σχήματος 5.3**

```
q=Plot[CDF[LogisticDistribution[100.29, (16.55*sqrt[3])/Pi], x],
{x, 50, 160}]
```

- **Κατασκευή σχήματος 5.4**

```
z=Sort[myList];edfList={};
For[i=1,i<= Length[z],i++,AppendTo[edfList,i/Length[z] ]]
newdata = Table[{z[[i]],edfList[[i]]},{i,Length[z]}];
z=Show[ListPlot[newdata],PlotRange→ {{50,160},{0,1}},
AxesOrigin→ {50,0}]
q=Plot[CDF[LogisticDistribution[100.29, (16.55*sqrt[3])/Pi], x],
{x, 50, 160}]
Show[z, q]
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Αντζουλάκος, Δημήτριος (2012). Ανάλυση Επιβίωσης, Πανεπιστημιακές Σημειώσεις, ΠΜΣ «Εφαρμοσμένη Στατιστική».
- Κούτρας, Μάρκος Β. (2004). *Εισαγωγή στις Πιθανότητες, Θεωρία και Εφαρμογές: Μέρος I*, Εκδόσεις Α.Σταμούλης.
- Κούτρας, Μάρκος Β. (2005). *Εισαγωγή στις Πιθανότητες, Θεωρία και Εφαρμογές: Μέρος II*, Εκδόσεις Α.Σταμούλης.
- Κούτρας, Μάρκος Β. (2011). *Ανάλυση Παλινδρόμησης και Ανάλυση Διακύμανσης*, Πανεπιστημιακές Σημειώσεις, ΠΜΣ «Εφαρμοσμένη Στατιστική».
- Μπούτσικας, Μιχαήλ (2004). *Μέθοδοι Προσομοίωσης και Υπολογιστικές Στατιστικές Τεχνικές*, Πανεπιστημιακές Σημειώσεις, ΠΜΣ «Εφαρμοσμένη Στατιστική».

Ξένη

- Abramowitz, M. and Stegun, I. A. (eds.) (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.
- Ahuja, J. C. and Nash, S. W. (1967). The generalized Gompertz-Verhulst family of distributions, *Sankhya A*, **29**, 141-156.
- Balakrishnan, N. (ed.) (1992). *Handbook of the Logistic Distribution*, New York: Dekker.
- Balakrishnan, N. and Leung, M. Y. (1988a). Order Statistics from the Type I generalized logistic distribution, *J. Statist. Plan. and Inference*, **26**, 221-236.
- Balakrishnan, N. and Leung, M. Y. (1988b). Means, variances, and covariances of order statistics, BLUE's for the Type I generalized logistic distribution, and some applications, *Commun. Statist.-Simula.*, **17**, 51-84.
- Count, E. W. (1943). Growth pattern of human physique: an approach to kinetic anthropometry, *Human Biol.*, **15**, 1-32.
- Davidson, R. R. (1980). Some properties of a family of generalized logistic distributions, *Statistical Climatology, Developments in Atmospheric Science 13*, S. Ikeda et al., (ed.), Elsevier, New York.
- Davis, H. T. (1935). *Tables of the Higher Mathematical Functions*, Vols. 1, 2. Principia Press, Bloomington.

- Eisen, M. (1979). *Mathematical models in Cell Biology and Cancer Chemotherapy*, Lecture Notes in Biomathematics, **30**, Springer-Verlag, Berlin and New York.
- Erkelens, J. (1968). A method of calculation for the logistic curve, *Statistica Neerlandica*, **22**, 213-217. (In Dutch).
- George, E. O. and Ojo, M. O. (1980). On a generalization of the logistic distribution, *Ann. Inst. Statist. Math. A*, **32**, 161-169.
- George, E. O. , El-Saidi, M., and Singh, K. (1986). A generalized logistic approximation of the Student *t* distribution, *Commun. Statist.-Simulation*, **15**, 1199-1208.
- Gumbel, E. J. (1961). Bivariate logistic distributions. *J. Amer. Statist. Assoc.*, **56**, 335-349.
- Harter, H. L. and Moore, A. H. (1967). Maximum-likelihood estimation, from censored samples, of the parameters of a logistic distribution. *J. Amer. Statist. Assoc.*, **62**, 675-684.
- Jensen, A. L. (1975). Comparison of logistic equation for population growth, *Biometrics*, **31**, 853-862.
- Johnson, N. L. and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2. Wiley, New York.
- Malthus T.R. (1798). *An Essay on the Principle of Population*, Chapter II, 18.
- Miller, R. S. and Botkin, D. B. (1974). Endangered species: models and predictions, *American Scientist*, **62**, 172-181.
- Mudholkar, G. S. and George, E. O. (1978). A remark on the shape of the logistic distribution, *Biometrika*, **65**, 667-668.
- Pearl, R. and Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation, *Proceedings of the National Academy of Sciences*, **6**, 275-288.
- Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves, *Biometrics*, **32**, 761-768.
- Reed, L. J. and J. Berkson (1929). The application of the logistic function to experimental data. *Journal of Physical Chemistry*, **33**, 760-779.
- Tan, W. Y. (1983). On the distribution of number of mutants at the hypoxanthine-guanine phosphoribosal transferase locus in Chinese hamster ovary cells, *Math.Biosciences*, **67**, 175-192.
- Verhulst, Pierre-François (1845). Mathematical Researches into the Law of Population Growth Increase. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, **18**, 1-42.

Πανεπιστήμιο Πειραιώς