ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΜΟΝΤΕΛΑ ΠΙΘΑΝΟΤΗΤΩΝ ΓΙΑ ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ ΚΙΝΔΥΝΩΝ

Βασιλική Καλαμίδα

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 Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Πειραιάς

Ιούλιος 2016

Η παρούσα διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμόν ….. συνεδρίαση του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου.

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής Μ. Κούτρας (Επιβλέπων)
- Αναπλ. Καθηγητής Γ. Πολίτης
- Επίκ. Καθηγητής Γ. Ψαρράκος

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

UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUADE PROGRAM IN ACTUARIAL SCIENCE AND RISK MANAGEMENT

PROBABILITY MODELS FOR FITTING RISK DATA

By

Vassiliki Kalamida

MSc Dissertation

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

Piraeus Greece

2016

ΕΥΧΑΡΙΣΤΙΕΣ

Με αφορμή την ολοκλήρωση της παρούσας διπλωματικής εργασίας θα ήθελα να ευχαριστήσω ιδιαιτέρως όλους τους καθηγητές του Πανεπιστημίου Πειραιώς οι οποίοι βρέθηκαν δίπλα μου τόσο σε προπτυχιακό όσο και σε μεταπτυχιακό επίπεδο και μετέδωσαν τις γνώσεις τους απλόχερα. Η προσήλωσή τους στο έργο τους είναι αξιοθαύμαστη και η συμβολή τους στις γνώσεις μου ήταν καθοριστική.

Ιδιατέρως θα ήθελα να ευχαριστήσω τον κο. Κούτρα για την καθοδήγηση και τη στήριξη του κατά της διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.

Κλείνοντας, εκφράζω επίσης θερμές ευχαριστίες στον κο. Πολίτη και τον κο. Ψαρράκο για την συμβολή τους στην ολοκλήρωση της εργασίας.

ΠΕΡΙΛΗΨΗ

Είναι γνωστό ότι υπάρχουν πολλά χαρακτηριστικά τα οποία δεν μπορούν να περιγραφούν ικανοποιητικά από τις κλασικές κατανομές όπως για παράδειγμα η κανονική ή η εκθετική. Χαρακτηριστικές τέτοιες περιπτώσεις αποτελούν οι αποδόσεις μιας επένδυσης ή οι αποζημιώσεις μιας ασφαλιστικής εταιρείας όπου η απόκλιση από τα κλασικά στατιστικά μοντέλα οφείλεται κυρίως στο γεγονός ότι συνήθως εμφανίζουν ακραίες τιμές.

Η παρούσα εργασία πραγματεύεται τη γενικευμένη βήτα παραγόμενη οικογένεια κατανομών (GBG) η οποία επεκτείνει κοινές κατανομές, που καλούνται γεννήτορες, σε τέτοιες που χαρακτηρίζονται από ευελιξία στη μοντελοποίηση δεδομένων τα οποία εμφανίζουν λοξότητα και κύρτωση.

Το πρώτο κεφάλαιο αποτελεί ένα εισαγωγικό μέρος στο οποίο αναφέρονται χαρακτηριστικά της κανονικής κατανομής που την καθιστούν ένα ελκυστικό εργαλείο στη μελέτη των κινδύνων, καθώς και η ανάγκη δημιουργίας νέων κατανομών οι οποίες να έχουν μεγαλύτερη ευελιξία και προσαρμοστικότητα σε πραγματικά δεδομένα διαχείρισης κινδύνων.

Το δεύτερο κεφάλαιο αναφέρεται σε μοντέλα κατανομών που έχουν προταθεί για την προσέγγιση μη κανονικών δεδομένων, συγκεκριμένα τη μετασχηματισμένη βήτα κατανομή και τη βήτα παραγόμενη οικογένεια κατανομών. Στη συνέχεια, αναλύεται η γενικευμένη βήτα παραγόμενη οικογένεια κατανομών και αναπτύσσονται διεξοδικά οι ιδιότητές της και βασικά στατιστικά μέτρα αυτής.

Στο τρίτο κεφάλαιο μελετώνται ως γεννήτορες κατανομές οι Pareto, Weibull και Λογιστική και εξειδικεύονται τα αποτελέσματα του δεύτερου κεφαλαίου για τις GBG κατανομές.

Το τέταρτο κεφάλαιο εξετάξει την εφαρμογή και προσαρμοστικότητα των γενικευμένων βήτα παραγόμενων κατανομών στη μοντελοποίηση πραγματικών δεδομένων.

Τέλος, στο παράρτημα παρατίθενται οι ορισμοί και οι ιδιότητες των συναρτήσεων Γάμμα και Βήτα καθώς και σχετικές αποδείξεις που δεν περιέχονται στο κυρίως κείμενο.

ABSTRACT

It is common knowledge that there are many features that cannot be described fully by traditional distributions such as the normal or the exponential one. Such typical cases are either the yield of an investment or the amount of compensation of an insurance company, which exhibit extreme values. Dealing with this kind of data, classical statistical models are insufficient in analysis.

This thesis deals with generalized beta generated (GBG) class of distributions which extends several common distributions, called parent generators, into distributions characterized by their flexibility in modeling data in practice. The study has the following structure.

The first chapter is an introductory section regarding the basic characteristics of normal distribution making it an attractive tool for studying hazards in general. Also the need for development models that are suitable to formulate the abovementioned data is discussed.

The second chapter refers to distributions models that have been proposed to approach the non-normal data, namely the transformed beta family and the beta generated class of distributions. Subsequently, the GBG class of distributions is analyzed thoroughly and its basic statistical measures are presented as well.

In chapter three, the results obtained by the analysis performed in chapter two are specified using as parent distributions the Pareto, the Weibull and the Logistic distributions. In each case, the distribution of the parent is analyzed in details and the corresponding GBG distribution with its own characteristics is provided.

The fourth chapter investigates the flexibility and applicability of GBG distributions in modeling a set of real data.

Finally, the appendix provides the definitions and relevant properties of Gamma and Beta functions, and considers the proofs that have not been detailed in the main body of this study.

ΠΕΡΙΕΧΟΜΕΝΑ

1 ΕΙΣΑΓΩΓΗ 1

1.1. Κανονική κατανομή 2

1.2 Ανάγκη αναζήτησης περισσότερο αποτελεσματικών κατανομών 3

2 ΜΟΝΤΕΛΑ ΚΑΤΑΝΟΜΩΝ 5

2.1 Τυπικοί μετασχηματισμοί 5

2.2 Οικογένεια μετασχηματισμένης βήτα κατανομής 6

2.3 Βήτα παραγόμενη οικογένεια κατανομών 9

2.4 Γενικευμένη βήτα παραγόμενη οικογένεια κατανομών 12

2.4.1 Αναπτύγματα της συνάρτησης κατανομής και της συνάρτησης πυκνότητας 14

2.4.2 Χαρακτηριστικά μεγέθη κατανομής 19

3 ΚΑΤΑΝΟΜΕΣ PARETO, WEIBULL, ΛΟΓΙΣΤΙΚΗ 39

3.1 Κατανομή Pareto$^{(II)}$ 39

3.1.1 Χαρακτηριστικά μεγέθη της κατανομής Pareto$^{(II)}$ 40

3.1.2 Εκτίμηση παραμέτρων 42

3.1.3 Μετασχηματισμοί και ειδικές περιπτώσεις 44

3.2 Γενικευμένη βήτα παραγόμενη Pareto$^{(II)}$ κατανομή 46

3.3 Κατανομή Weibull 51

3.3.1 Χαρακτηριστικά μεγέθη της κατανομής Weibull 52

3.3.2 Εκτίμηση παραμέτρων 54

3.3.3 Μετασχηματισμοί και ειδικές περιπτώσεις 56

3.4 Γενικευμένη βήτα παραγόμενη Weibull κατανομή 57

3.5 Λογιστική κατανομή 61

3.5.1 Χαρακτηριστικά μεγέθη της Λογιστικής κατανομής 62

3.5.2 Εκτίμηση παραμέτρων 64

3.5.3 Μετασχηματισμοί και ειδικές περιπτώσεις 66

3.6 Γενικευμένη βήτα παραγόμενη Λογιστική κατανομή 68

4 ΑΡΙΘΜΗΤΙΚΗ ΕΦΑΡΜΟΓΗ 73

4.1 Προσαρμογή των δεδομένων 75

4.2 Έλεγχοι λόγου πιθανοφάνειας 81

5 ΣΥΜΠΕΡΑΣΜΑΤΑ 87

ΠΑΡΑΡΤΗΜΑ 89

ΒΙΒΛΙΟΓΡΑΦΙΑ 105

ΣΧΗΜΑΤΑ

Σχήμα 1.1 Συνάρτηση πυκνότητας κανονικής κατανομής 2

Σχήμα 2.1 Οικογένεια κατανομών μετασχηματισμένης βήτα 8

Σχήμα 2.2 Γραφική παράσταση συνάρτησης πυκνότητας $BN(0,1,a,b)$ για διάφορες τιμές των παραμέτρων a, b 11

Σχήμα 2.3 Οικογένεια κατανομών GBG 14

Σχήμα 2.4 Η λοξότητα B (αριστερά) και η κύρτωση M (δεξιά) για την GBGN κατανομή σαν συνάρτηση του c, για $a = b = 0.5, 1, 10$ και 50. 24

Σχήμα 2.5 Η λοξότητα B (αριστερά) και η κύρτωση M (δεξιά) για την GBGStudent κατανομή με 10 βαθμούς ελευθερίας σαν συνάρτηση του c, για $a = b = 0.5, 1, 10$ και 50. . 24

Σχήμα 2.6 Η λοξότητα B (αριστερά) και η κύρτωση M (δεξιά) για την GBGStudent κατανομή με 5 βαθμούς ελευθερίας σαν συνάρτηση του c, για $a = b = 0.5, 1, 10$ και 50. ... 24

Σχήμα 3.1 Συνάρτηση πυκνότητας της $Par^{(II)}(\delta, \lambda)$ 40

Σχήμα 3.2 Συνάρτηση κινδύνου της $Par^{(II)}(\delta, \lambda)$ 40

Σχήμα 3.3 Συνάρτηση πυκνότητας της $GBGPar(\delta, \lambda, a, b, c)$ 47

Σχήμα 3.4 Συνάρτηση πυκνότητας της $W(\lambda, \gamma)$ για διάφορες τιμές του γ με $\lambda = 0.01^{\gamma}$ 51

Σχήμα 3.5 Συνάρτηση κινδύνου της $W(\lambda, \gamma)$ με $\lambda = 0.01^{\gamma}$ 52

Σχήμα 3.6 Συνάρτηση πυκνότητας της $GBGW(\lambda, \gamma, a, b, c)$ 58

Σχήμα 3.7 Συνάρτηση πυκνότητας της $Log(\mu, \sigma)$ 62

Σχήμα 3.8 Συνάρτηση κινδύνου της $Log(\mu, \sigma)$ 63

Σχήμα 3.9 Συνάρτηση πυκνότητας της $GBGLog(\mu, \sigma, a, b, c)$ 68

Σχήμα 4.1 Ιστόγραμμα και CDF της εμπειρικής κατανομής 73

Σχήμα 4.2 Ιστόγραμμα σε μεγέθυνση 73

Σχήμα 4.3 Εμπειρική συνάρτηση κατανομής για μεγέθη ζημιών έως 9000 74

Σχήμα 4.4 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των προσαρμοζόμενων κατανομών με γεννήτορα την Pareto$^{(II)}$ 76

Σχήμα 4.5 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των προσαρμοζόμενων κατανομών με γεννήτορα την Weibull 77

Σχήμα 4.6 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των προσαρμοζόμενων κατανομών με γεννήτορα τη Λογιστική 81

Σχήμα 4.7 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των καλύτερα προσαρμοζόμενων κατανομών για κάθε γεννήτορα 85

ΠΙΝΑΚΕΣ

Πίνακας 4.1 Περιγραφικά μέτρα 74

Πίνακας 4.2 Εμπ και πληροφοριακά κριτήρια όταν ο γεννήτορας είναι η Pareto$^{(II)}$ 76

Πίνακας 4.3 Εμπ και πληροφοριακά κριτήρια όταν ο γεννήτορας είναι η Weibull..... 77

Πίνακας 4.4 Εμπ και στατιστικές τιμές όταν ο γεννήτορας είναι η Pareto$^{(II)}$ 79

Πίνακας 4.5 Εμπ και στατιστικές τιμές όταν ο γεννήτορας είναι η Weibull 79

Πίνακας 4.6 Εμπ και σχετικά στατιστικά μέτρα όταν ο γεννήτορας είναι η Λογιστική 80

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Η διαχείριση κινδύνου είναι κεντρικός πυρήνας της στρατηγικής διαχείρισης κάθε οργανισμού. Είναι η διεργασία με την οποία οι οργανισμοί προσεγγίζουν μεθοδικά τους κινδύνους που σχετίζονται με τις δραστηριότητές τους, με σκοπό την επίτευξη κέρδους σε κάθε δραστηριότητα. Το επίκεντρο της καλής διαχείρισης κινδύνου είναι η αναγνώριση και ο χειρισμός τους που συνεπώς αυξάνουν την πιθανότητα επιτυχίας των συνολικών στόχων του οργανισμού.

Ως κίνδυνος μπορεί να οριστεί η αβεβαιότητα για γεγονότα και αποτελέσματα αυτών, τα οποία ενδέχεται να έχουν επίδραση στην οικονομική κατάσταση ενός οργανισμού. Οι κίνδυνοι αυτοί μπορούν να περιγραφούν από τη θεωρία πιθανότητων. Συνεπώς ο στοχαστικός παράγοντας στην εκτίμηση και ερμηνεία των κινδύνων είναι απαραίτητο στοιχείο της μελέτης και της ορθής και αποτελεσματικής διαχείρισης κινδύνων.

Είναι γνωστό ότι υπάρχουν πολλά χαρακτηριστικά τα οποία δεν μπορούν να περιγραφούν ικανοποιητικά από τις κλασικές κατανομές που γνωρίζουμε όπως η κανονική και η εκθετική. Τα δεδομένα κινδύνου, όπως οι αποδόσεις μιας επένδυσης ή οι αποζημιώσεις μιας ασφαλιστικής εταιρείας αποτελούν χαρακτηριστικές περιπτώσεις δεδομένων που εμφανίζουν ακραίες τιμές γεγονός που εξηγεί την απόκλισή τους από τα κλασικά στατιστικά μοντέλα.

Υπάρχουν τρεις βασικές προσεγγίσεις για την κατασκευή ζημιοκατανομών, η εμπειρική, η βασιζόμενη στις ροπές και η αναλυτική. Η εμπειρική μέθοδος χρησιμοποιείται όταν είναι διαθέσιμος ένας αρκετά μεγάλος αριθμός δεδομένων και με χρήση τεχνικών προσαρμογής, μπορεί να επιτευχθεί μια ικανοποιητικά στρωτή (ομαλή) και ακριβής εκτίμηση της συνάρτησης κατανομής των δεδομένων. Η μέθοδος των ροπών χρησιμοποιείται όταν δεν απαιτείται η ακριβής μορφή της ζημιοκατανομής και γίνεται η εκτίμηση κάποιων ροπών μικρής τάξης, όπως για παράδειγμα η μέση τιμή και η διακύμανση. Η αναλυτική μέθοδος είναι η πιο συχνά χρησιμοποιούμενη και συναντάται στη σχετική βιβλιογραφία του αναλογισμού. Η προσέγγιση αυτή αναζητά μαθηματικά μοντέλα κατανομών που να ταιριάζουν και να περιγράφουν αναλυτικά τα δεδομένα.

1.1. Κανονική κατανομή

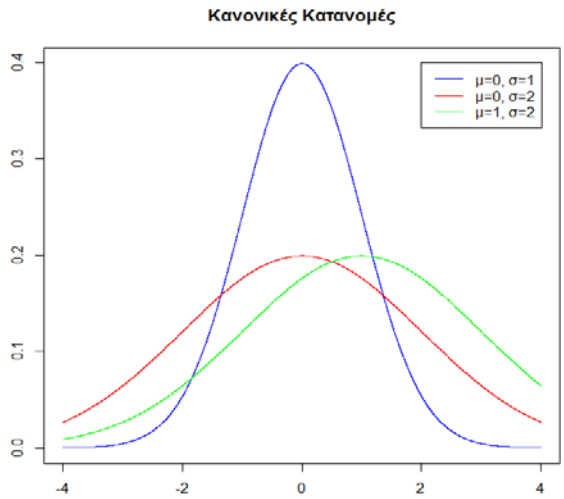
Η πιο συνηθισμένη κατανομή που χρησιμοποιείται μέχρι σήμερα για τη μελέτη των χαρτοφυλακίων είναι η κανονική κατανομή. Η κανονική κατανομή, γνωστή και ως Γκαουσιανή (Gaussian) κατανομή, αναφέρεται σε συνεχείς μεταβλητές αποτελώντας μία συνεχή συνάρτηση πυκνότητας πιθανότητας. Χρησιμοποιείται ως μία πρώτη προσέγγιση για να περιγραφούν τυχαίες μεταβλητές πραγματικών τιμών, οι οποίες τείνουν να συγκεντρώνονται γύρω από μια μέση τιμή.

Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής, γνωστής ως Γκαουσιανής συνάρτησης δίνεται από την ακόλουθη σχέση

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbf{R}$$

όπου, μ είναι η μέση τιμή και αποτελεί παράμετρο θέσης και σ^2 η διασπορά της κατανομής και αποτελεί παράμετρο κλίμακας.

Η γραφική παράσταση της παραπάνω συνάρτησης πυκνότητας απεικονίζεται στο Σχήμα 1.1 έχει σχήμα «καμπάνας» και αναφέρεται και ως κωδωνοειδής καμπύλη. Είναι συμμετρική και οι «ουρές» της πλησιάζουν τον οριζόντιο άξονα ασυμπτωτικά.



Σχήμα 1.1 Συνάρτηση πυκνότητας κανονικής κατανομής

Η κανονική κατανομή αποτελεί την πιο σημαντική κατανομή της στατιστικής μεθοδολογίας για τους εξής βασικούς λόγους:

- Την κανονική κατανομή ακολουθούν είτε με ακρίβεια είτε με μεγάλη προσέγγιση τα περισσότερα συνεχή φαινόμενα.
- Πολλές διακριτές κατανομές πιθανοτήτων μπορούν να προσεγγιστούν μέσω της κανονικής κατανομής.
- Η κανονική κατανομή αποτελεί σύμφωνα με το Κεντρικό Οριακό Θεώρημα, στο οποίο αναφέρεται ότι το άθροισμα ενός ικανοποιητικά μεγάλου αριθμού ανεξάρτητων και ισόνομων τυχαίων μεταβλητών προσεγγίζεται από την κανονική κατανομή, τη βάση της στατιστικής συμπερασματολογίας ή επαγωγικής στατιστικής.
- Τυχαία σφάλματα που εμφανίζονται σε διάφορες μετρήσεις έχουν κανονική κατανομή. Γι' αυτό το λόγο η κανονική κατανομή αναφέρεται πολλές φορές και ως κατανομή σφαλμάτων.

Στη διαχείριση κινδύνων, η κανονική κατανομή έχει πολλές εφαρμογές, καθώς αρκετές από τις υφιστάμενες μεθοδολογίες ξεκινούν με την υπόθεση της κανονικότητας των δεδομένων (π.χ. μέθοδος Value at Risk, Random Walk, κλπ). Οι ιδιότητές της την κατατάσσουν στις πρώτες θέσεις προτίμησης για την ερμηνεία της συμπεριφοράς των κινδύνων και αυτός είναι ο λόγος που η χρήση της είναι τόσο ευρέως διαδεδομένη.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση εξειδικευμένων κατανομών που έχουν προταθεί για την περιγραφή δεδομένων κινδύνου έναντι των «κλασικών» στατιστικών κατανομών.

1.2 Ανάγκη αναζήτησης περισσότερο αποτελεσματικών κατανομών

Παρά το γεγονός ότι είναι σύνηθες να γίνεται η υπόθεση ότι οι αποδόσεις ενός χαρτοφυλακίου είναι κανονικά κατανεμημένες, μια τέτοια προσέγγιση δεν είναι πάντα ικανοποιητική, καθώς περιπτώσεις στις οποίες παρατηρούνται παχιές ουρές ή ασυμμετρία στην κατανομή των αποδόσεων, δεν μπορούν να περιγραφούν αποτελεσματικά.

Τα δεδομένα κινδύνων συχνά δεν πληρούν τις προϋποθέσεις για την κανονικότητα έχοντας ως χαρακτηριστικά τους την ασυμμετρία, την ύπαρξη πολλαπλών κορυφών ή παχιές ουρές. Έτσι, τα συμπεράσματα από τη μελέτη της συμπεριφοράς των κινδύνων, χρησιμοποιώντας την κανονική κατανομή, δεν είναι πάντα αξιόπιστα, ενώ παράλληλα οι συνηθισμένες τεχνικές εξαγωγής συμπερασμάτων δεν μπορούν να εφαρμοστούν σε τέτοιου είδους δεδομένα. Η παρατήρηση αυτή μας οδήγησε στην ανάγκη αναζήτησης κατανομών που θα μπορούν να ερμηνεύσουν ικανοποιητικά τις συμπεριφορές αυτές.

Με βάση την αναλυτική μέθοδο θα γίνει η παρουσίαση διαφόρων προτάσεων κατανομών που μπορούν να ικανοποιήσουν προσεγγιστικά τα μη κανονικά δεδομένα. Θα γίνει εκτενής αναφορά στα βασικά μεγέθη και χαρακτηριστικά της κάθε κατανομής καθώς και η μεθοδολογία που ακολουθείται για την εκτίμηση των παραμέτρων της. Επιπλέον, θα παρουσιαστούν και οι κατανομές που προκύπτουν με κατάλληλους μετασχηματισμούς και θα μπορούσαν να αποτελέσουν ικανοποιητικά μοντέλα. Τέλος, θα γίνει εφαρμογή των τεχνικών αυτών σε πραγματικά δεδομένα και θα μπορέσουμε να συγκρίνουμε τις κατανομές ως προς την καταλληλότητά τους με βάση τα δεδομένα.

Η παρούσα μελέτη εστιάζει στην κατανομή Pareto που χρησιμοποιείται σε αναλογιστικές μελέτες λόγω βάρους στη δεξιά ουρά, στη Λογιστική κατανομή που μοιάζει μεν αρκετά με την κανονική κατανομή όμως δε διαθέτει βαριές ουρές και στην κατανομή Weibull που χαρακτηρίζεται επίσης από βαριές ουρές και ασυμμετρία και έχει μεγάλη ευελιξία προσαρμογής σε ποικιλία εφαρμογών.

ΚΕΦΑΛΑΙΟ 2

ΜΟΝΤΕΛΑ ΚΑΤΑΝΟΜΩΝ

Στο κεφάλαιο αυτό βασιζόμενοι στη διεθνή βιβλιογραφία, παρουσιάζουμε διάφορους τρόπους ανάπτυξης νέων μονοδιάστατων κατανομών στην προσπάθεια μοντελοποίησης της πολυπλοκότητας των δεδομένων που χειριζόμαστε. Αρχικά γίνεται αναφορά σε βασικούς μετασχηματισμούς τυχαίων μεταβλητών και εν συνεχεία, εστιάζοντας στο πεδίο των αναλογιστικών δεδομένων, θεωρούμε περισσότερο πολύπλοκα μοντέλα. Πιο συγκεκριμένα, τα μοντέλα που αναφερόμαστε είναι η μετασχηματισμένη βήτα (transformed beta) κατανομή, η βήτα παραγόμενη (beta generated-BG) οικογένεια κατανομών. Τέλος, θα παρουσιάσουμε αναλυτικά τη γενίκευση της τελευταίας, δηλαδή τη γενικευμένη βήτα παραγόμενη (generalized beta generated-GBG) οικογένεια κατανομών.

2.1 Τυπικοί μετασχηματισμοί

Θεωρούμε μια τυχαία μεταβλητή X που ακολουθεί μια κατανομή. Η ανάγκη δημιουργίας νέων τυχαίων μεταβλητών, έστω Y που εκπροσωπούν τις μετρήσεις που θέλουμε, οδήγησε στην ανάπτυξη μετασχηματισμών της X, συμβολικά $Y = g(X)$, και στη μελέτη της κατανομής της Y. Οι πλέον απλοί μετασχηματισμοί που έχουν χρησιμοποιηθεί είναι οι ακόλουθοι:

- Γραμμικός, $Y = aX + b$
  αν $b = 0$ έχουμε την περίπτωση $Y = aX$ που αναφέρεται ως μετασχηματισμός κλίμακας
  αν $a = 1$ έχουμε την περίπτωση $Y = X + b$ που αναφέρεται ως μετασχηματισμός μεταφοράς
- Ύψωσης σε δύναμη, $Y = X^a$
  αν a = −1 έχουμε την περίπτωση $Y = 1/X$ που δίνει την αντίστροφη κατανομή της X
- Λογαριθμικός, $Y = \log X$
- Εκθετικός, $Y = e^X$
- Odds ratio, $Y = X/(1 - X)$

Η μέθοδος της συνάρτησης κατανομής βρίσκει τη συνάρτηση κατανομής F_Y και στη συνέχεια με παραγώγιση προκύπτει η συνάρτηση πυκνότητας πιθανότητας f_Y. Συγκεκριμένα, για τη συνάρτηση κατανομής της τ.μ. Y έχουμε

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

και για τη σ.π.π .

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = \frac{d}{dy} f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Οι παραπάνω μετασχηματισμοί είναι αρκετά απλοί και ενδεχομένως να μην είναι ικανοποιητικοί για τη μελέτη μας. Στο παρόν κεφάλαιο, και εν συνεχεία στο τρίτο εφαρμόζουμε κάποιους από αυτούς τους μετασχηματισμούς καθώς και ορισμένους πιο περίπλοκους σε γνωστές κατανομές ενώ γίνεται αναφορά σε ποιες κατανομές μας οδηγούν.

Η επόμενη παράγραφος αναφέρεται στην οικογένεια κατανομών της μετασχηματισμένης βήτα κατανομής.

2.2 Οικογένεια μετασχηματισμένης βήτα κατανομής

Η οικογένεια της μετασχηματισμένης βήτα κατανομής (transformed beta family) έχει οδηγήσει στη δημιουργία χρήσιμων ζημιοκατανομών, ορισμένες εκ των οποίων είναι λοξές και διαθέτουν βαριά δεξιά ουρά. Οι νέες κατανομές έχουν μεγαλύτερη προσαρμοστικότητα σε δεδομένα χαρτοφυλακίων ζημιών σε σχέση με τις αρχικές κατανομές πάνω στις οποίες βασίστηκε η δημιουργία τους καθώς διαθέτουν περισσότερες παραμέτρους.

Σε πρώτο στάδιο ορίζουμε την κατανομή βήτα τύπου ΙΙ. Έστω η τυχαία μεταβλητή U η οποία ακολουθεί την κατανομή βήτα τύπου Ι με παραμέτρους a και b, συμβολικά $U \sim Beta^{(I)}(a, b)$, με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1, a > 0, b > 0 \tag{2.2.1}$$

και συνάρτηση κατανομής

$$F(x) = \frac{B(a,b;x)}{B(a,b)} = I_x(a,b) \tag{2.2.2}$$

όπου $B(\cdot,\cdot)$ είναι η συνάρτηση Βήτα, $B(\cdot,\cdot;\cdot)$ είναι η μη πλήρης συνάρτηση Βήτα και $I_x(a,b)$ είναι ο λόγος της συνάρτησης Βήτα (παράρτημα Π3, Π4, Π5).

Η τυχαία μεταβλητή X που δίνεται από το μετασχηματισμό $X = \frac{U}{1-U}$ ακολουθεί την κατανομή βήτα τύπου II με παραμέτρους a και b, συμβολικά $X \sim Beta^{(II)}(a,b)$, και έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{B(a,b)} \frac{x^{a-1}}{(1+x)^{a+b}}, \qquad x > 0, a > 0, b > 0 \tag{2.2.3}$$

και συνάρτηση κατανομής

$$F(x) = \frac{B\left(a, b; \frac{x}{1+x}\right)}{B(a,b)} = I_{\frac{x}{1+x}}(a,b). \tag{2.2.4}$$

Η μετασχηματισμένη βήτα (transformed beta) κατανομή, ή αλλιώς γενικευμένη βήτα τύπου II ή Pearson τύπου IV (Klugman, (2004)), έχει μελετηθεί ανεξάρτητα από τους Venter (1983) στην ασφαλιστική βιβλιογραφία, και McDonald (1984) στην οικονομική βιβλιογραφία ως κατανομή εισοδήματος, και προέρχεται από την κατανομή βήτα τύπου II. Συγκεκριμένα, αν θεωρήσουμε την τυχαία μεταβλητή Y που ακολουθεί την κατανομή βήτα τύπου II, με παραμέτρους τ και a, δηλαδή $Y \sim Beta^{(II)}(\tau, a)$, τότε η τυχαία μεταβλητή X που προκύπτει μέσω του μετασχηματισμού $X = \lambda Y^{1/\gamma}, \lambda > 0, \gamma > 0$, ακολουθεί τη μετασχηματισμένη βήτα κατανομή με συνάρτηση κατανομής

$$F_X(x) = P(X \le x) = P\left(\lambda Y^{1/\gamma} \le x\right) = P\left(Y \le \left(\frac{x}{\lambda}\right)^{\gamma}\right) = F_Y\left(\left(\frac{x}{\lambda}\right)^{\gamma}\right)$$

$$= \frac{B\left(\tau, a; \frac{x^{\gamma}}{\lambda^{\gamma} + x^{\gamma}}\right)}{B(\tau, a)} \tag{2.2.5}$$

και συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} F_Y\left(\left(\frac{x}{\lambda}\right)^{\gamma}\right) = \frac{\gamma x^{\gamma-1}}{\lambda^{\gamma}} f_Y\left(\left(\frac{x}{\lambda}\right)^{\gamma}\right)$$

$$= \frac{1}{B(\tau,\alpha)} \frac{\gamma x^{\gamma-1}}{\lambda^{\gamma}} \frac{\left(\frac{x}{\lambda}\right)^{\gamma(\tau-1)}}{\left[1 + \left(\frac{x}{\lambda}\right)^{\gamma}\right]^{\alpha+\tau}} = \frac{1}{B(\tau,\alpha)} \frac{\gamma}{x} \frac{\left(\frac{x}{\lambda}\right)^{\gamma\tau}}{\left[1 + \left(\frac{x}{\lambda}\right)^{\gamma}\right]^{\alpha+\tau}}$$

$$= \frac{\gamma x^{\gamma\tau-1} \lambda^{\alpha\gamma}}{B(\alpha,\tau)(\lambda^{\gamma} + x^{\gamma})^{\alpha+\tau}}, x > 0, a, \lambda, \gamma, \tau > 0. \tag{2.2.6}$$

Η μετασχηματισμένη βήτα κατανομή είναι μία κατανομή τεσσάρων παραμέτρων με μία παράμετρο κλίμακας $\lambda > 0$ και τρεις παραμέτρους μορφής $a > 0$, $\tau > 0$, και $\gamma > 0$, που επιτρέπουν στην κατανομή να έχει τεράστια ευελιξία. Περιλαμβάνει ως ειδικές περιπτώσεις μια πληθώρα περιπτώσεων που αναφέρουμε στη συνέχεια:

(i) για $\gamma = 1$ προκύπτει η Generalized Pareto με σ.π.π.

$$f(x) = \frac{\Gamma(a+\tau)}{\Gamma(a)\Gamma(\tau)} \frac{\lambda^{\alpha} x^{\tau-1}}{(\lambda + x)^{\alpha+\tau}}$$

(ii) για $\tau = 1$ προκύπτει η κατανομή Burr τύπου XII ή αλλιώς Singh-Maddala με σ.π.π.

$$f(x) = \frac{a\gamma\lambda^{\alpha\gamma} x^{\gamma-1}}{(\lambda^{\gamma} + x^{\gamma})^{\alpha+1}}$$

(iii) για $a = 1$ προκύπτει η κατανομή Inverse Burr ή αλλιώς Dagum ή Burr III με σ.π.π.

$$f(x) = \frac{\gamma\tau\lambda^{\gamma} x^{\gamma\tau-1}}{(\lambda^{\gamma} + x^{\gamma})^{\tau+1}}$$

(iv) για $\gamma = 1, \tau = 1$ προκύπτει η Pareto τύπου II ή Lomax με σ.π.π.

$$f(x) = \frac{a\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}$$

(v) για $\gamma = 1, a = 1$ προκύπτει η Inverse Pareto με σ.π.π.

$$f(x) = \frac{\lambda\tau x^{\tau-1}}{(\lambda + x)^{\tau+1}}$$

(vi) για $a = 1, \tau = 1$ προκύπτει η κατανομή Loglogistic ή Fisk στα οικονομικά, με σ.π.π.

$$f(x) = \frac{\gamma\lambda^{\gamma} x^{\gamma-1}}{(\lambda^{\gamma} + x^{\gamma})^{2}}$$

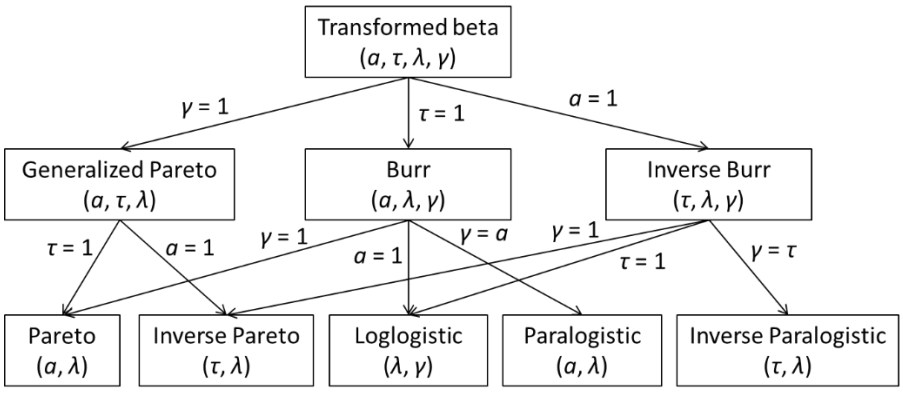
(vii) για $\tau = 1, \gamma = a$ προκύπτει η κατανομή Paralogistic με σ.π.π.

$$f(x) = \frac{a^2 x^{a-1} \lambda^{a^2}}{(\lambda^{a} + x^{a})^{\alpha+1}}$$

(viii) για $a = 1, \gamma = \tau$ προκύπτει η κατανομή Inverse Paralogistic με σ.π.π.

$$f(x) = \frac{\tau^2 \lambda^{\tau} x^{\tau^2-1}}{(\lambda^{\tau} + x^{\tau})^{\tau+1}}$$

Οι κατανομές που προκύπτουν από τη μετασχηματισμένη βήτα κατανομή παρουσιάζονται συνοπτικά στο Σχήμα 2.1.



Σχήμα 2.1 Οικογένεια κατανομών μετασχηματισμένης βήτα

Για τη μετασχηματισμένη βήτα κατανομή οι ροπές k τάξης δίνονται από τον τύπο

$$E(X^k) = \frac{\lambda^k \Gamma\left(\tau + \frac{k}{\gamma}\right) \Gamma\left(a - \frac{k}{\gamma}\right)}{\Gamma(a)\Gamma(\tau)}, \qquad -\tau\gamma < k < a\gamma. \tag{2.2.7}$$

Η κατανομή παρουσιάζει κορυφή (mode) στο σημείο

$$x = \lambda\left(\frac{\tau\gamma - 1}{a\gamma + 1}\right)^{1/\gamma}, \tag{2.2.8}$$

αν $\tau\gamma > 1$ και στο 0 διαφορετικά.

2.3 Βήτα παραγόμενη οικογένεια κατανομών

Έστω δύο τυχαίες μεταβλητές U και Y, όπου $U \sim Beta^{(I)}(a, b)$και Y μία τυχαία μεταβλητή που έχει συνάρτηση κατανομής F_Y και συνάρτηση πυκνότητας f_Y. Ας συμβολίσουμε με $\mathbf{p}$ το διάνυσμα παραμέτρων της τ.μ. Υ. Ορίζουμε μια νέα τυχαία μεταβλητή X μέσω του μετασχηματισμού $X = F_Y^{-1}(U)$. Για τη συνάρτηση κατανομής της X έχουμε

$$F_X(x; \mathbf{p}, a, b) = P(X \le x) = P(F_Y^{-1}(U) \le x) = P\big(U \le F_Y(x)\big)$$
$$= F_U(F_Y(x)) = \frac{B(a, b; F_Y(x))}{B(a, b)} = I_{F_Y(x)}(a, b). \tag{2.3.1}$$

Η συνάρτηση πυκνότητας πιθανότητας προκύπτει με παραγώγιση της συνάρτησης κατανομής

$$f_X(x; \mathbf{p}, a, b) = \frac{dF_X(x)}{dx} = \frac{d}{dx}\Big(F_U\big(F_Y(x)\big)\Big) = f_U\big(F_Y(x)\big) \cdot \frac{dF_Y(x)}{dx} = f_U\big(F_Y(x)\big) \cdot f_Y(x)$$
$$= \frac{1}{B(a, b)} f_Y(x)[F_Y(x)]^{a-1}[1 - F_Y(x)]^{b-1}. \tag{2.3.2}$$

Σύμφωνα με τα παραπάνω, η F_X και η f_X είναι αντίστοιχα η συνάρτηση κατανομής και η συνάρτηση πυκνότητας μιας κατανομής παραγόμενης από τη βήτα κατανομή (*beta generated-BG*) με γεννήτορα κατανομή (*parent distribution*) την F_Y. Συμβολίζουμε την F_X και την f_X αντίστοιχα ως F_{BG} και f_{BG}.

Από τη BG κατανομή, προκύπτουν οι εξής περιπτώσεις:

(i) Αν $a = b = 1$ έχουμε τη γεννήτορα κατανομή.

(ii) Αν η κατανομή γεννήτορας είναι η Ομοιόμορφη στο διάστημα (0, 1) τότε από την (2.3.2) προκύπτει η συνάρτηση πυκνότητας πιθανότητας της κατανομής βήτα τύπου Ι.

(iii) Αν ο γεννήτορας έχει σ.κ. $F(x) = x^c$, προκύπει η γενικευμένη βήτα (GB) κατανομή πρώτου είδους με σ.π.π. που δίνεται από την (2.4.1).

(iv) Αν a και b είναι ακέραιοι αριθμοί, έχουμε την περίπτωση της κατανομής των διατεταγμένων παρατηρήσεων η οποία δίνεται στην (2.4.51).

Η πρώτη κατανομή που μελετήθηκε από την οικογένεια των βήτα παραγόμενων κατανομών είχε ως γεννήτορα την τυπική κανονική κατανομή (Eugene et al., (2002)). Δηλαδή, θεωρούμε ότι $Y \sim N(0,1)$, η οποία έχει σ.π.π. $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ και σ.κ. $\Phi(x) = P(Y \leq x)$. Τότε, σύμφωνα με τα παραπάνω η τυχαία μεταβλητή $X = \Phi^{-1}(U)$ έχει συνάρτηση πυκνότητας

$$f_{BN}(x; 0,1, a, b) = \frac{1}{B(a,b)} \varphi(x)[\Phi(x)]^{a-1}[1 - \Phi(x)]^{b-1}$$

$$= \frac{1}{B(a,b)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} [\Phi(x)]^{a-1}[1 - \Phi(x)]^{b-1}, \quad -\infty < x < \infty. \qquad (2.3.3)$$

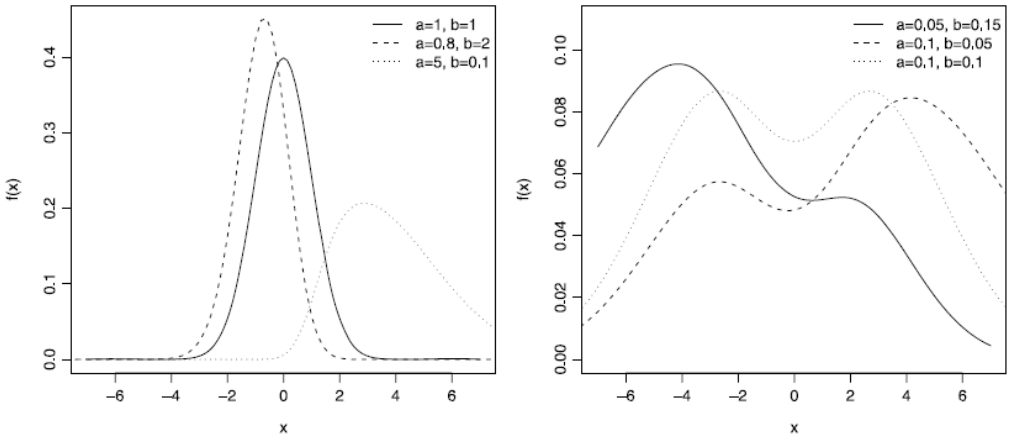
Στην περίπτωση που $X \sim BN(0,1,a,b)$ η τυχαία μεταβλητή που προκύπτει με το μετασχηματισμό $Y = \mu + \sigma X \sim BN(\mu, \sigma, a, b)$ έχει πυκνότητα

$$f_{BN}(y; \mu, \sigma, a, b) = \frac{1}{B(a,b)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \left[\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{a-1} \left[1 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{b-1}$$

δηλαδή έχει τη μορφή της βήτα παραγόμενης όταν η κατανομή γεννήτορας είναι η κανονική κατανομή $N(\mu, \sigma^2)$. Το αποτέλεσμα αυτό δικαιολογεί το γεγονός ότι οι παράμετροι θέσης και κλίμακας είναι πλεονάζουσες στην κατανομή γεννήτορα.

Οι παράμετροι a και b ελέγχουν τη λοξότητα της κατανομής μέσω των ουρών της κατανομής. Η βήτα-κανονική κατανομή είναι συμμετρική εάν $a = b$, έχει αρνητική λοξότητα αν $a < b$ και θετική λοξότητα αν $a > b$. Στην περίπτωση που $a = b > 1$, η *BN* κατανομή έχει θετική υπερβάλλουσα κύρτωση (excess kurtosis) και αν $a = b < 1$ έχει αρνητική υπερβάλλουσα κύρτωση. Εν τούτοις, οι τιμές της λοξότητας και της κύρτωσης στην υπό μελέτη κατανομή είναι περιορισμένες και ο μόνος τρόπος να επιτευχθεί έστω και μια μετρίου βαθμού κύρτωση είναι αν μέσω των τιμών των παραμέτρων αυξήσουμε τη λοξότητά της. Στη εργασία του ο Eugene δίνοντας τιμές στα a και b μεταξύ 0.05 και 100 έδειξε ότι η λοξότητα κινείται πάντα στο διάστημα (–1, 1). Επίσης, πέτυχε τη μεγαλύτερη τιμή της κύρτωσης 4.1825 για $a = 100$ και $b = 0.1$.

Στο Σχήμα 2.2 απεικονίζεται η σ.π.π. της βήτα παραγόμενης κατανομής με γεννήτορα την τυπική κανονική κατανομή για διάφορες τιμές των παραμέτρων a και b.



Σχήμα 2.2 Γραφική παράσταση συνάρτησης πυκνότητας $BN(0,1,a,b)$ για διάφορες τιμές των παραμέτρων a, b

Στη συνέχεια ο Jones (2004) αλλάζοντας τη γεννήτορα κατανομή από κανονική με μία οποιαδήποτε άλλη έδωσε μεγαλύτερη ευελιξία στις παραγόμενες κατανομές και τις ιδιότητές τους. Η μελέτη του εστιάστηκε σε συμμετρικές κατανομές γύρω από το μηδέν οι οποίες είχαν μόνο παράμετρο θέσης και κλίμακας και στήριγμα την ευθεία των πραγματικών. Τις βήτα παραγόμενες κατανομές μελέτησαν μεταξύ άλλων οι Nadarajah και Kotz (2004, 2005), ο Akinsete et al. (2008), οι Zografos και Balakrishnan (2009) και Barreto-Souza et al. (2010). Οι Jones και Larsen (2004) και Arnold et al. (2006) εισήγαγαν την κλάση των πολυδιάστατων BG κατανομών. Εφαρμογές των BG κατανομών μελετήθηκαν από τους Jones και Larsen (2004) σε δεδομένα θερμοκρασιών προσαρμόζοντας λοξή t και $logF$ κατανομές, από τον Akinsete et al. (2008) σε δεδομένα πλημμυρών προσαρμόζοντας βήτα-Pareto κατανομή, από το Razzaghi (2009) στη μοντελοποίηση δόσης-απόκρισης με χρήση βήτα-κανονικής κατανομής κ.ά.

Παρ' όλα αυτά, η οικογένεια BG έχει μόνο δύο παραμέτρους και συνεπώς δημιουργεί μια περιορισμένη ευελιξία στην κατανομή γεννήτορα. Για διάφορες επιλογές της κατανομής γεννήτορα ο υπολογισμός των ποσοστημορίων και των ροπών της BG κατανομής που προκύπτει μπορεί να γίνει αρκετά πολύπλοκος. Επίσης στην περίπτωση $a = b$, η λοξότητα είναι μηδέν για συμμετρική κατανομή γεννήτορα, η BG κατανομή που προκύπτει έχει μικρότερη κύρτωση από τη γεννήτορα κατανομή. Για παράδειγμα, χρησιμοποιώντας γεννήτορα την κατανομή Student (t) και για $a = b > 1$ προκύπτει ότι η κύρτωση συγκλίνει γρήγορα στο 3 καθώς a και b αυξάνουν, ενώ αν $a = b < 1$ η κύρτωση είναι άπειρη.

Ο Jones (2009) αλλάζοντας την κατανομή βήτα με την Kumaraswamy (1980) κατανομή, που αναφέρεται και ως «minimax» κατανομή, έδωσε την Kumaraswamy παραγόμενη (KwG) οικογένεια κατανομών, η οποία έχει ελκυστικές ιδιότητες στην προσομοίωση καθώς και μια απλή μορφή στην ποσοστιαία συνάρτησή της. Όμως και η KwG οικογένεια έχει δύο

παραμέτρους, ενώ για να πετύχουμε έλεγχο στις ουρές και στο κέντρο της κατανομής ενδεχομένως να χρειαστεί μια τρίτη παράμετρος.

Η ανάγκη λοιπόν για περαιτέρω έλεγχο του βάρους των ουρών των κατανομών καθώς και της λοξότητας και κύρτωσης αυτών οδήγησε στην ανάπτυξη της οικογένειας των γενικευμένων βήτα παραγόμενων κατανομών (Generalized beta generated family of distributions). Η τελευταία είναι μια περισσότερο ευέλικτη οικογένεια κατανομών καθώς έχει μια επιπλέον παράμετρο μορφής από τις κατανομές BG και KwG, η οποία δίνει τη δυνατότητα ελέγχου της λοξότητας και κύρτωσης της παραγόμενης κατανομής. Η GBG οικογένεια περιλαμβάνει μεταξύ άλλων ως περιπτώσεις της και τις δύο παραπάνω οικογένειες κατανομών, δηλαδή την BG και την KwG.

2.4 Γενικευμένη βήτα παραγόμενη οικογένεια κατανομών

Η γενικευμένη βήτα παραγόμενη (generalized beta generated – GBG) κατανομή προτάθηκε και μελετήθηκε από τους Alexander et al. (2011). Η ιδέα προέκυψε από τη γενικευμένη βήτα (GB) κατανομή πρώτου είδους, η οποία εμφανίστηκε από τον McDonald (1984), με συνάρτηση πυκνότητας πιθανότητας

$$f_{GB}(x;a,b,c) = c\frac{1}{B(a,b)}x^{ac-1}(1-x^c)^{b-1},\ \ 0<x<1, a>0, b>0, c>0. \qquad (2.4.1)$$

Ειδικές περιπτώσεις της GB κατανομής είναι

(i) Η κατανομή $Beta^{(I)}(a,b)$, η οποία προκύπτει για $c=1$.

(ii) Η κατανομή Kumaraswamy με σ.π.π.

$$f_{GB}(x;1,b,c) = cbx^{c-1}(1-x^c)^{b-1}, \qquad 0<x<1 \qquad (2.4.2)$$

η οποία προκύπτει για $a=1$.

Η γενικευμένη βήτα παραγόμενη κατανομή προκύπτει αν θεωρήσουμε μία κατανομή γεννήτορα με συνάρτηση κατανομής F, συνάρτηση πυκνότητας f και διάνυσμα παραμέτρων $\mathbf{p}$. Η σ.π.π. της GBG κατ' αναλογία της (2.4.1) δίνεται από τη σχέση

$$f_{GBG}(x;\mathbf{p},a,b,c) = c\frac{1}{B(a,b)}f(x;\mathbf{p})[F(x;\mathbf{p})]^{ac-1}\,(1-[F(x;\mathbf{p})]^c\,)^{b-1}. \qquad (2.4.3)$$

Εάν μια τυχαία μεταβλητή X έχει την παραπάνω συνάρτηση πυκνότητας, θα γράφουμε $X\sim GBG(F;\mathbf{p},a,b,c)$ ή απλά $X\sim GBG$.

Οι επιπλέον παράμετροι a, b και c του μοντέλου, πέραν αυτών του γεννήτορα, είναι παράμετροι σχήματος. Η εισαγωγή παραμέτρων θέσης ή κλίμακας στην κατανομή γεννήτορα είναι πλεονάζουσες. Αυτό μπορεί να φανεί αν θεωρήσουμε την τ.μ. $Y = \mu + \sigma X$. Τότε η πυκνότητα της Y μπορεί να δειχθεί ότι είναι

$$f_{GBG}(y; \mathbf{p}, \mu, \sigma, a, b, c) = \frac{c}{\sigma B(a,b)} f\left(\frac{y-\mu}{\sigma}\right)\left[F\left(\frac{y-\mu}{\sigma}\right)\right]^{ac-1}\left(1-\left[F\left(\frac{y-\mu}{\sigma}\right)\right]^{c}\right)^{b-1}.$$

Αν $\tilde{\mu}$ είναι μέση τιμή και $\tilde{\sigma}^2$ η διασπορά της X, τότε για την Y τα αντίστοιχα μεγέθη είναι

$$E(Y) = E(\mu + \sigma X) = \mu + \sigma E(X) = \mu + \sigma\tilde{\mu}$$

και

$$Var(Y) = Var(\mu + \sigma X) = \sigma^2 Var(X) = \sigma^2\tilde{\sigma}^2.$$

Ο τύπος (2.4.3) μπορεί να θεωρηθεί ως η σ.π.π. της τυχαίας μεταβλητής X που προκύπτει από το μετασχηματισμό

$$X = F^{-1}\left(U^{1/c}\right) \tag{2.4.4}$$

όπου $U \sim Beta^{(I)}(a, b)$ και F η συνάρτηση κατανομής μιας αυθαίρετης κατανομής γεννήτορα.

Η οικογένεια GBG περιέχει ως μέλη της διάφορες κατανομές που αναφέρουμε στη συνέχεια:

(i) Για $c = 1$ προκύπτει η βήτα παραγόμενη (BG) οικογένεια κατανομών με συνάρτηση πυκνότητας πιθανότητας

$$f_{GBG}(x; \mathbf{p}, a, b, 1) = \frac{1}{B(a,b)} f(x; \mathbf{p})[F(x; \mathbf{p}]^{a-1}[1 - F(x; \mathbf{p})]^{b-1}. \tag{2.4.5}$$

(ii) Για $a = 1$ προκύπτει η Kumaraswamy παραγόμενη (Kumaraswamy generated - KwG) κατανομή (Cordeiro και de Castro, 2011) με συνάρτηση πυκνότητας πιθανότητας

$$f_{GBG}(x; \mathbf{p}, 1, b, c) = cbf(x; \mathbf{p})[F(x; \mathbf{p}]^{c-1}(1 - [F(x; \mathbf{p})]^{c})^{b-1}. \tag{2.4.6}$$

(iii) Για $b = 1$ προκύπτει η εκθετικοποιημένη (exponentiated) F κατανομή, ή αλλιώς η κατανομή Lehmann τύπου I, με συνάρτηση πυκνότητας πιθανότητας

$$f_{GBG}(x; \mathbf{p}, \alpha, 1, c) = \alpha cf(x; \mathbf{p})[F(x; \mathbf{p})]^{\alpha c-1}. \tag{2.4.7}$$

(iv) Για $a = c = 1$ προκύπτει η εκθετικοποιημένη $1 - F$ κατανομή, ή αλλιώς η κατανομή Lehmann τύπου II, με συνάρτηση πυκνότητας πιθανότητας

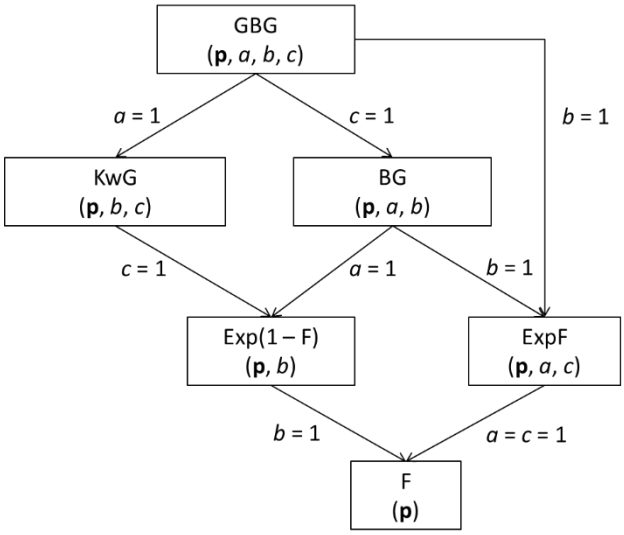
$$f_{GBG}(x; \mathbf{p}, 1, b, 1) = bf(x; \mathbf{p})[1 - F(x; \mathbf{p})]^{b-1}. \tag{2.4.8}$$

(v) Για $a = b = c = 1$ προκύπτει η κατανομή γεννήτορας

$$f_{GBG}(x; \mathbf{p}, 1,1,1) = f(x; \mathbf{p}). \tag{2.4.9}$$

(vi) Αν η κατανομή γεννήτορας είναι η ομοιόμορφη στο διάστημα (0, 1) τότε προκύπτει η GB κατανομή πρώτου είδους της Σχέσης (2.4.1).

Στο Σχήμα 2.3 απεικονίζονται διάφορα μέλη της GBG οικογένειας κατανομών



Σχήμα 2.3 Οικογένεια κατανομών GBG

Η συνάρτηση κατανομής της GBG δίνεται από τον τύπο

$$F_{GBG}(x;\mathbf{p},a,b,c) = P(X \le x) = P\left(F^{-1}\left(U^{1/c}\right) \le x\right) = P\left(U^{1/c} \le F(x;\mathbf{p})\right)$$
$$= P(U \le [F(x;\mathbf{p})]^c) = F_U([F(x;\mathbf{p})]^c)$$
$$= \int_0^{[F(x;\mathbf{p})]^c} \frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1}dt$$
$$= \frac{B(a,b;[F(x;\mathbf{p})]^c)}{B(a,b)}$$
$$= I_{[F(x;\mathbf{p})]^c}(a,b) \qquad (2.4.10)$$

όπου η $I_y(a,b) = \frac{B(a,b;y)}{B(a,b)}$ είναι ο λόγος της συνάρτησης Βήτα.

2.4.1 Αναπτύγματα της συνάρτησης κατανομής και της συνάρτησης πυκνότητας

Στην ενότητα αυτή θα παρουσιασθούν διάφοροι τρόποι με βάση τους οποίους μπορεί να εκφραστεί η συνάρτηση κατανομής και η συνάρτηση πυκνότητας της GBG κατανομής.

Αρχικά δίνονται κάποιες ποσότητες που θα φανούν χρήσιμες στην ανάλυση που θα ακολουθήσει.

Για $|z| < 1$ και $b > 0$ έναν πραγματικό μη ακέραιο αριθμό η ποσότητα $(1-z)^{b-1}$ μπορεί να αναπτυχθεί σε δυναμοσειρά ως εξής

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} z^j. \qquad (2.4.11\alpha)$$

Στην περίπτωση που το b είναι ακέραιος, ο δείκτης j στο άθροισμα σταματάει στο $b-1$.

Από τη γενική σχέση των διωνυμικών συντελεστών με τη συνάρτηση Γάμμα,

$$\binom{x}{y} = \frac{x!}{y!\,(x-y)!} = \frac{\Gamma(x+1)}{\Gamma(y+1)\Gamma(x-y+1)}$$

η παραπάνω σειρά γράφεται ως

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(b)}{\Gamma(b-j)j!}\, z^j. \qquad (2.4.11\beta)$$

Με χρήση της Σχέσης (2.4.11α) γράφουμε το ολοκλήρωμα που εμφανίζεται στη συνάρτηση κατανομής της κατανομής $Beta^{(I)}(a,b)$ σε μορφή αθροίσματος ως εξής

$$\int_0^x t^{a-1}(1-t)^{b-1}dt = \int_0^x t^{a-1} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} t^j dt = \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \int_0^x t^{a+j-1}dt.$$

Το ολοκλήρωμα του αθροίσματος είναι $\int_0^x t^{a+j-1}dt = \left[\frac{t^{a+j}}{a+j}\right]_0^x = \frac{x^{a+j}}{a+j}$ οπότε

$$\int_0^x t^{a-1}(1-t)^{b-1}dt = \sum_{j=0}^{\infty} \frac{(-1)^j \binom{b-1}{j}}{a+j} x^{a+j} = \sum_{j=0}^{\infty} \frac{(-1)^j}{a+j} \frac{\Gamma(b)}{j!\,\Gamma(b-j)} x^{a+j}. \qquad (2.4.11\gamma)$$

Για b πραγματικό μη ακέραιο το ανάπτυγμα του z^b σε σειρά, με χρήση της (2.4.11α) γράφεται διαδοχικά

$$z^b = [1-(1-z)]^{(b+1)-1} = \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} (1-z)^i = \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} \left(\sum_{k=0}^{i} (-1)^k \binom{i}{k} z^k \right)$$

$$= \sum_{i=0}^{\infty} \sum_{k=0}^{i} (-1)^{i+k} \binom{b}{i} \binom{i}{k} z^k.$$

Αντικαθιστώντας το $\sum_{i=0}^{\infty} \sum_{k=0}^{i}$ με $\sum_{k=0}^{\infty} \sum_{i=k}^{\infty}$ προκύπτει

$$z^b = \sum_{k=0}^{\infty} \sum_{i=k}^{\infty} (-1)^{i+k} \binom{b}{i} \binom{i}{k} z^k = \sum_{k=0}^{\infty} t_k(b) z^k \qquad (2.4.11\delta)$$

με $t_k(b) = \sum_{i=k}^{\infty} (-1)^{i+k} \binom{b}{i} \binom{i}{k}$.

Η συνάρτηση κατανομής και η συνάρτηση πυκνότητας της γενικευμένης βήτα παραγόμενης κατανομής μπορεί να πάρουν διάφορες μορφές τις οποίες και χρησιμοποιούμε ανάλογα με το

σκοπό μας. Οι διαφορετικές εκφράσεις συνοψίζονται στις περιπτώσεις όπου το b ή το a είναι ακέραιος, μέσω γραμμικού συνδυασμού εκθετικοποιημένων-F κατανομών και τέλος μέσω της υπεργεωμετρικής συνάρτησης. Αυτά αναπτύσσονται στη συνέχεια.

Όταν το b είναι ακέραιος, η συνάρτηση κατανομής της GBG που δίνεται στην (2.4.10), γράφεται (παράρτημα Π5)

$$F_{GBG}(x;\mathbf{p},a,b,c)=\sum_{j=0}^{b-1}\binom{a+b-1}{j}[F(x;\mathbf{p})]^{c(a+b-j-1)}(1-[F(x;\mathbf{p})]^{c})^{j}. \qquad (2.4.12)$$

Για να βρούμε μια ανάλογη σχέση για a ακέραιο, από τη συμμετρική σχέση του λόγου της συνάρτησης Βήτα, $I_y(a,b)=1-I_{1-y}(b,a)$, η συνάρτηση κατανομής της GBG γράφεται

$$F_{GBG}(x;\mathbf{p},a,b,c)=I_{[F(x;\mathbf{p})]^{c}}(a,b)=1-I_{1-[F(x;\mathbf{p})]^{c}}(b,a)$$

$$=1-\sum_{j=0}^{a-1}\binom{a+b-1}{j}[F(x;\mathbf{p})]^{cj}(1-[F(x;\mathbf{p})]^{c})^{a+b-j-1}. \qquad (2.4.13)$$

Από τη συνάρτηση κατανομής των Σχέσεων (2.4.12) και (2.4.13) μπορούμε με παραγώγιση να πάρουμε αντίστοιχες σχέσεις για τη συνάρτηση πυκνότητας πιθανότητας της GBG κατανομής.

Οι συναρτήσεις κατανομής και πυκνότητας πιθανότητας της GBG επίσης μπορούν να δοθούν ως γραμμικός συνδυασμός (μείξη) εκθετικοποιημένων-F κατανομών.

Η έννοια της μείξης είναι από τις πλέον σημαντικές για τη μοντελοποίηση των μεγεθών ατομικών ζημιών. Ενδεικτικά αναφέρουμε την περίπτωση στην οποία μια τ.μ. X είναι πεπερασμένη διακριτή μείξη των τ.μ. $X_1, X_2, \dots, X_n$ με αντίστοιχα βάρη μίξης $a_1, a_2, \dots, a_n$ όπου $a_i>0$ και $\sum_{i=1}^{n}a_i=1$. Αν f_i είναι η σ.π.π. και F_i είναι η σ.κ. της X_i, τότε η συνάρτηση πυκνότητας πιθανότητας, η συνάρτηση κατανομής, η συνάρτηση δεξιάς ουράς, η ροπή k-τάξης και η ροπογεννήτρια συνάρτηση της X δίνονται αντίστοιχα από τους τύπους

$$f_X(x)=\sum_{i=1}^{n}a_i f_i(x), \qquad F_X(x)=\sum_{i=1}^{n}a_i F_i(x), \qquad \bar{F}_X(x)=\sum_{i=1}^{n}a_i \bar{F}_i(x),$$

$$E\left(X^k\right)=\sum_{i=1}^{n}a_i E\left(X_i^k\right), \qquad M_X(t)=\sum_{i=1}^{n}a_i M_i(t).$$

Σχετικά, με την εκθετικοποιημένη-F κατανομή θεωρούμε ότι μια τυχαία μεταβλητή X θα την ακολουθεί, και συμβολίζουμε $X\sim Exp^{(a)}F$, όταν για μια αυθαίρετη κατανομή $F(x;\mathbf{p})$ και μία σταθερά $a>0$, η συνάρτηση πυκνότητας πιθανότητας και η συνάρτηση κατανομής της X είναι αντίστοιχα

$$g_a(x)=af(x;\mathbf{p})[F(x;\mathbf{p})]^{a-1} \quad \text{και} \quad G_a(x)=[F(x;\mathbf{p})]^{a}. \qquad (2.4.14)$$

Για παράδειγμα, στην περίπτωση της εκθετικής κατανομής με σ.π.π. $f(x) = \lambda e^{-\lambda x}$ και σ.κ. $F(x) = 1 - e^{-\lambda x}$, η εκθετικοποιημένη εκθετική κατανομή έχει σ.π.π. και σ.κ. που δίνονται αντίστοιχα από τις σχέσεις

$$g_a(x) = a\lambda e^{-\lambda x}\left(1 - e^{-\lambda x}\right)^{a-1} \quad \text{και} \quad G_a(x) = \left(1 - e^{-\lambda x}\right)^a. \tag{2.4.15}$$

Με χρήση της (2.4.11γ) η συνάρτηση κατανομής της GBG από την (2.4.10) για $b > 0$ πραγματικό μή-ακέραιο, γράφεται

$$\begin{aligned} F_{GBG}(x; \mathbf{p}, a, b, c) &= \frac{1}{B(a,b)} \int_0^{[F(x;\mathbf{p})]^c} t^{a-1}(1-t)^{b-1} dt \\ &= \frac{1}{B(a,b)} \sum_{j=0}^{\infty} \frac{(-1)^j \binom{b-1}{j}}{a+j} [F(x;\mathbf{p})]^{c(a+j)} \\ &= \sum_{j=0}^{\infty} \frac{(-1)^j \binom{b}{j}}{B(a,b+1)(a+j)} [F(x;\mathbf{p})]^{c(a+j)} \\ &= \sum_{j=0}^{\infty} w_j G_{c(a+j)}(x) \end{aligned} \tag{2.4.16}$$

όπου

$G_{c(a+j)}(x) = [F(x;\mathbf{p})]^{c(a+j)}$

και

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

Παραγωγίζοντας την (2.4.16) βρίσκουμε τη συνάρτηση πυκνότητας πιθανότητας

$$\begin{aligned} f_{GBG}(x; \mathbf{p}, a, b, c) &= \frac{d}{dx} F_{GBG}(x; \mathbf{p}, a, b, c) \\ &= \sum_{j=0}^{\infty} w_j \frac{d}{dx} G_{c(a+j)}(x) = \sum_{j=0}^{\infty} w_j g_{c(a+j)}(x). \end{aligned} \tag{2.4.17}$$

Συγκεκριμένα, είναι

$$g_{c(a+j)}(x) = \frac{d}{dx} G_{c(a+j)}(x) = c(a+j) f(x;\mathbf{p}) [F(x;\mathbf{p})]^{c(a+j)-1}$$

και με αντικατάσταση των βαρών w_j στην (2.4.17) προκύπτει

$$f_{GBG}(x; \mathbf{p}, a, b, c) = \frac{c}{B(a,b+1)} f(x;\mathbf{p}) \sum_{j=0}^{\infty} (-)^j \binom{b}{j} [F(x;\mathbf{p})]^{c(a+j)-1}. \tag{2.4.18}$$

Για τις σταθερές w_j ισχύει $\sum_{j=0}^{\infty} w_j = 1$, αφού για τη συνάρτηση πυκνότητας πιθανότητας της GBG είναι

$$1 = \int_x f_{GBG}(x; \mathbf{p}, a, b, c)dx = \int_x \sum_{j=0}^{\infty} w_j g_{c(a+j)}(x)\, dx = \sum_{j=0}^{\infty} w_j \int_x g_{c(a+j)}dx = \sum_{j=0}^{\infty} w_j \cdot 1.$$

Με βάση τα παραπάνω, η $X \sim GBG$ είναι μια μείξη των $Y_j \sim Exp^{[c(a+j)]}F(x; \mathbf{p})$ με αντίστοιχα βάρη w_j. Η δυνατότητα έκφρασης της GBG ως γραμμικό συνδυασμό εκθετικοποιημένων-F κατανομών, μας επιτρέπει την ανάπτυξη ιδιοτήτων της μέσω των ανάλογων ιδιοτήτων των εκθετικοποιημένων κατανομών. Για παράδειγμα, η συνάρτηση δεξιάς ουράς της GBG, είναι

$$\bar{F}_{GBG}(x; \mathbf{p}, a, b, c) = 1 - F_{GBG}(x; \mathbf{p}, a, b, c) = \sum_{j=0}^{\infty} w_j \bar{G}_{c(a+j)}(x) \tag{2.4.19}$$

όπου

$\bar{G}_{c(a+j)}(x) = 1 - G_{c(a+j)}(x) = 1 - [F(x; \mathbf{p})]^{c(a+j)}$.

Τέλος, η GBG μπορεί να εκφραστεί μέσω της υπεργεωμετρικής συνάρτησης Gauss,

$${}_2F_1(x; \alpha, \beta; \gamma) = \sum_{i=0}^{\infty} \frac{(\alpha)_i (\beta)_i}{(\gamma)_i} \frac{x^i}{i!}, \qquad |x| < 1$$

όπου με $(x)_i$ συμβολίσαμε το ανοδικό παραγοντικό i τάξης

$$(x)_i = \frac{\Gamma(x+i)}{\Gamma(x)} = x(x+1)\cdots(x+i-1).$$

Η υπεργεωμετρική συνάρτηση γράφεται εναλλακτικά και στη μορφή

$${}_2F_1(x; \alpha, \beta; \gamma) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{i=0}^{\infty} \frac{\Gamma(\alpha+i)\Gamma(\beta+i)}{\Gamma(\gamma+i)} \frac{x^i}{i!}.$$

Με βάση τη σχέση που συνδέει τη μη πλήρη συνάρτηση Βήτα με την υπεργεωμετρική συνάρτηση,

$$B(a, b; x) = \int_0^x t^{a-1}(1-t)^{b-1}dt = \frac{x^a}{a}\ {}_2F_1(x; a, 1-b; a+1),$$

η συνάρτηση κατανομής της GBG, της Σχέσης (2.4.10) εκφράζεται ως

$$F_{GBG}(x; \mathbf{p}, a, b; c) = \frac{B(a, b; [F(x; \mathbf{p})]^c)}{B(a, b)}$$
$$= \frac{[F(x; \mathbf{p})]^{ac}}{aB(a, b)}\ {}_2F_1([F(x; \mathbf{p})]^c; a, 1-b; a+1). \tag{2.4.20}$$

Από την (2.4.20), εκφράζοντας την GBG μέσω της υπεργεωμετρικής συνάρτησης, οι ιδιότητες της GBG για οποιαδήποτε κατανομή γεννήτορα F, μπορούν να εξαχθούν από τις ιδιότητες της

υπεργεωμετρικής συνάρτησης, που δίνονται για παράδειγμα στους Gradshteyen and Ryzhik (2000).

2.4.2 Χαρακτηριστικά μεγέθη κατανομής

Η συνάρτηση διακινδύνευσης ή συνάρτηση κινδύνου (hazard rate), $h(x)$, ορίζεται ως ο λόγος της συνάρτησης πυκνότητας πιθανότητας προς τη συνάρτηση δεξιάς ουράς της κατανομής,

$$h(x) = \lim_{dx \to 0} \frac{\frac{\bar{F}(x) - \bar{F}(x+dx)}{\bar{F}(x)}}{dx} = \frac{f(x)}{\bar{F}(x)}\,. \tag{2.4.21}$$

Η συνάρτηση κινδύνου εκφράζει την τάση του προς μελέτη αντικειμένου να αποτύχει στο χρονικό διάστημα $(x, x+dx]$ δεδομένου ότι έχει επιζήσει μέχρι τη χρονική στιγμή x. Όταν η συνάρτηση διακινδύνευσης είναι φθίνουσα τότε η στιγμιαία πιθανότητα αποτυχίας μειώνεται με το πέρασμα του χρόνου. Στην αντίθετη περίπτωση, όταν η συνάρτηση διακινδύνευσης είναι αύξουσα τότε η στιγμιαία πιθανότητα αποτυχίας αυξάνεται με την πάροδο του χρόνου.

Η συνάρτηση κινδύνου για την GBG κατανομή δίνεται από τη σχέση

$$\begin{aligned} h(x;\mathbf{p},a,b,c) &= \frac{f_{GBG}(x)}{\bar{F}_{GBG}(x)} \\ &= \frac{c\frac{1}{B(a,b)} f(x;\mathbf{p})[F(x;\mathbf{p})]^{ac-1}(1-[F(x;\mathbf{p})]^c)^{b-1}}{1-\frac{B(a,b;[F(x;\mathbf{p})]^c)}{B(a,b)}} \\ &= \frac{cf(x;\mathbf{p})[F(x;\mathbf{p})]^{ac-1}(1-[F(x;\mathbf{p})]^c)^{b-1}}{B(a,b)-B(a,b;[F(x;\mathbf{p})]^c)}. \end{aligned} \tag{2.4.22}$$

Στη συνέχεια αναπτύσσονται διάφορες εκφράσεις για τις ροπές s-τάξης περί την αρχή

$$\mu'_s = E(X^s). \tag{2.4.23}$$

Μια πρώτη έκφραση προκύπτει από τη Σχέση (2.4.4). Με αντικατάσταση της $X = F^{-1}\left(U^{1/c}\right)$ στην (2.4.23) και από τη σ.π.π. της $U \sim Beta^{(I)}(a,b)$, παίρνουμε

$$\mu'_s = E\left\{\left[F^{-1}\left(U^{1/c}\right)\right]^s\right\} = \frac{1}{B(a,b)}\int_0^1 \left[F^{-1}\left(x^{1/c}\right)\right]^s x^{a-1}(1-x)^{b-1}dx.$$

Θεωρώντας τον μετασχηματισμό $u = x^{1/c}$ είναι $x = u^c$ και $dx = cu^{c-1}du$. Επίσης για $0 < x < 1$ είναι $0 < u < 1$, οπότε

$$E(X^s) = \frac{1}{B(a,b)} \int_0^1 [F^{-1}(u)]^s u^{ca-c}(1-u^c)^{b-1} c u^{c-1} du$$

$$= \frac{c}{B(a,b)} \int_0^1 [F^{-1}(u)]^s u^{ca-1}(1-u^c)^{b-1} du. \qquad (2.4.24)$$

Μία προσεγγιστική έκφραση για τη μ_s' προκύπτει αν θεωρήσουμε το ανάπτυγμα Taylor της $F^{-1}(u)$ γύρω από το μέσο του γεννήτορα $\mu_F = E(X_F)$,

$$\mu_s' \approx \sum_{k=0}^{s} \binom{s}{k} [F^{-1}(\mu_F)]^{s-k} \left[F^{-1(1)}(\mu_F)\right]^k \sum_{i=0}^{k} (-1)^i \binom{k}{i} \mu_F^i \frac{B\left(a + \frac{k-i}{c}, b\right)}{B(\alpha, b)} \qquad (2.4.25)$$

όπου η πρώτη παράγωγος της $F^{-1}(u)$ είναι

$$F^{-1(1)}(u) = \frac{d}{du} F^{-1}(u) = \frac{1}{F'\left(F^{-1}(u)\right)} = \frac{1}{f(F^{-1}(u))} = [f(F^{-1}(u))]^{-1}.$$

Μία δεύτερη έκφραση για την μ_s' δίνεται μέσω των σταθμισμένων ροπών (probability weighted moments) οι οποίες όταν η τ.μ. Y ακολουθεί την κατανομή γεννήτορα και $s, r = 0,1,2, \ldots$, δίνονται από τον παρακάτω τύπο (Greenwood et al., (1979)),

$$\tau(s,r) = E(Y^s [F(Y)]^r) = \int_{-\infty}^{\infty} y^s [F(y)]^r f(y) dy. \qquad (2.4.26)$$

Από τη Σχέση (2.4.18) βρίσκουμε

$$\mu_s' = E(X^s) = \int_{-\infty}^{\infty} x^s f_{GBG}(x; \mathbf{p}, a, b, c) dx$$

$$= \int_{-\infty}^{\infty} x^s \frac{c}{B(a, b+1)} f(x; \mathbf{p}) \sum_{j=0}^{\infty} (-)^j \binom{b}{j} [F(x; \mathbf{p})]^{c(a+j)-1} \, dx$$

$$= \frac{c}{B(a, b+1)} \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} x^s [F(x; \mathbf{p})]^{c(a+j)-1} f(x; \mathbf{p}) \, dx$$

$$= \frac{c}{B(a, b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \int_{-\infty}^{\infty} x^s [F(x; \mathbf{p})]^{c(a+j)-1} f(x; \mathbf{p}) dx. \qquad (2.4.27)$$

Για a και c ακέραιους η (2.4.27) με χρήση της (2.4.26) γίνεται

$$\mu'_s = \frac{c}{B(a,b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \tau(s, c(a+j)-1). \tag{2.4.28}$$

Όταν το b είναι ακέραιος, ο δείκτης j στο άθροισμα σταματάει στο $b-1$.

Στην περίπτωση που ένα τουλάχιστον από τα a και c είναι πραγματικός μη-ακέραιος, χρησιμοποιούμε το ανάπτυγμα της $[F(x;\mathbf{p})]^q$ σε σειρά. Από τη (2.4.11δ) προκύπτει,

$$[F(x;\mathbf{p})]^q = \sum_{k=0}^{\infty} t_k(q)[F(x;\mathbf{p})]^k, \qquad \text{με} \quad t_k(q) = \sum_{i=k}^{\infty} (-1)^{i+k} \binom{q}{i}\binom{i}{k}. \tag{2.4.29}$$

Με βάση την παραπάνω από την (2.4.27) παίρνουμε

$$\mu'_s = E(X^s)$$

$$= \frac{c}{B(a,b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \int_{-\infty}^{\infty} x^s \left(\sum_{k=0}^{\infty} \sum_{i=k}^{\infty} (-1)^{i+k} \binom{c(a+j)-1}{i} \binom{i}{k} [F(x;\mathbf{p})]^k \right) f(x;\mathbf{p}) dx$$

$$= \frac{c}{B(a,b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \sum_{k=0}^{\infty} \sum_{i=k}^{\infty} (-1)^{i+k} \binom{c(a+j)-1}{i} \binom{i}{k} \int_{-\infty}^{\infty} x^s [F(x;\mathbf{p})]^k f(x;\mathbf{p}) dx$$

με χρήση της (2.4.26) οι ροπές γράφονται

$$\mu'_s = \frac{c}{B(a,b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \sum_{k=0}^{\infty} \sum_{i=k}^{\infty} (-1)^{i+k} \binom{c(a+j)-1}{i} \binom{i}{k} \tau(s,k)$$

$$= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{i=k}^{\infty} \frac{c}{B(a,b+1)} (-1)^{i+j+k} \binom{c(a+j)-1}{i} \binom{i}{k} \binom{b}{j} \tau(s,k)$$

$$= \sum_{k=0}^{\infty} \frac{c}{B(a,b+1)} \sum_{j=0}^{\infty} \sum_{i=k}^{\infty} (-1)^{i+j+k} \binom{c(a+j)-1}{i} \binom{i}{k} \binom{b}{j} \tau(s,k)$$

$$= \sum_{k=0}^{\infty} \upsilon_k \cdot \tau(s,k), \tag{2.4.30}$$

όπου

$$\upsilon_k = \frac{c}{B(a,b+1)} \sum_{j=0}^{\infty} \sum_{i=k}^{\infty} (-1)^{i+j+k} \binom{c(a+j)-1}{i} \binom{i}{k} \binom{b}{j}.$$

Μια τρίτη σχέση για τις ροπές $E(X^s)$ προκύπτει μέσω της ποσοστιαίας συνάρτησης της κατανομής γεννήτορα, $Q_F(u) = F^{-1}(u)$.

Αρχικά, από την (2.4.26), θέτοντας $u = F(y)$ είναι $y = F^{-1}(u)$, με $du = f(y)dy$ και για $-\infty < y < \infty$ είναι $0 < x < 1$ οπότε προκύπτει

$$\tau(s,r) = \int_0^1 [F^{-1}(u)]^s u^r du = \int_0^1 [Q_F(u)]^s u^r du. \quad (2.4.31)$$

Έτσι στη (2.4.28) για τις σταθμισμένες ροπές μπορεί να χρησιμοποιηθεί η παραπάνω σχέση.

Τέλος, μια τέταρτη έκφραση για τις ροπές μ_s' δίνεται μέσω των εκθετικοποιημένων-F κατανομών, μέσω της (2.4.17), δηλαδή

$$\mu_s' = \sum_{j=0}^{\infty} w_j E\left(Y_j^s\right), \quad (2.4.32)$$

όπου $Y_j \sim Exp^{[c(a+j)]}F(x;\mathbf{p})$.

Όπως ήδη έχει αναφερθεί, η εισαγωγή των παραμέτρων a, b και c στο υπό θεώρηση μοντέλο κατανομή, έχει σαν σκοπό τον έλεγχο της λοξότητας και της κύρτωσης της παραγόμενης GBG κατανομής. Στην ενότητα αυτή θα μελετήσουμε περισσότερο την επίδραση των παραμέτρων σε αυτά τα χαρακτηριστικά.

Η λοξότητα (skewness) μιας τυχαίας μεταβλητής X ορίζεται ως

$$\gamma_1 = skew(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \quad (2.4.33)$$

όπου $\mu = E(X)$ και $\sigma^2 = Var(X)$.

Η κατανομή έχει θετική, αρνητική ή μηδενική λοξότητα αν αντίστοιχα η ποσότητα $skew(X)$ είναι θετική, αρνητική ή μηδέν. Θετική ή αλλιώς δεξιά λοξότητα σημαίνει ότι οι περισσότερες τιμές της μεταβλητής βρίσκονται δεξιά της μέσης τιμής. Αρνητική ή αριστερή λοξότητα σημαίνει μεγαλύτερη έκταση της καμπύλης της κατανομής προς τα αριστερά της μέσης τιμής. Η λοξότητα συναρτήσει των ροπών της X δίνεται από τη σχέση

$$skew(X) = \frac{E(X^3) - 3\mu E(X^3) + 2\mu^3}{\sigma^3}. \quad (2.4.34)$$

Ένα μειονέκτημα που έχει ο παραπάνω τύπος της λοξότητας είναι ότι για κατανομές με βαριά ουρά για τις οποίες δεν υπάρχουν οι ροπές από κάποια τάξη και πάνω, το μέτρο αυτό έχει άπειρη τιμή. Ένας τύπος για τη λοξότητα που είναι λιγότερο ευαίσθητος στις ακραίες τιμές και μας επιτρέπει να αποφύγουμε το παραπάνω μειονέκτημα είναι βασισμένος στα τεταρτημόρια και δίνεται από τη σχέση [Kenney και Keeping, (1992)]

$$B = \frac{Q\left(\frac{3}{4}\right) - 2Q\left(\frac{1}{2}\right) + Q\left(\frac{1}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}. \quad (2.4.35)$$

Η κύρτωση (kurtosis) μιας τυχαίας μεταβλητής X ορίζεται ως

$$\beta_2 = kurt(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right], \quad (2.4.36)$$

ενώ συναρτήσει των ροπών της X δίνεται από τη σχέση

$$kurt(X) = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^3) - 3\mu^4}{\sigma^4}. \quad (2.4.37)$$

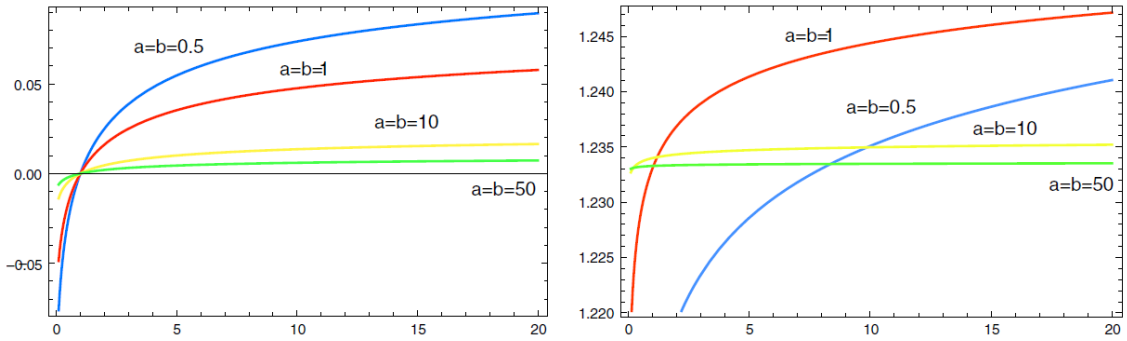
Η υπερβάλλουσα κύρτωση (excess kurtosis) ορίζεται ως η ποσότητα $\gamma_2 = kurt(X) - 3$.
Ο όρος κύρτωση δηλώνει τη σχέση του πλάτους της κατανομής γύρω από την κεντρική τιμή με τις ουρές της. Η τιμή $\beta_2 = 3$ ή $\gamma_2 = 0$, υποδηλώνει μεσόκυρτη κατανομή, δηλαδή συμμετρική κατανομή που έχει τη μορφή της κανονικής κατανομής, ενώ αν $\beta_2 > 3$ ή $\gamma_2 > 0$, έχουμε ένδειξη λεπτόκυρτης κατανομής. Τέτοιες κατανομές έχουν πιο οξεία κορυφή και πιο χοντρές ουρές (fat tails) από την κανονική κατανομή, που σημαίνει ότι τιμές γύρω από τη μέση τιμή καθώς και ακραίες τιμές έχουν μεγαλύτερη πιθανότητα εμφάνισης από ότι σε μια κανονική κατανομή. Ο δείκτης $\beta_2 < 3$ ή $\gamma_2 < 0$, υποδηλώνει πλατύκυρτη κατανομή. Τα χαρακτηριστικά αυτών των κατανομών είναι τα αντίθετα από αυτά των λεπτόκυρτων.

Όπως και για τη λοξότητα, ένα μειονέκτημα που έχει ο παραπάνω τύπος της κύρτωσης είναι ότι για κατανομές με βαριά ουρά για τις οποίες δεν υπάρχουν οι ροπές από κάποια τάξη και πάνω, το μέτρο αυτό επίσης έχει άπειρη τιμή. Αντίστοιχα με τη λοξότητα, υπάρχει τύπος για την κύρτωση που είναι λιγότερο ευαίσθητος στις ακραίες τιμές μέσω του οποίου ξεπερνάμε το προαναφερθέν μειονέκτημα. Ο τύπος είναι βασισμένος στα ογδοημόρια και δίνεται από την έκφραση [Moors, 1998]

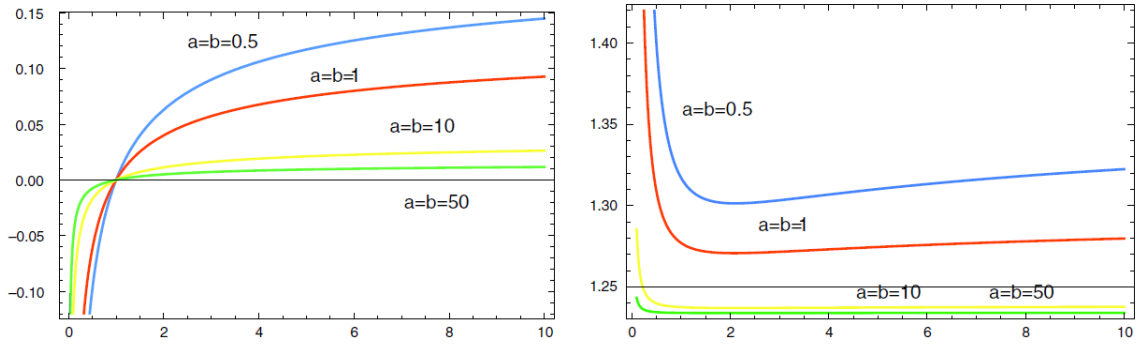
$$M = \frac{Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right) - Q\left(\frac{3}{8}\right) + Q\left(\frac{1}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}. \quad (2.4.38)$$

Στη συνέχεια εξετάζουμε τη γενικευμένη βήτα παραγόμενη κατανομή, θεωρώντας σαν γεννήτορα τις περιπτώσεις της τυπικής κανονικής κατανομής, συμβολικά $GBGN$, και της Student t κατανομής, συμβολικά $GBGStudent$, για την οποία θα αναφερθούμε όταν έχουμε 5 και 10 βαθμούς ελευθερίας. Και οι τρεις κατανομές έχουν λοξότητα $B = 0$, ενώ για το συντελεστή κύρτωσης έχουμε $M = 1.2331$ για την $N(0,1)$ κατανομή, $M = 1.27705$ για τη Student με 10 βαθμούς ελευθερίας και $M = 1.32688$ για τη Student με 5 βαθμούς ελευθερίας.

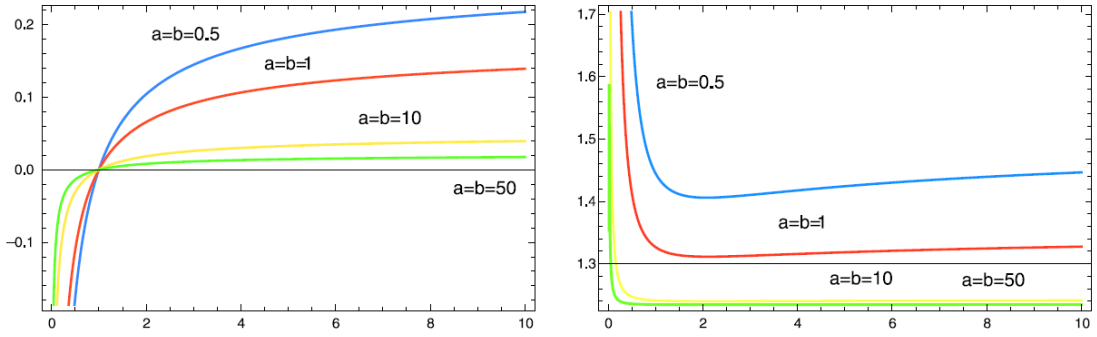
Τα μέτρα λοξότητας B και κύρτωσης M απεικονίζονται στο Σχήμα 2.4 για την $GBGN$ και στα Σχήματα 2.5 και 2.6 για την $GBGStudent$ με 10 και 5 βαθμούς ελευθερίας, αντίστοιχα. Όπως έχει αναφερθεί νωρίτερα στο κεφάλαιο, η εισαγωγή της επιπλέον παραμέτρου c στην κατανομή GBG, έχει σαν σκοπό τον έλεγχο της λοξότητας και κύρτωσης της κατανομής. Συνεπώς για να φανεί η επίδραση του c, διατηρούνται ίσες οι τιμές των a και b και ενδεικτικά εξετάστηκαν οι περιπτώσεις 0.5, 1, 10 και 50.



Σχήμα 2.4 Η λοξότητα B (αριστερά) και η κύρτωση M (δεξιά) για την GBGN κατανομή σαν συνάρτηση του c, για $a = b = 0.5, 1, 10$ και 50.



Σχήμα 2.5 Η λοξότητα B (αριστερά) και η κύρτωση M (δεξιά) για την GBGStudent κατανομή με 10 βαθμούς ελευθερίας σαν συνάρτηση του c, για $a = b = 0.5, 1, 10$ και 50.



Σχήμα 2.6 Η λοξότητα B (αριστερά) και η κύρτωση M (δεξιά) για την GBGStudent κατανομή με 5 βαθμούς ελευθερίας σαν συνάρτηση του c, για $a = b = 0.5, 1, 10$ και 50.

Από τα σχήματα παρατηρούμε ότι και για τις τρεις κατανομές GBG που εξετάστηκαν, ο συντελεστής λοξότητας B είναι μηδέν όταν $c = 1$, αρνητικός για $c < 1$ και θετικός για $c > 1$. Αν και τα τρία γραφήματα για τη λοξότητα φαίνονται να είναι παρόμοια, εν τούτοις οι οριζόντιοι και κατακόρυφοι άξονες έχουν διαφορετικές κλίμακες. Επίσης για τιμή $c \neq 1$, μπορούμε να συμπεράνουμε ότι α) η λοξότητα B αυξάνει καθώς αυξάνει η κύρτωση του γεννήτορα, με τις μεγαλύτερες τιμές να παρατηρούνται για την κατανομή Student με 5 βαθμούς ελευθερίας αφού η κατανομή αυτή έχει τη μεγαλύτερη κύρτωση από τις τρεις, και β) η λοξότητα μειώνεται καθώς αυξάνονται οι τιμές των a και b.

Όσον αφορά την κύρτωση όπως αποδίδεται με το συντελεστή M, βλέπουμε ότι είναι αρκετά περιορισμένη για τις GBG κατανομές. Για την GBGN, από το δεξί γράφημα του Σχήματος 2.4 φαίνεται ότι το M αυξάνει καθώς αυξάνει το c για όλες τις τιμές των a και b και παίρνει τη μεγαλύτερη τιμή όταν $a = b = 1$, εκτός αν το c είναι πολύ μικρό. Το M είναι ίσο με την κύρτωση του γεννήτορα όταν $c = 1$, μικρότερο από το αντίστοιχο του γεννήτορα για $c < 1$ και μεγαλύτερο για $c > 1$. Εντούτοις, καθώς το c αυξάνει, ο συντελεστής M συγκλίνει σε κάποιο όριο που είναι λίγο μεγαλύτερο από την τιμή κύρτωσης 1.2331 του γεννήτορα $N(0,1)$. Μπορούμε λοιπόν να συμπεράνουμε ότι όταν ο γεννήτορας είναι η κανονική κατανομή, η παράμετρος c λειτουργεί κυρίως σαν μια επιπλέον παράμετρος λοξότητας αντί σαν παράμετρος κύρτωσης.

Στην περίπτωση της GBGStudent κατανομής, η κύρτωση απεικονίζεται στο δεξί γράφημα των Σχημάτων 2.5 και 2.6. Βλέπουμε, ότι το μέτρο M μπορεί να γίνει μεγάλο και θετικό, αλλά μόνο για μικρές τιμές του c. Αρχικά, καθώς το c αυξάνει από μικρές τιμές, το M μειώνεται απότομα, παρουσιάζοντας ένα ελάχιστο γύρω από την τιμή $c = 1.5$. Όταν $a = b \leq 1$ η κύρτωση M της GBStudent είναι πάντα μεγαλύτερη από την τιμή M του γεννήτορα. Η τιμή M είναι πολύ κοντά στο μηδέν όταν $a = b = 10$ και όταν $a = b = 50$, εκτός όταν η παράμετρος c παίρνει εξαιρετικά μικρές τιμές. Για μικρές τιμές του c, η κύρτωση M μπορεί να γίνει αρκετά μεγάλη, αλλά αν συνδυαστεί με τις αρνητικά μεγάλες τιμές του συντελεστή λοξότητας B, συμπεραίνουμε ότι η παράμετρος c δρα ξανά κυρίως ως παράμετρος λοξότητας παρά κύρτωσης.

Η ροπογεννήτρια συνάρτηση της X, $M(t) = E(e^{tX})$, μπορεί να προκύψει είτε μέσω του αναπτύγματος Taylor, είτε μέσω των εκθετικοποιημένων-F κατανομών είτε τέλος μέσω της ποσοστιαίας συνάρτησης της κατανομής γεννήτορα.

Αναλυτικά, χρησιμοποιώντας το ανάπτυγμα Maclaurin (δηλαδή τη σειρά Taylor γύρω από το σημείο 0) της εκθετικής συνάρτησης προκύπτει ο παρακάτω τύπος

$$M(t) = E(e^{tX}) = E\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k\right) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k). \tag{2.4.39}$$

Σχετικά με τις ροπές $\mu_k' = E\left(X^k\right)$ που εμφανίζονται μπορεί να χρησιμοποιηθούν οι εκφράσεις που δόθηκαν παραπάνω.

Μέσω των εκθετικοποιημένων-F κατανομών από τη Σχέση (2.4.17), η ροπογεννήτρια συνάρτηση δίνεται από τον τύπο

$$M(t) = \sum_{j=0}^{\infty} w_j M_j(t), \tag{2.4.40}$$

όπου $M_j(t)$ είναι η ροπογεννήτρια συνάρτηση της $Y_j \sim Exp^{c(a+j)} F(x; \mathbf{p})$.

Ένας άλλος τρόπος υπολογισμού της ροπογεννήτριας είναι μέσω της ποσοστιαίας συνάρτησης της κατανομής γεννήτορα. Από την (2.4.18) είναι

$$\begin{aligned} M(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} f_{GBG}(x; \mathbf{p}, a, b, c) dx \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{c}{B(a, b+1)} f(x; \mathbf{p}) \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} [F(x; \mathbf{p})]^{c(a+j)-1} dx \\ &= \frac{c}{B(a, b+1)} \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} e^{tx} [F(x; \mathbf{p})]^{c(a+j)-1} f(x; \mathbf{p}) dx \\ &= \frac{c}{B(a, b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \int_{-\infty}^{\infty} e^{tx} [F(x; \mathbf{p})]^{c(a+j)-1} f(x; \mathbf{p}) dx. \end{aligned} \tag{2.4.41}$$

Για $r = 0,1, \ldots$ ορίζουμε τη συνάρτηση

$$\rho(t, r) = \int_{-\infty}^{\infty} e^{tx} [F(x; \mathbf{p})]^r f(x; \mathbf{p}) dx \ . \tag{2.4.42}$$

Θεωρούμε το μετασχηματισμό $u = F(x; \mathbf{p})$ οπότε είναι $y = F^{-1}(u; \mathbf{p})$ και $du = f(x; \mathbf{p}) dy$, ενώ για $-\infty < y < \infty$ είναι $0 < x < 1$, και το ολοκλήρωμα γίνεται

$$\rho(t, r) = \int_0^1 \exp[tF^{-1}(u; \mathbf{p})] u^r du = \int_0^1 \exp[tQ_{F(x;\mathbf{p})}(u)] u^r du$$

όπου $Q_{F(x;\mathbf{p})}(u) = F^{-1}(u; \mathbf{p})$ είναι η ποσοστιαία συνάρτηση της κατανομής γεννήτορα.

Για a και c θετικούς ακέραιους και η ποσότητα $c(a + j)$ είναι ακέραιος οπότε η Σχέση (2.4.41) με βάση την (2.4.42) γίνεται

$$M(t) = \frac{c}{B(a, b+1)} \sum_{j=0}^{\infty} (-1)^j \binom{b}{j} \rho(t, c(a+j)-1)). \tag{2.4.43}$$

Στο παραπάνω άθροισμα, όταν το b είναι ακέραιος, ο δείκτης j σταματάει στο $b-1$.

Η ποσοστιαία συνάρτηση $Q(u)$ της GBG κατανομής προκύπτει από τη λύση της εξίσωσης

$$F_{GBG}(x; \mathbf{p}, a, b, c) = u \tag{2.4.44}$$

οπότε

$$Q_{GBG}(u; \mathbf{p}, a, b, c) = F_{GBG}^{-1}(u).$$

Στην περίπτωσή μας από τη Σχέση (2.4.4) όπου

$$X = F^{-1}\left(U^{1/c}\right) \sim f_{GBG}(x),$$

η ποσοστιαία συνάρτηση είναι η αντίστροφη της συνάρτησης κατανομής του γεννήτορα F, δηλαδή

$$Q_{GBG}(u; \mathbf{p}, a, b, c) = F^{-1}\left(\left[Q_{\beta(a,b)}(u)\right]^{1/c}\right) \tag{2.4.45}$$

όπου $Q_{\beta(a,b)}(u) = I_u^{-1}(a, b)$ είναι η ποσοστιαία συνάρτηση της $U \sim Beta^{(I)}(a, b)$.

Μία προσεγγιστική έκφραση της $Q_{\beta(a,b)}(u)$ για θετικά a και b είναι η παρακάτω, και δίνεται στην ιστοσελίδα [http://functions.wolfram.com/06.23.06.0004.01],

$$Q_{\beta(a,b)}(u) = a_1 w + a_2 w^2 + a_3 w^3 + a_4 w^4 + O\left(w^{5/a}\right)$$

όπου

$u = [awB(\alpha, b)]^{1/a}, \qquad a > 0,$

$a_1 = 1,$

$a_2 = \dfrac{b-1}{a+1},$

$a_3 = \dfrac{(b-1)(a^2 + (3b-1)a + 5b - 4)}{2(a+1)^2(a+2)},$

$a_4 = \dfrac{(b-1)[a^4 + (6b-1)a^3 + (b+2)(8b-5)a^2 + (33b^2 - 30b + 4)a + b(31b-47) + 18]}{3(a+1)^3(a+2)(a+3)}$

Η μέση απόκλιση από το μέσο, για την περίπτωση των συμμετρικών κατανομών, ή τη διάμεσο για την περίπτωση λοξών κατανομών, χρησιμοποιείται σε κάποιο βαθμό ως μέτρο διασποράς ενός πληθυσμού.

Για την τυχαία μεταβλητή $X \sim GBG(\mathbf{p}, a, b, c)$, η μέση απόκλιση από το μέσο $\mu_1' = E(X)$ είναι

$$\delta_1(X) = E(|X - \mu_1'|) = \int_{-\infty}^{\infty} |x - \mu_1'| f_{GBG}(x) dx. \tag{2.4.46}$$

Αναλύοντας το απόλυτο $|x-\mu_1'| = \begin{cases} x-\mu_1', & x \geq \mu_1' \\ -x+\mu_1', & x < \mu_1' \end{cases}$, η προηγούμενη σχέση απλοποιείται ως εξής,

$$\delta_1(X) = \int_{-\infty}^{\mu_1'} (-x+\mu_1') f_{GBG}(x)dx + \int_{\mu_1'}^{\infty} (x-\mu_1') f_{GBG}(x)dx$$

$$= -\int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx + \mu_1' \int_{-\infty}^{\mu_1'} f_{GBG}(x)dx + \int_{\mu_1'}^{\infty} x f_{GBG}(x)dx - \mu_1' \int_{\mu_1'}^{\infty} f_{GBG}(x)dx$$

$$= -\int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx + \mu_1' \int_{-\infty}^{\mu_1'} f_{GBG}(x)dx + \left(\int_{-\infty}^{\infty} x f_{GBG}(x)dx - \int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx \right)$$

$$-\mu_1' \left(\int_{-\infty}^{\infty} f_{GBG}(x)dx - \int_{-\infty}^{\mu_1'} f_{GBG}(x)dx \right)$$

αλλά για τη σ.π.π. f_{GBG} ισχύει $\int_{-\infty}^{\infty} f_{GBG}(x)dx = 1$ και $\mu_1' = \int_{-\infty}^{\infty} x f_{GBG}(x)dx$ οπότε

$$\delta_1(X) = -\int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx + \mu_1' F_{GBG}(\mu_1') + \mu_1' - \int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx - \mu_1' + \mu_1' F_{GBG}(\mu_1')$$

$$= 2\mu_1' F_{GBG}(\mu_1') - 2\int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx$$

$$= 2\mu_1' F_{GBG}(\mu_1') - 2I(\mu_1') \qquad (2.4.47)$$

όπου $I(\mu_1') = \int_{-\infty}^{\mu_1'} x f_{GBG}(x)dx$.

Η μέση απόκλιση από τη διάμεσο M είναι

$$\delta_2(X) = E(|X-M|) = \int_{-\infty}^{\infty} |x-M| f_{GBG}(x)dx \qquad (2.4.48)$$

η οποία με όμοιο τρόπο απλοποιείται και δίνεται από τη σχέση

$$\delta_2(X) = \mu_1' - 2I(M) \qquad (2.4.49)$$

όπου $I(M) = \int_{-\infty}^{M} x f_{GBG}(x)dx$.

Η διάμεσος M υπολογίζεται από την ποσοστιαία συνάρτηση αν στον τύπο (2.4.45) θέσουμε $u = 0.5$.

Για την ποσότητα $I(d) = \int_{-\infty}^{d} x f_{GBG}(x)dx$ που προκύπτει και στις δύο περιπτώσεις θα υπολογίσουμε έναν γενικό τύπο. Από την (2.4.3) και το διωνυμικό ανάπτυγμα (2.4.11α) βρίσκουμε ότι

$$I(d) = \int_{-\infty}^{d} x \frac{c}{B(a,b)} f(x;\mathbf{p})[F(x;\mathbf{p})]^{ac-1}(1-[F(x;\mathbf{p})]^{c})^{b-1}dx$$

$$= \frac{c}{B(a,b)} \int_{-\infty}^{d} x f(x;\mathbf{p})[F(x;\mathbf{p})]^{ac-1} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} [F(x;\mathbf{p})]^{cj}\, dx$$

$$= \frac{c}{B(a,b)} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \int_{-\infty}^{d} x f(x;\mathbf{p})[F(x;\mathbf{p})]^{c(a+j)-1}dx.$$

Θεωρώντας τυχαίες μεταβλητές με θετικό στήριγμα και εφαρμόζοντας παραγοντική ολοκλήρωση βρίσκουμε

$$\int_{0}^{d} x f(x;\mathbf{p})[F(x;\mathbf{p})]^{c(a+j)-1}dx = \int_{0}^{d} x \left(\frac{[F(x;\mathbf{p})]^{c(a+j)}}{c(a+j)}\right)' dx$$

$$= \left[x \frac{[F(x;\mathbf{p})]^{c(a+j)}}{c(a+j)}\right]_{0}^{d} - \int_{0}^{d} \frac{[F(x;\mathbf{p})]^{c(a+j)}}{c(a+j)} dx$$

$$= \frac{d[F(d;\mathbf{p})]^{c(a+j)}}{c(a+j)} - \frac{1}{c(a+j)} \int_{0}^{d} [F(x;\mathbf{p})]^{c(a+j)}dx$$

οπότε η $I(d)$ γίνεται

$$I(d) = \sum_{j=0}^{\infty} \frac{(-1)^j \binom{b-1}{j}}{(a+j)B(a,b)} \left(d[F(d;\mathbf{p})]^{c(a+j)} - \int_{0}^{d} [F(x;\mathbf{p})]^{c(a+j)}\, dx \right)$$

ή

$$I(d) = \sum_{j=0}^{\infty} m_j(a,b) \left[d[F(d;\mathbf{p})]^{c(a+j)} - H_j(d) \right] \tag{2.4.50}$$

όπου $m_j(a,b) = \frac{(-1)^j \binom{b-1}{j}}{(a+j)B(a,b)}$ και $H_j(d) = \int_{0}^{d} [F(x;\mathbf{p})]^{c(a+j)}dx.$

Στην επόμενη παράγραφο εξετάζεται η στατιστική διατεταγμένου δείγματος (order statistic).

Έστω $(X_1, X_2, \dots, X_n)$ ένα τυχαίο δείγμα n παρατηρήσεων. Συμβολίζουμε με $X_{1:n}, 1 \le i \le n$, την παρατήρηση που αντιστοιχεί στην i κατ' αύξουσα σειρά μεγέθους παρατήρηση, η οποία ονομάζεται i-οστή διατεταγμένη παρατήρηση. Η n-διάστατη τυχαία μεταβλητή

$(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ λέγεται διατεταγμένο δείγμα το οποίο αντιστοιχεί στο παραπάνω τυχαίο δείγμα και ισχύει

$$X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}\,.$$

Η συνάρτηση πυκνότητας πιθανότητας της i-στής διατεταγμένης παρατήρησης δίνεται από τη σχέση (Κούτρας, (2005))

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}, \qquad i = 1, \dots, n, \tag{2.4.51}$$

και η συνάρτηση κατανομής από την

$$F_{i:n}(x) = I_{F(x)}(i, n-i+1) \tag{2.4.52}$$

ή χρησιμοποιώντας το διωνυμικό άθροισμα (παράρτημα Π5),

$$F_{i:n}(x) = \sum_{r=i}^{n} \binom{n}{r} [F(x)]^r [1-F(x)]^{n-r}. \tag{2.4.53}$$

Χρησιμοποιώντας τη Σχέση (2.4.11α) η σ.π.π. της (2.4.51) γράφεται

$$\begin{aligned} f_{i:n}(x) &= \frac{1}{B(i, n-i+1)} f(x)[F(x)]^{i-1} \sum_{r=0}^{n-i} (-1)^r \binom{n-i}{r} [F(x)]^r \\ &= \frac{1}{B(i, n-i+1)} f(x) \sum_{r=0}^{n-i} (-1)^r \binom{n-i}{r} [F(x)]^{r-i-1}. \end{aligned} \tag{2.4.54}$$

Στη συνέχεια η ιδέα είναι να αντικατασταθούν οι συναρτήσεις f και F με αυτές της GBG κατανομής, και με κατάλληλους υπολογισμούς να εκφραστεί η $f_{i:n}$ σαν γραμμικός συνδυασμός των εκάστοτε εκθετικοποιημένων κατανομών γεννητόρων.

Στην περίπτωση αυτή, η ανάπτυξη σχέσεων για τις ροπές r τάξης της i-οστής διατεταγμένης παρατήρησης, $E(X_{i:n}^r)$, είναι σχετικά εύκολη διαδικασία.

Οι L-ροπές χρησιμοποιούνται ως περιγραφικά στατιστικά αλλά και ως μέθοδος προσαρμογής κατανομών. Οι L-ροπές είναι ανάλογες των συνήθων ροπών και έχουν οριστεί από τον Hosking (1990) ως γραμμικός συνδυασμός αναμενόμενων τιμών των διατεταγμένων στατιστικών σύμφωνα με τη σχέση

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \quad r = 1, 2, \dots \tag{2.4.55}$$

Οι L-ροπές έχουν το πλεονέκτημα ότι υπάρχουν αν υπάρχει η μέση τιμή της κατανομής, ακόμη και αν κάποιες ροπές μεγαλύτερης τάξης δεν υπάρχουν. Επίσης η L-ροπές είναι ανθεκτικές σε φαινόμενα ακραίων τιμών (outliers). Οι δειγματικές εκτιμήσεις των L-ροπών δεν υψώνουν τις παρατηρήσεις του δείγματος σε τετραγωνικές ή κυβικές δυνάμεις, όπως συμβαίνει στις συνήθεις ροπές.

Οι πρώτες τέσσερις L-ροπές είναι

$$\lambda_1 = E(X_{1:1})$$

$$\lambda_2 = \frac{1}{2}[E(X_{2:2}) - E(X_{1:2})]$$

$$\lambda_3 = \frac{1}{3}[E(X_{3:3}) - 2E(X_{2:3}) + E(X_{1:3})]$$

$$\lambda_4 = \frac{1}{4}[E(X_{4:4}) - 3E(X_{3:4}) + 3E(X_{2:4}) - E(X_{1:4})].$$

Στατιστικές σχετικές με τη λοξότητα και την κύρτωση της κατανομής δίνονται αντίστοιχα από τις σχέσεις

$$(X) = \frac{\lambda_3}{\lambda_2} \text{ και } kurt(X) = \frac{\lambda_4}{\lambda_3}.$$

Οι L-ροπές μπορούν να εκφραστούν και ως συνάρτηση των σταθμισμένων ροπών μιας τ.μ. X με σ.κ. F. Από τη Σχέση (2.4.26) με

$$\beta_r = \tau(1, r) = E(X[F(X)]^r) = \int_x x[F(x)]^r f(x)dx$$

οι L-ροπές εκφρασμένες με όρους σταθμισμένων ροπών είναι

$$\lambda_r = \sum_{k=0}^{r-1} (-1)^{r-1-k} \binom{r-1}{k} \binom{r-1+k}{k} \beta_k, \quad r = 1, 2, \ldots \tag{2.4.56}$$

Συγκεκριμένα, οι πρώτες τέσσερις L-ροπές εκφρασμένες με όρους σταθμισμένων ροπών είναι

$$\lambda_1 = \beta_0$$

$$\lambda_2 = 2\beta_1 - \beta_0$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

Αντίστοιχα, οι σταθμισμένες ροπές β_r, $r = 1, 2, 3, 4$, εκφρασμένες σε όρους L-ροπών δίνονται από τις σχέσεις

$$\beta_0 = \lambda_1, \ \beta_1 = \frac{1}{2}(\lambda_1 + \lambda_2), \ \beta_2 = \frac{1}{6}\lambda_3 + \frac{1}{2}\lambda_2 + \frac{1}{3}\lambda_1 \text{ και } \beta_3 = \frac{1}{20}\lambda_4 + \frac{1}{4}\lambda_3 + \frac{9}{20}\lambda_2 + \frac{1}{4}\lambda_1.$$

Οι αμερόληπτες εκτιμήτριες των β_r δίνονται από τη σχέση (Landwehr et al., (1979))

$$\hat{\beta}_r = \frac{1}{n}\sum_{i=1}^{n} \frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)} x_{1:n}.$$

Στη συνέχεια θα εφαρμόσουμε τη μέθοδο της μέγιστης πιθανοφάνειας για την εκτίμηση των παραμέτρων της GBG κατανομής.

Έστω ένα τυχαίο δείγμα $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ μεγέθους n από την κατανομή GBG και $\boldsymbol{\theta} = (\mathbf{p}, a, b, c)$ το διάνυσμα των παραμέτρων της. Η συνάρτηση πιθανοφάνειας είναι

$$L(\boldsymbol{\theta} \mid \mathbf{x}) = \prod_{i=1}^{n} f_{GBG}(x_i; \boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{c}{B(a,b)} f(x_i; \mathbf{p})[F(x_i; \mathbf{p})]^{ac-1}(1 - [F(x_i; \mathbf{p})]^c)^{b-1}$$

$$= \frac{c^n}{[B(a,b)]^n} \prod_{i=1}^{n} f(x_i; \mathbf{p}) \left[\prod_{i=1}^{n} F(x_i; \mathbf{p})\right]^{ac-1} \left[\prod_{i=1}^{n} (1 - [F(x_i; \mathbf{p})]^c)\right]^{b-1}.$$

Για μεγαλύτερη ευκολία στους υπολογισμούς μας χρησιμοποιούμε το λογάριθμο της συνάρτησης πιθανοφάνειας. Η χρήση του λογαρίθμου δεν μας δημιουργεί κανένα πρόβλημα στο αποτέλεσμα εφ' όσον η συνάρτηση $\log L(\boldsymbol{\theta})$ είναι γνησίως αύξουσα και επομένως η τιμή που θα μεγιστοποιεί τις $L(\boldsymbol{\theta})$ και $\log L(\boldsymbol{\theta})$ θα είναι ίδια. Συνεπώς,

$$l(\boldsymbol{\theta}) := \log L(\boldsymbol{\theta} \mid \mathbf{x}) = n \log c - n \log B(a,b) + \sum_{i=1}^{n} \log f(x_i; \mathbf{p}) + (ac-1) \sum_{i=1}^{n} \log F(x_i; \mathbf{p})$$

$$+(b-1) \sum_{i=1}^{n} \log (1 - [F(x_i; \mathbf{p})]^c). \qquad (2.4.57)$$

Για τη μεγιστοποίηση της συνάρτησης $l(\boldsymbol{\theta})$ λύνουμε το σύστημα των μη γραμμικών εξισώσεων που προκύπτουν από την παραγώγιση της $l(\boldsymbol{\theta})$ ως προς τις παραμέτρους $\boldsymbol{\theta}$ και την εξίσωσή τους με το μηδέν. Συγκεκριμένα, οι πρώτες παράγωγοι είναι,

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a} = -n \frac{\partial}{\partial a} \log B(a,b) + c \sum_{i=1}^{n} \log F(x_i; \mathbf{p})$$

$$= -n[\psi(a) - \psi(a+b)] + c \sum_{i=1}^{n} \log F(x_i; \mathbf{p})$$

αφού

$$\frac{\partial}{\partial a} \log B(a,b) = \frac{\partial}{\partial a}\left(\log \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\right) = \frac{\partial}{\partial a}[\log \Gamma(a) + \log \Gamma(b) - \log \Gamma(a+b)]$$

$$= \frac{\partial}{\partial a} \log \Gamma(a) - \frac{\partial}{\partial a} \log \Gamma(a+b) = \frac{\Gamma'(a)}{\Gamma(a)} - \frac{\frac{\partial}{\partial a}\Gamma(a+b)}{\Gamma(a+b)} = \psi(a) - \psi(a+b),$$

όπου $\psi(y) = \dfrac{\Gamma'(y)}{\Gamma(y)}$ είναι η συνάρτηση δίγαμμα.

Παραγωγίζοντας ως προς b, c και $\mathbf{p}$ έχουμε αντίστοιχα

$$\frac{\partial l(\boldsymbol{\theta})}{\partial b} = -n[\psi(b) - \psi(a+b)] + \sum_{i=1}^{n} \log(1 - [F(x_i; \mathbf{p})]^c),$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial c} = \frac{n}{c} + a \sum_{i=1}^{n} \log F(x_i; \mathbf{p}) - (b-1) \sum_{i=1}^{n} \frac{[F(x_i; \mathbf{p})]^c \log F(x_i; \mathbf{p})}{1 - [F(x_i; \mathbf{p})]^c},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mathbf{p}} = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \mathbf{p}} f(x_i;\mathbf{p})}{f(x_i;\mathbf{p})} + (ac-1)\sum_{i=1}^{n} \frac{\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{F(x_i;\mathbf{p})} - c(b-1)\sum_{i=1}^{n} \frac{[F(x_i;\mathbf{p})]^{c-1}\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{1-[F(x_i;\mathbf{p})]^{c}}.$$

Εναλλακτικά, η μεγιστοποίηση της $l(\boldsymbol{\theta})$ μπορεί να επιτευχθεί απευθείας με χρήση κάποιου στατιστικού πακέτου.

Κάποιες αρχικές εκτιμήσεις των παραμέτρων a και b μπορούν να προκύψουν από τις εκτιμήτριες των c και $\mathbf{p}$.

Παρατηρούμε ότι αν $X \sim GBG(F;\mathbf{p},a,b,c)$ τότε η $Y = F(X;\mathbf{p}) \sim GB(a,b,c)$. Πράγματι από την (2.4.4) είναι $X = F^{-1}\left(U^{1/c}\right) \Rightarrow U^{1/c} = F(X)$, όπου $U \sim Beta(\mathrm{I})(a,b)$. Η συνάρτηση κατανομής και η συνάρτηση πυκνότητας πιθανότητας της $Y = F(X;\mathbf{p}) = U^{1/c}$ είναι αντίστοιχα

$$F_Y(y) = P(Y \le y) = P\left(U^{1/c} \le y\right) = P(U \le y^c) = F_U(y^c)$$

$$f_y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_U(y^c) = cy^{c-1}f_U(y^c) = cy^{c-1}\frac{1}{B(a,b)}(y^c)^{a-1}(1-y^c)^{b-1}$$

$$= \frac{c}{B(a,b)}y^{ac-1}(1-y^c)^{b-1}, \qquad 0 < y < 1.$$

Η παραπάνω είναι η συνάρτηση πυκνότητας πιθανότητας της γενικευμένης βήτα (GB) κατανομής πρώτου είδους (Σχέση 2.4.1) ή μιας GBG με γεννήτορα την $Ομοιόμορφη(0,1)$. Στη συνέχεια υπολογίζουμε τις ροπές,

$$E[F^s(X;\mathbf{p})] = \frac{B\left(a+\frac{s}{c},b\right)}{B(a,b)}, \tag{2.4.58α}$$

$$E\{[1-F^c(X;\mathbf{p})]^r\} = \frac{B(a,b+r)}{B(a,b)}, \tag{2.4.58β}$$

$$E\{F^s(X;\mathbf{p})[1-F^c(X;\mathbf{p})]^r\} = \frac{B\left(a+\frac{s}{c},b+r\right)}{B(a,b)}. \tag{2.4.58γ}$$

Ενδεικτικά για την απόδειξη της (2.4.58α) έχουμε

$$E[F^s(X;\mathbf{p})] = E(Y^s) = \int_y y^s f_Y(y)dy = \int_0^1 y^s \frac{c}{B(a,b)}y^{ac-1}(1-y^c)^{b-1}dy$$

$$= \frac{c}{B(a,b)}\int_0^1 y^{s+ac-1}(1-y^c)^{b-1}dy$$

θέτοντας $u = y^c$ είναι $y = u^{1/c}, dy = \frac{1}{c}u^{1/c-1}$ και για $0 < y < 1 \Rightarrow 0 < u < 1$ οπότε

$$E[F^s(X;\mathbf{p})] = \frac{c}{B(a,b)}\int_0^1 \left(u^{1/c}\right)^{s+ac-1}(1-u)^{b-1}\frac{1}{c}u^{1/c-1}\,du$$

$$= \frac{1}{B(a,b)}\int_0^1 u^{\frac{s}{c}+a-1}(1-u)^{b-1}\,du$$

$$= \frac{B\left(\frac{s}{c}+a,b\right)}{B(a,b)}\int_0^1 \frac{1}{B\left(\frac{s}{c}+a,b\right)}u^{\frac{s}{c}+a-1}(1-u)^{b-1} = \frac{B\left(\frac{s}{c}+a,b\right)}{B(a,b)}.$$

Με παρόμοιο τρόπο δείχνονται και οι άλλες δύο σχέσεις.

Θέτοντας στην (2.4.58α) $s = c$ και στην (2.4.58γ) $s = c$ και $r = 1$ παίρνουμε αντίστοιχα

$$E[F^c(X;\mathbf{p})] = \frac{B(a+1,b)}{B(a,b)} = \frac{\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{\frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{a}{a+b}$$

$$E\{F^c(X;\mathbf{p})[1-F^c(X;\mathbf{p})]\} = \frac{B(a+1,b+1)}{B(a,b)} = \frac{\frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+1+1)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{\frac{a\Gamma(a)b\Gamma(b)}{(a+b+1)\Gamma(a+b+1)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}}$$

$$= \frac{a\Gamma(a+b)}{(a+b+1)(a+b)\Gamma(a+b)} = \frac{a}{(a+b)(a+b+1)}.$$

Έστω $u = E[F^c(X;\mathbf{p})]$ και $v = E\{F^c(X;\mathbf{p})[1-F^c(X;\mathbf{p})]\}$.

Το παραπάνω σύστημα εξισώσεων γίνεται

$$u = \frac{a}{a+b} \text{ και } v = \frac{a}{(a+b)(a+b+1)}$$

με λύση

$$a = \frac{uv}{u(1-u)-v} \text{ και } b = \frac{v(1-u)}{u(1-u)-v}. \qquad (2.4.58\delta)$$

Στην επόμενη ενότητα αναφέρουμε πώς αξιοποιώντας τους εκτιμητές μέγιστης πιθανοφάνειας, μπορούμε να εφαρμόσουμε μεθόδους Στατιστικής Συμπερασματολογίας και ειδικότερα, εκτιμητικής των άγνωστων παραμέτρων με διαστήματα εμπιστοσύνης και ελέγχου υποθέσεων σχετικών με την GBG κατανομή (τις άγνωστες παραμέτρους της ή/και με τη μορφή της).

Για την κατασκευή διαστημάτων εμπιστοσύνης και τον έλεγχο υποθέσεων των παραμέτρων της κατανομής χρειαζόμαστε τον πίνακα πληροφορίας Fisher $\mathbf{J}_n(\boldsymbol{\theta})$, για το μέγεθος δείγματος n που είναι (Alexander et al., (2012))

$$\mathbf{J}_n(\boldsymbol{\theta}) = \begin{pmatrix} J_{aa} & J_{ab} & J_{ac} & J_{a\mathbf{p}} \\ J_{ba} & J_{bb} & J_{bc} & J_{b\mathbf{p}} \\ J_{ca} & J_{cb} & J_{cc} & J_{c\mathbf{p}} \\ J_{\mathbf{p}a} & J_{\mathbf{p}b} & J_{\mathbf{p}c} & J_{\mathbf{pp}} \end{pmatrix}$$

όπου τα στοιχεία του συμμετρικού πίνακα είναι

$$J_{aa} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial a^2} = -n\left[\psi'(a) - \frac{\partial}{\partial a}\psi(a+b)\right],$$

$$J_{ba} = J_{ab} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial b \partial a} = n\frac{\partial}{\partial b}\psi(a+b),$$

$$J_{ca} = J_{ac} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial c \partial a} = \sum_{i=1}^{n} \log F(x_i;\mathbf{p}),$$

$$J_{\mathbf{p}a} = J_{a\mathbf{p}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mathbf{p} \partial a} = c\sum_{i=1}^{n} \frac{\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{F(x_i;\mathbf{p})},$$

$$J_{bb} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial b^2} = -n\left[\psi'(b) - \frac{\partial}{\partial b}\psi(a+b)\right],$$

$$J_{cb} = J_{bc} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial c \partial b} = -\sum_{i=1}^{n} \frac{F^c(x_i;\mathbf{p}) \log F(x_i;\mathbf{p})}{1 - F^c(x_i;\mathbf{p})},$$

$$J_{\mathbf{p}b} = J_{b\mathbf{p}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mathbf{p} \partial b} = -c\sum_{i=1}^{n} \frac{F^{c-1}(x_i;\mathbf{p})\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{1 - F^c(x_i;\mathbf{p})},$$

$$J_{cc} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial c^2} = -\frac{n}{c^2} - (b-1)\sum_{i=1}^{n}\left\{\frac{F^c(x_i;\mathbf{p})\log^2 F(x_i;\mathbf{p})[1 - F^c(x_i;\mathbf{p})]}{[1 - F^c(x_i;\mathbf{p})]^2}\right.$$

$$\left. - \frac{F^c(x_i;\mathbf{p})\log F(x_i;\mathbf{p})[-F^c(x_i;\mathbf{p})\log F(x_i;\mathbf{p})]}{[1 - F^c(x_i;\mathbf{p})]^2}\right\}$$

$$= -\frac{n}{c^2} - (b-1)\sum_{i=1}^{n} \frac{F^c(x_i;\mathbf{p})\log^2 F(x_i;\mathbf{p})}{[1 - F^c(x_i;\mathbf{p})]^2},$$

$$J_{\mathbf{p}c} = J_{c\mathbf{p}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mathbf{p} \partial c} = a\sum_{i=1}^{n} \frac{\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{F(x_i;\mathbf{p})} - (b-1)\sum_{i=1}^{n} \frac{1}{[1 - F^c(x_i;\mathbf{p})]^2} \times$$

$$\left\{\left[cF^{c-1}(x_i;\mathbf{p})\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})\log F(x_i;\mathbf{p}) + \frac{F^c(x_i;\mathbf{p})}{F(x_i;\mathbf{p})}\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})\right][1 - F^c(x_i;\mathbf{p})]\right.$$

$$\left. + F^c(x_i;\mathbf{p})\log F(x_i;\mathbf{p}) cF^{c-1}(x_i;\mathbf{p})\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})\right\}$$

$$= a\sum_{i=1}^{n} \frac{\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{F(x_i;\mathbf{p})} - (b-1)\sum_{i=1}^{n} \frac{F^{c-1}(x_i;\mathbf{p})\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})}{[1 - F^c(x_i;\mathbf{p})]^2} \times$$

$$\{[c\log F(x_i;\mathbf{p}) + 1][1 - F^c(x_i;\mathbf{p})] + cF^c(x_i;\mathbf{p})\log F(x_i;\mathbf{p})\},$$

$$J_{\mathbf{pp}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mathbf{p}^2} = \sum_{i=1}^{n} \frac{\left(\frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}} f(x_i;\mathbf{p})\right) f(x_i;\mathbf{p}) - \left(\frac{\partial}{\partial \mathbf{p}} f(x_i;\mathbf{p})\right)^2}{f^2(x_i;\mathbf{p})}$$

$$+(ac-1)\sum_{i=1}^{n} \frac{\left(\frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}} F(x_i;\mathbf{p})\right) F(x_i;\mathbf{p}) - \left(\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})\right)^2}{F^2(x_i;\mathbf{p})}$$

$$-c(b-1)\sum_{i=1}^{n} \frac{F^{c-1}(x_i;\mathbf{p})}{[1-F^c(x_i;\mathbf{p})]^2} \times$$

$$\left\{ \left((c-1)F^{-1}(x_i;\mathbf{p}) \left(\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})\right)^2 + \frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}} F(x_i;\mathbf{p}) \right) [1-F^c(x_i;\mathbf{p})] \right.$$

$$\left. + cF^{-1}(x_i;\mathbf{p}) \left(\frac{\partial}{\partial \mathbf{p}} F(x_i;\mathbf{p})\right)^2 \right\}.$$

Εφ' όσον οι εκτιμητές μέγιστης πιθανοφάνειας των αγνώστων παραμέτρων $\boldsymbol{\theta} = (\mathbf{p}, a, b, c)$ δεν μπορούν να προκύψουν σε κλειστή μορφή, δεν είναι εύκολο να προσδιοριστεί η ακριβής κατανομή των εκτιμητών αυτών. Ωστόσο, εφαρμόζοντας προσέγγιση μεγάλου μεγέθους δείγματος, υπό συγκεκριμένες συνθήκες ομαλότητας που ικανοποιούν οι παράμετροι $\boldsymbol{\theta}$, στο εσωτερικό του παραμετρικού χώρου, έχουμε ότι για τους εκτιμητές $\widehat{\boldsymbol{\theta}}$ ασυμπτωτικά ισχύει,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \to N_{p+3}(0, \mathbf{I}(\boldsymbol{\theta})^{-1}), \text{καθώς } n \uparrow \infty$$

όπου $\mathbf{I}(\boldsymbol{\theta})$ είναι ο αναμενόμενος πίνακας πληροφορίας κατά Fisher. Δηλαδή, ο εμπ του $\boldsymbol{\theta}$ ακολουθεί ασυμπτωτικά πολυδιάστατη κανονική κατανομή με μέση τιμή $\boldsymbol{\theta}$, δηλαδή είναι ασυμπτωτικά αμερόληπτος, και πίνακα διασποράς ίσο με τον αντίστροφο του πίνακα πληροφορίας. Στην παραπάνω σχέση, μπορούμε να αντικαταστήσουμε το $\mathbf{I}(\boldsymbol{\theta})$ με τον παρατηρούμενο πίνακα πληροφορίας που εκτιμάταται στο $\widehat{\boldsymbol{\theta}}$, δηλαδή τον $\mathbf{J}_n\left(\widehat{\boldsymbol{\theta}}\right)$. Η πολυδιάστατη κανονική κατανομή $N_{p+3}\left(0, \mathbf{J}_n\left(\widehat{\boldsymbol{\theta}}\right)^{-1}\right)$ μπορεί να χρησιμοποιηθεί για την κατασκευή προσεγγιστικών διαστημάτων εμπιστοσύνης για κάθε μία παράμετρο. Το $100(1-\gamma)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης (Asymptotic Confidence Interval – ACI) για κάθε παράμετρο θ_i δίνεται από τη σχέση (Alexander et al., (2012), Barreto-Souza et al., (2010))

$$ACI_i = \left(\hat{\theta}_i - z_{\gamma/2}\sqrt{\hat{\jmath}^{\theta_i,\theta_i}}, \hat{\theta}_i + z_{\gamma/2}\sqrt{\hat{\jmath}^{\theta_i,\theta_i}}\right),$$

όπου $\hat{\jmath}^{\theta_i,\theta_i}$ είναι το i-στο διαγώνιο στοιχείου του πίνακα $\mathbf{J}_n\left(\widehat{\boldsymbol{\theta}}\right)^{-1}$ για $i = 1,2,\dots,p+3$ και $z_{\gamma/2}$ είναι το άνω $\left(1-\frac{\gamma}{2}\right)$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Η ασυμπτωτικά κανονική κατανομή είναι επίσης χρήσιμη στον έλεγχο καλής προσαρμογής των παραμέτρων της κατανομής, καθώς και στη σύγκριση της GBG κατανομής με κάποια από τις κατανομές που προκύπτουν σαν ειδικές περιπτώσεις της, με χρήση της στατιστικής του λόγου πιθανοφάνειας LR (likelihood ratio). Έτσι για τον έλεγχο της μηδενικής υπόθεσης $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ έναντι της εναλλακτικής $H_1: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, χρησιμοποιούμε την LR στατιστική (Alexander et al., (2012), Barreto-Souza et al., (2010)).

Για παράδειγμα, για να ελέγξουμε αν η προσαρμογή της GBG κατανομής υπερέχει στατιστικά από την προσαρμογή της BG κατανομής, ισοδύναμα έχουμε την υπόθεση

$$H_0: c = 1 \text{ έναντι της } H_1: c \neq 1.$$

Η στατιστική LR βασίζεται στους λογάριθμους πιθανοφάνειας των δύο μοντέλων και είναι,

$$LR = 2\{l(\hat{a}, \hat{b}, \hat{c}, \hat{\mathbf{p}}) - l(\tilde{a}, \tilde{b}, 1, \tilde{\mathbf{p}})\}$$

όπου $\hat{a}, \hat{b}, \hat{c}, \hat{\mathbf{p}}$ είναι οι εκτιμητές μέγιστης πιθανοφάνειας υπό την H_1 και $\tilde{a}, \tilde{b}, \tilde{\mathbf{p}}$ αντίστοιχα υπό την H_0.

ΚΕΦΑΛΑΙΟ 3

ΚΑΤΑΝΟΜΕΣ PARETO, WEIBULL, ΛΟΓΙΣΤΙΚΗ

Στο παρόν κεφάλαιο εφαρμόζουμε τις μεθοδολογίες του δευτέρου κεφαλαίου σε συγκεκριμένες κατανομές. Οι κατανομές με τις οποίες θα ασχοληθούμε είναι η Pareto δευτέρου είδους, η Weibull και η Λογιστική δύο παραμέτρων.

Αρχικά για κάθε μία από τις παραπάνω κατανομές αναφέρουμε τα χαρακτηριστικά της όπως τη συνάρτηση πυκνότητας πιθανότητας και τη συνάρτηση κατανομής της και βασικά μεγέθη της όπως η ένταση κινδύνου, οι ροπές και η ροπογεννήτρια συνάρτηση. Επιπλέον παρουσιάζεται η εκτίμηση των παραμέτρων της κατανομής μέσω της μεθόδου της μέγιστης πιθανοφάνειας, της μεθόδου των ροπών και της μεθόδου των ποσοστιαίων σημείων. Τέλος παρουσιάζονται ιδιότητες και μετασχηματισμοί των κατανομών.

Στη συνέχεια θεωρούμε τη γενικευμένη βήτα παραγόμενη (GBG) κατανομή με γεννήτορα κάθε μία από τις παραπάνω κατανομές. Δίνονται η σ.π.π., η σ.κ., η συνάρτηση κινδύνου, οι ροπές και άλλα μέτρα, και τέλος οι εκτιμήτριες μέγιστης πιθανοφάνειας των παραμέτρων.

3.1 Κατανομή $Pareto^{(II)}$

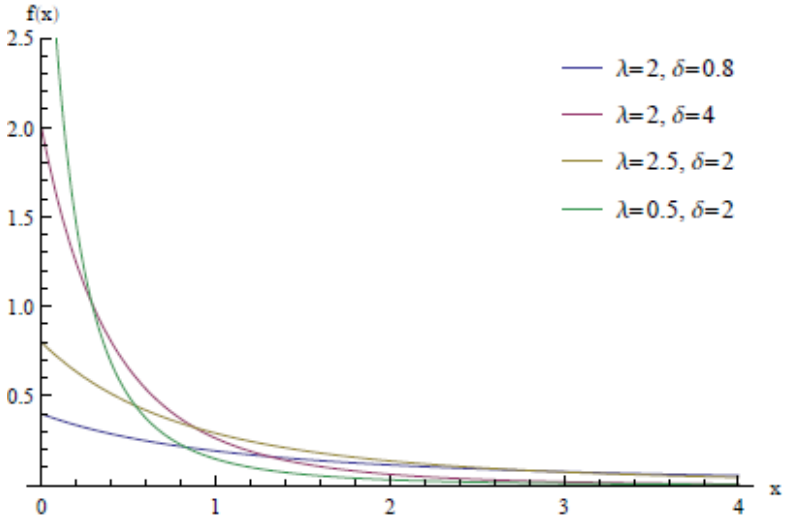
Η κατανομή Pareto, γνωστή και ως κατανομή νόμου δύναμης (power law distribution), χαρακτηρίζεται από σταθερότητα κλίμακας, θετική ασυμμετρία (δηλαδή μεγάλη δεξιά ουρά), μακρά απόσβεση και είναι μονοκόρυφη. Χρησιμοποιείται κατά κόρον σε αναλογιστικές μελέτες λόγω του βάρους στη δεξιά ουρά. Η ιδιότητα αυτή μπορεί να ερμηνεύσει ικανοποιητικά κινδύνους που εμφανίζονται στον ασφαλιστικό κλάδο, όπως τη σφοδρότητα ατυχημάτων για μεγάλες ζημιές του κλάδου γενικών ασφαλίσεων. Στο χαρακτηριστικό της θετικής ασυμμετρίας βασίζεται και ο νόμος Pareto (80-20) ο οποίος βρίσκει πολλές εφαρμογές στον κλάδο των οικονομικών, για παράδειγμα, το 80% του πλούτου μιας χώρας κατέχεται από το 20% του πληθυσμού της.

Θα λέμε ότι η τ.μ. *X* ακολουθεί την κατανομή $Pareto^{(II)}$, η οποία αναφέρεται επίσης και ως κατανομή Lomax, ή *American Pareto*, με παραμέτρους δ και λ και συμβολίζουμε $X \sim Par^{(II)}(\delta, \lambda)$ αν έχει συνάρτηση πυκνότητας πιθανότητας και συνάρτηση κατανομής αντίστοιχα

$$f(x;\delta,\lambda) = \frac{\delta\lambda^{\delta}}{(\lambda+x)^{\delta+1}}, \qquad x>0, \delta>0, \lambda>0, \tag{3.1.1}$$

$$F(x;\delta,\lambda) = 1 - \frac{\lambda^{\delta}}{(\lambda+x)^{\delta}}\ , \qquad x > 0, \delta > 0, \lambda > 0 \tag{3.1.2}$$

όπου το δ είναι παράμετρος μορφής και το λ είναι παράμετρος κλίμακας. Στο Σχήμα 3.1 απεικονίζεται η συνάρτηση πυκνότητας της Pareto για διάφορες τιμές των παραμέτρων της.



Σχήμα 3.1 Συνάρτηση πυκνότητας της $Par^{(II)}(\delta,\lambda)$

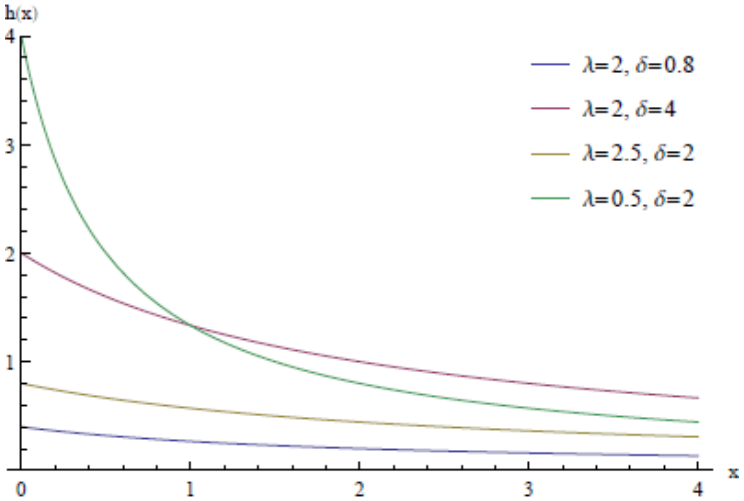
Στη συνέχεια αναφέρουμε τα χαρακτηριστικά μεγέθη της κατανομής.

3.1.1 Χαρακτηριστικά μεγέθη της κατανομής $Pareto^{(II)}$

Η συνάρτηση κινδύνου δίνεται από τον τύπο

$$h(x) = \frac{\delta}{\lambda + x} \tag{3.1.3}$$

και απεικονίζεται στο Σχήμα 3.2.



Σχήμα 3.2 Συνάρτηση κινδύνου της $Par^{(II)}(\delta,\lambda)$

Από το γράφημα βλέπουμε ότι η συνάρτηση διακινδύνευσης είναι φθίνουσα συνάρτηση ως προς x. Αυτό είναι μία επιπλέον ένδειξη ότι η κατανομή έχει βαριά δεξιά ουρά (Klugman, 2004). Συγκεκριμένα, ο ρυθμός μείωσης της συνάρτησης h όπως προκύπτει από την πρώτη παράγωγό της, είναι $h'(x) = -\frac{\delta}{(\lambda+x)^2}$.

Οι ροπές k τάξης της κατανομής δίνονται από τον τύπο

$$E(X^k) = \frac{\lambda^k k!}{(\delta-1)(\delta-2)\cdots(\delta-k)} = \lambda^k k! \frac{\Gamma(\delta-k)}{\Gamma(\delta)}, \qquad k < \delta. \tag{3.1.4}$$

Παρατηρούμε ότι οι ροπές της κατανομής δεν υπάρχουν για όλες τις τιμές της παραμέτρου δ.

Η μέση τιμή και η διασπορά της κατανομής χρησιμοποιώντας τον τύπο των ροπών με την κατάλληλη αντικατάσταση του k είναι αντίστοιχα

$$E(X) = \frac{\lambda}{\delta-1} \tag{3.1.5}$$

η οποία υπάρχει αν και μόνο αν $\delta > 1$, και

$$Var(X) = \frac{\delta\lambda^2}{(\delta-1)^2(\delta-2)} \tag{3.1.6}$$

η οποία είναι πεπερασμένη αν και μόνο αν $\delta > 2$.

Στην περίπτωση που $\delta \leq 1$ δεν υπάρχει η μέση τιμή της Pareto, ενώ αν $\delta \leq 2$ δεν υπάρχει η διασπορά της κατανομής.

Η λοξότητα και η κύρτωση της κατανομής, θέτοντας κατάλληλα τις απαιτούμενες ροπές στις Σχέσεις (2.4.34) και (2.4.37), προκύπτουν ότι είναι αντίστοιχα

$$skew(X) = \frac{2(\delta+1)}{\delta-3}\sqrt{\frac{\delta-2}{\delta}}, \qquad \delta > 3, \tag{3.1.7}$$

$$kurt(X) = \frac{3(\delta-2)(3\delta^2+\delta+2)}{(\delta-4)(\delta-3)\delta}, \qquad \delta > 4\,. \tag{3.1.8}$$

Παρατηρούμε ότι και η λοξότητα και η κύρτωση της Pareto είναι ανεξάρτητες της παραμέτρου λ. Η κατανομή έχει θετική λοξότητα, δηλαδή τείνει προς τα δεξιά, με $skew(X) \to 2$ καθώς $\delta \to \infty$ και $skew(X) \to \infty$ καθώς $\delta \to 3$. Σχετικά με την κύρτωση της κατανομής παρατηρούμε ότι $kurt(X) \to 9$ καθώς $\delta \to \infty$, και $kurt(X) \to \infty$ καθώς $\delta \to 4$.

Η ροπογεννήτρια συνάρτηση της Pareto υπάρχει για $t < 0$ αλλά δεν έχει αναλυτική μορφή, και απειρίζεται για $t > 0$. Αυτό είναι αναμενόμενο αποτέλεσμα, καθώς η κατανομή έχει βαριά ουρά. Ο τύπος της ροπογεννήτριας είναι

$$M_X(t) = \begin{cases} e^{-\delta t}\lambda(-\delta t)^{\lambda}[\Gamma(-\lambda) - \Gamma(-\lambda, -\delta t)], & t < 0 \\ 1, & t = 0 \\ \infty, & t > 0 \end{cases}. \tag{3.1.9}$$

Για μια θετική τυχαία μεταβλητή, $X > 0$, η περιορισμένη μέση της τιμή στο x ορίζεται ως η μέση τιμή της λογοκριμένης X στο x, και δίνεται από τον τύπο

$$E(X; x) = \int_0^x t f_X(t) dt + x[1 - F_X(x)] \tag{3.1.10}$$

ή εναλλακτικά μέσω της δεξιάς ουράς της κατανομής από τη σχέση

$$E(X; x) = \int_0^x [1 - F_X(t)] dt. \tag{3.1.11}$$

Για την $Par^{(II)}(\delta, \lambda)$, η περιορισμένη μέση τιμή είναι

$$E(X; x) = \frac{\lambda}{\delta - 1}\left[1 - \left(\frac{\lambda}{\lambda + x}\right)^{\delta - 1}\right], \ \delta > 1. \tag{3.1.12}$$

Η ποσοστιαία συνάρτηση της κατανομής δίνεται από τον τύπο

$$Q(p) = \lambda\left[\frac{1}{(1-p)^{1/\delta}} - 1\right], \quad p \in (0,1). \tag{3.1.13}$$

Για το 1ο, 2ο, και 3ο τεταρτημόριο θέτοντας στον παραπάνω τύπο $p = 0.25$, $p = 0.5$ και $p = 0.75$ αντίστοιχα προκύπτουν

$$Q_1 = \lambda\left[\left(\frac{4}{3}\right)^{1/\delta} - 1\right], Q_2 = \lambda\left(2^{1/\delta} - 1\right) \text{ και } Q_3 = \lambda\left(4^{1/\delta} - 1\right).$$

3.1.2 Εκτίμηση παραμέτρων

Έστω ένα τυχαίο δείγμα $\mathbf{X} = (X_1, X_2, \dots, X_n)$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από την κατανομή $Par^{(II)}(\delta, \lambda)$. Θα δούμε στη συνέχεια διάφορες μεθόδους εκτίμησης των παραμέτρων δ και λ.

Ξεκινώντας με τη μέθοδο της μέγιστης πιθανοφάνειας, η συνάρτηση πιθανοφάνειας της κατανομής είναι

$$L(\delta, \lambda|\mathbf{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\delta\lambda^{\delta}}{(\lambda + x_i)^{\delta+1}} = \frac{\delta^n \lambda^{\delta n}}{\left[\prod_{i=1}^{n}(\lambda + x_i)\right]^{\delta+1}}$$

Για μεγαλύτερη ευκολία στους υπολογισμούς μας όπως αναφέραμε ήδη, χρησιμοποιούμε το λογάριθμο της συνάρτησης πιθανοφάνειας. Επομένως,

$$l(\delta,\lambda)=\log L(\delta,\lambda|\mathbf{x})=n\log\delta+n\delta\log\lambda-(\delta+1)\sum_{i=1}^{n}\log(\lambda+x_i).$$

Παραγωγίζοντας την παραπάνω συνάρτηση ως προς δ και λ αντίστοιχα και θέτοντας ίσες με μηδέν τις σχέσεις που προκύπτουν βρίσκουμε

$$\frac{\partial l(\delta,\lambda)}{\partial\delta}=0\Rightarrow\frac{n}{\delta}+n\log\lambda-\sum_{i=1}^{n}\log(\lambda+x_i)=0\Rightarrow\frac{n}{\delta}=\sum_{i=1}^{n}\log(\lambda+x_i)-n\log\lambda$$

$$\Rightarrow\delta=\frac{n}{\sum_{i=1}^{n}\log(\lambda+x_i)-n\log\lambda}=\frac{n}{\sum_{i=0}^{n}[\log(\lambda+x_i)-\log\lambda]}$$

$$\Rightarrow\delta=\frac{n}{\sum_{i=1}^{n}\log\frac{\lambda+x_i}{\lambda}} \tag{3.1.14}$$

και

$$\frac{\partial l(\delta,\lambda)}{\partial\lambda}=0\Rightarrow\frac{n\delta}{\lambda}-(\delta+1)\sum_{i=1}^{n}\frac{1}{\lambda+x_i}=0\Rightarrow n\delta=\lambda\delta\sum_{i=1}^{n}\frac{1}{\lambda+x_i}+\lambda\sum_{i=1}^{n}\frac{1}{\lambda+x_i}$$

$$\Rightarrow\delta\left(n-\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}\right)=\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}\Rightarrow\delta\sum_{i=1}^{n}\left(1-\frac{\lambda}{\lambda+x_i}\right)=\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}$$

$$\Rightarrow\delta=\frac{\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}}{\sum_{i=1}^{n}\frac{x_i}{\lambda+x_i}}. \tag{3.1.15}$$

Εξισώνοντας τα δεύτερα μέλη των (3.1.14) και (3.1.15) προκύπτει

$$\frac{n}{\sum_{i=1}^{n}\log\frac{\lambda+x_i}{\lambda}}=\frac{\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}}{\sum_{i=1}^{n}\frac{x_i}{\lambda+x_i}}. \tag{3.1.16}$$

Η παραπάνω είναι μια μη γραμμική εξίσωση ως προς λ και για να λυθεί απαιτούνται αριθμητικές μέθοδοι (π.χ. Newton-Raphson). Συνεπώς καταλήγουμε στο συμπέρασμα ότι οι εκτιμήτριες συναρτήσεις των παραμέτρων δ,λ με τη μέθοδο της μέγιστης πιθανοφάνειας δεν μπορούν να δοθούν σε κλειστή μορφή. Λύνοντας αριθμητικά την (3.1.16) εκτιμάμε το λ και στη συνέχεια από την (3.1.14) ή από την (3.1.15) υπολογίζουμε και την παράμετρο δ.

Η εκτίμηση των παραμέτρων δ και λ μπορεί να πραγματοποιηθεί και με τη μέθοδο των ροπών. Σύμφωνα με την τελευταία, αρκεί να λύσουμε το σύστημα που ακολουθεί

$$\begin{cases}E(X)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i=m_1\\ E(X^2)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i^2=m_2\end{cases}\Rightarrow\begin{cases}\dfrac{\lambda}{\delta-1}=m_1\\ \dfrac{2\lambda^2}{(\delta-1)(\delta-2)}=m_2\end{cases}.$$

Μετά από πράξεις καταλήγουμε ότι οι εκτιμήτριες συναρτήσεις των δ και λ με τη μέθοδο ροπών είναι

$$\begin{cases} \tilde{\delta} = 2\dfrac{m_2 - m_1^2}{m_2 - 2m_1^2} = 2\dfrac{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}{n\sum_{i=1}^n X_i^2 - 2\left(\sum_{i=1}^n X_i\right)^2} \\ \tilde{\lambda} = \dfrac{m_1 m_2}{m_2 - 2m_1^2} = \dfrac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n X_i^2)}{n\sum_{i=1}^n X_i^2 - 2\left(\sum_{i=1}^n X_i\right)^2} \end{cases} \quad (3.1.17)$$

Η τρίτη μέθοδος εκτίμησης των παραμέτρων που θα αναφέρουμε είναι αυτή των ποσοστιαίων σημείων. Εφ' όσον έχουμε κατανομή με δύο παραμέτρους προς εκτίμηση θα χρειαστούμε δύο ποσοστιαία σημεία, έστω $Q(a)$ το a-στό ποσοστημόριο και $Q(\beta)$ το β-οστό ποσοστημόριο, με $0 < \alpha, \beta < 1$. Σχηματίζουμε το σύστημα εξισώσεων

$$\begin{cases} Q(\alpha) = \lambda\left[\dfrac{1}{(1-\alpha)^{1/\delta}} - 1\right] \\ Q(\beta) = \lambda\left[\dfrac{1}{(1-\beta)^{1/\delta}} - 1\right] \end{cases}. \quad (3.1.18)$$

Με διαίρεση κατά μέλη απαλοίφεται το λ και προκύπτει η παρακάτω εξίσωση ως προς δ

$$\left[\frac{1}{(1-\beta)^{1/\delta}} - 1\right] Q(\alpha) = \left[\frac{1}{(1-\alpha)^{1/\delta}} - 1\right] Q(\beta). \quad (3.1.19)$$

Η παραπάνω εξίσωση είναι μη γραμμική ως προς δ και απαιτεί αριθμητικές μεθόδους επίλυσης. Μόλις εκτιμήσουμε το δ, αντικαθιστούμε σε μία από τις (3.1.18) και βρίσκουμε το λ. Στην περίπτωση που χρησιμοποιηθεί η πρώτη, προκύπτει

$$\lambda = \frac{Q(\alpha)}{\dfrac{1}{(1-\alpha)^{\frac{1}{\delta}}} - 1}.$$

3.1.3 Μετασχηματισμοί και ειδικές περιπτώσεις

Στην παράγραφο αυτή θα αναπτύξουμε διάφορους μετασχηματισμούς τυχαίων μεταβλητών σχετικές με την κατανομή Pareto που μελετάμε, οι οποίοι μπορεί να οδηγήσουν είτε σε κατανομές Pareto με διαφορετικές παραμέτρους είτε σε νέες κατανομές.

Αρχικά, αξίζει να αναφέρουμε ότι η κατανομή $Par^{(II)}(\delta, \lambda)$ προκύπτει ως μεμιγμένη εκθετική με μικτική κατανομή την κατανομή γάμμα. Πράγματι, θεωρώντας

$$X|\Theta = \theta \sim f_{X|\Theta}(x|\theta) = \theta e^{-\theta x}, x > 0, \theta > 0$$

και

$$\Theta \sim f_\Theta(\theta) = \frac{\lambda^\delta}{\Gamma(\delta)}\theta^{\delta-1}e^{-\lambda\theta}, \theta > 0, \delta > 0, \lambda > 0,$$

η συνάρτηση πυκνότητας πιθανότητας της τ.μ. X θα δίνεται από τον τύπο

$$X\sim f_X(x)=\int_{\theta} f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)d\theta=\int_0^{\infty}\theta e^{-\theta x}\frac{\lambda^{\delta}}{\Gamma(\delta)}\theta^{\delta-1}e^{-\lambda\theta}d\theta=\frac{\lambda^{\delta}}{\Gamma(\delta)}\int_0^{\infty}\theta^{\delta+1-1}e^{-(\lambda+x)\theta}d\theta$$

$$=\frac{\lambda^{\delta}}{\Gamma(\delta)}\frac{\Gamma(\delta+1)}{(\lambda+x)^{\delta+1}}=\frac{\lambda^{\delta}}{\Gamma(\delta)}\frac{\delta\Gamma(\delta)}{(\lambda+x)^{\delta+1}}=\frac{\delta\lambda^{\delta}}{(\lambda+x)^{\delta+1}},x>0.$$

Η ειδική περίπτωση της κατανομής όπου οι τιμές των παραμέτρων είναι $\delta=\lambda=1$, δηλαδή $X\sim Par^{(II)}(1,1)$ είναι η κατανομή F ή αλλιώς κατανομή Snedecor, $F_{2,2}$. Η συνάρτηση πυκνότητας της $F_{n,k}$ κατανομής είναι

$$f(x)=\frac{\Gamma\left(\frac{n+k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{k}{2}\right)}\left(\frac{n}{k}\right)^{\frac{n}{2}}x^{n/2-1}\left(1+\frac{n}{k}x\right)^{-\frac{n+k}{2}},x>0. \qquad (3.1.20)$$

Α $X\sim Par^{(II)}(1,1)$ τότε ο μετασχηματισμός $X=\ln Y$, οδηγεί στην τυπική λογιστική κατανομή με σ.π.π. και σ.κ. αντίστοιχα ίσες με

$$f(x)=\frac{e^x}{(1+e^x)^2} \quad \text{και} \quad F(x)=\frac{e^x}{1+e^x},\ -\infty<x<\infty. \qquad (3.1.21)$$

Εάν η Y ακολουθεί την Pareto πρώτου είδους, ή αλλιώς *European Pareto*, δηλαδή $Y\sim Par^{(I)}(\lambda,\delta)$, η οποία έχει σ.π.π. και σ.κ. αντίστοιχα ίσες με

$$f(y)=\frac{\delta\lambda^{\delta}}{y^{\delta+1}},y>\lambda>0,\delta>0 \quad \text{και} \quad F(y)=1-\left(\frac{\lambda}{y}\right)^{\delta}, \qquad (3.1.22)$$

τότε ο μετασχηματισμός $X=Y-\lambda$ ακολουθεί την $Par^{(II)}(\delta,\lambda)$.

Συμμετρικά, αν η $X\sim Par^{(II)}(\delta,\lambda)$ τότε με το μετασχηματισμό $Y=X+\lambda$ οδηγούμαστε στην κατανομή Pareto τύπου I, $Par^{(I)}(\lambda,\delta)$.

Αν $U\sim Ομοιόμορφη(0,1)$ τότε με το μετασχηματισμό $X=\lambda\left(\frac{1}{U^{1/\delta}}-1\right),\delta,\lambda>0$ η τ.μ. X ακολουθεί την $Par^{(II)}(\delta,\lambda)$.

Αν θεωρήσουμε δύο τ.μ. Y_1 και Y_2 ανεξάρτητες και ισόνομες που ακολουθούν την εκθετική κατανομή με μέσο 1, δηλαδή έχουν συνάρτηση πυκνότητας πιθανότητας $f(y)=e^{-y},y>0$, τότε ο μετασχηματισμός $X=Y_1/Y_2$, οδηγεί στην κατανομή $Par^{(II)}(1,1)$.

Στη συνέχεια θεωρούμε ότι η τ.μ. $Y\sim Par^{(II)}(\delta,\lambda)$. Τότε έχουμε τα εξής αποτελέσματα.

Ο μετασχηματισμός κλίμακας $X=cY,c>0$, οδηγεί σε κατανομή $Par^{(II)}(\delta,c\lambda)$.

Ο μετασχηματισμός $X=\ln\left(1+\frac{Y}{\lambda}\right)$, οδηγεί στην εκθετική κατανομή, $X\sim\varepsilon(\delta)$ με σ.π.π. $f_X(x)=\delta e^{-\delta x}$.

Ο μετασχηματισμός $X=Y^{1/\gamma},\gamma>0$, οδηγεί σε κατανομή Burr τύπου XII, δηλαδή $X\sim Burr^{(XII)}(\delta,\lambda,\gamma)$ με σ.π.π.

$$f_X(x) = \frac{\delta\gamma\lambda^{\delta}}{(\lambda + x^{\gamma})^{\delta+1}} x^{\gamma-1}, x > 0, \delta > 0, \lambda > 0, \gamma > 0. \quad (3.1.23)$$

Επιπλέον, από το προηγούμενο αποτέλεσμα βλέπουμε ότι η ειδική περίπτωση για $\gamma = 1$ της κατανομής Burr, είναι η κατανομή Pareto, δηλαδή, $Burr^{(XII)}(\delta, \lambda, 1) \equiv Par^{(II)}(\delta, \lambda)$.

Αν $X_i \sim Par^{(II)}(\delta, \lambda), i = 1, \dots, n$, ανεξάρτητες και ισόνομες τ.μ. τότε η κατανομή της πρώτης διατεταγμένης παρατήρησης, $X_{1:n} := \min(X_1, \dots, X_n)$ είναι η $Par^{(II)}(n\delta, \lambda)$. Πράγματι, η σ.κ. της $X_{1:n}$ είναι

$$F_{1:n}(x) = P(X_{1:n} \le x) = 1 - P(X_{1:n} > x) = 1 - P(X_1 > x, \dots, X_n > x) = 1 - \prod_{i=1}^{n} P(X_i > x)$$

$$= 1 - \prod_{i=1}^{n} [1 - P(X_i \le x)] = 1 - \prod_{i=1}^{n} \left[1 - F_{X_i}(x)\right]$$

$$= 1 - \prod_{i=1}^{n} \left[1 - F_{X_1}(x)\right] = 1 - \left[1 - F_{X_1}(x)\right]^n$$

$$= 1 - \left\{1 - \left[1 - \left(\frac{\lambda}{\lambda + x}\right)^{\delta}\right]\right\}^n = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{n\delta}, x > 0$$

που είναι η σ.κ. της $Par^{(II)}(n\delta, \lambda)$.

3.2 Γενικευμένη βήτα παραγόμενη $Pareto^{(II)}$ κατανομή

Στην παράγραφο αυτή εξετάζουμε τη γενικευμένη Βήτα παραγόμενη κατανομή με γεννήτορα την κατανομή Pareto τύπου II.

Η συνάρτηση πυκνότητας πιθανότητάς της προκύπτει αν αντικαταστήσουμε στην (2.4.3) την σ.π.π. και σ.κ. της $Par^{(II)}(\delta, \lambda)$, οπότε παίρνουμε

$$f_{GBGPar}(x; \delta, \lambda, a, b, c) = c \frac{1}{B(a, b)} \frac{\delta\lambda^{\delta}}{(\lambda + x)^{\delta+1}} \left[1 - \left(\frac{\lambda}{\lambda + x}\right)^{\delta}\right]^{ac-1} \left\{1 - \left[1 - \left(\frac{\lambda}{\lambda + x}\right)^{\delta}\right]^c\right\}^{b-1},$$

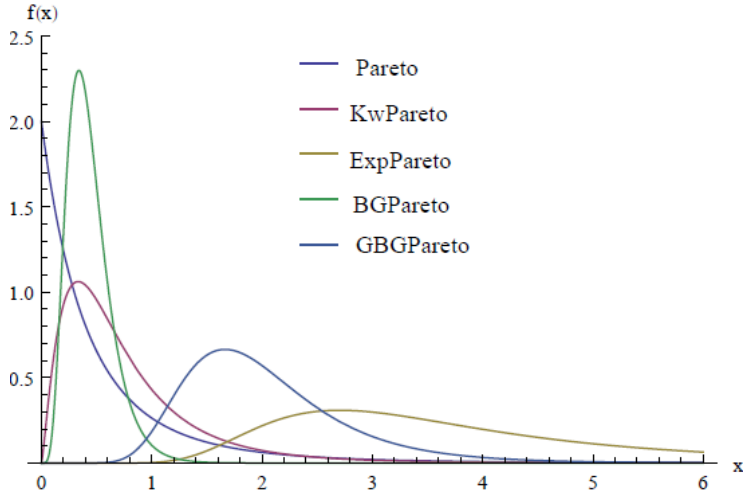
$$x > 0. \quad (3.2.1)$$

Σχετικά με την ασυμπτωτική συμπεριφορά της σ.π.π. παρατηρούμε ότι αν $x \to \infty$ η πυκνότητα τείνει στο 0 για $\delta > 0$.

Η συνάρτηση κατανομής της $GBGPar$ από την (2.4.16) είναι

$$F_{GBGPar}(x; \mathbf{p}, a, b, c) = \frac{1}{B(a, b)} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{(a + j) j! \, \Gamma(b - j)} \left[1 - \left(\frac{\lambda}{\lambda + x}\right)^{\delta}\right]^{c(a+j)}. \quad (3.2.2)$$

Στο Σχήμα 3.3 απεικονίζεται η συνάρτηση πυκνότητας της $GBGPar^{(II)}(\delta,\lambda)$ για διάφορες τιμές των παραμέτρων της. Συγκεκριμένα,, απεικονίζονται οι $Pareto-GBGP(4,2,1,1,1)$, $KwPareto-GBGP(4,2,1,1.5,2.3)$, $ExpPareto-GBGP(4,2,7.6,1,5)$, $BGPareto-GBGP(4,2,6.2,5.7,1)$, $GBGP(4,2,3,2.4,8)$.



Σχήμα 3.3 Συνάρτηση πυκνότητας της $GBGPar(\delta,\lambda,a,b,c)$

Στη συνέχεια θα βρούμε τις ροπές τις $GBGPar$ κατανομής υπολογίζοντας αρχικά τις ροπές της εκθετικοποιημένης $Par^{(II)}(\delta,\lambda)$. Έστω $Y\sim Exp^{(a)}Par^{(II)}(\delta,\lambda)$. Τότε από την (2.4.14) είναι

$$Y\sim g(y)=a\frac{\delta\lambda^{\delta}}{(\lambda+y)^{\delta+1}}\left[1-\left(\frac{\lambda}{\lambda+y}\right)^{\delta}\right]^{a-1},y>0 \tag{3.2.3}$$

η οποία με χρήση της (2.4.11α) γίνεται

$$Y\sim g(y)=a\frac{\delta\lambda^{\delta}}{(\lambda+y)^{\delta+1}}\sum_{j=0}^{\infty}(-1)^{j}\binom{a-1}{j}\left(\frac{\lambda}{\lambda+y}\right)^{\delta j}$$

$$=a\delta\sum_{j=0}^{\infty}(-1)^{j}\binom{a-1}{j}\frac{\lambda^{\delta(j+1)}}{(\lambda+y)^{\delta(j+1)+1}},\ y>0\,. \tag{3.2.4}$$

Για τις ροπές της Y έχουμε

$$E(Y^{k})=\int_{y}y^{k}g(y)dy=\int_{0}^{\infty}y^{k}a\delta\sum_{j=0}^{\infty}(-1)^{j}\binom{a-1}{j}\frac{\lambda^{\delta(j+1)}}{(\lambda+y)^{\delta(j+1)+1}}dy$$

$$=\frac{a\delta}{\lambda}\sum_{j=0}^{\infty}(-1)^{j}\binom{a-1}{j}\int_{0}^{\infty}y^{k}\left(\frac{\lambda}{\lambda+y}\right)^{\delta(j+1)+1}dy.$$

Για το ολοκλήρωμα θεωρούμε την αντικατάσταση

$$t=\frac{\lambda}{\lambda+y}\Rightarrow\lambda+y=\frac{\lambda}{t}\Rightarrow y=\frac{\lambda}{t}-\lambda\Rightarrow dy=-\frac{\lambda}{t^{2}}dt$$

και για τα όρια είναι $0 < y < \infty \Rightarrow 1 < t < 0$, οπότε

$$E(Y^k) = \frac{a\delta}{\lambda}\sum_{j=0}^{\infty}(-1)^j\binom{a-1}{j}\int_1^0\left(\frac{\lambda}{t}-\lambda\right)^k t^{\delta(j+1)+1}\left(-\frac{\lambda}{t^2}dt\right)$$

$$= a\delta\lambda^k\sum_{j=0}^{\infty}(-1)^j\binom{a-1}{j}\int_0^1\left(\frac{1-t}{t}\right)^k t^{\delta(j+1)-1}dt$$

$$= a\delta\lambda^k\sum_{j=0}^{\infty}(-1)^j\binom{a-1}{j}\int_0^1(1-t)^k t^{\delta(j+1)-k-1}dt$$

$$= a\delta\lambda^k\sum_{j=0}^{\infty}(-1)^j\binom{a-1}{j}\frac{\Gamma(\delta(j+1)-k)\Gamma(k+1)}{\Gamma(\delta(j+1)+1)}. \qquad (3.2.5)$$

Από τη Σχέση (2.4.32) με $Y_r \sim Exp^{[c(a+r)]}Par^{(II)}(\alpha,\delta)$ για τις ροπές της $X \sim GBGPar$ ισχύει

$$E(X^k) = \sum_{r=0}^{\infty} w_r E(Y_r^k)$$

$$= \sum_{r=0}^{\infty} w_r c(a+r)\delta\lambda^k\sum_{j=0}^{\infty}(-1)^j\binom{c(a+r)-1}{j}\frac{\Gamma(\delta(j+1)-k)\Gamma(k+1)}{\Gamma(\delta(j+1)+1)}, \qquad (3.2.6)$$

με

$$w_r = \frac{(-1)^r\binom{b}{r}}{B(a,b+1)(a+r)}.$$

Για τη συνάρτηση πυκνότητας της $X_{i:n}$ από την (2.4.54) αντικαθιστώντας τις (3.2.1) και (3.2.2) βρίσκουμε

$$f_{i:n}(x) = \sum_{r=0}^{n-i}\frac{(-1)^r\binom{n-i}{r}}{B(i,n-i+1)}\frac{c\delta\lambda^\delta}{B(a,b)(\lambda+x)^{\delta+1}}\left[1-\left(\frac{\lambda}{\lambda+x}\right)^\delta\right]^{ac-1}\times$$

$$\left[1-\left(1-\left(\frac{\lambda}{\lambda+x}\right)^\delta\right)^c\right]^{b-1}\left[\frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(b)}{(a+j)j!\,\Gamma(b-j)}\left[1-\left(\frac{\lambda}{\lambda+x}\right)^\delta\right]^{c(a+j)}\right]^{i+r-1}.$$

Η παραπάνω σ.π.π. αναλύεται περαιτέρω ως εξής. Για k θετικό ακέραιο από την ταυτότητα

$$\left(\sum_{i=0}^{\infty}a_i\right)^k = \sum_{m_1=0}^{\infty}\dots\sum_{m_k=0}^{\infty}a_{m_1}\dots a_{m_k}$$

παίρνουμε

$$f_{i:n}(x) = \sum_{r=0}^{n-i}\frac{(-1)^r\binom{n-i}{r}c\delta\lambda^\delta}{B(i,n-i+1)B(a,b)(\lambda+x)^{\delta+1}}\left[1-\left(\frac{\lambda}{\lambda+x}\right)^\delta\right]^{ac-1}\left[1-\left(1-\left(\frac{\lambda}{\lambda+x}\right)^\delta\right)^c\right]^{b-1}$$

$$\times\left[\frac{\Gamma(b)}{B(a,b)}\right]^{i+r-1}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\frac{(-1)^{\sum_{j=1}^{i+r-1}m_j}}{\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}\left[1-\left(\frac{\lambda}{\lambda+x}\right)^\delta\right]^{c\sum_{j=1}^{i+r-1}(a+m_j)}$$

ή

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\frac{(-1)^{r+\sum_{j=1}^{i+r-1}m_j}\binom{n-i}{r}[\Gamma(b)]^{i+r-1}B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}{B(i,n-i+1)[B(a,b)]^{i+r}\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}$$

$$\times\frac{c}{B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}\frac{\delta\lambda^{\delta}}{(\lambda+x)^{\delta+1}}\left[1-\left(\frac{\lambda}{\lambda+x}\right)^{\delta}\right]^{c\left[a+\sum_{j=1}^{i+r-1}(a+m_j)\right]-1}\left[1-\left(1-\left(\frac{\lambda}{\lambda+x}\right)^{\delta}\right)^{c}\right]^{b-1}$$

δηλαδή,

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\delta_{i,r}f_{i,r}(x) \tag{3.2.7}$$

όπου

$$f_{i,r}(x)=\frac{c}{B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}\frac{\delta\lambda^{\delta}}{(\lambda+x)^{\delta+1}}\left[1-\left(\frac{\lambda}{\lambda+x}\right)^{\delta}\right]^{c\left[a+\sum_{j=1}^{i+r-1}(a+m_j)\right]-1}$$

$$\times\left[1-\left(1-\left(\frac{\lambda}{\lambda+x}\right)^{\delta}\right)^{c}\right]^{b-1}$$

και οι συντελεστές $\delta_{i,r}$ είναι

$$\delta_{i,r}=\frac{(-1)^{r+\sum_{j=1}^{i+r-1}m_j}\binom{n-i}{r}[\Gamma(b)]^{i+r-1}B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}{B(i,n-i+1)[B(a,b)]^{i+r}\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}.$$

Παρατηρούμε ότι η $f_{i,r}$ είναι η συνάρτηση πυκνότητας μιας κατανομής $GBGPar\left(\delta,\lambda,a+\sum_{j=1}^{i+r-1}(a+m_j),b,c\right)$. Το αποτέλεσμα της Σχέσης (3.2.7) εκφράζει την πυκνότητα της διατεταγμένης στατιστικής $GBGPar$ σαν μείξη $GBGPar$ με συνέπεια διάφορες ιδιότητες της $X_{i:n}$ να μπορούν να δοθούν από αυτή τη διατύπωση της μείξης. Για παράδειγμα, οι s ροπές της $X_{i:n}$, για $b>0$ πραγματικό, μη ακέραιο, δίνονται από τον τύπο

$$E(X_{i:n}^{s})=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\delta_{i,r}E(X_{i:r}^{s}) \tag{3.2.8}$$

όπου $X_{i:r}\sim GBGPar\left(\delta,\lambda,a+\sum_{j=1}^{i+r-1}(a+m_j),b,c\right)$.

Στην περίπτωση που το b είναι θετικός ακέραιος, οι δείκτες $m_1,\dots,m_{i+r-1}$ σταματάνε στο $b-1$.

Για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας, έστω $\mathbf{x}=(x_1,\dots,x_n)$ ένα τυχαίο δείγμα μεγέθους n από την κατανομή και έστω $\boldsymbol{\theta}=(\delta,\lambda,a,b,c)$ το διάνυσμα των παραμέτρων της. Η συνάρτηση πιθανοφάνειας είναι

$$L(\boldsymbol{\theta}\mid\mathbf{x})=\prod_{i=1}^{n}f_{GBGPar}(x_i;\boldsymbol{\theta})$$

$$= \prod_{i=1}^{n} \frac{c}{B(a,b)} \frac{\delta\lambda^{\delta}}{(\lambda+x_i)^{\delta+1}} \left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{ac-1} \left\{1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c}\right\}^{b-1}$$

$$= \frac{c^n\delta^n\lambda^{n\delta}}{[B(a,b)]^n} \frac{1}{\left[\prod_{i=1}^{n}(\lambda+x_i)\right]^{\delta+1}} \left\{\prod_{i=1}^{n}\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]\right\}^{ac-1} \left\{\prod_{i=1}^{n}\left(1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c}\right)\right\}^{b-1}$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}|\mathbf{x}) = n\log c + n\log\delta + n\delta\log\lambda - n\log B(a,b) - (\delta+1)\sum_{i=1}^{n}\log(\lambda+x_i)$$

$$+(ac-1)\sum_{i=1}^{n}\log\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right] + (b-1)\sum_{i=1}^{n}\log\left\{1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c}\right\}.$$

Για τη μεγιστοποίηση της συνάρτησης $l(\boldsymbol{\theta})$ λύνουμε το σύστημα των μη γραμμικών εξισώσεων που προκύπτουν από την παραγώγιση της $l(\boldsymbol{\theta})$ ως προς τις παραμέτρους $\boldsymbol{\theta}$ και την εξίσωσή τους με το μηδέν. Οι πρώτες μερικοί παράγωγοι της $l(\boldsymbol{\theta})$ είναι

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a} = -n[\psi(a)-\psi(a+b)] + c\sum_{i=1}^{n}\log\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right],$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial b} = -n[\psi(b)-\psi(a+b)] + \sum_{i=1}^{n}\log\left\{1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c}\right\},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial c} = \frac{n}{c} + a\sum_{i=1}^{n}\log\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right] + (b-1)\sum_{i=1}^{n}\frac{\log\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]}{1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{-c}},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \delta} = \frac{n}{\delta} + n\log\lambda - \sum_{i=1}^{n}\log(\lambda+x_i) - (ac-1)\sum_{i=1}^{n}\frac{\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\log\left(\frac{\lambda}{\lambda+x_i}\right)}{1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}}$$

$$+c(b-1)\sum_{i=1}^{n}\frac{\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c-1}\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\log\left(\frac{\lambda}{\lambda+x_i}\right)}{1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c}},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = \frac{n\delta}{\lambda} - (\delta+1)\sum_{i=1}^{n}\frac{1}{\lambda+x_i} - \delta(ac-1)\sum_{i=1}^{n}\frac{\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta-1}\frac{x_i}{(\lambda+x_i)^2}}{1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}}$$

$$+\lambda^{\delta-1}\delta c(b-1)\sum_{i=1}^{n}\frac{\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c-1}\frac{x_i}{(\lambda+x_i)^{\delta+1}}}{1-\left[1-\left(\frac{\lambda}{\lambda+x_i}\right)^{\delta}\right]^{c}}.$$

3.3 Κατανομή Weibull

Η κατανομή Weibull ανήκει επίσης στην οικογένεια των εκθετικών κατανομών. Το βασικό της χαρακτηριστικό είναι η ευελιξία του σχήματος της συνάρτησης πυκνότητας πιθανότητάς της. Με την εναλλαγή τιμών στις παραμέτρους της, το σχήμα της κατανομής αλλάζει δραστικά και επιτρέπει την καλύτερη προσέγγιση μεγεθών που χαρακτηρίζονται από βαριές ουρές και ασυμμετρία.

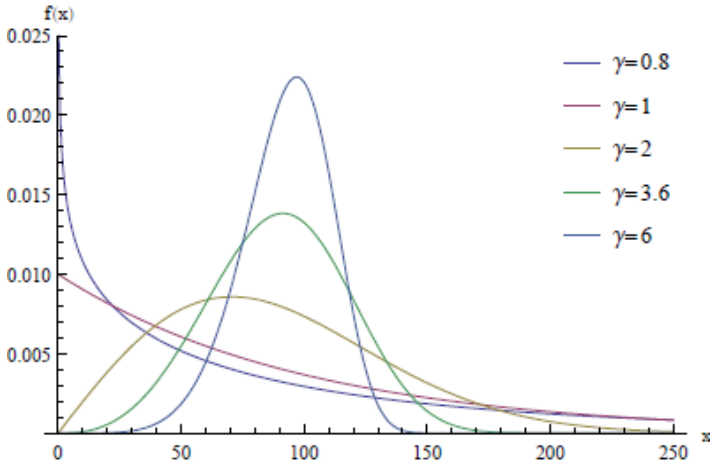
Η κατανομή Weibull μπορεί να είναι από θετικά ασύμμετρη, όπως η Pareto, έως τελείως συμμετρική, όπως η κανονική κατανομή. Η ευελιξία η οποία προσφέρεται προς χρήση, την κάνει να βρίσκει πολλές εφαρμογές σε πληθώρα επιστημών, όπως ανάλυση επιβίωσης, θεωρία ακραίων τιμών, πρόγνωση καιρού, πρόβλεψη μεγέθους ζημιών από ασφαλιζόμενο κίνδυνο, κ.α.

Θα λέμε ότι η τ.μ. X ακολουθεί την κατανομή Weibull με παραμέτρους κλίμακας λ και σχήματος γ και θα συμβολίζουμε $X \sim W(\lambda, \gamma)$ αν έχει συνάρτηση πυκνότητας πιθανότητας και συνάρτηση κατανομής αντίστοιχα

$$f(x) = \lambda\gamma x^{\gamma-1} e^{-\lambda x^{\gamma}}, x > 0, \lambda > 0, \gamma > 0, \tag{3.3.1}$$

$$F(x) = 1 - e^{-\lambda x^{\gamma}}, x > 0, \lambda > 0, \gamma > 0. \tag{3.3.2}$$

Στο Σχήμα 3.4 απεικονίζεται η συνάρτηση πυκνότητας της Weibull για διάφορες τιμές των παραμέτρων της.



Σχήμα 3.4 Συνάρτηση πυκνότητας της $W(\lambda, \gamma)$ για διάφορες τιμές του γ με $\lambda = 0.01^{\gamma}$

Για $\gamma = 1$ η κατανομή Weibull ταυτίζεται με την εκθετική κατανομή. Για $\gamma = 2$ η κατανομή Weibull ταυτίζεται με την κατανομή Rayleigh. Για τιμές $1 < \gamma < 3.6$ η Weibull προσεγγίζει την

Lognormal κατανομή. Για τις τιμές $3 < \gamma < 4$ προσεγγίζει την κανονική κατανομή, με την καλύτερη προσέγγιση να επιτυγχάνεται όταν $\gamma = 3.6$ οπότε έχει και την ελάχιστη λοξότητα. Για $\gamma = 6$ η κατανομή Weibull προσεγγίζει μια κανονική κατανομή με στενή και ψηλή καμπύλη.

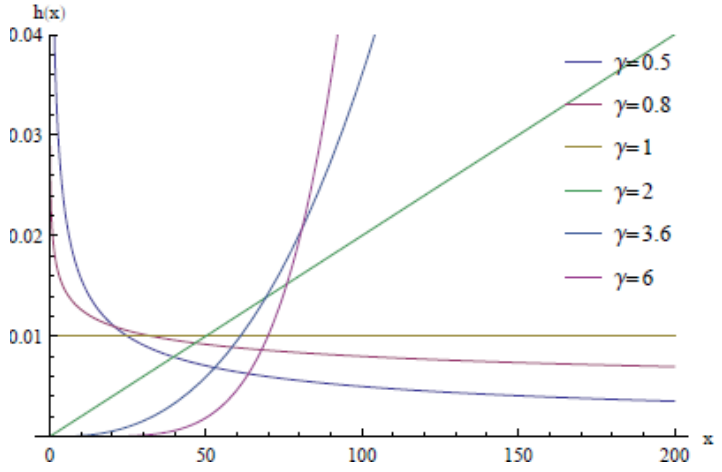
Το μέγιστο (mode), της κατανομής πετυχαίνεται στο σημείο $\left(\frac{\gamma-1}{\lambda\gamma}\right)^{1/\gamma}$ αν $\gamma > 1$ και στο 0 αν $\gamma = 1$. Στην επόμενη παράγραφο, αναπτύσσονται βασικά μεγέθη και χαρακτηριστικά της κατανομής Weibull.

3.3.1 Χαρακτηριστικά μεγέθη της κατανομής Weibull

Η συνάρτηση κινδύνου δίνεται από τη σχέση

$$h(x) = \frac{f(x)}{1 - F(x)} = \lambda\gamma x^{\gamma-1} \tag{3.3.3}$$

και απεικονίζεται στο Σχήμα 3.5.



Σχήμα 3.5 Συνάρτηση κινδύνου της $W(\lambda, \gamma)$ με $\lambda = 0.01^{\gamma}$

Αν $\gamma = 1$ έχουμε την ειδική περίπτωση της κατανομής Weibull, την εκθετική, η οποία έχει σταθερή συνάρτηση διακινδύνευσης $h(x) = \lambda$.

Επιπλέον, η πρώτη παράγωγος της συνάρτησης κινδύνου είναι $h'(x) = \lambda\gamma(\gamma - 1)x^{\gamma-2}$ από την οποία μπορούμε να διακρίνουμε τις περιπτώσεις:

Αν $\gamma > 1$ τότε $h'(x) > 0$ και η συνάρτηση διακινδύνευσης είναι γνησίως αύξουσα συνάρτηση, δηλαδή αυξάνει ο ρυθμός διακοπής.

Αν $\gamma < 1$ τότε $h'(x) < 0$ και η συνάρτηση διακινδύνευσης είναι γνησίως φθίνουσα συνάρτηση, δηλαδή ο ρυθμός διακοπής φθίνει.

Αν $\gamma = 2$, έχουμε την ειδική περίπτωση της κατανομής Weibull που ονομάζεται κατανομή *Rayleigh*, και η συνάρτηση διακινδύνευσης είναι μια ευθεία $h(x) = 2\lambda x$ με κλίση 2λ.
Σύμφωνα με τον Klugman (2004), κατανομές που έχουν αύξουσα συνάρτηση κινδύνου χαρακτηρίζονται ως κατανομές με ελαφριά ουρά.

Οι ροπές k τάξης της κατανομής δίνονται από τον τύπο

$$E(X^k) = \frac{\Gamma\left(\frac{k}{\gamma} + 1\right)}{\lambda^{k/\gamma}}, k > -\gamma. \tag{3.3.4}$$

Η μέση τιμή και η διασπορά της κατανομής χρησιμοποιώντας τον τύπο των ροπών με κατάλληλη αντικατάσταση του k προκύπτουν ότι είναι

$$\mu = E(X) = \frac{\Gamma\left(\frac{1}{\gamma} + 1\right)}{\lambda^{1/\gamma}}, \tag{3.3.5}$$

$$\sigma^2 = Var(X) = \lambda^{-2/\gamma}\left[\Gamma\left(\frac{2}{\gamma} + 1\right) - \left(\Gamma\left(\frac{1}{\gamma} + 1\right)\right)^2\right]. \tag{3.3.6}$$

Η λοξότητα και η κύρτωση της κατανομής προκύπτουν με την κατάλληλη αντικατάσταση των απαιτούμενων ροπών αντίστοιχα στους τύπους (2.4.34) και (2.4.37). Για τη λοξότητα είναι,

$$skew(X) = \frac{E(X^3) - 3\mu E(X^2) - 2\mu^3}{\sigma^3}$$

$$= \frac{\Gamma\left(\frac{3}{\gamma} + 1\right) + 2\left[\Gamma\left(\frac{1}{\gamma} + 1\right)\right]^3 - 3\Gamma\left(\frac{1}{\gamma} + 1\right)\Gamma\left(\frac{2}{\gamma} + 1\right)}{\left[\Gamma\left(\frac{2}{\gamma} + 1\right) - \left[\Gamma\left(\frac{1}{\gamma} + 1\right)\right]^2\right]^{3/2}}. \tag{3.3.7}$$

Βλέπουμε ότι η λοξότητα είναι ανεξάρτητη της παραμέτρου λ. Επιπλέον, για διάφορες τιμές του γ προκύπτουν τα εξής συμπεράσματα. Καθώς γ τείνει στο μηδέν ($\gamma \to 0$) η λοξότητα της κατανομής τείνει στο άπειρο. Καθώς το γ τείνει στο άπειρο ($\gamma \to \infty$) η λοξότητα της κατανομής τείνει στο μηδέν. Εάν $\gamma < 3.6$ η λοξότητα είναι θετική, οπότε η κατανομή είναι λοξή προς τα δεξιά. Εάν $\gamma = 3.6$ η λοξότητα είναι μηδέν, οπότε η κατανομή είναι προσεγγιστικά συμμετρική. Εάν $\gamma > 3.6$ η λοξότητα είναι αρνητική, οπότε η κατανομή είναι λοξή προς τα αριστερά.

Αντίστοιχα για την κύρτωση είναι,

$$kurt(X) = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}$$

$$= \frac{\Gamma\left(\frac{4}{\gamma}+1\right) - 4\Gamma\left(\frac{1}{\gamma}+1\right)\Gamma\left(\frac{3}{\gamma}+1\right) + 6\Gamma^2\left(\frac{1}{\gamma}+1\right)\Gamma\left(\frac{2}{\gamma}+1\right) - 3\Gamma^4\left(\frac{1}{\gamma}+1\right)}{\left[\Gamma\left(\frac{2}{\gamma}+1\right) - \Gamma^2\left(\frac{1}{\gamma}+1\right)\right]^2}. \quad (3.3.7)$$

Παρατηρούμε ότι και η κύρτωση της Weibull κατανομής είναι ανεξάρτητη της παραμέτρου λ.

Η ροπογεννήτρια συνάρτηση της $W(\lambda,\gamma)$ εξ ορισμού προκύπτει από τη σχέση

$$M_X(t) = E(e^{tX}) = \int_x e^{tx} f(x)dx = \int_0^\infty e^{tx}\,\lambda\gamma x^{\gamma-1}e^{-\lambda x^\gamma}dx. \quad (3.3.8)$$

Για $\gamma < 1$, το ολοκλήρωμα δεν συγκλίνει, οπότε η ροπογεννήτρια δεν υπάρχει για $t > 0$ δηλαδή, $M_X(t) = \infty, \text{για } t > 0$.

Για $\gamma > 1$, η ροπογεννήτρια υπάρχει δηλαδή, $M_X(t) < \infty, \text{για } t > 0$, αλλά δεν γνωρίζουμε το γενικό της τύπο και δεν μπορούμε να την υπολογίσουμε αναλυτικά.

Για $\gamma = 1$, έχουμε την εκθετική κατανομή με ροπογεννήτρια συνάρτηση $M_X(t) = \frac{\lambda}{\lambda - t}, t < \lambda$.

Η περιοριορισμένη μέση τιμή της $W(\lambda,\gamma)$ δίνεται από τον τύπο

$$E(X;x) = \lambda^{-1/\gamma}\,\gamma\left(\frac{1}{\gamma}+1, \lambda x^\gamma\right) + xe^{-\lambda x^\gamma}. \quad (3.3.9)$$

Για την ποσοστιαία συνάρτηση της κατανομής έχουμε τη σχέση

$$Q(p) = \left[-\frac{1}{\lambda}\ln(1-p)\right]^{1/\gamma}, p \in (0,1). \quad (3.3.10)$$

Θέτοντας στον τύπο $p = 0.25, p = 0.5$ και $p = 0.75$ προκύπτουν αντίστοιχα το 1ο, 2ο και 3ο τεταρτημόρια που είναι αντίστοιχα ίσα με

$$Q_1 = \left[-\frac{1}{\lambda}\ln\left(\frac{4}{3}\right)\right]^{1/\gamma}, \quad Q_2 = \left(\frac{\ln 2}{\lambda}\right)^{1/\gamma}, \quad Q_3 = \left(\frac{\ln 4}{\lambda}\right)^{1/\gamma}.$$

3.3.2 Εκτίμηση παραμέτρων

Στην παράγραφο αυτή δίνουμε τις εκτιμήτριες συναρτήσεις των παραμέτρων λ και γ με διάφορες μεθόδους. Θεωρούμε ένα τυχαίο δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από την κατανομή $W(\lambda,\gamma)$.

Ξεκινώντας με τη μέθοδο της μέγιστης πιθανοφάνειας, η συνάρτηση πιθανοφάνειας είναι η παρακάτω

$$L(\lambda,\gamma|\mathbf{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda\gamma x_i^{\gamma-1} e^{-\lambda x_i^{\gamma}} = \lambda^n \gamma^n \left(\prod_{i=1}^{n} x_i\right)^{\gamma-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i^{\gamma}\right).$$

Λογαριθμίζοντας τη συνάρτηση πιθανοφάνειας είναι

$$l(c,\gamma) = \log L(c,\gamma|\mathbf{x}) = n\log\lambda + n\log\gamma + (\gamma-1)\sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i^{\gamma}.$$

Παίρνουμε τις παραγώγους ως προς τις παραμέτρους και τις θέτουμε ίσες με το μηδέν

$$\frac{\partial l(\lambda,\gamma)}{\partial \lambda} = 0 \Rightarrow \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^{\gamma} = 0 \Rightarrow \frac{n}{\lambda} = \sum_{i=1}^{n} x_i^{\gamma} \Rightarrow \lambda = \frac{n}{\sum_{i=1}^{n} x_i^{\gamma}} \tag{3.3.11}$$

και

$$\frac{\partial l(\lambda,\gamma)}{\partial \gamma} = 0 \Rightarrow \frac{n}{\gamma} + \sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i^{\gamma} \log x_i = 0. \tag{3.3.12}$$

Αντικαθιστώντας την (3.3.11) στην (3.3.12) προκύπτει

$$\frac{n}{\gamma} + \sum_{i=1}^{n} \log x_i - \frac{n}{\sum_{i=1}^{n} x_i^{\gamma}} \sum_{i=1}^{n} x_i^{\gamma} \log x_i = 0 \tag{3.3.13}$$

η οποία είναι μη γραμμική εξίσωση ως προς γ και απαιτούνται αριθμητικές μέθοδοι για την επίλυσή της. Από την τελευταία αυτή εξίσωση βρίσκουμε την εκτίμηση μέγιστης πιθανοφάνειας $\hat{\gamma}$ της παραμέτρου γ και με αντικατάσταση στην (3.3.11) παίρνουμε την εκτίμηση $\hat{\lambda}$ της παραμέτρου λ.

Συνεχίζοντας, χρησιμοποιώντας τη μέθοδο των ροπών οι εκτιμήτριες συναρτήσεις των παραμέτρων λ και γ δίνονται από την επίλυση του συστήματος

$$\begin{cases} E(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i = m_1 \\ E(X^2) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 = m_2 \end{cases} \Rightarrow \begin{cases} \dfrac{\Gamma\left(\frac{1}{\gamma}+1\right)}{\lambda^{\frac{1}{\gamma}}} = m_1 \\ \dfrac{\Gamma\left(\frac{2}{\gamma}+1\right)}{\lambda^{\frac{2}{\gamma}}} = m_2 \end{cases} \Rightarrow \begin{cases} \lambda = \left[\dfrac{1}{m_1}\Gamma\left(\frac{1}{\gamma}+1\right)\right]^{\gamma} \\ \lambda = \left[\dfrac{1}{m_2}\Gamma\left(\frac{2}{\gamma}+1\right)\right]^{\frac{\gamma}{2}} \end{cases} \tag{3.3.14}$$

όπου εξισώνοντας τις δύο παραπάνω σχέσεις προκύπτει

$$\frac{1}{m_1}\Gamma\left(\frac{1}{\gamma}+1\right) = \left[\frac{1}{m_2}\Gamma\left(\frac{2}{\gamma}+1\right)\right]^{\frac{1}{2}}. \tag{3.3.15}$$

Η τελευταία είναι μια μη γραμμική εξίσωση ως προς γ, και για την επίλυσή της απαιτούνται αριθμητικές μέθοδοι. Βρίσκοντας το γ, και κάνοντας χρήση μιας από τις Σχέσεις (3.3.14) εκτιμάμε και την παράμετρο λ.

Τέλος, για την εκτίμηση των παραμέτρων εξετάζουμε και τη μέθοδο των ποσοστιαίων σημείων. Θεωρούμε δύο σημεία $Q(a)$ και $Q(\beta)$ ως το a-οστό και το β-οστό ποσοστημόρια αντίστοιχα, με $0 < \alpha, \beta < 1$, και σχηματίζουμε το σύστημα εξισώσεων,

$$\begin{cases} Q(\alpha) = \left(-\frac{1}{\lambda}\ln(1-a)\right)^{1/\gamma} \\ Q(\beta) = \left(-\frac{1}{\lambda}\ln(1-\beta)\right)^{1/\gamma} \end{cases}. \tag{3.3.16}$$

Διαιρώντας τις παραπάνω σχέσεις κατά μέλη προκύπτει η εκτίμηση της παραμέτρου γ,

$$\frac{Q(a)}{Q(\beta)} = \left(\frac{\ln(1-a)}{\ln(1-b)}\right)^{1/\gamma} \Rightarrow \ln\frac{Q(a)}{Q(\beta)} = \frac{1}{\gamma}\ln\left(\frac{\ln(1-a)}{\ln(1-b)}\right)$$

$$\Rightarrow \gamma = \frac{\ln\left(\frac{\ln(1-a)}{\ln(1-b)}\right)}{\ln\left(\frac{Q(a)}{Q(\beta)}\right)}. \tag{3.3.17}$$

Για την παράμετρο λ αντικαθιστώντας σε μια από τις Σχέσεις (3.3.16), έστω στην πρώτη, προκύπτει

$$[Q(\alpha)]^{\gamma} = -\frac{\ln(1-a)}{\lambda} \Rightarrow \lambda = -\frac{\ln(1-\alpha)}{[Q(\alpha)]^{\gamma}}. \tag{3.3.18}$$

Παρατηρούμε ότι η συγκεκριμένη μέθοδος είναι η μόνη από τις τρεις που εξετάσαμε, με την οποία καταφέραμε να εκτιμήσουμε τις παραμέτρους της κατανομής αναλυτικά.

3.3.3 Μετασχηματισμοί και ειδικές περιπτώσεις

Αρχικά δίνουμε κάποιες περιπτώσεις της κατανομής Weibull για συγκεκριμένες τιμές των παραμέτρων της.

Για $\gamma = 1$, η κατανομή που προκύπτει είναι η εκθετική $\varepsilon(\lambda)$ με πυκνότητα πιθανότητας

$$f(x;\lambda) = \lambda e^{-\lambda x}, \qquad x > 0. \tag{3.3.19}$$

Για $\gamma = 2$ και $\lambda = \frac{1}{2\sigma^2}$, η κατανομή που προκύπτει είναι η Rayleigh, $Ra(\sigma)$ με πυκνότητα πιθανότητας

$$f(x;\sigma) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}, \ \ x > 0. \tag{3.3.20}$$

Στη συνέχεια, αναφέρουμε σχετικούς μετασχηματισμούς που μπορούν να προκύψουν από την κατανομή Weibull, οι οποίοι είτε οδηγούν πάλι σε Weibull κατανομή με διαφορετικές παραμέτρους, είτε σε νέες κατανομές. Για την πρώτη περίπτωση έχουμε τα εξής:

Αν $Y \sim W(\lambda, \gamma)$ τότε ο μετασχηματισμός $X = aY, a > 0$ οδηγεί στην κατανομή $W\left(\frac{\lambda}{a^{\gamma}}, \gamma\right)$.

Αν $Y \sim W(\lambda, \gamma)$ τότε ο μετασχηματισμός $X = Y + \theta, \theta > 0$ οδηγεί στη Weibull κατανομή τριών παραμέτρων $W(\lambda, \gamma, \theta)$ όπου θ είναι παράμετρος θέσης [Weibull, (1951)], με σ.π.π. και σ.κ. αντίστοιχα ίσες με

$$f(x) = \lambda\gamma(x-\theta)^{\gamma-1}e^{-\lambda(x-\theta)^{\gamma}}, y > \theta, \qquad F(x) = 1 - e^{-\lambda(x-\theta)^{\gamma}}. \tag{3.3.21}$$

Παρατηρούμε ότι για $\theta = 0$ οδηγούμαστε στην $W(\lambda, \gamma)$.

Επιπλέον, ξεκινώντας από την $Y \sim W(\lambda, \gamma)$ μπορούμε εύκολα να καταλήξουμε στην εκθετική κατανομή. Συγκεκριμένα,
ο μετασχηματισμός $X = Y^{\gamma}$ οδηγεί στην εκθετική κατανομή $\varepsilon(\lambda)$, ενώ
ο μετασχηματισμός $X = \lambda Y^{\gamma}$ οδηγεί στην εκθετική κατανομή $\varepsilon(1)$.

Η Weibull κατανομή σχετίζεται και με την κατανομή Gumbel. Ειδικότερα, αν $Y \sim W(\lambda, \gamma)$, ο μετασχηματισμός $X = \mu + \sigma \ln(\lambda Y^{\gamma}), \mu \in \mathbf{R}, \sigma > 0$ οδηγεί στην κατανομή $Gumbel(\mu, \sigma)$, όπου μ είναι παράμετρος θέσης και σ είναι παράμετρος κλίμακας, και η οποία έχει σ.π.π.

$$f(x) = \frac{1}{\sigma}\exp\left[\frac{x-\mu}{\sigma} - \exp\left(\frac{x-\mu}{\sigma}\right)\right], \qquad x \in \mathbf{R}, \tag{3.3.22}$$

και σ.κ.

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right], \qquad x \in \mathbf{R}. \tag{3.3.23}$$

Αντίστροφα, θεωρώντας ότι η τυχαία μεταβλητή $X \sim Gumbel(\mu, \sigma)$, τότε ο μετασχηματισμός $Y = \left[\frac{1}{\lambda}\exp\left(\frac{X-\mu}{\sigma}\right)\right]^{1/\gamma}$ οδηγεί στην κατανομή $W(\lambda, \gamma)$.

Επίσης, η κατανομή $X \sim W(\lambda, \gamma)$ προκύπτει από την ομοιόμορφη $U \sim Uniform(0,1)$ με τον μετασχηματισμό $X = \left(\frac{-\ln U}{\lambda}\right)^{1/\gamma}$.

Τέλος, αναφέρουμε ότι για ένα τυχαίο δείγμα $(X_1, \dots, X_n)$ από την $W(\lambda, \gamma)$ κατανομή, η τυχαία μεταβλητή της πρώτης διατεταγμένης παρατήρησης $X_{1:n} := \min(X_1, \dots, X_n)$ ακολουθεί την κατανομή $W(n\lambda, \gamma)$.

3.4 Γενικευμένη βήτα παραγόμενη Weibull κατανομή

Η γενικευμένη βήτα παραγόμενη κατανομή με γεννήτορα την $W(\lambda, \gamma)$ προκύπτει αν αντικαταστήσουμε στην (2.4.3) τη συνάρτηση πυκνότητας πιθανότητας και τη συνάρτηση κατανομής από τις (3.3.1) και (3.3.2) οπότε βρίσκουμε

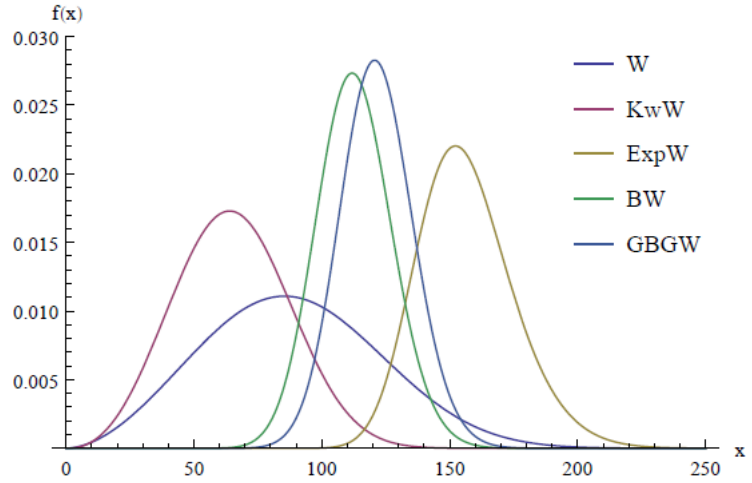
$$f_{GBGW}(x; \lambda, \gamma, a, b, c) = \frac{c}{B(a,b)}\lambda\gamma x^{\gamma-1}e^{-\lambda x^{\gamma}}\left(1 - e^{-\lambda x^{\gamma}}\right)^{ac-1}\left[1 - \left(1 - e^{-\lambda x^{\gamma}}\right)^{c}\right]^{b-1},$$

$$x > 0, \qquad (3.4.1)$$

Η συνάρτηση κατανομής της $GBGW$ που προκύπτει από την (2.4.16) παίρνει τη μορφή

$$F_{GBGW}(x;\lambda,\gamma,a,b,c) = \frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(b)}{(a+j)j!\,\Gamma(b-j)}\left(1-e^{-\lambda x^{\gamma}}\right)^{c(a+j)}. \qquad (3.4.2)$$

Στο Σχήμα 3.6 απεικονίζεται η συνάρτηση πυκνότητας της $GBGW(\lambda,\gamma)$ για διάφορες τιμές των παραμέτρων της. Συγκεκριμένα, διατηρώντας τις παραμέτρους του γεννήτορα σταθερές και ίσες με $\lambda = (1/100)^{\gamma}$ και $\gamma = 2.8$, απεικονίζονται οι $W-GBGW(\lambda,2.8,1,1,1,1)$, $KwW-GBGW(\lambda,2.8,1,3.2,1.2)$, $ExpW-GBGW(\lambda,2.8,4,1,5.2)$, $BW-GBGW(\lambda,2.8,9,3.2,1)$, $GBGW(\lambda,2.8,2.3,3,4.3)$.



Σχήμα 3.6 Συνάρτηση πυκνότητας της $GBGW(\lambda,\gamma,a,b,c)$

Για να δώσουμε εκφράσεις για τις ροπές της $GBGW$ χρησιμοποιούμε το αποτέλεσμα των Mudholkar and Srivastava (1993) σχετικά με τις ροπές της εκθετικοποιημένης Weibull. Έστω $Y\sim Exp^{(a)}W(\lambda,\gamma)$, τότε από την (2.4.14) είναι

$$Y\sim g(y) = a\lambda\gamma x^{\gamma-1}e^{-\lambda x^{\gamma}}\left(1-e^{-\lambda x^{\gamma}}\right)^{a-1}, \qquad y>0. \qquad (3.4.3)$$

Για $a\notin\mathbf{N}$ είναι

$$E\left(Y^k\right) = a\lambda^{-\frac{k}{\gamma}}\Gamma\left(\frac{k}{\gamma}+1\right)\sum_{i=0}^{\infty}\frac{(1-a)_i}{i!}(i+1)^{-\frac{k}{\gamma}-1}, \quad k=0,1,2,\ldots \qquad (3.4.4)$$

όπου $(x)_i = x(x+1)(x+2)\cdots(x+i-1)$ συμβολίζουμε το ανοδικό παραγοντικό i τάξης. Αν το $a\in\mathbf{N}$, τότε ο δείκτης i στο άθροισμα σταματάει στο $a-1$.

Από τη Σχέση (2.4.32) με $Y_j\sim Exp^{[c(a+j)]}W(\lambda,\gamma)$ για τις ροπές της $X\sim GBGW$ ισχύει

$$E(X^s) = \lambda^{-\frac{k}{\gamma}}\Gamma\left(\frac{k}{\gamma}+1\right)\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}w_j\frac{c(a+j)(1-c(a+j)_i)}{i!}(i+1)^{-\frac{k}{\gamma}-1}. \qquad (3.4.5)$$

Για τη συνάρτηση πυκνότητας της $X_{i:n}$ από την (2.4.54) αντικαθιστώντας τις (3.4.1) και (3.4.2) βρίσκουμε

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\frac{(-1)^r\binom{n-i}{r}}{B(i,n-i+1)}\frac{c\lambda\gamma x^{\gamma-1}e^{-\lambda x^{\gamma}}}{B(a,b)}\left(1-e^{-\lambda x^{\gamma}}\right)^{ac-1}\left[1-\left(1-e^{-\lambda x^{\gamma}}\right)^{c}\right]^{b-1}\times$$

$$\left[\frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(b)}{(a+j)j!\,\Gamma(b-j)}\left(1-e^{-\lambda x^{\gamma}}\right)^{c(a+j)}\right]^{i+r-1}$$

η οποία με χρήση της σχέσης

$$\left(\sum_{i=0}^{\infty}a_i\right)^k=\sum_{m_1=0}^{\infty}\dots\sum_{m_k=0}^{\infty}a_{m_1}\dots a_{m_k}$$

γίνεται

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\frac{(-1)^r\binom{n-i}{r}c\lambda\gamma x^{\gamma-1}e^{-\lambda x^{\gamma}}}{B(i,n-i+1)B(a,b)}\left(1-e^{-\lambda x^{\gamma}}\right)^{ac-1}\left[1-\left(1-e^{-\lambda x^{\gamma}}\right)^{c}\right]^{b-1}$$

$$\times\left[\frac{\Gamma(b)}{B(a,b)}\right]^{i+r-1}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\frac{(-1)^{\sum_{j=1}^{i+r-1}m_j}}{\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}\left(1-e^{-\lambda x^{\gamma}}\right)^{c\sum_{j=1}^{i+r-1}(a+m_j)}$$

ή

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\frac{(-1)^{r+\sum_{j=1}^{i+r-1}m_j}\binom{n-i}{r}[\Gamma(b)]^{i+r-1}B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}{B(i,n-i+1)[B(a,b)]^{i+r}\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}$$

$$\times\frac{c\lambda\gamma x^{\gamma-1}e^{-\lambda x^{\gamma}}}{B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}\left(1-e^{-\lambda x^{\gamma}}\right)^{c\left[a+\sum_{j=1}^{i+r-1}(a+m_j)\right]-1}\left[1-\left(1-e^{-\lambda x^{\gamma}}\right)^{c}\right]^{b-1}$$

δηλαδή,

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\delta_{i,r}f_{i,r}(x)\tag{3.4.6}$$

όπου

$$f_{i,r}(x)=\frac{c\lambda\gamma x^{\gamma-1}e^{-\lambda x^{\gamma}}}{B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}\left(1-e^{-\lambda x^{\gamma}}\right)^{c\left[a+\sum_{j=1}^{i+r-1}(a+m_j)\right]-1}\left[1-\left(1-e^{-\lambda x^{\gamma}}\right)^{c}\right]^{b-1}$$

και οι συντελεστές $\delta_{i,r}$ είναι

$$\delta_{i,r}=\frac{(-1)^{r+\sum_{j=1}^{i+r-1}m_j}\binom{n-i}{r}[\Gamma(b)]^{i+r-1}B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}{B(i,n-i+1)[B(a,b)]^{i+r}\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}.$$

Παρατηρούμε ότι η $f_{i,r}$ είναι η συνάρτηση πυκνότητας μιας κατανομής $GBGW\left(\lambda,\gamma,a+\sum_{j=1}^{i+r-1}(a+m_j),b,c\right)$. Το αποτέλεσμα της Σχέσης (3.4.6) εκφράζει την πυκνότητα της διατεταγμένης στατιστικής $GBGW$ σαν μείξη $GBGW$ με συνέπεια διάφορες

ιδιότητες της $X_{i:n}$ να μπορούν να δοθούν από αυτή τη διατύπωση της μείξης. Για παράδειγμα, οι s ροπές της $X_{i:n}$, για $b > 0$ πραγματικό, μη ακέραιο, δίνονται από τον τύπο

$$E(X_{i:n}^s) = \sum_{r=0}^{n-i} \sum_{m_1=0}^{\infty} \dots \sum_{m_{i+r-1}=0}^{\infty} \delta_{i,r} E(X_{i:r}^s) \tag{3.4.7}$$

όπου $X_{i:r} \sim GBGW\left(\lambda, \gamma, a + \sum_{j=1}^{i+r-1}(a + m_j), b, c\right)$.

Στην περίπτωση που το b είναι θετικός ακέραιος, οι δείκτες $m_1, \dots, m_{i+r-1}$ σταματάνε στο $b - 1$.

Για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας, έστω $\mathbf{x} = (x_1, \dots, x_n)$ ένα τυχαίο δείγμα μεγέθους n από την κατανομή *GBGW* και έστω $\boldsymbol{\theta} = (\lambda, \gamma, a, b, c)$ το διάνυσμα των παραμέτρων της. Η συνάρτηση πιθανοφάνειας είναι

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}) &= \prod_{i=1}^{n} f_{GBGW}(x_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \frac{c}{B(a,b)} \lambda\gamma x_i^{\gamma-1} e^{-\lambda x_i^\gamma} \left[1 - e^{-\lambda x_i^\gamma}\right]^{ac-1} \left[1 - \left(1 - e^{-\lambda x_i^\gamma}\right)^c\right]^{b-1} \\ &= \frac{c^n \lambda^n \gamma^n}{[B(a,b)]^n} \left(\prod_{i=1}^{n} x_i\right)^{\gamma-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i^\gamma\right) \left[\prod_{i=1}^{n} (1 - e^{-\lambda x_i^\gamma})\right]^{ac-1} \left\{\prod_{i=1}^{n} \left[1 - \left(1 - e^{-\lambda x_i^\gamma}\right)^c\right]\right\}^{b-1}. \end{aligned}$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι

$$l(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\mathbf{x}) = n\log c + n\log\lambda + n\log\gamma - n\log B(a,b) + (\gamma - 1)\sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i^\gamma$$

$$+(ac - 1)\sum_{i=1}^{n} \log\left(1 - e^{-\lambda x_i^\gamma}\right) + (b - 1)\sum_{i=1}^{n} \log\left[1 - \left(1 - e^{-\lambda x_i^\gamma}\right)^c\right].$$

Για τη μεγιστοποίηση της συνάρτησης $l(\boldsymbol{\theta})$ λύνουμε το σύστημα των μη γραμμικών εξισώσεων που προκύπτουν από την παραγώγιση της $l(\boldsymbol{\theta})$ ως προς τις παραμέτρους $\boldsymbol{\theta}$ και την εξίσωσή τους με το μηδέν. Οι ζητούμενες παράγωγοι είναι

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a} = -n[\psi(a) - \psi(a + b)] + c\sum_{i=1}^{n} \log\left(1 - e^{-\lambda x_i^\gamma}\right),$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial b} = -n[\psi(b) - \psi(a + b)] + \sum_{i=1}^{n} \log\left[1 - \left(1 - e^{-\lambda x_i^\gamma}\right)^c\right],$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial c} = \frac{n}{c} + a\sum_{i=1}^{n} \log\left(1 - e^{-\lambda x_i^\gamma}\right) - (b - 1)\sum_{i=1}^{n} \frac{\log\left(1 - e^{-\lambda x_i^\gamma}\right)}{\left(1 - e^{-\lambda x_i^\gamma}\right)^{-c} - 1},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^{\gamma} + (ac-1)\sum_{i=1}^{n} \frac{x_i^{\gamma}}{e^{\lambda x_i^{\gamma}} - 1} - c(b-1)\sum_{i=1}^{n} \frac{\left(1-e^{-\lambda x_i^{\gamma}}\right)^{c-1} x_i^{\gamma} e^{-\lambda x_i^{\gamma}}}{1-\left(1-e^{-\lambda x_i^{\gamma}}\right)^{c}},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i^{\gamma} \log x_i + \lambda(ac-1)\sum_{i=1}^{n} \frac{x_i^{\gamma} \log x_i}{e^{\lambda x_i^{\gamma}} - 1}$$

$$-\lambda c(b-1)\sum_{i=1}^{n} \frac{\left(1-e^{-\lambda x_i^{\gamma}}\right)^{c-1} e^{-\lambda x_i^{\gamma}} x_i^{\gamma} \log x_i}{1-\left(1-e^{-\lambda x_i^{\gamma}}\right)^{c}}.$$

3.5 Λογιστική κατανομή

Η λογιστική (logistic) κατανομή έχει παρόμοιο σχήμα με την κανονική κατανομή, αλλά διαθέτει πιο βαριές ουρές, και αυτό είναι το κύριο χαρακτηριστικό της που την κάνει ιδανικότερη κατανομή στην ερμηνεία δεδομένων κινδύνου, καθώς το ζητούμενο είναι να υπάρχει αργή απόσβεση. Επιπλέον, η λογιστική κατανομή είναι εύκολο να προκύψει από διάφορες άλλες κατανομές όπως η ομοιόμορφη ή η εκθετική, με αποτέλεσμα να είναι πολύ ευέλικτη.

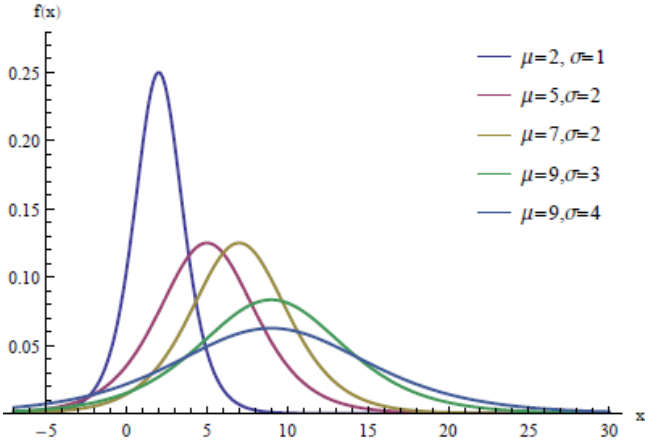
Η συνάρτηση πυκνότητας πιθανότητας και η συνάρτηση κατανομής της λογιστικής κατανομής με παραμέτρους μ και σ, θα συμβολίζουμε $X \sim Log(\mu, \sigma)$, είναι αντίστοιχα

$$f(x) = \frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}, \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0, \tag{3.5.1}$$

και

$$F(x) = \frac{e^{\frac{x-\mu}{\sigma}}}{1+e^{\frac{x-\mu}{\sigma}}} = \frac{1}{1+e^{-\frac{x-\mu}{\sigma}}}, \tag{3.5.2}$$

όπου το μ είναι παράμετρος θέσης και το σ είναι παράμετρος κλίμακας. Στο Σχήμα 3.7 απεικονίζεται η συνάρτηση πυκνότητας της $Log(\mu, \sigma)$ για διάφορες τιμές των παραμέτρων της.



Σχήμα 3.7 Συνάρτηση πυκνότητας της $Log(\mu, \sigma)$

Κάποιες από τις ιδιότητες της λογιστικής κατανομής είναι:

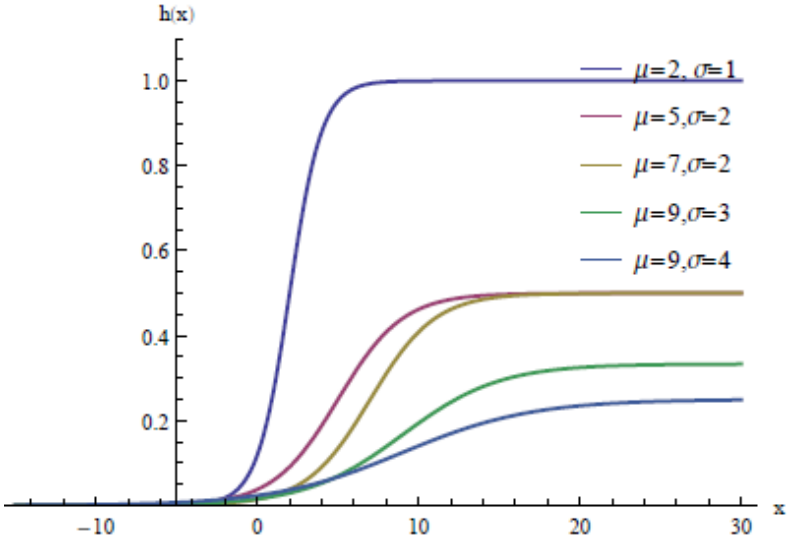
- Δεν έχει παράμετρο σχήματος. Αυτό σημαίνει ότι η γραφική παράσταση της συνάρτησης πυκνότητας της κατανομής, η οποία έχει κωδωνοειδή μορφή, δεν αλλάζει.
- Είναι συμμετρική γύρω από το μ αφού $f(\mu - x) = f(x + \mu)$ για κάθε x.
- Καθώς η παράμετρος μ αυξάνεται (μειώνεται), η συνάρτηση πυκνότητας πιθανότητας μετατοπίζεται προς τα δεξιά (αριστερά).
- Καθώς η παράμετρος σ αυξάνεται, η συνάρτηση πυκνότητας πιθανότητας απλώνεται περισσότερο γύρω από το μ, δηλαδή, γίνεται φαρδύτερη και πιο χαμηλή.
- Καθώς η παράμετρος σ μειώνεται, η σ.π.π. γίνεται πιο στενή, ψηλότερη και μαζεύεται πιο κοντά στο μ.
- Η μέγιστη τιμή της πετυχαίνεται στο $x = \mu$ και είναι ίση με $1/4\sigma$.
- Τα σημεία καμπής της σ.π.π. είναι εκεί που μηδενίζεται η δεύτερη παράγωγος της f και είναι τα $x = \mu + \sigma \ln(2 \pm \sqrt{3})$.
- Μοιάζει με την κανονική κατανομή $N(\mu, \sigma^2)$ αλλά διαφέρουν στις ουρές. Συγκεκριμένα η λογιστική έχει πιο παχιές ουρές, ένδειξη λεπτόκυρτης κατανομής.

3.5.1 Χαρακτηριστικά μεγέθη της Λογιστικής κατανομής

Η συνάρτηση διακινδύνευσης δίνεται από τη σχέση

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{1}{\sigma\left(1 + e^{-\frac{x-\mu}{\sigma}}\right)}. \tag{3.5.3}$$

Παρατηρούμε ότι καθώς $x \to \pm\infty$ είναι $0 \leq h(x) \leq \frac{1}{\sigma}$. Η συνάρτηση κινδύνου γραφικά δίνεται στο Σχήμα 3.8.



Σχήμα 3.8 Συνάρτηση κινδύνου της $Log(\mu, \sigma)$

Οι ροπές k τάξης της $Log(\mu, \sigma)$ κατανομής περί την αρχή και περί τον μέσο αντίστοιχα είναι

$$E(X^k) = E[(\mu + \sigma Y)^k] \tag{3.5.4}$$

και

$$E[(X - \mu)^k] = \sigma^k E(Y^k) \tag{3.5.5}$$

με Y να ακολουθεί την τυπική λογιστική κατανομή $Log(0,1)$ η οποία έχει ροπές

$$E(Y^k) = \begin{cases} 0, & k = 1,3,5, \dots \\ (2^k - 2)\pi^k |B_k|, & k = 2,4,6, \dots \end{cases} \tag{3.5.6}$$

όπου B_k είναι ο k-αριθμός Bernoulli.

Η μέση τιμή της $Log(\mu, \sigma)$ είναι

$$E(X) = \mu, \tag{3.5.7}$$

αφού η κατανομή είναι συμμετρική γύρω από την παράμετρο θέσης μ.

Για τη διασπορά της κατανομής, χρησιμοποιώντας τον τύπο των ροπών (3.5.5) για $k = 2$, βρίσκουμε

$$V(X) = \frac{\sigma^2 \pi^2}{3}. \tag{3.5.8}$$

Επίσης η λοξότητα και τη κύρτωση της κατανομής είναι σταθερές τιμές, ανεξάρτητες των παραμέτρων της και συγεκριμένα

$$skew(X) = 0 \tag{3.5.9}$$

$$kurt(X) = \frac{21}{5} \tag{3.5.10}$$

Η ροπογεννήτρια συνάρτηση της κατανομής μπορεί να δειχθεί ότι είναι

$$M_X(t) = e^{t\mu}B(\sigma t + 1, 1 - \sigma t) = e^{t\mu}\Gamma(\sigma t + 1)\Gamma(1 - \sigma t), \qquad t \in (-1,1) \tag{3.5.11}$$

όπου $B(\cdot,\cdot)$ και $\Gamma(\cdot)$ η συνάρτηση Βήτα και Γάμμα αντίστοιχα.

Η περιορισμένη μέση τιμή για την $Log(\mu, \sigma)$ είναι

$$E(X; x) = x + \sigma\left[-\ln\left(1 + e^{\frac{x-\mu}{\sigma}}\right) + \ln\left(1 + e^{\frac{-\mu}{\sigma}}\right)\right]. \tag{3.5.12}$$

Η ποσοστιαία συνάρτηση δίνεται από τον τύπο

$$Q(p) = \mu + \sigma\ln\left(\frac{p}{1-p}\right), \;\; p \in (0,1). \tag{3.5.13}$$

Ενώ για το 1ο, 2ο, και 3ο τεταρτημόριο θέτοντας στον παραπάνω τύπο $p = 0.25$, $p = 0.5$ και $p = 0.75$ αντίστοιχα προκύπτουν $Q_1 = \mu - \ln 3$, $Q_2 = \mu$ και $Q_3 = \mu + \sigma\ln 3$.

3.5.2 Εκτίμηση παραμέτρων

Έστω ένα δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από τη λογιστική κατανομή $Log(\mu, \sigma)$.

Αρχικά θα ασχοληθούμε με τις εκτιμήτριες συναρτήσεις μέγιστης πιθανοφάνειας των παραμέτρων μ και σ. Η συνάρτηση πιθανοφάνειας του δείγματος, είναι

$$L(\mu, \sigma\,|\mathbf{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{e^{\frac{x_i-\mu}{\sigma}}}{\sigma\left(1 + e^{\frac{x_i-\mu}{\sigma}}\right)^2} = \exp\left(\frac{1}{\sigma}\sum_{i=1}^{n}(x_i - \mu)\right)\frac{1}{\sigma^n\left[\prod_{i=1}^{n}\left(1 + e^{\frac{x_i-\mu}{\sigma}}\right)\right]^2}$$

και λογαριθμίζοντας προκύπτει,

$$l(\mu, \sigma) = \log L(\mu, \sigma\,|\mathbf{x}) = \frac{1}{\sigma}\sum_{i=1}^{n}(x_i - \mu) - n\log\sigma - 2\sum_{i=1}^{n}\log\left(1 + e^{\frac{x_i-\mu}{\sigma}}\right).$$

Παραγωγίζοντας ως προς μ και σ και θέτοντας τις παραγώγους ίσες με το μηδέν, παίρνουμε αντίστοιχα

$$\frac{\partial}{\partial\mu}l(\mu, \sigma) = 0 \Rightarrow \frac{1}{\sigma}\sum_{i=1}^{n}(-1) - 2\sum_{i=1}^{n}\frac{e^{\frac{x_i-\mu}{\sigma}}}{1 + e^{\frac{x_i-\mu}{\sigma}}}\left(-\frac{1}{\sigma}\right) = 0$$

$$\Rightarrow -n + 2\sum_{i=1}^{n}\frac{1}{e^{-\frac{x_i-\mu}{\sigma}} + 1} = 0 \tag{3.5.14}$$

και

$$\frac{\partial}{\partial\sigma}l(\mu,\sigma)=0\Rightarrow-\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)-\frac{n}{\sigma}-2\sum_{i=1}^{n}\frac{e^{\frac{x_i-\mu}{\sigma}}}{1+e^{\frac{x_i-\mu}{\sigma}}}\left(-\frac{x_i-\mu}{\sigma^2}\right)=0$$

$$\Rightarrow\sum_{i=1}^{n}x_i-n\mu+n\sigma-2\sum_{i=1}^{n}\frac{x_i-\mu}{e^{-\frac{x_i-\mu}{\sigma}}+1}=0\,. \tag{3.5.15}$$

Το σύστημα των (3.5.14) και (3.5.15) που προκύπτει είναι μη γραμμικό ως προς μ και σ και απαιτούνται αριθμητικές μέθοδοι για την επίλυσή του.

Η εκτίμηση των παραμέτρων μ και σ μπορεί να πραγματοποιηθεί επίσης με τη μέθοδο των ροπών. Η μέθοδος αυτή για τη συγκεκριμένη κατανομή είναι αρκετά πιο εύκολη εφ' όσον επιλυθεί το παρακάτω σύστημα

$$\begin{cases}E(X)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i=m_1\\ E(X^2)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i^2=m_2\end{cases}\Rightarrow\begin{cases}\mu=m_1\\ \dfrac{\sigma^2\pi^2}{3}+\mu^2=m_2\end{cases}\Rightarrow\begin{cases}\mu=m_1\\ \sigma=\sqrt{\dfrac{3(m_2-m_1^2)}{\pi^2}}\end{cases}.$$

Με αντικατάσταση των m_1 και m_2 καταλήγουμε ότι οι εκτιμήτριες συναρτήσεις με τη μέθοδο ροπών των μ και σ είναι αντίστοιχα

$$\mu=\bar{X}\qquad\text{και}\qquad\sigma=\frac{1}{\pi}\sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\right)}\,. \tag{3.5.16}$$

Τέλος, οι εκτιμήτριες συναρτήσεις των μ και σ μπορούν να προκύψουν και με τη μέθοδο των ποσοστιαίων σημείων. Προς τούτο θεωρούμε δύο σημεία, $Q(\alpha)$ και $Q(\beta)$ που αντιστοιχούν στο α-στό και β-οστό ποσοστημόρια αντίστοιχα με $0<a,\beta<1$. Στη συνέχεια θεωρούμε το σύστημα εξισώσεων,

$$\begin{cases}Q(a)=\sigma\ln\left(\dfrac{1-a}{a}\right)+\mu\\ Q(\beta)=\sigma\ln\left(\dfrac{1-\beta}{\beta}\right)+\mu\end{cases} \tag{3.5.17}$$

με αφαίρεση κατά μέλη προκύπτει

$$Q(\alpha)-Q(\beta)=\sigma\left[\ln\left(\frac{1-a}{a}\right)-\ln\left(\frac{1-b}{b}\right)\right]$$

ή

$$\sigma=\frac{Q(\alpha)-Q(\beta)}{\ln\left(\dfrac{b(1-a)}{a(1-b)}\right)} \tag{3.5.18}$$

και για το μ αντικαθιστούμε σε μία από τις (3.5.17), έστω στην πρώτη και προκύπτει

$$\mu = Q(\alpha) - [Q(\alpha) - Q(\beta)]\frac{\ln\left(\frac{1-\alpha}{\alpha}\right)}{\ln\left(\frac{\beta(1-\alpha)}{\alpha(1-\beta)}\right)}. \tag{3.5.19}$$

Παρατηρούμε ότι μόνο η μέθοδος μέγιστης πιθανοφάνειας δεν είναι σε θέση να μας δώσει άμεσα ένα αποτέλεσμα για την εκτίμηση των παραμέτρων της λογιστικής κατανομής.

3.5.3 Μετασχηματισμοί και ειδικές περιπτώσεις

Η ειδική περίπτωση της $Log(\mu, \sigma)$ με $\mu = 0$ και $\sigma = 1$ καλείται τυπική λογιστική κατανομή. Για την τ.μ. $X \sim Log(0,1)$ η σ.π.π. και η σ.κ. είναι αντίστοιχα

$$f(x) = \frac{e^x}{(1+e^x)^2}, \qquad -\infty < x < \infty \tag{3.5.20}$$

και

$$F(x) = \frac{e^x}{1+e^x} = \frac{1}{e^{-x}+1}, \qquad -\infty < x < \infty. \tag{3.5.21}$$

Στη συνέχεια αναφέρουμε κάποιους μετασχηματισμούς που σχετίζονται με την τυπική λογιστική κατανομή.

Αν η τ.μ. X ακολουθεί την τυπική λογιστική κατανομή $Log(0,1)$, τότε ο γραμμικός μετασχηματισμός $Y = \sigma X + \mu$, $\mu \in \mathbf{R}, \sigma > 0$ οδηγεί στη λογιστική κατανομή δύο παραμέτρων $Log(\mu, \sigma)$.

Αν η τ.μ. X ακολουθεί την τυπική λογιστική κατανομή, τότε με το μετασχηματισμό $\ln(e^X + 1)$ προκύπτει η εκθετική κατανομή $Exp(1)$. Αντίστροφα, αν η τ.μ. $X \sim Exp(1)$, τότε η τ.μ. $Y = \ln(e^X - 1)$ ακολουθεί την τυπική λογιστική κατανομή.

Η τυπική λογιστική κατανομή σχετίζεται με την Pareto κατανομή όπως αναφέρθηκε στην παράγραφο 3.1.3. Εκεί δόθηκε η σχέση της τυπικής λογιστικής με την Pareto δευτέρου είδους. Εδώ για λόγους ποικιλίας θα αναφέρουμε την Pareto πρώτου είδους. Αν $X \sim Log(0,1)$, τότε ο μετασχηματισμός $Y = e^X + 1$ οδηγεί στην κατανομή Pareto πρώτου είδους $Par^{(I)}(1,1)$. Από την άλλη μεριά, αν X ακολουθεί την $Par^{(I)}(1,1)$, τότε με το μετασχηματισμό $Y = \ln(X - 1)$ προκύπτει η τυπική λογιστική κατανομή $Log(0,1)$.

Για τη λογιστική κατανομή με παραμέτρους αναφέρουμε τους παρακάτω μετασχηματισμούς.

Αν X_1 και X_2 είναι ανεξάρτητες εκθετικά κατανεμημένες τυχαίες μεταβλητές με μέσους λ_1 και λ_2 αντίστοιχα, τότε ο μετασχηματισμός $\ln(X_1/X_2)$ οδηγεί στη λογιστική κατανομή

$Log\left(\ln\left(\frac{\lambda_1}{\lambda_2}\right),1\right)$. Όταν οι X_1 και X_2 είναι επιπλέον και ισόνομες δηλαδή $\lambda_1=\lambda_2$, τότε $\ln\left(\frac{X_1}{X_2}\right)\sim Log(0,1)$.

Αν $X_1\sim Gumbel(2\mu,\sigma)$ και $X_2\sim Gumbel(\mu,\sigma)$ είναι ανεξάρτητες τυχαίες μεταβλητές από την Gumbel κατανομή με σ.π.π. που δίνεται στην (3.3.22), τότε ο μετασχηματισμός X_1-X_2 οδηγεί στη λογιστική κατανομή $Log(\mu,\sigma)$. Στην περίπτωση που οι X_1 και X_2 είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από την τυπική Gumbel κατανομή $Gumbel(0,1)$ τότε ο μετασχηματισμός $X_1-X_2\sim Log(0,1)$.

Εάν η τυχαία μεταβλητή $X\sim Log(\mu,\sigma)$ τότε για το μετασχηματισμό $Y=\alpha X+b, a>0, b\in$ **R**, έχουμε

$$F_Y(y)=P(Y\le y)=P(aX+b\le y)=P\left(X\le\frac{y-b}{a}\right)=F_X\left(\frac{y-b}{a}\right)$$

και

$$f_Y(y)=\frac{d}{dy}F_Y(y)=\frac{1}{a}f_X\left(\frac{y-b}{a}\right)$$

$$=\frac{1}{a}\frac{\exp\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)}{\sigma\left[1+\exp\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)\right]^2}=\frac{\exp\left(\frac{y-b-a\mu}{a\sigma}\right)}{a\sigma\left[1+\exp\left(\frac{y-b-a\mu}{a\sigma}\right)\right]^2}, y\in\mathbf{R}.$$

Από τα παραπάνω προκύπτει ότι ο γραμμικός συνδυασμός $Y=\alpha X+b\sim Log(b+a\mu,a\sigma)$.

Επίσης, αν $X\sim Loglogistic(\gamma,\lambda)$ όπου το γ είναι παράμετρος μορφής και το λ παράμετρος κλίμακας, με σ.π.π.

$$f_X(x)=\frac{\frac{\gamma}{\lambda}\left(\frac{x}{\lambda}\right)^{\gamma-1}}{\left[1+\left(\frac{x}{\lambda}\right)^{\gamma}\right]^2}, x>0,\gamma>0,\lambda>0,$$

ο μετασχηματισμός $Y=lnX$ οδηγεί στη λογιστική κατανομή $Log(ln\lambda,1/\gamma)$. Αντίστροφα, αν $X\sim Log(\mu,\sigma)$, η τυχαία μεταβλητή Yμέσω του μετασχηματισμού $Y=e^X$ ακολουθεί την $Loglogistic(1/\sigma,e^{\mu})$.

Τέλος, αν $U\sim Ομοιόμορφη(0,1)$ τότε με το μετασχηματισμό $X=\mu-\sigma\ln\left(\frac{1}{U}-1\right), \sigma>0,\mu\in$ **R**, οδηγούμαστε στη λογιστική κατανομή $Log(\mu,\sigma)$. Αντίστροφα αν $X\sim Log(\mu,\sigma)$ τότε η τ.μ. $U=\frac{1}{1+e^{\frac{\mu-X}{\sigma}}}$ ακολουθεί την $Uniform(0,1)$.

3.6 Γενικευμένη βήτα παραγόμενη Λογιστική κατανομή

Η γενικευμένη βήτα παραγόμενη Λογιστική κατανομή προκύπτει θεωρώντας σαν κατανομή γεννήτορα την $Log(\mu,\sigma)$. Για τη συνάρτηση πυκνότητας πιθανότητάς της, $f_{GBGLog}(x;\mu,\sigma,a,b,c)$, αντικαθιστούμε στην (2.4.3) τις (3.5.1) και (3.5.2) και βρίσκουμε

$$f_{GBGLog}(x;\mu,\sigma,a,b,c)=\frac{c}{B(a,b)}\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{ac-1}\left[1-\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{c}\right]^{b-1}$$

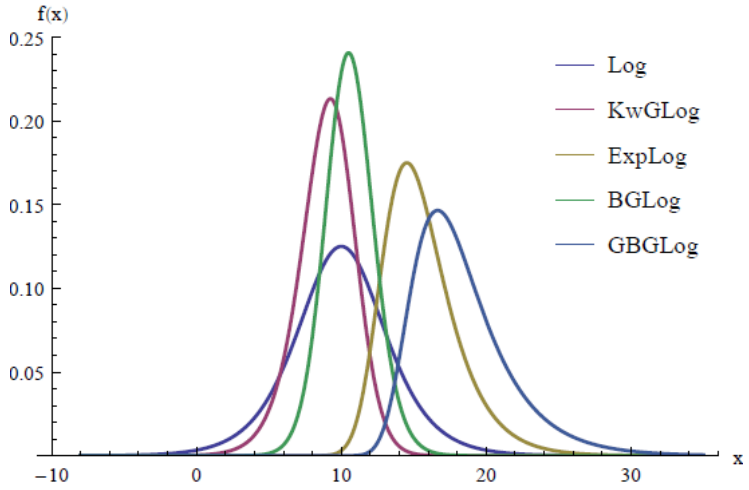
$$=\frac{c}{\sigma B(a,b)}e^{-\frac{x-\mu}{\sigma}}\left(e^{-\frac{x-\mu}{\sigma}}+1\right)^{-ac-1}\left[1-\left(e^{-\frac{x-\mu}{\sigma}}+1\right)^{-c}\right]^{b-1},$$

$$-\infty<x<\infty,-\infty<\mu<\infty,\sigma>0,a>0,b>0,c>0. \qquad (3.6.1)$$

Η συνάρτηση κατανομής της $GBGLog$ από την (2.4.16) είναι

$$F_{GBGLog}(x;\mu,\sigma,a,b,c)=\frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(b)}{(a+j)j!\,\Gamma(b-j)}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{c(a+j)}. \qquad (3.6.2)$$

Στο Σχήμα 3.9 απεικονίζεται η συνάρτηση πυκνότητας της $GBGLog(\mu,\sigma)$ για διάφορες τιμές των παραμέτρων της. Συγκεκριμένα, είναι $Log-GBGLog(10,2,1,1,1)$, $KwGLog-GBGLog(10,2,1,3.5,1.8)$, $ExpLog-GBGLog(10,2,4,1,2.4)$, $BGLog-GBGLog(10,2,3.6,2.8,1)$, $GBGLog(10,2,4,0.7,5)$.



Σχήμα 3.9 Συνάρτηση πυκνότητας της $GBGLog(\mu,\sigma,a,b,c)$

Για να δώσουμε εκφράσεις για τις ροπές της $GBGLog$ θα βρούμε αρχικά τη ροπογεννήτρια συνάρτηση της εκθετικοποιημένης λογιστικής κατανομής. Έστω $Y \sim Exp^{(a)}Log(\mu,\sigma)$, τότε από την (2.4.14) είναι

$$f_Y(y) = a\frac{e^{\frac{y-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{y-\mu}{\sigma}}\right)^2}\left(\frac{e^{\frac{y-\mu}{\sigma}}}{1+e^{\frac{y-\mu}{\sigma}}}\right)^{a-1} = \frac{ae^{-\frac{y-\mu}{\sigma}}}{\sigma\left(1+e^{-\frac{y-\mu}{\sigma}}\right)^{a+1}}, \quad -\infty < y < \infty. \tag{3.6.3}$$

Για τη ροπογεννήτρια συνάρτηση της Y έχουμε

$$M_Y(t) = E(e^{tY}) = \int_y e^{ty}f_Y(y)dy = \int_{-\infty}^{\infty} e^{ty}a\frac{e^{\frac{y-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{y-\mu}{\sigma}}\right)^2}\left(\frac{e^{\frac{y-\mu}{\sigma}}}{1+e^{\frac{y-\mu}{\sigma}}}\right)^{a-1} dy$$

θέτοντας $z = \frac{e^{\frac{y-\mu}{\sigma}}}{1+e^{\frac{y-\mu}{\sigma}}}$ είναι $dz = \frac{e^{\frac{y-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{y-\mu}{\sigma}}\right)^2}dy$ και $e^{\frac{y-\mu}{\sigma}} = \frac{z}{1-z} \Rightarrow e^y = \left(\frac{z}{1-z}\right)^{\sigma}e^{\mu}$ καθώς και για $-\infty < y < \infty$ είναι $0 < z < 1$ οπότε το ολοκλήρωμα γίνεται

$$M_Y(t) = \int_0^1 \left(\frac{z}{1-z}\right)^{\sigma t} e^{\mu t}az^{a-1}dz = ae^{\mu t}\int_0^1 z^{\sigma t+a-1}(1-z)^{-\sigma t+1-1}dz$$

αλλά το ολοκλήρωμα είναι η συνάρτηση Βήτα, και κάνοντας χρήση της σχέσης μεταξύ των συναρτήσεων Βήτα και Γάμμα, $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, η ροπογεννήτρια συνάρτηση γίνεται

$$M_Y(t) = ae^{\mu t}B(\sigma t + a, 1-\sigma t) = ae^{\mu t}\frac{\Gamma(\sigma t + a)\Gamma(1-\sigma t)}{\Gamma(a+1)}. \tag{3.6.4}$$

Έτσι η ροπογεννήτρια συνάρτηση της $X \sim GBGLog$ από την (2.4.40) με $Y_j \sim Exp^{[c(a+j)]}F$ είναι

$$M_X(t) = \sum_{j=0}^{\infty} w_j M_j(t) = \sum_{j=0}^{\infty} w_j c(a+j)e^{\mu t}B(\sigma t + c(a+j), 1-\sigma t) \tag{3.6.5}$$

και για τις ροπές της $X \sim GBGLog$ από τη Σχέση (2.4.32) ισχύει

$$\begin{aligned} E(X^s) &= \sum_{j=0}^{\infty} w_j E\left(Y_j^s\right) = \sum_{j=0}^{\infty} w_j \frac{d^s}{dt^s}M_{Y_j}(t)\Big|_{t=0} \\ &= \sum_{j=0}^{\infty} w_j \frac{d^s}{dt^s}c(a+j)e^{\mu t}B(\sigma t + c(a+j), 1-\sigma t)\Big|_{t=0} \end{aligned} \tag{3.6.6}$$

όπου από την (2.4.16) τα βάρη $w_j = \frac{(-1)^j\Gamma(b+1)}{B(a,b+1)(a+j)j!\Gamma(b-j)}$.

Για τη συνάρτηση πυκνότητας της $X_{i:n}$ από την (2.4.54) αντικαθιστώντας τις (3.5.1) και (3.5.2) βρίσκουμε

$$f_{i:n}(x) = \sum_{r=0}^{n-i}\frac{(-1)^r\binom{n-i}{r}}{B(i,n-i+1)}\frac{c}{B(a,b)}\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{ac-1}\left[1-\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^c\right]^{b-1}$$

$$\times\left[\frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(b)}{(a+j)j!\,\Gamma(b-j)}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{c(a+j)}\right]^{i+r-1}$$

η οποία με χρήση της ταυτότητας

$$\left(\sum_{i=0}^{\infty}a_i\right)^k=\sum_{m_1=0}^{\infty}\dots\sum_{m_k=0}^{\infty}a_{m_1}\dots a_{m_k}$$

γίνεται

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\frac{(-1)^r\binom{n-i}{r}}{B(i,n-i+1)}\frac{c}{B(a,b)}\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{ac-1}\left[1-\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^c\right]^{b-1}$$

$$\times\left[\frac{\Gamma(b)}{B(a,b)}\right]^{i+r-1}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\frac{(-1)^{\sum_{j=1}^{i+r-1}m_j}}{\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{c\sum_{j=1}^{i+r-1}(a+m_j)}$$

ή

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\frac{(-1)^{r+\sum_{j=1}^{i+r-1}m_j}\binom{n-i}{r}[\Gamma(b)]^{i+r-1}B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}{B(i,n-i+1)[B(a,b)]^{i+r}\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}$$

$$\times\frac{c}{B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{c\left[a+\sum_{j=1}^{i+r-1}(a+m_j)\right]-1}$$

$$\times\left[1-\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^c\right]^{b-1}$$

δηλαδή,

$$f_{i:n}(x)=\sum_{r=0}^{n-i}\sum_{m_1=0}^{\infty}\dots\sum_{m_{i+r-1}=0}^{\infty}\delta_{i,r}f_{i,r}(x)\qquad(3.6.7)$$

όπου

$$f_{i,r}(x)=\frac{c}{B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^{c\left[a+\sum_{j=1}^{i+r-1}(a+m_j)\right]-1}$$

$$\times\left[1-\left(\frac{1}{e^{-\frac{x-\mu}{\sigma}}+1}\right)^c\right]^{b-1}$$

και οι συντελεστές $\delta_{i,r}$ είναι

$$\delta_{i,r}=\frac{(-1)^{r+\sum_{j=1}^{i+r-1}m_j}\binom{n-i}{r}[\Gamma(b)]^{i+r-1}B\left(a+\sum_{j=1}^{i+r-1}(a+m_j),b\right)}{B(i,n-i+1)[B(a,b)]^{i+r}\prod_{j=1}^{i+r-1}(a+m_j)m_j!\,\Gamma(b-m_j)}.$$

Παρατηρούμε ότι η $f_{i,r}$ είναι η συνάρτηση πυκνότητας μιας κατανομής

$GBGLog\left(\mu, \sigma, a + \sum_{j=1}^{i+r-1}\left(a + m_j\right), b, c\right)$. Το αποτέλεσμα της (3.6.7) εκφράζει την πυκνότητα της διατεταγμένης στατιστικής $GBGLog$ σαν μείξη $GBGLog$ με συνέπεια διάφορες ιδιότητες της $X_{i:n}$ να μπορούν να δοθούν από αυτή τη διατύπωση της μείξης. Για παράδειγμα, οι s ροπές της $X_{i:n}$, για $b > 0$ πραγματικό, μη ακέραιο, δίνονται από τον τύπο

$$E(X_{i:n}^s) = \sum_{r=0}^{n-i} \sum_{m_1=0}^{\infty} \dots \sum_{m_{i+r-1}=0}^{\infty} \delta_{i,r} E(X_{i:r}^s) \tag{3.6.8}$$

όπου $X_{i:r} \sim GBGLog\left(\mu, \sigma, a + \sum_{j=1}^{i+r-1}\left(a + m_j\right), b, c\right)$.

Στην περίπτωση που το b είναι θετικός ακέραιος, οι δείκτες $m_1, \dots, m_{i+r-1}$ σταματάνε στο $b - 1$.

Για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας, θεωρούμε ένα τυχαίο δείγμα $\mathbf{x} = (x_1, \dots, x_n)$ μεγέθους n από την κατανομή *GBGLog* και έστω $\boldsymbol{\theta} = (\mu, \sigma, a, b, c)$ το ιάνυσμα των παραμέτρων της. Η συνάρτηση πιθανοφάνειας είναι

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}) &= \prod_{i=1}^{n} f_{GBGLog}(x_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \frac{c}{\sigma B(a,b)} e^{-\frac{x_i-\mu}{\sigma}} \left(e^{-\frac{x_i-\mu}{\sigma}} + 1\right)^{-ac-1} \left[1 - \left(e^{-\frac{x_i-\mu}{\sigma}} + 1\right)^{-c}\right]^{b-1} \\ &= \frac{c^n}{\sigma^n [B(a,b)]^n} \exp\left[-\frac{1}{\sigma}\sum_{i=1}^{n}(x_i - \mu)\right] \left[\prod_{i=1}^{n}\left(e^{-\frac{x_i-\mu}{\sigma}} + 1\right)\right]^{-ac-1} \left\{\prod_{i=1}^{n}\left[1 - \left(e^{-\frac{x_i-\mu}{\sigma}} + 1\right)^{-c}\right]\right\}^{b-1}. \end{aligned}$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι

$$\begin{aligned} l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}|\mathbf{x}) &= n\log c - n\log\sigma - n\log B(a,b) - \frac{1}{\sigma}\sum_{i=1}^{n}(x_i - \mu) \\ &\quad -(ac+1)\sum_{i=1}^{n}\log\left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right) + (b-1)\sum_{i=1}^{n}\log\left[1 - \left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right)^{-c}\right]. \end{aligned}$$

Για τη μεγιστοποίηση της συνάρτησης $l(\boldsymbol{\theta})$ λύνουμε το σύστημα των μη γραμμικών εξισώσεων που προκύπτουν αν θέσουμε τις παραγώγους της $l(\boldsymbol{\theta})$ ως προς τις παραμέτρους $\boldsymbol{\theta}$ ίσες με το μηδέν. Συγκεκριμένα για τις πρώτες μερικές παραγώγους έχουμε

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a} = -n[\psi(a) - \psi(a+b)] - c\sum_{i=1}^{n}\log\left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right),$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial b} = -n[\psi(b) - \psi(a+b)] + \sum_{i=1}^{n}\log\left[1 - \left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right)^{-c}\right],$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial c} = \frac{n}{c} - a\sum_{i=1}^{n}\log\left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right) + (b-1)\sum_{i=1}^{n}\frac{\log\left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right)}{\left(1 + e^{-\frac{x_i-\mu}{\sigma}}\right)^{c} - 1},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu} = \frac{n}{\sigma} - \frac{ac+1}{\sigma}\sum_{i=1}^{n}\frac{1}{e^{\frac{x_i-\mu}{\sigma}}+1} + \frac{c(b-1)}{\sigma}\sum_{i=1}^{n}\frac{\left(1+e^{-\frac{x_i-\mu}{\sigma}}\right)^{-c-1}e^{-\frac{x_i-\mu}{\sigma}}}{1-\left(1+e^{-\frac{x_i-\mu}{\sigma}}\right)^{-c}},$$

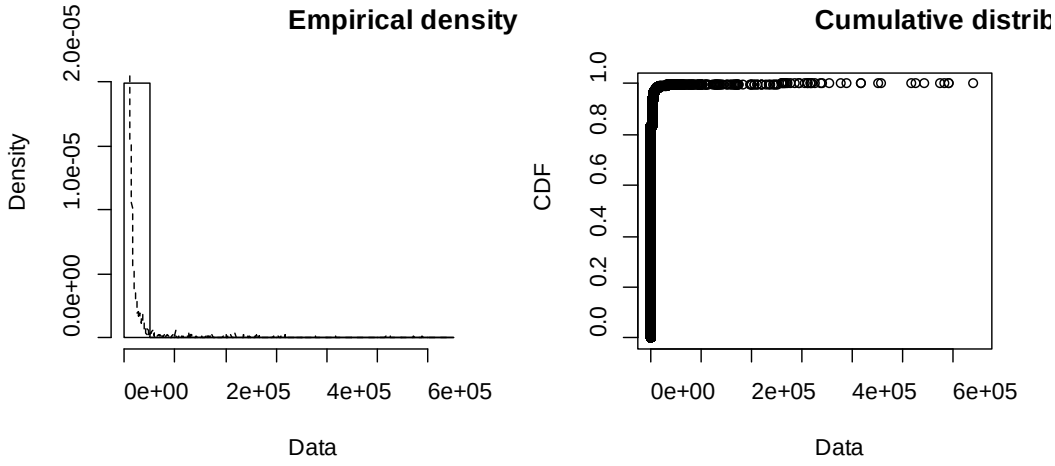
$$\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) - \frac{ac+1}{\sigma^2}\sum_{i=1}^{n}\frac{x_i-\mu}{e^{\frac{x_i-\mu}{\sigma}}+1}$$
$$+\frac{c(b-1)}{\sigma^2}\sum_{i=1}^{n}\frac{\left(1+e^{-\frac{x_i-\mu}{\sigma}}\right)^{-c-1}(x_i-\mu)e^{-\frac{x_i-\mu}{\sigma}}}{1-\left(1+e^{-\frac{x_i-\mu}{\sigma}}\right)^{-c}}.$$

ΚΕΦΑΛΑΙΟ 4

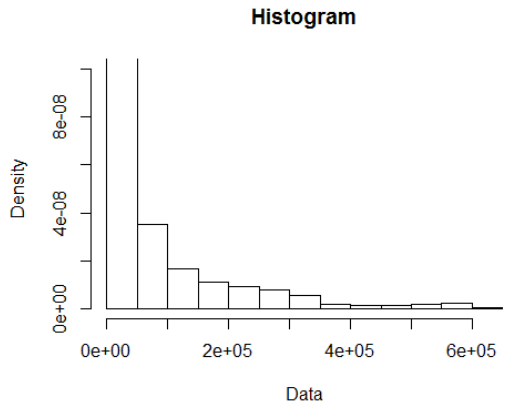
ΑΡΙΘΜΗΤΙΚΗ ΕΦΑΡΜΟΓΗ

Στο κεφάλαιο αυτό θα μελετήσουμε την ευελιξία προσαρμογής των διάφορων κατανομών που αναφέρθηκαν στο τρίτο κεφάλαιο, σε κάποιο σύνολο δεδομένων.

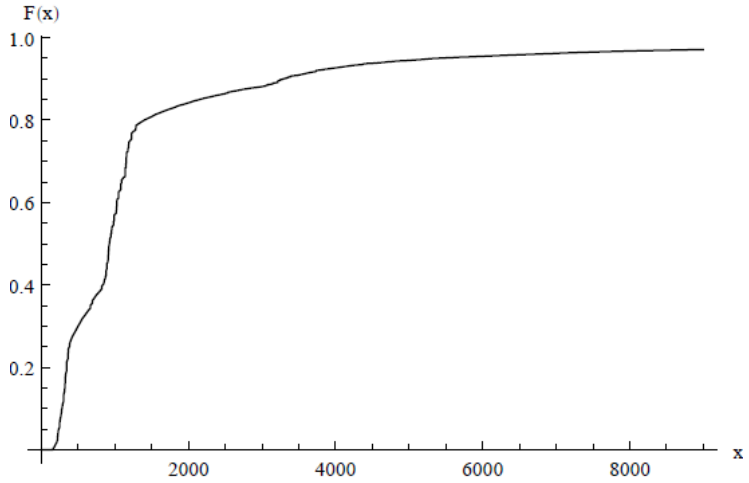
Τα δεδομένα που εξετάζουμε αφορούν μια ασφαλιστική εταιρεία του ελληνικού χώρου και αναφέρονται σε ζημιές περιουσίας σε ημερήσιο επίπεδο χρονικής διάρκειας τριών ετών. Συγκεκριμένα, πρόκειται για το χρονικό διάστημα 5/1/2006 έως 31/12/2008 διάρκειας 1092 ημερολογιακών ημερών με 734 εργάσιμες ημέρες. Το ύψος των ζημιών περιγράφεται από κάποια συνεχή κατανομή την οποία προσπαθούμε να προσεγγίσουμε μέσω των θεωρητικών κατανομών που μελετάμε.



Σχήμα 4.1 Ιστόγραμμα και CDF της εμπειρικής κατανομής



Σχήμα 3.2 Ιστόγραμμα σε μεγέθυνση



Σχήμα 4.3 Εμπειρική συνάρτηση κατανομής για μεγέθη ζημιών έως 9000

Το γράφημα της εμπειρικής συνάρτησης κατανομής και το ιστόγραμμα για το σύνολο των δεδομένων δίνονται στο Σχήμα 4.1 ενώ αντίστοιχες μεγεθύνσεις για το ιστόγραμμα στο Σχήμα 4.2 και για τη συνάρτηση κατανομής περιορισμένη σε μεγέθη ζημιών ως 9000, στο Σχήμα 4.3.

Από τα σχήματα βλέπουμε ότι υπάρχουν κίνδυνοι με χαμηλή συχνότητα αλλά μεγάλο κόστος καθώς και κίνδυνοι με υψηλή συχνότητα αλλά με χαμηλό κόστος ζημιών.

Στον Πίνακα 4.1 αναφέρουμε κάποια περιγραφικά στοιχεία των δεδομένων.

Περιγραφικά	**Τιμή**
Παρατηρήσεις	32494
Μέσος	2404.19
Διάμεσος	924.61
Τυπική απόκλιση	15491.42
Διακύμανση	239984168
Λοξότητα	23.35
Λοξότητα Bowley	– 0.3195
Κύρτωση	677.77
Κύρτωση Moors	1.525
Ελάχιστη τιμή	150.2
Μέγιστη τιμή	639846
Επικρατούσα τιμή (mode)	1139.31
Q1	379.33
Q2	924.61
Q3	1205.83

Πίνακας 4.1 Περιγραφικά μέτρα

Από τα παραπάνω περιγραφικά μέτρα βλέπουμε ότι η διακύμανση των δεδομένων είναι πολύ μεγάλη δηλαδή υπάρχουν μεγάλες αποκλίσεις από το μέσο καθώς και ανάμεσα στα μεγέθη των ζημιών κάτι που φαίνεται και από την ελάχιστη και μέγιστη τιμή του δείγματος. Η θετική τιμή του συντελεστή ασυμμετρίας (λοξότητα) δηλώνει ότι έχουμε δεδομένα με βαριά δεξιά ουρά. Η αρνητική τιμή του συντελεστή ασυμμετρίας Bowley ή τεταρτημοριακού συντελεστή, δείχνει ότι η διάμεσος είναι πιο κοντά στο τεταρτημόριο Q_3 από ότι στο τεταρτημόριο Q_1.

4.1 Προσαρμογή των δεδομένων

Στη συνέχεια θα εξετάσουμε τα διάφορα μοντέλα κατανομών που προκύπτουν για τους γεννήτορες που αναλύθηκαν στο κεφάλαιο 3. Θα βρούμε τους εκτιμητές μέγιστης πιθανοφάνειας των επιμέρους παραμέτρων και θα υπολογίσουμε διάφορα στατιστικά κριτήρια για την αξιολόγηση των διαφορετικών κατανομών ως προς την προσαρμογή τους στα δεδομένα. Το στατιστικό πακέτο που χρησιμοποιήθηκε ήταν το R. Λόγω των διαθέσιμων ρουτινών του R, αναλύουμε ξεχωριστά την περίπτωση των Pareto και Weibull με το πακέτο Newdistns και τη λογιστική κατανομή με το πακέτο mle2. Στο σημείο αυτό να τονίσουμε ότι οι μέθοδοι εκτίμησης των παραμέτρων είναι αριθμητικές και για τη σύγκλιση των αλγορίθμων απαιτούνται αρχικές τιμές εκκίνησης, δηλαδή η μεθοδολογία είναι δοκιμής και σφάλματος (trial and error).

Ξεκινώντας, θεωρούμε ως γεννήτορα κατανομή F την Weibull $W(\lambda, \gamma)$ και την Pareto $Par^{(II)}(\lambda, \delta)$ και εξετάζουμε τα μοντέλα γενικευμένης Βήτα παραγόμενης κατανομής $GBGF$, Βήτα παραγόμενης κατανομής BGF, Kumaraswamy παραγόμενης κατανομής $KwGF$ και του γεννήτορα F. Για κάθε μοντέλο, εκτιμούμε τις παραμέτρους με τη μέθοδο μέγιστης πιθανοφάνειας και για την αξιολόγησή του δίνουμε τις τιμές των επόμενων πληροφοριακών στατιστικών κριτηρίων:

Akaike Information Criterion: $AIC = -2logL + 2k$

Bayesian Information Criterion: $BIC = -2logL + klogN$

Consistent Akaike Information Criterion: $CAIC = -2logL + k(logN + 1)$

όπου, N ο αριθμός των παρατηρήσεων, k ο αριθμός των παραμέτρων στο υπό εκτίμηση μοντέλο, L η πιθανοφάνεια των δεδομένων στο εκτιμημένο μοντέλο.

Τα εν λόγω κριτήρια χρησιμοποιούνται για την επιλογή του καταλληλότερου μοντέλου (model selection), δηλαδή, υιοθετείται το μοντέλο εκείνο του οποίου η τιμή των κριτηρίων είναι η μικρότερη από όλες τις άλλες τιμές των εναλλακτικών μοντέλων. Η επιλογή της μικρότερης

τιμής για κάθε ένα από τα κριτήρια βασίζεται στο γεγονός ότι αυτά επιβάλλουν κάποια ποινή ως προς το πλήθος των παραμέτρων σε κάθε εξεταζόμενο μοντέλο-κατανομή, έτσι είναι λογικό να επιλέγεται στην εκάστοτε περίπτωση η κατανομή με τη μικρότερη τιμή.

Τα αποτελέσματα στην περίπτωση της Pareto δίνονται στον Πίνακα 4.2 και στην περίπτωση της Weibull στον Πίνακα 4.3. Σε παρένθεση δίνονται τα τυπικά σφάλματα των εκτιμήσεων.

Το στατιστικό πακέτο R μέσω του Newdistns, υλοποιεί την κατανομή $Par^{(II)}(\delta, \lambda)$ με την ονομασία Lomax, και με συνάρτηση πυκνότητας πιθανότητας της μορφής

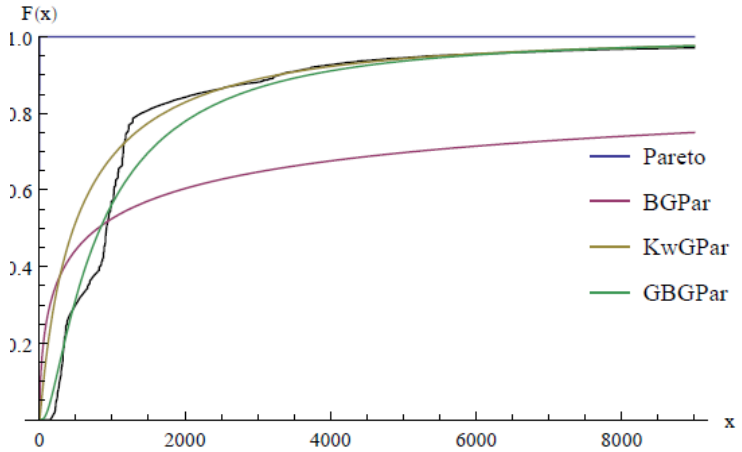
$$f(x) = \frac{s\left(\frac{1}{r}\right)^{s}}{\left(\frac{1}{r} + x\right)^{s}},$$

δηλαδή, η σχέση των παραμέτρων μεταξύ των δύο πυκνοτήτων είναι $\delta = s$ και $\lambda = \frac{1}{r}$.

Κατανομή	r	s	a	b	c	AIC	BIC	$CAIC$
Pareto	2700.2041 (51.645)	2.6129 (0.039)	1	1	1	540938.1	540954.9	540956.9
BGPar	2599.301	0.05502 (0.0005)	19.72 (0.169)	16.13 (0.189)	1	583965.4	583998.9	583965.4
KwGPar	4500.17	0.164 (0.0007)	1	42.369 (0.434)	28.734 (0.3805)	549806.7	549840.3	549806.7
GBGPar	2600.74 (42.041)	0.2274 (0.004)	18.936 (0.5189)	16.262 (0.8966)	16.777 (0.27184)	**532911.4**	**532953.4**	**532911.4**

Πίνακας 4.2 Εμπ και πληροφοριακά κριτήρια όταν ο γεννήτορας είναι η Pareto$^{(II)}$

Από τον παραπάνω πίνακα σύμφωνα με τα πληροφοριακά κριτήρια, βλέπουμε ότι η καλύτερα προσαρμοζόμενη κατανομή στα δεδομένα είναι η γενικευμένη Βήτα παραγόμενη Pareto κατανομή, ενώ η δεύτερη καλύτερη κατανομή είναι ο γεννήτορας.



Σχήμα 4.4 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των προσαρμοζόμενων κατανομών με γεννήτορα την Pareto$^{(II)}$

Για να δώσουμε μια γραφική σύγκριση ανάμεσα στις προσαρμοζόμενες κατανομές, στο Σχήμα 4.4 παρουσιάζονται η εμπειρική συνάρτηση κατανομής και οι θεωρητικές κατανομές των μοντέλων που προκύπτουν με γεννήτορα την Pareto.

Από το Σχήμα 4.4 φαίνεται ότι από τις τέσσερις εξεταζόμενες κατανομές, οι GBGPar και BGPar παρέχουν την πιο ικανοποιητική προσαρμογή, με αυτήν της GBGPar να είναι η καλύτερη στο μεγαλύτερο εύρος των δεδομένων.

Το στατιστικό πακέτο R, υλοποιεί την κατανομή $Weibull(\lambda, \gamma)$ με συνάρτηση πυκνότητας πιθανότητας της μορφής

$$f(x) = \left(\frac{r}{s}\right)\left(\frac{x}{s}\right)^{r-1} \exp\left[-\left(\frac{x}{s}\right)^{r}\right],$$

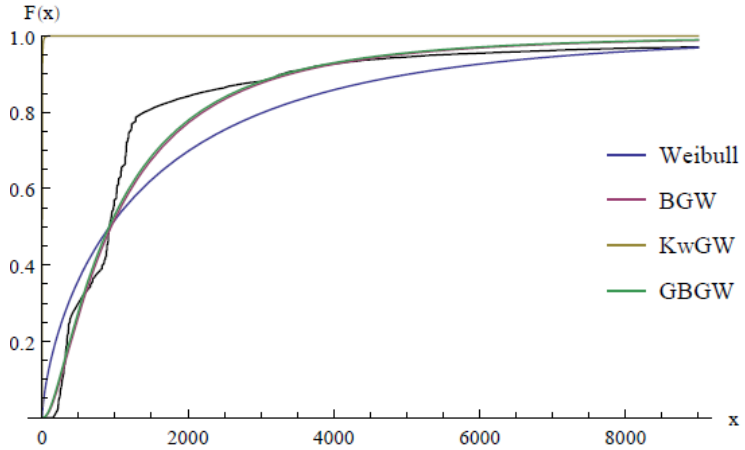
δηλαδή, η σχέση των παραμέτρων μεταξύ των δύο πυκνοτήτων είναι $\lambda = \left(\frac{1}{s}\right)^{r}$ και $\gamma = r$.

Κατανομή	r	s	a	b	c	AIC	BIC	$CAIC$
Weibull	0.70957 (0.0023)	1553.04 (12.905)	1	1	1	554490.1	554506.9	554508.9
BGW	0.3505 (0.0032)	83.2989	12.0544 (0.2312)	1.5602 (0.04639)	1	**535949.2**	**535982.8**	**535949.3**
KwGW	0.41056 (0.0032)	99.062 (0.6483)	1	9.001 (0.0725)	1.053 (0.0253)	536535.1	536568.7	536535.1
GBGW	0.1706 (0.0115)	82.6135 (33.313)	8.889 (0.4532)	11.9226 (0.8496)	3.4238 (0.3357)	536177	536218.9	536177

Πίνακας 4.3 Εμπ και πληροφοριακά κριτήρια όταν ο γεννήτορας είναι η Weibull

Με βάση τα παραπάνω αποτελέσματα η μικρότερη τιμή των πληροφοριακών κριτηρίων επιτυγχάνεται στη Βήτα παραγόμενη Weibull κατανομή, ενώ η δεύτερη καλύτερη κατανομή με μικρή διαφορά είναι η γενικευμένη Βήτα παραγόμενη Weibull.

Στο επόμενο σχήμα φαίνεται η προσαρμογή των υποψήφιων κατανομών με γεννήτορα την Weibull συγκρινόμενη με την εμπειρική συνάρτηση κατανομής.



Σχήμα 4.5 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των προσαρμοζόμενων κατανομών με γεννήτορα την Weibull

Στο Σχήμα 4.5 φαίνεται ότι κατανομές BGW και GBGW παρέχουν την πιο ικανοποιητική προσαρμογή στα δεδομένα με αυτήν της BGW να υπερέχει ελαφρά.

Για τη σύγκριση των στατιστικών κατανομών που εφαρμόστηκαν στα δεδομένα, εξετάστηκαν οι έλεγχοι καλής προσαρμογής των Cramer-von-Mises (CvM) και των Anderson-Darling (AD), (D' Agostino and Stephens, (1986)). Ο έλεγχος δηλαδή στον οποίο αναφέρονται είναι

H_0 : τα δεδομένα προσαρμόζουν την υπό μελέτη κατανομή F

H_1 : τα δεδομένα δεν προσαρμόζουν την υπό μελέτη κατανομή F

Τα εν λόγω τεστ είναι μη παραμετρικοί έλεγχοι οι οποίοι βασίζονται στη διάταξη των δεδομένων και στην εμπειρική συνάρτηση κατανομής. Για ένα δείγμα N παρατηρήσεων, $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$, έστω $\mathbf{x}_\delta = \{x_{(1)}, x_{(2)}, \dots, x_{(N)}\}$ είναι το διατεταγμένο σε αύξουσα τάξη δείγμα.

Το τεστ καλής προσαρμογής των *Cramer-von-Mises* βασίζεται στην απόσταση ανάμεσα στην αθροιστική συνάρτηση κατανομής $\hat{F}_N$ του δείγματος $\mathbf{x}$ και της υποψήφιας κατανομής, F. Η απόσταση δεν είναι η μέγιστη απόκλιση όπως στη περίπτωση του Kolmogorov-Smirnov, δηλαδή

$$D = \sup_x \left|\hat{F}_N(x) - F(x)\right|,$$

αλλά θεωρείται το τετράγωνο της απόστασης μεταξύ της αθροιστικής κατανομής του δείγματος και της θεωρητικής κατανομής και ολοκληρώνει σε όλο το πεδίο της κατανομής, δηλαδή είναι

$$D = \int_{-\infty}^{\infty} [F(x) - \hat{F}_N(x)]^2 \, dF.$$

Για ένα διατεταγμένο δείγμα $\mathbf{x}_\delta$, η απόσταση εκτιμάται από τη σχέση

$$W = \frac{1}{12N} + \sum_{i=1}^{N} \left[\frac{2i-1}{2N} - F(x_{(i)})\right]^2.$$

Το τεστ καλής προσαρμογής των *Anderson-Darling* είναι πιο ευαίσθητο στην ποιότητα της προσαρμογής στις ουρές της κατανομής, από το αντίστοιχο Kolmogorov-Smirnov τεστ που δίνει περισσότερο βάρος στο κέντρο της κατανομής, οπότε είναι περισσότερο κατάλληλο για δεδομένα με ακραίες τιμές όπως στην περίπωση που εξετάζουμε. Το AD τεστ βασίζεται στην απόσταση ανάμεσα στην αθροιστική συνάρτηση κατανομής $\hat{F}_N$ του δείγματος $\mathbf{x}$ και της υποψήφιας κατανομής, F, η οποία είναι τετραγωνικού τύπου, όπως στο Cramer-von-Mises, αλλά δίνει περισσότερο βάρος στις αποκλίσεις των ακραίων τιμών και δίνεται από τη σχέση

$$D = \int_{-\infty}^{\infty} \frac{\left[F(x) - \hat{F}_N(x)\right]^2}{F(x)[1 - F(x)]} \, dF(x).$$

Για ένα διατεταγμένο δείγμα $\mathbf{x}_\delta$, η απόσταση εκτιμάται από τη σχέση

$$A = -N - \sum_{i=1}^{N} \frac{2i-1}{N}\left[lnF\left(x_{(i)}\right) - \ln\left(1 - F\left(x_{(N+1-i)}\right)\right)\right].$$

Και στα δυο τεστ, αν η αντίστοιχη στατιστική A ή W είναι μεγαλύτερη από τη σχετική κριτική τιμή d_a απορρίπτεται η υποψήφια κατανομή σε επίπεδο σημαντικότητας a, ενώ αν A ή W είναι μικρότερη ή ίση από d_a δεχόμαστε την υποψήφια κατανομή σε επίπεδο σημαντικότητας a. Η κριτική τιμή d_a εξαρτάται από την υποψήφια κατανομή F και βρίσκεται σε ειδικούς πίνακες για κάποιες κατανομές. Εναλλακτικά, με χρήση του $pvalue$, η μηδενική υπόθεση απορρίπτεται σε ε.σ. a αν $pvalue < a$.

Επειδή και τα δύο τεστ βασίζονται σε αποστάσεις απόκλισης, όσο μικρότερες είναι οι τιμές των στατιστικών A και W τόσο καλύτερα η αντίστοιχη κατανομή προσαρμόζεται στα δεδομένα.

Στην περίπτωση της Pareto κατανομής τα αποτελέσματα δίνονται στον Πίνακα 4.4, ενώ όταν ο γεννήτορας είναι η Weibull κατανομή, στον Πίνακα 4.5.

Κατανομή	r	s	a	b	c	W	A	$-logL$
Pareto	2700.2041	2.6129	1	1	1	242.32	1476.66	270467.1
BGPar	2599.301	0.05502	19.72	16.13	1	103.279	569.851	291978.1
KwGPar	4500.17	0.164	1	42.369	28.734	124.0659	699.359	274899.4
GBGPar	2600.74	0.2274	18.936	16.262	16.777	**97.8**	**532.825**	266450.7

Πίνακας 4.4 Εμπ και στατιστικές τιμές όταν ο γεννήτορας είναι η Pareto(II)

Με βάση τα κριτήρια Anderson-Darling και Cramer-von Mises προκύπτει ότι η κατανομή που προσαρμόζεται καλύτερα στα δεδομένα είναι πάλι η γενικευμένη Βήτα παραγόμενη Pareto κατανομή, ενώ η δεύτερη καλύτερη κατανομή είναι η Βήτα παραγόμενη Pareto.

Κατανομή	r	s	a	b	c	W	A	$-logL$
Weibull	0.70957	1553.04	1	1	1	414.88	∞	277246
BGW	0.3505	83.2989	12.0544	1.5602	1	**128.6987**	**725.1082**	267970.6
KwGW	0.41056	99.062	1	9.001	1.053	132.0158	743.1873	268263.6
GBGW	0.1706	82.6135	8.889	11.9226	3.4238	133.1134	751.8321	268083.5

Πίνακας 4.5 Εμπ και στατιστικές τιμές όταν ο γεννήτορας είναι η Weibull

Στην περίπτωση που ο γεννήτορας είναι η Weibull, από τα παραπάνω αποτελέσματα, προκύπτει ότι οι τιμές των στατιστικών Anderson-Darling και Cramer-von Mises έχουν μικρότερες τιμές για τη Βήτα παραγόμενη κατανομή, δηλαδή η κατανομή αυτή προσαρμόζεται καλύτερα στα δεδομένα. Η δεύτερη καλύτερα προσαρμοζόμενη κατανομή είναι η Kumaraswamy παραγόμενη Weibull αν και η GBG Weibull πετυχαίνει παραπλήσιες τιμές για τις στατιστικές W και A.

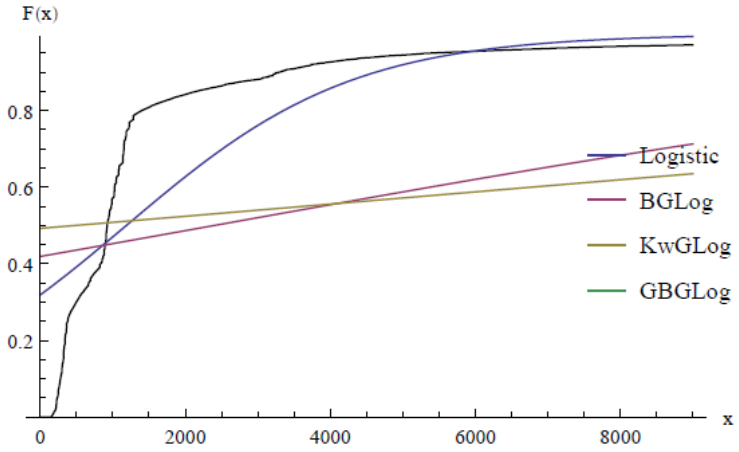
Στη συνέχεια θεωρούμε ως γεννήτορα κατανομή τη λογιστική $Log(\mu, \sigma)$. Χρησιμοποιώντας το πακέτο mle2 της R βρίσκουμε τους εκτιμητές μέγιστης πιθανοφάνειας, τα τυπικά σφάλματα αυτών, την τιμή του πληροφοριακού κριτηρίου AIC και την τιμή του λογαρίθμου της πιθανοφάνειας. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 4.6.

Κατανομή	μ	σ	a	b	c	AIC	$-logL$
Logistic	1187.218 (13.068382)	1563.665 (8.186711)	1	1	1	**624448.3**	312222.1
BGLog	11.40035 (0.00034555)	64993.47776 (0.0013204)	62.57944 (0.4900)	60.33634 (0.47237)	1	701422.4	350707.2
KwGLog	5.5185 (0.000025677)	64995 (0.00038657)	1	39.293 (0.42154)	5.8680 (0.017612)	735316.9	367654.5
GBGLog	10.006 (0.0001947)	65001 (0.0016632)	67.316 (1.709)	58.688 (0.046524)	0.092804 (0.019365)	701308.8	350649.4

Πίνακας 4.6 Εμπ και σχετικά στατιστικά μέτρα όταν ο γεννήτορας είναι η Λογιστική

Με βάση τα παραπάνω αποτελέσματα στην περίπτωση που ο γεννήτορας είναι η λογιστική κατανομή, βλέπουμε ότι η μικρότερη τιμή του κριτηρίου AIC επιτυγχάνεται για τη γεννήτορα κατανομή και η αμέσως μικρότερη τιμή, με μικρή διαφορά από τη Βήτα παραγόμενη λογιστική (BGLog), είναι για τη γενικευμένη Βήτα παραγόμενη λογιστική (GBGLog) κατανομή. Δηλαδή, η καλύτερα προσαρμοζόμενη κατανομή φαίνεται να είναι ο γεννήτορας.

Στο επόμενο σχήμα, δίνεται συγκριτικά η συνάρτηση κατανομής των δεδομένων και αυτή των προσαρμοζόμενων κατανομών με γεννήτορα τη Λογιστική.



Σχήμα 4.6 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των προσαρμοζόμενων κατανομών με γεννήτορα τη Λογιστική

Στο παραπάνω σχήμα, δεν εμφανίζεται η GBGLog κατανομή καθώς στη συγκεκριμένη περίπτωση αναπτύσσεται στο αρνητικό τμήμα του άξονα x. Από τις προσαρμοζόμενες κατανομές με γεννήτορα τη Λογιστική, φαίνεται ο γεννήτορας ως η καλύτερη καμπύλη. Παρ' όλα αυτά, κανένα από τα μοντέλα της Λογιστικής δεν είναι κατάλληλο για την περιγραφή των δεδομένων.

4.2 Έλεγχοι λόγου πιθανοφάνειας

Ένα πλεονέκτημα της εκτίμησης των παραμέτρων της κατανομής με τη μέθοδο της μεγίστης πιθανοφάνειας είναι το γεγονός ότι είναι δυνατός ο στατιστικός έλεγχος καλής προσαρμογής της κατανομής μέσω του ελέγχου του λόγου πιθανοφανειών (likelihood ratio tests). Για τη σύγκριση μεταξύ των διαφορετικών μοντέλων-κατανομών, δηλαδή κατανομών που έχουν τον ίδιο γεννήτορα αλλά διαφορετικό αριθμό παραμέτρων, θεωρούμε τον έλεγχο υποθέσεων μεταξύ του γενικού μοντέλου (GBG κατανομή) και ενός πιο περιορισμένου μοντέλου (κάποια υποπερίπτωση της GBG κατανομής). Συγκεκριμένα, οι υποθέσεις που διατυπώνονται είναι:

H_0 : τα δεδομένα ακολουθούν το μοντέλο με περιορισμούς

H_1 : τα δεδομένα ακολουθούν το μοντέλο χωρίς περιορισμούς

Έστω θ_0 το σύνολο των παραμέτρων στο μοντέλο με περιορισμούς και θ_1 το σύνολο των παραμέτρων στο μοντέλο χωρίς περιορισμούς. Επειδή τα μοντέλα είναι εμφωλευμένα, το θ_0 είναι υποσύνολο του θ_1. Θεωρώντας τις πιθανοθάνειες $L(\hat{\theta}_0)$ και $L(\hat{\theta}_1)$ των εκτιμημένων παραμέτρων $\hat{\theta}_0$ και $\hat{\theta}_1$ αντίστοιχα, η στατιστική του λόγου πιθανοφάνειας είναι

$$LR = -2log\frac{L(\hat{\theta}_0)}{L(\hat{\theta}_1)} = -2\left[logL(\hat{\theta}_0) - logL(\hat{\theta}_1)\right] = 2\left[l(\hat{\theta}_1) - l(\hat{\theta}_0)\right].$$

Η στατιστική $LR \sim \chi_p^2$ όπου οι βαθμοί ελευθερίας p είναι η διαφορά του αριθμού των εκτιμημένων παραμέτρων των δύο μοντέλων, δηλαδή $p = \#(\theta_1) - \#(\theta_0)$.

Η μηδενική υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας a αν $LR > \chi_{p,a}^2$ ή αν το αντίστοιχο $pvalue$ είναι μικρότερο του a.

Στην περίπτωση που ο γεννήτορας κατανομή είναι η Pareto, αρχικά εξετάζουμε τον έλεγχο μεταξύ γενικευμένης Βήτα παραγόμενης Pareto και Βήτα παραγόμενης Pareto, δηλαδή

$$H_0 : c = 1 \text{ έναντι } H_1 : \text{όχι } H_0.$$

Η τιμή της στατιστικής LR είναι

$$LR = 2[-266450.7 - (-291978.7)] = 51056$$

με $LR \sim \chi_{5-4}^2 \equiv \chi_1^2$ και το αντίστοιχο $pvalue$ είναι

$$pvalue = P(LR > 51056|H_0) = P(LR > 51056\big|LR \sim \chi_1^2) < 0.00001.$$

Βλέπουμε ότι για επίπεδο εμπιστοσύνης a = 1%, 5% και 10% το pvalue είναι μικρότερο, συνεπώς απορρίπτουμε τη μηδενική υπόθεση, δηλαδή μεταξύ των GBG Pareto και BG Pareto η πρώτη προσαρμόζεται καλύτερα στα δεδομένα. Συνεπώς, η επιπλέον παράμετρος λοξότητας c είναι καθοριστικής σημασίας για την επιλογή μοντέλου.

Στη συνέχεια προχωράμε στον έλεγχο μεταξύ γενικευμένης Βήτα παραγόμενης Pareto και Kumaraswamy παραγόμενης Pareto. Ο έλεγχος είναι

$$H_0 : a = 1 \text{ έναντι } H_1 : \text{όχι } H_0$$

με αντίστοιχες τιμές της στατιστικής

$$LR = 2[-266450.7 - (-274899.4)] = 16897.4$$

όπου $LR \sim \chi_{5-4}^2 \equiv \chi_1^2$ και

$$pvalue = P(LR > 16897.4|H_0) = P(LR > 16897.4\big|LR \sim \chi_1^2) < 0.00001.$$

Όπως και στον προηγούμενο έλεγχο, απορρίπτουμε τη μηδενική υπόθεση, δηλαδή η GBG Pareto εξακολουθεί να ταιριάζει καλύτερα στα δεδομένα σε σχέση με την Kumaraswamy παραγόμενη Pareto.

Τέλος, θεωρούμε τον έλεγχο μεταξύ γενικευμένης Βήτα παραγόμενης Pareto και του γεννήτορα,

$$H_0 : a = b = c = 1 \text{ έναντι } H_1 : \text{όχι } H_0$$

για το οποίο προκύπτει ότι

$$LR = 2[-266450.7 - (-270467.1)] = 8032.8$$

με $LR \sim \chi_{5-2}^2 \equiv \chi_3^2$ και το αντίστοιχο $pvalue$ είναι

$$pvalue = P(LR > 8032.8|H_0) = P(LR > 8032.8\big|LR{\sim}\chi_3^2) < 0.00001.$$

Και σε αυτή τη σύγκριση, προκύπτει ότι η GBG Pareto εξακολουθεί να ταιριάζει καλύτερα στα δεδομένα, σε σχέση με την απλή Pareto.

Συμπερασματικά, βλέπουμε ότι η γενικευμένη Βήτα παραγόμενη κατανομή με γεννήτορα την Pareto έχει καλύτερη προσαρμογή μεταξύ των παραπάνω συγκρινόμενων υποπεριπτώσεών της, κάτι που οφείλεται στη μεγαλύτερη ευελιξία της λόγω των επιπλέον παραμέτρων της.

Στην περίπτωση που ο γεννήτορας κατανομή θεωρείται η Weibull οι εξεταζόμενοι έλεγχοι έχουν ως εξής. Αρχικά για τον έλεγχο της αναγκαιότητας της ύπαρξης της επιπλέον παραμέτρου c, ο σχετικός έλεγχος μεταξύ Βήτα παραγόμενης κατανομής και γενικευμένης Βήτα παραγόμενης Weibull είναι

$$H_0: c = 1 \text{ έναντι } H_1 : \text{όχι } H_0$$

για τον οποίο προκύπτουν

$$LR = 2[-268083.5 - (-267970.6)] = -225.8$$

με $LR{\sim}\chi_{5-4}^2 \equiv \chi_1^2$. Η αρνητική τιμή της στατιστικής μας οδηγεί στο συμπέρασμα της μη απόρριψης της μηδενικής υπόθεσης, δηλαδή η κατανομή BG προσαρμόζεται καλύτερα στα δεδομένα.

Στη συνέχεια για τον έλεγχο υποθέσεων μεταξύ της Kumaraswamy παραγόμενης κατανομής γενικευμένης Βήτα παραγόμενης Weibull και,

$$H_0: a = 1 \text{ έναντι } H_1 : \text{όχι } H_0$$

προκύπτει ότι

$$LR = 2[-268083.5 - (-268263.6)] = 360.2$$

με $LR{\sim}\chi_{5-4}^2 \equiv \chi_1^2$ και αντίστοιχο $pvalue$

$$pvalue = P(LR > 360.2|H_0) = P(LR > 360.2\big|LR{\sim}\chi_1^2) < 0.00001$$

που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση δηλαδή, η γενικευμένη Βήτα παραγόμενη Weibull κατανομή προτιμάται στην προσαρμογή των δεδομένων.

Τέλος, ο έλεγχος ανάμεσα στην κατανομή γεννήτορα και τη γενικευμένη Βήτα παραγόμενη Weibull είναι

$$H_0: a = b = c = 1 \text{ έναντι } H_1 : \text{όχι } H_0.$$

Από τα δεδομένα προκύπτει,

$$LR = 2[-268083.5 - (-277243)] = 18319$$

με $LR{\sim}\chi_{5-2}^2 \equiv \chi_3^2$ και το αντίστοιχο pvalue είναι

$$pvalue = P(LR > 18319|H_0) = P(LR > 18319\big|LR{\sim}\chi_3^2) < 0.00001$$

που σημαίνει ότι απορρίπτουμε την μηδενική υπόθεση δηλαδή, η γενικευμένη Βήτα παραγόμενη Weibull κατανομή και σε αυτή την περίπτωση προσαρμόζεται καλύτερα στα δεδομένα.

Τέλος, ανάμεσα στα διαφορετικά μοντέλα κατανομών που εξετάστηκαν, με γεννήτορες την Pareto και τη Weibull, όπως προκύπτει από τις τιμές των σχετικών κριτηρίων που δίνονται στους Πίνακες 4.2-4.5, φαίνεται ότι η καλύτερα προσαρμοζόμενη κατανομή είναι η γενικευμένη Βήτα παραγόμενη κατανομή με γεννήτορα την Pareto, *GBGPar*.

Προχωρούμε τώρα σε σχετικούς ελέγχους υποθέσεων ανάμεσα στα διαφορετικά μοντέλα κατανομών όταν ο γεννήτορας είναι η Λογιστική κατανομή. Αρχικά εξετάζουμε τον έλεγχο μεταξύ γενικευμένης Βήτα παραγόμενης Λογιστικής και Βήτα παραγόμενης Λογιστικής, δηλαδή

$$H_0 : c = 1 \text{ έναντι της } H_1 : \text{όχι } H_0.$$

Η τιμή της στατιστικής LR είναι

$$LR = 2[-350649.4 - (-350707.2)] = 115.6$$

με $LR \sim \chi^2_{5-4} \equiv \chi^2_1$ και το αντίστοιχο $pvalue$ είναι

$$pvalue = P(LR > 115.6|H_0) = P(LR > 115.6\big|LR \sim \chi^2_1) < 0.00001.$$

Βλέπουμε ότι για επίπεδο εμπιστοσύνης a = 1%, 5% και 10% το pvalue είναι μικρότερο, συνεπώς απορρίπτουμε τη μηδενική υπόθεση, δηλαδή μεταξύ των GBGLog και BGLog η πρώτη προσαρμόζεται καλύτερα στα δεδομένα. Συνεπώς, η επιπλέον παράμετρος λοξότητας c είναι καθοριστικής σημασίας για την επιλογή μοντέλου.

Στη συνέχεια προχωράμε στον έλεγχο μεταξύ γενικευμένης Βήτα παραγόμενης λογιστικής και Kumaraswamy παραγόμενης λογιστικής. Ο έλεγχος είναι

$$H_0 : a = 1 \text{ έναντι της } H_1 : \text{όχι } H_0$$

με αντίστοιχες τιμές της στατιστικής

$$LR = 2[-350649.4 - (-367654.5)] = 34010.2$$

όπου $LR \sim \chi^2_{5-4} \equiv \chi^2_1$ και

$$pvalue = P(LR > 34010.2|H_0) = P(LR > 34010.2\big|LR \sim \chi^2_1) < 0.00001.$$

Όπως και στον προηγούμενο έλεγχο, απορρίπτουμε τη μηδενική υπόθεση, δηλαδή η GBGLog εξακολουθεί να ταιριάζει καλύτερα στα δεδομένα σε σχέση με την Kumaraswamy παραγόμενη λογιστική.

Τέλος, θεωρούμε τον έλεγχο μεταξύ γενικευμένης Βήτα παραγόμενης λογιστικής και του γεννήτορα,

$$H_0 : a = b = c = 1 \text{ έναντι της } H_1 : \text{όχι } H_0$$

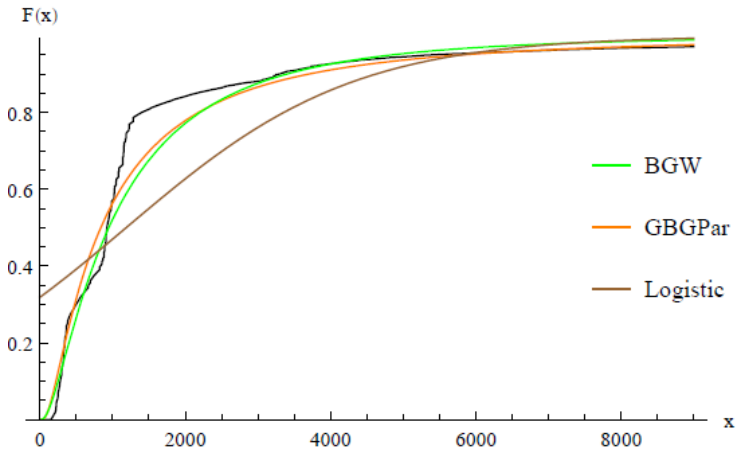
για το οποίο προκύπτει ότι

$$LR = 2[-350649.4 - (-312222.1)] = -76854.6$$

με $LR \sim \chi^2_{5-2} \equiv \chi^2_3$. Η αρνητική τιμή της στατιστικής μας οδηγεί στο συμπέρασμα της μη απόρριψης της μηδενικής υπόθεσης, δηλαδή η λογιστική κατανομή προσαρμόζεται καλύτερα στα δεδομένα σε σχέση με την GBGLog.

Τέλος, με βάση το AIC, το μόνο κοινό διαθέσιμο κριτήριο ανάμεσα στις διαφορετικές κατανομές γεννήτορες, βλέπουμε ότι η *GBGPar* προσαρμόζεται καλύτερα στα υπό μελέτη δεδομένα.

Στο σχήμα 4.7 φαίνεται η εμπειρική συνάρτηση κατανομής και οι θεωρητικές κατανομές των καλύτερα προσαρμοζόμενων κατανομών για κάθε έναν από τους τρεις εξεταζόμενους γεννήτορες.
Είναι προφανές ότι οι κατανομές GBGPar και BGW παρέχουν την καλύτερη προσαρμογή στα δεδομένα, με την πρώτη να υπερέχει της δεύτερης στο μεγαλύτερο τμήμα της εμπειρικής κατανομής.



Σχήμα 4.7 Σύγκριση της εμπειρικής συνάρτησης κατανομής και των καλύτερα προσαρμοζόμενων κατανομών για κάθε γεννήτορα

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εργασία αυτή μελετήθηκαν κατανομές που αποτελούν εν δυνάμει μοντέλα προσαρμογής μη κανονικών δεδομένων τα οποία συναντώνται για παράδειγμα στον αναλογιστικό κλάδο. Ειδικότερα, παρουσιάστηκε η γενικευμένη βήτα παραγόμενη (GBG) οικογένεια κατανομών, όσον αφορά μονοδιάστατες κατανομές, η οποία περιλαμβάνει και ενοποιεί τη βήτα παραγόμενη (BG), την Kumaraswamy παραγόμενη (KwG) και την εκθετικοποιημένη οικογένεια κατανομών. Η οικογένεια GBG έχει σαν πυρήνα της τη γενικευμένη βήτα κατανομή πρώτου είδους που μελετήθηκε από τον McDonald (1984) και επεκτείνει μια οποιαδήποτε κοινή κατανομή που καλείται γεννήτορας F, όπως την κανονική, Weibull, Laplace, Pareto, Gumbel κλπ. οι οποίες έχουν διαφορετικές μορφές. Αναπτύχθηκαν τύποι υπολογισμού της συνάρτησης πυκνότητας και συνάρτησης κατανομής της, καθώς και σχετικών στατιστικών μέτρων όπως οι συνήθεις και οι L ροπές, η ροπογεννήτρια συνάρτηση, η ποσοστιαία συνάρτηση, οι μέσες αποκλίσεις από το μέσο και τη διάμεσο, η στατιστική διατεταγμένου δείγματος. Διερευνήθηκε η συμβολή των παραμέτρων της GBG στην ερμηνεία της λοξότητας και κύρτωσης της παραγόμενης κατανομής. Δόθηκαν οι εκτιμήτριες μέγιστης πιθανοθάνειας και σχετική στατιστική συμπερασματολογία των παραμέτρων της GBG. Στη συνέχεια μελετήθηκαν διεξοδικά οι κατανομές-γεννήτορες Pareto, Weibull και Λογιστική και οι αντίστοιχες GBG κατανομές. Τέλος, η εφαρμογή των GBG κατανομών σε πραγματικά δεδομένα έδειξε την ευελιξία της GBG οικογένειας στη μοντελοποίησή τους.

ΠΑΡΑΡΤΗΜΑ

Π1. Συνάρτηση Γάμμα

Η συνάρτηση Γάμμα ή αλλιώς ολοκλήρωμα Euler δευτέρου είδους, δίνεται από τον τύπο

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt\,, a > 0, a \in \mathbf{R}.$$

Για τη συνάρτηση Γάμμα ισχύουν οι παρακάτω τύποι

$\Gamma(a+1) = a\Gamma(a), \qquad a > 0, a \in \mathbf{R}$

$\Gamma(n) = (n-1)!\,, \qquad n = 1,2,3, \dots$

$\Gamma(1) = 1,$

$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\,.$

Π2. Μη πλήρης ή ατελής συνάρτηση Γάμμα (incomplete gamma function)

Η κάτω μη πλήρης συνάρτηση Γάμμα δίνεται από τη σχέση

$$\gamma(a;x) = \int_0^x t^{a-1}e^{-t}dt\,, a > 0, x > 0.$$

Αντίστοιχα, η άνω μη πλήρης συνάρτηση Γάμμα δίνεται από το ολοκλήρωμα

$$\Gamma(a;x) = \int_x^\infty t^{a-1}e^{-t}dt\,, a > 0, x > 0.$$

Για τις μη πλήρεις συναρτήσεις Γάμμα ισχύουν οι σχέσεις

$\gamma(a;\infty) = \Gamma(a),$

$\Gamma(a;0) = \Gamma(a),$

$\gamma(a;x) + \Gamma(a;x) = \Gamma(a).$

Π3. Συνάρτηση Βήτα

Η συνάρτηση Βήτα ή αλλιώς ολοκλήρωμα Euler πρώτου είδους, δίνεται από τον τύπο

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt\,, a > 0, b > 0.$$

Η συνάρτηση Βήτα είναι συμμετρική, δηλαδή ισχύει

$$B(a,b) = B(b,a).$$

Όταν ένα από τα δύο ορίσματα της συνάρτησης είναι η μονάδα, τότε

$$B(c,1) = B(1,c) = \frac{1}{c}.$$

Η συνάρτηση Βήτα συνδέεται με τη συνάρτηση Γάμμα με τη σχέση

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Η συνάρτηση Βήτα εκφράζεται μέσω των διωνυμικών συντελεστών. Πράγματι, για n και k ακέραιους, με βάση την προηγούμενη σχέση, προκύπτει ότι

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} = \frac{1}{(n+1)B(k+1,n-k+1)}.$$

Π4. Μη πλήρης ή ατελής συνάρτηση Βήτα

Η μη πλήρης συνάρτηση Βήτα ορίζεται ως το ολοκλήρωμα

$$B(a,b;x) \text{ ή } B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt\,, a>0, b>0, 0<x<1.$$

Για $x=1$, προκύπτει η συνάρτηση Βήτα,

$$B(a,b;1) = B(a,b).$$

Η παράγωγός της είναι

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

Π5. Λόγος συνάρτησης Βήτα (incomplete beta function ratio)

Ο λόγος της συνάρτησης Βήτα ορίζεται ως

$$I_x(a,b) = \frac{B(a,b;x)}{B(a,b)} = \frac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}dt\,.$$

Η ποσότητα $I_x(a,b)$ είναι η αθροιστική συνάρτηση κατανομής της κατανομής $Beta^{(I)}(a,b)$

Για το λόγο ισχύουν

$I_0(a,b) = 0,$

$I_1(a,b) = 1,$

$I_x(a,b) = 1 - I_{1-x}(b,a),$

$$\sum_{j=k}^{n}\binom{n}{j}x^j(1-x)^{n-j} = I_x(k,n-k+1).$$

Η τελευταία σχέση, μπορεί να αποδειχθεί με ολοκλήρωση κατά παράγοντες, και συνδέει το λόγο της συνάρτησης Βήτα με τη δεξιά ουρά της διωνυμικής κατανομής. Συγκεκριμένα το άθροισμα είναι η πιθανότητα $P(Y \geq k)$, αν $Y \sim Binomial(n,x)$. Επίσης από την ίδια σχέση, για την αθροιστική συνάρτηση κατανομής της διωνυμικής, προκύπτει ότι

$$\sum_{j=0}^{k-1}\binom{n}{j}x^j(1-x)^{n-j} = 1 - I_x(k,n-k+1) = I_{1-x}(n-k+1,k).$$

Από τη σχέση αυτή θέτοντας $y = 1 - x$ βρίσκουμε

$$\sum_{j=0}^{k-1} \binom{n}{j} (1-y)^j y^{n-j} = I_y(n-k+1, k)$$

και με την αντικατάσταση $a = n - k + 1, b = k$ παίρνουμε

$$\sum_{j=0}^{b-1} \binom{a+b-1}{j} y^{a+b-j-1} (1-y)^j = I_y(a,b), \qquad \text{για } b \text{ ακέραιο.}$$

Π6. Αριθμοί Bernoulli B_n

Οι αριθμοί Bernoulli ανακαλύφθησαν από τον Jacob Bernoulli σε μια προσπάθεια υπολογισμού του αθροίσματος

$$S_k(n) = 1^k + 2^k + \cdots + n^k$$

και εμφανίζονται στη μοντέρνα θεωρία αριθμών και γεωμετρία.

Η ακολουθία των αριθμών Bernoulli $\{B_n\}_{n\geq 0}$ ορίζεται από τη δυναμοσειρά Taylor

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n$$

από την οποία προκύπτει η ακόλουθη αναδρομική σχέση υπολογισμού των B_n

$$\sum_{i=0}^{n-1} \binom{n}{i} B_i = \binom{n}{0} B_0 + \binom{n}{1} B_1 + \binom{n}{2} B_2 + \cdots + \binom{n}{n-1} B_{n-1} = 0, \text{ με } B_0 = 1.$$

Για τους αριθμούς Bernoulli ισχύει

$$B_{2n+1} = 0, \qquad n = 1,2,3,..$$

Οι αριθμοί Bernoulli μπορούν να υπολογιστούν επίσης και από τη σχέση

$$B_n = \lim_{x\to 0} \frac{d^n}{dx^n} \frac{x}{e^x - 1}.$$

Οι πρώτοι έντεκα αριθμοί Bernoulli είναι

$$B_0 = 1, B_1 = \frac{1}{2}, B_2 = \frac{1}{6}, B_3 = 0, B_4 = -\frac{1}{30}, B_5 = 0,$$
$$B_6 = \frac{1}{42}, B_7 = 0, B_8 = -\frac{1}{30}, B_9 = 0, B_{10} = \frac{5}{66}.$$

Π7. Αποδείξεις για την κατανομή Pareto τύπου II, $Par^{II}(\delta, \lambda)$

Π7.1 Ροπές

$$E\left(X^k\right) = \int_x x f_x(x) dx = \int_0^{\infty} x^k \frac{\delta \lambda^{\delta}}{(\lambda + x)^{\delta+1}} dx = \frac{\delta}{\lambda} \int_0^{\infty} x^k \left(\frac{\lambda}{\lambda + x}\right)^{\delta+1} dx$$

για το ολοκλήρωμα θεωρούμε την αντικατάσταση

$y = \frac{\lambda}{\lambda+x} \Rightarrow \lambda + x = \frac{\lambda}{y} \Rightarrow x = \frac{\lambda}{y} - \lambda \Rightarrow dx = -\frac{\lambda}{y^2} dy$ και για τα όρια είναι $0 < x < \infty \Rightarrow 1 <$

$y < 0$

οπότε

$$E(X^k) = \frac{\delta}{\lambda}\int_1^0 \left(\frac{\lambda}{y} - \lambda\right)^k y^{\delta+1}\left(-\frac{\lambda}{y^2}dy\right) = \delta\int_0^1 \left[\frac{\lambda(1-y)}{y}\right]^k y^{\delta-1}dy$$

$$= \delta\lambda^k \int_0^1 (1-y)^k y^{\delta-k-1}dy.$$

Όμως,

$$\int_0^1 x^{p-1}(1-x)^{q-1}dx = \frac{1}{B(p,q)} = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, p > 0, q > 0$$

οπότε

$$E(X^k) = \delta\lambda^k \frac{\Gamma(\delta-k)\Gamma(k+1)}{\Gamma(\delta+1)} = \delta\lambda^k \frac{\Gamma(\delta-k)k!}{\delta\Gamma(\delta)} = \frac{\lambda^k k!}{(\delta-1)(\delta-2)\cdots(\delta-k)}, k < \delta.$$

Π7.2 Μέση τιμή, διασπορά, λοξότητα και κύρτωση

Θέτοντας *k* = 1, 2, 3, 4 στον τύπο των ροπών προκύπτει αντίστοιχα

$$\mu = E(X) = \frac{\lambda}{\delta-1}, \delta > 1$$

$$E(X^2) = \frac{2\lambda^2}{(\delta-1)(\delta-2)}, \delta > 2$$

$$E(X^3) = \frac{6\lambda^3}{(\delta-1)(\delta-2)(\delta-3)}, \delta > 3$$

$$E(X^4) = \frac{24\lambda^4}{(\delta-1)(\delta-2)(\delta-3)(\delta-4)}, \delta > 4$$

οπότε για τη διασπορά είναι

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \frac{2\lambda^2}{(\delta-1)(\delta-2)} - \left(\frac{\lambda}{\delta-1}\right)^2 = \frac{\delta\lambda^2}{(\delta-1)^2(\delta-2)}, \delta > 2.$$

Για τη λοξότητα και κύρτωση, αντικαθιστούμε τις απαιτούμενες ροπές στους αντίστοιχους τύπους, και απλοποιούμε οπότε έχουμε

$$skew(X) = \frac{E(X^3) - 3E(X)\,E(X^2) + 2[E(X)]^3}{\sigma^3}$$

$$= \frac{\frac{6\lambda^3}{(\delta-1)(\delta-2)(\delta-3)} - 3\frac{\lambda}{\delta-1}\frac{2\lambda^2}{(\delta-1)(\delta-2)} - 2\left(\frac{\lambda}{\delta-1}\right)^3}{\left[\frac{\delta\lambda^2}{(\delta-1)^2(\delta-2)}\right]^{3/2}}$$

$$= \frac{2(\delta+1)}{\delta-3}\sqrt{\frac{\delta-2}{\delta}}, \qquad \delta > 3$$

και

$$kurt(X) = \frac{E(X^4) - 4E(X)\,E(X^3) + 6[E(X)]^2E(X^2) - 3[E(X)]^4}{\sigma^4}$$

$$= \frac{1}{\left[\frac{\delta\lambda^2}{(\delta-1)^2(\delta-2)}\right]^2}\left[\frac{24\lambda^4}{(\delta-1)(\delta-2)(\delta-3)(\delta-4)} - 4\frac{\lambda}{\delta-1}\,\frac{6\lambda^3}{(\delta-1)(\delta-2)(\delta-3)}\right.$$

$$\left. + 6\left(\frac{\lambda}{\delta-1}\right)^2\frac{2\lambda^2}{(\delta-1)(\delta-2)} - 3\left(\frac{\lambda}{\delta-1}\right)^4\right]$$

$$= \frac{3(\delta-2)(3\delta^2+\delta+2)}{(\delta-4)(\delta-3)\delta}, \delta > 4.$$

Π7.3 Ροπογεννήτρια συνάρτηση

Γνωρίζουμε ότι αν μια τ.μ. Y ακολουθεί την Pareto τύπου I κατανομή, δηλαδή $Y \sim Par^{(I)}(\lambda, \delta)$, με σ.π.π. $f(y) = \frac{\delta\lambda^\delta}{y^{\lambda+1}}, y \geq \lambda, \delta > 0$ τότε η τ.μ. X που προκύπτει με το μετασχηματισμό $X = Y - \lambda$ ακολουθεί την κατανομή Pareto τύπου II, δηλαδή $X \sim Par^{(II)}(\delta, \lambda)$.

Για τη ροπογεννήτρια συνάρτηση της Y είναι

$$M_Y(t) = \begin{cases} \delta(-\lambda t)^\delta[\Gamma(-\delta) - \Gamma(-\delta, -\lambda t)], & t < 0 \\ 1, & t = 0 \\ \infty, & t > 0 \end{cases}$$

οπότε για τη ροπογεννήτρια της X έχουμε

$$M_X(t) = E(e^{tX}) = E\left(e^{t(Y-\lambda)}\right) = E\left(e^{tY}e^{-\lambda t}\right) = e^{-\lambda t}E(e^{tY}) = e^{-\lambda t}M_Y(t)$$

$$= \begin{cases} e^{-\lambda t}\delta(-\lambda t)^\delta[\Gamma(-\delta) - \Gamma(-\delta, -\lambda t)], & t < 0 \\ 1, & t = 0 \\ \infty, & t > 0 \end{cases}.$$

Π7.4 Περιορισμένη μέση τιμή

$$E(X; x) = \int_0^x [1 - F_X(t)]dt = \int_0^x \left(\frac{\lambda}{\lambda+t}\right)^\delta dt = \lambda^\delta \int_0^x \frac{1}{(\lambda+t)^\delta}\,dt = \lambda^\delta \int_0^x (\lambda+t)^{-\delta}\,dt$$

$$= \lambda^\delta\left[\frac{1}{-\delta+1}(\lambda+t)^{-\delta+1}\right]_0^x = \frac{\lambda}{\delta-1}\left[1 - \left(\frac{\lambda}{\lambda+x}\right)^{\delta-1}\right].$$

Π7.5 Συνάρτηση quantile

Για να βρούμε την $Q(p)$ λύνουμε την εξίσωση $F(x) = p$ ως προς x οπότε παίρνουμε

$$F(x) = p \Rightarrow 1 - \left(\frac{\lambda}{\lambda+x}\right)^\delta = p \Rightarrow \left(\frac{\lambda}{\lambda+x}\right)^\delta = 1 - p \Rightarrow \frac{\lambda}{\lambda+x} = (1-p)^{1/\delta} - \lambda$$

$$\Rightarrow x = \frac{\lambda}{(1-p)^{1/\delta}}$$

οπότε $Q(p) = \lambda\left[\frac{1}{(1-p)^{\frac{1}{\delta}}} - 1\right]$.

Π8. Αποδείξεις για την κατανομή Weibull, $W(\lambda, \gamma)$

Π8.1 Ροπές

Έχουμε

$$E(X^k) = \int_x x^k f(x)dx = \int_0^\infty x^k \lambda\gamma x^{\gamma-1} e^{-\lambda x^\gamma} dx$$

οπότε θέτοντας $y = x^\gamma \Rightarrow x = y^{1/\gamma} \Rightarrow dx = \frac{1}{\gamma} y^{\frac{1}{\gamma}-1} dy$

και αλλάζοντας αντίστοιχα τα όρια του ολοκληρώματος είναι $0 < x < \infty \Rightarrow 0 < y < \infty$, παίρνουμε

$$E(X^k) = \int_0^\infty y^{k/\gamma} \lambda\gamma y^{\frac{1}{\gamma}(\gamma-1)} e^{-\lambda y} \frac{1}{\gamma} y^{\frac{1}{\gamma}-1} dy = \lambda \int_0^\infty y^{k/\gamma} e^{-\lambda y} dy$$

$$= \lambda \int_0^\infty y^{k/\gamma} e^{-\lambda y} dy = \lambda \frac{\Gamma\left(\frac{k}{\gamma}+1\right)}{\lambda^{\frac{k}{\gamma}+1}} = \frac{\Gamma\left(\frac{k}{\gamma}+1\right)}{\lambda^{\frac{k}{\gamma}}}, \quad k > -\gamma.$$

Π8.2 Μέση τιμή, διασπορά, λοξότητα και κύρτωση

Θέτοντας $k = 1,2,3,4$ στον τύπο των ροπών προκύπτει αντίστοιχα

$$\mu = E(X) = \frac{\Gamma\left(\frac{1}{\gamma}+1\right)}{\lambda^{1/\gamma}}$$

$$E(X^2) = \frac{\Gamma\left(\frac{2}{\gamma}+1\right)}{\lambda^{2/\gamma}}$$

$$E(X^3) = \frac{\Gamma\left(\frac{3}{\gamma}+1\right)}{\lambda^{3/\gamma}}$$

$$E(X^4) = \frac{\Gamma\left(\frac{4}{\gamma}+1\right)}{\lambda^{4/\gamma}}$$

οπότε για τη διασπορά είναι

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \frac{\Gamma\left(\frac{2}{\gamma}+1\right)}{\lambda^{2/\gamma}} - \left(\frac{\Gamma\left(\frac{1}{\gamma}+1\right)}{\lambda^{1/\gamma}}\right)^2 = \frac{\Gamma\left(\frac{2}{\gamma}+1\right) - \Gamma^2\left(\frac{1}{\gamma}+1\right)}{\lambda^{2/\gamma}}.$$

Για τη λοξότητα με αντικατάσταση στο σχετικό τύπο προκύπτει

$$skew(X) = \frac{E(X^3) - 3\mu E(X^2) - 2\mu^3}{\sigma^3}$$

$$= \frac{\frac{\Gamma\left(\frac{3}{\gamma}+1\right)}{\lambda^{3/\gamma}} - 3\frac{\Gamma\left(\frac{1}{\gamma}+1\right)}{\lambda^{1/\gamma}}\frac{\Gamma\left(\frac{2}{\gamma}+1\right)}{\lambda^{2/\gamma}} + 2\frac{\Gamma^3\left(\frac{1}{\gamma}+1\right)}{\lambda^{3/\gamma}}}{\left\{\lambda^{-2/\gamma}\left[\Gamma\left(\frac{2}{\gamma}+1\right) - \left[\Gamma\left(\frac{1}{\gamma}+1\right)\right]^2\right]\right\}^{3/2}}$$

$$= \frac{\Gamma\left(\frac{3}{\gamma}+1\right) + 2\left[\Gamma\left(\frac{1}{\gamma}+1\right)\right]^3 - 3\Gamma\left(\frac{1}{\gamma}+1\right)\Gamma\left(\frac{2}{\gamma}+1\right)}{\left[\Gamma\left(\frac{2}{\gamma}+1\right) - \left[\Gamma\left(\frac{1}{\gamma}+1\right)\right]^2\right]^{3/2}}.$$

Αντίστοιχα για την κύρτωση βρίσκουμε

$$kurt(X) = \frac{E(X^4) - 4E(X)\,E(X^3) + 6[E(X)]^2E(X^2) - 3[E(X)]^4}{\sigma^4}$$

$$= \frac{\frac{\Gamma\left(\frac{4}{\gamma}+1\right)}{\lambda^{4/\gamma}} - 4\frac{\Gamma\left(\frac{1}{\gamma}+1\right)}{\lambda^{1/\gamma}}\frac{\Gamma\left(\frac{3}{\gamma}+1\right)}{\lambda^{3/\gamma}} + 6\frac{\Gamma^2\left(\frac{1}{\gamma}+1\right)}{\lambda^{2/\gamma}}\frac{\Gamma\left(\frac{2}{\gamma}+1\right)}{\lambda^{2/\gamma}} - 3\frac{\Gamma^4\left(\frac{1}{\gamma}+1\right)}{\lambda^{4/\gamma}}}{\frac{\left[\Gamma\left(\frac{2}{\gamma}+1\right) - \Gamma^2\left(\frac{1}{\gamma}+1\right)\right]^2}{\lambda^{4/\gamma}}}$$

$$= \frac{\Gamma\left(\frac{4}{\gamma}+1\right) - 4\Gamma\left(\frac{1}{\gamma}+1\right)\Gamma\left(\frac{3}{\gamma}+1\right) + 6\Gamma^2\left(\frac{1}{\gamma}+1\right)\Gamma\left(\frac{2}{\gamma}+1\right) - 3\Gamma^4\left(\frac{1}{\gamma}+1\right)}{\left[\Gamma\left(\frac{2}{\gamma}+1\right) - \Gamma^2\left(\frac{1}{\gamma}+1\right)\right]^2}.$$

Π8.3 Περιορισμένη μέση τιμή

Έχουμε

$$E(X;x) = \int_0^x t f_X(t)dt + x\bar{F}_X(x) = \int_0^x t\,\lambda\,\gamma\,t^{\gamma-1}e^{-\lambda t^\gamma}dt + xe^{-\lambda x^\gamma}.$$

Θα μετασχηματίσουμε το ολοκλήρωμα σε μία κάτω μη πλήρη συνάρτηση Γάμμα.

Θέτουμε $y = \lambda t^\gamma \Rightarrow dy = \lambda\gamma t^{\gamma-1}dt$ οπότε είναι $t = \left(\frac{y}{\lambda}\right)^{\frac{1}{\gamma}}$ και $0 < t < x \Rightarrow 0 < y < \lambda x^\gamma$

οπότε έχουμε (με βάση το Π2)

$$E(X;x) = \int_0^{\lambda x^\gamma} \left(\frac{y}{\lambda}\right)^{\frac{1}{\gamma}} e^{-y}dy + xe^{-\lambda x^\gamma} = \lambda^{-\frac{1}{\gamma}}\int_0^{\lambda x^\gamma} y^{\frac{1}{\gamma}}\,e^{-y}dy + xe^{-\lambda x^\gamma}$$

$$= \lambda^{-1/\gamma}\gamma\left(\frac{1}{\gamma}+1;\lambda x^\gamma\right) + xe^{-\lambda x^\gamma}.$$

Π8.4 Συνάρτηση quantile

Για να βρούμε την $Q(p)$ λύνουμε την εξίσωση $F(x) = p$ ως προς x:

$F(x)=p \Rightarrow 1-e^{-\lambda x^{\gamma}}=p \Rightarrow e^{-\lambda x^{\gamma}}=1-p \Rightarrow -\lambda x^{\gamma}=\ln(1-p) \Rightarrow x^{\gamma}=-\frac{\ln(1-p)}{\lambda}$.

Επομένως, $Q(p)=\left[-\frac{1}{\lambda}\ln(1-p)\right]^{1/\gamma}$.

Π9. Αποδείξεις για τη Λογιστική κατανομή, $Log(\mu,\sigma)$

Για τη μελέτη-αποδείξεις της $Log(\mu,\sigma)$ θα χρησιμοποιήσουμε το γεγονός ότι προκύπτει με γραμμικό μετασχηματισμό της τυπικής λογιστικής κατανομής $Log(0,1)$, και θα εκκμεταλλευτούμε τις ιδιότητες της τελευταίας. Είναι γνωστό ότι:

$$\text{αν } Y\sim Log(0,1) \text{ τότε } X=\mu+\sigma Y\sim Log(\mu,\sigma).$$

Π9.1 Ένταση κινδύνου

$$h(x)=\frac{f(x)}{\bar{F}(x)}=\frac{\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}}{1-\frac{e^{\frac{x-\mu}{\sigma}}}{1+e^{\frac{x-\mu}{\sigma}}}}=\frac{\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)^2}}{\frac{1}{1+e^{\frac{x-\mu}{\sigma}}}}=\frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{\frac{x-\mu}{\sigma}}\right)}=\frac{1}{\sigma\left(e^{-\frac{x-\mu}{\sigma}}+1\right)}.$$

Π9.2 Ροπές

Για την $Y\sim Log(0,1)$ είναι

$$E(Y^k)=\begin{cases}0, & k=1,3,5,\dots\\ \left(2^k-2\right)\pi^k|B_k|, & k=2,4,6,\dots\end{cases}$$

όπου B_k είναι ο k-αριθμός Bernoulli. Επομένως, οι ροπές k τάξης περί το μέσο μ της X είναι

$$E[(X-\mu)^k]=E[(\sigma Y)^k]=\sigma^k E\left(Y^k\right)=\begin{cases}0, & k=1,3,5,\dots\\ \sigma^k\left(2^k-2\right)\pi^k|B_k|, & k=2,4,6,\dots\end{cases}$$

Π9.3 Μέση τιμή και διασπορά

Αφού η κατανομή είναι συμμετρική γύρω από την παράμετρο θέσης μ, για τη μέση τιμή είναι

$$E(X)=\mu.$$

Όσον αφορά τη διασπορά, από Π9.2 και Π6 βρίσκουμε

$$V(X)=E[(X-\mu)^2]=\sigma^2(2^2-2)\pi^2|B_2|=2\sigma^2\pi^2\frac{1}{6}=\frac{\sigma^2\pi^2}{3}.$$

Π9.4 Λοξότητα και κύρτωση

Για την $Y\sim Log(0,1)$ έχουμε $skew(Y)=0$ και $kurt(Y)=21/5$.

Για τη λοξότητα και την κύρτωση ενός γραμμικού μετασχηματισμού ισχύει

$$skew(aZ+b)=\begin{cases}skew(Z), & a>0\\ -skew(Z), & a<0\end{cases} \text{ και } kurt(aZ+b)=kurt(Z) \text{ για } a\in\mathbf{R}-\{0\}, b\in\mathbf{R}$$

οπότε για την $X=\mu+\sigma Y$ παίρνουμε

$$skew(X)=skew(\mu+\sigma Y)=skew(Y)=0$$

$$kurt(X) = kurt(\mu + \sigma Y) = kurt(Y) = 21/5.$$

Ενναλλακτικά, θα μπορούσαμε να χρησιμοποιήσουμε τις σχέσεις

$$skew(X) = \frac{E(X^3) - 3\mu E(X^2) - 2\mu^3}{\sigma^3}$$

και

$$kurt(X) = \frac{E(X^4) - 4E(X)\,E(X^3) + 6[E(X)]^2 E(X^2) - 3[E(X)]^4}{\sigma^4}$$

και αποτελέσματα για τις ροπές από το Π9.2.

Π9.5 Ροπογεννήτρια συνάρτηση

Η ροπογεννήτρια συνάρτηση της $Y \sim Log(0,1)$ δίνεται από τη σχέση

$$M_Y(t) = B(1+t, 1-t) = \Gamma(1+t)\Gamma(1-t), \qquad t \in (-1,1).$$

Επομένως, για την $X = \mu + \sigma Y$ θα έχουμε

$$M_X(t) = E\left(e^{tX}\right) = E\left(e^{t(\mu+\sigma Y)}\right) = E\left(e^{\mu t}e^{\sigma t Y}\right) = e^{\mu t}E\left(e^{\sigma t Y}\right) = e^{\mu t}M_Y(\sigma\, t)$$
$$= e^{\mu t}B(1+\sigma t, 1-\sigma t) = e^{\mu t}\Gamma(1+\sigma t)\Gamma(1-\sigma t), \qquad t \in (-1,1).$$

Π9.6 Περιορισμένη μέση τιμή

Είναι

$$1 - F(x) = 1 - \frac{1}{e^{-\frac{x-\mu}{\sigma}} + 1} = \frac{e^{-\frac{x-\mu}{\sigma}} + 1 - 1}{e^{-\frac{x-\mu}{\sigma}} + 1} = \frac{e^{-\frac{x-\mu}{\sigma}}}{e^{-\frac{x-\mu}{\sigma}} + 1} = \frac{1}{1 + e^{\frac{x-\mu}{\sigma}}}$$

οπότε

$$E(X;x) = \int_0^x [1 - F_X(t)]dt = \int_0^x \frac{1}{1 + e^{\frac{t-\mu}{\sigma}}}\,dt\,.$$

Θέτουμε $y = e^{\frac{t-\mu}{\sigma}} \Rightarrow dy = \frac{1}{\sigma}e^{\frac{t-\mu}{\sigma}}dt = \frac{1}{\sigma}ydt \Rightarrow dt = \frac{\sigma}{y}dy$ και για $0 < t < x$ είναι $e^{\frac{-\mu}{\sigma}} < t < e^{\frac{x-\mu}{\sigma}}$, άρα το ολοκλήρωμα γίνεται

$$E(X;x) = \sigma \int_{e^{\frac{-\mu}{\sigma}}}^{e^{\frac{x-\mu}{\sigma}}} \frac{1}{(1+y)y}dy.$$

Με ανάλυση σε απλά κλάσματα βρίσκουμε $\frac{1}{(1+y)y} = \frac{A}{y} + \frac{B}{1+y} = \frac{1}{y} - \frac{1}{1+y}$

οπότε

$$E(X;x) = \sigma \int_{e^{\frac{-\mu}{\sigma}}}^{e^{\frac{x-\mu}{\sigma}}} \left(\frac{1}{y} - \frac{1}{1+y}\right)dy = \sigma[\ln y - \ln(1+y)]_{e^{\frac{-\mu}{\sigma}}}^{e^{\frac{x-\mu}{\sigma}}}$$

$$= \sigma\left[\left(\frac{x-\mu}{\sigma} - \ln\left(1+e^{\frac{x-\mu}{\sigma}}\right)\right) - \left(\frac{-\mu}{\sigma} - \ln\left(1+e^{\frac{-\mu}{\sigma}}\right)\right)\right]$$

$$= x + \sigma\left[-\ln\left(1+e^{\frac{x-\mu}{\sigma}}\right) + \ln\left(1+e^{\frac{-\mu}{\sigma}}\right)\right]$$

με $e^{\frac{x-\mu}{\sigma}} > -1$ και $e^{\frac{-\mu}{\sigma}} > -1$.

Π9.7 Συνάρτηση quantile

Για να βρούμε την $Q(p)$ λύνουμε την εξίσωση $F(x) = p$ ως προς x:

$$F(x) = p \Rightarrow \frac{1}{e^{-\frac{x-\mu}{\sigma}}+1} = p \Rightarrow 1 + e^{-\frac{x-\mu}{\sigma}} = \frac{1}{p} \Rightarrow e^{-\frac{x-\mu}{\sigma}} = \frac{1}{p} - 1 \Rightarrow -\frac{x-\mu}{\sigma} = \ln\left(\frac{1-p}{p}\right)$$

$$\Rightarrow x = \mu - \sigma\ln\left(\frac{1-p}{p}\right) \Rightarrow x = \mu + \sigma\ln\left(\frac{p}{1-p}\right).$$

Επομένως, $Q(p) = \mu + \sigma\ln\left(\frac{p}{1-p}\right)$.

Π10. Κώδικας - αποτελέσματα R

```
# plot of histogram and the empirical CDF

> library("MASS")
> library("fitdistrplus")
> plotdist(t,histo=TRUE,demp=TRUE)
> plotdist(t,histo=TRUE, ylim=c(0,0.0000001))

# descriptive measures

> summary(t)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   150.2    379.3    924.6   2404.0   1206.0 639800.0

> length(t)
[1] 32494

> mean(t)
[1] 2404.193

> sd(t)
[1] 15491.42

> var(t)
[1] 239984168

> min(t)
[1] 150.2

> max(t)
[1] 639846

> median(t)
[1] 924.61

> library(moments)
> skewness(t)
[1] 23.34538

> kurtosis(t)
[1] 677.7666
```

```
> quantile(t)
       0%        25%        50%        75%       100%
   150.200    379.335    924.610   1205.810 639846.000

> q34 = quantile(t,prob=c(3/4))
> q12 = quantile(t,prob=c(1/2))
> q14 = quantile(t,prob=c(1/4))
> q78 = quantile(t,prob=c(7/8))
> q58 = quantile(t,prob=c(5/8))
> q38 = quantile(t,prob=c(3/8))
> q18 = quantile(t,prob=c(1/8))
> q68 = quantile(t,prob=c(6/8))
> q28 = quantile(t,prob=c(2/8))

# Bowley’s skewness
> B = (q34 - 2*q12 + q14)/(q34 - q14)

# Moors’ kurtosis
> M = (q78 - q58 - q38 + q18)/(q68 - q28)

> B
       75%
-0.3195196
> M
   87.5%
1.525377

# PARETO
# Estimate parent distribution PARETO

> library(actuar)
> p1 <- fitdist(t, "pareto")

> summary(p1)
Fitting of the distribution ' pareto ' by maximum likelihood
Parameters :
         estimate Std. Error
shape    2.612965  0.0389469
scale 2700.204132 51.6445322
Loglikelihood:  -270467.1   AIC:  540938.1   BIC:  540954.9
Correlation matrix:
          shape     scale
shape 1.0000000 0.9281585
scale 0.9281585 1.0000000

> gofstat(p1)
Goodness-of-fit statistics
                             1-mle-pareto
Kolmogorov-Smirnov statistic    0.1609977
Cramer-von Mises statistic    242.3204847
Anderson-Darling statistic   1476.6563942

Goodness-of-fit criteria
                               1-mle-pareto
Aikake's Information Criterion     540938.1
Bayesian Information Criterion     540954.9

# Estimate BG Lomax

> mbetag("lomax", t, starts = c(18, 15, 2600, 0.22), method="SANN")
$Estimates
              MLE    Std. Dev. Inf. 95% CI Sup. 95% CI
[1,] 1.972017e+01 0.1691501959 19.38864018 20.05169676
[2,] 1.613421e+01 0.1890623416 15.76365527 16.50476603
[3,] 2.599301e+03          NaN         NaN         NaN
[4,] 5.502165e-02 0.0004795364  0.05408178  0.05596152
```

```
$Measures
      AIC     CAIC      BIC     HQIC        W        A Min(-log(Likelihood))
 583965.4 583965.4 583998.9 583976.1 103.2789 569.8506              291978.7

$`Kolmogorov-Smirnov Test`
 KS Statistic KS p-value
    0.3225621          0

$`Convergence Status`

 "Algorithm Converged"

# Estimate Kumaraswamy Lomax
> mkumg("lomax", t, starts=c(39, 27, 4500, 0.16), method="SANN")
$Estimates
              MLE    Std. Dev. Inf. 95% CI Sup. 95% CI
[1,]   42.3689674 0.4339496376  41.5184417  43.2194931
[2,]   28.7335980 0.3804800024  27.9878709  29.4793251
[3,] 4500.1724156          NaN         NaN         NaN
[4,]    0.1640238 0.0006831867   0.1626848   0.1653628

$Measures
      AIC     CAIC      BIC     HQIC        W        A Min(-log(Likelihood))
 549806.7 549806.7 549840.3 549817.5 124.0659 699.3595              274899.4

$`Kolmogorov-Smirnov Test`
 KS Statistic KS p-value
    0.2176616          0

$`Convergence Status`

 "Algorithm Converged"

# Estimate GBG Lomax
> mgbg("lomax",t,starts=c(18,15,15,2600,0.22),method="SANN")
$Estimates
             MLE    Std. Dev.  Inf. 95% CI Sup. 95% CI
[1,]   18.935720  0.518938603   17.9186192   19.952821
[2,]   16.262100  0.896597097   14.5048025   18.019398
[3,]   16.777198  0.271844688   16.2443920   17.310004
[4,] 2600.740736 42.041172400 2518.3415518 2683.139919
[5,]    0.227405  0.004496531    0.2185919    0.236218

$Measures
      AIC     CAIC      BIC     HQIC    W        A Min(-log(Likelihood))
 532911.4 532911.4 532953.4 532924.9 97.8 532.8246              266450.7

$`Kolmogorov-Smirnov Test`
 KS Statistic KS p-value
    0.1369994          0

$`Convergence Status`

 "Algorithm Converged"


# WEIBULL

# Estimate parent distribution WEIBULL

> w1 <- fitdist(t, "weibull")

> summary(w1)
Fitting of the distribution ' weibull ' by maximum likelihood
Parameters :
        estimate   Std. Error
shape    0.70957  0.002313884
scale 1553.04026 12.905435035
```

```
Loglikelihood:  -277243   AIC:  554490.1   BIC:  554506.9
Correlation matrix:
          shape     scale
shape 1.0000000 0.3384988
scale 0.3384988 1.0000000

> gofstat(w1)
Goodness-of-fit statistics
                               1-mle-weibull
Kolmogorov-Smirnov statistic     0.2037876
Cramer-von Mises statistic     414.8799670
Anderson-Darling statistic             Inf

Goodness-of-fit criteria
                                 1-mle-weibull
Aikake's Information Criterion      554490.1
Bayesian Information Criterion      554506.9

# Estimate BG Weibull
> mbetag("weibull",t,starts=c(12,1.9,0.33,85),method="SANN")
$Estimates
            MLE   Std. Dev. Inf. 95% CI Sup. 95% CI
[1,] 12.0544820 0.231170321  11.6013965  12.5075675
[2,]  1.5601865 0.046396458   1.4692511   1.6511219
[3,]  0.3505011 0.004394097   0.3418889   0.3591134
[4,] 83.2989497         NaN         NaN         NaN

$Measures
      AIC     CAIC      BIC   HQIC        W        A Min(-log(Likelihood))
 535949.2 535949.3 535982.8 535960 128.6987 725.1082              267970.6

$`Kolmogorov-Smirnov Test`
 KS Statistic KS p-value
    0.1679213          0

$`Convergence Status`

 "Algorithm Converged"

# Estimate Kumaraswamy generated Weibull
> mkumg("weibull",t,starts=c(5.8,1.8,0.39,100),method="SANN")
$Estimates
            MLE   Std. Dev. Inf. 95% CI Sup. 95% CI
[1,]  9.0012254 0.072532266   8.8590647   9.1433860
[2,]  1.0530009 0.025288673   1.0034360   1.1025658
[3,]  0.4105547 0.003231583   0.4042209   0.4168885
[4,] 99.0624188 0.648342275  97.7916913 100.3331463

$Measures
      AIC     CAIC      BIC     HQIC        W        A Min(-log(Likelihood))
 536535.1 536535.1 536568.7 536545.8 132.0158 743.1873              268263.6

$`Kolmogorov-Smirnov Test`
 KS Statistic KS p-value
     0.178377          0

$`Convergence Status`

 "Algorithm Converged"


# Estimate GBG Weibull
> mgbg("weibull", t, starts=c(8, 12, 3.7, 0.17, 82), method="SANN")
$Estimates
            MLE   Std. Dev. Inf. 95% CI Sup. 95% CI
[1,]  8.8894125  0.45322579   8.0011063   9.7777187
[2,] 11.9226282  0.84962615  10.2573916  13.5878649
[3,]  3.4237726  0.33574023   2.7657338   4.0818114
```

```
[4,]  0.1706026  0.01149232   0.1480781   0.1931271
[5,] 82.6134611 33.31284779  17.3214792 147.9054430

$Measures
    AIC   CAIC      BIC     HQIC        W        A Min(-log(Likelihood))
 536177 536177 536218.9 536190.4 133.1134 751.8321              268083.5

$`Kolmogorov-Smirnov Test`
 KS Statistic KS p-value
    0.1582972          0

$`Convergence Status`

 "Algorithm Converged"
```

 # LOGISTIC

```
# Log(m,s)
# m : location parameter
# s : scale parameter
```

Estimate parent distribution Logistic

```
> l1 <- fitdist(t,"logis")
> summary(l1)
Fitting of the distribution ' logis ' by maximum likelihood
Parameters :
         estimate Std. Error
location 1187.218  13.068382
scale    1563.665   8.186711
Loglikelihood:  -312222.1   AIC:  624448.3   BIC:  624465
Correlation matrix:
           location      scale
location 1.00000000 0.08043332
scale    0.08043332 1.00000000

> gofstat(l1)
Goodness-of-fit statistics
                             1-mle-logis
Kolmogorov-Smirnov statistic   0.3400218
Cramer-von Mises statistic   967.3672742
Anderson-Darling statistic           Inf

Goodness-of-fit criteria
                               1-mle-logis
Aikake's Information Criterion    624448.3
Bayesian Information Criterion    624465.0
```

estimate BG, KwG, BGB when parent generator = logistic, using mle2

```
> library(stats4)
> library(bbmle)
```

estimate beta generated (BG) logistic (c = 1)

```
BGL <- function(m,s,a,b) {
R <- 1/beta(a,b)*dlogis(t, m, s)*(plogis(t,m,s))^(a-1)*(1-plogis(t,m,s))^(b-1)
      -sum(log(R))
}

> bgl1 <- mle2(BGL, start=list(m=12, s=65000, a = 14, b=60), method="SANN")
> bgl1

Call:
mle2(minuslogl = BGL, start = list(m = 12, s = 65000, a = 14,
    b = 60), method = "SANN")
```

```
Coefficients:
          m           s           a           b
   11.40035 64993.47776    62.57944    60.33634

Log-likelihood: -350707.2

> BIC(bgl1)
[1] NA

> AIC(bgl1)
[1] 701422.4

> logLik(bgl1)
'log Lik.' -350707.2 (df=4)

> summary(bgl1)
Maximum likelihood estimation

Call:
mle2(minuslogl = BGL, start = list(m = 12, s = 65000, a = 14,
    b = 60), method = "SANN")

Coefficients:
    Estimate Std. Error    z value     Pr(z)
m 1.1400e+01 3.4555e-04 3.2992e+04 < 2.2e-16 ***
s 6.4993e+04 1.3204e-03 4.9224e+07 < 2.2e-16 ***
a 6.2579e+01 4.9000e-01 1.2771e+02 < 2.2e-16 ***
b 6.0336e+01 4.7237e-01 1.2773e+02 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

-2 log L: 701414.4

# estimate KwG logistic (a=1)

KGL <- function(m,s,b,c) {
       R = c*b*dlogis(t, m, s) * (plogis(t,m,s))^(c-1) * (1-
(plogis(t,m,s))^c)^(b-1)
-sum(log(R))
}

> kgl1 <- mle2(KGL, start=list(m=10, s=65000, b = 40, c=7), method="SANN")
> kgl1

Call:
mle2(minuslogl = KGL, start = list(m = 10, s = 65000, b = 40,
    c = 7), method = "SANN")

Coefficients:
          m           s           b           c
    5.518507 64994.507240    39.292971     5.868022

Log-likelihood: -367654.5

> BIC(kgl1)
[1] NA

> AIC(kgl1)
[1] 735316.9

> logLik(kgl1)
'log Lik.' -367654.5 (df=4)

> summary(kgl1)
Maximum likelihood estimation

Call:
mle2(minuslogl = KGL, start = list(m = 10, s = 65000, b = 40,
```

```
    c = 7), method = "SANN")

Coefficients:
    Estimate Std. Error    z value     Pr(z)
m 5.5185e+00 2.5677e-05 2.1492e+05 < 2.2e-16 ***
s 6.4995e+04 3.8657e-04 1.6813e+08 < 2.2e-16 ***
b 3.9293e+01 4.2154e-01 9.3212e+01 < 2.2e-16 ***
c 5.8680e+00 1.7612e-02 3.3319e+02 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

-2 log L: 735308.9

# estimate GBG logistic
GBGL <- function(m,s,a,b,c) {
R <- c/beta(a,b)*dlogis(t, m, s)*(plogis(t,m,s))^(a*c-1)*(1-
(plogis(t,m,s))^c)^(b-1)
-sum(log(R))
}

> gbgl1 <- mle2(GBGL, start=list(m=10, s=65000, a = 62, b=60, c=1),
method="SANN")
> gbgl1

Call:
mle2(minuslogl = GBGL, start = list(m = 10, s = 65000, a = 62,
    b = 60, c = 1), method = "SANN")

Coefficients:
           m            s            a            b            c
1.000569e+01 6.500143e+04 6.731603e+01 5.868774e+01 9.280428e-01

Log-likelihood: -350649.4

> BIC(gbgl1)
[1] NA

> AIC(gbgl1)
[1] 701308.8

> logLik(gbgl1)
'log Lik.' -350649.4 (df=5)

> summary(gbgl1)
Maximum likelihood estimation

Call:
mle2(minuslogl = GBGL, start = list(m = 10, s = 65000, a = 62,
    b = 60, c = 1), method = "SANN")

Coefficients:
    Estimate Std. Error    z value     Pr(z)
m 1.0006e+01 1.9470e-04 5.1389e+04 < 2.2e-16 ***
s 6.5001e+04 1.6632e-03 3.9082e+07 < 2.2e-16 ***
a 6.7316e+01 1.8709e+00 3.5981e+01 < 2.2e-16 ***
b 5.8688e+01 4.6524e-01 1.2614e+02 < 2.2e-16 ***
c 9.2804e-01 1.9365e-02 4.7923e+01 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

-2 log L: 701298.8
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] R. B. D' Agostino, M. A. Stephens (1986), *Goodness-of-Fit Techniques*, Marcel Dekket, Inc., New York.

[2] A. Akinsete, F. Famoye, and C. Lee (2008). The beta-Pareto distribution, *Statistics*, vol. 42, no. 6, pp. 547-563.

[3] C. Alexander, G. M. Cordeiro, E. M. M. Ortega, J. M. Sarabia (2012). Generalized beta-generated distributions, *Computational Statistics & Data Analysis*, vol. 56, issue 6, pp. 1880-1897.

[4] B. C. Arnold, E. Castillo, J. M. Sarabia (2006). Families of multivariate distributions involving the Rosenblatt construction, *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1652-1662.

[5] W. Barreto-Souza, A. H. S. Santos, and G. M. Cordeiro (2010). The beta generalized exponential distribution, *Journal of Statistical Computation and Simulation*, vol. 80, no. 2, pp. 159-172.

[6] W. Barreto-Souza, G. M. Cordeiro, A. B. Simas (2011). Some results for beta Frechet distribution, *Communications on Statistics – Theory and Methods*, vol. 40, pp. 798-811.

[7] S. Chava, C. Stefanescu, S. Turnbull (2008), Modeling the loss distribution. Available online from http://faculty.london.edu/cstefanescu/research.html

[8] G. M. Cordeiro and M. de Castro (2011). A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, vol. 81, issue 7, pp. 883-898.

[9] N. Eugene, C. Lee, and F. Famoye (2002). Beta-normal distribution and its applications, *Communications in Statistics – Theory and Methods*, vol. 31, pp. 497-512.

[10] I. S. Gradshteyen, I. M. Ryzhik (2007). *Table of Integrals, Series, and Products*, 7th Edition, Academic Press, San Diego.

[11] J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, vol. 15, no. 5, pp. 1049-1054.

[12] A. K. Gupta, S. Nadarajah (2004). On the moments of the beta normal distribution, *Communications on Statistics – Theory and Methods*, vol. 33, no. 1, pp. 1-13.

[13] C. C. Hewitt Jr., B. Lefkowitz (1979). Methods for fitting distributions to insurance loss data, *Proceedings of the Casualty Actuarial Society*, LXVI, pp. 139-160.

[14] J. R. M. Hosking (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society,* vol. 52, no. 1, pp. 105-124.

[15] N. L. Johnson, S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd Ed., Hoboken, NJ: John Wiley and Sons, Inc.

[16] M. C. Jones (2004). Families of distributions arising from distributions of order statistics, *Test*, vol. 13, no. 1, pp. 1-43.

[17] M. C. Jones and P. V. Larsen (2004). Multivariate distributions with support above the diagonal, *Biometrika*, vol. 91, pp. 975-986.

[18] M. C. Jones (2009). Kumaraswamy's distribution: a beta-type distribution with some tractability advantages, *Statistical Methodology*, vol. 6, pp. 70-81.

[19] J. F. Kenney and E. S. Keeping (1962). *Mathematics of Statistics*, Part I, 3rd Ed., Princeton NJ.

[20] S. A. Klugman, H. H. Panjer, G. E. Willmot (2004). *Loss Models: From Data to Decisions*, 2nd Ed., John Wiley & Sons.

[21] P. Kumaraswamy (1980). A generalized probability density for double-bounded random processes, *Journal of Hydrology*, vol. 46, pp. 79-88.

[22] J. M. Landwehr, N. C. Matalas, and J. R. Wallis (1979). Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resources Research*, vol. 15, no. 5, pp. 1055-1064.

[23] R. Lourenzutti, D. Duarte, F. Castellares, and M. Azevedo (2012). The Beta Truncated Pareto Distribution. Available online from http://hdl.handle.net/1843/BUOS-92FNQK

[24] J. B. McDonald (1984). Some generalized functions for the size distribution of income, *Econometrica*, vol. 52, no. 3, pp. 647-663.

[25] J. J. A. Moors (1988). A quantile alternative for kurtosis, *Journal of the Royal Statistical Society, Series D (The Statistician)*, vol. 37, no. 1, pp. 25-32.

[26] A. L. Morais, G. M. Cordeiro (2010). A four-parameter generalized logistic distribution, In: 19º Simpósio Nacional de Probabilidade e Estatística, 2010, São Pedro. Available online from http://www.ime.unicamp.br/sinape/sites/default/files/Lemos_Cordeiro.pdf

[27] G. S. Mudholkar and D. K. Srivastava (1993). Exponentiated Weibull family for analyzing bathtub failure rate data, *IEEE Transactions on Reliability*, vol. 42, no. 2, pp. 299-302.

[28] S. Nadarajah and S. Kotz (2004). The beta Gumbel distribution, *Mathematical Problems in Engineering*, vol. 2004, no. 4, pp. 323-332.

[29] S. Nadarajah and S. Kotz (2006). The beta exponential distribution, *Reliability Engineering and System Safety*, vol. 91, no. 6, pp. 689-697.

[30] P. F. Paranaiba, E. M. M. Ortega, G. M. Cordeiro, R. R. Pescim (2011). The beta Burr XII distribution with application to lifetime data, *Computational Statistics and Data Analysis*, vol. 55, pp. 1118-1136.

[31] M. Razzaghi (2009). Beta-normal distribution in dose-response modeling and risk assessment for quantitative responses, *Environmental and Ecological Science*, vol. 16, pp. 21-36.

[32] G. O. Silva, E. M. M. Ortega, G. M. Cordeiro (2010). The beta modified Weibull distribution, *Lifetime Data Anal*, vol. 16, no. 3, pp. 409-430.

[33] G. Venter (1983). Transformed Beta and Gamma Distributions and Aggregate Losses, *PCAS*, 70, pp. 156-193.

[34] W. Weibull (1951). A statistical distribution function of wide applicability, *Journal of Applied Mechanics*, vol. 18, no. 3, pp. 293-297.

[35] K. Zografos and N. Balakrishnan (2009). On families of beta and generalized gamma generated distributions and associated inference, *Statistical Methodology*, vol. 6, pp. 344-362.

[36] Μ. Κούτρας (2004). *Εισαγωγή στις Πιθανότητες: Θεωρία και Εφαρμογές*, Μέρος Ι, Β΄ Έκδοση, Εκδ. Σταμούλη.

[37] Μ. Κούτρας (2005). *Εισαγωγή στις Πιθανότητες: Θεωρία και Εφαρμογές*, Μέρος ΙΙ, Β΄ Έκδοση, Εκδ. Σταμούλη.

[38] H. Maylor (2003). *Διαχείριση Έργων*, Τρίτη Αγγλική έκδοση, Εκδόσεις Κλειδάριθμος.

[39] P. J. Montana & B. H. Charnov (1993). *Μάνατζμεντ*, Δεύτερη Αμερικάνικη Έκδοση, Εκδόσεις Κλειδάριθμος.

[40] Ι. Χαλικιάς (2010). *Στατιστική: Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις*, 3η Έκδοση, Εκδόσεις Rosili.

[41] M. Spiegel, L. Stephens (2000). *Θεωρία και Προβλήματα Στατιστικής*, 3η έκδοση, εκδόσεις Τζιόλα.

[42] Γ. Παπαδόπουλος, *Η Κανονική Κατανομή*. Σημειώσεις μαθήματος Στατιστικής. Γεωπονικό Πανεπιστήμιο Αθηνών.

[43] Χ. Δαμιανού, Ν. Παπαδάτος, και Χ. Α. Χαραλαμπίδης (2003). *Εισαγωγή στις Πιθανότητες και τη Στατιστική*, Διδακτικές Σημειώσεις, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.