Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ

ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ ΓΙΑ ΤΗΝ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΤΩΝ ΕΙΣΟΔΗΜΑΤΩΝ ΚΑΙ ΑΛΛΩΝ ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΑΝΑΛΟΓΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Μαρία Καλκούνου

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

Σεπτέμβριος 2015

Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ

ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ ΓΙΑ ΤΗΝ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΤΩΝ ΕΙΣΟΔΗΜΑΤΩΝ ΚΑΙ ΑΛΛΩΝ ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΑΝΑΛΟΓΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Μαρία Καλκούνου

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

Σεπτέμβριος 2015

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' αριθμ. ……..

συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του

Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής Μ. Κούτρας (Επιβλέπων)

- Επίκουρος Καθηγητής Γ. Τζαβελάς

- Επίκουρος Καθηγητής Γ. Ψαρράκος

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

U N I V E R S I T Y O F P I R A E U S

DEPARTMENT OF STATISTICS

AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN

ACTUARIAL SCIENCE AND RISK MANAGEMENT

STOCHASTIC MODELS FOR THE DESCRIPTION OF FINANCIAL AND ACTUARIAL DATA

By

Maria Kalkounou

MSc Dissertation

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

Piraeus, Greece

September 2015

Στους γονείς μου Πέτρο και Όλγα και στην αδερφή μου Πάνυ.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μάρκο Κούτρα για την καθοδήγησή του κατά την διάρκεια της εκπόνησης της διπλωματικής μου εργασίας καθώς και τα υπόλοιπα μέλη της Τριμελούς Εξεταστικής Επιτροπής, κύριο Γεώργιο Τζαβελά και κύριο Γεώργιο Ψαρράκο για την συμμετοχή τους στην τριμελή εξεταστική επιτροπή. Ακόμη θα ήθελα να ευχαριστήσω και τους φίλους μου για όλη την βοήθειά τους, την Μαριάννα για τις πολύτιμες συμβουλές της σε όλη την διάρκεια αυτής της εργασίας , τον Δημήτρη για την βοήθεια και στήριξη που μου πρόσφερε, καθώς και την Μαρία για τα οικονομικά στοιχεία που μου πρόσφερε, που χωρίς αυτά δεν θα μπορούσε να υπάρξει πρακτικό κομμάτι.

Περίληψη

Σκοπός της παρούσας μελέτης είναι να δείξει ποια ή ποιες κατανομές είναι οι καλύτερες για την περιγραφή οικονομικών και αναλογιστικών μεγεθών όπως ο πλούτος, το εισόδημα, τα οικονομικά μεγέθη και οι ασφαλιστικές ζημιές. Για τον λόγο αυτό αναλύουμε κάποιες από τις πιο σημαντικές κατανομές την pareto, την λογαριθμοκανονική, την γάμμα, την βήτα και την dagum. Αναλύουμε τα χαρακτηριστικά της καθεμιάς και βρίσκουμε τους εκτιμητές τους με την μέθοδο της μέγιστης πιθανοφάνειας, την μέθοδο των ροπών και των ποσοστημοριών. Επίσης αναφέρουμε για κάθε κατανομή τα πεδία εφαρμογών της όπως είναι σε δεδομένα εισοδήματος, πλούτου, σε μεγέθη επιχειρήσεων και ασφαλιστικές ζημιές. Τέλος χρησιμοποιούμε δεδομένα ζημιών μιας ασφαλιστικής εταιρείας και εφαρμόζοντας την κάθε κατανομή στα δεδομένα αυτά βρίσκουμε ποια κάνει την καλύτερη προσαρμογή και μας δίνει αξιόπιστα αποτελέσματα.

Abstract

The purpose of this study is to show which distributions are the best for describing financial and actuarial sizes such as wealth, income, financials and insurance losses. For this reason, we analyze some of the most important distribution, the Pareto, the Lognormal, the Gamma, Beta and Dagum. We analyze the characteristics of each one and we find the estimators with the method of maximum likelihood, method of moments and the method of percentiles. Also we mention any distribution fields of applications such as data on income, wealth, in business fundamentals and insurance losses. Finally we use data loss from an insurance company and applying any allocation to these data we find what makes the best fit and gives reliable results.

Περιεχόμενα

Κατάλογος Διαγραμμάτων

Κατάλογος Πινάκων

ΚΕΦΑΛΑΙΟ 1

Κατανομές Οικονομικών και Αναλογιστικών Μεγεθών

1.1 Εισαγωγή

Στην παρούσα διπλωματική εργασία θα ασχοληθούμε με την μελέτη στοχαστικών μοντέλων τα οποία μπορούν να χρησιμοποιηθούν για την περιγραφή κατανομών των εισοδημάτων και άλλων οικονομικών και αναλογιστικών μεγεθών. Συγκεκριμένα αυτό ξεκίνησε από το τέλος του 19ου αιώνα, όταν ο Vilfredo Pareto ανέπτυξε ένα έντονο και εντυπωσιακό μοντέλο για την κατανομή του προσωπικού εισοδήματος, μέχρι τα τελευταία μοντέλα που αναπτύχθηκαν μερικά εκατοντάδες χρόνια αργότερα.

Γενικά τα ΄΄Οικονομικά μεγέθη κατανομών΄΄ περιλαμβάνουν τις κατανομές των προσωπικών εισοδημάτων, διαφόρων τύπων, της κατανομής του πλούτου και της κατανομής του μεγέθους των επιχειρήσεων. Επίσης περιλαμβάνουν εργασίες σχετικά με την κατανομή των αναλογιστικών ζημιών για τις οποίες υπάρχουν παρόμοια μοντέλα που ήταν σε χρήση από Σκανδιναβούς αναλογιστές (Meidell 1912, Hagstroem 1925). Είχε παρατηρηθεί ότι αρχικά στην ασφάλιση ζωής, το ασφαλισμένο ποσό είναι πιθανό να είναι ανάλογο με τα εισοδήματα των κατόχων ασφαλιστηρίων συμβολαίων αν και εν τέλει δεν υπάρχει σχεδόν κανένας συντονισμός μεταξύ των δύο περιοχών.

Σχετικά με την κατανομή του εισοδήματος, τον εικοστό αιώνα έγιναν πολλές προσπάθειες από ισχυρά έθνη όπως η Ρωσία (1917) και η Κίνα (1949) και σχεδόν όλες τι; χώρες της Ανατολικής Ευρώπης, για να πραγματοποιήσουν τις εκτεταμένες οικονομικές μεταρρυθμίσεις και τη θέσπιση οικονομικών καθεστώτων που θα μειώσουν δραστικά την εισοδηματική ανισότητα και να οδηγήσουν σε κάτι που να προσεγγίζει το ενιαίο σημείο κατανομής του εισοδήματος, όταν όλοι πληρώνονται με τον ίδιο μισθό. Το πιο ριζικό παράδειγμα είναι φυσικά το σχέδιο για την οικονομία της Λαϊκής Δημοκρατίας της Κίνας (PRC) που διακηρύχθηκε από τον Mao Tse Tung 1η Οκτωβρίου του 1949.

1.2 Ανάγκη περιγραφής της κατανομής εισοδημάτων

Όπως ο Okun (1975) το έθεσε <<το εισόδημα και ο πλούτος είναι δύο πλαισιωμένες επιδόσεις στο βιβλίο ρεκόρ της οικονομικής θέσης των ανθρώπων>>. Είναι για όλους αναμφισβήτητο ότι η κατανομή του μεγέθους του εισοδήματος είναι ζωτικής σημασίας για την αγορά σε σχέση με την κοινωνική και οικονομική πολιτική. Στις οικονομικές και κοινωνικές στατιστικές, η κατανομή μεγέθους του εισοδήματος είναι η βάση της συγκέντρωσης και των καμπυλών Lorenz, συνεπώς και στην καρδιά της μέτρησης της ανισότητας και τη γενικότερης κοινωνικής πρόνοιας αξιολογήσεων. Η κατανομή του εισοδήματος επηρεάζει επίσης την ζήτηση της αγοράς και της ελαστικότητας και κατά συνέπεια την συμπεριφορά των επιχειρήσεων και της ισορροπίας της αγοράς κατά μείζονα λόγο. Συχνά αναφέρεται ότι η κατανομή του εισοδήματος είναι ένας σημαντικός παράγοντας στον καθορισμό του ποσού της αποταμίευσης σε μια κοινωνία. Η γνώση της κατανομής μεγέθους των επιχειρήσεων είναι σημαντική για τους οικονομολόγους για την μελέτη της βιομηχανικής οργάνωσης, σε ρυθμιστικές αρχές της κυβέρνησης, καθώς και για τα δικαστήρια. Από το 2002 τεράστιες ανακατατάξεις στα εταιρικά ιδρύματα που απαιτούν μεγάλες επιχειρήσεις λαμβάνουν χώρα σε όλο τον κόσμο, ιδίως στις Ηνωμένες Πολιτείες και Γερμανία. Αυτό αναμφίβολα θα οδηγήσει σε δραστικές αλλαγές στο εγγύς μέλλον ως προς το μέγεθος της κατανομής των επιχειρήσεων και οι πρόσφατες συχνές συγχωνεύσεις και περιστασιακές κατανομές των επιχειρήσεων μπορεί να απαιτήσουν ακόμη και μια νέα μεθοδολογία.

Όσον αφορά τις ασφαλιστικές ζημιές είναι απαραίτητη η εκτίμηση κατανομών πιθανότητας που θα μπορούν να περιγράψουν τις διαδικασίες ζημιών που καλύπτονται από ασφαλιστήρια συμβόλαια. Για παράδειγμα, για να είναι σωστό το ασφάλιστρο που χρεώνεται σ' ένα συγκεκριμένο συμβόλαιο, θα πρέπει να βασίζεται στην υποκείμενη διαδικασία απώλειας για το συμβόλαιο. Πρακτικά είναι αδύνατον να γνωρίζουν την πραγματική διαδικασία, αλλά μια επαρκώς ακριβή εκτίμησή της μπορεί να παρέχει την βάση για ένα ικανοποιητικά ακριβές ασφάλιστρο. Κάποιος μπορεί να συζητήσει την διαδικασία απώλειας για έναν μεμονωμένα ασφαλισμένο με μία μόνο κάλυψη που παρέχεται από ένα ενιαίο συμβόλαιο, ή για μια ομάδα ασφαλισμένων με πολλαπλές καλύψεις που παρέχονται από πολλά συμβόλαια (βλ. [14]).

1.3 Αναζήτηση κατάλληλων κατανομών

Στις 10 Ιανουαρίου του 2000 ο James D. Wolfensohn, Πρόεδρος του Ομίλου της Παγκόσμιας Τράπεζας, εξέφρασε την πεποίθησή του ότι η <<καταπολέμηση της φτώχειας>> είναι <<το αληθινό κλειδί για την ασφάλεια και την ειρήνη>>.

Στη συνέχεια προχώρησε σε συζητήσεις για την κατανομή του εισοδήματος, θέματα τα οποία μαστίζουν την πλειοψηφία των αναπτυσσόμενων χωρών. Αυτό το ζήτημα έχει παρατηρηθεί σε όλη την ανθρώπινη ιστορία. Όπως οι αρχαίοι Έλληνες και Ρωμαίοι φιλόσοφοι απευθύνθηκαν σε πολλά συγγράμματα (δηλαδή ο Αριστοτέλης) έτσι και οι μεγάλοι στοχαστές του Μεσαίωνα και της περιόδου του Διαφωτισμού τους ακολούθησαν (βλ. [38]).

Με την ανάπτυξη των ποσοτικών μεθόδων, οι επιστήμονες από διάφορους επιστημονικούς κλάδους έχουν διερευνήσει το θέμα. Αυτή η ποσοτική προσέγγιση προσέφυγε στους μαθηματικούς, στατιστικούς και οικονομολόγους.

Το ερώτημα λοιπόν που γεννάται είναι το εξής : *Ποια μορφή συνάρτησης κατανομής πιθανότητας ταιριάζει πιο καλά στην περιγραφή κατανομής εσόδων;*

Με την πάροδο των χρόνων, το ερώτημα αυτό έχει μελετηθεί εκτενώς στα οικονομικά, καθώς και σε άλλες κοινωνικές επιστήμες. Σ' ένα από τα πρώτα συγγράμματα του Pareto που ασχολήθηκε εκτενώς με την ποσοτική εκτίμηση των κατανομών του εισοδήματος εμφανίστηκε το 1896 και 1897. Σε αυτές τις δημοσιεύσεις, ο Pareto καθόρισε μια νέα συνάρτηση πυκνότητας πιθανότητας, η οποία είναι σήμερα γνωστή ως κατανομή Pareto και πρότεινε την χρησιμοποίησή της ως αναπαράσταση της κατανομής εισοδήματος και πλούτου. Καθώς η έρευνα σε αυτόν τον τομέα συνεχιζόταν, νέες κατανομές ανακαλύφθηκαν και προτάθηκαν για την εκτίμηση κατανομών εισοδήματος. Για παράδειγμα το 1969, οι Brown και Aitchinson πρότειναν την χρήση της λογαριθμοκανονικής συνάρτησης πυκνότητας πιθανότητας. Υποστήριξαν ότι η λογαριθμοκανονική κατανομή είναι κατάλληλη για την κατανομή του εισοδήματος επειδή πληρεί έναν αριθμό συγκεκριμένων κριτηρίων.

Για παράδειγμα η συναρτησιακή μορφή της λογαριθμοκανονικής συνάρτησης πυκνότητας μπορεί εύκολα να προκύψει με βάση τις προβλέψεις της οικονομικής θεωρίας. Επιπλέον επειδή η λογαριθμοκανονική είναι μια κατανομή δύο παραμέτρων είναι σχετικά εύκολο να εκτιμήσει αυτές τις παραμέτρους και να παρέχει την οικονομική τους ερμηνεία. Ωστόσο, κατά τα επόμενα έτη, η έρευνα σχετικά με την κατανομή του εισοδήματος συνέχισε την εξέλιξή της. Το 1970, ο Thurow πρότεινε τη χρήση της κατανομής Βήτα Τύπου Ι, μια τριών παραμέτρων κατανομή ως μοντέλο για την κατανομή εισοδήματος. Κατά την εκτίμησή του, ο Thurow εξέτασε στοιχεία για τα εισοδήματα των αμερικάνικων οικογενειών του 1949 και χρησιμοποίησε ένα ανώτερο όριο της τάξεως των 15.000 δολλαρίων, λόγω του γεγονότος ότι ως επί το πλείστον ενδιαφερόταν για την εκτίμηση του εισοδήματος του χαμηλότερου άκρου. Βρήκε ότι η κατανομή Βήτα ταιριάζει με τις πραγματικές κατανομές εισοδήματος αρκετά καλά.

Αργότερα, οι Salem και Mount χρησιμοποίησαν την κατανομή Γάμμα (1974) για τον ίδιο σκοπό. Εφάρμοσαν μια δύο παραμέτρων κατανομή Γάμμα, η οποία είναι μια οριακή περίπτωση της κατανομής Βήτα και κατέληξαν στο συμπέρασμα ότι η κατανομή Γάμμα είναι μια καλύτερη συνάρτηση πυκνότητας πιθανότητας για την κατανομή εισοδήματος. Μετά από έναν μεγάλο αριθμό μελετών, κατέληξαν ότι υπάρχουν εναλλακτικές μέθοδοι που χρησιμοποιούν τις γενικεύσεις αυτών των κατανομών και παρέχουν σημαντικά καλύτερη εφαρμογή για την ανάλυση της κατανομής του εισοδήματος. Σε μια συγκεκριμένη μελέτη ο McDonald πρότεινε την χρήση της γενικευμένης κατανομής Βήτα τύπου Ι και ΙΙ ως κατάλληλη συνάρτηση πυκνότητας πιθανότητας για την κατανομή του εισοδήματος (1984). Πρότεινε ότι μιας και αυτές οι κατανομές περιλάμβαναν την κατανομή Βήτα τύπου Ι και ΙΙ, την Γάμμα, την Singh-Maddala, την λογαριθμική και κάποιες άλλες κατανομές, τότε αυτές θα εφαρμόζουν εξίσου καλά με τις κατανομές τύπου Ι Βήτα, τύπου ΙΙ Βήτα, Γάμμα και λογαρθμική.

Σε μια μεταγενέστερη μελέτη, οι McDonald και Xu (1995) έδειξαν ότι η γενικευμένη κατανομή Βήτα που περιλαμβάνει όλες τις παραπάνω κατανομές ως ειδικές περιπτώσεις είναι πολύ πιο ευέλικτη και έτσι προσαρμόζεται καλύτερα στις κατανομές εισοδήματος. Οι Bordley και McDonald (1996) εξέτασαν την σχετική απόδοση ορισμένων κατανομών. Σε αυτήν την έρευνα, εφάρμοσαν ένα φάσμα κατανομών σε δεδομένα οικογενειακού εισοδήματος των ΗΠΑ που λαμβάνονταν επί πέντε χρόνια. Στο τέλος αυτής της μελέτης, παρατήρησαν ότι από τα μοντέλα που εξέτασαν, η γενικευμένη κατανομή Βήτα τύπου ΙΙ είναι το μοντέλο που παρέχει την καλύτερη εφαρμογή σε κατανομές εισοδήματος (1996).

Βάσει λοιπόν των μελετών που έχουν γίνει όλα αυτά τα χρόνια, θα αναφέρουμε παρακάτω κάποιες από τις πιο σημαντικές κατανομές που υπάρχουν για την μελέτη γενικότερα των οικονομικών και αναλογιστικών δεδομένων, θα αναλύσουμε τα χαρακτηριστικά της κάθε κατανομής αλλά και θα τις συγκρίνουμε και μεταξύ τους εφαρμόζοντάς τες σε πραγματικά δεδομένα.

ΚΕΦΑΛΑΙΟ 2

Κατανομή Pareto

2.1 Ορισμός

Η κατανομή Pareto πήρε το όνομά της από τον Ιταλό οικονομολόγο και φιλόσοφο Vilfredo Pareto, ο οποίος έκανε αρκετές σημαντικές συνεισφορές για την οικονομία ιδιαίτερα στην μελέτη της κατανομής του εισοδήματος και ήταν ο πρώτος που ανακάλυψε ότι το εισόδημα ακολουθεί μια κατανομή Pareto. Έδωσε τον ακόλουθο ορισμό :

Εάν Χ είναι μια τυχαία μεταβλητή η οποία ακολουθεί την κατανομή Pareto (Τύπου Ι) τότε η πιθανότητα ότι το Χ είναι μεγαλύτερο από κάποιον αριθμό x, δίνεται από τον παρακάτω τύπο :

$$\bar{F}(\mathrm{x}) = \Pr(\mathrm{X} > \mathrm{x}) = \begin{cases} (\frac{x_0}{x})^{\alpha} & x \geq x_0 \\ 0 & x \geq x_0 \end{cases}$$

όπου x_0 είναι η ελάχιστη δυνατή τιμή της Χ και α είναι μια θετική παράμετρος ή αλλιώς δείκτης ουράς.

Όταν η κατανομή αυτή χρησιμοποιείται για να διαμορφώσει την κατανομή του πλούτου, ονομάζεται δείκτης Pareto.

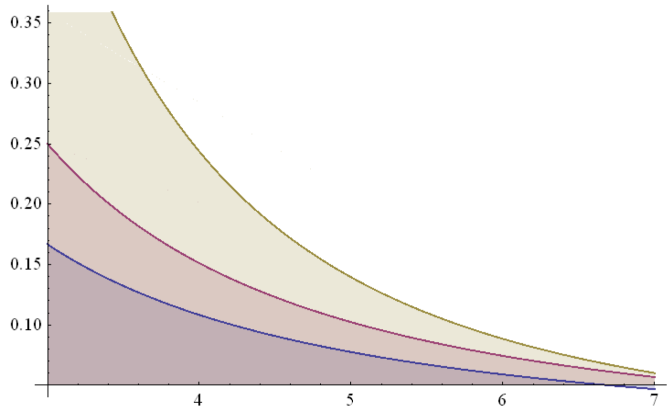
Επομένως η κλασική κατανομή Pareto ορίζεται ως

$$\mathrm{F(x)} = 1\text{-}(\frac{x}{x_0})^{-\alpha} \quad , \quad x \quad \geq x_0 > 0 \qquad (2.1)$$

και η συνάρτηση πυκνότητας είναι

$$\mathrm{f(x)} = \frac{\alpha {x_0}^{\alpha}}{x^{\alpha+1}} \quad , \quad x \quad \geq x_0 > 0 \qquad (2.2)$$

Στο **Διάγραμμα 2.1** απεικονίζεται η συνάρτηση πυκνότητας πιθανότητας της κατανομής Pareto.



Διάγραμμα 2.1.

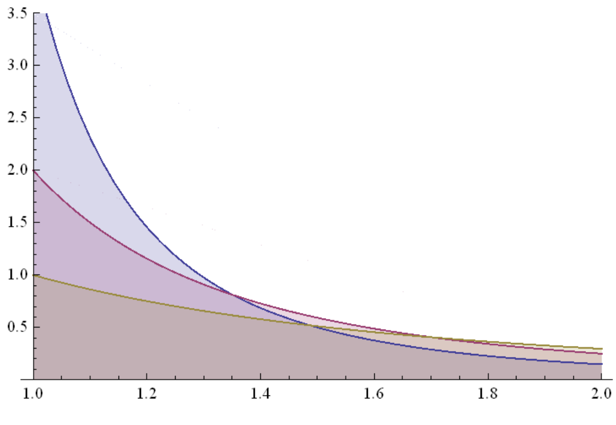
Συνάρτηση πυκνότητας πιθανότητας της Pareto για α = (0.17, 0.25, 0.75, 1.5)

Στο τέλος του 19ου αιώνα ο Vilfredo Pareto (1895-1897) πρότεινε τρεις παραλλαγές της κατανομής του. Η πρώτη ήταν η τύπου I (2.1), η δεύτερη ήταν η (2.3)

$$F(x) = 1- (1+\frac{x-\mu}{x_0})^{-\alpha} \quad , x \geq \mu \tag{2.3}$$

όπου αλλιώς ονομάζεται και τριών παραμέτρων κατανομή Pareto και η τρίτη ήταν η ειδική περίπτωση για μ=0 (2.4) ή αλλιώς κατανομή Pareto τύπου II

$$F(x) = 1- (1+\frac{x}{x_0})^{-\alpha} \quad , x \geq 0 \tag{2.4}$$



Διάγραμμα 2.2.

Συνάρτηση πυκνότητας πιθανότητας κατανομής Pareto τύπου II

2.2 Βασικά χαρακτηριστικά

Όπως αναφέραμε και στον ορισμό, α είναι ο δείκτης ουράς. Χαρακτηριστικό του δείκτη είναι ότι η ουρά είναι βαρύτερη όσο το α είναι μικρότερο, που σημαίνει ότι υπάρχουν μόνο ροπές μικρής τάξης. Συγκεκριμένα η κ-ροπή της κατανομής Pareto υπάρχει μόνο αν $\kappa < \alpha$ και ισούται με (βλ. [8])

$$\mathrm{E}(X^{\kappa}) = \frac{\alpha {x_0}^{\kappa}}{\alpha-\kappa} \quad , \tag{2.5}$$

και για κ=1 έχουμε την μέση τιμή

$$\mathrm{E}(\mathrm{X}) = \frac{\alpha x_0}{\alpha-\kappa} = \frac{\alpha x_0}{\alpha-1} \tag{2.6}$$

Ο τύπος της διακύμανσης είναι ο ακόλουθος

$$\mathrm{Var}(\mathrm{X}) = \frac{\alpha {x_0}^2}{\alpha(\alpha-1)^2(\alpha-2)} \ , \tag{2.7}$$

με συντελεστή διακύμανσης

$$\mathrm{CV} = \frac{1}{\sqrt{\alpha(\alpha-2)}} \quad , \tag{2.8}$$

Η ροπογεννήτρια της κατανομής Pareto έχει τον εξής τύπο

$$\mathrm{M}(\mathrm{t},\alpha,x_0) = \mathrm{E}(e^{tX}) = \alpha\,(-x_0\mathrm{t})^{\alpha}\,\Gamma(-\alpha,-x_0\mathrm{t}) \tag{2.9}$$

και ορίζεται μόνο για αρνητικές τιμές του t, δηλαδή $t \leq 0$.

2.3 Εκτίμηση παραμέτρων

Για την εκτίμηση των χαρακτηριστικών παραμέτρων Pareto ασχολήθηκαν οι Arnold (1983), Kotz, Johnson, Balakrishnan (βλ. [8]). Οι εκτιμητές παραμέτρων οι οποίοι θα αναφερθούν αναλυτικότερα παρακάτω είναι οι MLE (Maximum Likelihood Estimation-Εκτίμηση Μέγιστης Πιθανοφάνειας), Method of moments (Μέθοδο των ροπών) και οι Quantile methods (Μέθοδο ποσοστημορίων).

a. Εκτίμηση Μέγιστης Πιθανοφάνειας

Η συνάρτηση πιθανοφάνειας L για την Pareto κατανομή έχει την ακόλουθη μορφή (βλ. [24])

$$L(k,\alpha;\mathbf{x})=\prod_{i=1}^{n}\frac{\alpha k^{\alpha}}{x_i^{\alpha+1}}, \quad 0< k \leq \min\{x_i\}, \alpha > 0 \qquad (2.10)$$

Είναι σημαντικό να τονιστεί ότι στην ουσία η συνάρτηση πιθανότητας, μας δείχνει πόσο πιθανό είναι να έχουμε παρατηρήσει τα δεδομένα που στην πραγματικότητα έχουν παρατηρηθεί. Η μέγιστη συνάρτηση πιθανότητας εκτιμά τις τιμές των k και α που κάνουν όσο το δυνατόν μεγαλύτερο το L με τα δεδομένα που έχουμε. Επίσης η τιμή του k δεν μπορεί να είναι μεγαλύτερη από την μικρότερη τιμή του x. Έτσι μπορούμε να μεγιστοποιήσουμε το L βρίσκοντας τον λογάριθμό του (βλ. [37]).

$$\ln L(k,\alpha;\mathbf{x})=\sum_{i=1}^{n}\ln\left(\frac{\alpha k^{\alpha}}{x_i^{\alpha+1}}\right)=n\ln(\alpha)+\alpha\, n\ln(k)-(\alpha+1)\sum_{i=1}^{n}\ln(x_i)$$

$$\Rightarrow \frac{d}{d\alpha}=n/\alpha+n\ln(k)-\sum_{i=1}^{n}\ln(x_i)$$

Θέτοντας το παραπάνω ίσο με το μηδέν και λύνοντας ως προς α βρίσκουμε τον εκτιμητή

$$\widehat{\alpha}=n/\sum_{i=1}^{n}\ln\left(\frac{x_i}{\hat{k}}\right) \qquad (2.11)$$

β. Μέθοδος ροπών

Εκτιμάμε το α εξισώνοντάς το με την δειγματική μέση τιμή $\bar{x}$. Έτσι προκύπτει

$$\hat{\alpha}=\frac{\bar{X}}{\bar{X}-\hat{k}} \qquad (2.12)$$

όπου το $\widehat{k}$ είναι κάποια εκτίμηση του k. Επιπλέον η εκτίμηση του k από δείγμα n παρατηρήσεων επιτυγχάνεται ως ακολούθως: η πιθανότητα μια παρατήρηση να είναι καλύτερη του x είναι $(k/x)^{\alpha}$. Επίσης η πιθανότητα όλες οι τιμές του δείγματος $x_1,\ldots,x_n$ να είναι καλύτερες του x είναι $(k/x)^{\alpha n}$. Επομένως αυτή είναι η πιθανότητα η τιμή του μικρότερου δείγματος να είναι μεγαλύτερη από το x.

γ. Μέθοδος ποσοστημορίων

Έχουμε πιθανότητες p_1 και p_2 και τις καθορίζουμε με τα αντίστοιχα ποσοστημόρια x_1 και x_2

$$p_1=1-\left(\frac{k}{x_1}\right)^{\alpha} \qquad (2.13)$$

$$p_2=1-\left(\frac{k}{x_2}\right)^{\alpha} \qquad (2.14)$$

τότε παίρνουμε έναν εκτιμητή του α λύνοντας την παρακάτω εξίσωση

$$\hat{\alpha}=\frac{ln\frac{1-p_1}{1-p_2}}{ln\frac{X_1}{X_2}}$$

Όπου αν το αντικαταστήσουμε στις εξισώσεις (2.13) και (2.14) μας δίνει τον εκτιμητή του k (βλ. [37]).

2.4 Γενικευμένη κατανομή Pareto

Η Γενικευμένη κατανομή Pareto εισήχθηκε από τον Pikands (1975) και μελετήθηκε περαιτέρω από τους Davison, Smith (1984), Castillo (1997,2008) και άλλους. Ο γενικός ορισμός της είναι ο ακόλουθος :

Η συνάρτηση πυκνότητας πιθανότητας για την Γενικευμένη κατανομή Pareto με παράμετρο της μορφής $k\neq0$, *μέγεθος παραμέτρου σ και όριο θ δίνεται από τον τύπο*

$$y=f(x\backslash k,\sigma,\theta)=\left(\frac{1}{\sigma}\right)\left(1+k\frac{x-\theta}{\sigma}\right)^{-1-1/k} \quad (2.15)$$

για $\theta<x$, *όταν* $k>0$ *ή για* $\theta<x<\theta-\sigma/k$ *όταν* $k<0$

Για k=0 η πυκνότητα είναι

$$y = f(x\backslash 0,\sigma,\theta)=\left(\frac{1}{\sigma}\right)e^{-\frac{(x-\theta)}{\sigma}} \qquad \text{για } \theta<x \quad (2.16)$$

Αν $k=0$ και $\theta=0$, η γενικευμένη κατανομή Pareto είναι ισοδύναμη με την εκθετική κατανομή, αν $k>0$ και $\theta=\frac{\sigma}{k}$ τότε η γενικευμένη κατανομή Pareto είναι ισοδύναμη με την κατανομή Pareto και όταν $k=1$ είναι ισοδύναμη με την ομοιόμορφη.

2.5 Πεδία εφαρμογής

Την εφαρμογή της κατανομής Pareto την συναντάμε σε διάφορους τομείς όπως στο χρηματιστήριο για τις τυποποιημένες αποδόσεις τιμών σε μεμονωμένες μετοχές, στην υδρολογία, όπου εφαρμόζεται σε ακραία γεγονότα όπως η ετήσια μέγιστη βροχόπτωση μιας ημέρας, στην φυσική, για τα μεγέθη των σωματιδίων άμμου και άλλα πολλά.

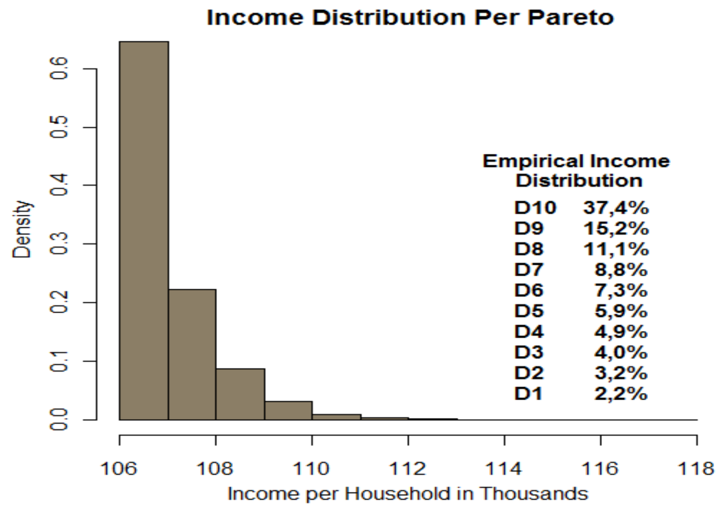
Στην παρούσα διπλωματική εργασία θα αναλύσουμε περαιτέρω την χρήση της κατανομής Pareto στα εξής πεδία: σε δεδομένα εισοδήματος (***income data***), πλούτου (***wealth data***), μεγεθών επιχειρήσεων (***firm sizes***) και ασφαλιστικών ζημιών (***insurance losses***).

α. Δεδομένα εισοδήματος

Οι κατανομές Pareto είναι χρήσιμες στην μοντελοποίηση και στην πρόβλεψη εργαλείων σε ένα ποικίλο φάσμα κοινωνικοοικονομικών συνθηκών και ιδιαίτερα είναι σαφές ότι στο επίκεντρο των συζητήσεων βρίσκεται το μέγεθος της κατανομής του εισοδήματος. Ο Pareto παρατήρησε ότι σε πολλούς πληθυσμούς ο αριθμός των ατόμων του πληθυσμού των οποίων το εισόδημα υπερβαίνει ένα δεδομένο επίπεδο x είχε προσεγγιστεί καλά από Cx-α για κάποιο πραγματικό C και α > 0. Έτσι έγινε προφανές ότι μια τέτοια προσέγγιση είναι δεκτή μόνο για μεγάλες τιμές του x.

Εκτός από τον Pareto, οι Ratz και van Scherrenberg (1981) κατέληξαν σ' ένα αρκετά έξυπνο αποτέλεσμα. Την περίοδο 1955-1978 αποφάσισαν να μελετήσουν την κατανομή των εισοδημάτων που είχαν καταχωρηθεί από επαγγελματίες μηχανικούς, σε όλους τους κλάδους, στην επαρχία του Οντάριο στον Καναδά. Έτσι μοντελοποίησαν την σχέση μεταξύ των παραμέτρων της κατανομής και των χρόνων εμπειρίας των μηχανικών με τεχνικές παλινδρόμησης και κατέληξαν σε μία αρνητική σχέση μεταξύ της μεταβλητής και της παραμέτρου α που σημαίνει ότι τα εισοδήματα των πιο έμπειρων μηχανικών έχουν μεγαλύτερο εύρος από αυτά των μηχανικών που βρίσκονταν στο ξεκίνημα της καριέρας τους.

Μία ακόμα μελέτη έγινε και από τους Cowell, Ferreira και Litchfield, οι οποίοι ασχολήθηκαν με μία ιδιάζουσα περίπτωση, αυτής του εισοδήματος της Βραζιλίας και αυτό γιατί παρουσιάζει μία από τις πιο άνισες κατανομές στον κόσμο, με το 51% του συνολικού εισοδήματος, το 10% ανήκει στους πλουσιότερους και μόνο το 2,1% πάει στους φτωχότερους 20% (1995). Με την εφαρμογή της Pareto σε εισοδήματα πάνω από 1000$ αποδείχθηκε ότι η ανισότητα μεταξύ των πολύ πλούσιων δεν ήταν πολύ ακραία το 1981 με το α να είναι στην περιοχή του 3,αργότερα όμως επιδεινώθηκε και την δεκαετία του 1980 μειώθηκε στο 2 (βλ. [8]).



Διάγραμμα 2.3

Κατανομή εισοδήματος για την Pareto

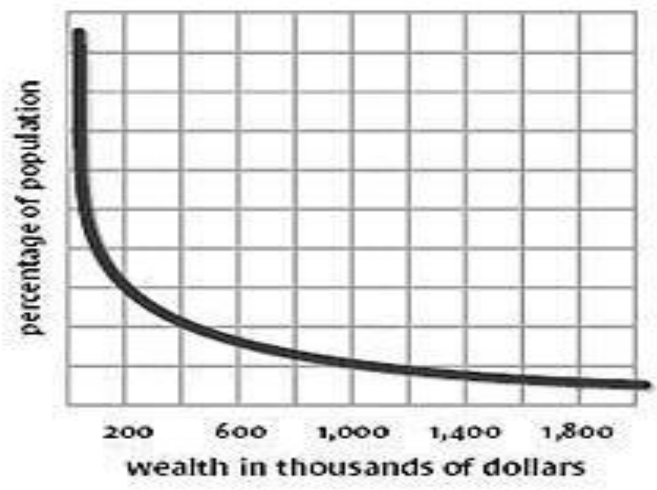
Πηγή: https://moorishwanderer.wordpress.com/tag/income-distribution/

Στο διάγραμμα **2.3** έχουμε την κατανομή του πραγματικού εισοδήματος στο Μαρόκο. Παρατηρούμε ότι το μέσο εισόδημα ανά νοικοκυριό είναι 106.000 και το 30% των νοικοκυριών πάνω από αυτήν την γραμμή καταλαμβάνει το 60% του συνολικού εγχώριου εισοδήματος. Επιβεβαιώνεται λοιπόν ότι τα εισοδήματα στο Μαρόκο ακολουθούν την κατανομή Pareto, το οποίο φαίνεται από την άνιση κατανομή του διαγράμματος.

β. Δεδομένα πλούτου

Ο Pareto χρησιμοποίησε αρχικά την κατανομή αυτή για να περιγράψει την κατανομή του πλούτου μεταξύ των ατόμων, δεδομένου ότι φαίνεται να δείχνουν αρκετά καλά τον τρόπο ότι ένα μεγαλύτερο μέρος του πλούτου της κάθε κοινωνίας ανήκει σε ένα μικρότερο ποσοστό των ανθρώπων σε αυτή την κοινωνία. Αυτή η διαπίστωση είναι ο Κανόνας 80-20 της διαχείρισης του χρόνου, ο οποίος εδράζεται στην αποκαλούμενη **Αρχή του Pareto.** Η Αρχή του Pareto, στο ευρύτερο πλαίσιό της, υποστηρίζει ότι 80% των αποτελεσμάτων προκύπτουν από 20% των μέσων ή αιτίων. Με άλλα λόγια, ο Κανόνας 80-20 σημαίνει ότι, σε κάθε κατάσταση, λίγοι παράγοντες (20%) είναι ζωτικοί και πολλοί (80%) είναι επουσιώδεις.

Ωστόσο, ο κανόνας 80-20 αντιστοιχεί σε μια συγκεκριμένη τιμή του α, και στην πραγματικότητα, τα στοιχεία του Pareto στους Βρετανικούς φόρους εισοδήματος του Cours d'économie politique δείχνει ότι περίπου το 30% του πληθυσμού είχε περίπου το 70% του εισοδήματος. (Σημαντικό είναι να σημειωθεί ότι η κατανομή Pareto δεν είναι ρεαλιστική για τον πλούτο για το κάτω άκρο. Στην πραγματικότητα, η καθαρή θέση μπορεί να είναι και αρνητική). Η κατανομή αυτή δεν περιορίζεται στην περιγραφή του πλούτου ή του εισοδήματος, αλλά σε πολλές περιπτώσεις στις οποίες η ισορροπία βρίσκεται στην κατανομή από το «μικρό» στο «μεγάλο».



Διάγραμμα 2.4

Ο πλούτος σε χιλιάδες δολλάρια

Πηγή : http://hbr.org/2002/04/wealth-happens/ar/

Στο διάγραμμα **2.4** απεικονίζεται ο πλούτος των νοικοκυριών των Ηνωμένων Πολιτειών, τα στοιχεία του οποίου έχουν ληφθεί από την απογραφή που διεξήχθη το 1998. Συγκεκριμένα δείχνει μια κατανομή των πλούσιων και των φτωχών που σχηματίζουν μια καμπύλη Pareto. Το υψηλότερο ποσοστό των νοικοκυριών πέφτει στα χαμηλότερα επίπεδα πλούτου, αλλά στο υψηλότερο σημείο η καμπύλη πέφτει σχετικά αργά εμφανίζοντας ´´fat-tailed´´ κατά το μοτίβο του Pareto.

γ. Μεγέθη επιχειρήσεων

Εμπειρικές μελέτες έχουν δείξει ότι η κατανομή του μεγέθους των επιχειρήσεων μπορεί να καταγραφεί από μία κατανομή Pareto. Ο νόμος Pareto είναι μια πολύ γνωστή ιδιότητα της κατανομής του μεγέθους των επιχειρήσεων. Αυτό μας λέει ότι η συχνότητα των επιχειρήσεων σε έναν πληθυσμό πάνω από ένα ορισμένο μέγεθος είναι αντιστρόφως ανάλογη προς το μέγεθος της επιχείρησης. Μια σειρά από μελέτες έχουν δοκιμάσει εμπειρικά αυτήν την υπόθεση και διατύπωσαν μοντέλα ικανά να παράγουν κατανομές όπως η Pareto. Οι μελέτες αυτές έχουν πολύ μεγάλη επιρροή στην βιομηχανική οικονομία και αλλού.

Συγκεκριμένα ο Steindl (1965) μελετώντας την κατανομή μεγέθους των επιχειρήσεων όρισε τους συντελεστές Pareto στο εύρος 1,0 και 1,5. Για όλες τις εταιρίες του Ηνωμένου Έθνους το 1931 και 1955, η παράμετρος α είναι περίπου ίση με 1,1. Ο Engwall (1968) μελέτησε τις μεγαλύτερες επιχειρήσεις (σύμφωνα με τις πωλήσεις) το 1965 ανάμεσα σε πέντε περιοχές : την Ευρώπη, την Σκανδιναβία, την Σουηδία, τις Ηνωμένες Πολιτείες αλλά και έξω από αυτές. Σε όλες τις περιπτώσεις χρησιμοποίησε την παράμετρο α μεταξύ 1 και 2 (βλ. [34]).

δ. Ασφαλιστικές ζημιές

Για την μελέτη των ασφαλιστικών ζημιών χρειαζόταν ένας τύπος μοντέλου με όχι πολύ λίγες παραμέτρους γιατί δεν θα μπορούσε να περιγράψει επαρκώς την κατανομή, αλλά όχι και με πολλές διότι θα ήταν δύσκολο να δουλέψουμε με αυτό. Στην βιβλιογραφία συναντάμε πολλές προσπάθειες να χρησιμοποιηθούν διάφορα μοντέλα για την περιγραφή των ζημιών, με επικρατέστερη την κατανομή Pareto τύπου II που αλλιώς ονομάζεται κατανομή Lomax ή Pearson τύπου VI. Απ' όλες τις ασφαλιστικές ζημιές η κατανομή Pareto είναι η πλέον κατάλληλη για την περιγραφή ζημιών πυρός. Συγκεκριμένα την περίοδο 1948-1952 οι Benckert και Sternberg χρησιμοποίησαν την κατανομή Pareto για την περιγραφή ζημιών που προκλήθηκαν από φωτιά σε τέσσερις τύπους σουηδικών σπιτιών, με παράμετρο α κοντά στο 0,5. Αργότερα χρησιμοποιήθηκε από τον Andersson στις Βόρειες χώρες (Φινλανδία, Νορβηγία και Σουηδία) με συντελεστές Pareto που είχαν εύρος 1,25-1,76.

ΚΕΦΑΛΑΙΟ 3

Κατανομή Lognormal

3.1 Ορισμός

Στην θεωρία των πιθανοτήτων, η κατανομή log-normal είναι η κατανομή μιας τυχαίας μεταβλητής της οποίας ο λογάριθμος κατανέμεται κανονικά. Εάν το Y είναι μια τυχαία μεταβλητή με κανονική κατανομή, τότε το X=exp(Y) έχει κατανομή log-normal. Ομοίως, εάν το X είναι λογαριθμική κανονική κατανομή, τότε το Y=ln(X) είναι κανονικά κατανεμημένο.

Η log-normal γράφεται και ως log normal αλλά και lognormal. Μερικές φορές αναφέρεται και ως **κατανομή Galton** ή κατανομή του Galton από τον στατιστικό Francis Galton.

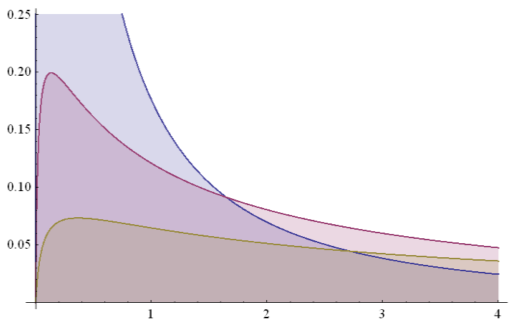
Μια τυχαία μεταβλητή X λέγεται ότι ακολουθεί την lognormal κατανομή με παραμέτρους μεr και σє(0,∞) αν ln(X) έχει την κανονική κατανομή με μέση τιμή μ και τυπική απόκλιση σ. Η lognormal κατανομή χρησιμοποιείται για να μοντελοποιήσει συνεχείς τυχαίες μεταβλητές, όταν η κατανομή θεωρείται ότι είναι ασύμμετρη, όπως ορισμένα εισοδήματα και μεταβλητές που αφορούν τη διάρκεια ζωής.

Η συνάρτηση κατανομής F της lognormal δίνεται από τον παρακάτω τύπο

$$\mathrm{F(x)} = \Phi\left[\frac{\ln(x)-\mu}{\sigma}\right],\ \mathrm{x}\epsilon(0,\infty) \qquad (3.1)$$

και η συνάρτηση πυκνότητας πιθανότητας με παραμέτρους μ και σ δίνεται από την έκφραση

$$\mathrm{f(x)}=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{[\ln(x)-\mu]^2}{2\sigma^2}\right),\quad \mathrm{x}\epsilon(0,\infty) \qquad (3.2)$$



Διάγραμμα 3.1

Συνάρτησης πυκνότητας πιθανότητας της Lognormal κατανομής

Εκτός από την κλασική κατανομή log-normal συναντάμε επιπλέον και την **κατανομή lognormal τριών παραμέτρων**. Η κατανομή log-normal τριών παραμέτρων χρησιμοποιείται συχνά στην υδρολογική ανάλυση των ακραίων πλημμύρων, στις βροχοπτώσεις έντονης διάρκειας και έντασης κ.λ.π. . Γι' αυτό τον λόγο θα δώσουμε μόνο τον τύπο της χωρίς να κάνουμε περαιτέρω ανάλυση.

$$F(x) = \frac{1}{(x-\lambda\sqrt{2})\pi\sigma}\exp\{-\frac{1}{2\sigma^2}[\ln(x-\lambda)-\mu]^2\},\ x > \lambda \qquad (3.3)$$

Εφόσον υπάρχει λ ϵ IR ώστε Z=log(X - λ) να ακολουθεί κανονική κατανομή.

3.2 Βασικά χαρακτηριστικά

Αν $Y \sim LN(\mu,\sigma^2)$ τότε

$$E(Y^r) = E(e^{rX}) = E(e^{\sigma r\left(\frac{X-\mu}{\sigma}\right)+r\mu}) = e^{r\mu}E(e^{\sigma r Z}) = e^{r\mu}e^{(\sigma r)^2/2} = e^{r\mu+\frac{1}{2}r^2\sigma^2},\quad r = 1,2,\ldots$$

απ' όπου προκύπτει άμεσα ότι η μέση τιμή είναι

$$E(T) = e^{\mu+\frac{1}{2}\sigma^2} \qquad (3.4)$$

και η ροπή δεύτερης τάξης

$$E(T^2) = e^{2\mu+2\sigma^2}$$

Επομένως

$$V(T) = E(T^2) - E(T)^2 = e^{2\mu+2\sigma^2} - (e^{\mu+\frac{1}{2}\sigma^2})^2 = e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \qquad (3.5)$$

ενώ ο συντελεστής διακύμανσης θα δίνεται από τον τύπο

$$\mathrm{CV}=e^{\mu+\frac{1}{2}\sigma^2}\sqrt{e^{\sigma^2}-1} \tag{3.6}$$

3.3 Εκτίμηση παραμέτρων

α. Εκτίμηση Μέγιστης Πιθανοφάνειας

Για να εκτιμήσουμε τους εκτιμητές μέγιστης πιθανοφάνειας θα ξεκινήσουμε με την συνάρτηση πιθανότητας. Η συνάρτηση πιθανότητας για την λογαριθμοκανονική κατανομή για X_is (i=1,2,… n) προέρχεται από την λήψη των πυκνοτήταν πιθανότητας του ατόμου X_is (βλ. [4]).

$$L(\mu,\sigma^2;\mathbf{x})=\prod_{i=1}^{n}[f(x_i\backslash\mu,\sigma^2)]$$

$$=\prod_{i=1}^{n}((2\pi\sigma^2)^{-1/2}\,x_i^{-1}\exp[\frac{-(\ln(x_{i)-\mu)}^2}{2\sigma^2}] = (2\pi\sigma^2)^{-n/2}\prod_{i=1}^{n}x_i^{-1}\exp[\sum_{i=1}^{n}\frac{-(\ln(x_{i)-\mu)}^2}{2\sigma^2}].$$

Στην συνέχεια βρίσκουμε τον λογάριθμο της τελευταίας σχέσης

$$L((\mu,\sigma^2;\mathbf{x})=\ln((2\pi\sigma^2)^{-n/2}\prod_{i=1}^{n}x_i^{-1}\exp[\sum_{i=1}^{n}\frac{-(\ln(x_{i)-\mu)}^2}{2\sigma^2}])$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2)-\sum_{i=1}^{n}\ln(x_i)-\sum_{i=1}^{n}\frac{-(\ln(x_{i)-\mu)}^2}{2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2)-\sum_{i=1}^{n}\ln(x_i)-\frac{\sum_{i=1}^{n}[ln(x_i)^2-2\ln(x_i)\mu+\mu^2}{2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2)-\sum_{i=1}^{n}\ln(x_i)-\frac{\sum_{i=1}^{n}[ln(x_i)^2}{2\sigma^2}+\frac{\sum_{i=1}^{n}2\ln(x_i)\mu}{2\sigma^2}-\frac{\sum_{i=1}^{n}\mu^2}{2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2)-\sum_{i=1}^{n}\ln(x_i)-\frac{\sum_{i=1}^{n}[ln(x_i)^2}{2\sigma^2}+\frac{\sum_{i=1}^{n}2\ln(x_i)\mu}{\sigma^2}-\frac{n\mu^2}{2\sigma^2}$$

Παραγωγίζοντας ως προς μ πρώτα και μετά ως προς σ^2 και θέτοντας ίσο με τον μηδέν, βρίσκουμε τις τιμές $\widehat{\mu}$ και $\widehat{\sigma^2}$ που μεγιστοποιούν την $L((\mu,\sigma^2;\mathbf{x})$.

$$\frac{\delta L}{\delta\mu}=\frac{\sum_{i=1}^{n}ln(X_i)}{\hat{\sigma}^2}-\frac{2n\hat{\mu}}{2\hat{\sigma}^2}=0 \qquad \Longrightarrow$$

$$\Longrightarrow \quad \frac{n\hat{\mu}}{\hat{\sigma}^2} = \frac{\sum_{i=1}^{n} \ln(x_i)}{\hat{\sigma}^2} \quad \Longrightarrow$$

$$\Longrightarrow \quad n\hat{\mu} = \frac{\sum_{i=1}^{n} \ln(x_i)}{\hat{\sigma}^2} \quad \Longrightarrow$$

$$\Longrightarrow \quad \hat{\mu}=\frac{\sum_{i=1}^{n} \ln(x_i)}{n}$$

οπότε βρίσκουμε την εμπ του

Επίσης

$$\frac{\delta L}{\delta \sigma^2} = -\frac{n}{2}\frac{1}{\hat{\sigma}^2} - \frac{\sum_{i=1}^{n}(\ln(x_i)-\widehat{\mu})^2}{2} - (\text{-}\hat{\sigma}^2)^{-2}$$

$$= -\frac{n}{2\hat{\sigma}^2} + \frac{\sum_{i=1}^{n}(\ln(x_i)-\widehat{\mu})^2}{2(\hat{\sigma}^2)^2} = 0$$

$$\Longrightarrow \quad \frac{n}{2\hat{\sigma}^2} = \frac{\sum_{i=1}^{n}(\ln(x_i)-\widehat{\mu})^2}{2\hat{\sigma}^4} \quad \Longrightarrow$$

$$\Longrightarrow \quad \mathrm{n} = \frac{\sum_{i=1}^{n}(\ln(x_i)-\widehat{\mu})^2}{\hat{\sigma}^2} \quad \Longrightarrow$$

$$\Longrightarrow \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\ln(x_i)-\widehat{\mu})^2}{n} \quad \Longrightarrow$$

$$\Longrightarrow \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\ln(x_i)-\frac{\sum_{i=1}^{n}\ln(x_i)}{n})^2}{n}$$

Έτσι, οι εκτιμήτριες μέγιστης πιθανοφάνειας είναι τελικά οι εξής

$$\hat{\mu}=\frac{\sum_{i=1}^{n} \ln(X_i)}{n} \tag{3.7}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\ln(X_i)-\frac{\sum_{i=1}^{n}\ln(X_i)}{n})^2}{n} \tag{3.8}$$

β. Μέθοδος ροπών

Μια άλλη δημοφιλής τεχνική εκτίμησης, είναι η μέθοδος των ροπών, που εξισώνει ένα δείγμα ροπών με μη παρατηρήσιμες ροπές του πληθυσμού, από τις οποίες μπορούμε να λύσουμε τις παραμέτρους που θέλουμε να εκτιμήσουμε.

Για να υπολογίσουμε με την μέθοδο των ροπών τις παραμέτρους $\bar{\mu}$ και $\bar{\sigma}^2$ πρέπει πρώτα να βρούμε τα E(X) και E(X^2) για X~Lognormal (μ,σ^2) (βλ. [4]). Άρα έχουμε

$$E(X^n)=\exp[n\mu+n^2\sigma^2/2]$$

$$=\exp[\mu+\sigma^2/2]$$

$$=\exp[2\mu+2\sigma^2]$$

Οπότε $E(X)=e^{\mu+(\frac{\sigma^2}{2})}$ και $E(X^2)=e^{2(\mu+\sigma^2)}$. Τώρα θέτουμε το E(X) ίσο με την πρώτη δειγματική ροπή m_1 και το $E(X^2)$ με την δεύτερη δειγματική ροπή m_2.

$$m_1=\frac{\sum_{i=1}^{n} x_i}{n}$$

$$m_2=\frac{\sum_{i=1}^{n} x_I^2}{n}$$

Θέτοντας $E(X) = m_1$ και $E(X^2) = m_2$ έχουμε

- $\Rightarrow e^{\tilde{\mu}+\widetilde{\sigma^2}/2} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \tilde{\mu}+\frac{\widetilde{\sigma^2}}{2} = \ln[\frac{\sum_{i=1}^{n} x_i}{n}] \Rightarrow \tilde{\mu}+\frac{\widetilde{\sigma^2}}{2} = \ln(\sum_{i=1}^{n} x_i)\text{- }\ln(n)$

  $\Rightarrow \tilde{\mu}= \ln(\sum_{i=1}^{n} x_i)\text{- }\ln(n)\text{- }\frac{\widetilde{\sigma^2}}{2}$

- $\Rightarrow e^{2(\tilde{\mu}+\widetilde{\sigma^2})} = \frac{\sum_{i=1}^{n} x_i^2}{n} \Rightarrow 2\tilde{\mu} + 2\widetilde{\sigma^2}=\ln[\frac{\sum_{i=1}^{n} x_i^2}{n}] \Rightarrow$

  $\Rightarrow 2\tilde{\mu} + 2\widetilde{\sigma^2}= \ln(\sum_{i=1}^{n} x_i^2)\text{-}\ln(n) \Rightarrow \tilde{\mu}= [\ln(\sum_{i=1}^{n} x_i^2)\text{-}\ln(n)\text{- }2\widetilde{\sigma^2}]\cdot\frac{1}{2}$

  $\Rightarrow \tilde{\mu} = \frac{\ln(\sum_{i=1}^{n} x_i^2}{2} - \frac{\ln(n)}{2} - \widetilde{\sigma^2}$

Στην συνέχεια εξισώνουμε τα δύο παραπάνω $\tilde{\mu}$

$$\Rightarrow \ln(\sum_{i=1}^{n} x_i)\text{- }\ln(n)\text{- }\frac{\widetilde{\sigma^2}}{2} = \frac{\ln(\sum_{i=1}^{n} x_i^2}{2} - \frac{\ln(n)}{2} - \widetilde{\sigma^2}$$

$$\Rightarrow 2\ln(\sum_{i=1}^{n} x_i)\text{-}2\ln(n)\text{- }\widetilde{\sigma^2}= \ln(\sum_{i=1}^{n} x_i^2)\text{-}\ln(n)\text{-}2\widetilde{\sigma^2}$$

$$\Rightarrow \widetilde{\sigma^2}= \ln(\sum_{i=1}^{n} x_i^2)\text{- }2\ln(\sum_{i=1}^{n} x_i)+\ln(n)$$

Τέλος αντικαθιστώντας το $\widetilde{\sigma^2}$ σε μία από τις παραπάνω εξισώσεις του $\tilde{\mu}$ έχουμε

$$\tilde{\mu}= \ln(\sum_{i=1}^{n} x_i)\text{- }\ln(n)\text{- }\frac{\widetilde{\sigma^2}}{2}$$

$$= \ln(\sum_{i=1}^{n} x_i)\text{- }\ln(n)\text{- }\frac{1}{2}[\ln(\sum_{i=1}^{n} x_i^2)\text{- }2\ln(\sum_{i=1}^{n} x_i)+\ln(n)$$

$$= \ln(\sum_{i=1}^{n} x_i)\text{- }\ln(n)\text{-}\frac{\ln(\sum_{i=1}^{n} x_i^2}{2} + \ln(\sum_{i=1}^{n} x_i)\text{ -}\frac{\ln(n)}{2}$$

$$=2\ln(\sum_{i=1}^{n} x_i)\text{- }\frac{3}{2}\ln(n)\text{)- }\frac{\ln(\sum_{i=1}^{n} x_i^2}{2}$$

Έτσι με την μέθοδο των ροπών οι εκτιμητές είναι οι εξής :

$$\tilde{\mu} = 2\ \ln(\sum_{i=1}^{n} X_i) - \frac{3}{2}\ln(\mathrm{n})\ ) - \frac{\ln(\sum_{i=1}^{n} X_i^2}{2} \qquad (3.9)$$

$$\widetilde{\sigma^2} = \ln(\sum_{i=1}^{n} X_i^2) - 2\ \ln(\sum_{i=1}^{n} X_i) + \ln(\mathrm{n}) \qquad (3.10)$$

γ. Εύρωστοι εκτιμητές Serfling

Μια άλλη μέθοδος εκτίμησης παραμέτρων σχεδιάστηκε από τον Serfling το 2002. Ο Serfling για την μέθοδο αυτήν έλαβε υπόψη του δύο διαφορετικά κριτήρια. Το πρώτο βασίστηκε στην ασυμπτωτική βελτιστοποίηση όσον αφορά την απόδοση διακύμανσης της τεχνικής εκτίμησης μέγιστης πιθανοφάνειας και το δεύτερο κριτήριο προκάλεσε στον Serfling ανησυχία όσον αφορά την ευρωστία, που την χωρίζει σε δύο μέρη, στο σημείο διάσπασης και στο μεγάλο σφάλμα ευαισθησίας. Το σημείο διάσπασης ενός εκτιμητή (BP: breakdown point) είναι το μεγαλύτερο μέρος των τιμών των δεδομένων που μπορεί να καταστραφεί χωρίς ο εκτιμητής να δώσει κάποια πληροφορία για την παράμετρο που μας απασχολεί. Το μεγάλο σφάλμα ευαισθησίας (GES: gross error sensitivity) μετρά την μέγιστη συνεισφορά στην εκτίμηση σφάλματος που μπορεί να προκληθεί από μία απομακρυσμένη παρατήρηση όταν ο δεδομένος εκτιμητής χρησιμοποιείται.

Ο Serfling ανέφερε επίσης ότι, όταν η αναμενόμενη αναλογία των ακραίων τιμών αυξάνεται τότε συνιστάται ένας εκτιμητής με υψηλό σημείο διάσπασης BP και χαμηλό GES.

Οι εξισώσεις των εκτιμητών μ και σ^2 για την κατανομή lognormal που δόθηκαν από τον Serfling είναι οι εξής :

$$\hat{\mu}_{s(k)} = \text{median}\ (\frac{\sum_{i=1}^{k} lnX_{k(i)}}{k}) \qquad (3.11)$$

$$\widehat{\sigma^2}_{s(m)} = \text{median}\ (\frac{\sum_{i=1}^{m} (lnX_{m(i)} - \frac{\sum_{j=1}^{m} lnX_{m(j)}}{m})^2}{m}) \qquad (3.12)$$

όπου X_k και X_m είναι ομάδες k και m τυχαία επιλεγμένων τιμών από ένα δείγμα μεγέθους n μεταβλητών λογαριθμοκανονικής κατανομής, λαμβάνοντας $\binom{n}{k}$ και $\binom{n}{m}$ φορές αντίστοιχα. Τα $X_{k(i)}$ ή τα $X_{m(i)}$ δείχνουν την κάθε i-τιμή της κάθε ομάδας από τα k ή m επιλεγμένα X_s. Ο Serfling επεσήμανε ότι αν τα $\binom{n}{k}$ και $\binom{n}{m}$ είναι μεγαλύτερα από 10^7, τότε είναι επαρκής για να υπολογιστεί ο εκτιμητής βασιζόμενος μόνο σε 10^7 τυχαία επιλεγμένες ομάδες. Αυτό συμβαίνει επειδή χρησιμοποιώντας παραπάνω από 10^7 ομάδες πιθανότατα δεν προσθέτει καμία πληροφορία που δεν έχει ήδη συλλεχθεί σχετικά με τα δεδομένα, ενώ περιορίζοντας τον αριθμό των ομάδων αποφεύγουμε την υπολογιστική επιβάρυνση.

Για την ταυτόχρονη εκτίμηση των μ και σ^2, ο Serfling συνιστά ότι για k=9 και m=9 αποδίδονται τα καλύτερα από κοινού αποτελέσματα σε σχέση με τις τιμές των BP και GES.

3.4 Γενικευμένη κατανομή Lognormal

Γνωστή και ως εκθετική κατανομή ισχύος, ή γενικευμένη κατανομή σφάλματος, ιδίως στην ιταλική λογοτεχνία, αυτή είναι μια παραμετρική οικογένεια των συμμετρικών κατανομών. Η γενικευμένη κατανομή προτάθηκε για πρώτη φορά από τον Subbotin (1923) Περιλαμβάνει όλες τις κανονικές και Laplace κατανομές, και ως οριακές περιπτώσεις περιλαμβάνει όλες τις συνεχείς ομοιόμορφες κατανομές για οριοθετημένα χρονικά διαστήματα της πραγματικής ευθείας.

Η συνάρτηση πυκνότητας πιθανότητας μιας γενικευμένης λογαριθμοκανονικής κατανομής με παραμέτρους μ, σ και s δίνεται από:

$$f(x)=\frac{s}{2x\sigma\Gamma(\frac{1}{s})}\exp\left(-\left|\frac{\log x-\mu}{\sigma}\right|^{s}\right), \qquad (3.13)$$

με $x > 0$, $-\infty < \mu < +\infty$, $\sigma > 0$ και $s \geq 1$

Σημαντικό είναι να σημειωθεί ότι το Γ υποδηλώνει την Γάμμα συνάρτηση. Αυτή η κατανομή έχει την κατανομή lognormal στην περίπτωση που πάρουμε το $\sqrt{s}=2$ και αλλάξουμε το σ με 2σ και για s=1 μας δίνει την κατανομή Laplace.

Η ικανότητα μιας κατανομής να ταιριάζει απόλυτα στα δεδομένα εξαρτάται από το σχήμα της. Το σχήμα μπορεί να ορίζεται από την τρίτη και τέταρτη ροπή και αυτές αντιπροσωπεύουν την ασυμμετρία των συντελεστών μιας δεδομένης κατανομής. Οι κατανομές logGN επιτρέπουν την μοντελοποίηση της κύρτωσης παρέχοντας έτσι μια πιο ευέλικτη προσαρμογή σε πειραματικά δεδομένα από την λογαριθμοκανονική κατανομή.

3.5 Πεδία εφαρμογής

Η lognormal κατανομή είναι συνήθως το καλύτερο μοντέλο για πραγματικά δεδομένα, το οποίο περιγράφεται και αναλυτικότερα σε διάφορες εργασίες όπως των Aitchison και Brown 1957, του Koch 1966,1969 και των Crow και Shimizu το 1988. Ο λόγος που προτιμάται από τις άλλες είναι γιατί ταιριάζει καλύτερα σε μικρούς συντελεστές διακύμανσης (βλ. [18]).

α. Δεδομένα εισοδήματος

Η λογαριθμοκανονική κατανομή είναι η πιο συχνά χρησιμοποιούμενη κατανομή για την μοντελοποίηση των εισοδημάτων και μισθών. Για τους σκοπούς της ανάλυσης των εισοδημάτων και άλλων χαρακτηριστικών, τα στατιστικά στοιχεία τα οποία συσχετίζονται με το εισόδημα χρησιμοποιούν κυρίως λογαριθμοκανονική κατανομή τριών παραμέτρων. Οι κατανομές lognormal δύο-παραμέτρων και τεσσάρων παραμέτρων χρησιμοποιούνται επίσης, αλλά όχι τόσο συχνά όσο η κατανομή τριών –παραμέτρων.

Η επιλογή της χρήσης της ανάλυσης της κατανομής του εισοδήματος είναι να διερευνηθεί η επίδραση των διαφορετικών ειδών φορολογίας σχετικά με τη διάρθρωση των εισοδημάτων. Η ανάλυση των εσόδων είναι σημαντική και χρήσιμη, διότι το επίπεδο του εισοδήματος συνδέεται με την ποιότητα της ζωής, παρέχει αντικειμενική άποψη και επιτρέπει τον ποσοτικό προσδιορισμό. Χρησιμοποιείται για να συγκρίνουμε τις διεθνείς ή διακρατικές περιφέρειες και μπορούμε να προβλέψουμε τη μελλοντική πρόοδο.

Υπάρχουν δύο παράγοντες που απαιτείται όταν κάνουμε ένα μοντέλο κατανομής του εισοδήματος. Το πρώτο είναι να πάρει την πιο ακριβή ομοιότητα των μοντέλων και της πραγματικότητας. Ο παράγοντας αυτός σημαίνει ν' αυξηθεί ο αριθμός των παραμέτρων του μοντέλου. Ο δεύτερος παράγοντας είναι να πάρει εύκολα την οικονομική ερμηνεία. Αυτό σημαίνει μείωση του αριθμού των παραμέτρων. Γι 'αυτό χρησιμοποιούμε πιο συχνά των τριών-παραμέτρων λογαριθμοκανονικής κατανομής.

Η λογαριθμοκανονική κατανομή είναι μία από τις πολλές κατανομές που χρησιμοποιούνται στην ανάλυση των εισοδημάτων. Η λογαριθμοκανονική κατανομή δύο-παραμέτρων ταιριάζει καλά σε ένα μεγάλο μέρος του φάσματος του μεσαίου εισοδήματος, αλλά δίνει μια κακή προσαρμογή στις ουρές. Ωστόσο, στην περιοχή μεσαίου εισοδήματος υπερβάλλει ασυμμετρίας.

Τα τελευταία 50 χρόνια γίνανε πολλές έρευνες με την χρήση της λογαριθμοκανονικής κατανομής πάνω σε δεδομένα εισοδήματος. Μία από τις πιο πρόσφατες ήταν από τον Kalecki (1945), ο οποίος βρήκε την λογαριθμοκανονική κατανομή δύο-παραμέτρων για όλο το φάσμα των εισοδημάτων να είναι αρκετά φτωχή, ενώ ένα μοντέλο δύο παραμέτρων για εισοδήματα πάνω από ένα ορισμένο όριο, που είναι, μια λογαριθμοκανονική κατανομή τριών-παραμέτρων να παρέχει μια καλή προσέγγιση.

Οι Gastwirth και Smith διαπίστωσαν με την χρήση μη παραμετρικών ορίων για τον συντελεστή Gini, ότι οι δείκτες που προέρχονται από λογαριθμοκανονική κατανομή δύο και τριών παραμέτρων δεν εμπίπτουν στα όρια των ΗΠΑ για το ατομικό ακαθάριστο προσαρμοσμένο εισόδημα για την περίοδο 1955-1969 και κατέληξαν στο συμπέρασμα ότι η λογαριθμοκανονική κατανομή είναι ακατάλληλη για την μοντελοποίηση τέτοιων δεδομένων.

Ενδιαφέρον έχει και η έρευνα των Kloek και van Dijk (1977) οι οποίοι ταίριαξαν την λογαριθμοκανονική κατανομή στην αυστραλιανή οικογένεια των διαθέσιμων εισοδημάτων και τα κατηγοριοποίησαν με βάση την ηλικία, το επάγγελμα και την εκπαίδευση του αρχηγού της οικογένειας.

β. Δεδομένα πλούτου

Από την στιγμή που ο Gibrat το (1931) έφερε στο προσκήνιο την λογαριθμοκανονική ως μοντέλο κατανομής εισοδήματος και μέχρι τις αρχές του 1970, προσέλκυσε το ενδιαφέρον μιας ευρείας ομάδας οικονομολόγων, στατιστικών και οικονομετρητών. Ο Sargan (1959) πρότεινε τη λογαριθμοκανονική ως ένα μοντέλο κατανομής του πλούτου. Με βάση τις αυστηρές και πολύ ρεαλιστικές παραδοχές κατέληξε στο συμπέρασμα ότι «η κατανομή του πλούτου τείνει να είναι περίπου λογαριθμοκανονική ". Στην οικονομική ιστορία, το λογαριθμοκανονικό μοντέλο ήταν και συνεχίζει να είναι μια κακή αναπαράσταση των παρατηρούμενων κατανομών εισοδήματος. Κατά μείζονα λόγο, αυτό είναι ένα εντελώς ακατάλληλο μοντέλο των παρατηρούμενων κατανομών πλούτου. Είναι ένα πολύ άκαμπτο μοντέλο για να είναι σε θέση να αντιπροσωπεύει με ακρίβεια το εισόδημα και τις κατανομές του πλούτου.

Ο Chesher το (1979), εκτίμησε ένα λογαριθμοκανονικό μοντέλο για την κατανομή του πλούτου στην Ιρλανδία με την καταγραφή του μεγέθους των ακινήτων. Είναι σαφές ότι η λογαριθμοκανονική κατανομή είναι ανώτερη από την κατανομή Pareto αυτών των δεδομένων, με x^2 και την πιθανότητα βελτίωσης περίπου 93%.

γ. Μεγέθη επιχειρήσεων

Τα μεγέθη των επιχειρήσεων συνήθως περιγράφονται από την λογαριθμοκανονική κατανομή αποτυπώνοντας με ακρίβεια το ανώτερο (μεγάλο μέγεθος) ουράς. Αρχίζοντας με τον Gibrat υπάρχει μια καθιερωμένη παράδοση που περιγράφει την κατανομή του μεγέθους της επιχείρησης σε βιομηχανικές χώρες από την λογαριθμοκανονική κατανομή. Αυτή η κατανομή είναι μια άμεση συνέπεια του <<νόμου της ανάλογης επίδρασης >>, σύμφωνα με τον οποίο η ανάπτυξη των επιχειρήσεων αντιμετωπίζεται ως μια τυχαία διαδικασία με τους ρυθμούς ανάπτυξης να είναι ανεξάρτητοι από το μέγεθος της εταιρίας.

Ο Quandt (1966), σε μια μελέτη που διερεύνησε για την κατανομή του μεγέθους της επιχείρησης (μέγεθος που μετράται σε όρους περιουσιακών στοιχείων) στις Ηνωμένες Πολιτείες, διαπίστωσε ότι η λογαριθμοκανονική κατανομή είναι πιο κατάλληλη από τις κατανομές Pareto των τύπων I-III για τα δεδομένα του. Πιο πρόσφατα, ο Stanley (1995) χρησιμοποίησε την λογαριθμοκανονική για να διαμορφώσει το μέγεθος της κατανομής των αμερικανικών επιχειρήσεων (από πωλήσεις), σημειώνοντας ότι το μοντέλο προβλέπει τις άνω ουρές.

δ. Ασφαλιστικές ζημιές

Ίσως ο πιο σημαντικός λόγος που η κατανομή lognormal είναι τόσο σημαντική στην ασφάλιση υγείας είναι διότι πραγματικά κάνει σωστή δουλειά. Αυτό ισχύει ιδιαίτερα στην Νέα Ζηλανδία όπου λόγω του νομοθετικού περιβάλλοντος, έχουμε πολύ μικρή έκθεση σε μεγάλες αξιώσεις για σωματική βλάβη. Σε άλλους τομείς και στην περίπτωση ορισμένων υλικών ζημιών, η κατανομή lognormal τείνει να υποδηλώνει το κόστος της ‘’μακράς ουράς’’. Αυτό αναφέρεται σε έγγραφο του Neil Christie στο Society το 1998: “An introduction to developing a claim size distribution for a reinsurance programme”.

Με την μελέτη της λογαριθμοκανονικής κατανομής ασχολήθηκαν πολλοί τα τελευταία 60 χρόνια. Ένα δείγμα από αυτές τις μελέτες είναι οι εξής (βλ. [8]) :

- Ο Benckert (1962) μελέτησε τις βιομηχανικές και μη βιομηχανικές ζημιές πυρός, την διακοπή επιχειρήσεων και την ασφάλιση ατυχήματος καθώς και την ασφάλιση αυτοκινήτων από τρίτους στην Σουηδία το 1948-1952.
- Ο Ferrara (1971) χρησιμοποίησε λογαριθμοκανονική κατανομή τριών παραμέτρων για την μοντελοποίηση βιομηχανικών ζημιών πυρός στην Ιταλία για την περίοδο 1963-1965.
- Οι Hogg και Klugman (1983) ταίριαξαν την λογαριθμοκανονική κατανομή με ένα μικρό σύνολο δεδομένων (35 παρατηρήσεις) των ζημιών του τυφώνα και βρήκαν ότι ταιριάζει τόσο καλά όσο και η κατανομή Weibull.
- Οι Burnecki, Kukla, και Weron (2000) χρησιμοποίησαν την λογαριθμοκανονική κατανομή, όταν μοντελοποίησαν ασφαλιστικές ζημιές περιουσίας και διαπίστωσαν ότι ξεπερνά την κατανομή Pareto για αυτά τα δεδομένα.

Συνολικά, φαίνεται ότι η κατανομή lognormal δεν είναι η καλύτερη επιλογή για την μοντελοποίηση εισοδήματος, εταιρικών μεγεθών, και ασφαλιστικών ζημιών.

ΚΕΦΑΛΑΙΟ 4

Κατανομή Γάμμα

4.1 Ορισμός

Το πρωτοποριακό έργο που σηματοδοτεί την αρχική χρήση της κατανομής Γάμμα (Gamma) ως μια κατανομή εισοδήματος, οφείλεται στον γάλλο στατιστικό Lucien March, ο οποίος το 1898 προσάρμοσε την κατανομή Γάμμα σε διάφορες γαλλικές, γερμανικές και αμερικάνικες κατανομές κέρδους. Περίπου 25 χρόνια αργότερα, το 1924 ο Amoroso όρισε την γενικευμένη κατανομή Γάμμα .

Ας πάρουμε δύο παραμέτρους $\alpha > 0$ και $\beta > 0$. Η συνάρτηση της Γάμμα Γ(α) ορίζεται από τον τύπο

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{dx}$$

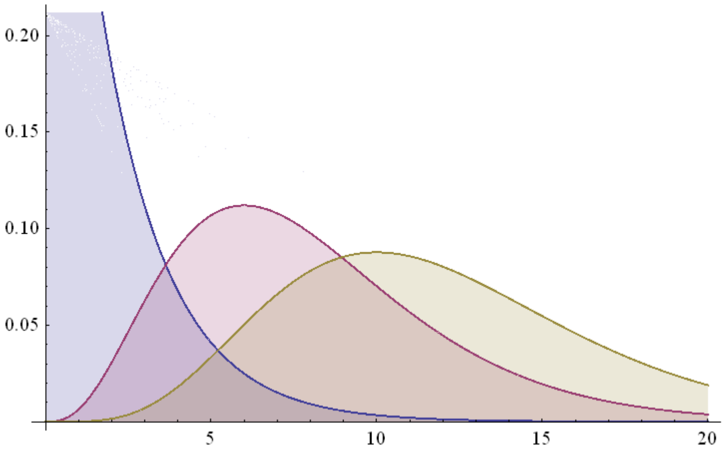
Αν διαιρέσουμε και τα δύο μέλη με Γ(α) τότε παίρνουμε

$$1=\int_0^\infty \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}=\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}\mathrm{dy}$$

όπου αντικαθιστούμε το x=βy. Επομένως αν ορίσουμε

$$f(x\backslash\alpha,\beta)=\begin{cases}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\ , & x\geq 0\\ 0, & x<0\end{cases} \tag{4.1}$$

τότε η $f(x\backslash\alpha,\beta)$ είναι μια συνάρτηση πυκνότητας πιθανότητας.



Διάγραμμα 4.1.

Συνάρτηση πυκνότητας πιθανότητας της κατανομής Γάμμα

Μερικές από τις ιδιότητες της συνάρτησης Γάμμα είναι οι εξής :

- Αν πάρουμε $\alpha > 1$ τότε ολοκληρώνοντας κατά μέλη μπορούμε να πάρουμε:

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{dx}=\int_0^\infty x^{a-1}d(-e^{-x})=$$

$$= x^{\alpha-1}(-e^{-x})|_0^\infty - \int_0^\infty (-e^{-x})(\alpha-1)x^{\alpha-2}dx=$$

$$=(\alpha\text{-}1)\int_0^\infty x^{(\alpha-1)-1}e^{-x}dx = (\alpha\text{-}1)\Gamma(\alpha\text{-}1)$$

- Για α=1 έχουμε

$$\Gamma(1)=\int_0^\infty e^{-x}dx = 1$$

Μπορούμε επίσης να γράψουμε Γ(2)= 1·1, Γ(3)=2·1, Γ(4)=3·2·1, Γ(5)=4·3·2·1

οπότε προκύπτει ότι Γ(n)= (n-1)!

4.2 Βασικά χαρακτηριστικά

Για να βρούμε την μέση τιμή και την διακύμανση της κατανομής Γάμμα θα υπολογίσουμε πρώτα την ροπή k- τάξης. Έχουμε

$$\mathrm{EX}^k = \int_0^\infty x^k \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}\mathrm{dx} = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{(\alpha+k)-1}e^{-\beta x}\mathrm{dx}=$$

$$=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+k)}{\beta^{\alpha+k}}\int_0^{\infty}\frac{\beta^{\alpha+k}}{\Gamma(\alpha+k)}x^{\alpha+k+1}e^{-\beta x}dx=$$

και επειδή το ολοκλήρωμα $\int_0^{\infty}\frac{\beta^{\alpha+k}}{\Gamma(\alpha+k)}x^{\alpha+k+1}e^{-\beta x}dx$ είναι ολοκλήρωμα της συνάρτηση πυκνότητας πιθανότητας της κατανομής Γ(α+k, β) θα είναι ίσο με 1. Άρα

$$=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma(\alpha+k)}{\beta^{\alpha+k}}=\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\beta^{k}}=\frac{(\alpha+k-1)\Gamma(\alpha+k-1)}{\Gamma(\alpha)\beta^{k}}$$

$$=\frac{(\alpha+k-1)(\alpha+k-2)\cdots a\Gamma(\alpha)}{\Gamma(\alpha)\beta^{k}}=\frac{(\alpha+k-1)\cdots\alpha}{\beta^{k}}.$$

Για k=1 έχουμε την μέση τιμή

$$\mathrm{E(X)}=\frac{\alpha}{\beta} \tag{4.2}$$

και για k= 2 έχουμε την ροπή δεύτερης τάξης

$$\mathrm{E(X^2)}=\frac{(\alpha+1)\alpha}{\beta^2} \tag{4.3}$$

Άρα ο τύπος την διακύμανσης είναι ο εξής :

$$\mathrm{Var(X)}=\mathrm{E(X^2)}-\mathrm{(EX)}^2=\frac{(\alpha+1)\alpha}{\beta^2}-\left(\frac{\alpha}{\beta}\right)^2=\frac{\alpha}{\beta^2} \tag{4.4}$$

4.3 Εκτίμηση παραμέτρων

α. Εκτίμηση Μέγιστης Πιθανοφάνειας

Για να βρούμε τον εκτιμητή μέγιστης πιθανοφάνειας ξεκινάμε από την πιθανότητα (βλ. [23])

$$L(\alpha,\beta;\mathbf{x})=\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}{x_1}^{\alpha-1}e^{-\beta x_1}\right)\ldots\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}{x_1}^{\alpha-1}e^{-\beta x_n}\right)=$$

$$=\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{n}(x_1x_{2\ldots}x_n)^{\alpha-1}e^{-\beta(x_{1+}x_{2+}\ldots+x_n}$$

$$\ln L(\alpha,\beta;\mathbf{x})=n(\alpha\ln\beta-\ln\Gamma(\alpha))+(\alpha-1)\sum_{i=1}^{n}ln x_i-\beta\sum_{i=1}^{n}x_i$$

Για τον προσδιορισμό των παραμέτρων που μεγιστοποιούν την πιθανότητα, λύνουμε τις εξισώσεις:

$$\frac{\vartheta}{\vartheta\alpha}\ln L(\hat{\alpha},\hat{\beta};\mathbf{x})=n(\ln\hat{\beta}-\frac{d}{d\alpha}\ln\Gamma(\hat{\alpha}))+\sum_{i=1}^{n}x_i=0$$

$$\frac{\vartheta}{\vartheta\beta}\ln L(\hat{\alpha},\hat{\beta};\mathbf{x})=n\frac{\hat{\alpha}}{\hat{\beta}}-\sum_{i=1}^{n}x_i=0 \qquad \text{ή} \qquad \hat{x}=\frac{\hat{\alpha}}{\hat{\beta}}$$

Αντικαθιστώντανς όπου $\hat{\beta}=\frac{\hat{\alpha}}{\hat{x}}$ στην πρώτη εξίσωση ,παίρνουμε μια σχέση για το $\hat{\alpha}$

$$n(\ln\hat{\alpha}-\ln x-\frac{d}{d\alpha}\ln\Gamma(\hat{\alpha}))+\sum_{i=1}^{n}\ln x_i=0$$

και καταλήγουμε σε μια δίγαμμα συνάρτηση

$$\psi(\alpha)=\frac{d}{d\alpha}\ln\Gamma(\alpha) \tag{4.5}$$

δηλαδή στην λογαριθμική παράγωγο της συνάρτησης Γάμμα.

β. Μέθοδος ροπών

Έστω $x_1,\ldots,x_n$ είναι ένα τυχαίο δείγμα από μια πιθανότητα κατανομής Γάμμα με παραμέτρους α και β . Θα βρω τους εκτιμητές για τις άγνωστες παραμέτρους α και β.

Για την κατανομή Γάμμα ισχύει :

$$E[X]=\alpha\beta \quad \text{και} \quad E[X^2]=\alpha\beta^2+\alpha^2\beta^2$$

Επειδή έχουμε δύο παραμέτρους , πρέπει να βρούμε τους δύο πρώτους εκτιμητές ροπών

$$\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}=\alpha\beta \qquad \text{και} \qquad \frac{1}{n}\sum_{i=1}^{n}{X_i}^2=\alpha\beta^2+\alpha^2\beta^2$$

Λύνοντας ως προς α και ως προς β καταλήγουμε

$$\alpha=(\bar{x}/\beta) \qquad \text{και} \qquad \beta=[\{(1/n)\sum_{i=1}^{n}{x_i}^2-\bar{x}^2\}/\bar{x}]$$

Επομένως οι εκτιμητές των α και β με την μέθοδο των ροπών είναι

$$\hat{\alpha}=\frac{\bar{X}}{\hat{\beta}} \tag{4.6}$$

και

$$\hat{\beta}=\frac{\frac{1}{n}\sum_{i=1}^{n}{X_i}^2-\bar{X}^2}{\bar{X}} \tag{4.7}$$

και συνεπάγεται ότι

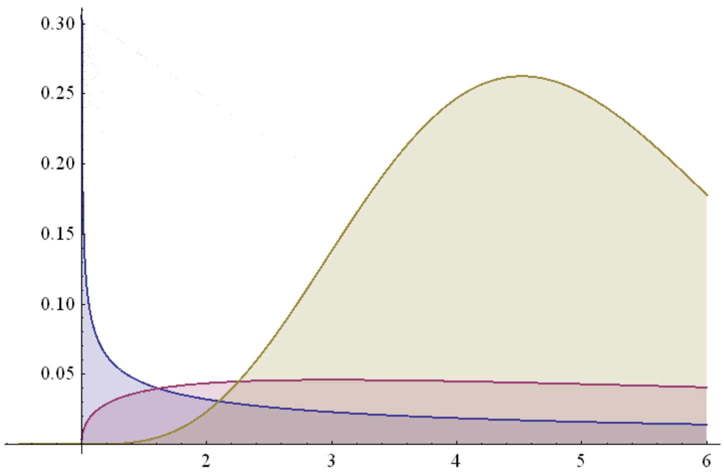
$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} - \bar{X}^2} = \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (4.8)$$

4.4 Γενικευμένη κατανομή Gamma

Η συνάρτηση πυκνότητας πιθανότητας της Γενικευμένης κατανομής Γάμμα (GG(α,β,γ)) δίνεται από τον τύπο (βλ. [31]).

$$f(y\backslash\alpha,\beta,\gamma)=\frac{\beta}{\gamma\Gamma(\alpha)}\left(\frac{y}{\gamma}\right)^{\alpha\beta-1}e^{-\left(\frac{y}{\gamma}\right)^{\beta}} \qquad y \geq 0,\ \ \beta, \alpha, \gamma > 0 \qquad (4.9)$$

όπου Γ(.) είναι η συνάρτηση Γάμμα, α και β είναι οι παράμετροι σχήματος και γ είναι η παράμετρος κλίμακας. Η GG οικογένεια είναι ευέλικτη και περιλαμβάνει πολλά γνωστά μοντέλα ως υποοικογένειες. Οι υποοικογένειες είναι η εκθετική (α=β=1), η γάμμα για (β=1), και η Weibull για (α=1). Η λογαριθμοκανονική κατανομή λαμβάνεται επίσης ως περιορισμένη κατανομή με α→∞. Με β=2 παίρνουμε μια υποοικογένεια της GG που είναι γνωστή ως γενικευμένη κανονική κατανομή.



Διάγραμμα 4.2.

Συνάρτησης πυκνότητας πιθανότητας της γενικευμένης κατανομής Γάμμα

4.5 Πεδία εφαρμογής

Όπως και οι περισσότερες κατανομές πιθανότητας, έτσι και η σημασία της κατανομής Γάμμα βρίσκεται σε πολυάριθμες εφαρμογές της σε μια ευρεία ποικιλία τομέων. Μερικές ιδιαίτερα ενδιαφέρουσες εφαρμογές περιλαμβάνουν τη μοντελοποίηση του ποσού της βροχόπτωσης σε μια περιοχή πάνω από ένα συγκεκριμένο χρονικό διάστημα . Η κατανομή Γάμμα δίνει επίσης την κατανομή πιθανότητας της ποσότητας του χρόνου που απαιτείται για n εμφανίσεις ενός γεγονότος σε μια διαδικασία Poisson.

α. Δεδομένα εισοδήματος και πλούτου

Ένα <<ιδανικό>> μοντέλο κατανομής του εισοδήματος είναι πιθανό να υπάρξει και η ανάπτυξη νέων μορφών κατανομών εισοδήματος θα μπορούσε να συνεχιστεί επ'αόριστον, διότι κανένα μοντέλο δεν είναι τέλεια προσαρμοσμένο σε κάθε σύνολο δεδομένων. Ένα μοντέλο όμως δύο παραμέτρων μπορεί να αναπαραστήσει με ακρίβεια τα δεδομένα και είναι επαρκές για να περιγράψει την κατανομή των εισοδημάτων. Εκτός λοιπόν από την λογαριθμοκανονική κατανομή που αναφέραμε στο Κεφάλαιο 3 ένα ακόμα πολυχρησιμοποιημένο μοντέλο δύο παραμέτρων είναι η κατανομή Γάμμα. Το πλεονέκτημα της χρήσης αυτής της κατανομής δεν είναι ότι περιλαμβάνει μόνο δύο παραμέτρους, αλλά επίσης και οι παράμετροι αυτοί έχουν μια σαφή οικονομική ερμηνεία. Συγκεκριμένα η παράμετρος α σχετίζεται άμεσα με την μέτρηση της ανισότητας με τέτοιο τρόπο που αν η τιμή του α αυξηθεί τότε η αριστερή ουρά μειώνεται και ως εκ τούτου και η ανισότητα μειώνεται.

Εκτός από την δύο παραμέτρων κατανομή χρησιμοποιήθηκε και η τεσσάρων παραμέτρων γενικευμένη Γάμμα κατανομή από τον Amaroso (1924-1925) για τα Πρώσικα εισοδήματα του 1912. Περίπου 50 χρόνια αργότερα ο Bartels (1977) εφάρμοσε την τριών παραμέτρων έκδοση φορολογικών εσόδων το 1969 για τις τρεις περιφέρειες στην Ολλανδία και διαπίστωσε ότι σε ορισμένες περιπτώσεις η κατανομή Γάμμα είναι η καλύτερη (βλ. [27]).

Ο McDonald (1984) εκτίμησε την γενικευμένη κατανομή Γάμμα για τα οικογενειακά εισοδήματα στις ΗΠΑ τις χρονιές 1970, 1975, και 1980. Ξεπερνά οκτώ άλλες κατανομές και είναι κατώτερη μόνο από τις γενικευμένες κατανομές Beta τύπου ΙΙ και Singh-Maddala. Σε μια εμπεριστατωμένη μελέτη που απασχολήθηκαν 15 μοντέλα κατανομής του εισοδήματος τύπου Beta και Γάμμα, οι Bordley, McDonald, και Mantrala (1996) ταίριαξαν την γενικευμένη κατανομή Γάμμα με τα οικογενειακά εισοδήματα των ΗΠΑ για το 1970, 1975, 1980, 1985, και 1990. Σε μια εφαρμογή που χρησιμοποίησε την περίοδο 1984-1993 γερμανικά εισοδήματα νοικοκυριών, η γενικευμένη κατανομή Γάμμα αποκαλύπτεται ως ανάρμοστο μοντέλο για αυτά τα δεδομένα από τους (Brachmann, Stich, και Trede, 1996). Ειδικότερα, δεν παρέχει καμία βελτίωση έναντι της κατανομής Γάμμα δύο παραμέτρων. Τα στοιχεία φαίνεται να απαιτούν ένα πιο ευέλικτο μοντέλο όπως γενικευμένη κατανομή Beta και η κατανομή Singh-Maddala.

β. Ασφαλιστικές ζημιές

Οι κατανομές Γάμμα διαδραματίζουν εξέχοντα ρόλο στην αναλογιστική επιστήμη. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι οι περισσότερες συνολικές κατανομές ασφαλιστικών αποζημιώσεων έχουν περίπου το ίδιο σχήμα όπως οι κατανομές Γάμμα : είναι στραμμένες προς τα δεξιά, είναι μη αρνητικές και μονότροπες. Επιπλέον οι κατανομές Γάμμα είναι καλά μελετημένες και έχουν υποστεί αρκετή ανάλυση. Σαν αποτέλεσμα υπάρχουν πολλά παραδείγματα εφαρμογής της κατανομής Γάμμα για την μοντελοποίηση ασφαλιστικών χαρτοφυλακίων π.χ. Hurlimann (2001), Melnick και Tenenbein (2000) και Rioux και Klugman (2004). Επίσης οι Herzog (1999) και Hossack et al. (1983) σημειώνουν ότι οι κατανομές Γάμμα παρέχουν ένα βολικό μοντέλο για το μέσο ποσοστό των αξιώσεων που υποβάλλονται από διάφορους ασφαλισμένους της ασφαλιστικής εταιρίας. Ο Bowers et al. (1997) χρησιμοποίησε μεταφρασμένες κατανομές Γάμμα ως μοντέλο για τις συνολικές ασφαλιστικές απαιτήσεις.

Στην <<παραδοσιακή>> θεωρία κινδύνου, οι επί μέρους ζημιές σε ένα χαρτοφυλάκιο υποτίθεται ότι είναι ανεξάρτητες, αν και στην πλειονότητα των περιπτώσεων η υπόθεση αυτή δεν είναι σύμφωνη με την πραγματικότητα. Για να κλείσει το χάσμα, πρέπει κανείς να καθορίσει το κατάλληλο πολυμεταβλητό μοντέλο πιθανοτήτων που θα ταιριάζει σε μια πραγματική επιχείρηση ασφαλίσεων ζωής. Θα πρέπει να σημειωθεί ότι εκτός από τις <<κλασικές>> προαναφερθείσες αναλογιστικές εφαρμογές, πολυπαραγοντικά μοντέλα με μονοδιάστατα περιθώρια κατανομής Γάμμα έχουν πρόσφατα χρησιμοποιηθεί στην μέτρηση του χρηματοοικονομικού κινδύνου. Δηλαδή, οι Furman και Landsman (2005) εξέτασαν το εξαρτώμενο από την ουρά μέτρο κινδύνου προσδοκίας (TCE) στην περίπτωση του Γάμμα πολυμεταβλητού χαρτοφυλακίου των κινδύνων.

ΚΕΦΑΛΑΙΟ 5

Κατανομή Βήτα

5.1 Ορισμός

Η κατανομή Βήτα χρησιμοποιείται ευρέως στη στατιστική μοντελοποίηση από οριοθετημένες τυχαίες μεταβλητές. Υπολογίζεται εύκολα και μπορεί να πάρει διάφορες μορφές. Παρ’ όλα αυτά η εφαρμογή της είναι περιορισμένη. Πρώτον ως κατανομή δύο παραμέτρων, μπορεί να παρέχει μόνο περιορισμένη ακρίβεια σε προσαρμοσμένα δεδομένα, δεύτερον η Βήτα δεν προσφέρει μια φυσική και βολική ερμηνεία για την εισαγωγή επεξηγηματικών μεταβλητών και τρίτον είναι ακατάλληλη για χρήση σε Μπευζιανή ανάλυση. Στην κατανομή Βήτα(p,q), οι παράμετροι (p,q) καθορίζουν τόσο το σχήμα, όσο και τις ροπές της κατανομής.

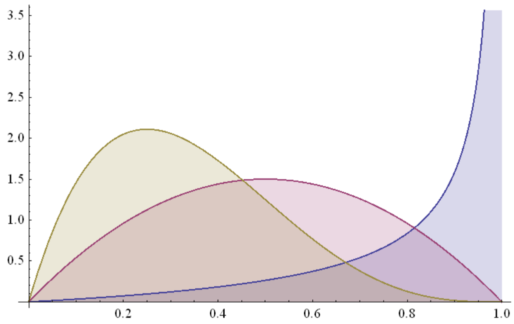
Πρόσφατη έρευνα έχει συμβάλλει σε τρεις γενικεύσεις της κατανομής Βήτα. Η πρώτη είναι των Armero και Bayarri (1994) που όρισαν την υπεργεωμετρική κατανομή Gauss (GH), μια σχετική κατανομή εισάχθηκε από τους McDonald και Xu (1995) ως Γενικευμένη κατανομή Βήτα (GB) και η τρίτη είναι του Gordy ο οποίος γενίκευσε την Βήτα σε μία εντελώς διαφορετική κατεύθυνση, την confluent hypergeometric κατανομή CH(p,q,s).

Y~Beta(α, β), α,β, > 0 είναι μία κατανομή Βήτα αν και μόνο αν έχει σ.π.π

$$f(y) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} & 0 \le y \le 1 \\ 0 & \text{αλλιώς} \end{cases} \quad (5.1)$$

Η συνάρτηση που εμφανίζεται στον παρανομαστή της Βήτα συνάρτησης πυκνότητας πιθανότητας ονομάζεται Βήτα συνάρτηση και ορίζεται ως

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad (5.2)$$



Διάγραμμα 5.1

Συνάρτηση πυκνότητας πιθανότητας της κατανομής Βήτα

➢ *Ιδιότητες της κατανομής Βήτα*

Η κανονική Ομοιόμορφη κατανομή (0, 1) είναι μια ειδική περίπτωση της Βήτα. Συγκεκριμένα για $\alpha = \beta = 1$ προκύπτει :

$$f(y) = \frac{y^0(1-y)^0}{B(1,1)} = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} = \frac{(2-1)!}{1} = 1 \qquad (5.3)$$

Σε γενικές γραμμές, μπορούμε να χρησιμοποιήσουμε μια κατανομή Βήτα για να έχουμε μια μη-ομοιόμορφη κατανομή σε κάθε πεπερασμένο χρονικό διάστημα σε πραγματική ευθεία. Για παράδειγμα, εάν $\theta_1 < X < \theta_2$, τότε ορίζεται

$$\Upsilon = \frac{X-\theta_1}{\theta_2-\theta_1}$$

και παρατηρούμε ότι $0 < \Upsilon < 1$. Τότε, είναι θεμιτό ότι Y~Beta(α, β).

5.2 Βασικά χαρακτηριστικά

Αν Y~Beta (α, β), τότε προκύπτει ότι η μέση τιμή ισούται με

$$E[Y] = \int_0^1 \frac{y^a(1-y)^{\beta-1}}{B(\alpha,\beta)}\,dy$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{\alpha}(1-y)^{\beta-1}\,dy$$

$$= \frac{\Gamma(\alpha+1)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+1)}$$

$$= \frac{\alpha}{\alpha+\beta} \tag{5.3}$$

$$E[Y^2] = \frac{1}{B(\alpha,\beta)} \int_0^1 y^{\alpha}(1-y)^{\beta-1}\,dy$$

$$= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)}$$

$$= \frac{(\alpha+1)\alpha}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$E[Y^n] = \frac{(a+n-1)(\alpha+n-2)\ldots(\alpha+1)\alpha}{(\alpha+\beta)(\alpha+\beta+1)\ldots(\alpha+\beta+n-1)}$$

$$= \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n)}$$

και η διακύμανση ισούται με

$$V(Y) = E[Y^2] - (E[Y])^2$$

$$= \frac{(\alpha+1)\alpha}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2}$$

$$= \frac{(\alpha+1)\alpha(\alpha+\beta)-\alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{5.4}$$

5.3 Εκτίμηση παραμέτρων

α. Εκτίμηση Μέγιστης Πιθανοφάνειας

Βρίσκοντας τον λογάριθμο της (5.1) εξίσωσης έχουμε:

$\ln L(\alpha,\beta;\mathbf{x})=n(\ln \Gamma(\alpha+\beta) - \ln\Gamma(\alpha)-\ln\Gamma(\beta))+(\alpha-1)\sum_{i=1}^{n} ln y_i+(\beta-1)\sum_{i=1}^{n}(1-y_i)$

Και το σύστημα των εξισώσεων της μέγιστης πιθανοφάνειας για την εύρεση των $\widehat{\alpha}$ και $\widehat{\beta}$ είναι :

$$\psi(\hat{\alpha})-\psi(\hat{\alpha}+\hat{\beta}) = n^{-1}\sum_{i=1}^{n} ln y_i \ ,$$

$$\psi(\hat{\beta})-\psi\left(\hat{\alpha}+\hat{\beta}\right) = n^{-1}\sum_{i=1}^{n} \ln(1-y_i)$$

όπου $\psi(\cdot)$ είναι η δίγαμμη εξίσωση , που ορίζεται ως

$$\psi(x)=\frac{d}{dx} log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)} \text{ για } x > 0$$

και $\quad y_i = \frac{x_i - p}{q}$

β. Μέθοδος ροπών

Η μέθοδος των ροπών μπορεί να αποκτηθεί επισημαίνοντας την πρώτη και δεύτερη ροπή της κατανομής Βήτα.

$$\mu_1 = \frac{n\alpha}{\alpha+\beta}$$

$$\mu_2 = \frac{n\alpha[n(1+\alpha)+\beta]}{(\alpha+\beta)(1+\alpha+\beta)}$$

εξισώνοντας αυτές τις ροπές με την πρώτη και δεύτερη ροπή ενός δείγματος έχουμε,

$$\widehat{\mu_1}=m_1$$
$$\widehat{\mu_2}=m_2$$

και λύνοντας ως προς α και β παίρνουμε

$$\hat{\alpha}=\frac{nm_1+m_2}{n\left(\frac{m_2}{m_1}-m_1-1\right)+m_1} \quad (5.5)$$

$$\hat{\beta}=\frac{(n-m_1)(n-\frac{m_2}{m_1})}{n\left(\frac{m_2}{m_1}-m_1-1\right)+m_1} \quad (5.6)$$

5.4 Γενικευμένη κατανομή Βήτα

Η γενικευμένη κατανομή Βήτα θεωρείται ως πρότυπο για την κατανομή εισοδήματος. Είναι γνωστό ότι σε ειδικές περιπτώσεις περιλαμβάνει την κατανομή Dagum μαζί με την κατανομή Singh-Maddala. Σχετικά μέτρα ανισότητας όπως ο συντελεστής Gini, ο δείκτης Pietra, ή ο δείκτης Theil εκφράζονται σε όρους των παραμέτρων της γενικευμένης κατανομής Βήτα.

Η κατανομή Βήτα μπορεί εύκολα να γενικευθεί από το διάστημα (0,1) σε ένα αυθαίρετο φραγμένο διάστημα χρησιμοποιώντας γραμμικό μετασχηματισμό. Έτσι, αυτή η γενίκευση είναι απλά η οικογένεια τοποθέτησης κλίμακας που σχετίζεται με το πρότυπο της κατανομής Βήτα.

Ας υποθέσουμε ότι το Z έχει την τυπική κατανομή Βήτα με αριστερή παράμετρο $\alpha > 0$ και δεξιά παράμετρο $\beta > 0$. Για $\gamma \epsilon R$ και $\delta \epsilon (0, \infty)$ τυχαία μεταβλητή $X = \gamma + \delta Z$

έχει την κατανομή Βήτα με αριστερή παράμετρο την α, δεξιά παράμετρο την β , θέση την παράμετρο γ και κλίμακα την παράμετρο δ.

Η X έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{B(\alpha,\beta)\delta^{\alpha+\beta-1}}(x-\gamma)^{\alpha-1}(\gamma+\delta-x)^{\beta-1}, \ x\epsilon\ (\gamma, \gamma+\delta) \quad (5.7)$$

Με X ορίζονται και τα ακόλουθα

- $E(X) = \gamma + \delta \frac{\alpha}{\alpha+\beta}$ (5.8)
- $Var(X) = \delta^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ (5.9)

5.5 Πεδία εφαρμογής

α. Δεδομένα εισοδήματος και πλούτου

Πολλές κατανομές έχουν θεωρηθεί ως περιγραφικά μοντέλα για την κατανομή εισοδήματος. Αυτές περιλαμβάνουν, ανάμεσα σε άλλες, την Lognormal, την Γάμμα, την Βήτα, την Singh-Maddala, την Pareto, και την Weibull . Σε πολλές εφαρμογές, η κατανομή Singh-Maddala ταιριάζει καλύτερα από την Γάμμα και η δεύτερη με την σειρά της ταιριάζει καλύτερα από την Lognormal. Ο Thurow υιοθέτησε την κατανομή Βήτα ως ένα μοντέλο για κατανομή εισοδήματος, και αυτό το μοντέλο περιλαμβάνει την κατανομή Γάμμα ως οριακή περίπτωση, συνεπώς η Βήτα παρέχει εξίσου καλή εφαρμογή με την Γάμμα.

Συγκεκριμένα ο Thurow το 1970 πρότεινε τη χρήση της κατανομής Βήτα τύπου I, μιας κατανομής τριών παραμέτρων ως το μοντέλο για την κατανομή εισοδήματος. Σ'αυτήν την εκτίμηση, ο Thurow εξέτασε στοιχεία εισοδήματος για τις αμερικάνικες οικογένειες το 1949 και χρησιμοποιήθηκε ένα ανώτερο όριο των 15.000 δολλαρίων λόγω του ότι ενδιαφερόταν κυρίως για την εκτίμηση της κατανομής εισοδήματος για το κάτω άκρο. Βρήκε ότι η κατανομή Βήτα ταιριάζει με τις πραγματικές κατανομές εισοδήματος αρκετά καλά.

Σε μια συγκεκριμένη μελέτη ο Mc Donald πρότεινε τη χρήση της γενικευμένης κατανομής Βήτα του πρώτου και δεύτερου τύπου ως κατάλληλη για την κατανομή του εισοδήματος (1984). Πρότεινε ότι, δεδομένου αυτές οι κατανομές περιλάμβαναν την Βήτα του πρώτου και δεύτερου τύπου (B1, B2), την Γάμμα (GA), την Singh-Maddala (SM), την λογαριθμοκανονική (LN), και κάποιες άλλες κατανομές σε ειδικές περιπτώσεις, παρέχουν τουλάχιστον τόσο καλή φόρμα, όπως οι B1, B2, GA, και LN.

Σε μια μεταγενέστερη μελέτη, οι McDonald και Xu (1995) έδειξαν ότι η γενικευμένη κατανομή Βήτα στην οποία υπάρχουν όλες οι παραπάνω κατανομές ως ειδικές περιπτώσεις είναι πολύ πιο ευέλικτη και έτσι καλύτερα "ταιριάζει" για να περιγράψει κατανομές εισοδήματος. Οι Bordley και McDonald (1996) έχουν επίσης εξετάσει τη σχετική απόδοση ορισμένων από αυτές τις κατανομές (βλ. [38]).

β.Ασφαλιστικές ζημιές

Στο βιβλίο <<Insurance : Mathematics and Economics>> των J.David Cummins, Georges Dionne και James B. McDonald, B.Michael Pritchett διερευνάται η χρήση μιας οικογένειας κατανομών πιθανότητας τεσσάρων παραμέτρων, η γενικευμένη κατανομή Βήτα τύπου II για τις διαδικασίες μοντελοποίησης ασφαλιστικών ζημιών. Η γενικευμένη κατανομή Βήτα τύπου II περιλαμβάνει πολλές κατανομές που χρησιμοποιούνται συχνά όπως η Lognormal, η Gamma και η Weibull. Η GB2 περιλαμβάνει επίσης την Burr και τις γενικευμένες κατανομές Γάμμα . Μέλη αυτής της οικογένειας κατανομών και οι αντίστροφές τους έχουν σημαντικές δυνατότητες για τη βελτίωση της προσαρμογής της κατανομής σε πολλές εφαρμογές που περιλαμβάνουν λεπτές ή βαριές ουρές κατανομών. Μέλη της οικογένειας της GB2 μπορεί να δημιουργηθεί ως μείγματα γνωστών κατανομών και παρέχει ένα πρότυπο ανομοιογένειας στις αξιώσεις των κατανομών. Επίσης παρουσιάζονται παραδείγματα, μοντέλων κατανομών ατομικών και των συνολικών ζημιών. Τα αποτελέσματα δείχνουν ότι οι φαινομενικά μικρές διαφορές στην μοντελοποίηση των ουρών μπορεί να οδηγήσουν σε μεγάλες διαφορές μεταξύ των ασφαλίστρων αντασφάλισης και των ποσοστημορίων στην κατανομή των συνολικών ασφαλιστικών ζημιών.

γ. Μεγέθη επιχειρήσεων

Σε μια πρόσφατη εργασία, οι Cabral και Mata (2003) [Από την εξέλιξη της κατανομής του μεγέθους της επιχείρησης: γεγονότα και θεωρία. American Economic Review 93, 1075-1090] για τις πορτογαλικές επιχειρήσεις, είχαν ως στόχο να εξακριβώσουν εάν η κατανομή του μεγέθους των νέων επιχειρήσεων (κάτω των 5 ετών) είναι αισθητά διαφορετική από εκείνη των παλαιότερων επιχειρήσεων (άνω των 30 ετών). Για την εκτέλεση της ανάλυση τους χρησιμοποίησαν έναν πολύ ολοκληρωμένο βιομηχανικό πίνακα, με περίπου 25k επιχειρήσεις για είκοσι χρόνια παρατηρήσεων. Όσον αφορά τα αποτελέσματα, εξακρίβωσαν μια σαφή διαφορά στην κατανομή του μεγέθους των επιχειρήσεων ανάλογα με την ηλικία, για τις οποίες έδωσαν μια καλή προσαρμογή χρησιμοποιώντας τη γενικευμένη κατανομή Βήτα του δεύτερου τύπου (βλ. [10]).

ΚΕΦΑΛΑΙΟ 6

Κατανομή Dagum

6.1 Ορισμός

Στη δεκαετία του 1970, ο Camilo Dagum ξεκίνησε την αναζήτηση για μια στατιστική κατανομή που να προσαρμόζεται πολύ καλά με τις εμπειρικές κατανομές εισοδήματος και πλούτου. Δεν ήταν ικανοποιημένος με τις κλασικές κατανομές που χρησιμοποιούνταν για να συνοψίσουν τα δεδομένα αυτά,η κατανομή Pareto (που αναπτύχθηκε από τον Ιταλό οικονομολόγο Vilfredo Pareto στα τέλη του $19^{ου}$ αιώνα (1895,1896,1897)) και η λογαριθμοκανονική κατανομή (διαδόθηκε από τον Γάλλο μηχανικό Robert Gibrat (1931)), κοίταξε για ένα μοντέλο ελαστικό στις βαριές ουρές που υπάρχουν στις εμπειρικές κατανομές εισοδήματος και πλούτου αλλά και να επιτρέπει μια εσωτερική λειτουργία. Η πρώτη πτυχή συλλήφθηκε από τον Pareto αλλά όχι από την λογαριθμοκανονική κατανομή, η τελευταία από την λογαριθμοκανονική κατανομή αλλά όχι από την κατανομή Pareto. Ο πειραματισμός από μια μετατοπισμένη κατανομή log-logistic (Dagum, 1975) αλλά και η γενίκευση της κατανομής προηγουμένως από τον Fisk (1961), τον έκαναν γρήγορα να συνειδητοποιήσει ότι ήταν αναγκαία μια περαιτέρω παράμετρος. Αυτό οδήγησε στην κατανομή Dagum τύπου I, μιας τριών παραμέτρων κατανομή και τεσσάρων παραμέτρων γενικεύσεις (Dagum, 1977, 1980) (βλ. [3]).

Χρειάστηκε πάνω από μια δεκαετία μέχρι ο Dagum να αρχίσει να εμφανίζεται στην οικονομική και οικονομετρική λογοτεχνία της αγγλικής γλώσσας. Η πρώτη εργασία, σε μια σημαντική εφημερίδα οικονομετρίας αξιοποιώντας τη κατανομή Dagum, εμφανίστηκε από τους Majumder και Chakravarty (1990). Στη στατιστική βιβλιογραφία, η κατάσταση είναι πιο ευνοϊκή, από το γεγονός ότι η περίφημη Εγκυκλοπαίδεια των Στατιστικών Επιστημών περιέχει, στο τόμο 4 (Kotz,1983), μια καταχώρηση με μοντέλα κατανομής του εισοδήματος, όπως ήταν αναμενόμενο συνταχθεί από τον Camilo Dagum (Dagum, 1983). Εκ των υστέρων, ο λόγος για αυτή τη μακρά καθυστέρηση είναι αρκετά προφανής: (1977) το έγγραφο του Dagum δημοσιεύθηκε στην Economie Appliqu'ee, μια γαλλική εφημερίδα με μόνο περιστασιακή συνεισφορές στην αγγλική γλώσσα και αρκετά περιορισμένης κυκλοφορίας στις αγγλόφωνες χώρες.

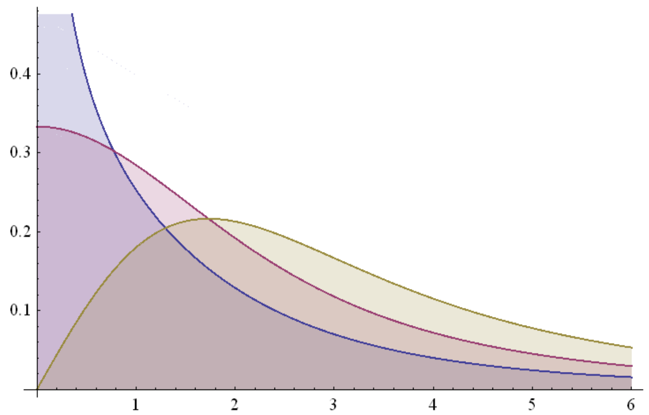
Η κατανομή Dagum είναι μια ειδική περίπτωση της γενικευμένης κατανομής Βήτα τύπου ΙΙ. Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Dagum έχει την εξής μορφή :

$$\mathrm{f(x)} = \frac{apx^{\alpha p-1}}{b^{\alpha p}[1+(\frac{x}{b})^{\alpha}]^{p+1}} \quad , \mathrm{x} > 0 \tag{6.1}$$

όπου α,b,p >0 και η συνάρτηση κατανομής είναι η

$$\mathrm{F(x)} = [1 + (\frac{x}{b})^{-\alpha}]^{-p} \quad , \mathrm{x} > 0 \tag{6.2}$$

όπου α,b,p > 0 με b να είναι η παράμετρος κλίμακας και α και p οι παράμετροι σχήματος.



Διάγραμμα 6.1

Συνάρτηση πυκνότητας πιθανότητας της κατανομής Dagum

6.2 Βασικά χαρακτηριστικά

Η κ-ροπή της Dagum κατανομής ισχύει για $\kappa < \alpha$ και $\kappa, \alpha, b > 0$ έτσι ώστε

$$E(X)^{\kappa} = \frac{b^{\kappa}\Gamma\left(p+\frac{\kappa}{\alpha}\right)\Gamma\left(1-\frac{\kappa}{\alpha}\right)}{\Gamma(p)} \tag{6.3}$$

όπου Γ(·) δηλώνει στην συνάρτηση Γάμμα.

Ειδικότερα η μέση τιμή και η διακύμανση ορίζονται ως :

$$E(X) = \frac{b\Gamma\left(p+\frac{1}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right)}{\Gamma(p)} \tag{6.4}$$

$$Var(X) = \frac{b^2\left[\Gamma(p)\Gamma\left(p+\frac{2}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)-\Gamma^2\left(p+\frac{1}{a}\right)\Gamma^2\left(1-\frac{1}{a}\right)\right]}{\Gamma^2(p)} \tag{6.5}$$

6.3 Εκτίμηση παραμέτρων

α. Εκτίμηση Μέγιστης Πιθανοφάνειας

Η αυξανόμενη διαθεσιμότητα μικροστοιχείων και η εκτίμηση πιθανοφάνειας από μεμονωμένες παρατηρήσεις προσελκύουν όλο και περισσότερο την προσοχή του Dagum, ο οποίος τις χρησιμοποιεί σε ένα από τα πρώτα παραδείγματά του, στα οικογενειακά εισοδήματα των ΗΠΑ για το 1969. Η λογαριθμική πιθανοφάνεια $l(\theta) \equiv \log L(\theta)$ για ένα πλήρες τυχαίο δείγμα μεγέθους n είναι

$$l(\alpha,b,p) = n\ln\alpha + n\ln p + (\alpha p-1)\sum_{i=1}^{n} \ln x_i - n\alpha p\ln b - (p+1)\sum_{i=1}^{n} \ln\{1+(\frac{x_i}{b})^{\alpha}\}$$

αποδίδοντας τις εξισώσεις πιθανότητας

$$\frac{n}{\alpha} + p\sum_{i=1}^{n} \ln\left(\frac{x_i}{b}\right) = (p+1)\sum_{i=1}^{n} \frac{\ln(\frac{x_i}{b})}{1+(\frac{b}{x_i})^{\alpha}}$$

$$\mathrm{np} = (\mathrm{p}+1)\sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{1+(\frac{\mathrm{b}}{\mathrm{x_i}})^{\alpha}}$$

$$\frac{n}{p} + \alpha\sum_{i=1}^{n} \ln\left(\frac{x_i}{b}\right) = \sum_{i=1}^{n} \ln\{1+\left(\frac{x_i}{b}\right)^{\alpha}\}$$

β. Πληροφορία Fisher

Για την εκτίμηση παραμέτρων μεγάλου μεγέθους δείγματος χρησιμοποιείται η πληροφορία Fisher :

$$\mathrm{I}(\theta) = [-\mathrm{E}(\frac{d^2 logL}{d\theta_i d\theta_j})_{i,j}] =: \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}$$

και παίρνει την μορφή

$$I_{11} = \frac{1}{\alpha^2(2+p)}[\mathrm{p}[\{\psi(\mathrm{p})-\psi(1)-1\}^2 + \psi'(\mathrm{p})+\psi'(1)]+2\{\psi(\mathrm{p})-\psi(1)\}]$$

$$I_{21} = I_{12} = \frac{p-1-p\{\psi(p)-\psi(1)\}}{b(2+p)}$$

$$I_{22} = \frac{\alpha^2 p}{b^2(2+p)}$$

$$I_{23} = I_{32} = \frac{a}{b(1+p)}$$

$$I_{31} = I_{13} = \frac{\psi(2)+\psi(p)}{\alpha(1+p)}$$

$$I_{33} = \frac{1}{p^2}$$

Όπου ψ είναι η δίγαμμη συνάρτηση.

Θα πρέπει να σημειωθεί ότι υπάρχουν πολλά παράγωγα της πληροφορίας Fisher στην στατιστική λογοτεχνία, η μία χρησιμοποιεί την παραμετροποίηση του Dagum που οφείλεται στον Latorre (1988) και μία δεύτερη που οφείλεται στον Zelterman (1987).

6.4 Κατανομές Dagum τύπου II και III

Ο Dagum το 1977 και το 1980 εισήγαγες άλλες δύο παραλλαγές της κατανομής του: τις κατανομές Dagum τύπου I και II. Η κατανομή τύπου II έχει συνάρτηση κατανομής την

$$F(x) = \alpha + (1\text{-}\alpha)[1 + \left(\frac{x}{b}\right)^{-a}]^{-p}, \quad x \geq 0, \qquad (6.6)$$

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

Η κατανομή τύπου II προτάθηκε ως μοντέλο για κατανομές εισοδήματος με μηδενικά και αρνητικά εισοδήματα αλλά κυρίως για να ταιριάζει σε κατανομές πλούτου οι οποίες συχνά παρουσιάζουν μηδενικό ακαθάριστο ενεργητικό αλλά και αρνητικό.

Υπάρχει επίσης και μια κατανομή Dagum τύπου III που ορίζεται ως

$$F(x) = \alpha + (1\text{-}\alpha)[1 + \left(\frac{x}{b}\right)^{-a}]^{-p}, \quad x \geq 0, \qquad (6.7)$$

όπου πάλι $\alpha, b, p > 0$ αλλά $\alpha < 0$.

6.5 Πεδία εφαρμογής

α .Δεδομένα εισοδήματος

Παρά το γεγονός ότι η κατανομή Dagum ήταν σχεδόν άγνωστη στη μεγάλη αγγλική γλώσσα των οικονομικών και οικονομετρικών εφημερίδων μέχρι και τη δεκαετία του 1990, υπάρχουν αρκετές εφαρμογές σε δεδομένα εισοδήματος και πλούτου, τα περισσότερα εκ των οποίων εμφανίστηκαν σε γαλλικές, ιταλικές και της λατινικής Αμερικής δημοσιεύσεις. Παραδείγματα περιλαμβάνουν τους Fattorini και Lemmi (1979), οι οποίοι χρησιμοποίησαν ιταλικά δεδομένα, οι Espinguet και Terraza (1983), οι οποίοι μελετούσαν γαλλικά κέρδη και ο Falcao Carneiro (1982) με μια εφαρμογή σε πορτογαλικά δεδομένα..

Ιδιαίτερο ενδιαφέρον παρουσιάζουν δημοσιεύσεις διάφορων κατανομών με τα ίδια δεδομένα, με το βλέμμα στραμμένο στην σχετική απόδοση. Από συγκριτικές μελέτες, όπως των McDonald και Xu (1995), Bordley,McDonald και Mantrala (1996), Bandourian, McDonald και Turley (2003) και Azzalini,Dal Capello και Kotz (2003) προκύπτει ότι η κατανομή Dagum συνήθως υπερτερεί των ανταγωνιστών της, εκτός από την GB2 γενικευμένη Βήτα κατανομή η οποία έχει μια επιπλέον παράμετρο. Οι Bandourian, McDonaldκαι Turley (2003), βρήκαν, σε μια μελέτη ότι χρησιμοποιώντας 82 σύνολα δεδομένων, η Dagum είναι η καλύτερη κατανομή τριών παραμέτρων σε όχι λιγότερες από 84% των περιπτώσεων.

Επίσης ο Dagum (1977, 1980) εφάρμοσε την κατανομή τύπου ΙΙ στα οικογενειακά εισοδήματα των ΗΠΑ των ετών 1960 και 1969, το μοντέλο του οποίου υπερέχει των εξισώσεων των κατανομών Γάμμα και Λογαριθμοκανονικής και στην εγκυκλοπαίδεια των Στατιστικών Επιστημών στις αρχές του 1983 τοποθέτησε την κατανομή Dagum τύπου Ι και ΙΙΙ . Αργότερα το 1989, οι Dagum και Lemmi εφάρμοσαν τις κατανομές Ι-ΙΙΙ του Dagum σε δεδομένα εισοδήματος της Ιταλίας από τις δειγματοληπτικές έρευνες της Banca d'Italia του 1977, 1980 και του 1984, όπου η εφαρμογή ήταν σε γενικές γραμμές αρκετά ικανοποιητική. Τα στοιχεία διαχωρίστηκαν ανά φύλο, περιοχή, και πηγή εισοδήματος.

Η Victoria-Feser (1995, 2000) εφάρμοσε ένα είδος κατανομής Dagum Ι για εισοδήματα νοικοκυριών που λαμβάνουν κοινωνικά επιδόματα χρησιμοποιώντας το 1979 στο Ηνωμένο Βασίλειο την Έρευνα Οικογενειακών Δαπανών (FES) και τα εισοδήματα από το 1985, του Ηνωμένου Βασιλείου FES.

Ο Bantilan (1995) μοντελοποίησε τα εισοδήματα από το οικογενειακό εισόδημα στις Φιλιππίνες για το 1957, το 1961, το 1965, το 1971, το 1985 και το 1988, χρησιμοποιώντας μια κατανομή τύπου I Dagum. Σημείωσε ότι το μοντέλο ταιριάζει με τα δεδομένα αρκετά καλά, ιδιαίτερα στις ουρές. Οι Bordley, McDonald, και Mantrala (1996) εφάρμοσαν στην τύπου I κατανομή Dagum στις ΗΠΑ τα οικογενειακά εισοδήματα για το 1970, 1975, 1980, 1985, και 1990. Για όλα τα σύνολα δεδομένων αποδείχθηκε ότι είναι το καλύτερο μοντέλο τριών παραμέτρων, όντας κατώτερο μόνο της GB2 κατανομής .

Οι Botargues και Petrecolla (1997, 1999) εκτίμησαν τα Dagum μοντέλα I-III για την κατανομή του εισοδήματος στην περιοχή Μπουένος Άιρες, για όλα τα έτη 1990 έως 1997. Διαπίστωσαν ότι τα μοντέλα Dagum ξεπερνούν λογαριθμοκανονική κατανομή και την Singh-Maddala (βλ. [7]).

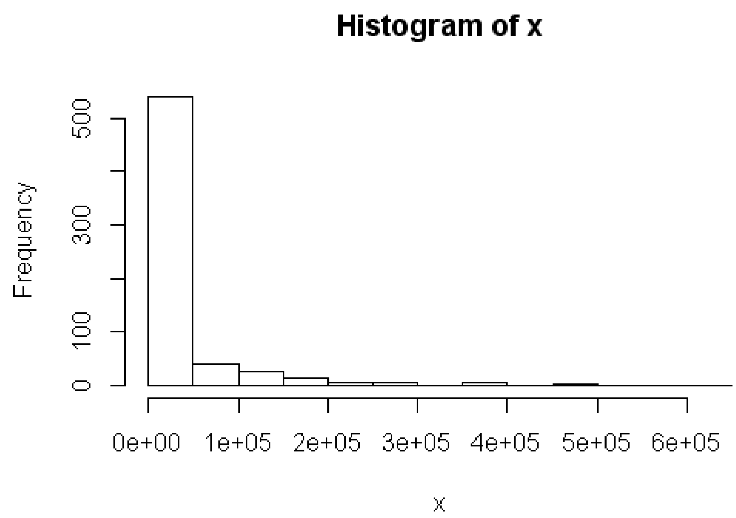
β . Αναλογιστικές ζημιές

Κατά την αναλογιστική βιβλιογραφία, ο Cummins και άλλοι (1990) εφάρμοσαν την κατανομή Dagum τύπου I (κάτω από το όνομα της Burr III) για τον υπολογισμό των συνολικών ζημιών πυρός από τους Cummins και Freifelder (1978). Η κατανομή αποδίδει αρκετά καλά, ωστόσο, η πλήρης ευελιξία της δεν απαιτείται και σε μια οριακή περίπτωση της μια παραμέτρου, η αντίστροφη εκθετική, είναι απολύτως επαρκής. Στην ίδια δημοσίευση αναφέρθηκαν επίσης δεδομένα σχετικά με τη σοβαρότητα των ζημιών πυρός και εφαρμόστηκαν στην GB2 και στις υποοικογένειες της τόσο ως ομαδοποιημένες αλλά και ως μεμονωμένες παρατηρήσεις. Παρά το γεγονός ότι η προσαρμογή του Dagum μοντέλου είναι πολύ καλή, αυτή της κατανομής Singh-Maddala είναι ελαφρώς καλύτερη.

ΚΕΦΑΛΑΙΟ 7

Αριθμητική εφαρμογή

Στο πρακτικό κομμάτι της παρούσας διπλωματικής χρησιμοποιήσαμε δεδομένα ζημιών που προέρχονται από χαρτοφυλάκιο ζημιών αυτοκινήτων εταιρίας ασφαλιστικών εργασιών με έδρα την Ελλάδα. Συγκεκριμένα τα δεδομένα μας αφορούν ασφαλιστικές ζημιές χρονικού διαστήματος 2009-2014. Όλοι οι υπολογισμοί έγιναν με τη βοήθεια της R και οι μέθοδοι που χρησιμοποιήσαμε για την εκτίμηση των παραμέτρων είναι οι μέθοδοι μέγιστης πιθανοφάνειας και ροπών πιθανότητας.



Διάγραμμα 7.1.

Ασφαλιστικές ζημιές των ετών 2009-2014

Στο παραπάνω ιστόγραμμα (**7.1**) που απεικονίζει τις τιμές του x, δηλαδή τις ασφαλιστικές ζημιές, παρατηρούμε ότι ακολουθούν κατανομή με βαριά ουρά. Στη συνέχεια θα προσπαθήσουμε να προσαρμόσουμε τα παραπάνω δεδομένα σε διάφορες κατανομές για τις οποίες θα γίνει τόσο στατιστικός (στατιστικός έλεγχος Kolmogorov-Smirnov και Anderson-Darling) όσο και γραφικός έλεγχος (σύγκριση της εκάστοτε θεωρητικής κατανομής με την εμπειρική κατανομή των δεδομένων μας).

Στατιστικός έλεγχος υποθέσεων

Στον στατιστικό έλεγχο υποθέσεων μας ενδιαφέρει να ελέγξουμε αν μια παράμετρος του πληθυσμού (για παράδειγμα η μέση τιμή ή η διακύμανση) ικανοποιεί μια υπόθεση (H_0→μηδενική) έναντι μιας εναλλακτικής υπόθεσης (H_1 → εναλλακτική). Σε αυτό το πρακτικό κομμάτι θα χρησιμοποιήσουμε δύο πολύ σημαντικούς στατιστικούς ελέγχους , αυτούς του Kolmogorov-Smirnov και του Anderson-Darling.

Kolmogorov-Smirnov test

Στη στατιστική, το Kolmogorov-Smirnov test είναι ένα μη παραμετρικό τεστ της ισότητας των συνεχών, μονοδιάστατων κατανομών πιθανοτήτων που μπορεί να χρησιμοποιηθεί για να συγκρίνει ένα δείγμα με αναφορά στην κατανομή πιθανοτήτων (one-sample K -S test), ή να συγκρίνει δύο δείγματα (two-sample K-S test). Το Kolmogorov-Smirnov test ποσοτικοποιεί μια απόσταση μεταξύ της εμπειρικής συνάρτησης κατανομής του δείγματος και της αθροιστικής συνάρτησης κατανομής της κατανομής αναφοράς, ή μεταξύ των εμπειρικών συναρτήσεων της κατανομής των δύο δειγμάτων. Η μηδενική κατανομή αυτής της στατιστικής υπολογίζεται με την μηδενική υπόθεση ότι τα δείγματα προέρχονται από την ίδια κατανομή (στην περίπτωση δύο δειγμάτων) ή ότι το δείγμα προέρχεται από την κατανομή αναφοράς (στην περίπτωση ενός δείγματος). Σε κάθε περίπτωση, οι κατανομές θεωρούνται κάτω από τη μηδενική υπόθεση ότι είναι συνεχείς κατανομές αλλιώς μη συνεχείς.

Το two-sample K-S test είναι ένα από τα πιο χρήσιμα και γενικευμένα για τη σύγκριση δύο δειγμάτων καθώς είναι ευαίσθητα σε διαφορές τόσο στην θέση, όσο και στο σχήμα των εμπειρικών αθροιστικών συναρτήσεων κατανομών των δύο δειγμάτων. Το τεστ Kolmogorov-Smirnov μπορεί να τροποποιηθεί για να χρησιμεύσει ως ένα τεστ καλής προσαρμογής. Στην ειδική περίπτωση των τεστ για κανονικότητα της κατανομής, τα δείγματα είναι τυποποιημένα και συγκρίσιμα με την τυπική κανονική κατανομή.

Anderson-Darling test

Το Anderson-Darling τεστ είναι ένα στατιστικό τεστ που ελέγχει αν ένα δείγμα των δεδομένων προέρχεται από μια δεδομένη κατανομή πιθανοτήτων. Στην βασική του μορφή, το τεστ υποθέτει ότι δεν υπάρχουν παράμετροι για να εκτιμηθούν στη κατανομή που δοκιμάζεται, στην οποία περίπτωση το τεστ και το σύνολο των κρίσιμων τιμών είναι χωρίς κατανομή. Ωστόσο, το τεστ πιο συχνά χρησιμοποιείται σε μια οικογένεια κατανομών, στην οποία οι παράμετροι αυτής της οικογένειας θα πρέπει να εκτιμώνται. Όταν εφαρμόζεται σε τεστ για το αν μια κανονική κατανομή περιγράφει επαρκώς ένα σύνολο δεδομένων, τότε είναι ένα από τα πιο ισχυρά εργαλεία .

Στον παρακάτω πίνακα (**7.1**) δίνονται κάποια περιγραφικά δεδομένα όπως η Μέση τιμή και η Τυπική απόκλιση. Όπως παρατηρούμε η τυπική απόκλιση είναι πολύ μεγαλύτερη από την μέση τιμή πράγμα που μαρτυρά την ύπαρξη κατανομής με βαριά ουρά.

Min	1o ποσοστημόριο	Median	Mean	3o ποσοστημόριο	Max
1000	12320	16120	40910	30090	627900

Πίνακας 7.1. Πίνακας περιγραφικών των δεδομένων μας

Διακύμανση	
Τυπική απόκλιση	72798,9
Κύρτωση	23,1604
Λοξότητα	4,42622

Πίνακας 7.2. Πίνακας λοξότητας και κυρτότητας των δεδομένων μας

Η κατανομή Pareto είναι μια από τις κατανομές που χρησιμοποιούνται για την περιγραφή δεδομένων απωλειών. Στο κεφάλαιο 2 ορίσαμε την κατανομή ως εξής :

$$\bar{F}(\mathrm{x}) = \mathrm{Pr}(\mathrm{X} > \mathrm{x}) = \begin{cases} 1 - (\frac{x_0}{x})^{\alpha} & x \geq x_0 \\ 0 & x \geq x_0 \end{cases}$$

Έχουμε λοιπόν την μηδενική και την εναλλακτική υπόθεση,

H_0 : η κατανομή κάνει καλή προσαρμογή στα δεδομένα

H_1 : η κατανομή δεν κάνει καλή προσαρμογή στα δεδομένα

Όταν το p-value είναι μικρότερο του **0.05** τότε απορρίπτουμε την μηδενική υπόθεση ενώ αν είναι μεγαλύτερο τότε δεχόμαστε την μηδενική υπόθεση.

Παρακάτω παραθέτουμε πίνακα στον οποίο φαίνεται η εκτίμηση παραμέτρων με τις δύο μεθόδους.

.

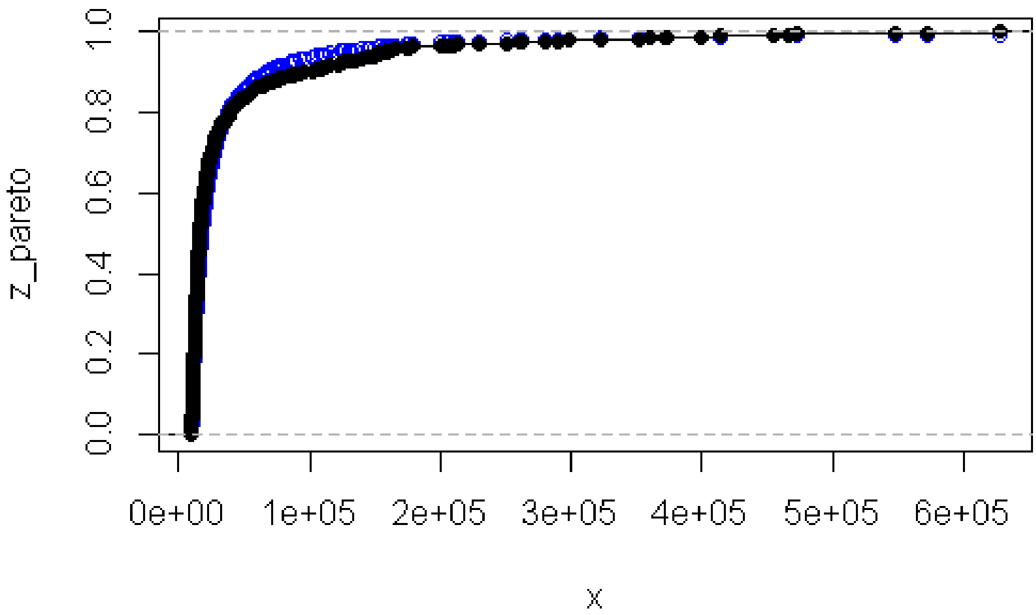
Κατανομή Pareto

Έχουμε τον τύπο του εκτιμητή μέγιστης πιθανοφάνειας από το κεφάλαιο 2 :

$$\hat{\alpha} = \mathrm{n} / \sum_{i=1}^{n} \ln(\frac{\mathrm{X_i}}{\hat{k}})$$

Μέθοδος μέγιστης πιθανοφάνειας	
Παράμετρος	**Τιμή**
Δείκτης ουράς	1.211
Παράμετρος κλίμακας	1000

Πίνακας 7.3. Εκτίμηση παραμέτρων για την κατανομή Pareto με τη μέθοδο μέγιστης πιθανοφάνειας



Διάγραμμα 7.2. Σύγκριση θεωρητικής και εμπειρικής αθροιστικής κατανομής για την κατανομή Pareto

Στο διάγραμμα **7.2.** απεικονίζεται η θεωρητική και εμπειρική αθροιστική κατανομή της Pareto κατανομής. Παρατηρούμε ότι υπάρχει μια μικρή απόσταση μεταξύ των δύο κατανομών που μας υποψιάζει ότι κατά πάσα πιθανότητα η κατανομή Pareto δεν κάνει καλή προσαρμογή στα δεδομένα μας.

Anderson-Darling GoF Test
p-value = 9.288e-07

Πίνακας 7.4 Στατιστικός έλεγχος Anderson –Darling για κατανομή Pareto

One-sample Kolmogorov-Smirnov test	
D=τιμή της ελεγχοσυνάρτησης	0.0696
p-value	0.003811

Πίνακας 7.5. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Pareto

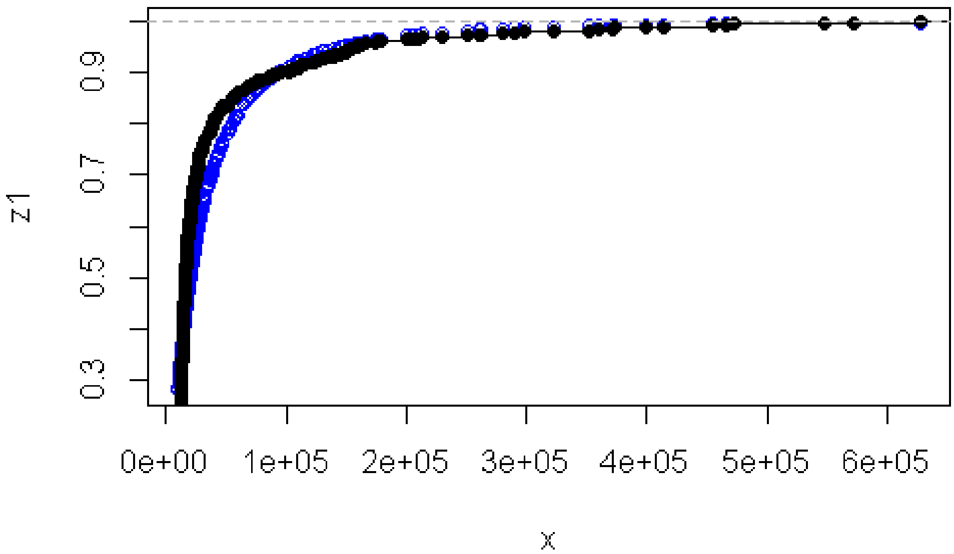
Παρατηρώντας τα p-value και στους δύο παραπάνω πίνακες (7.4 , 7.5) βλέπουμε ότι και τα δύο είναι < 0.05 άρα δεν κάνει η Pareto καλή προσαρμογή στα δεδομένα μας

Ο εκτιμητής της κατανομής Pareto με την μέθοδο των ροπών είναι ο εξής:

$$\hat{\alpha} = \frac{\bar{X}}{\bar{X} - \hat{k}}$$

Μέθοδος ροπών	
Παράμετρος	**Τιμή**
Δείκτης ουράς	1.323009
Παράμετρος κλίμακας	9988.29

Πίνακας 7.6. Εκτίμηση παραμέτρων για την κατανομή Pareto με τη μέθοδο ροπών



Διάγραμμα 7.3. Σύγκριση θεωρητικής και εμπειρικής αθροιστικής κατανομής για την κατανομή Pareto

Στο διάγραμμα 7.3. παρουσιάζεται μια πολύ μικρή απόσταση ανάμεσα στην θεωρητική και εμπειρική αθροιστική συνάρτηση.

Anderson-Darling GoF Test
p-value = 0.01459

Πίνακας 7.7. Στατιστικός έλεγχος Anderson-Darling για κατανομή Pareto

One-sample Kolmogorov-Smirnov test	
D	0.0539
p-value	0.04676

Πίνακας 7.8. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Pareto

Όπως είδαμε παραπάνω , έτσι με στην μέθοδο ροπών χρησιμοποιώντας τους δύο στατιστικούς ελέγχους παρατηρούμε ότι τα p-value < 0.05 άρα η Pareto δεν κάνει καλή προσαρμογή στα δεδομένα μας.

- **Λογαριθμοκανονική κατανομή**

Η δεύτερη κατανομή που αναφέρουμε στην παρούσα διπλώματικη (κεφάλαιο 3) για την περιγραφή απωλειών είναι η λογαριθμοκανονική κατανομή με συνάρτηση κατανομής

$$F(x) = \Phi[\frac{\ln(x)-\mu}{\sigma}],\ x\epsilon(0,\infty)$$

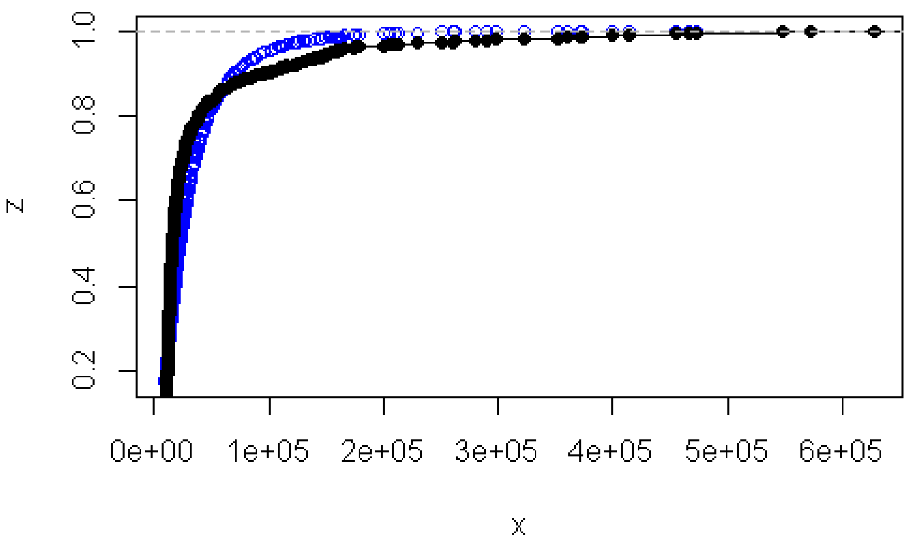
Και σε αυτήν την κατανομή χρησιμοποιήσαμε για την εκτίμηση παραμέτρων τις παρακάτω μεθόδους. Με την μέθοδο της μέγιστης πιθανοφάνειας καταλήξαμε στους παρακάτω εκτιμητές.

$$\hat{\mu}=\frac{\sum_{i=1}^{n}\ln(X_i)}{n}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\ln(X_i)-\frac{\sum_{i=1}^{n}\ln(X_i)}{n})^2}{n}$$

Μέθοδος μέγιστης πιθανοφάνειας	
Παράμετρος	**Τιμή**
Δείκτης ουράς	10.0355433
Παράμετρος κλίμακας	0.8789527

Πίνακας 7.9. Εκτίμηση παραμέτρων για την κατανομή Lognormal με τη μέθοδο μέγιστης πιθανοφάνειας



Διάγραμμα 7.4. Σύγκριση θεωρητικής και εμπειρικής αθροιστικής κατανομής για την κατανομή Lognormal

Anderson-Darling GoF Test
43.0318316

Πίνακας 7.10. Στατιστικός έλεγχος Anderson-Darling για κατανομή Lognormal

One-sample Kolmogorov-Smirnov test	
p-value	0.1796266

Πίνακας 7.11. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Lognormal

Στην lognormal κατανομή και με τους δύο ελέγχους Kolmogorov-Smirnov και Anderson-Darling παρατηρούμε ότι τα p-value είναι > 0.05 άρα η Lognormal κάνει καλή προσαρμογή στα δεδομένα μας.

Οι εκτιμητές με την μέθοδο των ροπών είναι οι εξής :

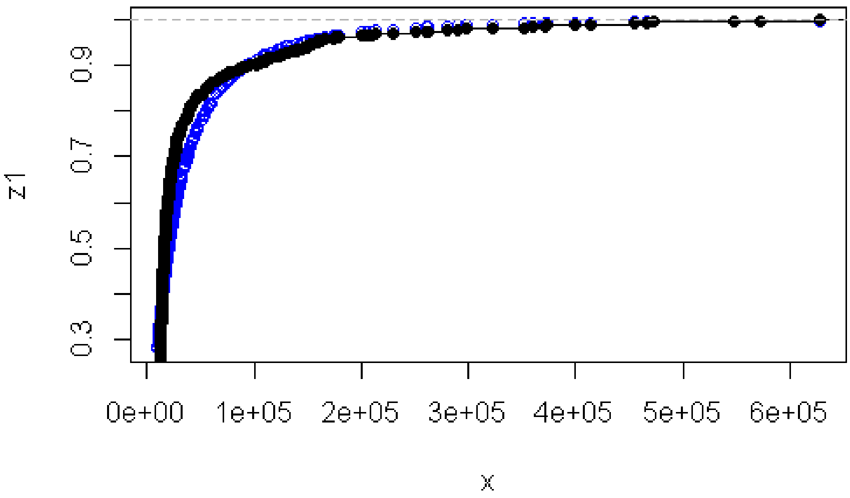
$$\tilde{\mu} = 2\ln\left(\sum_{i=1}^{n} X_i\right) - \frac{3}{2}\ln(n) - \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right)}{2}$$

$$\widetilde{\sigma^2} = \ln\left(\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\sum_{i=1}^{n} X_i\right) + \ln(n)$$

Μέθοδος ροπών	
Παράμετρος	**Τιμή**
Δείκτης ουράς	9.906213
Παράμετρος κλίμακας	1.194.103

Πίνακας 7.12. Εκτίμηση παραμέτρων για την κατανομή Lognormal με τη μέθοδο ροπών

Συγκρίνοντας τα δύο παραπάνω διαγράμματα καταλήγουμε στο συμπέρασμα ότι η κατανομή Lognormal προσαρμόζεται πιο καλά στα δεδομένα μας με την μέθοδο των ροπών.



Διάγραμμα 7.5. Σύγκριση θεωρητικής και εμπειρικής αθροιστικής κατανομής για την κατανομή Lognormal

Στο διάγραμμα 7.5. βλέπουμε ότι η γραμμή της θεωρητικής με την γραμμή της εμπειρικής αθροιστικής κατανομής σχεδόν συμπίπτουν.

Anderson-Darling GoF Test
49.8765814

Πίνακας 7.13. Στατιστικός έλεγχος Anderson-Darling για κατανομή Lognormal

One-sample Kolmogorov-Smirnov test	
p-value	0.2800283

Πίνακας 7.14. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Lognormal

Ομόιως με τα p-value των πινάκων 7.10 και 7.11 και εδώ τα p-value > 0.05 άρα και με την μέθοδο των ροπών καταλήγουμε στο ίδιο συμπέρασμα ότι η Lognormal κάνει καλή προσαρμογή στα δεδομένα μας.

- **Κατανομή Γάμμα**

Μια επίσης σημαντική κατανομή για την περιγραφή απωλειών είναι η κατανομή Γάμμα όπως αναφέραμε και στο κεφάλαιο 4, με συνάρτηση

$$\Gamma(\alpha)=\int_0^{\infty} x^{\alpha-1} e^{-x}\, \mathrm{dx}$$

$$\psi(\alpha)=\frac{d}{d\alpha}\ln\Gamma(\alpha)$$

Μέθοδος μέγιστης πιθανοφάνειας	
Παράμετρος	**Τιμή**
shape	3.18203,6e-01
rate	7.80169e-06

Πίνακας 7.15. Εκτίμηση παραμέτρων για την κατανομή Γάμμα με τη μέθοδο μέγιστης πιθανοφάνειας

Anderson-Darling GoF Test
135.69987

Πίνακας 7.16. Στατιστικός έλεγχος Anderson-Darling για κατανομή Γάμμα

One-sample Kolmogorov-Smirnov test	
p-value	0.50203

Πίνακας 7.17. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Γάμμα

Παρατηρούμε ότι και με τους δύο στατιστικούς ελέγχους τα p-value είναι > 0.05 άρα η κατανομή Γάμμα κάνει καλή προσαρμογή στα δεδομένα μας.

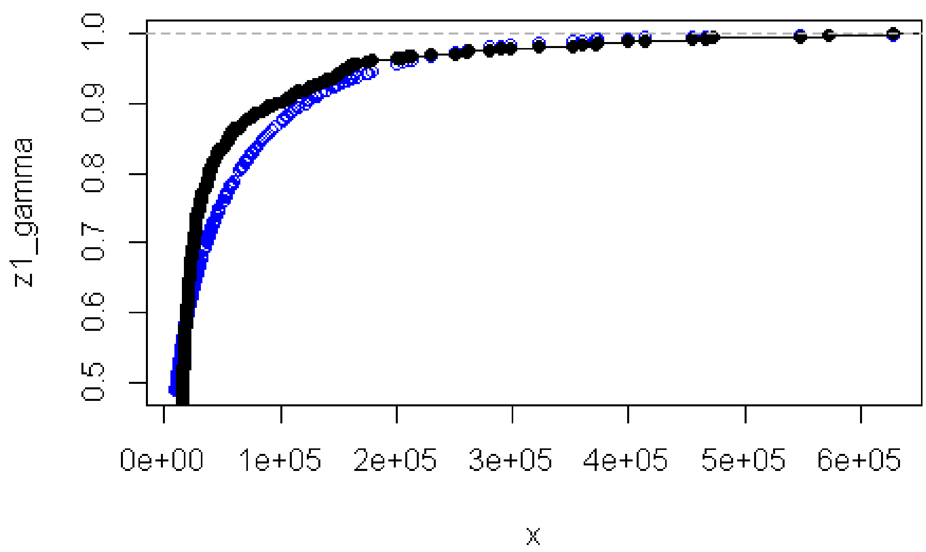
Οι εκτιμήτες με την μέθοδο των ροπών της κατανομής Γάμμα είναι οι εξής :

$$\hat{\alpha} = \frac{\overline{X}}{\hat{\beta}}$$

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i{}^2 - \overline{X}^2}{\overline{X}}$$

Μέθοδος ροπών	
Παράμετρος	**Τιμή**
shape	3.163030e-01
rate	7.731495e-06

Πίνακας 7.18. Εκτίμηση παραμέτρων για την κατανομή Γάμμα με τη μέθοδο ροπών



Διάγραμμα 7.6. Σύγκριση θεωρητικής και εμπειρικής αθροιστικής κατανομής για την κατανομή Γάμμα

Στο παραπάνω διάγραμμα βλέπουμε να υπάρχει μια μικρή απόσταση της θεωρητικής με την εμπειρική αθροιστική κατανομή, άρα η Γάμμα προσαρμόζεται αρκετά καλά στα δεδομένα μας.

Anderson-Darling GoF Test
136.65978

Πίνακας 7.19. Στατιστικός έλεγχος Anderson -Darling για κατανομή Γάμμα

One-sample Kolmogorov-Smirnov test	
p-value	0.48809

Πίνακας 7.20. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Γάμμα

Παρατηρούμε ότι και με τους δύο στατιστικούς ελέγχους τα p-value είναι > 0.05 άρα η κατανομή Γάμμα κάνει καλή προσαρμογή στα δεδομένα μας.

- **Κατανομή Βήτα**

Η κατανομή Βήτα χρησιμοποιείται επίσης ευρέως στη στατιστική για την περιγραφή δεδομένων απωλειών. Η συνάρτησή της όπως αναφέραμε και στο $5^{ο}$ κεφάλαιο της εργασίας είναι η εξής

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Και στην κατανομή Βήτα έγινε εκτίμηση παραμέτρων με τις μεθόδους μέγιστης πιθανοφάνειας και ροπών.

$$\hat{\alpha}=\frac{nm_1+m_2}{n\left(\frac{m_2}{m_1}-m_1-1\right)+m_1}$$

$$\hat{\beta}=\frac{(n-m_1)(n-\frac{m_2}{m_1})}{n\left(\frac{m_2}{m_1}-m_1-1\right)+m_1}$$

Μέθοδος μέγιστης πιθανοφάνειας	
Παράμετρος	**Τιμή**
shape 1	0.2305435
shape 2	3.3080154

Πίνακας 7.21. Εκτίμηση παραμέτρων για την κατανομή Βήτα με τη μέθοδο μέγιστης πιθανοφάνειας

Anderson-Darling GoF Test
Inf

Πίνακας 7.22. Στατιστικός έλεγχος Anderson-Darling για κατανομή Βήτα

One-sample Kolmogorov-Smirnov test	
p-value	0.5383119

Πίνακας 7.23. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Βήτα

Και η Βήτα κατανομή κάνει καλή προσαρμογή στα δεδομένα μας αφού τα p-value > 0.05.

Μέθοδος ροπών	
Παράμετρος	**Τιμή**
shape 1	0.5095889
shape 2	4.0251831

Πίνακας 7.24. Εκτίμηση παραμέτρων για την κατανομή Βήτα με τη μέθοδο ροπών

Anderson-Darling GoF Test
Inf

Πίνακας 7.25. Στατιστικός έλεγχος Anderson-Darling για κατανομή Βήτα

One-sample Kolmogorov-Smirnov test	
p-value	0.3047469

Πίνακας 7.26. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Βήτα

Και με την μέθοδο ροπών η Βήτα κατανομή κάνει καλή προσαρμογή στα δεδομένα μας αφού τα p-value > 0.05.

- **Κατανομή Dagum**

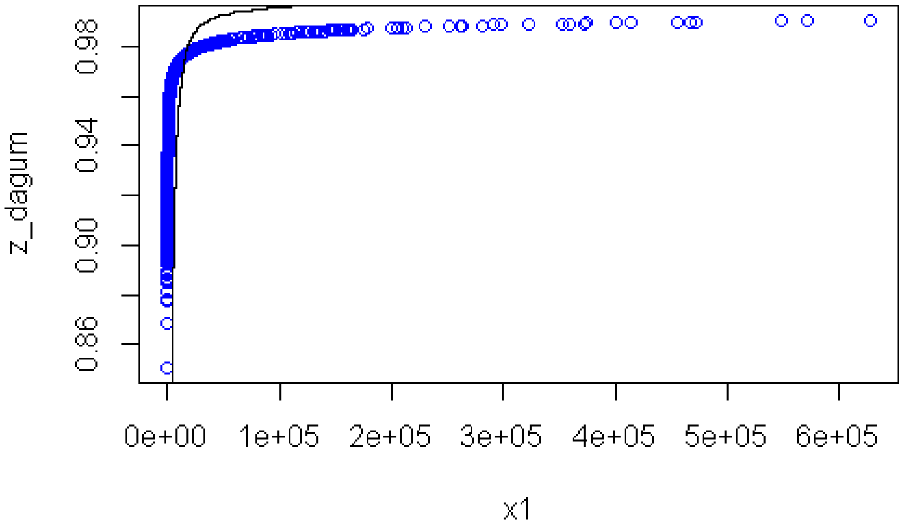
Η τελευταία κατανομή που αναφέρουμε στην παρούσα διπλωματική για την περιγραφή δεδομένων απωλειών είναι η κατανομή Dagum με συνάρτηση κατανομής

$$\mathrm{F(x)} = [1 + (\tfrac{x}{b})^{-\alpha}]^{-p} \quad , \ \mathrm{x} > 0$$

Σ' αυτήν την κατανομή χρησιμοποιήσαμε την μέθοδο ελαχίστων τετραγώνων για την εκτίμηση των παραμέτρων της.

Least Squares Coefficients			
	Estimate	**Std. Error**	**z value Pr(>|z|)**
(Intercept):1	6.78746	0.03953	171.685 < 2e-16
(Intercept):2	0.26295	0.01242	21.179 < 2e-16
(Intercept):3	0.19576	0.03323	5.891 3.84e-09

Πίνακας 7.27. Εκτίμηση παραμέτρων για την κατανομή Dagum με τη μέθοδο ελαχίστων τετραγώνων



Διάγραμμα 7.7. Σύγκριση θεωρητικής και εμπειρικής αθροιστικής κατανομής για την κατανομή Dagum

One-sample Kolmogorov-Smirnov test	
D	1
p-value	< 2.2e-16

Πίνακας 7.28. Στατιστικός έλεγχος Kolmogorov-Smirnov για κατανομή Dagum

Σύμφωνα με τα παραπάνω αποτελέσματα από το τεστ Kolmogorov-Smirnov έχουμε τον παρακάτω συγκεντρωτικό πίνακα.

Kolmogorov-Smirnov (p-value) **(α)**					
	Pareto	**Lognormal**	**Gamma**	**Beta**	**Dagum**
Μέθοδος μέγιστης πιθανοφάνειας	0.003811	0.1796266	0.50203	0.5383119	
Μέθοδος ροπών	0.04676	0.2800283	0.48809	0.3047469	
Μέθοδος ελαχίστων τετραγώνων					< 2.2e-16

Anderson-Darling (p-value) **(β)**					
	Pareto	**Lognormal**	**Gamma**	**Beta**	**Dagum**
Μέθοδος μέγιστης πιθανοφάνειας	9.288e-07	43.0318316	135.69987	inf	
Μέθοδος ροπών	0.01459	49.8765814	136.65978	inf	
Μέθοδος ελαχίστων τετραγώνων					

Πίνακας 7.29. Συγκεντρωτικοί πίνακες **(α)**, **(β)** των p-value των κατανομών

Παρατηρούμε ότι μόνο οι κατανομές Pareto και Dagum δεν κάνουν καλή προσαρμογή στα δεδομένα διότι p-value < **0,05** άρα απορρίπτεται η μηδενική υπόθεση.

Ανακεφαλαιώνοντας λοιπόν, καταλήξαμε στο συμπέρασμα χρησιμοποιώντας το Kolmogorov-Smirnov test, ότι καλύτερη προσαρμογή στα δεδομένα μας κάνουν οι κατανομές **Lognormal**, **Gamma** και **Beta**.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

[1] Σπυρίδων Ι. Χατζησπύρος, Πιθανότητες ΙΙ – Ροπογεννήτριες συναρτήσεις, 1-23.

Ξένη

[2] B.C. Arnold Pareto Distribution (2014), Vol **5**, 5931-5937

[3] Benjamin Sexto M., Humberto Vaquera H., Barry C. Arnold (2013), *Use* of the Dagum Distribution for Modeling Tropospheric Ozone levels, *Journal of Environmental Statistics*, Vol **5**, No **5**, 1-11.

[4] Brenda F. Ginos (2009), *Parameter Estimation for the Lognormal*, Department of Statistics Brigham Young University.

[5] Broderick O. Oluyede, Shujiao Huang and Mavis Pararai (2014), A New Class of Generalized Dagum Distribution with Applications to Income and Lifetime Data, Journal of Statistical and Econometric Methods, Vol.**3**, No.**2**, 125-151.

[6] Broderick O. Oluyede, Sasith Rajasooriya (2013), The Mc-Dagum distribution and its statistical properties with applications, *Asian journal of mathematics and applications*, Vol Article ID ama0085, 1-16.

[7] Camilo Dagum (2006), *Wealth distribution models: analysis and applications*, STATISTICA, Vol. **LXVI,** No. **3**, 235-268.

[8] Christian Kleiber and Samuel Kotz (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

[9] Daiane Aparecida Zuanetti, Carlos A. R. Diniz and Jose Galvao Leite (2006), A Lognormal model for insurance claims data, *REVSTAT – Statistical Journal*, Vol **4**, No **2**, 131-142.

[10] David Bowie, *What information does the distribution of firm sizes in the FT-SE 100 Index provide about long term equity returns?*, Department of Actuarial Mathematics and Statistics Heriot-Watt University, Edinburgh, EH14 4AS Scotland, 179-195.

[11] Duangkamon Chotikapanich (2008), *Modeling Income Distributions and Lorenz Curves* Jacques Silber, Bar Ilan University.

[12] Eckard Limpert, Werner A. Stahel, and Markus ABBT (2001), *Log-normal Distributions across the Sciences Keys and Clues*, BioScience Vol.**51** No.**5**, 341-352.

[13] Edward Furman (2008), *On a multivariate gamma distribution, ScienceDirect*, Statistics and Probability Letters.

[14] Gary Patrik, *Estimating casualty insurance loss amount distributions*, 57-109.

[15] George Vaphiades, *The application of Beta and Gamma distributions to under-reported income values*, «SPOUDAI», Vol. **46**, No **1-2**, University of Piraeus, 31-36.

[16] Héctor R. Gertel, Roberto Giuliodori, Paula F. Auerbach, Alejandro F. Rodríguez (2002), *A parametric estimation of personal income distribution in Argetina using the Dagum model*, 1-31.

[17] J. Martin & C.J. Perez, *Application of a generalized lognormal distribution to engineering data fitting*, Departamento de Matemαticas, Universidad de Extremadura, Cαceres, Spain, 869-874.

[18] Jacub Nedved (2011), *The use of the Lognormal distribution in analyzing incomes*, International Days of Statistics and Economics, Prague, 451-463.

[19] James B. McDonald (May, 1984), *Some generalized functions for the size distribution of income*, Econometrica, Vol. **52**, No. **3**, 647-663.

[20] James B. McDonald, Yexiao J. Xu (1993), *A* generalization of the beta distribution with applications, *Journal of Econometrics* , 133-152.

[21] James William McLeod (1998*), An Investigation of the CDF-Based Method of Moments*, University of Toronto.

[22] John C.B Cooper (2013), *UK Personal Income: an Application of the Pareto Distribution*, Applied Probabiliy Trust.

[23] Joseph C. Watkins (2011), Introduction *to Statistical Methodology, Topic 15: Maximum Likehood Estimation*, 182-197.

[24] Joseph Lee Petersen, *Estimating the Parameters of a Pareto Distribution-Introducing a Quantile Regression Method*.

[25] K. Krishnamoorthy (2006), Handbook *of statistical distributions with applications*, University of Louisiana at Lafayette U.S.A.

[26] Klaus L.P. Vasconcellos and Francisco Cribari-Neto (2005), Improved maximum likelihood estimation in a new class of beta regression models, *Brazilian Journal of Probability and Statistics* ,Vol.**19**, 13–31.

[27] Mercedes Prieto Alaiz and Maria-Pia Victoria-Feser (1996), *Modelling Income Distribution in Spain: A Robust Parametric Approach*, DARP Discussion Paper No.**20**, STICERD, London School of Economics, 1-31.

[28] Mette Rytgaard, *Estimation in the Pareto Distribution*, Nordisk Reinsurance A/S , Copenhagen, Denmark.

[29] Michael Fackler, *Reinventing Pareto: Fits for both small and large losses*, Munich Germany.

[30] Michael B. Gordy (1998), *A generalization of generalized beta distributions*, Board of Governors of the Federal Reserve System.

[31] Morteza Khodabin and Alireza Ahmadabadi (2010), *Some properties of generalized gamma distribution,* Mathematical Sciences, Vol. **4**, No. **1**, 9-28.

[32] OECD (2013), *Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*, 17-24.

[33] Oluwumi Adetan and Thomas J. Afullo, *Comparison of two methods to evaluate the Lognormal raindrop size distribution model in Durban*, School of Electrical, Electronic and Computer Engineering University of KwaZulu-Natal, South Africa.

[34] Orietta Marsili (2005), *Technology and the Size Distribution of Firms: Evidence from Dutch Manufacturing,* Review of Industrial Organization,Vol. **27**,303–328.

[35] Peter Davies B.Bus.Sc. (2002), *Some Simple Applications of the Normal and Log Normal Probability Distributions*, 1-13.

[36] Pietro Muliere and Davide Suverato (2014), *Income and Wealth Distributions in a Population of Heterogeneous Agents*, Munich Discussion Paper No. **2014-21**, Department of Economics University of Munich, 1-14.

[37] Richard E. Quandt (1964*), Old and new methods of estimation and the Pareto distribution*, Princeton University Econometric Research Program.

[38] Ripsy Bandourian (April 2000), *Income Distributions: A Comparison across Countries and Time*, Luxembourg Income Study Working Paper No. **231**, 1-15.

[39] Robert L. Axtell (2001), U.*S. Firm Sizes are Zipf Distributed, Center on Social and Economic Dynamics,* The Brookings Institution.

[40] Sasith Rajasooriya (2013), *Statistical Properties of the Mc-Dagum and Related Distributions*, Georgia Southern University, 1-31.

[41] Wael Abdul Lateef Jasim (2010), Bayes Estimator of the parameter Gamma distribution under Quadratic and LINEX Loss Function, Iraqi Journal of Statistical Science,Vol. **16**, 13-28.

[42] Wei-Hsiung Shen (1998), *Estimation of the parameters of a Lognormal distribution*, TAIWANESE JOURNAL OF MATHEMATICS, Vol. **2**, No. 2, 243-250.

Διαδικτυακοί τόποι

https://www.princeton.edu

http://www.math.uah.edu

http://learnpkpd.com

http://en.wikipedia.org

http://www.itl.nist.gov

http://contentdm.lib.byu.edu

http://www.quora.com

http://www.brighton-webs.co.uk