

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΑΝΑΛΟΓΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗΣ ΚΙΝΔΥΝΟΥ**

**ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ ΓΙΑ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ
ΑΠΑΙΤΗΣΕΩΝ: ΕΦΑΡΜΟΓΗ ΣΤΟΝ ΚΛΑΔΟ
ΑΥΤΟΚΙΝΗΤΟΥ**

Αναστασία Σ. Μαρούλη

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,
Ιούνιος 2014**

Πανεπιστήμιο Πειραιώς

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΑΝΑΛΟΓΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗΣ ΚΙΝΔΥΝΟΥ**

**ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ ΓΙΑ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ
ΑΠΑΙΤΗΣΕΩΝ: ΕΦΑΡΜΟΓΗ ΣΤΟΝ ΚΛΑΔΟ
ΑΥΤΟΚΙΝΗΤΟΥ**

Αναστασία Σ. Μαρούλη

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,
Ιούνιος 2014**

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 Αναλογιστική Επιστήμη και τη Διοικητική Κινδύνου.

Τα μέλη της τριμελούς επιτροπής ήταν:

Δημήτριος Αντζουλιάκος, Αναπληρωτής Καθηγητής, Επιβλέπων

Πολίτης Κωνσταντίνος, Επίκουρος Καθηγητής

Ψαρράκος Γεώργιος, Λέκτορας.

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

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

M.Sc. in Actuarial Science and Risk Management

Stochastic Models for Claim Claims: Application in Automobile Insurance

By

Anastasia S. Marouli

MSc Dissertation

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

Piraeus, June 2014

Πανεπιστήμιο Πειραιώς

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Δημήτριο Αντζουλάκο, για την πολύτιμη βοήθεια που μου παρείχε καθ' όλη την πορεία της διπλωματικής μου εργασίας. Η συμβολή του υπήρξε καταλυτική, ώστε η παρούσα διπλωματική εργασία να πάρει την τελική της μορφή. Τέλος οφείλω ένα μεγάλο ευχαριστώ στην οικογένειά μου και τον Άγγελο, που με στήριξαν όλο αυτό το διάστημα.

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Περίληψη

Με τον όρο στοχαστικό σύστημα εννοούμε κάθε πραγματικό σύστημα, η λειτουργία του οποίου επηρεάζεται σημαντικά από τον παράγοντα τύχη. Με άλλα λόγια, στοχαστικό είναι κάθε μοντέλο, η μελλοντική συμπεριφορά του οποίου δεν μπορεί να προβλεφθεί επακριβώς αλλά μόνο πιθανοθεωρητικά. Τα μοντέλα αυτά απαιτούν προσομοίωση του πραγματικού συστήματος, όπου οι σημαντικές σχέσεις μεταξύ των στοιχείων του έχουν αντικατασταθεί από αντίστοιχες μαθηματικές, ενώ οι μη σημαντικές έχουν αγνοηθεί. Οι στοχαστικές διαδικασίες είναι επαρκή μαθηματικά μοντέλα για την περιγραφή και τη μελέτη στοχαστικών συστημάτων.

Στην παρούσα διπλωματική γίνεται μελέτη συγκεκριμένων μοντέλων ως προς την προσαρμογή τους σε δεδομένα χαρτοφυλακίων του ασφαλιστικού κλάδου αυτοκινήτου. Στην αναλογιστική επιστήμη έχουν προταθεί διάφορες κατανομές που μπορούν να χρησιμοποιηθούν, αντί της κατανομής Poisson, για την περιγραφή του αριθμού των απαιτήσεων σε ένα ασφαλιστικό χαρτοφυλάκιο. Σε αυτή την οικογένεια των κατανομών ανήκει η Αρνητική Διωνυμική, η Poisson-Inverse Gaussian, η Strict Arcsine, η αρνητική διωνυμική – Pareto κατανομή, και άλλες. Στόχος της διπλωματικής είναι η κατάγραφή και η μελέτη των χαρακτηριστικών των παραπάνω κατανομών και η αξιολόγηση της προσαρμογής τους σε δεδομένα από τον κλάδο αυτοκινήτου.

Πανεπιστήμιο Πειραιώς

Abstract

By stochastic system we mean any real system, the function of which is greatly influenced by luck. In other words, each model is stochastic, the future behavior of which cannot be predicted precisely but only probabilistic. These models are a simulation of the real system, where significant relationships between elements are replaced by corresponding mathematical, while non- significant are ignored. Stochastic processes are adequate mathematical models for the description and study of stochastic systems

This thesis is a study of specific models with regard to their adaptation to data portfolios automobile insurance industry. In actuarial science, many different distributions have been suggested for the description of the number of claims of an insurance portfolio that can be applied instead of the Poisson distribution. In this family of distributions belong the Negative-Binomial distribution, the Poisson-Inverse Gaussian distribution, the strict-arc sine distribution, the Negative Binomial- Pareto distribution, the Poisson-Log Normal distribution and many others. The aim of this dissertation is to write down and study the basic characteristics of the aforementioned distributions and the evaluation of their fitness upon data from the automobile insurance industry.

Πανεπιστήμιο Πειραιώς

ΠΕΡΙΕΧΟΜΕΝΑ

1	Εισαγωγή.....	1
2.1	Το μοντέλο της κατανομής Poisson για τον αριθμό των απαιτήσεων	5
2.1.1	Μοντέλο Poisson - Ομοιογενές Χαρτοφυλάκιο	8
2.1.2	Το Μοντέλο Poisson για Ετερογενές Χαρτοφυλάκιο	10
2.2	Ετερογένεια και Μεμειγμένα Μοντέλα	11
2.3	Αρνητική Διωνυμική Κατανομή	16
2.4	Η κατανομή Poisson - Inverse Gaussian	18
2.5	Η κατανομή Negative Binomial-Pareto.....	20
2.6	Η κατανομή Strict Arcsine	22
2.7	Η κατανομή Poisson Lindley - Beta Prime	24
2.8	Η κατανομή Beta - Negative Binomial	26
2.9	Η κατανομή Generalized Poisson.....	30
3	Εφαρμογή για δεδομένα αριθμού απαιτήσεων στον κλάδο αυτοκινήτου	33
3.1	Εισαγωγή	33
3.2	Εκτιμητές μεγίστης πιθανοφάνειας.....	33
3.3	Μέθοδος Ροπών	35
3.4	Έλεγχος καλής προσαρμογής χ^2	35
3.4.1	Η κατανομή χ^2	37
3.4.2	Η μέθοδος.....	38
3.5	Εφαρμογή των δεδομένων του Lemaire σε διάφορα μοντέλα κατανομών.....	40
3.5.1	Προσαρμογή της κατανομής Poisson	40
3.5.2	Προσαρμογή της Αρνητικής Διωνυμικής κατανομής.....	44
3.5.3	Προσαρμογή της κατανομής Negative Binomial Pareto.....	48
3.5.4	Προσαρμογή της Poisson-Inverse Gaussian κατανομής.....	51
3.5.5	Εφαρμογή της Poisson-Inverse Gaussian κατανομής στα δεδομένα των Gossiaux & Lemaire (1981)	58
3.5.6	Προσαρμογή της Strict Arcsine Κατανομής.....	65
3.5.7	Προσαρμογή της Beta Negative Binomial κατανομής.....	68
4	Συμπεράσματα.....	71
	Παράρτημα.....	73
	Βιβλιογραφία.....	81

Πανεπιστήμιο Πειραιώς

1 ΕΙΣΑΓΩΓΗ

Η αναλογιστική επιστήμη ασχολείται με αβέβαια γεγονότα, στα οποία με σαφήνεια οι έννοιες των πιθανοτήτων και της στατιστικής παρέχουν ένα απαραίτητο εργαλείο για τη μέτρηση και τη διαχείριση των κινδύνων στον τομέα της ασφάλισης και της οικονομίας. Μια σημαντική πτυχή του επαγγέλματος της ασφάλισης είναι ο καθορισμός της τιμής, που συνήθως ονομάζεται ασφάλιστρο, η οποία πρέπει να πληρωθεί ως αντάλλαγμα για τη μεταβίβαση των κινδύνων. Είναι καθήκον του αναλογιστή να αξιολογήσει μια δίκαιη τιμή, δεδομένης της φύσης του κινδύνου. Η βιβλιογραφική έρευνα της αναλογιστικής επιστήμης καλύπτει ένα ευρύ φάσμα θεμάτων, μεταξύ των οποίων είναι η ταξινόμηση των κινδύνων σε κλάσεις (risk classification) και η εμπειρική τιμολόγηση (experience rating) στην ασφάλιση αστικής ευθύνης έναντι τρίτων στον κλάδο των αυτοκινήτων, όπου αποτελούν τις κινητήριες δυνάμεις της έρευνας που παρουσιάζεται σε αυτή τη διπλωματική. Τα ανωτέρω συνιστούν μια περιοχή της εφαρμοσμένης στατιστικής η οποία έχει δανειστεί εργαλεία από διαφορετικές εργαλειοθήκες της θεωρητικής στατιστικής, κυρίως εμπειρικές μεθόδους. Ωστόσο, η πολυπλοκότητα της τυπικής εφαρμογής, συμπεριλαμβανομένης της μη παρατηρήσιμης ετερογένειας κινδύνου (risk heterogeneity) σε ένα χαρτοφυλάκιο, της μη ισορροπημένης σχεδίασης και των μη παραμετρικών κατανομών, ενέπνευσε μια ανεξάρτητη θεμελιώδη έρευνα γνωστή ως «θεωρία αξιοπιστίας» (credibility theory), η οποία θεωρείται ο ακρογωνιαίος λίθος των σύγχρονων ασφαλιστικών μαθηματικών.

Επομένως ένα κρίσιμο ζήτημα στην ανάλυση δεδομένων είναι το πώς θα επιλέξει κανείς την κατάλληλη κατανομή ώστε να γίνει η καλύτερη δυνατή περιγραφή των εκάστοτε αποτελεσμάτων ενός πειράματος. Προς επίρρωση αυτού του σκοπού, γίνεται η μελέτη των χαρακτηριστικών κάποιων κατανομών, σύμφωνα με υποθέσεις που αφορούν τη φύση των πειραμάτων μας, ώστε να καταλήξουμε σε εκείνες που μοντελοποιούν καλύτερα τα δεδομένα μας. Η διαφορά μεταξύ αυτών των μοντέλων αναλύεται μέσω της μέσης τιμής και τη διακύμανσης του ετήσιου αριθμού καθώς και του κόστους των απαιτήσεων των ασφαλισμένων, οι οποίοι ανήκουν σε διαφορετικές κλάσεις κινδύνου. Ωστόσο, κατά την ταξινόμηση των κινδύνων σε κλάσεις πολλοί σημαντικοί παράγοντες δεν μπορούν να ληφθούν υπόψη εκ των προτέρων. Συνεπώς, παρά το a priori σύστημα, οι κλάσεις τιμολόγησης, δεν θα είναι απολύτως

ομοιογενείς και μπορεί να δημιουργηθεί μια δομή τιμολόγησης που είναι άδικη για τους ασφαλισμένους. Προκειμένου να μειωθεί το χάσμα μεταξύ του ατομικού ασφαλιστρού και του υποβόσκοντος κινδύνου και να αυξηθούν τα κίνητρα για την οδική ασφάλεια, το ατομικό ιστορικό ζημιών πρέπει να ληφθεί υπόψη στο πλαίσιο ενός a posteriori μοντέλου.

Πιο συγκεκριμένα, θα ασχοληθούμε με διάφορα χαρακτηριστικά που παρουσιάζουν αριθμητικά δεδομένα για τον αριθμό των απαιτήσεων όπως *zero inflation*, η *υπερδιασπορά* αλλά και η *ετερογένεια*. Αντίστοιχες στοχαστικές ιδιότητες θα πρέπει να αντιπροσωπεύουν αυτές τις πηγές μεταβλητότητας κατά τη μοντελοποίηση των αριθμητικών δεδομένων. Η κατανομή Poisson διαδραματίζει ένα σημαντικό ρόλο στη διαδικασία μοντελοποίησης παρόλο που ο δείκτης διασποράς της (index of dispersion) είναι ίσος με 1. Ο δείκτης αυτός χρησιμοποιείται σαν ένδειξη του κατά πόσο μπορεί να διαφέρει η διακύμανση από τη μέση τιμή, κάτι που στην Poisson δεν συμβαίνει. Ωστόσο παραμένει πολύτιμη κατανομή στον ασφαλιστικό κλάδο και στην αναλογιστική επιστήμη αφού αποτελεί τη βάση για κατανομές και στοχαστικές διαδικασίες που αντιπροσωπεύουν με ρεαλιστικότερο τρόπο τα δεδομένα μας. Η υπερδιασπορά καταγράφηκε ως χαρακτηριστικό για πρώτη φορά από τους *Greenwood & Yule* (1920), οι οποίοι πρότειναν ένα μοντέλο στο οποίο ο μέσος της Poisson ακολουθεί τη Γάμμα κατανομή. Στην πραγματικότητα αυτό οδήγησε στην Αρνητική Διωνυμική κατανομή (Negative Binomial distribution).

Επιπλέον μεμειγμένα μοντέλα Poisson, με επιπρόσθετες ελεύθερες παραμέτρους, εφαρμόστηκαν από τους *Karlis & Xekalaki* (2005). Η Αρνητική Διωνυμική κατανομή κατανομή έγινε ιδιαίτερα δημοφιλής αφού είναι πιο ευέλικτη κατανομή συγκριτικά με την Poisson κατανομή ιδιαίτερα στις περιπτώσεις που είναι υπό αμφισβήτηση οι αυστηροί περιορισμοί για την Poisson. Στη συνέχεια, ένα εναλλακτικό μοντέλο με πολλές κοινές μαθηματικές και στατιστικές ιδιότητες με αυτές της Αρνητικής Διωνυμικής κατανομής, είναι αυτό της Poisson Inverse Gaussian (P-IG) κατανομής, που παρουσιάστηκε από τον *Willmot* (1987). Εφαρμόζεται εξίσου καλά για τη μοντελοποίηση της συχνότητας των απαιτήσεων, εξαιτίας του εύκολου χειρισμού της αλλά και του γεγονότος ότι μπορεί ακόμη να χρησιμοποιηθεί και για τη μοντελοποίηση των ίδιων φυσικών φαινομένων όπως συμβαίνει και με την Αρνητική Διωνυμική κατανομή. Κάτι ανάλογο επεδίωξαν αργότερα και οι *Meng, Wang & Whitmore* (1999) με την κατανομή Negative Binomial-Pareto. Μία κατανομή στην

οποία η Αρνητική Διωνυμική κατανομή αποτυπώνει την υπερδιασπορά του χαρτοφυλακίου, και η Pareto προσομοιώνει την κατανομή του μέσου αριθμού των απαιτήσεων των συμμετεχόντων στο χαρτοφυλάκιο. Μια εξίσου σημαντική κατανομή είναι, η Strict Arcsine, που εισήχθηκε για πρώτη φορά από τους *Letac & Mora* (1990), ενώ οι *Kokonendji* και *Khoudar* δημοσίευσαν σχετικό άρθρο το 2004. Το 2012 μια νέα διακριτή κατανομή, εξαρτώμενη από δύο παραμέτρους εισάγεται από τους *E. Gómez-Déniz, A. Hernández Bastida, & M.P. Fernández Sánchez* (2012) ως αποτέλεσμα μείξης της κατανομής Lindley και της Beta Prime κατανομής. Ολοκληρώνοντας τη μελέτη κατανομών που βρίσκουν εφαρμογή στον ασφαλιστικό κλάδο, κρίνεται αναγκαία η αναφορά και στο μοντέλο της Beta Negative Binomial κατανομής.

Στο Κεφάλαιο 3 γίνεται εφαρμογή χ^2 τεστ καλής προσαρμογής, για τα περισσότερα από τα μοντέλα κατανομών που έχουμε αναφέρει παραπάνω, σε διάφορες ομάδες δεδομένων από τον κλάδο αυτοκινήτου που έχουμε συναντήσει σε ξένη κατά κύριο λόγο βιβλιογραφία. Αρχικά, γίνεται αναφορά στις μεθόδους ανεύρεσης εκτιμητών (ροπών, μεγίστης πιθανοφάνειας) αλλά και στην μεθοδολογία του ελέγχου καλής προσαρμογής. Κατόπιν παρουσιάζονται με τη βοήθεια του «στατιστικού πακέτου» R τα χ^2 τεστ για τις κατανομές Poisson, Negative Binomial, και Poisson Inverse Gaussian αναφορικά με τα δεδομένα ενός χαρτοφυλακίου ασφάλισης αυτοκινήτου που είχε συλλέξει ο *Lemaire* (1976) από μια Βελγική εταιρία. Κατά τον έλεγχο αυτό αναδείχθηκε ως επαρκέστερο μοντέλο αυτό της Poisson Inverse Gaussian κατανομής. Οι *Meng, Wei & Whitmore* (1999) θεωρούσαν ιδανική περίπτωση να έχουν πρόσβαση σε αριθμητικά δεδομένα που αφορούν τον κάθε ασφαλισμένο ξεχωριστά και μάλιστα σε πολλαπλές χρονικές περιόδους, προκειμένου να εξακριβώσουν την εγκυρότητα της Negative Binomial Pareto κατανομής στο ασφαλιστικό χαρτοφυλάκιο. Ωστόσο επειδή δεν υπήρχε η δυνατότητα αυτή, αρκέστηκαν και αυτοί στα δεδομένα του *Lemaire* (1976). Οι *Gossiaux & Lemaire* (1981) ανέλυσαν έξι ομάδες δεδομένων υπολογίζοντας τον αριθμό των απαιτήσεων που εγείρουν οι ασφαλισμένοι του κλάδου αυτοκινήτου ανά ασφαλιστικό συμβόλαιο στη Γερμανία το 1960. Τα δεδομένα αυτά επανεξετάζονται από τον *Willmot* (1987), δείχνοντας ότι η Αρνητική Διωνυμική εφαρμόζεται επαρκώς ικανοποιητικά σε μερικές περιπτώσεις, Παρόλα αυτά, η εφαρμογή του μοντέλου Poisson-Inverse Gaussian αποδεικνύεται ανώτερη και στις έξι ομάδες δεδομένων,

χρησιμοποιώντας είτε το στατιστικό έλεγχο καλής προσαρμογής του Pearson είτε τη μέθοδο της μέγιστης πιθανοφάνειας.

Στη συνέχεια πραγματοποιείται έλεγχος προσαρμογής με δεδομένα από ένα δείγμα που προέρχεται από τη Δημοκρατία της Κεντρικής Αφρικής κατά την περίοδο του 1984. Τα δεδομένα αυτά παρουσιάζουν υπερδιασπορά, και για το λόγο αυτό θα πρέπει να γίνει σε αυτά εφαρμογή μίας κατανομής που έχει το χαρακτηριστικό της υπερδιασποράς. Έτσι γίνεται συγκριτικός πίνακας για τις κατανομές Poisson Inverse Gaussian, Negative Binomial, Poisson, Strict Arcsine και Γενικευμένης Poisson κατανομής. Συμπερασματικά λοιπόν προκύπτει ότι η Strict Arcsine είναι μία κατανομή η οποία είναι αρκούτως ανταγωνιστική με την Αρνητική Διωνυμική και την Poisson-Inverse Gaussian.

Πανεπιστήμιο Πειραιώς

2 Στοχαστικά μοντέλα για τον αριθμό των απαιτήσεων

2.1 ΤΟ ΜΟΝΤΕΛΟ ΤΗΣ ΚΑΤΑΝΟΜΗΣ POISSON ΓΙΑ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ ΑΠΑΙΤΗΣΕΩΝ

Στην ενότητα αυτή, θα αναφερθούμε στην πιο δημοφιλή και ευρέως μελετημένη στοχαστική διαδικασία για τον αριθμό των απαιτήσεων, τη στοχαστική διαδικασία Poisson. Τι είναι όμως μια στοχαστική διαδικασία Poisson; Η διαδικασία Poisson καλύπτει ένα ευρύ φάσμα επιθυμητών θεωρητικών ιδιοτήτων. Ήδη από το 1903 ο Filip Lundberg στη διατριβή του είχε αναφερθεί σε αυτή, ως ένα στοχαστικό μοντέλο περιγραφής του αριθμού των απαιτήσεων. Λίγο αργότερα το 1930, ο Harald Cramér, διάσημος Σουηδός στατιστικολόγος, ανέπτυξε εκτενώς μια γενικευμένη θεωρία κινδύνου κάνοντας χρήση του συνόλου των απαιτήσεων S με αριθμό αφίξεων T_i που προέρχονται από μία διαδικασία Poisson. Τόσο λοιπόν για ιστορικούς λόγους όσο και λόγω των μαθηματικών της ιδιοτήτων η διαδικασία Poisson παίζει καθοριστικό ρόλο στα Ασφαλιστικά Μαθηματικά.

Η κατανομή Poisson

Η κατανομή Poisson ανακαλύφθηκε από τον *Siméon-Denis Poisson* (1781-1840) και δημοσιεύθηκε το 1833 στο έργο του με τίτλο “*Recherches sur la Probabilité des Jugements en Matières Civiles*” (*Έρευνα για την πιθανότητα των κρίσεων γύρω από εγκληματικά και αστικά θέματα*). Ουσιαστικά μια τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson απαριθμεί τον αριθμό των γεγονότων που συμβαίνουν σε ένα χρονικό διάστημα ή σε μια συγκεκριμένη περιοχή. Για παράδειγμα, ο αριθμός των απαιτήσεων που εγείρει ένας ασφαλισμένος οδηγός στην εταιρεία όπου είναι ασφαλισμένος μια συγκεκριμένη χρονική περίοδο. Ένα βασικό χαρακτηριστικό της κατανομής Poisson είναι η *ισοδιασπορά* (*equidispersion*), δηλαδή η διακύμανση και η μέση τιμή της ταυτίζονται.

Αναλυτικότερα, αν η τυχαία μεταβλητή X έχει συνάρτηση μάζας πιθανότητας που δίνεται από τον τύπο

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots$$

θα λέμε ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Poisson με παράμετρο $\lambda > 0$ (συμβ. $X \sim \text{Poisson}(\lambda)$).

Ακολουθεί η εύρεση ενός πολύ σημαντικού χαρακτηριστικού της κατανομής Poisson όπως είναι οι ροπές της, από τις οποίες μπορούμε να εξάγουμε συμπεράσματα τόσο για τη μέση της τιμή όσο και για τη διακύμανση της.

Ροπές της κατανομής Poisson

Αν $X \sim \text{Poisson}(\lambda)$, τότε η μέση τιμή που αποτελεί και την ροπή πρώτης τάξεως της κατανομής, δίνεται από την παρακάτω σχέση:

$$E(X) = \sum_{k=1}^{+\infty} k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k+1}}{k!} = \lambda.$$

Επιπλέον, για τη ροπή δεύτερης τάξης της κατανομής έχουμε ότι

$$E(X^2) = \sum_{k=1}^{+\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} (k+1) \frac{\lambda^{k+1}}{k!} = \lambda + \lambda^2.$$

Ως συνέπεια των δύο προηγούμενων αποτελεσμάτων είναι εύκολο πλέον να υπολογισθεί η διακύμανση της κατανομής με τον παρακάτω γνωστό τρόπο

$$V(X) = E(X^2) - [E(X)]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Εδώ παρατηρείται ξεκάθαρα πλέον το φαινόμενο της ισοδιασποράς (equidispersion) όπως ορίστηκε παραπάνω. Ο βαθμός ασυμμετρίας της κατανομής μιας τυχαίας μεταβλητής X , αποτιμάται με τη λοξότητα της (skewness) που συμβολίζεται με $\gamma[X]$. Η λοξότητα είναι η τρίτη κεντρική ροπή της X , κανονικοποιημένη από τη διασπορά της υψωμένη στη δύναμη $3/2$ προκειμένου να προκύψει αδιάστατο μέγεθος. Για την ακρίβεια η λοξότητα της X δίνεται από τον τύπο

$$\gamma[X] = \frac{E[(X - E[X])^3]}{(V[X])^{\frac{3}{2}}}.$$

Για οποιαδήποτε τυχαία μεταβλητή X με συμμετρική κατανομή, θα έχουμε ότι η λοξότητα της $\gamma[X]$ ότι θα είναι ίση με το 0. Οι κατανομές με θετική ασυμμετρία τείνουν να συγκεντρώνουν την περισσότερη μάζα πιθανότητάς τους σε ένα μικρό εύρος μικρών τιμών, ενώ η υπόλοιπη μάζα εκτείνεται σε ένα μεγάλο εύρος υψηλότερων τιμών.

Για την τυχαία μεταβλητή $X \sim \text{Poisson}(\lambda)$ έχουμε ότι $\gamma[X] = \frac{1}{\lambda}$. Ουσιαστικά, για μικρές τιμές του λ η κατανομή παρουσιάζει μεγάλη λοξότητα (ασυμμετρία) ενώ για μεγαλύτερες τιμές του λ η κατανομή είναι σχεδόν συμμετρική (περίπου για $\lambda > 15$).

Η ροπογεννήτρια συνάρτηση της κατανομής Poisson δίνεται από τη σχέση

$$M(t) = e^{\lambda(e^t - 1)}.$$

Πριν δώσουμε τον ορισμό μιας (στοχαστικής) διαδικασίας Poisson σημειώνουμε ότι για κάθε πραγματική συνάρτηση f με πεδίο ορισμού το σύνολο $[0, +\infty)$ θα γράφουμε

$$f(s, t) = f(t) - f(s), \quad 0 \leq s < t < \infty.$$

Τώρα είμαστε έτοιμοι να δώσουμε τον ορισμό μίας στοχαστικής διαδικασίας Poisson.

Ορισμός 1.1: Μία στοχαστική διαδικασία $(N(t))_{t \geq 0}$ θα καλείται Poisson αν ισχύουν οι παρακάτω συνθήκες:

1. Η $N(t)$ παίρνει μη αρνητικές ακέραιες τιμές $0, 1, 2, 3, \dots$, $N(0) = 0$ και για $s < t$ ισχύει ότι $N(s) \leq N(t)$.
2. Η $N(t)$ έχει στάσιμες προσαυξήσεις: Για κάθε ακέραιο k , $\Delta > 0$ και $s \leq t$ ισχύει ότι

$$P[N(t + \Delta) - N(t) = k] = P[N(s + \Delta) - N(s) = k]$$

3. Η $N(t)$ έχει ανεξάρτητες προσαυξήσεις: Για κάθε ακέραιο $k > 0$ και $0 \leq t_0 < t_1 < \dots < t_k$ οι τυχαίες μεταβλητές $N(t_j) - N(t_{j-1})$ για $j = 1, 2, \dots, k$ είναι ανεξάρτητες.

4. Για όλα τα $t \geq 0$ και $h > 0$, η τυχαία μεταβλητή $N(t, t+h] = N(t+h) - N(t)$ είναι μια τυχαία μεταβλητή Poisson με μέση τιμή $\mu(t, t+h)$ όπου $\mu(t) = \int_0^t \lambda(z) dz$ και $\lambda(z)$ η συνάρτηση έντασης της διαδικασίας.

Αν συνάρτηση έντασης $\lambda(z)$ της διαδικασίας είναι σταθερός αριθμός τότε η διαδικασία Poisson καλείται ομογενής.

2.1.1 ΜΟΝΤΕΛΟ POISSON - ΟΜΟΙΟΓΕΝΕΣ ΧΑΡΤΟΦΥΛΑΚΙΟ

Η κατανομή Poisson διαδραματίζει εξέχοντα ρόλο στη μοντελοποίηση αριθμητικών δεδομένων, κυρίως εξαιτίας της περιγραφικής της επάρκειας όταν υπάρχει ομοιογένεια στον υπό μελέτη πληθυσμό, δηλαδή όταν οι ασφαλισμένοι υπόκεινται όλοι στον ίδιο κίνδυνο. Σύμφωνα λοιπόν με τον *Lemaire* (1985), συμβολίζουμε με $N(t, t+\Delta t)$ τον αριθμό των απαιτήσεων στο χρονικό διάστημα $(t, t+\Delta t)$ και διατυπώνουμε τις επόμενες τρεις ιδιότητες:

1. $P [N (t, t + \Delta t) = 1] = \lambda \Delta t + o (\Delta t)$, όπου $\lambda > 0$
2. $P [N (t, t + \Delta t) > 1] = o (\Delta t)$
3. Αν τ και τ' είναι δύο ξεχωριστές χρονικές στιγμές, τότε

$$P [N(\tau) = k \cap N(\tau') = k'] = P [N(\tau) = k] \cdot P [N(\tau') = k']$$

(μία συνάρτηση $f(x)$ είναι $o(h)$ αν $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$).

Η πρώτη ιδιότητα δηλώνει ότι η πιθανότητα ενός ατυχήματος σε ένα μικρό χρονικό διάστημα $(t, t+\Delta t)$ είναι, αγνοώντας όρους μεγαλύτερης τάξης, ανάλογη με τη διάρκεια του χρονικού διαστήματος. Η δεύτερη ορίζει πως η πιθανότητα να συμβούν περισσότερα από ένα ατυχήματα σε αυτό το διάστημα είναι αμελητέα. Τέλος η τρίτη και τελευταία υπόθεση απαιτεί ο αριθμός των ατυχημάτων που αναφέρονται σε δύο διαφορετικά χρονικά διαστήματα να είναι ανεξάρτητες τυχαίες μεταβλητές. Εύκολα διαπιστώνεται πως η κατανομή Poisson ικανοποιεί και τις τρεις παραπάνω ιδιότητες. Αντιστρόφως μπορούμε να αποδείξουμε πως η κατανομή που ικανοποιεί και τις τρεις παραπάνω υποθέσεις είναι η Poisson. Πράγματι, χρησιμοποιώντας το συμβολισμό $p_k(t) = P[N(0, t) = k]$, θα έχουμε

$$\begin{aligned}
p_k(t + \Delta t) &= p_k(t)P[N(t, t + \Delta t) = 0] + p_{k-1}(t)P[N(t + \Delta t) = 1] \\
&+ \sum_{i=2}^k p_{k-i}(t)P[N(t, t + \Delta t) = i] \\
&= p_k(t)[1 - \lambda\Delta t + o(\Delta t)] + p_{k-1}(t)[\lambda\Delta t + o(\Delta t)] \\
&+ \sum_{i=2}^k p_{k-i}(t)o(\Delta t) = p_k(t)(1 - \lambda\Delta t) + p_{k-1}(t)\lambda\Delta t + o(\Delta t)
\end{aligned}$$

για $k = 0, 1, \dots$ (συνθήκη: $p_{-1}(t) = 0$).

Συνεπώς

$$\frac{p_k(t + \Delta t) - p_k(t)}{\Delta t} = -\lambda p_k(t) + \lambda p_{k-1}(t) + \frac{o(\Delta t)}{\Delta t}$$

Παίρνοντας το όριο για $\Delta t \rightarrow 0$ προκύπτει ότι

$$p'_k(t) = -\lambda p_k(t) + \lambda p_{k-1}(t), \quad k = 1, 2, \dots$$

και

$$p'_0(t) = -\lambda p_0(t), \quad k = 0.$$

Έτσι λύνοντας τις πιο πάνω εξισώσεις με βάση τις αρχικές υποθέσεις $p_0(0) = 1$ και $p_k(0) = 0$, για $k > 0$, έχουμε

$$p_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

ενώ για μια περίοδο μήκους 1 παίρνουμε

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Ανακεφαλαιώνοντας, βλέπουμε πως η Poisson είναι η μόνη κατανομή που επαληθεύει και τις τρεις ιδιότητες. Συνεπώς αν δεχτούμε πως και το πρότυπο των ατυχημάτων που προκαλούν οι οδηγοί τις ικανοποιεί, δεν μπορούμε παρά να

υιοθετήσουμε την Poisson ως την κατανομή που μοντελοποιεί των αριθμό των απαιτήσεων των ασφαλισμένων.

2.1.2 ΤΟ ΜΟΝΤΕΛΟ POISSON ΓΙΑ ΕΤΕΡΟΓΕΝΕΣ ΧΑΡΤΟΦΥΛΑΚΙΟ

Η κατανομή Poisson είναι άρρηκτα συνδεδεμένη με τον ασφαλιστικό κλάδο για την περιγραφή των απαιτήσεων σε ένα χαρτοφυλάκιο. Όμως η προσέγγιση της μοντελοποίησης μέσω της Poisson δυστυχώς δεν είναι ρεαλιστική για πολλά ασφαλιστικά δεδομένα. Η εμπειρία έχει αποδείξει, ότι η κατανομή του αριθμού των απαιτήσεων σε επαναλαμβανόμενες παρατηρήσεις για έναν ασφαλισμένο, τείνουν να έχουν μεγαλύτερη μεταβλητότητα από όσο μπορεί να καταγραφεί μέσω του μοντέλου Poisson. Συγκεκριμένα το μοντέλο της απλής κατανομής Poisson, αποτυγχάνει να περιγράψει ικανοποιητικά δεδομένα που η δειγματική τους διασπορά ξεπερνά το δειγματικό τους μέσο. Έτσι κατανομές που εμφανίζουν υπερδιασπορά (overdispersion) είναι πιθανές εναλλακτικές της απλής κατανομής Poisson.

Σε περιπτώσεις ανομοιογενούς πληθυσμού, δηλαδή που οι ασφαλισμένοι δεν υπόκεινται όλοι στον ίδιο κίνδυνο, είναι ευρέως διαδεδομένη η χρήση μεμειγμένων κατανομών Poisson ως υποκατάστατα της απλής κατανομής Poisson. Ιδιαίτερο ενδιαφέρον παρουσιάζουν πληθυσμοί που αποτελούνται από πεπερασμένους ομοιογενείς υποπληθυσμούς στους οποίους η κατανομή πιθανότητας των αποζημιώσεων μπορεί να θεωρηθεί ως κατανομή Poisson με διαφορετικά λ . Στην πραγματικότητα, το πρόβλημα της ετερογένειας προκύπτει από διαφορές στην οδική συμπεριφορά των ατόμων, οι οποίες δεν έχουν γίνει αντιληπτές από τον αναλογιστή. Ιδιαίτερα στον κλάδο της ασφάλισης αυτοκινήτου, είναι ήδη γνωστή από παλαιότερες μελέτες η τάση διαφοροποίησης μεταξύ ασφαλισμένων ως προς τον αριθμό των απαιτήσεων. Η τάση αυτή έχει αποδοθεί στην ποικιλία χαρακτηριστικών που δεν έχουν προσμετρηθεί στη διαμόρφωση του ασφαλιστρού, όπως διαφορετική προσωπικότητα ασφαλισμένων, κατάσταση υγείας, αλλά και φυσικούς, περιβαλλοντικούς παράγοντες κινδύνου.

Εκτός από τις γνωστές συνέπειες της απρατήρητης ετερογένειας στην αριθμητική ανάλυση των δεδομένων του χαρτοφυλακίου, όπως η υπερδιασπορά, υπάρχουν και αυτές που αφορούν στην κατασκευή του μεμειγμένου μοντέλου πιθανότητας που ακολουθεί. Τόσο το φαινόμενο των «πολλών μηδενικών» (excesses of zeros) όσο και

των βαριών ουρών είναι ενδεικτικά παραδείγματα ετερογένειας που δεν έχει ληφθεί υπόψη. Είναι συνήθης πρακτική να επιτρέπεται για μη παρατηρηθείσα ετερογένεια η χρήση τυχαίας μεταβλητής για την περιγραφή του μέσου μιας κατανομής Poisson, οδηγώντας σε ένα μοντέλο μεμειγμένης Poisson. Σε τέτοιες περιπτώσεις η αναμενόμενη ετήσια συχνότητα των απαιτήσεων καθίσταται τυχαία.

2.2 ΕΤΕΡΟΓΕΝΕΙΑ ΚΑΙ ΜΕΜΕΙΓΜΕΝΑ ΜΟΝΤΕΛΑ

Τα μεμειγμένα μοντέλα είναι διακριτοί ή συνεχείς σταθμισμένοι συνδυασμοί κατανομών που στοχεύουν στην αναπαράσταση ετερογενών πληθυσμών αποτελούμενων από δύο ή περισσότερους διαφορετικούς υποπληθυσμούς. Τέτοια μοντέλα ουσιαστικά χρησιμοποιούνται όταν ένας ετερογενής πληθυσμός δειγματικών μονάδων απαρτίζεται από αρκετούς υποπληθυσμούς που στον καθένα από αυτούς εφαρμόζεται ένα σχετικά απλούστερο μοντέλο. Η πηγή της ετερογένειας μπορεί να είναι το φύλο, η ηλικία, η γεωγραφική περιοχή κ.λπ.

- **Διακριτές Μείξεις**

Προκειμένου να προβούμε στη μαθηματική περιγραφή ενός μεμειγμένου μοντέλου, υποθέτουμε ότι η κατανομή του N (αριθμός απαιτήσεων) μπορεί να περιγραφεί από μία συνάρτηση μάζας πιθανότητας της ακόλουθης μορφής:

$$\Pr[N = k] = p(k|\boldsymbol{\psi}) = q_1 p_1(k|\xi_1) + q_2 p_2(k|\xi_2) + \dots + q_v p_v(k|\xi_v) \quad (1)$$

όπου $\boldsymbol{\psi} = (\mathbf{q}^T, \boldsymbol{\xi}^T)^T$, $\mathbf{q}^T = (q_1, \dots, q_v)$, $\boldsymbol{\xi}^T = (\xi_1, \dots, \xi_v)$.

Το μοντέλο αυτό αναφέρεται συχνά ως διακριτό ή πεπερασμένο μεμειγμένο μοντέλο. Εδώ το ξ_j είναι μια παράμετρος που χαρακτηρίζει τη συνάρτηση μάζας πιθανότητας $p(\cdot | \xi_j)$ και τα q_j είναι τα βάρη της μείξης. Χαρακτηριστικό παράδειγμα πεπερασμένης μείξης είναι οι διογκωμένες στο 0 κατανομές (*zero-inflated distributions*). Έχει παρατηρηθεί εμπειρικά ότι οι κατανομές που χαρακτηρίζονται ως κατανομές μέτρησης, συχνά υπερέχουν της κατανομής Poisson όσον αφορά το πλήθος των μηδενικών. Κατά συνέπεια, προκειμένου να προσαρμοστεί αυτό το χαρακτηριστικό, ο συνδυασμός της αρχικής κατανομής $\{p_k, k = 0, 1, \dots\}$ (είτε είναι η

κατανομή Poisson αυτή είτε όχι) και της εκφυλισμένης κατανομής με όλη την πιθανότητα συγκεντρωμένη στην αρχή (0), δίνει μια πεπερασμένη μείζη όπου

$$P[N = 0] = \omega + (1 - \omega)p_0$$

$$P[N = k] = (1 - \omega)p_k, \quad k = 1, 2, \dots$$

Η μείζη αυτού του είδους συνήθως αναφέρεται ως διογκωμένη στο 0 (*zero-inflated*) ή τροποποιημένη στο 0 (*zero-modified*) κατανομή, ή ως κατανομή με πρόσθετα μηδενικά (*distribution with added zeros*).

Το διάνυσμα των βαρών της μείζης $\mathbf{q}^T = (q_1, \dots, q_n)$ μπορεί να θεωρηθεί ως διακριτή συνάρτηση πιθανότητας που δηλώνει την αναλογία των n διαφορετικών υποπληθυσμών στο συνολικό πληθυσμό. Από την άλλη το διάνυσμα των παραμέτρων $\boldsymbol{\xi}^T = (\xi_1, \dots, \xi_n)$, σε ένα μοντέλο μείζης κατανομών Poisson, περιγράφει το μέσο ετήσιο αριθμό των απαιτήσεων στον εκάστοτε υποπληθυσμό. Έτσι στη γενική περίπτωση η σχέση (1) δηλώνει ότι υπάρχουν n κατηγορίες ασφαλισμένων, με ετήσια αναμενόμενη συχνότητα απαιτήσεων $\xi_1, \xi_2, \dots, \xi_n$ καθώς και ότι το ποσοστό των διαφόρων κατηγοριών στο χαρτοφυλάκιο είναι q_1, q_2, \dots, q_n . Ο αναλογιστής λαμβάνοντας υπόψη έναν ασφαλισμένο δεν γνωρίζει σε ποια κατηγορία ανήκει, αλλά ότι η πιθανότητα να προέρχεται από την j κατηγορία (υποπληθυσμό) είναι q_j . Οπότε, η συνάρτηση μάζας πιθανότητας (στο εξής σ.μ.π.) του αριθμού των απαιτήσεων για τον ασφαλισμένο οδηγό είναι ένας σταθμισμένος μέσος των σ.μ.π. των k κατηγοριών.

- **Συνεχείς Μείζεις**

Η αύξηση του αριθμού n των κατηγοριών της σχέσης (1), συχνά οδηγεί σε δραματική αύξηση του αριθμού των παραμέτρων q_j και ξ_j . Επομένως για μεγάλο n είναι προτιμότερο να καταφύγουμε σε συνεχείς μείζεις όπου το άθροισμα της σχέσης (1) αντικαθίσταται από ένα ολοκλήρωμα. Ειδικότερα επιτρέπουμε στο ξ να είναι συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας (στο εξής σ.π.π.) $g(\cdot)$ και έτσι το διακριτό μοντέλο μείζης αντικαθίσταται από το μοντέλο

$$p(k) = \int p(k|\xi)g(\xi)d\xi \quad (2)$$

που συνήθως αναφέρεται ως κατανομή μείξης. Όταν η $g(\cdot)$ εισέρχεται στο μοντέλο χωρίς παραμετρικές υποθέσεις ομιλούμε για ένα ημιπαραμετρικό μοντέλο μείξης. Συχνά όμως στην αναλογιστική επιστήμη η $g(\cdot)$ την παίρνουμε από κάποια συγκεκριμένη παραμετρική οικογένεια κατανομών και επομένως έχουμε ένα (πλήρες) παραμετρικό μοντέλο.

- **Μεμειγμένο Μοντέλο Poisson για τον Αριθμό των Απαιτήσεων**

Η κατανομή Poisson συνήθως δεν περιγράφει με ακρίβεια τις παρατηρήσεις ενός χαρτοφυλακίου ασφαλισμένων εξαιτίας της ετερογένειάς του. Χαρακτηριστικό παράδειγμα αυτής είναι οι διαφορετικές οδηγικές ικανότητες του καθενός. Ως φυσική απόρροια αυτού, η μέση τιμή λ της Poisson πολλαπλασιάζεται με μια θετική τυχαία επίδραση Θ . Προφανώς το Θ το επιλέγουμε έτσι ώστε $E[\Theta] = 1$ αφού θέλουμε να προσεγγίσουμε, σε μέσους όρους, τη συχνότητα του χαρτοφυλακίου. Δεσμεύοντας λοιπόν ως προς Θ έχουμε

$$\Pr[N = k | \Theta = \theta] = p(k | \lambda\theta) = e^{-\lambda\theta} \frac{(\lambda\theta)^k}{k!}, \quad k = 0, 1, \dots \quad (3)$$

όπου $p(\cdot | \lambda\theta)$ είναι η σ.μ.π. της Poisson με μέσο $\lambda\theta$. Η εξήγηση που δίνουμε σε αυτό το μοντέλο είναι ότι δεν έχουν όλοι οι οδηγοί του χαρτοφυλακίου την ίδια μέση συχνότητα εμφάνισης λ . Κάποιοι έχουν μεγαλύτερη συχνότητα ($\lambda\theta$ με $\theta \geq 1$) ενώ άλλοι μικρότερη ($\lambda\theta$ με $\theta \leq 1$). Με αυτό τον τρόπο ο ετήσιος αριθμός ατυχημάτων που προκαλούνται από έναν τυχαία επιλεγμένο ασφαλισμένο του χαρτοφυλακίου, κατανέμεται σύμφωνα με τους κανόνες που διέπουν μια μεμειγμένη Poisson. Σε αυτή την περίπτωση η πιθανότητα ένας τυχαία επιλεγμένος ασφαλισμένος να αναφέρει k απαιτήσεις στην εταιρεία, προσδιορίζεται μέσω της σχέσης (3). Στη γενική περίπτωση η Θ είναι συνεχής τυχαία μεταβλητή οπότε η σ.μ.π. που σχετίζεται με τα μεμειγμένα μοντέλα Poisson δίνεται από τη σχέση

$$\Pr[N = k] = E[p(k | \lambda\Theta)] = \int_0^\infty e^{-\lambda\theta} \frac{(\lambda\theta)^k}{k!} dF_\Theta(\theta) \quad (4)$$

όπου F_Θ δηλώνει τη συνάρτηση κατανομής (στο εξής σ.κ.) της Θ και ικανοποιεί τη σχέση $F_\Theta(0) = 0$. Αξίζει να αναφερθεί ότι το μεμειγμένο μοντέλο Poisson της

σχέσης (4) προσδιορίζει πως η μέση συχνότητα απαιτήσεων ενός ασφαλισμένου δεν μεταβάλλεται με το χρόνο απλά επιτρέπει σε κάποιους να έχουν υψηλότερη από άλλους. Στο εξής θα λέμε ότι η N ακολουθεί μεμειγμένη κατανομή Poisson με παράμετρο λ και επίπεδο κινδύνου (risk level) Θ (συμβ. $N \sim MPoi(\lambda, \Theta)$) όταν η σ.μ.π. δίνεται όπως στη σχέση (4).

- **Ιδιότητες των Μεμειγμένων Κατανομών Poisson**

Αν $N \sim MPoi(\lambda, \Theta)$ έχουμε ότι

$$E[N] = E[E(N|\Theta)] = E[\lambda\Theta] = \lambda E[\Theta]$$

και

$$E[N^2] = \int_0^{+\infty} (\lambda\theta + \lambda^2\theta^2) dF_{\Theta}(\theta) = \lambda E[\Theta] + \lambda^2 E[\Theta^2].$$

Έτσι η διασπορά της θα είναι ίση με

$$V[N] = \lambda E[\Theta] + \lambda^2 E[\Theta^2] - \lambda^2 (E[\Theta])^2 = \lambda + \lambda^2 V[\Theta]. \quad (5)$$

Παρατηρούμε εύκολα ότι η διασπορά του N ξεπερνά το μέσο του αφού

$$V[N] = \lambda + \lambda^2 V[\Theta] \geq \lambda = E[N]. \quad (6)$$

Για αυτό το λόγο στην περίπτωση που το Θ δεν εκφυλίζεται στη μονάδα έχουμε το φαινόμενο της υπερδιασποράς στις μεμειγμένες Poisson.

Η ασυμμετρία (λοξότητα) της κατανομής προκύπτει από τη σχέση

$$\gamma[N] = \frac{1}{(V[N])^{\frac{3}{2}}} \left(3V[N] - 2E[N] + \frac{\gamma[\Theta]}{\sqrt{V[\Theta]}} \frac{(V[N] - E[N])^2}{E[N]} \right). \quad (7)$$

Χρησιμοποιώντας την ανισότητα *Jensen*, δηλαδή ότι $E[\varphi(X)] \geq \varphi(E[X])$ για κάθε τυχαία μεταβλητή X και κάθε κυρτή συνάρτηση φ , παίρνουμε για την κατανομή της τυχαίας μεταβλητής N ($N \sim MPoi(\lambda, \Theta)$), ότι

$$\Pr[N = 0] = \int_0^{\infty} e^{-\lambda\theta} dF_{\Theta}(\theta) \geq e^{-\int_0^{\infty} \lambda\theta dF_{\Theta}(\theta)} = e^{-\lambda}.$$

Έτσι προκύπτει ότι οι μεμειγμένες κατανομές Poisson έχουν υπερβάλλοντα μηδενικά (excess of zeros) συγκριτικά με τις κατανομές Poisson με τον ίδιο μέσο. Αυτό έρχεται σε πλήρη αντιστοιχία με εμπειρικές μελέτες, στις οποίες οι αναλογιστές παρατηρούν περισσότερους ασφαλισμένους με μηδενικές απαιτήσεις από ότι δικαιολογείται από την κατανομή Poisson. Επίσης, αν $N \sim MPoi(\lambda, \theta)$ με $E[N] = \lambda$, τότε υπάρχουν δύο ακέραιοι $0 \leq k_0 < k_1$ τέτοιοι ώστε

$$Pr[N = k] \geq e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots, k_0,$$

$$Pr[N = k] \leq e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = k_0 + 1, \dots, k_1,$$

$$Pr[N = k] \geq e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq k_1 + 1.$$

Τα παραπάνω απέδειξε ο Shaked (1980). Ένα δεύτερο συμπέρασμα που εξάγεται από τα παραπάνω είναι ότι οι κατανομές που είναι μείξεις Poisson έχουν πιο παχιές δεξιές ουρές από την κατανομή Poisson με τον ίδιο μέσο.

- **Πιθανογεννήτρια μεμειγμένης κατανομής Poisson**

Η πιθανογεννήτρια συνάρτηση της μεμειγμένης κατανομής Poisson είναι στενά συνδεδεμένη με τη ροπογεννήτρια συνάρτηση. Η ροπογεννήτρια συνάρτηση είναι ένα σπουδαίο εργαλείο για τους στατιστικούς και τους αναλογιστές. Στο σημείο αυτό θα γίνει μια υπενθύμιση του τύπου της ροπογεννήτριας συνάρτησης μιας μη αρνητικής τυχαίας μεταβλητής X , συμβολίζεται με $M_X(t) = E[\exp(tX)]$, $t \geq 0$. Ακριβώς όπως η πιθανογεννήτρια συνάρτηση αποτελεί εργαλείο για την ανάλυση του αθροίσματος ανεξάρτητων διακριτών τυχαίων μεταβλητών, η ροπογεννήτρια συνάρτηση είναι ένα ισχυρό μέσο μελέτης αθροισμάτων συνεχών τυχαίων μεταβλητών. Στην περίπτωση της κατανομής $N \sim MPoi(\lambda, \theta)$ έχουμε ότι η πιθανογεννήτρια συνάρτηση της N , $\varphi_N(z)$, ικανοποιεί τη σχέση

$$\varphi_N(z) = \int_0^{+\infty} e^{-\lambda\theta} \sum_{k=0}^{+\infty} \frac{(z\lambda\theta)^k}{k!} dF_\theta(\theta) = \int_0^{+\infty} e^{\lambda\theta(z-1)} dF_\theta(\theta) = M_\theta(\lambda(z-1))$$

ή ισοδύναμα

$$M_{\theta}(t) = \varphi_N \left(1 + \frac{t}{\lambda}\right).$$

2.3 ΑΡΝΗΤΙΚΗ ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

Η Αρνητική Διωνυμική κατανομή (στο εξής N.B.), χρησιμοποιείται κατά κόρον αντί της Poisson στην περιγραφή δεδομένων. Υπάρχουν αρκετά μοντέλα που οδηγούν στη N.B. κατανομή. Ένα κλασσικό παράδειγμα προκύπτει από τη “ροπή πρόκλησης ατυχήματος” (accident proneness) η οποία αναπτύχθηκε μετά τους *Greenwood & Yule* (1920). Η θεωρία αυτή στην αρχική εκδοχή της υποστηρίζει πως ο αριθμός ατυχημάτων που θα αντιμετωπίσει ένας ασφαλισμένος ακολουθεί κατανομή Poisson, με μέσο που διαφέρει μεταξύ των ατόμων του υπό μελέτη πληθυσμού. Ο *Greenwood* και ο *Yule* εισήγαγαν τη N.B. ως μια μεμειγμένη κατανομή Poisson που παρουσιάζει υπερδιασπορά, οπότε μας βοηθά στη μοντελοποίηση στοιχείων που η διασπορά τους ξεπερνά εμφανώς το μέσο τους. Πρότειναν ένα μοντέλο όπου ο μέσος της Poisson ακολουθεί Γάμμα κατανομή. Στην πραγματικότητα αυτό οδήγησε στη γέννηση της N.B. κατανομής.

Μια τυχαία μεταβλητή Θ θα λέμε ότι ακολουθεί τη διπαραμετρική κατανομή Γάμμα (συμβ. $\Theta \sim \text{Gam}(a, \beta)$) αν έχει σ.π.π. που δίνεται από τον τύπο

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

Στην περίπτωση όπου $a = 1$, η Γάμμα κατανομή ταυτίζεται με την Εκθετική κατανομή με παράμετρο β ($\Theta \sim \text{Exp}(\beta)$) με σ.π.π.

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

Για τη συνάρτηση Γάμμα έχουμε ότι

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

$$\Gamma(a + 1) = a\Gamma(a).$$

Επίσης αν $a \in \mathbb{N}$, $\Gamma(a + 1) = a!$.

Έτσι δουλεύοντας με την τυχαία μεταβλητή $\theta \sim \text{Gam}(a, \tau)$, έχουμε ότι

$$f_{\theta}(\theta) = \frac{1}{\Gamma(\alpha)} \tau^{\alpha} \theta^{\alpha-1} e^{-\tau\theta}, \quad \alpha, \tau > 0$$

$$E[\theta] = \frac{\alpha}{\tau}, V[\theta] = \frac{\alpha}{\tau^2}, \gamma[\theta] = \frac{2}{\sqrt{\alpha}} \text{ και } M_{\theta}(t) = \left(\frac{\tau}{\tau-t}\right)^{\alpha}, \quad 0 \leq t \leq \tau.$$

Επίσης, από τη σχέση (2), προκύπτει με ολοκλήρωση η κατανομή p_k , $k = \{0, 1, 2, \dots\}$ για τον αριθμό των απαιτήσεων του χαρτοφυλακίου

$$\begin{aligned} p_k &= \int_0^{+\infty} p_k(\theta) f(\theta) d\theta = \int_0^{+\infty} \frac{\theta^k e^{-\theta}}{k!} \frac{\tau^{\alpha} e^{-\tau\theta} \theta^{\alpha-1}}{\Gamma(\alpha)} d\theta \\ &= \frac{\tau^{\alpha}}{k! \Gamma(\alpha) (1+\tau)^{k+\alpha}} \int_0^{+\infty} e^{-\theta(1+\tau)} [\theta(1+\tau)]^{k+\alpha-1} d[\theta(1+\tau)] \\ &= \frac{\Gamma(k+\alpha)}{\Gamma(k+1)\Gamma(\alpha)} \frac{\tau^{\alpha}}{(1+\tau)^{k+\alpha}} = \binom{k+\alpha-1}{k} \left(\frac{\tau}{1+\tau}\right)^{\alpha} \left(\frac{1}{1+\tau}\right)^k \\ &= \binom{k+\alpha-1}{k} p^{\alpha} q^k \end{aligned}$$

όπου θέσαμε

$$p = \frac{\tau}{1+\tau}, \quad q = 1-p = \frac{1}{1+\tau}, \quad \binom{k+\alpha-1}{k} = \frac{\Gamma(\alpha+k)}{\Gamma(k+1)\Gamma(\alpha)}.$$

Με αυτόν τον τρόπο οδηγούμαστε σε μια N.B. κατανομή με μέση τιμή $m = \frac{\alpha}{\tau}$ και διασπορά $\sigma^2 = \frac{\alpha}{\tau} \left(1 + \frac{1}{\tau}\right)$, (διασπορά υπερβαίνει το μέσο, ιδιότητα που παρουσιάζουν όλες οι συνεχείς μείξεις της Poisson. Επίσης ο συντελεστής λοξότητάς της δίνεται από τον τύπο

$$\frac{2 - \frac{\tau}{1+\tau}}{\sqrt{\frac{\alpha}{1+\tau}}}$$

και η ροπογεννήτριά της από τη σχέση

$$\left(\frac{\tau}{1+\tau-e^t}\right)^a, \quad t < \ln(1+\tau).$$

Κάνοντας χρήση του παρακάτω αναδρομικού τύπου μπορούν εύκολα να υπολογιστούν οι πιθανότητες της N.B. κατανομής:

$$p_{k+1} = \frac{k+a}{(k+1)(1+\tau)} p_k, \quad n = 1, 2, 3, \dots$$

με

$$p_0 = \left(\frac{\tau}{1+\tau}\right)^a.$$

2.4 Η ΚΑΤΑΝΟΜΗ POISSON - INVERSE GAUSSIAN

Ένα εναλλακτικό μοντέλο μεικτής Poisson κατανομής που ονομάζεται Poisson Inverse Gaussian κατανομή (στο εξής P-IG) εισήγαγε για πρώτη φορά ο *Holla* (1966), ενώ ο *Willmot* (1987) παρουσίασε πιθανές εφαρμογές του μοντέλου ως εναλλακτικό αυτού της N.B. κατανομής. Οι δύο κατανομές έχουν παρόμοιες μαθηματικές και στατιστικές ιδιότητες, ωστόσο η P-IG κατανομή είναι ιδανική για τη μοντελοποίηση δεδομένων που εμφανίζουν δεξιά ασυμμετρία.

Θα λέμε ότι μια τυχαία μεταβλητή X ακολουθεί την Inverse Gaussian κατανομή (συμβ. IG) με παραμέτρους $\mu > 0$ και $\beta > 0$ και θα συμβολίζουμε με $X \sim IGau(\mu, \beta)$, αν η σ.π.π. της είναι η παρακάτω

$$f(x) = \frac{\mu}{\sqrt{2\pi\beta x^3}} \exp\left(-\frac{1}{2\beta x}(x-\mu)^2\right), \quad x > 0.$$

Αν $X \sim IGau(\mu, \beta)$ τότε $E[X] = \mu$ και $V(X) = \mu\beta$. Για τη ροπογεννήτριά της έχουμε

$$\begin{aligned} M_X[t] &= \int_0^{+\infty} \frac{\mu}{\sqrt{2\pi\beta x^3}} \exp\left(-\frac{1}{2\beta x}(x-\mu)^2 + tx\right) dx \\ &= \exp\left(\frac{\mu}{\beta}\right) \int_0^{+\infty} \frac{\mu}{\sqrt{2\pi\beta x^3}} \exp\left(-\frac{1}{2\beta x}(x^2(1-2\beta t) + \mu^2)\right) dx. \end{aligned}$$

Θέτοντας $\xi = x\sqrt{1-2\beta t}$ παίρνουμε

$$\begin{aligned} M_X[t] &= \exp\left(\frac{\mu}{\beta}\right) \int_0^{+\infty} \frac{\mu}{\sqrt{2\pi\frac{\beta}{\sqrt{1-2\beta t}}}\xi^3} \exp\left(-\frac{1}{2\frac{\beta}{\sqrt{1-2\beta t}}}\xi^2 + \mu^2\right) d\xi \\ &= \exp\left(\frac{\mu}{\beta}(1-\sqrt{1-2\beta t})\right). \end{aligned}$$

Ολοκληρώνοντας τις σχέσεις (3)-(4) με $\theta \sim IGau(1, \tau)$, δηλαδή

$$f_\theta(\theta) = \frac{1}{\sqrt{2\pi\tau\theta^3}} \exp\left(-\frac{1}{2\tau\theta}(\theta-1)^2\right), \theta > 0$$

Προκύπτει ότι η σ.μ.π. της Poisson Inverse Gaussian κατανομής δίνεται από τον τύπο

$$\Pr[N = k] = \int_0^{+\infty} \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \frac{1}{\sqrt{2\pi\tau\theta^3}} \exp\left(-\frac{1}{2\tau\theta}(\theta-1)^2\right) d\theta.$$

- **Ροπές και πιθανότητες**

Για τη μέση τιμή και τη διακύμανση της N έχουμε

$$E[N] = \lambda, \quad V[N] = \lambda(1 + \lambda\tau).$$

Επίσης μπορεί εύκολα να αποδειχθεί ότι $\frac{\gamma[\theta]}{\sqrt{V[\theta]}} = 3$ από τη σχέση (7) για την P-IG κατανομή. Για αυτό το λόγο η λοξότητα της P-IG είναι μεγαλύτερη από αυτή της N.B. κατανομής με την ίδια διασπορά και μέση τιμή.

Η πιθανότητα στο 0 είναι

$$\Pr[N = 0] = \exp\left(\frac{1}{\tau}(1 - \sqrt{1 + 2\tau\lambda})\right)$$

ενώ γενικότερα ισχύει ο παρακάτω αναδρομικός τύπος υπολογισμού των πιθανοτήτων

$$\Pr[N = n] = \frac{2\lambda\tau}{1 + 2\lambda\tau} \left(1 - \frac{3}{2n}\right) \Pr[N = n - 1]$$

$$+ \frac{\lambda^2}{(1 + 2\lambda\tau)n(n-1)} \Pr[N = n-2], \quad n = 2,3,4, \dots$$

2.5 Η ΚΑΤΑΝΟΜΗ NEGATIVE BINOMIAL-PARETO

Όπως έχει διατυπωθεί και παραπάνω, οι ασφαλιστικές απαιτήσεις στον κλάδο αυτοκινήτου χαρακτηρίζονται από υπερδιασπορά συγκριτικά με το μοντέλο της Poisson. Επιπλέον, οι τάσεις των απαιτήσεων διαφέρουν μεταξύ των ασφαλισμένων ενός χαρτοφυλακίου. Έτσι οι Meng, Wei & Whitmore (1999) λαμβάνοντας υπόψη αυτά χαρακτηριστικά παρουσίασαν το μοντέλο της Negative Binomial-Pareto (στο εξής N.B.P.) κατανομής. Πιο συγκεκριμένα θεώρησαν το μοντέλο της Αρνητικής Διωνυμικής κατανομής για να αποτυπώσουν την υπερδιασπορά του χαρτοφυλακίου, και αυτό της Pareto ως κατάλληλο να περιγράψει την κατανομή του μέσου αριθμού των απαιτήσεων των συμμετεχόντων στο χαρτοφυλάκιο.

Έστω K ο αριθμός των απαιτήσεων ενός ασφαλισμένου σε μια συγκεκριμένη χρονική περίοδο που ακολουθεί τη $N.B.$ κατανομή με συνάρτηση μάζας πιθανότητας

$$p(k|\mu) = \frac{\Gamma(r+k)}{\Gamma(r)k!} \frac{r^r \mu^k}{(r+\mu)^{r+k}}, \quad k = 0,1, \dots \quad (8)$$

όπου μ ο μέσος αριθμός απαιτήσεων σε μια περίοδο, $E(K|\mu) = \mu$. Η παράμετρος $r > 0$ είναι μια άγνωστη σταθερά τέτοια που παίρνει την ίδια τιμή για όλους τους ασφαλισμένους. Η διασπορά της K δίνεται από τον τύπο:

$$\sigma^2(K|\mu) = \mu \left(1 + \frac{\mu}{r}\right) \quad (9)$$

Η ποσότητα μ/r της σχέσης (9), δηλώνει το επιπλέον ποσό της υπερδιασποράς της $N.B.$ κατανομής σε σχέση με την Poisson και όπως είναι προφανές είναι ανάλογη του μέσου αριθμού μ των απαιτήσεων και αντιστρόφως ανάλογη του r . Όταν $r \rightarrow \infty$ η σ.μ.π. $p(k|\mu)$ της σχέσης (8) προσεγγίζει την Poisson με μέσο μ .

Η συνάρτηση πυκνότητας πιθανότητας του μέσου μ μεταξύ των ασφαλισμένων του χαρτοφυλακίου όπως προαναφέρθηκε ακολουθεί Pareto κατανομή, δηλαδή

$$f(\mu) = \frac{\Gamma(s\zeta + sr + 1)}{\Gamma(s\zeta)\Gamma(sr + 1)} \frac{r^{sr+1} \mu^{s\zeta-1}}{(r+\mu)^{s\zeta+sr+1}} \quad (10)$$

και θα συμβολίζεται ως $Pareto(\zeta, r, s)$. Η παράμετρος $\zeta > 0$ είναι ο μέσος της κατανομής, δηλαδή $E(\mu) = \zeta$, το r είναι ίδιο με αυτό της (8) και το $s > 0$ είναι η παράμετρος που αποτιμά την ομοιογένεια της τάσης για απαιτήσεις μεταξύ των ασφαλισμένων (μεγαλύτερο s δηλώνει μεγαλύτερη ομοιομορφία).

Επιλέγουμε την Pareto κατανομή για αρκετούς λόγους. Είναι συζυγής με την N.B. κατανομή, παρουσιάζει λοξότητα προς τα δεξιά και είναι μονοκόρυφη γεγονός που την καθιστά ικανή να περιγράψει τη μεταβλητότητα της μέσης απαίτησης των μεμονωμένων ασφαλισμένων. Επιπλέον έχει αξιοσημείωτη μαθηματική ευελιξία στην προσαρμογή της σε δεδομένα. Αυτή η τρι-παραμετρική μορφή της κατανομής Pareto καλείται και γενικευμένη κατανομή Pareto. Οι *Klugman et al.* (1998) έδωσαν ιδιότητες της κατανομής αυτής.

Η περιθώρια συνάρτηση πιθανότητας της τ.μ. K (αριθμού των απαιτήσεων σε μια περίοδο για τυχαίο ασφαλισμένο του χαρτοφυλακίου), προκύπτει από τις σχέσεις (8) και (10) με ολοκλήρωση ως προς το μ , δηλαδή

$$p(k) = \int_0^{+\infty} p(k|\mu)f(\mu)d\mu$$

οπότε

$$p(k) = \frac{\Gamma(s\zeta + sr + 1)\Gamma(r + sr + 1)\Gamma(s\zeta + k)\Gamma(r + k)}{\Gamma(s\zeta)\Gamma(sr + 1)\Gamma(r)\Gamma(r + sr + s\zeta + k + 1)k!}, k = 0,1,2, \dots$$

Στο εξής θα αναφερόμαστε σε αυτή την περιθώρια κατανομή ως την κατανομή Negative Binomial-Pareto, και θα τη συμβολίζουμε $NBP(\zeta, r, s)$. Για τον αναδρομικό υπολογισμό πιθανοτήτων της συγκεκριμένης κατανομής ισχύουν οι εξής ιδιότητες

$$p(0) = \frac{\Gamma(s\zeta + sr + 1)\Gamma(sr + r + 1)}{\Gamma(sr + 1)\Gamma(s\zeta + sr + r + 1)}$$

και

$$\frac{p(k+1)}{p(k)} = \frac{(r+k)(s\zeta+k)}{(k+1)(s\zeta+sr+r+k+1)}, k = 0,1, \dots$$

Εν συνεχεία το μοντέλο αυτό θα εξεταστεί στο Κεφάλαιο 3 ως προς την προσαρμογή του σε περιπτώσεις όπου μελετώνται οι αξιώσεις που εγείρουν οι ασφαλισμένοι. Πιο συγκεκριμένα υποθέτουμε πως k_1, \dots, k_t είναι το πλήθος των απαιτήσεων σε t χρονικές περιόδους ενός ασφαλισμένου, και ότι κάθε ένας από τους παραπάνω αριθμούς απαιτήσεων ακολουθεί μια δική του αρνητική διωνυμική κατανομή (βλ. σχέση 8) ανεξάρτητη από τις άλλες. Η συνάρτηση πυκνότητας πιθανότητας $f(\mu)$ της σχέσης (10) λειτουργεί ως εκ των προτέρων κατανομή του μέσου αριθμού των απαιτήσεων του ασφαλισμένου. Κάνοντας χρήση του θεωρήματος Bayes προκειμένου να οδηγηθούμε στην εκ των υστέρων κατανομή του μέσου αριθμού απαιτήσεων για κάθε ασφαλισμένο οδηγούμαστε στην παρακάτω σχέση

$$f(\mu|k_1, \dots, k_t) \propto f(\mu) \prod_{i=1}^t p(k_i|\mu).$$

Θέτοντας $c = \sum_{i=1}^t k_i$, η εκ των υστέρων κατανομή $f(\mu|k_1, \dots, k_t)$ είναι μια Pareto κατανομή αλλά με παραμέτρους τώρα πια τις ζ_t , r , s_t , όπου,

$$\zeta_t = \frac{s\zeta + c}{s + t}$$

και

$$s_t = s + t.$$

Παρατηρούμε ότι η παράμετρος r συνεχίζει να παραμένει σταθερή. Το αποτέλεσμα της σχέσης $s_t = s + t$ μας δείχνει ότι η παράμετρος s_t παίζει το ρόλο του μετρητή του χρόνου για την εκ των προτέρων συνάρτηση πιθανότητας $f(\mu)$. Η παράμετρος ζ_t παίζει το ρόλο της τωρινής εκτίμησης του ρυθμού εμφάνισης απαιτήσεων του ασφαλισμένου, με το ζ να είναι ο ρυθμός εμφάνισης των απαιτήσεων της εκ των προτέρων συνάρτησης πυκνότητας $f(\mu)$.

2.6 Η ΚΑΤΑΝΟΜΗ STRICT ARCSINE

Η συγκεκριμένη κατανομή εισήχθηκε για πρώτη φορά από τους *Letac & Mora* (1990), ενώ οι *Kokonendji & Khoudar* δημοσίευσαν σχετικό άρθρο το 2004. Έχει διαπιστωθεί ότι η Strict Arcsine κατανομή (στο εξής SA) είναι μια κατανομή η οποία

παρουσιάζει υπερδιασπορά, ασυμμετρία προς τα δεξιά, και είναι λεπτόκυρτη. Συγκρίνοντας τη με την N.B. και την P-IG κατανομή, αξίζει να σημειωθεί ότι μπορεί να χρησιμοποιηθεί ως εναλλακτική τους. Η σ.μ.π. της δίνεται από τον τύπο :

$$P(x; p, a) = \frac{A(x; a)}{x!} p^x \exp\{-a \arcsin(p)\}, \quad x = 0, 1, 2, \dots, \quad (11)$$

όπου $0 < a$, $0 < p < 1$ και η ποσότητα $A(x; a)$ ορίζεται ως

$$A(x; a) = \begin{cases} \prod_{k=0}^{z-1} (a^2 + 4k^2), & \text{αν } x = 2z \text{ και } A(0; a) = 1, \\ a \prod_{k=0}^{z-1} (a^2 + (2k + 1)^2), & \text{αν } x = 2z + 1 \text{ και } A(1; a) = a. \end{cases}$$

Η διακριτή αυτή συνάρτηση πιθανότητας, που εξαρτάται από δύο παραμέτρους το p και το a . Θα συμβολίζουμε την τυχαία μεταβλητή X με σ.μ.π. την (11) ως $SA(p, a)$.

Η πιθανογεννήτρια συνάρτηση της κατανομής $SA(p, a)$ είναι η ακόλουθη

$$G_X(u) = \exp\{a[\arcsin(up) - \arcsin(p)]\}, \quad |up| < 1$$

από όπου μπορεί να προκύψει ότι η ροπογεννήτρια συνάρτηση μέσω της σχέσης $M_X(s) = G_X(e^s)$. Όπως απέδειξαν οι *Kokonendji & Khoudar* (2004), έχουμε ότι

$$E(X) = ap(1 - p^2)^{-\frac{1}{2}},$$

$$Var(X) = ap(1 - p^2)^{-\frac{3}{2}},$$

$$E(X - E(X))^3 = ap(1 - p^2)^{-\frac{5}{2}}(1 + 2p^2),$$

$$E(X - E(X))^4 - 3Var^2(X) = ap(1 - p^2)^{-\frac{7}{2}}(1 + 10p^2 + 4p^4).$$

Ο δείκτης διασποράς (Index of Dispersion) είναι

$$ID = \frac{Var(X)}{E(X)} = (1 - p^2)^{-1} > 1,$$

ο συντελεστής λοξότητας είναι

$$\gamma[X] = (ap)^{-\frac{1}{2}}(1-p^2)^{-\frac{1}{4}}(1+2p^2) > 0$$

και ο συντελεστής κύρτωσης είναι

$$\begin{aligned}\gamma_4 &= 3 + \frac{[E(X - E(X))^4 - 3Var^2(X)]}{Var^2(X)} \\ &= 3 + (ap)^{-1}(1-p^2)^{-\frac{1}{2}}(1+10p^2+4p^4) > 3\end{aligned}$$

Επιπλέον, για την SA κατανομή έχουμε ότι πως για $0 < a$ και $0 < p < 1$ είναι μονοκόρυφη ή δικόρυφη στο x και στο $x+1$.

2.7 Η ΚΑΤΑΝΟΜΗ POISSON LINDLEY - BETA PRIME

Μια νέα διακριτή κατανομή, εξαρτώμενη από δύο παραμέτρους, εισήχθη από τους *Gómez et al.* (2012) ως αποτέλεσμα μείξης της κατανομής Lindley και της Beta Prime κατανομής. Μπορεί να θεωρηθεί εναλλακτικό μοντέλο της αρνητικής διωνυμικής κατανομής, της Strict Arcsine, της Negative Binomial Pareto και της Poisson Inverse Gaussian. Ξεκινώντας με τη διακριτή κατανομή Poisson Lindley όπως παρουσιάστηκε από τον *Sankaran* (1970) και μελετήθηκε πρόσφατα από τους *Ghitany & Al-Mutairi* (2009), έχουμε ότι η συνάρτηση μάζας πιθανότητάς της δίνεται από τον τύπο :

$$\Pr(X = x) = \frac{\theta^2(\theta + x + 2)}{(1 + \theta)^{x+3}}$$

με $\theta > 0$ και $x = 0, 1, \dots$. Από εδώ και στο εξής θα συμβολίζεται ως P.L.

Η συγκεκριμένη κατανομή προέκυψε ως μείξη της Poisson με παράμετρο λ , $\lambda > 0$, όπου το λ ακολουθεί τη συνεχή Lindley κατανομή με συνάρτηση πυκνότητας πιθανότητας

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

και θα τη συμβολίζουμε ως L.

Η κατανομή Beta Prime είναι επίσης γνωστή ως ανεστραμμένη Βήτα ή κατανομή Βήτα δευτέρου είδους και έχει σ. π. π.

$$f(\theta; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \cdot \frac{\theta^{\alpha-1}}{(1+\theta)^{\alpha+\beta}} \quad (12)$$

με $\alpha > 0, \beta > 0, \theta > 0$, όπου $B(x,y)$ είναι η γνωστή συνάρτηση Βήτα και η συντομογραφία που θα χρησιμοποιήσουμε στη συνέχεια για αυτή την κατανομή είναι B.P.

Έτσι μια τυχαία μεταβλητή X θα λέμε ότι ακολουθεί την Poisson Lindley-Beta Prime ή $PLBP(\alpha, \beta)$ κατανομή εφόσον ισχύουν τα εξής:

$$X | \theta \sim PL(\theta)$$

$$\theta \sim BP(\alpha, \beta)$$

με $\alpha > 0, \beta > 0$. Αν λοιπόν $X \sim PLBP(\alpha, \beta)$ θα έχει τις παρακάτω ιδιότητες:

(α) Η συνάρτηση μάζας πιθανότητας δίνεται από τη σχέση

$$\begin{aligned} \Pr(X = x) &= \frac{a(1+a)\Gamma(\alpha+\beta)\Gamma(\beta+x)}{\Gamma(\beta)\Gamma(\alpha+\beta+x+3)} [(\beta+x)(2+x) + a + 2] \end{aligned} \quad (13)$$

για $x = 0, 1, 2, \dots$ και $\alpha > 0, \beta > 0$.

(β) Η παραγοντική ροπή k τάξης δίνεται από τη σχέση

$$\mu_{[k]}(X) = \frac{\Gamma(\alpha-k)\Gamma(\beta+k)}{k!\Gamma(\beta)} \times \frac{\alpha + \beta(\beta+1) + k^2}{\alpha + \beta}, \quad (14)$$

$k = 1, 2, \dots$

Οι ροπές στο μηδέν προκύπτουν από τη σχέση (14). Συγκεκριμένα

$$E(X) = \mu(\alpha, \beta) = \frac{\beta(2\beta + \alpha + 1)}{(\alpha + \beta)(\alpha - 1)}, \quad \alpha > 1$$

$$E(X^2) = \frac{\beta[6 + 2\beta(5 + 3\beta) + \alpha(\alpha + 4\beta + 1)]}{(\alpha + \beta)(\alpha - 1)(\alpha - 2)}, \quad \alpha > 2$$

Επιπρόσθετα, οι πιθανότητες της σχέσης (13) υπολογίζονται αναδρομικά κάνοντας χρήση του παρακάτω τύπου

$$\Pr(X = x) = \frac{(\beta + x)(x + 2) + a + 2}{(\beta + x - 1)(x + 1) + a + 2} \frac{\beta + x - 1}{a + \beta + x + 2} \Pr(X = x - 1)$$

για $x = 1, 2, \dots$, με αρχική συνθήκη

$$\Pr(X = 0) = \frac{a(a + 1)(2a + \beta + \alpha + 2)}{(\alpha + \beta + 2)(\alpha + \beta + 1)(\alpha + \beta)}$$

Η συγκεκριμένη κατανομή αποδεικνύεται πως είναι μονοκόρυφη με υψηλό ποσοστό μηδενικών τιμών και παρουσιάζει φυσικά υπερδιασπορά. Και τα δύο αυτά χαρακτηριστικά συναντώνται συχνά στον κλάδο ασφάλισης αυτοκινήτου.

2.8 Η ΚΑΤΑΝΟΜΗ ΒΕΤΑ - NEGATIVE BINOMIAL

Ολοκληρώνοντας τη μελέτη κατανομών που βρίσκουν εφαρμογή στον ασφαλιστικό κλάδο για την περιγραφή του αριθμού των απαιτήσεων σε ένα χαρτοφυλάκιο ασφαλιστηρίων συμβολαίων κρίνεται αναγκαία η αναφορά στο μοντέλο της Beta-Negative Binomial κατανομής. Πρόκειται για μία κατανομή τριών παραμέτρων, προερχόμενη από μείξη της βήτα και της αρνητικής διωνυμικής κατανομής. Όπως έχουμε ήδη δει, η Poisson κατανομή ενώ στην πράξη βρίσκει εφαρμογή για την μοντελοποίηση αριθμητικών δεδομένων, είναι δυνατόν να μην ταιριάζει με τις εμπειρικές παρατηρήσεις για ροπές μεγαλύτερης τάξης. Αυτό συμβαίνει εξαιτίας της μίας και μοναδικής ελεύθερης παραμέτρου της κατανομής Poisson η οποία δεν επιτρέπει στη διακύμανση να πάρει τιμή ανεξάρτητη από τη μέση τιμή της. Σε αντίθεση με την Poisson, η N.B. κατανομή έγινε ιδιαίτερα δημοφιλής ως πιο ευέλικτη κατά την εφαρμογή της. Βασικό χαρακτηριστικό της

αρνητικής διωνυμικής κατανομής είναι η καλύτερη εφαρμογή της σε δεδομένα που παρουσιάζουν υπερδιασπορά παρόλο που μπορεί να μην έχουν απαραίτητως και βαριά ουρά. Στο σημείο αυτό, αξίζει να σημειωθεί πως ενώ η ιδιαίτερα βαριά ουρά υποδηλώνει υπερδιασπορά το αντίστροφο δεν ισχύει απαραίτητα.

Στο δημοσίευμα του Wang (2011) γίνεται αναφορά σε μια μίξη της Poisson με την N.B. όπου υποθέτουμε πως η πιθανότητα επιτυχίας ακολουθεί την Βήτα κατανομή. Η κατανομή που προκύπτει από την υπόθεση αυτή καλείται Βήτα-Αρνητική διωνυμική, έχει μία επιπλέον ελεύθερη παράμετρο και καλή προσαρμοστικότητα σε αριθμητικά δεδομένα.

Όπως έχουμε ξαναδεί αν μια τυχαία μεταβλητή ακολουθεί την αρνητική διωνυμική κατανομή ($Y \sim NB(r, p)$) θα έχει συνάρτηση μάζας πιθανότητας

$$\Pr(Y = k) = \frac{\Gamma(k+r)}{k! \Gamma(r)} p^r (1-p)^k, \quad k = 0, 1, 2, \dots \quad (15)$$

όπου $0 < p < 1$ και το $r > 0$ δεν είναι απαραίτητα ακέραιος αριθμός. Επιπλέον μία τυχαία μεταβλητή P θα λέμε ότι ακολουθεί τη Βήτα κατανομή ($P \sim Beta(\alpha, \beta)$) αν έχει συνάρτηση πυκνότητας πιθανότητας

$$f(p) = \frac{p^{\alpha-1} (1-p)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < p < 1 \quad (16)$$

όπου $\alpha, \beta > 0$ και $B(\alpha, \beta)$ είναι ο συμβολισμός για την συνάρτηση Βήτα.

Μία τυχαία μεταβλητή Z θα ακολουθεί την Βήτα Αρνητική διωνυμική (BNB) κατανομή με παραμέτρους r, α, β εφόσον ικανοποιεί τη στοχαστική αναπαράσταση

$$Z|P = p \sim NB(r, p)$$

$$P \sim Beta(\alpha, \beta).$$

Για ευκολία χρησιμοποιούμε το συμβολισμό $Z \sim BNB(r, \alpha, \beta)$. Εφόσον η αρνητική διωνυμική είναι μία μείξη της κατανομής Poisson, μπορεί παρομοίως να χαρακτηριστεί και η κατανομή BNB ως τέτοιου είδους μοντέλο.

- **Συνάρτηση πιθανότητας και ροπές**

Στο κομμάτι αυτό, θα δοθούν ένας κλειστός τύπος και η παραγοντική ροπή της BNB κατανομής. Αν για μία τυχαία μεταβλητή Z , ισχύει ότι $Z|p \sim NB(r, p)$ και $p \sim Beta(\alpha, \beta)$, τότε σύμφωνα με τις σχέσεις (15) και (16) η συνάρτηση πιθανότητας της Z δίνεται από τη σχέση

$$\begin{aligned} \Pr(Z = k) = p_k &= \int_0^1 \Pr(Z = k|p) f(p) dp \\ &= \int_0^1 \frac{\Gamma(k+r)}{k! \Gamma(r)} p^r (1-p)^k \frac{p^{\alpha-1} (1-p)^{\beta-1}}{B(\alpha, \beta)} dp \\ &= \frac{\Gamma(k+r) B(\alpha+r, \beta+k)}{k! \Gamma(r) B(\alpha, \beta)}. \end{aligned}$$

Συνεπώς, η ροπογεννήτρια συνάρτηση, $g(t) = E(t^Z)$, μπορεί να βρεθεί από τον τύπο

$$g(t) = H(r, \beta; r + \alpha + \beta, t) \Pr(Z = 0)$$

όπου $H(a, b; c, t) = \sum_{n=0}^{\infty} ((a)_n (b)_n / (c)_n t^n / n!)$ είναι η υπεργεωμετρική συνάρτηση (βλέπε Abramowitz και Stegun, 1964 (σελίδα 558)) και $(x)_n = \Gamma(x+n)/\Gamma(x)$ για κάθε $x > 0$.

Λαμβάνοντας την παράγωγο της συνάρτησης $g(t)$ ως προς τη μεταβλητή t , παίρνουμε

$$g^{(k)}(t) = \frac{(r)_k (\beta)_k}{(r+\alpha+\beta)_k} H(r+k, \beta+k; r+\alpha+\beta+k, t) \Pr(Z=0), \alpha-k > 0 \quad (17)$$

Λόγω του ότι για την παραγοντική ροπή έχουμε ότι

$$m_{[k]}(Z) = E[\prod_{j=0}^{k-1} (Z-j)] = g^{(k)}(1), \quad \text{για } k \geq 1,$$

έπεται άμεσα από τη σχέση

$$H(a, b; c, 1) = \Gamma(c)\Gamma(c-a-b)/\Gamma(c-a)\Gamma(c-b), \text{ για } c-a-b > 0$$

ότι

$$m_{[k]}(Z) = \frac{\Gamma(r+k) B(\alpha-k, \beta+k)}{\Gamma(r) B(\alpha, \beta)}, \quad k < \alpha.$$

Οι ροπές περί το μηδέν μπορούν εύκολα να βρεθούν από τη σχέση $m_n = \sum_{k=1}^n S_n^k m_{[k]}$, όπου ο όρος S_n^k δηλώνει τους αριθμούς Stirling δευτέρου είδους.

Καθώς το r τείνει στο μηδέν ($r \rightarrow 0^+$), η BNB συγκλίνει σε μία κατανομή εκφυλισμένη στο μηδέν $\Pr(Z=0) = 1$. Επιπλέον, χρησιμοποιώντας τον τύπο του Stirling, $\Gamma(s+1) = \sqrt{2\pi s} \left(\frac{s}{e}\right)^s e^{\theta/12s}$ για $s > 0$ και $\theta \in (0,1)$, έχουμε ότι $\lim_{r \rightarrow +\infty} p_k = 0$, για οποιοδήποτε k . Πλέον γίνεται φανερό ότι οι παράμετροι r και β είναι ανταλλάξιμες.

Θεωρούμε μία ακολουθία BNB κατανομών όπου η παράμετρος a τείνει στο άπειρο με τέτοιο τρόπο ώστε η μέση τιμή της κατανομής Βήτα να παραμένει σταθερή. Δηλώνοντας τη μέση τιμή ως $\frac{\alpha}{\alpha+\beta} = \omega$, η παράμετρος β θα πρέπει να είναι ίση με $\frac{1-\omega}{\omega}$. Υπό αυτή την παραμετροποίηση, η συνάρτηση πιθανότητας της BNB της σχέσης (17) γίνεται

$$p_k = \binom{k+r-1}{k} \frac{\Gamma(a+r)\Gamma\left(\frac{1-\omega}{\omega}a+k\right)\Gamma\left(\frac{\alpha}{\omega}\right)}{\Gamma\left(\frac{\alpha}{\omega}+r+k\right)\Gamma(a)\Gamma\left(\frac{1-\omega}{\omega}a\right)}, \quad k = 0,1,2,\dots$$

Αν τώρα θεωρήσουμε ότι $a \rightarrow \infty$, από τον τύπο του Stirling θα έχουμε ότι:

$$p_k \rightarrow \binom{k+r-1}{k} \omega^r (1-\omega)^k$$

η οποία αποτελεί την συνάρτηση πιθανότητας της NB(r, ω). Με άλλα λόγια, η εναλλακτικά παραμετροποιημένη BNB κατανομή συγκλίνει στη NB και η παράμετρος a , είναι εκείνη η οποία ελέγχει την απόκλιση από την αρνητική διωνυμική κατανομή. Για το λόγο αυτό, καλούμε την παράμετρο a , παράμετρο ουράς.

2.9 Η ΚΑΤΑΝΟΜΗ GENERALIZED POISSON

Η μοντελοποίηση δεδομένων που αφορούν τη συχνότητα των απαιτήσεων για ένα χαρτοφυλάκιο ασφαλισμένων, αποτελεί ένα από τα σημαντικότερα κεφάλαια της θεωρίας κινδύνων. Παραδοσιακά, η κατανομή Poisson, κατά την οποία ο μέσος αριθμός των απαιτήσεων είναι ίσος με τη διακύμανση αυτών, καθώς και Αρνητική διωνυμική, κατά την οποία η διακύμανση του αριθμού των απαιτήσεων υπερβαίνει τον μέσο αυτών, χρησιμοποιούνται κατά κόρον λόγω των πολύ “βολικών” μαθηματικών τους ιδιοτήτων. Αρκετοί μελετητές, όπως οι *Gossiaux & Lemaire* (1981) όπως και ο *Wilmot* (1987), έχουν προβεί στην αναζήτηση εναλλακτικών κατανομών από τις δύο προαναφερθέντες. Ο *Consul* (1990), συνέκρινε τη Γενικευμένη Κατανομή Poisson, (GPD), η οποία προτάθηκε από τον ίδιο και τον *Jain* (1973), με αρκετές άλλες γνωστές κατανομές και, συμπέρανε ότι η Γενικευμένη Κατανομή Poisson είναι ένα λογικό μοντέλο για την περιγραφή δεδομένων που αφορούν τη συχνότητα των απαιτήσεων. Οι *Goovaerts & Kaas* (1991), παρουσίασαν ένα αναδρομικό σύστημα για τον υπολογισμό των συνολικών απαιτήσεων της κατανομής, υπό τις υπόθεση ότι οι απαιτήσεις είναι ανεξάρτητα και ομοιόμορφα κατανεμημένες διακριτές τυχαίες μεταβλητές, σε σχέση με τη συχνότητα απαίτησης της Γενικευμένης Κατανομής Poisson.

Σε αυτό το σημείο θα προσπαθήσουμε να παρουσιάσουμε κάποιες βασικές σχέσεις – ιδιότητες που αφορούν τη Γενικευμένη Κατανομή Poisson.

Η Γενικευμένη Κατανομή Poisson (GPD)

Οι *Consul & Jain* (1973), πρότειναν μία νέα γενίκευση της κατανομής Poisson η οποία και τροποποιήθηκε από τους *Consul & Shoukri* (1985), θα λέμε ότι μια τυχαία μεταβλητή N , ακολουθεί τη Γενικευμένη κατανομή Poisson (GPD) αν η συνάρτηση μάζας πιθανότητάς της δίνεται από τη σχέση :

$$Pr(N = n) = p_n(\lambda, \theta) = \begin{cases} \lambda(\lambda + n\theta)^{n-1} \frac{\exp(-\lambda - n\theta)}{n!}, & \text{για } n = 0, 1, 2, \dots \\ 0 & \text{, για } n > m \text{ όταν } \theta < 0 \end{cases}$$

και μηδέν διαφορετικά, όπου $\lambda > 0$, $\max(-1, -\lambda/m) \leq \theta < 1$ και $m(\geq 4)$ είναι ο μεγαλύτερος θετικός ακέραιος για τον οποίο η ποσότητα $\lambda + \theta m$ είναι θετική όταν το θ λαμβάνει αρνητικές τιμές.

Η γενικευμένη κατανομή Poisson εκφυλίζεται στην κατανομή Poisson όταν $\theta = 0$ ενώ συγχρόνως αποκτά τις ιδιότητες της υπερ-διασποράς και της υπο-διασποράς σύμφωνα με το αν $\theta > 0$ ή $\theta < 0$ αντίστοιχα. Χάριν απλότητας θα χρησιμοποιούμε στο εξής την περίπτωση όπου $\theta > 0$. Οι *Ambagaspitiya & Balakrishnan* (1994) παρουσίασαν την πιθανογεννήτρια καθώς και την ροπογεννήτρια της κατανομής σύμφωνα με τη συνάρτηση W του Lambert όπου το θ είναι θετικό όπως φαίνεται στις παρακάτω σχέσεις.

$$M_N(t) = \exp\left\{-\frac{\lambda}{\theta} [W(-\theta \exp(-\theta + t)) + \theta]\right\}$$

$$P_N(z) = \exp\left\{-\frac{\lambda}{\theta} [W(-\theta z \exp(-\theta)) + \theta]\right\}$$

όπου W είναι η συνάρτηση του Lambert για την οποία ισχύει

$$W(x) \exp(W(x)) = x$$

- **Κεντρικές ροπές**

Στην προσπάθεια μας να καταλήξουμε στις κεντρικές ροπές της γενικευμένης Poisson μας είναι απαραίτητες οι παραπάνω σχέσεις των *Ambagaspitiya & Balakrishnan* (1994) από τις οποίες οδηγούμαστε στις τέσσερις πρώτες ροπές της κατανομής.

$$\mu_1 = \lambda M$$

$$\mu_2 = \lambda M^3$$

$$\mu_3 = \lambda(3M - 2)M^4$$

$$\mu_4 = 3\lambda^2 M^6 + \lambda(15M^2 - 20M + 6)M^5$$

όπου $M = (1 - \theta)^{-1}$.

Έστω x_1, x_2, \dots, x_n τυχαίο δείγμα n αντικειμένων που ακολουθούν το μοντέλο της Γενικευμένης Poisson. Αν οι δειγματικές τιμές κατηγοριοποιηθούν σε κλάσεις συχνοτήτων και n_i είναι η συχνότητα της i -οστής κλάσης ($n_i = \text{πλήθος}\{x_j: 1 \leq j \leq n, x_j = i\}$), ο εκτιμητής μεγίστης πιθανοφάνειας του θ ($\hat{\theta}$), δίνεται από την επίλυση της εξίσωσης

$$\sum_{i=1}^k \frac{n_i i(i-1)}{\bar{x} + (i-\bar{x})\theta} - n\bar{x} = 0$$

όπου k , είναι το πλήθος των κλάσεων, $n = \sum_{i=1}^k n_i$ και \bar{x} είναι ο δειγματικός μέσος. Σημειώνουμε ότι ο για $k = 0$ ή $k = 1$ δεν παίρνουμε τιμή για τον εκτιμητή του θ . Επιπλέον με βάση τα παραπάνω ο εκτιμητής μεγίστης πιθανοφάνειας για το λ είναι

$$\hat{\lambda} = \bar{x}(1 - \hat{\theta})$$

3 ΕΦΑΡΜΟΓΗ ΓΙΑ ΔΕΔΟΜΕΝΑ ΑΡΙΘΜΟΥ ΑΠΑΙΤΗΣΕΩΝ ΣΤΟΝ ΚΛΑΔΟ ΑΥΤΟΚΙΝΗΤΟΥ

3.1 ΕΙΣΑΓΩΓΗ

Στο Κεφάλαιο 3 έγινε η μελέτη μιας σειράς υποψήφιας κατανομών για τη μοντελοποίηση του αριθμού των απαιτήσεων ενός ασφαλιστικού χαρτοφυλακίου. Το Κεφάλαιο 3 περιλαμβάνει προσαρμογές αυτών των κατανομών κυρίως σε δεδομένα χαρτοφυλακίου ασφάλισης αυτοκινήτου που είχε συλλέξει ο Lemaire (1976) από μια Βελγική εταιρία αλλά και σε δεδομένα που προέρχονται από διάφορες άλλες πηγές. Για την εξαγωγή ορισμένων αριθμητικών αποτελεσμάτων θα χρησιμοποιηθεί η R. Πριν προχωρήσουμε όμως στον έλεγχο ανεύρεσης της καταλληλότερης κατανομής ως προς τη συμβατότητά της με τα δεδομένα μας, θα αναλύσουμε τις τεχνικές που θα συμβάλλουν στην εξαγωγή του συμπεράσματος αυτού.

Το πρώτο βήμα στην ανάλυση δεδομένων είναι η επιλογή ενός εύλογου παραμετρικού μοντέλου για τις παρατηρήσεις και κατόπιν ο υπολογισμός των υποκείμενων παραμέτρων. Το πρόβλημα της εκτίμησης παραμέτρων είναι κλασικό στην στατιστική, και μπορεί να προσεγγιστεί με διάφορους τρόπους. Θα εξετάσουμε δύο συνήθεις και λογικές διαδικασίες, τη μέθοδο της μέγιστης πιθανοφάνειας και τη μέθοδο των ροπών. Αν και τα αποτελέσματα που λαμβάνουμε με αυτές τις δύο διαδικασίες είναι, συχνά, σχεδόν ίδια, οι προσεγγίσεις, είναι αρκετά διαφορετικές. Η μέθοδος της μέγιστης πιθανοφάνειας αλλά και μερικές άλλες μέθοδοι βλέπουν τις παραμέτρους ως ποσότητες οι τιμές των οποίων είναι καθορισμένες αλλά άγνωστες. Η καλύτερη εκτίμηση της τιμής τους καθορίζεται να είναι αυτή που μεγιστοποιεί την πιθανότητα να προκύψουν τα δείγματα τα οποία παρατηρήθηκαν.

3.2 ΕΚΤΙΜΗΤΕΣ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

Ο εκτιμητής μέγιστης πιθανοφάνειας (*maximum likelihood estimator*, M.L.E.) είναι η τιμή της παραμέτρου (ή της διανυσματικής παραμέτρου) η οποία καθιστά τα παρατηρηθέντα δεδομένα πιο πιθανά να έχουν συμβεί δοθέντος ότι έχουν προέλθει από τη διαδικασία παραγωγής των δεδομένων που έχουμε υποθέσει. Το μόνο που χρειαζόμαστε για να εξάγουμε τον εκτιμητή μέγιστης πιθανοφάνειας (στο εξής

Ε.Μ.Π.) είναι να κατασκευάσουμε ένα στατιστικό μοντέλο στη μορφή μιας συνάρτησης πιθανοφάνειας που δίνει την πιθανότητα να πάρουμε τα δεδομένα που έχουμε παρατηρήσει. Όσο μεγαλύτερη είναι η τιμή της πιθανοφάνειας τόσο καλύτερο το μοντέλο. Οι Ε.Μ.Π. έχουν πολλές επιθυμητές ασυμπτωτικές ιδιότητες: τη συνέπεια, την αποτελεσματικότητα, ασυμπτωτική κανονικότητα και η αρχή του αναλλοίωτου. Τα πλεονεκτήματα ενός Ε.Μ.Π. είναι ότι χρησιμοποιεί πλήρως όλες τις πληροφορίες σχετικά με τις παραμέτρους που περιέχονται στα δεδομένα και ότι είναι εξαιρετικά ευέλικτος.

Η συνάρτηση πιθανοφάνειας (likelihood function) ορίζεται ως η από κοινού συνάρτηση πυκνότητας (ή σ.π.π.) ενός τυχαίου δείγματος X_1, X_2, \dots, X_n ως συνάρτηση της άγνωστης παραμέτρου θ ($\theta \in \Theta$), και δίνεται από τη σχέση

$$L(\theta|\mathbf{x}) = L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta), \quad \theta \in \Theta.$$

Η $L(\theta|\mathbf{x})$ ονομάζεται συνάρτηση πιθανοφάνειας καθότι εκφράζει πόσο πιθανοφανείς ή διαφορετικά πόσο σύμφωνες με το συγκεκριμένο δείγμα είναι οι διάφορες τιμές της παραμέτρου θ .

Εκτιμητής μέγιστης πιθανοφάνειας $\hat{\theta}$ καλείται η σ.σ. $\hat{\theta} = \hat{\theta}(\mathbf{x})$ που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $L(\theta|\mathbf{x})$ ως προς θ , δηλαδή

$$L(\hat{\theta} | x_1, x_2, \dots, x_n) = \max_{\theta \in \Theta} L(\theta | x_1, x_2, \dots, x_n).$$

Συνήθως είναι ευκολότερο να εργαστεί κανείς με το λογάριθμο της πιθανοφάνειας από ότι με την ίδια την πιθανοφάνεια. Δεδομένου ότι ο λογάριθμος αυξάνεται μονότονα, το $\hat{\theta}$ που μεγιστοποιεί την $L(\theta|\mathbf{x})$, μεγιστοποιεί επίσης και τη συνάρτηση $l(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x})$. Οπότε χρησιμοποιούμε την $l(\theta|\mathbf{x})$ για ευκολία. Λόγω των ασυμπτωτικών ιδιοτήτων τους οι εκτιμητές μέγιστης πιθανοφάνειας είναι πιο επιθυμητοί. Επιπλέον, σε δείγματα μικρού μεγέθους συχνά εμφανίζουν μικρότερη διασπορά από άλλους εκτιμητές, όπως για παράδειγμα από αυτούς που προκύπτουν με τη μέθοδο των ροπών.

3.3 ΜΕΘΟΔΟΣ ΡΟΠΩΝ

Ένας άλλος τρόπος εύρεσης σημειακών εκτιμητών, είναι με τη μέθοδο των ροπών (*moments method*). Το σκεπτικό αυτής της μεθόδου είναι ότι εξισώνονται οι θεωρητικές ροπές με τις δειγματικές. Έτσι συνδέονται οι εκτιμώμενες παράμετροι με στατιστικές συναρτήσεις και από τη λύση των εξισώσεων που προκύπτουν, υπολογίζονται οι εκτιμητές.

Εάν λοιπόν

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

είναι η δειγματική ροπή k -τάξης, και $E[X^k]$ η αντίστοιχη θεωρητική ροπή, δημιουργούνται οι εξισώσεις

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k = E[X^k], \quad k = 1, 2, \dots$$

Παίρνουμε τόσες εξισώσεις όσες είναι οι παράμετροι της κατανομής που θέλουμε να εκτιμήσουμε. Η λύση του συστήματος των εξισώσεων δίνει τους ζητούμενους εκτιμητές.

3.4 ΈΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ χ^2

Τα παραμετρικά τεστ χρησιμοποιούν ποσοτικά δεδομένα και μπορούν να εφαρμοστούν μόνο όταν ικανοποιούνται ορισμένες προϋποθέσεις. Χαρακτηρίζονται από υποθέσεις σχετικές είτε για το σχήμα της κατανομής των πληθυσμών είτε για τις παραμέτρους της κατανομής. Τι γίνεται όμως αν ο ερευνητής έχει στη διάθεσή του ποιοτικά δεδομένα, ή δεν ικανοποιούνται οι προϋποθέσεις των παραμετρικών τεστ; Για αυτές τις περιπτώσεις η επαγωγική στατιστική παρέχει ένα άλλο σύνολο από στατιστικά τεστ ελέγχου υποθέσεων, τα οποία ονομάζονται μη-παραμετρικά (*non-parametric tests*) και χαρακτηρίζονται από ορισμένα πλεονεκτήματα όπως:

(α) Δεν απαιτείται η επαλήθευση της καταλληλότητας κάποιας υπόθεσης.

(β) Για μικρά δείγματα οι δοκιμασίες οδηγούν σε ακριβή συμπεράσματα και όχι σε προσεγγίσεις.

(γ) Τα δεδομένα που χρησιμοποιούνται είναι συχνότητες (counts) και όχι μετρήσεις (measurements).

Όταν λοιπόν ένα παραμετρικό τεστ δε μπορεί να χρησιμοποιηθεί, συνήθως υπάρχει ένα μη- παραμετρικό τεστ που μπορεί να το αντικαταστήσει. Αν και τα μη- παραμετρικά τεστ μπορούν να χρησιμοποιηθούν και σε περιπτώσεις που είναι δυνατή η εφαρμογή ενός παραμετρικού τεστ, συνήθως αυτό αποφεύγεται γιατί τα παραμετρικά τεστ δίνουν πιο αξιόπιστα αποτελέσματα. Το X^2 τεστ (Chi-Square test) είναι ίσως το πιο δημοφιλές μη-παραμετρικό τεστ. Αν και υπάρχουν διάφορες παραλλαγές του, ανάλογα με το διαθέσιμο αριθμό δειγμάτων και το στόχο της ανάλυσης, εμείς θα εξετάσουμε την πιο απλή μορφή του. Σε αυτές τις περιπτώσεις, ο ερευνητής έχει στη διάθεση του ένα δείγμα ποιοτικών δεδομένων οργανωμένο σε ονομαστικές κατηγορίες, και ο στόχος είναι να χρησιμοποιηθούν τα δεδομένα αυτά ώστε να προσδιοριστεί η αναλογία (ή το ποσοστό) του πληθυσμού που ανήκει στην κάθε κατηγορία. Για την επίτευξη αυτού του στόχου διατυπώνεται μια μηδενική υπόθεση, που είτε δηλώνει ότι δεν υπάρχει κάποια συγκεκριμένη προτίμηση στις διαθέσιμες ονομαστικές κατηγορίες (no-preference null-hypothesis), είτε δηλώνει ότι τα ποσοστά που προτιμώνται από τα υποκείμενα δε διαφέρουν από τα ποσοστά άλλων πληθυσμών οι οποίοι αποτελούν το σημείο αναφοράς (no difference from a comparison population).

Και στις δύο περιπτώσεις, αυτό που προσδιορίζει η μηδενική υπόθεση είναι ο αναμενόμενος αριθμός (expected frequency - f^e) των υποκειμένων που ανήκει σε κάθε ονομαστική κατηγορία. Ο έλεγχος υποθέσεων που ακολουθεί αξιολογεί αυτή τη μηδενική υπόθεση, συγκρίνοντας τον αριθμό των υποκειμένων που αναμένεται σε κάθε ονομαστική κατηγορία με τον αριθμό των υποκειμένων που παρατηρείται ότι ανήκει σε κάθε ονομαστική κατηγορία (observed frequency - f^o), με βάση τις μετρήσεις του δείγματος. Ο στατιστικός δείκτης ελέγχου που χρησιμοποιείται για αυτή την αξιολόγηση είναι το X^2 . Οι κρίσιμες τιμές για το στατιστικό δείκτη ελέγχου χ^2 δίνονται από την κατανομή χ^2 ενώ οι μαθηματικές σχέσεις υπολογισμού του είναι οι εξής:

$$\chi^2 = \sum \frac{(f^o - f^e)^2}{f^e}$$

όπου f_e και f_o η αναμενόμενη και η παρατηρούμενη συχνότητα. Επίσης, οι βαθμοί ελευθερίας της υποκείμενης κατανομή χ^2 είναι

$$df = C - 1$$

όπου C είναι ο αριθμός των ονομαστικών κατηγοριών.

Τα συμπεράσματα που προκύπτουν από την εφαρμογή του στατιστικού τεστ χ^2 είναι πιο αξιόπιστα όταν χρησιμοποιείται αντιπροσωπευτικό δείγμα, και η αναμενόμενη συχνότητα όλων των κατηγοριών είναι μεγαλύτερη από πέντε.

3.4.1 Η ΚΑΤΑΝΟΜΗ χ^2

Αν Y είναι μία τυχαία μεταβλητή από την κανονική κατανομή $N(\mu, \sigma^2)$ τότε η τυποποιημένη μεταβλητή

$$Z = \frac{Y - \mu}{\sigma}$$

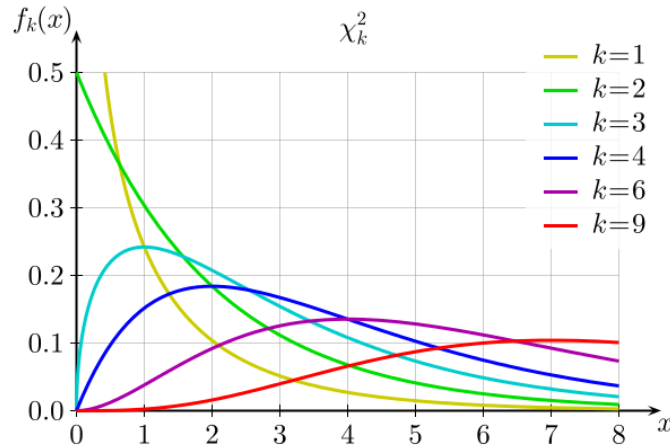
ακολουθεί μια κανονική κατανομή $N(0,1)$, ενώ η τυχαία μεταβλητή

$$\chi^2(1) = \left(\frac{Y - \mu}{\sigma}\right)^2 = Z^2$$

ακολουθεί μια νέα κατανομή που λέγεται χ^2 με 1 βαθμό ελευθερίας. Αν Y_1, Y_2, \dots, Y_k είναι ανεξάρτητες τυχαίες μεταβλητές από την κανονική κατανομή $N(\mu, \sigma^2)$ τότε η τυχαία μεταβλητή

$$\begin{aligned}\chi^2(k) &= \left(\frac{Y_1 - \mu}{\sigma}\right)^2 + \left(\frac{Y_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{Y_k - \mu}{\sigma}\right)^2 \\ &= Z_1^2 + Z_2^2 + \dots + Z_k^2\end{aligned}$$

ακολουθεί την χ^2 με k βαθμούς ελευθερίας (συμβ. $\chi^2(k)$). Η οικογένεια των κατανομών $\chi^2(k)$ είναι ασύμμετρη προς τα δεξιά όπως φαίνεται στην παρακάτω εικόνα



και έχει συνάρτηση πυκνότητας πιθανότητας που δίνεται από τον τύπο

$$f(x|k) = \frac{1}{\Gamma\left(\frac{k}{2}\right)!} \frac{1}{2^{\frac{k}{2}}} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x > 0$$

όπου k είναι οι βαθμοί ελευθερίας. Η αναμενόμενη τιμή της κατανομής $\chi^2(k)$ είναι k και η διακύμανση $2k$.

3.4.2 Η ΜΕΘΟΔΟΣ

Στη δοκιμασία καλής προσαρμογής (goodness-of-fit) υποθέτουμε ότι τα δεδομένα προέρχονται από μια πλήρως προσδιορισμένη κατανομή. Στη γενική περίπτωση υποθέτουμε ότι έχουμε ένα τυχαίο δείγμα με N παρατηρήσεις οι οποίες κατατάσσονται σε k κατηγορίες και συμβολίζουμε με $f_1^0, f_2^0, \dots, f_k^0$ τις παρατηρούμενες συχνότητες, έτσι ώστε

$$\sum_{i=1}^k f_i^0 = N \quad (18)$$

Για να ελέγξουμε τη μηδενική υπόθεση

$$H_0: \text{οι πιθανότητες των κατηγοριών } 1, 2, \dots, k \text{ είναι } \pi_1, \pi_2, \dots, \pi_k$$

με $\sum_{i=1}^k \pi_i = 1$ και εναλλακτική υπόθεση την «Δεν ισχύει η H_0 » εργαζόμαστε ως εξής:

Υπολογίζουμε το αναμενόμενο πλήθος των παρατηρήσεων που θα βρεθούν στην i κατηγορία κάτω από την H_0 που είναι

$$f_i^e = N\pi_i, \quad i = 1, 2, \dots, k$$

Έτσι

$$\sum_{i=1}^k f_i^e = \sum_{i=1}^k N\pi_i = N \sum_{i=1}^k \pi_i = N.$$

Αν οι παρατηρούμενες συχνότητες $f_1^0, f_2^0, \dots, f_k^0$ είναι κοντά στις αναμενόμενες συχνότητες $f_1^e, f_2^e, \dots, f_k^e$, τότε η μηδενική υπόθεση H_0 είναι αληθής.

Όπως είδαμε και παραπάνω, ένα μέτρο της απόκλισης μεταξύ των παρατηρούμενων και αναμενόμενων συχνοτήτων δίνεται από τη σ.σ.ε.

$$\chi^2 = \frac{\sum_{i=1}^k (f_i^0 - f_i^e)^2}{f_i^e}$$

που ακολουθεί την κατανομή χ^2 με $(k - 1)$ βαθμούς ελευθερίας. Η αφαίρεση της μονάδας από τους βαθμούς ελευθερίας επιβάλλεται από τη σχέση (18) σύμφωνα με την οποία όταν προσδιοριστούν οι συχνότητες στις $(k - 1)$ κατηγορίες η συχνότητα της τελευταίας κατηγορίας προσδιορίζεται από τη σχέση (18). Όπως είναι προφανές στην περίπτωση που δεν υπάρχουν διαφορές μεταξύ των f_i^0 και f_i^e τότε $\chi^2 = 0$. Καθώς οι διαφορές μεγαλώνουν, μεγαλώνει και η τιμή του χ^2 . Κατά συνέπεια η δοκιμασία καλής προσαρμογής χ^2 είναι πάντοτε μονόπλευρη όπου η δεξιά ουρά της κατανομής $\chi^2(k - 1)$ προσδιορίζει την περιοχή απόρριψης της δοκιμασίας. Για συγκεκριμένο επίπεδο σημαντικότητας α δεν μπορούμε να αποδεχθούμε την H_0 όταν $\chi^2 > \chi_a^2(k - 1)$, όπου $\chi_a^2(k - 1)$ είναι το άνω α ποσοστιαίο σημείο της κατανομής $\chi^2(k - 1)$ για το οποίο ισχύει

$$P(\chi^2(k - 1) > \chi_a^2(k - 1)) = \alpha.$$

Τέλος αξίζει να σημειωθεί ότι ενώ η κατανομή $\chi^2(k)$ είναι συνεχής, ο υπολογισμός της σ.σ.ε. χ^2 γίνεται με διακριτά δεδομένα. Το γεγονός αυτό δεν έχει ιδιαίτερη σημασία όταν τιμές των αναμενόμενων συχνοτήτων είναι $f_i^e \geq 5$. Στην περίπτωση που υπάρχουν κατηγορίες για τις οποίες ισχύει $f_i^e \leq 5$ τότε αυτές θα πρέπει να συγχωνευτούν με τις διπλανάς έτσι ώστε να ικανοποιηθεί η συνθήκη. Οι βαθμοί ελευθερίας του στατιστικού χ^2 που προκύπτει υπολογίζονται με βάση το νέο αριθμό των κατηγοριών. Εναλλακτικά, μπορούμε να αυξήσουμε το μέγεθος του δείγματος έως ότου γίνουν αποδεκτά όλα τα f_i^e

3.5 ΕΦΑΡΜΟΓΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΟΥ LEMAIRE ΣΕ ΔΙΑΦΟΡΑ ΜΟΝΤΕΛΑ ΚΑΤΑΝΟΜΩΝ.

Ακολούθως δίνονται τα δεδομένα ενός χαρτοφυλακίου από τον κλάδο ασφάλισης αυτοκινήτου που είχε συλλέξει ο Lemaire (1976) από μια Βελγική εταιρία. Ο αριθμός των παρατηρήσεων είναι $n = 106974$, με δειγματικό μέσο $\bar{x} = 0.1011$ και δειγματική διακύμανση $s^2 = 0.1074$ ($s = 0.3277$). Τα δεδομένα παρουσιάζονται στον παρακάτω πίνακα συχνοτήτων.

Πίνακας 1: Αριθμός απαιτήσεων χαρτοφυλακίου (Lemaire (1976))

k	n_k
0	96978
1	9240
2	704
3	43
4	9
> 4	0
Σύνολο	106974

3.5.1 ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ POISSON

Στην πρώτη εφαρμογή θα μελετήσουμε κατά πόσο μπορεί να μοντελοποιήσει η κατανομή Poisson στα παραπάνω δεδομένα. Υποθέτουμε λοιπόν πως πρόκειται για ένα ομοιογενές χαρτοφυλάκιο στο οποίο ο κάθε ασφαλισμένος εγείρει αριθμό απαιτήσεων τέτοιο που ακολουθεί κατανομή Poisson με μέσο λ .

Όπως ήδη ξέρουμε τόσο η μέθοδος μεγίστης πιθανοφάνειας όσο και αυτή των ροπών δίνουν σαν εκτιμητή της παραμέτρου λ της Poisson το $\hat{\lambda} = \bar{x} = 0.1010806$. Θέλουμε λοιπόν να ελέγξουμε αν τα δεδομένα μας προέρχονται από την κατανομή Poisson με $\lambda = 0.1010806$. Εργαζόμαστε με το R ως εξής (δείτε «Poisson δοκιμή λ » στο Παράρτημα) : Οι τιμές, $n_i = 0,1,2,3,4$ αποθηκεύονται στο διάνυσμα Obs1. Στη συνέχεια υπολογίζουμε το μέση τιμή των παρατηρήσεων (mean = 0.1010806) και τις πιθανότητες $p_i = P(X = i)$, $0 \leq i \leq 4$ (διάνυσμα prob1). Σημειώνεται επίσης ότι $p_0 + p_1 + \dots + p_4 = 1$. Κατόπιν υπολογίζουμε τις ποσότητες np_i (διάνυσμα Exp1).

Ο έλεγχος που θα κάνουμε είναι ο

$$H_0: P(E_0) = p_0, P(E_1) = p_1, \dots, P(E_4) = p_4$$

$$H_1: \text{Όχι } H_0$$

όπου $E_i, 0 \leq i \leq 4$ είναι το ενδεχόμενο να σημειωθούν i απαιτήσεις στο χαρτοφυλάκιο των ασφαλισμένων οδηγών. Μη απόρριψη της H_0 οδηγεί στο συμπέρασμα ότι τα δεδομένα προέρχονται από την κατανομή $P(0.1010806)$.

Εκτελώντας λοιπόν την παραπάνω δοκιμή στην R καταγράφονται τα παρακάτω αποτελέσματα.

Πίνακας 2: Έλεγχος καλής προσαρμογής της κατανομής Poisson στα δεδομένα του Πίνακα 1

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	$9.038601e - 01$	96689.54	96978
$X = 1$	$9.136276e - 02$	9773.440	9240
$X = 2$	$4.617503e - 03$	493.952	704
$X = 3$	$1.5558e - 04$	16.64302	43
$X = 4$	$4.012371e - 06$	0.4292194	9

Παρατηρούμε ότι $np_4 < 5$ οπότε ο έλεγχος καλής προσαρμογής δεν θα είναι αξιόπιστος.

Αποφασίζουμε να ομαδοποιήσουμε τις παρατηρήσεις 3, 4 και να κάνουμε τον εξής έλεγχο:

$$H_0: P(E_0) = p_0, P(E_1) = p_1, P(E_2) = p_2, P(E_3) = p_3$$

$$H_1: \text{όχι } H_0$$

Με βάση αυτό τον τρόπο ομαδοποίησης των δεδομένων έχουμε ένα καινούργιο πίνακα παρατηρήσεων.

Πίνακας 3: Ομαδοποιημένα δεδομένα του Πίνακα 1

k	n_k
0	96978
1	9240
2	704
3	52
Σύνολο	106974

Με τον ίδιο τρόπο, εκτελώντας το χ^2 τεστ στην R για τα παραπάνω δεδομένα (δείτε στο Παράρτημα (Poisson δοκιμή 2)) λαμβάνουμε τα ακόλουθα αποτελέσματα.

Πίνακας 4: Έλεγχος καλής προσαρμογής της κατανομής Poisson στα δεδομένα του Πίνακα 3

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	0.9038601459	96689.53525	96978
$X = 1$	0.0913627588	9773.43976	9240
$X = 2$	0.0046175029	493.95276	704
$X = 3$	0.0001595924	17.07224	52

Παρατηρούμε ότι $np_i > 5$ για $0 \leq i \leq 3$ οπότε προχωρούμε στην εκτέλεση του ελέγχου με τα ακόλουθα αποτελέσματα

$\chi^2 = 190.754$	$df = 3$	$p - value < 2.2e - 16$
--------------------	----------	-------------------------

Η προσαρμογή είναι εξαιρετικά φτωχή και η μηδενική υπόθεση απορρίπτεται από το χ^2 τεστ. Παρατηρούμε ότι δεν υπάρχει επαρκής μάζα πιθανότητας στη δεξιά ουρά της κατανομής Poisson. Το μοντέλο της Poisson δεν μπορεί να εφαρμοστεί στο συγκεκριμένο χαρτοφυλάκιο. Η υπόθεση της ομοιογένειας δεν φαίνεται είναι συμβατή με τη συγκεκριμένη στατιστική ανάλυση. Χρειαζόμαστε μια κατανομή της οποίας η διασπορά θα ξεπερνά το μέσο της, θα παρουσιάζει δηλαδή υπερδιασπορά.

- **Επιπλέον σχόλια για τα τεστ καλής προσαρμογής**

Όπως έχουμε δει και παραπάνω το πιο διαδομένο τεστ καλής προσαρμογής είναι το χ^2 τεστ. Αν συμβολίζουμε με m τον μέγιστο αριθμό απαιτήσεων στα παρατηρηθέντα

δεδομένα και με r ας συμβολίσουμε τον αριθμό των υπό εκτίμηση παραμέτρων. Τότε η σ.σ.ε. παίρνει τη μορφή

$$\chi^2 = \sum_{k=0}^m \frac{(n_k - np_k)^2}{np_k}$$

όπου n είναι το πλήθος των παρατηρήσεων και n_k ο συνολικός αριθμός των ασφαλιστηρίων συμβολαίων που έγειραν k απαιτήσεις.

Η παραπάνω σ.σ.ε. ακολουθεί ασυμπτωτικά την χ^2 κατανομή με $m - r$ βαθμούς ελευθερίας. Να σημειωθεί πως ο γνωστός κανόνας ότι «για κάθε παράμετρο που εκτιμάται χάνεται ένας βαθμός ελευθερίας» δεν επαληθεύεται σε εκτιμητές που δεν είναι ασυμπτωτικά ισοδύναμοι με τους πολυωνυμικούς εκτιμητές μέγιστης πιθανοφάνειας. Στην πραγματικότητα για τους τυπικούς εκτιμητές μέγιστης πιθανοφάνειας τα πραγματικά κρίσιμα σημεία του τεστ κυμαίνονται μεταξύ αυτών του χ_{m-r}^2 και του χ_m^2 τεστ. Η χρήση των κρίσιμων σημείων του χ_{m-r}^2 τεστ οδηγεί σε στρέβλωση του πραγματικού επιπέδου σημαντικότητας (Lemaire (1976)).

Διαδικασία ομαδοποίησης

Το χ^2 τεστ είναι ένα ασυμπτωτικό τεστ. Γι αυτό το λόγο είναι απαραίτητο να κάνουμε χρήση ομαδοποίησης των δεξιών ουρών των κατανομών έτσι ώστε η αναμενόμενη συχνότητα np_k να είναι αρκετά υψηλές. Κατά καιρούς έχουν προταθεί ποικίλες διαδικασίες ομαδοποίησης (D'Agostino & Stefens (1986)). Δύο από τις πιο διαδεδομένες είναι

- (A) (i) Όλα τα $np_k \geq 1$
- (ii) Τουλάχιστον το 80% των $np_k \geq 5$
- (B) Όλα τα $np_k \geq 5$

Στην περίπτωση των κατανομών που χρησιμοποιούνται για τον αριθμό των απαιτήσεων, το συμπέρασμα του τεστ είναι άρρηκτα συνδεδεμένο με τη μέθοδο ομαδοποίησης που θα χρησιμοποιήσουμε. Ωστόσο παρόλο που το μοντέλο της κατανομής Poisson όπως είδαμε απορρίφθηκε σύμφωνα και με τους δύο κανόνες ομαδοποίησης, τα περισσότερα μοντέλα που θα παρουσιαστούν στη συνέχεια θα απορριφθούν με βάση τον κανόνα (A) και θα γίνουν δεκτά με βάση τον κανόνα (B).

Είναι εμφανές λοιπόν πως μπορούμε να οδηγηθούμε σε εκ διαμέτρου αντίθετα συμπεράσματα ανάλογα με τη μέθοδο ομαδοποίησης που θα χρησιμοποιήσουμε.

3.5.2 ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ ΑΡΝΗΤΙΚΗΣ ΔΙΩΝΥΜΙΚΗΣ ΚΑΤΑΝΟΜΗΣ

Η απόρριψη του μοντέλου της κατανομής Poisson υποδεικνύει την ετερογένεια που παρουσιάζει το χαρτοφυλάκιο των ασφαλισμένων οδηγών. Έτσι προκύπτει η ανάγκη για ένα μοντέλο στο οποίο να αντανακλάται η διαφοροποίηση του κινδύνου. Η Αρνητική διωνυμική κατανομή ανταποκρίνεται σε αυτή την απαίτηση δεδομένου ότι όπως είδαμε και στο Κεφάλαιο 2 χαρακτηρίζεται από υπερδιασπορά. Υπενθυμίζουμε ότι αν μια τυχαία μεταβλητή ακολουθεί την Αρνητική διωνυμική κατανομή με παραμέτρους a και τ , τότε

$$P(X = k) = p_k = \binom{k + a - 1}{k} p^a q^k, \quad k = 0, 1, 2, \dots$$

όπου

$$p = \frac{\tau}{1 + \tau}, \quad q = 1 - p = \frac{1}{1 + \tau},$$

και

$$E[X] = m = \frac{a}{\tau}, \quad V[X] = \sigma^2 = \frac{a}{\tau} \left(1 + \frac{1}{\tau} \right).$$

Χρησιμοποιώντας λοιπόν τη μέθοδο των ροπών προκειμένου να εκτιμήσουμε τις παραμέτρους τ και a και με δεδομένο ότι $\bar{x} = 0.1011$ και $s^2 = 0.1074$ οδηγούμαστε στους εκτιμητές

$$\hat{\tau} = \frac{\bar{x}}{s^2 - \bar{x}} = 15.87526, \quad \hat{a} = \frac{\bar{x}^2}{s^2 - \bar{x}} = 1.604682.$$

Προσαρμόζουμε πάλι στα δεδομένα του Πίνακα 1 την αρνητική διωνυμική κατανομή με τη βοήθεια του R. Όπως φαίνεται στο Παράρτημα (Αρνητική διωνυμική 1) παίρνουμε τα ακόλουθα αποτελέσματα.

Πίνακας 5: Έλεγχος καλής προσαρμογής της αρνητικής διωνομικής στα δεδομένα του Πίνακα 1 με εκτιμητές από τη μέθοδο των ροτών.

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	$9.06625e - 01$	96985.461788	96978
$X = 1$	$8.621181e - 02$	9222.421903	9240
$X = 2$	$6.653357e - 03$	711.736253	704
$X = 3$	$4.737355e - 04$	50.677384	43
$X = 4$	$3.461282e - 05$	3.702672	9

Ο έλεγχος χ^2 σε αυτή την περίπτωση μας δίνει ένα όχι και τόσο αξιόπιστο αποτέλεσμα .

$\chi^2 = 8.86$	$df = 4$	$p - value = 0.0647$
-----------------	----------	----------------------

Το γεγονός αυτό μας οδηγεί σε σφάλμα κατά τον έλεγχο καλής προσαρμογής. Συνεπώς θα πρέπει να συνεχίσουμε με τη διαδικασία ομαδοποίησης των δεδομένων στις δύο τελευταίες τάξεις.

Μπορεί να χρησιμοποιήσαμε τους εκτιμητές που προέκυψαν από τη μέθοδο ροτών λόγω της ευκολίας που παρουσιάζει ο υπολογισμός τους, όμως αξίζει να παραθέσουμε και τα αποτελέσματα με τη μέθοδο μέγιστης πιθανοφάνειας. Η μέθοδος μέγιστης πιθανοφάνειας υπερέρχει ως τεχνική ακόμα και για μικρά δείγματα. Στο συγκεκριμένο παράδειγμα παίρνουμε ως Ε.Μ.Π. για την παράμετρο τ τον $\hat{\tau} = a/\bar{x} =$ όπου το a είναι η λύση της εξίσωσης

$$\sum_{k=0}^m n_k \left(\frac{1}{a} + \dots + \frac{1}{a+k-1} \right) = \sum_{k=0}^m n_k \ln \left(1 + \frac{\bar{x}}{a} \right).$$

Έτσι οι δύο καινούργιοι εκτιμητές που προκύπτουν από τη λύση της εξίσωσης αυτής είναι $\hat{a} = 1.6131$ και $\hat{\tau} = 16.1384$. Εφαρμόζοντας το χ^2 τεστ για τους

συγκεκριμένους εκτιμητές θα έχουμε διαφοροποιημένα αποτελέσματα σε σχέση με τη μέθοδο ροπών.

Πίνακας 6: Έλεγχος καλής προσαρμογής της αρνητικής διωνυμικής κατανομής στα δεδομένα του Πίνακα 1 με εκτιμητές μεγίστης πιθανοφάνειας.

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	$9.075745e - 01$	97086.874055	96978
$X = 1$	$8.542270e - 02$	9138.008013	9240
$X = 2$	$6.512220e - 03$	696.638214	704
$X = 3$	$4.576332e - 04$	48.954852	43
$X = 4$	$3.295067e - 05$	3.524865	9

Αποφασίζουμε να ομαδοποιήσουμε τις παρατηρήσεις 3, 4 και να κάνουμε τον εξής έλεγχο

$$H_0: P(E_0) = p_0, P(E_1) = p_1, P(E_2) = p_2, P(E_3) = p_3$$

$$H_1: \text{Όχι } H_0$$

Με βάση αυτό τον τρόπο ομαδοποίησης των δεδομένων έχουμε ένα καινούργιο πίνακα παρατηρήσεων.

Πίνακας 7: Έλεγχος καλής προσαρμογής της αρνητικής διωνυμικής κατανομής στα δεδομένα του Πίνακα 1 με εκτιμητές μεγίστης πιθανοφάνειας και ομαδοποίηση

	<i>prob2</i>	<i>Exp2</i>	<i>obs2</i>
$X = 0$	0.9064982843	96971.74746	96978
$X = 1$	0.0865167015	9255.03762	9240
$X = 2$	0.0065082982	696.21869	704
$X = 3$	0.0004767161	50.99623	52

Παρατηρούμε ότι $np_i > 5$ για $0 \leq i \leq 3$ οπότε ο έλεγχος είναι αξιόπιστος και η προσαρμογή της αρνητικής διωνυμικής κατανομής κρίνεται καλή.

$\chi^2 = 0.1316$	$df = 3$	$p - value = 0.9878$
-------------------	----------	----------------------

Η αρνητική διωνυμική κατανομή ήδη από το 1714, από τον Montmort, παρουσιάστηκε ως η κατανομή που δείχνει τον αριθμό των δοκιμών που απαιτούνται προκειμένου να λάβουμε ένα συγκεκριμένο αριθμό επιτυχιών. Είναι μια κατανομή που παρουσιάζει ευρύ φάσμα εφαρμογών και είναι ο πιο γνωστός αντιπρόσωπος των «κατανομών μεταδοτικότητας» (contagious distributions). Με τον όρο αυτό αναφερόμαστε σε κατανομές που χρησιμοποιούνται για την περιγραφή μολυσματικών ασθενειών. Ο Neyman ήταν αυτός που χρησιμοποίησε τον όρο πρώτη φορά το 1939. Μάλιστα ο Feller απέδειξε πως ουσιαστικά είναι δύο τα είδη μετάδοσης που ενέχονται σε τέτοιου είδους κατανομές. Το πρώτο είδος «πραγματική μετάδοση» (true contagion) οφείλεται στο γεγονός ότι κάθε «ευνοϊκό» φαινόμενο αυξάνει ή μειώνει την πιθανότητα να έρθουν ευνοϊκά αποτελέσματα. Ο άλλος τύπος, «φαινομενική μετάδοση», έγκειται στην ετερογένεια του πληθυσμού. Η αρνητική διωνυμική λοιπόν, είναι μια κατανομή που μπορεί να υποστηρίξει και τους δύο τύπους μετάδοσης. Η δική μας περίπτωση ωστόσο είναι μια τυπική μορφή φαινομενικής μετάδοσης, το ποσοστό των απαιτήσεων παραμένει σταθερό κατά τη διάρκεια της χρονιάς για κάθε ασφαλισμένο αλλά ποικίλει για το σύνολο του χαρτοφυλακίου. Συμπερασματικά, η «μετάδοση» προέρχεται από την ετερογένεια που παρουσιάζει το σύνολο των ασφαλισμένων. Πραγματική μετάδοση αφορά την περίπτωση που υπάρχει εξάρτηση μεταξύ των απαιτήσεων που εγείρονται και του ρυθμού των ατυχημάτων. Όλοι οι ασφαλισμένοι θεωρείται πως έχουν ίδια πιθανότητα να προκαλέσουν ατύχημα στην αρχή του έτους, όμως με την πρώτη απαίτηση που δημιουργείται αυξάνεται η πιθανότητα να προκαλέσουν μελλοντικές απαιτήσεις.

Στην περίπτωση που το $a \rightarrow \infty$ και το $\tau \rightarrow 0$ και ο μέσος της Αρνητικής Διωνυμικής κατανομής διατηρείται σταθερός, η Αρνητική διωνυμική κατανομή συγκλίνει στην Poisson. Έτσι η παράμετρος a μπορεί να ερμηνευθεί ως το μέτρο του βαθμού ομαδοποίησης κατά την εμφάνιση των απαιτήσεων διαχρονικά. Μετρά την επιρροή που μπορεί να ασκήσει η συχνότητα εμφάνισης μιας απαίτησης μια συγκεκριμένη χρονική στιγμή, σε μελλοντικές χρονικές περιόδους. Στην περίπτωση που το a είναι μεγάλο, ο βαθμός ομαδοποίησης είναι λιγότερο έντονος, οι απαιτήσεις

συμβαίνουν περισσότερο ή λιγότερο τυχαία και η κατανομή τους προσεγγίζει την κατανομή Poisson. Ουσιαστικά, η χρήση της αρνητικής διωνυμικής κατανομής κρίνεται απαραίτητη εφόσον είναι απαραίτητη και η ομαδοποίηση, δηλαδή το a μικρό.

Συνοψίζοντας, η αρνητική διωνυμική κατανομή αποτελεί μια αξιόλογη αναπαράσταση της οδηγικής συμπεριφοράς των ασφαλισμένων. Το συμπέρασμα αυτό επιβεβαιώνει πως ήταν έγκυρη η υπόθεση ότι η συμπεριφορά του κάθε οδηγού ικανοποιεί την κατανομή Poisson και συγχρόνως πως το σύνολο του χαρτοφυλακίου ακολουθεί τη Γάμμα κατανομή.

3.5.3 ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ NEGATIVE BINOMIAL PARETO

Οι Meng, Wei & Whitmore (1999), θεωρούσαν ιδανική περίπτωση να έχουν πρόσβαση σε αριθμητικά δεδομένα που αφορούν τον κάθε ασφαλισμένο ξεχωριστά και μάλιστα σε πολλαπλές χρονικές περιόδους, προκειμένου να εξακριβώσουν την εγκυρότητα της Negative Binomial-Pareto κατανομής, $NBP(\zeta, r, s)$, στο ασφαλιστικό χαρτοφυλάκιο. Ωστόσο επειδή δεν υπήρχε η δυνατότητα αυτή αρκέστηκαν στα δεδομένα του Lemaire όπως παρουσιάστηκαν στον Πίνακα 1. Η συχνότητα n_k δηλώνει το πλήθος των ασφαλισμένων που παρουσίασαν k απαιτήσεις. Με τη βοήθεια της συνάρτησης πιθανότητας της $NBP(\zeta, r, s)$ (δείτε Παράγραφο 2.4)

$$P(X = k) = p_k = \frac{\Gamma(s\zeta + sr + 1)\Gamma(r + sr + 1)\Gamma(s\zeta + k)\Gamma(r + k)}{\Gamma(s\zeta)\Gamma(sr + 1)\Gamma(r)\Gamma(r + sr + s\zeta + k + 1)k!}$$

έγινε η εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας των παραμέτρων της κατανομής και βρέθηκε ότι $\hat{\zeta} = 0.1011$, $\hat{r} = 3.736$ και $\hat{s} = 36.93$. Τα τυπικά σφάλματα για τις παραμέτρους που εκτιμήθηκαν με τη μέθοδο μέγιστης πιθανοφάνειας δείχνουν ότι το r και το s δεν έχουν υπολογιστεί με ακρίβεια. Με βάση τις τιμές των παραμέτρων που υπολόγισαν οι Meng, Wei & Whitmore (1999) κατέληξαν στον παρακάτω πίνακα.

Πίνακας 8: Έλεγχος καλής προσαρμογής της αρνητικής διωνυμικής-Pareto κατανομής στα δεδομένα του Πίνακα 1 με εκτιμητές μεγίστης πιθανοφάνειας.

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	$9.065754e - 01$	96980	96978
$X = 1$	$8.633780e - 02$	9235.9	9240
$X = 2$	$6.563277e - 03$	702.1	704
$X = 3$	$4.842298e - 04$	51.8	43
$X = 4$	$3.645745e - 05$	3.9	9
$X > 4$	$2.8044e - 06$	0.3	0

Τα τυπικά σφάλματα των εκτιμητών των παραμέτρων, όπως παρέχονται μέσω της μεθόδου μεγίστης πιθανοφάνειας, δείχνουν ότι οι ποσότητες r και s δεν εκτιμούνται με μεγάλη ακρίβεια.

Τα αποτελέσματα των εκτιμήσεων, παρέχουν ορισμένες πολύ χρήσιμες ερμηνείες σχετικά με την διαμόρφωση της πορείας της κατανομής. Δοθείσας της εκτίμησης της παραμέτρου r , μπορούμε να μετρήσουμε την έκταση της υπερδιασποράς (σε σχέση ως προς το μοντέλο της Poisson) για την κατανομή των απαιτήσεων που εγείρει ένας ασφαλισμένος, για ένα οποιοδήποτε δεδομένο μέσο ρυθμό απαιτήσεων μ . Η εκτιμώμενη μέση τιμή της προσαρμοσμένης κατανομής NBP, όπως δίνεται στον Πίνακα 8, είναι $\hat{\zeta} = 0.1011$ απαίτηση ανά χρόνο. Με έναν τέτοιο μέσο ρυθμό απαιτήσεων $\mu = 0,01011$ για ένα ασφαλισμένο, μπορούμε να υπολογίσουμε τον παράγοντα μ/r όπως αυτός δίνεται από τον τύπο (9) που υπολογίζει την διακύμανση του αριθμού των απαιτήσεων. Επομένως θα ισχύει ότι $\mu/\hat{r} = 0.1011/3.736 = 0.027$. Συνεπώς, από το αποτέλεσμα αυτό συμπεραίνουμε ότι η διακύμανση είναι αυξημένη κατά περίπου 3% σε σχέση με το μοντέλο της Poisson για τον συγκεκριμένο ασφαλισμένο. Αυτό είναι ένα μικρό ποσοστό. Για ένα ασφαλισμένο που παρουσιάζει μέσο ρυθμό απαιτήσεων ίσο με $\mu = 0.5$ ανά έτος, κάτι που αποτελεί μία αρκετά ακραία περίπτωση, η υπερδιασπορά θα είναι ίση με $\mu/\hat{r} = 0.5/3.736 = 0.13$. Ακόμα και σε αυτή την περίπτωση, ο αυξημένος μέσος ρυθμός απαιτήσεων έχει

πολύ μικρή επίπτωση στη διασπορά. Ένα ακόμα αποτέλεσμα που παρουσιάζει εξαιρετικό ενδιαφέρον και συμπεραίνεται από τους εκτιμητές παραμέτρων της NBP, είναι το εύρος της διακύμανσης του μέσου αριθμού απαιτήσεων μεταξύ των ασφαλισμένων. Ο συντελεστής της διακύμανσης (coefficient of variation) της συνάρτησης πυκνότητας $f(\mu)$ όπως παρουσιάζεται στη σχέση (10) δίνεται από τον τύπο

$$C.V. = \left(\frac{\zeta + r}{\zeta(rs - 1)} \right)^{\frac{1}{2}}.$$

Από την εκτίμηση των παραμέτρων μας, ο συντελεστής διακύμανσης C.V. εκτιμάται να είναι ίσος με 0.53 ή 53%. Η εκτίμηση αυτή, καταδεικνύει ότι υφίσταται μία σημαντική διαφοροποίηση ανάμεσα στους ασφαλισμένους, σε σχέση με την τάση προς δημιουργία απαιτήσεων. Επιπλέον, σημειώνουμε ότι αφού $\delta\hat{\zeta} = 3.731 > 1$, έπεται ότι η κορυφή της συνάρτησης πυκνότητας $f(\mu)$ είναι απομακρυσμένη από την τιμή της μηδέν.

Επίσης, γίνεται σύγκριση ως προς την προσαρμογή μεταξύ του μοντέλου της NBP και δύο άλλων μοντέλων τα οποία έχουν προταθεί στη βιβλιογραφία για την περιγραφή κατανομών απαιτήσεων. Συγκεκριμένα, αναφερόμαστε στο μοντέλο της αρνητικής διωνυμικής κατανομής και στο μοντέλο της Poisson-Inverse Gaussian κατανομής που θα παρουσιαστεί αναλυτικά σε επόμενη παράγραφο. Όπως έχει σημειωθεί νωρίτερα, το μοντέλο NB είναι μία ειδική περίπτωση του μοντέλου NBP, στο οποίο είτε θα ισχύει ότι $r = \infty$ είτε $s = \infty$. Οι προσαρμοσμένες κατανομές τόσο για το αρνητικό διωνυμικό μοντέλο όσο και για το μοντέλο P-IG έχουν παρθεί από τον Lemaire (1995). Ο Πίνακας 9 δίνει την τιμή της στατιστικής χ^2 τεστ καλής προσαρμογής για τα τρία προαναφερθέντα μοντέλα, μετά από ομαδοποίηση των κλάσεων, έτσι ώστε όλες οι αναμενόμενες συχνότητες np_k να είναι ίσες με 2 ή και περισσότερο και τουλάχιστον το 80% να είναι ίσες με 5 ή και περισσότερο. Οι στατιστικές συναρτήσεις ελέγχου χ^2 βασίζονται στους εκτιμητές μέγιστης πιθανοφάνειας οι οποίοι έχουν εξαχθεί από την πλήρη κατανομή συχνοτήτων. Ο πίνακας παρακάτω δίνει την τιμή του στατιστικού χ^2 για κάθε ένα μοντέλο, όπως επίσης και τους αντίστοιχους βαθμούς ελευθερίας. Οι σχετικοί έλεγχοι, σε επίπεδο σημαντικότητας στο 5%, υποδεικνύουν ότι όλα τα μοντέλα παρουσιάζουν έλλειψη ως προς την προσαρμογή τους στα δεδομένα. Το μοντέλο P-IG παρουσιάζει τη

μικρότερη τιμή χ^2 . Η επιλογή κάποιου εκ των τριών μοντέλων με βάση την προσαρμοστικότητα τους όπως προκύπτει από τις παρατηρηθείσες συχνότητες δεν είναι επαρκές κριτήριο αξιολόγησης.

Πίνακας 9: Σύγκριση των τριών μοντέλων ως προς την προσαρμογή τους στα δεδομένα του Πίνακα 1

Μοντέλο	χ^2 στατιστική συνάρτηση ελέγχου
Negative Binomial-Pareto	$\chi^2 = 6.74, df = 1, p - value = 0.009427$
Negative Binomial	$\chi^2 = 9.15, df = 2, p - value = 0.010306$
Poisson Inverse Gaussian	$\chi^2 = 6.25, df = 2, p - value = 0.043937$

Υπάρχει ένα χαρακτηριστικό του συνόλου των δεδομένων του Πίνακα 8 που μπορεί να εξηγήσει την προφανή έλλειψη προσαρμογής. Σημειώνουμε ότι οι συχνότητες $n_3 = 43, n_4 = 9$, και $\sum_{k=5}^{\infty} n_k = 0$ του Πίνακα 8 παρουσιάζουν ένα ασυνήθιστο πρότυπο. Το πρότυπο αυτό υπονοεί ότι διάφορες διαχειριστικές ή άλλες ενέργειες που έχουν ληφθεί από τον ασφαλιστή, περιλαμβανομένων ειδοποιήσεων ή αναστολής καλύψεων, μπορεί να έχουν αποτρέψει τον αριθμό απαιτήσεων να υπερβεί την τιμή 2 και κατ' αυτό τον τρόπο να οδηγηθεί σε ένα μη πραγματικό στατιστικό πρότυπο. Η απότομη πτώση της συχνότητας για τιμές απαιτήσεων μεγαλύτερες ή ίσες του 5 προκαλεί έκπληξη. Για την συγκεκριμένη περίπτωση δεν μπορούμε να προβούμε σε αιτιολόγηση της εξέλιξης του προτύπου.

3.5.4 ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ POISSON-INVERSE GAUSSIAN ΚΑΤΑΝΟΜΗΣ

Ένα εναλλακτικό μοντέλο από αυτό της αρνητικής διωνυμικής κατανομής είναι αυτό της Poisson-Inverse Gaussian κατανομής το οποίο έχει ήδη παρουσιαστεί στην Παράγραφο 2.3 του δεύτερου κεφαλαίου. Πριν παραθέσουμε όμως την εφαρμογή της συγκεκριμένης κατανομής στα δεδομένα του Πίνακα 1 κρίνεται αναγκαίο να μελετηθεί το μοντέλο αυτό όπως καταγράφηκε από τον Lemaire. Έτσι, σύμφωνα με τους Holla (1967), Sichel (1971) και Lemaire (1995), υιοθετώντας το μοντέλο

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

$$A \sim IGau(\mu, \beta).$$

για την Poisson Inverse Gaussian κατανομή, έχουμε ότι η πιθανογεννήτρια της δίνεται από τον τύπο (δείτε Παράγραφο 2.4)

$$P_X(z) = M_A(z-1) = \exp\left(\frac{\mu}{\beta}\left(1 - \sqrt{1 - 2\beta(z-1)}\right)\right).$$

Οι πιθανότητες p_k υπολογίζονται από τους παρακάτω αναδρομικούς τύπους

$$p_0 = \exp\left(\frac{\mu}{\beta}\left[1 - (1 + 2\beta)^{\frac{1}{2}}\right]\right),$$

$$p_1 = \mu p_0 (1 + 2\beta)^{-\frac{1}{2}},$$

και για $k = 2, 3, 4, \dots$

$$(1 + 2\beta)k(k-1)p_k = \beta(k-1)(2k-3)p_{k-1} + \mu^2 p_{k-2}.$$

Για τη μέση τιμή και τη διακύμανση της Poisson-Inverse Gaussian κατανομής έχουμε ότι $E[X] = \mu$ και $Var(X) = \sigma^2 = \mu(1 + \beta)$.

• Εκτιμητές Παραμέτρων

Υποθέτουμε ότι τα δεδομένα μας αποτελούνται από n ανεξάρτητες και ισόνομες παρατηρήσεις οι οποίες μπορούν να ομαδοποιηθούν έτσι ώστε n_k να είναι ο αριθμός των ασφαλιστηρίων συμβολαίων με k απαιτήσεις ($k = 0, 1, 2, \dots$). Συνεπώς, θα έχουμε ότι

$$\sum_{k=0}^{\infty} n_k = n, \quad \bar{X} = \frac{\sum_{k=0}^{\infty} kn_k}{n}, \quad S^2 = n^{-1} \sum_{k=0}^{\infty} k^2 n_k - \bar{X}^2$$

Επομένως ένας εκτιμητής για το μ είναι ο $\hat{\mu} = \bar{X}$ ο οποίος είναι και εκτιμητής μεγίστης πιθανοφάνειας. Μεγαλύτερο όμως ενδιαφέρον παρουσιάζει ο υπολογισμός του εκτιμητή της παραμέτρου β . Σύμφωνα με τη μέθοδο των ροπών εκτιμητής για το β είναι 0

$$\hat{\beta}_0 = S^2(\bar{X})^{-1} - 1 \quad (19)$$

ο οποίος είναι ένας απλός αλλά όχι ένας αποτελεσματικός εκτιμητής (Sichel (1982)). Στην περίπτωση που το ποσοστό των μηδενικών $n^{-1}n_0$ είναι μεγάλο, ένας πιο αποτελεσματικός αλλά εξίσου απλός εκτιμητής προκύπτει εξισώνοντας τη μέση τιμή και το ποσοστό των μηδενικών στην κατανομή. Έτσι, θα έχουμε

$$\hat{\beta}_1 = \frac{2\bar{X}\{\bar{X} + \log(n^{-1}n_0)\}}{\{\log(n^{-1}n_0)\}^2}. \quad (20)$$

Σύμφωνα με τον Sichel ο εκτιμητής $\hat{\beta}_1$ είναι σε αρκετές περιπτώσεις καλύτερος από τον $\hat{\beta}_0$ όμως αναμφίβολα αυτός που προκύπτει από τη μέθοδο μεγίστης πιθανοφάνειας είναι σε κάθε περίπτωση ανώτερος και από τους δύο. Η μέθοδος αυτή για την Poisson Inverse Gaussian κατανομή ταυτίζεται σε πολλά σημεία με αυτή για την Αρνητική διωνυμική κατανομή. Πιο συγκεκριμένα στην αρνητική διωνυμική κατανομή, ο εκτιμητής μεγίστης πιθανοφάνειας για το μέσο είναι ο δειγματικός μέσος και το ίδιο ισχύει για την Poisson-Inverse Gaussian κατανομή. Για τον εκτιμητή μεγίστης πιθανοφάνειας του β ο Willmot (1987) κάνει χρήση της μεθόδου Newton-Raphson όπως φαίνεται παρακάτω:

Η log-πιθανοφάνεια για το δείγμα είναι ίση με

$$l(\mu, \beta) = \sum_{k=0}^{\infty} n_k \log p_k \quad (21)$$

Συγχρόνως, είναι γνωστό πως η πιθανογεννήτρια συνάρτηση (probability generating function) δίνεται από τη σχέση

$$P(z) = \sum_{n=0}^{\infty} p_n z^n = e^{\mu\beta^{-1}\{1 - [1 - 2\beta(z-1)]^2\}^{\frac{1}{2}}}$$

Οπότε

$$P'(z) = \mu\{1 - 2\beta(z-1)\}^{-\frac{1}{2}}P(z)$$

και

$$\frac{\partial}{\partial \mu} P(z) = \frac{P(z)}{\beta} - \left(\frac{1 + 2\beta}{\mu\beta}\right)P'(z) + \frac{2}{\mu}zP'(z).$$

Οι συντελεστές του z^k είναι ίσοι και στα τα δύο μέλη της παραπάνω εξίσωσης και έτσι διαιρώντας με p_k έχουμε

$$\frac{\partial}{\partial \mu} \log p_k = \frac{1}{\beta} + \frac{2k}{\mu} - \frac{1+2\beta}{\mu\beta} t_k \quad (22)$$

όπου

$$t_k = \frac{(k+1)p_{k+1}}{p_k}.$$

Ομοίως,

$$\frac{\partial}{\partial \beta} P(z) = -\frac{\mu}{\beta} \frac{\partial}{\partial \mu} P(z) + \frac{z-1}{\beta} P'(z)$$

από όπου προκύπτει ότι

$$\frac{\partial}{\partial \beta} \log p_k = -\frac{\mu}{\beta} \left(\frac{\partial}{\partial \mu} \log p_k \right) + \frac{k-t_k}{\beta}. \quad (23)$$

Από τις σχέσεις (21) και (23) λοιπόν παρατηρούμε ότι

$$\frac{\partial}{\partial \beta} l(\mu, \beta) = -\frac{\mu}{\beta} \left(\frac{\partial}{\partial \mu} l(\mu, \beta) \right) + \frac{n}{\beta} \bar{X} - \frac{1}{\beta} \sum_{k=0}^{\infty} n_k t_k$$

Όμως επειδή οι μερικές παράγωγοι της log-πιθανοφάνειας ως προς μ και β είναι ίσοι με το μηδέν στους Ε.Μ.Π. καταλήγουμε στην επόμενη σχέση

$$\sum_{k=0}^{\infty} n_k \hat{t}_k = n\bar{X} \quad (24)$$

με \hat{t}_k τον εκτιμητή μεγίστης πιθανοφάνειας για το t_k . Επιπλέον, από τις σχέσεις (21) και (22) οδηγούμαστε στην ακόλουθη σχέση:

$$\frac{\partial}{\partial \mu} l(\mu, \beta) = \frac{n}{\beta} + \frac{2n}{\mu} \bar{X} - \left(\frac{1+2\beta}{\mu\beta} \right) \sum_{k=0}^{\infty} n_k t_k \quad (25)$$

Με τη βοήθεια της σχέσης (24) και εξισώνοντας την (25) με το μηδέν αποδεικνύεται ο ισχυρισμός που είχε αναφερθεί παραπάνω, ότι ο Ε.Μ.Π. για το μ είναι ο $\hat{\mu} = \bar{X}$.

Προκειμένου να βρει τον εκτιμητή για το β , ο Willmot (1987) όρισε την ακολουθία $\{p_k(\beta); k = 0, 1, 2, \dots\}$ των πιθανοτήτων της Poisson-Inverse Gaussian κατανομής όπου το μ έχει αντικατασταθεί με $\mu = \hat{\mu} = \bar{X}$. Επίσης θεώρησε ότι

$$t_k(\beta) = \frac{(k+1)p_{k+1}(\beta)}{p_k(\beta)}$$

Από τη σχέση (24) καταλήγουμε ότι το $\hat{\beta}$ αποτελεί τη λύση της παρακάτω εξίσωσης

$$c(\beta) = \sum_{k=0}^{\infty} n_k t_k(\beta) - n\bar{X} = 0 \quad (26)$$

Κάνοντας χρήση της σχέσης (23) αποδεικνύεται με ευθύ τρόπο ότι η παράγωγος της (26) είναι η εξής:

$$c'(\beta) = \beta^{-2}(1+\beta) \sum_{k=0}^{\infty} n_k t_k(\beta) [t_{k+1}(\beta) - t_k(\beta)] - \beta^{-1} \sum_{k=0}^{\infty} n_k t_k(\beta)$$

Οπότε ένας νέος εκτιμητής που προκύπτει για το β ύστερα από αυτούς τους υπολογισμούς είναι

$$\bar{\beta} - c(\bar{\beta})[c'(\bar{\beta})]^{-1}$$

όπου $\bar{\beta}$ είναι η προηγούμενη τιμή του $\hat{\beta}$. Ωστόσο ένας πρώτος εκτιμητής για το β ώστε να ξεκινήσει η στην αριθμητικά διαδικασία υπολογισμού του εκτιμητή προέρχεται από τις σχέσεις (19) και (20). Για παράδειγμα, αν το ποσοστό $n^{-1}n_0$ είναι μικρό κάποιος μπορεί να επιλέξει k τέτοιο ώστε το ποσοστό $n^{-1}n_k$ να είναι αρκετά μεγάλο, και μετά με τη βοήθεια της σχέσης που αφορά τον αναδρομικό υπολογισμό των πιθανοτήτων της Poisson Inverse Gaussian (Willmot 1986)

$$p_n = \frac{2\beta}{1+2\beta} \left(1 - \frac{3}{2n}\right) p_{n-1} + \frac{\mu^2}{n(n-1)(1+2\beta)} p_{n-2}, n = 2,3,4 \dots$$

αντικαθιστώντας το μ με το \bar{X} και το p_k με $n^{-1}F_k$ προκύπτει ο ακόλουθος εκτιμητής.

$$\hat{\beta}_2 = \frac{k(k-1)n_k - (\bar{X})^2 n_{k-2}}{(k-1)(2k-3)n_{k-1} - 2k(k-1)n_k}$$

- **Εφαρμογή στα δεδομένα του Πίνακα 1**

Έχοντας υπολογίσει τους εκτιμητές των παραμέτρων για την κατανομή Poisson Inverse Gaussian, είμαστε πλέον σε θέση να εξετάσουμε την προσαρμογή της στα δεδομένα του Πίνακα 1.

Όπως είδαμε παραπάνω οι εκτιμητές που προκύπτουν από τη μέθοδο των ροπών είναι

$$\hat{\mu} = \bar{X}$$

και

$$\hat{\beta} = \left(\frac{S^2}{\bar{X}} \right) - 1.$$

Στο συγκεκριμένο παράδειγμα που θα μελετήσουμε $\hat{\mu} = 0.101081$ και $\hat{\beta} = 0.062979$, με βάση τους οποίους παίρνουμε τα παρακάτω αποτελέσματα από τον έλεγχο καλής προσαρμογής.

Πίνακας 8: Έλεγχος καλής προσαρμογής της Poisson-Inverse Gaussian κατανομής στα δεδομένα του Πίνακα 1 με εκτιμητές από τη μέθοδο των ροπών

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	$9.06561407e - 01$	96978.5	96978
$X = 1$	$8.6379868e - 02$	9240.4	9240
$X = 2$	$6.52121e - 03$	697.5999	704
$X = 3$	$4.94512e - 04$	52.89993	43
$X = 4$	$3.9261e - 05$	4.199906	9
$X > 4$	$3.739e - 06$	0.3999758	0

Με βάση τον έλεγχο του χ^2 τεστ που ακολουθεί βλέπουμε πως έχουμε καλύτερα αποτελέσματα σε σχέση με την Αρνητική διωνυμική.

$\chi^2 = 7.8181$	$df = 5$	$p - value = 0.1666$
-------------------	----------	----------------------

Οι εκτιμητές μεγίστης πιθανοφάνειας για τις παραμέτρους μ και β είναι $\hat{\mu} = 0.101081$ και $\hat{\beta} = 0.062981$.

Εφαρμόζοντας λοιπόν το χ^2 τεστ για τους συγκεκριμένους εκτιμητές λαμβάνουμε τα ακόλουθα αποτελέσματα:

Πίνακας 9: Έλεγχος καλής προσαρμογής της Poisson Inverse Gaussian κατανομής στα δεδομένα του Πίνακα 1 με τη μέθοδο μέγιστης πιθανοφάνειας.

	<i>prob1</i>	<i>Exp1</i>	<i>obs1</i>
$X = 0$	$9.06573559e - 01$	96979.8	96978
$X = 1$	$8.6359302e - 02$	9238.2	9240
$X = 2$	$6.528689e - 03$	698.4	704
$X = 3$	$4.95447e - 04$	52.999	43
$X = 4$	$3.9261e - 05$	4.199906	9
$X > 4$	$3.739e - 06$	0.3999758	0

Το αποτέλεσμα το χ^2 τεστ είναι τα ακόλουθα

$\chi^2 = 7.7975$	$df = 5$	$p - value = 0.1678$
-------------------	----------	----------------------

3.5.5 ΕΦΑΡΜΟΓΗ ΤΗΣ POISSON-INVERSE GAUSSIAN ΚΑΤΑΝΟΜΗΣ ΣΤΑ ΔΕΔΟΜΕΝΑ ΤΩΝ GOSSIAUX & LEMAIRE (1981)

Έχουμε αναφέρει ότι οι μειγμένες κατανομές Poisson είναι ιδιαίτερα διαδεδομένες σε περιπτώσεις μοντελοποίησης δεδομένων από ετερογενή χαρτοφυλάκια και πιο συγκεκριμένα αποτελούν ένα καλό εργαλείο μέτρησης της ετερογένειας τους. Οι *Gossiaux & Lemaire* (1981) ανέλυσαν έξι ομάδες δεδομένων αναφορικά με τον αριθμό των απαιτήσεων που εγείρουν οι ασφαλισμένοι του κλάδου αυτοκινήτου ανά ασφαλιστικό συμβόλαιο στη Γερμανία το 1960. Τα δεδομένα παρουσιάζουν αξιοσημείωτη λοξότητα-ασυμμετρία, με αποτέλεσμα όπως ήδη έχουμε δει η Poisson να καθίσταται ακατάλληλο και ανεπαρκές μοντέλο για την περιγραφή τους. Με στόχο λοιπόν την ανεύρεση ενός καταλληλότερου μοντέλου χρησιμοποιήθηκε η μέθοδος μέγιστης πιθανοφάνειας η οποία εφαρμόστηκε τόσο για την αρνητική διωνυμική κατανομή όσο και για την Poisson-Inverse Gaussian κατανομή. Ο *Willmot* (1987) επανεξέτασε τα δεδομένα αυτά. Η αρνητική διωνυμική κατανομή, ενώ κρίθηκε επαρκέστερη συγκριτικά με την Poisson κατανομή, φάνηκε να υστερεί σε σχέση με την P-IG κατανομή η οποία είτε με βάση τον έλεγχο καλής προσαρμογής του Pearson, είτε σύμφωνα με τη μέγιστη τιμή της log-πιθανοφάνειας έδωσε καλύτερα αποτελέσματα και για τα έξι σετ δεδομένων.

Επιπρόσθετα, η προσαρμογή του εκτιμητή, είναι σημαντικά ανώτερη σε όλες τις ομάδες δεδομένων εκτός από εκείνη της ομάδας 4. Η προσαρμογή της κατανομής P-IG στην ομάδα δεδομένων 5 διαφαίνεται πως είναι οριακά καλή. Όμως, αν πραγματοποιηθεί ομαδοποίηση των κλάσεων συχνοτήτων 3 και 4 για τον υπολογισμό της στατιστικής του Pearson, όπως έπραξαν οι *Gossiaux & Lemaire* (1981), τότε υπό αυτή τη συνθήκη η προσαρμογή κρίνεται ως εξαιρετικά καλή. Οι ομαδοποιήσεις κλάσεων έγιναν με σκοπό οι αναμενόμενες συχνότητες να γίνουν επαρκώς μεγάλες έτσι ώστε η προσέγγιση του χ^2 ελέγχου να είναι αξιόπιστη.

Ομάδα δεδομένων 1:

Αριθμός Απαιτήσεων k	Αριθμός Ασφαλιστηρίων συμβολαίων n_k	NB	P-IG
0	103704	103723.61	103710.03
1	14075	13989.95	13989.95
2	1766	1857.08	1784.91
3	255	245.19	254.49
4	45	32.29	40.42
5	6	4.24	6.94
6	2	0.56	1.26
Σύνολο	119853	119852.92	119852.70
Maximum log- πιθανοφάνεια		-54615.315	-54609.758
Τιμή της σ.σ.ε. χ^2		12.37	0.78
Βαθμοί Ελευθερίας		3	3
p-value		0.006	0.855

Ομάδα δεδομένων 2:

Αριθμός Απαιτήσεων k	Αριθμός Ασφαλιστηρίων συμβολαίων n_k	NB	P-IG
0	370412	370438.94	370435.18
1	46545	46451.28	46476.38
2	3935	4030.50	3995.76
3	317	297.82	307.67
4	28	20.09	23.12
5	3	1.28	1.75
Σύνολο	421240	421239.91	421239.86
Maximum log- πιθανοφάνεια		-171136.966	-171134.472
Τιμή της σ.σ.ε. χ^2		7.94	2.74
Βαθμοί Ελευθερίας		2	2
p-value		0.019	0.254

Ομάδα δεδομένων 3:

Αριθμός Απαιτήσεων k	Αριθμός Ασφαλιστηρίων συμβολαίων n_k	NB	P-IG
0	7840	7847.01	7844.01
1	1317	1288.36	1306.12
2	239	256.53	238.23
3	42	54.07	53.27
4	14	11.71	13.75
5	4	2.58	3.89
6	4	0.57	1.17
7	1	0.13	0.37
Σύνολο	9461	9460.96	9460.81
Maximum log- πιθανοφάνεια		-5348.040	-5343.511
Τιμή της σ.σ.ε. χ^2		14.73	4.54
Βαθμοί Ελευθερίας		3	3
p-value		0.002	0.209

Ομάδα δεδομένων 4:

Αριθμός Απαιτήσεων k	Αριθμός Ασφαλιστηρίων συμβολαίων n_k	NB	P-IG
0	3719	3719.22	3718.58
1	232	229.90	234.54
2	38	39.91	34.86
3	7	8.42	8.32
4	3	1.93	2.45
5	1	0.46	0.80
Σύνολο	4000	3999.84	3999.55
Maximum log- πιθανοφάνεια		-1183.550	-1183.524
Τιμή της σ.σ.ε. χ^2		1.17	0.54
Βαθμοί Ελευθερίας		2	2
p-value		0.557	0.762

Ομάδα δεδομένων 5 (Πίνακας 1):

Αριθμός Απαιτήσεων K	Αριθμός Ασφαλιστηρίων συμβολαίων n_k	NB	P-IG
0	96978	96980.82	96978.53
1	9240	9230.90	9240.40
2	704	708.62	697.63
3	43	50.05	52.85
4	9	3.38	4.21
Σύνολο	106974	106973.77	106973.62
Maximum log- πιθανοφάνεια		-36104.099	-36103.574
Τιμή της σ.σ.ε. X^2		9.03	6.12
Βαθμοί Ελευθερίας		2	2
p-value		0.011	0.047

Ομάδα δεδομένων 6:

Αριθμός Απαιτήσεων K	Αριθμός Ασφαλιστηρίων συμβολαίων n_k	NB	P-IG
0	20592	20596.76	20595.74
1	2651	2631.03	2638.81
2	297	318.37	308.08
3	41	37.81	39.68
4	7	4.45	5.65
5	0	0.52	0.87
6	1	0.06	0.14
Σύνολο	23589	23589	23588.97
Maximum log- πιθανοφάνεια		-10223.420	-10221.868
Τιμή της σ.σ.ε. χ^2		3.60	0.76
Βαθμοί Ελευθερίας		2	2
p-value		0.165	0.684

Για την P-IG, μπορούμε να πούμε πως αποτελεί μία κατανομή η οποία είναι εξαιρετικά ανταγωνιστική της αρνητικής διωνυμικής κατανομής, όσον αφορά τόσο τις μαθηματικές και τις στατιστικές ιδιότητες που τη διέπουν, όσο επίσης και της εν δυνάμει ικανότητάς της να μπορεί να περιγράψει μια μεγάλη ποικιλία φυσικών φαινομένων.. Ένας επιπλέον λόγος που συνηγορεί σε αυτό, είναι οτι έχει αρκετά καλύτερη προσαρμογή σε δεδομένα απαιτήσεων αυτοκινήτων έναντι των περισσότερων ανταγωνιστών της.

3.5.6 ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ STRICT ARCSINE ΚΑΤΑΝΟΜΗΣ

Όταν τα δεδομένα παρουσιάζουν μεγάλο συντελεστή λοξότητας, και συνεπώς η κατανομή Poisson κρίνεται ακατάλληλη για την μοντελοποίηση των δεδομένων, τότε κατανομές που παρουσιάζουν υπερδιασπορά ή υποδιασπορά (*underdispersion*) σε σχέση με την κατανομή Poisson, μπορούν να εξετασθούν για τη δυνατότητα προσαρμογής τους στα δεδομένα αυτά. Τα προβλήματα υπερδιασποράς εντοπίζονται όταν η δειγματική διασπορά s^2 είναι μεγαλύτερη του δειγματικού μέσου \bar{x} .

Οι *Gossiaux & Lemaire* (1981), ανέλυσαν έξι ομάδες δεδομένων, μία από τις οποίες παρουσιάζεται στις πρώτες δύο στήλες του Πίνακα 11, στον οποίο δίνεται ο αριθμός των απαιτήσεων ανά ασφαλιστήριο συμβόλαιο του κλάδου αυτοκινήτου στη Γερμανία κατά τη διάρκεια του έτους 1960. Στα συγκεκριμένα δεδομένα προσαρμόζεται η κατανομή Poisson $P(a)$, η αρνητική διωνυμική κατανομή $NB(p, a)$, η strict arcsine κατανομή $SA(p, a)$, η Poisson-Inverse Gaussian κατανομή $PIG(p, a)$ και η γενικευμένη Poisson κατανομή $GP(p, a)$.

Πίνακας 11: Προσαρμογή κατανομών σε δεδομένα απαιτήσεων αυτοκινητοβιομηχανίας στη Γερμανία το 1960 1984 με $\bar{x} = .14$ και $s^2 = .16$.

Αριθμός Απαιτήσεων $n(k)$	Αριθμός Συμβολαίων $f(k)$	$P(a)$	$NB(p, a)$	$SA(p, a)$	$PIG(p, a)$	$GP(p, a)$
0	20592	20420.60	20605.14	20685.83	20587.68	20591.34
1	2651	2945.05	2615.44	2663.08	2650.45	2640.67
2	297	212.37	322.75	171.42	305.23	313.48
3	41	10.59	39.45	55.00	38.33	37.53
4	7	0.38	5.46	9.62	6.21	5.23
5	0	0.01	0.66	3.24	0.93	0.65
6	1	0.00	0.08	0.54	0.14	0.08
Σύνολο: n	23589	23589.00	23588.98	23588.73	23588.96	23588.98
$(\hat{p}, \hat{\alpha})$		(.14)	(.08, 1.6)	(.39, .33)	(.79, .33)	(.03, 4.6)
df		2	2	2	2	2
χ^2		196.18	3.12	98.33	0.48	1.92

Πίνακας 12: Προσαρμογή κατανομών σε δεδομένα απαιτήσεων αυτοκινητοβιομηχανίας στην Κεντρική Αφρικανική Δημοκρατία το 1984 με $\bar{x} = .37$ και $s^2 = .42$.

Αριθμός Απαιτήσεων $n(k)$	Αριθμός Συμβολαίων $f(k)$	$P(a)$	$NB(p, a)$	$SA(p, a)$	$PIG(p, a)$	$GP(p, a)$
0	6984	6874.81	7025.87	7012.25	6958.87	6983.55
1	2452	2576.05	2337.24	2461.02	2443.26	2403.39
2	433	482.42	519.56	431.77	502.87	510.22
3	100	60.22	96.18	76.24	80.72	86.43
4	26	6.07	18.69	16.16	12.48	14.70
≥ 5	5	0.43	2.46	2.56	1.80	1.71
Σύνολο: n	10000	10000	10000	10000	10000	10000
$(\hat{p}, \hat{\alpha})$		(0.37)	(0.05, 6.9)	(0.32, 1.1)	(0.77, .91)	(0.03, 12)
df		3	2	2	2	2
χ^2		131.39	25.04	15.61	34.01	27.77

Ο Πίνακας 12 παρουσιάζει παρόμοια δεδομένα από ένα δείγμα που προέρχεται από τη Δημοκρατία της Κεντρικής Αφρικής κατά την περίοδο του 1984. Τα δεδομένα των Πινάκων 11 και 12 παρουσιάζουν υπερδιασπορά, και για το λόγο αυτό θα πρέπει να γίνει στα δεδομένα αυτά εφαρμογή μίας κατανομής που και εκείνη θα έχει το χαρακτηριστικό της υπερδιασποράς σε αντίθεση με την Poisson, $P(a)$. Τέτοιες κατανομές όπως έχουμε δει είναι οι ακόλουθες:

- Αρνητική διωνυμική, $NB(p, a)$
- Strict Arcsine, $SA(p, a)$
- Poisson-Inverse Gaussian, $PIG(p, a)$
- Γενικευμένη Poisson, $GP(p, a)$

Εδώ, η μέθοδος εκτίμησης της πρώτης ροπής συνδυαζόμενη (εξαιρουμένης της κατανομής Poisson) με τη μέθοδο εκτίμησης παραμέτρων της αναλογίας των 1 και 0 (εξαιρουμένης της γενικευμένης κατανομής Poisson), χρησιμοποιούνται για κάθε μία από αυτές τις κατανομές στον αντίστοιχα τους έλεγχο τους καλής προσαρμογής, χ^2 . Πιο συγκεκριμένα, οι εκτιμήσεις των παραμέτρων δίνονται αντίστοιχα από τους παρακάτω εκτιμητές:

- $\hat{a}_p = \bar{x}$

- $(\hat{p}, \hat{a})_{NB} = (1 - [f(1)/f(0)]/\bar{x} ; [f(1)/f(0)]/\hat{p})$
- $(\hat{p}, \hat{a})_{SA} = (\{1 - ([f(1)/f(0)]/\bar{x})^2\}^{\frac{1}{2}} ; [f(1)/f(0)]/\hat{p})$
- $(\hat{p}, \hat{a})_{PIG} = (1 - \{[f(1)/f(0)]/[2\bar{x}]\}^2 ; [2f(1)/f(0)]/\hat{p})$
- $(\hat{p}, \hat{a})_{GP} = (1 + \ln [f(0)/n]/\bar{x} ; -\ln [f(0)/n]/\hat{p})$

Οι *Kokonendji & Khoudar* (2004), σημείωσαν ότι παρά την ανεπαρκή εκτίμηση των παραμέτρων, όπως παρουσιάζεται στον Πίνακα 11, λαμβάνουμε τα ίδια αποτελέσματα που οι *Gossiaux & Lemaire* (1981) και *Wilmot* (1987) παρουσίασαν έχοντας χρησιμοποιήσει διαφορετικές μεθόδους εκτίμησης για την αρνητική διωνυμική κατανομή και την Poisson-Inverse Gaussian κατανομή. Όμως, όπως μπορεί να παρατηρηθεί, η προσαρμογή της κατανομής Strict Arcsine στα δεδομένα του Πίνακα 11 φαίνεται να είναι οριακή ανάμεσα στις θεωρηθείσες κατανομές που παρουσιάζουν υπερδιασπορά. Παρατηρούμε ότι η αναμενόμενη συχνότητα για τις 2 απαιτήσεις είναι πιο υποεκτιμημένη από την Strict Arcsine κατανομή σε σχέση με τις άλλες κατανομές.

Παρόλα αυτά, από τον Πίνακα 12, παρατηρούμε ότι με το να προσαρμόσουμε την Strict Arcsine κατανομή στα δεδομένα, προκύπτει για παράδειγμα σημαντική βελτίωση σε σχέση με την προσαρμογή στα δεδομένα των άλλων κατανομών, και αυτό κρίνοντας από τη χαμηλή τιμή της σ.σ.ε. χ^2 . Όπως παρατηρείται από τους *Gossiaux & Lemaire* (1981), δεν υπάρχει ένας νόμος πιθανότητας που να παρουσιάζει την καλύτερη προσαρμογή σε όλες τις περιπτώσεις.

Συμπερασματικά, η Strict Arcsine είναι μία κατανομή η οποία είναι αρκούτως ανταγωνιστική με την αρνητική διωνυμική και την Poisson-Inverse Gaussian κατανομή. Επιπλέον η συνάρτηση μάζας πιθανότητάς της είναι αρκετά εύκολη στον χειρισμό της αλλά και στην εφαρμογή της σε αντίθεση με τις αντίστοιχες συνεχείς μικτές κατανομές, όπως είναι για παράδειγμα η μικτή Poisson ή η Poisson-Inverse Gaussian. Όπως μπορεί να παρατηρηθεί από τα δεδομένα του Πίνακα 12, συμβαίνει μία αρκετά σημαντική πτώση στη μάζα πιθανότητας στην τιμή 2 ενώ παρουσιάζεται μία σχετική ομοιομορφία στην τιμή 0. Το γεγονός αυτό μπορεί να αποτελέσει την αιτία είτε για καλή είτε για κακή προσαρμογή των δεδομένων αναλόγως την περίπτωση.

3.5.7 ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ ΒΕΤΑ ΝΕΓΑΤΙΒΕ ΒΙΝΟΜΙΑΛ ΚΑΤΑΝΟΜΗΣ.

Στην παράγραφο αυτή θα γίνει σύγκριση της Βήτα αρνητικής διωνυμικής κατανομής με την αρνητική διωνυμική κατανομή και από την σύγκριση αυτή θα γίνει αντιληπτό ότι η πρώτη είναι πιο ευέλικτη ως κατανομή.

Τα δεδομένα του Πίνακα 13 (οι δύο πρώτες στήλες) προέρχονται από δεδομένα απαιτήσεων οχημάτων ασφαλιστικής εταιρίας στην Κίνα τα οποία συναντάμε σε δημοσίευση των *Wang & Lei* (2000). Ο υπολογισμός των εκτιμητών μεγίστης πιθανοφάνειας των παραμέτρων της Βήτα Αρνητικής Διωνυμικής κατανομής μπορεί να γίνει με τον τρόπο που προτείνει ο *Wang* (2011) (ρουτίνα *fminsearch* του Matlab).

Πανεπιστήμιο Πειραιώς

Πίνακας 13: Συγκριτικός πίνακας προσαρμογής για τις κατανομές BNB, NB και Poisson για τον αριθμό απαιτήσεων που εγείρουν οχήματα ασφαλιστικής εταιρίας στην Κίνα.

Αριθμός Απαιτήσεων $n(k)$	Αριθμός Συμβολαίων $f(k)$	$BNB(r, a, \beta)$	$NB(r, p)$	$P(\lambda)$
0	27141	27147.40	27165.85	25528.60
1	5789	5746.84	5664.12	8107.98
2	1433	1505.27	1563.35	1287.56
3	457	445.75	466.68	136.31
4	155	144.67	144.55	10.82
5	56	50.65	45.76	0.68
6 ^a	27	18.98	14.69	0.04
7 ^a	2	7.44	4.76	0.01
8 ^a	1	3.08	1.56	0
9 ^a	1	1.33	0.51	0
$\geq 10^a$	0	0.59	0.17	0
Σύνολο	35072	35072	35072	35072
df		3	4	5
χ^2		4.4790	19.2848	27118.1
p-value		0.2142	0.0007	0
-log-likelihood		25417.91	25422.52	26712.72
AIC^1		50841.82	50849.04	53427.44
MLE's		$\hat{r} = 6.3664$ $\hat{a} = 17.3484$ $\hat{\beta} = 0.8157$	$\hat{r} = 0.6070$ $\hat{p} = 0.6565$	$\hat{\lambda} = 0.3176$

Οι αναμενόμενες συχνότητες έχουν ομαδοποιηθεί για τον υπολογισμό του χ^2 . Στον Πίνακα 13 λοιπόν έχουν καταγραφεί 35072 ασφαλιστήρια συμβόλαια αυτοκινήτων. Ο συντελεστής διασποράς έχει υπολογιστεί 1.547 δηλαδή μεγαλύτερος από 1 γεγονός που μαρτυρά υπερδιασπορά.

Συνεχίζοντας τώρα με το παράδειγμά μας, βλέπουμε στον πίνακα τόσο την p-value όσο και την τιμή χ^2 του τεστ με βάση τα οποία η κατανομή Negative Binomial Pareto

βλέπουμε να προσαρμόζεται καλύτερα από την Negative Binomial και εκείνη με τη σειρά της καλύτερα από την Poisson.

Πανεπιστήμιο Πειραιώς

4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Εν κατακλείδι, πολλά είναι τα μοντέλα εκείνα που μπορούν να αποτελέσουν χρήσιμα εργαλεία αναφορικά με την καλύτερη μοντελοποίηση της ανεξαρτησία που μπορεί να υπάρξει μεταξύ δύο συμβολαίων ενός χαρτοφυλακίου. Οι κατανομές που αναλύθηκαν στα παραπάνω κεφάλαια ως επί το πλείστον αποτελούν γενικευμένες μορφές της Poisson κατανομής και της αρνητικής διωνυμικής. Επιπλέον, κανένα από τα μοντέλα δεν φάνηκε να υπερέχει σε μεγάλο βαθμό έναντι των άλλων

Αναλυτικότερα, με εξαίρεση την κατανομή Poisson τα υπόλοιπα μοντέλα φάνηκαν να έχουν καλή προσαρμογή σε δεδομένα που παρουσιάζουν υπερδιασπορά. Στη βιβλιογραφία, οι κατανομές Negative Binomial, Poisson Inverse Gaussian και Generalized Poisson διαφέρουν. Η δικής μας μελέτη μας έδειξε ότι οι διαφορές των PIG και GP δεν είναι τόσο μεγάλες αφού συμπεριφέρονται κατά τον ίδιο περίπου τρόπο. Ωστόσο, το μοντέλο της PIG δείχνει να έχει καλύτερη προσαρμογή σε δεδομένα που παρουσιάζουν μεγαλύτερη λοξότητα και έχουν βαριά ουρά. Αυτό λοιπόν καθιστά την συγκεκριμένη κατανομή κατάλληλη για τέτοιου είδους δεδομένα.

Κατά τον έλεγχο συμβατότητας των δεδομένων του Lemaire όπως αυτά παρουσιάζονται στον Πίνακα 1, με την κατανομή Poisson παρατηρήσαμε ότι ο έλεγχος καλής προσαρμογής δεν ήταν αξιόπιστος. Αυτό γιατί δεν υπάρχει επαρκής μάζα πιθανότητας στη δεξιά ουρά της κατανομής Poisson. Το μοντέλο της Poisson δεν μπορεί να εφαρμοστεί στο συγκεκριμένο χαρτοφυλάκιο. Επομένως η υπόθεση της ομοιογένειας δεν είναι συμβατή με τη συγκεκριμένη στατιστική ανάλυση. Προκλήθηκε η ανάγκη για μια κατανομή της οποίας η διασπορά θα ξεπερνά το μέσο της θα παρουσιάζει δηλαδή υπερδιασπορά. Έτσι τέθηκε ως αναγκαίος ο έλεγχος της Αρνητικής Διωνυμικής κατανομής η οποία έδωσε καλύτερα συμπεράσματα αλλά όχι σε ικανοποιητικό βαθμό.

Στα ίδια δεδομένα ελέγχθηκε η προσαρμογή των κατανομών Negative Binomial Pareto, Negative Binomial και Poisson-Inverse Gaussian. Τα αποτελέσματα των εκτιμήσεων, μας έδωσαν ορισμένες πολύ χρήσιμες ερμηνείες σχετικά με την διαμόρφωση της πορείας της κατανομής. Ωστόσο σε συγκριτικό πίνακα (Πίνακας 9) που έγινε μεταξύ των Poisson, Negative Binomial Pareto και Poisson-Inverse Gaussian δόθηκε η τιμή του στατιστικού χ^2 τεστ καλής προσαρμογής για τα τρία

προαναφερθέντα μοντέλα, έχοντας υποβληθεί σε ομαδοποίηση των κλάσεων, έτσι ώστε όλες οι αναμενόμενες συχνότητες np_k να είναι ίσες με 2 ή και περισσότερο και τουλάχιστον το 80% να είναι ίσες με 5 ή και περισσότερο. Σε επίπεδο σημαντικότητας 5% όλα τα μοντέλα δεν παρουσιάζουν καλή προσαρμογή στα δεδομένα. Το μοντέλο P-IG παρουσιάζει τη μικρότερη τιμή χ^2 . Η επιλογή κάποιου εκ των τριών μοντέλων με βάση την προσαρμοστικότητα τους σε σχέση με τις παρατηρηθείσες συχνότητες, δεν μπορεί να αποτελέσει κριτήριο αξιολόγησης. Για την P-IG, μπορούμε να πούμε πως αποτελεί μία κατανομή η οποία είναι εξαιρετικά ανταγωνιστική της Αρνητικής Διωνυμικής κατανομής, όσον αφορά τόσο τις μαθηματικές και τις στατιστικές ιδιότητές που τη διέπουν, όσο επίσης και την εν δυνάμει ικανότητά της να λαμβάνεται υπόψη στην περιγραφή μιας μεγάλης ποικιλίας φυσικών φαινομένων. Εν κατακλείδι, λαμβάνοντας υπόψη και την εφαρμογή της κατανομής Strict Arcsine όπως την είδαμε στην Παράγραφο 3.4 μπορούμε να πούμε ότι παρουσιάζει ιδιαίτερος ικανοποιητική προσαρμογή ενώ παράλληλα αποτελεί πιο εύχρηστο μοντέλο συγκριτικά με την Poisson Inverse Gaussian κατανομή. Οι δυο αυτές αναδείχθηκαν καταλληλότερες σε σχέση με τις προηγούμενες. Ενώ τέλος βλέπουμε τη Negative Binomial Pareto να προσαρμόζεται καλύτερα από την Negative Binomial και εκείνη με τη σειρά της καλύτερα από την Poisson σε ένα συγκεκριμένο σετ δεδομένων.

ΠΑΡΑΡΤΗΜΑ

Poisson δοκιμή 1

```
> a<-c(0:4)
> obs1 <- c(96978,9240,704,43,9)
> acc <- rep(a1, obs1)
> table(acc)
acc
  0     1     2     3     4
96978 9240  704   43    9
> m <- a1*obs1;m
[1]  0 9240 1408  129  36
> mean <- sum(m)/sum(obs1); mean
[1] 0.1010806
> probl <- c(dpois(0:3,mean), 1-ppois(3,mean))
> Expl<-106974*probl
> ans1<-cbind(probl,Expl,obs1)
> row.names(ans1)<-c("X=0","X=1","X=2","X=3","X=4")
> ans1
```

	probl	Expl	obs1
X=0	9.038601e-01	9.668954e+04	96978
X=1	9.136276e-02	9.773440e+03	9240
X=2	4.617503e-03	4.939528e+02	704
X=3	1.555800e-04	1.664302e+01	43
X=4	4.012371e-06	4.292194e-01	9

```
> chisq.test(x=obs1,p=probl)
```

Chi-squared test for given probabilities

```
data:  obs1
X-squared = 332.1806, df = 4, p-value < 2.2e-16
```

Poisson δοκιμή 2

```
> acc2 <- rep(a2, obs2)
> table(acc2)
acc2
  0    1    2    3
96978 9240  704  52
> #m <- a2*obs2;m
> #mean <- sum(m)/sum(obs2); mean
> prob2 <- c(dpois(0:2,mean), 1-ppois(2,mean))
> Exp2<-106974*prob2
> ans2<-cbind(prob2,Exp2,obs2)
> row.names(ans2)<-c("X=0","X=1","X=2","X=3")
> ans2
```

	prob2	Exp2	obs2
X=0	0.9038601459	96689.53525	96978
X=1	0.0913627588	9773.43976	9240
X=2	0.0046175029	493.95276	704
X=3	0.0001595924	17.07224	52

```
> chisq.test(x=obs2,p=prob2)
```

Chi-squared test for given probabilities

```
data: obs2
X-squared = 190.754. df = 3. p-value < 2.2e-16
```


Αρνητική Διωνυμική δοκιμή 1 (Εκτιμητές από μέθοδο ροπών)

```
> a1 <- c(0:4)
> obs1 <- c(96978,9240,704,43,9)
> acc1 <- rep(a1, obs1)
> table(acc1)
acc1
  0    1    2    3    4
96978 9240  704  43    9
> m <- a1*obs1;m
[1]  0 9240 1408  129  36
> mean <- sum(m)/sum(obs1); mean
[1] 0.1010806
> sa <- (a1-mean)^2;
> sb <- sa*obs1
> s2 <- (1/(sum(obs1)-1))*sum(sb); s2
[1] 0.1074478
> a<- mean^2/(s2-mean);a
[1] 1.604682
> τ<- mean/(s2-mean);τ
[1] 15.87526
> p <- τ/(1+τ);p
[1] 0.9407417
> prob1 <- c(dnbinom(0:3,a,p), 1-pnbinom(3,a,p))
> Exp1<-106974*prob1
> ans1<-cbind(prob1,Exp1,obs1)
> row.names(ans1)<-c("X=0","X=1","X=2","X=3","X=4")
> ans1
```

	prob1	Exp1	obs1
X=0	9.066265e-01	96985.461788	96978
X=1	8.621181e-02	9222.421903	9240
X=2	6.653357e-03	711.736253	704
X=3	4.737355e-04	50.677384	43
X=4	3.461282e-05	3.702672	9

```
> sum(prob1)
[1] 1
> chisq.test(x=obs1,p=prob1)
```

Chi-squared test for given probabilities

```
data:  obs1
X-squared = 8.86, df = 4, p-value = 0.0647
```

Warning message:

```
In chisq.test(x = obs1, p = prob1) :
  Chi-squared approximation may be incorrect
```

Αρνητική Διωνυμική δοκιμή 1 (Εκτιμητές από μέθοδο μέγιστης πιθανοφάνειας)

```
> probl <- c(dnbinom(0:3,a,p), 1-pnbinom(3,a,p))
> Expl<-106974*probl
> ans1<-cbind(probl,Expl,obs1)
> row.names(ans1)<-c("X=0","X=1","X=2","X=3","X=4")
> ans1
```

	probl	Expl	obs1
X=0	9.075745e-01	97086.874055	96978
X=1	8.542270e-02	9138.008013	9240
X=2	6.512220e-03	696.638214	704
X=3	4.576332e-04	48.954852	43
X=4	3.295067e-05	3.524865	9

```
> sum(probl)
[1] 1
> chisq.test(x=obs1,p=probl)
```

Chi-squared test for given probabilities

data: obs1

X-squared = 10.5671, df = 4, p-value = 0.03189

Warning message:

In chisq.test(x = obs1, p = probl) :

Chi-squared approximation may be incorrect

Αρνητική Διωνυμική δοκιμή 2 (Εκτιμητές από μέθοδο ροπών)

```
acc2
  0    1    2    3
96978 9240 704  52
> m <- a2*obs2;m
[1] 0 9240 1408 156
> mean <- sum(m)/sum(obs2); mean
[1] 0.1009965
> sa <- (a2-mean)^2;
> sb <- sa*obs2
> s2 <- (1/(sum(obs2)-1))*sum(sb); s2
[1] 0.1068759
> a<- mean^2/(s2-mean);a
[1] 1.734927
> τ<- mean/(s2-mean);τ
[1] 17.17809
> p <- τ/(1+τ);p
[1] 0.9449887
>
> prob2 <- c(dnbinom(0:2,a,p), 1-pnbinom(2,a,p))
> Exp2<-106974*prob2
> ans2<-cbind(prob2,Exp2,obs2)
> row.names(ans2)<-c("X=0","X=1","X=2","X=3")
> ans2
      prob2      Exp2  obs2
X=0 0.9064982843 96971.74746 96978
X=1 0.0865167015  9255.03762  9240
X=2 0.0065082982   696.21869   704
X=3 0.0004767161    50.99623    52

> sum(prob2)
[1] 1
> chisq.test(x=obs2,p=prob2)

      Chi-squared test for given probabilities

data:  obs2
X-squared = 0.1316, df = 3, p-value = 0.9878
```

Poisson Inverse Gaussian 1 (Εκτιμητές από μέθοδο ροπών)

```
> probl <- c(9.06573559e-01,8.6359302e-02, 6.528689e-03,4.95447e-04,3.9261e-05,3.739e-06)
> Expl<-106974*probl
> sum(probl)
[1] 1
> ans1<-cbind(probl,Expl,obs1)
> row.names(ans1)<-c("X=0","X=1","X=2","X=3","X=4","X=5")
> ans1
```

	probl	Expl	obs1
X=0	0.906573559	9.697980e+04	96978
X=1	0.086359302	9.238200e+03	9240
X=2	0.006528689	6.984000e+02	704
X=3	0.000495447	5.299995e+01	43
X=4	0.000039261	4.199906e+00	9
X=5	0.000003739	3.999758e-01	0

```
> chisq.test(x=obs1,p=probl)
```

Chi-squared test for given probabilities

data: obs1

X-squared = 7.8181, df = 5, p-value = 0.1666

Warning message:

In chisq.test(x = obs1, p = probl) :

Chi-squared approximation may be incorrect

Poisson Inverse Gaussian 1 (Εκτιμητές από μέθοδο μέγιστης πιθανοφάνειας)

```
> probl <- c(9.06561407e-01,8.6379868e-02,6.52121e-03,4.94512e-04,3.9261e-05,3.739e-1
> Expl<-106974*probl
> ans1<-cbind(probl,Expl,obs1)
> row.names(ans1)<-c("X=0","X=1","X=2","X=3","X=4","X=5")
> ans1
```

	probl	Expl	obs1
X=0	0.906561407	9.697850e+04	96978
X=1	0.086379868	9.240400e+03	9240
X=2	0.006521210	6.975999e+02	704
X=3	0.000494512	5.289993e+01	43
X=4	0.000039261	4.199906e+00	9
X=5	0.000003739	3.999758e-01	0

```
> sum(probl)
[1] 1
> chisq.test(x=obs1,p=probl)
```

Chi-squared test for given probabilities

data: obs1

X-squared = 7.7975, df = 5, p-value = 0.1678

Warning message:

In chisq.test(x = obs1, p = probl) :

Chi-squared approximation may be incorrect

Negative Binomial Pareto (εκτιμητές από μέθοδο μεγίστης πιθανοφάνειας)

```
[1] 1
> Exp1<-106974*prob1
> ans1<-cbind(prob1,Exp1,obs1)
> row.names(ans1)<-c("X=0","X=1","X=2","X=3","X=4","X=5")
> ans1
      prob1      Exp1  obs1
X=0 9.065754e-01 9.698000e+04 96978
X=1 8.633780e-02 9.235900e+03 9240
X=2 6.563277e-03 7.021000e+02 704
X=3 4.842298e-04 5.180000e+01 43
X=4 3.645746e-05 3.900000e+00 9
X=5 2.804419e-06 2.999999e-01 0
> sum(prob1)
[1] 1
> chisq.test(x=obs1,p=prob1)

      Chi-squared test for given probabilities

data:  obs1
X-squared = 8.4712, df = 5, p-value = 0.1321

Warning message:
In chisq.test(x = obs1, p = prob1) :
  Chi-squared approximation may be incorrect
> █
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ambagaspittya, R.S. & Balakrishnan, N. (1994). *Manipulating Lagrangian distributions and associated compound distributions with Maple. Presented at the 29th Actuarial Research Conference, Aug. 25-27, 1994, Oregon State university, Corvallis.*

Antzoulakos D. & Chadjiconstantinidis S., *On Mixed & Compound Poisson Distributions, Scandinavian Actuarial Journal, 2004:3, 161 – 188.*

Arfwedson, G. (1955). *Research in collective risk theory. Part 2. Skandinavisk Aktuarietidskrift 38, 53-100.*

Bozdogan, H., (2000). *Akaike's information criterion and recent developments in information complexity. Journal of Mathematical Psychology 44, 62–91.*

D'Agostino, R. B., & Stefens M. A. (1986), *Goodness- of- fit Techniques, New York: Marcel Dekker.*

Consul, P.C. & Jain, G.C. (1973). *A generalization of Poisson distribution. Technometrics 15, 791-799.*

Consul., P.C. & Shoukri, M.M. (1985). *The generalized Poisson distribution when the sample mean is larger than the sample variance. Communications in Statistics -- Simulation and Computation 14, 1533-1547.*

Consul, P.C. (1989) *Generalized Poisson Distributions: Properties and Applications. Marcel Dekker Inc., New York Basel.*

Denuit M., X. Maréchal, S. Pitrebois & J.-F. Walhin, (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems.*

E. Gómez-Déniz, A. Hernández Bastida, & M.P. Fernández Sánchez (2012), *Bulletin of the Malaysian Mathematical Sciences Society. A Suitable Discrete Distribution for Modelling Automobile Claim Frequencies.*

Gossiaux, A.-M., & Lemaire, J. (1981). *Methodes d'ajustement de distributions de sinistres. Bulletin of the Swiss Association of Actuaries, 87–95.*

Ghitany, M.E. & Al-Mutairi, D.K. (2009). *Estimation methods for the discrete Poisson-Lindley distribution. Journal of Statistical Computation and Simulation*

Greenwood, M., & Yule, G.U. (1920). *An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society, Series A 83, 255–279.*

- Holla, M., (1966), *On a Poisson-inverse Gaussian distribution*. *Metrika*, 11, 115–121.
- Jorgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*. *Lecture Notes in Statistics*, 9. Springer-Verlag, New York.
- Karlis, D., Xekalaki, E., (2005). *Mixed Poisson distributions*. *International Statistical Review* 73, 35–5.
- Klugman, S. (1992). *Bayesian Statistics in Actuarial Science*. Kluwer, Boston.
- Klugman, S., Panjer, H., & Willmot, G. (2004). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., New York.
- Kokonendji, C. & Khoudar, M. (2004). *On strict arcsine distribution*. *Communications in Statistics-Theory and Methods*, 33, 5, 993-1006.
- Lemaire, J. (1976). *Driver versus company: Optimal behavior of the policyholder*. *Scandinavian Actuarial Journal*, 209–219.
- Lemaire Jean, (1985). *Automobile Insurance Actuarial Models, Poisson Model – Homogeneous Portfolio*.
- Lemaire Jean, (1985). *Automobile Insurance Actuarial Models, Negative Binominal Model – Heterogeneous Portfolio*.
- Letac, G. & Mora, M. (1990). *Natural real exponential families with cubic variance functions*. *Ann. Statist.* 18:1–37.
- Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publisher, Boston
- Meng Shengwang, Wei Yuan & G.A. Whitmore. "Accounting for individual over-dispersion in a bonus-malus automobile insurance system", *The ASTIN Bulletin*, Volume 29, No. 2, 1999, pp.327-337.
- Ord, J. & Whitmore, G. (1986). *The Poisson-Inverse Gaussian distribution as a model for species abundance*. *Communications in Statistics A, Theory and Methods* 15, 853-871.
- Sankaran, M. (1971). *The Discrete Poisson-Lindley Distribution*. *Biometrics*, 26, 145-149
- Shaked, M. & Shianthikumar, J.G. (1994). *Stochastic Orders and their Applications*. Academic Press. New York.
- Sichel, H. (1971). *On a family of discrete distributions particularly suited to represent long tailed frequency data*. *Proceedings of the Third Symposium on Mathematical Statistics (ed. N. Laubscher)*. Pretoria, CSIR.

Sichel, H. (1982). *Asymptotic efficiencies of three methods of estimation for the inverse Gaussian-Poisson distribution*. *Biometrika* 69, 467-472.

Wang, L.M., Lei, Y.L., 2000. *Simulation & EM algorithm for the distribution of number of claim in the heterogeneous portfolio*. *Communications on Applied Mathematics and Computational Science* 14(2), 71–78.

Wang, Zhaoliang (2011). *One mixed negative binomial distribution with application*", *Journal of Statistical Planning and Inference*, 141 (3), 1153-1160

Willmot, G. (1986). *Mixed compound Poisson distributions*. *Astin Bulletin* 16, S59-S79.

Willmot, G.E. (1987). *The Poisson-inverse Gaussian distribution as an alternative to the negative binomial*. *Scandinavian Actuarial Journal*, 113–127.

Willmot, G. (1988). *Parameter orthogonality for a family of discrete distributions*. *Journal of the American Statistical Association*, to appear.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ