ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ **ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ**

ΥΠΟΛΟΓΙΣΤΙΚΗ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΜΕ ΤΟ ΠΑΚΕΤΟ actuar ΤΗΣ ΓΛΩΣΣΑΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R

Βασιλική Σ. Μαρκουτσά

Διπλωματική Εργασία

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

Πειραιάς

Σεπτέμβριος 2022

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 Διοικητική Κινδύνου

Τα μέλη της Επιτροπής ήταν:

- Αντζουλάκος Δημήτριος, Αναπληρωτής Καθηγητής του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης (Επιβλέπων)
- Τζαβελάς Γεώργιος, Αναπληρωτής Καθηγητής του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης
- Ψαρράκος Γεώργιος, Αναπληρωτής Καθηγητής του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης

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

UNIVERSITY OF PIRAEUS

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
ACTUARIAL SCIENCE AND RISK MANAGEMENT

COMPUTATIONAL ACTUARIAL SCIENCE WITH THE R actuar PACKAGE

By

Vasiliki S. Markoutsa

MSc Dissertation

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

Piraeus, Greece
September 2022

Στην οικογένειά μου

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές ευχαριστίες και την ιδιαίτερη εκτίμηση μου, προς τον καθηγητή του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς και επιβλέποντα στην διπλωματική μου εργασία κ. Δημήτριο Αντζουλάκο, για την καθοδήγησή του καθ' όλη τη διάρκεια εκπόνησης της εργασίας αυτής με στόχο το καλύτερο δυνατό αποτέλεσμα καθώς και την καθοριστική συμβολή του στην ολοκλήρωσή της.

Θα ήθελα να ευχαριστήσω επίσης και τα άλλα δύο μέλη της τριμελούς εξεταστικής επιτροπής, τους κύριους Γεώργιο Τζαβελλά, Αναπληρωτή Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, και Γεώργιο Ψαρράκο, Αναπληρωτή καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, για την επίβλεψη τους.

Τέλος, θα ήθελα να εκφράσω ένα θερμό ευχαριστώ στην οικογένειά μου, καθώς και σε όλους τους συναδέλφους και τους φίλους μου, για την στήριξη, την υπομονή και την βοήθεια καθ’ όλη την διάρκεια των σπουδών μου.

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η μελέτη ειδικών θεμάτων της θεωρίας κινδύνων και η αριθμητική επίλυση προβλημάτων που σχετίζονται με αυτά με χρήση της γλώσσας προγραμματισμού R. Συγκεκριμένα, η εν λόγω μελέτη εστιάζεται κυρίως στο μοντέλο συλλογικού κινδύνου και την πιθανότητα χρεοκοπίας ενός χαρτοφυλακίου.

Η μελέτη μας θα ξεκινήσει δίνοντας μία σύντομη περιγραφή σε βασικές έννοιες της θεωρίας πιθανοτήτων καθώς και στις κλάσεις κατανομών που έχουν εφαρμογή στο μοντέλο συλλογικό κινδύνου. Στη συνέχεια, δίνεται έμφαση στην παρουσίαση της θεωρίας που διέπει το μοντέλο συλλογικό κινδύνου όπου δίνεται έμφαση στις μεθόδους υπολογισμού της κατανομής των συνολικών αποζημιώσεων με χρήση του πακέτου `actuar` της γλώσσας προγραμματισμού R. Τέλος, εστιάζουμε στο κεντρικό πρόβλημα της θεωρίας χρεοκοπίας που είναι ο ακριβής υπολογισμός της πιθανότητας χρεοκοπίας.

Όλα τα θέματα τα οποία παρουσιάζονται στην παρούσα εργασία μελετώνται και ερμηνεύονται τόσο σε θεωρητικό όσο και σε πρακτικό επίπεδο. Έτσι, σε κάθε κεφάλαιο παρουσιάζεται η σχετική θεωρία και δίνονται και επιλύονται αριθμητικά παραδείγματα για την καλύτερη κατανόησή της.

ABSRACT

The subject of this thesis is the study of special topics of risk theory and the numerical solution of related problems using the programming language R. In particular, this study focuses mainly on the collective risk model and the probability of default of a portfolio.

Our study will start by giving a brief description of basic concepts of probability theory and the classes of distributions that apply to the collective risk model. Next, we emphasize the theory underlying the collective risk model where we focus on the methods of computing the distribution of total compensation using the `actuar` package of the R programming language. Finally, we focus on the central problem of default theory which is the exact computation of the probability of default.

All the topics presented in this paper are studied and interpreted both at the theoretical and practical level. Thus, in each chapter the relevant theory is presented and numerical examples are given and solved for a better understanding.

Πίνακας περιεχομένων

1 Κεφάλαιο: ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό δίνεται μία σύντομη αναφορά σε βασικές έννοιες της θεωρίας πιθανοτήτων που θα μας απασχολήσουν στην παρούσα εργασία. Ιδιαίτερη μνεία γίνεται στις σημαντικότερες διακριτές και συνεχείς κατανομές, αλλά και σε άλλες αρκετά δημοφιλείς κατανομές που χρησιμοποιούνται ευρέως στην αναλογιστική επιστήμη και τη θεωρία κινδύνων, όπως η λογαριθμοκανονική, η κατανομή Weibull, η κατανομή Pareto, κλπ.

Επιπλέον σημαντικό αντικείμενο μελέτης του κεφαλαίου αυτού είναι οι κλάσεις κατανομών $R(a,b,0)$ και στην κλάση $R(a,b,1)$. Συγκεκριμένα, αναλύονται διεξοδικά οι ανωτέρω κλάσεις και αναγνωρίζονται οι παράμετροι a και b για διάφορες γνωστές κατανομές.

1.1 Τυχαία μεταβλητή

1.1.1 Ορισμός

Έστω Ω ο δειγματικός χώρος ενός πειράματος τύχης. Μια πραγματική συνάρτηση $X{:}\,\Omega \rightarrow \mathrm{R}$ λέγεται τυχαία μεταβλητή (τ.μ.), αν για κάθε υποσύνολο I του R το σύνολο $X^{-1}(I)$ είναι ενδεχόμενο του Ω, δηλαδή το σύνολο

$$X^{-1}(I) = \{\omega \in \Omega{:}\, X(\omega) \in I\} \subseteq \Omega$$

είναι ενδεχόμενο του Ω.

Αν στο δειγματικό χώρο Ω έχει οριστεί μια συνάρτηση πιθανότητας P_Ω, τότε μπορούμε να γράψουμε ότι για κάθε $I \subseteq R$

$$P(X \in I) = P_\Omega\big(X^{-1}(I)\big).$$

1.2 Συνάρτηση κατανομής

1.2.1 Ορισμός

Έστω X μια τυχαία μεταβλητή που ορίζεται σε κάποιο δ.χ. Ω. Η πραγματική συνάρτηση $F{:}\,R \rightarrow R$ που δίνεται από τον τύπο

$$F(x) = P(X \leq x) = P_\Omega(X^{-1}\big((-\infty, x]\big), \quad x \in R$$

ονομάζεται (αθροιστική) συνάρτηση κατανομής της X.

1.2.2 Ιδιότητες

Ισχύουν οι ακόλουθες ιδιότητες:

- ✓ Αν $x_1 < x_2$,τότε $F(x_1) \leq F(x_2)$, δηλαδή η F είναι αύξουσα συνάρτηση.
- ✓ $\lim_{x\to\infty} F(x) = 1$ και $\lim_{x\to-\infty} F(x) = 0$ δηλαδή $0 \leq F(x) \leq 1$
- ✓ Για κάθε φθίνουσα ακολουθία πραγματικών αριθμών $\{x_\nu\}_{\nu\geq 1}$ με $\lim_{\nu\to\infty} x_\nu = x$,

ισχύει $\lim_{\nu\to\infty} F(x_\nu) = F(x)$.

Η τελευταία ιδιότητα δηλώνει ότι η συνάρτηση μιας τυχαίας μεταβλητής X είναι συνεχής από δεξιά

1.3 Διακριτές τυχαίες μεταβλητές

Στη θεωρία πιθανοτήτων, μία θεμελιώδης διάκριση για τις τυχαίες μεταβλητές που χρησιμοποιούμε στην πράξη είναι ανάμεσα σε διακριτές και συνεχείς. Μία τυχαία μεταβλητή X λέγεται διακριτή (ή απαριθμητή), αν το σύνολο τιμών της R_X είναι πεπερασμένο ή το πολύ (απείρως) αριθμήσιμο. Άρα για διακριτές τ.μ. έχουμε ότι

$R_X = \{x_1, x_2, x_3, \dots, x_\nu\}$ ή $R_X = \{x_1, x_2, x_3, \dots\}$,

αντιστοίχως.

1.3.1 Ορισμός

Σε κάθε διακριτή τυχαία μεταβλητή X, αντιστοιχεί μια πραγματική συνάρτηση $f: R \to R$ με τύπο

$$f(x_i) = P(X = x_i), \quad x_i \in R_X$$

και $f(x) = 0$ για $x \notin R_X$, η οποία ονομάζεται συνάρτηση πιθανότητας (σ.π.) της τυχαίας μεταβλητής X.

1.3.2 Ιδιότητες

Η συνάρτηση πιθανότητας f μιας διακριτής τυχαίας μεταβλητής X ικανοποιεί τις επόμενες συνθήκες:

- ✓ $f(x) = 0, \quad x \notin R_X$
- ✓ $f(x_i) \geq 0, \quad x_i \in R_X$
- ✓ $\sum_{x_i \in R_X} f(x_i) = 1$

Σημειώνεται ότι ισχύει και το αντίστροφο του ανωτέρου, δηλαδή αν για μια συνάρτηση $f: R \to R$ ικανοποιούνται και οι τρεις παραπάνω συνθήκες, τότε η f

αποτελεί συνάρτηση πιθανότητας μιας τυχαίας μεταβλητής X με σύνολο τιμών το R_X.

1.3.3 Μέση Τιμή

Έστω X μια διακριτή τ.μ. με συνάρτηση πιθανότητας $f(x)$ και g μία πραγματική συνάρτηση. Τότε η μέση τιμή της διακριτής τυχαίας μεταβλητή $g(X)$ δίνεται από τη σχέση

$$E[g(X)] = \sum_{x \in R_X} g(x)f(x).$$

1.4 Συνεχείς τυχαίες μεταβλητές

1.4.1 Ορισμός

Μία τυχαία μεταβλητή X λέμε ότι είναι απόλυτα συνεχής (ή απλά συνεχής), αν υπάρχει μία μη αρνητική συνάρτηση $f: R \to [0, \infty)$ με την ιδιότητα ότι για κάθε υποσύνολο A των πραγματικών αριθμών, το οποίο μπορεί να γραφεί ως ένωση πεπερασμένου ή απείρως αριθμήσιμου πλήθους διαστημάτων, ισχύει ότι

$$P(X \in A) = \int_A f(x)\, dx.$$

Στην περίπτωση αυτή η συνάρτηση f ονομάζεται συνάρτηση πυκνότητας πιθανότητας ή απλά συνάρτηση πυκνότητας της X.

1.4.2 Ιδιότητες

Ισχύουν οι ακόλουθες ιδιότητες:

- ✓ $f(x) \geq 0, \quad x \in R.$
- ✓ $\int_{-\infty}^{\infty} f(x)dx = 1.$

1.4.3 Μέση Τιμή

Έστω X μια συνεχής τ.μ. με συνάρτηση πυκνότητας $f(x)$ και g μία πραγματική συνάρτηση. Τότε η μέση τιμή της συνεχούς τυχαίας μεταβλητή $g(X)$ δίνεται από την σχέση

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx\,.$$

Ολοκληρώνοντας τη σύντομη παρουσίαση των διακριτών και συνεχών τυχαίων μεταβλητών σημειώνουμε ότι

$$F(t) = \sum_{x \in R_X : x \leq t} f(x) \qquad (X \text{ διακριτή τ.μ.})$$

και

$$F(t) = \int_{-\infty}^{t} f(x)dx \qquad (X \text{ συνεχής τ.μ.})$$

Η συνάρτηση που εκφράζει την πιθανότητα του συμπληρωματικού ενδεχομένου σε σχέση με τη συνάρτηση F συμβολίζεται $\bar{F}$. Συγκεκριμένα, η συνάρτηση

$$\bar{F}(x) = 1 - F(x) = P(X > x), \quad x \in R$$

αναφέρεται ως συνάρτηση επιβίωσης ή ως ουρά της κατανομής της τ.μ. X.

Μία άλλη σημαντική έννοια που συνδέεται με μία τυχαία μεταβλητή X είναι η διακύμανση της X, η οποία αποτελεί το κύριο μέτρο μεταβλητότητας των τιμών της τυχαίας μεταβλητής, και δίνεται από τον τύπο

$$Var(X) = E\big[\big(X - E(X)\big)\big]^2 = E(X^2) - [E(X)]^2.$$

Σημειώνεται ότι, η τετραγωνική ρίζα της διακύμανσης αναφέρεται ως η τυπική απόκλιση της X.

Ο συντελεστής μεταβλητότητας μια μεταβλητής X, που συμβολίζεται με CV_X, ορίζεται ως το πηλίκο της τυπικής απόκλισης προς τη μέση τιμή της X, δηλαδή

$$CV_X = \frac{\sqrt{Var(X)}}{E(X)}.$$

Ο συντελεστής ασυμμετρίας μια τυχαίας μεταβλητής X με μέση τιμή $\mu = E(X)$ συμβολίζεται με γ και ορίζεται από την σχέση

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

Για τυχαίες μεταβλητές των οποίων η κατανομή είναι συμμετρική (όπως η κανονική και η ομοιόμορφη) έχουμε $\gamma = 0$, ενώ αν $\gamma > 0$ ($\gamma < 0$) λέμε ότι η κατανομή της X παρουσιάζει θετική (αντίστοιχα αρνητική) ασυμμετρία.

1.5 Συνελίξεις

Η συνέλιξη δύο συναρτήσεων f και g συμβολίζεται με $f * g$ και ορίζεται ως εξής:

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(x-t)g(t)dt.$$

Έστω τώρα δύο ανεξάρτητες συνεχείς τυχαίες μεταβλητές X και Y, με αντίστοιχες συναρτήσεις πυκνότητας $f(x)$ και $g(x)$ που ορίζονται στο διάστημα $[0,\infty)$. Τότε η συνάρτηση πυκνότητας της τυχαίας μεταβλητής $X+Y$, δίνεται από τον τύπο

$$(f * g)(x) = \int_0^x f(x-t)g(t)dt = \int_0^x f(t)g(x-t)dt\,, \quad x \geq 0.$$

Αντίστοιχα, ορίζεται και η συνέλιξη για «διακριτές» συναρτήσεις $f(x)$ και $g(x)$ (δηλαδή για ακολουθίες). Ειδικά στην περίπτωση που οι f, g ορίζονται στο σύνολο $\{0, 1, 2, \dots\}$, έχουμε

$$(f * g)(x) = \sum_{t=0}^{x} f(x-t)g(t) = \sum_{t=0}^{x} f(t)g(x-t)\,, \quad x = 0, 1, 2, \dots .$$

Αν οι συναρτήσεις $f(x)$ και $g(x)$ είναι συναρτήσεις πιθανότητας των διακριτών τ.μ. X, Y, αντίστοιχα, τότε η παραπάνω συνάρτηση $(f * g)(x)$ είναι η συνάρτηση πιθανότητας του αθροίσματος $X+Y$.

Τα παραπάνω, μπορούν να γενικευθούν για περισσότερες από δύο συναρτήσεις. Για παράδειγμα,

$$(f * g * h)(x) = ((f * g) * h)(x) = \big(f * (g * h)\big)(x).$$

Ειδικότερα, έχουμε και τον ακόλουθο συμβολισμό

$$\underbrace{(f * f * \dots * f)}_{n}(x) = f^{*n}(x).$$

1.6 Σύνθετες κατανομές

Οι σύνθετες κατανομές χρησιμοποιούνται εκτενώς στη Θεωρία Κινδύνου και ειδικότερα στο πρότυπο συλλογικού κινδύνου.

Η γενική μορφή το μοντέλου συλλογικού κινδύνου περιγράφεται από τη σύνθετη κατανομή

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

όπου:

– η τυχαία μεταβλητή $X_{i,}$ παριστάνει το ύψος της i ατομικής ζημιάς (δεν μπορεί να πάρει αρνητικές τιμές), με συνάρτηση πιθανότητας/πυκνότητας f

– η τυχαία μεταβλητή N, παριστάνει το πλήθους των ζημιών (κινδύνων) ή και απαιτήσεων.

– η τυχαία μεταβλητή S, παριστάνει το ύψος της συνολικής ζημιάς

Η τυχαία μεταβλητή S θα μελετηθεί διεξοδικά στο επόμενο κεφάλαιο.

1.7 Οι κυριότερες διακριτές και συνεχείς κατανομές

Στη συνέχεια παρατίθενται ο ορισμός και βασικά χαρακτηριστικά των κυριότερων διακριτών και συνεχών κατανομών που χρησιμοποιούνται στην θεωρία των κινδύνων.

Σημειώνεται ότι, όταν μία κατανομή πιθανότητας χρησιμοποιείται ως πρότυπο μοντέλο για το μέγεθος των απαιτήσεων σε ένα ασφαλιστικό χαρτοφυλάκιο, τότε στην ασφαλιστική ορολογία αυτή ονομάζεται ως ζημιοκατανομή.

1.7.1 Διακριτές κατανομές

Διωνυμική

Μία τυχαία μεταβλητή X θα λέμε ότι ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και p (συμβολικά $X \sim B(n, p)$) όταν η συνάρτηση πιθανότητας της X είναι

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

Εδώ το n είναι θετικός ακέραιος, p ένας θετικός αριθμό στο διάστημα (0,1) και $q = 1 - p$. Επίσης, η μέση τιμή και η διακύμανση δίνονται από τις σχέσεις

$$E(X) = np, \qquad Var(X) = npq.$$

Για $n = 1$, η διωνυμική κατανομή ονομάζεται συνήθως κατανομή Bernoulli.

Γεωμετρική

Θεωρούμε μία ακολουθία (ανεξάρτητων) δοκιμών Bernoulli με πιθανότητα επιτυχίας p (και αποτυχίας $q - 1$) σταθερή για όλες τις δοκιμές. Έστω X ο αριθμός των δοκιμών μέχρι την εμφάνιση της πρώτης επιτυχίας. Η κατανομή της τυχαίας μεταβλητής X καλείται γεωμετρική κατανομή με παράμετρο p και συμβολίζεται $G(p)$.

Η συνάρτηση πιθανότητας της γεωμετρικής κατανομής $G(p)$ (ή $G_1(p)$) δίνεται από τον τύπο

$$f(x) = P(X = x) = q^{x-1} p, \quad x = 1, 2, \dots .$$

Επίσης, η μέση τιμή και η διακύμανση δίνονται από τις σχέσεις

$$E(X) = \frac{1}{p}, \qquad Var(X) = \frac{q}{p^2}.$$

Με $G_0(p)$ συμβολίζουμε την κατανομή της τ.μ. $Y = X - 1$ με σύνολο τιμών $R_Y = \{0,1,2,\dots\}$.

Αρνητική Διωνυμική

Η κατανομή αυτή αποτελεί γενίκευση της γεωμετρικής κατανομής. Σε μία ακολουθία δοκιμών Bernoulli με πιθανότητα επιτυχίας p, θεωρούμε τώρα την τ.μ. X η οποία παριστάνει τον αριθμό δοκιμών μέχρι να εμφανιστούν r επιτυχίες , όπου r είναι ένας θετικός ακέραιος. Τότε το πεδίο τιμών της X είναι το σύνολο $R_X = \{r, r+1, r+2, \dots\}$, ενώ η αντίστοιχη συνάρτηση πιθανότητας δίνεται από την σχέση

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}, \qquad x = r, r+1, \dots .$$

Η κατανομή της παραπάνω τυχαίας μεταβλητής X καλείται αρνητική διωνυμική κατανομή με παραμέτρους r και p, και συμβολίζεται $Nb{\sim}(r,p)$ (ή $NB_r(r,p)$). Επίσης, η μέση τιμή και η διακύμανση δίνονται από τις σχέσεις

$$E(X) = \frac{r}{p}, \qquad Var(X) = \frac{rq}{p^2}.$$

Με $NB_0(r,p)$ συμβολίζουμε την κατανομή της τ.μ. $Y = X - r$ με σύνολο τιμών $R_Y = \{0,1,2,\dots\}$.

Poisson

Μία τυχαία μεταβλητή X λέμε ότι ακολουθεί την κατανομή Poisson με παράμετρο λ (όπου $\lambda > 0$) όταν η συνάρτηση πιθανότητας της X δίνεται από την σχέση

$$f(x) = P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \dots .$$

Συμβολικά γράφουμε $X \sim P(\lambda)$. Επίσης, η μέση τιμή και η διακύμανση δίνονται από τις σχέσεις

$$E(X) = \lambda, \qquad Var(X) = \lambda.$$

1.7.2 Συνεχείς κατανομές

Κανονική κατανομή

Μία τυχαία μεταβλητή X λέμε ότι ακολουθεί την κανονική κατανομή με παραμέτρους $\mu \in R$ και $\sigma^2 > 0$, όταν η συνάρτηση πυκνότητας της X δίνεται από την σχέση

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty.$$

Στη περίπτωση αυτή γράφουμε συμβολικά ότι $X \sim N(\mu, \sigma^2)$. Όταν έχουμε $\mu = 0$ και $\sigma^2 = 1$, τότε λέμε ότι η X είναι μία τυποποιημένη κανονική τυχαία μεταβλητή με κατανομή $X \sim N(0,1)$.

Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \mu, \qquad Var(X) = \sigma^2.$$

Για την αθροιστική συνάρτηση της κανονικής κατανομής δεν υπάρχει αναλυτικός τύπος.

Εκθετική κατανομή

Μία τυχαία μεταβλητή X λέμε ότι ακολουθεί την εκθετική κατανομή με παράμετρο β και συμβολικά γράφουμε $X \sim Exp(\beta)$ όταν έχει συνάρτηση πυκνότητας

$$f(x) = \beta e^{-\beta x}, \quad x \geq 0$$

ενώ $f(x) = 0$ για $x < 0$.

Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \frac{1}{\beta}, \qquad Var(X) = \frac{1}{\beta^2}.$$

Η αθροιστική συνάρτηση κατανομής της εκθετικής κατανομής δίνεται από τον τύπο

$$F(x) = 1 - e^{-\beta x}, \qquad x \geq 0.$$

Ομοιόμορφη κατανομή

Έστω α, β δύο πραγματικοί αριθμοί με $\alpha < \beta$. Θα λέμε ότι μία μεταβλητή X ακολουθεί την ομοιόμορφη κατανομή στο διάστημα $[\alpha, \beta]$ όταν η συνάρτηση πυκνότητας της X δίνεται από την σχέση

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & x \in [\alpha, \beta] \\ 0, & x \notin [\alpha, \beta]. \end{cases}$$

Συμβολικά, γράφουμε $X \sim Uni(\alpha, \beta)$. Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \frac{\alpha + \beta}{2}, \qquad Var(X) = \frac{(\beta - \alpha)^2}{12}.$$

Η αθροιστική συνάρτηση κατανομής, $F(x) = P(X \leq x)$, για μία μεταβλητή X με κατανομή $U(\alpha, \beta)$ δίνεται από τον τύπο

$$F(x) = \begin{cases} 0, & x < \alpha \\ \dfrac{x - \alpha}{\beta - \alpha}, & \alpha \leq x \leq \beta \\ 1, & x > \beta. \end{cases}$$

Γάμμα κατανομή

Η κατανομή Γάμμα αποτελεί γενίκευση της εκθετικής κατανομής που είδαμε παραπάνω. Μία μεταβλητή X ακολουθεί την κατανομή Γάμμα με παραμέτρους $\alpha > 0$ και $\beta > 0$ όταν έχει συνάρτηση πυκνότητας

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

Εδώ $\Gamma(\cdot)$ είναι η συνάρτηση Γάμμα, που ορίζεται από τη σχέση

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

για κάθε $\alpha > 0$. Στην περίπτωση που το α είναι ένας θετικός ακέραιος, τότε ισχύει $\Gamma(\alpha) = (\alpha - 1)! = 1 \cdot 2 \cdot 3 \cdots (\alpha - 1)$.

Συμβολικά, γράφουμε $X \sim Ga(\alpha, \beta)$. Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \frac{\alpha}{\beta}, \qquad Var(X) = \frac{\alpha}{\beta^2}.$$

Είναι φανερό ότι για $\alpha = 1$, η κατανομή Γάμμα συμπίπτει με την εκθετική κατανομή με παράμετρο β.

Γενικότερα, στην περίπτωση όπου η παράμετρος α είναι ένας θετικός ακέραιος η κατανομή αναφέρεται συνήθως ως κατανομή Erlang. Επίσης, η αθροιστική συνάρτηση της κατανομής Γάμμα δεν είναι γενικά εύκολο να υπολογιστεί. Ωστόσο, για την κατανομή Erlang, η αθροιστική της συνάρτηση υπολογίζεται από συγκεκριμένους αναλυτικούς τύπους.

Η κατανομή Pareto

Μία τυχαία μεταβλητή X λέμε ότι ακολουθεί την κατανομή Pareto με παραμέτρους α και β αν η συνάρτηση πυκνότητάς της δίνεται από τη σχέση

$$f(x) = \frac{\alpha\beta^{\alpha}}{(\beta + x)^{\alpha+1}}, \qquad x \geq 0.$$

Συμβολικά, γράφουμε $X \sim Pa(\alpha, \beta)$. Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \frac{\beta}{\alpha - 1}, \qquad Var(X) = \frac{\alpha\beta^2}{(\alpha - 1)^2(\alpha - 2)}.$$

Για την ύπαρξη της μέσης τιμής αναγκαία και ικανή συνθήκη είναι η τιμή της παραμέτρου α να είναι $\alpha > 1$, ενώ η διακύμανση είναι πεπερασμένη αν και μόνο αν $\alpha > 2$. Η αθροιστική συνάρτηση κατανομής, $F(x) = P(X \leq x)$, για μία μεταβλητή X με κατανομή Pareto δίνεται από τον τύπο

$$F(x) = P(X \leq x) = 1 - \left(\frac{\beta}{x + \beta}\right)^{a}, \quad x \geq 0$$

Η κατανομή Weibull

Λέμε ότι μια τυχαία μεταβλητή X ακολουθεί την Weibull κατανομή με παραμέτρους $\alpha, \beta > 0$, αν η συνάρτηση πυκνότητάς της δίνεται από τον τύπο

$f_X(x) = \frac{\alpha}{\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\}, \quad x \geq 0.$

Συμβολικά θα γράφουμε $X \sim Wei(\alpha, \beta)$. Για την κατανομή αυτή ισχύει ότι

$$E(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right),$$

$$V(X) = \beta^2\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right).$$

Αν θέσουμε $\tau = \beta^{-\alpha}$ και $\gamma = \alpha$, τότε η συνάρτηση πυκνότητας μπορεί να γραφτεί στην εναλλακτική μορφή

$f_X(x) = \tau\gamma \cdot x^{\gamma-1}\, exp\{-\tau x^{\gamma}\}, \quad x \geq 0.$

Η αθροιστική συνάρτηση κατανομής που αντιστοιχεί στην παραπάνω πυκνότητα είναι

$$F(x) = 1 - e^{-\tau x^{\gamma}}, \ \ x \geq 0,$$

και

$$\bar{F}(x) = 1 - F(x) = e^{-\tau x^{\gamma}}, \quad x \geq 0.$$

Περιπτώσεις της τιμής της παραμέτρου γ:

- για $\gamma = 1$, προκύπτει η εκθετική κατανομή
- για $\gamma < 1$, η κατανομή έχει πιο βαριά ουρά από την εκθετική κατανομή. Η ουρά της κατανομής συγκλίνει στο μηδέν πιο αργά από την εκθετική κατανομή
- για $\gamma > 1$, η κατανομή έχει πιο ελαφριά ουρά από την εκθετική κατανομή, δηλαδή η ουρά της κατανομής συγκλίνει πιο γρήγορα στο μηδεν.

Λογαριθμοκανονική κατανομή

Μία τυχαία μεταβλητή X λέμε ότι έχει λογαριθμοκανονική κατανομή με παραμέτρους $\mu \in R$ και $\sigma^2 > 0$, αν ο φυσικός λογάριθμος της, $Y = ln\, X$, ακολουθεί την κανονική κατανομή $X \sim N(\mu, \sigma^2)$.

Παρατηρούμε αρχικά ότι το πεδίο τιμών της X είναι το διάστημα $(0, \infty)$, δηλαδή η X παίρνει μόνο θετικές τιμές. Συμβολικά, γράφουμε $X \sim LN(\mu, \sigma^2)$ και η πυκνότητα της X είναι

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{-(lnx - \mu)^2}{2\sigma^2}\right), \qquad 0 < x < \infty.$$

Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \qquad Var(X) = [\exp(\sigma^2 - 1)] \cdot \exp\,(2\mu + \sigma^2)$$

Πιθανότητες που αναφέρονται στην τυχαία μεταβλητή $X \sim LN(\mu, \sigma^2)$ υπολογίζονται με βάση το γεγονός ότι η $Y = ln\, X$ ακολουθεί την κανονική κατανομή.

1.8 Ροπογεννήτριες - Πιθανογεννήτριες

1.8.1 Ορισμός

Έστω X μία διακριτή τυχαία μεταβλητή, η οποία παίρνει μη-αρνητικές ακέραιες τιμές με συνάρτηση πιθανότητας f. Η πιθανογεννήτρια συνάρτηση P_X της τυχαίας μεταβλητής X, ορίζεται ως

$$P_X(s) = E(s^X) = \sum_{x=0}^{\infty} P(X = x)s^x = \sum_{x=0}^{\infty} f(x)s^x\,.$$

Είναι γνωστό ότι η πιθανογεννήτρια συγκλίνει απόλυτα στο διάστημα $[-1, 1]$. Με παραγώγιση της πιθανογεννήτριας συνάρτησης προκύπτουν οι παραγοντικές ροπές της X, αφού

$$\frac{d^r}{ds^r} P_X(s)\bigg|_{s=1} = E(X(X-1)(X-2)\dots(X-r+1)), \qquad r \geq 0.$$

Από τα παραπάνω, προκύπτει ότι

$$E(X) = P_X'(1)$$

$$E[X(X-1)] = P_X''(1)$$

$$Var(X) = E(X^2) - E^2(X) = P_X''(1) + P_X'(1) - \left(P_X'(1)\right)^2.$$

Αν $P_X(s)$ είναι η πιθανογεννήτρια συνάρτηση μίας διακριτής τυχαίας μεταβλητής X, τότε η συνάρτηση πιθανότητας της X ικανοποιεί τη σχέση

$$f(r) = P(X = r) = \frac{1}{r!} \cdot \frac{d^r}{ds^r} P_X(s)\bigg|_{s=0}, \quad r = 0, 1, 2, \dots$$

Τέλος, έστω X μία διακριτή τυχαία μεταβλητή με πιθανογεννήτρια P_X και a, b δύο πραγματικοί αριθμοί. Τότε, η πιθανογεννήτρια του γραμμικού συνδυασμού $Y = aX + b$ δίνεται από τον τύπο

$$P_Y(s) = P_{aX+b}(s) = s^b P_X(s^a).$$

1.8.2 Ορισμός

Η ροπογεννήτρια συνάρτηση $M_X(s)$ μιας τυχαίας μεταβλητής X, δίνεται από τον τύπο

$$M_X(s) = E(e^{sX})$$

υπό την προϋπόθεση ότι η μέση τιμή $E(e^{sX})$ υπάρχει για κάθε s, που ανήκει τουλάχιστον σε ένα διάστημα της μορφής $(-\delta, \delta)$, $\delta > 0$.

Τονίζουμε, όπως φαίνεται και από τον παραπάνω ορισμό, ότι η ροπογεννήτρια συνάρτηση μιας τυχαίας μεταβλητής δεν υπάρχει πάντα.

- Αν η τυχαία μεταβλητή X είναι διακριτή, με συνάρτηση πιθανότητας f, τότε η ροπογεννήτριά της δίνεται από τον τύπο

$$M_X(s) = \sum_{x \in R_X} e^{sx} f(x).$$

- Αν η τυχαία μεταβλητή X είναι συνεχής, με συνάρτηση πυκνότητας f, τότε η ροπογεννήτριά της δίνεται από τον τύπο

$$M_X(s) = \int_{-\infty}^{+\infty} e^{sx} f(x) dx.$$

Στην τελευταία περίπτωση σημειώνουμε, ότι η ποσότητα

$$L_X(s) = M_X(-s) = \int_{-\infty}^{+\infty} e^{-sx} f(x) dx$$

είναι ο μετασχηματισμός Laplace της συνάρτησης $f(x)$.

Με παραγώγιση της ροπογεννήτριας συνάρτησης, προκύπτουν οι ροπές r −τάξης, μ_r, της X, αφού

$$\left. \frac{d^r}{ds^r} M_X(s) \right|_{s=0} = E(X^r) = \mu_r, \qquad r \geq 0$$

λόγω του ότι

$$\frac{d^r}{ds^r} M_X(s) = E(X^r e^{sX}).$$

Από τα παραπάνω, προκύπτει ότι

$$E(X) = M_X'(0)$$

$$E(X^2) = M_X''(0)$$

$$Var(X) = E(X^2) - E^2(X) = M_X''(0) - \left(M_X'(0)\right)^2.$$

Έστω X μία τυχαία μεταβλητή, με ροπογεννήτρια M_X και a, b δύο πραγματικοί αριθμοί. Τότε, η ροπογεννήτρια του γραμμικού συνδυασμού $Y = aX + b$ δίνεται από τον τύπο

$$M_Y(s) = M_{aX+b}(s) = e^{bs} M_X(as).$$

1.9 Κλάση κατανομών $R(a, b, 0)$

1.9.1 Ορισμός

Η διακριτή τυχαία μεταβλητή N, με $R_N = \{0,1,2, \ldots\}$, ανήκει στην κλάση κατανομών $R(a, b, 0)$, αν η συνάρτηση πιθανότητάς της $p_n = \Pr(N = n)$, ικανοποιεί την αναδρομική σχέση

$$p_n = p(n) = \left(a + \frac{b}{n}\right) p_{n-1}, \qquad n = 1,2, \dots$$

όπου, a και b είναι κατάλληλες σταθερές, και προφανώς $p(n) = 0$ για $n < 0$. Ο παραπάνω τύπος είναι γνωστός ως αναδρομικός τύπος του Panjer και μελετήθηκε από τον Panjer (1981).

Οι κατανομές που ικανοποιούν την σχέση $p_n = p(n) = \left(a + \frac{b}{n}\right) p_{n-1}$, $n = 1, 2, \dots$ είναι η Poisson, η διωνυμική, η αρνητική διωνυμική, και προφανώς η γεωμετρική.

1.9.2 Μέλη της κλάσης $\boldsymbol{R(a, b, 0)}$

- **Η κατανομή Poisson**

Αν η τ.μ. $N \sim P(\lambda)$, έπεται ότι

$$\frac{p_n}{p_{n-1}} = \frac{e^{-\lambda} \dfrac{\lambda^n}{n!}}{e^{-\lambda} \dfrac{\lambda^{n-1}}{(n-1)!}} = \frac{\lambda}{n} \Leftrightarrow p_n = \left(\frac{\lambda}{n}\right) p_{n-1}$$

οπότε $a = 0$ και $b = \lambda$. Ακόμη, $p_0 = e^{-\lambda}$.

- **Η διωνυμική κατανομή**

Αν η τ.μ. $N \sim B(m, p)$, έπεται ότι

$$\frac{p_n}{p_{n-1}} = \frac{\binom{m}{n} p^n q^{m-n}}{\binom{m}{n-1} p^{n-1} q^{m-n+1}} = \frac{\dfrac{m!}{n!\,(m-n)!} p}{\dfrac{m!}{(n-1)!\,(m-n+1)!} q} = \frac{p(m-n+1)}{nq}$$

$$= -\frac{p}{q} + \frac{p(m+1)}{qn} \Leftrightarrow p_n = \left(-\frac{p}{q} + \frac{p(m+1)}{nq}\right) p_{n-1}$$

οπότε $a = -\frac{p}{q}$ και $b = (m+1)\,\frac{p}{q}$. Ακόμη, $p_0 = q^m$.

- **Η αρνητική διωνυμική κατανομή**

Αν η τ.μ. $N \sim NB_0(r, p)$ ($R_N = \{0,1,2, \dots\}$) έπεται ότι

$$\frac{p_n}{p_{n-1}} = \frac{\binom{r+n-1}{n} p^r q^n}{\binom{r+(n-1)-1}{n-1} p^r q^{n-1}} = \frac{\frac{(r+n-1)!}{n!\,(r+n-1-n)!} q}{\frac{(r+n-2)!}{(n-1)!\,(r+n-2-n+1)!}} = \frac{r+n-1}{n} q$$

$$= q + \frac{r-1}{n} q \quad \Leftrightarrow p_n = \left(q + \frac{r-1}{n} q\right) p_{n-1}$$

οπότε $a = q$ και $b = (r-1)q$. Ακόμη, $p_0 = p^r$.

- **Η γεωμετρική κατανομή**

Η γεωμετρική κατανομή $G_0(p)$ ($R_N = \{0,1,2,\dots\}$) είναι ειδική περίπτωση της αρνητικής διωνυμικής για $r = 1$, συνεπώς από τα αποτελέσματα που βρήκαμε προηγουμένως προκύπτει ότι $a = q$, $b = 0$ και $p_0 = p$.

Τα παραπάνω αποτελέσματα δίνονται συνοπτικά στον ακόλουθο πίνακα.

Πίνακας 1.7.1. Η οικογένεια κατανομών του Panjer.

ΚΑΤΑΝΟΜΗ	a	b	p_0	p_n
$P(\lambda)$	0	λ	$e^{-\lambda}$	$e^{-\lambda}\frac{\lambda^n}{n!}$
$B(m,p)$	$-\frac{p}{q}$	$\frac{(m+1)p}{q}$	q^m	$\binom{m}{n} p^n(1-p)^{m-n}$
$NB_0(r,p)$	q	$(r-1)q$	p^r	$\binom{r+n-1}{n}(1-p)^n p^r$
$G_0(p)$	q	0	p	$(1-p)^n p$

Οι Sundt & Jewell (1981) έδειξαν, ότι οι παραπάνω τέσσερις κατανομές (μαζί με την εκφυλισμένη κατανομή) είναι οι μοναδικές που ικανοποιούν τον αναδρομικό τύπο του Panjer, δηλαδή είναι τα μοναδικά μέλη της οικογένειας $R(a,b,0)$.

1.9.3 Ιδιότητες της κλάσης κατανομών $\boldsymbol{R(a,b,0)}$

Αν η τ.μ. N ανήκει στην κλάση κατανομών $(a,b,0)$, τότε

- $M'_N(t) = \frac{(a+b)e^t}{1-ae^t} M_N(t)$.

- $P_N'(t) = \frac{a+b}{1-at} P_N(t)$.
- $E(N) = \frac{a+b}{1-a}$
- $E(N^2) = \frac{(a+b)(1+a+b)}{(1-a)^2}$
- $Var(N) = \frac{a+b}{(1-a)^2}$
- $\gamma = \frac{\alpha+1}{\sqrt{a+b}}$

Αν τώρα χρησιμοποιήσουμε τα αποτελέσματα που αποδείξαμε παραπάνω, μπορούμε να υπολογίσουμε δύο ακόμα βασικές ποσότητες που είναι η σχετική διακύμανση και ο συντελεστής μεταβλητότητας.

Σχετική διακύμανση καλείται το τετράγωνο του συντελεστή μεταβλητότητας. Η ποσότητα αυτή μας δίνει τη διακύμανση των τιμών ως προς το μέσο και είναι ίση με:

$$CV^2 = \frac{Var(N)}{E^2(N)}.$$

Συνεπώς για την κλάση κατανομών $(a, b, 0)$ θα έχουμε ότι

$$CV^2 = \frac{\frac{a+b}{(1-a)^2}}{\frac{(a+b)^2}{(1-a)^2}} = \frac{1}{a+b}.$$

Επίσης ο συντελεστής μεταβλητότητας είναι

$$D(N) = \frac{Var(N)}{E(N)} = \frac{\frac{a+b}{(1-a)^2}}{\frac{a+b}{1-a}} = 1-a.$$

Σημειώνεται ότι ο συντελεστής μεταβλητότητας είναι πάντα μεγαλύτερος του μηδενός διότι το α παίρνει τιμές μικρότερες της μονάδας ($a < 1$).

Επίσης, για τις σύνθετες κατανομές, συμπεραίνουμε ότι:

- Αν η τυχαία μεταβλητή N ανήκει στην κλάση $R(a, b, 0)$, τότε για τη μέση τιμή και τη διακύμανση της τυχαίας μεταβλητής $S_N = S$ έχουμε ότι:

$$E(S) = \frac{a+b}{1-a} E(X)$$

$$Var(S) = \frac{a+b}{1-a} Var(X) + \frac{a+b}{(1-a)^2} E^2(X).$$

Έτσι, προκύπτουν τα ακόλουθα αναλυτικά αποτελέσματα:

- Αν η τυχαία μεταβλητή $N \sim P(\lambda)$, δηλαδή $a = 0$ και $b = \lambda$, τότε

$$P'_N(t) = \lambda P_N(t)$$
$$E(N) = Var(N) = \lambda$$
$$E(S) = \lambda E(X)$$
$$Var(S) = \lambda E(X^2)$$

- Αν η τυχαία μεταβλητή $N \sim B(m,p)$, δηλαδή $a = -\frac{p}{q}$ και $b = \frac{(m+1)p}{q}$, τότε

$$P'_N(t) = \frac{mp}{q+pt} P_N(t)$$
$$E(N) = mp, \quad Var(N) = mpq$$
$$E(S) = mpE(Y)$$
$$Var(S) = mpVar(X) + mpqE^2(X) = mpE(X^2) - mp^2E^2(X).$$

- Αν η τυχαία μεταβλητή $N \sim NB_0(r,p)$, δηλαδή $a = q$ και $b = (r-1)q$, τότε

$$P'_N(t) = \frac{rq}{1-qt} P_N(t)$$
$$E(N) = \frac{rq}{p}, \quad Var(N) = \frac{rq}{p^2},$$
$$E(S) = \frac{rq}{p} E(X),$$
$$Var(S) = \frac{rq}{p} Var(X) + \frac{rq}{p^2} E^2(Y) = \frac{rq}{p} E(X^2) + \frac{rq^2}{p^2} E^2(X).$$

1.10 Κλάση κατανομών $R(a, b, 1)$

1.10.1 Ορισμός

Η διακριτή τυχαία μεταβλητή N, με $R_N = \{0,1,2,\ldots\}$, ανήκει στην κλάση κατανομών $R(a,b,1)$, αν η συνάρτηση πιθανότητάς της $p_n = \Pr(N = n)$, ικανοποιεί την αναδρομική σχέση

$$p_n = p(n) = \left(a + \frac{b}{n}\right) p_{n-1}, \qquad n = 2,3,\ldots$$

Ο παραπάνω τύπος δόθηκε για πρώτη φορά από τους Sundt & Jewell (1981).

Στην οικογένεια κατανομών $R(a, b, 1)$ ανήκουν όλες οι κατανομές που ανήκουν στην κλάση $R(a, b, 0)$ και επιπλέον ανήκουν οι περικομμένες αλλά και οι τροποποιημένες στο μηδέν Poisson, Διωνυμική, Αρνητική Διωνυμική, Γεωμετρική και η λογαριθμική κατανομή (Klugman et al, 2004).

1.10.2 Μέλη της κλάσης $\boldsymbol{R(a,b,1)}$

Από τους ορισμούς των δύο κλάσεων κατανομών $R(a,b,0)$ και $R(a,b,1)$, παρατηρούμε ότι τα μέλη των οικογενειών, των κατανομών αυτών, έχουν συναρτήσεις πιθανότητας που ικανοποιούν την ίδια αναδρομική σχέση. Όμως, για την κλάση $R(a,b,0)$ η αναδρομική σχέση ικανοποιείται για $n = 1,2,3, \ldots$, ενώ για την κλάση $R(a,b,1)$ η αναδρομική σχέση ικανοποιείται για $n = 2,3,4, \ldots$.
Συνεπώς, οι παράμετροι a και b της κλάσης κατανομών $R(a,b,1)$ δεν καθορίζουν πλήρως την κατανομή της N, αφού τώρα η πιθανότητα p_0 μπορεί να οριστεί ελεύθερα. Επίσης, η κλάση κατανομών $R(a,b,1)$ περιέχει ως μέλη της όλες τις κατανομές της κλάσης $R(a,b,0)$, δηλαδή περιέχει ως μέλη της τις κατανομές

- $Poisson$,
- $Διωνυμική$,
- $Αρνητική\ Διωνυμική$,
- $Γεωμετρική$.

Επιπλέον, περιέχει ως μέλη της όλες τις περικομμένες στο μηδέν κατανομές (zero-truncated) και όλες τις τροποποιημένες στο μηδέν κατανομές (zero-modified) που αντιστοιχούν στην κλάση κατανομών $R(a,b,0)$. Άρα, η κλάση κατανομών $R(a,b,1)$ περιέχει ως μέλη της και τις ακόλουθες κατανομές:

- $Zero - truncated\ Διωνυμική$,
- $Zero\ - truncated\ Poisson$,
- $Zero\ - truncated\ Αρνητική\ Διωνυμική$,
- $Zero\ - truncated\ Γεωμετρική$.
- $Zero\ - modified\ Διωνυμική$,
- $Zero\ - modified\ Poisson$,
- $Zero\ - modified\ Αρνητική\ Διωνυμική$,
- $Zero - modified\ Γεωμετρική$.

Σημειώνουμε, ότι για μια διακριτή τ.μ. X με συνάρτηση πιθανότητας

$$p_k = P(X = k),\ \ k = 0, 1, 2, \ldots.$$

η αντίστοιχη zero-truncated κατανομή (κατανομή αποκομμένη στο 0), $X^{(T)}$, έχει συνάρτηση πιθανότητας που ικανοποιεί τη σχέση

$$p_k^{(T)} = P\left(X^{(T)} = k\right) = \begin{cases} 0, & k = 0 \\ \dfrac{1}{1-p_0} p_k, & k = 1, 2, \ldots \end{cases}$$

Η αντίστοιχη zero-modified κατανομή (κατανομή τροποποιημένη στο 0), $X^{(M)}$, έχει συνάρτηση πιθανότητας, που ικανοποιεί τη σχέση

$$p_k^{(M)} = P\left(X^{(M)} = k\right) = \begin{cases} p_0^{(M)}, & k = 0 \\ \dfrac{1-p_0^{(M)}}{1-p_0} p_k = (1-p_0^{(M)}) p_k^{(T)}, & k = 1, 2, \ldots \end{cases}$$

Αν $P(t)$, $P^{(T)}(t)$ και $P^{(M)}(t)$ είναι η πιθανογεννήτρια της τυχαίας μεταβλητής X, $X^{(T)}$ και $X^{(M)}$, αντίστοιχα, τότε

$$P^{(T)}(t) = \frac{P(t) - p_0}{1 - p_0}$$

και

$$P^{(M)}(t) = \left(1 - \frac{1-p_0^{(M)}}{1-p_0}\right) + \frac{1-p_0^{(M)}}{1-p_0} P(t).$$

Ακολουθεί ο συνοπτικός πίνακας με τα μέλη της κλάσης $R(a, b, 1)$.

Πίνακας 1.2.:Πίνακας με τις κατανομές που ανήκουν στην κλάση κατανομών $\boldsymbol{R(a,b,1)}$

Κατανομή	$\boldsymbol{p_0}$	$\boldsymbol{p_1}$	$\boldsymbol{a}$	$\boldsymbol{b}$
Geometric $[G_0(p)]$ $p_n = (1-q)q^n$	$1-q$	$(1-q)q$	q	0
ZT Geometric	0	$1-q$	q	0
ZM Geometric	p_0^M	$(1-p_0^M)(1-q)$	q	0
Negative binomial $[Nb_0(r,p)]$ $p_n = \binom{n+r-1}{n} \times (1-q)^r q^n$	$(1-q)^r$	$rq(1-q)^r$	q	$(r-1)q$
ZT Negative binomial	0	$\dfrac{rq(1-q)^r}{1-(1-q)^r}$	q	$(r-1)q$

or ETNB				
ZM Negative binomial or ZM ETNB	p_0^M	$(1-p_0^M)\dfrac{rq(1-q)^r}{1-(1-q)^r}$	q	$(r-1)q$
Poisson $[P(\lambda)]$ $p_n = e^{-\lambda}\dfrac{\lambda^n}{n!}$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	0	λ
ZT Poisson	0	$\dfrac{\lambda}{e^{\lambda}-1}$	0	λ
ZM Poisson	p_0^M	$(1-p_0^M)\dfrac{\lambda}{e^{\lambda}-1}$	0	λ
Binomial $[B(m,p)]$ $p_n = \binom{m}{n}(1-q)^n q^{m-n}$	q^m	$m(1-q)q^{m-1}$	$-\dfrac{(1-q)}{q}$	$\dfrac{(m+1)(1-q)}{q}$
ZT Binomial	0	$\dfrac{m(1-q)q^{m-1}}{1-q^m}$	$-\dfrac{(1-q)}{q}$	$\dfrac{(m+1)(1-q)}{q}$
ZM Binomial	p_0^M	$(1-p_0^M)\times \times\dfrac{m(1-q)q^{m-1}}{1-q^m}$	$-\dfrac{(1-q)}{q}$	$\dfrac{(m+1)(1-q)}{q}$
Logarithmic [LS(q)] $p_n = -\dfrac{q^n}{nln(1-q)}$	0	$-\dfrac{q}{ln(1-q)}$	q	$-q$
ZM Logarithmic	p_0^M	$-(1-p_0^M)\dfrac{q}{ln(1-q)}$	q	$-q$

1.10.3 Ιδιότητες της κλάσης κατανομών $\boldsymbol{R(a,b,1)}$

Αν η τ.μ. N ανήκει στην κλάση κατανομών $(a,b,0)$, τότε

- $P_N'(t) = \frac{(a+b)P_N(t)+p_1-(a+b)p_0}{1-at}$
- $E(N) = \frac{(a+b)(1-p_0)+p_1}{1-a}$
- $Var(N) = \frac{[1-p_1+(a+b)p_0][a+b+p_1-(a+b)p_0]}{(1-a)^2}$

Επίσης, αν η τυχαία μεταβλητή N ανήκει στην κλάση $R(a,b,1)$ τότε για τη μέση τιμή και τη διακύμανση της τυχαίας μεταβλητής $S_N = S$, έχουμε

$$E(S) = \frac{(a+b)(1-p_0)+p_1}{1-a}E(X)$$

$$Var(S) = \frac{(a+b)(1-p_0)+p_1}{1-a}Var(X) + \frac{[1-p_1+(a+b)p_0][a+b+p_1-(a+b)p_0]}{(1-a)^2}E^2(X).$$

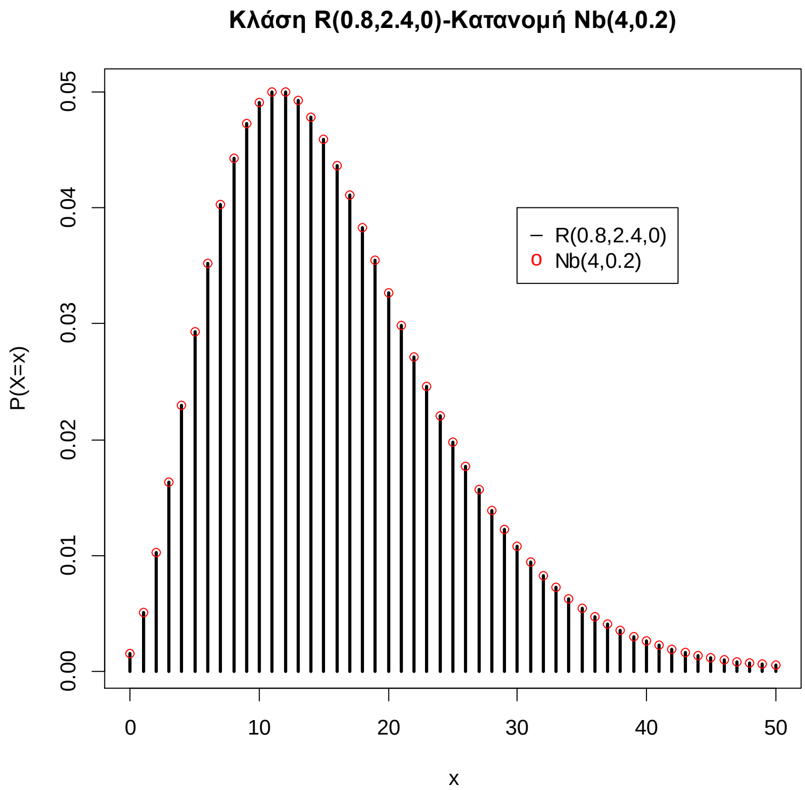
Παρακάτω, ακολουθούν δύο παραδείγματα της κλάσης κατανομών $R(a,b,0)$ και $R(a,b,1)$.

Παράδειγμα 1.1

Να οριστεί η συνάρτηση `pmfnbinom` που να υπολογίζει τη συνάρτηση πιθανότητας της κατανομής $R(0.8,2.4,0)$ με $p_0 = 0.2^4$. Να δοθεί η γραφική παράσταση της συνάρτησης πιθανότητας για $n = 0,1,\ldots,50$. Στο ίδιο σχήμα να γίνει γραφική παράσταση της συνάρτησης πιθανότητας της κατανομής $Nb_0(4,0.2)$.

Λύση

```
###Paradeigma1
p<-0.2
q<-0.8
r<-4
p0<-p^r
a<-q
b<-(r-1)*q
pmfnbinom<-function(n) {
   if(n==0) return(p0)
   return((a+b/n)*pmfnbinom(n-1))
}
y<-sapply(0:50,pmfnbinom)
plot(0:50,y,type="h",lwd=3,xlab="x",ylab="P(X=x)",main="K
λάση R(0.8,2.4,0)-Κατανομή Nb(4,0.2)")
points(0:50,dnbinom(0:50,4,0.2),col="red")
legend(30,0.04,c("R(0.8,2.4,0)","Nb(4,0.2)"),pch=c("_","o
"),col=c("black","red"))
```



Παράδειγμα 1.2

Για $r = -0.5$ και $q = 0.5$ να βρεθεί η συνάρτηση πιθανότητας της εκτεταμένης αρνητικής διωνυμικής κατανομής ETNB για $n = 0, 1, \dots, 8$ καθώς επίσης και της αντίστοιχης τροποποιημένης στο 0 κατανομής ZM.ETNB με $p_0{}^{(M)} = 0.6$ με χρήση των συναρτήσεων `pmfETnbinom` και `pmfEMnbinom`, αντίστοιχα.

Λύση

```
#Paradeigma2
q<-0.5
p<-1-q
r<- -0.5
a<-q
b<-(r-1)*q
p0<-0
p1<-(r*q*p^r)/(1-p^r)
pmfETnbinom<-function(n) {
   if(n==0) return(p0)
   if(n==1) return(p1)
```

```
  return((a+b/n)*pmfETnbinom(n-1))
}
n<-0:8
ETNB<-sapply(n,pmfETnbinom)
ETNB
pM0<-0.6
pM1<-(1-pM0)*p1
pmfEMnbinom<-function(n) {
  if(n==0) return(pM0)
  if(n==1) return(pM1)
  return((a+b/n)*pmfEMnbinom(n-1))
}
ZM.ETNB<-sapply(n,pmfEMnbinom)
ZM.ETNB
m<-cbind(n,ETNB,ZM.ETNB)
rownames(m)<-rep("",nrow(m))
round(m,digits=5)

n     ETNB  ZM.ETNB
 0 0.00000 0.60000
 1 0.85355 0.34142
 2 0.10669 0.04268
 3 0.02667 0.01067
 4 0.00834 0.00333
 5 0.00292 0.00117
 6 0.00109 0.00044
 7 0.00043 0.00017
 8 0.00017 0.00007
```

2 Κεφάλαιο: Θεωρία κινδύνου και το πακέτο actuar

Το συλλογικό πρότυπο που εξετάζουμε στο παρόν κεφάλαιο είναι ένα από τα σημαντικότερα μοντέλα στη θεωρία κινδύνου, αλλά και την αναλογιστική επιστήμη γενικότερα.

Βασικό αντικείμενο μελέτης αποτελεί η κατανομή των συνολικών απαιτήσεων που φτάνουν σε μία ασφαλιστική εταιρεία, σε σχέση με ένα συγκεκριμένο χαρτοφυλάκιο της εταιρείας, μέσα σε ένα συγκεκριμένο χρονικό διάστημα. Η κεντρική θέση που κατέχει το συλλογικό πρότυπο στα αναλογιστικά μαθηματικά και η ανάγκη για την ανάλυση σε βάθος της κατανομής των συνολικών απαιτήσεων έχουν οδηγήσει σε μία σειρά από αποτελέσματα τα οποία και θα εξετάσουμε στο κεφάλαιο αυτό.

Επίσης, γίνεται έρευνα σχετικά με κάποιες μεθόδους οι οποίες μπορούν να χρησιμοποιηθούν για την προσέγγιση των συνολικών αποζημιώσεων στην περίπτωση που οι ατομικές ζημιές περιγράφονται από μια συνεχή τυχαία μεταβλητή (διακριτοποίηση).

Για όλα τα παραπάνω θα δοθούν αντίστοιχα παραδείγματα υπολογισμού της κατανομής των συνολικών αποζημιώσεων με ένα από τα σημαντικότερα πακέτα με εφαρμογή στην θεωρία κινδύνου στο περιβάλλον της γλώσσας R, το πακέτο `actuar`.

2.1 Συνελίξεις

Στη θεωρία πιθανοτήτων, αλλά και στα αναλογιστικά μαθηματικά, ασχολούμαστε συχνά με προβλήματα που αφορούν αθροίσματα ανεξάρτητων τυχαίων μεταβλητών, δηλαδή με αθροίσματα της μορφής

$$S_n = X_1 + X_2 + \cdots + X_n,$$

όπου n είναι ένας γνωστός ακέραιος, και X_i $(1 \leq i \leq n)$ είναι ανεξάρτητες τυχαίες μεταβλητές. Ήδη στο Κεφάλαιο 1 δόθηκε ο μαθηματικός όρος συνέλιξη που σχετίζεται με το παραπάνω άθροισμα.

Είναι γνωστό ότι η ροπογεννήτρια συνάρτηση $M_{S_n}(t)$ της τυχαίας μεταβλητής S_n, δίνεται από τον τύπο

$$M_{S_n}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t)$$

όπου $M_{X_i}(t)$ η ροπογεννήτρια συνάρτηση της τυχαίας μεταβλητής X_i $(1 \leq i \leq n)$. Αντίστοιχος τύπος ισχύει και για τις αντίστοιχες πιθανογεννήτριες συναρτήσεις (όταν οι τ.μ. X_i $(1 \leq i \leq n)$ είναι διακριτές), δηλαδή

$$P_{S_n}(t) = P_{X_1}(t)P_{X_2}(t) \cdots P_{X_n}(t).$$

Ας εξετάσουμε τώρα, τη περίπτωση του αθροίσματος n ανεξάρτητων και ισόνομων μη αρνητικών, ακέραιων διακριτών τυχαίων μεταβλητών $X_1, X_2, \dots, X_n$ με κοινή συνάρτηση πιθανότητας f, κοινή συνάρτηση κατανομής F, κοινή ροπογεννήτρια συνάρτηση $M_X(s)$, και κοινή πιθανογεννήτρια $P_X(s)$.

Έστω

$$G_{S_n}(x) = P(S_n \leq x) = F^{*n}(x), \quad x = 0, 1, \dots \, .$$

Τότε

$$F^{*n}(x) = \begin{cases} F(x), & n = 1 \\ \sum_{t=0}^{x} F^{*(n-1)}(x-t)\, f(t), & n = 2, 3, \dots \end{cases}$$

όπου

$$F^{*0}(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Επίσης, έστω

$$g_{S_n}(x) = P(S_n = x) = f^{*n}(x), \quad x = 0, 1, \dots \, .$$

Τότε

$$f^{*n}(x) = \begin{cases} f(x), & n = 1 \\ \sum_{t=0}^{x} f^{*(n-1)}(x-t)\, f(t), & n = 2, 3, \dots \end{cases}$$

όπου

$$f^{*0}(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Επίσης

$$M_{S_n}(t) = E(e^{tS_n}) = [M_X(t)]^n$$

και

$$P_{S_n}(t) = E(t^{S_n}) = [P_X(t)]^n.$$

Στην περίπτωση που οι τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ είναι συνεχείς, ισχύουν όλες οι παραπάνω σχέσεις, μόνο που τα αθροίσματα θα πρέπει να αντικατασταθούν με ολοκληρώματα στον ορισμό των ποσοτήτων $F^{*n}(x)$ και $f^{*n}(x)$ για $n \geq 2$.

Η συνάρτηση `convolve` του πακέτου `stats` υπολογίζει συνελίξεις ακολουθιών/διανυσμάτων.

Παράδειγμα 2.1

Η διακριτή τυχαία μεταβλητή X έχει σύνολο τιμών $R_X = \{1, 2\}$ και συνάρτηση πιθανότητας f τέτοια ώστε

$$P(X = 1) = f(1) = 0.4, \qquad P(X = 2) = f(2) = 0.6.$$

Έστω $X_1, X_2, \ldots, X_{10}$ είναι τυχαίο δείγμα μεγέθους 10 από την παραπάνω κατανομή και έστω

$$Y = X_1 + X_2 + \cdots + X_{10}.$$

Να υπολογιστούν οι πιθανότητες $P(Y = y)$ για $y = 10,11, \ldots ,20$.

Λύση

```
library(stats)
conv <- c(0, 0.4, 0.6)
fx <- c(0, 0.4, 0.6)
n <- 10
for(i in 2:n) {conv <- convolve(fx,rev(conv),type="o")}
m <- cbind(10:20,conv[11:21])
rownames(m) <- rep("",nrow(m))
colnames(m) <- c("y","      P(Y=y)")
round(m, digits=7)


  y       P(Y=y)
 10    0.0001049
 11    0.0015729
 12    0.0106168
 13    0.0424673
 14    0.1114767
 15    0.2006581
 16    0.2508227
 17    0.2149908
 18    0.1209324
```

```
19      0.0403108
20      0.0060466
```

2.2 Διακριτοποίηση μιας συνεχούς τυχαίας μεταβλητής

Όπως αναφέραμε και στο προηγούμενο κεφάλαιο σε αρκετές εφαρμογές χρειάζεται να μετατρέψουμε μία συνεχή τυχαία μεταβλητή σε διακριτή διατηρώντας ωστόσο τις ιδιότητές της σχετικά με την κατανομή της πιθανότητάς της στο σύνολο των τιμών της. Η συνάρτηση `discretize` του πακέτου `actuar` προσφέρει 4 τεχνικές για τη διακριτοποίηση μιας συνεχούς τυχαίας μεταβλητής X.

Έστω λοιπόν $F(x) = F_X(x) = P(X \leq x)$ η συνάρτηση κατανομή μιας συνεχούς τ.μ. X που θέλουμε να διακριτοποιήσουμε στο διάστημα $[a, b]$. Έστω $g_y = P(Y = y)$ η πιθανότητα στο σημείο y της αντίστοιχης διακριτής τ.μ. Y.

2.2.1 Upper Discretization

Με αυτή τη μέθοδο έχουμε ότι στο διάστημα $[a, b]$ η τ.μ. Y παίρνει τις τιμές στο σύνολο $R_{[a,b]} = \{a,\ a+h,\ a+2h,\ \dots,\ a+(k-1)h\}$ όπου $h = (b-a)/k$ (δηλαδή $b = a + kh$). Οι πιθανότητες $g_y = P(Y = y)$ στο σύνολο $R_{[a,b]}$ υπολογίζονται με τους τύπους

$$g_a = P(Y = a) = F_X(a+h) - F_X(a),$$
$$g_{a+h} = P(Y = a+h) = F_X(a+2h) - F_X(a+h),$$
$$g_{a+2h} = P(Y = a+2h) = F_X(a+3h) - F_X(a+2h),$$
$$\dots$$
$$g_{a+(k-1)h} = P(Y = a+(k-1)h) = F_X(a+kh) - F_X(a+(k-1)h).$$

Επομένως ισχύει ο γενικός τύπος

$$g_y = P(Y = y) = F_X(y+h) - F_X(y), \quad y = a,\ a+h,\ a+2h,\ \dots,\ a+(k-1)h.$$

Η συγκεκριμένη μέθοδος εγγυάται ότι

$$P(a < X \leq b) = F_X(b) - F_X(a) = \sum_{y \in R_{[a,b]}} g_y.$$

Επίσης η συνάρτηση κατανομής της τ.μ. Y είναι πάντα πάνω από τη συνάρτηση κατανομής της τ.μ. X.

2.2.2 Lower Discretization

Με αυτή τη μέθοδο έχουμε ότι στο διάστημα $[a,b]$ η τ.μ. Y παίρνει τις τιμές στο σύνολο $R_{[a,b]} = \{a,\ a+h,\ a+2h,\ \dots,\ b\}$ όπου $h = (b-a)/k$ (δηλαδή $b = a + kh$). Οι πιθανότητες $g_y = P(Y = y)$ στο σύνολο $R_{[a,b]}$ υπολογίζονται με τους τύπους

$$
\begin{gathered}
g_a = P(Y = a) = F_X(a), \\
g_{a+h} = P(Y = a+h) = F_X(a+h) - F_X(a), \\
g_{a+2h} = P(Y = a+2h) = F_X(a+2h) - F_X(a+h), \\
\dots \\
g_b = P(Y = b) = F_X(b) - F_X(b-h).
\end{gathered}
$$

Επομένως ισχύει ο γενικός τύπος

$$
g_y = P(Y = y) = \begin{cases} F_X(a), & y = a \\ F_X(y) - F_X(y-h), & y = a+h,\ a+2h,\ \dots,\ b \end{cases}
$$

Η συγκεκριμένη μέθοδος εγγυάται ότι

$$
P(X \le b) = F_X(b) = \sum_{y \in R_{[a,b]}} g_y.
$$

Επίσης η συνάρτηση κατανομής της τ.μ. Y είναι πάντα κάτω από τη συνάρτηση κατανομής της τ.μ. X.

2.2.3 Midpoint or Rounding method

Με αυτή τη μέθοδο έχουμε ότι στο διάστημα $[a,b]$ η τ.μ. Y παίρνει τις τιμές στο σύνολο $R_{[a,b]} = \{a,\ a+h,\ a+2h,\ \dots,\ a+(k-1)h\}$ όπου $h = (b-a)/k$ (δηλαδή $b = a + kh$). Οι πιθανότητες $g_y = P(Y = y)$ στο σύνολο $R_{[a,b]}$ υπολογίζονται με τους τύπους

$$
\begin{gathered}
g_a = P(Y = a) = F_X(a + h/2), \\
g_{a+h} = P(Y = a+h) = F_X(a+h+h/2) - F_X(a+h-h/2), \\
g_{a+2h} = P(Y = a+2h) = F_X(a+2h+h/2) - F_X(a+2h-h/2), \\
\dots
\end{gathered}
$$

$$
\begin{aligned}
g_{a+(k-1)h} &= P(Y = a+(k-1)h) \\
&= F_X(a+(k-1)h+h/2) - F_X(a+(k-1)h-h/2).
\end{aligned}
$$

Επομένως ισχύει ο γενικός τύπος

$$g_y = P(Y = y) = \begin{cases} F_X(a + h/2), & y = a \\ F_X(y + h/2) - F_X(y - h/2), & y = a + h,\ a + 2h,\ \dots,\ a + (k-1)h \end{cases}$$

Η συγκεκριμένη μέθοδος εγγυάται ότι

$$P(X \le b - h/2) = F_X(b - h/2) = \sum_{y \in R_{[a,b]}} g_y.$$

Επίσης η συνάρτηση κατανομής της τ.μ. X περνάει ακριβώς στο μέσο των βημάτων της συνάρτησης κατανομής της τ.μ. Y.

2.2.4 Unbiased method or Mean Preserving

Με αυτή τη μέθοδο έχουμε ότι στο διάστημα $[a, b]$ η τ.μ. Y παίρνει τις τιμές στο σύνολο $R_{[a,b]} = \{a,\ a+h,\ a+2h,\ \dots,\ b\}$ όπου $h = (b-a)/k$ (δηλαδή $b = a + kh$). Οι πιθανότητες $g_y = P(Y = y)$ στο σύνολο $R_{[a,b]}$ υπολογίζονται με τον γενικό τύπο

$$g_y = P(Y = y) = \begin{cases} \dfrac{E(X \wedge a) - E(X \wedge a + h)}{h} + 1 - F(a), & y = a \\[2ex] \dfrac{2E(X \wedge y) - E(X \wedge y - h) - E(X \wedge y + h)}{h}, & a < y < b \\[2ex] \dfrac{E(X \wedge b) - E(X \wedge b - h)}{h} - 1 + F(b) & y = b \end{cases}$$

όπου $X \wedge u = \min(X, u)$.

Η συγκεκριμένη μέθοδος εγγυάται ότι

$$P(a < X \le b) = F_X(b) - F_X(a) = \sum_{y \in R_{[a,b]}} g_y$$

και

$$\sum_{y \in R_{[a,b]}} y g_y = \int_a^b x f_X(x) dx.$$

<u>Παράδειγμα 2.2</u>

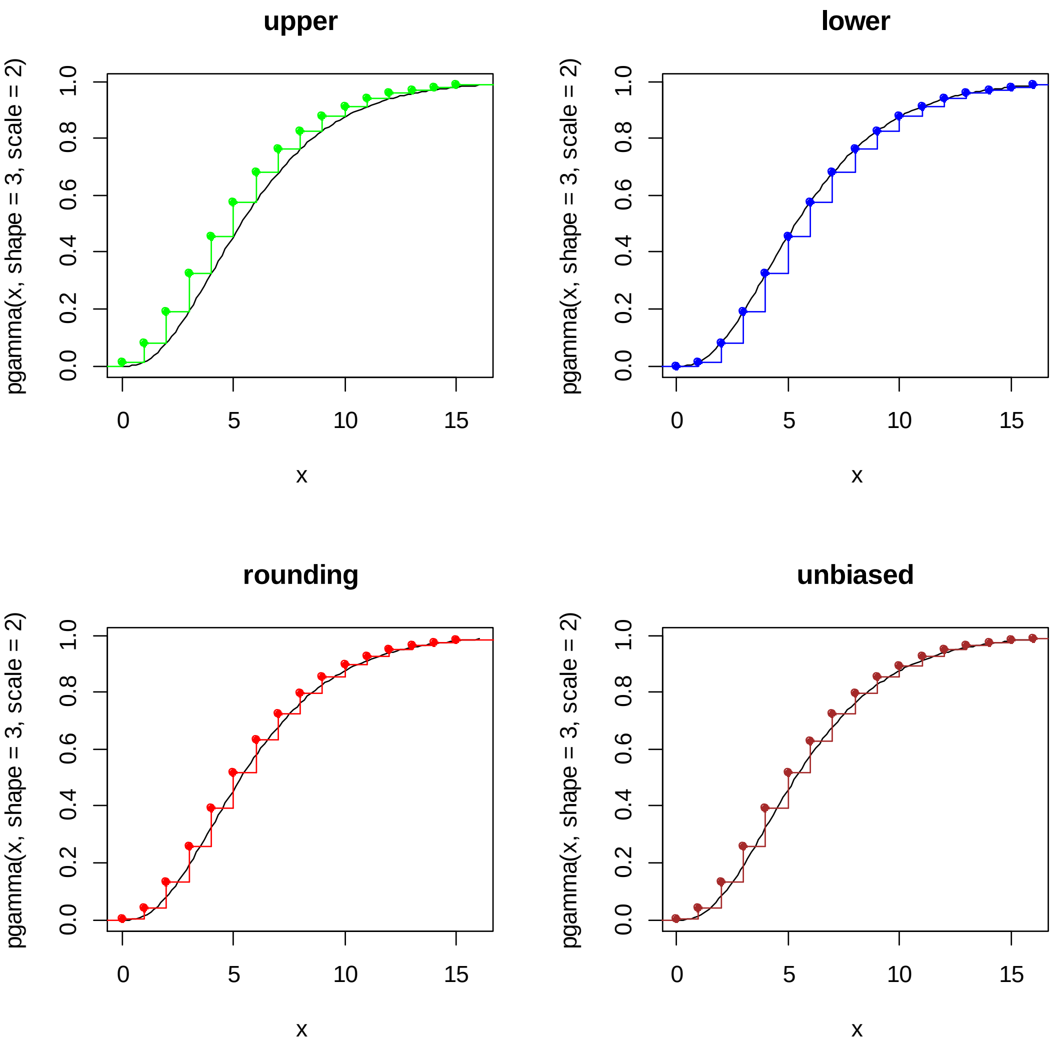
Να διακριτοποιηθεί η κατανομή Γάμμα με παραμέτρους $shape = 3$ και $scale = 2$ (δηλαδή η $Ga(3, 1/2)$) με όλες τις μεθόδους στο διάστημα [0,16] με βήμα $h = 1$ και να δοθούν οι τιμές $g_y = P(Y = y)$ για $y = 0, 1, 2, \cdots, 15$. Να δοθούν επίσης γραφικές παραστάσεις των αντίστοιχων συναρτήσεων κατανομής.

Λύση

```
library(actuar)
a <- 0; b <- 16; h <- 1
fu <- discretize(pgamma(x,shape=3,scale=2), from=a, to=b,
step=h, method="upper")
fl <- discretize(pgamma(x,shape=3,scale=2), from=a, to=b,
step=h, method="lower")
fr <- discretize(pgamma(x,shape=3,scale=2), from=a, to=b,
step=h, method="rounding")
fb <- discretize(pgamma(x,shape=3,scale=2), from=a, to=b,
step=h, method="unbiased", lev = levgamma(x, shape=2,
scale=2))
y <- 0:(b-1)
m <- cbind(y,fu[1:b],fl[1:b],fr[1:b],fb[1:b])
rownames(m) <- rep("",nrow(m))
colnames(m) <- c("y"," g(y)-Upper"," g(y)-Lower"," g(y)-
Rounding", " g(y)-Unbiased")
round(m, digits=5)
```

y	g(y)-Upper	g(y)-Lower	g(y)-Rounding	g(y)-Unbiased
0	0.01439	0.00000	0.00216	0.00388
1	0.06591	0.01439	0.03834	0.03892
2	0.11085	0.06591	0.09103	0.09014
3	0.13217	0.11085	0.12450	0.12350
4	0.13286	0.13217	0.13463	0.13393
5	0.12062	0.13286	0.12788	0.12750
6	0.10234	0.12062	0.11189	0.11176
7	0.08274	0.10234	0.09250	0.09251
8	0.06453	0.08274	0.07336	0.07345
9	0.04893	0.06453	0.05636	0.05648
10	0.03628	0.04893	0.04223	0.04236
11	0.02641	0.03628	0.03102	0.03112
12	0.01893	0.02641	0.02240	0.02249

```
13      0.01340     0.01893       0.01595       0.01602
14      0.00938     0.01340       0.01123       0.01128
15      0.00650     0.00938       0.00782       0.00786

# Συναρτήσεις κατανομής των X, Y στο διάστημα [a,b]
par(mfrow=c(2,2))
curve(pgamma(x, shape=3, scale =2), xlim = c(0, b),
main="upper")
x <- seq(a,b,h)
plot(stepfun(head(x, -1), diffinv(fu)), pch = 19, add =
TRUE, col="green")
curve(pgamma(x, shape=3, scale =2), xlim = c(0, b),
main="lower")
plot(stepfun(x, diffinv(fl)), pch = 19, add = TRUE,
col="blue")
curve(pgamma(x, shape=3, scale =2), xlim = c(0, b),
main="rounding")
plot(stepfun(head(x, -1), diffinv(fr)), pch = 19, add =
TRUE, col="red")
curve(pgamma(x, shape=3, scale =2), xlim = c(0, b),
main="unbiased")
plot(stepfun(x, diffinv(fb)), pch = 19, add = TRUE,
col="brown")
```



2.3 Το μοντέλο συλλογικού κινδύνου

Η γενική μορφή του μοντέλου συλλογικού κινδύνου περιγράφεται από τη σύνθετη κατανομή

$$S_N = S = \begin{cases} X_1 + X_2 + \cdots + X_N, & N = 1,2, \ldots \\ 0, & N = 0 \end{cases}$$

όπου:

(i) η τυχαία μεταβλητή X_i παριστάνει το ύψος της i ατομικής ζημιάς (δεν μπορεί να πάρει αρνητικές τιμές), με συνάρτηση πιθανότητας/πυκνότητας f και συνάρτηση κατανομής F

(ii) η τυχαία μεταβλητή N, παριστάνει το πλήθους των ζημιών (κινδύνων) ή και απαιτήσεων/αξιώσεων με συνάρτηση πιθανότητας p

(iii) η τυχαία μεταβλητή S_N, παριστάνει το ύψος της συνολικής ζημιάς με συνάρτηση πιθανότητας/πυκνότητας g και συνάρτηση κατανομής G.

Θα χρησιμοποιήσουμε τους ακόλουθους συμβολισμούς:

- $F(x) = P\,(X \leq x)$ η συνάρτηση κατανομής της X,
- $f(x) = P(X = x)$ η συνάρτηση πιθανότητας (ή ανάλογα πυκνότητας) της X,
- $G(x) = P(S \leq x)$ η συνάρτηση κατανομής της S_N,
- $g(x) = P(S = x)$ η συνάρτηση πιθανότητας (ή ανάλογα πυκνότητας) της S_N,

Μερικές φορές γράφουμε απλώς $S = \sum_{i=1}^{N} X_i$ και είναι σαφές ότι $S = 0$ αν $N = 0$, δηλαδή αν δεν έρθει καμία απαίτηση, το ποσό που θα καταβάλει η εταιρεία είναι μηδέν.

Ειδικότερα για την τυχαία μεταβλητή N, που είναι πάντοτε μια μη-αρνητική ακέραια (διακριτή) τυχαία μεταβλητή, συμβολίζουμε τη συνάρτηση πιθανότητάς ως

$p_n = p(n) = P(N = n), \quad n = 0,1, \dots .$

2.3.1 Συνάρτηση κατανομής της *S*

Η συνάρτηση κατανομής της τ.μ. S ικανοποιεί τη σχέση

$$G(x) = P(S \leq x) = \sum_{n=0}^{\infty} p_n\, F^{*n}(x), \qquad x \geq 0,$$

όπου $F^{*n}(x) = P(X_1 + X_2 + \cdots + X_n \leq x)$, $n \geq 1$. Υπενθυμίζουμε ότι

$$F^{*n}(x) = \begin{cases} \displaystyle\int_0^x F^{*(n-1)}(x-y) f(y) dy, & X \text{ συνεχής τ.μ.} \\ \displaystyle\sum_{y=0}^{x} F^{*(n-1)}(x-y)\, f(y)\,, & X \text{ διακριτή τ.μ.} \end{cases}$$

όπου

$$F^{*0}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases}$$

2.3.2 Κατανομή της S

Αν η τυχαία μεταβλητή X είναι διακριτή τότε και η τυχαία μεταβλητή S είναι διακριτή με συνάρτηση πιθανότητας που δίνεται από τον τύπο

$$g(x) = P(S = x) = \sum_{n=0}^{\infty} p_n f^{*n}(x), \quad x = 0,1,2,\dots$$

όπου

$$f^{*n}(x) = \begin{cases} f(x), & n = 1 \\ \sum_{y=0}^{x} f^{*(n-1)}(x-y)\, f(y), & n = 2,3,\dots \end{cases}$$

και

$$f^{*0}(x) = \begin{cases} 1, & x = 0, \\ 0, & x \neq 0. \end{cases}$$

Ειδικότερα από τον παραπάνω τύπο της $g(x)$ προκύπτει ότι

$$g(0) = P_N(f(0))$$

όπου $P_N(\cdot)$ είναι η πιθανογεννήτρια της τ.μ. N.

Αν τη τυχαία μεταβλητή X είναι συνεχής, τότε η τυχαία μεταβλητή S είναι τυχαία μεταβλητή μικτού τύπου με συνεχές μέρος στο διάστημα $(0, \infty)$ που δίνεται από τον τύπο

$$g(x) = \sum_{n=1}^{\infty} p_n f^{*n}(x), \quad x > 0$$

όπου

$$f^{*n}(x) = \begin{cases} f(x), & n = 1 \\ \int_0^x f^{*(n-1)}(x-y) f(y) dy, & n = 2,3,\dots \end{cases}$$

και

$$f^{*0}(x) = \begin{cases} 1, & x = 0, \\ 0, & x \neq 0. \end{cases}$$

Το διακριτό μέρος έχει πιθανότητα μόνο στο 0 ίση με

$$g(0) = P(N = 0) = p_0.$$

Φυσικά αν $p_0 = P(N = 0) = 0$, τότε προφανώς η τυχαία μεταβλητή S είναι μια συνήθης συνεχής τυχαία μεταβλητή.

2.3.3 Ιδιότητες της S

Κατά τα συνήθη ας συμβολίσουμε με $M_Y(t)$και $P_Y(t)$ τη ροπογεννήτρια και την πιθανογεννήτρια συνάρτηση μιας τυχαίας μεταβλητής Y (η $P_Y(t)$ προφανώς υπάρχει μόνο αν η τ.μ. Y είναι διακριτή). Τότε

- $M_S(t) = M_N[\log M_X(t)]$.
- $M_S(t) = P_N[M_X(t)]$.
- $P_S(t) = P_N[P_X(t)]$.

Επίσης για τη μέση τιμή, τη διακύμανση και για τις ροπές 3ης τάξης της S έχουμε ότι

- $E(S) = E(N)E(X)$
- $Var(S) = E(N)Var(X) + Var(N)E^2(X)$.
- $E(S^3) = E(N^3)E^3(X) + 3E(N^2)E(X)Var(X) + E(N)E\left[\left(X - E(X)\right)^3\right]$,
- $E\left[\left(S - E(S)\right)^3\right] = E[\left(N - E(N)\right)^3]E^3(X) + E[\left(X - E(X)\right)^3]E(N) + 3Var(N)E(X)Var(X)$.

2.4 Υπολογισμός της κατανομής της $\boldsymbol{S}$ με ειδικούς τύπους

Ο ακριβής υπολογισμός της κατανομής της S μέσω κλειστών τύπων συνήθως δεν είναι εφικτός εκτός μερικών απλών περιπτώσεων. Ακόμη όμως και σε αυτές τις περιπτώσεις οι τύποι τις περισσότερες φορές δεν είναι ελκυστικοί. Στη συνέχεια, ενδεικτικά, αναφέρουμε τρεις περιπτώσεις στις οποίες η κατανομή του ύψους των ζημιών είναι μια αρκετά απλή συνεχής κατανομή, όπως είναι η εκθετική.

Γεωμετρική - Εκθετική

Έστω ότι το πλήθος των ζημιών είναι τ.μ. και ακολουθεί τη γεωμετρική κατανομή $G_0(p)$ με συνάρτηση πιθανότητας

$$p_n = P(N = n) = pq^n, \qquad n = 0, 1, 2, \ldots.$$

Το ύψος των αποζημιώσεων είναι τ.μ. για την οποία ισχύει $X \sim Exp(\lambda)$. Τότε

$$G(x) = 1 - qe^{-p\lambda x}, \qquad x \geq 0.$$

Η κατανομή της τ.μ. S είναι μικτού τύπου, διότι έχει μάζα πιθανότητας στο σημείο μηδέν και είναι συνεχής στο διάστημα $(0, \infty)$ και δίνεται από την σχέση

$$g(x) = \begin{cases} p, & x = 0 \\ qp\lambda e^{-p\lambda x}, & x > 0 \end{cases}$$

Διωνυμική - Εκθετική

Έστω ότι το πλήθος των ζημιών είναι τ.μ. και ακολουθεί τη διωνυμική κατανομή, $B(m,p)$ με συνάρτηση πιθανότητας

$$p_n = P(N = n) = \binom{m}{n} p^n q^{m-n}, \quad n = 0,1,2,\ldots,m\,.$$

Το ύψος των αποζημιώσεων είναι τ.μ. για την οποία ισχύει $X \sim Exp(\lambda)$. Τότε

$$G(x) = 1 - e^{-\lambda x} \sum_{n=1}^{m} \binom{m}{n} p^n q^{m-n} \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!}, \quad x \geq 0.$$

Η κατανομή της τ.μ. S είναι μικτού τύπου, διότι έχει μάζα πιθανότητας στο σημείο μηδέν και είναι συνεχής στο διάστημα $(0,\infty)$ και δίνεται από την σχέση

$$g(x) = \begin{cases} q^m, & x = 0 \\ e^{-\lambda x} \sum_{n=1}^{m} \binom{m}{n} p^n q^{m-n} \frac{\lambda^n}{(n-1)!} x^{n-1}, & x > 0 \end{cases}$$

Αρνητική Διωνυμική – Εκθετική

Έστω ότι το πλήθος των ζημιών είναι τ.μ. και ακολουθεί τη αρνητική διωνυμική κατανομή $Nb_0(r,p)$ με συνάρτηση πιθανότητας

$$p_n = P(N = n) = \binom{r+n-1}{n} p^r q^n, \quad x = 0, 1, 2, \ldots$$

Το ύψος των αποζημιώσεων είναι τ.μ. για την οποία ισχύει $X \sim Exp(\lambda)$. Τότε

$$G(x) = 1 - e^{-p\lambda x} \sum_{j=0}^{r-1} \frac{(p\lambda x)^j}{j!} \sum_{n=j+1}^{r} \binom{r}{n} q^n p^{r-n}, \quad x \geq 0.$$

Η κατανομή της τ.μ. S είναι μικτού τύπου, διότι έχει μάζα πιθανότητας στο σημείο μηδέν και είναι συνεχής στο διάστημα $(0,\infty)$ και δίνεται από την σχέση

$$g(x) = \begin{cases} p^r, & x = 0 \\ \sum_{t=1}^{r} \binom{r}{t} \frac{p^r (q\lambda)^t}{(t-1)!} x^{t-1} e^{-p\lambda x}, & x > 0. \end{cases}$$

2.5 Υπολογισμός της κατανομής της $\boldsymbol{S}$ με τη συνάρτηση aggregateDist του πακέτου actuar

Ο υπολογισμός της συνάρτησης πιθανότητας αλλά και της συνάρτησης κατανομής της S είναι αρκετά χρονοβόρος και επίπονος λόγω των αρκετών πράξεων αφού οι σχέσεις είναι συνήθως αναδρομικές. Ιδιαίτερα στην περίπτωση που η κατανομή των αποζημιώσεων είναι συνεχής τυχαία μεταβλητή, ο υπολογισμός της κατανομής της S είναι ακόμη πιο δύσκολος. Υπάρχουν βεβαίως μερικές περιπτώσεις που η κατανομή της S δίνεται από αναλυτικό τύπο, αλλά αυτή η περίπτωση είναι η εξαίρεση και όχι ο κανόνας. Για τον λόγο αυτό στη συνέχεια θα δειχθεί πώς μπορεί να υπολογιστεί η κατανομή της S με τη χρήση της γλώσσας προγραμματισμού R.

Αρχικά η συνάρτηση που θα χρησιμοποιηθεί είναι η `aggregateDist`, η οποία μας δίνει την αθροιστική συνάρτηση κατανομής της τ.μ. S και βρίσκεται στο πακέτο `actuar`. Η συνάρτηση αυτή παρέχει πέντε διαφορετικές μεθοδολογίες για τον υπολογισμό της κατανομής του ύψους (μεγέθους) των συνολικών ζημιών σε ένα συγκεκριμένο χρονικό διάστημα.

2.5.1 Μέθοδος Convolution

Εφαρμόζεται όταν τα μεγέθη των αποζημιώσεων περιγράφονται από μια διακριτή τ.μ. X με σ.κ $F(x)$ και συνάρτηση πιθανότητας $f(x)$. Στο όρισμα `model.freq` της συνάρτησης `aggregateDist` δίνουμε το διάνυσμα των πιθανοτήτων της τ.μ. N και στο όρισμα `model.sev` δίνουμε το διάνυσμα των πιθανοτήτων της τ.μ. X. Η συνάρτηση `aggregateDist` δίνει ως αποτέλεσμα τη σ.κ. της S την οποία υπολογίζει με χρήση του γνωστού τύπου

$$G(x) = P(S \leq x) = \sum_{n=0}^{\infty} p_n F^{*n}(x), \quad x \geq 0,$$

χρησιμοποιώντας την συνάρτηση `convolve`.

2.5.2 Μέθοδος Recursive

Εφαρμόζεται όταν τα μεγέθη των αποζημιώσεων περιγράφονται από μια διακριτή τ.μ. X με συνάρτηση πιθανότητας $f(x)$ και η τ.μ. N ανήκει στην κλάση κατανομών $R(a, b, 1)$. Τότε η συνάρτηση πιθανότητας των συνολικών απαιτήσεων υπολογίζεται με χρήση του αναδρομικού σχήματος

$$g(x) =$$

$$= \begin{cases} P_N(f(0)) & x = 0, \\ \dfrac{p_1 - (a+b)p_0}{1 - af(0)} f(x) + \dfrac{1}{1 - af(0)} \sum_{y=1}^{x} \left(a + b\dfrac{y}{x}\right) f(y) g(x-y), & x = 1,2, \dots . \end{cases}$$

Αξίζει να σημειώσουμε ότι υπό τις παραπάνω προϋποθέσεις

$$E(S) = \frac{p_1 + (a+b)(1-p_0)}{1-a} E(X)$$

και

$$Var(S) = \frac{p_1 + (a+b)(1-p_0)}{1-a} \left[Var(X) + E(X)^2 \frac{[1 - p_1 + (a+b)p_0]}{(1-a)} \right].$$

Ειδικότερα για την κλάση $R(a, b, 0)$ το παραπάνω αναδρομικό σχήμα παίρνει τη μορφή

$$g(x) = \begin{cases} P_N(f(0)) & x = 0, \\ \dfrac{1}{1 - af(0)} \sum_{y=1}^{x} \left(a + b\dfrac{y}{x}\right) f(y) g(x-y), & x = 1,2, \dots . \end{cases}$$

Στο όρισμα `model.freq` της συνάρτησης `aggregateDist` δίνουμε την κατανομή της τ.μ. N και στο όρισμα `model.sev` δίνουμε το διάνυσμα των πιθανοτήτων της τ.μ. X.

Μια σημαντική περίπτωση είναι η περίπτωση της σύνθετης κατανομής Poisson η οποία προκύπτει όταν η τ.μ. N ικανοποιεί στο αναδρομικό σχήμα

$$g(x) = P(S = x) = \frac{\lambda}{x} \sum_{k=1}^{x} k f(k) g(x-k), \;\; x = 1, 2, 3, \dots$$

με αρχική συνθήκη την

$$g(0) = P_N\big(f(0)\big) = e^{\lambda(f(0)-1)}.$$

Ο αριθμός των όρων του αθροίσματος στον παραπάνω τύπο είναι x και συνεπώς όσο αυξάνει το x τόσο αυξάνουν και οι όροι του αθροίσματος. Οι όροι του αθροίσματος σταθεροποιούνται στην περίπτωση που τα μεγέθη των αποζημιώσεων περιγράφονται από μία διακριτή τ.μ. X με πεπερασμένο σύνολο τιμών, δηλαδή

$$f(k) = P(X = k), \qquad k = 0, 1, 2, \dots, r.$$

Τότε η συνάρτηση πιθανότητας της σύνθετης κατανομής Poisson παίρνει τη μορφή

$$g(x) = P(S = x) = \frac{\lambda}{x} \sum_{k=1}^{\min(x,r)} k f(k) g(x-k), \ \ x = 1, 2, 3, \dots$$

Για παράδειγμα αν $r = 2$, τότε

$$g(0) = e^{\lambda(f(0)-1)},$$

$$g(1) = \frac{\lambda}{1}[1f(1)g(0)],$$

$$g(2) = \frac{\lambda}{2}[1f(1)g(1) + 2f(2)g(0)]$$

$$g(3) = \frac{\lambda}{3}[1f(1)g(2) + 2f(2)g(1) + 3f(3)g(0)],$$

$$g(4) = \frac{\lambda}{4}[1f(1)g(3) + 2f(2)g(2) + 3f(3)g(1)],$$

$$g(5) = \frac{\lambda}{5}[1f(1)g(4) + 2f(2)g(3) + 3f(3)g(2)],$$

$$g(6) = \frac{\lambda}{6}[1f(1)g(5) + 2f(2)g(4) + 3f(3)g(3)],$$

κ.ο.κ.

<u>Παράδειγμα 2.3</u>

Αν το πλήθος των απαιτήσεων N σε ένα έτος έχει την κατανομή Poisson με παράμετρο $\lambda = 4$ και το μέγεθος των ατομικών ζημιών X έχει την ακόλουθη συνάρτηση πιθανότητας

$$f(x) = \begin{cases} 0.25, & k = 1 \\ 0.5, & k = 2 \\ 0.25, & k = 3 \end{cases}$$

να βρεθεί η συνάρτηση πιθανότητας $g(x)$, $x = 0,1,\dots,20$, των ετήσιων συνολικών αποζημιώσεων S χρησιμοποιώντας το αναδρομικό σχήμα

$$g(x) = P(S = x) = \frac{\lambda}{x} \sum_{k=1}^{\min(x,r)} k f(k) g(x-k), \ \ x = 1, 2, 3, \dots \ .$$

__Λύση__

Κατασκευάζουμε τη συνάρτηση `Panjer.Poisson` που υλοποιεί το παραπάνω αναδρομικό σχήμα και παίρνουμε τα ακόλουθα αποτελέσματα.

```
Panjer.Poisson <- function(f,lambda)
{if (sum(f)>1|any(f<0))stop("f parameter not a destiny")
   if(lambda*sum(f)>727)stop("Underflow")
   g <- exp(-lambda*sum(f))
   r <- length(f)
   x <-0
   repeat{x<-x+1
      m <- min(x,r)
      last <-(lambda/x)*sum(1:m*head(f,m)*rev(tail(g,m)))
      g <- c(g,last)
      cumul <- sum(g)
      if(cumul>0.9999999) break
}
return(g)}

y <- Panjer.Poisson(c(0.25,0.5,0.25),4)
x <- 0:20
m <- cbind(x,y[1:21])
rownames(m)<-rep("",nrow(m))
colnames(m)<-c("x","g(x)")
round(m,digits=6)


  x      g(x)
  0 0.018316
  1 0.018316
  2 0.045789
  3 0.058000
  4 0.074026
  5 0.088678
  6 0.093130
  7 0.095703
```

```
 8 0.091782
 9 0.083776
10 0.073801
11 0.062205
12 0.050728
13 0.040073
14 0.030686
15 0.022878
16 0.016615
17 0.011775
18 0.008159
19 0.005532
20 0.003675
```

2.5.3 Μέθοδος Normal Approximation I (method="normal")

Η συγκεκριμένη μέθοδος βασίζεται στο Κεντρικό Οριακό Θεώρημα (ΚΟΘ) και δίνει ως αποτέλεσμα την πιθανότητα

$$G(x) = P(S \leq x) \approx \Phi\left(\frac{x - \mu_s}{\sigma_s}\right).$$

2.5.4 Μέθοδος Normal Approximation II (method="npower")

Δίνεται ως αποτέλεσμα η πιθανότητα

$$G(x) = P(S \leq x) \approx \Phi\left(-\frac{3}{\gamma_s} + \sqrt{\frac{9}{{\gamma_s}^2} + 1 + \frac{6(x - \mu_s)}{\gamma_s \sigma_s}}\right).$$

Στον παραπάνω τύπο γ_s είναι ο συντελεστής ασυμμετρίας της τ.μ. S

$$\gamma_s = \frac{E[(S - \mu_s)^3]}{{\sigma_s}^3}$$

ο οποίος υπολογίζεται από τη σχέση

$$\gamma_s = \frac{\gamma_N ({\sigma_N}^2)^{3/2} {\mu_X}^3 + 3{\sigma_N}^2 \mu_X {\sigma_X}^2 + \mu_N \gamma_X ({\sigma_N}^2)^{3/2}}{({\sigma_N}^2)^{3/2}}.$$

Η προσέγγιση ισχύει μόνο όταν $x > \mu_s$ και δίνει καλά αποτελέσματα όταν $\gamma_s <$ 1.

2.5.5 Simulation Method (method=”simulation”)

Με αυτή τη μέθοδο γίνεται προσομοίωση ενός τυχαίου δείγματος από την S και στη συνέχεια βρίσκεται η εμπειρική συνάρτηση κατανομής της S.

Για την καλύτερη κατανόηση των παραπάνω μεθόδων υπολογισμού της συνάρτησης κατανομή, ακολουθούν ορισμένες εφαρμογές αυτών.

2.6 Εφαρμογές της συνάρτησης `aggregateDist`

Παράδειγμα 2.4

Η κατανομή της N δίνεται στο παρακάτω πλαίσιο

n	0	1	2	3	4	5	6	7
p_n	0.01	0.1	0.15	0.2	0.25	0.15	0.1	0.04

Η κατανομή της X δίνεται στο παρακάτω πλαίσιο

x	1	2	3	4	5	6	7	8	9
$f(x)$	0.15	0.2	0.3	0.125	0.075	0.05	0.05	0.025	0.025

Να χρησιμοποιηθεί η μέθοδος των συνελίξεων για να βρεθεί η συνάρτηση πιθανότητας $g(x)$, και η συνάρτηση κατανομής $G(x)$ $x = 0,1,\ldots,30$, των συνολικών αποζημιώσεων S. Να δοθούν και γραφικές παραστάσεις.

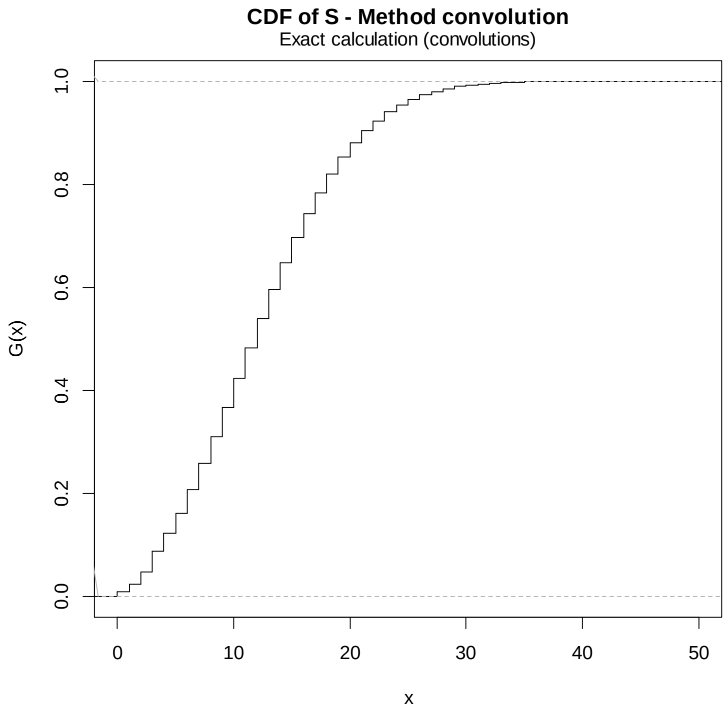
Λύση

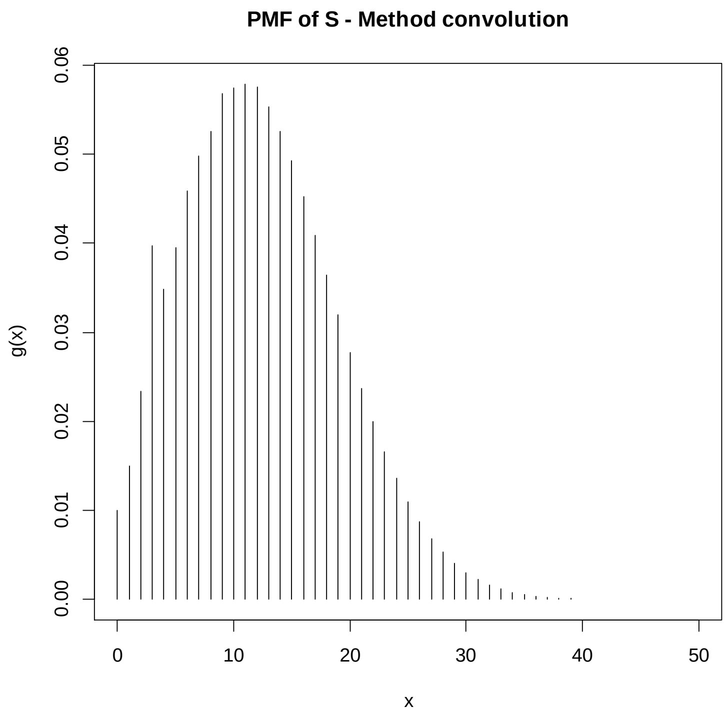
Έχουμε τον ακόλουθο κώδικα:

```
library(actuar)

# Frequency: N
fN <- c(0.01, 0.1, 0.15, 0.2, 0.25, 0.15, 0.1, 0.04)

# Severity: X
fX <- c(0, 0.15, 0.2, 0.3, 0.125, 0.075, 0.05, 0.05,
0.025, 0.025)
```

```
# Method convolution
Gs.conv <- aggregateDist("convolution", model.freq = fN,
model.sev = fX)

# Table cdf and pmf
c <- 30
x <- seq(0,c,1)
gs.conv <- diff(Gs.conv)
d <- c+1
mat <- cbind(x,Gs.conv(x),gs.conv[1:d])
rownames(mat) <- rep("", nrow(mat))
colnames(mat) <- c("x","       G(x)_Convolution", "
g(x)_Convolution")
round(mat,digits=8)

# Plot cdf
plot(Gs.conv,                          do.points=FALSE,
verticals=TRUE,xlim=c(0,50),xlab="x",          ylab="G(x)",
main="CDF of S - Method convolution")

# Plot pmf
plot(knots(Gs.conv),    gs.conv,    type="h",xlim=c(0,50),
xlab="x",   ylab="g(x)",   main="PMF   of   S   -   Method
convolution")

  x   G(x)_Convolution   g(x)_Convolution
  0          0.0100000         0.01000000
  1          0.0250000         0.01500000
  2          0.0483750         0.02337500
  3          0.0880500         0.03967500
  4          0.1228766         0.03482656
  5          0.1623380         0.03946139
  6          0.2082400         0.04590208
```

7	0.2580500	0.04980996
8	0.3105665	0.05251652
9	0.3673843	0.05681781
10	0.4248457	0.05746134
11	0.4827440	0.05789839
12	0.5403007	0.05755667
13	0.5955797	0.05527896
14	0.6481866	0.05260689
15	0.6974641	0.04927747
16	0.7427260	0.04526194
17	0.7836137	0.04088775
18	0.8200653	0.03645153
19	0.8520441	0.03197887
20	0.8797449	0.02770078
21	0.9034320	0.02368712
22	0.9233877	0.01995567
23	0.9399692	0.01658150
24	0.9535588	0.01358960
25	0.9645359	0.01097711
26	0.9732793	0.00874343
27	0.9801425	0.00686317
28	0.9854483	0.00530576
29	0.9894872	0.00403894
30	0.9925153	0.00302806





<u>**Παράδειγμα 2.5**</u>

Το πλήθος των απαιτήσεων N σε μία συγκεκριμένη χρονική περίοδο ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p = 1/2$, δηλαδή

$$p_k = P(N = k) = p(1 - p)^{k-1}, \qquad k = 0, 1, 2, \dots$$

Η κατανομή του μεγέθους X των αποζημιώσεων δίνεται στο παρακάτω πλαίσιο

x	1	3	5
$f(x)$	0.25	0.5	0.25

Να χρησιμοποιηθεί η μέθοδος των συνελίξεων και η αναδρομική για να βρεθεί η συνάρτηση πιθανότητας $g(x)$, και η συνάρτηση κατανομής $G(x)$ $x = 0,1, \dots, 20$, των συνολικών αποζημιώσεων S.

Λύση

Έχουμε τον ακόλουθο κώδικα:

```
library(actuar)

# Frequency: N
p <- 1/2
q <- 1-p
prb <- 1-1e-15
limN <- qgeom(prb, prob=p)
fN <- dgeom(0:limN, prob=p)

# Severity: X
fX <- c(0,0.25, 0, 0.5, 0, 0.25)

# Method convolution
Gs.conv <- aggregateDist("convolution", model.freq =fN,
model.sev = fX)

# Method recusrsive
Gs.rec <- aggregateDist("recursive", model.freq
="geometric", model.sev = fX, prob = p)

# Table cumulative distribution function
c <- 20
x <- seq(0,c,1)
mat <- cbind(x,Gs.conv(x),Gs.rec(x))
rownames(mat) <- rep("", nrow(mat))
```

```
colnames(mat)   <-   c("x",   "      G(x)_Convolution",   "
G(x)_Recursive")
round(mat,digits=8)

# Table probability mass function
gs.conv <- diff(Gs.conv)
gs.rec <- diff(Gs.rec)
d <- c+1
mat <- cbind(x,gs.conv[1:d],gs.rec[1:d])
rownames(mat) <- rep("", nrow(mat))
colnames(mat)   <-   c("x","      g(x)_Convolution",   "
g(x)_Recursive")
round(mat,digits=8)

  x   G(x)_Convolution   G(x)_Recursive
  0          0.5000000        0.5000000
  1          0.5625000        0.5625000
  2          0.5703125        0.5703125
  3          0.6962891        0.6962891
  4          0.7276611        0.7276611
  5          0.7960358        0.7960358
  6          0.8438892        0.8438892
  7          0.8586905        0.8586905
  8          0.8933814        0.8933814
  9          0.9136026        0.9136026
 10          0.9283774        0.9283774
 11          0.9448787        0.9448787
 12          0.9538468        0.9538468
 13          0.9629979        0.9629979
 14          0.9707947        0.9707947
 15          0.9758582        0.9758582
 16          0.9808416        0.9808416
 17          0.9845347        0.9845347
 18          0.9874061        0.9874061
```

```
19          0.9899855        0.9899855
20          0.9918642        0.9918642
>
> # Table probability mass function
> gs.conv <- diff(Gs.conv)
> gs.rec <- diff(Gs.rec)
> d <- c+1
> mat <- cbind(x,gs.conv[1:d],gs.rec[1:d])
> rownames(mat) <- rep("", nrow(mat))
>  colnames(mat)  <-  c("x","     g(x)_Convolution",   "
g(x)_Recursive")
> round(mat,digits=8)

  x   g(x)_Convolution   g(x)_Recursive
  0         0.50000000       0.50000000
  1         0.06250000       0.06250000
  2         0.00781250       0.00781250
  3         0.12597656       0.12597656
  4         0.03137207       0.03137207
  5         0.06837463       0.06837463
  6         0.04785347       0.04785347
  7         0.01480126       0.01480126
  8         0.03469089       0.03469089
  9         0.02022124       0.02022124
 10         0.01477480       0.01477480
 11         0.01650126       0.01650126
 12         0.00896812       0.00896812
 13         0.00915108       0.00915108
 14         0.00779685       0.00779685
 15         0.00506349       0.00506349
 16         0.00498336       0.00498336
 17         0.00369315       0.00369315
 18         0.00287140       0.00287140
 19         0.00257937       0.00257937
```

```
20          0.00187864          0.00187864
```

Παράδειγμα 2.6

Το πλήθος των απαιτήσεων N σε μία συγκεκριμένη χρονική περίοδο ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p = 1/4$, δηλαδή

$$p_k = P(N = k) = \frac{1}{4}\left(\frac{3}{4}\right)^{k-1}, \qquad k = 0, 1, 2, \dots$$

Το μέγεθος μιας αποζημίωσης X ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p = 1/5$.

Να χρησιμοποιηθούν όλες οι μέθοδοι της συνάρτησης `aggregateDist` για να βρεθεί η συνάρτηση πιθανότητας $g(x)$, και η συνάρτηση κατανομής $G(x)$, $x = 0,1, \dots ,20$, των συνολικών αποζημιώσεων S.

Λύση

Έχουμε τον ακόλουθο κώδικα:

```
library(actuar)

# Frequency: N
p <- 1/4
q <- 1-p
prb <- 1-1e-15
limN <- qgeom(prb, prob=p);limN
fN <- dgeom(0:limN, prob=p)

# Severity: X
P <- 1/5
Q <- 1-P
limX <- qgeom(prb, prob=P);limX
fX <- dgeom(0:limX,P)

# Method convolution
Gs.conv  <-  aggregateDist("convolution",  model.freq  =fN,
model.sev = fX)

# Method recusrsive
```

```
Gs.rec <- aggregateDist("recursive", model.freq
="geometric", model.sev = fX, prob = p)

# Method normal
meanN <- q/p
varN <- q/p^2
meanX <- Q/P
varX <- Q/P^2
meanS <- meanN*meanX
varS <- meanN*varX+varN*meanX^2
Gs.norm <- aggregateDist("normal", moments =
c(meanS,varS))

# Method npower
skewN <- (2-p)/sqrt(1-p)
skewX <- (2-P)/sqrt(1-P)
skewS <-
(skewN*varN^(3/2)*meanX^3+3*varN*meanX*varX+meanN*skewX*v
arX^(3/2))/(varS^(3/2))
Gs.npower <- aggregateDist("npower", moments =
c(meanS,varS,skewS))

# Method simulation
mfreq <- expression(data = rgeom(prob=p))
msev <- expression(data = rgeom(prob=P))
Gs.sim <- aggregateDist("simulation", nb.simul = 100000,
mfreq, msev)


# Table cumulative distribution function
c <- 20
x <- seq(0,c,1)
mat <- cbind(x,Gs.conv(x),Gs.rec(x), Gs.norm(x),
Gs.npower(x), Gs.sim(x))
```

```
rownames(mat) <- rep("", nrow(mat))
colnames(mat)  <-  c("x",  "      G(x)_Convolution",  "
G(x)_Recursive",  "      G(x)_Normal","     G(x)_NPower","
G(x)_Simulation")
round(mat,digits=5)

# Table probability mass function
gs.conv <- diff(Gs.conv)
gs.rec <- diff(Gs.rec)
y <- seq(0,c+1,1)
gs.norm <- c(Gs.norm(0),diff(Gs.norm(y)))
gs.npower <- c(Gs.npower(0),diff(Gs.npower(y)))
gs.sim <- diff(Gs.sim)
d <- c+1
mat          <-          cbind(x,gs.conv[1:d],gs.rec[1:d],
gs.norm[1:d],gs.npower[1:d],gs.sim[1:d])
rownames(mat) <- rep("", nrow(mat))
colnames(mat)   <-   c("x","      g(x)_Convolution",   "
g(x)_Recursive","     g(x)_Normal","     g(x)_NPower",   "
g(x)_Simulation")
round(mat,digits=8)
> round(mat,digits=5)
```

x	G(x)_Convolution	G(x)_Recursive	G(x)_Normal	G(x)_NPower	G(x)_Simulation
0	0.29412	0.29412	0.22485	NA	0.29615
1	0.33564	0.33564	0.24418	NA	0.33651
2	0.37472	0.37472	0.26437	NA	0.37648
3	0.41150	0.41150	0.28538	NA	0.41345
4	0.44612	0.44612	0.30715	NA	0.44833
5	0.47870	0.47870	0.32962	NA	0.47959
6	0.50936	0.50936	0.35273	NA	0.50991
7	0.53823	0.53823	0.37639	NA	0.53822
8	0.56539	0.56539	0.40053	NA	0.56514
9	0.59095	0.59095	0.42505	NA	0.59020
10	0.61502	0.61502	0.44987	NA	0.61428
11	0.63766	0.63766	0.47489	NA	0.63665
12	0.65898	0.65898	0.50000	NA	0.65826
13	0.67904	0.67904	0.52511	0.64702	0.67823
14	0.69792	0.69792	0.55013	0.66493	0.69733
15	0.71569	0.71569	0.57495	0.68201	0.71490

```
16          0.73241      0.73241   0.59947   0.69830       0.73170
17          0.74815      0.74815   0.62361   0.71383       0.74687
18          0.76297      0.76297   0.64727   0.72863       0.76156
19          0.77691      0.77691   0.67038   0.74272       0.77555
20          0.79003      0.79003   0.69285   0.75614       0.78882
```

```
# Table probability mass function
gs.conv <- diff(Gs.conv)
gs.rec <- diff(Gs.rec)
y <- seq(0,c+1,1)
gs.norm <- c(Gs.norm(0),diff(Gs.norm(y)))
gs.npower <- c(Gs.npower(0),diff(Gs.npower(y)))
gs.sim <- diff(Gs.sim)
d <- c+1
mat          <-          cbind(x,gs.conv[1:d],gs.rec[1:d],
gs.norm[1:d],gs.npower[1:d],gs.sim[1:d])
rownames(mat) <- rep("", nrow(mat))
colnames(mat)    <-    c("x","        g(x)_Convolution",    "
g(x)_Recursive","     g(x)_Normal","     g(x)_NPower",   "
g(x)_Simulation")
round(mat,digits=5)
```

```
 x  g(x)_Convolution  g(x)_Recursive  g(x)_Normal  g(x)_NPower  g(x)_Simulation
 0           0.29412         0.29412      0.22485           NA          0.29615
 1           0.04152         0.04152      0.01933           NA          0.04036
 2           0.03908         0.03908      0.02019           NA          0.03997
 3           0.03678         0.03678      0.02101           NA          0.03697
 4           0.03462         0.03462      0.02177           NA          0.03488
 5           0.03258         0.03258      0.02247           NA          0.03126
 6           0.03066         0.03066      0.02311           NA          0.03032
 7           0.02886         0.02886      0.02366           NA          0.02831
 8           0.02716         0.02716      0.02414           NA          0.02692
 9           0.02557         0.02557      0.02452           NA          0.02506
10           0.02406         0.02406      0.02482           NA          0.02408
11           0.02265         0.02265      0.02501           NA          0.02237
12           0.02131         0.02131      0.02511           NA          0.02161
13           0.02006         0.02006      0.02511           NA          0.01997
14           0.01888         0.01888      0.02501      0.01791          0.01910
15           0.01777         0.01777      0.02482      0.01708          0.01757
16           0.01672         0.01672      0.02452      0.01629          0.01680
17           0.01574         0.01574      0.02414      0.01553          0.01517
18           0.01481         0.01481      0.02366      0.01480          0.01469
19           0.01394         0.01394      0.02311      0.01409          0.01399
```

20	0.01312	0.01312	0.02247	0.01342	0.01327

Παράδειγμα 2.7

Το πλήθος των απαιτήσεων N σε μία συγκεκριμένη χρονική περίοδο ακολουθεί τη διωνυμική κατανομή με παραμέτρους $n = 3$ και $p = 0.6$, και η κατανομή του μεγέθους X των αποζημιώσεων ακολουθεί την ομοιόμορφη κατανομή στο διάστημα (0,20).

Να γίνει διακριτοποίηση της κατανομής του μεγέθους X των αποζημιώσεων με τη μέθοδο "rounding" στο διάστημα (0,20) με βήμα $h = 1/1000$. Στη συνέχεια, να βρεθεί η συνάρτηση κατανομής $G(x)$, $x = 0,1,\dots,20$, της τ.μ. S με την βοήθεια των μεθόδων convolution και simulation.

Λύση

Έχουμε τον ακόλουθο κώδικα:

```
library(actuar)

# Frequency: N
n <- 3
p <- 0.6
fN <- dbinom(0:n, size = n, prob = p)

# Severity: X
a <- ua <- 0
b <- ub <- 20
c <- 1000
h <- 1/c
x <- seq(a,b,h)
fX <- discretize(punif(x,min=ua, max=ub), from=a, to=b,
step=h, method="rounding")

# Method convolution
Gs.conv <- aggregateDist("convolution", model.freq =fN,
model.sev = fX, x.scale=h)
```

```
# Method simulation
mfreq <- expression(data = rbinom(size=n, prob=p))
msev <- expression(data = runif(min=ua,max=ub))
Gs.sim <- aggregateDist("simulation", nb.simul = 100000,
mfreq, msev)

# Table cumulative distribution function
c <- 30
x <- seq(0,c,1)
mat <- cbind(x,Gs.conv(x),Gs.sim(x))
rownames(mat) <- rep("", nrow(mat))
colnames(mat)  <-  c("x",  "     G(x)_Convolution",  "
G(x)_Simulation")
round(mat,digits=5)

  x   G(x)_Convolution   G(x)_Simulation
  0            0.06401           0.06300
  1            0.07895           0.07800
  2            0.09500           0.09395
  3            0.11219           0.11184
  4            0.13054           0.13017
  5            0.15007           0.14980
  6            0.17082           0.17058
  7            0.19281           0.19311
  8            0.21608           0.21717
  9            0.24063           0.24135
 10            0.26651           0.26797
 11            0.29374           0.29416
 12            0.32235           0.32235
 13            0.35236           0.35322
 14            0.38380           0.38578
 15            0.41670           0.41872
 16            0.45109           0.45269
 17            0.48699           0.48986
```

18	0.52442	0.52862
19	0.56343	0.56700
20	0.60401	0.60634
21	0.63073	0.63295
22	0.65685	0.65873
23	0.68233	0.68416
24	0.70710	0.70914
25	0.73112	0.73198
26	0.75433	0.75576
27	0.77668	0.77743
28	0.79811	0.79871
29	0.81856	0.81936
30	0.83799	0.83826

Παράδειγμα 2.8

Το πλήθος των απαιτήσεων N ακολουθεί την αποκομμένη στο σημείο 0 γεωμετρική κατανομή με παράμετρο $p = 0.5$, και η κατανομή του μεγέθους X των αποζημιώσεων ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda = 1/10$ (δηλαδή $Exp(1/10)$).

Να γίνει διακριτοποίηση της κατανομής του μεγέθους X των αποζημιώσεων με τη μέθοδο "`rounding`" στο διάστημα (0,20) με βήμα $h = 1/20$. Στη συνέχεια, να βρεθεί η συνάρτηση κατανομής $G(x)$, $x = 0,1,\ldots,30$, της τ.μ. S με την βοήθεια των μεθόδων `convolution` και `recurscive`.

Να συγκριθούν τα αποτελέσματα με εκείνα της κατανομής τη S που είναι η $Exp(\lambda p)$.

Λύση

Έχουμε τον ακόλουθο κώδικα:

```
library(actuar)

# Frequency: N
q <- 0.5; p <- 1-q
a <- q; b <- 0
p0 <- 0; p1 <- p
```

```
pmfZTG <- function(n) {
       if (n==0) return(p0)
       if (n==1) return(p1)
       return((a+b/n)*pmfZTG(n-1))
}

n <- 0:50
fN <- sapply(n,pmfZTG)

# Severity: X
ra <- 1/10  # mean=10
a <- 0
b <- floor(qexp(0.99999,rate=ra));b
c <- 20; h <- 1/c
x <- seq(a,b,h)
fX <- discretize(pexp(x,rate=ra), from=a, to=b, step=h,
method="rounding")

# Method convolution
Gs.conv <- aggregateDist("convolution", model.freq =fN,
model.sev = fX, x.scale=h)

# Method recusrsive
Gs.rec <- aggregateDist("recursive", model.freq ="zero-
truncated geometric", model.sev = fX, prob = p, x.scale =
h, maxit=100000, tol=0.0001)

# Table cumulative distribution function
c <- 30
x <- seq(0,c,1)
Gs.exact <- pexp(x,p*ra)
mat <- cbind(x,Gs.conv(x),Gs.rec(x), Gs.exact)
rownames(mat) <- rep("", nrow(mat))
```

```
colnames(mat)   <-   c("x",   "     G(x)_Convolution",   "
G(x)_Recursive", "  G(x)_Exact")
round(mat,digits=8)

  x   G(x)_Convolution   G(x)_Recursive   G(x)_Exact
  0         0.00125000       0.00125000   0.00000000
  1         0.04995961       0.04995961   0.04877058
  2         0.09629363       0.09629363   0.09516258
  3         0.14036791       0.14036791   0.13929202
  4         0.18229266       0.18229266   0.18126925
  5         0.22217272       0.22217272   0.22119922
  6         0.26010780       0.26010780   0.25918178
  7         0.29619277       0.29619277   0.29531191
  8         0.33051785       0.33051785   0.32967995
  9         0.36316888       0.36316888   0.36237185
 10         0.39422750       0.39422750   0.39346934
 11         0.42377138       0.42377138   0.42305019
 12         0.45187438       0.45187438   0.45118836
 13         0.47860678       0.47860678   0.47795422
 14         0.50403543       0.50403543   0.50341470
 15         0.52822391       0.52822391   0.52763345
 16         0.55123270       0.55123270   0.55067104
 17         0.57311934       0.57311934   0.57258507
 18         0.59393855       0.59393855   0.59343034
 19         0.61374240       0.61374240   0.61325898
 20         0.63258041       0.63258041   0.63212056
 21         0.65049967       0.65049967   0.65006225
 22         0.66754500       0.66754500   0.66712892
 23         0.68375903       0.68375903   0.68336323
 24         0.69918228       0.69918228   0.69880579
 25         0.71385333       0.71385333   0.71349520
 26         0.72780887       0.72780887   0.72746821
 27         0.74108379       0.74108379   0.74075974
 28         0.75371128       0.75371128   0.75340304
```

29	0.76572292	0.76572292	0.76542971
30	0.77714875	0.77714875	0.77686984

2.7 Προσομοίωση κατανομών με χρήση του πακέτου actuar

Το πακέτο actuar παρέχει διάφορες λειτουργίες για την δημιουργία τυχαίων δειγμάτων από σύνθετα αναλογιστικά μοντέλα χρησιμοποιώντας κυρίως τις εξής τις εξής συναρτήσεις:

1. `rmixture (n, probs, models)`: Για την προσομοίωση από διακριτές μείξεις κατανομών.
2. `rcompound (n, model.freq, model.sev, SIMPLIFY=TRUE)`: Για την προσομοίωση σύνθετων μοντέλων .
3. `rcomphierarc (nodes , model.freq = NULL , model.sev = NULL , weights = NULL)`: Για την προσομοίωση σύνθετων μοντέλων όπου ο αριθμός των κινδύνων και το μέγεθος της ζημιάς έχουν ιεραρχική δομή.

2.7.1 Προσομοίωση διακριτών μείξεων κατανομών

Μια τυχαία μεταβλητή X αποτελεί διακριτή μείξη (k-point mixture) των τυχαίων μεταβλητών $X_1, X_2, \dots, X_n$ με βάρη μείξης $p_1, p_2, \dots, p_n$, αν η συνάρτηση κατανομής της $F(x)$ δίνεται από τον τύπο

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + \cdots p_n F_n(x) = \sum_{i=1}^{n} p_i F_i(x),$$

όπου $p_i \geq 0$ με $\sum_{i=1}^{n} p_i = 1$.

Επίσης η συνάρτηση πιθανότητας (ή πυκνότητας) της X ικανοποιεί τη σχέση

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \cdots p_n f_n(x) = \sum_{i=1}^{n} p_i f_i(x).$$

Παράδειγμα 2.9

Έστω ότι η τ.μ. X η οποία προκύπτει από μείξη τριών εκθετικών κατανομών με παραμέτρους $\lambda_1 = 2$, $\lambda_2 = 4$ και $\lambda_3 = 5$, και βάρη $p_1 = 3/14$, $p_2 = 4/14$ και $p_3 = 7/14$. Να παραχθεί δείγμα μεγέθους 10 παρατηρήσεων από την τ.μ. X.

<u>Λύση</u>

Η συνάρτηση πυκνότητας της X είναι η

$$f(x) = \frac{3}{14}(2e^{-2x}) + \frac{4}{14}(4e^{-4x}) + \frac{7}{14}(5e^{-5x}) = \frac{6e^{-2x} + 16e^{-4x} + 35e^{-5x}}{14}.$$

Για την παραγωγή των 10 παρατηρήσεων έχουμε τον ακόλουθο κώδικα

```
rmixture(10, probs = c(3,4,7), models = expression(rexp(2),
rexp(4), rexp(5)))

 [1] 0.02089735 0.12095276 0.15120735 0.08803977 0.21331267
0.02814927
 [7] 0.13220854 0.18851220 0.26238903 0.54935223
```

2.7.2 Προσομοίωση σύνθετων κατανομών

Σε πολλές περιπτώσεις οι αναλογιστές καλούνται να υπολογίσουν πιο σύνθετα μοντέλα συλλογικού προτύπου. Στην προκειμένη περίπτωση, το πακέτο `actuar` διαθέτει την συνάρτηση `rcompound` η οποία παράγει τυχαίο δείγμα από την τυχαία μεταβλητή S όταν η κατανομή των δύο τυχαίων μεταβλητών N και X. Στην περίπτωση που η N έχει την κατανομή Poisson τότε μπορεί να χρησιμοποιηθεί και η συνάρτηση `rcomppois`.

Παράδειγμα 2.10

Το πλήθος των απαιτήσεων N σε μία συγκεκριμένη χρονική περίοδο ακολουθεί τη κατανομή *Poisson* με παράμετρο $\lambda = 6$ και η κατανομή μεγέθους X των αποζημιώσεων ακολουθεί κανονική κατανομή με μέση τιμή $\mu = 8$ και διακύμανση $\sigma^2 = 4$.

Να παραχθεί ένα τυχαίο δείγμα 10000 παρατηρήσεων της S. Στη συνέχεια να υπολογιστούν προσεγγιστικά οι πιθανότητες $P(S = 0)$ και $P(S \leq 20.5)$.

Λύση

Έχουμε τον ακόλουθο κώδικα

```
library(actuar)
simu <- 10000
S          <-          rcompound(simu,model.freq=rpois(3),
model.sev=rnorm(8,2))

# P(S=0)
```

```
a<-S[S==0]; length(a)/simu
[1] 0.0493
# Ακριβής τιμή P(N=0)
dpois(0,3)
[1] 0.04978707

# P(S<=20.5)
a<-S[S<=20.5]; length(a)/simu
[1] 0.4423
```

Εναλλακτικά, με τη συνάρτηση `rcomppois`, έχουμε τον ακόλουθο κώδικα

```
S <- rcomppois(simu, 3, rnorm(8,2))

# P(S=0)
a<-S[S==0]; length(a)/simu
[1] 0.0481

# P(S<=20.5)
a<-S[S<=20.5]; length(a)/simu
[1] 0.4546
```

2.7.3 Προσομοίωση σύνθετων ιεραρχικών μοντέλων

Τα ιεραρχικά μοντέλα πιθανοτήτων χρησιμοποιούνται ευρέως για δεδομένα που παρουσιάζουν μια πολυεπίπεδη μορφή στην δομή τους. Το κύριο χαρακτηριστικό τους είναι η ύπαρξη ενός κανόνα πιθανότητας σε κάθε επίπεδο της ταξινόμησής τους το οποίο εξαρτάται από κάποιο προηγούμενο αποτέλεσμα. Μία απλή μορφή ιεραρχικού μοντέλου θα μπορούσε να γραφεί ως:

$$X_t|\Lambda,\Theta \sim Poisson(\Lambda),$$
$$\Lambda|\Theta \sim Gamma(3,\Theta),$$
$$\Theta \sim Gamma(2,2),$$

όπου X_t είναι τα πραγματικά δεδομένα και οι τυχαίες μεταβλητές $\Lambda\, και\, \Theta$ θεωρούνται γενικά ως παράμετροι αβεβαιότητας.

Τυχαίο δείγμα μεγέθους n από το παραπάνω μοντέλο προκύπτει με τον ακόλουθο κώδικα:

```
rpois(n, rgamma(n, shape=3, rgamma(n, shape=2, rate=2)))
```

Στην περίπτωση της προσομοίωσης ιεραρχικών μοντέλων με χρήση της γλώσσα προγραμματισμού R πρέπει να οριστούν τα εν λόγω μοντέλα με τρόπο τέτοιο ώστε να ανταποκρίνονται στα παρακάτω κριτήρια:

1. Πρέπει να είναι απλό και σύμφωνα με τα πρότυπα της εν λόγω έκδοσης της R και προγενέστερων.
2. Να επιτρέπει οποιοδήποτε αριθμό επιπέδων και κόμβων
3. Ανεξάρτητα από το επίπεδο να επιτρέπεται η χρήση παραμέτρων υψηλότερων από την ιεραρχική δομή.

Τα ιεραρχικά μοντέλα επιτυγχάνονται μέσω αντικατάστασης μιας ή περισσότερων παραμέτρων μιας κατανομής σε ένα δεδομένο επίπεδο από οποιονδήποτε συνδυασμό των παραπάνω επιπέδων. Η εν λόγω λειτουργία μπορεί να προσομοιώσει δεδομένα στην περίπτωση όπου οι τ.μ. N και X είναι ιεραρχικές και να επιστρέψει αποτελέσματα στην μορφή ενός ενιαίου χαρτοφυλακίου σύμφωνα με τα συμβόλαια και τα έτη που δόθηκαν, όπως παρουσιάζεται στο παράδειγμα που έπεται.

Παράδειγμα 2.11

Έστω το παρακάτω σύνθετο ιεραρχικό μοντέλο

$$S_{ijt} = X_{ijt1} + \cdots + X_{ijtN_{ijt}},$$

όπου $i = 1,2,\cdots,I,\ j = 1,2,\cdots,J_i,\ t = 1,2,\cdots,n_{ij}$ και

$$N_{ijt}|\Lambda_{ij},K_i \sim Poisson(w_{ijt}\Lambda_{ij}),$$

$$\Lambda_{ij}|K_i \sim Gamma(K_i,1),$$

$$K_i \sim Exp(2)$$

και

$$X_{ijtu}|\Theta_{ij},P_i \sim Lognormal(\Theta_{ij},1),$$

$$\Theta_{ij}|P_i \sim N(P_i,1),$$

$$P_i \sim N(2,0.1).$$

Έστω ότι $I = 2$, $J_1 = 4$, $J_2 = 3$ και $n_{11} = \cdots = n_{14} = 4$, $n_{21} = n_{22} = n_{23} = 5$ και τα βάρη $w_{ij} \sim U(0.5, 2.5)$. Να γίνει προσομοίωση του ανωτέρω μοντέλου.

Λύση

Η προσομοίωση του μοντέλου με τη συνάρτηση `rcomphierarc` επιτυγχάνεται με τον παρακάτω κώδικα:

```
library(actuar)
nodes <- list(cohort = 2,contract = c(4, 3),year = c(4, 4, 4,
4, 5, 5, 5))
mf <- expression(cohort = rexp(2),contract = rgamma(cohort,
1),year = rpois(weights * contract))
ms <- expression(cohort = rnorm(2, sqrt(0.1)),contract =
rnorm(cohort, 1),year = rlnorm(contract, 1))
wijt <- runif(31, 0.5, 2.5)
pf <- rcomphierarc(nodes = nodes,model.freq = mf, model.sev =
ms,weights = wijt)
s<-aggregate (pf)
round (s,digits=3)
```

cohort	contract	year.1	year.2	year.3	year.4	year.5
1	1	0.000	11.511	13.260	20.064	NA
1	2	0.000	23.338	0.000	0.000	NA
1	3	300.945	192.397	0.000	162.344	NA
1	4	0.000	87.992	77.080	153.293	NA
2	1	0.000	0.000	0.000	0.000	0.000
2	2	0.000	0.000	0.000	0.000	0.000
2	3	21.387	0.000	15.647	0.000	25.122

3 Κεφάλαιο: Θεωρία χρεοκοπίας και το πακέτο actuar

Η θεωρία χρεοκοπίας στο συλλογικό πρότυπο της θεωρίας κινδύνου είναι ένας κλάδος της αναλογιστικής επιστήμης που μελετά την φερεγγυότητα του ασφαλιστή βασιζόμενη σε μαθηματικά μοντέλα του πλεονάσματος. Η θεωρητική θεμελίωση της θεωρίας χρεοκοπίας αλλά και του συλλογικού προτύπου γενικότερα, η οποία είναι γνωστή στη διεθνή βιβλιογραφία και ως classical Poisson risk model, εισήχθη από τον Σουηδό αναλογιστή Fillip Lundberg το 1903.

Βασικός στόχος του κλασικού μοντέλου αλλά και των μετέπειτα επεκτάσεων του είναι ο υπολογισμός της πιθανότητας χρεοκοπίας. Η μελέτη των συνολικών αποζημιώσεων ενός ασφαλιστή έχει σημαντικό ρόλο στη θεωρία της χρεοκοπίας. Εδώ οι αποζημιώσεις αυτές δεν εξετάζονται σε σταθερό χρόνο, αλλά δυναμικά καθώς εξελίσσονται με την πάροδο του χρόνου και με τη βοήθεια του συλλογικού προτύπου μακράς χρονικής περιόδου. Στο πρότυπο αυτό, κεντρικό πρόβλημα είναι το πρόβλημα χρεοκοπίας, στο οποίο εξετάζεται η μεταβολή του πλεονάσματος που προκύπτει από τα έσοδα (ασφάλιστρα) μείον τα έξοδα (αποζημιώσεις) για έναν ασφαλιστή, τόσο σε διακριτό όσο και σε συνεχή χρόνο. Βασικά εργαλεία για τη θεωρία της χρεοκοπίας αποτελούν οι σύνθετες κατανομές πιθανότητας (ιδιαίτερα η σύνθετη γεωμετρική και η σύνθετη Poisson) και η θεωρία στοχαστικών ανελίξεων.

Κάποιες από τις ποσότητες οι οποίες θα μελετηθούν στο κεφάλαιο αυτό είναι η πιθανότητα, ο χρόνος χρεοκοπίας, το πλεόνασμα πριν τη χρεοκοπία και το έλλειμμα τη στιγμή της χρεοκοπίας. Τέλος, θα εισάγουμε την έννοια των κλιμακωτών υψών και της σωρευτικής απώλειας καθώς και του χρόνου χρεοκοπίας.

3.1 Η στοχαστική ανέλιξη των συνολικών αποζημιώσεων

Στο συλλογικό μοντέλο της θεωρία κινδύνων χρησιμοποιούμε την τυχαία μεταβλητή $S = X_1 + X_2 + \cdots + X_N$, η οποία περιγράφει τις συνολικές απαιτήσεις, που προέρχονται από ένα χαρτοφυλάκιο, προς μία ασφαλιστική εταιρεία. Στη θεωρία χρεοκοπίας εξετάζουμε τις συνολικές αποζημιώσεις από ένα χαρτοφυλάκιο ασφαλισμένων μιας ασφαλιστικής εταιρείας όπως αυτές εξελίσσονται στο χρόνο. Γι’ αυτό το λόγο χρησιμοποιούμε τη στοχαστική διαδικασία (στοχαστική ανέλιξη) $\{S(t): t \geq 0\}$ που ορίζεται από την σχέση

$$S(t) = \begin{cases} \sum_{i=1}^{N(t)} X_i, & N(t) \geq 1 \\ 0, & N(t) = 0 \end{cases}$$

και είναι μία σύνθετη ανέλιξη. Η απαριθμήτρια ανέλιξη $\{N(t): t \geq 0\}$ μετρά τον αριθμό των απαιτήσεων προς την εταιρεία στο χρόνο. Το σημαντικότερο παράδειγμα μιας τέτοιας ανέλιξης, είναι αυτή στην οποία η $\{N(t)\}$ είναι μία ανέλιξη Poisson, οπότε τότε λέμε ότι η $\{S(t): t \geq 0\}$ ακολουθεί μία σύνθετη ανέλιξη Poisson.

3.2 Η στοχαστική ανέλιξη πλεονάσματος (surplus)

Η ασφαλιστική εταιρεία είναι μια επιχείρηση η οποία έχει σκοπό και το κέρδος. Μας ενδιαφέρουν, δηλαδή, τα έσοδα της εταιρείας τα οποία μεταβάλλονται στο χρόνο. Έσοδα αυτά προέρχονται από την είσπραξη των συνολικών ασφαλίστρων που εισπράττει η εταιρεία. Η ασφαλιστική εταιρεία θα πρέπει να έχει κεφάλαια επαρκή για την αποτελεσματική άσκηση των δραστηριοτήτων της αλλά να έχει και πρόσθετα κεφάλαια (αποθεματικά) για την αντιμετώπιση απρόοπτων εξελίξεων που θα μπορούσαν να θέσουν σε κίνδυνο την ομαλή λειτουργία της επιχείρησης. Στην ορολογία των ασφαλίσεων, τα αποθεματικά αυτά καλούνται πλεόνασμα.

Έστω τώρα, μία συνάρτηση $P(t)$ η οποία δηλώνει τα συνολικά ασφάλιστρα που εισρέουν στην εταιρεία στο χρονικό διάστημα $[0, t]$. Επισημαίνεται ότι τα ασφάλιστρα καθορίζονται με ακρίβεια από τον ασφαλιστή, δηλαδή, και σύμφωνα με τον Grandell (1991), δεν υπάρχει αβεβαιότητα ως προς την εξέλιξή τους στο χρόνο, για το λόγο αυτό η $P(t)$ είναι μία αύξουσα (μαθηματική) συνάρτηση και όχι μία στοχαστική ανέλιξη.

Η στοχαστική ανέλιξη του πλεονάσματος $\{U(t), t \geq 0\}$ ορίζεται για κάθε $t \geq 0$ από τη σχέση

$$U(t) = u + P(t) - S(t)$$

όπου u είναι το αρχικό αποθεματικό, $P(t)$ τα συνολικά ασφάλιστρα που εισπράττονται το χρονικό διάστημα $[0, t]$ και $S(t)$ η σύνθετη ανέλιξη για τις συνολικές αποζημιώσεις στο ίδιο χρονικό διάστημα.

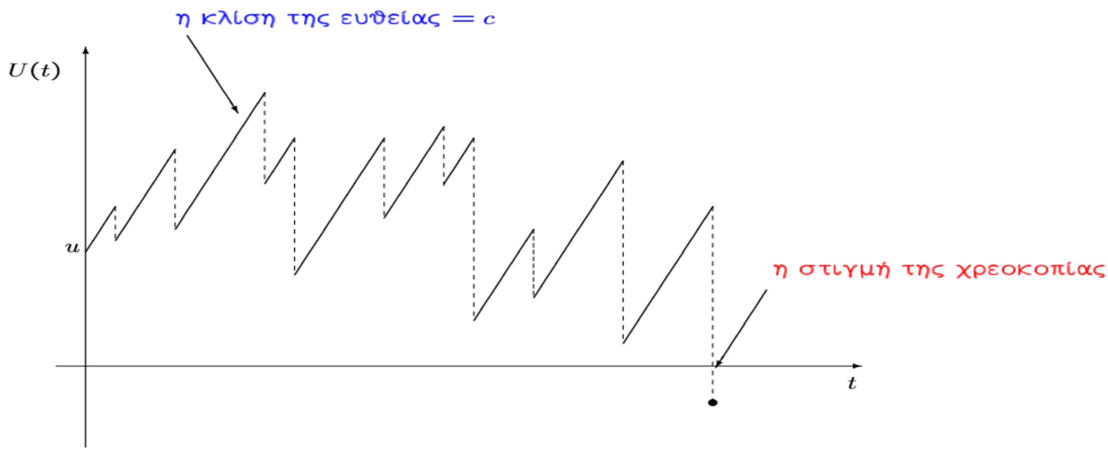
Το $U(t)$ καλείται αποθεματικό ή πλεόνασμα τη χρονική στιγμή t, ενώ το $U(0) = u$ ($u \geq 0$) λέγεται αρχικό αποθεματικό.

3.3 Το κλασσικό μοντέλο της θεωρίας κινδύνων

Στο κλασσικό μοντέλο της θεωρίας κινδύνων ισχύουν οι ακόλουθες υποθέσεις:

- $P(t) = ct$ για κάποιο $c > 0$ και c είναι μία σταθερά που εκφράζει το ασφάλιστρο που πληρώνεται στη μονάδα του χρόνου και ονομάζεται ένταση του ασφαλίστρου (premium rate). Δηλαδή ο ρυθμός είσπραξης ασφαλίστρων παραμένει σταθερός διαχρονικά, με άλλα λόγια το ασφάλιστρο που πληρώνεται από τον ασφαλισμένο στη μονάδα του χρόνου δεν μεταβάλλεται. Άρα από την σχέση αυτή προκύπτει ότι η $P(t)$ είναι μία γραμμική συνάρτηση.
- Οι τυχαίες μεταβλητές X_i που δηλώνουν το μέγεθος των αποζημιώσεων, είναι ανεξάρτητες και ισόνομες και είναι επίσης ανεξάρτητες από τον αριθμό των αποζημιώσεων $N(t)$ σε ένα διάστημα $[0, t]$.
- Η απαριθμήτρια ζημιών $\{N(t): t \geq 0\}$ είναι μία ανέλιξη Poisson με ένταση λ, όπου η σταθερά λ εκφράζει τον αναμενόμενο αριθμό ζημιών στη μονάδα του χρόνου, έτσι ώστε η ανέλιξη η $\{S(t): t \geq 0\}$ να είναι μία σύνθετη ανέλιξη Poisson.

Η διαδικασία του πλεονάσματος περιγράφεται από το παρακάτω διάγραμμα:



Σχήμα 3.1. Η ανέλιξη του πλεονάσματος (Πολίτης (2017))

Μία βασική υπόθεση που κάνουμε πάντοτε στο κλασικό μοντέλο της θεωρίας κινδύνων είναι ότι

$$c > \lambda\mu_1$$

όπου λ ο μέσος ρυθμός αποζημιώσεων στη μονάδα του χρόνου και μ_1 η μέση αποζημίωση. Δηλαδή το γινόμενο αυτό, δηλώνει τη μέση τιμή των εξόδων για τον ασφαλιστή στη μονάδα του χρόνου. Άρα, η παραπάνω ανίσωση απαιτεί τα έσοδα να υπερβαίνουν κατά μέσο όρο τα έξοδα σε κάθε χρονική μονάδα. Για το λόγο αυτό, αναφέρεται ως η συνθήκη του καθαρού κέρδους (net profit condition).

Από την παραπάνω βασική υπόθεση, προκύπτει ότι *το* c ικανοποιεί μια σχέση της μορφής

$$c = (1+\theta)\lambda\mu_1, \quad \theta > 0.$$

Το θ καλείται περιθώριο ασφαλείας ή συντελεστής ασφαλείας (premium loading factor), και δίνεται προφανώς από τον τύπο

$$\theta = \frac{c}{\lambda\mu_1} - 1.$$

Είναι προφανές ότι, ο συντελεστής θ παίρνει πάντα θετικές τιμές. Στην πραγματικότητα, ο συντελεστή αυτός εκφράζει πόσο μεγαλύτερα είναι τα έσοδα από τα έξοδα της εταιρείας κατά μέσο όρο σε ένα χαρτοφυλάκιο. Έτσι, το θ μπορεί να θεωρηθεί ότι εκφράζει το αναμενόμενο ποσοστό κέρδους για τον ασφαλιστή. Για το λόγο αυτό, το θ στην πράξη παίρνει συνήθως τιμές μεταξύ 0 και 1 γιατί διαφορετικά το συγκεκριμένο χαρτοφυλάκιο δεν είναι ανταγωνιστικό.

3.4 Πιθανότητα χρεοκοπίας

Πρέπει να σημειωθεί ότι στην πραγματικότητα τα ασφάλιστρα δεν είναι το μόνο έσοδο για έναν ασφαλιστικό οργανισμό, όπως και οι αποζημιώσεις δεν είναι το μόνο έξοδο. Έτσι, η μαθηματική χρεοκοπία δεν ταυτίζεται απαραίτητα με την πραγματική, είναι όμως ένα βασικό μέτρο που χρησιμοποιείται για να εξετάσουμε τη φερεγγυότητα του χαρτοφυλακίου και να διαμορφώσουμε ανάλογη οικονομική πολιτική.

Στη θεωρία κινδύνων, χρεοκοπία σημαίνει ότι κάποια τυχαία μελλοντική στιγμή T, έχουμε για πρώτη φορά πλεόνασμα αρνητικό, δηλαδή $U(T) < 0$ και $U(T-) > 0$.

Ειδικότερα έχουμε τα ακόλουθα:

- ✓ Χρόνος Χρεοκοπίας T

Ο ασφαλιστής, λοιπόν, θέλει να γνωρίζει την πιθανότητα χρεοκοπίας μέσα σε ένα πεπερασμένο ή άπειρο χρόνο. Παρακάτω, θα εξετάσουμε τις αντίστοιχες σχέσεις για την πιθανότητα χρεοκοπίας ως προς τον χρόνο T.

Αναλυτικότερα , για $t \geq 0$ ο χρόνος χρεοκοπίας μπορεί να οριστεί ως εξής :

$$T = \begin{cases} inf\{t \geq 0 : U(t) < 0\} \\ \\ \infty, \quad\quad U(t) \geq 0 \text{ για κάθε } t \geq 0. \end{cases}$$

Είναι, δηλαδή, η χρονική στιγμή κατά την οποία το πλεόνασμα γίνεται για πρώτη φορά αρνητικό, δηλαδή μετατρέπεται σε έλλειμμα.

Επίσης, είναι προφανές ότι ο χρόνος χρεοκοπίας είναι μια ελαττωματική τυχαία μεταβλητή αφού με θετική πιθανότητα μπορεί να πάρει την τιμή άπειρο, καθώς ισχύει η συνθήκη του καθαρού κέρδους.

$$P(T = \infty) > 0.$$

- ✓ Πλεόνασμα πριν τη χρεοκοπία

Αποτελεί το μέγεθος του πλεονάσματος την χρονική στιγμή ακριβώς πριν την έλευση του ζημιογόνου γεγονότος, δηλαδή

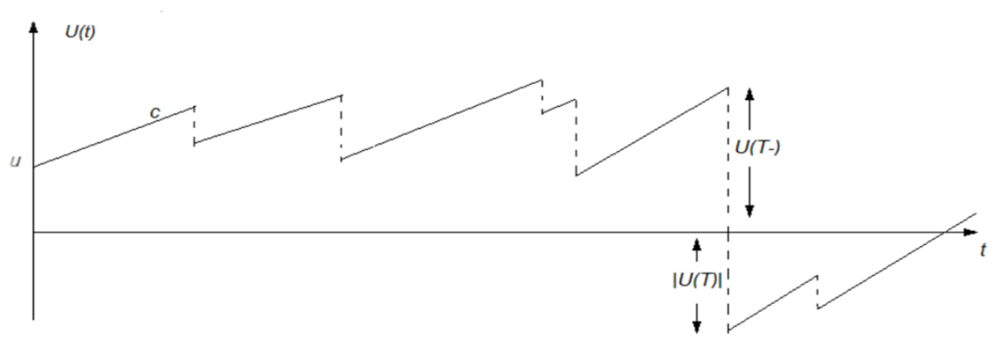
$U(T-) = \lim_{t \to T-} U(t)$.

Η τυχαία μεταβλητή $U(T-)$ παίρνει μόνο θετικές τιμές.

- ✓ Έλλειμμα τη στιγμή της χρεοκοπίας

Το εν λόγω έλλειμμα αναφέρεται στη σφοδρότητα της χρεοκοπίας (severity of ruin), δηλαδή το μέγεθος της πτώσης του πλεονάσματος κάτω από το μηδέν, δηλαδή

$$|U_T| = -U_T$$



Σχήμα 3.2. Στοχαστική διαδικασία πλεονάσματος έναντι της τ.μ. του ελλείματος

Παρατηρούμε ότι, το πλεόνασμα $U(t)$ παρουσιάζει άλματα τις χρονικές στιγμές επέλευσης των ζημιογόνων γεγονότων. Τα άλματα αυτά είναι του ίδιου μεγέθους με τα αντίστοιχα άλματα της διαδικασίας των συνολικών αποζημιώσεων.

Τώρα για $u \geq 0$, η πιθανότητα χρεοκοπίας σε άπειρο χρόνο ορίζεται ως

$$\psi(u) = P(T < \infty \mid U(0) = u) = P(U(t) < 0 \text{ για κάποιο } t \geq 0 | U(0) = u).$$

Ισχύει ότι:

- Η $\psi(u)$ είναι φθίνουσα συνάρτηση ως προς u.
- $\lim\limits_{u \to \infty} \psi(u) = 0$

Στην περίπτωση που δεν ισχύει η συνθήκη του καθαρού κέρδους, τότε η χρεοκοπία είναι βέβαιη, δηλαδή

$$\psi(u) = 1, \quad \text{για κάθε } u \geq 0.$$

Για $u \geq 0$, η πιθανότητα χρεοκοπίας σε πεπερασμένο χρόνο ορίζεται ως

$$\psi(u, t) = P(T < t \mid U(0) = u) = P(U(\tau) < 0) \text{ για κάποιο } 0 < \tau \leq t)$$

Ισχύει ότι:

- Η $\psi(u, t)$ είναι αύξουσα συνάρτηση ως προς t, ενώ είναι φθίνουσα ως συνάρτηση του u.
- $\lim\limits_{t \to \infty} \psi(u, t) = \psi(u)$.

Η πιθανότητα μη χρεοκοπίας $\delta(u)$ είναι η πιθανότητα

$$\begin{aligned}\delta(u) &= 1 - \psi(u) \\ &= Pr(T = \infty | U(0) = u) = Pr(U(t) \geq 0 \text{ για κάθε } t \geq 0 | U(0) = u).\end{aligned}$$

Διαπιστώνουμε ότι:

- Η $\delta(u)$ είναι αύξουσα συνάρτηση ως προς u
- $\lim_{u\to\infty} \delta(u) = 1$
- Η $\delta(u)$ είναι δεξιά συνεχής συνάρτηση.

Επισημαίνεται ότι η $\delta(u)$ αντιστοιχεί σε μία μεικτή κατανομή, αφού $\delta(0) > 0$, δηλαδή η πιθανότητα μη χρεοκοπίας με μηδέν αρχικό απόθεμα είναι θετική, δηλαδή έχει μάζα πιθανότητας στο σημείο $u = 0$, ενώ η $\delta(u)$ είναι συνεχής (δηλαδή έχει πυκνότητα) στο $(0, \infty)$.

Στο κλασικό μοντέλο, η πιθανότητα μη χρεοκοπίας $\delta(u)$ ικανοποιεί την ακόλουθη ολοκληροδιαφορική εξίσωση

$$\delta'(u) = \frac{\lambda}{c}\delta(u) - \int_0^u \delta(u-x)f(x)dx$$

όπου f είναι η συνάρτηση πυκνότητας των αποζημιώσεων.

Επίσης η $\delta(u)$ ικανοποιεί την εξίσωση

$$\delta(u) = \delta(0) + \frac{\lambda}{c}\int_0^u \delta(u-x)\bar{F}(x)dx, \qquad u \geq 0$$

όπου $\bar{F}(x) = 1 - F(x)$ η ουρά της κατανομής των αποζημιώσεων, είτε η κατανομή αυτή είναι διακριτή είτε συνεχής.

Παρατηρούμε ότι από την παραπάνω ολοκληροδιαφορική εξίσωση και τη συνθήκη του καθαρού κέρδους προκύπτει ότι

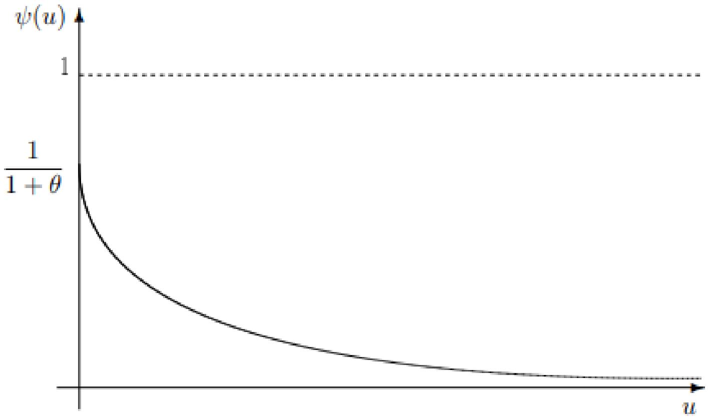
$$\delta(0) = 1 - \frac{\lambda\mu_1}{c} > 0$$

ενώ συναρτήσει του περιθωρίου ασφαλείας θ, ο παραπάνω τύπος γίνεται

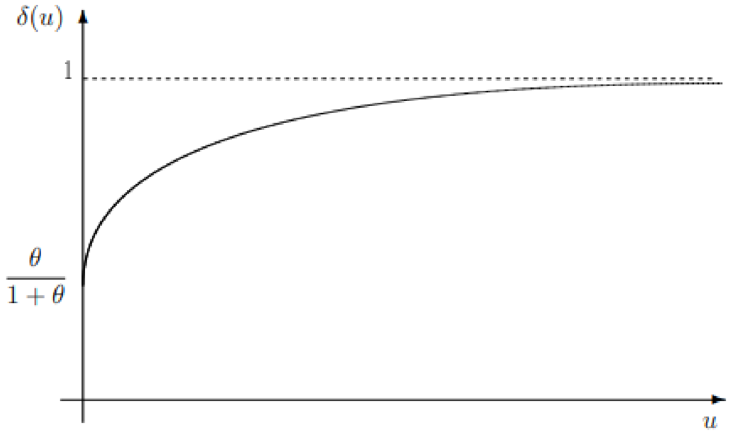
$$\delta(0) = \frac{\theta}{1+\theta}$$

άρα και η πιθανότητα χρεοκοπίας με αρχικό αποθεματικό μηδέν είναι

$$\psi(0) = \frac{1}{1+\theta}.$$



Σχήμα 3.3. Γραφική παράσταση της πιθανότητας χρεοκοπίας



Σχήμα 3.4. Γραφική παράσταση της πιθανότητας μη χρεοκοπίας

Μια ακόμη τυχαία μεταβλητή που σχετίζεται με τη στοχαστική διαδικασία του πλεονάσματος και παρουσιάζει έντονο ενδιαφέρον, είναι το μέγεθος της κάθετης πτώσης του πλεονάσματος κάτω από το αρχικό αποθεματικό.

Έστω οι τυχαίες μεταβλητές L_i $(i = 1,2, \dots)$ που εκφράζουν τις σταδιακές πτώσεις του πλεονάσματος από την τιμή του αρχικού αποθεματικού u μέχρι τη στιγμή της χρεοκοπίας T. Στην περίπτωση μη χρεοκοπίας οι τ.μ. L_i εκφράζουν το μέγεθος της πτώση του πλεονάσματος από την τιμή του αρχικού αποθεματικού u μέχρι την ελάχιστη τιμή που παίρνει η στοχαστική διαδικασία πλεονάσματος $\{U(t), t \geq 0\}$.

Ορίζουμε μία τυχαία μεταβλητή η οποία συμβολίζεται με K η οποία εκφράζει το πλήθος των κλιμακωτών υψών, σταδιακών πτώσεων του πλεονάσματος από την τιμή του αρχικού αποθεματικού u μέχρι τη στιγμή της χρεοκοπίας T, δηλαδή το πλήθος των μεταβλητών L_i (οι οποίες παίρνουν γνησίως θετική τιμή) σε μία ανέλιξη πλεονάσματος. Είναι σαφές ότι η μεταβλητή K είναι διακριτή, εφόσον παίρνει ακέραιες και μη αρνητικές τιμές.

Η κατανομή της K στο κλασικό μοντέλο της θεωρίας κινδύνων ακολουθεί τη γεωμετρική κατανομή με πιθανότητα επιτυχίας δ(0), δηλαδή

$$P(K=k)=\big(\psi(0)\big)^k\delta(0)=\left(\frac{1}{1+\theta}\right)^k\left(\frac{\theta}{1+\theta}\right),\qquad k=0,1,\dots\ .$$

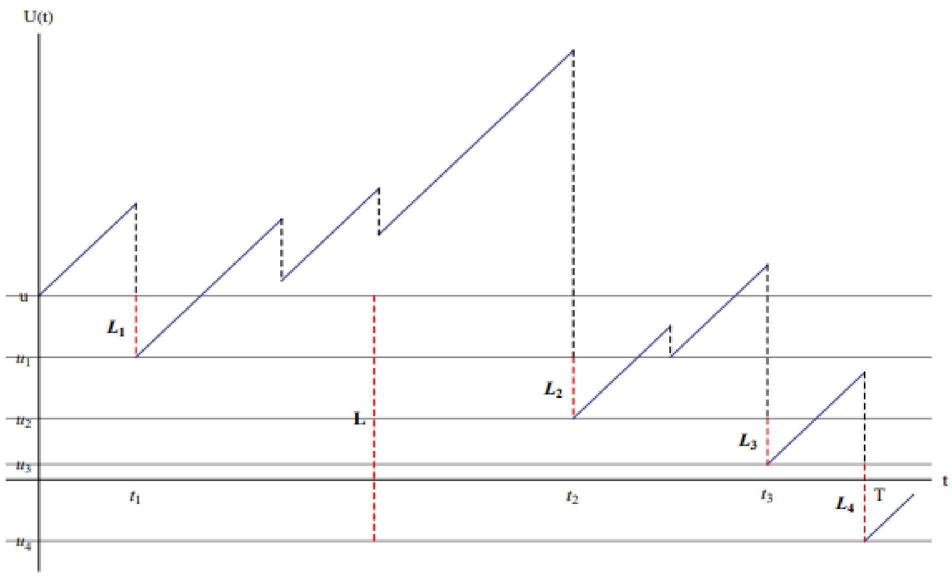
Όταν οι τ.μ. L_i $(i=1,2,\dots)$ είναι μεταξύ τους ανεξάρτητες και ισόνομες και ανεξάρτητες της τ.μ. K τότε η σύνθετη τυχαία μεταβλητή L που ορίζεται από τον τύπο

$$L=\begin{cases}0, & K=0\\ L_1+L_2+\cdots+L_K, & K\geq 1\end{cases}$$

ονομάζεται μέγιστη σωρευτική απώλεια.

Η τυχαία μεταβλητή L εκφράζει τη συνολική πτώση του πλεονάσματος κάτω από την τιμή του αρχικού αποθεματικού u έως την ελάχιστη τιμή της στοχαστικής διαδικασίας του πλεονάσματος $\{U(t),t\geq 0\}$.

Στο ακόλουθο σχήμα παρατίθεται η αναπαράσταση της ακολουθίας των κλιμακωτών υψών και παρουσιάζεται και η μέγιστη συσσωρευτική απώλεια, τα οποία έχουν ως εξής:



Σχήμα 3.5. Μέγιστη σωρευτική απώλεια στη διαδικασία πλεονάσματος, Πολίτης, Κ.(2012)

Η κατανομή της τ.μ. L είναι μια σύνθετη γεωμετρική κατανομή με

$$P(L=0)=Pr(K=0)=\delta(0)=\frac{\theta}{1+\theta}$$

και

$$P(L\leq u))=\delta(u),$$
$$P(L>u))=\psi(u)=\bar{F}_L(u).$$

Για την κατανομή των συνεχών τυχαίων μεταβλητών L_i έχουμε ότι

$$F_{L_i}(x)=\frac{1}{\mu}\int_0^x F_X(y)dy\,,\qquad i=1,2,\dots$$

3.5 Ο συντελεστής προσαρμογής R

Στο κλασσικό μοντέλο της θεωρίας κινδύνων ο συντελεστής προσαρμογής R ορίζεται να είναι η μοναδική θετική λύση της εξίσωσης

$$\lambda M_X(r)-\lambda-cr=0$$

ως προς r, ή ισοδύναμα της εξίσωσης

$$M_X(r)=1+(1+\theta)\mu_1 r$$

όπου

$$M_X(r)=E(e^{rX})=\int_0^\infty e^{rX}f(x)dx$$

είναι η ροπογεννήτρια της κατανομής των αποζημιώσεων. Η εξίσωση αυτή είναι γνωστή ως εξίσωση του Lundberg ή εξίσωση του συντελεστή προσαρμογής. Ισοδύναμες εξισώσεις οι οποίες μπορεί να είναι χρήσιμες στον υπολογισμό του R είναι η

$$\int_0^\infty e^{rX}\bar{F}(x)=\frac{c}{\lambda}$$

και η

$$\int_0^\infty e^{rX}dH(x)=\frac{c}{\lambda\mu_1}$$

όπου

$$H(x)=\int_0^x\frac{\bar{F}(y)}{\mu}dy.$$

Από την σχέση

$$M_X(r) = 1 + (1+\theta)\mu_1 r$$

αντιλαμβανόμαστε ότι ο συντελεστής προσαρμογής R είναι ανεξάρτητος της παραμέτρου λ. Δηλαδή, εξαρτάται μόνο από τη μέση τιμή των αποζημιώσεων και το περιθώριο ασφαλείας θ, και απαραίτητη προϋπόθεση για την ύπαρξή του είναι η ύπαρξη της ροπογεννήτριας της κατανομής των αποζημιώσεων.

Ο συντελεστής προσαρμογής R είναι μία ποσότητα με ιδιαίτερη σημασία στη θεωρία κινδύνων γενικά, αλλά και τη θεωρία χρεοκοπίας ιδιαίτερα.

Οι κύριοι λόγοι είναι οι δύο παρακάτω:

1. Η ανισότητα Lundberg

Η ανισότητα Lundberg δίνει πληροφορίες για την πιθανότητα χρεοκοπίας με τη βοήθεια ενός άνω φράγματος συναρτήσει του αρχικού αποθεματικού και του συντελεστή προσαρμογής. Η ανισότητα είναι η ακόλουθη

$$\psi(u) \leq e^{-Ru}, \quad u \geq 0.$$

2. Ο ασυμπτωτικός τύπος (Cramér – Lundberg)

Η πιθανότητα χρεοκοπίας $\psi(u)$ ικανοποιεί την ακόλουθη ασυμπτωτική σχέση

$$\psi(u) \sim C e^{-Ru}$$

καθώς $u \to \infty$, ή ισοδύναμα

$$\lim_{u\to\infty} \frac{\psi(u)}{e^{-Ru}} = C$$

όπου C μία θετική σταθερά που δίνεται από τον τύπο

$$C = \frac{\theta \mu_1}{R \int_0^\infty x e^{Rx} \bar{F}(x) dx}.$$

Από τον ασυμπτωτικό τύπο των Cramér – Lundberg προκύπτει ότι

$$\lim_{u\to\infty} \frac{\psi(u)}{C e^{-Ru}} = 1.$$

Παρατηρήσεις:

1. Για δεδομένη τιμή R η πιθανότητα χρεοκοπίας θα ελαττώνεται αναλογικά με το πόσο μεγεθύνεται το αρχικό αποθεματικό

2. Για δεδομένη τιμή του αρχικού αποθεματικού η πιθανότητας χρεοκοπίας θα μικραίνει αναλογικά με το πόσο μεγαλώνει η τιμή του R.

3.6 Ένα άνω φράγμα για το συντελεστή προσαρμογής R

Στο κλασσικό μοντέλο της θεωρίας χρεοκοπίας εντοπίζουμε περιπτώσεις όπου ο συντελεστής προσαρμογής υπάρχει, ωστόσο δεν μπορεί να υπολογιστεί με ακρίβεια. Έτσι για την διευκόλυνση των υπολογισμών υπολογίζουμε ένα άνω φράγμα για το R.

Χρησιμοποιούμε την αρχική εξίσωση της παραγράφου σε συνδυασμό με την εξίσωση της έντασης ασφαλίστρου, και έχουμε ότι

$$\lambda + cR = \lambda M_X(R)$$

όπου υποθέτουμε ότι οι αποζημιώσεις έχουν πυκνότητα f. Συνεπώς

$$\lambda + cR = \lambda \int_0^\infty e^{Rx} f(x)dx > \lambda \int_0^\infty \left(1 + Rx + \frac{1}{2}R^2x^2\right) f(x)dx$$

$$= \lambda \left[\int_0^\infty f(x)dx + \int_0^\infty Rxf(x)dx + \int_0^\infty \frac{1}{2}R^2x^2f(x)dx\right]$$

$$= \lambda\left(1 + R\mu_1 + \tfrac{1}{2}R^2\mu_2\right).$$

Επομένως

$$2(c + \lambda\mu_1)R > \lambda R^2\mu_2$$

ή ισοδύναμα

$$R < \frac{2(c + \lambda\mu_1)}{\lambda\mu_2}.$$

Χρησιμοποιώντας τη σχέση $c = (1 + \theta)\lambda\mu_1$, $\theta > 0$ καταλήγουμε στο άνω φράγμα του συντελεστή προσαρμογής που είναι

$$R < \frac{2\theta\mu_1}{\mu_2}.$$

3.7 Εφαρμογές

3.7.1 Συντελεστής Προσαρμογής *R*

Η συνάρτηση `adjCoef` του πακέτου `actuar` υπολογίζει τον συντελεστή προσαρμογής R. Η σύνταξη της συνάρτησης `adjCoef` είναι η ακόλουθη:

```
adjCoef(mgf.claim, mgf.wait = mgfexp, premium.rate,
upper.bound, h, reinsurance = c("none", "proportional",
"excess-of-loss"), from, to, n = 101)
```

όπου

- `mgf.claim:` μία έκφραση γραμμένη ως συνάρτηση του x ή των x και y, ή εναλλακτικά το όνομα μιας συνάρτησης η οποία δίνει ως αποτέλεσμα τη ροπογεννήτρια συνάρτηση των απαιτήσεων.
- `mgf.wait:` Μία έκφραση γραμμένη ως συνάρτηση του x ή εναλλακτικά το όνομα μιας συνάρτησης η οποία δίνει ως αποτέλεσμα η οποία δίνει ως αποτέλεσμα τη ροπογεννήτρια συνάρτηση των ενδιάμεσων χρόνων αναμονής των απαιτήσεων.
- `premium.rate:` Αν το όρισμα `reinsurance = "none",` είναι η ένταση ασφαλίστρου. Διαφορετικά μία έκφραση γραμμένη ως συνάρτηση του y ή εναλλακτικά το όνομα μια συνάρτησης η οποία δίνει ως αποτέλεσμα τη συνάρτηση έντασης του ασφαλίστρου.
- `upper.bound:` ένα άνω όριο για τον συντελεστή προσαρμογής,

Παράδειγμα 3.1

Η απαριθμήτρια ζημιών είναι μία ανέλιξη Poisson με ένταση $\lambda = 4$ και η ένταση ασφαλίστρου $c = 7$. Ακόμα, υποθέτουμε ότι η κατανομή του ποσού της ατομικής απώλειας είναι η ακόλουθη

$$P(X = 1) = 0.6, \qquad P(X = 2) = 0.4.$$

Να προσδιοριστεί ο συντελεστής προσαρμογής R.

<u>**Λύση**</u>

Έχουμε τον ακόλουθο κώδικα

```
library(actuar)
mgfx <- function(x) 0.6 * exp(x) + 0.4 * exp(2 * x)
adjCoef(mgf.claim=mgfx(x), mgf.wait = mgfexp(x, 4),
premium.rate = 7, upper = 0.3182)
[1] 0.2702897
```

Παράδειγμα 3.2

Έστω ότι η κατανομή των αποζημιώσεων είναι η εκθετική κατανομή με παράμετρο $\beta = 2$, και η απαριθμήτρια ζημιών είναι μία ανέλιξη Poisson με ένταση $\lambda = 3$. Να προσδιοριστεί ο συντελεστής προσαρμογής R για ένταση ασφαλίστρου $c = 2.4$

<u>Λύση</u>

Έχουμε τον ακόλουθο κώδικα

```
c=2.4
adjCoef(mgf.claim = mgfexp(x,3), mgf.wait = mgfexp(x, 2),
premium.rate = c, upper = 3)
[1] 2.166667
```

Παράδειγμα 3.3

Η κατανομή των ενδιάμεσων χρόνων αναμονής μεταξύ διαδοχικών απαιτήσεων είναι η ανάστροφη Γάμμα με παράμετρο μορφής a και παράμετρο κλίμακας s, δηλαδή η κατανομή με συνάρτηση πυκνότητας

$$f(x) = \frac{\left(\frac{s}{x}\right)^a \exp\left(-s/x\right)}{x\Gamma(a)}, \quad x > 0.$$

Η κατανομή των αποζημιώσεων X είναι μια γενικευμένη Erlang κατανομή (ή υποεκθετική κατανομή) με παραμέτρους $\lambda_1, \lambda_2, \lambda_3$, δηλαδή έχει ροπογεννήτρια συνάρτηση

$$M_X(t) = E(e^{tX}) = \prod_{i=1}^{3}\left(\frac{\lambda_i}{\lambda_i - t}\right), \quad t < \min(\lambda_1, \lambda_2, \lambda_3).$$

Να προσδιοριστεί ο συντελεστής προσαρμογής R για ένταση ασφαλίστρου $c = 1.1$, $a = 2$, s =11/6, $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$.

<u>Λύση</u>

Έχουμε τον ακόλουθο κώδικα

```
library(actuar)
mgf <- function(x) 1/(1 - x) * 2/(2 - x) * 3/(3 - x)
adjCoef(mgf.claim=mgf, mgf.wait = mgfinvgamma(x, shape=2,
scale=11/6), premium.rate=1.1, upper = 1)
[1] 0.03428817
```

3.7.2 Πιθανότητα χρεοκοπίας

Η κύρια δυσκολία για τον υπολογισμό της πιθανότητας χρεοκοπίας σε άπειρο χρόνο έγκειται στην έλλειψη συγκεκριμένων τύπων. Ωστόσο, εάν οι χρόνοι μεταξύ των απαιτήσεων ακολουθούν την εκθετική κατανομή $Exp(\lambda)$ και τα ποσά των απαιτήσεων ακολουθούν την εκθετική κατανομή $Exp(\beta)$, τότε

$$\psi(u) = \frac{\lambda}{c\beta}\exp\left(-\left(\beta - \frac{\lambda}{c}\right)u\right).$$

Στη γενική περίπτωση για τον υπολογισμό της πιθανότητας χρεοκοπίας θα χρησιμοποιήσουμε την συνάρτηση `ruin` του πακέτου `actuar` που έχει την ακόλουθη σύνταξη;

```
ruin(claims = c("exponential", "Erlang", "phase-type"), par.claims, wait = c("exponential", "Erlang", "phase-type"), par.wait, premium.rate = 1, tol = sqrt(.Machine$double.eps), maxit = 200L, echo = FALSE)
```

όπου

- `claims`: η κατανομή των απαιτήσεων
- `wait`: η κατανομή των ενδιάμεσων χρόνων αναμονής μεταξύ διαδοχικών απαιτήσεων
- `par.claims, par.wait`: λίστα που περιέχει τις παραμέτρους των κατανομών
- `premium.rate`: η ένταση του ασφαλίστρου
- `tol, maxit, echo`: αντίστοιχα το επίπεδο ανοχής των κριτηρίων διακοπής και ο μέγιστος αριθμός των επαναλήψεων

Παράδειγμα 3.4

Θέλουμε να βρούμε την πιθανότητα χρεοκοπίας στην περίπτωση που η κατανομή των αποζημιώσεων είναι μια μείξη εκθετικών κατανομών, δηλαδή

$$f(x) = \sum_{i=1}^{n} \pi_i b_i e^{-b_i x}, \quad x \geq 0,$$

όπου $\pi_i \geq 0$ και $\sum_{i=1}^{n} \pi_i = 1$. Σε αυτή την περίπτωση η πιθανότητα χρεοκοπίας (Gerber et. al (1987)) δίνεται από τον τύπο

$$\psi(u)=\sum_{k=1}^{n} C_k e^{-r_k u}, \quad u \geq 0$$

όπου $r_1, r_2, \dots, r_n$ οι ρίζες της εξίσωσης του Lundberg, και

$$C_k=\sum_{j=1}^{n}\frac{C_{jk}}{b_j}, \qquad C_{jk}=\frac{\pi_j/(b_j-r_k)}{\sum_{m=1}^{n}\frac{\pi_m}{(b_m-r_k)^2}}.$$

Για παράδειγμα έστω ότι η συνάρτηση πυκνότητας της τυχαίας μεταβλητής X είναι η

$$f(x)=\left(\frac{1}{2}\right)3e^{-3x}+\left(\frac{1}{2}\right)7e^{-7x}, \quad x \geq 0,$$

δηλαδή $(\pi_1,\pi_2)=\left(\frac{1}{2},\frac{1}{2}\right)$, και $(b_1,b_2)=(3,7)$. Η μέση τιμή και η ροπογεννήτρια συνάρτηση των αποζημιώσεων είναι

$$\mu=\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{7}\right)=\frac{5}{21}$$

και

$$M_X(r)=\left(\frac{1}{2}\right)\left(\frac{3}{3-r}\right)+\left(\frac{1}{2}\right)\left(\frac{7}{7-r}\right).$$

Συνεπώς από την εξίσωση του Lundberg, $M_X(r)=1+(1+\theta)r\mu$, για περιθώριο ασφαλείας $\theta=2/5$, προκύπτει ότι

$$1+\left(1+\frac{2}{5}\right)\left(\frac{5}{21}\right)r=\left(\frac{1}{2}\right)\left(\frac{3}{3-r}\right)+\left(\frac{1}{2}\right)\left(\frac{7}{7-r}\right)$$

ή ισοδύναμα

$$2r^3+14r^2+12r=0.$$

Η μία ρίζα είναι το μηδέν η οποία απορρίπτεται και οι άλλες δύο ρίζες είναι $r_1=1$ και $r_2=6$. Για τα C_1, C_2 έχουμε

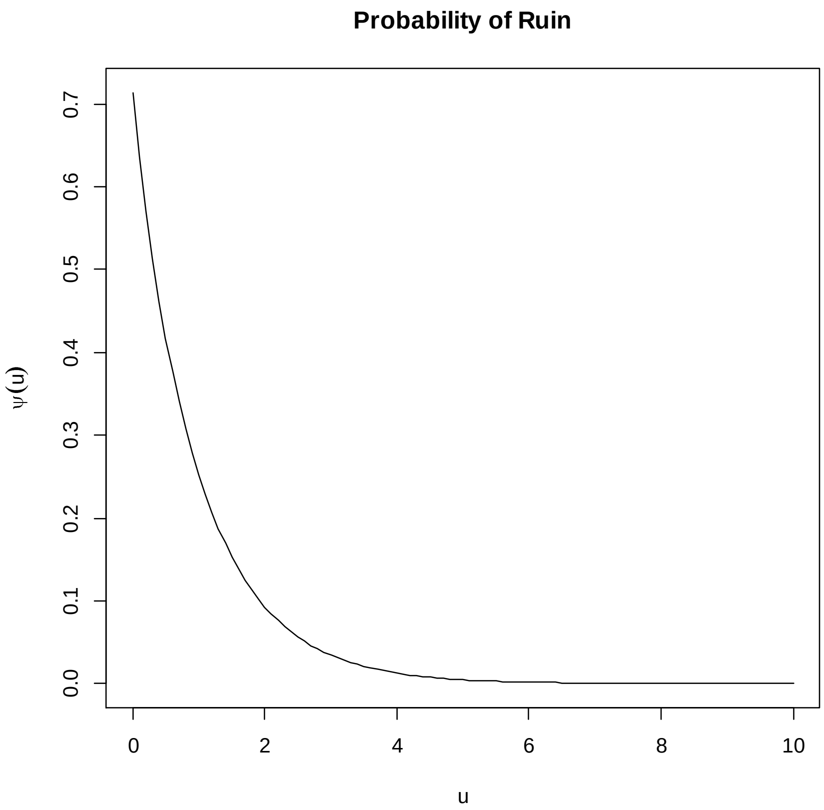
$$C_1=\frac{24}{35}, \quad C_2=\frac{1}{35}.$$

Έτσι η πιθανότητα χρεοκοπίας είναι

$$\psi(u) = \frac{24}{35}e^{-u} + \frac{1}{35}e^{-6u}, \qquad u \geq 0.$$

Με τη συνάρτηση `ruin` έχουμε τον ακόλουθο κώδικα:

```
psi <- ruin(claims = "exp", par.claims = list(rate = c(3,
7), w = c(0.5,0.5)), wait = "exp",
par.wait = list(rate = 3), premium.rate = 1)
u <- 0:10
psi(u)
 [1] 7.142857e-01 2.523310e-01 9.280151e-02 3.413970e-02
1.255930e-02
 [6] 4.620307e-03 1.699716e-03 6.252905e-04 2.300315e-04
8.462387e-05
[11] 3.113138e-05
(24 * exp(-u) + exp(-6 * u))/35
[1]  7.142857e-01  2.523310e-01  9.280151e-02  3.413970e-02
1.255930e-02
 [6] 4.620307e-03 1.699716e-03 6.252905e-04 2.300315e-04
8.462387e-05
[11] 3.113138e-05
plot(psi, from = 0, to = 10)
```



Παράδειγμα 3.4

Αντίστοιχα με το παραπάνω παράδειγμα έστω ότι θέλουμε να βρούμε την πιθανότητα χρεοκοπίας στην περίπτωση που η κατανομή των αποζημιώσεων είναι μείξη κατανομών Erlang, δηλαδή

$$f(x) = \sum_{i=1}^{n} \pi_i b_i \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad x \geq 0,$$

όπου $\pi_i \geq 0$ και $\sum_{i=1}^{n} \pi_i = 1$.

Για παράδειγμα έστω ότι η συνάρτηση πυκνότητας της τυχαίας μεταβλητής X είναι η

$$f(x) = \left(\frac{1}{3}\right) \lambda^2 x e^{-\lambda \chi} + \left(\frac{2}{3}\right) \lambda^4 x^3 e^{-\lambda x}$$

δηλαδή $(\pi_1, \pi_2) = \left(\frac{1}{3}, \frac{2}{3}\right)$, και $(b_1, b_2) = (1, 3)$.

Για τον υπολογισμό της πιθανότητας χρεοκοπίας θα χρησιμοποιήσουμε ομοίως την συνάρτηση `ruin` όπως φαίνεται στον ακόλουθο κώδικα:

```
> psi <- ruin(claims = "Erlang",
```

```
par.claims = list(shape = c(2, 4), rate = c(1, 3),
w = c(1, 2)/3),
wait = "Erlang",
par.wait = list(shape = 2, rate = 1),
premium = 1.2)
psi(0:10)
[1]    0.57635887    0.43252506    0.31954822    0.23785313
0.17755251 0.13260513
 [7]    0.09903136    0.07395051    0.05521804    0.04122935
0.03078404
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

[1] Αντζουλάκος Δ., 2016. Σημειώσεις Π.Μ.Σ. «Αναλογιστική Επιστήμη & Διοικητική Κινδύνου», Τμήμα Στατιστικής & Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

[2] Κούτρας Μ. (2004). Εισαγωγή στις πιθανότητες. Σταμούλη Α.Ε. Πειραιάς.

[3] Πολίτης, Κ. (2012). Εισαγωγή στη θεωρία συλλογικού κινδύνου. Εκδόσεις Σταμούλη.

[4] Κουτσόπουλος Κ.Ι., 1999. Αναλογιστικά Μαθηματικά, Μέρος Ι, Θεωρία των Κινδύνων, Εκδόσεις Συμμετρία

[5] Χατζηκωνσταντινίδης Ε. (2016). Θεωρία Κινδύνου. Σημειώσεις ΠΜΣ Αναλογιστικής Επιστήμης και Διοικητικής Κινδύνου, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Ξένη

[6] Goulet V., Auclair S., Dutang C., Milhaud X., Ouellet T., Parent A., Pigeon M., Pouliot L.P., 2018,Package "actuar", Actuarial Functions and Heavy Tailed Distributions, CRAN.

[7] Klugman, S. A., Panjer, H. H., Willmot, G. E., 2012, Loss Models, From Data to Decisions, Fourth Edition, Wiley.

[8] Charpentier, A. (2014). Computational Actuarial Science with R, CRC Press

Διπλωματικές Εργασίες

[9] Μπιλάς Ε., 2018. Η κλάση κατανομών (α,β,0) του Panjer: Θεωρία, γενικεύσεις και εφαρμογές στη θεωρία συλλογικού κινδύνου, Διπλωματική Εργασία, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, ΠΜΣ στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου, Πανεπιστήμιο Πειραιώς.

[10] Μαμάνδρα Ε., 2016. Αναδρομικός υπολογισμός της κατανομής των συνολικών αποζημιώσεων: Μια επισκόπηση, Διπλωματική Εργασία, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, ΠΜΣ στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου, Πανεπιστήμιο Πειραιώς.

[11] Κουτσογιαννάκης Γ., 2019. Κατανομές τύπου φάσεων και οι εφαρμογές τους στην αναλογιστική επιστήμη, Διπλωματική Εργασία, Τμήμα Στατιστικής

και Ασφαλιστικής Επιστήμης, ΠΜΣ στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου, Πανεπιστήμιο Πειραιώς.

Ιστοσελίδες

https://cran.r-project.org/web/packages/actuar/vignettes/risk.pdf

https://cran.r-project.org/

https://cran.r-project.org/web/packages/actuar/vignettes/simulation.pdf

https://cran.r-project.org/web/packages/actuar/index.html