

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ
ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



Στοχαστική Προσομοίωση Monte Carlo

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κέντας Γεώργιος

ΑΜ:09105187

Επιβλέπων Καθηγητής:

Φουσκάκης Δημήτριος

ΑΘΗΝΑ 2012

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε υπό την επίβλεψη του Καθηγητή κ. Δημήτρη Φουσκάκη, τον οποίο ευχαριστώ θερμά για τη συνεχή καθοδήγηση και τις πολύτιμες συμβουλές του καθ' όλη την διάρκεια της μελέτης αυτής.

Ευχαριστώ θερμά τους φίλους μου, που ήταν δίπλα μου πάντα, για όλη την βοήθεια που μου προσέφεραν, όποτε το χρειάστηκα, καθώς και για την απαραίτητη ψυχική δύναμη που αντλούσα από αυτούς κάθε στιγμή!

Τέλος θα ήθελα να ευχαριστήσω ολόψυχα τους γονείς μου για την υπομονή τους, τους κόπους και τις θυσιές που έκαναν για μένα και την ηθική στήριξη που μου προσέφεραν. Στην οικογένειά μου αφιερώνω την εργασία αυτή, γιατί είναι πάντα θερμοί υποστηρικτές μου σε κάθε βήμα και κάθε προσπάθειά μου.

Π ε ρ ί λ η ψ η

Στην εργασία αυτή στο πρώτο κεφάλαιο θα δούμε μια σύντομη ιστορική αναδρομή της εξέλιξης των μεθόδων Monte Carlo καθώς και χρήσεις αυτών. Στο δεύτερο κεφάλαιο θα παρουσιάσουμε, χωρίς να επικεντρωθούμε σε αυτό, βασικές μεθόδους παραγωγής «ψευδοτυχαίων» αριθμών. Ακολούθως θα δούμε τις βασικές μεθόδους προσομοίωσης από κατανομές και θα τις εφαρμόσουμε σε γνωστές μας κατανομές. Στο τρίτο κεφάλαιο θα δούμε μεθόδους προσομοίωσης βασικών στοχαστικών ανελίξεων και θα αναλύσουμε τα αποτελέσματα που παίρνουμε από την προσομοίωση. Στα τέταρτο κεφάλαιο θα δούμε την Monte Carlo ολοκλήρωση και θα παρουσιάσουμε τους βασικούς MCMC αλγόριθμους, Metropolis-Hastings και Gibbs, με χρήση παραδειγμάτων. Στο πέμπτο και τελευταίο κεφάλαιο θα δούμε τεχνικές ελάττωσης της διασποράς. Τέλος ακολουθεί η βιβλιογραφία που περιλαμβάνει τις πηγές που χρησιμοποιήσαμε για την συγγραφή αυτής της διπλωματικής. Σε όλα τα παραδείγματα δίνεται και κώδικας γραμμένος στην γλώσσα προγραμματισμού R.

A b s t r a c t

In this thesis we focus on Stochastic Simulation and Monte Carlo methods, which are useful in problems that are difficult to be solved analytically. Monte Carlo methods have a variety of applications in many fields such as mathematics, physics, biology chemistry and economics and can be used to simulate complicated systems.

In the first chapter we make an introduction to Monte Carlo methods, present their historical evolution and some applications. In the second chapter we present algorithms for producing “pseudorandom numbers” and we analyze the basic Monte Carlo methods for producing random numbers from a given distribution. At the end of the chapter we use these methods to produce random numbers from commonly used distributions, both continuous and discrete. In the third chapter we simulate from basic stochastic processes (Poisson Process, Brown Movement and Random Walk). In the fourth chapter we introduce Monte Carlo Integration and basic MCMC algorithms (Metropolis-Hastings and Gibbs Sampler). Then we use these algorithms to simulate from distributions, and analyze the results. In the final chapter we present techniques that decrease the variance of the estimators. In every method analyzed in this thesis we provide examples and computer coding in R statistical language.

Περιοχόμενα

ΚΕΦΑΛΑΙΟ 1 - ΕΙΣΑΓΩΓΗ.....	5
ΚΕΦΑΛΑΙΟ 2 – ΜΕΘΟΔΟΙ ΠΡΟΣΟΜΟΙΩΣΗΣ MONTE CARLO.....	8
1.1 Εισαγωγή.....	8
1.2 Παραγωγή «Τυχαίων Αριθμών».....	8
1.2.1 Πολλαπλασιαστική Μέθοδος – Multiplicative Congruential Method	9
1.2.2 Μεικτή μέθοδος – Mix congruential method.....	11
1.3 Μέθοδοι Παραγωγής Τυχαίων Αριθμών από Κατανομές	12
1.3.1 Μέθοδος της Αντιστροφής - Inversion.....	12
1.3.2 Μέθοδος της Σύθεσης – Composition Method.....	17
1.3.3 Μέθοδος της Απόρριψης – Acceptance-Rejection Method.....	20
1.4 Προσομοίωση από Γνωστές Κατανομές	27
1.4.1 Προσομοίωση Συνεχών Κατανομών	27
1.4.2 Προσομοίωση Διακριτών Κατανομών.....	44
1.4.3 Προσομοίωση Άλλων Γνωστών Κατανομών.....	49
ΚΕΦΑΛΑΙΟ 3 – ΠΡΟΣΟΜΟΙΩΣΗ ΒΑΣΙΚΩΝ ΣΤΟΧΑΣΤΙΚΩΝ ΑΝΕΛΙΞΕΩΝ.....	51
2.1 Εισαγωγή.....	51
2.2 Προσομοίωση Ανέλιξης Poisson.....	52
2.2.1 Ομογενής Ανέλιξη Poisson.....	52
2.2.2 Μη Ομογενής Ανέλιξη Poisson	56
2.3 Προσομοίωση Κίνησης Brown.....	64
2.4 Προσομοίωση Τυχαίου Περίπατου με Απορροφητικά Φράγματα	67
ΚΕΦΑΛΑΙΟ 4 – MONTE CARLO ΜΕ ΤΗΝ ΧΡΗΣΗ ΜΑΡΚΟΒΙΑΝΩΝ ΑΛΥΣΙΔΩΝ.....	71
3.1 Εισαγωγή.....	71
3.2 Ολοκλήρωση Monte Carlo	72
3.3 Αλγόριθμος Metropolis-Hastings.....	76
3.4 Δειγματολήπτης Gibbs	81

ΚΕΦΑΛΑΙΟ 5 – ΤΕΧΝΙΚΕΣ ΕΛΑΤΤΩΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ.....	88
4.1 Εισαγωγή.....	88
4.2 Αντιθετικές Τυχαίες Μεταβλητές.....	90
4.3 Ρυθμιστικές Μεταβλητές.....	94
4.4 Δειγματοληψία Σπουδαιότητας.....	97
Βιβλιογραφία.....	102

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Πολλές φορές η επιστημονική κοινότητα έρχεται αντιμέτωπη με προβλήματα που δεν λύνονται αναλυτικά. Στην προσπάθεια επίλυσης τους χρησιμοποιούνται πολλές φορές προσεγγιστικές μέθοδοι, όπως για παράδειγμα η αριθμητική ανάλυση. Με την εξέλιξη όμως των υπολογιστών τα τελευταία χρόνια οι επιστήμονες έστρεψαν την προσοχή τους σε νέους δρόμους, και επέβαλαν την ανάπτυξη της στατιστικής και διαφόρων υπολογιστικών τεχνικών. Οι μέθοδοι προσομοίωσης Monte Carlo, που είναι και το θέμα αυτής της διπλωματικής είναι μια κατηγορία υπολογιστικών αλγόριθμων που μιμούνται την συμπεριφορά διαφόρων φυσικών και μαθηματικών συστημάτων.

Τα ερεθίσματα για την ανάπτυξη αυτών των υπολογιστικών τεχνικών, φαίνεται να δόθηκαν πολύ νωρίς από προβλήματα της θεωρίας πιθανοτήτων, όπως η βελόνα του Buffon, αλλά πιο φανερά η μελέτη τους άρχισε από την περίοδο ανάπτυξης των ηλεκτρονικών υπολογιστών. Η πρώτη γνωστή χρήση αυτών των υπολογιστικών τεχνικών έγινε από τον Enrico Fermi, γνωστό και ως τον πατέρα της βόμβας υδρογόνου, σε μια εργασία του (την οποία δεν δημοσίευσε) για την εκτίμηση των ιδιοτήτων των νετρονίων. Αργότερα οι μέθοδοι Monte Carlo έπαιξαν καθοριστικό ρόλο στο περίφημο Manhattan Project που αφορούσε την κατασκευή της πρώτης ατομικής βόμβας, κατά την διάρκεια του Β' Παγκοσμίου Πολέμου.

Η αρχή της σε βάθος μελέτης αυτών των αλγορίθμων έγινε από τον πολωνό μαθηματικό Stanislaw Ulam. Ο ίδιος αναφέρει ότι η έμπνευση του ήρθε το 1946 καθώς ήταν άρρωστος και ενώ έπαιζε ένα είδος πασιέντζας με πολύ λίγες πιθανότητες νίκης. Προσπάθησε λοιπόν να υπολογίσει αυτή την πιθανότητα. Η σύνθετη όμως δομή του παιχνιδιού έκανε αδύνατο τον αναλυτικό υπολογισμό αυτής της πιθανότητας. Έτσι σκέφτηκε να παίξει εκατό γύρους και να παρατηρήσει «εμπειρικά» πόσο συχνά κερδίζει στην πράξη. Έπειτα διερωτήθηκε αν μπορεί να εκτιμήσει μια συνάρτηση

πιθανότητας χρησιμοποιώντας δείγματα από την εμπειρική συνάρτηση πιθανότητας. Αργότερα τον ίδιο χρόνο περιέγραψε τις σκέψεις του στον John von Newman, με τον οποίο και άρχισαν να εργάζονται στο θέμα.

Αφού ήθελαν να κρατήσουν την εργασία τους κρυφή έπρεπε να διαλέξουν ένα κωδικό όνομα. Έτσι διάλεξαν την κωδική ονομασία Monte Carlo, που παραπέμπει στο καζίνο Monte Carlo στο Μονακό, που ο θείος του Ulam δανειζόταν χρήματα για να τζογάρει. Αρχικά χρησιμοποιούσαν λίστες από πραγματικά τυχαίους αριθμούς, οι οποίες παράγονταν με τεχνικές όπως το ρίξιμο ζαριών ή ανακατεύοντας τράπουλες. Αφού αυτό έκανε χρονοβόρα την διαδικασία, ο Von Newman ανέπτυξε ένα αλγόριθμο ο οποίος κατασκεύαζε «ψευδοτυχαίους» αριθμούς.

Η βασική αρχή λειτουργίας των Monte Carlo αλγόριθμων στην αρχή της εξέλιξης τους, βασιζόταν στην εξαγωγή ενός πλήθους δειγμάτων ανεξάρτητων και ισόνομων, ώστε να υπάρχει ασυμπτωτική σύγκλιση της εκτιμώμενης κατανομής στην πραγματική βάση των νόμων των μεγάλων αριθμών. Οι πρώτες χρήσεις της μεθόδου στα πρώτα χρόνια της εξέλιξής της, εκτός από το στρατιωτικό πεδίο, εστιάζονταν στην εκτίμηση μέσων τιμών και τυπικών αποκλίσεων καθώς και στον υπολογισμό τιμών από την εκ των υστέρων κατανομή σε πιθανοκρατικά προβλήματα. Αργότερα το 1953 ο Nicolas Metropolis χρησιμοποίησε δείγμα από Μαρκοβιανή αλυσίδα της οποίας η στάσιμη κατανομή ήταν η κατανομή που ήθελε να προσομοιώσει. Αυτή ήταν η αρχή μιας νέας κατηγορίας αλγόριθμων, των Monte Carlo Markov Chain (MCMC).

Σήμερα οι αλγόριθμοι Monte Carlo χρησιμοποιούνται σε ένα ευρύ επιστημονικό φάσμα εκτός των μαθηματικών όπως τη στατιστική φυσική, τη βιολογία, τη γενετική τα χρηματοοικονομικά και αλλού. Ορισμένα παραδείγματα όπου μπορεί να αξιοποιηθεί η μέθοδος της προσομοίωσης είναι:

1. Συστήματα ουρών αναμονής (π.χ. τράπεζες, νοσοκομεία, δίκτυα υπολογιστών).
2. Έλεγχος αποθεμάτων αποθηκών.
3. Αξιοπιστία πολύπλοκων συστημάτων.
4. Κίνηση μετοχών σε μια χρηματιστηριακή αγορά.
5. Εξάπλωση ασθενειών.
6. Μελέτη μετεωρολογικών φαινομένων.
7. Προσομοίωση πειραμάτων πυρηνικής φυσικής.
8. Μελέτη αντοχής κτηρίων σε σεισμούς.

Με την μέθοδο της προσομοίωσης είναι δυνατή και η μελέτη κάποιων πολύπλοκων στατιστικών προβλημάτων όπως π.χ.:

1. Προσεγγιστικός υπολογισμός της μέσης τιμής μιας τ.μ. που ακολουθεί κατανομή με πολύπλοκη συνάρτηση πυκνότητας πιθανότητας.
2. Προσεγγιστική εύρεση της κατανομής κάποιας πολύπλοκης στατιστικής συνάρτησης.
3. Ασυμπτωτική συμπεριφορά πολύπλοκων στοχαστικών μοντέλων.
4. Προσεγγιστικός υπολογισμός του p -τιμής ή της ισχύς ενός ελέγχου μιας υπόθεσης σε ένα σύνθετο μοντέλο.

Ιδιαίτερα χρήσιμη είναι η βοήθειά τους στον υπολογισμό ολοκληρωμάτων, κυρίως πολλαπλών και όταν η υπό ολοκλήρωση ποσότητα έχει σύνθετη μορφή. Ακόμα δουλεύουν αποτελεσματικά σε περίπλοκης δομής πολυδιάστατα προβλήματα που οι συνηθισμένες αριθμητικές μέθοδοι αδυνατούν.

ΚΕΦΑΛΑΙΟ 2

ΜΕΘΟΔΟΙ ΠΡΟΣΟΜΟΙΩΣΗΣ MONTE CARLO

2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε την παραγωγή τυχαίων αριθμών με την χρήση ηλεκτρονικών υπολογιστών. Έχει ιδιαίτερο ενδιαφέρον η ιστορία και ο τρόπος ο οποίος γίνεται αυτή η παραγωγή, αλλά ξεφεύγει από τους στόχους αυτής της διπλωματικής και για αυτό θα κάνουμε μια μικρή αναφορά μόνο χωρίς να δώσουμε ιδιαίτερη έμφαση σε αυτό το θέμα. Ακολούθως θα παρουσιάσουμε τις τεχνικές – αλγόριθμους με τις οποίες μπορούμε να προσομοιώσουμε τιμές από μια κατανομή της οποίας ξέρουμε την συνάρτηση κατανομής, και στην συνέχεια θα χρησιμοποιήσουμε αυτές τις μεθόδους για να φτιάξουμε αλγόριθμους για την προσομοίωση τιμών από γνωστές μας κατανομές, διακριτών και συνεχών.

2.2 Παραγωγή «Τυχαίων Αριθμών»

Η βάση για την μελέτη ενός στοχαστικού φαινομένου μέσω προσομοίωσης αποτελείται από μια διαδικασία η οποία παράγει «τυχαίους αριθμούς». Με τον όρο τυχαίοι αριθμοί εννοούμε το αποτέλεσμα μιας πεπερασμένης ακολουθίας X_1, X_2, \dots, X_n ανεξάρτητων τυχαίων μεταβλητών η κάθε μια εκ των οποίων κατανέμεται ομοιόμορφα στο $(0,1)$. Δηλαδή,

$$F_{X_i} = P(X_i \leq x) = x, \quad x \in (0,1).$$

Η παραγωγή μιας μεγάλης ακολουθίας τυχαίων αριθμών είναι αρκετά δύσκολη. Στα πρωταρχικά χρόνια της προσομοίωσης, οι τυχαίοι αριθμοί παράγονταν με τεχνικές όπως το ρίξιμο νομίσματος ή ζαριών, ανακατεύοντας τράπουλες, γυρίζοντας ρουλέτα και άλλα οι οποίες πιστευόταν ότι ήταν ο μόνος τρόπος να παράγει κανείς πραγματικούς τυχαίους αριθμούς. Αργότερα χρησιμοποιήθηκαν συσκευές όπως δίοδοι θορύβων και μετρητές Geiger. Αυτοί οι τρόποι όμως εγκαταλείφθηκαν για διάφορους λόγους. Για το λόγο αυτό καταφύγαμε στην παραγωγή *ψευδοτυχαίων αριθμών* με την χρήση απλών αλγορίθμων σε ηλεκτρονικό υπολογιστή.

Οι ψευδοτυχαίοι αριθμοί, όπως φαίνεται και από το όνομα τους, δεν είναι στην πραγματικότητα τυχαίοι αλλά παράγονται από κάποιες ντετερμινιστικές αναγωγικές γεννήτριες ξεκινώντας από κάποια αρχική τιμή (seed). Παρόλα αυτά μια ακολουθία τέτοιων αριθμών, αναλόγως από την διαδικασία που χρησιμοποιούμε, μπορεί να περάσει αρκετούς από τους ελέγχους τυχειότητας. Στη συνέχεια θα παρουσιάσουμε περιληπτικά κάποιες από αυτές τις επαναληπτικές διαδικασίες, χωρίς όμως να ασχοληθούμε με την αξιολόγηση της κάθε διαδικασίας.

Εδώ είναι χρήσιμο να προσέξουμε ότι αν $U \sim U(0,1)$ τότε $a + (b-a)U(0,1) \sim U(a,b)$ και άρα θα περιοριστούμε στην παραγωγή τιμών από την $U(0,1)$.

2.2.1 Πολλαπλασιαστική Μέθοδος – *Multiplicative Congruential Method*

Αλγόριθμος:

- I. **Θέτουμε** $x_0 = seed$.
- II. **Υπολογίζουμε το** $x_i = ax_{i-1} \bmod m$.
- III. **Θέτουμε** $U_i = \frac{x_i}{m}$.

Από τον ορισμό αυτού του αλγόριθμου μπορούμε να δούμε ότι:

$$x_i \in \{0, 1, \dots, m-1\} \text{ και}$$

$$U_i = \frac{x_i}{m} \in \left\{0, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}\right\} \subseteq [0, 1).$$

Όπως φαίνεται από τον αλγόριθμο μετά από ένα αριθμό επαναλήψεων $k \leq m$ θα εμφανιστεί ένας αριθμός $x_k = x_r$, ($r < k$) που έχει ήδη εμφανιστεί στην θέση r και άρα από εκεί και ύστερα θα επαναληφθεί η αρχική ακολουθία. Το $d = k - r$ ονομάζεται περίοδος της γεννήτριας. Επομένως θα πρέπει η επιλογή των a και m να γίνει έτσι ώστε για κάθε αρχική τιμή το πλήθος των βημάτων μέχρι την επανάληψη της διαδικασίας να είναι αρκετά μεγάλο. Το m πρέπει να είναι ένας αρκετά μεγάλος πρώτος αριθμός που να μπορεί να χειριστεί ο επεξεργαστής μας αλλά και το λειτουργικό μας. Μια καλή επιλογή για ένα 32-bit υπολογιστή είναι $m = 2^{32} - 1$ και $a = 7^5 - 1$. Ένα θεώρημα που μας βοηθά στην επιλογή των a, m είναι το ακόλουθο:

ΘΕΩΡΗΜΑ: Μια πολλαπλασιαστική γεννήτρια με $m = 2^b \geq 16$ έχει μέγιστη περίοδο $d = m/4$ αν και μόνο αν ισχύουν τα ακόλουθα:

- i. $a \bmod 8 = 3$ ή $a \bmod 8 = 5$.
- ii. x_0 περιττός.

Παράδειγμα

Ας υποθέσουμε για παράδειγμα ότι $a = 5^{2q+1}$, $q = 0, 1, 2, \dots$. Τότε:

$$\begin{aligned} a \bmod 8 &= (4+1)^{2q+1} \bmod 8 = \left(4^{2q+1} + \binom{2q+1}{1} 4^{2q} + \binom{2q+1}{2} 4^{2q-1} + \dots + \binom{2q+1}{2q-1} 4^2 + \binom{2q+1}{2q} 4 + 1 \right) \bmod 8 \\ &= \left(8 \cdot 4^{2q-2} + \binom{2q+1}{1} 8 \cdot 4^{2q-2} + \binom{2q+1}{2} 8 \cdot 4^{2q-3} + \dots + \binom{2q+1}{2q-1} 8 + 8q + 5 \right) \bmod 8 = 5 \bmod 8 = 5. \end{aligned}$$

Τότε μια πολλαπλασιαστική γεννήτρια της μορφής:

$$x_i = 5^{2q+1} x_{i-1} \bmod 2^b, \quad q=0, 1, 2, \dots \text{ και } 2^b \geq 16,$$

έχει περίοδο $d = \frac{2^b}{4} = 2^{b-2}$. Άρα αν $a = 5^{17}$ και $b = 43$ η πολλαπλασιαστική γεννήτρια θα έχει περίοδο $d = 2^{41} = 2.2 \cdot 10^{12}$. Αν υποθέσουμε ότι ο ηλεκτρονικός υπολογιστής γεννά 1000 τυχαίους αριθμούς το δευτερόλεπτο τότε θα έχουμε επανάληψη κάποιου αριθμού x_i μετά από 69 χρόνια.

2.2.2 Μεικτή μέθοδος – Mix congruential method

Αλγόριθμος:

- I. **Θέτουμε** $x_0 = seed$.
- II. **Υπολογίζουμε** το $x_i = (ax_{i-1} + c) \bmod m$.
- III. **Θέτουμε** $U_i = \frac{x_i}{m}$.

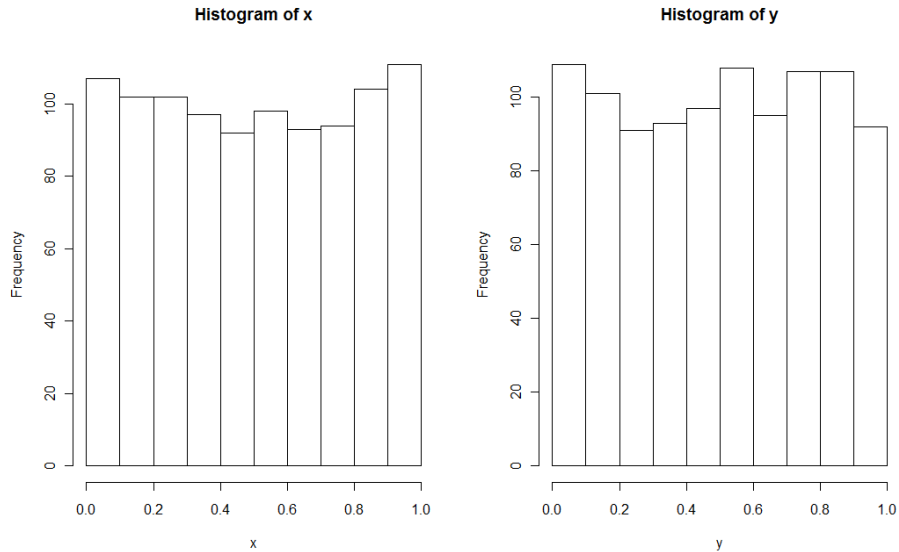
Επιλέγουμε τα a, m και c με το ίδιο σκεπτικό όπως και πριν. Ένα θεώρημα που μας βοηθά στην επιλογή τους είναι το ακόλουθο.

ΘΕΩΡΗΜΑ: Η μεικτή γεννήτρια έχει μέγιστη περίοδο m αν και μόνο αν ισχύουν τα ακόλουθα:

- i. Τα c και m είναι πρώτοι μεταξύ τους δηλαδή $\text{MKΔ}(c, m) = 1$.
- ii. Για κάθε πρώτο παράγοντα p του m πρέπει $a \equiv 1 \pmod p$.
- iii. Αν το 4 διαιρεί το m ακριβώς τότε πρέπει $a \equiv 1 \pmod 4$.

Ο ακόλουθος κώδικας είναι γραμμένος σε R και χρησιμοποιεί την παραπάνω γεννήτρια με $a = 7^5$, $m = 2^{31} - 2$ και $c = 0$. Παραθέτουμε επίσης τα ιστογράμματα των τιμών που δημιουργήσαμε με αυτή την μέθοδο καθώς και των τιμών που παράγει από μόνη της R με την εντολή *runif*, η οποία χρησιμοποιεί μια γεννήτρια που ονομάζεται *Mersenne-Twister* (Matsumoto και Nishimura, 1998), με περίοδο $2^{19937} - 1$ (Η ανάλυση της μεθόδου *Mersenne-Twister* είναι αρκετά πολύπλοκη και ξεφεύγει από τους στόχους της παρούσας διπλωματικής).

```
x<-rep(0,1000); seed<-runif(1)
x[1]<-seed; a<-7^5; m<-2^31-2;c<-0
for (i in 2:1000){
  x[i]<-((x[i-1]*a+c)%m )
}
x<-x/m; x[1]<-seed
y<-runif(1000)
par(mfrow=c(1,2))
hist(x);hist(y)
```



Σχήμα 1

Το διάνυσμα x περιέχει τις τιμές που δημιουργήσαμε εμείς ενώ το y τις τιμές που δημιούργησε η έτοιμη συνάρτηση της R. Από εδώ και στο εξής θα χρησιμοποιούμε την συνάρτηση *runif* χωρίς να αμφισβητούμε την τυχαιότητα της.

2.3 Μέθοδοι Παραγωγής Τυχαίων Αριθμών από Κατανομές

Αφού είδαμε πως γίνεται η παραγωγή τυχαίων αριθμών από την ομοιόμορφη κατανομή τώρα θα μελετήσουμε πως χρησιμοποιώντας ένα δείγμα από την ομοιόμορφη κατανομή μπορούμε να προσομοιώσουμε δείγματα άλλων κατανομών. Θα μελετήσουμε την μέθοδο της αντιστροφής (*inversion*), την μέθοδο της σύνθεσης (*composition*) και την μέθοδο της απόρριψης (*rejection*).

2.3.1 Μέθοδος της Αντιστροφής - *Inversion*

Έστω X μια τυχαία μεταβλητή με σ.κ.π. F . Αφού η F είναι μη φθίνουσα η αντίστροφη συνάρτηση F^{-1} μπορεί να οριστεί ως:

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}, 0 \leq y \leq 1.$$

Τότε αν $U \sim U(0,1)$ και $X = F^{-1}(U)$ η X έχει σ.κ.π. την F :

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

Έτσι αν θέλουμε να προσομοιώσουμε τυχαίες τιμές από την F αρκεί να παράξουμε τυχαίες τιμές στο $U \sim U(0,1)$ και να υπολογίσουμε το $X = F^{-1}(U)$.

Αλγόριθμος Αντιστροφής για Συνεχής Κατανομές

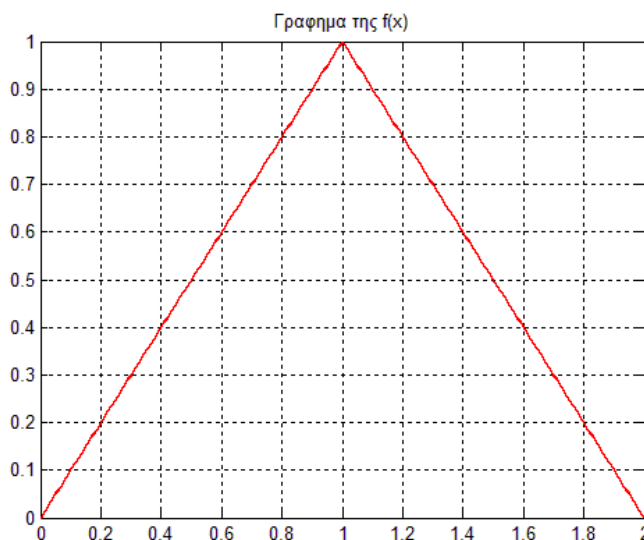
1. Παράγουμε ένα δείγμα με n τυχαίους αριθμούς από την $U \sim U(0,1)$.
2. Υπολογίζουμε το $X = F^{-1}(U)$.

Παράδειγμα

Έστω η τ.μ. X με σ.π.π

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2-x, & 1 \leq x \leq 2 \end{cases}$$

το γράφημα της οποίας φαίνεται παρακάτω στο Σχήμα 2.



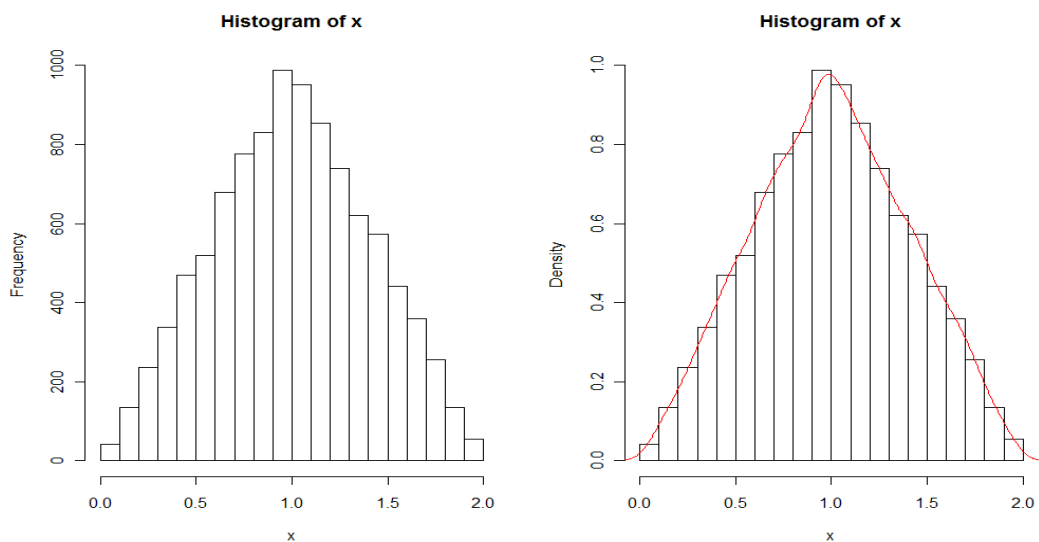
Σχήμα 2

Τότε η σ.κ.π. και η αντίστροφη της υπολογίζονται ως εξής:

$$F(x) = \begin{cases} 0, & x < 0 \\ \int_0^x x dx = \frac{x^2}{2} & 0 < x < 1 \\ \frac{1}{2} + \int_1^x (2-x) dx = 2x - \frac{x^2}{2} - 1, & 1 \leq x \leq 2 \\ 1, & x > 2 \end{cases} \quad \text{και} \quad F^{-1}(u) = \begin{cases} \sqrt{2u}, & 0 < u < \frac{1}{2} \\ 2 - \sqrt{2-2u}, & \frac{1}{2} \leq u \leq 1 \end{cases}$$

Χρησιμοποιήσαμε τον πιο κάτω κώδικα για να προσομοιώσουμε 10000 τιμές από την κατανομή μας και τα αποτελέσματα φαίνονται στο Σχήμα 3.

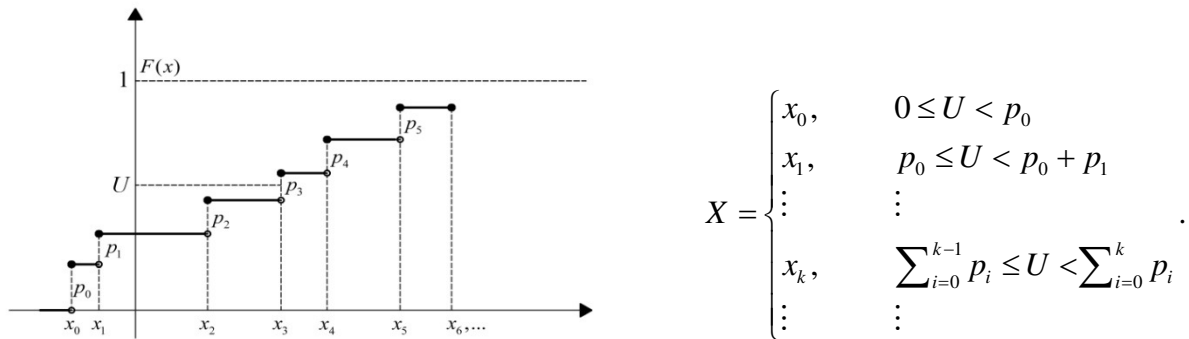
```
u<-runif(10000)
x<-rep(0,10000)
for (i in 1:10000) {
  if (u[i]<1/2) {
    x[i]<-sqrt(2*u[i])
  }
  else {
    x[i]<-2-sqrt(2-2*u[i])
  }
}
par(mfrow=c(1,2))
hist(x,xlim=c(0,2))
hist(x,freq=FALSE,xlim=c(0,2))
lines(density(x), col="red")hist(y)
```



Σχήμα 3

Συγκρίνοντας το ιστόγραμμα του δείγματος που προσομοιώσαμε και την εκτιμήτρια της σ.π.π, που λαμβάνουμε από την R με τη χρήση kernels (Σχήμα 3) με το γράφημα της κατανομής (Σχήμα 2) φαίνεται ότι η inverse δημιουργεί ένα αρκετά ικανοποιητικό δείγμα από την συγκεκριμένη κατανομή.

Η μέθοδος της αντιστροφής μπορεί να χρησιμοποιηθεί και στις **διακριτές τυχαίες μεταβλητές**. Έστω $U \sim U(0,1)$. Θεωρούμε την διακριτή τ.μ με συνάρτηση μάζας πιθανότητας $\{p_0, p_1, \dots, p_k, \dots\}$ καθώς και την τυχαία μεταβλητή X που ορίζεται ως εξής:



Τότε, $P(X = x_k) = P\left(\sum_{i=0}^{k-1} p_i \leq U \leq \sum_{i=0}^k p_i\right) = \sum_{i=0}^k p_i - \sum_{i=0}^{k-1} p_i = p_k$ και άρα η τυχαία μεταβλητή X έχει συνάρτηση κατανομής πιθανότητας $\{p_0, p_1, \dots, p_k, \dots\}$ όπως ακριβώς και η αρχική μας κατανομή.

Ο επόμενος αλγόριθμος προσομοιώνει τιμές από την διακριτή κατανομή με συνάρτηση πιθανότητας $\{p_0, p_1, \dots, p_k, \dots\}$.

Αλγόριθμος Αντιστροφής για Διακριτές Κατανομές

1. Παράγουμε ένα αριθμό από $U \sim U(0,1)$.
2. Αν $U < p_0$ τότε $X = x_0$ και σταματάμε, αλλιώς πάμε στο επόμενο βήμα.
3. Αν $U < p_0 + p_1$ τότε $X = x_1$ και σταματάμε, αλλιώς πάμε στο επόμενο βήμα.
4. Αν $U < p_0 + p_1 + p_2$ τότε $X = x_2$ και σταματάμε, αλλιώς πάμε στο επόμενο βήμα.

Κ.Ο.Κ....

Παράδειγμα

Έστω η διακριτή κατανομή με συνάρτηση πιθανότητας $\{p_1 = 0.1, p_2 = 0.2, p_3 = 0.3, p_4 = 0.4\}$. Υπολογίζουμε:

$$p_1 + p_2 = 0.3,$$

$$p_1 + p_2 + p_3 = 0.6,$$

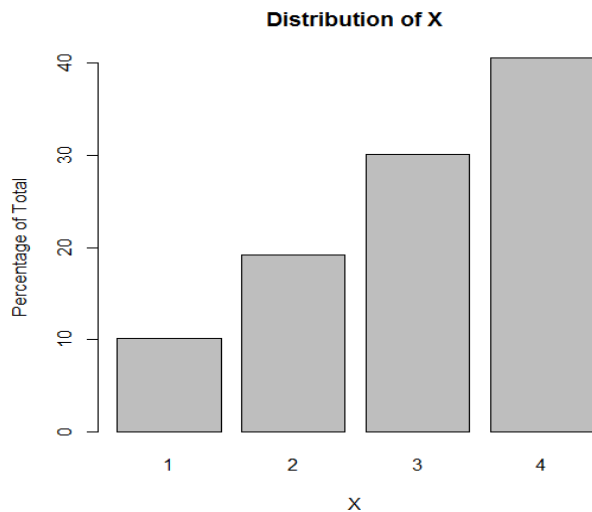
$$p_1 + p_2 + p_3 + p_4 = 1.$$

Χρησιμοποιήσαμε τον πιο κάτω κώδικα για να προσομοιώσουμε 1000 τιμές από την κατανομή μας και τα αποτελέσματα φαίνονται στο τομεόγραμμα του Σχήματος 4.

```

p<-c(0.1,0.3,0.6,1); xi<-c(1,2,3,4); x<-rep(0,1000)
for (i in 1:10000) {
  u<-runif(1)
  j<-1
  x[i]<-1
  while (u>p[j]){
    j<-j+1
    x[i]<-xi[j]
  }
}
xx<-table(x)
barplot(xx/100,main="Distribution of X",ylab="Percentage of Total",xlab="X")

```



Σχήμα 4

Όπως φαίνεται και στο τομεόγραμμα του Σχήματος 4, η μέθοδος της αντιστροφής έχει αρκετά καλά αποτελέσματα και στις διακριτές κατανομές.

Γενικά η μέθοδος της αντιστροφής προϋποθέτει ότι μπορεί να βρεθεί η αντίστροφη $F^{-1}(U)$ της σ.κ.π. F είτε αναλυτικά είτε αλγοριθμικά. Για πολλές κατανομές αυτό είναι πολύ δύσκολο. Για κάποιες άλλες κατανομές είναι επίσης δύσκολο η ακόμη και ακατόρθωτο να υπολογιστεί το ολοκλήρωμα $F(x) = \int_{-\infty}^x f(t)dt = u$. Ακόμα και στην περίπτωση που μπορέσουμε να βρούμε τον αντίστροφο μετασχηματισμό $F^{-1}(U)$, η μέθοδος της αντιστροφής σε μερικές περιπτώσεις δεν είναι αποδοτική.

2.3.2 Μέθοδος της Σύνθεσης – Composition Method

Η μέθοδος της σύνθεσης χρησιμοποιείται για την παραγωγή τυχαιών αριθμών από μίξεις κατανομών. Θα ξεκινήσουμε την μελέτη της μεθόδου για **διακριτές** τυχαίες μεταβλητές. Ας υποθέσουμε για παράδειγμα ότι επιθυμούμε να παράγουμε ένα δείγμα από μια κατανομή με συνάρτηση μάζας πιθανότητας της μορφής

$$p_i = P(X = i) = ap_i^{(1)} + (1-a)p_i^{(2)},$$

όπου $p_i^{(1)}$ και $p_i^{(2)}$ είναι συναρτήσεις μάζας πιθανότητας δύο τυχαιών μεταβλητών. Έστω τώρα $U \sim U(0,1)$. Τότε

$$P(X = i | U < a)P(U < a) + P(X = i | U \geq a)P(U \geq a) = ap_i^{(1)} + (1-a)p_i^{(2)} = P(X = i).$$

Στην πιο γενική μορφή είναι $p_i = \sum_j a_j p_i^{(j)}$, όπου $F_j = p_i^{(j)}$ συναρτήσεις μάζας πιθανότητας των τ.μ. X_j και $\sum_j a_j = 1$, $a_j \in (0,1)$. Ορίζουμε την διακριτή τ.μ. Y με τιμές $j = 1, 2, \dots$ και με σ.μ.π. $P(Y = j) = a_j$. Τότε η σ.μ.π. της X γράφεται ως εξής:

$$P(X = i) = \sum_j P(X = i | Y = j)P(Y = j) = \sum_j a_j p_i^{(j)}.$$

Έτσι προκύπτει ο ακόλουθος αλγόριθμος:

Αλγόριθμος Σύνθεσης για Διακριτές Κατανομές

1. Παράγουμε ένα τυχαίο αριθμό y από την κατανομή $P[Y = j] = a_j$.
2. Εάν $y = j$ τότε παράγουμε ένα τυχαίο αριθμό από την κατανομή F_j .

Η μέθοδος της σύνθεσης λειτουργεί και στις **συνεχείς** κατανομές όπου η σ.π.π είναι η

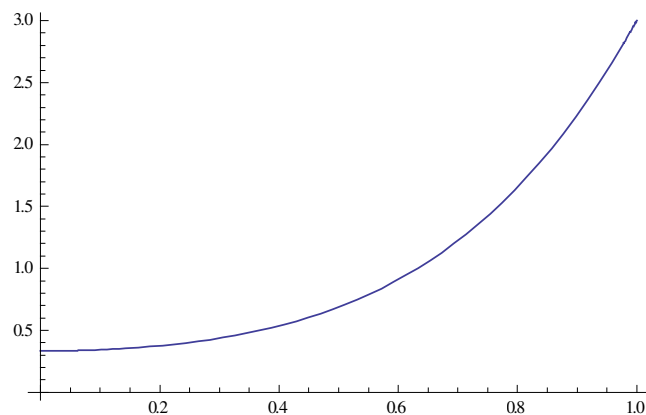
$$f(x) = \sum_{i=1}^n a_i f_i, \quad \sum_{i=1}^n a_i = 1 \text{ και } f_i \text{ σ.π.π. Ο αλγόριθμος διαμορφώνεται ως εξής}$$

Αλγόριθμος Σύνθεσης για Συνεχείς Κατανομές

1. Παράγω μια τυχαία μεταβλητή I που είναι ίση με i με πιθανότητα a_i .
2. Προσομοιώνω $X \sim F_i$.

Παράδειγμα

Έστω η τυχαία μεταβλητή X που ακολουθεί κατανομή με σ.π.π $f(x) = \frac{1+3x^2+5x^4}{3}$, $x \in (0,1)$.



Σχήμα 5

Η $f(x)$ μπορεί να γραφτεί και σαν $f(x) = \frac{1}{3} + \frac{1}{3}3x^2 + \frac{1}{3}5x^4 = \frac{1}{3}f_1(x) + \frac{1}{3}f_2(x) + \frac{1}{3}f_3(x)$, όπου κάθε f_i , $i=1,2,3$ είναι σ.π.π με πεδίο ορισμού ίδιο με αυτό της f . Έτσι χρησιμοποιώντας την μέθοδο της αντιστροφής και διαλέγοντας f_i με πιθανότητα $\frac{1}{3}$ έχουμε:

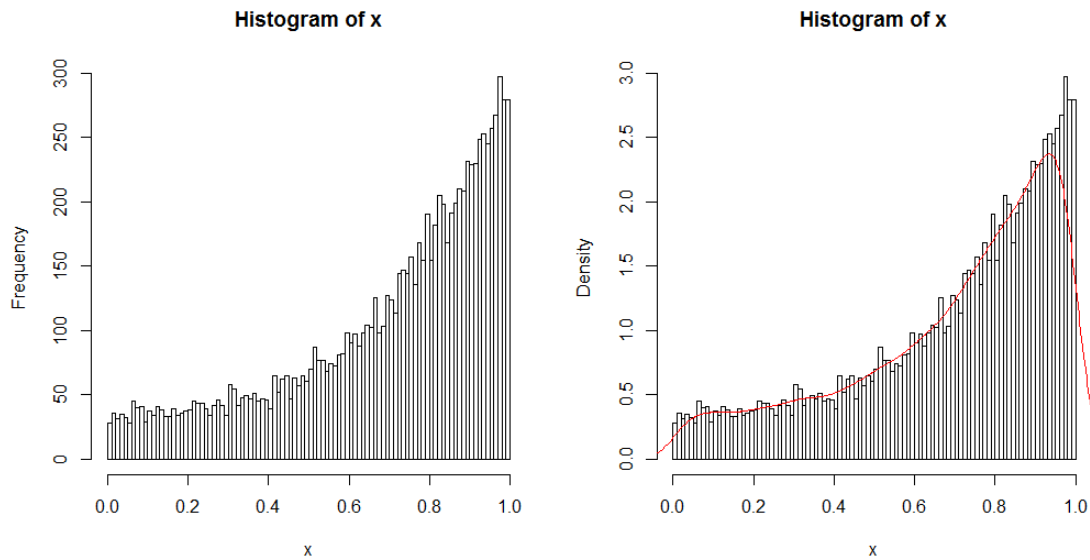
$$F_1(x) = x \Rightarrow F^{-1}_1(U) = u,$$

$$F_2(x) = x^3 \Rightarrow F^{-1}_2(U) = \sqrt[3]{U},$$

$$F_3(x) = x^5 \Rightarrow F^{-1}_3(U) = \sqrt[5]{U}.$$

Έτσι διαμορφώνεται ο πιο κάτω κώδικας στην R που προσομοιώνει 10000 τιμές από την κατανομή $f(x)$, τα αποτελέσματα του οποίου φαίνονται στο Σχήμα 6.

```
n<-10000
x<-rep(0,n)
for (i in 1:n) {
  a<-runif(1)
  u<-runif(1)
  if (a<=1/3){
    x[i]<-u
  } else {
    if (a<=2/3) {
      x[i]<-u^(1/3)
    } else {
      x[i]<-u^(1/5)
    }
  }
}
par(mfrow=c(1,2))
hist(x,xlim=c(0,1),nclass=100)
hist(x,freq=FALSE,xlim=c(0,1),nclass=100)
lines(density(x), col="red",xlim=c(0,1))
```

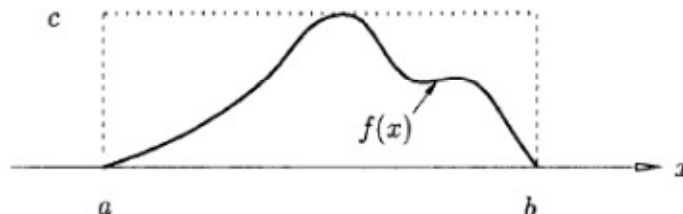


Σχήμα 6

2.3.3 Μέθοδος της Απόρριψης – Acceptance-Rejection Method

Τόσο η μέθοδος της αντιστροφής όσο και η μέθοδος της σύνθεσης είναι μέθοδοι οι οποίες χρησιμοποιούν άμεσα την σ.κ.π. της τυχαίας μεταβλητής που θέλουμε να προσομοιώσουμε. Η μέθοδος της απόρριψης είναι πιο εύκαμπτη από τις προηγούμενες μεθόδους διότι χρησιμοποιεί ένα ενδιάμεσο βήμα το οποίο μπορούμε να διαμορφώσουμε κατάλληλα ώστε να είναι ευκολότερη η παραγωγή τυχαίων αριθμών από κατανομές με δύσχρηστη συνάρτηση πιθανότητας.

Θα ξεκινήσουμε την μελέτη της μεθόδου για **συνεχείς** τυχαίες μεταβλητές. Ας υποθέσουμε ότι θέλουμε να προσομοιώσουμε τιμές από την κατανομή f όπως φαίνεται στο Σχήμα 7. Η συνάρτηση κατανομής f είναι φραγμένη σε ένα διάστημα $[a, b]$ και μηδέν έξω από το διάστημα αυτό.



Σχήμα 7

Για να προσομοιώσουμε τιμές από αυτήν την κατανομή ακολουθούμε τα ακόλουθα βήματα:

1. Παράγω $X \sim U(a,b)$,
2. Παράγω $Y \sim U(0,c)$,
3. Αν $Y \leq f(X)$ τότε δέχομαι το X . Διαφορετικά επιστρέφουμε στο βήμα 1.

Μπορούμε να δούμε ότι από τα βήματα 1 και 2 το διάνυσμα (X,Y) είναι ομοιόμορφα κατανεμημένο στο $[a,b] \times [0,c]$. Το βήμα 3 μας βεβαιώνει ότι τα επιτρεπτά (X,Y) είναι ομοιόμορφα κατανεμημένα κάτω από το γράφημα της f . Αυτό σημαίνει ότι τα αποδεκτά X ακολουθούν την κατανομή f .

Μπορούμε να γενικεύσουμε την μέθοδο που περιγράψαμε πιο πάνω ως ακολούθως. Έστω g μια αυθαίρετη συνάρτηση πυκνότητας πιθανότητας και ο **φάκελος** $\varphi(x) = Cg(x)$ καλύπτει την f . Καλούμε την g **κατανομή εισήγησης** (proposal pdf) και ισχύει ότι $\varphi(x) \geq f(x) \geq 0$ για κάθε x . Προφανώς για να έχει νόημα η μέθοδος πρέπει να είναι εύκολο να παράξουμε τιμές από την g . Τότε ο αλγόριθμος της απόρριψης έχει ως ακολούθως.

Αλγόριθμος Απόρριψης για Συνεχείς Κατανομές

1. Παράγουμε $Y \sim g(x)$.
2. Παράγουμε $U \sim U(0,1)$.
3. Αν $U \leq \frac{f(Y)}{Cg(Y)}$ τότε θέτω $X = Y$. Διαφορετικά επιστρέφουμε στο βήμα 1.

Ο τυχαίος αριθμός X που παράγεται από τον παραπάνω αλγόριθμο έχει συνάρτηση πυκνότητας πιθανότητας f . Αρχικά παρατηρούμε ότι σε κάθε επανάληψη δεχόμαστε ή απορρίπτουμε την τιμή Y ανεξάρτητα από τις προηγούμενες επαναλήψεις και με πιθανότητα:

$$P\left(U \leq \frac{f(Y)}{Cg(Y)}\right) = \int_R \frac{f(y)}{Cg(y)} g(y) dx = \int_R \frac{f(y)}{C} dx = \frac{1}{C}.$$

Και άρα ,

$$\begin{aligned}
 P(X \leq x) &= \sum_{n=1}^{\infty} P(\eta \text{ Y έγινε δεκτή στην } n\text{-οστή επανάληψη και } Y \leq x) \\
 &= \sum_{n=1}^{\infty} P(n-1 \text{ απορρίψεις, } \eta \text{ } n\text{-οστή επανάληψη δεκτή και } Y \leq x) \\
 &= \sum_{n=1}^{\infty} \left(1 - \frac{1}{C}\right)^{n-1} P\left(U \leq \frac{f(y)}{Cg(y)}, Y \leq x\right) \\
 &= \sum_{n=1}^{\infty} \left(1 - \frac{1}{C}\right)^{n-1} \frac{1}{C} \int_{-\infty}^x f(y) dy = \int_{-\infty}^x f(y) dy.
 \end{aligned}$$

Το πλήθος N των επαναλήψεων του αλγορίθμου (μέχρι την αποδοχή μιας τιμής) προφανώς ακολουθεί τη γεωμετρική κατανομή με πιθανότητα επιτυχίας $1/C$. Επομένως χρειάζονται κατά μέσο όρο $E(N) = C$ επαναλήψεις και άρα θα πρέπει να πάρουμε όσο το δυνατό μικρότερο C . Το μικρότερο C που μπορούμε να πάρουμε είναι: $C = \sup \left\{ \frac{f(x)}{g(x)}, x: g(x) \neq 0 \right\}$.

Χρησιμοποιώντας το παράδειγμα που χρησιμοποιήσαμε στην μέθοδο της αντιστροφής θα προσπαθήσουμε να προσομοιώσουμε τιμές με την μέθοδο της απόρριψης. Θα χρησιμοποιήσουμε σαν κατανομή εισήγησης την ομοιόμορφη κατανομή η οποία ορίζεται στο διάστημα $(0,2)$. Η συνάρτηση του παραδείγματος είναι:

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2-x, & 1 \leq x \leq 2 \end{cases} .$$

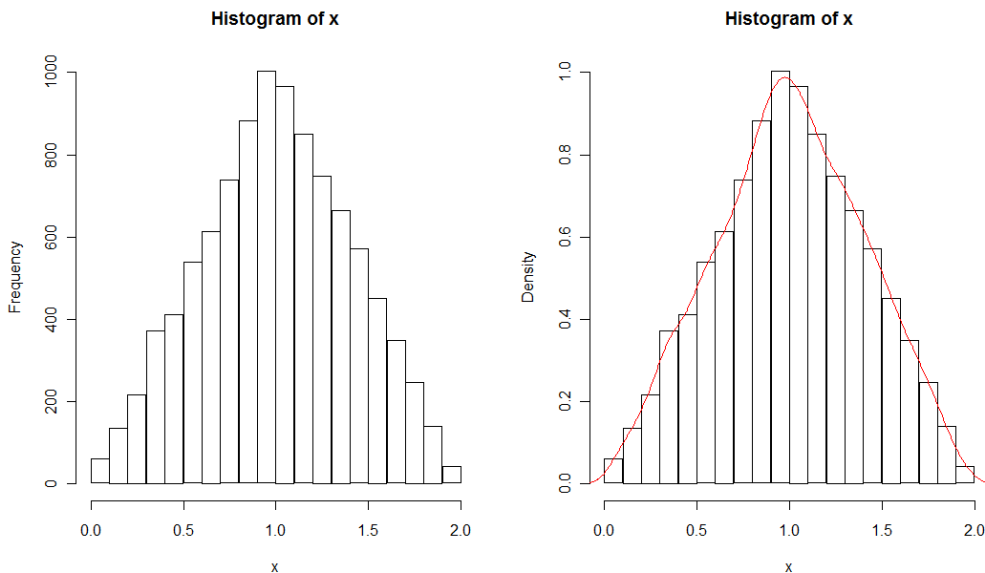
Έτσι έχουμε $g(x) = 1/2$ και $C = \sup \left\{ \frac{f(x)}{g(x)}, x: g(x) \neq 0 \right\} = \sup \{2f(x), x\} = 2$.

Χρησιμοποιήσαμε τον πιο κάτω κώδικα για να προσομοιώσουμε 10000 τιμές από την κατανομή μας και τα αποτελέσματα φαίνονται στο ιστόγραμμα του Σχήματος 8.


```

x<-rep(0,10000)
i<-1
while (i<10000){
  y=2*runif(1)
  u=runif(1)
  if (y<1) {
    f<-y }
  else {
    f<-2-y
  }
  if (u<f) {
    x[i]<-y
    i<-i+1
  }
}
par(mfrow=c(1,2))
hist(x,xlim=c(0,2))
hist(x,freq=FALSE,xlim=c(0,2))
lines(density(x), col="red")

```



Σχήμα 8

Συγκρίνοντας το ιστόγραμμα και την εκτιμήτρια της σ.π.π, που λαμβάνουμε από την R, του Σχήματος 6 με το γράφημα της κατανομής στο Σχήμα 2 και το δείγμα που προσομοιώσαμε με την μέθοδο της αντιστροφής στο Σχήμα 3 βλέπουμε ότι και οι δύο μέθοδοι δημιουργούν ένα πολύ καλό δείγμα από την συγκεκριμένη κατανομή.

Η μέθοδος της απόρριψης μπορεί να μας προσομοιώσει τιμές από κατανομές σύνθετης μορφής σ.κ.π. που δεν μπορούμε να υπολογίσουμε την αντίστροφη συνάρτηση. Η επιλογή της κατανομής εισήγησης για να έχει πρακτικό ενδιαφέρον η μέθοδος πρέπει να πληροί τα ακόλουθα κριτήρια:

1. Πρέπει να είναι εύκολο να παράξουμε μια τυχαία μεταβλητή από αυτή.
2. Η **αποδοτικότητα** της μεθόδου, το οποίο είναι ίσο με την πιθανότητα αποδοχής και υπολογίστηκε $\frac{1}{C}$, πρέπει να είναι μεγάλο, δηλαδή πρέπει το C να είναι όσο πιο κοντά στο 1 γίνεται. Αυτό συμβαίνει όταν η $g(x)$ είναι κοντά στην $f(x)$.

Όπως και στην μέθοδο της αντιστροφής η μέθοδος της απόρριψης μπορεί να χρησιμοποιηθεί και σε **διακριτές** κατανομές. Έστω λοιπόν ότι θέλουμε να προσομοιώσουμε τιμές από μια κατανομή με σ.μ.π. $p_j = P(X = j), j = 0, 1, \dots$. Έστω επίσης ότι μπορούμε να παράξουμε εύκολα τυχαίους αριθμούς από μια κατανομή $q_j = P(X = j), j = 0, 1, \dots$ με την μόνη προϋπόθεση αν $q_j = 0$ τότε $p_j = 0$. Σύμφωνα με αυτή την μέθοδο παράγουμε ένα τυχαίο αριθμό Y από την $q_j = P(Y = j), j = 0, 1, \dots$ και στην συνέχεια τον αποδεχόμαστε με πιθανότητα ανάλογη του πηλίκου p_Y / q_Y . Πιο συγκεκριμένα, αν υπάρχει μια σταθερά $C < \infty$ για την οποία ισχύει $\frac{p_j}{q_j} \leq C$ για κάθε $j = 0, 1, \dots$, ο αλγόριθμος έχει ως εξής.

Αλγόριθμος Απόρριψης για Διακριτές Κατανομές

1. Παράγουμε $Y \sim q_j$.
2. Παράγουμε $U \sim U(0, 1)$.
3. Αν $U \leq \frac{p_Y}{Cq_Y}$ τότε θέτω $X = Y$. Διαφορετικά επιστρέφουμε στο βήμα 1.

Ο τυχαίος αριθμός X που παράγεται από τον πιο πάνω αλγόριθμο έχει συνάρτηση μάζας πιθανότητας $p_j = P(X = j), j = 0, 1, \dots$. Αρχικά, όπως και στην περίπτωση της

συνεχής κατανομής παρατηρούμε ότι σε κάθε επανάληψη δεχόμαστε ή απορρίπτουμε την τιμή Y ανεξάρτητα από τις προηγούμενες επαναλήψεις και μάλιστα με πιθανότητα:

$$P\left(U \leq \frac{p_Y}{Cq_Y}\right) = \sum_i P\left(U \leq \frac{p_Y}{Cq_Y} | Y = i\right) P(Y = i) = \sum_i P\left(U \leq \frac{p_i}{Cq_i}\right) q_i = \sum_i \frac{p_i}{Cq_i} q_i = \frac{1}{C},$$

και άρα:

$$\begin{aligned} P(X = i) &= \sum_{n=1}^{\infty} P(\eta \text{ } Y \text{ έγινε δεκτή στην } n\text{-οστή επανάληψη και } Y = i) \\ &= \sum_{n=1}^{\infty} P(n-1 \text{ απορρίψεις, } \eta \text{ } n\text{-οστή επανάληψη δεκτή και } Y = i) \\ &= \sum_{n=1}^{\infty} \left(1 - \frac{1}{C}\right)^{n-1} P(\eta \text{ } n\text{-οστή επανάληψη δεκτή με } Y = i) \\ &= \sum_{n=1}^{\infty} \left(1 - \frac{1}{C}\right)^{n-1} P\left(U \leq \frac{p_Y}{Cq_Y} | Y = i\right) P(Y = i) \\ &= \sum_{n=1}^{\infty} \left(1 - \frac{1}{C}\right)^{n-1} P\left(U \leq \frac{p_i}{Cq_i}\right) q_i = C \frac{p_i}{Cq_i} q_i = p_i. \end{aligned}$$

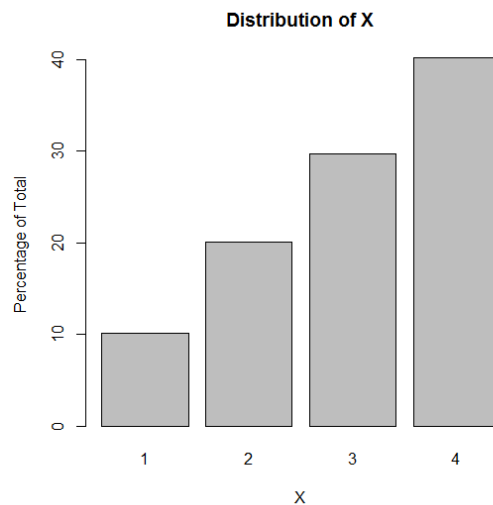
Όπως και στην συνεχή περίπτωση το πλήθος των N επαναλήψεων του αλγορίθμου μέχρι την αποδοχή μιας τιμής ακολουθεί γεωμετρική κατανομή με πιθανότητα επιτυχίας $\frac{1}{C}$ και άρα χρειάζονται κατά μέσο όρο $E(N) = C$ επαναλήψεις. Συνεπώς για να αποφύγουμε πολλές επαναλήψεις πρέπει να πάρουμε όσο το δυνατό μικρότερο C . Το μικρότερο C που μπορούμε να πάρουμε δίνεται από την σχέση $C = \sup \left\{ \frac{p_j}{q_j}, j = 0, 1, \dots \right\}$.

Συνεχίζοντας με το παράδειγμα που χρησιμοποιήσαμε πριν στην μέθοδο της αντιστροφής με την κατανομή $\{p_1 = 0.1, p_2 = 0.2, p_3 = 0.3, p_4 = 0.4\}$ και χρησιμοποιώντας σαν κατανομή εισήγησης την $\{q_j = 0.25, j = 1, \dots, 4\}$ υπολογίζουμε με βάση τον αλγόριθμο που αναπτύξαμε:

$$C = \max \left\{ \frac{p_i}{q_i}, i = 1, \dots, 4 \right\} = \max \{4p_i, i = 1, \dots, 4\} = 4 \cdot 0.4 = 1.6 \quad \text{και} \quad Cq_Y = 1.6 \cdot 0.25 = 0.4.$$

Έτσι χρησιμοποιούμε τον παρακάτω κώδικα στην R για προσομοίωση 1000 τιμών από την κατανομή μας.

```
p<-c(0.1, 0.2, 0.3, 0.4)
C<-1.6
x<-rep(0,10000)
i<-1
while (i<10001) {
  u<-runif(1)
  y<-floor(4*u+1)
  u2<-runif(1)
  if (u2<p[y]/0.4) {
    x[i]<-y
    i<-i+1
  }
}
xx<-table(x)
barplot(xx/100,main="Distribution of X",ylab="Percentage of Total",xlab="X")
```



Σχήμα 9

Στο Σχήμα 9 φαίνονται τα αποτελέσματα του αλγόριθμου της απόρριψης που είναι όπως αναμένονταν από την αρχική μας κατανομή $\{p_j\}$. Ο πιο πάνω αλγόριθμος απαιτεί κατά μέσο όρο $C=1.6$ επαναλήψεις για κάθε τυχαίο αριθμό ενώ ο αντίστοιχος αλγόριθμος αντιστροφής για την ίδια κατανομή χρειάζεται με βέλτιστη διάταξη των

τιμών $1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.1 = 2$ επαναλήψεις. Αυτό μας επιβεβαιώνει ότι ο αλγόριθμος της απόρριψης είναι πιο γρήγορος από αυτόν της αντιστροφής.

2.4 Προσομοίωση από Γνωστές Κατανομές

Σε αυτή την παράγραφο θα παρουσιάσουμε αλγόριθμους προσομοίωσης των πιο γνωστών κατανομών χρησιμοποιώντας τις μεθόδους που αναλύσαμε στην προηγούμενη παράγραφο. Θα δούμε ξεχωριστά μεθόδους για συνεχείς και διακριτές κατανομές και θα προσπαθήσουμε να χρησιμοποιήσουμε τις κατάλληλες μεθόδους που είναι αποτελεσματικές και εύκολες στην εφαρμογή τους ανάλογα με την κάθε κατανομή.

2.4.1 Προσομοίωση Συνεχών Κατανομών

2.4.1.1 Εκθετική Κατανομή

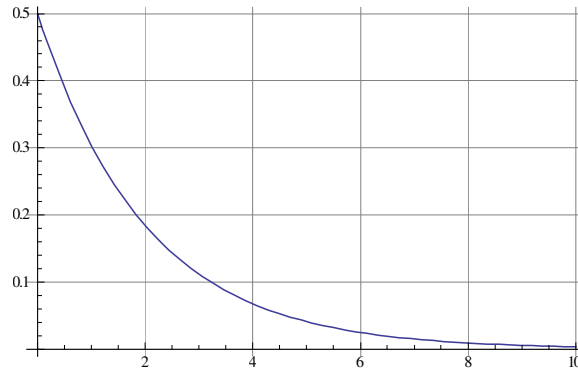
Έστω ότι η τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή με παράμετρο λ $X \sim \text{Exp}(\lambda)$. Η σ.π.π. είναι $f(x) = \lambda e^{-\lambda x}$ και η σ.κ.π. είναι $F(x) = 1 - e^{-\lambda x}$. Σε αυτή την περίπτωση είναι πολύ εύκολη η εφαρμογή της **μεθόδου της αντιστροφής**. Έτσι λύνοντας $u = F(x)$ παίρνουμε

$$F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u).$$

Έτσι έχουμε τον πιο κάτω αλγόριθμο:

1. Παράγουμε $U \sim U(0,1)$,
2. Υπολογίζουμε το $X = -\frac{1}{\lambda} \ln(1-U)$.

Θα αναπτύξουμε τον κώδικα στην R χρησιμοποιώντας σαν παράμετρο $\lambda = \frac{1}{2}$ και προσομοιώνοντας 10000 τιμές από την κατανομή.

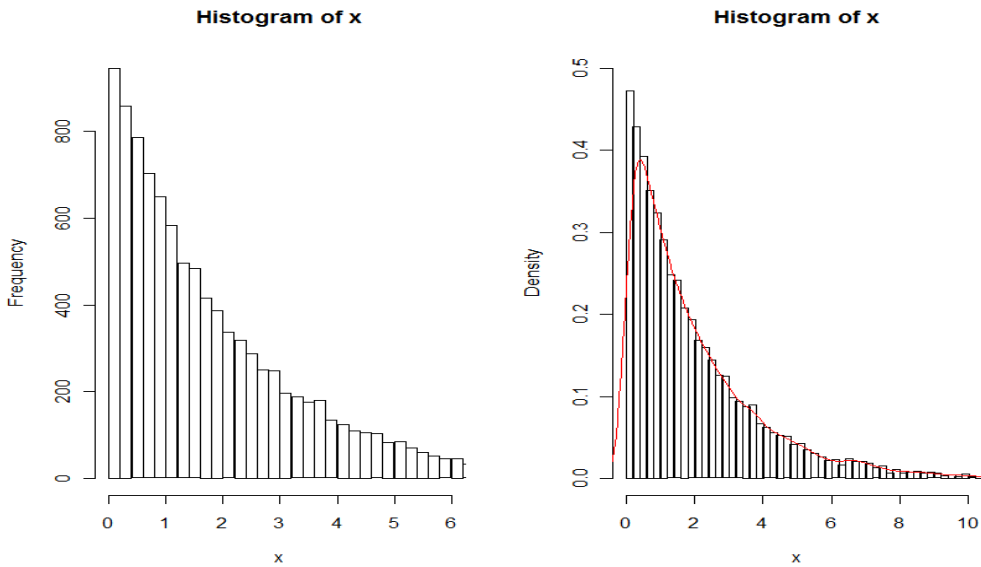


Σχήμα 10

(Θεωρητική Παράσταση της Εκθετικής με Κατανομής με Παράμετρο ½)

```

u<-runif(10000); x<-rep(0,10000)
for (i in 1:10000) {
  x[i]<--2*log(u[i])
}
par(mfrow=c(1,2))
hist(x,xlim=c(0,6),nclass=100)
hist(x,freq=FALSE,xlim=c(0,10),ylim=c(0,0.5),nclass=100)
lines(density(x), col="red")
mean(x); var(x)
    
```



Σχήμα 11

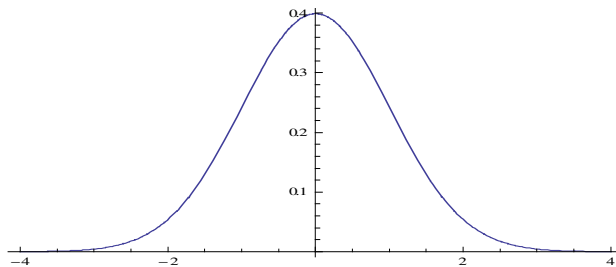
Ο πιο πάνω αλγόριθμος μας επιστρέφει το γράφημα που φαίνεται στο Σχήμα 11 που όπως βλέπουμε είναι πάρα πολύ κοντά στο γράφημα της εκθετικής κατανομής με παράμετρο $\frac{1}{2}$. Επίσης μας επιστρέφει δειγματικό μέσο 2.016813 και διασπορά 4.067048 τιμές που είναι πολύ κοντά στις θεωρητικές τιμές της εκθετικής κατανομής με παράμετρο $\lambda = 0.5$, $\mu = 2$ και $\sigma^2 = 4$.

2.4.1.2 Κανονική Κατανομή

Η παραγωγή τυχαίων αριθμών από την κατανομή $N(\mu, \sigma^2)$ είναι πολύ χρήσιμη αφού τέτοιοι αριθμοί εμφανίζονται σε πολλές εφαρμογές. Θα παρουσιάσουμε τρεις μεθόδους για την προσομοίωση τιμών από την κατανομή αυτή, αρχικά με την μέθοδο της απόρριψης και ακολούθως με τις πολικές μεθόδους *Box Muller* και *Polar Maraglia*.

Αρχικά παρατηρούμε ότι αρκεί να επιτύχουμε την παραγωγή τυχαίων αριθμών από την $N(0,1)$. Πράγματι αν $Z \sim N(0,1)$ τότε $\mu + \sigma Z \sim N(\mu, \sigma^2)$. Έτσι στις παραγράφους που ακολουθούν για χάριν απλότητας θα παράξουμε τιμές από την $N(0,1)$ η οποία έχει σ.π.π. την

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in R.$$



Σχήμα 12

(Θεωρητική Παράσταση της $N(0,1)$)

Θα υπολογίσουμε την σ.π.π. f της $|Z|$:

$$f(x) = \frac{d}{dx} P(|Z| \leq x) = \frac{d}{dx} P(-x \leq Z \leq x) = \frac{d}{dx} (\Phi(x) - \Phi(-x)) = \frac{d}{dx} (2\Phi(x) - 1) = 2\varphi(x).$$

Άρα καταλήγουμε ότι:

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \geq 0.$$

Για να παράξουμε τυχαίους αριθμούς από την τελευταία μπορούμε να χρησιμοποιήσουμε την μέθοδο της απόρριψης με κατανομή εισήγησης την εκθετική κατανομή με παράμετρο a και πρέπει να υπολογίσουμε το c έτσι ώστε να ικανοποιείται η παρακάτω σχέση:

$$h(x) = \frac{\frac{2}{\sqrt{2\pi}} e^{-x^2/2}}{ae^{-ax}} = \frac{2e^{ax-x^2/2}}{a\sqrt{2\pi}} \leq c, \quad x > 0.$$

Έτσι μπορούμε να υπολογίσουμε:

$$\frac{d}{dx} h(x) = \frac{2}{a\sqrt{2\pi}} e^{ax-x^2/2} (a-x) = 0,$$

$$\Rightarrow x = a.$$

$$\frac{d^2}{dx^2} h(a) = -\frac{2}{a\sqrt{2\pi}} e^{a^2/2} < 0,$$

$$\Rightarrow c = \max h(x) = h(a) = \frac{2e^{a^2/2}}{a\sqrt{2\pi}}.$$

Για να έχουμε ελάχιστο αριθμό επαναλήψεων θα πάρουμε $a = 1$, που ελαχιστοποιεί την παράσταση του c που υπολογίσαμε. Έστω λοιπόν ότι $W \equiv |Z|$ και I μια δίτιμη τ.μ. που παίρνει τιμές $\{1, -1\}$, είναι ανεξάρτητη από την W και $P(I = 1) = 1 - P(I = -1) = \frac{1}{2}$. Τότε η τ.μ. $X = IW \sim N(0,1)$:

$$\begin{aligned} P(X \leq x) &= P(IW \leq x) = P(IW \leq x | I = 1)P(I = 1) + P(IW \leq x | I = -1)P(I = -1) = \\ &= \frac{1}{2}P(W \leq x) - \frac{1}{2}P(W > -x) = \frac{1}{2}(P(W \leq x) + 1 - P(W < -x)) = \begin{cases} \frac{1}{2}(1 + P(W \leq x)), & x \geq 0 \\ \frac{1}{2}(1 - P(W \leq -x)), & x < 0 \end{cases}, \end{aligned}$$

και άρα,

$$f_x(x) = \frac{d}{dx} P(X \leq x) = \begin{cases} \frac{1}{2} f_w(x), & x \geq 0 \\ \frac{1}{2} f_w(x), & x < 0 \end{cases} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in R.$$

Έτσι ο αντίστοιχος αλγόριθμος παραγωγής τυχαίων αριθμών X από την $N(0,1)$ είναι ο εξής:

1. Παράγουμε $U_1 = U(0,1)$ και με την μέθοδο της αντιστροφής θέτουμε $Y = -\ln U_1$ ($Y \sim E(1)$).
2. Παράγουμε $U_2 = U(0,1)$.
3. Εάν $U_2 \leq \frac{f(Y)}{cg(Y)} = e^{-(Y-1)^2/2}$ τότε θέτουμε $W = Y$ και συνεχίζουμε, διαφορετικά επιστρέφουμε στην αρχή.
4. Παράγουμε $U_3 = U(0,1)$.
5. Αν $U_3 \geq 0.5$ τότε θέτουμε $Z = W$, διαφορετικά θέτουμε $Z = -W$.

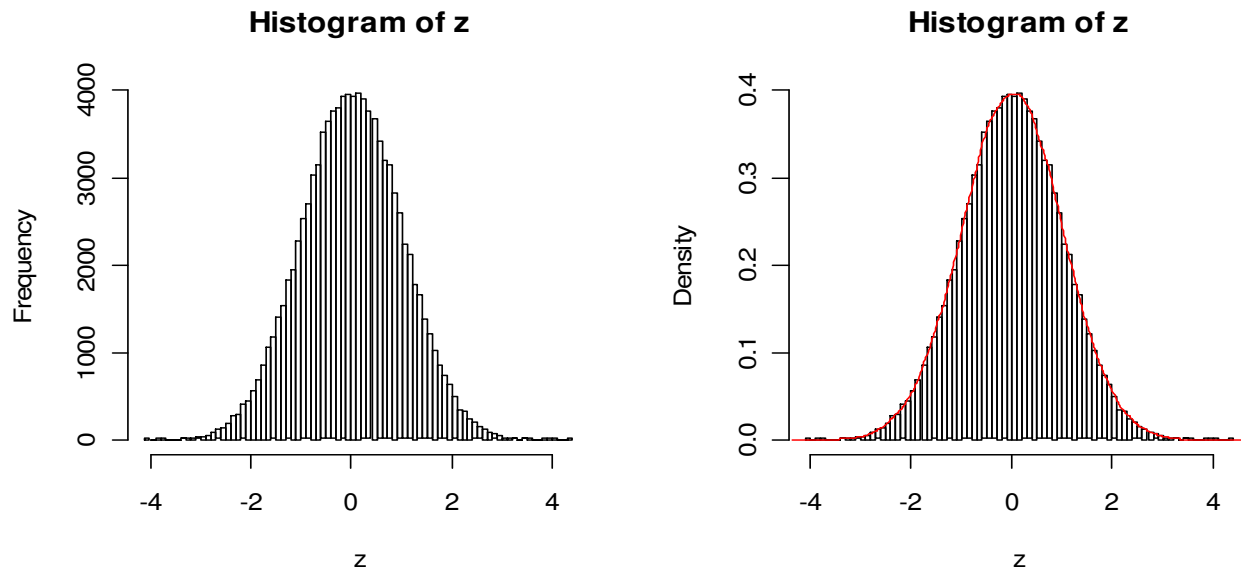
Ο αντίστοιχος κώδικας στην R που παράγει 100000 τιμές από την κανονική κατανομή $N(0,1)$ είναι ο εξής:

```

z<-rep(0,100000)
i<-1
while (i<100000){
  y=-log(runif(1))
  u2=runif(1)
  a=exp(-(y-1)^2/2)
  if (u2<=a) {
    w<-y
    u3=runif(1)
    if(u3>=0.5) {
      z[i]=w }
    else {
      z[i]=-w }
    i<-i+1 }
}
par(mfrow=c(1,2))
hist(z,nclass=100)
hist(z,freq=FALSE,nclass=100)
lines(density(z), col="red")
mean(z); var(z)

```

Ο πιο πάνω κώδικας μας επιστρέφει τα γραφήματα που φαίνονται στο Σχήμα 13 καθώς και δειγματικό μέσο 0.00071 και διασπορά 0.9974 που είναι πολύ κοντά στις θεωρητικές τιμές της κανονικής κατανομής $N(0,1)$, $\mu=0$ και $\sigma^2=1$.



Σχήμα 13

Μια άλλη μέθοδος παραγωγής τιμών από την κανονική κατανομή είναι η *Box-Muller* που είναι γνωστή και ως πολική μέθοδος. Η μέθοδος αυτή βασίζεται στον μετασχηματισμό σε πολικές συντεταγμένες μιας διδιάστατης κανονικής κατανομής.

Έστω X, Y ανεξάρτητες τ.μ. από την $N(0,1)$ με από κοινού σ.π.π. την:

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{\sqrt{2\pi}} e^{-(x^2+y^2)/2}.$$

Αν (X, Y) είναι καρτεσιανές συντεταγμένες σημείου στο επίπεδο τότε οι πολικές συντεταγμένες $(r, \theta) = (R^2, \Theta)$ είναι $R^2 = X^2 + Y^2$, $\Theta = \arctan(X / Y)$ ή αντίστροφα $X = R \cos \theta$, $Y = R \sin \theta$. Τότε ο μετασχηματισμός που ορίζεται είναι ο $(r, \theta) = (x^2 + y^2, \arctan(x / y))$ με αντίστροφό τον $(x, y) = (\sqrt{r} \cos \theta, \sqrt{r} \sin \theta)$.

Τότε η από κοινού σ.π.π. του ζεύγους (R^2, Θ) θα είναι:

$$\begin{aligned}
 f_{(R^2, \Theta)} &= f_{(X, Y)}(\sqrt{r} \cos \theta, \sqrt{r} \sin \theta) \left| \begin{array}{cc} \frac{\partial}{\partial r} \sqrt{r} \cos \theta & \frac{\partial}{\partial \theta} \sqrt{r} \cos \theta \\ \frac{\partial}{\partial r} \sqrt{r} \sin \theta & \frac{\partial}{\partial \theta} \sqrt{r} \sin \theta \end{array} \right| = \\
 &= \frac{1}{2\pi} e^{-\frac{1}{2}(r \cos^2 \theta + r \sin^2 \theta)} \left| \begin{array}{cc} \frac{\cos \theta}{2\sqrt{r}} & -\sqrt{r} \sin \theta \\ \frac{\sin \theta}{2\sqrt{r}} & \sqrt{r} \cos \theta \end{array} \right| = \frac{1}{2\pi} e^{-r/2} \frac{1}{2} (\cos^2 \theta + \sin^2 \theta) = \\
 &= \frac{1}{2} e^{-r/2} \frac{1}{2\pi} = f_{R^2}(r) f_{\Theta}(\theta), \quad r > 0, \theta \in [0, 2\pi).
 \end{aligned}$$

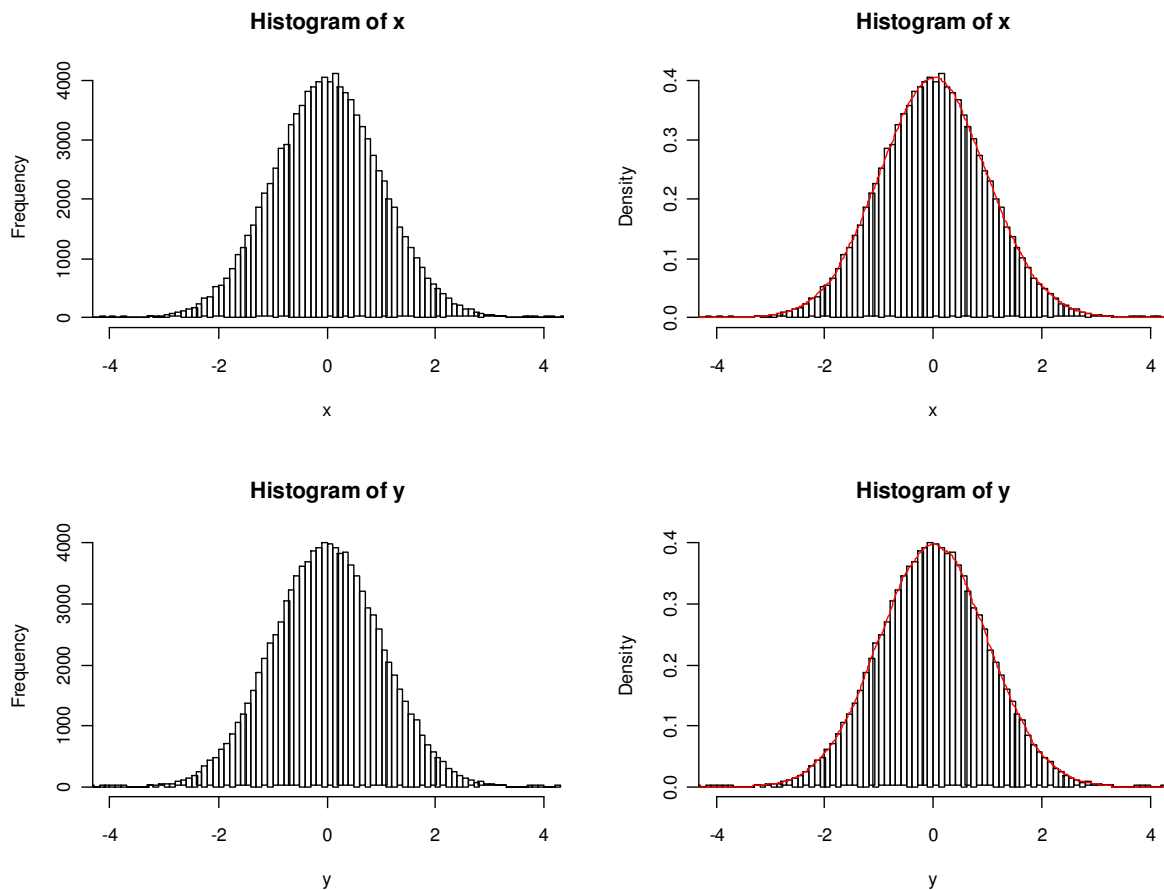
Άρα $R^2 \sim E\left(\frac{1}{2}\right)$ και $\Theta \sim U(0, 2\pi)$. Συνεπώς αν προσομοιώσουμε σημεία με πολικές συντεταγμένες που να ακολουθούν τις κατανομές που προσδιορίσαμε (χρησιμοποιώντας την μέθοδο της αντιστροφής για να παράγουμε τιμές από την εκθετική κατανομή $E\left(\frac{1}{2}\right)$), τότε τα σημεία αυτά θα έχουν καρτεσιανές συντεταγμένες $(X, Y) = (R \cos \theta, R \sin \theta)$ και θα κατανέμονται στο επίπεδο σύμφωνα με τη διδιάστατη κανονική κατανομή. Έτσι προκύπτει ο παρακάτω αλγόριθμος ο οποίος προσομοιώνει τιμές από δύο τ.μ. που ακολουθούν κανονική κατανομή:

1. Παράγουμε $U_1 = U(0,1)$ και θέτουμε $R^2 = -2 \ln U_1$.
2. Παράγουμε $U_2 = U(0,1)$ και θέτουμε $\Theta = 2\pi U_2$.
3. Θέτουμε $X = R \cos \Theta$ και $Y = R \sin \Theta$.

Ο αντίστοιχος κώδικας στην R που παράγει 100000 τιμές από δύο κανονικές κατανομές $N(0,1)$ είναι ο εξής:

```

u1<-runif(100000); u2<-runif(100000); r<-rep(0,100000); t<-rep(0,100000)
x<-rep(0,100000); y<-rep(0,100000)
for (i in 1:100000) {
  r[i]<-2*log(u1[i]); t[i]<-2*pi*u2[i]; x[i]<-cos(t[i])*sqrt(r[i]); y[i]<-sin(t[i])*sqrt(r[i])
}
par(mfrow=c(2,2))
hist(x,xlim=c(-4,4),nclass=100); hist(x,freq=FALSE,xlim=c(-4,4),nclass=100); lines(density(x), col="red")
mean(x);var(x)
hist(y,xlim=c(-4,4),nclass=100); hist(y,freq=FALSE,xlim=c(-4,4),nclass=100); lines(density(y), col="red")
mean(y);var(y)
    
```



Σχήμα 14

Ο πιο πάνω κώδικας μας επιστρέφει τα γραφήματα που φαίνονται στο Σχήμα 14 καθώς και δειγματικούς μέσους για x, y 0.00053 και 0.002 αντίστοιχα όπως και διασπορές 0.991 και 1.005.

Η μέθοδος Box-Muller απαιτεί λιγότερα βήματα συγκρινόμενη με την μέθοδο που αναπτύξαμε πιο πάνω χρησιμοποιώντας την μέθοδο της αντιστροφής. Ένα σοβαρό μειονέκτημα του όμως είναι ότι προϋποθέτει τον υπολογισμό ενός ημίτονου και ενός συνημίτονου που θεωρούνται χρονοβόροι. Μια τροποποίηση της πιο πάνω μεθόδου είναι η μέθοδος *Polar Marsgalia*, που αποφεύγει τους συγκεκριμένους υπολογισμούς.

Η μέθοδος προτείνει να παράγουμε ζεύγη ανεξάρτητων τ.μ. (V_1, V_2) τέτοια ώστε $V_1, V_2 \sim U(-1,1)$ και $V_1^2 + V_2^2 \leq 1$. Δηλαδή παράγουμε σημεία που κατανέμονται ομοιόμορφα μέσα στον μοναδιαίο κύκλο. Αφού κατανέμονται ομοιόμορφα η από κοινού σ.π.π. θα είναι της μορφής:

$$f_{V_1, V_2}(V_1, V_2) = \begin{cases} C, & V_1^2 + V_2^2 \leq 1 \\ 0, & V_1^2 + V_2^2 > 1 \end{cases}, \text{ για κάποια σταθερά } C \in \mathbb{R}.$$

Αφού η f_{V_1, V_2} είναι σ.π.π. πρέπει:

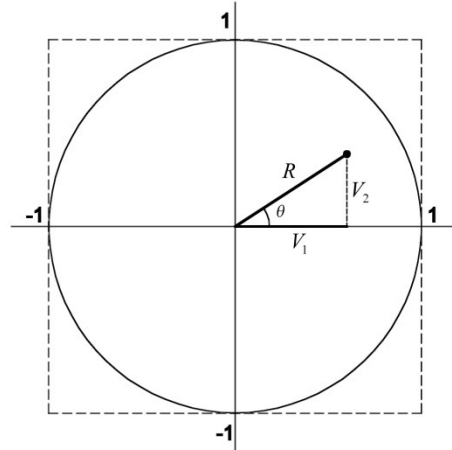
$$\begin{aligned} \int_{-1}^1 \int_{-\sqrt{1-V_2^2}}^{\sqrt{1-V_2^2}} C dV_1 dV_2 &= 1, \\ \Rightarrow 2C \int_{-1}^1 \sqrt{1-V_2^2} dV_2 &= 2C \int_{-\pi/2}^{\pi/2} \sqrt{1-(\sin u)^2} \cos u du = 1, \\ \Rightarrow 2C \int_{-\pi/2}^{\pi/2} (\cos u)^2 du &= C \int_{-\pi/2}^{\pi/2} (\cos 2u + 1) du = C \left[\frac{\sin 2u}{2} + u \right]_{-\pi/2}^{\pi/2} = C\pi = 1, \\ \Rightarrow C &= \frac{1}{\pi}. \end{aligned}$$

Αν τώρα (R^2, Θ) οι πολικές συντεταγμένες του σημείου (V_1, V_2) η από κοινού σ.π.π. των (R^2, Θ) είναι:

$$\begin{aligned} f_{(R^2, \Theta)} &= f_{(V_1, V_2)}(\sqrt{r} \cos \theta, \sqrt{r} \sin \theta) \begin{vmatrix} \frac{\partial}{\partial r} \sqrt{r} \cos \theta & \frac{\partial}{\partial \theta} \sqrt{r} \cos \theta \\ \frac{\partial}{\partial r} \sqrt{r} \sin \theta & \frac{\partial}{\partial \theta} \sqrt{r} \sin \theta \end{vmatrix} = \\ &= \frac{1}{\pi} \begin{vmatrix} \frac{\cos \theta}{2\sqrt{r}} & -\sqrt{r} \sin \theta \\ \frac{\sin \theta}{2\sqrt{r}} & \sqrt{r} \cos \theta \end{vmatrix} = \frac{1}{\pi} \cdot \frac{1}{2} (\cos^2 \theta + \sin^2 \theta) = \\ &= \frac{1}{2\pi} = f_{R^2}(r) f_{\Theta}(\theta), \quad r \in (0, 1), \theta \in (0, 2\pi). \end{aligned}$$

Άρα $R^2 \sim U(0, 1)$ και $\Theta \sim U(0, 2\pi)$. Στην μέθοδο Box Muller είδαμε ότι χρειάζεται να υπολογίσουμε το ημίτονο και το συνημίτονο μιας ομοιόμορφα κατανεμημένης γωνίας στο $(0, 2\pi)$. Η Θ είναι μια τέτοια γωνία και μπορεί να υπολογιστεί όπως φαίνεται και στο Σχήμα 15 από τις σχέσεις:

$$\cos \Theta = \frac{V_1}{\sqrt{V_1^2 + V_2^2}} \quad \text{και} \quad \sin \Theta = \frac{V_2}{\sqrt{V_1^2 + V_2^2}}.$$



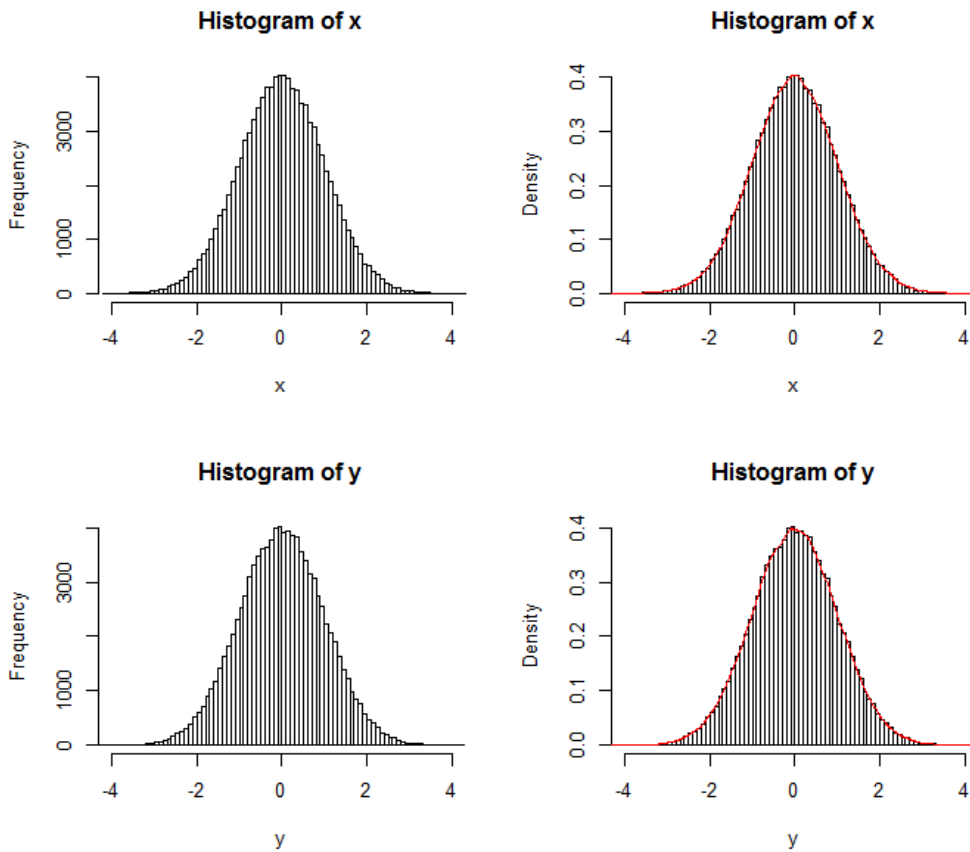
Σχήμα 15

Επομένως παράγοντας ομοιόμορφα κατανεμημένα σημεία (V_1, V_2) στον μοναδιαίο κύκλο, μπορούμε να αντικαταστήσουμε στον αλγόριθμο Box Muller το U_1 από το $R^2 = V_1^2 + V_2^2$ και τα ημίτονα και συνημίτονα από τις πιο πάνω σχέσεις. Έτσι καταλήγουμε στον πιο κάτω αλγόριθμο:

1. Παράγουμε $V_1 \sim U(-1,1)$ και $V_2 \sim U(-1,1)$.
2. Αν $V_1^2 + V_2^2 > 1$ επιστρέφουμε στο βήμα 1, αλλιώς προχωράμε στο βήμα 3.
3. Θέτουμε $R^2 = V_1^2 + V_2^2$, $X = \frac{V_1}{R} \sqrt{-2 \ln R^2}$ και $Y = \frac{V_2}{R} \sqrt{-2 \ln R^2}$.

Ο αντίστοιχος κώδικας στην R που παράγει 100000 τιμές από δύο κανονικές κατανομές $N(0,1)$ είναι ο εξής:

```
x<-rep(0,100000); y<-rep(0,100000)
for (i in 1:100000) {
  v1<-runif(1,-1,1); v2<-runif(1,-1,1)
  while ((v1^2+v2^2)>1) {
    v1<-runif(1,-1,1); v2<-runif(1,-1,1)
  }
  r<-sqrt(v1^2+v2^2); x[i]<-v1*sqrt(-2*log(r^2))/r; y[i]<-v2*sqrt(-2*log(r^2))/r
}
par(mfrow=c(2,2))
hist(x,xlim=c(-4,4),nclass=100); hist(x,freq=FALSE,xlim=c(-4,4),nclass=100); lines(density(x), col="red")
mean(x); var(x)
hist(y,xlim=c(-4,4),nclass=100); hist(y,freq=FALSE,xlim=c(-4,4),nclass=100); lines(density(y), col="red")
mean(y); var(y)
```



Σχήμα 16

Ο κώδικας μας επιστρέφει τα γραφήματα του Σχήματος 16. που όπως βλέπουμε είναι πάρα πολύ κοντά στο γράφημα της $N(0,1)$. Επίσης επιστρέφει και δειγματικούς μέσους για X και Y , 0.00013 και -0.004 αντίστοιχα, όπως και διασπορές 0.999 και 1.001. Η μέθοδος Polar Maraglia χρειάζεται κατά μέσο όρο $\frac{4}{\pi}$ επαναλήψεις (η πιθανότητα αποδοχής κάποιου σημείου είναι $\frac{\pi}{4}$) για να παράγει δύο τιμές από δύο ανεξάρτητες κανονικές κατανομές. Αν και έχει μεγαλύτερη πολυπλοκότητα από την μέθοδο Box Muller, αποφεύγει τους υπολογισμούς των ημίτονων και συνημίτονων και αυτό την καθιστά γρηγορότερη.

2.4.1.3 Κατανομή Γάμμα

Θα ξεκινήσουμε την μελέτη παραγωγής αριθμών από την κατανομή γάμμα $G(a, \lambda)$ παρατηρώντας ότι αν $X \sim G(a, 1)$ τότε $X/\lambda \sim G(a, \lambda)$ και οπότε θα αρκεστούμε στην παραγωγή αριθμών από την $G(a, 1)$. Πράγματι έστω $X \sim G(a, 1)$. Τότε η σ.π.π. της X

Θα είναι $f_x = \frac{x^{a-1}e^{-x}}{\Gamma(a)}$ και $f_{x/\lambda} = \frac{d}{dx}P\left(\frac{X}{\lambda} \leq x\right) = \frac{d}{dx}P(X \leq \lambda x) = \frac{d}{dx}F_x(\lambda x) =$
 $= f_x(\lambda x) \frac{d}{dx}(\lambda x) = \lambda f_x(\lambda x) = \lambda \frac{(\lambda x)^{a-1} e^{-\lambda x}}{\Gamma(a)} = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$ που είναι η σ.π.π. της $G(a, \lambda)$.

Θα προσπαθήσουμε να χρησιμοποιήσουμε την μέθοδο της απόρριψης, και αφού η σ.π.π. της κατανομής γάμμα λαμβάνει γνήσια θετικές τιμές στο $(0, \infty)$ θα εξετάσουμε αν μπορούμε να χρησιμοποιήσουμε σαν βοηθητική κατανομή την $E(\theta)$. Αρκεί να υπάρχει σταθερά $c < \infty$ που να ικανοποιεί την ακόλουθη σχέση:

$$\sup\{h(x)\} = \sup\left\{\frac{f(x)}{g(x)}\right\} = \sup\left\{\frac{x^{a-1}e^{-x}}{\Gamma(a)\theta e^{-\theta x}}\right\} = \sup\left\{\frac{x^{a-1}e^{(\theta-1)x}}{\Gamma(a)\theta}\right\} = c. \text{ για } x > 0$$

Όπως φαίνεται και από τον αριθμητή της πιο πάνω παράστασης αν $a < 1$ η $h(x)$ δεν μπορεί να φραχθεί άνω. Έτσι για $\alpha > 1$:

$$\frac{d}{dx}h(x) = \frac{1}{\Gamma(a)\theta} \left[(\alpha-1)x^{\alpha-2}e^{(\theta-1)x} + x^{\alpha-1}(\theta-1)e^{(\theta-1)x} \right] = \frac{x^{\alpha-2}e^{(\theta-1)x}}{\Gamma(a)\theta} (a-1+x(\theta-1)) = 0.$$

Άρα το πιθανό μέγιστο λαμβάνεται για $x_0 = \frac{a-1}{1-\theta}$ και επειδή θέλουμε $x > 0$ πρέπει $\theta < 1$.

Υπολογίζοντας και την δεύτερη παράγωγο της παράστασης στο $x_0 = \frac{a-1}{1-\theta}$ βρίσκουμε

$$\frac{d^2}{dx^2}h(x_0) = \frac{e^{1-a} \left(\frac{1-a}{\theta-1}\right)^\alpha (\theta-1)^3}{\Gamma(a)\theta(\alpha-1)^2}$$

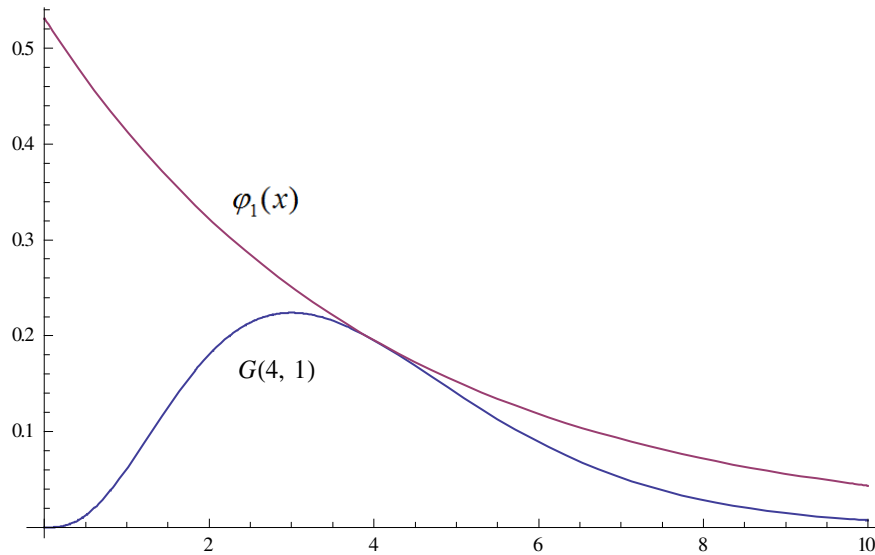
που για $\theta < 1$ και $\alpha > 1$ είναι αρνητική και άρα το x_0 είναι

μέγιστο και $c(\theta) = h(x_0) = \left(\frac{a-1}{1-\theta}\right)^{\alpha-1} \frac{e^{-(a-1)}}{\Gamma(a)\theta}$. Πρέπει τώρα να βρούμε τώρα το $\theta < 1$ που να ελαχιστοποιεί το $h(x_0)$:

$\frac{d}{d\theta}c(\theta) = \frac{a\theta-1}{\theta^2}(1-\theta)^{-\alpha} = 0$ και άρα $\theta_0 = \frac{1}{\alpha}$ και $\frac{d^2}{d\theta^2}c(\theta_0) = (\alpha-1)^{-\alpha} \alpha^{3+\alpha} > 0$ που σημαίνει ότι στο θ_0 υπάρχει ελάχιστο. Έτσι καταλήγουμε:

$$c = \sup\{h(x) : x > 0\} = \frac{a^\alpha e^{-(a-1)}}{\Gamma(a)}.$$

Στο Σχήμα 17 φαίνονται η κατανομή $G(6,1)$ καθώς και η συνάρτηση $\varphi_1(x) = \frac{4^4 e^{-(4-1)}}{\Gamma(4)} f_{E(1/4)}$ που είναι οι φάκελος της αντίστοιχης κατανομής. Όπως εξηγήσαμε και πιο πάνω όταν $a \leq 1$ δεν μπορούμε να χρησιμοποιήσουμε την εκθετική κατανομή σαν κατανομή εισήγησης. Πιο κάτω θα δούμε την μέθοδο των *Ahrens and Dieters* για $a \leq 1$.



Σχήμα 17

Έτσι ο αλγόριθμος που μας προσομοιώνει τιμές από την $G(a,l)$ είναι ο ακόλουθος:

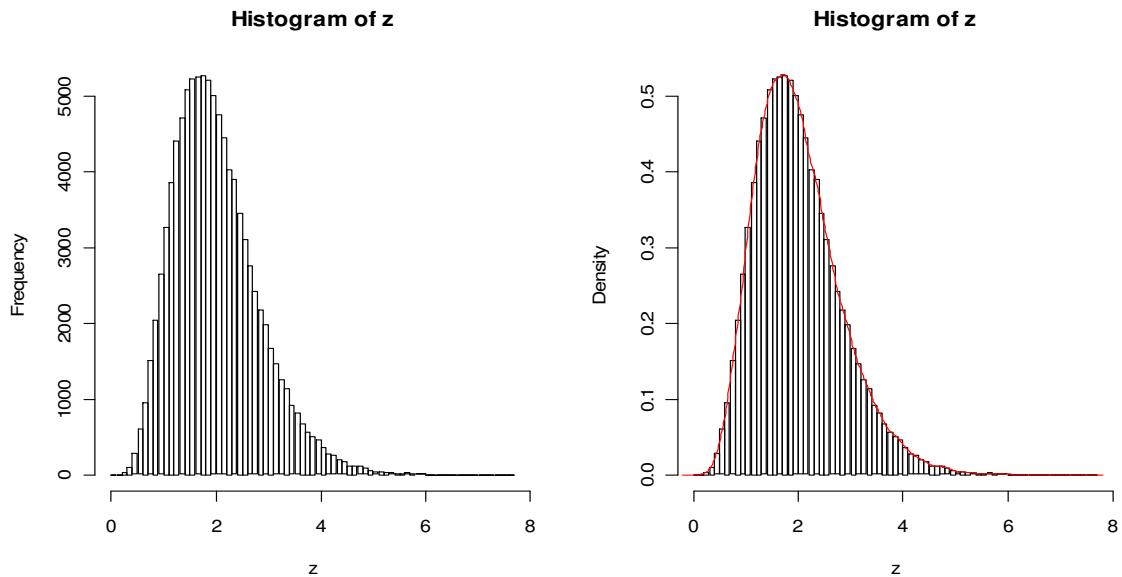
1. Παράγουμε $U_1 = U(0,1)$ και θέτουμε $y = -a \ln U_1$ (χρησιμοποιούμε την μέθοδο της αντιστροφής για την παραγωγή $y \sim E(1/a)$).
2. Παράγουμε $U_2 = U(0,1)$.
3. Εάν $U_2 \leq \frac{f(y)}{cg(y)} = \left(ye^{1-\frac{y}{a}} / a \right)^{a-1}$ τότε δεχόμαστε το y/l διαφορετικά επιστρέφουμε στο 1.

Ο αντίστοιχος κώδικας στην R που προσομοιώνει 100000 τιμές από την $G(6,3)$ είναι ο εξής:

```

z<-rep(0,100000)
i<-1
a=6
l=3
while (i<100000){
  y=-a*log(runif(1))
  u2=runif(1)
  b=(y*exp(1-y/a)/a)^(a-1)
  if (u2<=b) {
    z[i]=y/l
    i<-i+1 }
}
par(mfrow=c(1,2))
hist(z,nclass=100)
hist(z,freq=FALSE,nclass=100)
lines(density(z), col="red")
mean(z)
var(z)

```



Σχήμα 18

Ο πιο πάνω κώδικας μας επιστρέφει τα γραφήματα στο σχήμα 13 καθώς και δειγματικό μέσο 1.999 και διασπορά 0.66 που είναι πάρα πολύ κοντά στις θεωρητικές τιμές της κατανομής $\mu = 2$ και $\sigma^2 = \frac{2}{3}$.

Παρακάτω παραθέτουμε την μέθοδο των *Ahrens and Dieter, 1974* για την προσομοίωση τιμών από την κατανομή Γάμμα με παράμετρο $a \leq 1$.

Αρχικά παρατηρούμε ότι για $0 < a \leq 1$ ισχύουν οι πιο κάτω ανισότητες:

$$f(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x} \leq g(x) = \begin{cases} \frac{1}{\Gamma(a)} x^{a-1}, & 0 \leq x \leq 1 \\ \frac{1}{\Gamma(a)} e^{-x}, & 1 \leq x < \infty \end{cases}$$

Η συνάρτηση:

$$h(x) = \begin{cases} \frac{ea}{e+a} x^{a-1}, & 0 \leq x \leq 1 \\ \frac{ea}{e+a} e^{-x}, & 1 \leq x < \infty \end{cases},$$

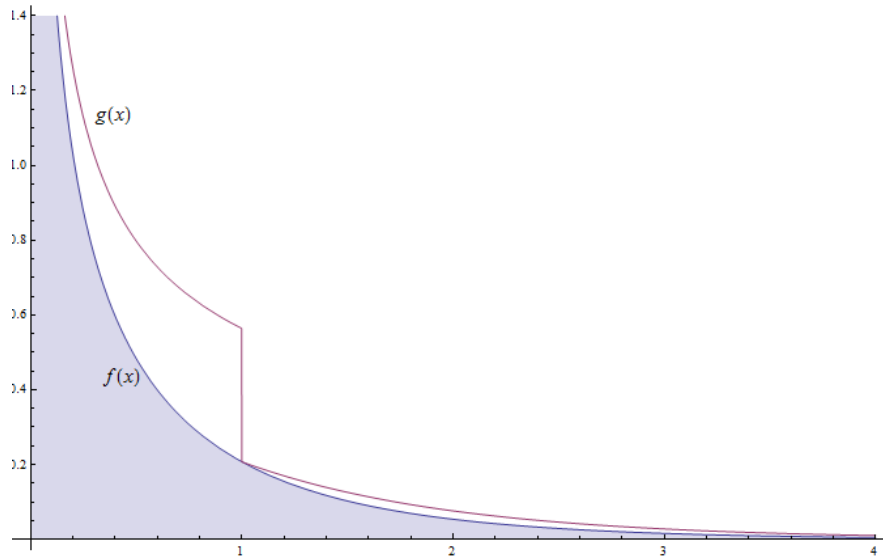
είναι συνάρτηση πυκνότητας πιθανότητας (αφού είναι θετική και όπως φαίνεται παρακάτω ολοκληρώνει στην μονάδα) και ανάλογη της $g(x)$: $h(x) = \frac{ea}{e+a} \Gamma(a) g(x)$.

$$\int_0^{\infty} h(x) dx = \int_0^1 h(x) dx + \int_1^{\infty} h(x) dx = \frac{e}{e+a} + \frac{e}{e+a} = 1.$$

Έτσι μπορούμε να χρησιμοποιήσουμε την μέθοδο της απόρριψης με φάκελο την $g(x)$ και κατανομή εισήγησης την $h(x) = Cg(x)$ με $C = \frac{ea}{e+a} \Gamma(a)$. Απομένει λοιπόν να προσομοιώσουμε τιμές από την h που μπορεί να επιτευχθεί χρησιμοποιώντας την μέθοδο της αντιστροφής. Θέτοντας $b = \frac{e+a}{e}$ και κάνοντας τις απαραίτητες πράξεις καταλήγουμε:

$$H(x) = \begin{cases} \frac{e}{e+a} x^a, & 0 \leq x \leq 1 \\ 1 - \frac{a}{e+a} e^{1-x}, & 1 \leq x < \infty \end{cases} \quad \text{και} \quad H^{-1}(p) = H^{-1}(bu) = \begin{cases} p^{\frac{1}{a}}, & 0 \leq p \leq 1 \\ -\ln\left(\frac{b-p}{a}\right), & 1 \leq p < \infty \end{cases},$$

όπου για ευκολία στις πράξεις υπολογίσαμε την αντίστροφη $H^{-1}(bu)$. Στο Σχήμα 19 φαίνονται η κατανομή γάμμα και η συνάρτηση φάκελος $g(x)$ όπως τις εισαγάγαμε παραπάνω.



Σχήμα 19

Έτσι καταλήγουμε στον πιο κάτω αλγόριθμο:

1. Παράγουμε $U = U(0,1)$ και θέτουμε $b = \frac{(e+a)}{e}$ και $p = bU$.
2. Αν $p \leq 1$ θέτουμε $x = p^{1/a}$, και παράγουμε $U_2 = U(0,1)$. Αν $U_2 > e^{-x}$ τότε επιστρέφουμε στο 1 αλλιώς δεχόμαστε το x .
3. Αν $p > 1$ θέτουμε $x = -\ln\left(\frac{b-p}{a}\right)$, και παράγουμε $U_2 = U(0,1)$. Αν $u_2 > x^{a-1}$ τότε επιστρέφουμε στο 1 αλλιώς δεχόμαστε το x .

Ο μέσος αριθμός επαναλήψεων που χρειάζεται ο αλγόριθμος είναι $n = \frac{1}{C} = \frac{e+a}{e\alpha\Gamma(\alpha)} = \frac{e+a}{e\Gamma(\alpha+1)}$.

Ο πιο κάτω πίνακας δείχνει τον μέσο αριθμό επαναλήψεων για διάφορες τιμές του α .

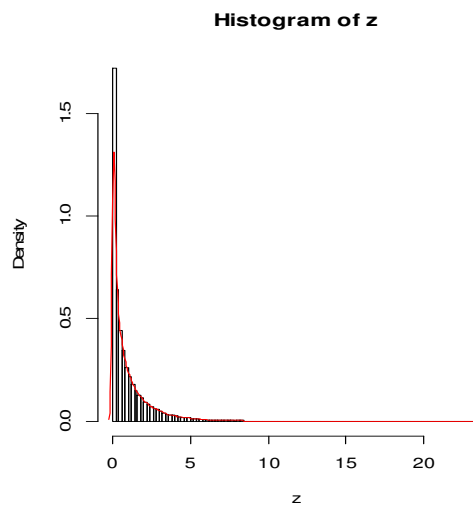
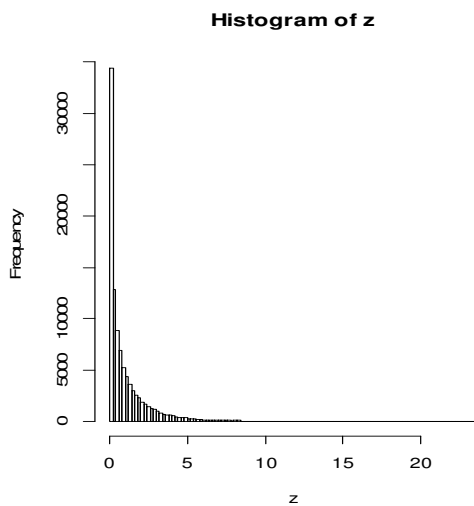
α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
n	1.000	1.090	1.169	1.237	1.239	1.336	1.366	1.384	1.390	1.384	1.368

Ο αντίστοιχος κώδικας στην R που προσομοιώνει 100000 τιμές από την $G\left(\frac{1}{2}, \frac{1}{2}\right)$ είναι ο εξής:

```

z<-rep(0,100000) ; i<-1
a=0.5; b=(exp(1)+a)/exp(1)
l=0.5
while (i<100000){
  p=b*runif(1)
  if (p<1) {
    x<-p^(1/a)
    u2=runif(1)
    if(u2<=exp(-x)) {
      z[i]=x/l
      i<-i+1 } }
  else {
    x<-log((b-p)/a)
    u2=runif(1)
    if(u2<=x^(a-1)) {
      z[i]=x/l
      i<-i+1 }
  }
}
par(mfrow=c(1,2))
hist(z,nclass=100)
hist(z,freq=FALSE,nclass=100)
lines(density(z), col="red")
mean(z)

```



Σχήμα 20

Ο πιο πάνω κώδικας μας επιστρέφει τα γραφήματα στο Σχήμα 20 καθώς και δειγματικό μέσο 0,999 και διασπορά 2,001 που είναι πάρα πολύ κοντά στις θεωρητικές τιμές της κατανομής $\mu = 1$ και $\sigma^2 = 2$.

2.4.2 Προσομοίωση Διακριτών Κατανομών

2.4.2.1 Γεωμετρική Κατανομή

Η γεωμετρική κατανομή με παράμετρο p έχει συνάρτηση μάζας πιθανότητας:

$$p_i = P(X = i) = p(1 - p)^{i-1} = pq^{i-1}, i = 1, 2, 3, \dots$$

Η τ.μ. X εκφράζει το πλήθος των δοκιμών μέχρι την εμφάνιση της πρώτης επιτυχίας σε μια ακολουθία ανεξάρτητων δοκιμών Bernoulli με πιθανότητα επιτυχίας p . Για να προσομοιώσουμε τιμές από την συγκεκριμένη κατανομή μπορούμε εύκολα να χρησιμοποιήσουμε την μέθοδο της αντιστροφής. Σύμφωνα με την μέθοδο πρέπει να υπολογίσουμε τα αθροίσματα:

$$F_k = \sum_{i=1}^k p_i = \sum_{i=1}^k pq^{i-1} = p \sum_{i=0}^{k-1} q^i = p \frac{1 - q^k}{1 - q} = 1 - q^k,$$

και δεδομένου ενός τυχαίου αριθμού $U \sim U(0,1)$ θα λαμβάναμε:
 $X = \{k \geq 1 : 1 - q^{k-1} \leq U < 1 - q^k\}$.

Η μέθοδος της αντιστροφής όμως απαιτεί κατά μέσο όρο $E(N) = \sum_{i=1}^{\infty} iP(N=i) = \sum_{i=1}^{\infty} ip_i = \dots = \frac{1}{p}$ βήματα. Άρα για μικρά p η μέθοδος δεν θα ήταν αποδοτική. Αυτό μπορούμε να το αποφύγουμε εάν παρατηρήσουμε ότι:

Αν $X = \left\lceil \frac{\ln U}{\ln q} \right\rceil + 1$ όπου $\lceil \cdot \rceil$ το ακέραιο μέρος της παράστασης τότε:

$$\begin{aligned} P(X = k) &= P\left(\left\lceil \frac{\ln U}{\ln q} \right\rceil = k - 1\right) = P\left(k - 1 \leq \frac{\ln U}{\ln q} < k\right) = P(k \ln q < \ln U \leq (k - 1) \ln q) \\ &= P(q^k < U \leq q^{k-1}) = q^{k-1} - q^k = pq^{k-1}, \text{ οπότε } X \sim Ge(p). \end{aligned}$$

Τότε για κάθε τυχαίο αριθμό που παράγουμε παίρνουμε ακριβώς ένα αριθμό από την γεωμετρική κατανομή.

Έτσι προκύπτει ο ακόλουθος αλγόριθμος:

1. Παράγουμε $U = U(0,1)$.
2. Θέτουμε $X = \left\lceil \frac{\ln U}{\ln q} \right\rceil + 1$.

Ο αντίστοιχος κώδικας στην R που μας προσομοιώνει 10000 τιμές από την γεωμετρική κατανομή με παράμετρο $p = 0.1$ είναι:

```
u<-runif(10000)
x<-rep(0,10000)
p<-0.1
q<-1-p
for (i in 1:10000) {
  x[i]<-floor(log(u[i])/log(q))+1
}
mean(x)
var(x)
```

που μας επιστρέφει δειγματικό μέσο 10.13 και διασπορά 89.74 που είναι πολύ κοντά στις θεωρητικές $\mu = \frac{1}{p} = 10$ και $\sigma^2 = \frac{q}{p^2} = 90$.

2.4.2.2 Κατανομή Poisson

Η κατανομή Poisson με μέση τιμή λ , έχει συνάρτηση μάζας πιθανότητας:

$$p_i = P(X = i) = \frac{e^{-\lambda} \lambda^i}{i!}, \quad i = 1, 2, 3, \dots$$

Η τ.μ. X που ακολουθεί κατανομή Poisson εκφράζει το πλήθος των επιτυχιών σπάνιων ενδεχομένων σε ένα μεγάλο αριθμό δοκιμών σε ένα συγκεκριμένο χρονικό ή χωρικό διάστημα.

Από την συνάρτηση μάζας πιθανότητας παρατηρούμε ότι:

$$P_{i+1} = \frac{e^{-\lambda} \lambda^{i+1}}{(i+1)!} = \frac{e^{-\lambda} \lambda^i \lambda}{(i+1)i!} = \frac{\lambda}{i+1} \frac{e^{-\lambda} \lambda^i}{i!} = \frac{\lambda}{i+1} P_i.$$

Μα βάση την πιο πάνω παρατήρηση και χρησιμοποιώντας την μέθοδο της αντιστροφής καταλήγουμε στον αλγόριθμο προσομοίωσης τιμών από την κατανομή Poisson:

1. Παράγουμε $U = U(0,1)$.
2. Θέτουμε $p = e^{-\lambda}$, $F = p$ και $i = 0$.
3. Αν $U < F$ τότε $X = i$ και σταματώ αλλιώς πηγαίνουμε στο βήμα 4.
4. Θέτω $p = p \frac{\lambda}{i+1}$, $F = F + p$ και $i = i+1$ και επιστρέφουμε στο βήμα 3.

Ο αντίστοιχος κώδικας στην R που προσομοιώνει 10000 τιμές από την κατανομή Poisson με $\lambda = 5$ στην R είναι:

```
u<-runif(10000)
x<-rep(0,10000)
l<-5
for (k in 1:10000) {
  p<-exp(-l)
  f<-p
  i<-0
  while (u[k]>f){
    p<-p*l/(i+1)
    f<-f+p
    i<-i+1
  }
  x[k]<-i
}
mean(x)
var(x)
library(vcd) ## loading vcd package
gf <- goodfit(x,type="poisson",method="MinChisq")
summary(gf)
```

Ο κώδικας επιστρέφει επίσης το αποτέλεσμα του ελέγχου καλής προσαρμογής του Pearson:

Goodness-of-fit test for poisson distribution

	χ^2	df	$P(> \chi^2)$
<i>Pearson</i>	14.73276	15	0.4708327

Η p -τιμή του παραπάνω ελέγχου είναι $p \approx 0.47$, οπότε σε επίπεδο σημαντικότητας 5% δεν έχουμε ενδείξεις εναντίον της υπόθεσης ότι το παραγόμενο δείγμα προέρχεται από την κατανομή Poisson. Ο κώδικας στην R μας επιστρέφει επίσης μέση τιμή 4.9935 και διασπορά 5.093367 που είναι πάρα πολύ κοντά στις θεωρητικές τιμές της εν λόγω κατανομής.

2.4.2.3 Διωνυμική Κατανομή

Η συνάρτηση μάζας πιθανότητας της διωνυμικής κατανομής $B(n, p)$ με παραμέτρους n, p είναι:

$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, 2, \dots, n.$$

Αν η τ.μ. $X \sim B(n, p)$ τότε μπορούμε να πούμε ότι η X εκφράζει το πλήθος των επιτυχιών ενός ενδεχομένου με πιθανότητα επιτυχίας p σε n ανεξάρτητες δοκιμές Bernoulli. Για την παραγωγή δείγματος από την συγκεκριμένη κατανομή μπορούμε και πάλι με ευκολία να χρησιμοποιήσουμε την μέθοδο της αντιστροφής. Η μέθοδος της αντιστροφής για την διωνυμική κατανομή υπολογίζεται ότι έχει πολυπλοκότητα $O(np)$ και μάλιστα είναι η πιο συμφέρουσα μέθοδος. Για σκοπούς όμως της εργασίας αυτής θα παρουσιάσουμε μια άλλη μέθοδο χρησιμοποιώντας την πιο κάτω ιδιότητα της διωνυμικής κατανομής.

Αν οι ανεξάρτητες τμ $Z_1, Z_2, \dots, Z_n \sim Bernoulli$ με παράμετρο p τότε η

$$Z = \sum_{i=1}^n Z_i \sim B(n, p).$$

Η μέθοδος τώρα είναι προφανής. Τα οφέλη αυτής της μεθόδου είναι ότι είναι εύκολη στην εφαρμογή και επίσης αποφεύγουμε τους λογαριθμικούς και εκθετικούς υπολογισμούς. Παρόλα αυτά η συγκεκριμένη μέθοδος έχει πολυπλοκότητα $O(n)$. Ο αλγόριθμος έχει ως εξής:

1. Θέτουμε $Z = 0, i = 1$.
2. Αν $i > n$ τότε επιστρέφω το Z και σταματώ αλλιώς πηγαίνουμε στο βήμα 3.
3. Παράγω $U \sim U(0,1)$ και θέτω $Z = Z + \lfloor U + p \rfloor, i = i + 1$ και επιστρέφουμε στο βήμα 2.

Ο αντίστοιχος κώδικας στην R που προσομοιώνει 10000 τιμές από την $B(10,0.1)$ είναι:

```
z<-rep(0,10000)
p<-0.1
n<-10
for (k in 1:10000) {
x<-0
for (i in 1:n){
  u<-runif(1)
  x<-x+floor(u+p)
}
z[k]<-x
}
mean(z)
var(z)
library(vcd) ## loading vcd package
gf <- goodfit(z,type= "binomial", par = list(prob = 0.1, size = 10))
summary(gf)
```

Ο κώδικας επιστρέφει επίσης το αποτέλεσμα του ελέγχου καλής προσαρμογής του Pearson:

Goodness-of-fit test for binomial distribution

	χ^2	<i>df</i>	$P(> \chi^2)$
<i>Pearson</i>	9.248449	10	0.5086985

Η p -τιμή του παραπάνω ελέγχου είναι $p \approx 0.51$, οπότε σε επίπεδο σημαντικότητας 5% δεν έχουμε ενδείξεις εναντίον της υπόθεσης ότι το παραγόμενο δείγμα προέρχεται από την διωνυμική κατανομή. Ο κώδικας στην R μας επιστρέφει επίσης μέση τιμή 1.0042 και διασπορά 0.8888 που είναι πάρα πολύ κοντά στις θεωρητικές τιμές της εν λόγω κατανομής ($\sigma^2 = np(1-p) = 0.9, \mu = np = 1$).

2.4.3 Προσομοίωση Άλλων Γνωστών Κατανομών

Με όσα αναλύσαμε πιο πάνω μπορούμε τώρα να δείξουμε πως προσομοιώνουμε από κάποιες άλλες κατανομές (διακριτές και συνεχείς), χρησιμοποιώντας τις κατανομές που παρουσιάσαμε στις πιο πάνω παραγράφους και κάποιες ιδιότητες αυτών των κατανομών:

- i. Κατανομή X_n^2 : Αν $Z_1, Z_2, \dots, Z_n \sim N(0,1)$, ανεξάρτητες, τότε η $Q \equiv \sum_{i=1}^n Z_i^2 \sim X_n^2$:
 1. Παράγουμε $Z_1, Z_2, \dots, Z_n \sim N(0,1)$ με μια από τις μεθόδους που δείξαμε στην παράγραφο 2.4.1.2.
 2. Επιστρέφω το άθροισμα $Q = Z_1^2 + Z_2^2 + \dots + Z_n^2$.

- ii. Κατανομή Cauchy: Αν $Z_1, Z_2 \sim N(0,1)$, ανεξάρτητες, τότε η $X = \frac{Z_1}{Z_2} \sim Cauchy(0,1)$:
 1. Παράγουμε $Z_1, Z_2 \sim N(0,1)$ με μια από τις μεθόδους που δείξαμε στην παράγραφο 2.4.1.2.
 2. Επιστρέφω το πηλίκο $X = \frac{Z_1}{Z_2}$.

- iii. Αρνητική Διωνυμική Κατανομή: Αν $\lambda \sim G(a,b)$ και $X \sim Poisson(\lambda)$, τότε η $X \sim NB(a,b)$:
 1. Παράγουμε $\lambda \sim G(a,b)$ με την μέθοδο που δείξαμε στην παράγραφο 2.4.1.3.
 2. Παράγουμε $X \sim E(\lambda)$, με την μέθοδο της αντιστροφής, και επιστρέφω το X .

- iv. Κατανομή Student με n βαθμούς ελευθερίας: Αν $Z \sim N(0,1)$ και $X \sim X_n^2$, τότε η $T = \frac{Z}{\sqrt{X/n}} \sim T_n$:
 1. Παράγουμε $X_1, X_2, \dots, X_n \sim N(0,1)$ και θέτουμε $X = \sum_{i=1}^n X_i^2$, με μια από τις μεθόδους που δείξαμε στην παράγραφο 2.4.1.2.
 2. Παράγουμε $Z \sim N(0,1)$.

3. Επιστρέφω το $T = \frac{Z}{\sqrt{X/n}}$.

v. Κατανομή F του Snedecor με m και n βαθμούς ελευθερίας: Αν $X \sim X_m^2$ και

$Y \sim X_n^2$ τότε $F = \frac{X}{Y} \sim F(m,n)$:

1. Παράγουμε $X_1, X_2, \dots, X_m \sim N(0,1)$ και θέτουμε $X = \sum_{i=1}^m X_i$ με μια από τις μεθόδους που δείξαμε στην παράγραφο 2.4.1.2.

2. Παράγουμε $X_1, X_2, \dots, X_n \sim N(0,1)$ και θέτουμε $Y = \sum_{i=1}^n X_i$ με μια από τις μεθόδους που δείξαμε στην παράγραφο 2.4.1.2.

3. Επιστρέφω το πηλίκο $F = \frac{X}{Y}$.

vi. Κατανομή Γάμμα $G(n, \lambda)$, $n \in N$: Αν $X_1, X_2, \dots, X_n \sim E(\lambda)$ ανεξάρτητες, τότε

$\sum_{i=1}^n X_i \sim G(n, \lambda)$:

1. Παράγω $X_1, X_2, \dots, X_n \sim E(\lambda)$ με την μέθοδο της αντιστροφής.

2. Επιστρέφω το άθροισμα $X = X_1 + X_2 + \dots + X_n$.

vii. Κατανομή Βήτα: Αν $X \sim G(a, \lambda)$ και $Y \sim G(b, \lambda)$ τότε $B = \frac{X}{X+Y} \sim B(a,b)$:

1. Παράγω $X \sim G(a, \lambda)$ και $Y \sim G(b, \lambda)$, με την μέθοδο της απόρριψης, όπως αναλύθηκε στην παράγραφο 2.4.1.3.

2. Επιστρέφω το πηλίκο $\frac{X}{X+Y}$.

ΚΕΦΑΛΑΙΟ 3

ΠΡΟΣΟΜΟΙΩΣΗ ΣΤΟΧΑΣΤΙΚΩΝ ΑΝΕΛΙΞΕΩΝ

3.1 Εισαγωγή

Είναι γνωστό ότι, αν το αποτέλεσμα ενός τυχαίου πειράματος είναι ένας αριθμός στο R , τότε αυτό να μπορεί να εκφραστεί χρησιμοποιώντας μία τυχαία μεταβλητή X . Υπάρχουν όμως τυχαία πειράματα στα οποία το αποτέλεσμα δεν μπορεί να περιγραφεί από έναν μόνο αριθμό, αλλά από k αριθμούς. Σε αυτή την περίπτωση συνήθως χρησιμοποιούμε τυχαία διανύσματα $\mathbf{X} = (X_1, X_2, \dots, X_k) \in R^k$ για να εκφράσουμε το αποτέλεσμα.

Υπάρχουν όμως και πειράματα με ακόμη «συνθετότερο» σύνολο αποτελεσμάτων. Για παράδειγμα, η παρακολούθηση της εξέλιξης ενός τυχαίου πειράματος στο χρόνο (ή στο χώρο) δεν μπορεί να εκφρασθεί με μία ή με k μόνο τυχαίες μεταβλητές (π.χ. η παρακολούθηση μιας διαδικασίας αφίξεων πελατών σε ένα κατάστημα, ή η παρακολούθηση της αυξομείωσης της τιμής μιας μετοχής κατά τη διάρκεια του χρόνου). Σε τέτοιες περιπτώσεις επιθυμούμε να γνωρίζουμε την τιμή μιας ή περισσότερων ποσοτήτων (π.χ. πλήθος αφίξεων ή τιμή μετοχής) κάθε χρονική στιγμή t . Ο απλούστερος τρόπος να εκφράσουμε ένα τέτοιο «σύνθετο» αποτέλεσμα είναι μέσω μιας οικογένειας τυχαίων μεταβλητών έτσι ώστε η τ.μ. $X(t)$ να εκφράζει την τιμή της ποσότητας που μελετούμε την χρονική στιγμή t . Η οικογένεια αυτή των τ.μ. η οποία περιγράφει την εξέλιξη ενός τυχαίου πειράματος στο χρόνο (ή στο χώρο) καλείται *στοχαστική διαδικασία* ή *στοχαστική ανέλιξη*. Η προσομοίωση στοχαστικών ανελίξεων έχει ιδιαίτερο ενδιαφέρον και σε αυτό το κεφάλαιο θα παρουσιάσουμε την προσομοίωση ανέλιξης Poisson, της κίνησης Brown και του απλού τυχαίου περίπατου με απορροφητικά φράγματα.

3.2 Προσομοίωση Ανέλιξης Poisson

Το μοντέλο της διαδικασίας Poisson εμφανίζεται αρκετά συχνά στην πράξη (π.χ. τηλεφωνήματα που φθάνουν σε ένα τηλεφωνικό κέντρο, πελάτες που φθάνουν σε ένα οποιοδήποτε κέντρο εξυπηρέτησης, σωματίδια που εκπέμπονται σε χρόνο t από κάποια πηγή, δυστυχήματα που συμβαίνουν σε χρόνο t και ανάλογα τυχαία φαινόμενα).

Όταν μια στοχαστική ανέλιξη απαριθμεί ένα πλήθος συμβάντων που συνέβησαν μέχρι και τον χρόνο t (όπως για παράδειγμα το πλήθος των πελατών που μπαίνει σε μια τράπεζα κατά την διάρκεια του χρόνου) καλείται *απαραριθμήτρια στοχαστική ανέλιξη*.

Η απαραριθμήτρια στοχαστική ανέλιξη $\{X(t), t \geq 0\}$ είναι ανέλιξη Poisson αν:

- 1) $X(0) = 0$.
- 2) Έχει ανεξάρτητες και στάσιμες προσαυξήσεις (στην ομογενή περίπτωση).
- 3) Για κάθε $0 \leq s < t$ η τυχαία μεταβλητή $X(t) - X(s)$ είναι Poisson με παράμετρο $\lambda(t-s)$, $\lambda = \lambda(t) > 0$ ή αντίστοιχα:
- 3') $P(X(t+h) - X(t) = 1) = \lambda h + o(h)$ και

$$P(X(t+h) - X(t) = 0) = 1 - \lambda h + o(h) \text{ όπου } h > 0 \text{ πολύ μικρό και } \lim_{h \rightarrow 0} \frac{o(h)}{h} = 0.$$

Από τις δύο τελευταίες σχέσεις έχουμε ότι $P(X(t+h) - X(t) = k) = o(h)$, $k = 2, 3, \dots$

3.2.1 Ομογενής Ανέλιξη Poisson

Όταν η παράμετρος της διαδικασίας είναι σταθερή στον χρόνο, δηλαδή $\lambda(t) = \lambda = c$, τότε η ανέλιξη Poisson καλείται ομογενής. Σε μια ανέλιξη Poisson μας ενδιαφέρουν εκτός από την κατανομή της $\{X(t), t \geq 0\}$ και οι εξής τυχαίες μεταβλητές που σχετίζονται με αυτή:

W_n : Χρόνος αναμονής μέχρι την πραγματοποίηση του n -οστού γεγονότος, και

T_n : Ο ενδιάμεσος χρόνος μεταξύ του $(n-1)$ -οστού και n -οστού γεγονότος.

Οι πιο πάνω τυχαίες μεταβλητές συνδέονται μεταξύ τους με τις εξής προφανής σχέσεις:

$$W_n = T_1 + T_2 + \dots + T_n = \sum_{i=1}^n T_i,$$

$$T_n = W_n - W_{n-1}.$$

Αποδεικνύεται ότι η συνάρτηση πυκνότητας πιθανότητας της W_n είναι η $f_{W_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, t > 0$ καθώς και ότι οι ενδιάμεσοι χρόνοι $T_i, i=1,2,\dots,n$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με κοινή κατανομή την εκθετική κατανομή με παράμετρο λ (την παράμετρο της αντίστοιχης Poisson). Επίσης από την τρίτη ιδιότητα της ανέλιξης Poisson ισχύει:

$$P(X(t+s) - X(t) = n) = e^{-\lambda(t+s-t)} \frac{(\lambda(t+s-t))^n}{n!} = e^{-\lambda s} \frac{(\lambda s)^n}{n!},$$

δηλαδή ο αριθμός των γεγονότων σε ένα χρονικό διάστημα μήκους s ακολουθεί κατανομή Poisson με παράμετρο λs .

Με βάση τα παραπάνω, για να προσομοιώσουμε τιμές από την κατανομή Poisson με παράμετρο λ , μπορούμε να προσομοιώσουμε μια ομογενή ανέλιξη Poisson με την ίδια παράμετρο λ και να μετρήσουμε τον αριθμό των συμβάντων που συμβαίνουν σε χρονικό διάστημα μοναδιαίου μήκους. Για να προσομοιώσουμε μια ομογενή ανέλιξη Poisson με παράμετρο λ για το χρονικό διάστημα $[0, t_0]$ αρκεί να παράξουμε ανεξάρτητες τιμές από μια εκθετική κατανομή με παράμετρο λ ($T_1, T_2, \dots, T_k, \dots$), χρησιμοποιώντας την μέθοδο της αντιστροφής και ελέγχοντας πότε $T_1 + T_2 + \dots + T_k + T_{k+1} > t_0$. Τότε X_k μια ανεξάρτητη τιμή από την κατανομή Poisson με παράμετρο λt_0 . Η ανέλιξη $\{X(t), t \in [0, t_0]\}$ καθορίζεται από τους διαδοχικούς χρόνους πραγματοποίησης συμβάντων, δηλαδή τους χρόνους:

$$T_1, T_1 + T_2, T_1 + T_2 + T_3, \dots$$

Έτσι καταλήγουμε στον παρακάτω αλγόριθμο:

- 1) Θέτω $t=0$ και $X=0$.
- 2) Παράγω $U \sim U(0,1)$ και θέτω $T = -\frac{1}{\lambda} \ln(U)$.
- 3) Θέτω $t=t+T$ και αν $t > t_0$ σταματώ αλλιώς προχωρώ στο βήμα 4.
- 4) Θέτω $X = X + 1$ και $S(X) = t$.
- 5) Επιστρέφω στο βήμα 2.

Στον παραπάνω αλγόριθμο t είναι ο χρόνος, X το πλήθος των συμβάντων μέχρι τον χρόνο t , και $S(X_i)$ οι διαδοχικοί χρόνοι των συμβάντων. Το τελικό X όπως αναφέραμε θα είναι τυχαίος αριθμός από την κατανομή Poisson με παράμετρο λt_0 .

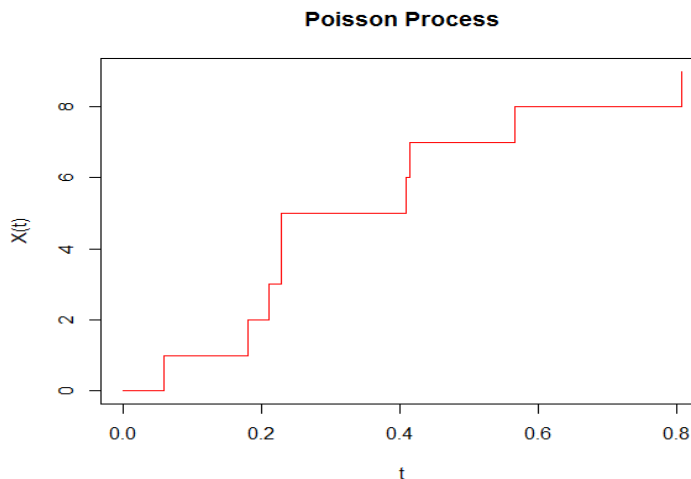
Παρακάτω δίνουμε τον αλγόριθμο προσομοίωσης ανέλιξης Poisson με $\lambda=5$ για τον χρόνο $[0,1]$.

```

t<-0; X<-0
t0<-1; l<-5
T<-log(runif(1))/l
t<-t+T; s<-{0}
while (t<=t0){
    X<-X+1
    s<-append(s,t)
    T<-log(runif(1))/5
    t<-t+T
}
X; s
x<-c(0:(length(s)-1))
plot(s,x, col="red",type='s',main="Poisson Process",cex=0.8,xlab="t",ylab="X(t)")
    
```

Ο αλγόριθμος αυτός μας επιστρέφει

$X=9$ και $s = [0.00000000 0.05816643 0.17942653 0.21090287 0.22771882 0.22796383 0.40854034 0.41432912 0.56605212 0.80742651]$, όπου X είναι ο τελικός αριθμός των συμβάντων και s οι διαδοχικοί χρόνοι των συμβάντων. Αναπαριστώντας γραφικά την ανέλιξη Poisson έχουμε:



Σχήμα 21

Επαναλαμβάνοντας τον πάνω αλγόριθμο και αποθηκεύοντας σε ένα διάνυσμα τους τελικούς αριθμούς X , που λαμβάνουμε από την κάθε επανάληψη, προσομοιώνουμε τιμές από την κατανομή Poisson με $\lambda = \lambda t_0 = 5$.

Συγκεκριμένα ο κώδικας στην R είναι:

```

y<-{}
s2<-{}
for (i in 1:10000){
t<-0
X<-0
t0<-1
l<-5
T<--log(runif(1))/l
t<-t+T
s<-{}
while (t<=t0){
  X<-X+1
  s<-append(s,t)
  s2<-append(s2,t)
  T<--log(runif(1))/5
  t<-t+T
}
y<-append(y,X)
}
hist(s2,freq=FALSE,nclass=30)
mean(y)
var(y)
library(vcd) ## loading vcd package
gf <- goodfit(y,type= "poisson",method= "MinChisq")
summary(gf)

```

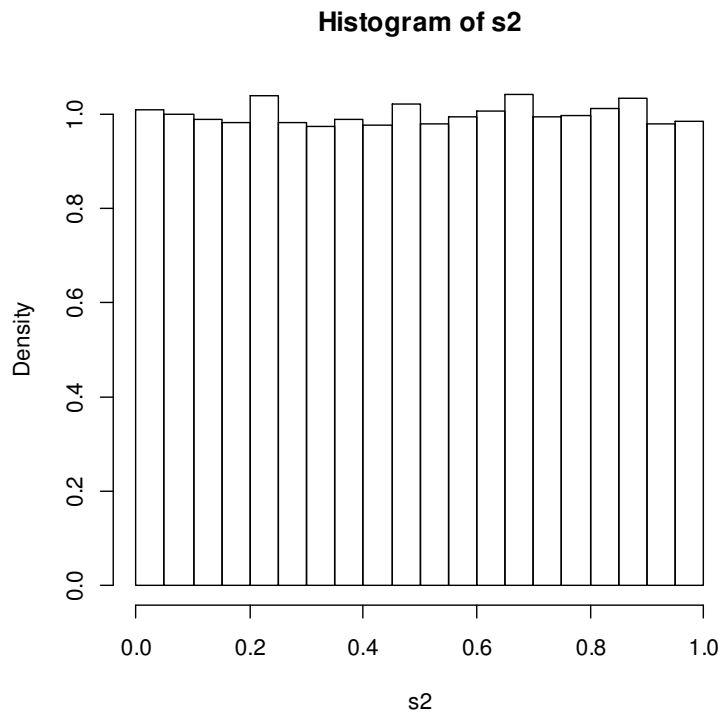
Ο έλεγχος καλής προσαρμογής του Pearson μας επιστρέφει:

Goodness-of-fit test for poisson distribution

	χ^2	df	$P(> \chi^2)$
<i>Pearson</i>	13.86065	14	0.4601433

Με ανάλογα επιχειρήματα όπως και στο προηγούμενο κεφάλαιο σε επίπεδο σημαντικότητας 5% δεν έχουμε ενδείξεις εναντίον της υπόθεσης ότι το παραγόμενο

δείγμα ακολουθεί την κατανομή Poisson. Επίσης ο κώδικας μας επιστρέφει δειγματικό μέσο 4.97 και διασπορά 4.82, πολύ κοντά στις θεωρητικές τις εν λόγω κατανομής $\mu = \sigma^2 = \lambda = 5$. Στον κώδικα συμπεριλάβαμε επίσης και μια μεταβλητή την s_2 η οποία αποθηκεύει όλες τις χρονικές τιμές των συμβάντων για να μελετήσουμε την παράμετρο της διαδικασίας (την συχνότητα των συμβάντων στα διάφορα υποδιαστήματα του $[0,1]$). Ακολούθως αναπαριστούνται με ένα ιστόγραμμα. Όπως φαίνεται και στο Σχήμα 22 τα συμβάντα κατανέμονται ομοιόμορφα και άρα η παράμετρος είναι σταθερή.



Σχήμα 22

3.2.2 Μη Ομογενής Ανέλιξη Poisson

Ο ορισμός της μη ομογενούς στοχαστικής ανέλιξης Poisson διαφέρει από τον ορισμό της ομογενούς ανέλιξης στο ότι δεν έχει στάσιμες προσαιξήσεις. Σε αυτή την περίπτωση έχουμε ότι η παράμετρος είναι $\lambda(t)$, δηλαδή είναι συνάρτηση του χρόνου t . Είναι γεγονός ότι κατά την μαθηματική μοντελοποίηση πολλών φυσικών φαινομένων, θεωρείται η ομογενής διαδικασία Poisson αντί της μη ομογενούς η οποία συνήθως περιγράφει ορθότερα το φαινόμενο. Αυτό συμβαίνει διότι η μη ομογενής δεν μπορεί εύκολα να μελετηθεί αναλυτικά. Για το λόγο αυτό είναι πολύ χρήσιμη η μελέτη της μέσω προσομοίωσης. Στη συνέχεια θα παρουσιάσουμε δύο μεθόδους προσομοίωσης της μη ομογενούς ανέλιξης Poisson.

3.2.2.1 Μέθοδος εκλέπτυνσης της ομογενούς ανέλιξης Poisson

Θεωρούμε τα ακόλουθα:

- 1) $p(t), t > 0$: μια δεδομένη συνάρτηση με τιμές στο $[0,1]$,
- 2) $\{X(t), t \geq 0\}$ μια ομογενής διαδικασία Poisson με παράμετρο λ και
- 3) $\{M(t), t \geq 0\}$ μια διαδικασία η οποία καταμετρά το i συμβάν της $\{X(t), t \geq 0\}$ που συμβαίνει σε χρόνο S_i με πιθανότητα $p(S_i)$ και κάθε συμβάν της ομογενούς $\{X(t), t \geq 0\}$ καταμετράται η όχι στην $\{M(t), t \geq 0\}$ ανεξάρτητα από τα υπόλοιπα.

Για μικρό $h > 0$ ώστε $p(t)$ σταθερή στο $[t, t+h]$ χρησιμοποιώντας το θεώρημα ολικής πιθανότητας και την ιδιότητα (4) της ανέλιξης Poisson όπως ορίστηκε πιο πάνω έχουμε:

$$\begin{aligned}
 P(M(t+h) - M(t) = k) &= \sum_{i=k}^{\infty} P[(M(t+h) - M(t)) = k \mid X(t+h) - X(t) = i] \cdot P(X(t+h) - X(t) = i) \\
 &= \sum_{i=k}^{\infty} \binom{i}{k} p(t)^k (1-p(t))^{i-k} \cdot P(X(t+h) - X(t) = i) \\
 &= \begin{cases} (1-\lambda h + o(h)) + (1-p(t))(\lambda h + o(h)) + \sum_{i=2}^{\infty} (1-p(t))^i o(h) & , k = 0 \\ p(t)(\lambda h - o(h)) + \sum_{i=2}^{\infty} i p(t) (1-p(t))^{i-1} o(h) & , k = 1 \\ \sum_{i=2}^{\infty} \binom{i}{k} p(t)^k (1-p(t))^{i-k} o(h) & , k \geq 2 \end{cases} \\
 &= \begin{cases} 1 - p(t)\lambda h + o(h) & , k = 0 \\ p(t)\lambda h + o(h) & , k = 1 \\ o(h) & , k \geq 2. \end{cases}
 \end{aligned}$$

Έτσι η διαδικασία $\{M(t), t \geq 0\}$ είναι μια μη ομογενής διαδικασία με παράμετρο $\lambda(t) = \lambda p(t)$. Έτσι μπορούμε να προσομοιώσουμε μια μη ομογενή διαδικασία με παράμετρο $\lambda(t)$ βασιζόμενοι σε μια ομογενή διαδικασία με παράμετρο λ . Συγκεκριμένα προσομοιώνουμε την ομογενή διαδικασία και δεχόμαστε το συμβάν που πραγματοποιήθηκε στον χρόνο S_i με πιθανότητα $p(S_i) = \frac{\lambda(S_i)}{\lambda}$. Έτσι καταλήγουμε στον

πιο κάτω αλγόριθμο προσομοίωσης της μη ομογενούς ανέλιξης Poisson για το χρονικό διάστημα $[0, t_0]$:

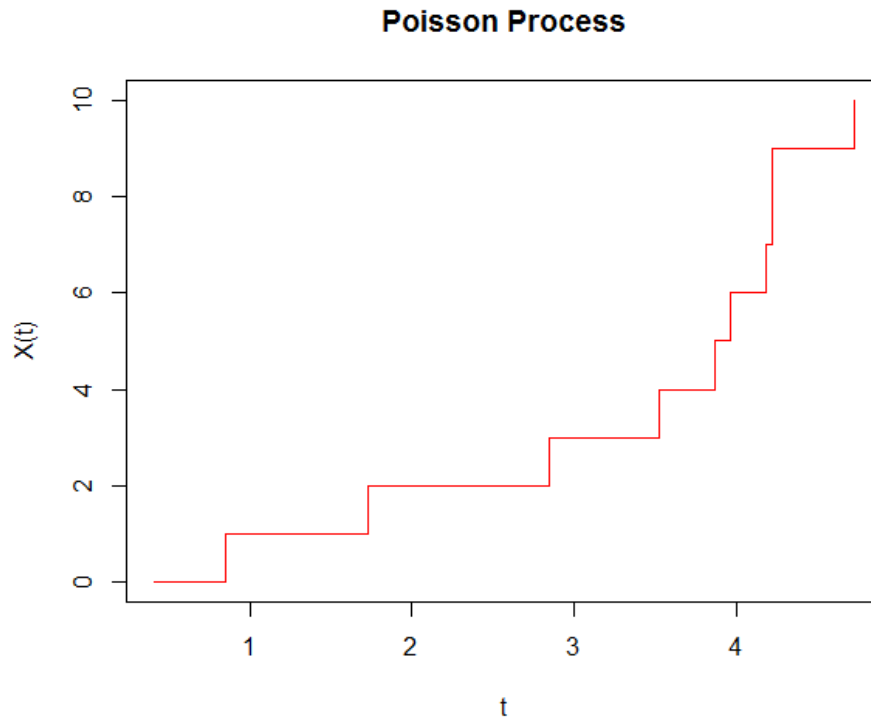
- 1) Θέτω $\lambda = \sup\{\lambda(t) : t \in [0, t_0]\}$ (θέτω *supremum* για να εξασφαλίσω ότι $p(t) = \frac{\lambda(t)}{\lambda} \in [0, 1]$)
- 2) Θέτω $t = 0$ και $X = 0$.
- 3) Παράγω $U \sim U(0,1)$ και θέτω $T = -\frac{1}{\lambda} \ln(U)$.
- 4) Θέτω $t = t + T$ και αν $t > t_0$ σταματώ αλλιώς προχωρώ στο βήμα 5.
- 5) Παράγω $U_2 \sim U(0,1)$ και αν $U_2 < \frac{\lambda(t)}{\lambda}$ θέτω $X = X + 1$ και $S(X) = t$.
- 6) Επιστρέφω στο βήμα 3.

Ο αντίστοιχος κώδικας στην R που προσομοιώνει την μη ομογενή ανέλιξη με $\lambda(t) = t$ στο $[0, 5]$ είναι ($\lambda = \sup\{\lambda(t) : t \in [0, 5]\} = 5$):

```
t<-0
X<-0
t0<-5
l<-5
T<-log(runif(1))/l
t<-t+T
s<-{}
while (t<=t0){
  u<-runif(1)
  if (u<t/l) {
    X<-X+1
    s<-append(s,t)}
  T<-log(runif(1))/l
  t<-t+T
}
X
s
x<-c(0:(length(s)-1))
plot(s,x, col="red",type='s',main="Poisson Process",cex=0.8,xlab="t",ylab="X(t)")
(s, color=c("red"),main="Poisson Process",cex=0.8,xlab="t",ylab="X(t)")
```

Ο αλγόριθμος αυτός μας επιστρέφει

$X=11$ και $s=[0.4109084\ 0.8452551\ 1.7312196\ 2.8472005\ 3.5261394\ 3.8678601\ 3.9631016\ 4.1873404\ 4.2194634\ 4.2230656\ 4.7307792]$, όπου X είναι ο τελικός αριθμός των συμβάντων και s οι διαδοχικοί χρόνοι των συμβάντων. Αναπαριστώντας γραφικά την ανέλιξη Poisson έχουμε:



Σχήμα 23

Επαναλαμβάνοντας την διαδικασία πολλές φορές οι τελικοί αριθμοί $X(t_0)$ θα ακολουθούν κατανομή Poisson με παράμετρο $\lambda = \int_{t=0}^{t=t_0} \lambda(t)dt$. Στην συγκεκριμένη περίπτωση $\lambda = \int_{t=0}^{t=5} \lambda(t)dt = \int_{t=0}^{t=5} tdt = \frac{25}{2}$. Όπως και στην περίπτωση της ομογενούς, στον κώδικα της R συμπεριλήφθηκε και η μεταβλητή s_2 που καταγράφει όλους τους χρόνους των συμβάντων που γίνονται αποδεκτά για να μελετήσουμε την παράμετρο. Ο κώδικας και τα αποτελέσματα είναι τα ακόλουθα:

```

y<-{}
s2<-{}
for (i in 1:10000){
t<-0
X<-0
t0<-5
l<-5
T<--log(runif(1))/l
t<-t+T
s<-{}
while (t<=t0){
  u<-runif(1)
  if (u<t/l){
    X<-X+1
    s<-append(s,t)
    s2<-append(s2,t)}
  T<--log(runif(1))/5
  t<-t+T
}
y<-append(y,X)
}
hist(s2,freq=FALSE,nclass=30)
mean(y)
var(y)
library(vcd) ## loading vcd package
gf <- goodfit(y,type= "poisson",method= "MinChisq")
summary(gf)

```

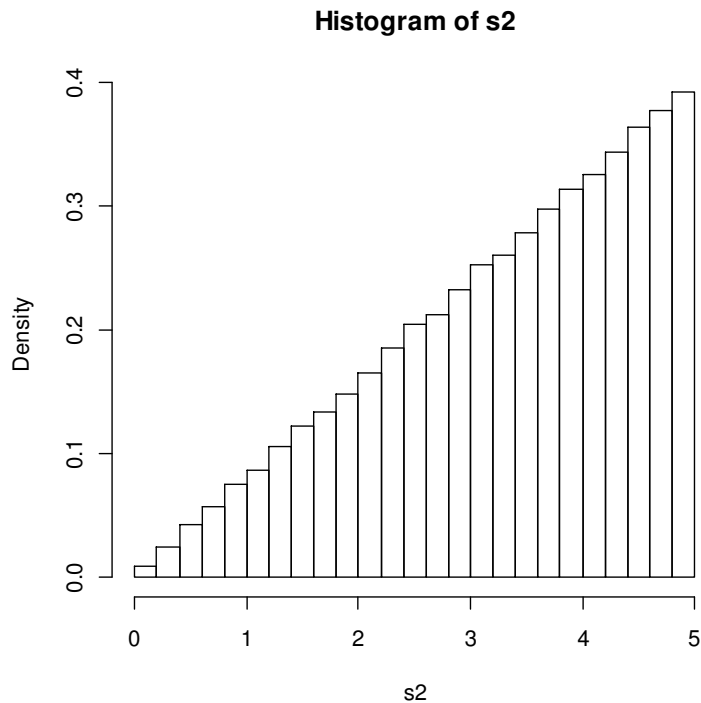
Ο έλεγχος καλής προσαρμογής του Pearson επιστρέφει:

Goodness-of-fit test for poisson distribution

	χ^2	df	$P (> \chi^2)$
<i>Pearson</i>	26.05312	26	0.4601876

Με ανάλογα επιχειρήματα όπως και στο προηγούμενο κεφάλαιο σε επίπεδο σημαντικότητας 5% δεν έχουμε ενδείξεις εναντίον της υπόθεσης ότι το παραγόμενο δείγμα ακολουθεί την κατανομή Poisson. Επίσης ο κώδικας μας επιστρέφει δειγματικό μέσο 12.5223 και διασπορά 12.45015, πολύ κοντά στις θεωρητικές τις εν λόγω κατανομής ($\mu = \sigma^2 = \lambda = 12.5$). Ακολουθώντας αναπαριστούνται τα συμβάντα που έγιναν

αποδεκτά στα διάφορα υποδιαστήματα $[0,5]$ με ένα ιστόγραμμα. Όπως φαίνεται και στο Σχήμα 24 η παράμετρος είναι αυξάνεται γραμμικά.



Σχήμα 24

Ένα σοβαρό μειονέκτημα αυτού του αλγόριθμου είναι ότι όταν το $\lambda(t)$ είναι αρκετά μικρότερο του λ , ο αλγόριθμος απορρίπτει πολλούς από τους τυχαίους χρόνους που παράγονται από την βοηθητική ομογενή διαδικασία.

Μια τεχνική για να παρακάμψουμε αυτό το μειονέκτημα είναι να διαμερίσουμε κατάλληλα το διάστημα $[0, t_0]$ και να προσομοιώσουμε την μη ομογενή διαδικασία Poisson ξεχωριστά στην κάθε διαμέριση. Στο τέλος μπορούμε να ενώσουμε όλα τα σύνολα με τους χρόνους από τα ξεχωριστά διαστήματα σε ένα σύνολο το οποίο μπορεί να θεωρηθεί ότι περιέχει τους χρόνους συμβάντων μιας διαδικασίας Poisson στο $[0, t_0]$. Αυτό έχουμε το δικαίωμα να το κάνουμε λόγω της ιδιότητας της διαδικασίας Poisson ότι έχει ανεξάρτητες προσαυξήσεις και το τι συμβαίνει σε ένα διάστημα χρόνου δεν επηρεάζει κάποιο άλλο. Έτσι για παράδειγμα στο παράδειγμα μας αν είχαμε μια διαμέριση στο διάστημα $[0, 0.5]$ το $\lambda_1 = \sup\{\lambda(t) : t \in [0, 0.5]\} = 0.5 < 5$ και η πιθανότητα

αποδοχής π.χ. του 0.1 θα ήταν $p = \frac{0.1}{0.5} = \frac{1}{5}$ δεκαπλάσια δηλαδή από την αντίστοιχη του προηγούμενου αλγόριθμου που έχει πιθανότητα αποδοχής $p = \frac{0.1}{5} = \frac{1}{50}$.

3.2.2.2 Μέθοδος Παραγωγής Ενδιάμεσων Χρόνων

Όπως και στην ομογενή περίπτωση η δεύτερη μέθοδος προσομοίωσης μη ομογενούς διαδικασίας Poisson βασίζεται στην παραγωγή των ενδιάμεσων χρόνων $T_1, T_2, T_3, \dots, T_i$ όπως αυτοί ορίστηκαν στην παράγραφο 3.2.1. Στην μη ομογενή περίπτωση οι ενδιάμεσοι χρόνοι T_i δεν είναι ανεξάρτητες τυχαίες μεταβλητές με κοινή κατανομή την εκθετική. Μπορούμε όμως γνωρίζοντας τους πρώτους T_i να παράγουμε τον αμέσως

επόμενο T_{i+1} . Έστω $S(i) = \sum_{k=1}^i T_k$, τότε

$$\begin{aligned} F_{S(i)}(t) &= P[T_{i+1} \leq t | S(i) = s] = P[X(t+s) - X(s) \geq 1] \\ &= 1 - P[X(t+s) - X(s) = 0] = 1 - e^{-\int_s^{s+t} \lambda(t) dt} \frac{\left(\int_s^{s+t} \lambda(t) dt \right)^0}{0!} = 1 - e^{-\int_s^{s+t} \lambda(t) dt}. \end{aligned}$$

Έτσι $T_i \sim F_{S(i-1)}$ και η παραγωγή των T_i γίνεται αναδρομικά, με την μέθοδο της αντιστροφής. Έτσι καταλήγουμε στον παρακάτω αλγόριθμο:

- 1) Θέτω $t=0$, $i=1$, $S(0)=0$ και $X=0$.
- 2) Παράγω $T \sim F_{S(i-1)}$.
- 3) Θέτω $t=t+T$ και αν $t > t_0$ σταματώ αλλιώς προχωρώ στο βήμα 4.
- 4) Θέτω $X=X+1$, $i=i+1$ και $S(i)=t$.
- 5) Επιστρέφω στο βήμα 2.

Για την προσομοίωση του παραδείγματος της προηγούμενης παραγράφου έχουμε:

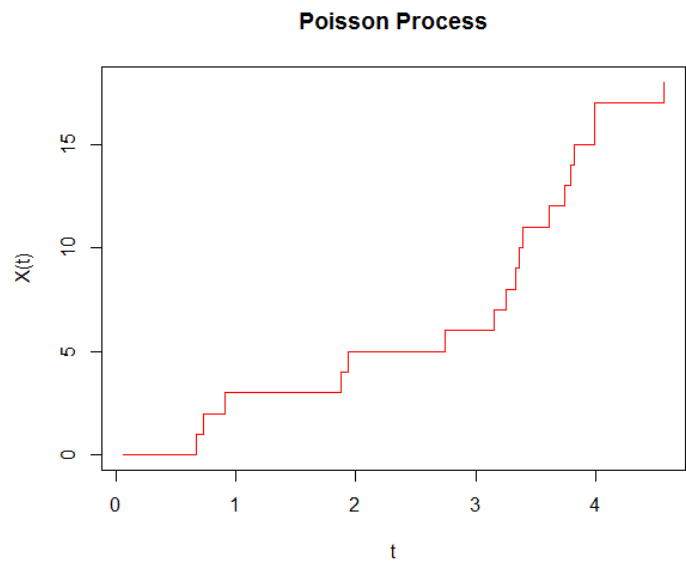
$\lambda(t) = t$, $F_{S(i)}(t) = 1 - \int_{s(i)}^{s(i)+t} t dt = 1 - e^{-\frac{1}{2}((s(i)+t)^2 - t^2)}$ και η παραγωγή των $F_{S(i)}$ μπορεί να γίνει εύκολα με την μέθοδο της αντιστροφής $F_{S(i)}^{-1}(U) = \sqrt{S(i)^2 - 2 \ln(1-U)} - S(i)$.

Ο αντίστοιχος κώδικας στην R είναι:

```
t<-0
X<-0
t0<-5
T<-(t^2-2*log(runif(1)))^0.5
t<-t+T
s<-{}
while (t<t0){
  T<-(t^2-2*log(runif(1)))^0.5-t
  s<-append(s,t)
  X<-X+1
  t<-t+T
}
X
s
x<-c(0:(length(s)-1))
plot(s,x, col="red",type='s',main="Poisson Process",cex=0.8,xlab="t",ylab="X(t)")
```

που επιστρέφει

$X = 19$ και $s = [0.06420357 \ 0.66931026 \ 0.72752961 \ 0.91386415 \ 1.88024919 \ 1.93286024 \ 2.74812151 \ 3.15430517 \ 3.25248345 \ 3.33431951 \ 3.36787565 \ 3.39021559 \ 3.60928080 \ 3.74734084 \ 3.79353765 \ 3.82128117 \ 3.98864286 \ 3.99060293 \ 4.57167038]$ όπου X είναι ο τελικός αριθμός των συμβάντων και s οι διαδοχικοί χρόνοι των συμβάντων. Αναπαριστώντας γραφικά την ανέλιξη Poisson έχουμε:



Σχήμα 25

3.3 Προσομοίωση Κίνησης Brown

Στην περίπτωση της ανέλιξης Poisson σε κάθε απειροστό διάστημα χρόνου η $X(t)$ παραμένει σταθερή ή αυξάνεται κατά μια μονάδα ανεξάρτητα από το παρελθόν. Το 1827 ο άγγλος βοτανολόγος Brown παρατήρησε ότι όταν ένα μόριο βυθιστεί σε ένα υγρό ή αέριο κινείται άτακτα. Το 1905 ο Einstein απέδειξε ότι το παραπάνω φαινόμενο μπορεί να εξηγηθεί υποθέτοντας ότι το μόριο βομβαρδίζεται συνεχώς από τα μόρια του μέσου (υγρού ή αερίου) προκαλώντας του μετατοπίσεις. Έτσι η μετατόπιση του σε ένα χρονικό διάστημα μπορεί να θεωρηθεί ότι κατανέμεται κανονικά, ως άθροισμα πολλών μετατοπίσεων, σύμφωνα με το Κεντρικό Οριακό Θεώρημα. Αυτή η κίνηση ονομάστηκε κίνηση Brown. Η $X(t)$ αλλάζει απειροστή τιμή σε απειροστό διάστημα χρόνου. Για αυτό τον λόγο παρουσιάζει ιδιαίζουσες μορφές κίνησης.

Μια στοχαστική ανέλιξη $\{X(t) : t \geq 0\}$ ονομάζεται κίνηση Brown με παραμέτρους μ (drift parameter) και σ^2 (volatility) αν έχει τις ακόλουθες ιδιότητες:

- 1) $X(0) = 0$,
- 2) Έχει ανεξάρτητες προσαυξήσεις,
- 3) Για κάθε $t, y > 0$ η $X(t+y) - X(y) \sim N(t\mu, t\sigma^2)$.

Προφανώς μια τέτοια ανέλιξη δεν μπορεί να προσομοιωθεί με ακρίβεια διότι σε ένα πεπερασμένο χρονικό διάστημα λαμβάνει άπειρες διαφορετικές τιμές. Θα αρκεστούμε στο να προσομοιώσουμε τιμές σε κάποια χρονικά διαστήματα πολύ μικρά. Για να προσομοιώσουμε λοιπόν την κίνηση Brown στο διάστημα $t \in [0, a]$ θα διαμερίσουμε το διάστημα μας σε n διαστήματα ως εξής:

$$t_0 = 0, t_1 = \Delta, t_2 = 2\Delta, \dots, t_i = i\Delta, \dots, t_n = n\Delta, \text{, παίρνοντας το } \Delta \text{ πολύ μικρό.}$$

Με αυτή την διαμέριση παρατηρούμε ότι:

$$X(t_0) = X(0) = 0,$$

$$X(t_1) = X(t_0) + (X(t_1) - X(t_0)), \text{ όπου } (X(t_1) - X(t_0)) \sim N((t_1 - t_0)\mu, (t_1 - t_0)\sigma^2),$$

$$X(t_2) = X(t_1) + (X(t_2) - X(t_1)), \text{ όπου } (X(t_2) - X(t_1)) \sim N((t_2 - t_1)\mu, (t_2 - t_1)\sigma^2),$$

⋮

$$X(t_{i+1}) = X(t_i) + (X(t_{i+1}) - X(t_i)), \text{ όπου } (X(t_{i+1}) - X(t_i)) \sim N((t_{i+1} - t_i)\mu, (t_{i+1} - t_i)\sigma^2).$$

Έτσι αναδρομικά ξεκινώντας από $X(0) = 0$, παράγουμε $X(t_1), X(t_2), \dots, X(t_i), \dots, X(t_n)$. Ο σχετικός αλγόριθμος έχει ως εξής:

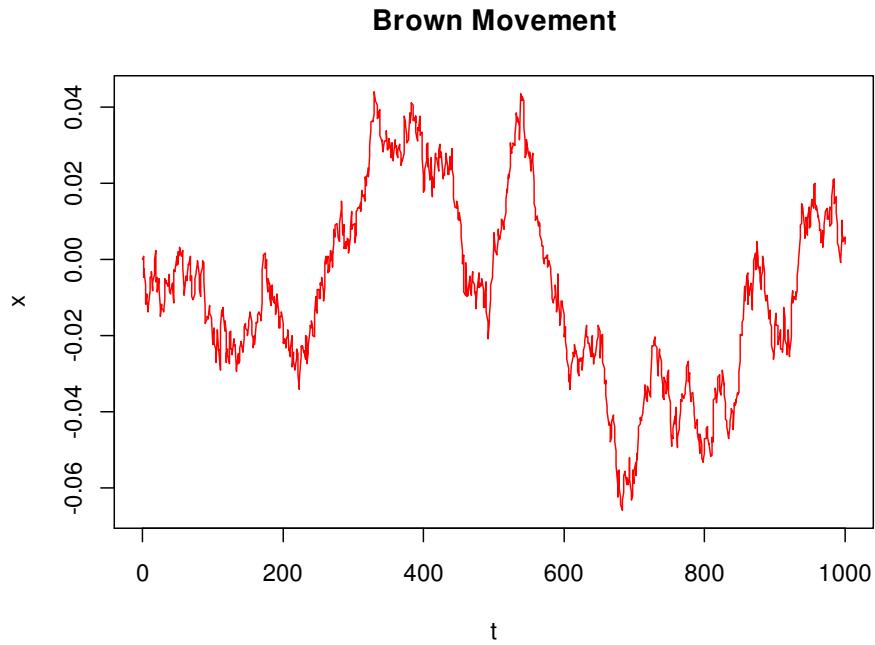
- 1) Θέτω $i = 1$ και $X_0 = 0$.
- 2) Παράγω $Z \sim N(0,1)$ με κάποια από τις μεθόδους που αναλύσαμε στο προηγούμενο κεφάλαιο.
- 3) Θέτω $X_i = X_{i-1} + \mu(t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}}Z$.
- 4) Θέτω $i = i + 1$ και αν $i \leq n$ επιστρέφω στο βήμα 2.

Ο αντίστοιχος κώδικας στην R που προσομοιώνει μια κίνηση Brown $\{X(t) : t > 0\}$ με παραμέτρους $\mu = 0.1, \sigma = 0.1$ στο διάστημα $t \in [0,1]$ χωρίζοντας το σε $n = 1000$ διαστήματα, δηλαδή $\Delta = \frac{1}{n} = 0.001$ είναι ο ακόλουθος:

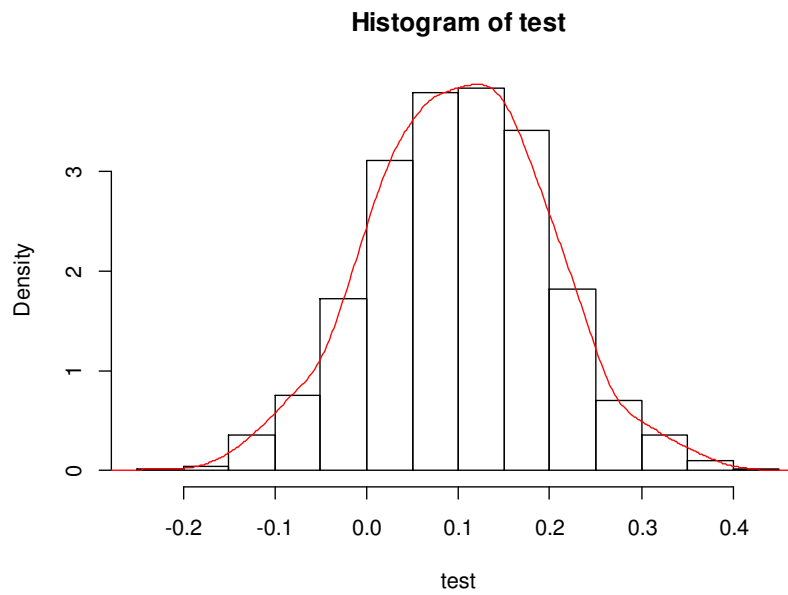
```
test<-{}
for (j in 1:1000){
n<-1000
t<-c(0:n)
x<-rep(0,n+1)
m=0.1
s=0.1
d=1/n
for (i in 2:(n+1)){
  z=rnorm(1)
  x[i]=x[i-1]+m*d+s*d^0.5*z
}
test<-append(test,x[1001])
plot(t,x, type = "l", col = "red", lwd=1, main="Brown Movement")
}
hist(test,freq=FALSE,nclass=10)
lines(density(test))
mean(test)
var(test)
```

Γραφικά η κίνηση Brown φαίνεται στο Σχήμα 26 στην επόμενη σελίδα. Ο πιο πάνω κώδικας επαναλαμβάνει τον πιο πάνω αλγόριθμο 1000 φορές και αποθηκεύει τις $X(t_{1000}) = X(1)$ σε μια μεταβλητή test. Αναμένουμε οι τιμές που θα πάρουμε να κατανέμονται σύμφωνα με την κανονική κατανομή με μέσες τιμή $\mu = 0.1$ και

$\sigma^2 = 0.1^2 = 0.01$. Πράγματι ο κώδικας επιστρέφει μέση τιμή 0.101 και διασπορά 0.0095. Το ιστόγραμμα της μεταβλητής test φαίνεται στο Σχήμα 27.



Σχήμα 26



Σχήμα 27

3.4 Προσομοίωση Τυχαίου Περιπάτου με Απορροφητικά Φράγματα

Ας υποθέσουμε ότι υπάρχουν δύο παίκτες A και B που παίζουν ένα τυχερό παιχνίδι με χρήματα. Ο A έχει α ευρώ και ο B έχει β ευρώ. Ο παίκτης A κερδίζει ένα ευρώ με πιθανότητα p ενώ ο παίκτης B με πιθανότητα $q=1-p$. Το παιχνίδι σταματά όταν ένας από τους δύο χάσει το αρχικό του ποσό.

Το πιο πάνω παράδειγμα είναι ένας απλός τυχαίος περίπατος $\{X(t):t > 0\}$ με απορροφητικά φράγματα τα $-\alpha$ και β . Δηλαδή $X_n = -\alpha$ ή $X_n = \beta$ για κάποιο n . Αποδεικνύεται με την βοήθεια του κεντρικού οριακού θεωρήματος ότι ο τυχαίος περίπατος θα σταματήσει σε ένα από τα δύο απορροφητικά φράγματα με πιθανότητα την μονάδα. Στην προσομοίωση που θα ακολουθήσει θα υπολογίσουμε και την πιθανότητα απορρόφησης a_0 και b_0 . Ξεκινώντας με $X(0) = 0$, έστω η τ.μ.:

$$I = \begin{cases} 0 & \text{εάν η απορρόφηση γίνει στο } -\alpha. \\ 1 & \text{εάν η απορρόφηση γίνει στο } \beta. \end{cases}$$

Αποδεικνύεται ότι για $\lambda = \frac{q}{p}$:

$$a_0 = P[I = 0 | X(0) = 0] = \begin{cases} \frac{b}{a+b} & , \lambda = 1 \\ \frac{\lambda^b - 1}{\lambda^b - \lambda^{-\alpha}} & , \lambda \neq 1 \end{cases} \quad \text{και} \quad b_0 = P[I = 1 | X(0) = 0] = \begin{cases} \frac{a}{a+b} & , \lambda = 1 \\ \frac{1 - \lambda^{-\alpha}}{\lambda^b - \lambda^{-\alpha}} & , \lambda \neq 1. \end{cases}$$

Παρατηρούμε ότι $a_0 + b_0 = 1$. Δηλαδή προκύπτει, όπως προαναφέραμε, ότι ο απλός τυχαίος περίπατος σταματά με πιθανότητα την 1. Στην προσομοίωση που θα κάνουμε θα επαληθεύσουμε και αυτή την συνθήκη.

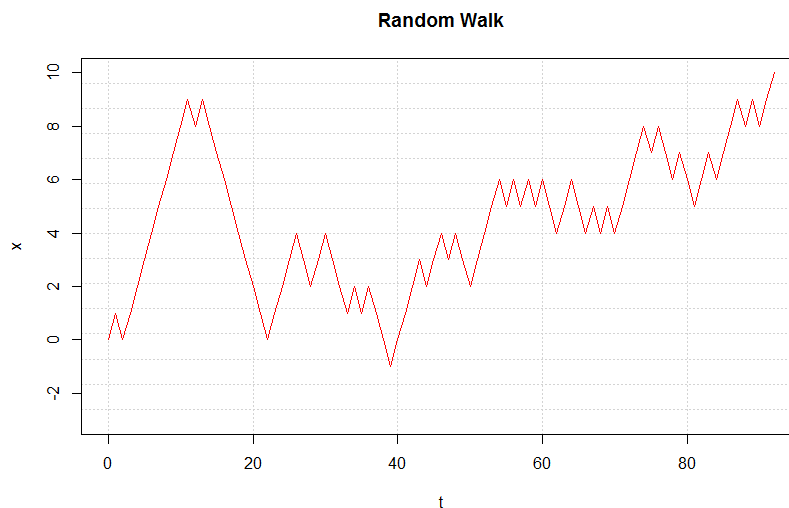
Ο σχετικός αλγόριθμος που προσομοιώνει την κίνηση έχει ως εξής:

- 1) Θέτω $i=1$ και $X_0 = 0$.
- 2) Παράγω $U \sim U(0,1)$.
- 3) Αν $p < U$ θέτω $X[i] = X[i-1] + 1$, αλλιώς θέτω $X[i] = X[i-1] - 1$.
- 4) Αν $X(i) = \beta$ θέτω $I = 1$ και σταματώ.
- 5) Αν $X(i) = -\alpha$ θέτω $I = 0$ και σταματώ.
- 6) Επιστρέφω στο βήμα 2.

Ο αντίστοιχος κώδικας στην R που προσομοιώνει τον εν λόγω περίπατο με $\alpha = 3$, $\beta = 10$ και $p = 0.6$ είναι ο ακόλουθος.

```
a<-3
b<-10
p<-0.6
x<-{0}
i<-1
while (x[i]!=-a && x[i]!=b) {
  u=runif(1)
  if (u<p) {
    x<-append(x,x[i]+1)
  } else {
    x<-append(x,x[i]-1)
  }
  i<-i+1
}
if (x[length(x)]==b) {
  l<-1
} else {
  l<-0
}
t<-0:(length(x)-1)
plot(t,x,type='l',col='red',ylim = c(-a,b), panel.first = grid(ny=15), main = "Random Walk")
```

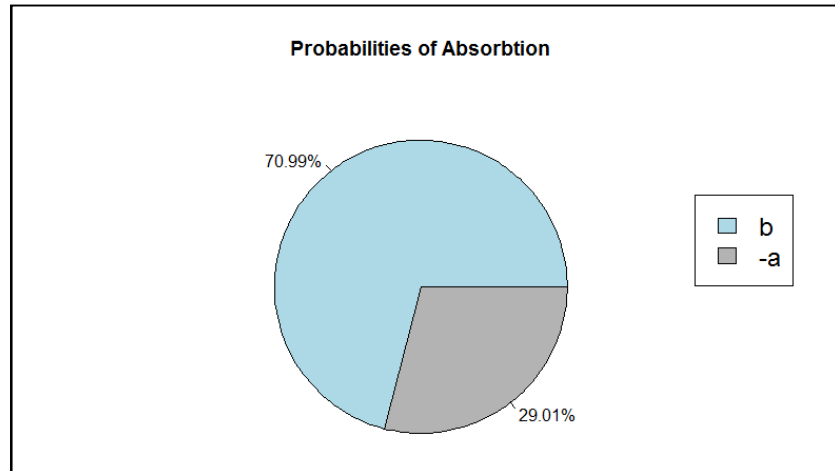
Ο τυχαίος περίπατος που παράγεται φαίνεται στο Σχήμα 28:



Σχήμα 28

Επαναλαμβάνοντας τον αλγόριθμο 1000 φορές και αποθηκεύοντας σε μια λίστα που έγινε η απορρόφηση κάθε φορά, βλέπουμε στο Σχήμα 29 τα ποσοστά απορρόφησης. Ο σχετικός αλγόριθμος και τα αποτελέσματα φαίνονται παρακάτω:

```
Y<-{}
T<-{}
for(j in 1:1000){
a<-3
b<-10
p<-0.6
x<-{0}
i<-1
while (x[i]!=-a && x[i]!=b) {
  u=runif(1)
  if (u<p) {
    x<-append(x,x[i]+1)
  } else {
    x<-append(x,x[i]-1)
  }
  i<-i+1
}
if (x[length(x)]==b) {
  l<-1
} else {
  l<-0
}
Y<-append(Y,l)
T<-append(T,length(x)-1)
t<-0:(length(x)-1)
plot(t,x,type='l',col='red',ylim = c(-a,b), panel.first = grid(ny=15), main = "Random Walk")
}
k<-c(sum(Y),1000-sum(Y))
klabels<-round(k/sum(k)*100 ,1 )
klabels<-paste(klabels, "%", sep="")
pie(k,labels=klabels,main="Probabilities of Absorbtion",col=c("lightblue","grey70"))
legend(1.5, 0.5, c("b","-a"), cex=1.5,fill=c("lightblue","grey70"))
```



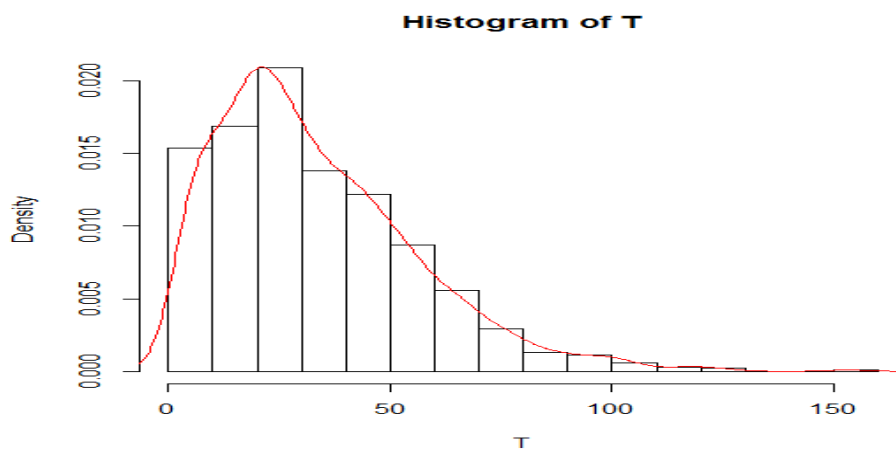
Σχήμα 29

Οι θεωρητικές τιμές πιθανότητας απορρόφησης για τα δεδομένα μας υπολογίζονται:

$$a_0 = P[I = 0 | X(0) = 0] = \frac{\lambda^b - 1}{\lambda^b - \lambda^{-\alpha}} = \frac{\left(\frac{2}{3}\right)^{10} - 1}{\left(\frac{2}{3}\right)^{10} - \left(\frac{2}{3}\right)^{-3}} = 0.29 \text{ και } b_0 = 1 - a_0 = 0.71,$$

και συμπίπτουν με αυτές που υπολογίστηκαν από την προσομοίωση.

Στον κώδικα συμπεριλάβαμε και μια μεταβλητή που αποθηκεύει τους χρόνους που χρειάστηκε για να γίνει η απορρόφηση. Με αυτό τον τρόπο μπορούμε να εκτιμήσουμε και την κατανομή του χρόνου απορρόφησης που είναι πολύ δύσκολο να γίνει αναλυτικά. Συγκεκριμένα μας δίνει μέσο 31.83, διασπορά 507.62, ελάχιστο χρόνο 3 και μέγιστο 211. Η εκτίμηση της κατανομής φαίνεται στο Σχήμα 30.



Σχήμα 30

ΚΕΦΑΛΑΙΟ 4

ΜΕΘΟΔΟΙ MONTE CARLO ΜΕ ΤΗΝ ΧΡΗΣΗ ΜΑΡΚΟΒΙΑΝΩΝ ΑΛΥΣΙΔΩΝ (MCMC)

4.1 Εισαγωγή

Το 1954 οι Nikolas Metropolis, Arianna Rosenbluth, Augusta Teller και Edward Teller επιχείρησαν να προσομοιώσουν δείγμα από μια κατανομή χρησιμοποιώντας μια κατάλληλη μαρκοβιανή αλυσίδα της οποίας η κατανομή ισορροπίας ισούταν με την κατανομή που θέλαμε να πάρουμε δείγμα. Από την εργασία τους αυτή προέκυψε ο αλγόριθμος Metropolis που είναι ιστορικά η πρώτη εμφάνιση των Markov Chain Monte Carlo (MCMC) αλγορίθμων. Έτσι λοιπόν δημιουργήθηκε μια νέα κλάση αλγορίθμων που με τη βοήθειά της παίρνουμε δείγμα από κάποιες κατανομές, στην οποία έχουμε εξάρτηση ενός βήματος. Δηλαδή η πιθανότητα με βάση την οποία ο αλγόριθμος προχωρά από σημείο σε σημείο εξαρτάται μόνο από την τρέχουσα θέση και όχι από το παρελθόν.

Εξέλιξη του αλγόριθμου Metropolis είναι ο Metropolis – Hastings αλγόριθμος (1973), από τον Hasting. Αυτός είναι ίσως ο πιο βασικός αλγόριθμος, γιατί είναι απλός και εύχρηστος, και οι περισσότεροι αλγόριθμοι που προέκυψαν μετέπειτα είναι ειδικές περιπτώσεις του. Ένας ακόμη πολύ χρήσιμος MCMC αλγόριθμος, ο δειγματολήπτης του Gibbs, προέκυψε το 1984 από μια εργασία των Geman, S. και Geman, D.

Σήμερα οι MCMC αλγόριθμοι χρησιμοποιούνται σε ένα ευρύ επιστημονικό φάσμα εκτός των μαθηματικών, όπως στη στατιστική φυσική, τη βιολογία, τη γενετική, τη χημεία, την πληροφορική, τα χρηματοοικονομικά και αλλού. Στη στατιστική και στις πιθανότητες συγκεκριμένα, χρησιμοποιούνται για εύρεση βασικών χαρακτηριστικών

κάποιων μοντέλων, όπως η μέση τιμή, η διασπορά, και άλλων παραμέτρων. Ιδιαίτερα χρήσιμη είναι η βοήθειά τους στον υπολογισμό ολοκληρωμάτων, κυρίως πολλαπλών και όταν η υπό ολοκλήρωση ποσότητα έχει σύνθετη μορφή. Ακόμα δουλεύουν αποτελεσματικά σε περίπλοκης δομής πολυδιάστατα προβλήματα που οι συνηθισμένες αριθμητικές μέθοδοι αδυνατούν. Σε αυτό το κεφάλαιο θα αναλύσουμε τους πιο βασικούς Markov Chain Monte Carlo αλγόριθμους και θα δώσουμε κάποια παραδείγματα.

4.2 Ολοκλήρωση Monte Carlo

Πολλά ολοκληρώματα είναι δύσκολο ή ακόμη και αδύνατο να υπολογιστούν με αναλυτική μέθοδο. Το πρόβλημα γίνεται ακόμη δυσκολότερο όταν η συνάρτηση που θέλουμε να ολοκληρώσουμε είναι ορισμένη σε χώρους πολλών διαστάσεων. Θα ξεκινήσουμε με την Monte Carlo ολοκλήρωση σε μια διάσταση.

Έστω $f(x): R \supseteq A \rightarrow R$ και θέλουμε να υπολογίσουμε το ολοκλήρωμα $\int_a^b f(x)dx$, $[a,b] \subseteq A$. Θεωρούμε ακόμη την $p(x)$ η οποία είναι μια οποιαδήποτε συνάρτηση πυκνότητας πιθανότητας με $p(x) = 0, \forall x \notin [a,b]$. Τότε το ολοκλήρωμα μπορεί να γραφτεί:

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{p(x)} p(x)dx = \int_a^b h(x)p(x)dx.$$

Έτσι το ζητούμενο ολοκλήρωμα δίνεται σαν η μέση τιμή της τ.μ $h(Z) = \frac{f(Z)}{p(Z)}$, όπου Z ορίζεται στο διάστημα $[a,b]$ και έχει σ.π.π. την $p(x)$. Δηλαδή

$$I = \int_a^b f(x)dx = \int_a^b h(x)p(x)dx = E(h).$$

Έστω $h_i = h(Z_i) = \frac{f(Z_i)}{p(Z_i)}$, $i = 1, 2, \dots, N$ και Z_i ανεξάρτητες και ισόνομες τ.μ. με σ.π.π. την $p(x)$. Επειδή οι Z_i είναι ανεξάρτητες και ισόνομες τ.μ. και οι h_i θα είναι ανεξάρτητες και ισόνομες και άρα:

$$P \left[N\mu - 3\sigma < \sum_{i=1}^N h_i < N\mu + 3\sigma \right] = P \left[NI - 3\sigma < \sum_{i=1}^N h_i < NI + 3\sigma \right] \approx 0.997,$$

$$P\left[\frac{1}{N}\left|\sum_{i=1}^N h_i - I\right| < 3\frac{\sigma}{N}\right] \approx 0.997.$$

Έτσι για μεγάλο N θα ισχύει :

$$I_N \equiv \frac{1}{N} \sum_{i=1}^N h_i \approx I,$$

με πιθανότητα το σφάλμα να είναι μικρότερο του $3\frac{\sigma}{N}$ να είναι σχεδόν 0.997.

Έστω για **παράδειγμα** ότι θέλουμε να υπολογίσουμε το ολοκλήρωμα $\int_0^1 e^{x^2} dx$, που είναι αδύνατο να υπολογιστεί αναλυτικά. Για $p(x)$ επιλέγω την ομοιόμορφη στο $[0,1]$. Τότε:

$$I = \int_0^1 e^{x^2} dx = \int_0^1 p(x) \frac{e^{x^2}}{p(x)} dx = \int_0^1 p(x) h(x) dx.$$

Ο αλγόριθμος που θα χρησιμοποιήσουμε είναι ο ακόλουθος:

- 1) Παράγω $U_1, U_2, \dots, U_N \sim U(0,1)$.
- 2) Για κάθε $U_i, i = 1, 2, \dots, N$ υπολογίζω το $H_i = e^{U_i^2}$.
- 3) Η εκτίμηση του ολοκληρώματος \hat{I} είναι ο δειγματικός μέσος.

Ο αντίστοιχος κώδικας στην R είναι:

```
u<-runif(10000)
h<-exp(u^2)
I<-mean(h)
```

Ο κώδικας επιστρέφει $\hat{I} = 1.4644$. Μια σύγκριση με το αποτέλεσμα που επιστρέφει η Mathematica (1.4626) δείχνει πόσο ακριβής είναι η ολοκλήρωση Monte Carlo.

Ακολούθως θα δούμε την ολοκλήρωση Monte Carlo με εμπλοκή Μαρκοβιανών αλυσίδων. Έστω η τ.μ. X με συνάρτηση κατανομής $\pi(x)$ και συνάρτηση πυκνότητας πιθανότητας $f(x)$.

Έστω επίσης η $g(x)$ για την οποία ισχύει $E(|g(x)|) < \infty$ και θέλουμε να υπολογίσουμε το ολοκλήρωμα:

$$\mu = \int g(x)f(x)dx = E(g(x)).$$

Ας υποθέσουμε ότι παίρνουμε τυχαίο δείγμα $\{X_1, X_2, \dots, X_n\}$ από την κατανομή ισορροπίας π. Σύμφωνα με τον Ισχυρό Νόμο των Μεγάλων Αριθμών όταν $n \rightarrow \infty$ θα έχουμε:

$$\widehat{\mu}_{MC} = \frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow{\sigma.\beta} \mu = \int g(x)f(x)dx \text{ και}$$

Αυτή η μέθοδος είναι πολύ χρήσιμη στις πολλαπλές ολοκληρώσεις. Έστω ότι θέλουμε να υπολογίσουμε το ολοκλήρωμα:

$$\mu = \int_{[0,1]^k} g(\mathbf{x})d\mathbf{x} = \int_0^1 \int_0^1 \dots \int_0^1 g(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k.$$

Τότε αν $U_1, U_2, \dots, U_k \sim U(0,1)$ ανεξάρτητες ισόνομες έχουμε:

$$\mu = E(g(\mathbf{U})) = E(g(U_1, U_2, \dots, U_k)) \text{ και}$$

$$\frac{1}{n} \sum_{i=1}^n g(\mathbf{U}_i) \xrightarrow{\sigma.\beta} \mu.$$

Η πιο πάνω μέθοδος προφανώς ισχύει για οποιαδήποτε άκρα με κατάλληλο μετασχηματισμό.

Για παράδειγμα θα υπολογίσουμε τον αριθμό π. Ξέρουμε ότι η εξίσωση κύκλου ακτίνας 1 είναι $x^2 + y^2 = 1$ και το εμβαδό του τεταρτοκύκλιου $\frac{\pi r^2}{4} = \frac{\pi}{4}$. Επίσης το εμβαδό του εκφράζεται από το ολοκλήρωμα:

$$I_1 = \int_0^1 \int_0^1 I_{(x^2+y^2 \leq 1)} dx dy, \text{ όπου } I_{(x^2+y^2 \leq 1)} = \begin{cases} 1 & , x^2 + y^2 \leq 1 \\ 0 & , x^2 + y^2 > 1. \end{cases}$$

Προφανώς το $\pi = 4I_1$.

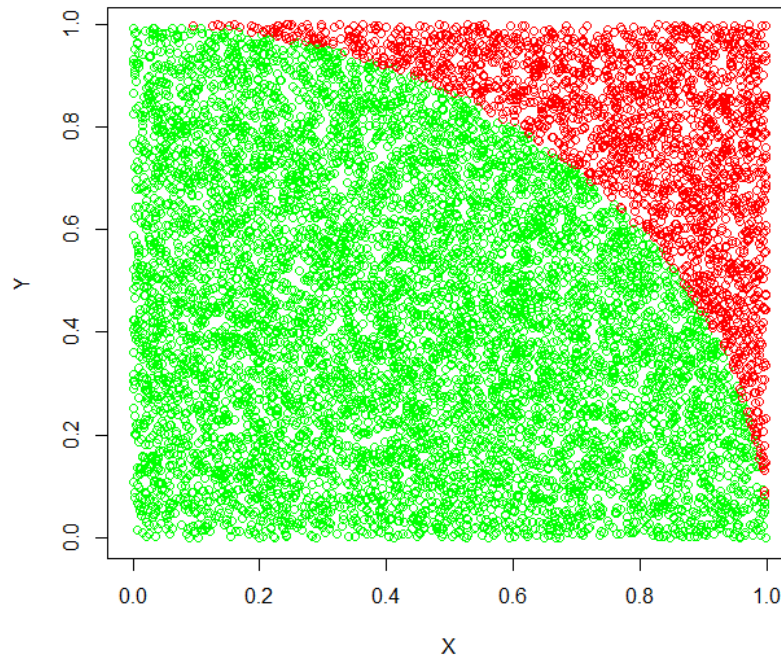
Ο αλγόριθμος που θα χρησιμοποιήσουμε είναι ο ακόλουθος:

- 1) Παράγω $X \sim U(0,1)$ και $Y \sim U(0,1)$.
- 2) Αν $X^2 + Y^2 \leq 1$ τότε θέτω $R[i] = 1$ αλλιώς θέτω $R[i] = 0$, και $i = i + 1$.
- 3) Αν $i < N$ τότε επιστρέφω στο βήμα 1 αλλιώς πάω στο βήμα 4.
- 4) Θέτω $\pi = \overline{4R}$.

Ο αντίστοιχος κώδικας στην R είναι:

```
n<-10000
R<-{}
X<-{}
Y<-{}
xrej<-{}
yrej<-{}
for (i in 1:n) {
  x<-runif(1, min=0, max=1)
  y<-runif(1, min=0, max=1)
  if ((x^2+y^2)<=1){
    R<-append(R,(1))
    X<-append(X,x)
    Y<-append(Y,y)
  } else {
    R<-append(R,0)
    xrej<-append(xrej,x)
    yrej<-append(yrej,y)
  }
}
sample=4*R
p<-mean(sample)
p;
plot(X,Y,col="green")
points(xrej,yrej,col="red")
```

Που μας επιστρέφει $\hat{\pi} = 3.131$. Στο Σχήμα 31 φαίνονται τα σημεία που προσομοιώθηκαν από τις ομοιόμορφες κατανομές. Με πράσινο είναι τα σημεία που δεχόμαστε ενώ με κόκκινο τα σημεία που απορρίπτουμε.



Σχήμα 31

Σε όσα αναφέραμε παραπάνω το δείγμα μας αποτελείτο από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Αυτό όμως δεν είναι συνήθως εφικτό, οπότε εδώ υπεισέρχονται οι Μαρκοβιανές αλυσίδες. Αρχικά δημιουργούμε έναν Markov Chain αλγόριθμο από κατάλληλη εργοδική Μαρκοβιανή αλυσίδα. Αφού τον τρέξουμε για αρκετό χρόνο, ώστε να περάσει η **burn in** περίοδος, δηλαδή η αρχική περίοδος κατά την οποία είμαστε ακόμα μακριά από την επιθυμητή κατανομή, παίρνουμε τις n τελευταίες τιμές που δημιουργούνται από τον αλγόριθμο. Τότε αυτές είναι ένα εξαρτημένο δείγμα από την ζητούμενη κατανομή. Παρακάτω θα παρουσιάσουμε τους δύο πιο σημαντικούς MCMC αλγόριθμους.

4.3 Αλγόριθμος Metropolis-Hastings

Ο στόχος του αλγόριθμου είναι η κατασκευή Μαρκοβιανής αλυσίδας με κατανομή ισορροπίας $p(\mathbf{x}): R^n \rightarrow A \subseteq R^n$ και η εξαγωγή δείγματος από αυτήν. Από το δείγμα που θα πάρουμε μπορούμε να εκτιμήσουμε ικανοποιητικά κάποια μέτρα της κατανομής ισορροπίας, όπως η μέση τιμή και η διασπορά (Εργοδικό Θεώρημα).

Έστω ότι θέλουμε να παράγουμε δείγμα από μια κατανομή στόχο $p(\mathbf{x}): R^n \rightarrow A \subseteq R^n$ της οποίας δεν γνωρίζουμε την σταθερά κανονικοποίησης C . Γνωρίζουμε δηλαδή την συνάρτηση $\pi(\mathbf{x}) = Cp(\mathbf{x})$ χωρίς να γνωρίζουμε το C .

Αρχικά επιλέγουμε μια κατανομή εισήγησης $g(\mathbf{x}|y)$ που δίνει την πιθανότητα μετάβασης από το y στο \mathbf{x} . Ο αλγόριθμος ξεκινά παίρνοντας τυχαία ένα σημείο \mathbf{x}_0 . Έστω ότι βρισκόμαστε στην i -οστή επανάληψη και έχει επιλεγεί το σημείο \mathbf{x}_i . Ο αλγόριθμος ακολουθεί ως εξής:

- 1) Επιλέγουμε ένα υποψήφιο \mathbf{x}^* από την κατανομή $g_i(\mathbf{x}^*|\mathbf{x}_i)$.
- 2) Υπολογίζουμε την πιθανότητα αποδοχής $\alpha_i = \min \left\{ \frac{\pi(\mathbf{x}^*) \cdot g_i(\mathbf{x}_i|\mathbf{x}^*)}{\pi(\mathbf{x}_i) \cdot g_i(\mathbf{x}^*|\mathbf{x}_i)}, 1 \right\}$ (όπου $g(\mathbf{x}_i|\mathbf{x}_j)$ είναι η πιθανότητα μετάβασης από το \mathbf{x}_j στο \mathbf{x}_i).
- 3) Παίρνω $U \sim U(0,1)$ και αν $U \leq \alpha_i$ τότε $\mathbf{x}_{i+1} = \mathbf{x}^*$ αλλιώς $\mathbf{x}_{i+1} = \mathbf{x}_i$.
- 4) Επιστρέφουμε στο 1.

Ο αλγόριθμος τερματίζει όταν συμπληρωθεί ο αριθμός των επαναλήψεων που έχουμε ορίσει. Ο πυρήνας μετάβασης της αλυσίδας που έχει οριστεί από την πιθανότητα αποδοχής εξασφαλίζει την αντιστρεψιμότητας της αλυσίδας:

$$\begin{aligned} \pi(\mathbf{x}_i)g(\mathbf{x}_{i+1}|\mathbf{x}_i) &= \pi(\mathbf{x}_i)g(\mathbf{x}_{i+1}|\mathbf{x}_i) \frac{\pi(\mathbf{x}_{i+1})g(\mathbf{x}_i|\mathbf{x}_{i+1})}{\pi(\mathbf{x}_{i+1})g(\mathbf{x}_i|\mathbf{x}_{i+1})} \\ &= \pi(\mathbf{x}_{i+1})g(\mathbf{x}_i|\mathbf{x}_{i+1}) \frac{\pi(\mathbf{x}_i)g(\mathbf{x}_{i+1}|\mathbf{x}_i)}{\pi(\mathbf{x}_{i+1})g(\mathbf{x}_i|\mathbf{x}_{i+1})} = \pi(\mathbf{x}_{i+1})g(\mathbf{x}_i|\mathbf{x}_{i+1}). \end{aligned}$$

Αν επιπλέον η αλυσίδα που δημιουργείται είναι αδιαχώριστη και απεριοδική, τότε η στάσιμη κατανομή της είναι η κατανομή στόχος.

Παράδειγμα

Θεωρούμε την τ.μ. $X \sim G(a,1)$, $a=4$ που έχει σ.π.π. $f(x) = \frac{x^{a-1}e^{-x}}{\Gamma(a)}$. Το γράφημα της συνάρτησης f φαίνεται στο Σχήμα 17 (παράγραφος 2.4.1.3).

Σαν κατανομή εισήγησης θα χρησιμοποιήσουμε μια συμμετρική κατανομή που παράγεται από τον δειγματολήπτη τυχαίου περιπάτου. (Random Walk Sampler).

Σύμφωνα με τον δειγματολήπτης τυχαίου περιπάτου η τ.μ. X^* για μια δεδομένη κατάσταση x δίνεται από τον τύπο $X^* = x_i + Z$, όπου η τ.μ. Z παράγεται συνήθως από μια συμμετρική κατανομή όπως για παράδειγμα την $N(0, \Sigma)$. Και επειδή η κατανομή εισήγησης είναι συμμετρική και η πιθανότητα αποδοχής που ορίσαμε πιο πάνω γίνεται:

$$\alpha_i = \min \left\{ \frac{\pi(x^*)}{\pi(x_i)}, 1 \right\}.$$

Στο παράδειγμά θα χρησιμοποιήσουμε $Z = N(0, \sigma^2) = N(0, 4)$. Ο αλγόριθμος μας διαμορφώνεται ως εξής:

- 1) Θέτω $x_0 = 0.1, s = 2, i = 1, alpha = 4$.
- 2) Παράγω έναν τυχαίο αριθμό $Z \sim N(0, 1)$ και θέτω $x^* = x_i + sZ$.
- 3) Υπολογίζω τα $\pi(x^*)$ και $\pi(x_i)$ όπου $\pi(x)$ η κατανομή στόχος δηλαδή η $f(x) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}$. (Επειδή το πεδίο ορισμού της κατανομής Γάμμα είναι $[0, \infty)$ αν στο βήμα 2 το υποψήφιο σημείο υπολογιστεί $x^* < 0$ θέτουμε $\pi(x^*) = 0$).
- 4) Υπολογίζω την πιθανότητα αποδοχής $\alpha_i = \min \left\{ \frac{\pi(x^*)}{\pi(x_i)}, 1 \right\}$.
- 5) Παράγω $U \sim U(0, 1)$ και αν $U < \alpha_i$ θέτω $x_{i+1} = x^*$ αλλιώς θέτω $x_{i+1} = x_i$.
- 6) Θέτω $i = i + 1$ και αν $i \leq N$ επιστρέφω στο βήμα 2, διαφορετικά σταματώ.

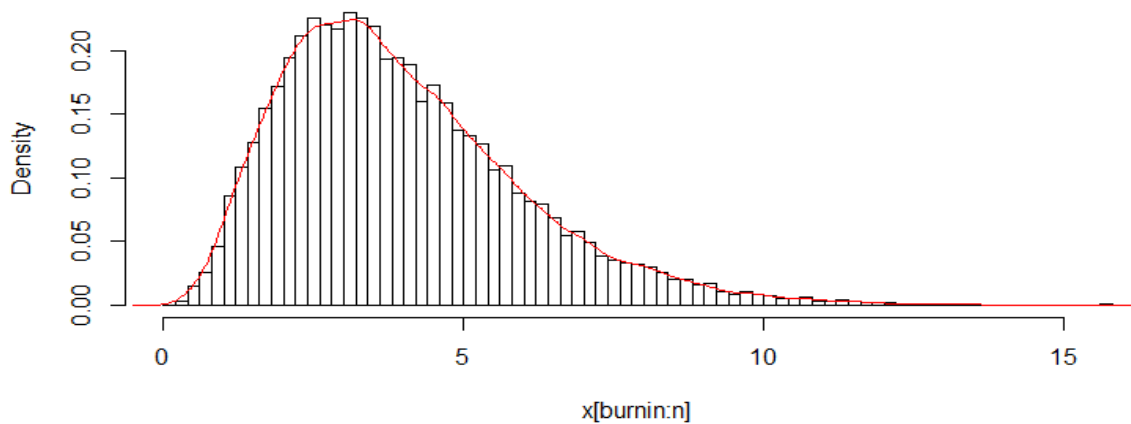
Στον κώδικα της R επιλέξαμε $N=40000$ και burn in περίοδο τις πρώτες 5000 τιμές. Σε αυτό το σημείο πρέπει να αναφέρουμε ότι η επιλογή του σ είναι αρκετά σημαντική. Αν το σ είναι πολύ μικρό τότε τα x_i που παράγονται θα είναι ισχυρά θετικά συσχετισμένα και άρα η διακύμανση της εκτίμησης που θέλουμε να κάνουμε (π.χ. ο μέσος μ της κατανομής στόχου) θα είναι πολύ μεγάλος. Σε αντίθεση αν το σ είναι πολύ μεγάλο θα οδηγήσει σε μια πολύ μικρή πιθανότητα αποδοχής και πολλά από τα σημεία που παράγονται θα απορρίπτονται που και πάλι θα οδηγήσει σε χαμηλή αποδοτικότητα του αλγόριθμου. Παρακάτω δίνεται ο κώδικας στην R για $\sigma = 2$, ένα ιστόγραμμα του δείγματος που παράγουμε, καθώς και τα τελευταία 5000 βήματα της μαρκοβιανής αλυσίδας για $\sigma = 2, \sigma = 0.1, \sigma = 10$.


```

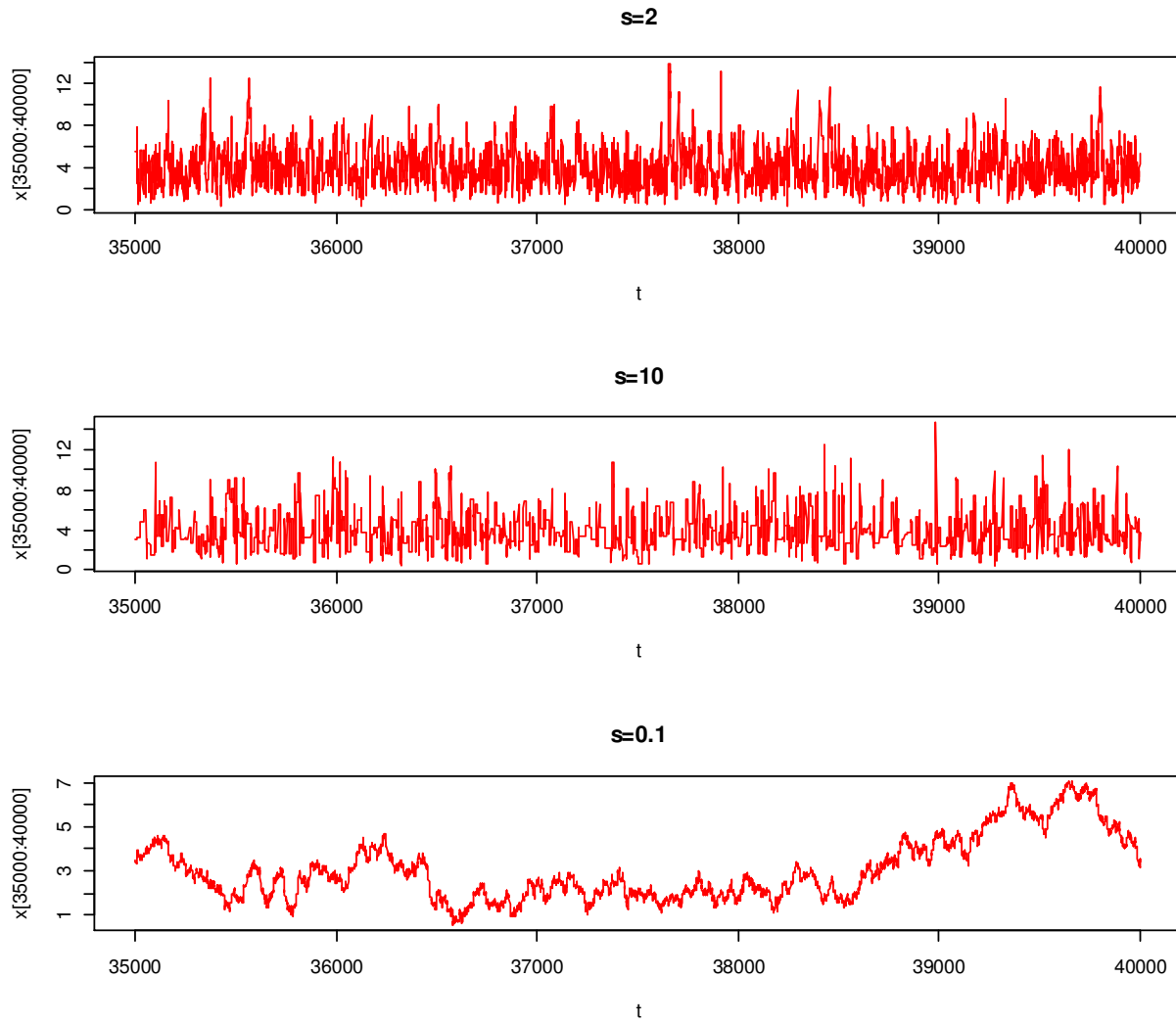
n=40000
burnin=5000
x<-{0.1}
s<-2
alpha<-4
for (i in 1:n) {
  y<-x[i]+s*rnorm(1,0,1)
  if (y<0){
    fy<-0
  }else {
    fy<-y^(alpha-1)*exp(-y)/gamma(alpha)
  }
  fx<-x[i]^(alpha-1)*exp(-x[i])/gamma(alpha)
  a<-min((fy)/(fx),1)
  u<-runif(1)
  if (u<a){
    x<-append(x,y)
  } else {
    x<-append(x,x[i])
  }
}
par(mfrow=c(2,1))
hist(x[burnin:n],freq=FALSE,nclass=75,main="Metropolis Hastings - s=2",)
lines(density(x[burnin:n]),col="red")
t<-c(35000:40000)
plot(t,x[35000:40000],type='l',col="red")
mean(x[burnin:n])
var(x[burnin:n])

```

Metropolis Hastings G(4,1)



Σχήμα 32



Σχήμα 33

Στο ιστόγραμμα του Σχήματος 32 φαίνεται ότι η $G(4,1)$ προσομοιώνεται αρκετά καλά με τον αλγόριθμο Metropolis-Hastings. Ο κώδικας μας επιστρέφει επίσης μέση τιμή 4.03 και διακύμανση 4.06, πολύ κοντά στις θεωρητικές τιμές της κατανομής. Στο Σχήμα 33 φαίνεται αυτό που αναφέραμε και πιο πάνω για τις διάφορες τιμές του σ . Για $\sigma = 0.1$ οι τιμές της αλυσίδας που δημιουργείται αλλάζουν πολύ συχνά ενώ για $\sigma = 10$ μένουν στάσιμες για πολλές επαναλήψεις. Για $\sigma = 2$ έχουμε μια ενδιάμεση κατάσταση και για αυτό επιλέξαμε αυτήν για την προσομοίωση.

4.4 Δειγματολήπτης Gibbs

Ο δειγματολήπτης Gibbs χρησιμοποιεί διαφορετική μεθοδολογία από αυτήν του Metropolis Hastings. Η ιδέα είναι να πάρουμε δείγμα $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ από την κατανομή στόχο $f(\mathbf{x})$, όπου $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, παίρνοντας δείγμα από τις δεσμευμένες κατανομές $f(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Ένα μεγάλο πλεονέκτημα του είναι ότι μπορούμε να πάρουμε δείγμα από μια κατανομή και να εκτιμήσουμε ικανοποιητικά κάποια μέτρα αυτής, όπως η μέση τιμή και η διασπορά, χωρίς να την γνωρίζουμε. Το πλεονέκτημα αυτό είναι ακόμα μεγαλύτερο σε περιπτώσεις που η μορφή της κατανομής είναι ιδιαίτερα σύνθετη και ο αναλυτικός υπολογισμός κάποιων μέτρων αδύνατος.

Ο αλγόριθμος ξεκινά όπως και στον Metropolis Hastings με την επιλογή κάποιου αρχικού $\mathbf{x}_0 = (x_1^0, x_2^0, \dots, x_n^0)$. Έστω ότι βρισκόμαστε στην i επανάληψη του αλγορίθμου και έχει επιλεγεί το σημείο $\mathbf{x}_{i-1} = (x_1^{i-1}, x_2^{i-1}, \dots, x_n^{i-1})$ στην $i-1$ επανάληψη. Ο αλγόριθμος συνεχίζει ως εξής:

- 1) Επιλέγω ένα υποψήφιο x_1^* τέτοιο ώστε: $x_1^* \sim f(x_1 | x_2^{i-1}, x_3^{i-1}, \dots, x_n^{i-1})$.
- 2) Επιλέγω ένα υποψήφιο x_2^* τέτοιο ώστε: $x_2^* \sim f(x_2 | x_1^*, x_3^{i-1}, \dots, x_n^{i-1})$.
- 3) Επιλέγω ένα υποψήφιο x_3^* τέτοιο ώστε: $x_3^* \sim f(x_3 | x_1^*, x_2^*, x_4^{i-1}, \dots, x_n^{i-1})$.
- ⋮
- ⋮
- 4) Επιλέγω ένα υποψήφιο x_k^* τέτοιο ώστε: $x_k^* \sim f(x_k | x_1^*, x_2^*, \dots, x_{k-1}^*, x_{k+1}^{i-1}, \dots, x_n^{i-1})$.
- ⋮
- ⋮
- 5) Επιλέγω ένα υποψήφιο x_n^* τέτοιο ώστε: $x_n^* \sim f(x_n | x_1^*, x_2^*, \dots, x_{n-1}^*)$.
- 6) Θέτω $\mathbf{x}_i = (x_1^*, x_2^*, \dots, x_n^*)$.
- 7) Θέτω $i = i+1$. Ο αλγόριθμος τελειώνει όταν τελειώσουν οι επαναλήψεις που έχουμε ορίσει.

Στην πραγματικότητα ο δειγματολήπτης Gibbs είναι μια ειδική περίπτωση του αλγορίθμου Metropolis-Hastings, που για κατανομή εισήγησης χρησιμοποιούμε τις δεσμευμένες κατανομές $f(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Έστω ότι επιλέγουμε υποψήφιο σημείο της k διάστασης του $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Ορίζω $\mathbf{x}_k^* = (x_1, x_2, \dots, x_{k-1}, x_k^*, x_{k+1}, \dots, x_n)$ και $\mathbf{x}_{-k} = \mathbf{x}_{-k}^* = (x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$. Τότε με τον συμβολισμό που χρησιμοποιήσαμε στον αλγόριθμο Metropolis-Hastings, και παρατηρώντας ότι $\pi(\mathbf{x}) = \pi_{\mathbf{x}_k}(\mathbf{x}_{-k})\pi(x_k | \mathbf{x}_{-k})$,

όπου $\pi_{\mathbf{x}_{-k}}(\mathbf{x}_{-k})$ είναι η περιθώρια κατανομή ως προς \mathbf{x}_{-k} , η πιθανότητα αποδοχής του αλγόριθμου όπως ορίζεται στο βήμα 2 του αλγόριθμου υπολογίζεται:

$$\alpha = \min \left(\frac{\pi(\mathbf{x}^*)g(\mathbf{x}|\mathbf{x}^*)}{\pi(\mathbf{x})g(\mathbf{x}^*|\mathbf{x})}, 1 \right) = \min \left(\frac{\pi_{\mathbf{x}_{-k}}(\mathbf{x}_{-k})\pi(x_k^*|\mathbf{x}_{-k})\pi(x_k|\mathbf{x}_{-k})}{\pi_{\mathbf{x}_{-k}}(\mathbf{x}_{-k})\pi(x_k|\mathbf{x}_{-k})\pi(x_k^*|\mathbf{x}_{-k})}, 1 \right) = 1.$$

Έτσι ο δειγματολήπτης Gibbs είναι μια ειδική περίπτωση του Metropolis Hastings, για την οποία η πιθανότητα αποδοχής του υποψήφιου σημείου είναι η μονάδα.

Για να καταλάβουμε καλύτερα γιατί ο δειγματολήπτης Gibbs εξάγει δείγμα από την από κοινού κατανομή $f(\mathbf{x})$ θα δούμε ένα παράδειγμα στις δύο διαστάσεις.

Έστω X, Y τυχαίες μεταβλητές που ακολουθούν την κατανομή Bernoulli με από κοινού συνάρτηση μάζας κατανομής:

$$\begin{bmatrix} f_{X,Y}(0,0) & f_{X,Y}(1,0) \\ f_{X,Y}(0,1) & f_{X,Y}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}, \text{ όπου } p_i > 0 \text{ και } \sum_{i=1}^4 p_i = 1.$$

Οι περιθώριες κατανομές των X και Y είναι αντίστοιχα:

$$f_X = [f_X(0), f_X(1)] = [p_1 + p_3, p_2 + p_4],$$

$$f_Y = [f_Y(0), f_Y(1)] = [p_1 + p_2, p_3 + p_4].$$

Οι δεσμευμένες πιθανότητες μπορούν να γραφούν υπό μορφή πινάκων ως εξής:

$$A_{X|Y} = \begin{bmatrix} f_{X|Y}(0|0) & f_{X|Y}(1|0) \\ f_{X|Y}(0|1) & f_{X|Y}(1|1) \end{bmatrix} = \begin{bmatrix} \frac{p_1}{p_1 + p_2} & \frac{p_2}{p_1 + p_2} \\ \frac{p_3}{p_3 + p_4} & \frac{p_4}{p_3 + p_4} \end{bmatrix},$$

$$A_{Y|X} = \begin{bmatrix} f_{Y|X}(0|0) & f_{Y|X}(1|0) \\ f_{Y|X}(0|1) & f_{Y|X}(1|1) \end{bmatrix} = \begin{bmatrix} \frac{p_1}{p_1 + p_3} & \frac{p_3}{p_1 + p_3} \\ \frac{p_2}{p_2 + p_4} & \frac{p_4}{p_2 + p_4} \end{bmatrix}.$$

Ο δειγματολήπτης Gibbs είναι μια μέθοδος παραγωγής δείγματος από την $f_{X,Y}$ με έμμεσο τρόπο, παράγοντας δείγματα από τις δεσμευμένες κατανομές $f_{X|Y}$ και $f_{Y|X}$. Έτσι ξεκινώντας από μια συγκεκριμένη κατάσταση X_0 και Y_0 παράγονται εναλλάξ από τις $f_{X|Y}$ και $f_{Y|X}$ μια ακολουθία:

$$X_0, Y_0, X_1, Y_1, \dots, X_i, Y_i, \dots, X_n, Y_n.$$

Αν ενδιαφερόμαστε στην παραγωγή μόνο από την περιθώρια κατανομή f_X τότε θεωρούμε τα $X_0, X_1, \dots, X_i, \dots, X_n$ από την πιο πάνω ακολουθία. Τότε αυτή μπορεί να θεωρηθεί σαν μια Μαρκοβιανή αλυσίδα με την ιδιότητα:

$$P[X_i = x_i | X_{i-1} = x_{i-1}] = \sum_y P[X_i = x_i | Y_i = y] P[Y_i = y | X_{i-1} = x_{i-1}].$$

Ο πίνακας μετάβασης της συγκεκριμένης αλυσίδας θα είναι $A_{X|X} = A_{Y|X} A_{X|Y}$. Αν συμβολίσουμε την περιθώρια κατανομή του X_k με $f_k = [f_k(0), f_k(1)]$ τότε

$$f_k = f_0 (A_{X|X})^k = f_0 (A_{X|X})^{k-1} A_{X|X} = f_{k-1} A_{X|X}.$$

Αποδεικνύεται (Hoel, Port, and Stone (1978)) ότι εάν ο πίνακας $A_{X|X}$ έχει όλα του τα στοιχεία θετικά τότε η πιο πάνω σχέση συνεπάγεται ότι για οποιαδήποτε αρχική κατανομή f_0 και καθώς $k \rightarrow \infty$ η f_k συγκλίνει στην **μοναδική** κατανομή f που ικανοποιεί την πιο πάνω σχέση. Έτσι:

$$\begin{aligned} f_X A_{X|X} &= [p_1 + p_3 \quad p_2 + p_4] \times \begin{bmatrix} \frac{p_1}{p_1 + p_3} & \frac{p_3}{p_1 + p_3} \\ \frac{p_2}{p_2 + p_4} & \frac{p_4}{p_2 + p_4} \end{bmatrix} \times \begin{bmatrix} \frac{p_1}{p_1 + p_2} & \frac{p_2}{p_1 + p_2} \\ \frac{p_3}{p_3 + p_4} & \frac{p_4}{p_3 + p_4} \end{bmatrix} \\ &= [p_1 + p_2 \quad p_3 + p_4] \times \begin{bmatrix} \frac{p_1}{p_1 + p_2} & \frac{p_2}{p_1 + p_2} \\ \frac{p_3}{p_3 + p_4} & \frac{p_4}{p_3 + p_4} \end{bmatrix} = [p_1 + p_3, \quad p_2 + p_4] = f_X. \end{aligned}$$

Και άρα καθώς $k \rightarrow \infty$ η κατανομή της X_k θα συγκλίνει στην κατανομή της f_X .

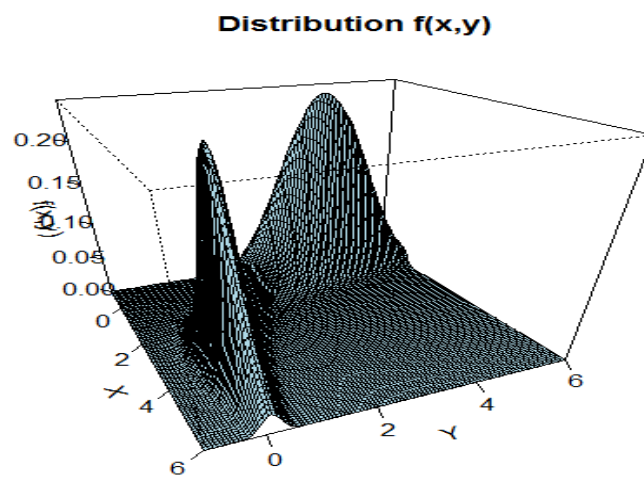
Παράδειγμα

Έστω το τυχαίο διάνυσμα $\mathbf{X} = (X, Y)$ που ακολουθεί τη διδιάστατη κατανομή

$$f(\mathbf{x}) = f(x, y) = c \cdot \exp\left(-\left(x^2 y^2 + x^2 + y^2 - 8x - 8y\right)/2\right),$$

όπου $c = 1/20216.335877$ η σταθερά κανονικοποίησης. Θέλουμε να εκτιμήσουμε την παράμετρο $l = E[X]$ η οποία έχει θεωρητική τιμή $l \approx 1.85997$.

Μια γραφική αναπαράσταση της κατανομής φαίνεται στο Σχήμα 34.



Σχήμα 34

Η από κοινού σ.π.π. των X και Y μπορεί να μετασχηματιστεί ως ακολούθως:

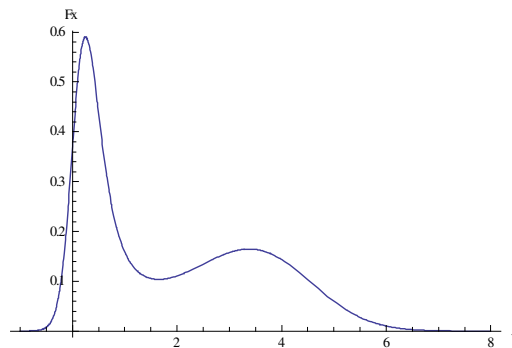
$$\begin{aligned} f(x, y) &= c \cdot \exp\left(-\left(x^2 y^2 + x^2 + y^2 - 8x - 8y\right)/2\right) = c(y) \cdot \exp\left(-\left(x^2 (y^2 + 1) - 8x\right)/2\right) \\ &= c(y) \cdot \exp\left(-\frac{1+y^2}{2} \left(x^2 - \frac{8x}{1+y^2} + \left(\frac{4}{1+y^2}\right)^2\right) + \frac{8}{1+y^2}\right) \\ &= c_2(y) \cdot \exp\left(-\frac{1+y^2}{2} \left(x - \frac{4}{1+y^2}\right)^2\right). \end{aligned}$$

Όπου $c(y) = c \exp\left(-\frac{y^2 - 8y}{2}\right)$ και $c_2(y) = c \exp\left(-\frac{y^2 - 8y}{2} + \frac{8}{1+y^2}\right)$ είναι συναρτήσεις που εξαρτώνται μόνο από το y . Έτσι για δεδομένο y συμπεραίνουμε ότι η δεσμευμένη

κατανομή $f_{x|y}$ είναι η $N\left(\frac{4}{1+y^2}, \frac{1}{1+y^2}\right)$. Αντίστοιχα λόγω συμμετρίας για δεδομένο x η δεσμευμένη κατανομή $f_{y|x}$ είναι η $N\left(\frac{4}{1+x^2}, \frac{1}{1+x^2}\right)$.

Σημείωση: Οι περιθώριες κατανομές f_x και f_y είναι αρκετά δύσκολο να υπολογιστούν αναλυτικά και για αυτό χρησιμοποιήσαμε το πιο πάνω “τέχνασμα” για να βρούμε τις δεσμευμένες κατανομές. Ωστόσο για σκοπούς σύγκρισης των τιμών που θα προσομοιώσουμε, υπολογίσαμε την περιθώρια f_x μέσω του Mathematica

($f_x = \frac{ce^{\frac{4x-x^2}{2} + \frac{8}{1+x^2}} \sqrt{2\pi}}{\sqrt{1+x^2}}$) και η κατανομή της φαίνεται στο Σχήμα 35:



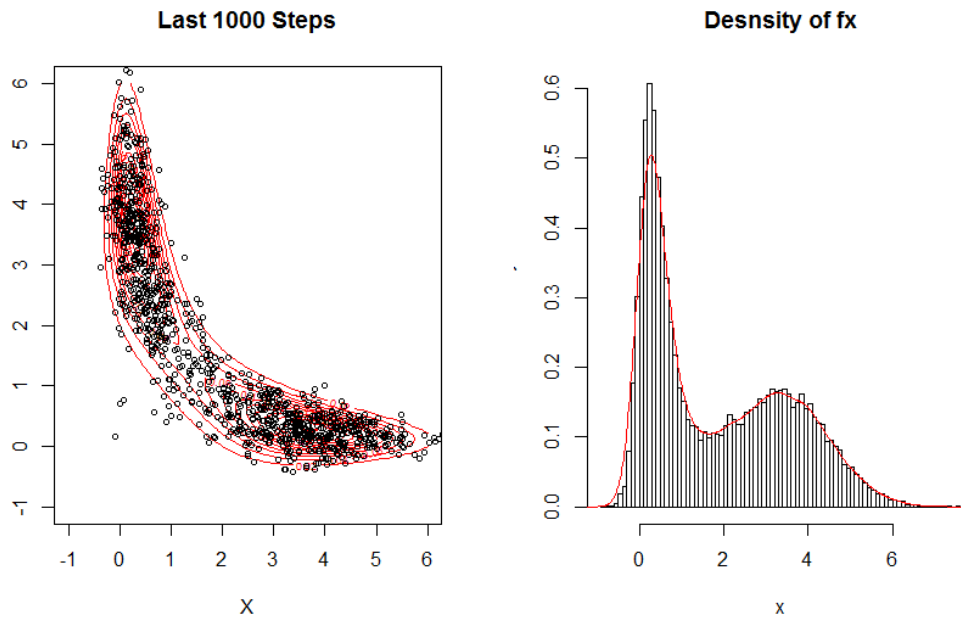
Σχήμα 35

Έτσι χρησιμοποιώντας την μεθοδολογία του δειγματολήπτη Gibbs έχουμε τον πιο κάτω αλγόριθμο:

- 1) Θέτουμε αρχικό $\mathbf{x}[0] = (x_0, y_0)$ και $i = 1$.
- 2) Θέτουμε $m = \frac{4}{1+y[i-1]^2}$ και $s = \sqrt{\frac{1}{1+y[i-1]^2}}$.
- 3) Προσομοιώνουμε $Z \sim N(0,1)$ και θέτουμε $x^* = m + sZ$.
- 4) Θέτω $m = \frac{4}{1+x^{*2}}$ και $s = \sqrt{\frac{1}{1+x^{*2}}}$.
- 5) Προσομοιώνουμε $Z \sim N(0,1)$ και θέτουμε $y^* = m + sZ$.
- 6) Θέτουμε $\mathbf{x}[i] = (x[i], y[i]) = (x^*, y^*)$, $i = i+1$ και αν $i < N$ σταματούμε αλλιώς επιστρέφουμε στο βήμα 2.
- 7) Επιστρέφω το \bar{x} .

Εδώ είναι χρήσιμο να παρατηρήσουμε ότι σε αντίθεση με τον αλγόριθμο του Metropolis Hastings, στον αλγόριθμο Gibbs δεν απορρίπτεται κανένα σημείο. Ακολουθεί ο κώδικας στην R για $N=30000$ και τα αποτελέσματα του κώδικα.

```
n<-30000
x<-rep(0,n)
y<-rep(0,n)
x[1]<-1
y[1]<-1
f <- function(x, y)
{
  c<-1/20216.335877
  z<-c*exp(-(x^2*y^2+x^2+y^2+8*x+8*y)/2)
}
for (i in 1:n) {
  s<-1/(1+y[i]^2)
  m<-4/(1+y[i]^2)
  xexp<-m+sqrt(s)*rnorm(1)
  s<-1/(1+xexp^2)
  m<-4/(1+xexp^2)
  yexp<-m+sqrt(s)*rnorm(1)
  x[i+1]<-xexp
  y[i+1]<-yexp
}
z<-f(x,y)
mean(x)
par(mfrow=c(1,2))
f <- function(x, y)
{
  c<-1/20216.335877
  z<-c*exp(-(x^2*y^2+x^2+y^2-8*x-8*y)/2)
}
y1 <- seq(-1,6, length=100)
x1 <-seq(-1,6, length=100)
z<-outer(x1,y1,f)
contour(x1,y1,z,col="red", main= "Last 1000 Steps",xlab="X", ylab="Y")
points(x[29000:30000],y[29000:30000],cex=0.7)
hist(x,freq=FALSE,nclass=100,main="Desnsity of fx")
lines(density(x), col="red")
```

Σχήμα 36

Στο Σχήμα 36 αριστερά φαίνονται τα τελευταία 1000 βήματα που προσομοιώθηκαν από τον αλγόριθμο μαζί με ένα contour plot (ισοϋψείς καμπύλες) της θεωρητικής κατανομής. Στα δεξιά φαίνεται η εκτιμήτρια της κατανομής f_x . Ο κώδικας μας επιστρέφει την εκτίμηση για το $\hat{l} = \bar{X} = 1.8585$ που είναι μια πάρα πολύ καλή προσέγγιση στην πραγματική τιμή του l .

ΚΕΦΑΛΑΙΟ 5

ΤΕΧΝΙΚΕΣ ΕΛΑΤΤΩΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

5.1 Εισαγωγή

Όπως είδαμε και στα προηγούμενα κεφάλαια για να εκτιμήσουμε μια παράμετρο θ κάποιας κατανομής (ή κάποιου στοχαστικού μοντέλου) F , προσομοιώνουμε n τυχαίους αριθμούς $X_1, X_2, \dots, X_n \sim F$ και εκτιμούμε την παράμετρο με την βοήθεια αυτών.

Έστω ότι θέλουμε να υπολογίσουμε το ολοκλήρωμα:

$$\theta = \int_a^b f(x)dx = \int_a^b \varphi(x)p(x)dx,$$

όπου $p(x)$ είναι σ.π.π. και $\int_a^b p(x)dx = 1$. Τότε αν $X_1, X_2, \dots, X_n \sim p(x)$ και $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$ τότε:

$$E(\hat{\theta}) = E\left[\frac{1}{n} \sum_{i=1}^n \varphi(X_i)\right] = \frac{1}{n} \sum_{i=1}^n E[\varphi(X_i)] = \frac{1}{n} n E[\varphi(X)] = \int_a^b \varphi(x) p(x) dx = \theta.$$

Δηλαδή η $\hat{\theta}$ είναι μια αμερόληπτη εκτιμήτρια του θ , ενώ η διασπορά της είναι:

$$V(\hat{\theta}) = V\left[\frac{1}{n} \sum_{i=1}^n \varphi(X_i)\right] = \frac{1}{n^2} \sum_{i=1}^n V[\varphi(X_i)] = \frac{1}{n} V[\varphi(X)] = \frac{1}{n} \int_a^b (\varphi(x) - \theta)^2 p(x) dx = \frac{C}{n}.$$

Άρα διαφορετικές επιλογές της $\varphi(x)$ και $p(x)$ μπορούν να οδηγήσουν σε διαφορετικά C και συνεπώς σε διαφορετικές διασπορές της εκτιμήτριας.

Έστω για παράδειγμα ότι θέλουμε να εκτιμήσουμε την τιμή του ολοκληρώματος:

$$\theta = \int_2^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx.$$

Η θεωρητική τιμή του ολοκληρώματος είναι $\theta = 1 - F(2) \approx 0.1476$, όπου $F(x)$ είναι η σ.κ.π. της κατανομής Cauchy. Το ολοκλήρωμα μπορεί να μετασχηματιστεί ως ακολούθως:

$$\theta = \int_2^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx = \int_{-\infty}^{\infty} I_{x>2} \frac{1}{\pi} \frac{1}{1+x^2} dx = \int_{-\infty}^{\infty} \varphi(x) p(x) dx.$$

Τότε παράγοντας τιμές X_1, X_2, \dots, X_n από την $p(x)$, που είναι η κατανομή Cauchy, και εκτιμώντας την τιμή του ολοκληρώματος από τον δειγματικό μέσο των $\varphi(X_i)$ έχουμε:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \varphi(X_i) \Rightarrow n\hat{\theta} = \sum_{i=1}^n \varphi(X_i) \sim \text{Bin}(n, \theta).$$

Άρα η διασπορά της εκτιμήτριας είναι:

$$\text{Var}(\hat{\theta}) = \frac{n\theta(1-\theta)}{n^2} \approx \frac{0.126}{n}.$$

Αν τώρα μετασχηματίσουμε το αρχικό μας ολοκλήρωμα ως:

$$\theta = \int_2^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{2} \int_{-\infty}^{\infty} I_{|x|>2} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{2} \int_{-\infty}^{\infty} \varphi(x) p(x) dx,$$

τότε παράγοντας τιμές X_1, X_2, \dots, X_n από την $p(x)$, που είναι η κατανομή Cauchy, εκτιμούμε την τιμή του θ από την:

$$\hat{\theta} = \frac{1}{2n} \sum_{i=1}^n \varphi(X_i) \Rightarrow 2n\hat{\theta} = \sum_{i=1}^n \varphi(X_i) \sim \text{Bin}(n, 2\theta).$$

Άρα η διασπορά της εκτιμήτριας γίνεται:

$$\text{Var}(\hat{\theta}) = \frac{n \cdot 2\theta(1-2\theta)}{4n^2} \approx \frac{0.052}{n},$$

δηλαδή 2.4 φορές μικρότερη από πριν. Στη συνέχεια του κεφαλαίου θα δώσουμε κάποιες τεχνικές ελάττωσης της διασποράς της εκτιμήτριας.

5.2 Αντιθετικές Τυχαίες Μεταβλητές

Όπως γνωρίζαμε, αν $X_1, X_2, X_3, \dots, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές τότε

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}, \text{ όπου } \sigma^2 \text{ είναι η διασπορά των } X_i. \text{ Ας υποθέσουμε ότι οι } X_i, i=1,2,\dots,n$$

δεν είναι ανεξάρτητες τότε:

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \left[\sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{cov}(X_i, X_j) \right] = \frac{\sigma^2}{n} + \frac{2}{n^2} \sum_{i < j} \text{cov}(X_i, X_j).$$

Αν λοιπόν οι τ.μ. $X_i, i=1,2,\dots,n$ είναι αρνητικά συσχετισμένες θα ισχύει $\text{cov}(X_i, X_j) < 0$ και άρα

$$\text{Var}(\bar{X}_n) < \frac{\sigma^2}{n}.$$

Η μέθοδος των αντιθετικών τυχαίων μεταβλητών είναι μια μέθοδος παραγωγής αρνητικά συσχετισμένων τυχαίων μεταβλητών. Έστω λοιπόν $U_1, U_2, \dots, U_n \sim U(0,1)$, ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την ομοιόμορφη κατανομή στο $(0,1)$

και h μια μονότονη συνάρτηση τέτοια ώστε οι τυχαίες μεταβλητές $X_1, X_2, X_3, \dots, X_n$ να είναι τέτοιες ώστε $X_1 = h(U_1), \dots, X_n = h(U_n)$. Τότε η εκτιμήτρια του μέσου δίνεται από

$$\text{τον δειγματικό μέσο των } X_i, \hat{\mu} = \frac{1}{n} \sum_{i=1}^n h(U_i).$$

Η ιδέα είναι να παραχθούν εξαρτημένα ζεύγη τυχαίων μεταβλητών με τον ακόλουθο τρόπο:

$$X_1^{*(a)} = h(U_1), X_1^{*(b)} = h(1-U_1), X_2^{*(a)} = h(U_2), X_2^{*(b)} = h(1-U_2), \dots \text{ΚΟΚ.}$$

Έτσι το δείγμα μας χωρίζεται στις δύο μέρη :

$$X_i^* = \begin{cases} X_i^{*(a)} = h(U_i) \\ X_i^{*(b)} = h(1-U_i) \end{cases}, i = 1, 2, \dots, n.$$

Είναι προφανές ότι για την παραγωγή δείγματος μεγέθους n χρειαζόμαστε $n/2$ τυχαίους αριθμούς από την ομοιόμορφη κατανομή $U(0,1)$.

Για να αποδείξουμε ότι οι $X_i^{*(a)}$ και $X_i^{*(b)}$ είναι αρνητικά συσχετισμένες χρησιμοποιούμε το παρακάτω θεώρημα, το οποίο παραθέτουμε χωρίς απόδειξη.

ΘΕΩΡΗΜΑ: Αν Y τυχαία μεταβλητή τότε για κάθε ζεύγος αύξουσων συναρτήσεων $f, g : R \rightarrow R$ θα ισχύει:

$$\text{cov}(f(Y), g(Y)) \geq 0.$$

Αν h αύξουσα, τότε για $Y = U_i \sim U(0,1)$, $f(x) = h(x)$ και $g(x) = -h(1-x)$ (η $-h(1-x)$ είναι αύξουσα συνάρτηση του x) από το θεώρημα έχουμε:

$$\begin{aligned} \text{cov}(h(U_i), -h(1-U_i)) &\geq 0, \\ \Rightarrow \text{cov}(h(U_i), h(1-U_i)) &\leq 0. \end{aligned}$$

Αντίστοιχα αν h φθίνουσα, τότε για $Y = U(0,1)$, $f(x) = -h(x)$ (η $-h(x)$ είναι αύξουσα συνάρτηση του x) και $g(x) = h(1-x)$ από το θεώρημα έχουμε:

$$\begin{aligned} \text{cov}(-h(U_i), h(1-U_i)) &\geq 0, \\ \Rightarrow \text{cov}(h(U_i), h(1-U_i)) &\leq 0. \end{aligned}$$

Έτσι πετύχαμε ένα δείγμα που αποτελείται από αρνητικά συσχετισμένα ζεύγη. Τότε η εκτίμηση του μέσου δίνεται ως εξής:

$$\hat{\mu}^{(*)} = \frac{1}{n} \sum_{i=1}^{n/2} (X_i^{*(a)} + X_i^{*(b)}).$$

Βλέπουμε τώρα ότι:

$$\begin{aligned} \text{Var}[\hat{\mu}^{*}] &= \frac{n/2}{n^2} (\text{Var}[h(X^{*(a)})] + \text{Var}[h(X^{*(b)})] + 2\text{cov}(X^{*(a)}, X^{*(b)})) \\ &= \text{Var}(\hat{\mu}) + \frac{\text{cov}(X^{*(a)}, X^{*(b)})}{n} < V(\hat{\mu}). \end{aligned}$$

Στην γενική περίπτωση αν X_i παράγονται με βάση m ανεξάρτητους τυχαίους αριθμούς U_1, U_2, \dots, U_m έτσι ώστε:

$$X_1 = h(\mathbf{U}_1), X_2 = h(\mathbf{U}_2), \dots, X_n = h(\mathbf{U}_n),$$

όπου $\mathbf{U}_i = (U_1^i, U_2^i, \dots, U_m^i)$ και η $h(\mathbf{U}_i): R^m \rightarrow R$ είναι μονότονη κατά σημείο συνάρτηση, τότε όμοια με παραπάνω μπορούμε να θέσουμε:

$$X_1^{*(a)} = h(\mathbf{U}_1), X_1^{*(b)} = h(1 - \mathbf{U}_1), X_2^{*(a)} = h(\mathbf{U}_2), X_2^{*(b)} = h(1 - \mathbf{U}_2), \dots \kappa.ο.κ.$$

Παράδειγμα

Θέλουμε να υπολογίσουμε την τιμή του ολοκληρώματος $I = \int_0^1 e^x dx$. Με την μέθοδο που αναπτύξαμε στην παράγραφο (4.2) θα μπορούμε να παράγουμε n τυχαίους αριθμούς $U_i \sim U(0,1)$, να θέσουμε $X_i = e^{U_i}$ και να εκτιμήσουμε το ολοκλήρωμα με τον δειγματικό μέσο:

$$\hat{I} = \overline{X_n} = \frac{1}{n} \sum_i^n e^{U_i}.$$

Η συγκεκριμένη εκτιμήτρια θα είχε διασπορά:

$$Var(\overline{X_n}) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(e^{U_i}) = \frac{1}{n} Var(e^U),$$

$$\text{όπου } Var(e^U) = (E[e^{2U}] - E[e^U]^2) = \int_0^1 e^{2U} dU - \left(\int_0^1 e^U dU\right)^2 = \frac{e^2 - 1}{2} - (e - 1)^2 \approx 0.2420.$$

Προσέχοντας τώρα ότι η $h(x) = e^x$ είναι γνήσια θετική συνάρτηση μπορούμε να χρησιμοποιήσουμε την μέθοδο των αντιθετικών τυχαίων μεταβλητών και θα έχουμε χρησιμοποιώντας τον συμβολισμό που καθορίσαμε πιο πάνω

$$X_i^* = \begin{cases} X_i^{*(a)} = h(U_i) = e^{U_i} \\ X_i^{*(b)} = h(1 - U_i) = e^{1 - U_i} \end{cases}, i = 1, 2, \dots, \frac{n}{2}.$$

$$Var(\overline{X_n^*}) = Var\left(\frac{1}{n} \sum_{i=1}^{n/2} (X_i^{*(a)} + X_i^{*(b)})\right) = \frac{1}{n^2} \sum_{i=1}^{n/2} Var(e^{U_i} + e^{1 - U_i}) = \frac{1}{2n} Var(e^U + e^{1 - U}),$$

Όπου,

$$\text{Var}(e^U + e^{1-U}) = \text{Var}(e^U) + \text{Var}(e^{1-U}) + 2\text{cov}(e^U, e^{1-U}),$$

$$\text{Var}(e^{1-U}) = E[e^{2(1-U)}] + (E[e^{1-U}])^2 = \frac{e^2 - 1}{2} + (e - 1)^2,$$

$$\text{cov}(e^U, e^{1-U}) = E[e^U e^{1-U}] - E[e^U] E[e^{1-U}] = e - (e - 1)(e - 1) = 3e - e^2 - 1,$$

$$\Rightarrow \text{Var}(e^U + e^{1-U}) = 2 \left(\frac{e^2 - 1}{2} - (e - 1)^2 \right) + 2(3e - e^2 - 1) \approx 0.01565.$$

Η σχετική αποτελεσματικότητα ορίζεται σαν $\frac{\text{Var}(\overline{X}_n)}{\text{Var}(\overline{X}_n^*)}$ και θα είναι ίση:

$$\frac{\text{Var}(\overline{X}_n)}{\text{Var}(\overline{X}_n^*)} \approx \frac{1/n \cdot 0.2420}{1/2n \cdot 0.01565} = \frac{2 \cdot 0.2420}{0.01565} = 30.93.$$

Δηλαδή η εκτιμήτρια του μέσου που βασίζεται στις αντιθετικές μεταβλητές \overline{X}_n^* θα έχει διασπορά 30.93 φορές μικρότερη από τη διασπορά της εκτιμήτριας \overline{X}_n .

Ακολουθεί κώδικας στην R ο οποίος υπολογίζει το συγκεκριμένο ολοκλήρωμα με τις δύο μεθόδους και επιστρέφει την σχετική αποτελεσματικότητα. (η σχετική αποτελεσματικότητα που εκτιμήθηκε από τον κώδικα είναι 30.9265).

```

mxant<-rep(0,1000) ; mx<-rep(0,1000)
for (i in 1:1000) {
  n<-10000; u1<-runif(n/2) ; u2<-runif(n/2); u<-append(u1,u2)
  xa<-exp(u1);  xb<-exp(1-u1)
  xant<-append(xa,xb) # to x xrisimopointas tis antithetikes metablites
  x<-exp(u)      #to x me tin klassiki monte carlo oloklirosi
  mxant[i]<-mean(xant)
  mx[i]<-mean(x)
}
s1<-var(mxant);s2<-var(mx)
s2/s1
    
```

5.3 Ρυθμιστικές Μεταβλητές

Ας υποθέσουμε ότι και πάλι θέλουμε να εκτιμήσουμε την μέση τιμή μ μιας τυχαίας μεταβλητής X και ότι τώρα εκτός από την σ.π.π. της X μπορούμε να παράγουμε τιμές και από την σ.π.π. μιας άλλης τυχαίας μεταβλητής Y η οποία έχει γνωστή μέση τιμή μ_y , και είναι εξαρτημένη από την X .

Θεωρούμε την τυχαία μεταβλητή:

$$X_a = X - a(Y - \mu_y), \quad a \in R.$$

(η οποία έχει και αυτή μέση τιμή $E[X_a] = E[X] - a(E[Y] - \mu_y) = E[X] = \mu$)

Η διασπορά της X_a θα είναι:

$$\text{Var}(X_a) = \text{Var}(X - a(Y - \mu_y)) = \text{Var}(X) + a^2 \text{Var}(Y) - 2a \text{cov}(X, Y),$$

και το a^* για το οποίο ελαχιστοποιείται η διασπορά είναι:

$$\frac{d\text{Var}(X_a)}{da} = 2a^* \text{Var}(Y) - 2 \text{cov}(X, Y) = 0,$$

$$a^* = \frac{\text{cov}(X, Y)}{\text{Var}(Y)}.$$

Έτσι η διασπορά της X_a για το συγκεκριμένο a^* είναι:

$$\begin{aligned} \text{Var}(X_{a^*}) &= \text{Var}(X) + \left(\frac{\text{cov}(X, Y)}{\text{Var}(Y)} \right)^2 \text{Var}(Y) - 2 \frac{\text{cov}(X, Y)}{\text{Var}(Y)} \text{cov}(X, Y) \\ &= \text{Var}(X) - \frac{\text{cov}(X, Y)^2}{\text{Var}(Y)} \\ &= \text{Var}(X) - (a^*)^2 \text{Var}(Y). \end{aligned}$$

Η τυχαία μεταβλητή Y καλείται ρυθμιστική μεταβλητή ή μεταβλητή. Μπορούμε λοιπόν να προσομοιώσουμε τα ζεύγη τιμών $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ και να εκτιμήσουμε την μέση τιμή μ από τον δειγματικό μέσο των $(X_{a^*})_i$ $i = 1, 2, \dots, n$:

$$\widehat{\mu}_{\alpha^*} = \overline{X_{\alpha^*}} = \frac{1}{n} \sum_{i=1}^n (X_i - a^* (Y_i - \mu_y)).$$

Η διασπορά της πιο πάνω εκτιμήτριας είναι:

$$\text{Var}(\overline{X_{\alpha^*}}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n (X_i - a^* (Y_i - \mu_y))\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}((X_i - a^* (Y_i - \mu_y))) = \frac{\text{Var}(X) - (a^*)^2 \text{Var}(Y)}{n},$$

και η σχετική αποτελεσματικότητα όπως ορίστηκε στην παράγραφο (5.2)

$$\begin{aligned} \frac{\text{Var}(\overline{X})}{\text{Var}(\overline{X_{\alpha^*}})} &= \frac{\frac{1}{n}(\text{Var}(X))}{\frac{1}{n}(\text{Var}(X) - (a^*)^2 \text{Var}(Y))} \\ &= \frac{\text{Var}(X)}{\text{Var}(X) \left[1 - \frac{(\text{cov}(X, Y))^2}{\text{Var}(X)\text{Var}(Y)}\right]} \\ &= \frac{1}{1 - (\rho(X, Y))^2} > 1. \end{aligned}$$

Άρα η νέα εκτιμήτρια $\widehat{\mu}_{\alpha^*} = \overline{X_{\alpha^*}}$ είναι αποτελεσματικότερη της αρχικής και μάλιστα όσο μεγαλύτερη είναι η εξάρτηση των τυχαίων μεταβλητών X και Y τόσο πιο αποτελεσματική θα είναι.

Στην πράξη συνήθως οι ποσότητες $\text{cov}(X, Y)$ και $\text{Var}(Y)$ δεν είναι γνωστές. Συνεπώς μπορούν να εκτιμηθούν από τις πιο κάτω αμερόληπτες εκτιμήτριες:

$$\begin{aligned} \widehat{\text{cov}(X, Y)} &= S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y}), \\ \widehat{\text{Var}(Y)} &= S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y})^2, \\ \widehat{a^*} &= \frac{S_{XY}}{S_Y^2}. \end{aligned}$$

Σαν παράδειγμα θα θεωρήσουμε το ολοκλήρωμα που είδαμε και στις αντιθετικές μεταβλητές $I = \int_0^1 e^x dx$ το οποίο αρχικά εκτιμάτε από την εκτιμήτρια:

$$\hat{I} = \bar{X} = \frac{1}{n} \sum_{i=1}^n e^{U_i}, \quad U_i \sim U(0,1), \quad i=1,2,\dots,n.$$

Πρώτα θα πρέπει να επιλέξουμε τη ρυθμιστική μεταβλητή Y η οποία πρέπει να είναι συσχετισμένη με την $X = e^U$ και με γνωστή μέση τιμή μ_Y . Με μια πρώτη ματιά φαίνεται ότι μια καλή επιλογή είναι η $Y = U$, η οποία είναι θετικά συσχετισμένη με την $X = e^U$. Ακολούθως υπολογίζουμε την βέλτιστη τιμή της σταθεράς a :

$$a^* = \frac{\text{cov}(e^U, U)}{\text{Var}(U)} = \frac{E[e^U U] - E[e^U]E[U]}{\text{Var}(U)} = \frac{\int_0^1 u e^u du - \frac{1}{2} \int_0^1 e^u du}{\frac{1}{12}} = \frac{1 - \frac{1}{2}(e-1)}{\frac{1}{12}} = 6(3-e).$$

Άρα η νέα μας εκτιμήτρια γίνεται:

$$\hat{I}^* = \frac{1}{n} \sum_{i=1}^n (X_i - a^*(Y_i - \mu_Y)).$$

Ακολούθως υπολογίζουμε την σχετική αποτελεσματικότητα της εκτιμήτριας:

$$\begin{aligned} \frac{\text{Var}(\hat{I})}{\text{Var}(\hat{I}^*)} &= \frac{1}{1 - \rho^2(e^U, U)} = \frac{1}{1 - \frac{\text{cov}^2(e^U, U)}{\text{Var}(e^U)\text{Var}(U)}} = \frac{\text{Var}(e^U)\text{Var}(U)}{\text{Var}(e^U)\text{Var}(U) - \text{cov}^2(e^U, U)} \\ &= \frac{\left(\frac{e^2-1}{2} - (e-1)^2\right) \frac{1}{12}}{\left(\frac{e^2-1}{2} - (e-1)^2\right) \frac{1}{12} - \frac{1}{4}(3-e)^2} \approx 61.4268. \end{aligned}$$

Συνεπώς η νέα εκτιμήτρια του ολοκληρώματος θα έχει διασπορά 61.4268 φορές μικρότερη από την διασπορά της αρχικής εκτιμήτριας. Ακολουθεί κώδικας στην R ο οποίος υπολογίζει το συγκεκριμένο ολοκλήρωμα με τις δύο μεθόδους και επιστρέφει την σχετική αποτελεσματικότητα καθώς και 95% διαστήματα εμπιστοσύνης.

```

mx<-rep(0,1000)
my<-rep(0,1000)
for (i in 1:1000) {
  n<-10000
  u<-runif(n)
  x<-exp(u)
  a=6*(3-exp(1))
  y<-x-a*(u-0.5)
  mx[i]<-mean(x)
  my[i]<-mean(y)
}
var(mx)/var(my)
dex1<-mean(mx)-sqrt(var(mx))*1.96
dex2<-mean(mx)+sqrt(var(mx))*1.96
dex<-c(dex1,dex2)
dey1<-mean(my)-sqrt(var(my))*1.96
dey2<-mean(my)+sqrt(var(my))*1.96
dey<-c(dey1,dey2)
dex ; dey

```

Ο κώδικας μας επιστρέφει σχετική αποτελεσματικότητα 61.0565 ενώ τα διαστήματα εμπιστοσύνης είναι (όπου d είναι το εύρος του διαστήματος εμπιστοσύνης):

Χωρίς την ρυθμιστική μεταβλητή: $I \in (1.7085, 1.7278)$, $d = 0.0193$.

Με την ρυθμιστική μεταβλητή: $I^* \in (1.7170, 1.7194)$, $d^* = 0.0024$.

Τα αποτελέσματα επιβεβαιώνουν όσα αναλύσαμε πιο πάνω, και μας δείχνουν πόσο πιο αποτελεσματική και ακριβής είναι μια εκτίμηση χρησιμοποιώντας τεχνικές ελάττωσης της διακύμανσης.

5.4 Δειγματοληψία Σπουδαιότητας

Έστω τώρα ότι θέλουμε να υπολογίσουμε το ολοκλήρωμα $I = \int_a^b f(x)dx$. Στην παράγραφο 4.2 είδαμε ότι αν $p(x)$ είναι συνάρτηση πυκνότητας πιθανότητας και X τυχαία μεταβλητή που ακολουθεί κατανομή με σ.π.π. την $p(x)$ τότε:

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{p(x)} p(x)dx = \int_a^b h(x) p(x)dx,$$

και μπορούμε να εκτιμήσουμε το I παράγοντας X_1, X_2, \dots, X_n από την $p(x)$ και χρησιμοποιώντας την εκτιμήτρια:

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n h(X_i).$$

Εναλλακτικά αν $g(x)$ μια άλλη σ.π.π. με $\{x: p(x) > 0\} = \{x: g(x) > 0\}$ τότε:

$$I = \int_a^b h(x)p(x)dx = \int_a^b \frac{h(x)p(x)}{g(x)} g(x)dx = \int_a^b \psi(x)g(x)dx.$$

Έτσι μπορούμε να υπολογίσουμε το ολοκλήρωμα παράγοντας τυχαίους αριθμούς X_1, X_2, \dots, X_n από την $g(x)$ και να χρησιμοποιήσουμε την εκτιμήτρια:

$$\hat{I}_a = \frac{1}{n} \sum_{i=1}^n \psi(X_i).$$

Η διασπορά της \hat{I}_a ισούται με:

$$\begin{aligned} \text{Var}(\hat{I}_a) &= \frac{1}{n} \int_a^b [\psi(x) - I]^2 g(x)dx \\ &= \frac{1}{n} \int_a^b \left[\frac{h(x)p(x)}{g(x)} - I \right]^2 g(x)dx \\ &= \frac{1}{n} \int_a^b \left[\frac{f(x)}{g(x)} - I \right]^2 g(x)dx \\ &= \frac{1}{n} V_g \left(\frac{f(x)}{g(x)} \right). \end{aligned}$$

Άρα η διασπορά μηδενίζεται αν $\frac{f(x)}{g(x)}$ ισούται με το I , και επειδή αυτό προϋποθέτει την γνώση εκ των υστέρων της τιμής που προσπαθούμε να εκτιμήσουμε, για την ελάττωση της διασποράς προσπαθούμε να επιτύχουμε ισότητα του $\frac{f(x)}{g(x)}$ με μια σταθερά, ή ισοδύναμα προσπαθούμε να βρούμε μια $g(x)$ που να μιμείται όσο το δυνατό καλύτερα το σχήμα της αρχικής $f(x)$. Η συνάρτηση $g(x)$ καλείται *συνάρτηση*

σπουδαιότητας (importance function) και η διαδικασία παραγωγής X_1, X_2, \dots, X_n από την $g(x)$ καλείται δειγματοληψία σπουδαιότητας (importance sampling).

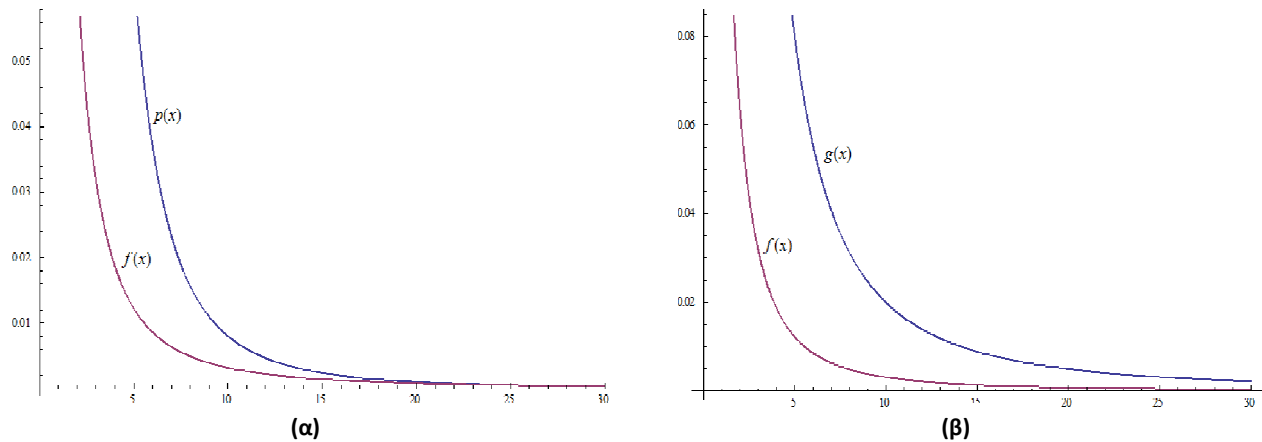
Σαν παράδειγμα θα εκτιμήσουμε το ολοκλήρωμα:

$$I = \int_2^{\infty} f(x) dx = \int_2^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx.$$

Θα εκτιμήσουμε το ολοκλήρωμα δύο φορές. Την πρώτη φορά θα προσομοιώσουμε τιμές από την σ.π.π. $p(x) = \frac{8}{x^3}$ ($\int_2^{\infty} \frac{8}{x^3} dx = 1$) και την άλλη την σ.π.π. $g(x) = \frac{2}{x^2}$

($\int_2^{\infty} \frac{2}{x^2} dx = \left. -\frac{2}{x} \right|_2^{\infty} = 0 + 1 = 1$). Τα γραφήματα των $f(x)$, $g(x)$ και $p(x)$ φαίνονται στο

Σχήμα 37.



Σχήμα 37

Για την παραγωγή τιμών από τις g και p θα χρησιμοποιήσουμε την μέθοδο της αντιστροφής:

$$P(x) = \int_2^x \frac{8}{t^3} dt = 1 - \frac{4}{x^2} \text{ και άρα } P^{-1}(U) = \sqrt{\frac{4}{1-U}}.$$

$$G(x) = \int_2^x \frac{2}{t^2} dt = 1 - \frac{2}{x} \text{ και άρα } G^{-1}(U) = \frac{2}{1-U}.$$

Το ολοκλήρωμα μας γίνεται

$$\int_2^{\infty} h(x)p(x)dx = \int_2^{\infty} \frac{x^3}{8\pi(1+x^2)} \frac{8}{x^3} dx \text{ στην πρώτη περίπτωση και}$$

$$\int_2^{\infty} \psi(x)g(x)dx = \int_2^{\infty} \frac{x^2}{2\pi(1+x^2)} \frac{2}{x^2} dx \text{ στην δεύτερη περίπτωση.}$$

Οπότε οι εκτιμήσεις μας:

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n h(X_i), \text{ όπου } X_i \sim p(x) \text{ και}$$

$$\hat{I}_a = \frac{1}{n} \sum_{i=1}^n \psi(X_i), \text{ όπου } X_i \sim g(x).$$

Ο κώδικας στην R είναι ο ακόλουθος:

```
ma<-rep(0,1000)
m<-rep(0,1000)
for (i in 1:1000) {
  n<-10000
  u<-runif(n)
  x<-sqrt(4/(1-u))
  sample<-x^3/(8*pi*(1+x^2))
  m[i]<-mean(sample)
  xa<-2/(1-u)
  samplea<-xa^2/(2*pi*(1+xa^2))
  ma[i]<-mean(samplea)
}
sna<-var(ma)
sn<-var(m)
de1<-mean(m)-sqrt(sn)*1.96
de2<-mean(m)+sqrt(sn)*1.96
de<-c(de1,de2)
de
de1a<-mean(ma)-sqrt(sna)*1.96
de2a<-mean(ma)+sqrt(sna)*1.96
dea<-c(de1a,de2a)
dea
sn/sna
```

Ο άνω κώδικας μας επιστρέφει ένα 95% διαστήματα εμπιστοσύνης για κάθε εκτιμήτρια $I \in (0.1418, 0.1538)$, $I_a \in (0.1474, 0.1478)$ και σχετική αποτελεσματικότητα 1718.34. Η διασπορά της εκτιμήτριας δηλαδή όταν χρησιμοποιήσαμε την $g(x)$ είναι 1718.34 φορές μικρότερη από την διασπορά όταν χρησιμοποιήσαμε την $p(x)$. Βλέπουμε λοιπόν πόσο σημαντική είναι η επιλογή κατάλληλης κατανομής για την Monte Carlo ολοκλήρωση.

Βιβλιογραφία

Ελληνική Βιβλιογραφία

- [1] Δελλαπόρτας, Π. (1994) *Προσομοίωση και Στοχαστικά Μοντέλα*, Οικονομικό Πανεπιστήμιο Αθηνών.
- [2] Κουίκογλου, Β.Σ. (2002) *Προσομοίωση*, Πολυτεχνείο Κρήτης.
- [3] Μπούτσικας, Μ. (2004) *Μέθοδοι Προσομοίωσης και Υπολογιστικές Στατιστικές Τεχνικές*, Πανεπιστήμιο Πειραιά.
- [4] Κοκολάκης, Γ.Ε. και Σπηλιώτης, Ι. (1999) *Εισαγωγή στην Θεωρία Πιθανοτήτων και Στατιστική*, Εκδόσεις Συμεών.
- [5] Χρυσάφινου, Ο. (2004) *Εισαγωγή στις Στοχαστικές Ανελίξεις*, Εκδόσεις Σοφία, Θεσσαλονίκη.
- [6] Κοκολάκης, Γ.Ε. (2008) *Σημειώσεις Στοχαστικών Ανελίξεων*, Εκδόσεις ΕΜΠ, Αθήνα .

Αγγλική Βιβλιογραφία

- [1] Ahrens, J.H. and Dieter, U. (1974) Computer methods for sampling from Gamma Beta Poisson and Binomial distributions, *Computing* **B.12**, 223-224.
- [2] Rubinstein, R.Y. and Kroese, D.P.(2008) *Simulation and the Monte Carlo Method*, John Wiley & Sons Inc.
- [3] Ripley, B.D. (1987) *Stochastic Simulation*, John Wiley & Sons Inc.
- [4] Fishman, G.S. (1999) *Monte Carlo Concepts, Algorithms and Applications*, Springer-Verlag New York Inc.
- [5] Geyer, J.C. (2005) *Markov Chain Monte Carlo Lecture Notes*, University of Minnesota School of Statistics.

- [6] McLeish, D.L (2004) *Monte Carlo Simulation and Finance*, Wiley Finance.