

Ιόνιο Πανεπιστήμιο
Τμήμα Πληροφορικής



-ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ-

ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΛΟΥ ΤΥΧΑΙΟΥ ΠΕΡΙΠΑΤΟΥ

Ελευθέριος Χατζηωακείμ

ΑΜ: Π2013086

Επιβλέπων: Μάρκος Αυλωνίτης

Κέρκυρα 2024

Επιβλέπων

Μάρκος Αυλωνίτης, *Καθηγητής*

Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Θεμιστοκλής Έξαρχος, *Αναπληρωτής Καθηγητής*

Ιόνιο Πανεπιστήμιο

Αριστείδης Βραχάτης, *Επίκουρος Καθηγητής*

Ιόνιο Πανεπιστήμιο

ΠΕΡΙΛΗΨΗ

Στην παρούσα εργασία παρουσιάζεται το ζήτημα της προσομοίωσης τυχαίων διαδικασιών συγκεκριμένα του “Απλού Τυχαίου Περιπάτου”. Η προσομοίωση σε υπολογιστικά συστήματα έχει διευκολύνει σε μέγιστο βαθμό την διενέργεια πειραμάτων στον υπολογιστή και την παραγωγή δειγμάτων και συμπερασμάτων αυτών καθώς σε ελάχιστο χρόνο δίνει αποτελέσματα για αυτά και απελευθερώνει τον ανθρώπινο παράγοντα για περιττές ενέργειες στο πειραματικό σκέλος. Στο πρώτο κεφάλαιο παρουσιάζονται εισαγωγικά θέματα γύρω από την στατιστική και τις πιθανότητες έτσι ώστε να εισαχθεί ο αναγνώστης ομαλά στα επόμενα κεφάλαια. Θέματα όπως ο ορισμός τι είναι η τυχαιότητα καθώς στοιχεία που αποτελούν βάση στις εξισώσεις πιθανοτήτων και ιδιότητες αυτών όπως η μέση τιμή και η διακύμανση. Στο δεύτερο κεφάλαιο παρουσιάζονται οι διάφορες εξισώσεις πιθανοτήτων που μοντελοποιούν τα τυχαία φαινόμενα και διαδικασίες όπως η διαδικασία διακλάδωσης που μπορούμε να την συσχετίσουμε με το φαινόμενο του “Απλού Τυχαίου Περιπάτου” καθώς εμφανίζουν παρόμοιες ιδιότητες κατά την κατασκευή τους. Στο τρίτο κεφάλαιο παρουσιάζονται μαθηματικές τεχνικές που εξηγούν την φύση των τυχαίων αριθμών και τι κατανομή ακολουθούν αυτοί. Επίσης κάποια απλά παραδείγματα προσομοίωσης, δίνοντας έμφαση στην χρήση συγκεκριμένων συναρτήσεων που παράγουν τυχαία αποτελέσματα, τεχνική σημαντική στην διαδικασία της προσομοίωσης. Στο τέταρτο κεφάλαιο ο αναγνώστης εισέρχεται στο πειραματικό σκέλος και το μοντέλο του “Απλού Τυχαίου Περιπάτου” δηλαδή το κύριο στοιχείο αυτού του πονήματος. Παρουσιάζεται το θεωρητικό μοντέλο του καθώς και η κατασκευή του μοντέλου προσομοίωσης που θα το μιμηθεί. Γίνεται ξεκάθαρο μέσω παραδείγματος κώδικα, τα τρία κυρίαρχα στοιχεία της προσομοίωσης : Η συλλογή δεδομένων-δειγμάτων, τεχνικές για την κατασκευή του μοντέλου και η επικύρωση αυτού που θα δείξει κατά πόσο το μοντέλο εμφανίζεται ως αντιπροσωπευτικού του πραγματικού συστήματος. Τέλος παρουσιάζονται διάφορα αποτελέσματα και διαγράμματα, εικόνες που απεικονίζουν το ταίριασμα του υπολογιστικού μοντέλου με το σύστημα.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	3
ΚΕΦΑΛΑΙΟ 1	6
ΠΙΘΑΝΟΤΗΤΕΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ	6
1.1 ΕΝΝΟΙΑ ΣΤΑΤΙΣΤΙΚΗΣ	6
1.2. ΤΥΧΑΙΟΤΗΤΑ	6
1.3. ΠΛΑΝΗ ΤΟΥ ΤΖΟΓΑΔΟΡΟΥ	7
1.4. ΑΝΑΠΑΡΑΣΤΑΣΗ ΔΕΔΟΜΕΝΩΝ	10
1.5. ΔΕΙΓΜΑ	10
1.6. ΠΛΗΘΥΣΜΟΣ	11
1.7. ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΑ ΔΙΑΚΡΙΤΑ ΜΑΘΗΜΑΤΙΚΑ	11
1.8. ΣΥΝΔΥΑΣΜΟΙ	11
ΚΕΦΑΛΑΙΟ 2	13
ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ ΚΑΙ ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ	13
2.1 ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ	13
2.1.1. ΔΟΚΙΜΗ BERNULLI	13
2.1.2 ΓΕΩΜΕΤΡΙΚΗ ΚΑΤΑΝΟΜΗ	13
2.1.3. ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ	14
2.2 ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ	15
2.3 ΑΛΥΣΙΔΕΣ MARKOV	16
2.4. ΔΙΑΔΙΚΑΣΙΑ ΔΙΑΚΛΑΔΩΣΗΣ	17
ΚΕΦΑΛΑΙΟ 3	20
ΠΡΟΣΟΜΟΙΩΣΗ ΣΥΣΤΗΜΑΤΩΝ	20
3.1 ΠΡΟΣΟΜΟΙΩΣΗ	20
3.2 ΠΡΟΣΟΜΟΙΩΣΗ MONTE CARLO	21
3.2.1 ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ	23
3.3 ΑΛΓΟΡΙΘΜΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ	25
3.3.1. ΑΛΓΟΡΙΘΜΟΣ BOX-MULLER	25
3.3.2. ΔΕΙΓΜΑΤΟΛΗΨΙΑ INVERSE TRANSFORM	28
3.3.3 ΚΑΤΑΣΚΕΥΗ ΚΑΙ ΠΡΟΣΟΜΟΙΩΣΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΔΙΑΚΛΑΔΩΣΗΣ	29
ΚΕΦΑΛΑΙΟ 4	31
ΤΥΧΑΙΟΙ ΠΕΡΙΠΑΤΟΙ	31
4.1 ΟΡΙΣΜΟΣ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	31

4.2 ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΛΟΥ ΤΥΧΑΙΟΥ ΠΕΡΙΠΑΤΟΥ.....	32
4.2.1. ΣΥΜΠΕΡΑΣΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ	34
4.3. ΠΡΟΒΛΗΜΑ ΚΑΤΑΣΤΡΟΦΗΣ ΤΟΥ ΠΑΙΚΤΗ	36
4.4 ΠΡΟΣΟΜΟΙΩΣΗ ΔΙΑΔΙΚΑΣΙΑΣ ΚΑΤΑΣΤΡΟΦΗΣ ΠΑΙΚΤΗ	37
4.4.1 ΠΕΡΙΓΡΑΦΗ ΠΡΟΣΟΜΟΙΩΣΗΣ	39
4.4.2 ΣΤΑΤΙΣΤΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ ΚΑΤΑΣΤΡΟΦΗΣ ΠΑΙΚΤΗ	40
4.5 ΑΞΙΟΛΟΓΗΣΗ ΜΟΝΤΕΛΟΥ ΠΡΟΣΟΜΟΙΩΣΗΣ	40
ΒΙΒΛΙΟΓΡΑΦΙΑ	44

ΚΕΦΑΛΑΙΟ 1

ΠΙΘΑΝΟΤΗΤΕΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ

1.1 ΕΝΝΟΙΑ ΣΤΑΤΙΣΤΙΚΗΣ

Γενικά η στατιστική ανάλυση έχει δύο σκοπούς. Στην πρώτη περίπτωση να βρει μοτίβα σε αμετάβλητες πηγές δεδομένων και σχέσεις που να έχουν κάποιο νόημα και η δεύτερη αφορά την καταπολέμηση της αποφένιας την τάση να ανακαλύπτεις μοτίβα που να αποτελούν προϊόν ανθρώπινης φαντασίας για κάποιες μεταβλητές.

Στην στατιστική μιλάμε για απλούς μηχανισμούς τύχης που παρουσιάζουν παραδείγματα όπως το στρίψιμο ενός νομίσματος, το ρίξιμο ενός ζαριού με αφηρημένα μοντέλα για αυτά. Γενικά υπάρχουν παιχνίδια που εξαρτώνται από το τράβηγμα ενός ή περισσοτέρων καρτών από μία τράπουλα με 52 κάρτες. Αυτά επίσης αποτελούν μηχανισμό τύχης βασισμένα στην κίνηση του τραβήγματος αν τα χαρτιά είναι καλά ανακατεμένα. Μερικές βασικές έννοιες στην στατιστική απεικονίζονται μέσω των παιχνιδιών αυτών[1].

Στο παράδειγμα του δίκαιου ζαριού με X να παίρνει τις τιμές 1, 2, 3, 4, 5, 6 που συμβολίζουν τις πλευρές του ζαριού το $P(x) = \frac{1}{6}$. κάθε περίπτωση. Η αντιστοίχιση αυτή των τιμών με το $P(X)$ αφορά ένα πιθανοτικό μοντέλο για αυτόν τον μηχανισμό τύχης. Με την τιμή $\frac{1}{6}$ που παίρνουν όλες οι πλευρές εννοούμε ότι προκύπτουν ίσα από το ρίξιμο του δίκαιου ζαριού. Επίσης με αυτό το $\frac{1}{6}$ εννοούμε την σχετική συχνότητα εμφάνισης της κάθε πλευράς του ζαριού κατά το ρίξιμο του.

1.2. ΤΥΧΑΙΟΤΗΤΑ

Το ρίξιμο ζαριού το οποίο αποτελεί ένα τυχαίο πείραμα βασίζεται σε έναν ευρετικό κανόνα που ακολουθούν αυτά. Τα τυχαία πειράματα πρέπει να ακολουθούν μια επαναληπτική δομή και με αποτελέσματα που έχουμε προσδιορίσει και δεν μπορούν να προβλεφθούν εκ των προτέρων. Παρόλο που αναφέρουμε τον κανόνα αυτόν θα πρέπει να προσέξουμε κάποια συγκεκριμένα σημεία. Σε ένα τυχαίο πείραμα θα πρέπει να προσέξουμε συγκεκριμένα σημεία:

- Αβεβαιότητα
- Προσδιορισμός των αποτελεσμάτων
- Η επαναληπτικότητα
- Ενέργεια/Δραστηριότητα του πειράματος

Έστω ένα συγκεκριμένο παράδειγμα ριζίματος ζαριού όπου στηρίζεται στους κανόνες αυτούς. Ρίχνοντας το “ενεργούμε” και επίσης παρατηρούμε συγκεκριμένα αποτελέσματα. Έστω αυτά είναι :

Να προσγειωθεί το ζάρι σε μια καρέκλα ή στο πάτωμα. Πετώντας το ζάρι δεν μπορούμε να ξέρουμε εκ των προτέρων ποιο θα εμφανιστεί από τα δύο. Για αυτό λόγο των παραπάνω θεωρείται ως τυχαίο πείραμα. Είναι συχνό μια τυχαία μεταβλητή να αναπαρίστανται μέσω μιας γραμμικής εξίσωσης. Μια γραμμική εξίσωση που είναι της μορφής :

$$Y = b + (a - b)X \quad (1.1)$$

όπου α, β αποτελούν συντελεστές. Σε αυτή την περίπτωση όπου TM η Y, η μεταβλητή X είναι επίσης TM που σημαίνει ότι μπορεί να πάρει μόνο δύο τιμές: Την τιμή 1 με πιθανότητα και την 0 με πιθανότητα 1-p. Όταν X = 1, τότε η Y είναι:

$$Y = b + (a - b)1 \quad (1.2)$$

$$Y = b + a - b \quad (1.3)$$

$$Y = a \quad (1.4)$$

Οπότε όταν X = 1 δηλαδή εκφράζει το ένα αποτέλεσμα ενός πειράματος τότε Y = α. Όταν X = 0 ισχύει:

$$Y = b \quad (1.5)$$

Αυτό σημαίνει ότι η Y εξαρτάται απευθείας από την τιμή της X και η σχέση τους είναι ο γραμμικός συνδυασμός που αναφέραμε.

Έστω τώρα το X = αποτέλεσμα ενός παιχνιδιού και X = 1 αν νίκη και 0 αν ήττα.

Υπάρχουν περιπτώσεις ανθρώπων που πιστεύουν αυτά τα (τυχαία) αποτελέσματα ακολουθούν κάποια μοτίβα με αποτέλεσμα να μην υπάρχει ανεξαρτησία. Η πεποίθηση αυτή είναι γνωστή ως πλάνη του τζογαδόρου.

1.3. ΠΛΑΝΗ ΤΟΥ ΤΖΟΓΑΔΟΡΟΥ

Υπάρχει μια παρανόηση όσον αφορά τους μηχανισμούς τυχαιοποίησης από την πλευρά του παίκτη (σε μια περίπτωση παιχνιδιού ρουλέτας). Υπάρχει η αίσθηση ότι κάποια νούμερα ή χρώματα σε μια δοκιμή (συνεχιζόμενα παιχνίδια) έχουν μεγαλύτερη πιθανότητα εμφάνισης (για παράδειγμα μεγαλύτερη από 1/2) οπότε παρατηρείται διάφοροι παίκτες να σημειώνουν χρώματα ή τα “αγαπημένα τους νούμερα. Δηλαδή αν έχει έρθει συνεχόμενα μαύρο υπάρχει η πεποίθηση ότι θα έρθει το κόκκινο σε μια επόμενη δοκιμή λόγω μιας υποκειμενικής κατάστασης ισορροπίας. Ουσιαστικά τα άτομα βασίζονται στην λογική της ευρετικής αντιπροσώπευσης όπου τα άτομα δεν αντιλαμβάνονται την τυχαία φύση της ρουλέτας για την παραγωγή τυχαίων αποτελεσμάτων αλλά με βάση προηγούμενα αποτελέσματα (που αποτελούν μεμονωμένα περιστατικά στην πραγματικότητα)τα οποία θεωρούνται “αντιπροσωπευτικά”, παίρνουν ανάλογες αποφάσεις[2]. Για παράδειγμα, αν ένα άτομο ακούσει για μια σειρά από περιπτώσεις που ένας επενδυτής έκανε μεγάλα κέρδη στην αγορά μετοχών, μπορεί να υποθέσει ότι η επένδυση στην αγορά μετοχών είναι συνήθως κερδοφόρα, παρά το γεγονός ότι δεν έχει λάβει υπόψη του τη συνολική στατιστική στην αγορά.

Η ευρετική αντιπροσώπευση μπορεί να οδηγήσει σε λάθη στην εκτίμηση πιθανοτήτων καθώς τα άτομα μπορεί να αγνοήσουν σημαντικά στατιστικά δεδομένα ή να υπερεκτιμήσουν τη σημασία μεμονωμένων περιστατικών που φαίνονται "αντιπροσωπευτικά" αλλά δεν είναι πραγματικά αντικειμενικά ενδεικτικά της πραγματικότητας.

Έστω ότι ένας ερευνητής θέλει να μετρήσει την επίδραση της νίκης ή της ήττας στην ψυχολογία των παικτών. Χρησιμοποιεί την μεταβλητή για να καταγράψει την ευτυχία τους μετά το παιχνίδι όπου α στην εξίσωση μεγάλη ευτυχία και b μικρή ευτυχία. Έτσι αν X = 1 τότε Y = α δηλαδή θα έχουμε μεγάλη ευτυχία.

Αυτό μας επιτρέπει να μετατρέπουμε ένα απλό “ναι” ή “όχι” αποτέλεσμα σε μια πιο ποσοτική μέτρηση που μπορεί να αναλυθεί στατιστικά.

Στον μαθηματικό κόσμο γενικά το αντικείμενο της τυχαία μεταβλητής αποτελεί η ανάθεση τιμών που να αντικατοπτρίζουν το αποτέλεσμα ενός πειράματος. Επιστρέφοντας στο παράδειγμα με το στρίψιμο ενός νομίσματος μια μεταβλητή θα θεωρείται Τυχαία όταν θα πάρει τις τιμές για παράδειγμα $X=1$ εάν στο αποτέλεσμα έχει έρθει Κορώνα ή θα ναι $X=0$ εάν είναι Γράμματα. Η X δηλαδή μπορεί να εκφραστεί ως ο αριθμός της Κορώνας στο στρίψιμο που κάνουμε (Στο πρώτο στρίψιμο). Κατά την αρχική παρατήρηση του πραγματικού κόσμου στο παράδειγμα ανατίθεται στο X του προβλήματος, τα γράμματα Η και Τ που σημαίνουν Heads (Κορώνα) ή Τ (Γράμματα). Έστω τώρα ότι κάνουμε 10 επαναλήψεις του στρίψιματος ενός νομίσματος και έρχονται τυχαία τα αποτελέσματα:

H, H, T, H, T, H, H, H, T, H,

Το X θα είναι δηλαδή :

1, 1, 0, 1, 0, 1, 1, 1, 0, 1

Που σημαίνει αν πάρουμε το άθροισμα αυτής της ακολουθίας του στρινίματος δηλαδή το $\sum X$ αυτό ισούται με: $\sum X = 7$ που σημαίνει 7 φορές Κορώνα σε 10 στρινίματα.

Όσον αφορά τον ορισμό της Κατανομής Πιθανότητας μιας Τυχαίας Μεταβλητής, αφορά την αντιστοίχιση πιθανών τιμών της Τυχαίας Μεταβλητής με τιμές των πιθανοτήτων που κατανέμονται σε αυτές. Αθροίζοντας όλες τις τιμές πιθανοτήτων για τις αντίστοιχες τιμές Τυχαίας Μεταβλητής το αποτέλεσμα θα είναι $\sum P(X)=1$.

Μιλώντας για τον ορισμό της μέσης τιμής μιας τυχαίας μεταβλητής αναφερόμαστε στο άθροισμα των πιθανών αποτελεσμάτων επί τις πιθανότητες τους. Ουσιαστικά αποτελεί ένας σταθμισμένος μέσος όρος όπου πολλαπλασιάζεται η τιμή της τυχαίας μεταβλητής με το “βάρος” της δηλαδή την αντίστοιχη πιθανότητα. Ο συμβολισμός της είναι :

$$E(X) = \sum xP(x) \quad (1.6)$$

Ένα παράδειγμα με βάση την προηγούμενη γραμμική εξίσωση $Y = b + (a-b) X$ όπου η Y εξαρτάται από τις τιμές της X , για την μέση τιμή της θα ισχύει το εξής:

$$E[Y] = ap + b(1 - p) \quad (1.7)$$

Για το παραπάνω ισχύει η ιδιότητα της γραμμικότητας της αναμενόμενης τιμής δηλαδή:

$$E[Y] = E[b] + E[(a - b)X] \quad (1.8)$$

όπου a, b σταθερές.

Άρα η αναμενόμενη τιμή της $E[b]$ είναι η b η σταθερά καθώς και $a-b$ για την $E[(a-b)]$. Οπότε η αναμενόμενη τιμή του γραμμικού συνδυασμού των τυχαίων μεταβλητών αυτών είναι ίση με τον γραμμικό συνδυασμό των αναμενόμενων τυχαίων μεταβλητών. Όταν μας ζητάνε τον μέσο (εννοώντας την μέση τιμή) μπορεί συχνά να συναντήσουμε το μ που είναι το ίδιο με το $E(X)$.

Η μέση τιμή αποτελεί παράμετρος (δηλαδή ένας αριθμός) για τον χαρακτηρισμό ενός πιθανοτικού μοντέλου όμως δεν είναι αρκετή από μόνη της. Έστω ένα μη δίκαιο ζάρι όπου ισχύει το εξής:

X	1	2	3	4	5	6
P(X)	0.3	0.1	0.1	0.1	0.1	0.3

Πίνακας 1.1: Παράδειγμα πιθανοτήτων - Τιμών TM

Παρατηρούμε ότι οι μεγαλύτερες τιμές της συνολικής πιθανότητας όπου $\sum P(X)=1$ για κάθε x , βρίσκονται στις ακραίες τιμές της. Οπότε μας ενδιαφέρει ένα στατιστικό μέτρο που να δείχνει τον βαθμό διασκορπισμού των τιμών από τον μέσο όρο. Αυτό γίνεται γράφοντας τον υπολογισμό X-μ που δείχνει την απόκλιση της τιμής της τυχαίας μεταβλητής από τον μέσο[1]. Για να μην βγάζουμε πολλές αποκλίσεις, είναι συνετό να συνδυάσουμε τις τιμές αυτές αθροίζοντας δηλαδή:

$$\sum X - \mu P(X) \quad (1.9)$$

Πολλαπλασιάζοντας με το βάρος της πιθανότητας κάθε τιμής.

Όμως από μόνο του δεν είναι αρκετό αυτό καθώς για λόγους ευανάγνωστους και για θεωρήσεις που δεν είναι σχετικές, στους υπολογισμούς των αποκλίσεων δεν μας ενδιαφέρουν τα πρόσημα που θα προκύψουν δηλαδή αν το X είναι μεγαλύτερο ή μικρότερο του μ στην ποσότητα X-μ. Για αυτό παίρνουμε την απόλυτη τιμή της απόκλισης των τιμών:

$$\sum |X - \mu| P(X) \quad (1.10)$$

Όπου μέσω αυτού προκύπτει για την Διασπορά της τυχαίας μεταβλητής η φόρμουλα:

$$\sum (X - \mu)^2 \quad (1.11)$$

Πριν περάσουμε στην πιθανότητα το P(X) να αποτελεί την πιθανότητα της τυχαίας μεταβλητής, θα αναφέρουμε τον συμβολισμό P(A) όπου «A» λέγεται γεγονός και είναι ένα σύνολο από πιθανά αποτελέσματα.

Οι τρεις κανόνες πιθανότητας για ένα γεγονός είναι οι εξής:

1. $P(not A) = 1 - P(A)$
2. $P(A OR B) = P(A) + P(B) - P(A AND B)$,
3. $P(A AND B) = P(A|B)P(B)$.

Ο τελευταίος κανόνας αφορά την δεσμευμένη πιθανότητα ενός γεγονότος με βάση ένα άλλο. Με αυτό εννοούμε ότι κάνουμε τη πρόβλεψη για το γεγονός μας A(την πιθανοτική πρόβλεψη) έχοντας δεδομένο ότι έχει γίνει το γεγονός B. Η εμφάνιση του B αποτελεί πληροφορία σχετική με το A. Για παράδειγμα έστω ότι τραβάμε δύο κάρτες από μια τράπουλα [1]. Έστω H1 και H2 γεγονότα που δείχνουν το τράβηγμα μιας κάρτας καρδιά από την τράπουλα 52 καρτών. Συγκεκριμένα το H2 εκφράζει η δεύτερη κάρτα να είναι καρδιά[1]. Η πιθανότητα αυτή υπολογίζεται ως εξής: $P(H2/H1) = \frac{12}{51} \cdot \frac{13}{52}$. Το αποτέλεσμα αυτό

καθορίζεται από το γεγονός του πρώτου τραβήγματος και επειδή το σύνολο των καρδιών αποτελούν 13/52 τότε ισχύει στην πρόβλεψη του $H_2 = \frac{12}{51}$ καθώς απομένουν 12.

1.4. ΑΝΑΠΑΡΑΣΤΑΣΗ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα αναπαρίστανται με Γραφήματα για την σύνοψη τους και για τα διάφορα συμπεράσματα που εξάγονται από αυτά. Μία αρχή που αντιστοιχεί σε αυτά είναι ότι οι αριθμητικές ποσότητες δεν χαρακτηρίζονται μόνο σε σχέση με το ύψος αλλά και από τον χώρο[1]. Ένα παράδειγμα που το ανθρώπινο μάτι τείνει να επικεντρώνεται αφορά στην απεικόνιση του διαγράμματος δύο ορθογωνίων όπου το πρώτο έχει ύψος 2 και το δεύτερο 4. Σε αυτή την περίπτωση μας ενδιαφέρει πέρα από το ύψος του σχήματος αλλά και το εμβαδόν (δηλαδή ο χώρος) που καταλαμβάνει.

Όσον αφορά για την συλλογή δεδομένων σε ένα πείραμα που θέλουμε να μελετήσουμε, είναι σημαντικό να θεωρήσουμε την ερώτηση από που προέκυψαν αυτά τα δεδομένα ή αν τα σύλλεξε μόνος του αυτός που κάνει το πείραμα όπου είναι μια διαδικασία που μπορεί να αποβεί χρονοβόρα. Γενικά υπάρχουν τρόποι διεξαγωγής πειραμάτων για την συλλογή δεδομένων όπως αν κάνει κάποιος μια δημοσκόπηση ρωτώντας ένα δείγμα ανθρώπων (δηλαδή ένα μικρότερο γκρουπ) από έναν πληθυσμό για κάποια χαρακτηριστικά τους. Ένα τέτοιο παράδειγμα αποτελεί ένα τυχαίο πείραμα όπου παράγει τυχαία αποτελέσματα. Όμως τα δεδομένα μπορεί να αποβούν ασήμαντα αν είναι για ένα χαρακτηριστικό μια παράθεση αριθμών χωρίς να λαμβάνονται περαιτέρω άλλα χαρακτηριστικά-πληροφορίες που να καθορίζουν το ερώτημα που θέλουμε. Ένα παράδειγμα που βασίζεται στο παράδοξο του Σίμσον είναι το εξής: Έστω ένα σύνολο από δεδομένα, συγκεκριμένα 340 για ασθενείς που ακολουθούν μια αγωγή. Χωρίζουμε σε δύο υποομάδες τους ασθενείς δηλαδή:

Η ομάδα πρώτης αγωγής αποτελείται από 160 ασθενείς όπου 60 ασθενείς θα ανακάμψουν και στην δεύτερη ομάδα αγωγής (340-160=230) θα ανακάμψουν 65.

Παίρνοντας το ποσοστό ανάκαμψης για τις δύο ομάδες ισχύει ότι μεγαλύτερο ποσοστό ανάκαμψης έχει η πρώτη (καθώς $\frac{60}{160} > \frac{65}{230} = 0,37\% > 0,28\%$). Παίρνοντας το συνολικό ποσοστό $\frac{60}{160} + \frac{65}{230} = \frac{125}{390} = 32\%$, όπου συνειδητοποιούμε ότι το συνολικό ποσοστό ανάκαμψης είναι διαφορετικό από τα προηγούμενα. Ένα αίτιο για αυτό μπορεί να είναι ότι δεν λάβαμε άλλα χαρακτηριστικά υπόψιν όπως για παράδειγμα της σοβαρής αρχική κατάσταση του ασθενή όπου στην πρώτη περίπτωση μπορεί να έχει ασθενείς με πχ. A_1 =Σοβαρότητα νόσου υψηλή και στην δεύτερη A_2 =Σοβαρότητα νόσου χαμηλή. Αυτό το χαρακτηριστικό μπορεί να επηρεάσει το παράδοξο αποτέλεσμα και για αυτό είναι σημαντικό να διεξάγεται λεπτομερής ανάλυση των δεδομένων, ιδίως όταν υπάρχουν πολλαπλοί παράγοντες που ενδέχεται να επηρεάζουν τα αποτελέσματα.

1.5. ΔΕΙΓΜΑ

Ο ορισμός του τυχαίου δείγματος μιας κατανομής αναφέρεται σε ένα πείραμα μια σειρά από αποτελέσματα ενός πειράματος τα οποία εκφράζονται ως σύνολο με τιμές. Για παράδειγμα στο ρίξιμο του ζαριού όπου η τυχαία μεταβλητή που μοντελοποιούμε δηλώνει το πάνω μέρος του ζαριού σαν αποτέλεσμα (στο ρίξιμο) τότε ένα σύνολο θα μπορούσε να ναι το εξής: (2,1,5,6,2). Στην περίπτωση αυτή μιλάμε και για το μέγεθος του δείγματος όπου είναι $n=5$ στις 5 ανεξάρτητες επαναλήψεις του πειράματος αυτού. Με τον όρο ανεξαρτησία εννοούμε σε περίπτωση που ένα αποτέλεσμα δεν επηρεάζεται από ένα άλλο. Στην περίπτωση μας για παράδειγμα έστω το ζητούμενο η πιθανότητα

1.6. ΠΛΗΘΥΣΜΟΣ

Η διαδικασία λήψης ενός δείγματος από ένα πληθυσμό μας είναι ήδη γνωστή από μια δημοσκόπηση γνώμης και από μια έρευνα πάνω στους ψηφοφόρους. Πέρα από την λήψη αυτή πρώτα θα πρέπει να καθορίσουμε τον πληθυσμό καθ' αυτόν όπου αυτό θα γίνει κάνοντας τις κατάλληλες ερωτήσεις για να τον προσδιορίσουμε με σαφήνεια. Παραδείγματα πληθυσμού θα μπορούσαν να ήταν οι εγγεγραμμένοι ψηφοφόροι μιας περιοχής καθώς επίσης ένα προϊόν μιας ημέρας από την γραμμή παραγωγής. Δηλαδή ένας πληθυσμός θα μπορούσε να μην περιέχει άμεσα στην έρευνα τον ανθρώπινο παράγοντα. Ένας πληθυσμός υπό εξέταση θα πρέπει να είναι σαφής στον προσδιορισμό του για το τι αντικείμενα περιλαμβάνει. Επίσης στο παράδειγμα των εγγεγραμμένων επειδή οι ψηφοφόροι μπορούν να αλλάζουν από ώρα σε ώρα ή στιγμή θα πρέπει να προσδιορίσουμε κι άλλους παράγοντες για να είναι πιο ακριβής[1]. Ένα ανάλογο παράδειγμα ερώτησης θα μπορούσε να ήταν:

“Ποιο ποσοστό ανθρώπων ηλικίας 18-30 αθλούνται συχνά; Ή σε τι αναλογία άνθρωποι πάνω από 60 αθλούνται συχνά.”

Στην περίπτωση του δείγματος από έναν πληθυσμό λέμε για μέγεθος δείγματος όταν παίρνουμε ένα υποσύνολο από τον πληθυσμό που θα εξετάσουμε. Για το πείραμα τυχαίου δείγματος έστω ότι σε μια τράπουλα με 52 χαρτιά η οποία αποτελεί τον πληθυσμό μας και τραβάμε 5 κάρτες στο χέρι μας, αυτό το τράβηγμα αποτελεί το δείγμα μας το οποίο είναι τυχαίο καθώς όταν το επανατοποθετήσουμε πίσω και έχει ανακατευτεί καλά η τράπουλα, δεν γνωρίζουμε από πριν ποιες κάρτες συγκεκριμένα θα τραβήξουμε σε ένα επόμενη επανάληψη[1].

Γενικά θέλουμε για την παραγωγή τυχαίων δειγμάτων να προέρχονται από έναν τυχαίο μηχανισμό. Για παράδειγμα στην περίπτωση της τράπουλας και του δείγματος που αναφέραμε ο μηχανισμός είναι το ανακάτεμα της. Προκύπτει όμως θέμα πολλές φορές κατά πόσο ο μηχανισμός είναι τυχαίος. Σε μια περίπτωση δειγματοληψίας όπου για τους συμμετέχοντες μιας έρευνας, καθορίζονται ποσοστά-στόχοι(“quota”), μπορούν να εμφανιστούν στρεβλώσεις στα αποτελέσματα. Ένα σχετικό παράδειγμα αποτελούν τα αποτελέσματα παλιών δημοσκοπήσεων στην Αμερική για τον Πρόεδρο όπου δημοσκοπικές εταιρίες έδιναν μεγάλο προβάδισμα για νίκη στον εν τέλη ηττημένο της διαδικασίας. Αυτές οι ανακρίβειες προέκυψαν καθώς αυτή η διαδικασία της δειγματοληψίας στηριζόταν κατά πολύ στην ανθρώπινη κρίση και όχι στην τυχαία επιλογή της έρευνας καθώς διάφορες προκαταλήψεις και ανθρώπινες επιλογές καθόρισαν ένα μη αντιπροσωπευτικό δείγμα[1].

1.7. ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΑ ΔΙΑΚΡΙΤΑ ΜΑΘΗΜΑΤΙΚΑ

Ο κανόνας της μέτρησης βασίζεται στην πολλαπλασιαστική αρχή. Για παράδειγμα έχουμε να κάνουμε κάτι με m τρόπους και κάτι άλλο με n τρόπους. Δηλαδή έστω ότι μπορούμε να πάμε από την δουλειά σε ένα χώρο αναψυχής με 3 τρόπους και από τον χώρο αναψυχής στο σπίτι με 2 τρόπους. Άρα ισχύει $3 \times 2 = 6$ διαφορετικοί τρόποι συνδυασμών.

1.8. ΣΥΝΔΥΑΣΜΟΙ

Για την πλήρη κατανόηση του ορισμού της διωνυμικής κατανομής γίνεται αναφορά σε στοιχεία συνδυαστικής σε πιθανοτικά προβλήματα για την κατανόηση του. Έστω ότι το πείραμα μας είναι το

τράβηγμα 5 καρτών από μια τράπουλα 52 καρτών όπου 13 κάρτες είναι ίδιου χρώματος και έχουμε συνολικά 4 διαφορετικά χρώματα. Ο χώρος καταστάσεων δηλαδή το σύνολο των δυνατών αποτελεσμάτων S στην προκειμένη περίπτωση ο αριθμός των δυνατών (διαφορετικών ομάδων 5αδων) συνδυασμών μιας 5αδας καρτών, ισούται με:

$$|S| = \binom{52}{5} \quad (1.12)$$

όπου δηλώνει (το διωνυμικό ανάπτυγμα το πλήθος των συνδυασμών 5 καρτών από 52 κάρτες.

την πιθανότητα ενός γεγονότος να έχουμε τράβηγμα 5αδής αποτελούμενη από: 3 ρηγάδες και 2 διαφορετικές κάρτες βρίσκεται ως εξής:

$$P(E) = \frac{1}{\binom{52}{5}} \quad (1.13)$$

Καθώς αυτό το αποτέλεσμα συνδυασμών καρτών είναι μοναδικό σε σχέση με το σύνολο των δυνατών συνδυασμών[3].

ΚΕΦΑΛΑΙΟ 2

ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ ΚΑΙ ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ

2.1 ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ

2.1.1. ΔΟΚΙΜΗ BERNOULLI

Όταν αναφερόμαστε σε ένα πείραμα δοκιμής Bernoulli, εννοούμε ένα πείραμα με δύο δυνατά αποτελέσματα όπου το αποτέλεσμα που μας ενδιαφέρει το αποκαλούμε επιτυχία και συμβολίζεται ως $P(E)=p$ και το άλλο ως αποτυχία και συμβολίζεται με $P(A)=1-q$ (η πράξη αυτή είναι καθώς δεν έχουμε άλλες πιθανότητες στον ορισμό μας). Για να γίνουμε πιο συγκεκριμένοι μέσω ενός παραδείγματος ας αναφέρουμε την παρατήρηση ενός φαναριού σε έναν συγκεκριμένο χρόνο. Η ΤΜ μπορεί να οριστεί ως : Η ένδειξη ή το χρώμα που εναλλάσσεται με τυχαίο τρόπο. Οπότε αν θέλουμε την πιθανότητα να είναι κόκκινο το φανάρι αυτό θα γίνει για παράδειγμα: $P(X=\text{κόκκινο το φανάρι}) = P(X=1) = p$, όπου p η πιθανότητα επιτυχίας.

2.1.2 ΓΕΩΜΕΤΡΙΚΗ ΚΑΤΑΝΟΜΗ

Μια γεωμετρική ΤΜ μπορεί να οριστεί ως ο αριθμός των ρίψεων μέχρι την πρώτη επιτυχία. Έστω για παράδειγμα το X =ο αριθμός των ατόμων που εξετάζονται μέχρι να βρεθεί αυτός που έχει ένα σπάνιο τύπο αίματος. Αυτά τα πειράματα έχουν δύο δυνατά αποτελέσματα(το καθένα όπου τα δεδομένα να έχει τον σπάνιο τύπο ή να μην τον έχει)και είναι ανεξάρτητα και επαναλαμβανόμενα πειράματα Bernoulli. Στο παράδειγμα που αναφέραμε έστω οι δοκιμές είναι τρεις ως όπου βρέθηκε ως συγκεκριμένος τύπος δηλαδή: $P(X=3) = P(\text{"ο 1 να μην το έχει"}) \cdot P(\text{"ο 2 να μην το έχει"}) \cdot P(\text{"να ο τύπος αίματος"}) = (1 - p) * (1 - p) * p = q^2 \cdot p$.

Στην περίπτωση αυτή ισχύει η πολλαπλασιαστική αρχή για τις πιθανότητες λόγω της ανεξαρτησίας που υπάρχει σε κάθε πείραμα.

Οι γενικοί τύποι για την κατανομή αυτή έχουν ως εξής:

$$(2.1) \quad P(X = x) = q^{x-1} \cdot p$$

$$\mu = \frac{1}{p} \quad (2.2)$$

$$\sigma^2 = \frac{q}{p^2} \quad (2.3)$$

2.1.3. ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

Σε αυτή την περίπτωση θα δωθεί έμφαση σε αυτή την κατανομή λόγω της μετέπειτα μοντελοποίησης μας στο πρόβλημα του Απλού Τυχαίου Περιπάτου. Όπως και η ΓΚ έτσι και η ΔΚ βασίζεται στις ανεξάρτητες επαναλήψεις μιας δοκιμής Bernoulli όπου έχουμε τα αποτελέσματα “επιτυχία” και “αποτυχία”. Στην περίπτωση αυτή για την ΔΚ έχουμε έναν αριθμό n που δηλώνει τον αριθμό των επαναλήψεων του πειράματος. Η διωνυμική μεταβλητή για την μοντελοποίηση πραγματικών προβλημάτων θα ορίζεται ως:

Ο αριθμός των επιτυχιών σε « n » ανεξάρτητες δοκιμές

Μπορούμε επίσης να το πούμε ως και το ενδεχόμενο για παράδειγμα $\{x$ επιτυχίες σε n δοκιμές $\}$. Μια “πραγματική” μοντελοποίηση μιας κατάστασης θα μπορούσε να είναι η παραγωγή 1500 ηλεκτρικών εξαρτημάτων σε μια ημέρα [1, σ. 105]. Αυτή η κατάσταση μπορεί να “μιμηθεί” το στρίψιμο 1500 δίκαιων νομισμάτων, πειράματα που βασίζονται στο μοντέλο της ΔΚ.

Μια άλλη περίπτωση που εφαρμόζεται το διωνυμικό μοντέλο μπορεί να ζητάμε την πιθανότητα του αποτελέσματος μιας βολής στο μπάσκετ. Το αν έβαλε ή έχασε την βολή, στην επόμενη βολή δεν έχει επιρροή οπότε οι επαναλήψεις είναι ανεξάρτητες. Με βάση ακόμα και με αυτό το παράδειγμα δηλαδή τις n εκτελέσεις βολών καταλαβαίνουμε ότι η διαφορά από την δοκιμή Bernoulli είναι ο συνολικός αριθμός n των επαναλήψεων των δοκιμών.

Όπως έχουμε αναφέρει η ΤΜ μοντελοποιείται με βάση τα αποτελέσματα του πειράματος. Έστω στο πείραμα με το μπάσκετ με τα αποτελέσματα να ναι ως ακολουθίες επιτυχιών ή αποτυχιών ως εξής:

A, E, E, A, E, A, E, E, E, E

Όπου η ΤΜ θα ναι ο αριθμός των επιτυχιών δηλαδή $X = 7$ στο σύνολο δοκιμών $n = 10$. Οπότε για συντομογραφία θα γράφουμε ότι $X \sim B(n, p)$ με n και p παραμέτρους που έχουμε αναφέρει. Να γενικεύσουμε όπως στο παράδειγμα με το μπάσκετ ότι η πιθανότητα επιτυχίας p παραμένει ίδια σε κάθε δοκιμή λόγω της ανεξαρτησίας όπου κάθε αποτέλεσμα δεν επηρεάζει ένα άλλο με το p να είναι σταθερό.

Η ΔΜ μπορεί να προκύψει από την περίπτωσης της κατανομής Bernoulli. Έστω το άθροισμα

$$X = X_1, X_2, X_3, \dots, X_n,$$

που η κάθε περίπτωση της X_i ακολουθεί την ΒΜ με πιθανότητα επιτυχίας p να πάρει την τιμή x και $1-p$ να είναι μηδέν. Το άθροισμα αυτό είναι ο αριθμός των επιτυχιών στις n (ανεξάρτητες) δοκιμές με p .

Η εξίσωση της διωνυμικής κατανομής έχει ως εξής:

$$P(X = x) = \binom{n}{x} p^x q^{1-x} \quad (2.4)$$

Ο αριθμός των επιτυχιών της ΤΜ αυτής αν θέλουμε να δείξουμε σε πιο ενδεχόμενο αντιστοιχείται είναι το εξής: $\{x$ επιτυχίες σε n δοκιμές $\}$. Αυτό αποτελεί την ένωση των A_i ενδεχομένων που είναι ακολουθίες από x επιτυχίες και $n-x$ αποτυχίες δηλαδή $\{x$ επιτυχίες σε n δοκιμές $\} = \cup_{i=1}^x A_i$. Για παράδειγμα έχουμε $n=3$

δοκιμές και $A_1=(1,0,0)$, $A_2=(0,1,0)$ και $A_3=(0,0,1)$ όπου φαίνονται τα σύνολα με τους διαφορετικούς τρόπους που διατάσσονται οι επιτυχίες '1'. Τότε το $\Sigma P(A_i)$,είναι ίσο με την εξίσωση της ΔΜ.

Η μέση τιμή και η διακύμανση της $X \sim B(n,p)$ είναι:

$$E(X) = n \cdot p \quad (2.5)$$

$$Var(X) = n \cdot p \cdot q. \quad (2.6)$$

Το αποτέλεσμα της $E(X)$ προκύπτει από το άθροισμα της ΤΜ X που ακολουθεί διωνυμική κατανομή. Να θυμίσουμε: $X = X_1 + X_2 + X_3 \dots + X_n \sim B(n, p)$. Άρα ισχύει:

$E(X) = E(X_i) = E(X_1, X_2, X_3, \dots, X_n)$ και επειδή είναι ξένα μεταξύ τους ισχύει ότι

$E(X) = E(X_i) = E(X_1) + E(X_2) + \dots E(X_n)$. Επειδή είναι πολλά bernoulli πειράματα και η μέση τιμή Bernoulli είναι $E(X) = p$ άρα το αποτέλεσμα είναι το παραπάνω. Παρόμοια διαδικασία ακολουθείται και για την διακύμανση όπου

$$Var(X_i) = pq \quad (2.7)$$

$$\sigma^2 = Var(X) = Var(X_1 + X_2 + X_3 \dots, X_n) = Var(X_1) + Var(X_2) + \dots, Var(X_n) \quad (2.8)$$

2.2 ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ

Οι Papoulis και Pillai (2007) υποστηρίζουν:

...Μία ΤΜ x είναι ένας κανόνας ανάθεσης ενός αριθμού $x(\zeta)$ σε κάθε έκβαση ζ του πειράματος S . Μια στοχαστική διαδικασία $x(t)$ είναι ένας κανόνας αντιστοίχισης μιας συνάρτησης $x(t, \zeta)$ σε κάθε ζ . Έτσι, μια στοχαστική διαδικασία είναι μια οικογένεια από χρονικές συναρτήσεις που εξαρτώνται από την παράμετρο ζ ή, ισοδύναμα, μία συνάρτηση των t και ζ . Το πεδίο τιμών για το ζ είναι το σύνολο όλων των πειραματικών εκβάσεων και το πεδίο τιμών για το t είναι ένα σύνολο R από πραγματικούς αριθμούς.

Εάν το R είναι ο πραγματικός άξονας, τότε η $x(t)$ είναι μια διαδικασία συνεχούς χρόνου. Εάν το R είναι ένα σύνολο από ακραίους, τότε η $x(t)$ είναι μια διαδικασία διακριτού χρόνου. Μια διαδικασία διακριτού χρόνου είναι, επομένως, μία ακολουθία από ΤΜ. Μία τέτοια διαδικασία θα συμβολίζεται με $x_n, \dots$, ή προκειμένου να αποφευχθούν οι διπλοί δείκτες, με $x[n]$.

Θα λέμε ότι η $x(t)$ είναι μια διαδικασία διακριτής κατάστασης όταν οι τιμές της είναι μετρήσιμες. Διαφορετικά, είναι μια διαδικασία συνεχούς κατάστασης.

.... Θα χρησιμοποιήσουμε το συμβολισμό $x(t)$ για να αναπαραστήσουμε μία στοχαστική διαδικασία παραλείποντας, όπως κάναμε και στην περίπτωση των ΤΜ, την εξάρτηση της από το ζ . Έτσι η $x(t)$ έχει τις ακόλουθες ερμηνείες:

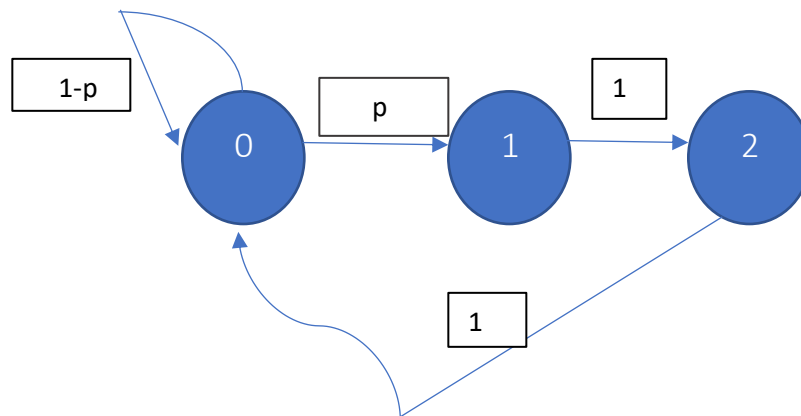
1. Αποτελεί μία οικογένεια (ή ένα σύνολο) συναρτήσεων $x(t, \zeta)$. Σύμφωνα με την ερμηνεία αυτή, τα t και ζ είναι μεταβλητές.

2. Είναι μόνο συνάρτηση του χρόνου (ή ένα δείγμα μιας γνωστής διαδικασίας). Στην περίπτωση αυτή, το t είναι μεταβλητή και το ζ είναι σταθερό.
3. Εάν το t είναι σταθερό και το ζ είναι μεταβλητή, τότε η $x(t)$ είναι ΤΜ που αντιστοιχεί στην κατάσταση μιας γνωστής διαδικασίας το χρόνο t .
4. Εάν τα t και ζ είναι σταθερά, τότε η $x(t)$ είναι ένας αριθμός.

Για να κατανοηθεί καλύτερα, ένα απλοποιημένο παράδειγμα αφορά την μοντελοποίηση ενός μηχανήματος.

Έστω μηχάνημα στην αρχή κάθε περιόδου όπου είτε **λειτουργεί** είτε είναι **χαλασμένο**. Αν λειτουργεί με πιθανότητα p χαλάει στην διάρκεια της περιόδου. Αν χαλάσει (μπαίνει σε διαδικασία επισκευής) η επισκευή διαρκεί 2 περιόδους. Έστω στοχαστική ανέλιξη $\{X_n, n = 0, 1, \dots\}$ όπου n η συγκεκριμένη ημέρα. Παρατηρώντας το σύστημα στην αρχή της περιόδου (χρόνου) σημαίνει ότι είναι διακριτού χρόνου και με βάση τα παραπάνω, αποτελείται από τρεις καταστάσεις. Η μοντελοποίηση του χώρου καταστάσεων αναπαρίστανται ως εξής: $X_n = \{0, 1, 2\}$

Όπου στην πρώτη λειτουργεί κανονικά στην αρχή της περιόδου n . Στην επόμενη κατάσταση σημαίνει ότι έχει πάθει βλάβη και είναι στην 1η μέρα επισκευής και στην τελευταία ότι έχει πάθει βλάβη και βρίσκεται στην 2η μέρα επισκευής.



Εικόνα 2.1: Διάγραμμα μεταβάσεων για το παράδειγμα της στοχαστικής διαδικασίας που δείχνει την επικοινωνία των καταστάσεων μέσω των πιθανοτήτων τους.

2.3 ΑΛΥΣΙΔΕΣ MARKOV

Οι διαδικασίες Markov αντιπροσωπεύουν την απλούστερη γενίκευση των ανεξάρτητων διαδικασιών όπου η έκβαση του πειράματος κάθε χρονική στιγμή εξαρτάται μόνο από την αμέσως προηγούμενη έκβαση και από καμία πριν από αυτή. Συνεπώς, σε μια διαδικασία Markov $x(t)$, εφόσον το παρόν είναι καθορισμένο, το παρελθόν δεν έχει καμία επίδραση στο μέλλον. Αυτό σημαίνει ότι αν $t_{n-1} < t_n$, από την (15-1) συνεπάγεται ότι αν

$$t_1 < t_2 < \dots < t_n, \text{ τότε } P[x(t_n) \leq x_n | x(t_{n-1}), \dots, x(t_1)] = P[x(t_n) \leq x_n | x(t_{n-1})]$$

Η αλυσίδα Markov είναι μια ειδική διαδικασία Markov όπου το σύστημα μπορεί να καταλαμβάνει ένα πεπερασμένο ή ένα μετρήσιμο άπειρο πλήθος καταστάσεων $e_1, e_2, \dots, e_j, \dots$, έτσι ώστε μελλοντική εξέλιξη της διαδικασίας, όταν αυτή βρεθεί σε συγκεκριμένη κατάσταση, να εξαρτάται αποκλειστικά από την παρούσα κατάσταση και όχι από το πώς έφθασε σε αυτή. Τόσο οι αλυσίδες Markov όσο και οι διαδικασίες Markov μπορούν να είναι διακριτού χρόνου ή συνεχούς χρόνου, ανάλογα με το αν ο δείκτης του χρόνου είναι διακριτός ή συνεχής[4].

Σε μια αλυσίδα Markov διακριτού χρόνου $\{x_n\}$ με πεπερασμένο ή άπειρο σύνολο από καταστάσεις $e_1, e_2, \dots, e_i, \dots$, έστω $x_n = x(t_n)$ κατάσταση του συστήματος στο $t = t_n$ Εάν $t_n = nT$, τότε για $n \geq m \geq 0$, η ακολουθία $x_m \rightarrow x_{m+1} \rightarrow \dots x_n, \dots$ αναπαριστά την εξέλιξη του συστήματος. Έστω ότι $p_i(m) = P\{x_m = e_i\}$

Αντιπροσωπεύει την πιθανότητα τη χρονική στιγμή $t = t_m$ το σύστημα να βρίσκεται στην κατάσταση e_i και $p_{ij}(m, n) = P\{x_n = e_j | x_m = e_i\}$ πιθανότητα το σύστημα να μεταβεί στην κατάσταση e_j στο $t = t_n$ δεδομένου ότι τη χρονική στιγμή $t = t_m$ βρισκόταν στην κατάσταση e_i (ανεξάρτητα από τη συμπεριφορά της πριν από το t_m). Οι αριθμοί $p_{ij}(m, n)$ εκφράζουν τις πιθανότητες μετάβασης της αλυσίδας Markov από την κατάσταση e_i τη χρονική στιγμή t_m στην e_j τη χρονική στιγμή t_n .

Θα παρουσιάσουμε ένα παράδειγμα ΜΑΔΧ για καλύτερη κατανόηση.

Έστω δύο αθλητικές ομάδες η Α και η Β που παίζουν διαδοχικά παιχνίδια **ανεξάρτητα** μεταξύ τους. Η Α κερδίζει με πιθανότητα p και η Β αντίστοιχα με πιθανότητα $1-p$. Έστω ότι συμμετέχουν σε ένα πρωτάθλημα, με κανόνα ότι τερματίζει το πρωτάθλημα όταν κάποια ομάδα κερδίσει 2 παιχνίδια στην σειρά. Η εξέλιξη του πρωταθλήματος θα μοντελοποιηθεί με μια ΜΑΔΧ. Έστω $\{X_n, n = 1, 2, \dots, n\}$ η ΜΑΔΧ σε κάθε βήμα n αντιστοιχεί ένα παιχνίδι. Οι τιμές (που ανήκουν στον χώρο καταστάσεων) που μοντελοποιούν τα γεγονότα είναι οι εξής: $X_n = 0$ αρχή αγώνα, $X=1$ η Α κέρδισε το τελευταίο παιχνίδι, $X=2$ η Α κέρδισε το τελευταίο παιχνίδι, $X=-1$ η Β κέρδισε το τελευταίο παιχνίδι, $X=-2$, η Β κέρδισε το τελευταίο παιχνίδι.

Μπορούμε να διαπιστώσουμε ότι ακολουθεί την μαρκοβιανή ιδιότητα καθώς η πιθανότητα νίκης για την Α ομάδα με βάση τον προ υπάρχον κανόνα είναι $P(X_n = 2 | X_{n-1} = 1) = p_{12}$. Καθώς μας ενδιαφέρει να έχει κερδίσει 2 παιχνίδια στην σειρά για να τελειώσει το πρωτάθλημα, στην δεδομένη κατάσταση που έχει παιχτεί το τελευταίο παιχνίδι στο οποίο νίκησε, δεν μας ενδιαφέρει να έχει περαιτέρω μνήμη από το παρελθόν καθώς μας ενδιαφέρει το γεγονός 2 σερί παιχνίδια νίκης, οπότε αποτελεί περιττό για παράδειγμα αν υπήρχε κι άλλο παρελθοντικό γεγονός νίκης είτε της ομάδας Α είτε της Β.

2.4. ΔΙΑΔΙΚΑΣΙΑ ΔΙΑΚΛΑΔΩΣΗΣ

Παρακάτω θα μιλήσουμε για ένα μαθηματικό (στοχαστικό) μοντέλο που βασίζεται στην διαδικασία Galton-Watson. Το μοντέλο αυτό επινοήθηκε καθώς υπήρξε γενικά αντικείμενο παρατηρήσεων, η εξαφάνιση των επωνύμων ατόμων που είχαν διακεκριμένες θέσεις. Ο Galton ξεκίνησε την παρατήρηση του για απλά άτομα και τον καθορισμό της πιθανότητας εξαφάνισης τους.

Γενικά μια διαδικασία διακλάδωσης (μια περίπτωση τέτοια είναι η Galton-Watson) που επηρεάζεται από γεγονότα τύχης είναι η εξής: Έστω αντικείμενα ίδιου είδους που παράγουν άλλα αντικείμενα. Η διαδικασία

τους μπορεί να είναι διαδικασία ανθρώπινης αναπαραγωγής μέχρι την παραγωγή ενός πληθυσμού βακτηρίων. Έστω ένα σύνολο αντικειμένων σε μία γενιά που έχει παιδιά. Τα παιδιά τους ανήκουν στην επόμενη γενιά και το ίδιο συμβαίνει και στις επόμενες. Ένα γεγονός που μας βοηθά να καταλάβουμε ότι αποτελεί διακριτή διαδικασία είναι ότι μας ενδιαφέρει η παρατήρηση και μόνο των γενεών της εκάστοτε διαδικασίας. Επομένως ορίζετε θέτοντας ως $Z_0, Z_1, Z_2, \dots, Z_n$ τις τυχαίες μεταβλητές για τις συνεχόμενες γενιές[5].

Για να εμβαθύνουμε στην διαδικασία μοντελοποίησης της Galton-Watson διαδικασίας, οι παραπάνω τυχαίες μεταβλητές δηλώνουν τον αριθμό των ατόμων στην εκάστοτε γενιά. Αν γνωρίζουμε για παράδειγμα το μέγεθος στην $Z_n = i$ γενιά n , τότε αν επιθυμούμε να κάνουμε μια **πρόβλεψη** για τις επόμενες γενιές των τυχαίων μεταβλητών τότε η **πιθανότητα** αυτή θα εξαρτάται και μόνο από την δεδομένη κατάσταση-πληροφορία (δηλαδή την i) και όχι από το παρελθόν της τροχιάς αυτής[5]. Δηλαδή οι $Z_1, Z_2, \dots, Z_n$ αποτελούν μια μαρκοβιανή αλυσίδα με πιθανότητα μετάβασης:

$$P_{ij} = P(Z_{n+1} = j | Z_n = i) \quad (2.9)$$

Να αναφέρουμε στο επόμενο κεφάλαιο που θα προσομοιώσουμε μια διαδικασία διακλάδωσης θα κάνουμε υπόθεση ότι ακολουθεί κατανομή Poisson (διακριτή συνάρτηση κατανομής), όντας διακριτή διαδικασία για την διευκόλυνση υπολογισμών. Όμως χρήσιμο αποτελεί να αναδείξουμε σε αυτό το σημείο, έναν άλλο τρόπο υπολογισμού πιθανοτήτων για να έχουμε μια πλήρη εικόνα κατανόησης του μοντέλου Galton-Watson.

Πρέπει για αρχή να υποθέσουμε ότι ο πληθυσμός **ξεκινάει** με ένα άτομο δηλαδή $Z_0 = 1$. Μετά από το πέρασμα μιας μονάδας χρόνου(γενιάς) ο πληθυσμός είναι Z_1 όπου αν είναι μηδέν τότε μιλάμε για την πιθανότητα αφανισμού σαν γεγονός. Αν είναι θετικός ο αριθμός τότε **κάθε άτομο** στην γενιά αυτή γεννά παιδιά. Για παράδειγμα η $X_{1,1}$ τυχαία μεταβλητή αποτελεί τον (τυχαίο) αριθμό των παιδιών που έχει το πρώτο άτομο στην πρώτη γενιά, η $X_{1,2}$ τα παιδιά του 2 ατόμου και ούτω καθεξής[6]. Έτσι η παραπάνω μεταβλητή Z_n που ορίσαμε ως ο αριθμός των ατόμων στην γενιά n καθορίζεται από τις X ως εξής:

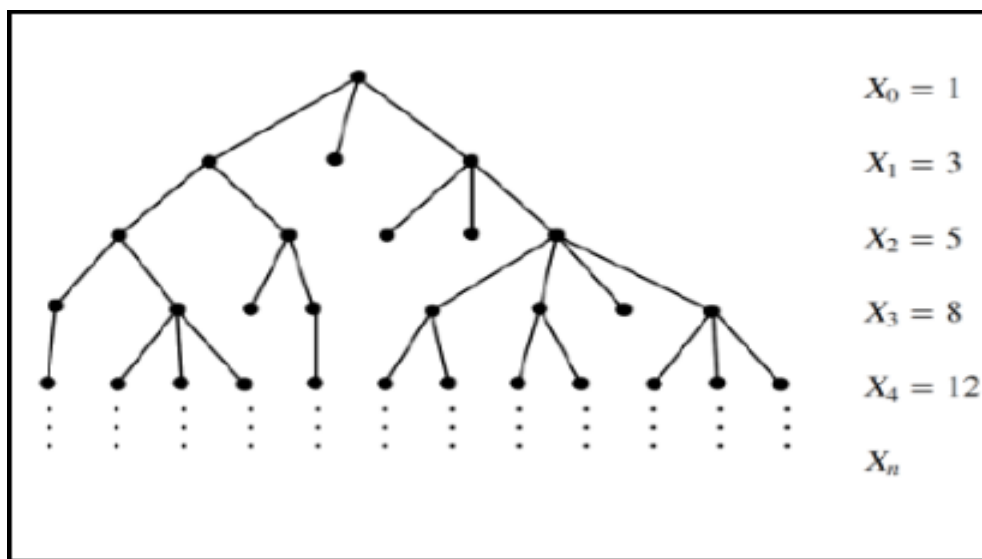
$$Z_n = \sum_{i=1}^{Z_{n-1}} X_i \quad (2.10)$$

Η τυχαία μεταβλητή Z_n γράφεται από μια κατανομή μάζας πιθανότητας η οποία ονομάζεται κατανομή απογόνων. Επειδή κάθε άτομο σε μια γενιά αναπαράγεται ανεξάρτητα από κάθε άλλο, η συνάρτηση κατανομής p_2 που δείχνει την πιθανότητα (για παράδειγμα) ένα άτομο να έχει 2 απογόνους είναι ίδια για κάθε άτομο στον πληθυσμό και ανεξάρτητη της γενιάς στην οποία είμαστε λόγω της μαρκοβιανής ιδιότητας.

Όπως αναφέραμε για το απλοϊκό μοντέλο γίνεται η χρήση της πιθανογεννήτριας $G(s)$. Η γεννήτρια αυτή αποτελεί μαθηματικό εργαλείο που χρησιμοποιείται στην περιγραφή των πιθανοτήτων διαφόρων αποτελεσμάτων της τυχαίας μεταβλητής. Ο τύπος της είναι ο εξής:

$$G(s) = \sum_{k=0}^{\infty} p_k s^k \quad (2.11)$$

Όπου p_k όπως αναφέραμε η πιθανότητα να έχουμε k απογόνους.



Εικόνα 2.2: Δεντρική Δομή της εξέλιξης πληθυσμού στην Διαδικασία Διακλάδωσης

ΚΕΦΑΛΑΙΟ 3

ΠΡΟΣΟΜΟΙΩΣΗ ΣΥΣΤΗΜΑΤΩΝ

3.1 ΠΡΟΣΟΜΟΙΩΣΗ

Προσομοίωση σημαίνει, η μεταφορά ενός φυσικού προβλήματος σε ένα άλλο σύστημα υπολογιστή όπου μεταφέροντας το (μαθηματικό ή σχεδιαστικό) μοντέλο του παρατηρούμε διάφορα για τις παραμέτρους του και τα χαρακτηριστικά του καθώς μελετάμε την συμπεριφορά αυτού μακροπρόθεσμα. Βασικό στοιχείο που πρέπει να γίνει ξεκάθαρο είναι ότι προσομοιάζουμε στοιχεία που είναι πραγματικά. Η χρησιμότητα της Προσομοίωσης έγκειται στο γεγονός ότι πολλές πειραματικές μελέτες μπορούν με την χρήση του υπολογιστή να γίνουν με μεγαλύτερη ακρίβεια και σε μικρότερη διάρκεια. Αποτελεί ορισμός της Προσομοίωσης η αποτύπωση της συμπεριφοράς μιας διαδικασίας ή ενός συστήματος μέσω ενός προγράμματος προσομοίωσης. Για την διαδικασία που θα παρουσιάσουμε παρακάτω θα αναφέρουμε ορισμένες ορολογίες που βοηθούν στον προσδιορισμό της εκάστοτε διαδικασίας κατά την προσομοίωση της.

Στην προσομοίωση έχουμε τις οντότητες που συμμετέχουν σε αυτή. Οι οντότητες, είναι στοιχεία που επηρεάζουν άλλες οντότητες ή επηρεάζονται από αυτές, κινούνται στο σύστημα και υπολογίζονται διάφορα χαρακτηριστικά τους. Ένα παράδειγμα που μπορεί να βασιστούμε σε αυτό αποτελεί το μοντέλο ουράς αναμονής. Εδώ οι οντότητες είναι τα διαφορετικά στοιχεία που μπαίνουν στο σύστημα, εισέρχονται σε μια ουρά, βγαίνουν από αυτήν πηγαίνοντας στην επεξεργασία τους(και στους διάφορους υπολογισμούς που δέχονται για τα χαρακτηριστικά τους) και τελικά απομακρύνονται από το σύστημα. Επίσης γίνεται διάκριση αν το μοντέλο είναι ντετερμινιστικό ή στοχαστικό[7].

Αποτελεί θεμελιώδης σημασία η γνώση των κατανομών που ακολουθούν τα γεγονότα του συστήματος που θέλουμε να προσομοιώσουμε και οι παραμέτρους τους έτσι ώστε με την ακρίβεια που θα έχουμε στα δεδομένα της εισόδου μας θα έχουμε αξιόπιστα και ακριβή αποτελέσματα. Οπότε στην περίπτωση να μην γνωρίζουμε για τις τυχαίες μεταβλητές μας (τυχαία γεγονότα) την κατανομή που ακολουθούν προσθέτει προβλήματα στην εκτέλεση της προσομοίωσης. Σε αυτή την περίπτωση έχουμε την δημιουργία κατανομών που ακολουθούν οι τυχαίες μεταβλητές για το τρέξιμο της προσομοίωσης, τεχνική που βασίζεται στην γνώση των ειδικών[8].

Σε σχέση με τα πραγματικά αποτελέσματα ενός συστήματος, πολλές φορές χρειάζεται να αναλύσουμε και να προβλέψουμε πιθανά σενάρια για να κατανοήσουμε την συμπεριφορά του υπό διάφορες συνθήκες. Αυτό επιτυγχάνεται με την παραγωγή τυχαίων αριθμών μιας τυχαίας μεταβλητής που ακολουθεί μια συγκεκριμένη κατανομή. Η κατανομή περιγράφει στην είσοδο των τιμών αυτών πώς αυτές είναι διασπαρμένες ή διανεμημένες. Ένα παράδειγμα αποτελεί μια τυχαία σειρά αριθμών για το γεγονός (η ΤΜ μας) της τυχαίας άφιξης πελατών σε ένα σύστημα καθώς και μια τυχαία σειρά για τον χρόνο εξυπηρέτησης των πελατών από έναν εξυπηρετητή.

Οι ψευδό-τυχαίοι αριθμοί που παράγονται από τις γεννήτριες τυχαίων αριθμών πρέπει να έχουν ορισμένες επί πλέον ιδιότητες, οι οποίες είναι απαραίτητες για τη χρήση

τους σε προσομοίωση. Οι ιδιότητες αυτές συνοψίζονται ως εξής:

1. Οι τυχαίοι αριθμοί θα πρέπει να είναι κατανεμημένοι ομοιόμορφα, να ακολουθούν, δηλαδή, την ομοιόμορφη κατανομή που περιγράφηκε παραπάνω.
2. Οι τυχαίοι αριθμοί θα πρέπει να είναι ανεξάρτητοι και ταυτόσημα κατανεμημένοι στο διάστημα $[0,1]$.
3. Οι τυχαίοι αριθμοί θα πρέπει να είναι επαναλήψιμοι. Θα πρέπει δηλαδή να υπάρχει η δυνατότητα αναπαραγωγής ακριβώς της ίδιας σειράς τυχαίων αριθμών.
4. Οι αριθμοί θα πρέπει να είναι διαθέσιμοι ανεξάρτητα από το υπολογιστικό σύστημα παραγωγής τους. Αυτό σημαίνει ότι η παραγωγή τους δεν θα πρέπει να εξαρτάται από συγκεκριμένο υλικό (hardware) αλλά από συγκεκριμένη μέθοδο, η οποία να μπορεί να εφαρμοσθεί σε οποιοδήποτε υπολογιστικό σύστημα.
5. Η παραγωγή των τυχαίων αριθμών θα πρέπει να χαρακτηρίζεται από μεγάλη υπολογιστική ταχύτητα και ελάχιστο αποθηκευτικό χώρο. Η ιδιότητα αυτή είναι απαραίτητη λόγω της ανάγκης δημιουργίας μεγάλου πλήθους τυχαίων αριθμών κατά τη διάρκεια μιας προσομοίωσης και της χρήσης τους για τον υπολογισμό των χαρακτηριστικών των οντοτήτων του μοντέλου που προσομοιώνεται. Η παραγωγή, επομένως, των τυχαίων αριθμών δεν θα πρέπει να αποτελεί τον κύριο υπολογιστικό όγκο μιας προσομοίωσης ούτε τις κύριες αποθηκευτικές απαιτήσεις της.
6. Όπως είναι προφανές, η ιδιότητα αυτή αποκλείει μεθόδους όπως ο υπολογισμός του π για την παραγωγή τυχαίων αριθμών ή τη χρήση των πινάκων της RAND.
7. Οι τυχαίοι αριθμοί θα πρέπει να έχουν πολύ μεγάλη περίοδο (ή κύκλο επαναληψιμότητας). Η αναγκαιότητα της ιδιότητας αυτής θα γίνει περισσότερο κατανοητή στην επόμενη ενότητα[7].

Η Επικύρωση του μοντέλου προσομοίωσης είναι σημαντική όσον αφορά την ακριβή αναπαράσταση σε σχέση με το θεωρητικό μοντέλο που μοντελοποιεί ένα πραγματικό πρόβλημα. Για παράδειγμα για ένα πρόγραμμα προσομοίωσης ουρών για να παράγουμε αξιόπιστα και ακριβή αποτελέσματα, ειδικά όταν αναλύουμε παράγοντες όπως ο μέσος χρόνος αναμονής στο σύστημα, χρειάζεται να επεξεργαστεί μεγάλο αριθμό πελατών δηλαδή θα πρέπει να τρέξει για μεγάλο χρόνο η προσομοίωση έτσι ώστε να βγουν τα κατάλληλα συμπεράσματα[9].

Σε τέτοιες περιπτώσεις οι αναλυτές προγραμμάτων προσομοίωσης τρέχουν ένα απλοποιημένο πρόγραμμα προσομοίωσης με γνωστές θεωρητικές λύσεις. Δηλαδή μπορούν να χρησιμοποιήσουν θεωρητικά μοντέλα με λύσεις έτσι ώστε να προβούν σε συγκρίσεις με το απλοποιημένο πρόγραμμα για την ορθότητα του. Για παράδειγμα για ένα πρόγραμμα προσομοίωσης logistics, αυτό αναπαρίσταται με μοντέλα και μαθηματικούς τύπους από την θεωρία ουρών και συγκεκριμένα τον κανόνα M/M/n με χρόνους εξυπηρέτησης και άφιξης εκθετικούς. Το απλοποιημένο πρόγραμμα θα χρησιμοποιήσει τις παραμέτρους αυτές (λόγου αυτό απλοποιημένο πρόγραμμα) συγκρίνοντας τα θεωρητικά αποτελέσματα με τις μετρήσεις της προσομοίωσης για την συνολική αξιολόγηση του μοντέλου προσομοίωσης[9].

3.2 ΠΡΟΣΟΜΟΙΩΣΗ MONTE CARLO

Αφορά για τεχνικές ή μεθόδους που χρησιμοποιούν τυχαία δείγματα για την προσομοίωση της συμπεριφοράς συστημάτων. Λόγω αδυναμία επίλυσης προβλημάτων επειδή δεν μπορούν να αντιμετωπιστούν με τις παραδοσιακές αναλυτικές μεθόδους, χρησιμοποιούνται γενικά αλγόριθμοι όπου κάνουν χρήση της ομοιόμορφης κατανομής όπου $U_i \sim Uniform(0,1)$ με U_i σύνολο τυχαίων μεταβλητών και $U_i = \{U_1, U_2, \dots, U_n\}$. Οι τυχαίες τιμές που θα επιλέξουμε σε αυτό το διάστημα έχουν την ίδια πιθανότητα εμφάνισης καθώς η συνάρτηση πυκνότητας πιθανότητας είναι $f(x) = \frac{1}{b-a}$. Την διαδικασία

αυτή την κάνουμε για να δημιουργήσουμε ένα σύνολο δεδομένων που να ακολουθεί την πολυπλοκότητα και τις ιδιότητες μιας F συνάρτησης κατανομής που θα μας δίνεται, χρησιμοποιώντας τις απλούστερες ιδιότητες της ομοιόμορφης κατανομής.

Έστω για ένα τυχαίο (μεταβλητό) X και θέλουμε να βρούμε την μέση του τιμή (δηλαδή μια πραγματική παράμετρος). Κάποιες φορές είναι δύσκολος ο υπολογισμός αυτός γι αυτό χρησιμοποιούμε την μέθοδο Monte Carlo για την εκτίμηση αυτού του μέσου όρου. Έστω για παράδειγμα ότι η X είναι η ζημιά που προκαλείται από ένα φυσικό φαινόμενο σε χιλιάδες ευρώ και ότι έχει την ακόλουθη πιθανότητα κατανομής:

- Πιθανότητα 0,1 για ζημιά 10 χιλιάδες ευρώ.
- Πιθανότητα 0,5 για ζημιά 5 χιλιάδες ευρώ.
- Πιθανότητα 0,4 για ζημιά 2 χιλιάδες ευρώ.

Η εκτίμηση όπως είπαμε θα γίνει με την μέθοδο Monte Carlo. Με μια γλώσσα προγραμματισμού (π.χ. Python) θα πραγματοποιήσουμε την προσομοίωση για διαφορετικό αριθμό επαναλήψεων.

Ο αριθμός επαναλήψεων είναι πολύ καθοριστικό στοιχείο στην προσομοίωση με την μέθοδο Monte Carlo καθώς ανάλογο τον στόχο μας (για υπολογισμό) χρειάζεται διαφορετικός αριθμός επαναλήψεων. Στην περίπτωση που θέλουμε να εκτιμήσουμε διάφορες τιμές μέσων όρων, ο αριθμός των επαναλήψεων πρέπει να είναι μικρός της τάξεως 100 επαναλήψεις. Ενώ στην περίπτωση που θέλουμε να αναλύσουμε γεγονότα με πολύ μικρή πιθανότητα εμφανίσεις (σπάνια) χρειάζονται χιλιάδες επαναλήψεις για να τα παρατηρήσουμε. Όμως στην περίπτωση πολλών επαναλήψεων όταν τρέχει μια προσομοίωση, γεννιούνται θέματα με τον επεξεργαστή καθώς απαιτείται περισσότερη χρήση μνήμης και ταχύτητας[8].

Όπως αναφέραμε η βάση της προσομοίωσης είναι η παραγωγή τυχαίων ανεξάρτητων αριθμών που ακολουθούν μια συγκεκριμένη κατανομή με πεπερασμένο μέσο όρο και τυχαία απόκλιση. Όμως σε ένα πρόγραμμα ο υπολογιστής δεν παράγει τυχαίους αριθμούς ακριβώς αλλά ψευδό-τυχαίους μέσω ενός αλγορίθμου. Πριν την παραγωγή μιας σειράς τυχαίων αριθμών ορίζουμε ένα seed με μια συγκεκριμένη τιμή. Η χρήση του ίδιου seed δείχνει ότι αναπαράγεται η ίδια σειρά τυχαίων αριθμών. Το seed δηλαδή εξασφαλίζει με την χρήση του στην συνάρτηση-γεννήτορα ότι η ακολουθία των αριθμών που θα παραχθεί θα είναι προβλέψιμη και επαναλαμβανόμενη.

Ο ψευδό-κώδικας της παραπάνω διαδικασίας φαίνεται παρακάτω:

ΨΕΥΔΟΚΩΔΙΚΑΣ

ΑΛΓΟΡΙΘΜΟΣ Γεννήτορας_με_ιδιο_seed

1: Αρχικοποίησε το seed του γεννήτορα τυχαίων αριθμών έτσι ώστε να επαναλαμβάνονται οι τυχαίες σειρές τιμών $seed \leftarrow 0$

2: Εκκίνησε τον γεννήτορα τυχαίων αριθμών με το δοθέν seed.

3: Αρχικοποίησε τον Γεννήτορα(seed)

4: Παραγωγή τυχαίας σειράς αριθμών

ΓΙΑ i από 1 μέχρι 4

5: $τυχαίος_αριθμός[i] \leftarrow \text{ΠαραγωγήΤυχαίουΑριθμού}()$

Εικόνα 3.1: Αλγόριθμος για την παραγωγή ίδιων τυχαίων σειρών

Η συνάρτηση Παραγωγή Τυχαίου Αριθμού() στον παραπάνω ψευδο-κώδικα σε πολλές προγραμματιστικές γλώσσες εμφανίζεται με την εντολή `rand`. Η συνάρτηση `rand` (γεννήτρια συνάρτηση) παράγει ψευδό-τυχαίους αριθμούς στο διάστημα $(0,1)$ που σημαίνει ότι είναι τυχαίοι αριθμοί που ακολουθούν την ομοιόμορφη κατανομή. Ακολουθούν ορισμένα παραδείγματα της `rand` στο MATLAB:

- `rand`= Παραγωγή τυχαίου αριθμού στο διάστημα $(0,1)$
- `rand(m,n)`= Πίνακας $m \times n$ με τυχαίους αριθμούς στο $(0,1)$.

3.2.1 ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ

Όπως αναφέραμε τη χρήση της `rand` την συναντάμε σε διάφορες προγραμματιστικές γλώσσες (με παραλλαγές) που μας βοηθά στην προσομοίωση φαινομένων και πειραμάτων. Το ακόλουθο παράδειγμα αποτελεί την προσομοίωση για το στρίψιμο ενός ζαριού, στην γλώσσα Python, με την προσομοίωση να τρέχει για μεγάλο αριθμό επαναλήψεων. Θα χρησιμοποιήσουμε την βιβλιοθήκη `random` που περιλαμβάνει την συνάρτηση `random.randint(a,b)` που παράγει ένα **τυχαίο** αριθμό ανάμεσα στο α το κατώτερο όριο και το β το ανώτερο όριο.

Η τυχαία μεταβλητή μοντελοποιείται ως το αποτέλεσμα του στρίψιματος που είναι το *heads* και το *tails*. Όλα τα παραπάνω αναπαρίστανται ως εξής:

```
from random import randint

N = 1000000

Heads = 0

Tails = 0

for k in range(N):

    i= randint(1, 2)

    If i==1:

        Heads +=1

    else:

        Tails += 1

Print N, Heads, Tails
```

Ο παραπάνω κώδικας θα μπορούσε να έχει την εξής μοντελοποίηση σε ένα τέτοιο πιθανοτικό πρόβλημα:

K, Γ, Γ, K, Γ, K, Γ, K ...

Όπου η παραπάνω αποτελεί μια **τυχαία** σειρά των δύο αποτελεσμάτων.

Ένα άλλο παράδειγμα μοντελοποίησης παρόμοιο με το προηγούμενο αποτελεί το ρίξιμο ενός (δίκαιου) ζαριού το οποίο έχει 6 πλευρές αριθμημένες από το 1 έως το 6 για κάθε πλευρά. Αυτό μπορεί να μοντελοποιηθεί πάλι στην Python με τη συνάρτηση *randint* όπου επιλέγεται ένας αριθμός **τυχαίος** από το 1 έως το 6. Αυτό γίνεται ως εξής:

```
from random import randint

N = 6000000

count = 0

for k in range(N):

    i= randint(1, 6)

    If i==5:

        Count+=1

print count
```

Έστω παράδειγμα με 2 τυχαίες μεταβλητές που θα μοντελοποιήσουμε που μαθηματικά θα λέγαμε ως X_1 = αποτέλεσμα ρίψης του πρώτου ζαριού και X_2 = αποτέλεσμα ρίψης δεύτερου ζαριού. Στόχος μας είναι να δείξουμε ότι θα νικήσουμε **αν** φέρουμε το ίδιο αποτέλεσμα (δίκαιου) ζαριού στην ίδια επανάληψη[10]. Αυτό αναπαριστάται παρακάτω ως εξής:

```
# Initialize number of dice, simulate & record outcome
np.random.seed(123)

die, probabilities, num_dice = [1,2,3,4,5,6], [1/6, 1/6, 1/6, 1/6, 1/6, 1/6], 2

outcomes = np.random.choice(die, size=num_dice, p=probabilities)

#Win if the two dice show the same number

If outcomes[0] == outcomes[1]:

    answer='win'

else:

    answer='lose'

print("the dice shows {} and {}. You {}".format(outcomes[0], outcomes[1], answer))
```

Out [0]: The dice shows 5 and 2. You lose!

Για την πλήρη κατανόηση του, στη μεταβλητή *outcomes* από την συνάρτηση `np.random.choice` (από την βιβλιοθήκη `numpy`), επιστρέφεται ένας πίνακας με δύο θέσεις λόγω της παραμέτρου `size` όπου έπειτα εξισώνονται για να προκύψει ο στόχος που έχουμε θέσει.

Όμως όπως έχουμε ήδη αναφέρει σε ένα πείραμα για να βγάλεις συμπεράσματα για την συμπεριφορά των αποτελεσμάτων πρέπει να το επαναλάβουμε πολλές φορές. Οπότε αν θέλαμε να δούμε πόσες φορές θα νικάγαμε σε (για παράδειγμα) 100 επαναλήψεις, θα ορίζαμε μια μεταβλητή `sim` και `win` και θα αρχικοποιούσαμε με 0 την δεύτερη όπου αργότερα ανάλογα το `outcome` θα αυξάνεται[10].

3.3 ΑΛΓΟΡΙΘΜΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Παρακάτω παρουσιάζονται οι αλγόριθμοι για την παραγωγή τιμών μιας τυχαίας μεταβλητής X με συγκεκριμένη κατανομή από την ομοιόμορφη που οι ιδιότητες της μας βοηθούν γι' αυτό.

3.3.1. ΑΛΓΟΡΙΘΜΟΣ BOX-MULLER

Πριν αναφερθούμε στον αλγόριθμο πρέπει να κατανοήσουμε την διαδικασία του Γκαουσιανού ολοκληρώματος:

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \quad (3.1)$$

Έστω ότι έχουμε δύο συναρτήσεις της μορφής αυτής με μεταβλητές x και y αντίστοιχα;

$$\left[\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right] \left[\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right] = \quad (3.2)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} dx dy = \quad (3.3)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy \quad (3.4)$$

Στην περίπτωση αυτή αναλύουμε την συνάρτηση στον xy -χώρο όπου θα κάνουμε μετατροπή σε πολικές συντεταγμένες για την βολικότητα των πράξεων. Ισχύει ότι και r είναι η ακτίνα, η απόσταση του σημείου από την αρχή των αξόνων και θ η γωνία που περιγράφει την κατεύθυνσή από τον θετικό x -άξονα προς το σημείο. Ακόμα τα όρια των ολοκληρωμάτων αλλάζουν σε r και θ αντίστοιχα όπου ολοκληρώνουμε πάνω σε όλες τις τιμές του r και

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta \quad (3.5)$$

Στο επόμενο σημείο ακολουθεί ο υπολογισμός αυτού μέσω αλλαγής μεταβλητής. Όμως θα περάσουμε τώρα στον αλγόριθμο που βασίζεται στην παραπάνω διαδικασία. Στην περίπτωση του Box-Muller μιλάμε για την παραγωγή τιμών για τις τυχαίες μεταβλητές που ακολουθούν κανονική κατανομή μέσω των ομοιόμορφα κατανεμημένων. Ο αλγόριθμος περιλαμβάνει μια μη γραμμική μετατροπή μέσω των τριγωνομετρικών συναρτήσεων. Έστω ότι έχουμε τις συναρτήσεις πυκνότητας πιθανότητας για την X και τη Y :

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3.6)$$

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad (3.7)$$

Παίρνοντας την από κοινού κατανομή πιθανότητας $f(x, y)$ και επειδή οι X και Y είναι ανεξάρτητες μεταβλητές ισχύει:

$$f(x, y) = f(x)f(y) \quad (3.8)$$

$$= \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right] \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \right] \quad (3.9)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2+y^2)}{2}} \quad (3.10)$$

Όπως και από πάνω θέτοντας $R^2 = X^2 + Y^2$ κάνουμε τον μετασχηματισμό δηλαδή;

$$f(x, y) \rightarrow f(r, \theta) \quad (3.11)$$

$$f(r, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} \quad (3.12)$$

Ερχόμαστε στο σημείο που θα χρησιμοποιήσουμε ένα σετ τυχαίων μεταβλητών (U_1, U_2) που ακολουθούν ομοιόμορφη κατανομή για τις τιμές που λαμβάνουν έτσι ώστε να παραχθούν οι κανονικές τιμές-δείγματα. Έστω :

$$U_1 = \frac{\Theta}{2\pi} \rightarrow \Theta = 2\pi U_1 \quad (3.13)$$

$$U_2 = e^{-\frac{R^2}{2}} \rightarrow R = \sqrt{(-2 \ln U_2)} \quad (3.3.14)$$

Σύμφωνα με την (3.3.13) η U_1 ακολουθεί ομοιόμορφη στο $(0,1)$ και είναι εξίσου πιθανές οι τιμές σε αυτό το διάστημα, τότε και η θ ακολουθεί ομοιόμορφη στο διάστημα $(0,2\pi)$ που σημαίνει ότι η πιθανότητα να πάρει οποιαδήποτε γωνία μέσα σε αυτό το διάστημα είναι ίδια. Αυτό αποδεικνύεται και από τον τύπο (3.3.12) όπου η $f(r, \theta)$ είναι το γινόμενο δύο διαφορετικών συναρτήσεων των ανεξάρτητων μεταβλητών r, θ όπου η μεταβλητή θ έχει συνάρτηση πυκνότητας πιθανότητας:

$$f(\theta) = \frac{1}{2\pi} \quad (3.15)$$

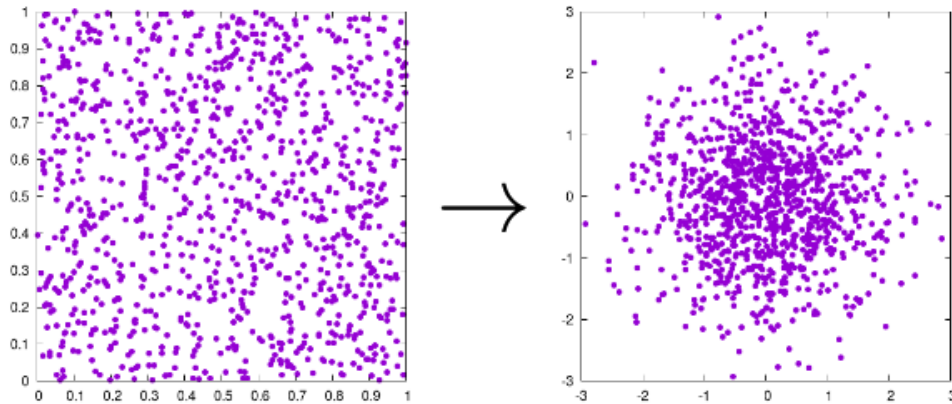
Παρατηρούμε ότι φαίνεται η θ να ακολουθεί ομοιόμορφη καθώς η συνάρτηση είναι σταθερή ως προς το διάστημα $(0,2\pi)$ όπως θα ήταν η $f(y)$ αν δεν κάναμε τον μετασχηματισμό στο διάστημα (a, β) με $f(y) = \frac{1}{b-a}$. Τώρα μπορούμε να μετατρέψουμε τις πολικές συντεταγμένες πίσω στις καρτεσιανές χρησιμοποιώντας:

$$X = R \cos \theta \quad (3.16)$$

$$Y = R \sin \theta \quad (3.3.17)$$

$$X = R \cos \theta = \sqrt{-2 \ln(U_2) \cos(2\pi U_1)} \quad (3.18)$$

$$Y = R \sin \theta = \sqrt{-2 \ln(U_2) \sin(2\pi U_1)} \quad (3.19)$$



Εικόνα 3.2: Γραφική απεικόνιση της συνάρτησης πυκνότητας των μετασχηματισμένων X, Y

3.3.2. ΔΕΙΓΜΑΤΟΛΗΨΙΑ INVERSE TRANSFORM

Έστω ότι θέλουμε να πάρουμε δείγματα X που έχουν μια συνάρτηση κατανομής F (αθροιστική συνάρτηση κατανομής). Αυτό πραγματοποιείται χρησιμοποιώντας (όπως έχουμε αναφέρει και πιο πάνω) τιμές μιας μεταβλητής που ακολουθεί την ομοιόμορφη κατανομή. Έστω ότι ισχύει $y = F(x)$ και y παίρνει τιμές στο $(0,1)$. Ένα πρώτο ερώτημα είναι τι κατανομή ακολουθεί η y . Αυτό αποδεικνύεται ως εξής:

$$P(Y \leq y) = P(F(x) \leq y) \quad (3.20)$$

$$= P(x \leq F^{-1}(y)) \quad (3.21)$$

$$= F(F^{-1}(y)) \quad (3.22)$$

$$= y \quad (3.23)$$

Με αυτό τον τρόπο παρατηρούμε ότι όντως η y ακολουθεί ομοιόμορφη και θα χρησιμεύσει στον ορισμό της αντίστροφης της $F(x)$ έτσι ώστε να παράγουμε δείγματα για το X . Η αντίστροφη ορίζεται ως εξής:

$$F^{-1}(y) = \min \{x | F(x) \geq y\} \quad (3.24)$$

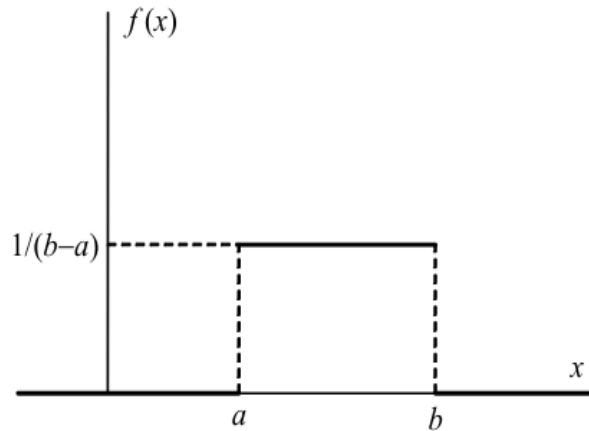
Δηλαδή είναι το ελάχιστο x για την τιμή της $F(x)$ που είναι ίση με την y ή μεγαλύτερη. Η $F^{-1}(y)$ κατανέμεται σύμφωνα με την $F(x)$. Για τον υπολογισμό των δειγμάτων θα εφαρμόσουμε τον τύπο $x = F^{-1}(y)$.

Ένα παράδειγμα που θα μας βοηθήσει να καταλάβουμε την χρήση της αντίστροφης είναι: Έστω ότι ο χρόνος αναμονής για την άφιξη ενός λεωφορείου ακολουθεί μια εκθετική κατανομή η οποία είναι κοινή για τέτοια είδη δεδομένων. Η εκθετική κατανομή είναι χρήσιμη για την περιγραφή του χρόνου ανάμεσα σε συνεχόμενα τυχαία γεγονότα και έχει την παρακάτω συνάρτηση κατανομής:

$$F(x; \lambda) = 1 - e^{-\lambda x} \quad (3.25)$$

Όπου x είναι ο χρόνος αναμονής και λ αποτελεί ο ρυθμός αύξησης που είναι το αντίστροφο του μέσου χρόνου αναμονής. Έχοντας τον μέσο χρόνο αναμονής ότι είναι 20 λεπτά άρα $\lambda = \frac{1}{20}$ ανά λεπτό. Για την προσομοίωση της άφιξης των λεωφορείων:

- Παράγουμε μια τυχαία τιμή U από την ομοιόμορφη κατανομή $U(0,1)$ για παράδειγμα.
- Μέσω της αντίστροφης συνάρτησης κατανομής $F^{-1}(U; \lambda)$ η οποία για την εκθετική κατανομή είναι: $X = F^{-1}(U; \lambda) = -\frac{1}{\lambda} \ln(1 - U)$
- Βάζοντας την τιμή έχουμε: $X = -\frac{1}{\frac{1}{20}} \ln(1 - 0.5) = -20 \ln(0.5)$. Επομένως $X \approx 13.86$



Εικόνα 3.3: Διάγραμμα ομοιόμορφης κατανομής στο (α.β).

3.3.3 ΚΑΤΑΣΚΕΥΗ ΚΑΙ ΠΡΟΣΟΜΟΙΩΣΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΔΙΑΚΛΑΔΩΣΗΣ

Για την προσομοίωση της διαδικασίας διακλάδωσης δεδομένου την πιθανότητα απογόνων

$$p_k = P(Z_1 = k)$$

με $k = 0, 1, 2, 3$ όπου ανήκουν στο σύνολο των ακεραίων, υποθέτουμε ότι είναι γνωστή η συνάρτηση κατανομής F της p_k και ισχύει όπως δείξαμε και στο προηγούμενο κεφάλαιο ότι:

$$\eta = F(\gamma) \quad (3.26)$$

Με την γ να ακολουθεί $U(0,1)$ και εφαρμόζοντας τις διάφορες τιμές γ_i στην F , παράγονται οι τυχαίες τιμές της η_i όπου είναι ανεξάρτητες μεταξύ τους και ακολουθούν την κατανομή μάζας $\{p_k\}$. **Οι τιμές αυτές αποτελούν τα παιδιά του κάθε ατόμου της γενιάς.** Δηλαδή $\eta_1 = 2$ έχουμε 2 παιδιά για το πρώτο άτομο. Οι γ_i αποτελούν τυχαίες τιμές που παράγονται από τον υπολογιστή όπως θα δούμε. Ακολουθεί η προσομοίωση στην Python, της επόμενης γενιάς βασισμένη στην δεδομένη (τωρινή) που αποτελείται από τα παιδιά του κάθε ατόμου.

```
def transformation_function(probfunc):
    Cumulative_fu=np.cumsum(probfunc)
    r=np.random.rand()
    return np.where(cumulative > r)[0][0]

lambda_po=1.5
max_k=15
pmf=np.array([np.exp(-lambda_po)* lambda**k / np.math.factorial(k)
```

```

def simulate_branching_process(num_generations):
    Z=[1]
    for n in range(num_generations):
        Z_n=Z[-1]
        offspring_counts=[transformation_function(pmf) for _ in range(Zn)]
        Zn_plus_1=sum(offspring_counts)
        Z_append(Zn_plus_1)
    return Z

num_generations=10
generations_sizes=simulate_branching_process(num_generations)
print("Generation_sizes", generation_sizes)

```

Όπως βλέπουμε από τον παραπάνω κώδικα στην Python:

- Χρησιμοποιώντας την συνάρτηση `transformation_function()` ουσιαστικά εφαρμόζουμε την μέθοδο `Inverse transform` για τυχαία δειγματοληψία. Η συνάρτηση δέχεται σαν όρισμα έναν πίνακα πιθανοτήτων απογόνων.
- Υπολογίζει με την συνάρτηση `cumsum()` τον αθροιστικό πίνακα πιθανοτήτων έχοντας πάρει σαν όρισμα τις πιθανότητες Poisson(τον πίνακα τους)
- Παράγει με την `rand()` μια τυχαία τιμή στο διάστημα (0,1) για την προσομοίωση των τυχαίων τιμών των απογόνων μας (δηλαδή της νέα τυχαίας μεταβλητής η_i μετά την εφαρμογή των `rand` τιμών)
- Εφαρμόζεται η συνάρτηση `where()` (ανήκει στην βιβλιοθήκη NumPy) όπου επιστρέφεται ένας boolean πίνακας που οι τιμές του καθορίζονται από την συνθήκη, αν οι πιθανότητες της `cumulative` είναι μεγαλύτερες από τον τυχαίο αριθμό `r`. Στην περίπτωση που έχουμε για παράδειγμα το `True` στο index 2 του boolean πίνακα τότε επιστρέφεται από την `where` ως τιμή της τυχαίας μεταβλητής το 2.
- Με την `simulate_branching_process()` προσομοιώνουμε την διαδικασία για να δούμε την αύξηση του πληθυσμού των ατόμων. Ξεκινάμε με μια λίστα που αποθηκεύεται ένα άτομο για την πρώτη γενιά (εννοώντας το πρώτο index) και κατασκευάζουμε μια `for` για τις διάφορες γενιές πληθυσμών. Έχοντας στην τωρινή γενιά τα άτομα του πληθυσμού που προσδιορίζονται από την τελευταία θέση της λίστας (`[-1]` δηλώνει προς το τέλος) βρίσκουμε μετά με την δομή της `list comprehension` τα παιδιά του κάθε ατόμου της τωρινής γενιάς. Αυτό φαίνεται ξεκάθαρα καθώς εφαρμόζεται `Zn` φορές(μέσω `for`, `range` για κάθε άτομο του τωρινού πληθυσμού) η συνάρτηση της παραγωγής τυχαίων αριθμών για το κάθε άτομο.
- Στο τέλος ενημερώνεται μέσω της `append` η λίστα των ατόμων όπου τοποθετούνται τα νέα στοιχεία(της `Z_plus_1`) στην τελευταία θέση.

ΚΕΦΑΛΑΙΟ 4

ΤΥΧΑΙΟΙ ΠΕΡΙΠΑΤΟΙ

4.1 ΟΡΙΣΜΟΣ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Ο “Απλός Τυχαίος Περίπατος” αποτελεί παράδειγμα Στοχαστική Διαδικασίας ανεξάρτητων προσανξήσεων (λόγω των τυχαίων μεταβολών) και μοντελοποιείται ως εξής. Έστω μια ακολουθία $X_i = \pm 1$ με ίση πιθανότητα $p=q=0.5$ να πάρει μία από τις δύο τιμές. Η συνολική ποσότητα S_n στο βήμα n είναι:

$$S_n = X_1 + X_2 + \dots + X_n \quad (4.1)$$

με μέση τιμή των κάθε X_i να είναι $E[X_i] = 1 \cdot 0.5 + (-1) \cdot 0.5$ επομένως $E[X_i] = 0$ Η μέση τιμή της ποσότητας S_n είναι:

$$E(S_n) = E(X_1 + X_2 + \dots + X_n) \quad (4.2)$$

$$E(S_n) = E[X_1] + E[X_2] + \dots + E[X_n] \quad (4.3)$$

$$E(S_n) = 0 \quad (4.4)$$

Για την διακύμανση του S_n ισχύουν:

$$Var(S_n) = Var(X_1 + X_2 + \dots + X_n) \quad (4.5)$$

$$Var(S_n) = nVar(X_i) \quad (4.6)$$

$$Var(X_i) = E(X_i^2) - (E(X_i))^2 = 1 \quad (4.7)$$

$$Var(S_n) = n \quad (4.8)$$

Η Μαρκοβιανή ιδιότητα ισχύει για τον Απλό Τυχαίο Περίπατο δηλαδή:

$$X_n = \sum_{i=1}^{n-1} X_{n-1} \pm 1 \quad (4.9)$$

όπου η θέση του περιπάτου για την μελλοντική στιγμή n εξαρτάται από την παροντική $n-1$ και μόνο μαζί με το τυχαίο βήμα.

Έστω η θέση του περιπάτου είναι στο σημείο k . Για να καταλήξουμε στο σημείο αυτό έχουμε περάσει N_+ συνολικά θετικά βήματα και N_- αρνητικά βήματα. Αυτά πρέπει να ικανοποιούν:

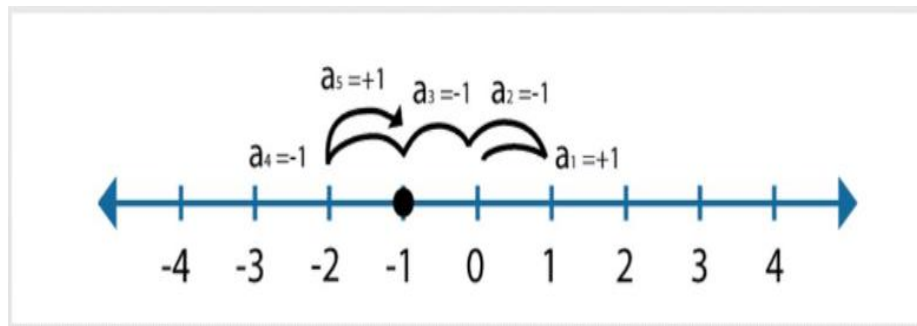
$$N_+ + N_- = k \quad (4.10)$$

$$N_+ + N_- = n \quad (4.11)$$

$$N_+ = \frac{n+k}{2} \quad (4.12)$$

Άρα ισχύει για τον διωνυμικό συντελεστή: $\binom{n}{N_+} = \binom{n}{\frac{n+k}{2}}$.

Επομένως στην περίπτωση του συμμετρικού Απλού Τυχαίου Περιπάτου για την διωνυμική κατανομή ισχύει:

$$P(S_n) = \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}} \quad (4.13)$$


Εικόνα 4.1: Απεικόνιση τυχαίας κίνησης μοναδιαίου βήματος του “Απλού Τυχαίου Περιπάτου.”

4.2 ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΛΟΥ ΤΥΧΑΙΟΥ ΠΕΡΙΠΑΤΟΥ

Στόχος της προσομοίωσης αποτελεί να φανεί σε δεδομένη χρονική στιγμή που σταματάει ο Απλός Τυχαίος Περιπάτος (σημείο) καθώς και τα συνολικά μονοπάτια του για να φανεί η τυχαιότητα του. Ακόμα υπολογίζεται η πιθανότητα με βάση την συνάρτηση παραγωγής Ιστογράμματος στην Python, για τα δείγματα που παράγουμε και απεικονίζεται ο τύπος κατανομής που εκφράζεται μέσω αυτής της εμπειρικής κατανομής.

Για την προσομοίωση του Απλού Τυχαίου Περιπάτου όσον αφορά την μίμηση της συμπεριφοράς του πραγματικού αυτού συστήματος, ακολουθούν οι εξής υλοποιήσεις στην Python:

- Κατασκευή του μοντέλου μέσω της `simulate_random_walk` όπου η `position` αντιστοιχεί στο άθροισμα της S_n και με την `cumsum()` παράγεται ο πίνακας που απεικονίζει σε κάθε θέση (χρονική στιγμή-`num_steps`) τη συνολική ποσότητα των `steps` με βάση την προηγούμενη θέση και το τωρινό τυχαίο `step` (δικαιολογεί την σχέση $X_n = \sum_{i=1}^{n-1} X_i \pm 1$).
- Η προσομοίωση του μοντέλου πολλές φορές, απεικονίζοντας τα διάφορα τυχαία μονοπάτια για ένα συγκεκριμένο αριθμό `steps` έτσι ώστε να φανούν οι τυχαίες διαδρομές που παίρνει στην διάρκεια του χρόνου.
- Η εμπειρική κατανομή που υπολογίζει μέσω του Ιστογράμματος πόσο συχνά παρατηρούνται οι διάφορες τυχαίες τιμές της `final_positions[]` που δείχνει το σύνολο των τυχαίων τιμών της τελευταίας θέσης του πίνακα `position` μέσω της `append`. Ουσιαστικά αποτελεί μια εκτίμηση που

θα αντιστοιχούσε στον μαθηματικό συμβολισμό $P(X_i = \text{αριθμος βημάτων})$ με το i να αποτελεί το `num_steps`.

- Δεδομένα της θεωρητικής κατανομής στην περίπτωση μας η κανονική κατανομή. Τα δεδομένα της κατανομής όπως ο μέσος και η διακύμανση ακολουθούν τις μαθηματικές ιδιότητες του Απλού τυχαίου περιπάτου.
- Ένα διάγραμμα (μέσω ενός `plt.figure()`) όπου συγκρίνεται μέσω της οπτικοποίησης η εμπειρική κατανομή και η κανονική για την Επικύρωση του μοντέλου. Επίσης δίνονται τα δεδομένα για την μέση τιμή και διακύμανση που είναι τα ίδια παρόλο την κανονική προσέγγιση της διωνυμικής κατανομής τους.
- Στην σύγκριση μέσω της οπτικοποίησης επικαλύπτεται το ιστόγραμμα από την κανονική κατανομή, με δεδομένα της κανονικής να προέρχονται από την `np.linspace(min(final_positions), max(final_positions), 1000)` όπου το εύρος των τιμών σχετίζεται με των εμπειρικών δεδομένων για να έχει σημασία η σύγκρισή τους.
- Όμως θέλουμε και μια μέθοδος σύγκρισης που να ποσοτικοποιεί το ταίριασμα του μοντέλου προσομοίωσης πέρα από την εμφανή σύγκριση των διαγραμμάτων.

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

def simulate_random_walk(num_steps):
    # Generate random steps (-1 or 1)
    steps = np.random.choice([-1, 1], size=num_steps)
    # Compute the position of the walker at each step
    position = np.cumsum(steps)
    return position
```

Εικόνα 4.2: Προσομοίωση της διαδικασίας

```
def plot_random_walks(num_simulations, num_steps):
    plt.figure(figsize=(12, 6))
    for _ in range(num_simulations):
        position = simulate_random_walk(num_steps)
        plt.plot(position, alpha=0.7)
    plt.xlabel('Step')
    plt.ylabel('Position')
    plt.title('Multiple Random Walk Simulations')
    plt.grid(True)
    plt.show()
```

Εικόνα 4.3: Προσομοίωση τυχαίων μονοπατιών

```
def validate_random_walk(num_simulations, num_steps):
    final_positions = []
    for _ in range(num_simulations):
        position = simulate_random_walk(num_steps)
        final_positions.append(position[-1])

    # Plot histogram of final positions
    plt.figure(figsize=(12, 6))
    plt.hist(final_positions, bins=30, density=True, alpha=0.6, color='g', label='Simulated Data')

    # Plot the theoretical normal distribution
    mu = 0
    sigma = np.sqrt(num_steps)
    x = np.linspace(min(final_positions), max(final_positions), 1000)
    plt.plot(x, stats.norm.pdf(x, mu, sigma), 'k', linewidth=2, label='Theoretical Normal Distribution')

    plt.xlabel('Final Position')
    plt.ylabel('Probability Density')
    plt.title('Distribution of Final Positions in Random Walk')
    plt.legend()
    plt.grid(True)
    plt.show()

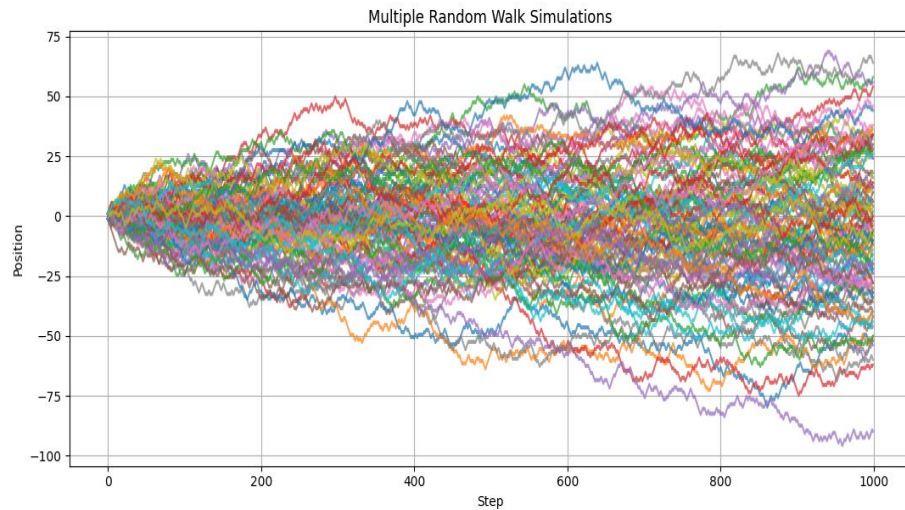
    # Perform a statistical test (Kolmogorov-Smirnov test)
    ks_statistic, p_value = stats.ks_test(final_positions, 'norm', args=(mu, sigma))
    print(f"KS Statistic: {ks_statistic}, P-value: {p_value}")
```

Εικόνα 4.4: Συνάρτηση στατιστικής αξιολόγησης του μοντέλου

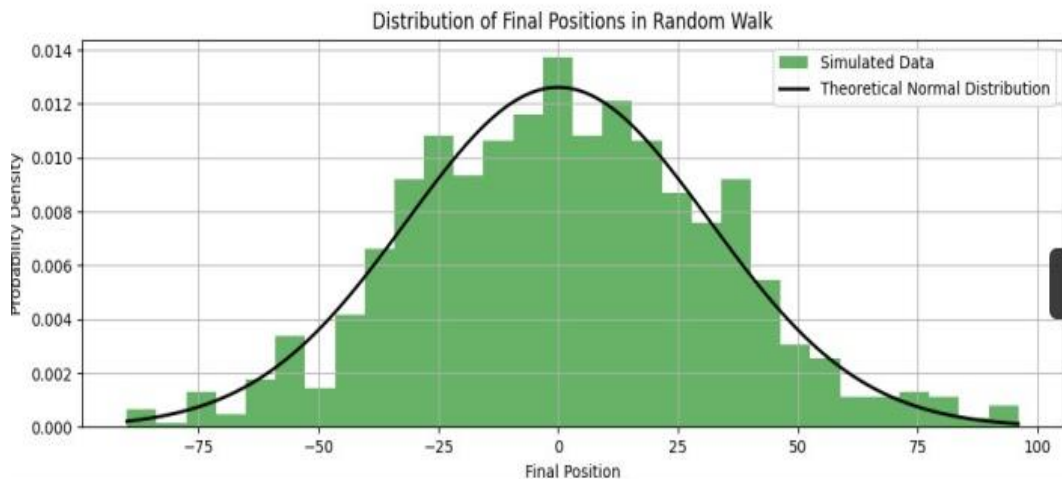
4.2.1. ΣΥΜΠΕΡΑΣΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ

Για την οπτικοποίηση της εμπειρικής κατανομής χρειάζονται πολλά σημεία-δεδομένα για καλύτερα στατιστικά συμπεράσματα και να φανεί ότι τα δεδομένα διαχέονται. Για αυτό η προσομοίωση γίνεται για ένα μεγάλο αριθμό. Για αυτό μας ενδιαφέρει να συλλέξουμε πολλά final_positions-δείγματα.

- Έστω η προσομοίωση με num_simulations=1000 φορές και num_steps=1000. Ακολουθούν εικόνες για ασφαλή συμπεράσματα:



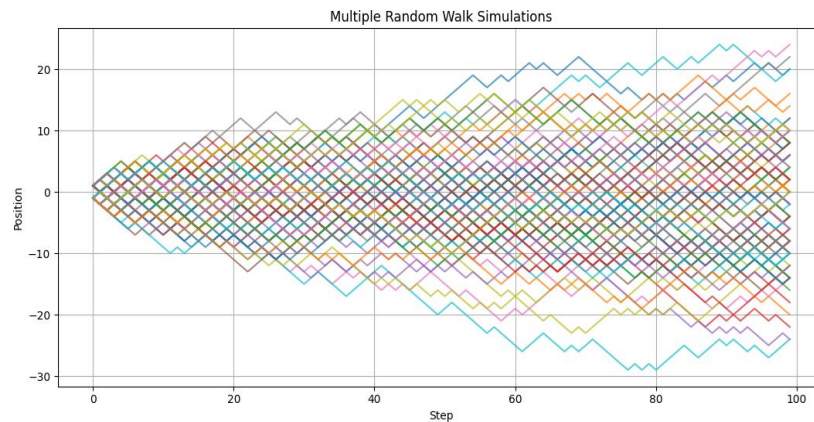
Εικόνα 4.5: Προσομοίωση πολλών τυχαίων μονοπατιών με μεγάλη διάρκεια



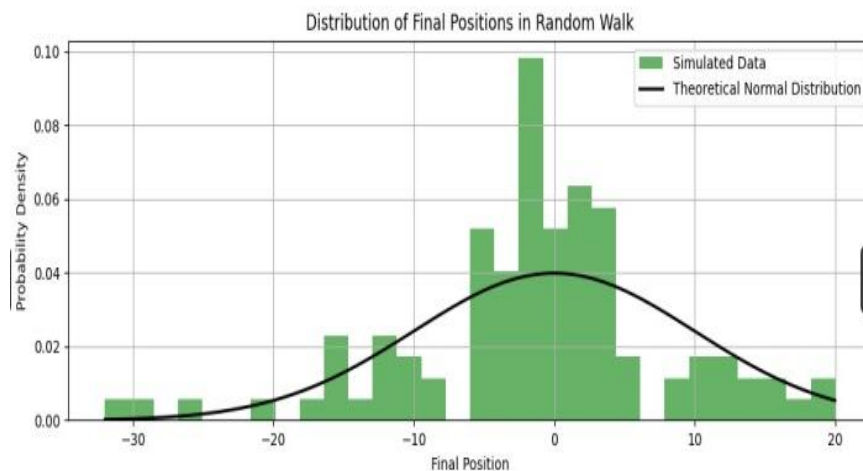
Εικόνα 4.6: Επικάλυψη-σύγκριση της κανονικής με την εμπειρική κατανομή (δείγματος).

Παρατηρούμε στην Εικόνα 4.6 ότι η κανονική κατανομή στα δεδομένα-δείγματα της προσομοίωσης επικαλύπτει την εμπειρική κατανομή και το σύστημα προσομοίωσης έχει παρόμοια χαρακτηριστικά με τα θεωρητικά μέρη του πραγματικού συστήματος.

Σε αντίθεση η δοκιμή της συνάρτησης `validaterandomwalk(100,100)` και της `plotrandomwalks(100,100)` δηλαδή έχοντας συλλέξει λιγότερα δείγματα έχει τα εξής αποτελέσματα:



Εικόνα 4.7: Μονοπάτια τυχαίων περιπάτων για συγκεκριμένο `num_simulations`



Εικόνα 4.8: Σύγκριση κατανομών δείγματος και κανονικής

Στη προσομοίωση αυτή οπτικά στην Εικόνα 4.7 παρόλο που είναι εμφανής τα χαρακτηριστικά των κάθε ατομικών μονοπατιών, στην περίπτωση για 1000 φορές, αποτελεί λόγω περισσότερων δεδομένων, πιο εμφανές χαρακτηριστικά όπως οι διακυμάνσεις. Για την εικόνα 4.8 παρατηρείται ότι η εμπειρική κατανομή λόγω λιγότερο δειγμάτων και προσομοιώσεων, έχει ένα μη συνεκτικό διακριτό διάγραμμα και δεν “μιμείται” την κανονική κατανομή όπως στην περίπτωση δοκιμής 1000 φορές.

Για την εικόνα 4.6 δεν καθιστά αρκετός ο έλεγχος κανονικότητας μέσω οπτικής σύγκρισης των διαγραμμάτων. Στην περίπτωση αυτή γίνεται ένα τεστ κανονικότητας για να εξακριβωθεί ότι τα τυχαία δεδομένα μας ακολουθούν την κατανομή αυτή. Ο έλεγχος αυτός ονομάζεται έλεγχος καλής προσαρμογής. Στην περίπτωση όπου num_simulations=1000 παίρνουμε τα εξής αποτελέσματα στατιστικού ελέγχου:

KS Statistic: 0.029085182977075408, P-value: 0.35916446401955937

Εικόνα 4.9: Τεστ κανονικότητας της κατανομής δείγματος

Τα αποτελέσματα που επιστρέφει η stats.tests αφορά το KS Statistic και το P-value. Η πρώτη τιμή δείχνει την διαφορά της εμπειρικής αθροιστικής κατανομής με την θεωρητική κατανομή της. Παρατηρούμε ότι αποτελεί πολύ μικρή τιμή που δείχνει την ομοιότητα των δύο. Για την P-value η τιμή 0.35 είναι πολύ μεγαλύτερη από τις κλασικές τιμές του επιπέδου σημαντικότητας $\alpha=0.05$ (ή 0.1). Οπότε για την στατιστική απόφαση λόγω του $0.35 > 0.05$ σημαίνει ότι η υπόθεση που έχει γίνει ότι ακολουθεί η εμπειρική κατανομή την κανονική, δεν απορρίπτεται καθώς δεν διαπιστώνονται στατιστικά σημαντικές αποκλίσεις όσον αφορά την κανονικότητα.

4.3. ΠΡΟΒΛΗΜΑ ΚΑΤΑΣΤΡΟΦΗΣ ΤΟΥ ΠΑΙΚΤΗ

Το πρόβλημα αυτό αποτελεί μια ειδική περίπτωση απλού τυχαίου περιπάτου με κλασικό παράδειγμα του τα συνεχή στοιχήματα έχοντας ένα αρχικό ποσό. Έστω παίκτης A με ποσό(π.χ χρήματα ή μάρκες) $X_0 = n$ παίζοντας σε ένα καζίνο εναντίον ενός παίκτη (θα μπορούσε να είναι και το ίδιο το καζίνο) που ξεκινάει με ένα ποσό κ. Σε κάθε παιχνίδι-δοκιμή ο παίκτης A είτε χάνει 1 είτε κερδίζει 1 με πιθανότητα p, q αντίστοιχα. Το παιχνίδι σταματάει όταν ο A έχει συσσωρεύσει όλο το ποσό δηλαδή όταν $X_j = n + \kappa$ ή όταν ο παίκτης χρεοκοπήσει δηλαδή $X_j = 0$ έχοντας ξεκινήσει με το συγκεκριμένο ποσό n. Σε τέτοιες καταστάσεις μας ενδιαφέρει η πιθανότητα $P(X_n = 0) = ?$ (εδώ το n συμβολίζει το ποσό και όχι την δοκιμή). Λόγω των συνεχόμενων παιχνιδιών μπορούμε το p_n να το αναλύσουμε ως προς επόμενο παιχνίδι όπου λόγω του νόμου Ολικής Πιθανότητας προκύπτει: $p_n = pP_{n+1} + qP_{n-1}$ με την πιθανότητα p να αποτελεί την νίκη του παίκτη A παίρνοντας 1. Στην παρούσα φάση δεν θα προβούμε σε αποδείξεις του υπολογισμού της πιθανότητας της χρεοκοπίας του παίκτη A όμως ο τύπος υπολογισμού είναι ο εξής:

$$P(\text{νικη παικτη A} | i \text{ ποσο}) = P(i) = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} \quad (4.9)$$

Αυτό ισχύει στην περίπτωση όπου $p \neq q$ και για κάθε πιθανότητα σε ένα παιχνίδι. Στην περίπτωση όπου $p = q$ ισχύει το εξής:

$$P(i) = \frac{i}{N} \quad (4.10)$$

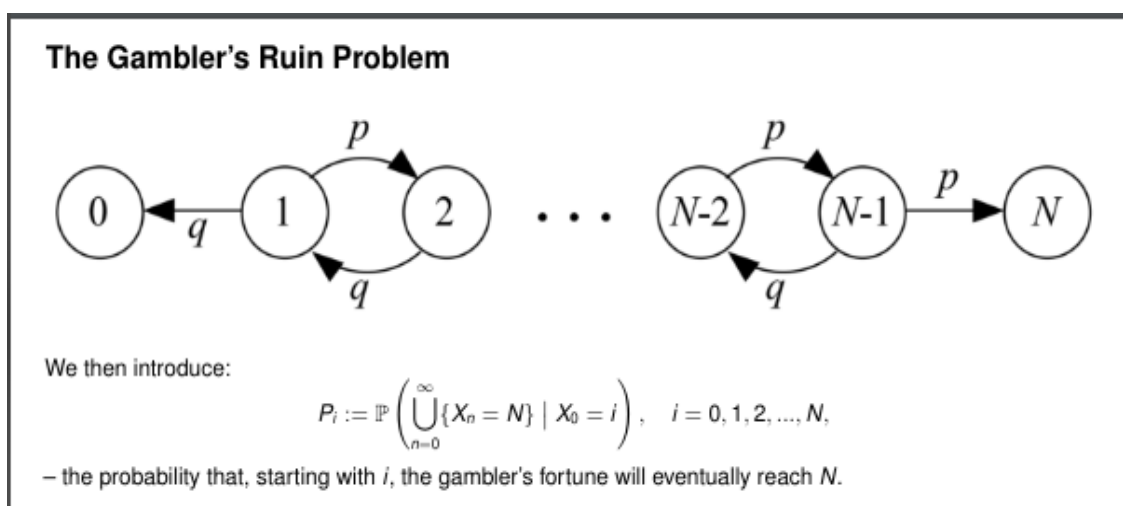
Και στις δύο περιπτώσεις το i αφορά το αρχικό ποσό που ξεκίνησε ο παίκτης το παιχνίδι των στοιχημάτων και το N αφορά το συνολικό ποσό που θα τερματίσει το παιχνίδι.

Για την μέση τιμή ισχύει:

$$E(X) = 1 + 0,5E[X - 1] + 0,5E[X + 1] \quad (4.11)$$

$$E(X) = Ax + B \quad (4.12)$$

με $X=A$ και $N-X=B$. Με τον τύπο να ακολουθεί την λογική της αναδρομής που αναφέρθηκε.



Εικόνα 4.10: Γράφος της στοχαστικής αλυσίδας “Καταστροφή του παίκτη”

4.4 ΠΡΟΣΟΜΟΙΩΣΗ ΔΙΑΔΙΚΑΣΙΑΣ ΚΑΤΑΣΤΡΟΦΗΣ ΠΑΙΚΤΗ

Η αναπαράσταση της καταστροφής του παίκτη θα γίνει μέσω της Python όπως και στα προηγούμενα παραδείγματα για να υπάρξει πιο σαφής η κατανόηση συγκεκριμένων συναρτήσεων. Στο παράδειγμα προσομοιώνουμε σε πολλές δοκιμές τα σενάρια στοιχημάτων (σε κάθε δοκιμή πολλά στοιχήματα) όπου παίζουν δύο παίκτες a και b με αρχικά ποσά ο καθένας θέλοντας να νικήσουν οι παίκτες, μηδενίζοντας ο ένας τον άλλον. Παρακάτω παρατίθεται ο κώδικας που το αποδεικνύει:

```
def gamblers_ruin_simulation(a_stakes, b_stakes, num_game):
    a_wins=0
    b_wins=0
    flips=[]
    steps=[]
```

```

for _ in range(num_game):
    step_num=0
    leader='0'
    flip_num=-1
    a=a_stakes
    b=b_stakes
    while a >0 and b >0:
        step_num+=1
        If np.random.rand() <=0.5:
            a+=1
            b-=1
        else:
            a-=1
            b+=1
        If a> a_stakes and leader !='A'
            leader='A'
            flip_num+=1
        elif a <a_stakes and leader !='B'
            leader='B'
            flip_num+=1
        If b==0:
            a_wins+=1
        else:
            b_wins+=1
        flips.append(flip_num)
        steps.append(step_num)
    steps_ave=np.mean(steps)
    prob_a_win=a_wins/num_game

```

```

prob_b_win=b_wins/num_game

flip_ave=np.mean(flips)

print(f'a start with {a_stakes}')

print(f'b start with {b_stakes}')

print(f'Average number of steps={step_ave}')

print(f'Expected number of steps={a_stakes*b_stakes}')

print(f'Average chance of A winning={prob_a_win:.3f}')

print(f'Expected chance of A winning={a_stakes/(a_stakes+b_stakes:.3f}')

print(f'Average chance of B winning={prob_b_win:.3f}')

print(f'Expected chance of B winning={b_stakes/(a_stakes+b_stakes:.3f}')

Print(f'Average number of flips={flip_ave}')

```

4.4.1 ΠΕΡΙΓΡΑΦΗ ΠΡΟΣΟΜΟΙΩΣΗΣ

- Καταρχήν ορίζεται η συνάρτηση με τα ποσά του κάθε παίκτη και των αριθμών των παιχνιδιών έτσι ώστε να μπορούμε να προβούμε σε δοκιμές αλλάζοντας τα.
- Ορίζεται ο συνολικός αριθμός των νικών του παίκτη a και b ως μηδέν για αρχικές συνθήκες καθώς δεν έχει ξεκινήσει το παιχνίδι (άρα δεν έχει νίκες κανένας) καθώς και την λίστα flips που δείχνει σε κάθε σημείο ποιος προηγείται/επικρατεί. Επίσης μια λίστα steps που καθορίζει την διάρκεια του παιχνιδιού που είναι κενή καθώς το παιχνίδι δεν έχει αρχίσει.
- Οι δοκιμές προσομοιώνονται με την επαναληπτική δομή for που κάθε ένα game_num αποτελεί μια δοκιμή/παιχνίδι ξεκινώντας με κάποιες συνθήκες. Δηλαδή έχουν οριστεί τα ποσά του καθενός να ισούνται με τα αρχικά καθώς έχει δηλωθεί μια μεταβλητή steps_num=0 (που συσχετίζεται με την αρχική λίστα steps) για να δείξει ότι δεν έχει ξεκινήσει η έναρξη των στοιχημάτων και της διαδικασίας γενικότερα και απλά έχουμε τα αρχικά ποσά (Θα μοντελοποιώντας μαθηματικά.
- Κάθε δοκιμή είναι ανεξάρτητη από την προηγούμενη. Έτσι με την while που ακολουθεί θα παιχτούν εντός μιας δοκιμής συνεχή στοιχήματα/πονταρίσματα τα οποία θα σταματήσουν με βάση την συνθήκη που έχουμε ορίσει (έχει οριστεί με την λογική ήττας ενός από των δύο δηλαδή να μην έχει άλλο ποσό).
- Στην αρχή κάθε while αυξάνεται το step_num επειδή σε κάθε βήμα θα έχουμε (όπως και στο θεωρητικό πρόβλημα με βήματαX₁,X₂) μια αλλαγή στα ποσά του παίκτη a και b.
- Ουσιαστικά το step και η τυχαία αλλαγή προσομοιώνονται με τα συνεχή στριψίματα ενός νομίσματος που συμβαίνει με την συνάρτηση rand της βιβλιοθήκης numpy. Αν ο τυχαίος αριθμός που παράγεται από την rand είναι μικρότερος από το 0.5 τότε συμβαίνει το γεγονός a +=1 δηλαδή η νίκη του a αλλιώς το αντίθετο.
- Εντός της while έπειτα ελέγχουμε αν το ποσό μετά τα στοιχήματα και την αλλαγή του αν είναι μεγαλύτερο του αρχικού ποσού και αν είναι τότε με την μεταβλητή leader='A' καθορίζουμε ποιος έχει το προβάδισμα από τους δύο. Η ίδια συνθήκη ανάλογα το ποσό επηρεάζει και τον παίκτη B.

- Η προσαύξηση των νικών του A (παρόμοια του B) έρχεται στην περίπτωση αν $b=0$ που γίνεται ± 1 πριν τερματιστεί ο βρόχος για το συγκεκριμένο `game_num`.
- Η μεταβλητή `a_wins` μαθηματικά μπορούμε να την φανταστούμε ως το δείγμα $X_1 + X_2 + \dots + X_n$.
- Στη συνέχεια προστίθεται στο τέλος της λίστας `flips` μέσω της `append()`, όσα `flip_num` συνέβησαν έτσι ώστε να φανούν οι συνολικές αλλαγές στο προβάδισμα του παιχνιδιού. Παρόμοια και για τα `step_num`.
- Δηλαδή τα `step_num` προσθέτονται με τη σειρά στη λίστα `steps` που δηλώνει σε κάθε θέση το κάθε παιχνίδι δείχνοντας πόσα βήματα χρειάστηκαν για να τελειώσει.
- Για το στατιστικό κομμάτι τα εμπειρικά αποτελέσματα τα λαμβάνουμε εφαρμόζοντας για την πιθανότητα νίκης του παίκτη a , τον τύπο του δειγματικού μέσου που είναι ένας πολύ καλός εκτιμητής πιθανοτήτων με θεωρητικό (μαθηματικό) τύπο: $\frac{X_1+X_2+\dots+X_n}{n} = P(\text{νίκη } a)$.
- Έπειτα η αναμενόμενη διάρκεια παιχνιδιού υπολογίζεται με την συνάρτηση `mean(steps)` στην λίστα `steps`.
- Καθώς και ο μέσος αριθμός αλλαγών στο προβάδισμα μέσω του `mean` στην λίστα `flips`.

4.4.2 ΣΤΑΤΙΣΤΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ ΚΑΤΑΣΤΡΟΦΗΣ ΠΑΙΚΤΗ

- Ασχοληθήκαμε με την περίπτωση όπου $p = q$ ως αποτέλεσμα ο θεωρητικός τύπος υπολογισμού της πιθανότητας στηρίζεται σε αυτή την περίπτωση.
- Το μοντέλο συμπεραίνουμε ότι μιμείται την ειδική περίπτωση διαδικασίας του “Απλού Τυχαίου Περιπάτου” η οποία είναι Απλός Τυχαίος Περίπατος με Απορροφητικά Φράγματα δηλαδή σημεία-δεδομένα που θα σταματήσει το παιχνίδι (εκεί η πιθανότητα παραμονής είναι 1). Αυτό απεικονίζεται στο πρόγραμμα μέσω των συνθηκών όπου αν φτάσει στο μηδέν ($a > 0$) το παιχνίδι σταματάει.
- Για την προσομοίωση της διαδικασίας χρησιμοποιήθηκε η μέθοδος Monte Carlo για μια συνεχής δειγματοληψία. Αυτό πραγματοποιήθηκε μέσω της συνεχούς εκτέλεσης της συνάρτησης `np.random.rand()` που χρησιμοποιήθηκαν τυχαίοι αριθμοί για την λύση του προβλήματος.
- Στο στατιστικό κομμάτι που είναι το πιο σημαντικό για την Επαλήθευση του μοντέλου, το μοντέλο παρήγαγε μέσους που καθορίζουν την συμπεριφορά του, μέσω στατιστικών μεθόδων και το μόνο που χρειάζεται για την αξιολόγηση του είναι η σύγκριση με τα θεωρητικά αποτελέσματα.
- Έχοντας καλή γνώση του θεωρητικού μοντέλου και της τυπολογίας που το περιγράφουν, υπολογίσαμε επίσης τις θεωρητικές παραμέτρους (της οποίες τις δείχνουμε στην οθόνη μέσω της `print`).
- Έγινε υπολογισμός το ένα κάτω από το άλλο έτσι ώστε να φαίνεται ο μέσος δειγματικός και η μέση θεωρητική τιμή.

4.5 ΑΞΙΟΛΟΓΗΣΗ ΜΟΝΤΕΛΟΥ ΠΡΟΣΟΜΟΙΩΣΗΣ

Πριν όμως βγάλουμε τα σαφείς συμπεράσματα μας, πρέπει να αναφέρουμε τον τύπο του Νόμου των Μεγάλων Αριθμών πάνω στο οποίο εξηγούνται οι συγκλίσεις των εμπειρικών δεδομένων μας (της προσομοίωσης). Έστω ο δειγματικός μέσος $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ στην περίπτωση που $n \rightarrow \infty$ τότε $\bar{X} \rightarrow \mu$ που σημαίνει η σύγκλιση του ως προς την μέση τιμή μ . Στην λογική αυτή χρησιμοποιούνται οι έτοιμες συναρτήσεις για τους δειγματικούς μέσους των `steps` και `flips` καθώς και ο υπολογισμός της πιθανότητας νίκης του a και του b με τον δειγματικό τύπο (όπως προαναφέραμε). Παρατηρούμε ότι τα

εμπειρικά δεδομένα υπακούουν στον τύπο δηλαδή υπάρχει η σύγκλιση τους με τους θεωρητικούς υπολογισμούς όπου διαπιστώνεται και η Επικύρωση του μοντέλου.

Ακολουθούν παραδείγματα εικόνων που εξηγούν αναλυτικότερα την σύγκριση και τις τιμές των παραμέτρων καθώς και αλλαγές από παραμετροποιήσεις.

Στην περίπτωση της συνάρτησης `gamblers_ruin_simulation(10,10,100)` παρατηρούμε τα εξής:

```
Average chance of A winning = 0.430
Expected chance of A winning = 0.500
Average chance of B winning = 0.570
Expected chance of B winning = 0.500
```

Εικόνα 4.11: Παράδειγμα 1 προσομοίωσης για παρατήρηση δεδομένων.

```
Average number of steps = 86.64
Expected number of steps = 100
```

Εικόνα 4.12: Παράδειγμα 1 των δεδομένων των δειγματικών μέσων.

Θα μπορούσαμε να πούμε ότι τα δεδομένα στην σύγκριση έχουν έναν φανερό βαθμό απόκλισης. Έστω η κλήση της συνάρτησης ως `gamblers_ruin_simulation(10,10,1000)`:

```
Average chance of A winning = 0.513
Expected chance of A winning = 0.500
Average chance of B winning = 0.487
Expected chance of B winning = 0.500
```

Εικόνα 4.13: Παράδειγμα 2 προσομοίωσης για παρατήρηση δεδομένων.

```
Average number of steps = 97.008
Expected number of steps = 100
```

Εικόνα 4.14: Παράδειγμα 2 των δεδομένων των δειγματικών μέσων

Με την κλήση της `gamblers_ruin_simulation(10,10,10000)` απεικονίζονται τα εξής:

```
Average chance of A winning = 0.496
Expected chance of A winning = 0.500
Average chance of B winning = 0.504
Expected chance of B winning = 0.500
```

Εικόνα 4.15: Παράδειγμα 3 προσομοίωσης για παρατήρηση δεδομένων

```
Average number of steps = 100.1264
Expected number of steps = 100
```

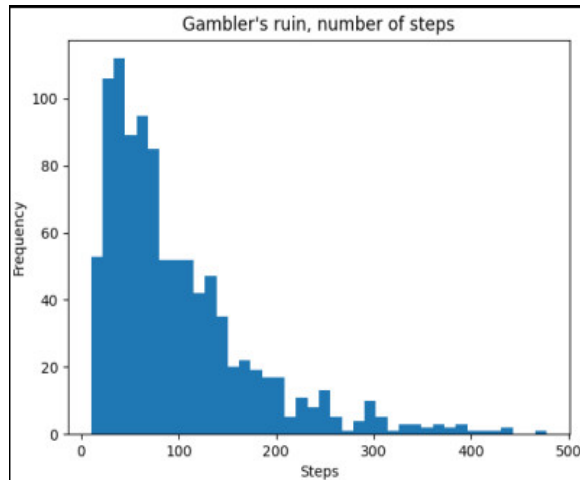
Εικόνα 4.16: Παράδειγμα 3 των δεδομένων των δειγματικών μέσων

Παρατηρούμε όσο μεγαλώνει το `game_num` (οι φορές της προσομοίωσης) τα δεδομένα συγκλίνουν ολοένα και περισσότερο στα θεωρητικά αποτελέσματα, φανερώνοντας την σωστή εφαρμογή του “Νόμου των Μεγάλων Αριθμών” και συνεπώς την ακρίβεια του μοντέλου να μιμείται το πραγματικό.

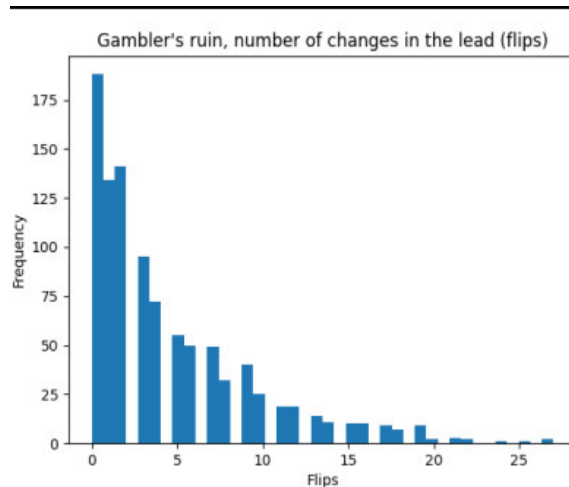
Οι σχέσεις της συχνότητας με τις ανατροπές στα παιχνίδια (αλλαγές στο προβάδισμα-flips) καθώς και με την διάρκεια του παιχνιδιού (μέσω της `steps`) αναλύονται μέσω διαγραμμάτων. Παρουσιάζοντας την κατανομή των αριθμών βημάτων μπορούμε να βγάλουμε συμπέρασμα για την χρονική διάρκεια του παιχνιδιού. Το διάγραμμα παρουσιάζει έμμεσα την πιθανότητα καθώς κάνουμε χρήση του ορισμού της συχνότητας που δείχνει τον αριθμό των παιχνιδιών με ένα συγκεκριμένο αριθμό `step_count`.

Στον άξονα τον `x` παρουσιάζετε ο αριθμός των βημάτων και στο `y` η συχνότητα των παιχνιδιών.

- Τρέχοντας την `gamblers_ruin_simulation()` 1000 φορές έχουμε τα εξής ιστογράμματα:



Εικόνα 4.17: Διάγραμμα 1 κατανομής των προσομοιώσεων



Εικόνα 4.18: Διάγραμμα 2 κατανομής των προσομοιώσεων

Συμπεραίνουμε από το πρώτο διάγραμμα ότι η συχνότητα (πιθανότητα) το παιχνίδι να τελειώσει νωρίς είναι μεγάλη (<100) και ότι όσο μεγαλώνουν τα βήματα άρα και περισσότερη διάρκεια παιχνιδιού, μειώνεται η συχνότητα. Για την λοξότητα της συνάρτησης κατανομής προς τα δεξιά, παρατηρούμε ότι παρόλο που τα περισσότερα παιχνίδια τείνουν να τελειώνουν νωρίς, υπάρχει ένα σημαντικό υποσύνολο κοντά στα 300 steps που η συχνότητα ανεβαίνει δείχνοντας ότι σε κάποια παιχνίδια παρατείνεται η διαδικασία ολοκλήρωσης τους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] J. Creighton, *A first course in probability models and Statistical Inference*. New York: Springer-Verlag, 1994, pp. 3-105
- [2] G. Keren and C. Lewis, “The two fallacies of Gamblers: Type I and Type II”, *Organisational Behaviour and Human Decision Processes*, vol 60, no.1, pp. 75-89, October 1994. Doi:10.1006/obhd.1994.1075.[Online].
<https://www.sciencedirect.com/science/article/abs/pii/S0749597884710752>
- [3] A. Pappoulis & S.U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 2002, pp.373-695
- [4] W.D. Wallis, *A Beginner's Guide to Discrete Mathematics*. Boston: Birkhäuser, 2003, pp. 186
- [5] T.E. Harris, *The Theory of Branching Process*. California: The Rand Corporation, 1994, pp. 2-4
- [6] G. Zitikovic. “Introduction to Stochastic Process”, Lecture Notes, University of Texas, Austin, USA, 2014
- [7] Μ. Ρουμελιώτης, *Προσομοίωση και Μοντελοποίηση*. Ελληνικό Ανοικτό Πανεπιστήμιο, 2001, pp. 11-162
- [8] J. P.C Kleijnen, “Simulation experiments in practice: Statistical design and regression analysis”, *Journal of Simulation*, vol 2, no. 1, pp. 19-27, 2008. doi: 10.1057/palgrave.jos.4250032
- [9] P. L. Bonate, “A Brief Introduction to Monte Carlo Simulation”, *Clin. Pharmacokinet*, vol. 40, no. 1, pp. 15-22, 2001. doi: 10.2165/00003088-200140010-00002.
- [10] M. Wagle, “Probability and Simulation”, Medium, Dec. 17, 2019. [Online]. Available: <https://medium.com/@manilwagle/probability-and-simulation-6a28fc1f1cb0>.