

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**

**Σχολή Χρηματοοικονομικής και Στατιστικής**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΚΛΑΔΩΤΕΣ ΑΝΕΛΙΞΕΙΣ: ΘΕΩΡΙΑ ΚΑΙ  
ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΕΠΙΔΗΜΙΟΛΟΓΙΑ**

Γεωργία Δ. Κόχειλα

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής  
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των  
απαιτήσεων για την απόκτηση του Μεταπτυχιακού  
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική.

Πειραιάς  
Σεπτέμβριος 2022

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. .... συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της Επιτροπής ήταν:

- Αναπληρωτής Καθηγητής Πολίτης Κωνσταντίνος (Επιβλέπων)
- Αναπληρωτής Καθηγητής Μπερσίμης Σωτήρης
- Αναπληρωτής Καθηγητής Τζαβελάς Γεώργιος

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

**UNIVERSITY OF PIRAEUS**  
**School of Finance and Statistics**



**Department of Statistics and Insurance Science**

**POSTGRADUATE PROGRAM IN**  
**APPLIED STATISTICS**

**Branching processes: theory and applications in**  
**epidemiology**

By

Georgia D. Kocheila

MSc Dissertation

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

Piraeus, Greece  
September 2022





*Στους γονείς μου,  
Θωμαΐς και Δημήτρη  
& στην αδερφή μου,  
Ελένη*



# Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, κύριο Πολίτη Κωνσταντίνο, για την συνεργασία και το αμέριστο ενδιαφέρον για την ολοκλήρωση της διπλωματικής μου εργασίας

Επιπλέον, θα ήθελα να απευθύνω θερμές ευχαριστίες στους καθηγητές Μπερσίμη Σωτήρη και Τζαβελά Γεώργιο, για τον χρόνο που διέθεσαν για την διόρθωση της διπλωματικής μου εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου και τους φίλους μου για την στήριξη και την εμπιστοσύνη τους.





## Περίληψη

Αφορμή για την παρούσα διπλωματική εργασία αποτέλεσε το ξαφνικό ξέσπασμα της πανδημίας του κορονοϊού *Covid-19*. Ανά τα χρόνια, πολλές είναι οι επιδημίες οι οποίες έπληξαν την ανθρωπότητα και κατέστησαν επιτακτική την ανάγκη για δημιουργία μαθηματικών μοντέλων, τα οποία μπορούν να δώσουν μια σαφή εικόνα για τη δυναμική των επιδημικών κυμάτων και να συμβάλουν στην έγκαιρη καταπολέμησή τους.

Στο Κεφάλαιο 2 γίνεται εισαγωγή στην επιδημιολογία όπου αναλύονται τα είδη των επιδημιολογικών μελετών, βασικοί επιδημιολογικοί δείκτες αλλά και μέτρα κινδύνου. Στο επόμενο κεφάλαιο, το Κεφάλαιο 3 αναπτύσσουμε την μεθοδολογία των στοχαστικών διαδικασιών, την κατηγοριοποίησή τους και την ανάλυση αυτών. Στο Κεφάλαιο 4 παρουσιάζεται αναλυτικά η τυπική θεωρία κλαδωτών ανελίξεων, τα μοντέλα κλαδωτών ανελίξεων συνεχούς χρόνου, οι κλαδωτές Ανελίξεις με ένα αναπαραγωγικό γεγονός ανά κύκλο ζωής και στην τελευταία ενότητα περιγράφονται τα επιδημιολογικά μοντέλα που στηρίζονται στις κλαδωτές ανελίξεις. Τέλος, στο 5<sup>ο</sup> κεφάλαιο γίνεται ανάλυση των μοντέλων *SIR* και *SEIR*, καθώς επίσης και εφαρμογή ενός μοντέλου *SIR* σε δεδομένα που αφορούν το πρώτο κύμα εξάπλωσης της πανδημίας του κορονοϊού στην Ελλάδα.



## Abstract

The reason behind this master's thesis was the sudden outbreak of the Covid-19 coronavirus pandemic. Over the years there have been many epidemics, which have affected humanity and necessitated the creation of mathematical models, which can provide an insight into the dynamics of epidemics and contribute to their prompt treatment

In Chapter 2 there is an introduction to epidemiology, where the types of epidemiological studies, basic epidemiological indicators and risk measures are analyzed. Thereinafter, Chapter 3 deals with stochastic processes, their categorization (discrete and continuous time) and their analysis. In Chapter 4, the standard branching processes, continuous time branching models as well as the branching with one reproductive event per life cycle are presented in detail, whereas in the last section there is a description of epidemiological models based on branching. Finally, in the 5th chapter there is an analysis of the *SIR* and *SEIR* models, and an *SIR* model is applied to data concerning the first wave of the spread of the coronavirus pandemic in Greece.



# Περιεχόμενα

Περίληψη .....	10
Abstract.....	xii
1.1 Πανδημίες και μαθηματικά μοντέλα .....	1
1.2 Στόχοι της εργασίας .....	2
1.3 Μεθοδολογία.....	2
2.1 Εισαγωγικές Έννοιες.....	3
2.1.1 Επιδημιολογία.....	3
2.2 Είδη Ιατρικών και Επιδημιολογικών Μελετών .....	4
2.3 Σχέσεις Μεταβλητών και Αιτιότητα .....	9
2.3.1 Ορισμός συγχυτικός ή συνεπιδρών παράγοντας: .....	9
2.3.2 Κριτήρια Αξιολόγησης Αιτιολογικών Σχέσεων.....	11
2.4 Δείκτες Νοσηρότητας.....	12
2.4.1 Επιπολασμός .....	12
2.4.2 Επίπτωση .....	12
2.4.3 Ρυθμός Επίπτωσης .....	13
2.4.4 Διάρκεια Ασθένειας.....	13
2.4.5 Θνησιμότητα.....	13
2.5 Μέτρα Κινδύνου.....	14
2.5.1 Σχετική Πιθανότητα ( <i>Odds</i> ) .....	14
2.5.2 Λόγος Σχετικών Πιθανοτήτων ( <i>Odds Ratio</i> ) .....	15
2.5.3 Σχετικός Κίνδυνος ( <i>Relative Risk</i> ).....	15
2.5.4 Οφειλόμενος ή Αποδιδόμενος Κίνδυνος ( <i>Attributable Risk</i> ) .....	16
2.6 Κλινικές Δοκιμές ( <i>Clinical Trials</i> ).....	17
2.6.1 Φάσεις Κλινικών Δοκιμών .....	18
2.6.2 Τύποι Κλινικών Δοκιμών.....	18
3.1 Εισαγωγή.....	20
3.2 Εισαγωγικές Έννοιες και ορισμοί.....	21
3.2.1 Κατηγοριοποίηση Στοχαστικών Ανελίξεων.....	21
3.2.2 Εξάρτηση τυχαίων μεταβλητών μιας στοχαστικής ανάλυσης .....	22
3.2.2.1 Στάσιμες Στοχαστικές Ανελίξεις.....	22

3.2.2.2	Μαρκοβιανή Αλυσίδα ή Μαρκοβιανή Ανέλιξη .....	23
3.3	Στοχαστικές διαδικασίες διακριτού χρόνου.....	24
3.3.1	Μαρκοβιανές Ανελίξεις διακριτού χρόνου και χώρου καταστάσεων .....	24
3.3.2	Ταξινόμηση των καταστάσεων μίας Μαρκοβιανής αλυσίδας σε διακριτό χρόνο .....	28
3.4	Στοχαστικές διαδικασίες συνεχούς χρόνου .....	29
3.4.1	Στοχαστική ανέλιξη <i>Poisson</i> .....	30
3.4.2	Απλή ανέλιξη Γεννήσεως-Θανάτου .....	33
4.1	Θεωρητικό πλαίσιο διαδικασιών <i>Bienaymé-Galton-Watson</i> .....	36
4.2.	Στοιχειώδης θεωρία και έννοιες κλαδωτών ανελίξεων.....	38
4.2.1	Τα μοντέλα κλαδωτών ανελίξεων συνεχούς χρόνου .....	48
4.2.2	Κλαδωτές Ανελίξεις με ένα αναπαραγωγικό γεγονός ανά κύκλο ζωής.....	50
4.2.3	Επιδημιολογικά Μοντέλα Στηριζόμενα σε Κλαδωτές Ανελίξεις .....	54
Κεφάλαιο 5.....		56
Εφαρμογή.....		56
5.1	Εισαγωγή.....	56
5.2	Μοντέλο <i>SIR</i> .....	57
5.3	Αριθμητική εφαρμογή με δεδομένα κορωνοϊού ( <i>Covid - 19</i> ), για την περίπτωση της Ελλάδας.....	60
5.4	Εφαρμογή του υπολογιστικού πακέτου <i>covid19.analytics</i> της <i>R</i> .....	61
5.5	Μοντέλο <i>SEIR</i> .....	69
5.6	Σύγκριση Μοντέλων <i>SIR</i> & <i>SEIR</i> .....	72
BIBΛΙΟΓΡΑΦΙΑ.....		74





# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

### 1.1 Πανδημίες και μαθηματικά μοντέλα

Υπάρχει μακρά παράδοση στη χρήση μαθηματικών μοντέλων για τη εξαγωγή γνώσεων και συμπερασμάτων σχετικά με τη δυναμική μετάδοσης των μολυσματικών ασθενειών και την αξιολόγηση του πιθανού αντίκτυπου διαφορετικών στρατηγικών παρέμβασης (*Heesterbeek et al.*, 2015). Η αυξανόμενη χρήση μαθηματικών μοντέλων για την πρόβλεψη επιδημιών έχει τονίσει τη σημασία του σχεδιασμού αξιόπιστων μοντέλων που αποτυπώνουν τα βασικά χαρακτηριστικά μετάδοσης συγκεκριμένων παθογόνων παραγόντων και κοινωνικών πλαισίων.

Τα επιδημιολογικά μοντέλα χωρίζονται σε δύο μεγάλες κατηγορίες: στατιστικά μοντέλα που βασίζονται σε μεγάλο βαθμό σε δεδομένα και μηχανιστικά μοντέλα που βασίζονται σε υποκείμενες θεωρητικές αρχές που αναπτύχθηκαν από επιστήμονες σχετικά με τον τρόπο εξάπλωσης της νόσου. Προκειμένου να δημιουργηθούν προβλέψεις επιδημίας που είναι χρήσιμες για τη λήψη αποφάσεων για τη δημόσια υγεία, υπάρχει ανάγκη να σχεδιαστούν αξιόπιστα μοντέλα που να αποτυπώνουν τα βασικά χαρακτηριστικά μετάδοσης για συγκεκριμένα παθογόνα και κοινωνικά πλαίσια (*Adiga et al.*, 2020).

Τα μοντέλα που βασίζονται σε δεδομένα χρησιμοποιούν στατιστικές μεθόδους και μεθόδους μηχανικής μάθησης για την πρόβλεψη των αποτελεσμάτων, όπως ο αριθμός των περιπτώσεων, η θνησιμότητα και οι νοσοκομειακές απαιτήσεις. Αυτός είναι ένας πολύ ενεργός τομέας έρευνας και έχει αναπτυχθεί μια ευρεία κατηγορία τεχνικών, συμπεριλαμβανομένων μεθόδων αυτόματης παλινδρόμησης χρονοσειρών, Μπεϋζιανών τεχνικών και βαθιάς μάθησης (*Adhikari et al.*, 2019). Μηχανιστικά μοντέλα εξάπλωσης της νόσου σε πληθυσμοί (*Marathe & Vullikanti*, 2013) χρησιμοποιούν μηχανιστικές (γνωστές και ως διαδικαστικές ή αλγοριθμικές) μεθόδους για να περιγράψουν την εξέλιξη μιας επιδημίας μέσω ενός πληθυσμού. Το πιο συνηθισμένο από αυτά είναι τα μοντέλα τύπου *SIR*. Τα υβριδικά μοντέλα που συνδυάζουν μηχανιστικά μοντέλα με προσεγγίσεις μηχανικής μάθησης βάσει δεδομένων αρχίζουν επίσης να γίνονται δημοφιλή,

## 1.2 Στόχοι της εργασίας

Στόχος της εργασίας είναι η παρουσίαση βασικών εννοιών που αφορούν τις στοχαστικές διαδικασίες και συγκεκριμένα το πεδίο των κλαδωτών ανελίξεων καθώς και την εφαρμογή τους στην επιδημιολογία πληθυσμών. Όπως και σε άλλες πανδημίες, οι ερευνητές και οι υπεύθυνοι χάραξης πολιτικής για τη δημόσια υγεία ενδιαφέρονται για ερωτήματα όπως πώς είναι πιθανό να προχωρήσει και πώς μπορεί να ελεγχθεί η πανδημία, ποια είναι η δυνητική παρέμβαση εξισορροπώντας παράλληλα τη δημόσια υγεία και τον οικονομικό αντίκτυπο, γιατί ορισμένες χώρες τα πήγαν καλύτερα από άλλες χώρες κατά τη διάρκεια της πανδημίας και άλλα. Ειδικότερα, τα μοντέλα και οι προβολές/προβλέψεις τους έχουν λάβει πρωτοφανή προσοχή. Με ένα πλήθος πλαισίων μοντελοποίησης, υποκείμενες υποθέσεις, διαθέσιμα σύνολα δεδομένων και την περιοχή/χρονικό πλαίσιο που μοντελοποιείται, αυτές οι προβλέψεις ποικίλλουν ευρέως. Μία επεξήγηση δύο εξ αυτών των μοντέλων, με σύγκριση της εφαρμογής τους, ενδεχομένως θα ωφελήσει τους ερευνητές και θα χρησιμεύσει επίσης ως ιστορικό αρχείο για μελλοντικές προσπάθειες.

## 1.3 Μεθοδολογία

Για το θεωρητικό μέρος της παρούσας εργασίας έλαβε χώρα βιβλιογραφική επισκόπηση για την διερεύνηση των κατάλληλων δεδομένων τόσο για το θεωρητικό πλαίσιο που περιλαμβάνει την επιδημιολογία και τις ανάλογες μελέτες, όσο και για τα μαθηματικά μοντέλα που χρησιμοποιούνται σε αυτόν τον κλάδο. Χρησιμοποιήθηκε βασική βιβλιογραφία για τη θεωρία που αναφέρεται σε μεταβλητές, δείκτες αλλά και στοχαστικές διαδικασίες και διερευνήθηκε η διεθνής βιβλιογραφία για τις έρευνες που έχουν λάβει χώρα για την πανδημία της *Covid - 19*, με έμφαση στις έρευνες που διεξήχθησαν στην Ελλάδα.

Εν συνεχεία, για το κεφάλαιο όπου εφαρμόζεται το μοντέλο *SIR*, αρχικά αναφέρεται η χρησιμότητά του ιστορικά σε επιδημιολογικές μελέτες, αλλά και οι λόγοι που προτιμάται έναντι άλλων. Ταυτόχρονα στο ίδιο κεφάλαιο εφαρμόζεται το εν λόγω μοντέλο στα επιδημιολογικά δεδομένα της Ελλάδας, αλλά και γίνεται αναφορά στην εφαρμογή του μοντέλου *SEIR*, το οποίο είναι επαυξημένο κατά μία πληθυσμιακή ομάδα, και επιχειρείται σύγκριση και των δύο μοντέλων.

# ΚΕΦΑΛΑΙΟ 2

## Εισαγωγή στην Επιδημιολογία

### 2.1 Εισαγωγικές Έννοιες

Προτού προχωρήσουμε στην ανάπτυξη του κεφαλαίου θα χρειαστεί να ορίσουμε δυο σημαντικές ορολογίες της επιστήμης της Επιδημιολογίας. Αρχικά με τον ως Υγεία, ορίζουμε την κατάσταση της πλήρους σωματικής, ψυχικής και κοινωνικής ευεξίας και όχι απλά την απουσία νόσου από τον οργανισμό (Π.Ο.Υ. -1946). Επιπλέον, ως Νόσος χαρακτηρίζεται οποιαδήποτε κατάσταση εμποδίζει την φυσιολογική λειτουργικότητα του ανθρώπινου οργανισμού.

#### 2.1.1 Επιδημιολογία

Ως επιδημιολογία ορίζεται η επιστήμη, η οποία μελετά την κατανομή και την εξέλιξη διαφόρων νοσημάτων ή χαρακτηριστικών στον ανθρώπινο πληθυσμό και των παραγόντων που τις διαφοροποιούν ή μπορεί να τις επηρεάσουν. Εξετάζει χαρακτηριστικά που συνδέονται με τον περιβαλλοντικό όσο και με τον γενετικό έλεγχο. (Ντζούφρας & Περπέρογλου, 2009).

Αντικείμενο της επιδημιολογίας είναι και η περιγραφική κατανομή των παραγόντων που επηρεάζουν τον πληθυσμό. Ενδεικτικά: φύλο, ηλικία, γεωγραφική περιοχή και επάγγελμα, ο κλάδος αυτός ονομάζεται «περιγραφική επιδημιολογία». Επιπρόσθετα, οι αιτίες που προκαλούν ασθένειες ή αλλαγές στον ανθρώπινο πληθυσμό είναι αντικείμενο της «επαγωγικής» ή «αναλυτικής επιδημιολογίας». Στην επιδημιολογία κάθε νόσος περιγράφεται κυρίως από την συχνότητα εμφάνισης, το χρόνο διάρκειας και την εξέλιξη της (Ντζούφρας & Περπέρογλου, 2009).

Ως στόχοι της επιδημιολογίας ορίζονται οι ακόλουθοι:

- Η μέτρηση και περιγραφή της νοσηρότητας & θνησιμότητας ενός πληθυσμού
- Ο έλεγχος της χρονικής εξέλιξης μιας νόσου

- Η ανακάλυψη των αιτιακών παραγόντων εκείνων που προκαλούν μια νόσο
- Η μελέτη των συνθηκών και των αιτιών που προκαλούν επιδημίες
- Η πρόληψη μέσω της κατανομής του ιστορικού των νοσημάτων
- Η ταξινόμηση των νοσημάτων
- Ο προγραμματισμός, η οργάνωση και η αξιολόγηση των υπηρεσιών υγείας.

## 2.2 Είδη Ιατρικών και Επιδημιολογικών Μελετών

Οι Ιατρικές και Επιδημιολογικές μελέτες χωρίζονται σε δύο μεγάλες κατηγορίες. Η πρώτη κατηγορία αφορά τις Πειραματικές ή Παρεμβατικές μελέτες (*Experimental – Intervention Studies*). Σε αυτήν συμπεριλαμβάνονται οι κλινικές δοκιμές (*Clinical Trials*), οι οποίες με τη σειρά τους χωρίζονται σε Τυχαιοποιημένες και Ήμι-πειραματικές μελέτες. Στην δεύτερη κατηγορία εντάσσονται οι μελέτες Παρατήρησης ή Μη Πειραματικές ή Μη Παρεμβατικές μελέτες (*Observational Medical Surveys*), ενώ παράλληλα αυτές κατηγοριοποιούνται σε Περιγραφικές ή Διατμηματικές - Συγχρονικές μελέτες (*Descriptive surveys, cross-sectional studies*) και Αναλυτικές ή Περιγραφικής επιδημιολογίας μελέτες (*Analytic / Aetiologic studies*). Στις αναλυτικές μελέτες εντάσσονται οι Προοπτικές Μελέτες ή Μελέτες κοορτής ή Μελέτες παρακολούθησης (*Prospective or Cohort or Follow-up studies*) και οι Αναδρομικές μελέτες ή μελέτες μαρτύρων- ασθενών (*Retrospective or Case -Control studies*). (Ντζούφρας & Περπέρογλου, 2009).

Ο διαχωρισμός των Μελετών σε Πειραματικές Μελέτες και Μελέτες Παρατήρησης βασίζεται στο αν ο ερευνητής παρεμβαίνει ή όχι στη διαδικασία. Στις Πειραματικές Μελέτες ο ερευνητής παρεμβαίνει στο πείραμα ελέγχοντας την κατανομή των ατόμων στις ομάδες ελέγχου. Αντίθετα, στις Μελέτες Παρατήρησης ο ερευνητής δεν παρεμβαίνει στο πείραμα, απλώς παρατηρεί και καταγράφει. Βασικός σκοπός στην προκειμένη περίπτωση είναι η διερεύνηση των αιτιών που προκαλούν μία νόσο και η κατασκευή διαγνωστικών-προγνωστικών μοντέλων.

Εν συνεχεία στις Πειραματικές Μελέτες ο ερευνητής παρεμβαίνει ενεργητικά στον καθορισμό των ομάδων, της θεραπείας ή άλλων χαρακτηριστικών. Οι Πειραματικές Μελέτες χωρίζονται σε τέσσερις υποκατηγορίες:

1. Εκούσια και σχεδιασμένη εφαρμογή ενός πιθανού αιτιολογικού μοντέλου σε μια ομάδα και παρακολούθηση των ευρημάτων.

2. Εκούσια απομάκρυνση ενός παράγοντα όπου η απομάκρυνση γίνεται με συγκεκριμένο τρόπο και όχι κατ' επιλογή των συμμετεχόντων.
3. Πειράματα φύσης στα οποία εξετάζονται διαφορετικές ομάδες που δημιουργήθηκαν από ένα γεγονός.
4. Σχεδιασμένη τροποποίηση παθογενών μηχανισμών σε μια ομάδα ανθρώπων.

Οι Αιτιολογικές Μελέτες ή Μελέτες Επαγωγικής Επιδημιολογίας έχουν ως στόχο τον έλεγχο μιας επιδημιολογικής υπόθεσης ενώ παράλληλα αποσκοπούν στην διερεύνηση της αιτιότητας ενός νοσήματος. Αυτές οι μελέτες αποτελούν βασικό πυλώνα της σύγχρονης Επιδημιολογίας. Σε αυτή την κατηγορία μπορούμε να εντάξουμε τόσο μελέτες παρατήρησης (όπως οι προοπτικές και αναδρομικές μελέτες) όσο και παρεμβατικές μελέτες (κλινικές δοκιμές). Η διαφορά μεταξύ των μελετών παρατήρησης και των κλινικών δοκιμών είναι ότι στις μελέτες παρατήρησης βασικός στόχος είναι ο εντοπισμός παραγόντων κινδύνου που αυξάνουν την εμφάνιση μιας νόσου, συνεπώς ο στόχος είναι διαγνωστικός και προληπτικός. Εν αντιθέσει με τις κλινικές όπου ο βασικός στόχος είναι ο εντοπισμός νέων φαρμάκων και θεραπειών που είτε περιορίζουν τα συμπτώματα μίας νόσου είτε οδηγούν στην πλήρη ίαση (θεραπεία).

Στις Μελέτες Παρατήρησης, ο ερευνητής δε συμμετέχει αλλά παρατηρεί και καταγράφει. Μπορούν να χωριστούν σε τρεις βασικές κατηγορίες:

- i. Διατμηματικές – σύγχρονες μελέτες

Σε αυτές τις μελέτες δεν λαμβάνουμε υπόψη τη χρονική αλληλουχία μεταξύ των μεταβλητών (*temporal relationship*). Εξετάζουμε ένα τυχαίο δείγμα σε συγκεκριμένο τόπο και χρόνο. Προκειμένου να επιτευχθεί αυτό χρησιμοποιούνται οι συνηθισμένοι τρόποι δειγματοληψίας όταν διαθέτουμε μεγάλους πληθυσμούς, είτε γίνεται απογραφικά όταν ο πληθυσμός είναι περιορισμένος. Οι διατμηματικές-σύγχρονες μελέτες έχουν ως στόχο τη μελέτη των σχέσεων μεταξύ μεταβλητών (ασθενειών και παραγόντων κινδύνου). Σε πολλές από αυτές τις μελέτες κύριος στόχος είναι η εκτίμηση του επιπολασμού μίας ασθένειας. Οι διατμηματικές- σύγχρονες μελέτες έχουν εφαρμογή και για συμπερασματολογία που αφορά τα υγιή άτομα που συμμετέχουν στη μελέτη και κατ' επέκταση τους πληθυσμούς υγιών ατόμων που αντιπροσωπεύουν.

- ii. Προοπτικές μελέτες

Για την συγκεκριμένη κατηγορία μελετών έχουν χρησιμοποιηθεί διαφορετικοί όροι όπως μελέτες κοορτής (*Cohort Studies*), παρακολούθησης (*Follow-up*) και διαχρονικές-διαμήκεις (*Longitudinal*). Η ονομασία κοορτή (*Cohort*) είναι στρατιωτικός όρος και αναφέρεται στο ένα δέκατο μιας Ρωμαϊκής λεγεώνας. Στις προοπτικές μελέτες έχει την έννοια της ομάδας ή σειράς ατόμων την οποία παρακολουθούμε κατά τη διάρκεια του χρόνου. Όσον αφορά τις διαχρονικές-διαμήκεις (*Longitudinal*) έρχονται σε αντιδιαστολή με τις συγχρονικές-διατμηματικές μελέτες και αφορούν μελέτες που γίνονται σε πολλά διαφορετικά χρονικά και γεωγραφικά σημεία, πρωτεύοντος αναφέρονται στην επαναλαμβανόμενη χρονικά διενέργεια μιας μελέτης στον ίδιο αριθμό ατόμων. Βασικό τους χαρακτηριστικό αποτελούν οι επαναλαμβανόμενες μετρήσεις στα ίδια άτομα σε διαφορετικές χρονικές στιγμές.

Το κύριο χαρακτηριστικό των προοπτικών μελετών είναι ότι επιλέγουμε μία ή περισσότερες ομάδες ατόμων και τις παρακολουθούμε στο χρόνο με σκοπό να παρατηρήσουμε την εξέλιξη τους ως προς κάποιες ασθένειες ή πιθανούς παράγοντες κινδύνου. Η πιο απλή εφαρμογή προοπτικής μελέτης αποτελείται από μία συγχρονική μελέτη όπου καταγράφονται τα βασικά στοιχεία του ασθενή ή του ατόμου που εξετάζουμε και έπειτα καταγράφουμε το χρόνο θανάτου ή τον τελευταίο χρόνο όπου γνωρίζαμε ότι το άτομο ήταν στη ζωή (μελέτες θνησιμότητας ή χρόνου επιβίωσης).

Οι προοπτικές μελέτες αποτελούν ένα αρκετά αξιόπιστο τρόπο διερεύνησης αιτιολογικών σχέσεων στην επιδημιολογία. Τα κύρια χαρακτηριστικά τους είναι ή εκμηδένιση των σφαλμάτων λόγω καταγραφής ή ανάκλησης μνήμης. Είναι σημαντικό ότι λαμβάνεται υπόψη ο χρόνος και η χρονική αλληλουχία των σχέσεων και μπορεί να εκτιμηθεί η μεταβλητότητα που οφείλεται στα ίδια τα άτομα.

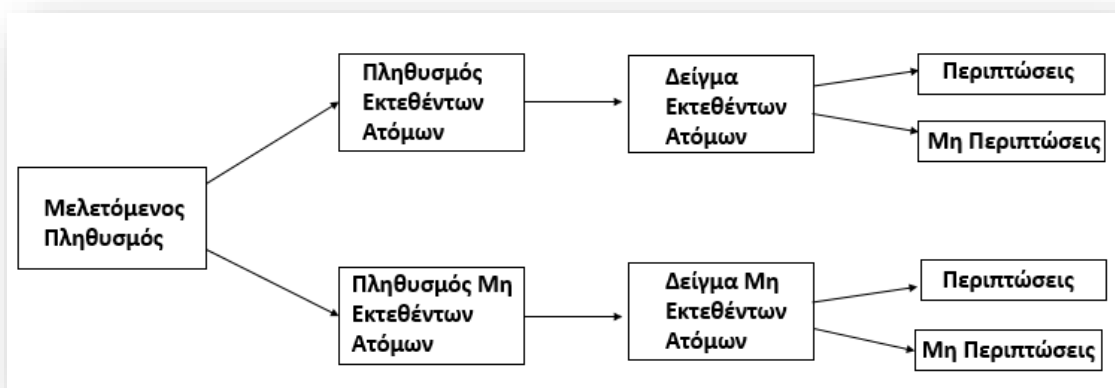
Μία σημαντική κατηγορία προοπτικών μελετών είναι αυτή που βασίζεται στα ιστορικά αρχεία των ατόμων (*historical record cohort studies*). Σε αυτές τις μελέτες επιλέγουμε την ομάδα που μας ενδιαφέρει και εξετάζουμε την πορεία της μέσα από «αντικειμενικά» ιστορικά αρχεία, στα οποία δεν αντιμετωπίζουμε σφάλμα ανάκλησης μνήμης. Παρ' όλα αυτά ενδέχεται η αντιμετώπιση σφαλμάτων τα οποία οφείλονται στην αξιοπιστία και την ακρίβεια των ιατρικών αρχείων (Ντζούφρας & Περπέρογλου, 2009).

Πλεονέκτημα των Προοπτικών Μελετών αποτελεί το γεγονός πως μπορούμε να μελετήσουμε την επίδραση σπάνιων παραγόντων κινδύνου. Λαμβάνοντας μία ομάδα με άτομα που έχουν εκτεθεί στον κίνδυνο και μια δεύτερη ομάδα με άτομα που δεν έχουν εκτεθεί στον κίνδυνο (ομάδα ελέγχου). Είναι ιδιαίτερα αξιόπιστες καθώς περιορίζονται τα σφάλματα λόγω

ανάκλησης μνήμης, καταγραφής και προκατάληψης. Σε αντιδιαστολή με τις αναδρομικές μελέτες μπορούμε να υπολογίσουμε απόλυτα και σχετικά μέτρα κινδύνου. Μπορούμε να μελετήσουμε τη σχέση της νόσου που μας ενδιαφέρει σε σχέση με άλλα νοσήματα.

Ωστόσο, οι Προοπτικές Μελέτες έχουν και αρκετά σημαντικά μειονεκτήματα. Αρχικά, για την διεξαγωγή μιας τέτοιας μελέτης απαιτείται μεγάλος αριθμός ατόμων, ακόμη και για ασθένειες με συχνή εμφάνιση. Κάτι τέτοιο είναι αδύνατο να συμβεί όταν η υπό μελέτη νόσος είναι πολύ σπάνια στον πληθυσμό. Επιπλέον, είναι αρκετά χρονοβόρες και έχουν ιδιαίτερα μεγάλο κόστος, τόσο ως προς ανθρώπινο δυναμικό που απαιτείται, όσο ως προς τον εξοπλισμό που χρειάζεται για την πραγματοποίησή τους. Τέλος, η ανάλυση των δεδομένων είναι ιδιαίτερα περίπλοκη λόγω του όγκου δεδομένων αλλά και τις φύσης τους (επαναλαμβανόμενες μετρήσεις σε μεγάλο αριθμό ατόμων, πλήθος ελλιπών τιμών). Το παραπάνω έχει ως αποτέλεσμα την ανάγκη απασχόλησης εξειδικευμένου προσωπικού για μεγάλο χρονικό διάστημα.

Ένα σημαντικό χαρακτηριστικό που μπορούμε να συναντήσουμε σε μια προοπτική μελέτη είναι το γεγονός ότι οι συμμετέχοντες εισέρχονται και εξέρχονται από αυτή σε διαφορετική χρονική στιγμή. Επιπλέον, πολύ συχνά θα χρειαστεί να αντιμετωπιστεί το φαινόμενο όπου έχουμε μη πλήρεις τιμές, αυτή η περίπτωση ονομάζεται λογοκρισία (*Censoring*) και τα δεδομένα αυτά λογοκριμένα (*Censored*). Τέλος, τα δύο κύρια χαρακτηριστικά μιας προοπτικής μελέτης είναι οι ομάδες-κοορτές έκθεσης στον κίνδυνο και ο χρόνος παρακολούθησης.



Σχήμα 2.1. Σχηματική Παράσταση Προοπτικής Μελέτης

Οι Αναδρομικές Μελέτες συλλέγουν, από το παρόν, δεδομένα που αφορούν το παρελθόν, ανατρέχοντας έτσι χρονικά στο παρελθόν. Οι συμμετέχοντες ερωτώνται για συνήθειες που είχαν

στο παρελθόν. Συνεπώς, οι μελέτες αυτές βασίζονται σε αναμνηστική ανάκληση πληροφοριών (*Recall Information*) και για αυτό το λόγο είναι πολύ πιθανό να εμπεριέχουν σφάλμα μέτρησης.

Οι πιο συνηθισμένες αναδρομικές μελέτες είναι αυτές των μαρτύρων - ασθενών (*Case - Control Studies*). Η σύνθεσή τους συνήθως αποτελείται από δύο ομάδες, την ομάδα των περιπτώσεων ή ασθενών (*Cases*, ατόμων που νοσούν) και των μαρτύρων (*Controls*, ατόμων που δεν νοσούν).

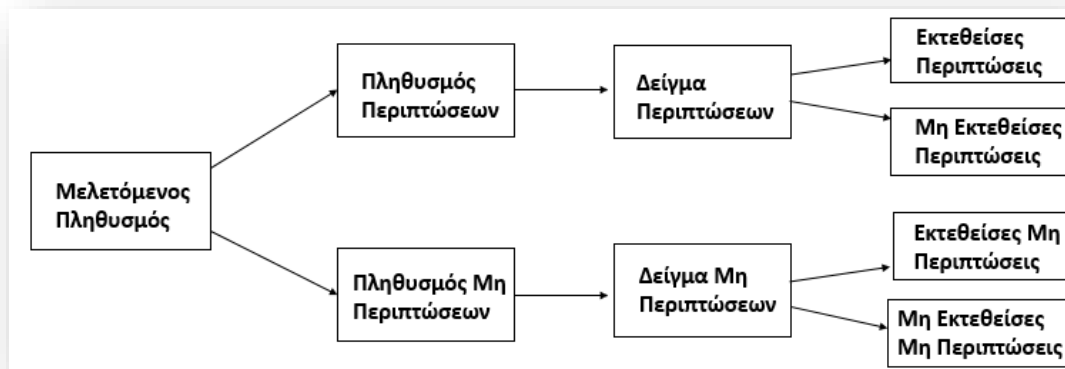
Οι περιπτώσεις αποτελούν την ομάδα των ασθενών (*Disease Group*) ενώ οι μάρτυρες αποτελούν την ομάδα ελέγχου (*Control Group*). Τα άτομα που νοσούν δεν επιλέγονται τυχαία αλλά συνήθως απογραφικά από μία ή περισσότερες νοσοκομειακές μονάδες που ειδικεύονται στη νόσο που θέλουμε να εξετάσουμε. Οι μάρτυρες μπορούν να επιλεγούν από το γενικό πληθυσμό, συνήθως όμως επιλέγονται από άλλα τμήματα των ίδιων νοσοκομειακών μονάδων από τα οποία έχουμε επιλέξει τους ασθενείς. Ο λόγος επιλογής των μαρτύρων από το ίδιο νοσοκομείο είναι ότι με αυτόν τον τρόπο διατηρούμε διάφορους κοινωνικό-οικονομικούς συσχετικούς παράγοντες σταθερούς. Στόχος αυτών των μελετών είναι να εξετάσουμε αν υπάρχει διαφοροποίηση της έκθεσης στον κίνδυνο για τις δύο ομάδες.

Τα στάδια σχεδιασμού των Αναδρομικών Μελετών είναι τα παρακάτω:

1. Η επιλογή των ασθενών
2. Η επιλογή των μαρτύρων
3. Ο τρόπος ορισμού και μέτρησης του παράγοντα κινδύνου
4. Η ανάλυση της σχέσης νόσου-παράγοντα κινδύνου

Και σε αυτό το είδος των μελετών συναντάμε πολλά πλεονεκτήματα. Αρχικά, απαιτούν σχετικά μικρό δείγμα για την διεξαγωγή τους ενώ είναι χρονικά περιορισμένες. Συγκριτικά με άλλου είδους μελέτες έχουν μικρό κόστος και σχετικά εύκολη ανάλυση. Επιπλέον, μελετούν ικανοποιητικά την σχέση νόσου και παράγοντα κινδύνου, ενώ τέλος μπορούν να χρησιμοποιηθούν για την μελέτη σπάνιων ασθενειών.





Σχήμα 2.2. Σχηματική Παράσταση Αναδρομικής Μελέτης

Στα μειονεκτήματα της μεθόδου, είναι πως ως μελέτες είναι επιρρεπείς σε σφάλματα μέτρησης, ενώ μπορούμε να υπολογίσουμε μόνο τα σχετικά μέτρα κινδύνου και όχι τα απόλυτα. Επιπλέον, απαιτείται πολύ προσοχή στην επιλογή της ομάδας ελέγχου. Αν η επιλογή του δείγματος δεν γίνει προσεκτικά, τότε αυξάνεται η πιθανότητα μεροληπτικής επιλογής των συμμετεχόντων (*Selection Bias*). Σε σύγκριση με τις προοπτικές μελέτες, διαπιστώνουμε συχνότερη εμφάνιση πλασματικών σχέσεων. Τέλος, δεν διαθέτουν ισχυρό αποδεικτικό κύρος, καθώς εξετάζουν τις σχέσεις ανάποδα όσον αφορά την χρονική τους αλληλουχία.

Παρόλα αυτά οι αναδρομικές μελέτες είναι ένα πολύ σημαντικό εργαλείο το οποίο έχει επικρατήσει στην ιατρική έρευνα. Οι αναδρομικές μελέτες μπορούν να λειτουργούν ως διερευνητικές και επικουρικές των μεγάλων προοπτικών μελετών.

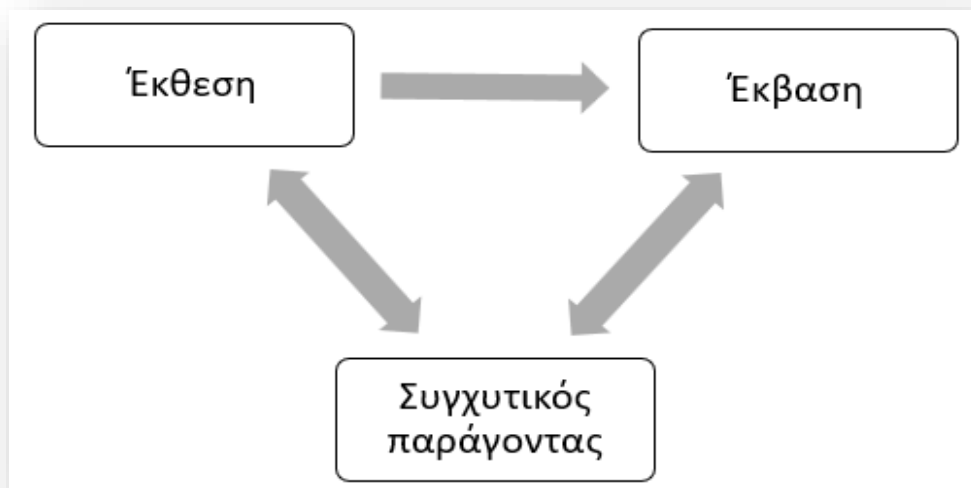
## 2.3 Σχέσεις Μεταβλητών και Αιτιότητα

Στην εμφάνιση μίας ασθένειας συνεπιδρούν πολλοί διαφορετικοί παράγοντες. Συνεπώς, η σχέση της νόσου με τον παράγοντα κινδύνου πιθανότατα επηρεάζεται από άλλους παράγοντες οι οποίοι ονομάζονται συγχυτικοί.

### 2.3.1 Ορισμός συγχυτικός ή συνεπιδρών παράγοντας:

Ως Συγχυτικό παράγοντα (*Confounding Factor*) ορίζουμε μια μεταβλητή που σχετίζεται και με την ασθένεια αλλά και με τον παράγοντα κινδύνου που εξετάζουμε. Οι συγχυτικοί παράγοντες πρέπει να λαμβάνονται υπόψη πριν να εξετάσουμε τη σχέση ασθένειας – παράγοντα κινδύνου. Ένας συγχυτικός παράγοντας ονομάζεται αιτιολογική δίοδος αν ο παράγοντας κινδύνου επηρεάζει το συγχυτικό παράγοντα και αυτός με τη σειρά του τη νόσο. Αν η ισχύς της

σχέσης νόσου – παράγοντα κινδύνου αλλάζει για διαφορετικά επίπεδα ενός συγχυτικού παράγοντα τότε υπάρχει αλληλεπίδραση παράγοντα κινδύνου και συγχυτικού παράγοντα πάνω στη νόσο ή αλλιώς διαφοροποίηση της επίδρασης του συγχυτικού παράγοντα στη νόσο για διαφορετικά επίπεδα του συγχυτικού παράγοντα (*Effect Modification*). Αυτός ο συγχυτικός παράγοντας ονομάζεται διαφοροποιητής επιδράσεων ή τροποποιητικός παράγοντας (*Effect Modifier*) του παράγοντα κινδύνου στη νόσο.



Σχήμα 2.3. Επίδραση Συγχυτικού Παράγοντα

Οι συγχυτικοί παράγοντες πρέπει να ελέγχονται και να αδρανοποιούνται αυτό πραγματοποιείται με κατάλληλο σχεδιασμό της μελέτης ή με ειδική στατιστική ανάλυση. Ακόμα κι όταν καταφέρουμε να εξαλείψουμε τους συγχυτικούς παράγοντες είναι σημαντικό να εντοπίσουμε τις σχέσεις αιτιότητας. Προκειμένου να επιτευχθεί αυτό, έχουν οριστεί συγκεκριμένα κριτήρια τα οποία μπορούμε να εφαρμόσουμε για να ελέγξουμε αν οι στατιστικά σημαντικές συσχετίσεις είναι σχέσεις αιτιότητας. Όταν μία σχέση πληροί τα ακόλουθα κριτήρια, τότε έχουμε μία σχέση αιτιότητας (Ντζούφρας & Περπέρογλου, 2009).

Τα κριτήρια αυτά είναι αναφορικά τα ακόλουθα και αναλύονται παρακάτω:

- Συνέπεια (*Consistency*).
- Ισχύς (*Strength*).
- Ειδικότητα (*Specificity*).

- Χρονική αλληλουχία (*Temporality*).
- Λογική αλληλουχία (*Coherency*).

### 2.3.2 Κριτήρια Αξιολόγησης Αιτιολογικών Σχέσεων

**α) Συνέπεια:** Μία σχέση ονομάζεται συνεπής αν εμφανιστεί επαναλαμβανόμενα σε μελέτες που έχουν πραγματοποιηθεί με διαφορετικό σχεδιασμό και σε διαφορετικούς πληθυσμούς. Ενδεικτικό παράδειγμα συνεπούς σχέσης είναι η θετική σχέση καπνίσματος – καρκίνου του πνεύμονα η οποία έχει επιβεβαιωθεί σε μεγάλο αριθμό μελετών τα τελευταία χρόνια.

**β) Ισχύς:** Μία σχέση ονομάζεται ισχυρή όταν η επίδραση της παρουσίας του παράγοντα κινδύνου είναι μεγάλη. Η επίδραση υπολογίζεται με την αναλογική αύξηση της πιθανότητας θανάτου ή εμφάνισης της νόσου. Ενδεικτικό παράδειγμα, η πιθανότητα εμφάνισης καρκίνου του πνεύμονα είναι αυξημένη στους καπνιστές και αυξάνει ανάλογα με τον αριθμό των τσιγάρων που καπνίζει κάθε άτομο.

**γ) Ειδικότητα:** Μία σχέση παράγοντα κινδύνου – νόσου ονομάζεται ειδική αν η παρουσία του παράγοντα κινδύνου οδηγεί με μεγάλη πιθανότητα στην εμφάνιση της νόσου, ενώ η απουσία του οδηγεί με μεγάλη πιθανότητα στην αποφυγή εμφάνισης της νόσου. Αν ο παράγοντας μίας νόσου είναι «ειδικός» είναι πολύ πιθανό και η σχέση να είναι αιτιολογική. Η ύπαρξη μιας στατιστικής σχέσης αιτιότητας δεν συνεπάγεται ότι θα επαληθεύεται σε όλες τις περιπτώσεις. Ενδεικτικό παράδειγμα, αν κάποιος είναι καπνιστής δεν είναι βέβαιο ότι θα εμφανίσει τη νόσο.

**δ) Χρονική αλληλουχία:** Είναι σημαντικό ο παράγοντας κινδύνου να προϋπάρχει χρονικά της εμφάνισης της νόσου. Αυτό δεν είναι πάντα εύκολο, ο καθαρισμός της χρονικής αλληλουχίας είναι δύσκολος σε περιπτώσεις όπου η νόσος εντοπίζεται πολύ αργότερα από την πραγματική στιγμή εμφάνισης ή σε περιπτώσεις όπου το άτομο άμεσα στον κίνδυνο αλλά έμμεσα ( ενδεικτικό παράδειγμα το παθητικό κάπνισμα) .

**ε) Λογική Αλληλουχία:** Μία σχέση έχει λογική αλληλουχία όταν δεν αντικρούει αποδεδειγμένες αλήθειες των φυσικών και βιολογικών επιστημών.

## 2.4 Δείκτες Νοσηρότητας

### 2.4.1 Επιπολασμός

Ως επιπολασμό (*Prevalence*) μιας νόσου ορίζουμε το ποσοστό / αναλογία ατόμων στον πληθυσμό που έχουν τη νόσο σε μια συγκεκριμένη χρονική στιγμή ή κατά τη διάρκεια ενός χρονικού διαστήματος (Τζαβελάς, 2020).

$$\text{Επιπολασμός} = \frac{\text{Αριθμός κρουσμάτων σε ένα πληθυσμό σε συγκεκριμένη χρονική περίοδο}}{\text{Μέγεθος πληθυσμού κατά τη συγκεκριμένη χρονική περίοδο}}$$

Ωστόσο, στο σημείο αυτό πρέπει να αναφέρουμε ότι η κλίμακα του χρόνου χρειάζεται να οριστεί προσεκτικά. Ο χρόνος μπορεί να οριστεί ως :

- 1) Ηλικία ενός ατόμου
- 2) Χρόνος από την έκθεση σε συγκεκριμένο παράγοντα κινδύνου
- 3) Ημερολογιακός χρόνος
- 4) Χρόνος από τη διάγνωση

Σε ορισμένες εφαρμογές επιλέγεται ένα διαφορετικό είδος κλιμακοποίησης του χρόνου. Ενδεικτικά στις μελέτες μολυσματικών ασθενειών μια κλίμακα για το χρόνο είναι ο αριθμός των διακεκριμένων επαφών με ένα μολυσμένο αντικείμενο ή πρόσωπο.

### 2.4.2 Επίπτωση

Αθροιστική επίπτωση (*Cummulative Incidence*) μίας νόσου για το χρονικό διάστημα  $[t_0 - t_1]$  είναι η πιθανότητα ένα άτομο που δεν έχει εκδηλώσει προηγουμένως την ίδια νόσο να εμφανίσει για πρώτη φορά συμπτώματα της νόσου μέσα σε αυτό το διάστημα.

$$\text{Επίπτωση} = \frac{\text{Αριθμός νέων περιστατικών}}{\text{Συνολικός πληθυσμός (που κινδ. να νοσήσει στην αρχή του χρ. διαστήματος)}}$$

Οι εμπειρικοί δείκτες επίπτωσης διακρίνονται σε δείκτες επίπτωσης ποσοστού και σε δείκτες επίπτωσης-πυκνότητας. Κοινό τους χαρακτηριστικό είναι ότι, οι περιπτώσεις νοσήματος, των οποίων μετράτε η συχνότητα, είναι συμβάντα (επεισόδια ή μεταβολές στο χρόνο) (Σύψα, 2016).

### 2.4.3 Ρυθμός Επίπτωσης

Η αθροιστική επίπτωση λαμβάνει υπόψη το χρόνο ως προς τα επεισόδια της νόσου που εμφανίζονται μέσα σε ένα χρονικό διάστημα. Δε λαμβάνει υπόψη το χρόνο διάρκειας μιας νόσου ο οποίος είναι σημαντικός παράγοντας για τη κατανόηση της επίδρασης της νόσου στον πληθυσμό.

Ως Ρυθμό επίπτωσης (*incidence rate*), ορίζουμε το λόγο του αριθμού των νέων επεισοδίων μιας νόσου σε ένα χρονικό διάστημα  $[t_0 - t_1]$  προς «ανθρωποχρόνο κινδύνου» για την ίδια περίοδο. Ο «ανθρωποχρόνος κινδύνου» είναι ο χρόνος παρακολούθησης αθροιστικά για όλα τα άτομα του πληθυσμού. Συνεπώς, ο παρονομαστής είναι χρόνος σε αντίθεση με την αθροιστική επίπτωση όπου παρονομαστής είναι ο αριθμός των ατόμων του πληθυσμού (Σύψα, 2016).

### 2.4.4 Διάρκεια Ασθένειας

Ως Διάρκεια Ασθένειας (*Duration*), ορίζουμε το χρονικό διάστημα από τη πρώτη εμφάνιση της νόσου ως την τελική θεραπεία. Όταν αναφερόμαστε σε ένα πληθυσμό ή ένα δείγμα μας ενδιαφέρει η μέση διάρκεια της νόσου για τα άτομα που έχουν εμφανίσει τη νόσο. Όταν μια ασθένεια είναι θανατηφόρα τότε ως διάρκεια ζωής ονομάζεται ο χρόνος μέχρι την κατάληξη του ασθενή. Η διάρκεια είναι σημαντικό μέτρο της επίδρασης μίας ασθένειας σε ένα πληθυσμό καθώς ασθένειες με μεγάλη μέση διάρκεια οδηγούν σε υψηλό επιπολασμό ακόμα και αν ο ρυθμός επίπτωσης είναι μικρός.

### 2.4.5 Θνησιμότητα

Ως «θνησιμότητα» (*Mortality Rate*) ορίζουμε τον λόγο των θανάτων σε μία περίοδο, διά του μέσου συνολικού πληθυσμού εκείνης της περιόδου (Τζαβελάς, 2020).

$$\text{Δείκτης Θνησιμότητας} = \frac{\text{Θάνατοι κατά τη διάρκεια μιας χρονικής περιόδου}}{\text{Μέγεθος πληθυσμού}}$$

Ενδεικτικά, ο αριθμός των θανάτων που έχουν καταγραφεί από την αρχή της πανδημίας μέχρι σήμερα σύμφωνα με την ιστοσελίδα ([ourworldindata.org/coronavirus](http://ourworldindata.org/coronavirus), n.d.) ανέρχεται στους

33010. Ο πληθυσμός της Ελλάδας είναι 10.432.481 εκατομμύρια. Συνεπώς, ο δείκτης θνησιμότητας για τη περίοδο από την αρχή της πανδημίας μέχρι σήμερα υπολογίζεται ως 0.0031.

## 2.5 Μέτρα Κινδύνου

Προκειμένου να συνδέσουμε τους παράγοντες κινδύνου με την έκβαση μιας νόσου χρησιμοποιούμε τα μέτρα κινδύνου. Συνήθως οι παράγοντες κινδύνου είναι κατηγορικές μεταβλητές (π.χ. το φύλο), όπως επίσης και η κύρια μεταβλητή απόκρισης στις περισσότερες Ιατρικές μελέτες είναι δίτιμη και αφορά τη παρουσία ασθένειας ή όχι. Τα κύρια μέτρα κινδύνου είναι τα ακόλουθα :

### 2.5.1 Σχετική Πιθανότητα (*Odds*)

Ως Σχετική Πιθανότητα ορίζουμε το λόγο της πιθανότητας εμφάνισης ενός ενδεχομένου έναντι της πιθανότητας μη εμφάνισης του.

$$Odds = \frac{P(A)}{1 - P(A)}$$

Ερμηνεία σχετικής πιθανότητας στην περίπτωση που το ενδεχόμενο  $A$  δηλώνει την εμφάνιση κάποιας νόσου σ' ένα άτομο :

- $Odds = 1$  συνεπάγεται ίση πιθανότητα εμφάνισης και μη εμφάνισης της νόσου.
- $Odds > 1$  συνεπάγεται η εμφάνιση της νόσου είναι πιο πιθανή από την πιθανότητα μη εμφάνισης της νόσου.
- $Odds < 1$  συνεπάγεται η εμφάνιση της νόσου είναι λιγότερο πιθανή από την πιθανότητα μη εμφάνισης της νόσου.

### 2.5.2 Λόγος Σχετικών Πιθανοτήτων (*Odds Ratio*)

Ορισμός : Ως «λόγο σχετικών πιθανοτήτων», ορίζουμε το λόγο της σχετικής πιθανότητας μιας ομάδας να έχει εκτεθεί σε έναν παράγοντα κινδύνου έναντι της σχετικής πιθανότητας μη έκθεσης στον παράγοντα κινδύνου.

$$\mathbf{Odds\ Ratio} = \frac{\text{σχετική πιθανότητα εμφάνισης της νόσου για τους εκτεθειμένους}}{\text{σχετική πιθανότητα εμφάνισης της νόσου των μη εκτεθειμένων}}$$

Ερμηνεία Λόγου σχετικών πιθανοτήτων :

- Αν  $OR = \alpha$  τότε η σχετική πιθανότητα της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον κίνδυνο ( $X = 1$ ) είναι ίση με  $\alpha$  φορές την ίδια σχετική πιθανότητα όταν δεν εκτεθεί στον κίνδυνο ( $X = 2$ ).
- Αν  $OR = \alpha > 1$  τότε η σχετική πιθανότητα της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον κίνδυνο ( $X = 1$ ) είναι  $\alpha - 1$  φορές μεγαλύτερη της ίδιας σχετικής πιθανότητας όταν δεν εκτεθεί στον κίνδυνο ( $X = 2$ ).
- Αν  $OR = \alpha < 1$  τότε η σχετική πιθανότητα της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον κίνδυνο ( $X = 1$ ) είναι  $1 - \alpha$  φορές μικρότερη της ίδιας σχετικής πιθανότητας όταν δεν εκτεθεί στον κίνδυνο ( $X = 2$ ).

### 2.5.3 Σχετικός Κίνδυνος (*Relative Risk*)

Ως Σχετικό Κίνδυνο, ορίζουμε το λόγο της επίπτωσης ή της θνησιμότητας δύο ομάδων με διαφορετική έκθεση στον παράγοντα κινδύνου.

$$\mathbf{Relative\ Risk} = \frac{\text{πιθανότητα ενδεχομένου για τους εκτεθειμένους}}{\text{πιθανότητα ενδεχομένου για τους μη εκτεθειμένους}}$$

Ερμηνεία σχετικού κινδύνου :

- Η πιθανότητα εμφάνισης της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου ( $X = 1$ ) είναι ίση με  $a$  φορές την ίδια πιθανότητα όταν δεν εκτεθεί στον παράγοντα κινδύνου ( $X = 2$ ).
- Αν  $a > 1$  τότε : Η πιθανότητα εμφάνισης της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου ( $X = 1$ ) είναι  $a - 1$  φορές μεγαλύτερη από την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο ( $X = 2$ ).
- Αν  $a < 1$  τότε : Η πιθανότητα εμφάνισης της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου ( $X = 1$ ) είναι  $1 - a$  φορές μικρότερη από την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο ( $X = 2$ ).
- Αν  $a = 1$  τότε : Η πιθανότητα εμφάνισης της νόσου ( $Y = 1$ ) όταν ένα άτομο εκτεθεί στον παράγοντα κινδύνου ( $X = 1$ ) είναι ίση με την ίδια πιθανότητα όταν δεν εκτεθεί στον κίνδυνο ( $X = 2$ ) (δηλαδή η μεταβλητή  $X$  δεν επηρεάζει την εμφάνιση της νόσου συνεπώς δεν αποτελεί παράγοντα κινδύνου).

Ο λόγος σχετικών πιθανοτήτων μπορεί να υπολογιστεί και στις προοπτικές (μελέτες κοορτής) και στις αναδρομικές (μαρτύρων - ασθενών) μελέτες, αντίθετα ο σχετικός κίνδυνος μπορεί να υπολογιστεί μόνο σε προοπτικές μελέτες. Τέλος, ο λόγος σχετικών πιθανοτήτων είναι ασυμπτωματικά ίσος με το σχετικό κίνδυνο όταν η πιθανότητα εμφάνισης της νόσου είναι μικρή, όπως στις σπάνιες ασθένειες (Ντζούφρας & Περπέρογλου, 2009).

#### **2.5.4 Οφειλόμενος ή Αποδιδόμενος Κίνδυνος (*Attributable Risk*)**

Ως Αποδιδόμενο κίνδυνο ορίζουμε τη διαφορά μεταξύ των ρυθμών επίπτωσης (ή των δεικτών θνησιμότητας) των ομάδων με άτομα εκτεθειμένα και μη εκτεθειμένα σε ένα παράγοντα κινδύνου. Όταν αναφερόμαστε σε ένα πίνακα συνάφειας  $2 \times 2$  τότε ο αποδιδόμενος κίνδυνος ορίζεται ως η διαφορά των δεσμευμένων πιθανοτήτων. Την πιθανότητα εμφάνισης της νόσου αν κάποιος ασθενής έχει εκτεθεί στον κίνδυνο μείον την πιθανότητα εμφάνισης της νόσου αν ο ασθενής δεν έχει εκτεθεί στον κίνδυνο. Το μειονέκτημα του συγκεκριμένου δείκτη είναι ότι υπολογίζει τις απόλυτες διαφορές. Ενδεικτικά, οι πιθανότητες 0.10 και 0.20 θα μας δώσουν τον



ίδιο αποδιδόμενο κίνδυνο με τις πιθανότητες 0.6 και 0.7 αγνοώντας το γεγονός ότι στην πρώτη περίπτωση η πιθανότητα της μίας ομάδας είναι διπλάσια από την άλλη.

## 2.6 Κλινικές Δοκιμές (*Clinical Trials*)

Ως Κλινικές Δοκιμές ορίζουμε τις συγκριτικές μελέτες στις οποίες οι επιστήμονες παρεμβαίνουν στη φυσική πορεία της νόσου δίνοντας φάρμακα ή θεραπευτικές αγωγές σε τουλάχιστον μία από τις ομάδες μελέτης και μετά εξετάζουν την επίδραση των αγωγών στη νόσο.

Τα βασικότερα στοιχεία μιας Κλινικής Δοκιμής είναι τα ακόλουθα:

1) Θεραπευτική αγωγή (*treatment*) ή τεχνική παρέμβασης (*intervention technique*) : είναι ο τρόπος παρέμβασης στον ασθενή, με σκοπό τη βελτίωση της κατάστασής του.

2) Ομάδα Ελέγχου ή Ψευδοφάρμακου ή Αδρανούς Θεραπείας (*control/placebo groups*): Ψευδοφάρμακο ή αδρανής θεραπεία είναι η αγωγή κατά την οποία δεν γίνεται ουσιαστική θεραπεία. Οι ομάδες αναφοράς οι οποίες λαμβάνουν ψευδοφάρμακο στην πραγματικότητα δεν λαμβάνουν καμία θεραπεία αυτό συμβαίνει κυρίως για ψυχολογικούς λόγους ώστε ο ασθενής να έχει τη εντύπωση ότι του χορηγείται η ίδια θεραπεία με τις υπόλοιπες ομάδες. Η ομάδα που λαμβάνει τη θεραπεία ονομάζεται ομάδα ελέγχου ή ομάδα αναφοράς ( *reference treatment group*).

3) Μόνη ή διπλή τύφλωση / άγνοια (*single or double blindness*) : είναι η κατάσταση κατά την οποία οι ασθενείς δεν γνωρίζουν την αγωγή που λαμβάνουν. Στην περίπτωση της διπλής τύφλωσης, ούτε οι ασθενείς αλλά ούτε και οι γιατροί δεν γνωρίζουν την αγωγή που λαμβάνει ο κάθε ασθενής.

4) Τυχαιοποίηση (*randomization/random allocation*) : είναι η κατάσταση των αγωγών στα άτομα που συμμετέχουν στη κλινική δοκιμή. Σκοπός αυτής της διαδικασίας είναι η εξάλειψη μεροληψίας επιλογής, με άλλα λόγια η κατανομή των ασθενών είναι τυχαία και δεν βασίζεται στις εκτιμήσεις/ προτιμήσεις του γιατρού. Οι ομάδες πρέπει να έχουν παρόμοια χαρακτηριστικά και παρόμοιο μέσο επίπεδο κατάστασης της νόσου (Ντζούφρας & Περπέρογλου, 2009).

## 2.6.1 Φάσεις Κλινικών Δοκιμών

Οι Κλινικές Δοκιμές χωρίζονται σε 4 σημαντικές Φάσεις, οι οποίες αναλύονται παρακάτω:

- Φάση I (*Phase I Trials – Human pharmacology*): Οι μελέτες αυτές γίνονται σε μικρό αριθμό ανθρώπων εφόσον έχουν προηγηθεί εργαστηριακές δοκιμές σε ζώα. Σκοπός της πρώτης φάσης είναι να εκτιμήσουμε το μέγιστο ύψος της δόσης που μπορεί να γίνει αποδεκτό από τον οργανισμό χωρίς σημαντικές παρενέργειες οι οποίες συνήθως μετριοούνται από το επίπεδο τοξικότητας (*toxicity level*). Στην φάση αυτή λοιπόν μας ενδιαφέρει κυρίως η κλινική φαρμακολογία και τοξικότητα.

- Φάση II (*Phase II Trials – Therapeutic Exploratory*): Οι μελέτες αυτές είναι μελέτες ασφάλειας και αποτελεσματικότητας των φαρμάκων/θεραπειών. Βασίζονται πολύ στη μελέτη φάσης I και εξετάζουν την αποτελεσματικότητα του φαρμάκου. Στην φάση αυτή λοιπόν μας ενδιαφέρει κυρίως η διερεύνηση της αποτελεσματικότητας της θεραπείας.

- Φάση III (*Phase III Trials - Therapeutic Confirmatory*): Οι μελέτες αυτές αποτελούν συνέχεια της φάσης II. Στην φάση αυτή λοιπόν μας εστιάζουμε στην πλήρη αξιολόγηση της αποτελεσματικότητας της θεραπείας.

- Φάση IV (*Phase IV Trials - Therapeutic Use*): Οι μελέτες αυτές αφορούν παρακολούθηση (*screening*) του φαρμάκου μετά την κυκλοφορία του. Εδώ σκοπός είναι ο εντοπισμός παρενεργειών ή προβλημάτων που δεν μπορούσαν να εντοπιστούν στη μελέτη φάσης III. Στο στάδιο αυτό η μελέτη είναι μη παρεμβατική και μοιάζει περισσότερο με επιδημιολογική μελέτη. (Ντζούφρας & Περπέρογλου, 2009).

## 2.6.2 Τύποι Κλινικών Δοκιμών

Ως προς τον τύπο των Κλινικών Δοκιμών ορίζουμε τους κάτωθι:

- Κλινική ή θεραπευτική δοκιμή: Συμμετέχουν ασθενείς και οι αγωγές είναι θεραπευτικές. Σκοπός του συγκεκριμένου τύπου δοκιμών είναι η θεραπεία της νόσου. Το μέγεθος δείγματος είναι 20 έως 100 άτομα.

- Κύριες μελέτες πρόληψης (*primary prevention studies*): Συμμετέχουν υγιής άτομα, οι αγωγές που λαμβάνουν οι συμμετέχοντες περιλαμβάνουν μέτρα πρόληψης/προφύλαξης και στόχος είναι η αποφυγή εμφάνισης της νόσου υγείας.

- Δευτερεύουσες μελέτες πρόληψης (*secondary prevention studies*): Συμμετέχουν ασθενείς οι οποίοι συνήθως έχουν εμφανίσει ένα επίπεδο της νόσου, οι αγωγές που λαμβάνουν οι συμμετέχοντες περιλαμβάνουν μέτρα πρόληψης/προφύλαξης και στόχος είναι η αποφυγή 2<sup>ου</sup> επεισοδίου της νόσου ή επιβάρυνση της κατάστασης. Ενδεικτικό παράδειγμα αυτού του τύπου δοκιμών αποτελεί η χημειοθεραπεία σε καρκινοπαθείς.

# ΚΕΦΑΛΑΙΟ 3

## Στοχαστικές Διαδικασίες

### 3.1 Εισαγωγή

Η θεωρία στοχαστικών ανελίξεων ή στοχαστικών διαδικασιών αναπτύχθηκε εξαιτίας της ανάγκης μαθηματικής περιγραφής και μοντελοποίησης συστημάτων ή φαινομένων, τα οποία εξελίσσονται χρονικά με τυχαία και μη προσδιοριστικό τρόπο. Η αριθμητική προσομοίωση φυσικών μοντέλων, που υποστηρίζεται από τη διαθεσιμότητα των αυξανόμενων υπολογιστικών πόρων, παίζει σήμερα πολύ σημαντικό ρόλο στις διαδικασίες σχεδιασμού και στη λήψη αποφάσεων που μπορεί να έχουν σημαντικές συνέπειες τόσο από οικονομικής όσο και από ανθρώπινης άποψης. Η ποσοτικοποίηση και η διάδοση της αβεβαιότητας σε φυσικά συστήματα εμφανίζονται ως κρίσιμο μονοπάτι για τη βελτίωση της πρόβλεψης της απόκρισής τους. Η απόκτηση αξιόπιστων προβλέψεων απαιτεί αυστηρό έλεγχο του εννοιολογικού μοντέλου. Αυτό το μοντέλο πρέπει να σχεδιαστεί προκειμένου να αναπαράγει με ακρίβεια τις παρατηρήσεις ή πιο συγκεκριμένα, ορισμένες παρατηρήσιμες ποσότητες ενδιαφέροντος. Τα φυσικά μοντέλα, τα οποία γενικά είναι αρκετά καλά κατακτημένα μέσα στο ντετερμινιστικό πλαίσιο, πρέπει στη συνέχεια να συμπληρωθούν από μια σχετική μοντελοποίηση αβεβαιοτήτων. Στη συνέχεια, πρέπει να προταθεί ένα αριθμητικό μοντέλο καθώς και τεχνικές επίλυσης (*ad hoc*), προκειμένου να διαδοθούν οι αβεβαιότητες και να εκτιμηθεί ο αντίκτυπός τους στην απόκριση του εννοιολογικού μοντέλου. (Δημητράκος, 2022)

Οι ακολουθίες ανεξάρτητων και πανομοιότυπα κατανεμημένων τυχαίων μεταβλητών είναι στοχαστικές διαδικασίες, αλλά δεν είναι πάντα ενδιαφέρουσες ως στοχαστικά μοντέλα επειδή συμπεριφέρονται λίγο πολύ με τον ίδιο τρόπο. Προκειμένου να εισαχθεί περισσότερη μεταβλητότητα, μπορεί κανείς να επιτρέψει κάποια εξάρτηση από το παρελθόν με τον τρόπο ντετερμινιστικών εξισώσεων επανάληψης (*Brémaud*, 2020).

Γενικά διακρίνονται δύο τύποι αβεβαιοτήτων: οι εγγενείς αβεβαιότητες, που σχετίζονται με μια φυσική μεταβλητότητα των εξεταζόμενων φυσικών φαινομένων, και τις γνωσιολογικές αβεβαιότητες, που προκύπτουν από την έλλειψη γνώσης. Φυσικά, οι τελευταίες είναι από τη

φύση τους αναγώγιμες. Ωστόσο, λόγω της πολυπλοκότητας των φυσικών φαινομένων, της έλλειψης παρατηρήσεων ή ακόμη και της αναγκαιότητας παροχής απλουστευμένων μαθηματικών μοντέλων, η αντιμετώπιση επιστημικών αβεβαιοτήτων φαίνεται επίσης απαραίτητη για τη βελτίωση της προβλεψιμότητας του μοντέλου (*Grinstead & Snell, 2012*).

Στο παρόν κεφάλαιο μετά τη σύντομη εισαγωγή, θα παρουσιαστούν κάποιες εισαγωγικές έννοιες απαραίτητες για την κατανόηση των δεδομένων, ενώ στη συνέχεια θα αναφερθούν οι βασικοί ορισμοί και θα παρουσιαστούν οι δύο βασικές θεματικές ενότητες, οι στοχαστικές διαδικασίες διακριτού και συνεχούς χρόνου.

## 3.2 Εισαγωγικές Έννοιες και ορισμοί

Συμβολίζουμε με  $X(t)$  τη κατάσταση του συστήματος κατά τη χρονική στιγμή  $t$  όπου  $t \geq 0$ . Θεωρούμε ότι για κάθε  $t$  η κατάσταση  $X(t)$  είναι μια τυχαία μεταβλητή η οποία ορίζεται σ' ένα χώρο πιθανότητας  $(\Omega, F, P)$  κοινό για όλα τα  $t$ . Η  $X(t)$  είναι για συγκεκριμένο  $t$  μια απεικόνιση  $X(\omega, t)$  με πεδίο ορισμού το δειγματικό χώρο  $\Omega$  ενός πειράματος τύχης με τιμές στο  $R$ , ή στο  $R^k$  γενικότερα. Για συγκεκριμένο  $\omega \in \Omega$  έχουμε τη συνάρτηση  $X(\omega, t) = x(t), t \geq 0$  η οποία αποτελεί μια διαδρομή ή τροχιά απ' όλες τις δυνατές διαδρομές ή τροχιές που μπορούν να προκύψουν από το χώρο πιθανότητας  $(\Omega, F, P)$  και την οικογένεια των τυχαίων μεταβλητών  $\{X(t), t \geq 0\}$ . Γενικεύοντας την ερμηνεία του  $t$  ως στοιχείου ενός παραμετρικού συνόλου  $T$  και χρησιμοποιώντας τον όρο τυχαία μεταβλητή είτε πρόκειται για μονοδιάστατες, είτε πρόκειται για πολυδιάστατες τυχαίες μεταβλητές. Ως στοχαστική ανέλιξη (*Stochastic Process*) ορίζουμε κάθε οικογένεια τυχαίων μεταβλητών  $\{X(t), t \in T\}$  σ' ένα κοινό χώρο πιθανότητας  $(\Omega, F, P)$  (*Μπούτσικα, 2020*).

### 3.2.1 Κατηγοριοποίηση Στοχαστικών Ανελίξεων

Οι στοχαστικές ανελίξεις χωρίζονται γενικά σε διαδικασίες διακριτού και συνεχούς χρόνου με διακριτό και συνεχές πεδίο τιμών, ανάλογα με τη φύση του χρόνου και του χώρου των τιμών τους. Συγκεκριμένα, ως χώρο καταστάσεων  $S$  ορίζουμε το σύνολο των δυνατών τιμών των τυχαίων μεταβλητών  $X(t)$  όπου  $(t \in T)$  και το σύνολο  $T$  καλείται παραμετρικός χώρος.

Εάν το σύνολο  $T$  είναι αριθμήσιμο τότε έχουμε στοχαστική ανέλιξη διακριτού χρόνου, ενώ αν το σύνολο  $T$  είναι μη αριθμήσιμο τότε έχουμε στοχαστική ανέλιξη συνεχούς χρόνου.

Επιπλέον με βάση το χώρο καταστάσεων  $S$ , οι στοχαστικές ανελίξεις έχουν διακριτό χώρο καταστάσεων αν το σύνολο αυτό είναι πεπερασμένο ή αριθμήσιμο, ενώ αν το σύνολο αυτό είναι υπεραριθμήσιμο, η στοχαστική ανέλιξη έχει συνεχή χώρο καταστάσεων. Σύμφωνα με τα παραπάνω προκύπτουν τέσσερις περιπτώσεις κατηγοριοποίησης των στοχαστικών ανελίξεων. (Μπούτσικας, 2020)

- i. Στοχαστική ανέλιξη διακριτού χρόνου, διακριτού πεδίου τιμών
- ii. Στοχαστική ανέλιξη συνεχούς χρόνου, διακριτού πεδίου τιμών
- iii. Στοχαστική ανέλιξη διακριτού χρόνου, συνεχούς πεδίου τιμών
- iv. Στοχαστική ανέλιξη συνεχούς χρόνου, συνεχούς πεδίου τιμών.

### 3.2.2 Εξάρτηση τυχαίων μεταβλητών μιας στοχαστικής ανελίξης

Μας ενδιαφέρει να μελετήσουμε την εξάρτηση μεταξύ των τυχαίων μεταβλητών  $X(t), t \in T$  της στοχαστικής ανελίξης. Η εξάρτηση αυτή καθορίζεται από την από κοινού συνάρτηση κατανομής (σ.κ.) οποιουδήποτε πεπερασμένου υποσυνόλου  $k$  τυχαίων μεταβλητών της στοχαστικής ανελίξης.

$$F_{t_1, t_2, \dots, t_k}(x_1, \dots, x_k) := P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_k} \leq x_k) \quad x_1, x_2, \dots, x_k \in S,$$

για κάθε  $k \geq 1$ .

Δύο στοχαστικές ανελίξεις  $\{X(t), t \in T\}, \{X(s), s \in T\}$  καλούνται στοχαστικά ανεξάρτητες αν κάθε ζεύγος πεπερασμένων υποοικογενειών τους,  $\underline{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_k})$  και  $\underline{Y} = (Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$  είναι στοχαστικά ανεξάρτητες.

Αναλόγως με τη στοχαστική εξάρτηση των μελών τους, οι στοχαστικές ανελίξεις μπορούν να κατηγοριοποιηθούν σε Μαρκοβιανές ανελίξεις, Στοιχηματικές ανελίξεις (*Martingales*), Στάσιμες ανελίξεις, Απαριθμήτριες στοχαστικές ανελίξεις, Κλαδωτές ανελίξεις, Ανανεωτικές ανελίξεις, Τυχαίους περιπάτους, ανελίξεις Διάχυσης, ανελίξεις *Levy* κ.α. Στη συνέχεια θα παρουσιαστούν οι βασικότερες κατηγορίες εξ' αυτών.

#### 3.2.2.1 Στάσιμες Στοχαστικές Ανελίξεις

Μία στοχαστική ανέλιξη  $X = \{X(t), t \in T\}$  καλείται αυστηρώς (ή ισχυρώς) στάσιμη (*strictly stationary*) αν τα τυχαία διανύσματα  $(X_{t_1}, X_{t_2}, \dots, X_{t_k}), (X_{t_1+s}, X_{t_2+s}, \dots, X_{t_k+s})$ , έχουν την ίδια από κοινού συνάρτηση κατανομής, για κάθε  $t_1, \dots, t_k \in T$  και  $t_1 + s, \dots, t_k + s \in T$ .

Ωστόσο, στην πράξη, είναι αρκετά δύσκολο να εξασφαλισθεί η αυστηρή στασιμότητα, κι έτσι στην πλειοψηφία των περιπτώσεων αρκούμαστε σε μια ασθενέστερη έννοια στασιμότητας.

Μία στοχαστική ανέλιξη  $X = \{X(t), t \in T\}$  καλείται ασθενώς στάσιμη (*weakly stationary*) αν:

$$E(X_t) = \mu, \forall t \in T$$

$$Cov(X_{s+t}, X_t) = \gamma(s), \forall s + t \in T$$

Από τις παραπάνω σχέσεις παρατηρούμε ότι η μέση τιμή μιας ασθενώς στάσιμης στοχαστικής ανέλιξης παραμένει σταθερή στον χρόνο, ενώ η συνδιακύμανση της δεν εξαρτάται από την χρονική στιγμή  $t$ , αλλά μόνο από τη χρονική απόσταση των δύο τυχαίων μεταβλητών.

### 3.2.2.2 Μαρκοβιανή Αλυσίδα ή Μαρκοβιανή Ανέλιξη

Μια από τις σημαντικότερες κατηγορίες της οικογένειας των στοχαστικών ανελιξεων είναι οι στοχαστικές διαδικασίες, οι οποίες που έχουν την λεγόμενη αμνήμονα ή Μαρκοβιανή ιδιότητα. Συγκεκριμένα, δεδομένου του παρόντος, δηλαδή δεδομένου του  $X_y$ , θεωρώντας ότι το παρόν είναι ο χρόνος  $y$ , οι μελλοντικές της τιμές ( $X_{y+t}, t > 0$ ) είναι ανεξάρτητες από τις παρελθοντικές ( $X_s, s < y, s \in T$ ). Επομένως οι μελλοντικές της καταστάσεις εξαρτώνται μόνο από το παρόν και όχι από το παρελθόν. Επισημαίνεται πως δεν είναι απαραίτητο να γνωρίζουμε όλες τις καταστάσεις στο παρελθόν για να εφαρμοστεί η Μαρκοβιανή ιδιότητα, καθώς επίσης πως ως παρελθόν δεν θεωρείται απαραίτητα η προηγούμενη χρονική στιγμή αλλά μπορεί να ληφθεί η πλησιέστερη στο παρόν χρονική στιγμή.

Μία στοχαστική ανέλιξη  $\{X(t), t \in T\}$  ορισμένη σε ένα χώρο πιθανότητας  $(\Omega, \mathcal{F}, P)$  με τιμές σε ένα χώρο  $S$  (με σύνολο ενδεχομένων  $\mathcal{S}$ ) έχει την Μαρκοβιανή ιδιότητα αν, για κάθε ενδεχόμενο  $A \in \mathcal{S}$  ισχύει ότι:

$$P(X_{y+t} \in A \mid X_y = x_y, X_s = x_s, s < y) = P(X_{y+t} \in A \mid X_y = x_y, X_s),$$

για κάθε  $y \in T$  και κάθε  $t \in T$  με  $t \geq 0$ .

Μία στοχαστική ανέλιξη  $\{X(t), t \in T\}$  με τη παραπάνω ιδιότητα καλείται αλυσίδα *Markov* (ή Μαρκοβιανή αλυσίδα). Η ιδιότητα αυτή ικανοποιείται από πολλές στοχαστικές ανελιξεις (π.χ. τυχαίος περίπατος, ανέλιξη *Poisson*, κίνηση *Brown*) που χρησιμοποιούνται για την μοντελοποίηση στοχαστικών φαινομένων.

### 3.3 Στοχαστικές διαδικασίες διακριτού χρόνου

Αν και η βασική τους θεωρία δεν είναι υπερβολικά περίπλοκη, είναι εξαιρετικά αποτελεσματικό να μοντελοποιούνται ως κατηγορικές ακολουθίες δεδομένων. Οι ομοιογενείς μαρκοβιανές αλυσίδες (αλυσίδες *Markov*) διακριτού χρόνου διαθέτουν το απαιτούμενο χαρακτηριστικό αφού μπορούν πάντα να αναπαρασταθούν, τουλάχιστον όσον αφορά την κατανομή, ως επαναληπτική στοχαστική εξίσωση:

$$X_{n+1} = f(X_n, Z_n + 1),$$

όπου το  $\{Z_n\}, n \geq 1$  είναι *iid* ακολουθία, ανεξάρτητη από την αρχική κατάσταση  $X_0$ .

Η πιθανολογική εξάρτηση από το παρελθόν είναι μόνο μέσω της προηγούμενης κατάστασης, αλλά αυτή η περιορισμένη ποσότητα μνήμης αρκεί για να δημιουργήσει μια μεγάλη ποικιλία συμπεριφορών. Για αυτόν τον λόγο, οι αλυσίδες *Markov* έχουν βρει εφαρμογές σε πολλούς τομείς, συμπεριλαμβανομένης της βιολογίας όπως η εξαφάνιση ενός είδους ή το φαινόμενο δημιουργίας μίας ασθένειας, της φυσικής, της κοινωνιολογίας, της επιχειρησιακής έρευνας (έλεγχος αποθεμάτων, θεωρία ουρών αναμονής κ.α.) και της μηχανικής, όπου παρέχουν ποιοτικές και ποσοτικές απαντήσεις καθώς και πολύτιμες γνώσεις για το σχεδιασμό συστημάτων (*Ching et al.*, 2008. *Brémaud*, 2020, Χρυσοφίνου, 2012).

#### 3.3.1 Μαρκοβιανές Ανεπίξεις διακριτού χρόνου και χώρου καταστάσεων

Όπως αναφέρθηκε παραπάνω, μια Μαρκοβιανή ανέλιξη, ανάλογα με τον χρόνο μελέτης, χωρίζεται σε δυο βασικές κατηγορίες. Συγκεκριμένα:

(α) Αν  $X$  είναι Μαρκοβιανή ανέλιξη διακριτού χρόνου  $T$  και χώρου  $S$ , τότε η πιθανότητα μετάβασης (ή μεταπήδησης) 1ης τάξης της Μαρκοβιανή ανέλιξης από την κατάσταση  $i$  στο χρόνο  $n$  στην κατάσταση  $j$  στο χρόνο  $n + 1$  συμβολίζεται με:

$$p_{ij}(n, n + 1) := P(X_{n+1} = j | X_n = i), i, j \in S, n \in T.$$

(β) Αν η πιθανότητα μετάβασης  $p_{ij}(n, n + 1)$  δεν εξαρτάται από το χρόνο  $n$ , δηλαδή είναι σταθερή ως προς αυτόν, για κάθε  $i, j \in S$ , τότε η Μαρκοβιανή ανέλιξη καλείται (χρονικά) ομογενής και η παραπάνω πιθανότητα μετάβασης 1ης τάξης συμβολίζεται απλούστερα με:

$$p_{ij} := P(X_{n+1} = j | X_n = i), i, j \in S, n \in T.$$



Στην παρούσα ενότητα εξετάζονται οι ανελίξεις σε διακριτό χρόνο.

Σύμφωνα με τον Δημητράκο (2022) στον παραπάνω ορισμό, το σύνολο  $T$  των χρονικών στιγμών της στοχαστικής ανελίξης έχει ληφθεί ίσο με  $\{0, 1, \dots\}$ . Θα αναφερθούν οι Μαρκοβιανές αλυσίδες σε διακριτό χρόνο με χρονικά ομογενείς πιθανότητες μετάβασης. Αυτό σημαίνει ότι:

$$P\{X_{n+1} = j | X_n = i\} = p_{ij}, i, j \in I, \text{ για οποιαδήποτε τιμή της χρονικής στιγμής } n.$$

Οι πιθανότητες μετάβασης  $p_{ij}$  καλούνται πιθανότητες μετάβασης πρώτης τάξης και ικανοποιούν τις εξής σχέσεις:

$$p_{ij} \geq 0, i, j \in I \text{ και } \sum_{j \in I} p_{ij} = 1, i \in I (*)$$

Η Μαρκοβιανή αλυσίδα  $\{X_n, n = 0, 1, \dots\}$  προσδιορίζεται πλήρως από τη συνάρτηση πιθανότητας της αρχικής κατάστασης  $X_0$  και των πιθανοτήτων μετάβασης πρώτης τάξης  $p_{ij}$ . Η ισότητα του ορισμού δύναται να ερμηνευτεί ως εξής: Η μελλοντική πιθανοτική συμπεριφορά της αλυσίδας εξαρτάται μόνο από την παρούσα κατάστασή της και όχι από το παρελθόν της (Δημητράκος, 2022). Η σχέση του ορισμού είναι γνωστή ως Μαρκοβιανή ιδιότητα (*Markovian property*). Οι πιθανότητες μετάβασης πρώτης τάξης μπορούν να χρησιμοποιηθούν για να υπολογιστεί η πιθανότητα να μεταβεί η ανελίξη από την κατάσταση  $i$  στην κατάσταση  $j$  μετά από  $n$  βήματα. Οι πιθανότητες μετάβασης  $n$  τάξης ορίζονται ως εξής:

$$p_{ij}^{(n)} = \{X_n = j | X_0 = i\}, n = 1, 2, \dots \text{ και } i, j \in I.$$

Σημειώνεται ότι ισχύει  $p_{ij}^{(1)} = p_{ij}$  και επιπλέον ορίζουμε:

$$p_{ij}^{(0)} = 1 \text{ αν } j = i \text{ και } p_{ij}^{(0)} = 0, \text{ αν } j \neq i.$$

Επιπλέον, έχουμε τα ακόλουθα στοιχεία (Privault, 2018):

- Οι πιθανότητες μετάβασης πρώτης τάξης μπορούν να χρησιμοποιηθούν για τον πλήρη υπολογισμό της κατανομής πιθανοτήτων της διαδικασίας ως ακολούθως:

$$\begin{aligned} & P(Z_n = i_n, Z_{n-1} = i_{n-1}, \dots, Z_0 = i_0) \\ &= P(Z_n = i_n | Z_{n-1} = i_{n-1}, \dots, Z_0 = i_0)P(Z_{n-1} = i_{n-1}, \dots, Z_0 = i_0) = \end{aligned}$$

$$\begin{aligned}
& P(Z_n = i_n | Z_{n-1} = i_{n-1})P(Z_{n-1} = i_{n-1} | Z_{n-2} = i_{n-2}, \dots, Z_0 = i_0) \times P(Z_{n-2} = i_{n-2}, \dots, Z_0 = i_0) \\
& = P(Z_n = i_n | Z_{n-1} = i_{n-1})P(Z_{n-1} = i_{n-1} | Z_{n-2} = i_{n-2}) \\
& \quad \times P(Z_{n-2} = i_{n-2} | Z_{n-3} = i_{n-3}) P(Z_{n-3} = i_{n-3}, \dots, Z_0 = i_0)
\end{aligned}$$

που βάσει επαγωγικού συλλογισμού υποδεικνύει είτε ότι:

$$\begin{aligned}
& P(Z_n = i_n, Z_{n-1} = i_{n-1}, \dots, Z_0 = i_0) \\
& = P(Z_n = i_n | Z_{n-1} = i_{n-1}) \dots P(Z_1 = i_1 | Z_0 = i_0)P(Z_0 = i_0)
\end{aligned}$$

είτε ότι:

$$\begin{aligned}
& P(Z_n = i_n, Z_{n-1} = i_{n-1}, \dots, Z_1 = i_1 | Z_0 = i_0) \\
& P(Z_n = i_n | Z_{n-1} = i_{n-1}) \dots P(Z_1 = i_1 | Z_0 = i_0), \mu\epsilon i_0, i_1, \dots, i_n \in S
\end{aligned}$$

- Επιπλέον, με τον νόμο της ολικής πιθανότητας που εφαρμόζεται στα γεγονότα  $A_k = \{Z_2 = i_2 \text{ και } Z_1 = k\}, k \in S$  κάτω από το μέτρο πιθανότητας  $P(\cdot | Z_0 = i_0)$  προκύπτει ότι:

$$\begin{aligned}
P(Z_2 = i_2 | Z_0 = i_0) &= \sum_{i_1 \in S} P(Z_2 = i_2 \text{ and } Z_1 = i_1 | Z_0 = i_0) \\
&= \sum_{i_1 \in S} P(Z_2 = i_2 | Z_1 = i_1) P(Z_1 = i_1 | Z_0 = i_0) \\
&\quad \mu\epsilon i_0, i_2 \in S \quad \text{και}
\end{aligned}$$

$$\begin{aligned}
P(Z_1 = i_1) &= \sum_{i_0 \in S} P(Z_1 = i_1, Z_0 = i_0) \\
&= \sum_{i_0 \in S} P(Z_1 = i_1 | Z_0 = i_0) P(Z_0 = i_0), i_1 \in S
\end{aligned}$$

Εδώ είναι ενδιαφέρον να παρατεθεί ο νόμος της ολικής πιθανότητας (*Privault, 2018*) ο οποίος εφαρμόζεται παραπάνω :

$$\mathbb{P}(B) \sum_{n=1}^{\infty} \mathbb{P}(B \cap A_n) = \sum_{n=1}^{\infty} \mathbb{P}(B | A_n) \mathbb{P}(A_n)$$

όπου  $\{A_n: n = 1, 2, \dots\}$  είναι μία διαμέριση του δειγματικού χώρου. Οι πιθανότητες μετάβασης πρώτης τάξης συγκεντρώνονται σε ένα πίνακα που καλείται πίνακας μετάβασης πρώτης τάξης της αλυσίδας (*One - Step Transition Probability Matrix*). Ο πίνακας αυτός συνήθως συμβολίζεται με  $P$  και καλείται στοχαστικός πίνακας (*Stochastic Matrix*), εφόσον το άθροισμα των τιμών σε κάθε γραμμή ισούται με τη μονάδα. Στην ειδική περίπτωση κατά την οποία  $I = \{0, 1, \dots, N\}$ , δηλαδή όταν το σύνολο  $I$  είναι πεπερασμένο, σύμφωνα με τις εξισώσεις *Chapman-Kolmogorov*, οι οποίες παρέχουν μια μέθοδο υπολογισμού των πιθανοτήτων μετάβασης  $n$ -τάξης  $p_{ij}^{(n)}$  μίας Μαρκοβιανής αλυσίδας, ο πίνακας που προκύπτει είναι ο ακόλουθος. Έχουμε ότι:  $P_{(n)} = P^n, n = 0, 1, \dots$ , όπου,  $P_{(n)}$  όπως φαίνεται παρακάτω ισούται με:

$$\begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & \dots & p_{0N}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} & \dots & p_{1N}^{(n)} \\ \dots & \dots & \dots & \dots \\ p_{N0}^{(n)} & p_{N1}^{(n)} & \dots & p_{NN}^{(n)} \end{bmatrix}, \quad n = 0, 1, \dots,$$

Επίσης  $P_{(1)} = P$ . Ο πίνακας  $P$  ονομάζεται πίνακας μεταβάσεων της Μαρκοβιανής αλυσίδας. Ένα σημαντικό ζήτημα αποτελεί η εύρεση της ασυμπτωτικής συμπεριφοράς των πιθανοτήτων μετάβασης  $n$ -τάξεως καθώς το  $n$  τείνει στο άπειρο.

Για τις εξισώσεις *Chapman-Kolmogorov* ισχύει ότι:

$$p_{ij}^{(n+1)} = \sum_{k \in I} p_{ik}^{(n)} p_{kj}, \text{ για κάθε } n = 1, 2, \dots \text{ και κάθε } i, j \in I,$$

διότι διαδοχικά ισχύει ότι:

$$\begin{aligned} p_{ij}^{(n+1)} &= P\{X_{n+1} = j | X_0 = i\} = \\ &= \sum_{k \in I} P\{X_{n+1} = j | X_n = k, X_0 = i\} P\{X_n = k | X_0 = i\} \\ &= \sum_{k \in I} P\{X_{n+1} = j | X_n = k\} P\{X_n = k | X_0 = i\} \\ &= \sum_{k \in I} p_{ik}^{(n)} p_{kj}, \quad i, j \in I, n = 0, 1, \dots \end{aligned}$$

Η δεύτερη ισότητα προκύπτει από την εφαρμογή του Θεωρήματος Ολικής Πιθανότητας, αν θεωρηθεί δεδομένη η κατάσταση στην οποία βρίσκεται η αλυσίδα τη χρονική στιγμή  $n$  (πιο

συγκεκριμένα, θεωρείται δεδομένο το ενδεχόμενο  $\{X_n = k\}$ . Η τρίτη ισότητα είναι αποτέλεσμα της Μαρκοβιανής ιδιότητας (Δημητράκος, 2022).

### 3.3.2 Ταξινόμηση των καταστάσεων μίας Μαρκοβιανής αλυσίδας σε διακριτό χρόνο

Σύμφωνα με τον Δημητράκο (2022), για κάθε  $i, j \in I$  και κάθε  $n = 1, 2, \dots$  ορίζουμε τη πιθανότητα  $f_{ij}^{(n)}$  ως εξής:

$$f_{ij}^{(n)} = P\{X_n = j, X_k \neq j, k = 1, \dots, n-1 | X_0 = i\}$$

θεωρώντας ότι  $f_{ij}^{(0)} = 0$ . Η πιθανότητα  $f_{ij}^{(n)}$  είναι η πιθανότητα να μεταβεί η αλυσίδα για πρώτη φορά στην κατάσταση  $j$ , κατά τη χρονική στιγμή  $t = n$ , αν η αρχική κατάσταση είναι η κατάσταση  $i$ .

Ορίζουμε επίσης τις πιθανότητες  $f_{ij}$  ως εξής:

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}, \text{ για κάθε } i, j \in I.$$

Η πιθανότητα  $f_{ij}$  είναι η πιθανότητα να μεταβεί η ανάλιξη κάποτε στην κατάσταση  $j$ , αν η αρχική κατάσταση είναι η κατάσταση  $i$ . Δίνουμε τους ακόλουθους ορισμούς.

1. Μία κατάσταση  $i$  καλείται **έμμονη** ή **επαναληπτική** (*Recurrent State*) αν ισχύει ότι:  $f_{ii} = 1$ . Μία μη-έμμονη κατάσταση καλείται **μεταβατική** (*Transient State*).
2. Ο μέσος χρόνος επανόδου σε μία έμμονη κατάσταση είναι:
$$\mu_{ii} = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$$
3. Μία έμμονη κατάσταση  $i$  καλείται **θετική έμμονη** (*Positive Recurrent State*), αν ισχύει ότι:  $\mu_{ii} < \infty$ . Αν  $\mu_{ii} = \infty$ , η έμμονη κατάσταση  $i$  καλείται μηδενική έμμονη (*Null Recurrent State*).
4. Μια κατάσταση  $i$  λέγεται **περιοδική** με περίοδο  $d$  αν  $p_{ii}^{(n)} = 0$  οποτεδήποτε  $n$  δεν διαιρείται με το  $d$  και ο  $d$  είναι ο μέγιστος ακέραιος μ' αυτή την ιδιότητα, δηλαδή  $p_{ii}^{(n)} = 0$  για κάθε  $n$  και  $p_{ii}^{(nd)} > 0$  και ο θετικός ακέραιος  $d$  είναι ο μέγιστος με την ιδιότητα  $p_{ii}^{(nd)} > 0$ . Μια κατάσταση  $i$  λέγεται **απεριοδική** (*Aperiodic*) αν έχει περίοδο  $d = 1$ .

5. Μία *έμμομη* κατάσταση  $i$  καλείται *περιοδική* (*Periodic State*) με περίοδο  $d$  αν η επάνοδος στην κατάσταση  $i$  μπορεί να γίνει μόνο σε  $d, 2d, 3d, \dots$  βήματα και  $d > 1$  είναι ο μεγαλύτερος ακέραιος με αυτήν την ιδιότητα.
6. Μία *απεριοδική, θετική έμμομη* κατάσταση καλείται *εργοδική* (*Ergodic State*).
7. Ένα *μη-κενό σύνολο καταστάσεων*  $C$  λέγεται *κλειστό* (*Closed Set of States*) αν ισχύει ότι:  $p_{ij} = 0$  για κάθε  $i \in C$  και  $j \notin C$ .
8. Ένα *μη-κενό σύνολο καταστάσεων*  $C$  καλείται *ανάγωγο* αν είναι κλειστό και δεν περιέχει κανένα γνήσιο υποσύνολο του που να είναι κλειστό.
9. Η κατάσταση  $j$  είναι *προσιτή* (*Accessible*) από την κατάσταση  $i$  αν ισχύει ότι:  $p_{ij}^{(n)} > 0$ , για κάποιο  $n \geq 0$ .
10. Οι καταστάσεις  $i$  και  $j$  επικοινωνούν (μεταξύ τους) αν ισχύει ότι:  $i \rightarrow j$  και  $j \rightarrow i$ .

### 3.4 Στοχαστικές διαδικασίες συνεχούς χρόνου

Σε αυτή την παράγραφο ξεκινάμε τη μελέτη στοχαστικών διεργασιών συνεχούς χρόνου, οι οποίες είναι οικογένειες  $\{X_t : t \in R_+\}$  τυχαίων μεταβλητών. Ο στόχος είναι από τις μαρκοβιανές αλυσίδες διακριτού χρόνου να περάσουμε στις αντίστοιχες συνεχούς χρόνου, με την κύρια διαφορά μεταξύ των δύο να είναι η αντικατάσταση του πίνακα μετάβασης. Στις εν λόγω αλυσίδες, η μετάβαση από μια κατάσταση της διαδικασίας σε μια άλλη λαμβάνει χώρα σε ακέραιες χρονικές στιγμές  $n = 0, 1, \dots$ . Εν τούτοις, συνήθως σε πραγματικές πρακτικές εφαρμογές, η εναλλαγή των καταστάσεων μιας Μαρκοβιανής αλυσίδας μπορεί να πραγματοποιηθεί οποιαδήποτε χρονική στιγμή  $t \geq 0$ . Εφόσον κατέστη απαραίτητο να αναλυθούν οι αναφερόμενες εφαρμογές, αναδύθηκε η έννοια των Μαρκοβιανών αλυσίδων σε συνεχή χρόνο  $\{X(t), t \geq 0$  (*Continuous - Time Markov Chains*) (Privault, 2018, Δημητράκος, 2022).

Δύο θεμελιώδη παραδείγματα είναι αυτά των διαδικασιών *Poisson* και των διαδικασιών γέννησης και θανάτου, ακολουθούμενα από την κατασκευή συνεχούς χρόνου *Markov* αλυσίδων και των γεννητριών τους γενικότερα. Από την άποψη των προσομοιώσεων, η χρήση αλυσίδων *Markov* συνεχούς χρόνου δεν προκαλεί ιδιαίτερη δυσκολία καθώς οποιαδήποτε προσομοίωση συνεχούς χρόνου βασίζεται στην πραγματικότητα σε δείγματα διακριτού χρόνου. Από θεωρητικής σκοπιάς, ωστόσο, η αυστηρή αντιμετώπιση της ιδιότητας *Markov* συνεχούς χρόνου είναι πολύ πιο απαιτητική από την αντίστοιχη διακριτού χρόνου, κυρίως λόγω της χρήσης της ισχυρής ιδιότητας *Markov* (Privault, 2018, Δημητράκος, 2022).

Ο ορισμός λοιπόν των παραπάνω δίδεται από τον Δημητράκο (2022) ως εξής:

Η στοχαστική ανέλιξη (διαδικασία) σε συνεχή χρόνο  $\{X(t), t \geq 0\}$  με διακριτό χώρο καταστάσεων το πεπερασμένο ή απείρως αριθμήσιμο σύνολο  $I$  καλείται Μαρκοβιανή αλυσίδα σε συνεχή χρόνο αν:

$$P\{X(t_n) = i_n \mid X(t_0) = i_0, \dots, X(t_{n-1}) = i_{n-1}\} = P\{X(t_n) = i_n \mid X(t_{n-1}) = i_{n-1}\},$$

για οποιαδήποτε  $0 \leq t_0 < \dots < t_{n-1} < t_n$  και  $i_0, i_1, \dots, i_{n-1}, i_n \in I$ .

### 3.4.1 Στοχαστική ανέλιξη *Poisson*

Θεωρούμε μια ανέλιξη σε συνεχή χρόνο  $\{X(t), t \geq 0\}$ , όπου η τυχαία μεταβλητή  $X(t)$  αναπαριστά τον αριθμό των αφίξεων (ή των γεγονότων) που συμβαίνουν κατά το χρονικό διάστημα  $(0, t)$ . Τα γεγονότα λαμβάνουν χώρα κατά τυχαίο τρόπο. Ο χώρος καταστάσεων είναι το σύνολο  $I = \{0, 1, 2, \dots\}$ . Θα πρέπει να ισχύουν τα παρακάτω δεδομένα (Δημητράκος 2022):

(i) Οι αριθμοί των γεγονότων που συμβαίνουν κατά τη διάρκεια ξένων χρονικών διαστημάτων είναι ανεξάρτητα κατανεμημένοι. Δηλαδή η ανέλιξη έχει «ανεξάρτητες προσαυξήσεις».

(ii) Η κατανομή της τυχαίας μεταβλητής  $X(t+h) - X(t)$ , δηλαδή ο αριθμός των αφίξεων που συμβαίνουν κατά το χρονικό διάστημα  $(t, t+h)$  εξαρτάται μόνο από το  $h$  και όχι από το  $t$ . Δηλαδή υπάρχει «χρονική στασιμότητα»

(iii) Ισχύει ότι  $P\{X(t+h) - X(t) = 1\} = \lambda h + o(h)$ , καθώς  $h \rightarrow 0$ , για κάποιο θετικό αριθμό  $\lambda$ . Με  $o(h)$  αναπαριστάται μία συνάρτηση του  $h$  τέτοια ώστε  $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$ .

(iv)  $P\{X(t+h) - X(t) \geq 2\} = o(h)$ , καθώς  $h \rightarrow 0$ .

Η ανέλιξη  $\{X(t), t \geq 0\}$ , η οποία πληροί τις προϋποθέσεις που παρουσιάστηκαν καλείται ανέλιξη *Poisson* με ρυθμό  $\lambda > 0$  (*Poisson process with rate  $\lambda$* ).

Από τα αξιώματα (iii) και (iv), προκύπτει ότι:

$$P\{X(t+h) - X(t) \geq 0\} = 1 - \lambda h + o(h), \text{ καθώς } h \rightarrow 0.$$

Έστω  $p_n(t) = P\{X(t) = n\}$ ,  $n \geq 0$ .

Το ζητούμενο είναι να προσδιοριστεί η κατανομή της τυχαίας μεταβλητής  $X(t), t \geq 0$ . Οι προδρομικές εξισώσεις (*Forward Equations*) λαμβάνονται ως εξής:

Καθώς  $h \rightarrow 0$  για  $n = 0$ , ισχύει ότι:  $p_0(t+h) = p_0(t)(1 - \lambda h + o(h))$  ή ισοδύναμα:

$$\frac{p_0(t+h) - p_0(t)}{h} = \lambda p_0(t) + \frac{o(h)}{h} \quad \text{ή} \quad \frac{dp_0(t)}{dt} = -\lambda p_0(t) \quad (3.1)$$

Καθώς  $h \rightarrow 0$  για  $n \geq 1$ , ισχύει ότι:  $p_n(t+h) = p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h))$  ή ισοδύναμα:

$$\frac{dp_n(t)}{dt} = -\lambda p_n(t) + \lambda p_{n-1}(t) \quad (3.2)$$

Ορίζουμε την **πιθανο-γεννήτρια** της τυχαίας μεταβλητής  $X(t)$ , ως εξής:

$$\varphi(z, t) = \sum_{n=0}^{\infty} p_n(t) z^n, \quad |z| < 1$$

Πολλαπλασιάζοντας τις εξισώσεις (2) με  $z^n$  και προσθέτοντας στην Εξίσωση (1), έχουμε:

$$\sum_{n=0}^{\infty} \frac{dp_n(t)}{dt} z^n = -\lambda \sum_{n=0}^{\infty} p_n(t) z^n + \lambda \sum_{n=1}^{\infty} p_{n-1}(t) z^n, \quad \text{δηλαδή} \quad \frac{d\varphi(z, t)}{dt} = -\lambda(1-z)\varphi(z, t).$$

Η λύση της παραπάνω διαφορικής εξίσωσης υπολογίζεται ως εξής:

$$\int \frac{d\varphi}{\varphi} = -\lambda(1-z) \int dt$$

Ισοδύναμα, ισχύει ότι:  $\ln \varphi = -\lambda(1-z)t + c$ , όπου  $c$  σταθερά, δηλαδή

$$\varphi(z, t) = f(z) \exp\{-\lambda(1-z)t\}$$

για κάποια συνάρτηση  $f(z)$ , διότι η τιμή της σταθεράς μπορεί να εξαρτάται από την τιμή της μεταβλητής  $z$ .

Η αρχική συνθήκη είναι  $\varphi(z, 0) = 1$  από την οποία έπεται ότι:

$$\varphi(z, t) = \exp\{-\lambda t(1-z)\}.$$

Η ροπογεννήτρια της τυχαίας μεταβλητής  $X(t)$  ορίζεται ως:

$$M(z, t) = \sum_{n=0}^{\infty} e^{zn} P\{X(t) = n\} = \varphi(e^z, t).$$

Άρα,  $M(z, t) = \exp\{\lambda t(e^z - 1)\}$ .

Παρατηρούμε ότι ο τύπος της ροπογεννήτριας  $M(z, t)$  συμπίπτει με τον τύπο της ροπογεννήτριας της κατανομής *Poisson* με παράμετρο  $\lambda t$ , διότι αν  $X(t) \sim \text{Poiss}(\lambda t)$ ,  $\lambda > 0$  οπότε μπορεί να γραφεί ότι:

$$\begin{aligned} M_{x(t)}(z, t) &= E(e^{tX(t)}) = \sum_{x=0}^{\infty} e^{tx} P(X(t) = x) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda t} \frac{(\lambda t)^x}{x!} = e^{-\lambda t} \sum_{x=0}^{\infty} \frac{(\lambda t e^t)^x}{x!} \\ &= e^{-\lambda t} e^{\lambda t e^t} = \exp\{\lambda t(e^t - 1)\}. \end{aligned}$$

Άρα, η τυχαία μεταβλητή  $X(t)$  ακολουθεί την κατανομή *Poisson* με παράμετρο  $\lambda t$ , δηλαδή  $X(t) \sim \text{Poiss}(\lambda t)$ .

Έστω  $T_1$  ο χρόνος μέχρι την πρώτη άφιξη. Ισχύει ότι:

$$P\{T_1 > t\} = P\{X(t) = 0\} = e^{-\lambda t}.$$

Άρα, η συνάρτηση πυκνότητας της τυχαίας μεταβλητής  $T_1$  είναι

$$f_{T_1}(t) = \lambda e^{-\lambda t}, t \geq 0. \text{ Δηλαδή, } T_1 \sim \text{Expo}(\lambda).$$

Για  $n \geq 2$ , έστω  $T_n$  ο χρόνος που μεσολαβεί μεταξύ της  $n - 1$  - οστής και της  $n$  - οστής αφίξεως. Ισχύει ότι:

$$\begin{aligned} P\{T_{n+1} > t \mid T_1 = t_1, T_2 = t_2, \dots, T_n = t_n\} &= P\{X(s + t) - X(s) = 0\} \\ &= P\{X(t) - X(0) = 0\} = P\{X(t) = 0\} = e^{-\lambda t} \end{aligned}$$

όπου,  $s = t_1 + t_2 + \dots + t_n$ . Η πρώτη ισότητα είναι συνέπεια του Αξιώματος (i) ενώ η δεύτερη ισότητα είναι συνέπεια του Αξιώματος (ii).

Συνεπώς, ο χρόνος  $T_{n+1}$  είναι ανεξάρτητος των  $T_1, \dots, T_n$  και ακολουθεί την *Expo*( $\lambda$ ). Συμπεραίνουμε ότι η ακολουθία  $T_n, n \geq 1$ , είναι μία ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών που ακολουθούν την Εκθετική κατανομή με παράμετρο ίση με  $\lambda > 0$  (Δημητράκος, 2022).



### 3.4.2 Απλή ανέλιξη Γεννήσεως-Θανάτου

Έστω ότι σε ένα χρονικό διάστημα μήκους  $h$ , όπου  $h \rightarrow 0$ , ένα συγκεκριμένο άτομο ενός πληθυσμού γεννάει ένα νέο άτομο

- Γεννάει ένα νέο άτομο με πιθανότητα  $\lambda h + o(h)$  ή
- πεθαίνει με πιθανότητα  $\mu h + o(h)$ ,

όπου  $\lambda$  και  $\mu$  θετικοί αριθμοί. Τα διαφορετικά άτομα συμπεριφέρονται ανεξάρτητα το ένα από το άλλο και ανεξάρτητα από την προηγούμενη ιστορία της ανέλιξης.

Αν τη χρονική στιγμή  $t$  υπάρχουν  $n$  άτομα στον πληθυσμό τότε η πιθανότητα να συμβεί ακριβώς μία γέννηση στο χρονικό διάστημα  $(t, t + h)$ , όπου  $h \rightarrow 0$ , είναι ίση με

$$\binom{n}{1} (\lambda h + o(h))(1 - \lambda h + o(h))^{n-1} = n\lambda h + o(h).$$

Ομοίως, η πιθανότητα να συμβεί ακριβώς ένας θάνατος είναι ίση με  $n\mu h + o(h)$ .

Η πιθανότητα να συμβούν περισσότερα από 1 γεγονότα στο χρονικό διάστημα  $(t, t + h)$ , όπου  $h \rightarrow 0$ , είναι ίση με  $o(h)$ .

Η αντίστοιχη Μαρκοβιανή αλυσίδα σε συνεχή χρόνο  $X(t), t \geq 0$ , όπου η τυχαία μεταβλητή  $X(t)$  αναπαριστά το μέγεθος του πληθυσμού τη χρονική στιγμή  $t$ , καλείται απλή ανέλιξη γεννήσεως-θανάτου (*Simple Birth - Death Process*) με ρυθμούς γέννησης και θανάτου ανά άτομο  $\lambda$  και  $\mu$ , αντίστοιχα.

Ο χώρος των καταστάσεων της διαδικασίας είναι το σύνολο των μη-αρνητικών ακεραίων αριθμών. Για  $n \geq 1$  και  $h \rightarrow 0$ , έχουμε:

$$P\{X(t + h) - X(t) = 1 | X(t) = n\} = \lambda n h + o(h)$$

$$P\{X(t + h) - X(t) = -1 | X(t) = n\} = \mu n h + o(h)$$

$$P\{X(t + h) - X(t) = 0 | X(t) = n\} = 1 - n\lambda h - \mu n h + o(h)$$

$$P\{X(t + h) - X(t) = j | X(t) = n\} = o(h), \text{ για } j \notin \{0, 1, -1\}$$

Η κατάσταση 0 είναι μία κατάσταση απορρόφησης (*absorbing state*) και αντιστοιχεί στην εξάλειψη του πληθυσμού. Στον πίνακα 1 παρουσιάζονται οι μεταβάσεις της ανέλιξης και των αντίστοιχων ρυθμών.

ΜΕΤΑΒΑΣΗ	ΡΥΘΜΟΣ
$x \rightarrow x + 1$	$\lambda x, x \geq 1$
$x \rightarrow x - 1$	$\mu x, x \geq 1$

Πίνακας 3.4.1.1. Μετάβαση της ανέλιξης και των αντίστοιχων ρυθμών.

Αυτό που μελετάται είναι η κατανομή της τυχαίας μεταβλητής  $X(t)$ .

Οι πιθανότητες είναι οι εξής:

$$p_n(t) = P\{X(t) = n | X(0) = \xi\}.$$

Οι προδρομικές εξισώσεις, για  $n \geq 1$  και  $h \rightarrow 0$ , λαμβάνονται ως εξής:

$$p_n(t+h) = p_n(t)(1 - \lambda nh - \mu nh + o(h)) + p_{n-1}(t)(\lambda(n-1)h + o(h)) + p_{n+1}(t)(\mu(n+1)h + o(h)) + o(h).$$

Για  $n = 0$  και  $h \rightarrow 0$ ,

$$p_0(t+h) = p_0(t) + p_1(t)(\mu h + o(h)) + o(h).$$

Προκύπτουν οι εξής διαφορικές εξισώσεις:

$$\frac{dp_n(t)}{dt} = -(\lambda + \mu)np_n(t) + \lambda(n-1)p_{(n-1)}(t) + \mu(n+1)p_{n+1}(t), n \geq 1 \quad (\mathbf{A})$$

$$\frac{dp_0(t)}{dt} = \mu p_1(t) \quad (\mathbf{B})$$

Για τη λύση των Εξισώσεων (A) και (B) χρησιμοποιείται η τεχνική της πιθανο - γεννήτριας. Θεωρούμε την πιθανο - γεννήτρια της τυχαίας μεταβλητής  $X(t)$ , δηλαδή τη συνάρτηση:

$$\varphi(z, t) = \sum_{n=0}^{\infty} p_n(t) z^n, |z| < 1.$$

Πολλαπλασιάζοντας τις Εξισώσεις (A) επί  $z^n$  και προσθέτοντας στην Εξίσωση (B), λαμβάνεται:

$$\begin{aligned} & \sum_{n=0}^{\infty} \left[ \frac{dp_n(t)}{dt} \right] z^n = \\ & = -(\lambda + \mu)z \sum_{n=1}^{\infty} np_n(t)z^{(n-1)} + \mu \sum_{n=0}^{\infty} (n+1)p_{n+1}(t)z^n + \lambda z^2 \sum_{n=1}^{\infty} (n-1)p_{n-1}(t)z^{n-2} \end{aligned}$$

Λύνοντας το παρακάτω σύστημα εξισώσεων προκύπτει ότι : (Δημητράκος, 2022).

$$\frac{\partial t}{1} = -\frac{\partial z}{(\lambda z - \mu)(z - 1)} = \frac{\partial \varphi}{0}.$$

Πρέπει  $\partial \varphi = 0$  δηλαδή  $\varphi = const$ . Αυτή είναι μία από τις δύο ανεξάρτητες λύσεις.

Για να βρεθεί η δεύτερη λύση πρέπει να διακριθούν δύο περιπτώσεις:

1. Περίπτωση:  $\lambda \neq \mu$ , όπου με κατάλληλες πράξεις προκύπτει τελικά,

$$\varphi(z, t) = \frac{[(\mu(z-1)e^{(\lambda-\mu)t} + \mu - \lambda z)]}{[\lambda(z-1)e^{(\lambda-\mu)t} + \mu - \lambda z]^{\xi}}$$

2. Περίπτωση:  $\lambda = \mu$ , τότε  $(z, t) \rightarrow 1$ , καθώς  $t \rightarrow \infty$ . Αν  $\lambda < \mu$  τότε  $(z, t) \rightarrow 1$ , καθώς  $t \rightarrow \infty$ .

Αν  $\lambda > \mu$  τότε  $\varphi(z, t) \rightarrow \left(\frac{\mu}{\lambda}\right)^{\xi}$ , καθώς  $t \rightarrow \infty$ . Δηλαδή  $p_0(t) = \varphi(0, t) \rightarrow 1$  καθώς  $t \rightarrow \infty$ ,  $\lambda \leq \mu$  και  $p_0(t) = \varphi(0, t) \rightarrow \left(\frac{\mu}{\lambda}\right)^{\xi}$ , καθώς  $t \rightarrow \infty$ , αν  $\lambda > \mu$ . Αυτό σημαίνει ότι, αν  $\lambda \leq \mu$ , ο πληθυσμός εξαλείφεται τελικά με πιθανότητα 1 ενώ αν  $\lambda > \mu$ , ο πληθυσμός εξαλείφεται τελικά με πιθανότητα  $\left(\frac{\mu}{\lambda}\right)^{\xi}$ .

# ΚΕΦΑΛΑΙΟ 4

## Κλαδωτές Ανελίξεις

### 4.1 Θεωρητικό πλαίσιο διαδικασιών *Bienaymé-Galton-Watson*

Η σύγχρονη μαθηματική θεωρία των πιθανοτήτων έχει τις ρίζες της στις προσπάθειες ανάλυσης των τυχερών παιχνιδιών από τον *Gerolamo Cardano* τον 16<sup>ο</sup> καθώς επίσης και από τους *Pierre de Fermat*, *Blaise Pascal* («πρόβλημα των πόντων») και *Christiaan Huygens* τον 17<sup>ο</sup> αιώνα. Τον 19<sup>ο</sup> αιώνα, αυτό που σήμερα καλείται ο κλασικός ορισμός της πιθανότητας ολοκληρώθηκε από τον *Pierre Laplace* (Χρυσαφίνου, 2012).

Κατά τα τέλη του 19<sup>ου</sup> αιώνα η θεωρία πιθανοτήτων άρχισε να ξεφεύγει από τα στενά πλαίσια της πρόβλεψης των αποτελεσμάτων τυχερών παιχνιδιών και να βρίσκει εφαρμογές σε νέους κλάδους επηρεάζοντας καταλυτικά την εξέλιξή τους. Από τους πλέον σημαντικούς, είναι αυτός της αναπαραγωγής πληθυσμών (Χρυσαφίνου, 2012).

Το πρόβλημα που αποτέλεσε τη βάση για την ανάπτυξη του συγκεκριμένου κλάδου ήταν η καταγραφή της εξαφάνισης των επωνύμων των ευρωπαϊκών αριστοκρατικών οικογενειών, όπως αυτή αποτυπώθηκε από τον ο *Francis Galton* το 1867.

Ο *Galton* προσπαθώντας να λύσει το πρόβλημα της εξαφάνισης των οικογενειακών ονομάτων το έθεσε ως ένα ανοικτό πρόβλημα στους αναγνώστες του *'Educational Times'*:

“*Problem 4001*: Ένα μεγάλο έθνος, από το οποίο εμείς ασχολούμαστε μόνο με τους ενήλικους άνδρες οι οποίοι είναι  $N$  σε αριθμό και ο καθένας από αυτούς φέρει ξεχωριστό επώνυμο, αποικίζουν μια περιοχή. Ο νόμος του πληθυσμού τους είναι τέτοιος ώστε σε κάθε γενιά το  $a_0\%$  των ενήλικων ανδρών δεν έχουν αρσενικά παιδιά που να έχουν ενήλικη ζωή, το  $a_1\%$  έχει ένα τέτοιο αρσενικό παιδί, το  $a_2\%$  έχει δύο και συνεχίζοντας ως  $a_5\%$  που έχει πέντε.

Βρείτε (1) ποιο ποσοστό των επωνύμων θα εξαφανιστεί μετά από  $r$  γενιές και (2) πόσες περιπτώσεις επωνύμων θα υπάρχουν που να αποτελούνται από  $m$  άτομα.”

Ο *Galton* δεν έλαβε ικανοποιητική απάντηση από τους αναγνώστες του περιοδικού και επειδή δεν κατάφερε να βρει την απάντηση μόνος του ζήτησε την βοήθεια του μαθηματικού *Henry William Watson*.

Η διατύπωση που πρότεινε ο *Watson* βασίζεται σε πιθανογεννήτριες συναρτήσεις (probability generating functions) και ισχύει μέχρι σήμερα. Αρχικά κατήγγησε τον περιορισμό  $k \leq 5$ , θέτοντας παράλληλα  $p_k$  αντί του προτεινόμενου από τον *Galton*  $\alpha_k$ .

Έτσι αν η αρχική συνάρτηση είναι:

$$f(s) = \sum_{k=0}^{\infty} p_k \cdot s^k \quad (0 \leq s \leq 1) \quad (4.1)$$

$$\text{και } f_1(s) = f(s), \quad f_{n+1}(s) = f(f_n(s)) = f_n(f(s)) \quad (n = 1, 2 \dots) \quad (4.2)$$

τότε οι δυναμοσειρές για το  $f_n$  έχουν ως συντελεστές τους όρους της κατανομής πιθανοτήτων για τον συνολικό αριθμό των αρσενικών στη γενιά  $n^{th}$ , ενώ το πρώτο αρσενικό αποτελεί τη μηδενική γενιά (*Kendall*, 1966).

Επιπρόσθετα ο *Watson* παρατήρησε πως η πιθανότητα εξάλειψης  $q_n$ , ικανοποιεί την εξίσωση:

$$q_1 = p_0 \text{ και } q_{n+1} = f(q_n)$$

$$\text{και αν } q_n \rightarrow q, \text{ όταν } n \rightarrow \infty, \text{ τότε } q = f(q) \quad (4.3)$$

και έχει πάντοτε ρίζα (root)  $q = 1$  (*Kendall*, 1966).

Με βάση αυτό ο *Watson* πίστευε ότι η εξαφάνιση της ανδρικής γενεολογικής γραμμής ήταν αναπόφευκτη. Ωστόσο, λίγα χρόνια αργότερα, ο *Agner Krarup Erlang* εξετάζοντας το πρόβλημα του *Watson*, αναθεώρησε τη διατύπωση της εξίσωσης (4.3) και ερμήνευσε την πιθανότητα εξαφάνισης  $q$  ώστε να ισούται με: την πιθανότητα να μην υπάρχει άρρεν απόγονος στην πρώτη γενιά, συν (ii) την πιθανότητα να υπάρχει ένας άρρεν απόγονος στην πρώτη γενιά, ακολουθούμενη από εξαφάνιση, συν (iii) την πιθανότητα δύο αρσενικών στην πρώτη γενιά, ακολουθούμενη από την τελική εξαφάνιση και ούτω καθεξής (*Kendall*, 1966).

$$q = p_0 + p_1q + p_2q^2 + \dots \quad (4.4)$$

Σύμφωνα με τις προαναφερθείσες πληροφορίες, η γενεαλογία και γενικά ο τομέας της κοινωνιολογίας, έχουν χρησιμοποιήσει εκτενώς τις κλαδωτές ανελίξεις, επομένως οι επιστήμες αυτές θεωρούνται μια κλασική περιοχή εφαρμογής των διαδικασιών αυτών (*Gonzalez and del Puerto* (2010), (*Lange*, 2010)).

## 4.2. Στοιχειώδης θεωρία και έννοιες κλαδωτών ανελίξεων

Μια κλαδωτή ανέλιξη, μπορεί, όπως έχει ήδη αναφερθεί, να περιγράψει ένα ευρύ φάσμα βιολογικών προβλημάτων που εκτείνονται από την επικράτηση ή την εξάλειψη ενός ονόματος, ενός χαρακτηριστικού, ενός συγκεκριμένου γονιδίου κλπ. με εξαιρετικά ενδιαφέρουσες κατά περίπτωση προεκτάσεις (Χρυσάφινου, 2012).

Οι εφαρμογές των κλαδικών ανελίξεων σε άλλους κλάδους των φυσικών επιστημών είναι πολυάριθμες και η μελέτη τους είναι πέρα από τους σκοπούς του συγκεκριμένου κεφαλαίου. Μπορούν ωστόσο να αναφερθούν χαρακτηριστικά εφαρμογές στον κλάδο της Φυσικής, που σχετίζονται με πολύπλοκα προβλήματα πυρηνικής σχάσης, αλυσιδωτών αντιδράσεων νετρονίων κλπ (Χρυσάφινου, 2012).

Όπως έχει ήδη αναφερθεί, μια κλαδωτή ανέλιξη είναι μια στοχαστική διαδικασία που αποτελείται από συλλογές τυχαίων μεταβλητών με κύριο σκοπό να χρησιμεύσει ως μαθηματικό μοντέλο ενός πληθυσμού στον οποίο κάθε άτομο στη γενιά  $n$  παράγει κάποιο τυχαίο αριθμό απογόνων στη γενιά  $n + 1$ . Τα άτομα αυτά σε επόμενη γενιά παράγουν άλλους απογόνους κοκ. Όλα τα άτομα αναπαράγονται ανεξάρτητα αλλά με την ίδια κατανομή απογόνων (Μητροφάνη, 2021).

Μια κλαδωτή ανέλιξη έχει μια κατανομή «έκρηξης ή κατάρρευσης»: είτε ο πληθυσμός θα εκραγεί (θα μεγαλώνει απεριόριστα) είτε θα αποτύχει εντελώς, δηλαδή θα εξαληφθεί. Επιπλέον, ένα σημαντικό τους χαρακτηριστικό είναι ότι από τη στιγμή που το μέγεθος του πληθυσμού αρχίσει να μειώνεται, η εξαφάνιση της κλαδωτής ανελίξεως θα συμβεί σύντομα στα πρώτα στάδια. Ένα από τα κρίσιμα ερωτήματα στη θεωρία των κλαδωτών ανελίξεων είναι η πιθανότητα εξαφάνισης, η μη ύπαρξη δηλαδή ατόμων του πληθυσμού μετά από κάποιο πεπερασμένο αριθμό γενεών, οι συνθήκες κάτω από τις οποίες ο πληθυσμός εξαφανίστηκε ή η πιθανότητα επιβίωσης του πληθυσμού αυτού (*Athreya and Ney, 1972*).

Η ταξινόμηση των κλαδωτών ανελίξεων θα μπορούσε να γίνει σε διάφορες βάσεις, με την πλειονότητα των επιστημόνων να χρησιμοποιεί τις ακόλουθες: i) ανάλογα με την παράμετρο χρόνου και ii) την κρισιμότητά τους (*Kimmel and Axelrod, 2015*). Με βάση την κρισιμότητά τους, υπάρχουν τρεις περιπτώσεις (υποκρίσιμη, κρίσιμη και υπερκρίσιμη).

Πρώτα ωστόσο πρέπει να πραγματοποιηθεί μια εισαγωγή στις κλαδωτές ανελίξεις κατά *Galton Watson / Galton Watson Process (GWP)*, ώστε να υπάρχουν διαθέσιμες περισσότερες

πληροφορίες σχετικά με αυτήν την ταξινόμηση, επομένως, η ανάλυση των συνθηκών κρισιμότητας ακολουθεί αμέσως μετά (*Haccou et al.*, 2005; *Kimmel and Axelrod*, 2015).

Με βάση τη διάρκεια ζωής (παράμετρος χρόνου) έχουμε βαθμιδωτή ταξινόμηση. Εάν οι διάρκειες ζωής όλων των ατόμων είναι πανομοιότυπες και ίσες με 1, προκύπτει ότι η διαδικασία μπορεί να περιγραφεί χρησιμοποιώντας έναν δείκτη διακριτού χρόνου, πανομοιότυπο με τον αριθμό των διαδοχικών γενεών (*Athreya and Ney*, 1972; *Kemeny and Snell*, 1972; *Kimmel and Axelrod*, 2015; *Pazsit et al.*, 2007).

Αυτή είναι η κλασική προσέγγιση κατά *Galton-Watson* [η απλούστερη μορφή Κλαδωτών Ανελιξιών Διακριτού Χρόνου / *Discrete Time Branching Processes* (ΚΑΔΧ/DTBP)]. Από την άλλη πλευρά, εάν οι διάρκειες ζωής των ατόμων του πληθυσμού δεν είναι σταθερές (μία μονάδα χρόνου όπως στην κλασική προσέγγιση), υπάρχουν δύο γενικές κατηγορίες (*Athreya and Ney*, 1972; *Haccou et al.*, 2005):

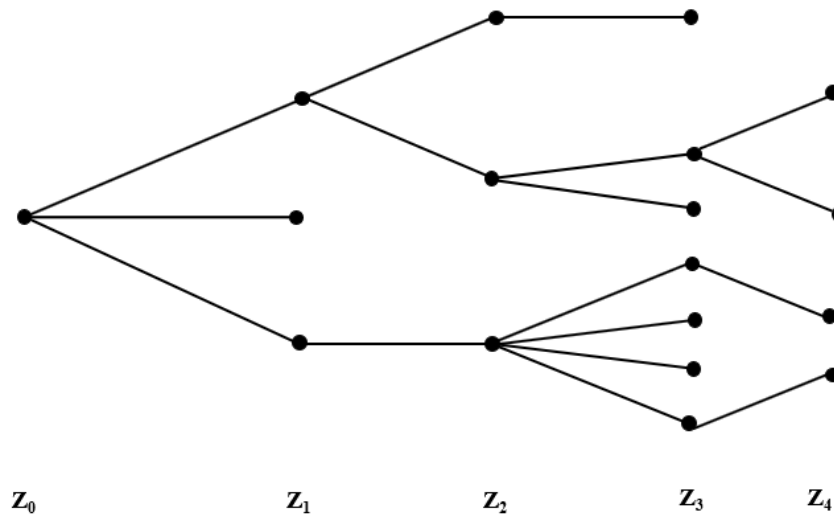
- Διαδικασία *Markov*: Εάν η κατανομή της διάρκειας ζωής των ατόμων του πληθυσμού είναι εκθετική (καλείται διαδικασία *Markov* συνεχούς χρόνου)
- Ηλικιοεξαρτώμενη Διαδικασία (ή *Bellman-Harris*): Εάν η διάρκεια ζωής των ατόμων του πληθυσμού είναι μια αυθαίρετη μη αρνητική τυχαία μεταβλητή.

Στην πλειονότητα των στοχαστικών διεργασιών, υπάρχουν κλαδωτές ανελιξεις διακριτού αλλά και συνεχούς χρόνου. Στα μοντέλα κλαδωτών ανελιξεων διακριτού χρόνου, ο χρόνος αντιπροσωπεύεται από ακέραιους αριθμούς που υποδεικνύουν περιόδους αναπαραγωγής (*Haccou et al.*, 2005). Έτσι, στις απλούστερες μορφές, δεν υπάρχουν επικαλύψεις μεταξύ των γενεών και η αναπαραγωγή συμβαίνει μόνο σε ξεχωριστές περιόδους (*Haccou et al.*, 2005).

Τα μοντέλα κλαδωτών ανελιξεων συνεχούς χρόνου (ΚΑΣΧ) είναι πιο περίπλοκα από τα μοντέλα διακριτού χρόνου, ωστόσο, καθώς μια κλαδωτή ανέλιξη διακριτού χρόνου έχει μικρά χρονικά βήματα, μπορεί να θεωρηθεί ως μια προσέγγιση ενός μοντέλου συνεχούς ανέλιξης. Στην πραγματικότητα, τα μοντέλα κλαδωτών ανελιξεων διακριτού χρόνου χρησιμοποιούνται συνήθως για την εξαγωγή χρήσιμων πληροφοριών σχετικά με το αντίστοιχο μοντέλο συνεχούς χρόνου (*Haccou et al.*, 2005).

Οι παλαιότερες και απλούστερες κλαδωτές ανελιξεις διακριτού χρόνου είναι οι απλές διαδικασίες κλαδωτών ανελιξεων κατά *Galton-Watson*, οι οποίες ονομάζονται επίσης διεργασίες *Bienaymé-Galton-Watson* ή απλά κλαδωτές ανελιξεις. Όπως αναφέρθηκε ήδη, έχουμε μόνο έναν τύπο ατόμων του πληθυσμού και δεν υπάρχουν επικαλύψεις μεταξύ των

γενεών, όπως επιβεβαιώνεται από το Σχήμα 4.1 που είναι μια απλή απεικόνιση της διαδικασίας (Μητροφάνη, 2021).



Σχήμα 4.1. Δέντρο Κλαδωτής ανέλιξης Διακριτού Χρόνου

Στο μοντέλο αυτό όπως φαίνεται και από το παραπάνω σχήμα, δεν υπάρχουν επικαλύψεις μεταξύ των γενεών και ο πληθυσμός συγκροτείται από έναν μόνο τύπο (ή μεμονωμένο τύπο) ατόμων (Haccou et al., 2005).

Το μέγεθος του πληθυσμού στη  $n$ -οστή γενιά συμβολίζεται με  $Z_n$ . Κάθε άτομο ζει ακριβώς μια μονάδα χρόνου, παράγει έναν τυχαίο αριθμό απογόνων  $\xi$  και πεθαίνει. Όλα τα άτομα αναπαράγονται ανεξάρτητα. Ο πληθυσμός ξεκινά με ένα άτομο ( $Z_0 = 1$ ) στη  $n = 0$  γενιά, με το μέγεθος του πληθυσμού της επόμενης γενιάς να μπορεί να προσδιοριστεί αθροίζοντας τους αριθμούς των απογόνων (Schinazi, 1999).

Ο αριθμός των απογόνων του  $i^{ου}$  ατόμου του πληθυσμού συμβολίζεται ως  $\xi_i$  με  $i = 1, 2, 3 \dots$  να είναι ανεξάρτητες και πανομοιότυπα κατανομημένες τυχαίες μεταβλητές και η κατανομή μεγέθους οικογένειας είναι  $P(\xi = k) = p_k$ , να είναι η πιθανότητα παραγωγής  $k$  αριθμού απογόνων (Haccou et al., 2005). Επιπλέον, το μέγεθος του πληθυσμού της  $(n+1)$  γενιάς ( $Z_{n+1}$ ) προκύπτει ως το άθροισμα των απογόνων που προκύπτουν από τα άτομα του πληθυσμού της  $n$ -οστής γενιάς:

$$Z_{n+1} = \xi_1 + \xi_2 + \dots + \xi_{Z_n}$$

Άρα, ο αριθμός των ατόμων τη στιγμή  $n$ ,  $Z_n$ , είναι ίσος με τον συνολικό αριθμό των απογόνων των ατόμων  $1, 2, 3, \dots, Z_{n-1}$ . Τότε η κλαδωτή ανέλιξη είναι  $\{Z_0, Z_1, Z_2, \dots\} = \{Z_n: n \in \mathbb{N}\}$



$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i \quad (4.5)$$

Η κατάσταση (*State*) της κλαδωτής ανέλιξης τη στιγμή  $n$  είναι  $Z_n$ , παίρνει τιμές  $0, 1, 2, \dots$  αλλά  $Z_0 = 1$  πάντα. Μόλις το  $Z_n$  πάρει την τιμή 0, αυτό σημαίνει ότι κανένα άτομο δεν είναι ζωντανό, τότε η ανέλιξη είναι νεκρή και αυτή η κατάσταση θεωρείται ως απορροφητική (*Absorbing*). Ωστόσο, η προαναφερθείσα διατύπωση και οι παραδοχές συνθέτουν τον απλούστερο τύπο μιας κλαδωτής ανέλιξης.

Υπάρχουν πιο πολύπλοκα μοντέλα κλαδωτών ανελιξεων, όπως οι ανελιξεις πολλαπλών τύπων, όπου ο πληθυσμός αποτελείται από περισσότερα από ένα είδη. Μία από τις σημαντικότερες διαφορές στην περίπτωση αυτή είναι ότι τα άτομα του πληθυσμού του ίδιου είδους θεωρείται ότι έχουν τον ίδιο κατανομημένο αριθμό απογόνων, κάτι που δεν ισχύει για τα άτομα διαφορετικών ειδών (*Haccou et al., 2005*). Άλλες πιο περίπλοκες περιπτώσεις μοντέλων κλαδωτών ανελιξεων είναι εκείνες στις οποίες κάθε άτομο δεν ζει απαραίτητα μόνο μία μονάδα χρόνου, δεν έχουν δηλαδή όλα τα άτομα την ίδια διάρκεια ζωής ενώ παράλληλα μπορούν να δώσουν απογόνους σε διαφορετικές ηλικίες (*Haccou et al., 2005*).

Δύο είναι τα βασικά ζητήματα/βασικές παράμετροι που οφείλουν να αναλυθούν σχετικά με τους πληθυσμούς και τις διαδικασίες αναπαραγωγής: το αναμενόμενο πληθυσμιακό μέγεθος και η πιθανότητα εξαφάνισης (*Athreya and Ney, 1972; Haccou et al., 2005; Schinazi, 1999*). Στη θεωρία πιθανοτήτων και τη στατιστική, ο αριθμητικός μέσος του μεγέθους του πληθυσμού ή η αναμενόμενη τιμή είναι ένα μέτρο της κεντρικής τάσης είτε μιας κατανομής πιθανοτήτων (*probability distribution*) είτε της τυχαίας μεταβλητής που χαρακτηρίζεται από αυτή την κατανομή. Έτσι, φαίνεται αρκετά λογικό να υπολογίσουμε τον αριθμητικό μέσο του  $Z_n(E(Z_n))$ , γιατί επιθυμούμε να γνωρίζουμε ποιος αναμένεται να είναι ο πληθυσμός σε  $n$  γενεές (*Schinazi, 1999*).

Από την άλλη πλευρά, η διακύμανση είναι το τετράγωνο της τυπικής απόκλισης μιας τυχαίας μεταβλητής από τον αριθμητικό της μέσο. Διαισθητικά, μετρά πόσο μακριά είναι διεσπαρμένο ένα σύνολο (τυχαίων) αριθμών από τη μέση τιμή. Όπως αναφέραμε παραπάνω για το  $Z_n$ , ο πληθυσμός είτε θα απογειωθεί είτε θα αποτύχει εντελώς. Στην πραγματικότητα, εάν το μέγεθος του πληθυσμού φθίνει, η εξαφάνιση είναι πιθανό να συμβεί πολύ γρήγορα, μέσα στις πρώτες γενιές. Αυτό εξηγεί γιατί μας ενδιαφέρει το  $Var(Z_n)$ . Η υψηλή διακύμανση υποδηλώνει ότι ο πληθυσμός δεν θα εξελιχθεί γρήγορα γύρω από τις μέσες τιμές και αντίστοιχα, η χαμηλή

διακύμανση μας ειδοποιεί ότι ο πληθυσμός πιθανότατα θα απογειωθεί γύρω από τη μέση τιμή (Μητροφάνη, 2021).

Τόσο το  $E(Z_n)$  όσο και το  $Var(Z_n)$  μπορούν να εκφραστούν με όρους αριθμητικού μέσου και διακύμανσης του αριθμού των απογόνων. Έτσι, έστω  $E(\xi) = m$  και  $Var(\xi) = \sigma^2$ . Αυτοί είναι ο αριθμητικός μέσος και διακύμανσης για ένα μεμονωμένο άτομο. Για να υπολογίσουμε τον αναμενόμενο αριθμό μεγέθους πληθυσμού, θέτουμε  $\{Z_0, Z_1, Z_2, \dots\}$  μια κλαδωτή ανέλιξη με  $Z_0 = 1$ ,  $\xi$  τον αριθμό των απογόνων και με αρχική υπόθεση ότι  $E(\xi) = m$ , έχουμε:

$$E(Z_n) = E(Z_n|Z_0 = 1) = m^n \quad (4.6)$$

Για την απόδειξη του παραπάνω ισχυρισμού επαγωγικά και σύμφωνα με τη βιβλιογραφία (Haccou et al., 2005; Schinazi, 1999), πρέπει πρώτα να αναφερθεί πως:

$$E(Z_1|Z_0 = 1) = E(\xi) = m = m^1$$

Η παραπάνω εξίσωση είναι αληθής για  $n = 1$ . Ας υποθέσουμε ότι ο τύπος λειτουργεί για το  $n$ . Δεσμεύοντας ως προς την τιμή της μεταβλητής  $Z_n$ , λαμβάνεται:

$$E(Z_{n+1}|Z_0 = 1) = \sum_{k \geq 1} E(Z_{n+1}|Z_n = k)P(Z_n = k|Z_0 = 1) \quad (4.7)$$

Καθώς η στοχαστική ανέλιξη διακριτού χρόνου  $\{Z_n\}$  είναι Μαρκοβιανή, προκύπτει σύμφωνα με την μαρκοβιανή ιδιότητα ότι:

$$E(Z_{n+1}|Z_0 = 1, Z_n = k) = E(Z_{n+1}|Z_n = k) \text{ για κάθε } k \geq 1. \text{ Έτσι ισχύει ότι:}$$

$$E(Z_{n+1}|Z_n = k) = E(\sum_{i=1}^k \xi_i) = km \text{ καθώς έγινε η υπόθεση πως } E(\xi) = m$$

Έτσι η εξίσωση (4.7) γίνεται:

$$E(Z_{n+1}|Z_0 = 1) = \sum_{k \geq 1} km \cdot P(Z_n = k|Z_0 = 1) = mE(Z_n|Z_0 = 1)$$

Καθώς είναι γνωστό πως:

$$\sum_{k \geq 1} k \mathbb{P}(Z_n = k|Z_0 = 1) = E(Z_n|Z_0 = 1).$$

και από την επαγωγική υπόθεση ισχύει ότι:  $E(Z_n|Z_0 = 1) = m^n$ , εξάγεται ότι:

$$E(Z_{n+1}|Z_0 = 1) = mm^n = m^{n+1}$$

$$E(Z_{n+2}|Z_0 = 1) = mm^{n+1} = m^{n+2} \text{ κοκ}$$

Έτσι, από την επαγωγική ανάλυση προκύπτει ότι  $E(Z_n|Z_0 = 1) = m^n$ , στο οποίο ο αναμενόμενος αριθμός μεγέθους πληθυσμού  $E(Z_n)$  σχετίζεται με τη μέση τιμή του αριθμού των απογόνων.

Για να υπολογίσουμε τη διακύμανση, έστω  $\{Z_0, Z_1, Z_2, \dots\}$  μια κλαδωτή ανέλιξη με  $Z_0 = 1$ , έστω το  $Y$  συμβολίζει την τυχαία μεταβλητή που δηλώνει το πλήθος των απογόνων  $\zeta$  και γίνεται η υπόθεση πως  $E(Y) = m$  και  $Var(Y) = \sigma^2$ . Τότε,

$$Var(Z_n) = \begin{cases} \sigma^2 n & \text{αν } m = 1 \\ \sigma^2 m^{n-1} \left( \frac{m^n - 1}{m - 1} \right) & \text{αν } m \neq 1 \end{cases} \quad (4.8)$$

Τα ακόλουθα παραδείγματα δίνουν απλές εφαρμογές των προαναφερθέντων τύπων:

1. Μέγεθος οικογένειας  $Y \sim Geometric(p=0.3)$

$$\text{Στην περίπτωση αυτή έχουμε, } m = E(\xi) = \frac{q}{p} = \frac{0.7}{0.3} = 2.33$$

$$\sigma^2 = \frac{q}{p^2} = \frac{0.7}{(0.3)^2} = 7.78$$

Για τον αναμενόμενο πληθυσμό της  $10^{15}$  γενιάς ( $n=10$ ) ισχύει:

$$E(Z_{10}) = m^{10} = 2.33^{10} = 4,715.87$$

Καθώς  $m = 2.33 \neq 1$

$$Var(Z_{10}) = (7.78)^2 (2.33)^2 \left( \frac{2.33^{10} - 1}{2.33 - 1} \right) = 43.4 \times 10^7$$

2. Μέγεθος οικογένειας  $Y \sim Geometric(p=0.5)$

Στην περίπτωση αυτή, η μέση τιμή και η διακύμανση της  $Y$  είναι αντίστοιχα ίσες με  $X$

$$m = E(\xi) = \frac{q}{p} = \frac{0.5}{0.5} = 1.00$$

και

$$\sigma^2 = \frac{q}{p^2} = \frac{0.5}{(0.5)^2} = 2.00$$

Για τον αναμενόμενο πληθυσμό της  $10^{15}$  γενιάς ( $n=10$ ) ισχύει:

$$E(Z_{10}) = m^{10} = 1^{10} = 1$$

και

$$Var(Z_{10}) = \sigma^2 n = 2 \times 10 = 20$$

Καθώς αυξάνει το  $n$ , η διακύμανση αυξάνει ενώ μέση τιμή παραμένει αμετάβλητη.

Μέσα σε αυτά τα παραδείγματα μπορούμε να εντοπίσουμε πώς ο αριθμητικός μέσος και η διακύμανση μπορούν να χρησιμοποιηθούν ως εργαλεία πρόβλεψης προκειμένου να εντοπισθεί τι συμβαίνει με την εξαφάνιση του πληθυσμού πριν υπολογίσουμε την πιθανότητα εξαφάνισης. Μια χαμηλή διακύμανση προειδοποιεί για το γεγονός ότι η διαδικασία συγκεντρώνεται στενά γύρω από τις μέσες τιμές. Είναι όμως αρκετή η διακύμανση ενός πληθυσμού για να προβλεφθεί η εξαφάνιση ενός πληθυσμού; Η απάντηση είναι όχι, γιατί πρέπει να γνωρίζουμε τη μέση τιμή γύρω από την οποία συγκεντρώνεται η διαδικασία (Μητροφάνη, 2021).

Στο παράδειγμα 1, ο μέσος όρος και η διακύμανση είναι υψηλές, επομένως οι τιμές κατανέμονται ευρέως γύρω από τον "υψηλό" μέσο όρο. Στο παράδειγμα 2, ο μέσος όρος και η διακύμανση είναι χαμηλές, πράγμα που σημαίνει ότι οι τιμές κατανέμονται κοντά στο ένα. Έτσι, ο πληθυσμός του παραδείγματος 2 είναι πιο πιθανό να εξαφανιστεί από τον πληθυσμό του παραδείγματος 1 (Μητροφάνη, 2021).

Επιπλέον, αυτή η υπόθεση εξηγεί, πρώτον, το γεγονός ότι εάν ο πληθυσμός εξαλειφθεί, είναι πιθανό αυτό να συμβεί πολύ γρήγορα, και δεύτερον γιατί ο προσδιορισμός της τιμής  $Var(Z_n)$  είναι πολύ σημαντικός. (Schinazi, 1999).

Εάν η συνάρτηση απογόνων ( $PGF$ ) είναι  $G(z)$ , τότε ο αναμενόμενος αριθμός μεγέθους πληθυσμού είναι  $G'(1)$ . Όπως έχει ήδη αναφερθεί, μία από τις κύριες ιδιότητες των κλαδωτών ανελίξεων είναι ότι η πιθανότητα τελικής εξαφάνισης εκτιμάται επίσης με τη χρήση των συναρτήσεων απογόνων (Μητροφάνη, 2021).

Με τον όρο εξάλειψη εννοούμε ότι η τυχαία ακολουθία  $\{Z_n\}$  είναι σε κάθε περίπτωση μηδέν για ένα πεπερασμένο αριθμό τιμών. Εφόσον το  $Z_n$  έχει ακέραια τιμή, η εξάλειψη είναι ένα γεγονός που αναγκάζει το  $Z_n$  να τείνει στο 0 ( $Z_n \rightarrow 0$ ). Ο πληθυσμός εξαφανίζεται έως τη γενιά  $n$  αν  $Z_n = 0$ , και μετά από αυτό το διάστημα ο πληθυσμός εξαφανίζεται για πάντα:  $Z_t = 0$  για κάθε  $t > n$  (Haccou et al., 2005).

Επιπλέον ισχύει ότι, αν ένας πληθυσμός εξαφανιστεί, παραμένει εξαφανισμένος. Έτσι, αν υποθέσουμε ότι η εξαφάνιση συμβαίνει τη στιγμή  $n=2$ , έχουμε ότι  $\{Z_2 = 0\}$  και αυτό επιβάλλει το  $\{Z_3 = 0\}$  κ.ο.κ. Η πιθανότητα τελικής εξαφάνισης αναφέρεται στην πιθανότητα ότι ο πληθυσμός θα εξαφανιστεί έως τη γενιά  $n$ , για οποιαδήποτε τιμή του  $n$  και αν το  $q_n$  αντιπροσωπεύει την πιθανότητα εξαφάνισης στη γενιά  $n$ , τότε (Haccou et al., 2005; Schinazi, 1999):

$$P(\text{ultimate extinction}) = q = \lim_{n \rightarrow \infty} q_n$$

Λαμβάνοντας υπόψη την αρχική διατύπωση των Watson-Erlang (σχέσεις (3.1)-(3.4)), η συνάρτηση απογόνων διατυπώνεται ως εξής:

$$G(q) = p_0 + p_1q + p_2q^2 + \dots \quad (4.9)$$

όπου  $p_i$  είναι η πιθανότητα να έχουμε  $i = 1, 2, 3, \dots$  άτομα και  $q$  είναι η πιθανότητα τελικής εξαφάνισης. Έστω  $q_n$  η πιθανότητα εξαφάνισης στην  $n$ -οστή χρονική περίοδο (γενιά) και  $q_0 = 0$ , καθώς  $n \rightarrow \infty$  έπεται  $q_n \rightarrow q$  ( $0 \leq q \leq 1$ ).

Προκειμένου ο πληθυσμός να εξαφανιστεί κατά την περίοδο  $n$ , όλα τα άτομα θα πρέπει να πεθάνουν τουλάχιστον έως την αμέσως προηγούμενη περίοδο  $n - 1$ . Επομένως,

$$q_n = p_0 + p_1q_{n-1} + p_2q_{n-1}^2 + \dots \quad (4.10)$$

Σύμφωνα με τη συνάρτηση απογόνων (σχέση (4.9)), η σχέση (4.10) μπορεί να γραφτεί ως  $q_n = G(q_{n-1})$  και σύμφωνα με τον Watson, καθώς  $q_n \rightarrow q$ , η πιθανότητα εξαφάνισης πρέπει να ικανοποιεί το  $G(q) = q$ . Ενώ  $\sum p_i = 1$ , η λύση  $q=1$  είναι πάντα μια ρίζα της εξίσωσης  $G(q) = q$ . Επιπλέον, η πρώτη και η δεύτερη παράγωγος της συνάρτησης απογόνων σχηματίζονται αντίστοιχα ως εξής, σε σχέση με το  $q$ :

$$G'(q) = p_1 + 2p_2q + 3p_3q^2 + \dots \quad (4.11)$$

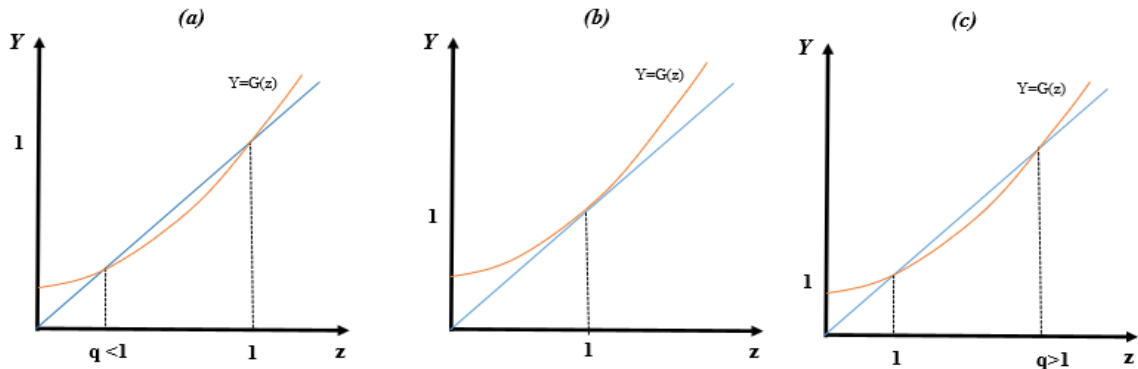
$$G''(q) = 2p_2 + 3 \cdot 2p_3q + 4 \cdot 3p_4q^2 + \dots \quad (4.12)$$

Είναι προφανές ότι και οι δύο παράγωγοι είναι μη αρνητικές για το  $q \geq 0$  και αποδεικνύεται ότι το  $G(q)$  είναι μια μη φθίνουσα συνάρτηση με κοίλα προς τα πάνω.

Έτσι, τα  $y = z$  και  $y = G(z)$  θα πρέπει να τέμνονται το πολύ σε δύο σημεία και το ένα από αυτά είναι το  $(1,1)$  καθώς το  $q = 1$  είναι πάντα μια ρίζα της εξίσωσης. Κατ'αυτὸν τὸν τρόπο,

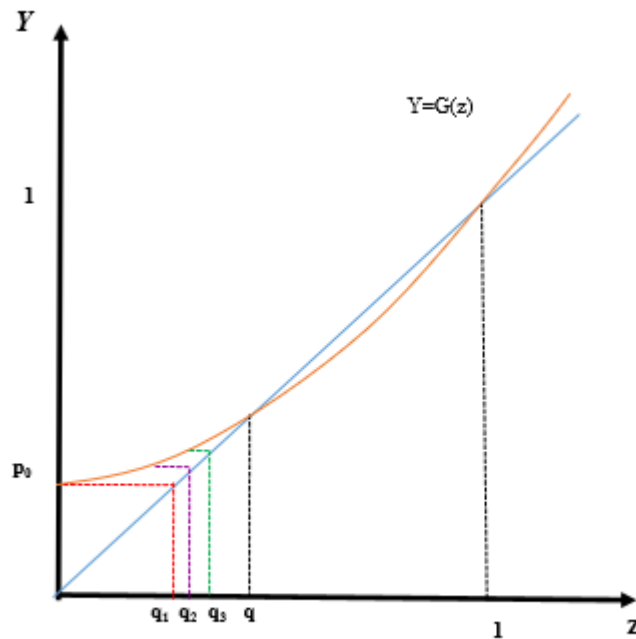
υπάρχουν τρία σενάρια, που απεικονίζονται στο Σχήμα. (4.2) (Athreya and Ney, 1972; Fazlollahtabar and Niaki, 2018):

- 2 ρίζες  $(q,1)$ , όπου  $0 \leq q < 1$  (Σχήμα (4.2α))
- 1 ρίζα  $q = 1$  (Σχήμα (4.2β))
- 2 ρίζες  $(1,q)$ , όπου  $q > 1$  (Σχήμα (4.2γ))



Σχήμα 4.2. Γραφήματα σεναρίων σημείων τομής των  $y = z$  και  $y = G(z)$ .

Ενώ το  $q$  αναφέρεται στην πιθανότητα εξαφάνισης πρέπει να είναι  $0 \leq q \leq 1$ , επομένως στις περιπτώσεις (b) και (c) η μόνη λύση είναι  $q = 1$ . Ωστόσο, υπενθυμίζοντας ότι γενικά  $G'(1) = m$ , η πρώτη παράγωγος της συνάρτησης απογόνων στο 1 δίνει τον αναμενόμενο αριθμό μεγέθους πληθυσμού, την περίπτωση (a), που παρουσιάζεται στο σχήμα 4.3 παρακάτω, αντιστοιχεί σε μια υπερκρίσιμη περίπτωση  $G'(1) = m > 1$  (Μητροφάνη, 2021).



Σχήμα 4.3. Σημεία τομής των  $y = z$  και  $y = G(z)$  σε ένα σενάριο υπερκρίσιμης περίπτωσης.

Από τα παραπάνω, και επειδή  $q_0 = 0, q_1 = G(q_0), \dots, q_n = G(q_{n-1})$ , τα σημεία  $(q_i, G(q_i))$  θα βρίσκονται πάντα πάνω από την ευθεία  $y = z$  (Fazlollahtabar and Niaki, 2018). Επομένως, υπάρχει μια ακόμη λύση που θα πρέπει να είναι μικρότερη από 1 ( $q < 1$ ) που υπολογίζεται ως η μικρότερη ρίζα της εξίσωσης  $G(z) = z$ .

Η πιθανότητα εξαφάνισης σχετίζεται, όπως έχει ήδη αναφερθεί με τον αριθμητικό μέσο ( $m$ ). Ορισμένες τιμές του  $m$  εγγυώνται τη λήξη της κλαδωτής ανέλιξης με πιθανότητα 1. Εάν ο μέσος αριθμός απογόνων ανά άτομο  $m$  είναι 1 ή λιγότερος, τότε η τελική εξαφάνιση είναι βέβαιη (Μητροφάνη, 2021). Άλλα τιμές του  $m$  εγγυώνται ότι η πιθανότητα εξαφάνισης θα είναι μικρότερη από 1. Εάν ο μέσος αριθμός απογόνων ανά άτομο  $m$  είναι μεγαλύτερος από 1, τότε η τελική εξαφάνιση του πληθυσμού δεν είναι βέβαιη αλλά μπορεί να συμβεί (Lange, 2010).

Σύμφωνα με την προαναφερθείσα ανάλυση, έστω ότι  $\{Z_0, Z_1, Z_2, \dots\}$  μια κλαδωτή ανέλιξη με  $Z_0 = 1$ , και  $m$  ο αναμενόμενος αριθμός μεγέθους αναπαραγωγής. Έστω επίσης  $q$  η πιθανότητα της τελικής εξαφάνισης. Τότε, σύμφωνα με τη βιβλιογραφία (Athreya and Ney, 1972; Haccou et al., 2005):

- Εάν  $m > 1$ , τότε  $q < 1$  : η εξαφάνιση δεν είναι εγγυημένη (υπερκρίσιμη περίπτωση)
- Εάν  $m < 1$ , τότε  $q = 1$  : η εξαφάνιση είναι εγγυημένη (υποκρίσιμη περίπτωση)

- Εάν  $m = 1$ , τότε  $q = 1$  : η εξαφάνιση δεν είναι εγγυημένη εκτός εάν υπάρχει μόνο μια πληθυσμιακή μονάδα (κρίσιμη περίπτωση)

Μόλις διαμορφωθεί η συνάρτηση απογόνων, μπορούμε να ορίσουμε τα ακόλουθα δεδομένα σχετικά με μια κλαδωτή ανέλιξη:

- Η πιθανότητα τελικής εξαφάνισης ως τη μικρότερη ρίζα του  $G(z) = z$ .
- Η πιθανότητα του πληθυσμού να εξαφανιστεί σε μια συγκεκριμένη γενιά ( $q_n$ ).
- Οι συνθήκες εξαφάνισης (ο αναμενόμενος αριθμός πληθυσμού) ως  $G'(1)$ .

#### 4.2.1 Τα μοντέλα κλαδωτών ανελιξεων συνεχούς χρόνου

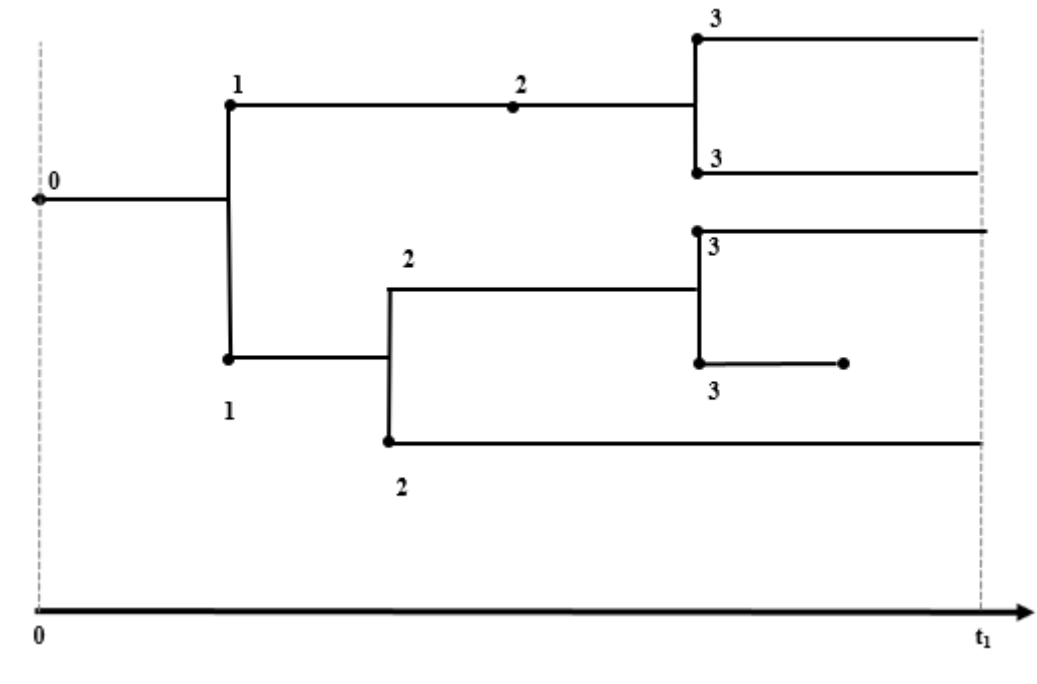
Όπως αναφέρθηκε και στην προηγούμενη παράγραφο, σε μια διαδικασία κλαδωτής ανέλιξης, οι μελλοντικές καταστάσεις είναι ανεξάρτητες από το παρελθόν. Αυτό σημαίνει ότι οι κλαδωτές αυτές ανελιξεις ικανοποιούν την ιδιότητα των διαδικασιών (αλυσίδων) *Markov*, όπου το μέλλον εξαρτάται μόνο από το παρόν και καθόλου από το παρελθόν. Επιπλέον, στην κλασική προσέγγιση *Galton-Watson*, κάθε άτομο ζει ακριβώς για μία μονάδα χρόνου. Υποθέτουμε ότι η διάρκεια ζωής των ατόμων του πληθυσμού κατανέμεται εκθετικά και κατ' αυτόν τον τρόπο έχουμε μια διαδικασία *Markov* συνεχούς χρόνου. Οι αλυσίδες *Markov* είναι οι αντίστοιχες των κλαδωτών ανελιξεων κατά *Galton-Watson (GWBP)* σε συνεχή χρόνο (*Haccou et al., 2005*).

Η κύρια διαφορά μεταξύ των κλαδωτών ανελιξεων συνεχούς και διακριτού χρόνου είναι ότι οι γεννήσεις και οι θάνατοι συμβαίνουν σε τυχαίους χρόνους στις κλαδωτές ανελιξεις συνεχούς χρόνου. Επιπλέον, οι τελευταίες έχουν την ιδιότητα *Markov* εάν και μόνο εάν οι χρόνοι γέννησης και θανάτου κατανέμονται εκθετικά (*Haccou et al., 2005*).

Επιπλέον, οι κλαδωτές ανελιξεις διακριτού χρόνου είναι πού πιο απλές από τις αντίστοιχες συνεχούς. Στις ΚΑΔΧ, ο πληθυσμός, κάθε φορά, περιέχει άτομα μόνο από μία γενιά. Σε συνεχόμενο χρόνο αυτό δεν είναι αλήθεια καθώς άτομα από διαφορετικές διαφορετικής γενιές μπορεί να είναι ζωντανά ταυτόχρονα (*Μητροφάνη, 2021*).

Αυτό συμβαίνει λόγω της διαφοράς στη διάρκεια ζωής. Το Σχήμα 4.4 δείχνει ένα οικογενειακό δέντρο μιας κλαδωτής ανέλιξης συνεχούς χρόνου. Το άτομο πεθαίνει μετά από ένα μεμονωμένο γεγονός αναπαραγωγής, που είναι η απλούστερη μορφή ΚΑΣΧ. Οι αριθμοί στους κλάδους (χρόνοι ζωής) δηλώνουν τη δημιουργία κάθε ατόμου. Σε αυτό το γενικό παράδειγμα, τη στιγμή  $t_1$  ο πληθυσμός αποτελείται από ένα σωματίδιο δεύτερης γενιάς και τρία σωματίδια τρίτης γενιάς (*Μητροφάνη, 2021*).





Σχήμα 4.4. Δέντρο Κλαδωτής Ανέλιξης Συνεχούς Χρόνου.

Στην κλαδωτή ανέλιξη συνεχούς χρόνου μια ενσωματωμένη διαδικασία παραγωγής χρησιμοποιείται συνήθως για την εξαγωγή ορισμένων θεμελιωδών αποτελεσμάτων από τις κλαδωτές ανελιξεις διακριτού χρόνου και την προσάρτησή τους στην αντίστοιχη ανέλιξη συνεχούς χρόνου. Η ενσωματωμένη διαδικασία μετράει τον αριθμό των πληθυσμιακών μονάδων διαφορετικών ηλικιών (*Haccou et al.*, 2005).

Έτσι, αυτή η διαδικασία αποτελείται από τυχαίες μεταβλητές με ακέραιο αριθμό και συνήθως συμβολίζεται ως  $Z_n$ . Για παράδειγμα, στο σχήμα 4.4 είχαμε  $Z_0 = 1, Z_1 = 2, Z_2 = 3$ . Βιβλιογραφικά, έχει παρατηρηθεί πως πολλοί συγγραφείς δεν χρησιμοποιούν διαφορετικό σύμβολο για την ενσωματωμένη διαδικασία, απλώς ονομάζουν κλαδωτή ανέλιξη διακριτού χρόνου  $Z_n$  (*Haccou et al.*, 2005).

Η διαδικασία ενσωματωμένης παραγωγής είναι μια κλαδωτή ανέλιξη διακριτού χρόνου. Υποθέτοντας ότι  $\xi_1, \xi_2, \dots, \xi_n$  είναι ο συνολικός αριθμός απογόνων των διαφόρων ατόμων στη  $n$ -οστή γενιά σε όλη τους τη ζωή, υπενθυμίζουμε την αντίστοιχη σχέση

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i$$

Επιπλέον, η υπόθεση ότι όλα τα άτομα αναπαράγονται ανεξάρτητα το ένα από το άλλο και έχουν την ίδια κατανομή (αναπαραγωγής), οδηγεί στο αποτέλεσμα ότι η ενσωματωμένη διαδικασία είναι μια κλαδωτή ανέλιξη κατά *Galton-Watson* (*Athreya and Ney, 1972*).

Μια διαδικασία ενσωματωμένης παραγωγής δεν μεταφέρει άλλες χρήσιμες πληροφορίες σχετικά με την αρχική κλαδωτή ανέλιξη συνεχούς χρόνου, αλλά χρησιμοποιείται ευρέως, όπως αναφέρθηκε προηγουμένως, για την απλοποίηση της αρχικής διαδικασίας και μέσω αυτής ελαχιστοποιούμε την υπολογιστική προσπάθεια ώστε να επιτευχθούν θεμελιώδη αποτελέσματα. Για παράδειγμα, η πιθανότητα εξαφάνισης σε μια κλαδωτή ανέλιξη συνεχούς χρόνου καθορίζεται από την πολύ απλούστερη ενσωματωμένη διαδικασία. Εξάλλου, η εξάλειψη της αρχικής ανέλιξης συμβαίνει εάν και μόνο εάν η ενσωματωμένη διαδικασία εξαλειφθεί επίσης (*Haccou et al., 2005*).

#### **4.2.2 Κλαδωτές Ανελίξεις με ένα αναπαραγωγικό γεγονός ανά κύκλο ζωής**

Οι ανελίξεις αυτές συνήθως θεωρούνται ως διεργασίες διαχωρισμού, επειδή το άτομο μετά από έναν κύκλο ζωής αντικαθίσταται από τους απογόνους του. Σε αυτή την ενότητα θα αναλύσουμε τις ανελίξεις εκείνες στις οποίες η αναπαραγωγή είναι ανεξάρτητη από τη διάρκεια ζωής του προγόνου και τα άτομα αναπαράγονται ανεξάρτητα το ένα από το άλλο καθώς και τις ανελίξεις εκείνες όπου η αναπαραγωγή είναι ηλικιοεξαρτώμενη (*Haccou et al., 2005*).

##### **i. Κλαδωτές ανελίξεις κατά *Markov* και διαδικασίες γέννησης και θανάτου**

Οι μαρκοβιανές κλαδωτές ανελίξεις είναι αντίστοιχες του κλαδωτών ανελίξεων συνεχούς χρόνου. Οι ΚΑΣΧ είναι διαχωριζόμενες μαρκοβιανές κλαδωτές ανελίξεις με ορισμένα χαρακτηριστικά. Πρώτον, υπάρχει ένας τύπος ατόμων του πληθυσμού, η αναπαραγωγή γίνεται μεμονωμένα μεταξύ των πληθυσμιακών μονάδων και τα άτομα αναπαράγονται με την ίδια κατανομή. Η ιδέα είναι ότι ένα άτομο ζει μια διάρκεια ζωής η οποία κατανέμεται εκθετικά και στη συνέχεια χωρίζεται σε  $\xi$  άτομα/απογόνους. Η διαδικασία αναπαραγωγής κάθε ατόμου είναι ανεξάρτητη μεταξύ τους και επιπλέον, όπως αναφέρθηκε προηγουμένως, η κατανομή αναπαραγωγής είναι ανεξάρτητη από την ηλικία του γονέα (*Μητροφάνη, 2021*).

Μαρκοβιανές κλαδωτές ανελίξεις χαρακτηρίζονται από την παράμετρο  $\lambda$  του εκθετικά κατανομημένου χρόνου ζωής και την κατανομή των τιμών  $\xi$  των απογόνων. Εδώ, η παράμετρος  $\lambda$  μετρά την πιθανότητα διάσπασης ανά μονάδα χρόνου και είναι σταθερή λόγω της εκθετικής κατανομής χρόνου ζωής. Η παράμετρος  $\lambda$  ονομάζεται, διαφορετικά από πολλούς συγγραφείς,

ως ποσοστό θανάτου, κίνδυνος ή ένταση (*Athreya and Ney, 1972; Haccou et al., 2005; Schinazi, 1999*).

Η στοχαστική διαδικασία  $(Z_t)_{t \geq 0}$ , που δίνει το μέγεθος του πληθυσμού ανά πάσα στιγμή  $t \geq 0$ , καθορίζεται από τη δομή που περιγράφεται παραπάνω και την αρχική συνθήκη  $Z_0$ . Η εκθετική κατανομή έχει μια μορφή συνάρτησης πυκνότητας πιθανότητας  $\lambda e^{-\lambda t}, t > 0$  για κάποιο  $\lambda > 0$ . Στα μαθηματικά σύμβολα, μια πραγματική, μη αρνητική τυχαία μεταβλητή  $T$  ακολουθεί μια εκθετική κατανομή εάν και μόνο εάν (*Haccou et al., 2005*):

$$P(T > t + u | T > u) = P(T > t) \quad \text{για κάθε } t, u \geq 0$$

Με άλλα λόγια, η κατανομή πιθανοτήτων του υπόλοιπου χρόνου ζωής είναι ανεξάρτητη από το μέλλον/επόμενο, δεδομένου ότι έχει επιτευχθεί μια καθορισμένη ηλικία. Τα άτομα με εκθετικά κατανομημένο χρόνο ζωής δεν γερνούν. Η ιδιότητα μη γήρανσης, επομένως, εξάγεται από την Μαρκοβιανή ιδιότητα όπως στις κλαδωτές ανελίξεις *Galton-Watson*, μας ενδιαφέρει η αναμενόμενη τιμή του μεγέθους του πληθυσμού (*Μητροφάνη, 2021*).

Επιπλέον, αν παραβλέψουμε τον κίνδυνο να συμβαίνουν περισσότεροι από ένας θάνατοι σε κάθε γενιά, υποθέτουμε ότι το  $Z_u$  είναι  $\xi$  ή 1. Αυτό εξαρτάται από το αν ο πρόγονος της διαδικασίας έχει πεθάνει έως τη χρονική στιγμή  $u$  ή όχι (και στη συνέχεια απέκτησε  $\xi$  παιδιά). Όπως στις κλαδωτές ανελίξεις *Galton-Watson* ίσχυε πως  $m = E[\xi]$ , αντίστοιχα εδώ εξάγεται πως (*Athreya and Ney, 1972; Haccou et al., 2005; Schinazi, 1999*):

- Αν  $m > 1$ , το μέγεθος του πληθυσμού έχει εκθετική αύξηση.
- Αν  $m < 1$ , το μέγεθος του πληθυσμού έχει εκθετική μείωση.

Οι μαρκοβιανές κλαδωτές ανελίξεις συνήθως σχετίζονται με διαδικασίες γέννησης και θανάτου που είναι μοντέλα συνεχούς χρόνου. Αυτές οι διεργασίες ορίζονται συνήθως ως διεργασίες Markov με ακέραια τιμή  $Z_t$  με την ιδιότητα ότι η ένταση της προόδου/ανόδου κατά ένα βήμα από το μέγεθος του πληθυσμού  $Z_t = j$  είναι  $jb$  και η αντίστοιχη ένταση της μείωσης κατά ένα βήμα είναι  $jd$  (*Μητροφάνη, 2021*).

Με  $d$  συμβολίζουμε το ρυθμό θανάτων και με  $b$  το ρυθμό αναπαραγωγής. Η ιδέα πίσω από την παραπάνω θεώρηση είναι ότι αν το μέγεθος του πληθυσμού είναι  $Z_t = j$ , έχουμε  $j$  άτομα πληθυσμού την παρούσα χρονική στιγμή, το καθένα με ποσοστό γεννήσεων  $b$ . Έτσι, ολόκληρος ο πληθυσμός αυξάνεται με το ποσοστό  $jb$  και μειώνεται με ποσοστό  $jd$ . Σε ατομικό επίπεδο, η διαδικασία γέννησης-θάνατος είναι μια δυαδική διαχωριζόμενη μαρκοβιανή κλαδωτή ανέλιξη (*binary splitting Markov branching process*). Οι χρόνοι ζωής κατανομούνται εκθετικά με την

παράμετρο  $(b+d)$  (Haccou et al., 2005; Schinazi, 1999). Έτσι, ο μέσος χρόνος ζωής είναι  $1/(b + d)$ , ενώ η κατανομή των απογόνων είναι:

$$P(\xi = 0) = \frac{d}{b + d} \text{ και } P(\xi = 2) = \frac{b}{b + d}$$

Ως εκ τούτου, αυτό το μοντέλο δημιουργεί την υπόθεση ότι ένα άτομο έχει μια πιθανότητα ανά μονάδα χρόνου να πεθάνει χωρίς απογόνους, ίση με  $\frac{d}{b+d}$  και μια πιθανότητα ανά μονάδα χρόνου να δημιουργήσει απόγονο ίση με  $\frac{b}{b+d}$ . Επιπλέον, οι διαδικασίες γέννησης και θανάτου αναφέρονται μερικές φορές ως γραμμικές διεργασίες (σε αντίθεση με τις μαρκοβιανές κλαδωτές ανελιξιζιστον *Markov*) με ρυθμό ανάπτυξης  $\mu_j$ , αν  $X_t = j$ . Αυτό αντιστοιχεί σε πληθυσμιακά εξαρτώμενες κλαδωτές ανελιξιζεις συνεχούς χρόνου (Athreya and Ney, 1972; Haccou et al., 2005; Schinazi, 1999).

ii. Δυαδικές κλαδωτές ανελιξιζεις συνεχούς χρόνου

Ορίζουμε μια δυαδική κλαδωτή ανέλιξη συνεχούς χρόνου με τις ακόλουθες ιδιότητες:

- Ο αριθμός των ατόμων τη στιγμή  $t$  συμβολίζεται με  $Z_t$  και ξεκινάμε με έναν μόνο τύπο αναπαραγωγικού ατόμου,  $Z_0 = 1$ .
- Ανεξαρτησία μεταξύ ατόμων
- Κάθε άτομο γεννά ένα νέο άτομο με πιθανότητα  $\lambda$  ή πεθαίνει με πιθανότητα 1.

Κάθε άτομο του πληθυσμού έχει δύο ανεξάρτητες εκθετικές τυχαίες μεταβλητές που συνδέονται με αυτό. Η μία τυχαία μεταβλητή έχει παράμετρο  $\lambda$  (γέννηση) και η άλλη έχει παράμετρο 1 (θάνατος). Ο κανόνας είναι ότι εάν η τυχαία μεταβλητή με  $\exp(\lambda)$  συμβαίνει πριν από την τυχαία μεταβλητή με ρυθμό 1, τότε το άτομο αντικαθίσταται από δύο άτομα (χωρίζεται στα δύο). Διαφορετικά, το άτομο πεθαίνει χωρίς απογόνους (Μητροφάνη, 2021).

Όπως συμβαίνει στις διαδικασίες γέννησης-θανάτου, παρατηρούμε ότι το πρόβλημα επιβίωσης για μια κλαδωτή ανέλιξη συνεχούς χρόνου είναι πανομοιότυπο με το πρόβλημα επιβίωσης για την αντίστοιχη κλαδωτή ανέλιξη διακριτού χρόνου. Τώρα, γνωρίζουμε ότι μια ΚΑΔΧ επιβιώνει εάν  $E(\xi_1) > 1$ . Η αναμενόμενη τιμή του  $\xi_1$  μπορεί να υπολογισθεί ακόμα κι αν δεν γνωρίζουμε την κατανομή του (Haccou et al., 2005).

Ως επόμενο βήμα, πρέπει να οριστεί μια διαφορική εξίσωση για το  $M_t$  (το αναμενόμενο μέγεθος του πληθυσμού τη στιγμή  $t$  δηλαδή  $E(Z_t)$ ) ώστε να εντοπισθεί τι θα συμβεί στον πληθυσμό στο μέλλον. Τη χρονική στιγμή 0 έχουμε ένα μόνο άτομο. Πρέπει να δούμε τι θα συμβεί σε ένα μικρό χρονικό διάστημα  $h$ . Έχουμε τρία πιθανά σενάρια:

1. Το άτομο δίνει απογόνους μεταξύ των χρονικών στιγμών 0 και  $h$ , με πιθανότητα  $\lambda h + o(h)$  (καθώς ένα άτομο γεννά/ χωρίζεται σε δύο άτομα με ρυθμό  $\lambda$ )

2. Το άτομο πεθαίνει χωρίς απογόνους, με πιθανότητα  $1 \cdot h + o(h)$  (γιατί με ρυθμό 1, ένα άτομο πεθαίνει)

3. Δεν συμβαίνει τίποτα απολύτως, με πιθανότητα  $1 - (\lambda + 1)h + o(h)$

Γενικά, μπορεί να υποθεθεί ότι, αν  $m = E(\xi) > 1$ , η διαδικασία επιβιώνει, διαφορετικά (αν  $m = E(\xi) < 1$ ) ο πληθυσμός πεθαίνει. Μια αξιοσημείωτη σημείωση εδώ είναι ότι, λαμβάνοντας υπόψη μια κλαδωτή ανέλιξη συνεχούς χρόνου με  $Z_0 = 1$  και υποθέτοντας ότι  $Z_1 = 3$ , καθένα από αυτά τα τρία άτομα του πληθυσμού ήταν δυνατό να εμφανιστεί σε διαφορετικούς χρόνους μεταξύ 0 και 1 (Μητροφάνη, 2021).

Ωστόσο, υποθέτοντας ότι  $(Z_n)_{n \geq 0}$  είναι μια κλαδωτή ανέλιξη διακριτού χρόνου, τίθεται ο ισχυρισμός ότι κάθε μια από αυτές τις πληθυσμιακές μονάδες, που έχουν την ίδια κατανομή απογόνων, ξεκινά την κλαδωτή ανέλιξη τη στιγμή 1. Αυτό προκύπτει από την ιδιότητα Markov (η ιδιότητα χωρίς μνήμη της εκθετικής κατανομής). Όπως αναφέραμε προηγουμένως, τα άτομα με εκθετικά κατανομημένο χρόνο ζωής δεν γερνούν. Σημασία έχει μόνο ότι ένα άτομο είναι παρόν τη στιγμή 1 και καθόλου πόσο χρόνων είναι τη στιγμή 1 (Athreya and Ney, 1972; Haccou et al., 2005).

iii. Κλαδωτές ανελίξεις κατά *Bellman-Harris* (εξαρτώμενες από την ηλικία)

Οι κλαδωτές ανελίξεις *Bellman-Harris* αποτελούν διαδικασίες διακλάδωσης σε συνεχή χρόνο, όπου τα άτομα έχουν μια αυθαίρετη κατανομή διάρκειας ζωής. Το ποσοστό κινδύνου δεν είναι σταθερό σε αυτή την περίπτωση. Έτσι, το ποσοστό κινδύνου είναι  $\lambda(\alpha)$ , όπου το  $\alpha$  υποδηλώνει την ηλικία του προγόνου. Έτσι, οι ανελίξεις αυτές χαρακτηρίζονται συνήθως ως ηλικιοεξαρτώμενες ονομάζονται, επειδή η αναπαραγωγή εξαρτάται από την ηλικία του προγόνου (Μητροφάνη, 2021).

Επιπλέον, στις διεργασίες *Bellman-Harris* υπάρχει ένας μοναδικός τύπος ατόμων που αναπαράγονται ανεξάρτητα το ένα από το άλλο και η αναπαραγωγή γίνεται με την ίδια κατανομή για κάθε άτομο. Μια χρήσιμη σημείωση σχετικά με το ποσοστό κινδύνου, που έχει παρατηρηθεί βιβλιογραφικά (Haccou et al., 2005), είναι ότι το ποσοστό επικινδυνότητας δεν επηρεάζει την αναμενόμενη αναπαραγωγή κατά τη διάρκεια ζωής, επομένως για την κρισιμότητα της διαδικασίας δεν έχει σημασία αν έχουμε μια διαδικασία διακλάδωσης *Markov* ή *Bellman-Harris*, αρκεί να έχουμε την ίδια κατανομή του αριθμού απογόνων. Ωστόσο, το

αναμενόμενο μέγεθος πληθυσμού σε μια δεδομένη στιγμή δεν είναι το ίδιο. Επιπλέον, η ηλικιακή δομή του πληθυσμού ως συνάρτηση του χρόνου διαφέρει στους δύο τύπους διαδικασιών (Haccou *et al.*, 2005).

### 4.2.3 Επιδημιολογικά Μοντέλα Στηριζόμενα σε Κλαδωτές Ανελίξεις

Στη σχετική βιβλιογραφία που σχετίζεται με τα μαθηματικά μοντέλα πληθυσμών που εξετάζουν τις επιδημίες, τα ντετερμινιστικά μοντέλα αποτελούν ισχυρά εργαλεία όταν ο εξεταζόμενος πληθυσμός είναι αρκετά μεγάλος. Αντίθετα, τα στοχαστικά μοντέλα μπορούν να παρέχουν χρήσιμες πληροφορίες όταν το μέγεθος του πληθυσμού είναι μάλλον μικρό (Allen, 2015). Η θεωρία των κλαδωτών ανελίξεων είναι απλή: αυτές οι διαδικασίες χρησιμοποιούνται συχνά για τη μοντελοποίηση φυσικών συστημάτων στα οποία ένα άτομο ενός τύπου ζει για μια μονάδα χρόνου και με το θάνατό του παράγει  $n$  πανομοιότυπα αντίγραφα του εαυτού του. Πολυπλοκότερες διαδικασίες αυτής της απλής διαδικασίας περιγράφονται από τις κλαδωτές ανελίξεις πολλαπλών τύπων (Haccou *et al.*, 2005).

Η συγκεκριμένη κατασκευή της προαναφερθείσας διαδικασίας δικαιολογεί ότι αυτές οι ανελίξεις εφαρμόζονται συνήθως στη βιολογία, τη φυσική και την επιδημιολογία. Έτσι, κάθε άτομο δημιουργεί ένα μέγεθος οικογένειας, όπου τα μεγέθη της οικογένειας είναι ανεξάρτητα και πανομοιότυπα καταναμημένα (*independent and identically distributed/i.i.d*) τυχαίες μεταβλητές (Athreya and Ney, 1972).

Σχετικά με τη μελέτη πληθυσμών όπως έχει ήδη αναφερθεί, δύο ζητήματα είναι σημαντικά: ο αναμενόμενος αριθμός πληθυσμού και η εξαφάνιση (ακριβέστερα η πιθανότητα εξαφάνισης) (Athreya and Ney, 1972; Haccou *et al.*, 2005). Εάν το  $k$  είναι η πιθανότητα εξαφάνισης στη χρονική περίοδο  $k^{th}$ , τότε η μικρότερη θετική ρίζα της συνάρτησης δημιουργίας πιθανότητας  $h(d) = d$  ως  $d_k$ , τείνει στο  $d$ , ανιχνεύεται ως η πιθανότητα τελικής εξαφάνισης  $d$  (Μητροφάνη, 2021).

Επιπλέον, η πρώτη παράγωγος της πιθανογεννήτριας συνάρτησης, που συνήθως σχηματίζεται ως  $Q(s) = q_k s^k$ , χρησιμοποιείται για τον υπολογισμό του αναμενόμενου αριθμού μεγέθους πληθυσμού ( $m$ ) (Athreya and Ney, 1972; Haccou *et al.*, 2005). Σύμφωνα με τη σχετική βιβλιογραφία, το αναμενόμενο μέγεθος ενός πληθυσμού ( $m$ ) θα μπορούσε να χρησιμοποιηθεί ως μέτρο για την κατηγοριοποίηση μιας κλαδωτής ανέλιξης σε περιπτώσεις προβλέποντας την μελλοντική κατάσταση του συστήματος, ως εξής:

- Αν  $m > 1$  τότε  $d < 1$ : η εξάλειψη δεν είναι εγγυημένη (υπερκρίσιμη περίπτωση),
- Αν  $m < 1$ , τότε  $d = 1$ : η εξάλειψη είναι εγγυημένη (υποκρίσιμη περίπτωση) και τέλος
- αν  $m = 1$ , τότε  $d = 1$  : η εξάλειψη δεν είναι εγγυημένη εκτός εάν το μέγεθος της οικογένειας είναι ίσο με 1 (κρίσιμη περίπτωση)

# Κεφάλαιο 5

## Εφαρμογή

### 5.1 Εισαγωγή

Η δημιουργία και η εξάπλωση μολυσματικών ασθενειών είναι ένα σύνθετο φαινόμενο με διάφορους παράγοντες όπως το περιβάλλον στο οποίο το παθογόνο και πιθανοί δέκτες βρίσκονται, ο πληθυσμός που εκτίθεται σε αυτό, καθώς και οι δυναμικές του πληθυσμού που μελετάται. Ο ρόλος των μαθηματικών στην επιδημιολογία είναι να μοντελοποιήσει την δημιουργία και την εξάπλωση ενός ιού. Μια κυρίαρχη μέθοδος για να γίνει αυτό, είναι να χωριστεί ο πληθυσμός σε μικρότερες ομάδες υπό ορισμένες παραδοχές, οι οποίες αντιπροσωπεύουν την ευπάθειά τους έναντι στον ιό που βρίσκεται στο σύστημα. Τέτοιου είδους μοντέλα καλούνται διαχωριστικά μοντέλα (*Compartmental models*) στον τομέα της επιδημιολογίας και χρησιμεύουν για την κατανόηση των συστημάτων και των δυναμικών που δρα ο ιός (*Compartmental models in epidemiology, Wikipedia*).

Ιστορικά, υπάρχουν πολλές επιδημίες κατά τη διάρκεια των ετών. Για παράδειγμα ήδη από τον Θουκυδίδη περιγράφεται η επιδημία πανούκλας στην αρχαία Αθήνα (430 – 428 π.χ.) όταν 1050 από τους 4000 στρατιώτες πεθαίνουν. Ακόμη, η επιδημία πανούκλας στην Ευρώπη το 14ο αιώνα, ή αλλιώς «μαύρος θάνατος» (*black death*) εκείνη την εποχή, κατά τη διάρκεια της οποίας από ένα πληθυσμό περίπου 85 εκατ. κατοίκων, πεθαίνει περίπου το 1/3. Με πιο σύγχρονη τη νόσο του κορωνοϊού 2019 (*Covid-19*), η οποία εντοπίστηκε αρχικά τον Δεκέμβριο του 2019 και έχει εξαπλωθεί γρήγορα παγκοσμίως προκαλώντας τη συνεχιζόμενη πανδημία. Αυτή η νέα πανδημία του κορωνοϊού κρίθηκε ως η μεγαλύτερη παγκόσμια απειλή και στις 30 Ιανουαρίου 2020 ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) υποστήριξε ότι αυτή η νέα κατάσταση αποτελεί έκτακτη ανάγκη για τη δημόσια υγεία διεθνώς, λόγω των χιλιάδων μολυσμένων περιπτώσεων και θανάτων που αναφέρθηκαν σε όλο τον κόσμο (*Ndairou, et al. (2020), Ivorra, et al. (2020)*). Το πρώτο κρούσμα στην Ελλάδα επιβεβαιώθηκε στις 26 Φεβρουαρίου (2020).



Τα μαθηματικά μοντέλα μετάδοσης μολυσματικών ασθενειών περιγράφουν αποτελεσματικά και απλά την εξέλιξη των ασθενειών παρέχοντας ποσοτικά δεδομένα στην επιδημιολογία (Ivorra, et al. (2020)· Wang(2020); Ahmed, et al. (2020)). Επιπλέον, η χρήση μαθηματικών και στατιστικών εργαλείων επιτρέπει τη διεξαγωγή μακροπρόθεσμων ή βραχυπρόθεσμων προβλέψεων σχετικά με την εξάπλωση της νόσου στον πληθυσμό, ώστε να υποστηρίζονται οι αποφάσεις για στρατηγικές παρέμβασης, διαχείριση επιδημιών και υγειονομικής περίθαλψης (Ivorra, et al. (2020)· Wang (2020) Ahmed, et al. (2020); Overton, et al. (2020)). Υπάρχουν πολλά μαθηματικά μοντέλα στην επιδημιολογία, στο παρόν κεφάλαιο θα ασχοληθούμε με τα επιδημιολογικά μοντέλα *S-I-R* (Susceptible- Infected - Recovered) και *S-E-I-R* (*Susceptible - Exposed - Infected – Recovered*) με κύριο στόχο αυτής της εφαρμογής την ανάπτυξη ενός απλού μαθηματικού μοντέλου χρησιμοποιώντας τις κλαδωτές ανελίξεις για την ανάλυση του αρχικού σταδίου της έξαρσης του *Covid - 19* στην Ελλάδα.

## 5.2 Μοντέλο *SIR*

Στην επιδημιολογία το μοντέλο *SIR* χρησιμοποιείται πιο συχνά ως το πιο ρεαλιστικό μοντέλο για την μετάδοση μιας νόσου. Οι *Kermack* και *McKendrick* (*KM*) ορίζουν το ντετερμινιστικό μοντέλο επιδημίας με σταθερό πληθυσμό  $N$  ατόμων και τρεις καταστάσεις (*SIR*).

- *S (Susceptibles)*: Άτομα του πληθυσμού, τα οποία είναι υγιή αλλά μπορούν να μολυνθούν. Κοινώς, η ευπαθής ομάδα.
- *I (Infectives)*: Άτομα του πληθυσμού, τα οποία νοσούν (μολυσμένοι).
- *R (Removed)*: Άτομά που δεν μπορούν πλέον να προσβληθούν από τη μόλυνση και, κατά συνέπεια, δεν επιδρούν στη δυναμική μετάδοσης με κανέναν τρόπο όταν έρχονται σε επαφή με άλλα μέλη του πληθυσμού. Αυτό μπορεί να συμβεί είτε με τον εντοπισμό και την απομόνωσή τους (π.χ. καραντίνα), είτε με την ανοσοποίησή τους με κάποια τεχνητή μέθοδο (π.χ. εμβολιασμός) ή με την ανοσοποίησή τους εφόσον ιαθούν, είτε με θάνατο εξαιτίας της ασθένειας. Αν και από ιατρική και κοινωνική πλευρά ο τρόπος που ένα άτομο εισήλθε στην *R* κατηγορία είναι ουσιαστικός (αν όλοι εμβολιαστούν πριν νοσήσουν ή ιαθούν ο πληθυσμός επιβιώνει ποσοτικά ανεπηρέαστος, ενώ αν όλοι πεθάνουν, εφόσον νοσήσουν, ο πληθυσμός εξαλείφεται) από τη μαθηματική πλευρά της μοντελοποίησης πολλές φορές μια τέτοια διάκριση είναι ασήμαντη.



Σχήμα 5.1: Αναπαράσταση του Επιδημικού Μοντέλου *SIR*

Με βάση την παραπάνω περιγραφή, ο πληθυσμός ( $N$ ), σε ένα *SIR* υπόδειγμα, μπορεί να χωριστεί σε τρία ξένα μεταξύ τους σύνολα, έτσι ώστε να ισχύει η ισότητα:

$$N = S + I + R$$

Οι διαφορικές εξισώσεις που χρησιμοποιούνται για να περιγράψουν το επιδημικό μοντέλο *SIR* είναι οι ακόλουθες:

$$\frac{dS}{dt} = -\beta SI \quad (5.1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (5.2)$$

$$\frac{dR}{dt} = \gamma I \quad (5.3)$$

Όπου  $\beta, \gamma$  θετικές σταθερές, τις οποίες ορίζουμε ως ρυθμό μετάδοσης και ως ρυθμό ανάκαμψης αντίστοιχα.

Η υπόθεση για την (5.1) είναι ότι τα ευπαθή άτομα μολύνονται με ρυθμό ανάλογο με τον αριθμό των επαφών μεταξύ ατόμων του  $S$  και του  $I$ , υποθέτοντας ότι υπάρχει ομοιόμορφη ανάμειξη του πληθυσμού, ενώ για την (5.3) υποθέτουμε ότι ο ρυθμός με οποίο τα άτομα αδυνατούν να μεταδώσουν τη νόσο είναι ανάλογος με τον αριθμό των μολυσμένων. Εφ' όσον έχουμε υποθέσει ότι ο πληθυσμός παραμένει σταθερός, έχουμε ότι:

$$N = S(t) + I(t) + R(t) \quad (4.4)$$

Εάν  $S = 0$  σε κάποια δεδομένη χρονική στιγμή, τότε το υπόδειγμά

οδηγείται με πιθανότητα ένα προς μία σταθερή λύση:  $(S, I, R) = (0, 0, N)$ .

Εάν  $I = 0$ , τότε καμία μεταβολή δεν θα επέλθει στα μεγέθη  $S$  και  $R$ . Οπότε, εάν θεωρήσουμε ότι  $I_0 = 0$ , τότε το υπόδειγμά θα παραμείνει με πιθανότητα 1 στο σημείο όπου βρίσκεται τη χρονική στιγμή 0 και επομένως υπάρχει μια σταθερή λύση:  $(S, I, R) = (S_0, 0, R_0)$ . Στη συνέχεια θα αναφέρουμε περιπτώσεις όπου οι  $S, I$  λαμβάνουν αποκλειστικά θετικές τιμές. Σε μια προσπάθεια προσδιορισμού του πρόσημου των παραπάνω παραγώγων

Παρατηρείται ότι  $\frac{dS}{dt} < 0$ , σε οποιαδήποτε χρονική στιγμή. Δεν ισχύει το ίδιο για την  $\frac{dI}{dt}$  η οποία λαμβάνει θετικές τιμές αν και μόνο ισχύει  $S > \frac{\gamma}{\beta}$ . Αυτό σημαίνει ότι ο αριθμός των μολυσματικών ατόμων βαίνει αυξανόμενος, μόνο εάν ισχύει η προηγούμενη συνθήκη για τους ευαίσθητους. Επιπλέον, εφόσον το  $S$  θα βαίνει μειούμενο σε όλες τις χρονικές περιόδους, τελικά, η συνθήκη θα πάψει να ισχύει έπειτα από πεπερασμένο αριθμό περιόδων και το  $I$ , έπειτα από τη λήψη μιας μέγιστης τιμής θα αρχίσει να βαίνει μειούμενο έως ότου μηδενιστεί. Αντίθετα, εάν εξ αρχής ισχύει ότι  $S < \frac{\gamma}{\beta}$ , τότε το  $I$  βαίνει καθ' όλη την εξέλιξη του υποδείγματος μειούμενο έως ότου μηδενιστεί, προτού εξαντλήσει το σύνολο των  $S$ . Στην ουσία η πρώτη περίπτωση αποτελεί ένα μοντέλο που εκφράζει μια επιδημία, ενώ το δεύτερο περιγράφει μία μη επιδημική κατάσταση, όπου η μόλυνση εξαλείφεται αφ' εαυτής.

Είναι σημαντικό να προσδιοριστεί ο βασικός ρυθμός αναπαραγωγής. Ας υποθέσουμε έναν ιό που δεν έχει ακόμη προσβάλει κανένα μέλος του πληθυσμού. Συνεπώς το σύνολο του πληθυσμού βρίσκεται στην κατηγορία των ευαίσθητων και κανείς δεν έχει μολυνθεί ή ανοσοποιηθεί ( $S(0) = N$ ). Εάν κάτω από αυτές τις συνθήκες, μια μικρή ομάδα ατόμων μολυνθεί, αυτό θα έχει επίδραση στα υποσύνολα του πληθυσμού βασισμένη στο βασικό ρυθμό αναπαραγωγής. Συνεπώς, ο βασικός ρυθμός αναπαραγωγής προκύπτει ως:  $R_0 = \frac{\beta S_0}{\gamma}$  (Brauer, 2008).

Είναι σαφές ότι οι προσδιοριστικές συναρτήσεις του  $SIR$  μοντέλου δεν είναι αύξουσες (συνεπώς ένα κρούσμα είναι αδύνατο να οδηγήσει σε επιδημία), εάν δεν ισχύει η συνθήκη

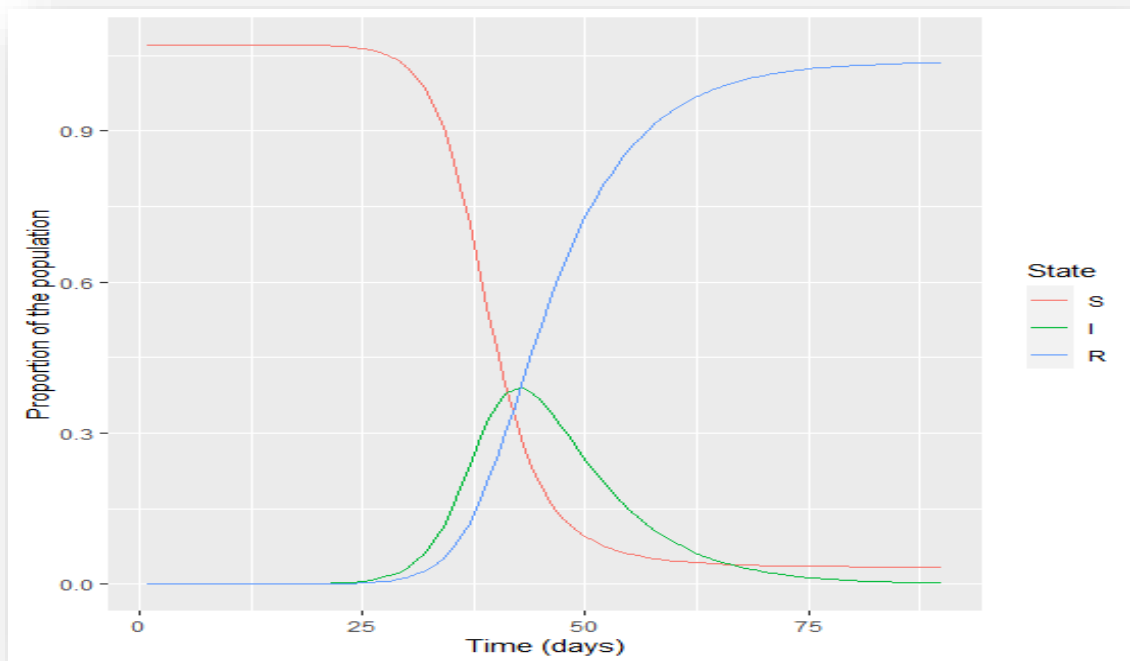
$R_0 > 1$ . Σε αυτή την περίπτωση ο αριθμός των μολυσμένων και μολυσματικών ατόμων  $I_t$  δεν μπορεί να είναι αύξων και δεν απαιτείται ιδιαίτερη ανησυχία και πρόληψη.

Σύμφωνα με τον *Watmough* (2009), το φαινόμενο αυτό μπορεί να περιγράψει μαθηματικά ως μία οριακή τιμή (threshold) την οποία πρέπει να υπερβεί το  $R_0$  για να καταστεί το μοντέλο επιδημικό. Με άλλα λόγια ένα ορισμένο  $R_0 < 1$  συσχετίζεται (*ceteris paribus*) και με ένα δεδομένο άνω όριο μολύνσεων πριν την οριστική εξάλειψη της απειλής (όχι ωστόσο και της ασθένειας). Έτσι, μπορούμε να ισχυριστούμε ότι η επίδραση ενός εμβολίου, μιας θεραπείας ή ενός μέτρου πρόληψης επί του βασικού ρυθμού αναπαραγωγής αποτελεί ένα καλό μέτρο για την αξιολόγησή τους.

### **5.3 Αριθμητική εφαρμογή με δεδομένα κορωνοϊού (Covid - 19), για την περίπτωση της Ελλάδας**

Το πρώτο κρούσμα COVID-19 στην Ελλάδα επιβεβαιώθηκε στις 26 Φεβρουαρίου, όταν ιδιώτης επέστρεψε από μια επίσκεψη στη Βόρεια Ιταλία. Τα περισσότερα από τα επόμενα κρούσματα τον Φεβρουάριο και τις αρχές Μαρτίου σχετίζονταν με άτομα που είχαν επιστρέψει από στην Ιταλία, το Ισραήλ και την Αίγυπτο ή με τις επαφές αυτών. Μέχρι τις 22 Μαρτίου, όπου η ελληνική κυβέρνηση ανακοίνωσε το πρώτο γενικό lockdown, αναφέρθηκαν 624 κρούσματα και 2 θάνατοι. (Μητροφάνη, 2021).

Στο σχήμα 5.2 που ακολουθεί βλέπουμε τις καταστάσεις του μοντέλου *SIR* αναφορικά με την περίπτωση της Ελλάδας για τις πρώτες 90 ημέρες από την έναρξη της πανδημίας του Covid 19. Έχουμε ότι το  $R_0 > 1$ , συνεπώς κάθε άτομο που μολύνεται θα μολύνει περισσότερα από ένα άτομα, έτσι η επιδημία θα εξαπλωθεί, δηλαδή  $\frac{dI}{dt} > 0$ . Παρατηρούμε ότι ο αριθμός των ευπαθών ατόμων μειώνεται αρχικά και έπειτα σταθεροποιείται ενώ παράλληλα ο αριθμός των μολυσματικών αυξάνεται το ίδιο χρονικό διάστημα, το οποίο δηλώνει και την εξάπλωση της επιδημίας. Ο αριθμός των διαγραμμένων αυξάνεται και τελικά σταθεροποιείται.



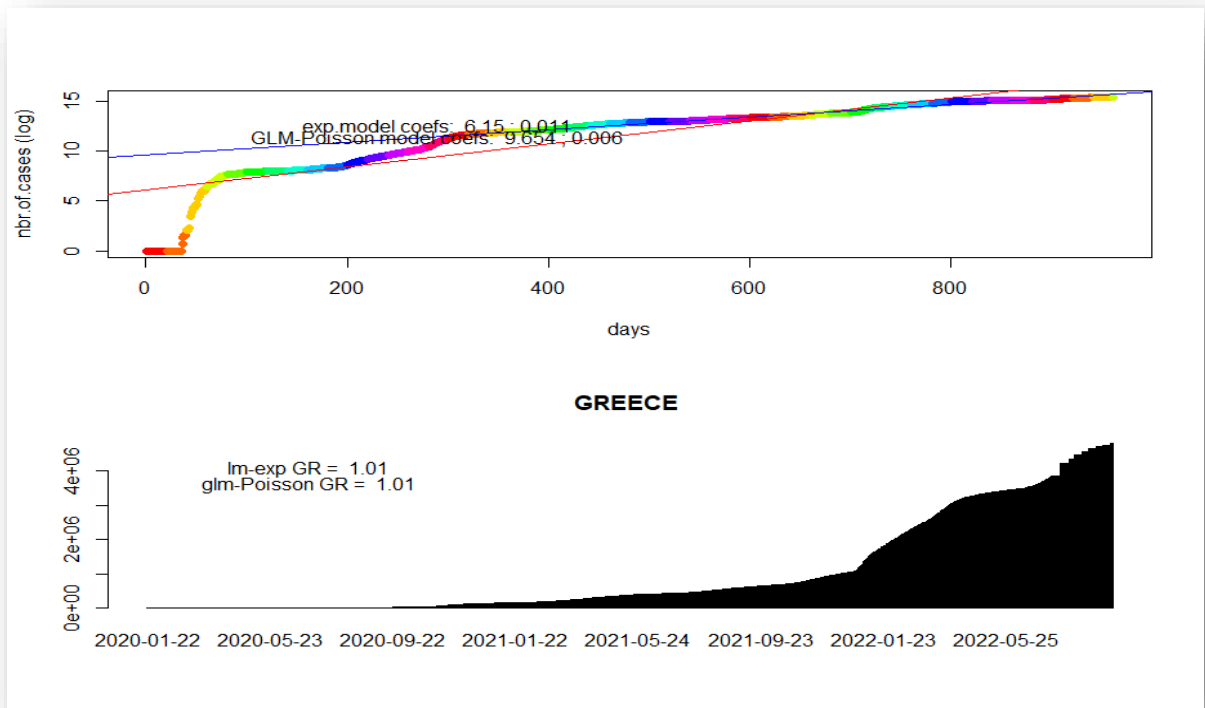
Σχήμα 5.2: : Καταστάσεις του Μοντέλου SIR με  $R_0 > 1$

#### 5.4 Εφαρμογή του υπολογιστικού πακέτου *covid19.analytics* της *R*

Η *R* είναι μια γλώσσα προγραμματισμού ανοικτού κώδικα και περιβάλλον που παρέχει στον χρήστη τη δυνατότητα να κάνει υπολογιστική στατιστική και γραφήματα. Από την έναρξη της πανδημίας έχει εισάγει στο επίσημο *CRAN* περισσότερα από δώδεκα πακέτα που σχετίζονται με την πανδημία *Covid 19*. Συγκεκριμένα, σε αυτή την ενότητα θα χρησιμοποιήσουμε το πακέτο *covid19.analytics R*, το οποίο επιτρέπει στους χρήστες να λαμβάνουν ζωντανά παγκόσμια δεδομένα για τον *Covid - 19*, καθώς και σε εργαλεία ανάλυσης και οπτικοποίησης για την επεξεργασία των δεδομένων. Αυτό συμβαίνει με την πρόσβαση και την ανάκτηση των δημόσιων δεδομένων που είναι διαθέσιμα και δημοσιεύονται από πολλές πηγές όπως το “*COVID - 19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*”.

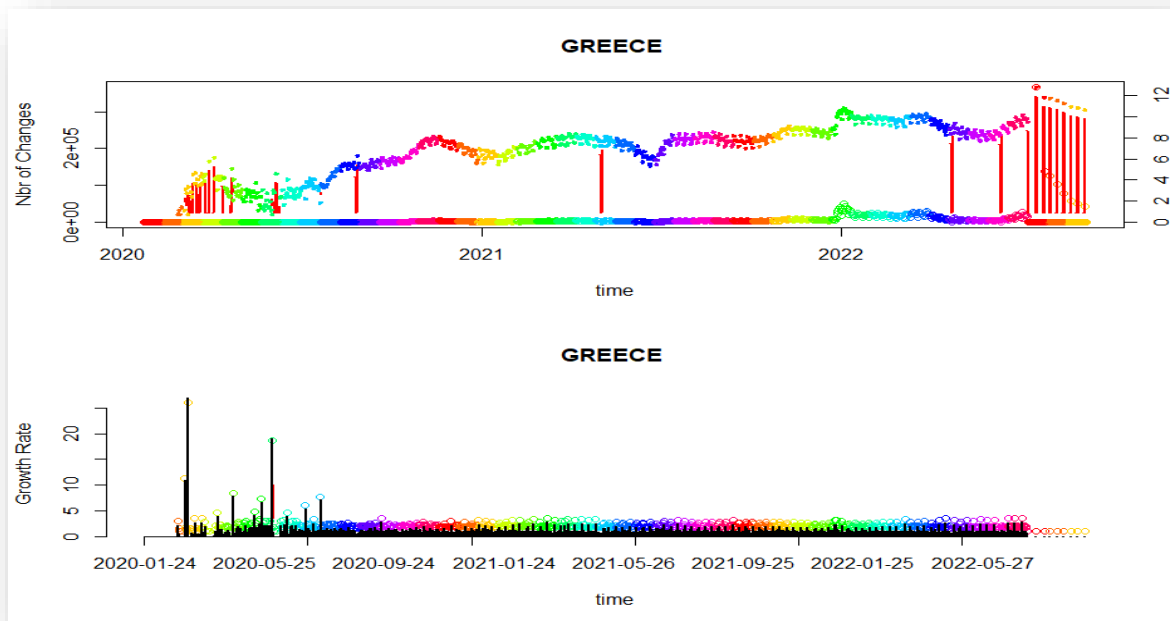
Το πακέτο *covid19.analytics* επιτρέπει στους χρήστες να διερευνούν τον συνολικό αριθμό κρουσμάτων ανά τοποθεσία (Σχήμα 5.3) μέσω της συνάρτησης *totals.per.location*. Επιπλέον, είναι δυνατό να εκτιμηθούν οι παγκόσμιες τάσεις ανά τοποθεσία με χρήση των συναρτήσεων *single.trend*, *mtrends* και *itrends*. Τέλος, όπως αναφέραμε και παραπάνω ένα από τα νέα

χαρακτηριστικά που προστέθηκαν από το πακέτο *covid19.analytics*, είναι η δυνατότητα μοντελοποίησης της εξάπλωσης του ιού ενσωματώνοντας πραγματικά δεδομένα. Η συνάρτηση *generate.SIR.model* υλοποιεί ένα απλό μοντέλο SIR για καθορισμένο σύνολο δεδομένων και μια συγκεκριμένη τοποθεσία. Η συνάρτηση *generate.SIR.model* επιχειρεί να εκτιμήσει τις κατάλληλες τιμές για τις παραμέτρους  $\beta$  και  $\gamma$  από την έναρξη της πανδημίας χρησιμοποιώντας πραγματικά δεδομένα.



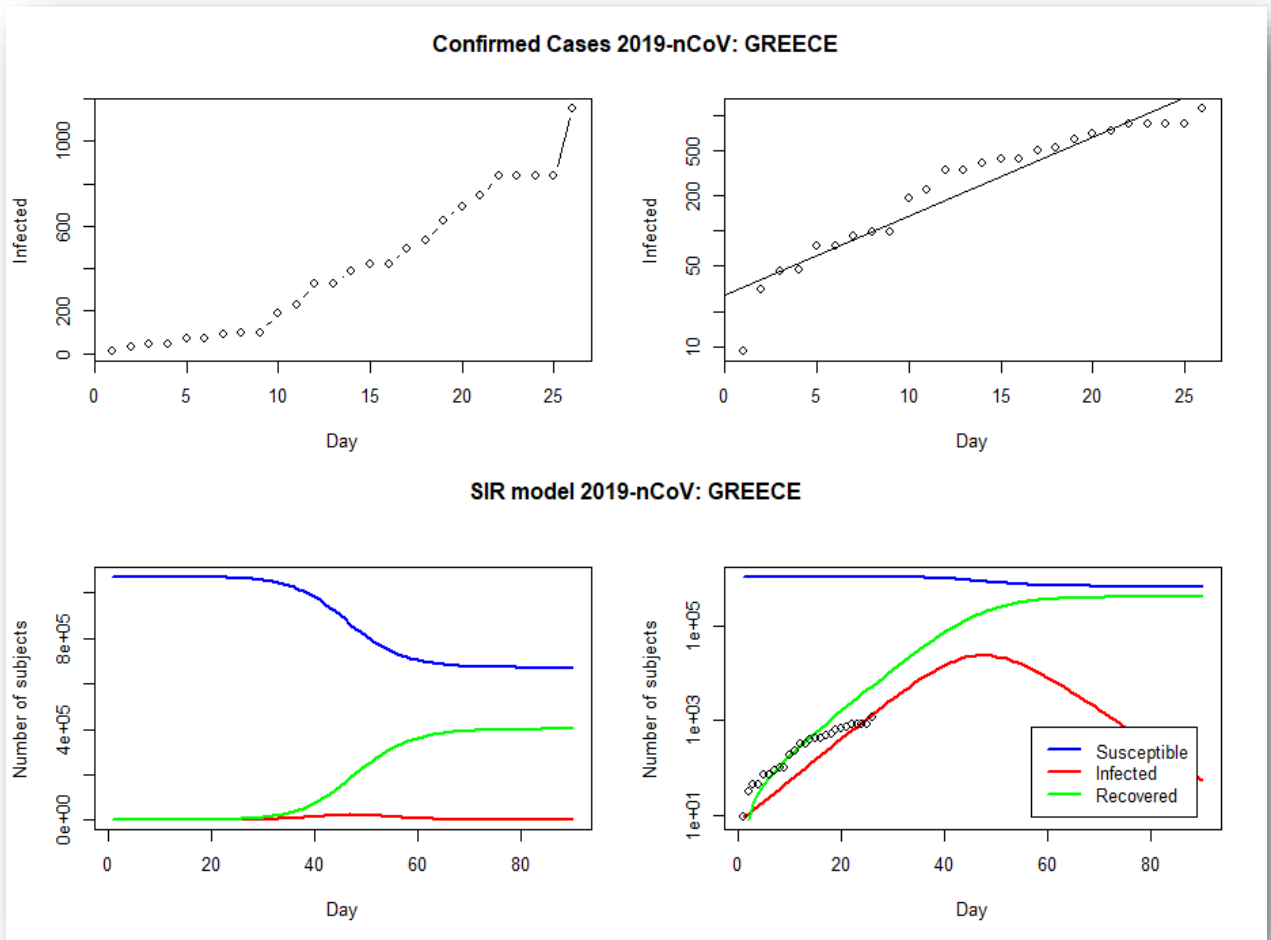
Σχήμα 5.3 : Γραφική απεικόνιση του αριθμού των κρουσμάτων (*in log*) και του ρυθμού αύξησης του *COVID - 19* στην Ελλάδα

Το πρώτο γράφημα δείχνει τον αριθμό των κρουσμάτων σε λογαριθμική κλίμακα στον κάθετο άξονα και τον αριθμό των ημερών στον οριζόντιο άξονα. Παρατηρείται απότομη αύξηση του αριθμού των κρουσμάτων από την ημέρα 30. Η ημέρα 50 παρουσίασε μια σταθερή αύξηση στον αριθμό των περιπτώσεων σε σύγκριση με την αύξηση της ημέρας 30. Στο δεύτερο γράφημα παρουσιάζεται ο αριθμός των κρουσμάτων σε γραμμική κλίμακα. Παρατηρείται ότι υπήρξε μία απότομη αύξηση στα κρούσματα το Δεκέμβριο του 2021.



Σχήμα 5.4 : Αριθμός αλλαγών (διαφορά στον αριθμό των κρουσμάτων) και ρυθμός ανάπτυξης του *COVID* - 19 στη Ελλάδα

Το σχήμα 5.4 προκύπτει από την συνάρτηση *growth.rate*. Τα γραφήματα εμφανίζουν την ημερήσια μεταβολή των κρουσμάτων τόσο σε γραμμική όσο και σε λογαριθμική κλίμακα ως συνάρτηση του χρόνου. Παρατηρείται από το πρώτο γράφημα ότι υπήρξε μία σταθερή διαφορά στον αριθμό των κρουσμάτων από την έναρξη της πανδημίας μέχρι τα τέλη Μαρτίου το 2020, ενώ η διαφορά παρέμεινε σταθερή περίπου μέχρι τις αρχές του καλοκαιριού όπου και σημειώθηκε απότομη μείωση στον αριθμό των κρουσμάτων. Τα κρούσματα αυξήθηκαν ξανά στις αρχές του Φθινοπώρου και συνέχισαν να αυξάνονται μέχρι τα τέλη του 2021. Στο δεύτερο γράφημα του σχήματος 5.4 παρατηρείται ότι ο ρυθμός ανάπτυξης του ιού στην Ελλάδα κορυφώθηκε στα τέλη Μαΐου 2020 με την επόμενη κορυφή να εμφανίζεται τον Ιούλιο του 2020 εμφανώς μικρότερη από αυτή του Μαΐου.



Σχήμα 5.5 Γραφική απεικόνιση του μοντέλου *SIR* για την περίπτωση της Ελλάδας.

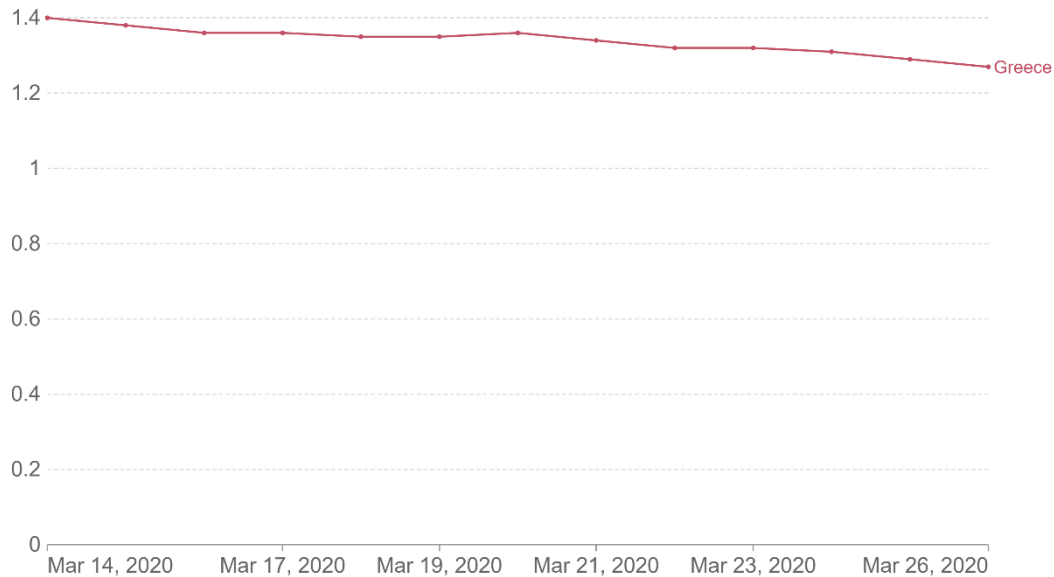
Στο πρώτο γράφημα απεικονίζεται ο αριθμός των κρουσμάτων σε γραμμική κλίμακα ενώ στο δεύτερο γράφημα σε λογαριθμική. Στο τρίτο γράφημα παρουσιάζεται το μοντέλο *SIR* σε γραμμική κλίμακα ενώ στο τέταρτο γράφημα σε λογαριθμική, με τις κουκίδες να αποτυπώνουν τον πραγματικό αριθμό των κρουσμάτων. Το πακέτο *covid19.analytics* υπολογίζει την τιμή βήτα, την τιμή γάμμα και τον βασικό ρυθμό αναπαραγωγής, με τις παραμέτρους του μοντέλου να είναι  $\beta = 1$  για τον ρυθμό μετάδοσης της νόσου,  $\gamma = 0.79$  για τον ρυθμό ανάρρωσης και το  $R_0 = 1.25$ . Από την τιμή που λαμβάνεται για το  $R_0$ , προκύπτει ότι κατά μέσο όρο ένα μολυσματικό άτομο μολύνει 1.25 ευπαθή άτομα κατά την περίοδο μόλυνσης του. Η τιμή που προκύπτει για το  $R_0$  ανήκει στο εύρος μεταξύ 1.4 - 2.5, το οποίο προτείνεται και από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ).



## Estimate of the effective reproduction rate (R) of COVID-19



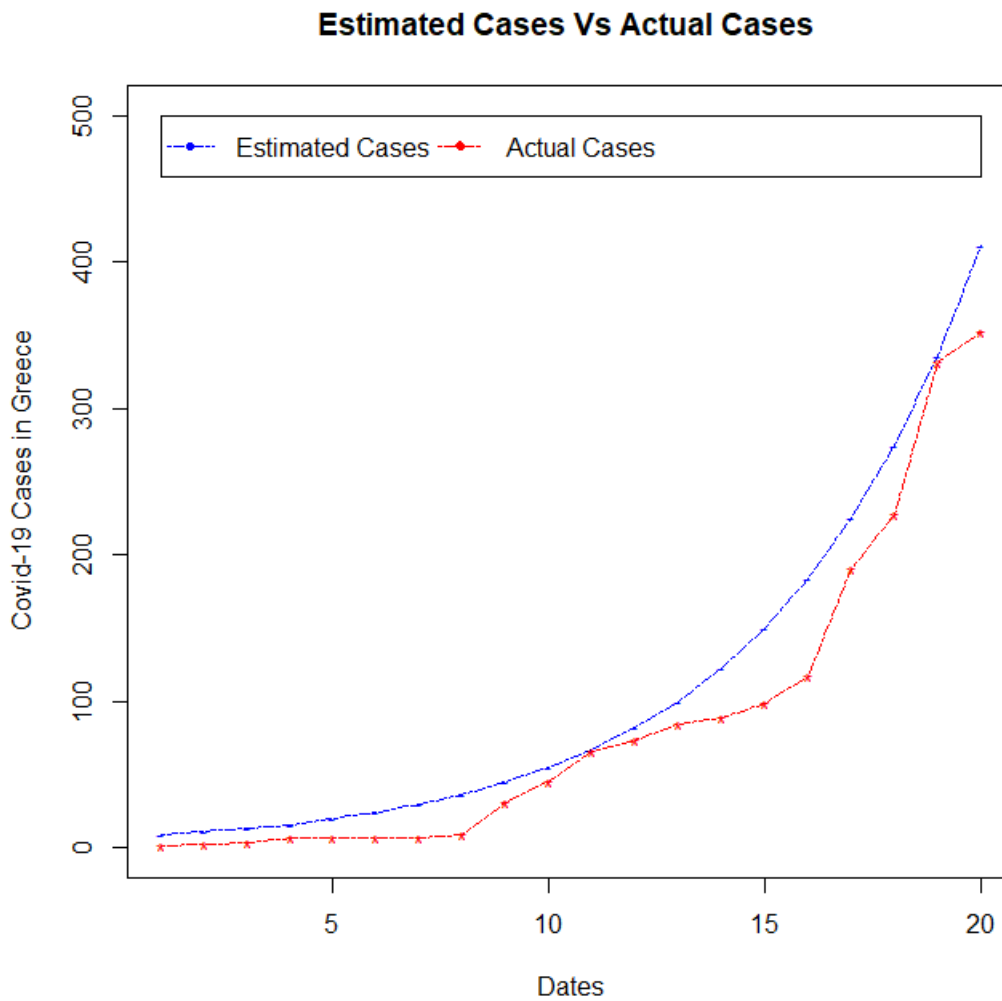
The reproduction rate represents the average number of new infections caused by a single infected individual. If the rate is greater than 1, the infection is able to spread in the population. If it is below 1, the number of cases occurring in the population will gradually decrease to zero.



Source: Arroyo-Marioli F, Bullano F, Kucinskis S, Rondón-Moreno C (2021) Tracking R of COVID-19: A new real-time estimation using the Kalman filter. CC BY

Σχήμα 5.6 Ρυθμός αναπαραγωγής στην Ελλάδα 14/03/20 -26/03/20 (ourworldindata.org)

Η εκτίμηση του βασικού ρυθμού αναπαραγωγής από πακέτο *covid19.analytics* για τις 16/02/2020 ήταν  $R_0 = 1.25$ . Σύμφωνα με την ιστοσελίδα Our World in Data η εκτίμηση του ρυθμού αναπαραγωγής για τις 16/03/2020 ήταν  $R_0 = 1.36$ . Παρατηρούμε ότι δεν υπάρχει σημαντική απόκλιση στην εκτίμηση του μοντέλου από την πραγματική τιμή του βασικού ρυθμού αναπαραγωγής.



Σχήμα 5.7. Γραφική απεικόνιση των εκτιμήσεων για τον αριθμό των κρουσμάτων μέσω του μοντέλου *SIR* σε αντιπαραβολή με τον πραγματικό αριθμό κρουσμάτων για τις πρώτες είκοσι ημέρες της πανδημίας στην Ελλάδα.

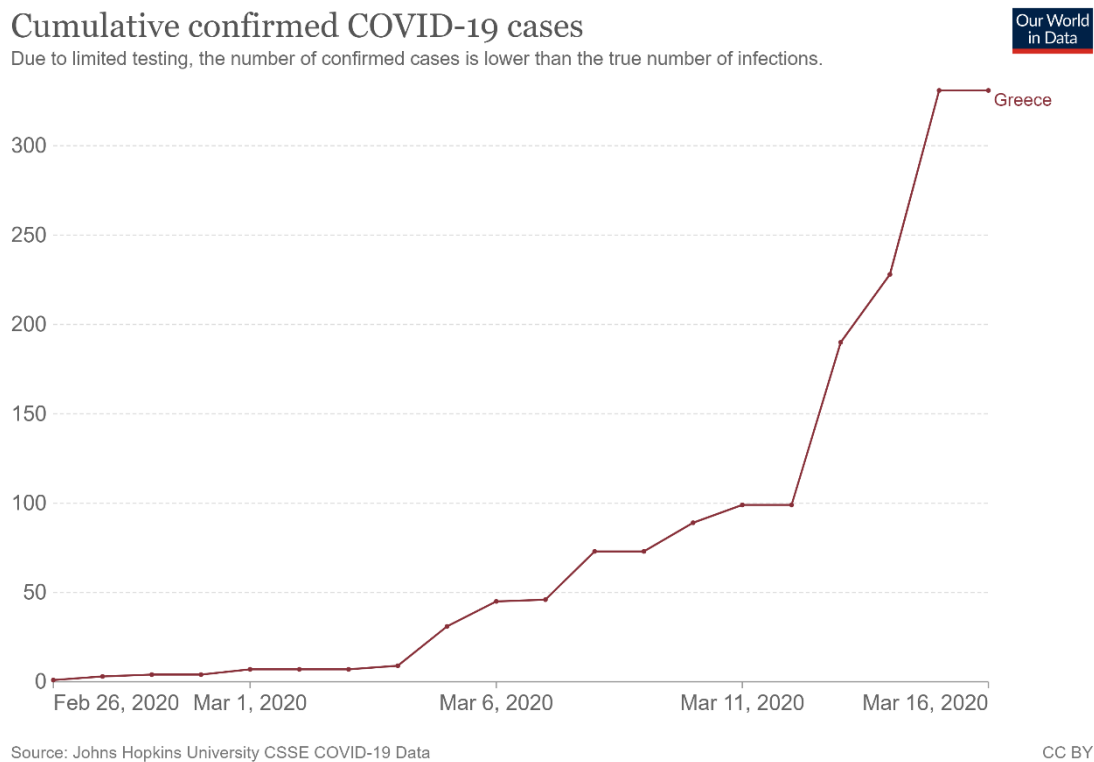
Date	Estimation of cases	Actual cases
26/2/2020	9	1
27/2/2020	11	3
28/2/2020	13	4

29/2/2020	16	7
1/3/2020	20	7
2/3/2020	24	7
3/3/2020	30	7
4/3/2020	36	9
5/3/2020	45	31
6/3/2020	55	45
7/3/2020	67	66
8/3/2020	82	73
9/3/2020	100	84
10/3/2020	123	89
11/3/2020	150	99
12/3/2020	184	117
13/3/2020	225	190
14/3/2020	275	228
15/3/2020	336	331
16/3/2020	411	352

Πίνακας 5.1: Εκτιμήσεις κρουσμάτων μέσω της συνάρτησης *generate.SIR.model* σε αντιπαραβολή με τα πραγματικά κρούσματα για τις πρώτες είκοσι ημέρες της πανδημίας στην Ελλάδα

Στο σχήμα 5.7 και στον πίνακα 5.1 παρουσιάζονται οι προβλέψεις των θετικών κρουσμάτων *Covid – 19*, μέσω του μοντέλου *SIR* σε αντιπαραβολή με τα πραγματικά κρούσματα που καταγράφηκαν τις πρώτες είκοσι ημέρες της πανδημίας στην Ελλάδα. Παρατηρούμε ότι δεν

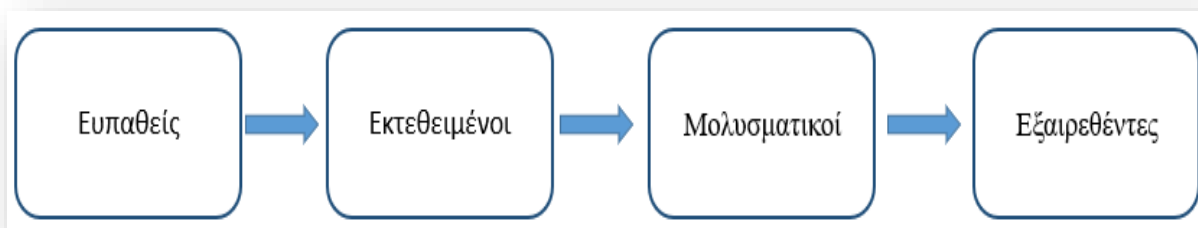
υπάρχουν μεγάλες αποκλίσεις μεταξύ των εκτιμημένων και των πραγματικών τιμών. Ο αριθμός των κρουσμάτων αυξάνεται εκθετικά μετά την 8 ημέρα. Θα παραθέσουμε μερικούς από τους παράγοντες οι οποίοι επηρέασαν την εξέλιξη του αριθμού των κρουσμάτων και οι οποίοι δεν λαμβάνονται υπόψη στα μοντέλα *SIR*. Τα πεδία του πίνακα που είναι μαρκαρισμένα αφορούν τις Κυριακές, ημέρες που ο αριθμός των κρουσμάτων που δηλωνόταν ήταν μικρότερος από τα πραγματικά κρούσματα καθώς δεν διενεργούνταν μεγάλος αριθμός εργαστηριακών τεστ. Είναι επίσης σημαντικό να αναφέρουμε ότι από τις 9 Μαρτίου 2020 η κυβέρνηση ξεκίνησε σταδιακά να λαμβάνει προληπτικά μέτρα αναχαίτησης της επιδημίας του κορονοϊού, με αποκορύφωση το πρώτο καθολικό *lockdown* στις 23 Μαρτίου. Συνεπώς, τόσο η κυβερνητική δράση, όσο και η αντίδραση του κοινού έχουν ισχυρές επιπτώσεις στην εξέλιξη της πανδημίας και κατ' επέκταση στον αριθμό των κρουσμάτων.



Σχήμα 5.8 Αθροιστικός αριθμός κρουσμάτων covid19 στην Ελλάδα 26/2/20-16/3/20(ourworldindata.org)

## 5.5 Μοντέλο SEIR

Το μοντέλο *SEIR* είναι επαυξημένο του *SIR* κατά μία πληθυσμιακή ομάδα. Η νέα αυτή πληθυσμιακή ομάδα *E* (*Exposed*) αφορά τα άτομα τα οποία έχουν εκτεθεί στον ίο, έχουν ασθενήσει αλλά για κάποιο χρονικό διάστημα δεν μπορούν να μεταδώσουν την ασθένεια. Επιπλέον, πρέπει να αναφέρουμε ότι το μοντέλο *SEIR* δεν λαμβάνει υπόψη θανάτους ή γεννήσεις κατά την περίοδο εκτίμησης.



*Σχήμα 5.9: Αναπαράσταση του Επιδημικού Μοντέλου SEIR*

Με βάση την παραπάνω περιγραφή, ο πληθυσμός ( $N$ ), σε ένα *SEIR* υπόδειγμα, μπορεί να χωριστεί σε τέσσερα ξένα μεταξύ τους σύνολα, έτσι ώστε να ισχύει η ισότητα:

$$N = S + E + I + R$$

$$\frac{dS}{dt} = -\beta SI \quad (5.5)$$

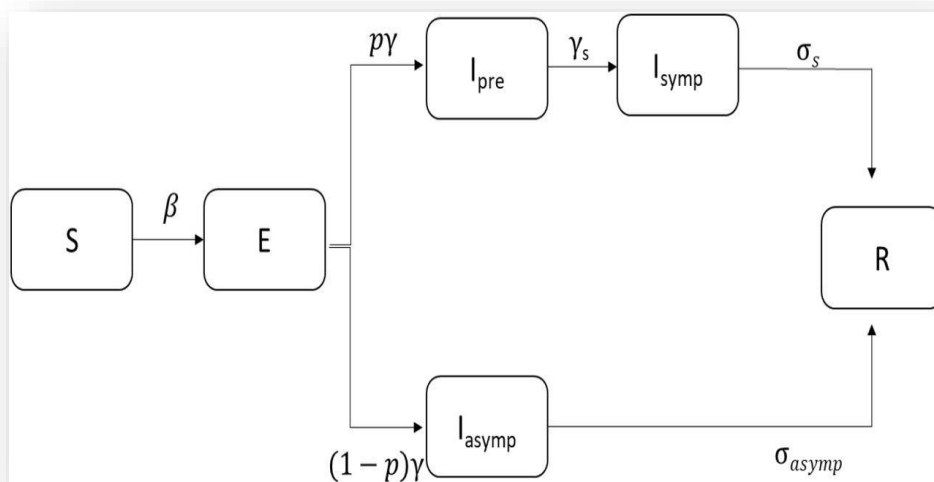
$$\frac{dI}{dt} = \sigma E - \gamma I \quad (5.6)$$

$$\frac{dR}{dt} = \gamma I \quad (5.7)$$

$$\frac{dE}{dt} = \beta SI - \sigma E \gamma I \quad (5.8)$$

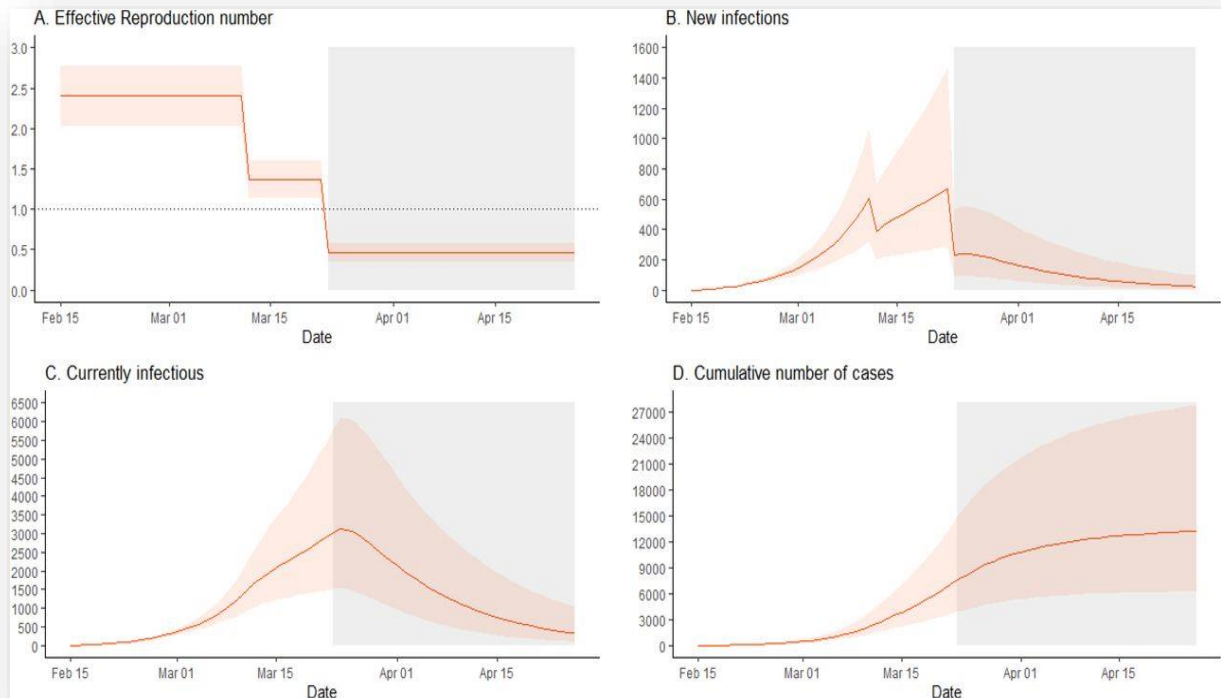
Η παράμετρος  $\sigma$  εκφράζει τον ρυθμό εκδήλωσης συμπτωμάτων της ασθένειας από τη στιγμή της απόκτησής της.

Στη μελέτη που πραγματοποιήθηκε από την Σύψα και συνεργάτες το 2020 (Sypsa et. al, 2020) επιλέχθηκε ένα πιο σύνθετο μοντέλο *SEIR* για την προσομοίωση της επιδημίας στην Ελλάδα (Φεβρουάριος-Απρίλιος 2020). Ο πληθυσμός χωρίζεται σε 6 πληθυσμιακές ομάδες τους ευπαθείς, τους εκτεθειμένους, τους μολυσματικούς πριν την εμφάνιση συμπτωμάτων, τους μολυσματικούς και κλινικά άρρωστους, τους πραγματικά ασυμπτωματικούς και μερικώς μολυσματικούς και τέλος αυτούς που έχουν αναρρώσει. Έγινε η υπόθεση ότι ένα ποσοστό  $p = 80\%$  των μολυσμένων περιπτώσεων θα αναπτύξει συμπτώματα και ότι η μεταδοτικότητα μπορεί να εμφανιστεί 1,5 ημέρες πριν από την έναρξη των συμπτωμάτων. Προκειμένου να εκτιμηθεί ο ρυθμός αναπαραγωγής  $R_t$ , παρουσία μέτρων κοινωνικής αποστασιοποίησης, αξιολογήθηκε η μείωση του  $R_0$  χρησιμοποιώντας δεδομένα κοινωνικών επαφών που συλλέχθηκαν από μια έρευνα που εφαρμόστηκε κατά τη διάρκεια του *lockdown*. Το αντίκτυπο των μέτρων μοντελοποιήθηκε πολλαπλασιάζοντας το ποσοστό μόλυνσης  $\beta$  με μια παράμετρο  $\sigma$  που αντιστοιχεί στη μείωση του  $R_t$  σε δύο μεγάλες περιόδους μέτρων: την περίοδο των αρχικών μέτρων, συμπεριλαμβανομένου του κλεισίματος σχολείων, χώρων διασκέδασης και καταστημάτων (με εξαίρεση σούπερ μάρκετ, παντοπωλεία και φαρμακεία) μέχρι την ημέρα πριν από το *lockdown* (11 Μαρτίου-22 Μαρτίου) και την περίοδο καραντίνας (23 Μαρτίου-26 Απριλίου).



Σχήμα 5.10. Τροποποιημένο μοντέλο *SEIR* (Sypsa et. al, 2020)

Στο σχήμα 5.9 ο πληθυσμός ταξινομούνται σε Ευαίσθητους ( $S$ ), Εκτεθειμένους ( $D$ ), Μολυσματικούς (πριν εμφανίσουν συμπτώματα:  $I_{pre}$ , κλινικά άρρωστους:  $I_{asympt}$  ή ασυμπτωματικούς :  $I_{asympt}$ ) και σε αυτούς που έχουν αναρρώσει ( $R$ ). Υποθέσαμε ότι ένα ποσοστό  $p$  των εκτεθειμένων περιπτώσεων θα αναπτύξει συμπτώματα και ότι η μεταδοτικότητα μπορεί να εμφανιστεί πριν από την έναρξη των συμπτωμάτων.



Σχήμα 5.11. Το πρώτο κύμα της επιδημίας SARS - CoV-2 στην Ελλάδα (Sypsa et. al, 2020)

Στο σχήμα 5.8. απεικονίζεται το πρώτο κύμα της επιδημίας SARS - CoV - 2 στην Ελλάδα (15 Φεβρουαρίου - 26 Απριλίου 2020). Το γκρι ορθογώνιο υποδεικνύει την περίοδο περιορισμών όλων των άσκοπων μετακινήσεων στη χώρα (*lockdown*). Στο πρώτο γράφημα του σχήματος 4.8 παρατηρούμε την πορεία του  $R_t$  με την πάροδο του χρόνου. Το  $R_0$  υπολογίστηκε 2.38 (95% CI: 2.01, 2.80) για τις 18 Φεβρουαρίου 2020. Η αρχική πτώση προήλθε από τα πρώτα μέτρα κοινωνικής αποστασιοποίησης (κλείσιμο σχολείων, καταστημάτων, εστιατορίων κ.λπ.) και η δεύτερη πτώση από την εφαρμογή περιορισμών όλων των άσκοπων μετακινήσεων σε όλη τη χώρα (*lockdown*). Η εκτιμώμενη  $R_t$  κατά τη διάρκεια του *lockdown* είναι 0,46 (95% CrI: 0.35, 0.57).

## 5.6 Σύγκριση Μοντέλων *SIR* & *SEIR*

Τα μοντέλα *SIR* αποτελούν τα βασικότερα μοντέλα επιδημιολογίας και η χρήση τους, λόγω της ευκολίας της εφαρμογής τους είναι ευρέως διαδεδομένη. Ωστόσο, κατά την μελέτη ενός εν δυνάμει επιδημικού κύματος, οι ερευνητές θα πρέπει να λαμβάνουν υπόψιν τα ξεχωριστά βιολογικά χαρακτηριστικά του κάθε ιού. Ένα από τα μοντέλα – τροποποίηση που έχουν προταθεί είναι τα μοντέλα *SEIR*. Η διαφοροποίηση των δύο μοντέλων έγκειται στο γεγονός πως το μοντέλο *SEIR* θεωρεί πως ένα άτομο πριν νοσήσει, διανύει μια περίοδο κατά την οποία ενώ έχει έρθει σε επαφή με τον ιό, δεν νοσεί ακόμη, αλλά παράλληλα είναι σε θέση κι αυτός με την σειρά του να μεταδώσει τον ιό, πριν εκδηλώσει τα συμπτώματα της νόσου. Για την εφαρμογή των διαθέσιμων δεδομένων *Covid – 19* που έχουμε στην διάθεσή μας για τη περίοδο από 26 Φεβρουαρίου 2020 έως 26 Μαρτίου 2020, εφαρμόσαμε το μοντέλο *SIR* χρησιμοποιώντας το πακέτο *covid19.analytics* και συγκεκριμένα τη συνάρτηση *generate.SIR.model*. Παρατηρήσαμε ότι δεν προκύπτουν μεγάλες διαφορές στην εκτίμηση του αριθμού των νέων κρουσμάτων και του  $R_0$  σε σχέση με τις εκτιμήσεις του ενισχυμένου μοντέλου *SEIR* που χρησιμοποιήθηκε στη μελέτη που πραγματοποιήθηκε από Σύψα και συνεργάτες το 2020 (*Sypsa et. al, 2020*) για το διάστημα από την αρχή της πανδημίας στη χώρα μέχρι και τις 16 Μαρτίου. Συγκεκριμένα η εκτίμηση που δίνεται από το μοντέλο *SIR* για τον αριθμό των νέων κρουσμάτων και το  $R_0$  για τις 16 Μαρτίου είναι 411 και 1.25 αντίστοιχα. Για το μοντέλο *SEIR* όπως παρατηρούμε από τα 2 πρώτα γραφήματα στο σχήμα 5.8 ο αριθμός των νέων κρουσμάτων και το  $R_0$  για τις 16 Μαρτίου είναι 420 και 1.4 αντίστοιχα. Συνεπώς, για την συγκεκριμένη χρονική περίοδο της μελέτης μας, τα δυο μοντέλα συγκλίνουν ως προς τα συμπεράσματά τους.





## ΒΙΒΛΙΟΓΡΑΦΙΑ

### Ξενόγλωσση

Adiga, A., Dubhashi, D., Lewis, B. (2020). Mathematical Models for COVID-19 Pandemic: A Comparative Analysis. *J Indian Inst Sci*, **100**, 793–807.

Adhikari B., Xu X., Ramakrishnan N., Prakash B.A. (2019) Epiddeep: exploiting embeddings for epidemic forecasting, Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, *Association for Computing Machinery*, 577–586.

Ahmed, Ayub, Bashdar Salam, Mahmud Mohammad, Ali Akgul, and Sarbaz H. A. Khoshnaw. (2020). Analysis coronavirus disease (COVID-19) model using numerical approaches and logistic model. *AIMS Bioengineering*, **7**, 130-146.

Allen, L.J.S. (2015). Stochastic Population and Epidemic Models: Persistence and Extinction, Springer International Publishing.

Athreya K.B., and Ney P.E. (1972). Branching Processes, Grundlehren der Mathematischen Wissenschaften, Springer, **196**.

Bertozi, A.L., Franco E., Mohler G., Short M.B., Sledge D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 16732–16738.

Brémaud, P. (2020). Discrete-Time Markov Chains. In: Markov Chains. Texts in Applied Mathematics, **31**, 63 - 109.

Chrysafinou O. (2012). An Introduction To Branching Processes, Sophia Publications, 1-410, ISBN: 9789606706639.

Ching W.K., Ng M. K., Fung E. S. (2008). Higher-order multivariate markov chains and their applications. *Linear Algebra and its Applications*, **428**, 492–507.

Dobson I., Carreras B.A., Newman D.E. (2005). Branching Process Models for the Exponentially Increasing Portions of Cascading Failure Blackouts. Thirty – eight Annual Hawaii International Conference on System Sciences, pp. 64a–64a.

Dorman K.S., Sinsheimer J.S., Lange K. (2004). In the Garden of Branching Processes. *SIAM Rev.* **46**, 202–229. <https://doi.org/10.1137/S0036144502417843>.

Fazlollahtabar H., and Niaki S.T.A. (2018). Modified branching process for the reliability analysis of complex systems: Multiple-robot systems. *Communications in Statistics - Theory and Methods* **47**, 1641–1652. <https://doi.org/10.1080/03610926.2017.1324985>.

Green M. D., Freedman D. M., Gordis L. (2007): Reference Guide on Epidemiology. Reference Manual on Scientific Evidence, **3**, Washington, D.C.: The National Academies Press.

Greenland S., and Thomas D.C. (1982): On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology*, **116**, 547-55.

Grinstead C.M., Snell J.L. (2012). Introduction to probability. American Mathematical Society.

Haccou, P., Jagers, P., and Vatutin, V.A. (2005). Branching processes: variation, growth, and extinction of populations, Cambridge, New York, Cambridge University Press.

Heesterbeek, H. Anderson, R.M. Andreasen, V. Bansal, S. De Angelis, D. et al. (2015) Modeling infectious disease dynamics in the complex landscape of global health *Science*, **347**.

Hughes-Hallett D., Gleason A.M., Lock P.F., Flath D.E., Gordon S.P., Lomen D.O., Lovelock D., McCallum W., Osgood B., G., Quinney D., Pasquale A., Rhea K., Tecosky-Feldman J., Thrash J., B., Tucker, T., W. (2002). Applied Calculus, *Wiley*, Toronto, **2**.

Ivorra, B., M.R. Ferrández, M. Vela-Pérez, and A.M. Ramos. (2020). "Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) considering its particular characteristics. The case of China." *Communications in Nonlinear Science and Numerical Simulation*.

Jacob, C. (2010). Branching Processes: Their Role in Epidemiology. *IJERPH*, **7**, 1186–1204. <https://doi.org/10.3390/ijerph7031204>.

Jagers, P. (1975). Branching processes with biological applications, London, John Wiley, New York.

Kemeny, J.G., and Snell, J.L. (1972). Mathematical models in the social sciences , Cambridge, Mass, MIT Press.

Kendall, D.G. (1966). Branching Processes Since 1873. *Journal of the London Mathematical Society*, **41**, 385–406. <https://doi.org/10.1112/jlms/s1-41.1.385>.

Kermack, W, McKendrick, A, Earn, D., 2008, Contributions to the mathematical theory of epidemics - “A Light Introduction to Modelling Recurrent Epidemics”, in Brauer, F., van den Driessche, P. & Wu, J. (eds.), *Mathematical Epidemiology*, Springer, P.K. Maini, Oxford.

Kim, J., and Dobson, I. (2010). Approximating a Loading-Dependent Cascading Failure Model With a Branching Process. *IEEE Trans. Rel.*, **59**, 691–699.

Kimmel, M., and Axelrod, D.E. (2015). Biological Background. In *Branching Processes in Biology*, New York, NY: Springer New York, 19–36.

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., Sun, F., Jit, M., Munday, J.D., et al. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases* **20**, 553–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).

Lange, K. (2010). Discrete-Time Markov Chains. In *Applied Probability*, New York, NY: Springer New York, pp. 151–185.

Macken, C.A., and Perelson, A.S. (1985). *Branching Processes Applied to Cell Surface Aggregation Phenomena* (Berlin, Heidelberg: Springer Berlin Heidelberg).

Ma, Z., Zhou, Y. & Wu, J. (eds.), *Modeling and Dynamics of Infectious Diseases*, Higher Education Press & World Scientific, P. R. China.

Marathe M, Vullikanti A (2013) Computational epidemiology. *Commun ACM*, **56**, 88–96.

Merrill, R. M. (2010): *Introduction to Epidemiology* (Fifth Edition). London: Jones & Bartlett.

Mitrofani I. A. (2021). *Applications on Branching Processes*. (Master Thesis, University of the Aegean).

National Public Health Organisation. Epidemiological surveillance of COVID-19 - Daily report. 26 April 2020.

Ndaïrou, Faïçal, Iván Area, Juan J. Nieto, and Delfim F.M. Torres. (2020). "Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan." *Chaos, Solitons and Fractals* ((135)).

Ourworldindata.org <https://ourworldindata.org/coronavirus#explore-the-global-situation>

Pazsit, I., Lenard Pal, TotalBoox, and TBX (2007). Neutron Fluctuations. (Elsevier Science).

Perreira-Maxwell, 1998, AZ of Medical Statistics.

Ponce, M. (2021), “covid19.analytics: An R Package to Obtain, Analyze and Visualize Data from the Coronavirus Disease Pandemic”. University of Toronto.

Privault, N. (2018). *Discrete-Time Markov Chains*. In: *Understanding Markov Chains*. Springer Undergraduate Mathematics Series. Springer, Singapore.

Schinazi, R.B. (1999). Discrete Time Markov Chains. In *Classical and Spatial Stochastic Processes*, (Boston, MA: Birkhäuser Boston), 1–41.

Sypsa, V., Sotirios Roussos, Dimitrios Paraskevis, Theodore Lytras, S Sotirios Tsiodras, Angelos Hatzakis. (2020). «Modelling the SARS-CoV-2 first epidemic wave in Greece: social contact patterns for impact assessment and an exit strategy from social distancing measures»

<https://www.medrxiv.org/content/10.1101/2020.05.27.20114017v2.full>

Wang, Jin . (2020). "Mathematical models for COVID-19: applications, limitations, and potentials." *J Public Health Emerg*, **4**.

Watmough, J., 2009, “The Basic Reproduction Number and the Final Size of an Epidemic”, *Modelling and Dynamics of Infectious Diseases*, 237 – 252.

## Ελληνική

Δημητράκος Θ.(2022). Πανεπιστημιακές σημειώσεις *Στοχαστικές ανελίξεις*, Σάμος, Πανεπιστήμιο Αιγαίου.

Κοκολάκης Γ. (2007) .Πανεπιστημιακές σημειώσεις *Στοχαστικές ανελίξεις*, Αθήνα, Εθνικό Μετσόβιο Πολυτεχνείο.

Μπούτσικας, Μ. (2020). Πανεπιστημιακές σημειώσεις *Εισαγωγή στις στοχαστικές ανελίξεις*, Πειραιάς, Πανεπιστήμιο Πειραιώς.

Ντζούφρας, Ι. (2009): *Εισαγωγή στη Βιοστατιστική και την Επιδημιολογία*, Αθήνα, Οικονομικό Πανεπιστήμιο Αθηνών.

Ξεκαλάκη, Ε. (2001) Πανεπιστημιακές σημειώσεις *Βιοστατιστική και Επιδημιολογία*, Αθήνα, Οικονομικό Πανεπιστήμιο Αθηνών.

Πολίτης Κ. (2015) .Πανεπιστημιακές σημειώσεις *Θεωρία χρεοκοπίας*, Πειραιάς: Πανεπιστήμιο Πειραιώς.

Πολίτης Κ. (2012). *Εισαγωγή στη θεωρία συλλογικού κινδύνου*, Εκδόσεις Σταμούλη, Αθήνα.

Πρίσκα Ε.(2020) Πανεπιστημιακές σημειώσεις *Επιδημιολογία*, Αθήνα, Πανεπιστήμιο Δυτικής Αττικής

Σύψα Β. (2016) Πανεπιστημιακές σημειώσεις *Στατιστική και την Επιδημιολογία*, Αθήνα, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

Τζαβελάς Γ. (2020) Πανεπιστημιακές σημειώσεις *Βιοστατιστική και Επιδημιολογία*, ΠΜΣ στην *Εφαρμοσμένη Στατιστική* Πειραιάς: Πανεπιστήμιο Πειραιώς.

Χρυσ αφίνου Ο. (2012). *Εισαγωγή στις στοχαστικές ανελίξεις*, Εκδόσεις Σοφία,Θεσσαλονίκη.