ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ


Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
URL: http://www.semfe.ntua.gr


# ΜΑΡΚΟΒΙΑΝΕΣ ΑΛΥΣΙΔΕΣ ΑΠΟΦΑΣΕΩΝ ΚΑΙ ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ


Αγγλιά Κατερίνα,09115091

Επιβλέπων: Ιωάννης Κολέτσος, Αναπληρωτής Καθηγητής


Η τριμελής εξεταστική επιτροπή

| | | |
|---|---|---|
| Ιωάννης Κολέτσος | Κοκκίνης Βασίλειος | Στεφανέας Πέτρος |
| Αναπλ.καθηγητής | Αναπλ.καθηγητής | Αναπλ.καθηγητής |

**Αθήνα, Φεβρουάριος 2023**

*Συγγραφέας*:

Κατερίνα Αγγλιά

*Επιβλέπων*:

Κολέτσος Ιωάννης,
Αναπληρωτής Καθηγητής Ε.Μ.Π

16 Φεβρουαρίου 2023

Ακαδημαϊκή χρονιά 2022/2023

**ΕΥΧΑΡΙΣΤΙΕΣ**

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της διπλωματικής μου εργασίας κ.Ιωάννη Κολέτσο για την σημαντική βοήθεια του στην συγγραφή αυτού του έργου.Επίσης,θα ήθελα να ευχαριστήσω όλους τους καθηγητές της σχολής εφαρμοσμένων μαθηματκών και φυσικών επιστημών για την πολύτιμη συμβολή τους στην εκπαιδευτική μου πορεία εως σήμερα.Τέλος, ευχαριστώ την οικογένεια και τους δικούς μου ανθρώπους για τη συμπαραστασή τους.

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ACKNOWLEDGEMENTS**

I would like to sincerely thank my thesis supervisor Mr. Ioanni Koletsos for his important help in writing this project. I would also like to thank all the professors of the Faculty of Applied Mathematics for their valuable contribution in my educational path until today and finally the family and my own my people for their support.

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**Περίληψη**

Η λήψη αποφάσεων αποτελεί μία διαδικασία έντονης σημασίας στην επίλυση προβλημάτων που συναντώνται σε ποικίλους τομείς.Στο παρόν έργο, παρουσιάζονται επιστημονικοί κλάδοι και μέθοδοι μεγάλης χρησιμότητας οι οποίοι οργανώνουν και μελετούν τον τρόπο που λαμβάνουμε αποφάσεις κατά τη λειτουργεία συγκεκριμένων συστημάτων .

Στο πρώτο κεφάλαιο, παρουσιάζεται το αντικείμενο της επιχειρησιακής έρευνας,γίνεται μία ιστορική αναδρομή και παρουσιάζονται οι βασικές τεχνικές της.Στο δεύτερο κεφάλαιο, γίνεται μία εισαγωγή στην θεωρία αποφάσεων και σε βασικές της μεθόδους ενώ παρουσιάζονται τα δένδρα αποφάσεων.Στο τρίτο κεφάλαιο,εισάγουμε το μαθηματικό αντικείμενο των στοχαστικών διαδικασιών και μιας ειδικής κατηγορίας αυτών,των μαρκοβιανών αλυσίδων και ερχόμαστε αντιμέτωποι με το ερώτημα του πως μπορούμε να σχεδιάσουμε την λειτουργία τους .Πλέον, κάνουμε λόγο για μαρκοβιανές αλυσίδες αποφάσεων όπου η μετάβαση απο μία κατάσταση με μία άλλη επιφέρει κάποιο κόστος( ή κέρδος )το οποίο εξαρτάται από την απόφαση που πάρθηκε.Στόχος μας είναι η ελαχιστοποίηση(η αντίστοιχα μεγιστοποίηση) του μέσου κόστους ανά μονάδα χρόνου.Οι μέθοδοι με τους οποίους προσεγγίζουμε το παραπάνω πρόβλημα είναι ο γραμμικός προγραμματισμός,ο αλγόριθμος βελτίωσης της πολιτικής και η μέθοδος των διαδοχικών προσεγγίσεων.Σημαντική παρατήρηση σε αυτό το σημείο είναι πως οι πολιτικές των συστημάτων μας είναι στάσιμες δηλαδή δεν εξαρτώνται απο την τιμή του χρόνου.

Λέξεις κλειδιά: επιχειρησιακή έρευνα,θεωρία αποφάσεων,στοχαστικές διαδικασίες ,μαρκοβιανά συστήματα,μαρκοβιανές αλυσίδες αποφάσεων

**Abstract**

Decision-making is a process of intense importance in solving problems encountered in various fields. In this thesis, scientific disciplines and methods of great utility are presented which organize and study the way we make decisions during the operation of specific systems.

In the first chapter, the object of operational research is presented, a historical review is made and its basic techniques are presented. In the second chapter, there is an introduction to decision theory and its basic methods, while decision trees are presented.In the third chapter, we introduce the mathematical subject of stochastic processes and a special category of them, Markovian chains, and we are faced with the question of how we can design their operation. Now, we are discussing about Markovian decision chains where the transition from one state to another entails some cost (or profit) which depends on the decision made. Our goal is to minimize (respectively maximize) the mean cost per unit time.The methods with which we approach the above problem are linear programming, the policy improvement algorithm and the method of successive approximations. An important observation at this point is that the policies of our systems are stationary, that is, they do not depend on the value of time.

Key words : operational research, decision theory, stochastic processes, Markovian systems, Markovian decision processes

# Contents

# 1 Επειχηριασιακή έρευνα

## 1.1 Εισαγωγή στην επιχειρησιακή έρευνα

Έχετε σκεφτεί ποτέ πως θα μπορούσαν να αντιμετωπιστούν προβλήματα της καθημερινότητας σας όπως ποιες είναι οι βέλτιστες αγορές ρούχων και επίπλων,ποια είναι η επιλογή της κατάλληλης λωρίδας ταμείου σε ένα μαγαζί ή η λιγότερο χρονοβόρα διαδρομή για τη δουλειά;Σίγουρα η απάντηση δεν είναι μονοδιάσταστη αλλά περιέχει την έννοια της επιχειρησιακής έρευνας.Η ΕΕ (συχνά αναφέρεται ως επιστήμη διαχείρισης) είναι μια επιστημονική προσέγγιση στη λήψη αποφάσεων που επιδιώκει να σχεδιάσει και να λειτουργήσει καλύτερα ένα σύστημα, συνήθως υπό συνθήκες που απαιτούν την κατανομή πόρων.Τα προαναφερθέντα προβλήματα μπορούν να διαχειρστούν μέσω βασικών μεθόδων της όπως η λήψη αποφάσεων,η θεωρία ουρών και η θεωρία δικτύων που θα αναλυθούν στη συνέχεια.

Βέβαια, η επιχιερησιακή έρευνα δεν μένει στα καθημερινά προβλήματα αλλά αναλαμβάνει ζητήματα μεγαλύτερων διαστάσεων σε τομείς επιχειρήσεων και βιομηχανιών και άλλων επιστημονικών κλάδων.

## 1.2 Ορισμός

**Ορισμός Ackoff και Sasienni**

Επιχειρησιακή Έρευνα μπορεί να θεωρηθεί ότι είναι: η εφαρμογή επιστημονικών μεθόδων από μικτές ομάδες σε προβλήματα που αφορούν τον έλεγχο οργανωμένων συστημάτων (αποτελούμενων από ανθρώπους και μηχανές) κατά τρόπο ώστε να παρέχουν λύσεις που εξυπηρετούν κατά τον καλύτερο δυνατό τρόπο τους σκοπούς του οργανισμού ως συνόλου

**Ορισμός R. Watson-Watt, A.P. Rowe**

"Η Επιχειρησιακή Έρευνα αποσκοπεί στο να ερευνήσει ποσοτικά εάν ένας οργανισμός παίρνει από τη λειτουργία του εξοπλισμού του τη βέλτιστη δυνατή συνεισφορά σε σχέση με τον ολικό αντικειμενικό σκοπό του, ποιες αλλαγές σε εξοπλισμό και μεθόδους απαιτούνται για τη βελτίωση των αποτελεσμάτων με το μικρότερο δυνατό κόστος σε προσπάθεια και χρόνο και τέλος σε ποιο βαθμό μεταβολές στους επιμέρους αντικειμενικούς σκοπούς (τακτικοί αντικειμενικοί σκοποί) θα συνεισέφεραν στην πιο οικονομική και έγκαιρη εκτέλεση του ολικού στρατηγικού αντικειμενικού σκοπού".

**Ορισμός Ελληνικής Εταιρίας Επιχειρησιακών Ερευνών**

Επιχειρησιακή Έρευνα είναι η επιστημονική προετοιμασία των αποφάσεων της Διοικήσεως με την επιστημονική ανάλυση των δεδομένων και τη δημιουργία μαθηματικών προτύπων.

## 1.3 Ορολογία

- OR: Operational Research (Ευρωπαϊκή ονομασία)
- OR: Operations Research (Αμερικάνικη ονομασία)
- MS: Management Science (εναλλακτική ονομασία)
- OR/MS: Operations Research and Management Science
- IE: Industrial Engineering (σπανιότερα χρησιμοποιούμενο )
- DS: Decision Science (σπανιότερα χρησιμοποιούμενο)
- EE: Επιχειρησιακή Έρευνα (Ελληνική ονομασία)

## 1.4 Η προέλευση της ΕΕ και ιστορική αναδρομή

Από την έλευση της βιομηχανικής επανάστασης ο κόσμος έχει έρθει σε επαφή με μία αξιοσημείωτη ανάπτυξη των οργανισμών τόσο όσον αφορά το μέγεθος όσο και την πολυπλοκότητα τους.Πιο συγκεκριμένα,η τεχνολογική ανάπτυξη έχει επειφέρει αύξηση του καταμερισμού της εργασίας και των αρμοδιοτήτων διαχείρησης τα οποία εν γένει εμφάνισαν θεαματικά αποτελέσματα.Ωστόσο,η προαναφερθείσα εξειδίκευση έχει δημιουργήσει και κάποια προβλήματα,προβλήματα που σχετίζονται με την ανεξάρτητη εξέλιξη των διάφορων τμημάτων των οργανισμών με ένα δικό τους σύστημα αξιών και στόχων τα οποία σε κάποιες περιπτώσεις είναι δύσκολο να εξισορροπηθούν με την ταυτόχρονη εξέλιξη του οργανισμού ως σύνολο.Για περάδειγμα ,γίνεται όλο και πιο δύσκολο να κατανεμηθούν οι οικονομικοί πόροι στις διάφορες δραστηριότητες.Τέτοιου είδους προβλήματα ήταν αυτά που οδήγησαν στην εμφάνιση και άνθηση της επιχειρησιακής έρευνας.

Ιστορικά,η πρώτη κοινά αποδεκτή εφαρμογή της εμφανίζεται στις στρατιωτικές επιχειρήσης στις αρχές του Β΄ παγκοσμιού πολέμου .Η βρετανική και στη συνέχεια η στρατιωτική διοίκηση των ΗΠΑ κάλεσαν ένα μεγάλο αριθμό επιστημόνων με στόχο την αντιμετώπιση διαφόρων προβλημάτων τακτικής με συνυπάρχουσο πρόβλημα την όσο δυνατόν βέλτιστη αξιοποίηση των λιγοστών πόρων λόγω πολέμου.

Αλλά, η χρήση της δεν παρέμεινε σε αυτό το σημειό . Μετά την λήξη του πολέμου και στις αρχές του 1950 οι σύμβουλοι των επιχειρήσεων που είχαν υπηρετήσει είτε στις είτε με τις ομάδες της ΕΕ στον πόλεμο διαπιστώνοντας την εφαρμογή τους στο πλαίσιο των επιχειρήσεων την εισήγαναν σε διάφορους τομείς όπως στις βιομηχανίες και σε κυβερνητικούς οργανισμούς.Μάλιστα,πολλά απο τα τυπικά εργαλεία της ΕΕ όπως ο γραμμικός προγραμματισμός,ο δυναμικός προγραμματισμός ,η θεωρία ουρών και η θεωρία αποθεμάτων ήταν ήδη καλά ανεπτυγμένα πρίν απο τα τέλη της δεκαετίας του 1950.

### 1.4.1 Τα πρώτα προβλήματα που επιλύθηκαν αποτελεσματικά

**1941** Δημιουργήθηκε τομέας επιχειρησιακής έρευνας με στόχο τον εντοπισμό και εξολόθρευση γερμανικών υποβρυχίων. Έπρεπε να λυθούν προβλήματα όπως

- Οργάνωση της συντήρησης και επιθεώρηση
- Σύγκριση τύπων αεροσκαφών
- Βελτίωση της πιθανότητας επιτυχίας της επίθεσης

**Οργάνωση της συντήρησης και επιθεώρησης**
**Δεδομένα**: Κάθε 350 ώρες πτήσης, ένα αεροσκάφος απαιτεί 7 μικρές επιθεωρήσεις (διάρκειας 2-5 ημερών) και μία μεγάλη (διάρκειας 15 ημερών)
**Λύση**: Κάθε συνεργείο μπορεί να εξυπηρετεί όλα τα αεροσκάφη(σχέση 1 προς Ν)
**Αποτελέσματα**: Αύξηση των ωρών πτήσεων κατά 61%

**Βελτίωση της Πιθανότητας Επιτυχίας της Επίθεσης**
**Δεδομένα:** – Απαιτούνται 170 ανθρωποώρες συντήρησης για 1 ώρα πτήσης
– Απαιτούνται 200 ώρες πτήσης για 1 επίθεση σε υποβρύχιο
– Πιθανότητα επιτυχίας (1941): 2-3%(1.1-1.7 εκ. ανθρωποώρες)
– Επιτυχία δεν σημαίνει και βύθιση του υποβρυχίου!
– Η χρήση των πυρομαχικών είχε σημαντικό ρόλο και ενέπλεκε επιπλέον παραμέτρους, όπως:
• ακτίνα δράσης, λάθη σκόπευσης, προσανατολισμός βόμβας, ύψος απελευθέρωσης της βόμβας.
**Αποτελέσματα**: Το 1945 η πιθανότητα επιτυχίας της επίθεσης αυξήθηκε στο 40%(από 3%max το 1941)
**Η βελτίωση που πραγματοποιήθηκε**: – Απαιτούνται 170 ανθρωποώρες συντήρησης για 1 ώρα πτήσης
– Απαιτούνται 200 ώρες πτήσης για 1 επίθεση σε υποβρύχιο
– Πιθανότητα επιτυχίας (1945): 40%
– Απαιτούμενες ανθρωποώρες: $170 \times 200/0.40 = 85,000$
– Εξοικονόμηση ανθρωποωρών: $1,100,000 - 85,000 = 1,015,000$
– Με κόστος 10 €ανά ανθρωποώρα, εξοικονομήθηκαν $10,150,000$€

## 1.5 1.4.1 Τεχνικές επιχειρησισκής έρευνας

**1.Γραμμικός προγραμματισμός(Linear Programming)**

Ένα πρόβλημα γραμμικού προγραμματισμού(LP)είναι ένα πρόβλημα βελτιστοποίησης για το οποίο κάνουμε τα εξής:

- Προσπαθούμε να μεγιστοποιήσουμε (ή να ελαχιστοποιήσουμε)μια γραμμική συνάρτηση η οποία έχει κατασκευαστεί υπολογίζοντας τις μεταβλητές και τους περιορισμούς του προβλήματος.Η συνάρτηση αυτή ονομάζεται αντικειμενική συνάρτηση.

- Οι τιμές των μεταβλητών απόφασης πρέπει να ικανοποιούν ενα σύνολο περιορισμών και κάθε περιορισμός πρέπει να είναι μία γραμμική εξίσωση ή ανίσωση.

- Κάθε μεταβλητή σχετίζεται με έναν περιορισμό προσήμου.Για οποιαδήποτε μεταβλητή $x_i$ ο περιορισμός προσήμου προσδιορίζει ότι η μεταβλητή θα πρέπει είτε να είναι μη αρνητική ή να μην υπάρχει περιορισμός προσήμου.

**2.Μη γραμμικός προγραμματισμός(Nonlinear Programming)**

Όπως προαναφέρθηκε στον γραμμικό προγραμματισμό θελουμε να μεγιστοποιήσουμε(η ελαχιστοποιήσουμε)μια γραμμική συνάρτηση που υπόκειται σε γραμμικούς περιορισμούς.Αλλά, σε πολλά ενδιαφέροντα προβλήματα η αντικειμενική συνάρτηση μπορεί να μην είναι γραμμική ή ορισμένοι από τους περιορισμούς μπορεί να μην είναι γραμμικοί.Υπάρχουν συγκεκριμένες τεχνικές για την επίλυση τέτοιων μη γραμμικών προβλημάτων προγραμματισμού.

**3.Θεωρία παιγνίων(Game Theory):** Η θεωρία παιγνίων είναι μια χρήσιμη στρατηγική για την λήψη αποφάσεων σε περιπτώσεις όπου δυο ή περισσότεροι φορείς έχουν αντικρουόμενα συμφέροντα.Συχνότερα μελετώνται προβλήματα στα οποία υπαρχουν δυο υπεύθυνοι λήψης αποφάσεων αλλά εν γένει μελετάμε το παιχνίδι ν-παικτών.

**4.Θεωρία Αποφάσεων(Decision Analysis):**

Βασικό στοιχείο της επιχειρησιακής έρευνας πάνω στην οποία βασίζεται η δομή της είναι η λήψη αποφάσεων. Με τον όρο απόφαση εννοείται το αποτέλεσμα μιας διαδικασίας, όπου ο λήπτης της απόφασης εξετάζει ένα εύρος λύσεων για ένα πρόβλημα και επιλέγει αυτήν την λύση που είναι βέλτιστη.Η λήψη της απόφασης γίνεται με βάση τους στόχους που εξετάζονται στο συγκεκριμένο πρόβλημα και των εναλλακτικών λύσεων που προσφέρονται και πραγματοποιείται με τη χρήση δεδομένων κριτηρίων και περιορισμών που ορίζονται σε κάθε περίπτωση.

**5.Δυναμικός προγραμματισμός (Dynamic programming)**:

Ο Δυναμικός Προγραμματισμός είναι γενικά μία μέθοδος που λύνει προβλήματα βελτιστοποίησης, τα οποία συμμετέχουν στη λήψη μιας ακολουθίας αποφάσεων, προσδιορίζοντας, για κάθε απόφαση, υπο-προβλήματα που μπορούν να λυθούν με τέτοιο τρόπο, ώστε μία βέλτιστη λύση του πραγματικού προβλήματος να μπορεί να βρεθεί από βέλτιστες λύσεις των υπο-προβλημάτων

**6.Ουρές αναμονής(Queuing Theory):**

Στην καθημερινότητα μας συναντάμε ουρές αναμονής σε διάφορες καταστάσεις .Για παράδειγμα,οχήματα που πρέπει να περιμένουν για να εκφορτωθούν μπορεί να καθυστερήσουν επόμενες αποστολές,αεροπλάνα που περιμένουν να απογειωθούν ή να προσγειωθούν ενδέχεται να διαταράξουν τα μεταγενέστερα δρομολόγια.Η θεωρία ουρών λοιπόν, είναι η μελέτη της αναμονής σε όλες αυτές τις διάφορες μορφές.Οι τεχνικές για κάθε μοντέλο υποδεικνύουν πως αυτό θα λειτουργεί βέλτιστα.

**7.Μοντέλα αποθέματος(Inventory Theory)**:

Το μοντέλο αποθεμάτων λαμβάνει αποφάσεις που ελαχιστοποιούν το συνολικό κόστος αποθεμάτων. Το μοντέλο αυτό μειώνει με επιτυχία το συνολικό κόστος αγοράς, μεταφοράς και αποθεματοποίησης

**8.Προσομοίωση(simulation):**

Η προσομοίωση μπορεί να οριστεί ως μια τεχνική που μιμείται τη λειτουργία ενός συστήματος του πραγματικού κόσμου καθώς εξελίσσεται με την πάροδο του χρόνου. Αυτό γίνεται συνήθως με την ανάπτυξη ενός μοντέλου προσομοίωσης. Ένα μοντέλο προσομοίωσης συνήθως έχει τη μορφή ενός συνόλου υποθέσεων σχετικά με τη λειτουργία του συστήματος.Οι υποθέσεις αυτές ,εκφράζονται ως μαθηματικές ή λογικές σχέσεις μεταξύ των αντικειμένων ενδιαφέροντος στο σύστημα. Σε αντίθεση με τις ακριβείς μαθηματικές λύσεις που είναι διαθέσιμες στα περισσότερα αναλυτικά μοντέλα, η διαδικασία προσομοίωσης περιλαμβάνει την εκτέλεση ή την εκτέλεση του μοντέλου μέσα στο χρόνο, συνήθως σε υπολογιστή, για τη δημιουργία αντιπροσωπευτικών δειγμάτων των μέτρων απόδοσης

# 2 Θεωρία αποφάσεων

Πολλά εργαλεία της επιχειρησιακής έρευνας, όπως ο γραμμικός προγραμματισμός και ο ακέραιος προγραμματισμός, έχουν σχεδιαστεί για την λήψη αποφάσεων σε περιβάλλοντα όπου οι επιπτώσεις των διαφόρων εναλλακτικών αποφάσεων είναι γνωστές σε έναν εύλογο βαθμό βεβαιότητας. Πως μπορούμε όμως να διαχειριστούμε καταστάσεις οι οποίες χαρακτηρίζονται απο αβεβαιότητα; Ας πάρουμε για παράδειγμα έναν κατασκευαστή ο οποίος εισάγει στην αγορά ένα νέο προϊόν. Ποια θα είναι η αντίδραση των πελατών; Ποια η ποσότητα που θα πρέπει να παραχθεί; Το προϊόν θα πρέπει να δοκιμαστεί σε μία μικρότερη αγορά πριν ξεκινήσει η πλήρης διανομή του; Αυτά είναι κάποια ενδεικτικά ερωτήματα τα οποία χαρακτηρίζονται από κάποιο βαθμό αβεβαιότητας και μπορούν να προσεγγιστούν από ένα ιδιαίτερα χρήσιμο εργαλείο της επιχειρισιακής έρευνας, την θεωρία αποφάσεων.

Η θεωρία αποφάσεων χωρίζεται σε δύο κατηγορίες. Στην λήψη αποφάσεων χωρίς πειραματισμό (decision making without experimentation) και στην λήψη αποφάσεων με πειραματισμό (decision making with experimentation).Η δεύτερη κατηγορία αναφέρεται σε καταστάσεις, στην οποία η λήψη αποφάσεων δεν λαμβάνεται επιτόπου, αλλά διενεργούνται πριν την τελική απόφαση κάποιοι πειραματισμοί, οι οποίοι περιορίζουν τον βαθμό αβεβαιότητας της τελικής απόφασης, όμως επιφέρουν και κάποια κόστη.

## 2.1 Λήψη απόφασεων χωρίς πειραματισμό

**Πλαίσιο της θεωρίας αποφάσεων χωρίς πειραματισμό**

Σε γενικές γραμμές, ο υπεύθυνος λήψης των αποφάσεων θα πρέπει να επιλέξει μία ενέργεια από ένα σύνολο δυνατών **ενεργειών**. Το σύνολο αυτό περιέχει όλλες τις εφικτές και προς εξέταση εναλλακτικές λύσεις του προβλήματος. Βέβαια, ο υπεύθυνος επιλέγει μια ενέργεια η οποία χαρακτηρίζεται από το στοιχείο της αβεβαιότητας. Υπάρχουν τυχαίοι παράγοντες οι οποίοι την επηρεάζουν και δεν μπορούν να ελεγχθούν από τον υπεύθυνο αποφάσεων. Αυτοί οι τυχαίοι παράγοντες καθορίζουν την κατάσταση του συστήματος την στιγμή που εκτελείται η ενέργεια που αποφασίστηκε. Κάθε μία από αυτες τις δυνατές καταστάσεις χαρακτηρίζεται ως **κατάσταση της φύσης (state of nature)**.

Για κάθε συνδυασμό μιας ενέργειας και μίας κατάστασης, ο διαχειριστής των αποφάσεων (agent) γνωρίζει το αναμενόμενο κέρδος ή ζημία **(payoff)**, το οποίο αποτελεί ένα ποσοτικό μέτρο της αξίας της λαμβάνουσας απόφασης. Συνήθως χρησιμοποιείται για αυτόν τον σκοπό ένας πίνακας (payoff table).

Μία σημαντική παρατήρηση σε αυτό το σημείο είναι πως κατά την διαδικασία λήψης των αποφάσεων διατίθονται δεδομένα τα οποία βοηθούν στην εκτίμηση της σχετικής πιθανότητας των πιθανών καταστάσεων της φύσης (συνήθως μέσω κατανομής πιθανότητας). Έτσι, αν θεωρήσουμε την κατάσταση της φύσης ως μία τυχαία μεταβλητή, τότε η κατανομή πιθανότητας αναφέρεται ως prior κατανομή και οι αντίστοιχες πιθανότητες των καταστάσεων prior πιθανότητες.

**Παράδειγμα**

Η GOFERBROKE COMPANY κατέχει μια έκταση γης που μπορεί να περιέχει πετρέλαιο. Ένας γεωλόγος ανέφερε στη διοίκηση ότι πιστεύει ότι υπάρχει 1 πιθανότητα πετρελαίου στις 4. Λόγω αυτής της προοπτικής, μια άλλη εταιρεία πετρελαίου προσφέρθηκε να αγοράσει τη γη για 90.000\$. Ωστόσο, η Goferbroke εξετάζει το ενδεχόμενο να κρατήσει τη γη για να εξορύξει μόνη της για πετρέλαιο. Το κόστος της γεώτρησης είναι 100.000\$. Εάν βρεθεί πετρέλαιο, τα αναμενόμενα έσοδα θα προκύψουν είναι 800.000\$, επομένως το αναμενόμενο κέρδος της εταιρείας (μετά την αφαίρεση του κόστους γεώτρησης) θα είναι 700.000 \$. Θα προκύψει απώλεια 100.000\$ (το κόστος γεώτρησης) εάν η γη δεν έχει πετρέλαιο.Ωστόσο, πριν αποφασισθεί εάν θα πραγματοποιήθεί γεώτρηση ή πώληση, μια άλλη επιλογή είναι να πραγματοποιήθεί μια λεπτομερής σεισμική έρευνα της γης για να ληφθεί μια καλύτερη εκτίμηση της πιθανότητας εύρεσης πετρελαίου.Το κόστος της έρευνας είναι 30.000\$.

### Λύση με το κριτήριο μέγιστης/ελάχιστης αποπληρωμής

Η λογική πίσω από το συγκεκριμένο κριτήριο είναι πως δίνεται το πιο πιθανό όφελος που προκύπτει από την επιλεγμένη απόφαση. Παρόλα αυτά, επειδή επικεντρώνεται στη χείριστη περίπτωση που μπορεί να συμβεί θεωρείται ως απαισιόδοξο σενάριο και δεν προτιμάται.

> **Κριτήριο μέγιστης/ελάχιστης αποπληρωμής(maximin payoff criterion**:
> Για κάθε πιθανή ενέργεια, βρείτε την ελάχιστη απόδοση σε όλες τις πιθανές καταστάσεις της φύσης. Στη συνέχεια, βρείτε το μέγιστο από αυτές τις ελάχιστες αποδόσεις. Επιλέξτε την ενέργεια της οποίας η ελάχιστη απόδοση δίνει αυτό το μέγιστο

Ξεκινώντας,ελέγχουμε την πρώτη γραμμή του πίνακα για την ενέργεια εξόρυξη (drill)και επιλέγουμε το ελάχιστο στοιχείο της.Συνεχίζουμε την ίδια διαδικασία για την δέυτερη γραμμή του πίνακα και την ενέργεια πώληση(sell). Τέλος, ανάμεσα στα δυο ελάχιστα στοιχεία επιλέγουμε το μέγιστο και κρατάμε ως βέλτιστη την ενέργεια που του αντιστοιχεί.Ο παρακάτω πίνακας παρουσιάζει την διαδικασία στο περιβάλλον του excel.

| | Maximin Payoff Criterion for the Goferbroke Co. Problem | | | | |
|---|---|---|---|---|---|
| 1 | Maximin Payoff Criterion for the Goferbroke Co. Problem | | | | |
| 2 | | | | | |
| 3 | | | State of Nature | Minimum | |
| 4 | Alternative | Oil | Dry | in Row | |
| 5 | Drill | 700 | -100 | -100 | |
| 6 | Sell | 90 | 90 | 90 | Maximin |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |

| | H | I |
|---|---|---|
| 5 | =MIN(C5:G5) | =IF(H5=MAX($H$5:$H$9),"Maximin","") |
| 6 | =MIN(C6:G6) | =IF(H6=MAX($H$5:$H$9),"Maximin","") |
| 7 | =MIN(C7:G7) | =IF(H7=MAX($H$5:$H$9),"Maximin","") |
| 8 | =MIN(C8:G8) | =IF(H8=MAX($H$5:$H$9),"Maximin","") |
| 9 | =MIN(C9:G9) | =IF(H9=MAX($H$5:$H$9),"Maximin","") |

Figure 1: maximin criterion

### Λύση με το κριτήριο της μέγιστης πιθανοφάνειας

Σύμφωνα με το κριτήριο της μέγιστης πιθανοφάνειας καθορίζεται η πιο πιθανή κατάσταση της φύσης για την οποία προσδιορίζεται η δράση εκείνη για την οποία αποδίδεται η μέγιστη αποπληρωμή.

> **Κριτήριο μέγιστης πιθανόφάνειας**:
> Προσδιορίστε την πιο πιθανή κατάσταση της φύσης (αυτή με τη μεγαλύτερη prior πιθανότητα). Για αυτή τη φυσική κατάσταση, βρείτε την ενέργεια με τη μέγιστη απόδοση. Επιλέξτε αυτήν την ενέργεια

Εφαρμόζοντας αυτό το κριτήριο στο παράδειγμα, λαμβάνουμε πως η κατάσταση χωρίς πετρέλαιο (dry)έχει μεγαλύτερη πιθανότητα ίση με 0.75. Στη στήλη χωρίς πετρέλαιο (dry), η πώληση έχει τη μέγιστη απόδοση ίση με 90, οπότε η επιλογή είναι να πουληθεί η γη.Ο παρακάτω πίνακας παρουσιάζει την διαδικασία στο περιβάλλον του excel.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Maximum Likelihood Criterion for the Goferbroke Co. Problem | | | | | | | |
| 2 | | | | | | | | |
| 3 | | | | | State of Nature | | | |
| 4 | | Alternative | Oil | Dry | | | | |
| 5 | | Drill | 700 | -100 | | | | |
| 6 | | Sell | 90 | 90 | | | | Maximum |
| 7 | | | | | | | | |
| 8 | | | | | | | | |
| 9 | | | | | | | | |
| 10 | | Prior Probability | 0.25 | 0.75 | | | | |
| 11 | | | | Maximum | | | | |

Figure 2: maximum likelihood

Με την εφαρμογή του συγκεκριμένου κριτηρίου, όμως, αγνοούνται σημαντικές πληροφορίες όπως, πως καμία φυσική κατάσταση δεν θεωρείται πιο πιθανή από κάποια άλλη. Αυτό αποτελεί μειονέκτημα ειδικά στα προβλήματα με πολλές πιθανές φυσικές καταστάσεις, όπου η πιθανότητα της βέλτιστης είναι ιδιαίτερα μικρή με αποτέλεσμα να είναι επισφαλής η εστίαση σε αυτήν.

**Ο κανόνας απόφασης του Bayes**

Το τρίτο κριτήριο μας είναι ο κανόνας απόφασης του Bayes ο οποίος είναι και ο πιο συχνός σε χρήση.

> **Κανόνας απόφασης Bayes**:
> Χρησιμοποιώντας τις διαθέσιμες εκτιμήσεις των πιθανοτήτων των αντίστοιχων φυσικών καταστάσεων , υπολογίστε την αναμενόμενη αξία της ανταμοιβής για καθεμία από τις πιθανές ενέργειες. Επιλέξτε τη δράση με τη μέγιστη αναμενόμενη απόδοση

Για το παραδειγμά μας υπολογίζουμε διαδοχικά

$$E[Payoff(driill)] = 0.25(700) + 0.75(-100) = 100 \tag{1}$$

$$E[Payoff(sell)] = 0.25(90) + 0.75(90) = 90 \tag{2}$$

από όπου προκύπτει πως η ενέργεια εξόρυξης(1) δίνει την μεγαλύτερη απόδοση και άρα επιλέγεται.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Bayes' Decision Rule for the Goferbroke Co. Problem | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | State of Nature | | | Expected | |
| 4 | | Alternative | Oil | Dry | | | | Payoff | |
| 5 | | Drill | 700 | -100 | | | | 100 | Maximum |
| 6 | | Sell | 90 | 90 | | | | 90 | |
| 7 | | | | | | | | | |
| 8 | | | | | | | | | |
| 9 | | | | | | | | | |
| 10 | | Prior Probability | 0.25 | 0.75 | | | | | |

| | H | I |
|---|---|---|
| 5 | `=SUMPRODUCT(C5:G5,C10:G10)` | `=IF(H5=MAX($H$5:$H$9),"Maximum","")` |
| 6 | `=SUMPRODUCT(C6:G6,C10:G10)` | `=IF(H6=MAX($H$5:$H$9),"Maximum","")` |
| 7 | `=SUMPRODUCT(C7:G7,C10:G10)` | `=IF(H7=MAX($H$5:$H$9),"Maximum","")` |
| 8 | `=SUMPRODUCT(C8:G8,C10:G10)` | `=IF(H8=MAX($H$5:$H$9),"Maximum","")` |
| 9 | `=SUMPRODUCT(C9:G9,C10:G10)` | `=IF(H9=MAX($H$5:$H$9),"Maximum","")` |

Figure 3: Bayes' decision rule

Ένα μεγάλο πλεονέκτημα του κανόνα του Bayes σε αντιδιαστολή με τα προηγούμενα κριτήρια είναι πως αυτός συνδυάζει όλες τις πληροφορίες που παρέχονται συμπεριλαμβανομένων των αποδόσεων της κάθε ενέργειας(payoff) αλλά και των αντίστοιχων πιθανοτήτων(prior).

Πολλοί θεωρούν πως αυτές οι εκτιμήσεις των πιθανοτήτων περιέχουν έναν μεγάλο βαθμό υποκειμενικότητας και θα πρέπει να αποφεύγονται καθώς δεν υπάρχει ακριβής τρόπος πρόβλεψης του μέλλοντος. Παρολαυτά, ο ισχυρισμός μας είναι πως σε πολλές περιπτώσεις η αποκτηθείσα εμπειρία καθώς και τα συλλεχθέντα δεδομένα μας επιτρέπουν να δειεξάγουμε εύλογες εκτιμήσεις των πιθανοτήτων.Η παραπάνω παρατήρηση περί υποκειμενικότητας θα πρέπει να αντιμετωπιστεί διενεργόντας περισσότερη έρευνα(experimentation).Συχνά, για αυτόν τον σκοπό χρησιμοποιείται η ανάλυση ευασισθησίας(decision analysis).

**Ανάλυση ευαισθησίας με το κριτήριο του Bayes**

Στο παράδειγμα μας,θα εφαρμόσουμε την ανάλυση ευαισθησίας για τα δεδομένα που αφορούν τις πιθανότητες των καταστάσεων της φύσης ενώ εν γένει θα μπορούσε να εφαρμοστεί και στις αποδόσεις των ενεργειών(payoffs) Η GOFEBROKE COMPANY θεωρεί πως οι πραγματικές τιμές πιθανότητας που αντιστοιχούν στη κατάσταση ύπαρξη πετρελαίου βρίσκονται σε ένα εύρος απο 0.15 έως 0.35.Έτσι,εφόσον οι πιθανότητες των δύο ενδεχομένων αθροίζουν στο ένα η αντίστοιχη πιθανότητα μη ύπαρξης πετρελαίου βρίσκεται στο εύρος απο 0.85 έως 0.65.

Η ανάλυση ευαισθησίας ξεκινά εφαρμόζοντας το κριτήριο του Bayes δύο φορές μία όταν η πιθανότητα για πετρέλαιο βρίσκεται στο κάτω άκρο(0.15) και μία οταν βρίσκεται στο άνω άκρο(0.35).Η ανάλυση τότε δείχνει (Figure 4 )πως οι διαφορές είναι μεγάλης κλίμακας και θα πρέπει να διενεργηθεί περαιτέρω πειραματισμός.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Bayes' Decision Rule for the Goferbroke Co. Problem | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | State of Nature | | | Expected | |
| 4 | | Alternative | Oil | Dry | | | | Payoff | |
| 5 | | Drill | 700 | -100 | | | | 20 | |
| 6 | | Sell | 90 | 90 | | | | 90 | Maximum |
| 7 | | | | | | | | | |
| 8 | | | | | | | | | |
| 9 | | | | | | | | | |
| 10 | | Prior Probability | 0.15 | 0.85 | | | | | |

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Bayes' Decision Rule for the Goferbroke Co. Problem | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | State of Nature | | | Expected | |
| 4 | | Alternative | Oil | Dry | | | | Payoff | |
| 5 | | Drill | 700 | -100 | | | | 180 | Maximum |
| 6 | | Sell | 90 | 90 | | | | 90 | |
| 7 | | | | | | | | | |
| 8 | | | | | | | | | |
| 9 | | | | | | | | | |
| 10 | | Prior Probability | 0.35 | 0.65 | | | | | |

Figure 4: sensitivity analysis for 2 values

Συγκεκριμένα, όταν η πιθανότητα είναι 0.15 η ενέργεια επιλογής καθορίζεται ως πώληση της γης και μάλιστα με μία απόδοση 90 σε αντιδιαστολή με την απόδοση 20 για εξόρυξη ενώ όταν η πιθανότητα είναι 0.35 η ενέργεια επιλογής είναι η εξόρυξη με μία απόδοση 180 έναντι 90 για την ενέργεια της πώλησης.Για την αντιμετώπιση αυτού του προβλήματος θεωρούμε ως p:την πιθανότητα(prior)της ύπαρξης πετρελαίου και έτσι η αναμενόμενη τιμή της απόδοσης για την εξόρυξη δίνεται ως:

$$E[Payoff(drill)] = 700p - 100(1-p) = 800p - 100 \tag{3}$$

Βάση αυτής της προυπόθεσης, μπορούμε να κατασκευάσουμε την γραφική παράσταση της αναμενόμενης απόδοσης έναντι του p η οποία αποτελείται ουσιαστικά απο την ευθεία η οποία δημιουργείται απο τα 2 σημεία του πίνακα που βρίσκονται στα κελια C10 και H5.

Το σημείο τομής των δύο ευθειών είναι και το σημείο εναλλαγής των δυο αποφάσεων.Για να το βρούμε έχουμε διαδοχικά

$$E[Payoff(drill)] = E[Payoff(sell)] \Rightarrow 800p - 100 = 90 \Rightarrow p = 0.2375 \tag{4}$$

Συμπερασματικά, εάν

- $p < 0.2375$ θα πρέπει να παρθεί η απόφαση της ενέργειας πώληση

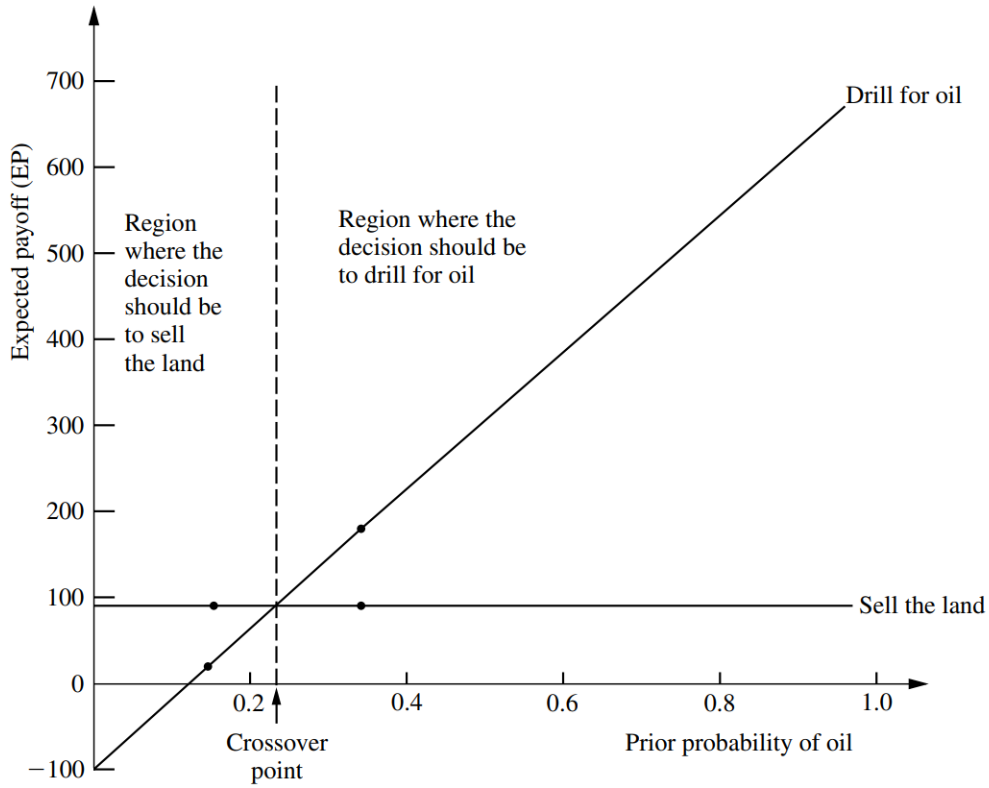



Figure 5: expected payoff as p changes

- $p > 0.2375$ θα πρέπει να παρθεί η απόφαση εξόρυξης της γης

## 2.2 Λήψη αποφάσεων με πειραματισμό

Στο προηγούμενο κεφάλαιο καταλήξαμε στην τελική απόφαση χρησιμοποιώντας τις εκτιμήσεις των prior πιθανοτήτων.Πλέον όμως, μπορούμε να εκτελέσουμε κάποιους πειραματισμούς με στόχο την βελτίωση των εκτιμώμενων πιθανοτήτων ως προς το βαθμό της αβεβαιότητας.Αυτές οι βελτιωμένες πιθανότητες ονομάζονται μεταγενέστερες(posterior probabilities)

Επιστρέφοντας στο παράδειγμα μας, γνωρίζουμε πως υπάρχει η δυνατότητα έρευνας(η οποία δεν λήφθηκε υπόψη στη προηγούμενη προσέγγιση μας χωρίς πειραματισμό στο προηγούμενο κεφάλαιο) η οποία πρόκειται να μας δώσει μια καλύτερη εκτίμηση πιθανότητας αλλά επιφέρει και ένα κόστος 30.000\$. Η έρευνα αυτή χρησιμοποιεί σεισμικές δονήσεις οι οποίες δείχνουν εάν η γεωλογική δομή είναι ευνοική στην ύπαρξη πετρελαίου ή όχι.Τα δυνατά αποτελέσματα είναι

1. USS:Μη ευνοικές σεισμικές δονήσεις

2. FSS:Ευνοικ ές σεισμικές δονήσεις

Βασιζόμενοι σε παλαιότερεες ενδείξεις ,αν υπάρχει πετρέλαιο,η πιθανότητα μη ευνοικών σεισμικών δονήσεων είναι P(USS/κατάσταση=πετρέλαιο)=0.4 ενώ αν δεν υπάρχει πετρέλαιο η πιθανότητα μη ευνοικών σεισμικών δονήσεων είναι P(USS/κατάσταση=χωρίς πετρέλαιο)=0.8 και έτσι, P(FSS/κατάσταση =πετρέλαιο)=0.6 και P(FSS/κατάσταση=χωρίς πετρέλαιο)=0.2 αντίστοιχα.

Ορίζοντας ως:

- n=ο αριθμός των πιθανών καταστάσεων της φύσης
- P(κατάσταση =κατάσταση i)=prior πιθανότητα ότι η πραγματική κατάσταση της φύσης είναι η κατάσταση i
- πληροφορία =πληροφορία του πειραματισμού
- πληροφορία j=μια πιθανή τιμή του πειραματισμού
- P(κατάσταση=κατάσταση i/πληροφορία =πληροφορία j)= =posterior πιθανότητα ότι η κατάσταση της φύσης είναι η κατάσταση i δεδομένου ότι η πληροφορία =πληροφορία j για κάθε i=1,2,. . . n

και χρησιμοποιώντας το θεώρημα του Bayes' ,

$$P(\text{κατάσταση=κατάσταση i/πληροφορία =πληροφορία j}) =$$
$$= \frac{P(\text{ πληροφορία=πληροφορία j/κατάσταση=κατάσταση i})\text{P(κατάσταση =κατάσταση i)}}{\sum_{k=1}^{n} P(\text{πληροφορία =πληροφορία j/κατάσταση=κατάσταση k})\text{P(κατάσταση=κατάσταση i)}}$$
,

μπορούμε πλέον να υπολογίσουμε τις posterior πιθανότητες ως εξής:

1.
$$P(\text{κατάσταση=πετρέλαιο /Πληροφορία=USS}) = \frac{0.4(0.25)}{0.4(0.25) + 0.8(0.75)} = \frac{1}{7}$$

2.
$$P(\text{κατάσταση= χωρίς πετρέλαιο /πληροφορία=USS}) = \frac{6}{7}$$

3.
$$P(\text{κατάσταση=πετρέλαιο /πληροφορία=FSS}) = \frac{0.6(0.25)}{0.6(0.25) + 0.2(0.75)} = \frac{1}{2}$$

4.
$$P(\text{κατάσταση= χωρίς πετρέλαιο /πληροφορία=FSS}) = \frac{1}{2}$$

Το παρακάτω δενδροδιάγραμμα αποτελεί έναν οργανωμένο τρόπο αναπαράστασης των παραπάνω πιθανοτήτων

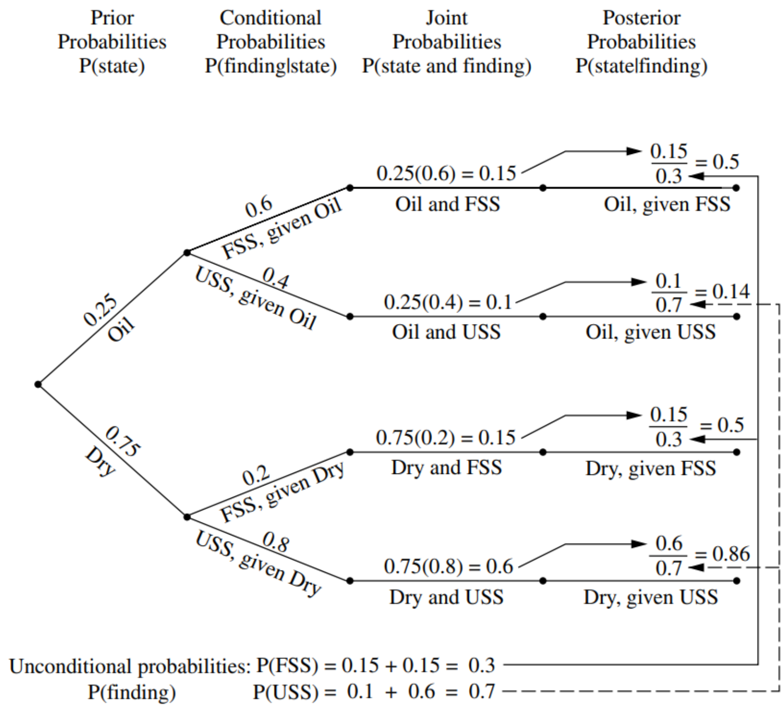



Figure 6: probability tree

Πλέον,χρησιμοποιώντας τις αποδόσεις(σε χιλιάδες δολαρίων),είμαστε σε θέση να υπολογίσουμε τις αναμενόμενες τιμές των αποδόσεων εάν η πληροφορία είναι ευνοικές σεισμικές δονήσεις(FSS) ή μη ευνοικές σεισμικές δονήσεις(USS)

1. E[payoff(εξόρυξη/πληροφορία=FSS)=$\frac{1}{2}(700) + \frac{1}{2}(-100) - 30 = 270$
2. E[payoff(πώληση/πληροφορία=FSS)=$\frac{1}{2}(90) + \frac{1}{2}(90) - 30 = 60$
3. E[payoff(εξόρυξη/πληροφορία=USS)=$\frac{1}{7}(700) + \frac{6}{7}(-100) - 30 = -15.7$
4. E[payoff(πώληση/πληροφορία=USS)=$\frac{1}{7}(90) + \frac{6}{7}(90) - 30 = 60$

Έτσι, αν η πληροφορία της έρευνας είναι μη ευνοικές σεισμικές δονήσεις ως βέλτιστη επιλογή φαίνεται η ενέργεια πώλησης της γης με απόδοση 90 ενώ στην περίπτωση που έχουμε ευνοικές σεισμικές δονήσεις προτιμάται η εξόρυξη με απόδοση 300. Για την πραγματοποίηση του παραπάνω πειράματος βέβαια, είναι απαραίτητος ο προσδιορισμός της δυνητικής του αξίας.Αυτή μπορεί να γίνει με δύο τρόπους.Ο πρώτος είναι ο προσδιορισμός της αναμενόμενης αξίας την τέλειας πληροφορίας.Θεωρούμε δηλαδή πως το πείραμα δεν περιορίζεται απο κάποιο βαθμό αβεβαιότητας και υπολογίζουμε την βελτίωση στην αναμενόμενη απόδοση(payoff) χωρίς να λαμβάνεται υπόψιν το κόστος του πειράματος.Στη συνέχεια υπολογίζεται η αξία της τέλειας πληροφορίας απο τη σχέση

Αναμενόμενη αξία της τέλειας πληροφορίας(EVPI)=
=αναμενόμενη απόδοση με χρήση της τέλειας πληροφορίας
−αναμενόμενη απόδοση χωρίς την τέλεια πληροφορία

Για το παραδειγμά μας έχουμε

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Expected Payoff with Perfect Information for Goferbroke | | | | | | |
| 2 | | | | | | | |
| 3 | | | | State of Nature | | | |
| 4 | | Alternative | Oil | Dry | | | |
| 5 | | Drill | 700 | -100 | | | |
| 6 | | Sell | 90 | 90 | | | |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | Prior Probability | 0.25 | 0.75 | | | |
| 11 | | Maximum Payoff | 700 | 90 | | | |
| 12 | | | | | | | |
| 13 | | Expected Payoff with Perfect Information = | | | | 242.5 | |

| | C | D | E | F | G |
|---|---|---|---|---|---|
| 11 | =MAX(C5:C9) | =MAX(D5:D9) | =MAX(E5:E9) | =MAX(F5:F9) | =MAX(G5:G9) |

| | F |
|---|---|
| 13 | =SUMPRODUCT(C10:G10,C11:G11) |

Αναμενόμενη απόδοση για την τέλεια πληροφόρηση=0.25(700)+0.75(90)=242.5 και άρα

$$EVPI = 242.5 - 100 = 142.5$$

Εφόσον το 142.5 είναι μεγαλύτερο του 30 φαίνεται πως αξίζει να προχωρήσουμε στην σεισμική έρευνα(πειραματισμό) Ο δεύτερος τρόπος υπολογίζει κατευθείαν την τιμή πειραματισμού σε αντιδιαστολή με τον προηγούμενο υπολογισμό ενός άνω ορίου.Συγκεκριμένα θα πρέπει αρχικά να ληφθεί η αναμενόμενη απόδοση με πειραματισμό.Για τον προσδιορισμό αυτού του μεγέθους θα πρέπει να βρεθούν οι posterior πιθανότητες ,η βέλτιστη πολιτική με πειραματισμό και η αντίστοιχη αναμενόμενη απόδοση(χωρίς το κόστος του πειράματος)

$$\text{Αναμενόμενη απόδοση με πειραματισμό}= \sum_j (\Pi\lambda\eta\rho o\varphi o\rho\iota\alpha = \pi\lambda\eta\rho o\varphi o\rho\iota\alpha \quad j)\text{E[payoff/Πληροφορια=πληροφορια j]}$$

Για το παραδειγμά μας γνωρίζουμε πως

1. P(USS)=0.7
2. P(FSS)=0.3
3. P(Payoff/πληροφορία=USS)=90
4. P(Payoff/πληροφορία=FSS)=270

και έτσι Αναμενόμενη απόδοση με πειραματισμό=0.7(90)+0.3(300)=153 Τελικά,

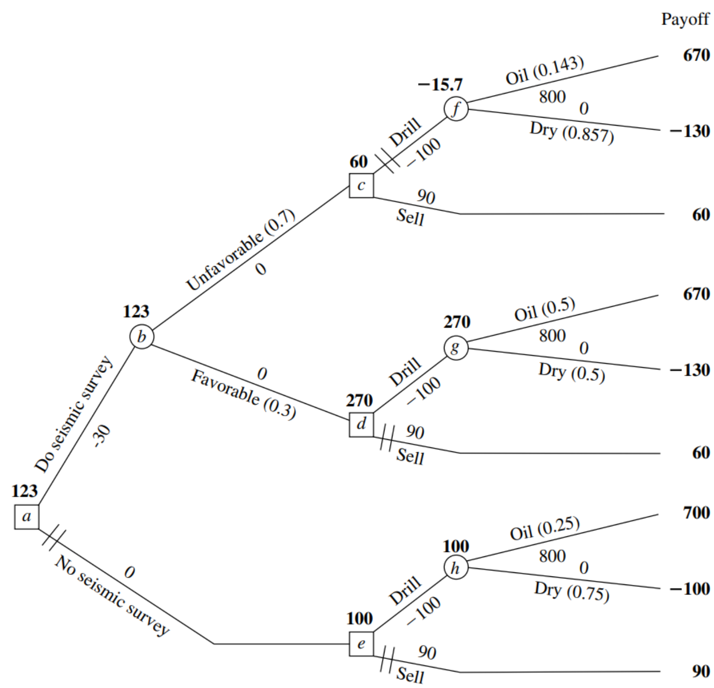
Αναμενόμενη αξία του πειραματισμού(EVE)=Αναμενόμενη απόδοση με πειραματισμό-Αναμενόμενη απόδοση χωρίς πειραματισμό

Άρα,

$$EVE = 153 - 100 = 53$$

## 2.3 Δένδρα αποφάσεων

Τα δένδρα αποφάσεων αποτελούν μια οπτικοποίηση των παραπάνω διαδικασιών λήψης απόφασης.Ένα από τα βασικά πλεονεκτήματα τους είναι πως βοηθούν στα προβλήματα όπου είναι αναγκαία η λήψη διαδοχικών αποφάσεων καθώς οπτικοποιούν την εξάρτηση τους. Η γραφική απεικόνιση των δένδρων αποφάσεων αποτελείται από κλαδιά και κόμβους.Οι κανόνες που ακολουθούνται είναι οι εξής: Με κύκλο συμβολίζονται οι κόμβοι που δηλώνουν την φυσική κατάσταση, με τετράγωνα συμβολίζονται οι κόμβοι της απόφασης, ενώ τα κλαδιά που προέρχονται από έναν κόμβο απόφασης αντανακλούν όλες τις εναλλακτικές αποφάσεις που μπορεί να ληφθούν και αφορούν στο συγκεκριμένο σημείο. Στο παράδειγμα μας το δένδρο απόφασης αναπαρίσταται ως εξής:




**Ανάλυση δένδρου αποφάσεων**:

1. Ξεκινάμε από τη δεξιά πλευρά του δέντρου αποφάσεων και μετακινούμαστε αριστερά κάθε φορά. Για κάθε στήλη, εκτελούμε είτε το βήμα 2 είτε το βήμα 3, ανάλογα με το αν οι κόμβοι σε αυτή τη στήλη είναι κόμβοι ευκαιρίας ή κόμβοι αποφάσεων.

2. Για κάθε κόμβο ευκαιρίας, υπολογίζουμε την αναμενόμενη απόδοση πολλαπλασιάζοντας την αναμενόμενη απόδοση κάθε κλάδιου (εμφανίζεται με έντονη γραφή στα δεξιά κάθε κλαδιού) με την πιθανότητα αυτού του κλαδιού και στη συνέχεια άθροιζουμε αυτα τα αποτελέσμτα. Καταγράφουμε αυτήν την αναμενόμενη απόδοση για κάθε κόμβο απόφασης με έντονη γραφή δίπλα στο κόμβο και προσδιορίζουμε αυτήν την ποσότητα ως την αναμενόμενη ανταμοιβή για τον κλάδι που οδηγεί σε αυτό το κόμβο.

3. Για κάθε κόμβο απόφασης, συγκρίνουμε τις αναμενόμενες αποδόσεις των υποτμημάτων του και επιλέγουμε την λύση της οποίας το υποτμήμα έχει τη μεγαλύτερη αναμενόμενη απόδοση. Σε κάθε περίπτωση, καταγράφουμε την επιλογή στο δέντρο αποφάσεων εισάγοντας μια διπλή παύλα ως φράγμα σε κάθε κλάδι που απορρίφθηκε.

Για να ξεκινήσουμε τη διαδικασία, λάμβάνουμε υπόψη την πιο δεξιά στήλη κόμβων, δηλαδή, τους κόμβους ευκαιρίας Ξεκινάμε απο την πιο δεξιά στήλη κόμβων f,g και h.Εφόσον αποτελούν κόμβους ευκαιρίας υπολογίζουμε τις αναμενόμενες αποδόσεις ως εξής:

- κόμβος f: $EP = \frac{1}{7}(670) + \frac{6}{7}(-130) = -15.7$
- κόμβος g:$EP = \frac{1}{2}(670) + \frac{1}{2}(-130) = 270$
- κόμβος h:$EP = \frac{1}{4}(700) + \frac{3}{4}(-100) = 100$

Αυτές οι αναμενόμενες αποδόσεις τοποθετούνται πάνω από τους κόμβους.Στην συνέχεια,συνεχίζουμε μία στήλη αριστερά στους κόμβους απόφασης c,d και e εκτελούμε το βήμα 3

- κόμβος c:

  Η εξόρυξη έχει αναμενόμενη απόδοση -15.7

  Η πώληση έχει αναμενόμενη απόδοση 60

  εφόσον $60 > -15.7$ επιλέγω την πώληση

- κόμβος d:

  Η εξόρυξη έχει αναμενόμενη τιμή 270

  Η πώληση έχει αναμενόμενη τιμή 60

  Αφού $270 > 60$ επιλέγω την εξόρυξη

- κόμβος e:

  Η εξόρυξη έχει αναμενόμενη τιμή 100

  Η πώληση έχει αναμενόμενη τιμή 90

  Άρα επιλέγω την εξόρυξη

Σύνεχίζω στον κόμβο b ο οποίος είναι κόμβος ευκαιρίας με

$$EP = 0.7(60) + 0.3(270) = 123$$

Τελικά, φθάνουμε στον κόμβο απόφασης a
κόμβος a:

- κάνουμε σεισμική έρευνα EP=123
- δεν κάνουμε σεισμική έρευνα EP=100

Άρα επιλέγουμε να κάνουμε σεισμική έρευνα.

**Βέλτιστη πολιτική**

Η βέλτιστη πολιτική βάση της παραπάνω ανάλυσης προκύπτει ως να γίνει η σεισμική έρευνα.Εάν το αποτέλεσμα είναι μη ευνοικό να γίνει πώληση της γης ενώ αν το αποτέλεσμα είναι ευνοικό να γίνει εξόρυξη πετρελαίου.Η αναμενόμενη απόδοση ειναι 123$.

# 3 Εισαγωγή στις στοχαστικές διαδικασίες και τα μαρκοβιανά συστήματα

Οι μαρκοβιανές αλυσίδες αποφάσεων αποτελούν το σημέιο τομής δύο ιδιαίτερα εφαρμοσμένων κλάδων της μαθηματικής επιστήμης :των στοχαστικών διαδικασιών και της επιχειρησιακής έρευνας. Στη συνέχεια, θα ερθουμε σε επαφή με τις στοχαστικές διαδικασίες και συγκεκριμένα με ενα ιδιαίτερα χρήσιμο εργαλείο αυτών ,τις μαρκοβιανες διαδικασίες και τα εργαλεία που μας προσφέρονται και υπεισέρχονται στο χωρο της επιχειρησιακής έρευνας

## 3.1 Στοχαστικές διαδικασίες

Υπάρχουν συχνά φαινόμενα τα οποία εξελίσσονται στο χρόνο με κάποια τυχαιότητα .Ας πάρουμε για παράδειγμα τη ρίψη ενός ζαριού ή την τιμή μια μετοχής μια επιχείρησης και την εξέλιξη αυτών με τη πάροδο του χρόνου.Τα μαθηματικά εργαλεία τα οποία μοντελοποιούν τέτοια φαινόμενα ονομάζονται στοχαστικές διαδικασίες.

Υπάρχουν αρκετοί τρόποι για να ορίσουμε μια στοχαστική διαδικασία .Πάντως σίγουρα βασικά χαρακτηριστικά της είναι :

- ένα σύνολο X που ονομάζεται χώρος καταστάσεων και στον οποίο ανήκουν οι τιμές που παίρνει η στοχαστική διαδικασία . Εάν ο χώρος κατάστασεων είναι διακριτός, η διαδικασία ορίζεται ως μία αλυσίδα και οι καταστάσεις συνήθως προσδιορίζονται με το σύνολο των φυσικών αριθμών 0, 1, 2,... ή ένα υποσύνολο αυτού. Ένα παράδειγμα χώρου διακριτών καταστάσεων είναι ο αριθμός των πελατών σε μια εγκατάσταση εξυπηρέτησης. Ένα παράδειγμα ενός χώρου συνεχούς κατάστασης είναι το χρονικό διάστημα που περιμένει ένας πελάτης για να εξυπηρετηθεί.Στη συνέχεια θα ασχοληθούμε μονο με διαδικασίες διακριτού χώρου καταστάσεων.

- ένα σύνολο T που συνήθως αλλά όχι κατα κανόνα είναι ένα σύνολο χρόνων. Εάν το σύνολο χρόνου είναι διακριτό, π.χ,$T = 0, 1, 2, ...$, τότε έχουμε μια διαδικασία διακριτού χρόνου, διαφορετικά εάν το T είναι συνεχές, π.χ,$T = t : 0 \leq t < \infty$ καλούμε την διαδικασία ως διαδικασία συνεχούς χρόνου.

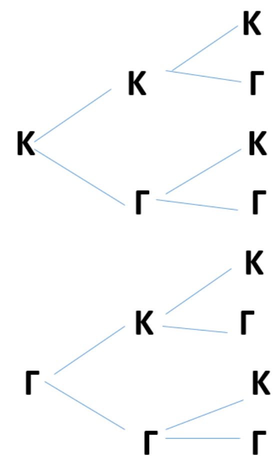
Μία **στοχαστική διαδικασία** ορίζεται ως μία συλλογή τυχαίων μεταβλητών $x_t$ όπου ο δείκτης $t$ διατρέχει ένα δεδομένο σύνολο T.Συχνά το T λαμβάνεται ως το σύνολο των μη αρνητικών ακεραίων και το $x_t$ αντιπροσωπεύει ένα μετρήσιμο χαρακτηριστικό ενδιαφέροντος την στιγμή $t$.
Η τρέχουσα κατάσταση του συστήματος μπορεί να εμπίπτει σε οποιαδήποτε απο τις αμοιβαία αποκλειούμενες $M+1$ κατηγορίες οι οποίες ονομάζονται καταστάσεις του συστήματος **states**. Η τυχαία μεταβλητή $X_t$ αντιπροσωπεύει την κατάσταση του συστήματος τη στιγμή $t$, άρα οι μόνες πιθανές τιμές της είναι $0, 1, ..., M$. Το σύστημα παρατηρείται σε συγκεκριμένες χρονικές στιγμές, με ετικέτα $t = 0, 1, 2, ....$ Έτσι, η στοχαστική διαδικασία $\{X_t\} = \{X_0, X_1, X_2, ...\}$ παρέχει μια μαθηματική αναπαράσταση του τρόπου με τον οποίο η κατάσταση του φυσικού συστήματος εξελίσσεται με την πάροδο του χρόνου. Αυτό το είδος διαδικασίας αναφέρεται ως διαδικασία διακριτού χρόνου και είναι αυτή η οποία θα μας απασχολήσει κυρίως στη συνέχεια

## 3.2 Παραδείγματα στοχαστικών διαδικασιών

**Παράδειγμα 1**
Στρίβουμε ένα κέρμα 3 φορές .Ξεκινόντας από το 0 κάνουμε ένα βήμα προς τα αριστερά κάθε φορά που φέρνουμε γράμματα και ένα βήμα δεξιά κάθε φορά που φέρνω κεφαλή. Μπορούμε να περιγράψουμε τη θέση μας ως μια στοχαστική διαδικασία ορισμένη στο σύνολο χρόνων $T = \{0, 1, 2, 3\}$ με τιμές στον χώρο καταστάσεων $Z$. Πράγματι, αν για $k = 0, 1, 2, 3$ συμβολίζουμε με $X_k$ τη θέση μας μετά από k στριψίματα, τότε οι $(X_k)_{k \in T}$ είναι τυχαίες μεταβλητές με τιμές στον $Z$, που λαμβάνει τιμές στον δειγματικό χώρο,

$$\Omega = \{KKK, KK\Gamma, K\Gamma K, K\Gamma\Gamma, \Gamma KK, \Gamma K\Gamma, \Gamma\Gamma K, \Gamma\Gamma\Gamma\}$$

.




Έχουμε λοιπόν $X_0(\omega) = 0$ για κάθε $\omega \in \Omega$ και

$$x_1(\omega) = \begin{cases} 1, & \text{εάν } \omega \in (KKK, KK\Gamma, K\Gamma K, K\Gamma\Gamma) \\ -1, & \text{εάν } \omega \in (\Gamma KK, \Gamma K\Gamma, \Gamma\Gamma K, \Gamma\Gamma\Gamma) \end{cases}$$

$$x_2(\omega) = \begin{cases} 2, & \text{εάν } \omega \in (KKK, KK\Gamma) \\ 0, & \text{εάν } \omega \in (\Gamma KK, \Gamma K\Gamma, K\Gamma K, K\Gamma\Gamma) \\ -2, & \text{εάν } \omega \in (\Gamma\Gamma K, \Gamma\Gamma\Gamma) \end{cases}$$

$$x_3(\omega) = \begin{cases} 3, & \text{εάν } \omega \in (KKK) \\ -3, & \text{εάν } \omega \in (\Gamma\Gamma\Gamma) \\ -1, & \text{εάν } \omega \in (K\Gamma\Gamma, \Gamma K\Gamma, \Gamma\Gamma K) \\ 1, & \text{εάν } \omega \in (K\Gamma K, \Gamma KK, KK\Gamma) \end{cases}$$

**Παράδειγμα 2**

Ένα κατάστημα φωτογραφικών μηχανών διαθέτει ένα συγκεκριμένο μοντέλο κάμερας που μπορεί να παραγγελθεί εβδομαδιαία. Έστω $D_1, D_2, ...$ αντιπροσωπεύουν τη ζήτηση για αυτήν την κάμερα (ο αριθμός μονάδων που θα πωλούνταν εάν το απόθεμα δεν εξαντληθεί) κατά τη διάρκεια της πρώτης εβδομάδας,δεύτερης εβδομάδας, . . . , αντίστοιχα. Υποθέτουμε ότι τα $D_i$ είναι ανεξάρτητα και πανομοιότυπα κατανεμημένες τυχαίες μεταβλητές με κατανομή Poisson με μέσο όρο 1. Έστω οτι το $X_0$ αντιπροσωπεύει τον αριθμό των καμερών που υπάρχουν στην αρχή, το $X_1$ τον αριθμό των καμερών που υπάρχουν στο τέλος της εβδομάδας 1, $X_2$ ο αριθμός των φωτογραφικών μηχανών στο χέρι στο τέλος της εβδομάδας 2, και ούτω καθεξής. Ας υποθέσουμε ότι το $Q_0 = 3$. Το βράδυ του Σαββάτου το κατάστημα κάνει μια παραγγελία που παραδίδεται στην ώρα του για τα επόμενα εγκαίνια του καταστήματος τη Δευτέρα. Το κατάστημα χρησιμοποιεί την ακόλουθη πολιτική παραγγελιών: Εάν δεν υπάρχουν κάμερες στο απόθεμα, το κατάστημα παραγγέλνει 3 κάμερες. Ωστόσο, εάν υπάρχουν κάμερες σε απόθεμα, δεν γίνεται παραγγελία. Οι πωλήσεις χάνονται όταν η ζήτηση υπερβαίνει το απόθεμα σε ετοιμότητα.

Έτσι, $X_t$ για $t = 0, 1, ...$ είναι μια στοχαστική διαδικασία της μορφής που μόλις περιγράφηκε. Οι πιθανές καταστάσεις της διαδικασίας είναι οι ακέραιοι αριθμοί 0, 1, 2, 3, που αντιπροσωπεύουν τον πιθανό αριθμό των καμερών στο χέρι στο τέλος της εβδομάδας. Οι τυχαίες μεταβλητές $(X_t)$ είναι εξαρτημένες και μπορούν να αξιολογηθούν επαναληπτικά από την έκφραση:

$$x_{t+1} = \begin{cases} \max(3 - D_{t+1}, 0), & \text{εάν } x_t = 0 \\ max(x_t - D_{t+1}, 0), & \text{εάν } x_t \geq 1 \end{cases}$$

για t=0,1,2.......

## 3.3 Μαρκοβιανές αλυσίδες

Μία ειδική περίπτωση των στοχαστικών διαδικασιών διακριτού χρόνου είναι οι μαρκοβινές διαδικασίες ή μαρκοβιανές αλυσίδες.Για να οπτικοποιήσουμε την ιδιότητα εκείνη την οποία τις διακρίνει από τις υπόλοιπες στοχαστικές διαδικασίες ας πάρουμε ως παράδειγμα έναν βάτραχο ο οποίος πηδά απο νούφαρο σε νούφαρο σε μία λίμνη .Το προς τα που θα πηδήξει ο βάτραχος εξαρτάται μόνο από το ποιες πληροφορίες μπορεί να συλλέξει από το τρέχον νούφαρο στο οποίο βρίσκεται καθώς δεν θυμάται τιποτα για τις καταστάσεις που επισκέφτηκε πριν απο την τρέχουσα θέση.

**Μαρκοβιανή ιδιότητα:**
Μία στοχαστική διαδικασία διακριτού χρόνου είναι μαρκοβιανή εάν για όλα τα t=0,1,2... και για όλες τις καταστάσεις ικανοποιείται η μαρκοβιανή ιδιότητα δηλαδή:

$$P(X_{t+1} = i_{t+1}/X_t = i_t, X_{t-1} = i_{t-1}, .., X_0 = i_0) = P(X_{t+1} = i_{t+1}/X_t = i_t)$$

Με λόγια, αυτή η μαρκοβιανή ιδιότητα λέει ότι η πιθανότητα οποιουδήποτε μελλοντικού «γεγονότος», δεδομένου οποιουδήποτε παρελθόντικού «γεγονότος» και της παρούσας κατάστασης $X_t = i_t$, είναι ανεξάρτητη από το παρελθόν και εξαρτάται μόνο από την παρούσα κατάσταση. Οι δεσμευμένες πιθανότητες $p_{ij} = P(X_{t+1} = j/X_t = i)$ για μία μαρκοβιανή αλυσίδα ονομάζονται **πιθανότητες μετάβασης 1ης τάξης** από την κατάσταση i στην κατάσταση j κατά το t+1 βήμα .Εάν οι εν λόγω πιθανότητες δεν εξαρτώνται απο το χρονικό βήμα t ονομάζονται **στάσιμες** εννοώντας πως

$$p_{ij} = P(X_{t+1} = j/X_t = i) = P(X_1 = j/X_0 = i)$$

και τότε κάνουμε λόγω για χρονικά ομοιογενείς αλυσίδες $X_t$.Επίσης ,η ύπαρξη των στάσιμων πιθανοτήτων μετάβασης 1ης τάξης υπονοούν ότι για κάθε i,j και n (n=0,1,2,...)

$$p_{ij}^{(n)} = P(X_{t+n} = j/X_t = i) = P(Xn = j/X_0 = i)$$

οι οποίες ονομάζονται **πιθανότητες μετάβασης n-τάξης** και δηλώνουν την πιθανότητα το σύστημα να βρεθεί στη κατάσταση j μετά από ακριβώς n βήματα(μονάδες χρόνου )δεδομένου οτι ξεκίνησε στην κατάσταση ι σε οποιοδήποτε χρόνο t.

Ο πίνακας P ο οποίος σχηματίζεται τοποθετόντας τις pij σε γραμμή i και στήλη j για κάθε i και j ονομάζεται **πίνακας πιθανοτήτων μετάβασης** και αναπαρίσταται ως:

$$\mathbf{P}^{(n)} = \begin{array}{c c} \text{State} & \begin{array}{cccc} 0 & 1 & \cdots & M \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ M \end{array} & \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & \cdots & p_{0M}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} & \cdots & p_{1M}^{(n)} \\ \cdots & \cdots & \cdots & \cdots \\ p_{M0}^{(n)} & p_{M1}^{(n)} & \cdots & p_{MM}^{(n)} \end{bmatrix} \end{array}$$

Σημειώστε ότι η πιθανότητα μετάβασης σε μια συγκεκριμένη γραμμή και στήλη είναι για τη μετάβαση από την κατάσταση της γραμμής στην κατάσταση της στήλης. Όταν n =1, αφήνουμε τον εκθέτη n και απλά αναφέρομαστε σε πίνακα μετάβασης. Σε αυτό το σημείο σημειώνεται πως τα στοιχεία του πίνακα μετάβασης ικανοποιούν τις παρακάτω ιδιότητες:

- $0 \leq p_{i,j}^{(n)} \leq 1$ για κάθε i και j;n=0,1,2,...
- $\sum_{j=0}^{M} p_{i,j}^{(n)} = 1$ για κάθε i;n=0,1,2,...

### 3.3.1 Παράδειγμα

Ένα μοντέλο καιρικών συνθηκών.

Σκεφτείτε μια ομοιογενή αλυσίδα Markov διακριτού χρόνου που περιγράφει το ημερήσιο μοτίβο καιρού στο Μπέλφαστ της Βόρειας Ιρλανδίας . Απλοποιούμε την κατάσταση λαμβάνοντας υπόψη μόνο τρεις τύπους καιρού: βροχερό, συννεφιασμένο και ηλιόλουστο. Αυτές οι τρεις καιρικές συνθήκες περιγράφουν τις τρεις καταστάσεις της αλυσίδας Markov: η κατάσταση 1 (R) αντιπροσωπεύει μια (κυρίως) βροχερή μέρα. κατάσταση 2 (C), μια (κυρίως) συννεφιασμένη μέρα. και κατάσταση 3 (S), μια (κυρίως) ηλιόλουστη μέρα. Ο καιρός παρατηρείται καθημερινά. Σε κάθε δεδομένη βροχερή μέρα, η πιθανότητα ότι θα βρέξει την επόμενη μέρα υπολογίζεται στο 0,8, η πιθανότητα η επόμενη μέρα να είναι συννεφιασμένη είναι 0,15, ενώ η πιθανότητα αύριο να έχει ηλιοφάνεια είναι μόνο 0,05. Ομοίως, οι πιθανότητες μπορούν να υπολογιστούν όταν μία ημέρα είναι συννεφιασμένη ή ηλιόλουστη, όπως φαίνεται στο Σχήμα 1.

Από όπου μπορεί εύκολα να υπολογιστεί και ο πίνακας πιθανοτήτων μετάβασης

$$P = \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \end{pmatrix}$$

Τα στοιχεία του πίνακα αντιπροσωπεύουν δεσμευμένες πιθανότητες.Για παράδειγμα το στοιχείο $p_{32}$ δηλώνει ότι η πιθανότητα αύριο να είναι συννεφιασμένη μέρα δεδομένου πως σήμερα είναι ηλιόλουστη μέρα είναι 0.3.Επίσης χρησιμοποιώντας πως για χρονικά ομοιογενείς αλυσίδες έχουμε πως

$$Prob(X_{n+m} = a, X_{n+m-1} = b, ..., X_{n+2} = k, X_{n+1} = j / X_n = i) = p_{ij} p_{jk} .... p_{cb} p_{ba} \tag{5}$$

και έτσι μπορούμε να υπολογίσουμε πιο περίπλοκες δεσμευμένες πιθανότητες.

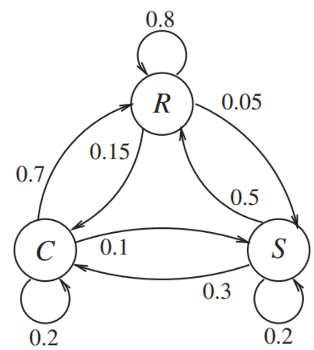



Figure 7: Σχήμα 1

Για παράδειγμα,μπορούμε να βρούμε την πιθανότητα αύριο να είναι συννεφιασμένη μέρα και μεθαύριο βροχερή μέρα δεδομένου ότι σήμερα είναι ηλιόλουστη μέρα.Σε αυτήν την περίπτωση για κάθε n=0,1,2,... ισχύει ότι:

$$Prob\{X_{n+2} = R, X_{n+1} = C/X_n = S\} = p_{SC}p_{CR} = (0.3)(0.7) = 0.21 \qquad (6)$$

## ΠΑΡΑΔΕΙΓΜΑ ΔΙΑΔΙΚΑΣΙΑΣ ΓΙΑ ΤΗΝ ΟΠΟΙΑ ΔΕΝ ΙΣΧΥΕΙ Η ΜΑΡΚΟΒΙΑΝΗ ΙΔΙΟΤΗΤΑ

Ένας μπασκετμπολίστας προπονείται στις ελεύθερες βολές. Κάθε φορά που εκτελεί μια βολή, η πιθανότητα να ευστοχήσει είναι 8/10, αν έχει ευστοχήσει και στις δύο προηγούμενες προσπάθειές του, 5/10, αν έχει αστοχήσει και στις δύο προηγούμενες προσπάθειες και 7/10 σε κάθε άλλη περίπτωση.Ορίζοντας ως $x_n, n \in N$ την στοχαστική διαδικασία όπου $x_n = 1$ αν η n-οστη προσπάθεια είναι εύστοχη και $x_n = 0$ αν είναι άστοχη μπορούμε να παρατηρήσουμε πως η πιθανότητα ευστοχίας στην επόμενη προσπάθεια εξαρτάται απο το αποτέλεσμα των δυο τελευταίων βολών.Συγκεκριμένα,

$$P[x_{n+1} = 1/x_n = x_{n-1} = 1] = \frac{8}{10}$$

και

$$P[x_{n+1} = 1/x_{n-1} = 0, x_n = 1] = \frac{7}{10}$$

όπου $\frac{8}{10}$ διαφορετικό του $\frac{7}{10}$.

Το πρόβλημα εδώ είναι ότι ο χώρος καταστάσεων της αλυσίδας $X = \{0, 1\}$ είναι πολύ μικρός για να κωδικοποιήσει την πληροφορία που χρειαζόμαστε για να περιγράψουμε την κατανομή της επόμενης κατάστασης. Μπορούμε όμως, μεγαλώνοντας τον χώρο καταστάσεων, να περιγράψουμε την προπόνηση του αθλητή ως μια μαρκοβιανή αλυσίδα. Αν ορίσουμε τη στοχαστική διαδικασία $Y_n\{n \in N\}$, με $Y_n = (X_n, X_{n+1})$, αυτή είναι μια μαρκοβιανή αλυσίδα στον χώρο καταστάσεων $X = \{(0,0),(0,1),(1,0)(1,1)\}$ με πίνακα πιθανοτήτων μετάβασης

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 3/4 & 7/10 \\ 3/10 & 7/10 & 0 & 0 \\ 0 & 0 & 2/10 & 8/10 \end{pmatrix}$$

### 3.3.2 Ταξινόμηση καταστάσεων μαρκοβιανών αλυσίδων

Είναι προφανές πως οι πιθανότητες μετάβασης που σχετίζονται με τις καταστάσεις παίζουν σημαντικό ρόλο στη μελέτη των μαρκοβιανών αλυσίδων.Στην συνέχεια περιγράφονται κάποιες έννοιες και ορισμοί που αφορούν αυτές τις καταστάσεις.

Δεδομένων δύο καταστάσεων i και j, μια **διαδρομή(path)** από το i στο j είναι μια ακολουθία μεταβάσεων που ξεκινά απο το i και τελειώνει σε j, έτσι ώστε κάθε μετάβαση στην ακολουθία να έχει θετική πιθανότητα να συμβεί

Η κατάσταση $j$ ονομάζεται **προσβάσιμη** από την κατάσταση $i$ αν $p_{ij}^n > 0$ για κάποιο $n \geq 0$.

Εάν η κατάσταση $j$ είναι προσβάσιμη απο την κατάσταση $i$ και η κατάσταση $i$ είναι προσβάσιμη απο την κατάσταση $j$ τότε λέμε πως οι καταστάσεις $i, j$ **επικοινωνούν**

Ένα σύνολο καταστάσεων S σε μια αλυσίδα Markov είναι ένα **κλειστό σύνολο**εάν καμία κατάσταση εκτός του S δεν είναι προσβάσιμη από καμία κατάσταση στο S

Η σχέση επικοινωνίας είναι μία σχέση ισοδυναμίας η οποία έχει τις παρακάτω ιδιότητες:

- Οποιαδήποτε κατάσταση επικοινωνεί με τον εαυτό της καθώς,

$$p_{ii}^0 = P\{X_0 = i/X_0 = i\} = 1$$

- Εάν η κατάσταση i επικοινωνεί με την κατάσταση j, τότε και η κατάσταση j επικοινωνεί με την κατάσταση i
- Εάν η κατάσταση i επικοινωνεί με την κατάσταση j και η κατάσταση j επικοινωνεί με την κατάσταση k, τότε η κατάσταση i επικοινωνεί με την κατάσταση k

Ως αποτέλεσμα αυτών των τριών ιδιοτήτων επικοινωνίας, οι καταστάσεις μπορούν να διαμεριστούν σε μία ή περισσότερες ξεχωριστές κλάσεις έτσι ώστε εκείνες οι καταστάσεις που επικοινωνούν μεταξύ τους να βρίσκονται στην ίδια κλάση. (Μια κλάση μπορεί να αποτελείται από μια ενιαία κατάσταση). Αν υπάρχει μόνο μία κλάση,δηλαδή όλες οι καταστάσεις επικοινωνούν, η αλυσίδα Markov χαρακτηρίζεται ως **μη υποβιβάσιμη (irreducible)**.

## ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΕΣ ΚΑΙ ΠΑΡΟΔΙΚΕΣ ΚΑΤΑΣΤΑΣΕΙΣ

Συχνά, μας ενδιαφέρει αν μία διαδικασία η οποία υπεισέρχεται σε μία κατάσταση θα επιστρέψει ή όχι στην κατάσταση αυτή. Μία κατάσταση λέμε ότι είναι παροδική αν κατά τη είσοδο σε αυτήν η διαδικασία ίσως δεν μπορεί να γυρίσει ποτέ πίσω σε αυτήν.

Η κατάσταση i είναι **παροδική(transient)** αν και μόνο αν υπάρχει κατάσταση $j(j \neq i)$ που είναι προσβάσιμη από την i αλλά δεν ισχύει το αντίστροφο δηλαδή η i δεν είναι προσβάσιμη απο την j.

Με άλλα λόγια μια κατάσταση i είναι παροδική εάν υπάρχει τρόπος να φύγουμε απο την κατάσταση i που δεν επιστρέφει ποτε πίσω σε αυτήν.Συνεπώς,η διαδικασία θα επισκεφθεί μια παροδική κατάσταση ένα πεπερασμένο αριθμό φορών.

Εάν μία κατάσταση δεν είναι παροδική τότε είναι επαναλαμβανόμενη.

Μία κατάσταση ονομάζεται **επαναλαμβανόμενη(recurrent)** εάν εφόσον η διαδικασία εισαχθεί σε αυτήν την κατάσταση είναι βέβαιο πως θα γυρίσει πισω σε αυτήν ξανά.

Μια διαφορετική προσέγγιση σε αυτές τις έννοιες επιτυγχάνεται μέσω του ορισμού του συνολικού αριθμού επισκέψεων σε μία κατάσταση. Αν θεωρήσω μία μαρκοβιανή αλυσίδα $\{X_n\}_{n \in N_0}$ με τιμές σε ένα αριθμήσιμο χώρο καταστάσεων X ,για $x \in X$ ορίζουμε ως το σύνολικό πλήθος επισκέψεων της αλυσίδας στην κατάσταση x

$$V(x) = \sum_{k=0}^{\infty} \mathbb{1}\{x_k = x\}$$

Θα ονομάζουμε μια κατάσταση $x \in X$ επαναληπτική αν $P(V(x) < \infty) = 0$. Θα ονομάζουμε μια κατάσταση $x \in X$ παροδική , αν $P(V(x) < \infty) = 1$.

Μία κατάσταση ονομάζεται **απορροφητική(absorbing)** εάν αφού η διαδικασία εισαχθεί σε αυτήν , δεν θα εξέλθει απο εκείνη ξανά .Έτσι η κατάσταση i είναι απορροφητική αν και μόνο αν $p_{ii} = 1$

Μια κατάσταση i είναι **περιοδική** με περίοδο $k > 1$ αν k είναι ο μικρότερος αριθμός έτσι ώστε όλα μονοπάτια που οδηγούν από την κατάσταση i πίσω στην κατάσταση i έχουν μήκος που είναι πολλαπλάσιο του k. Αν μία επαναληπτική κατάσταση δεν είναι περιοδική, αναφέρεται ως απεριοδική

Τέλος,Εάν όλες οι καταστάσεις σε μια αλυσίδα είναι επαναλαμβανόμενες, μη περιοδικές και επικοινωνούν μεταξύ τους,η αλυσίδα λέγεται ότι είναι **εργοδική**.

### 3.3.3 Παραδείγματα

Το σχήμα 2 απεικονίζει τις διάφορες καταστάσεις μιας διαδικασίας. Δεν υπάρχουν πιθανότητες στα μονοπάτια όμως θεωρείται έμμεσα ότι το άθροισμα των πιθανοτήτων σε μονοπάτια που εξέρχονται από οποιαδήποτε κατάσταση ισούται με ένα.

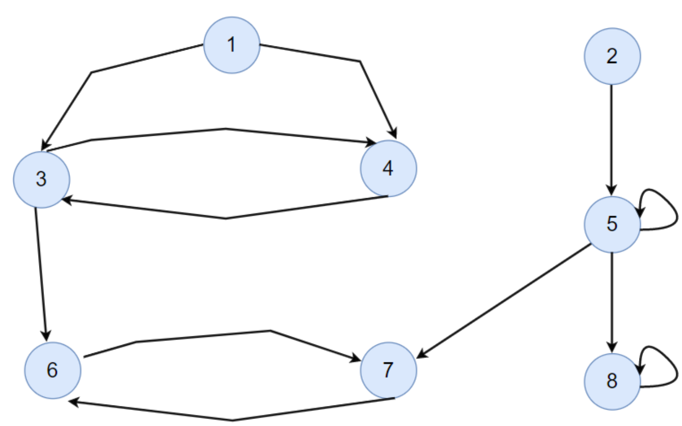



Figure 8: Σχήμα 2

Στο σχήμα παρατηρούμε ότι οι καταστάσεις:

- 1 και 2 είναι παροδικές.

  Η αλυσίδα μέσω της 1 μπορεί να περάσει στις καταστάσεις 3 και 4 όμως δεν μπορεί να γυρίσει πίσω στην 1.Οι ίδιες παρατηρήσεις ισχύουν και για την 2.Επίσης η αλυσίδα μπορεί να υπάρξει σε αυτές τις καταστάσεις μόνο στα πρώτα χρονικά βήματα λειτουργίας της.Αυτές οι καταστάσεις ονομάζονται εφήμερες

- 3 και 4 είναι επίσης παροδικές.

  Η αλυσίδα μπορεί να εναλλαχθεί μεταξύ αυτών των 2 όμως αν από την 3 επιλέξει την μετάβαση στην κατάσταση 6 δεν μπορεί να γυρίσει πίσω

- 6 και 7 είναι επαναλαμβανόμενες.

  Εάν η αλυσίδα φτάσει σε μία από αυτές κάθε δυνατή μετέπειτα μετάβαση θα είναι απο τη μία κατάσταση στην άλλη.Επίσης, παρατηρούμε πως αν η αλυσίδα είναι στην 6 επιστρέφει στην 6 κάθε δεύτερη εναλλαγή.Το ίδιο συμβαίνει και για την κατάσταση 7 ,το οποίο μας οδηγεί στο συμπέρασμα οτι αυτές είναι περιοδικές με περίοδο 2.

- 5 είναι παροδική .

  Η αλυσίδα Markov θα εισέλθει σε αυτήν την κατάσταση σε πρώτο χρόνο απο την κατάσταση 2,εάν η κατάσταση 2 είναι η αρχική κατάσταση την αλυσίδας.Με το που βρεθεί στην κατάσταση 5 η αλυσίδα Markov μπορεί να παραμείνει σε αυτήν έναν ορισμένο αριθμό χρονικών βημάτων,όμως εν τέλει θα προχωρήσει σε μία απο τις καταστάσεις 7 ή 8.

- τέλος, η 8 είναι επαναλαμβανόμενη.

  Εάν η αλυσίδα φτάσει σε εκείνη θα παραμείνει εκέι για πάντα.Σημείωνεται πως για αυτό η κατάσταση ορίζεται ως απορροφητική.

### 3.3.4 Στάσιμη κατανομή

**Το παράδειγμα του αναψυκτικού (της κολα)**

Υποθέτουμε ότι ολόκληρη η βιομηχανία cola παράγει μόνο 2 είδη. Δεδομένου ότι η τελευταία αγορά ενός ατόμου ήταν το είδος 1 υπάρχει πιθανότητα 90% το άτομο να αγοράσει ξανά το είδος αυτό. Δεδομένου όμως οτι ένα άτομο επέλεξε το είδος 2 υπάρχει πιθανότητα 80% να το αγοράσει ξανά μετέπειτα. Ερωτήσεις:

1)στην περίπτωση που ένα άτομο αγοράσει το είδος 2, ποια η πιθανότητα να αγοράσει το είδος 1 σε δυο αγορές απο τώρα;

2)στην περίπτωση που ένα άτομο αγοράσει το είδος 1 ποια η πιθανότητα να αγοράσει το είδος αυτό σε τρεις αγορες απο τώρα;

**Απαντήσεις**:

Θεωρούμε ως:

- κατάσταση 1: το άτομο αγόρασε τελευταία φορά το είδος 1
- κατάσταση 2 :το άτομο αγόρασε τελευταία φορά το είδος 2

Αν ορίσουμε ως $X_n$ να είναι το είδος κόλα που αγόρασε ένα άτομο κατά την n-ιοστή μελλοντική αγορά (παρούσα αγορά = $X_0$), τότε η $X_0, X_1, \ldots$ μπορεί να περιγραφεί ως μια Markov αλυσίδα με τον ακόλουθο πίνακα μετάβασης P.

$$\begin{array}{cc} cola1 & cola2 \end{array}$$
$$P = \begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix} \begin{array}{c} cola1 \\ cola2 \end{array}$$

1.)

Βρίσκοντας τον πινακα μετάβασης $P^2$ ο οποιος ορίζει τις πιθανότητες μεταβασης σε 2 αγορές απο τώρα.

$$P^2 = \begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix} * \begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix} = \begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix}$$

βρίσκουμε πως η απάντηση του ερωτήματος 1 είναι η πιθανότητα 0.34.Το οποίο σημαίνει πως σε δυο μελλοντικές αγορές,ένας αγοραστής του είδους 2 θα αγοράσει το είδος 1.

2.)

Στη συνέχεια ,με την ίδια λογική υπολογίζουμε τον πίνακα πιθανοτήτων μετάβασης

$$P^3 = \begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix} * \begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix} = \begin{pmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{pmatrix}$$

από όπου και προκύπτει πως η ζητούμενη πιθανότητα είναι 0.781. Με την ίδια λογική μπορούμε να υπολογίσουμε και του πίνακες για διάφορες τιμές των n.Για παράδειγμα ,βρίσκουμε πως

$$P^4 = \begin{pmatrix} 0.75 & 0.25 \\ 0.51 & 0.49 \end{pmatrix}$$

$$P^5 = \begin{pmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{pmatrix}$$

$$P^{10} = \begin{pmatrix} 0.68 & 0.32 \\ 0.65 & 0.35 \end{pmatrix}$$

$$P^{20} = \begin{pmatrix} 0.67 & 0.33 \\ 0.67 & 0.33 \end{pmatrix}$$

$$P^{30} = \begin{pmatrix} 0.67 & 0.33 \\ 0.67 & 0.33 \end{pmatrix}$$

Φαίνεται πως για μεγάλα n οι πιθανότητες το άτομο να αγοράσει την κόλα είδους 1 είναι 0.67 ανεξαρτήτως της αρχικής του κατάστασης δηλαδή αν ήταν αγοραστής είδους 1 ή 2.Ομοίως η πιθανότητα για μεγάλα n το άτομο να αγοράσει το είδος 2 προσεγγίζει το 0.33 ανεξαρτήτως αρχικής κατάστασης.

Μέσω αυτού του παραδείγματος γίνεται μία εισαγωγή στην έννοια της στάσιμης κατανομής η οποία μπορεί να χρησιμοποιηθεί για να περιγράψει την μακροπρόθεσμη συμπεριφορά των εργοδικών αλυσιδών markov.

> Ας θεωρήσουμε ως P τον πίνακα μετάβασης για μία εργοδική αλυσίδα s-καταστάσεων.Τότε,υπάρχει διάνυσμα $\pi = [\pi_1\pi_2....\pi_s]$ (steady state diastribution) τέτοιο ώστε $\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & ... & \pi_s \\ \pi_1 & \pi_2 & ... & \pi_s \\ ... & ... & ... & ... \\ \pi_1 & \pi_2 & ... & \pi_s \end{pmatrix}$ και $\lim_{n\to\infty} p_{ij}(n) = \pi_j$ Επίσης η σταθερή κατανομή ικανοποιεί τις παρακάτω συνθήκες:
>
> $$\pi_1 + \pi_2 + ... + \pi_s = 1$$
>
> $$\pi_j = \sum_{k=1}^{k=s} \pi_\kappa p_{kj}$$
>
> όπου σε μορφή πίνακα μπορεί να γραφτεί ως π=πP

Αν χρησιμοποιήσουμε τα παραπάνω μπορούμε επίσης να βρούμε την σταθερή κατανομή του παραδείγματος της κόλα ως εξής:

$$[\pi_1\pi_2] = [\pi_1\pi_2] \begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix}$$

απο όπου μαζί με την συνθήκη $\pi_1 + \pi_2 = 1$ προκύπτει το σύστηκα των τριών εξισώσεων

$$\pi_1 + \pi_2 = 1$$

$$\pi_1 = 0.9\pi_1 + 0.2\pi_2$$

$$\pi_2 = 0.1\pi_1 + 0.8\pi_2$$

του οποίου η λύση προκύπτει ως $\pi = [\pi_1 \quad \pi_2] = [2/3 \quad 1/3]$ όπως και ήταν αναμενόμενο.

# 4 Εισαγωγή στις μαρκοβιανές διαδικασίες αποφάσεων

Στο προηγούμενο κεφάλαιο ήρθαμε σε επαφή με τις μαρκοβιανές αλυσίδες διακριτού χρόνου και σχεδιάσαμε τον πίνακα μετάβασης (1ης τάξης) ο οποίος δεδομένης της τρέχουσας κατάστασης παρέχει τις πιθανότητες για τις επερχόμενες καταστάσεις στις οποίες θα μεταβεί η αλυσίδα.Σκοπός μας, στις προηγούμενες σελίδες ήταν με την απόκτηση του πίνακα πιθανότητων μετάβασης να περιγράψουμε τη συμπεριφορά της αλυσίδας ώστε να εκτιμήσουμε την αποδοση της.(π.χ βρίσκοντας την στάσιμη κατανομή της)

Τώρα είμαστε αντιμέτωποι με το ερώτημα του πως μπορούμε να **σχεδιάσουμε** την λειτουργία του συστήματος ώστε να εκτιμήσουμε την απόδοση του. Συγκεκριμένα, από δω και στο εξής ένας υπέυθνος λήψης απόφασης(decision maker ,agent) έρχεται αντιμέτωπος με το πως να επηρεάσει τη συμπεριφορά ενός πιθανολογικού συστήματος καθώς αυτό εξελίσσεται με την πάροδο του χρόνου. Αυτό το κάνει παίρνοντας αποφάσεις ή επιλέγοντας ενέργειες. Στόχος του είναι να επιλέξει μια ακολουθία ενεργειών που αναγκάζουν το σύστημα να εξελιχθεί βέλτιστα σε σχέση με ορισμένα προκαθορισμένα κριτήρια απόδοσης. Εφόσον το σύστημα που μοντελοποιούμε είναι σε εξέλιξη, η κατάσταση του συστήματος πριν από την επόμενη απόφαση εξαρτάται από τη τρέχουσα απόφαση. Κατά συνέπεια, οι αποφάσεις δεν πρέπει να λαμβάνονται χωρίς σκέψη, αλλά πρέπει να προβλέπονται οι ευκαιρίες και το κόστος (ή ανταμοιβές) που σχετίζονται με μελλοντικές καταστάσεις του συστήματος.Η διαδικασία αυτή επιλογής της βέλτιστης λειτουργίας της μαρκοβιανής αλυσίδας ονομάζεται μαρκοβιανή διαδικασία αποφάσεων.

## 4.1 Ορισμός μαρκοβιανών διαδικασιών απόφασης και μέσου αναμενόμενου κόστους)

Μία **Μ.Δ.Α(Μαρκοβιανή Διαδικασία Απόφασης)** είναι μία συλλογή $< T, S, A, p_{ij}(a_n), C(i, an) >$

όπου:

- $T$ το σύνολο των χρονικών περιόδων απόφασης**(decision epochs)**.

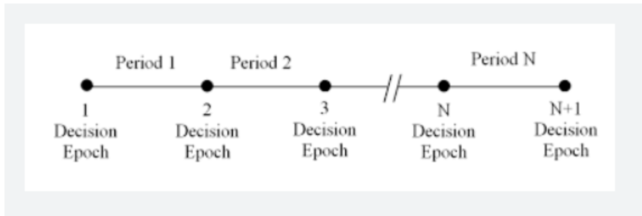
  Εν γένει αποφάσεις λαμβάνονται σε χρονικά σημέια τα οποία αναφέρονται ως χρονικοί περίοδοι απόφασης.Το σύνολο αυτό μπορεί να είναι είτε διακριτό $T = 0, 1, 2, 3..N$ είτε συνεχές $T = [0, \infty]$. Οτάν το σύνολο είναι διακριτό τότε αποφάσεις λαμβάνονται σε όλες τις χρονικές περιόδους απόφασης ενώ αντίθετα όταν είναι συνεχές αποφάσεις λαμβάνονται

  - σε όλες τιε χρονικές περιόδους απόφασης(συνεχόμενα)
  - σε τυχαία σημεία του χρόνου στα οποία συμβαίνουν γεγονότα ενδιαφέροντος(π.χ αφίξεις σε ουρές αναμονης)
  - σε επιλεγμένα χρονικά σημεία τα οποία θεωρούνται σημεία -ευκαιρία παρατήρησης.

  Επίσης το σύνολο του χρόνου μπορεί να είναι πεπερασμένο $T = 1, 2, 3, ....N$ με $N \leq \infty$ ή μη φραγμένο $T = 1, 2, ....$ .Επίσης, το χρονικό σύνολο μπορεί να είναι διάστημα με $T = [0, N]$ ή $T = [0, \infty]$.Και στις δύο περιπτώσεις όταν το N είναι πεπερασμένο ορίζουμε ένα πρόβλημα πεπερασμένου χρονικού ορίζοντα(finite horizon problem).Διαφορετικά, ορίζουμε ένα πρόβλημα αορίστου χρονικού ορίζοντα.Εν γένει, στα προβλήματα περερασμένου χρονικού ορίζοντα η τελευταία απόφαση λαμβάνεται στην χρονική περίοδο απόφασης N-1 .Παρόλα αυτά, συμπεριλαμβάνουμε την χρονική περίοδο αποφασης N λόγω υπολογισμού την τελικής καταστάσεως του συστήματος. Στην παρούσα εργασία θα επικεντρωθούμε σε μοντέλα διακριτού χρόνου .

- S ο πεπερασμένος **χώρος καταστάσεων** και A το **σύνολο αποφασεων**.

  Θεωρούμε πως σε κάποια χρονική περίοδο απόφασης το σύστημα παρατηρείται σε κάποια κατάσταση $\iota \in S$ τότε μπορούμε να επιλέξουμε μία απόφαση a από ένα σύνολο δυνατών αποφάσεων A(ι) τ.ω A=$\cup_{i \in S} A(i)$




- $p_{ij}(a_n)$ **οι πιθανότητες μετάβασης και** $c(i, a_n)$ **τα κόστη αποφάσεων** αντίστοιχα.

  Κατά την επιλογή μιας απόφασης $a \in A(i)$ στην κατάσταση ι σε κάποια χρονική περίοδο απόφασης έχουμε τα εξής αποτελέσματα:

  - .Έχουμε ένα άμεσο ή αναμενόμενο κόστος c(a) ή c(i,a)(συχνά αναφερόμαστε αντίστοιχα σε έσοδα ως αρνητικά κόστη ή σε ανταμοιβές).Το αναμενόμενο κόστος πολλές φορές εξαρτάται από την κατάσταση στην οποία θα βρεθεί το σύστημα στην επόμενη χρονικη περίοδο απόφασης όπου

  $$c(i, a) = \sum_{j \in S} c(i, a, j) p_{ij}(a) \forall i \in S$$

  - Η κατάσταση του συστήματος την επόμενη χρονική περίοδο απόφασης καθορίζεται απο την κατανομή πιθανότητας $p_{ij}(an)$ όπου $\sum_{j \in S} p_{ij}(an) = 1$

Τέλος, η μαρκοβιανή ιδιότητα αποδίδεται σε αυτό το είδος συστήματος καθώς οι πιθανότητες μετάβασης και οι συναρτήσεις κόστους εξαρτώνται μόνο απο την τρέχουσα κατάσταση και απόφαση .

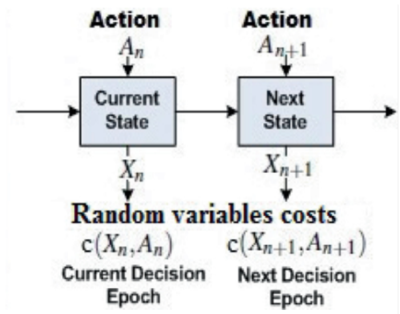



Figure 9: Evolution of the system

Στη συνέχεια, ερχόμαστε σε επαφή με ένα παράδειγμα και τους τρόπους επίλυσης του .

**Παράδειγμα**

Ένας κατασκευαστής έχει στη διάθεση του ένα μηχάνημα το οποίο παίζει βασικό ρόλο σε μια απο τις παραγωγικές διαδικασίες μιας βιομηχανίας.Το μηχάνημα αυτό λόγω χρήσης υποβαθμίζεται(σε ποιότητα και απόδοση) και έτσι κάθε τέλος μιας εβδομάδας πραγματοποιείται έλεγχος ο οποίος το κατηγοροποιεί σε μία από τις παρακάτω κατηγορίες-καταστάσεις(Figure 3).Επίσης, βάση ερευνών παλαιότερων στατιστικών δεδομένων δημιουργήθηκε ο πίνακας πιθανοτήτων μετάβασης ο οποίος φαίνεται παρακάτω και δίνει πληροφορίες για το πως εξελίσσεται το σύστημα απο μήνα σε μήνα(δίνει πιθανότητες μετάβασης απο μία κατάσταση σε ένα μήνα (γραμμή)σε μία αλλη κατάσταση τον επόμενο μήνα(στήλη).

| ΚΑΤΑΣΤΑΣΗ(STATE) | CONDITION(ΒΑΘΜΟΣ ΥΠΟΒΑΘΜΙΣΗΣ) |
|---|---|
| **0** | Καλό σαν καινούργιο |
| **1** | Λειτουργικό-μικρή αλλοίωση |
| **2** | Λειτουργικό –σημαντική αλλοίωση |
| **3** | Μη λειτουργικό |

Figure 10: καταστάσεις συστήματος

$$
P_0 = \begin{array}{c} \\ \\ \\ \\ \end{array}
\begin{array}{cc}
\begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \\
\begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array}
\end{array}
$$

Σε αυτό το σημείο είναι σημαντικό να γίνει η εξής παρατήρηση.Οι επόμενες μεταβάσεις του συστήματος εξαρτώνται μόνο απο την τρέχουσα κατάσταση στην οποία βρίσκεται το σύστημα(βάση ερευνών) και έτσι χαρακτηρίζεται απο τη σημαντική μαρκοβιανή ιδιότητα.Μπορούμε άρα πλέον να ορίσουμε την κατάσταση του συσήματος στο τέλος κάθε μήνα ως μια τυχαια μεταβλητή $X_t$ και την συλλογή αυτών $\{X_t, t = 0, 1, 2, ...\}$ (των τ.μ) ως μια μαρκοβιανή αλυσίδα διακριτού χρόνου της οποίας ο πίνακας πιθανοτήτων μετάβασης ειναι ο παραπάνω($P_0$)

Μέσω του πίνακα πιθανοτήτων μετάβασης παρατηρούμε ότι η κατάσταση 3 ειναι απορροφητική και άρα το μηχάνημα θα πρέπει να αντικατασταθεί ώστε να μην μείνει σε αυτήν την κατάσταση.

Η διδικασία αποκατάστασης επιφέρει **συνολικό κόστος 6.000$** όπου

- **4000$** το κόστος αντικατάστασης του μηχανήματος και
- **2.000$** το κόστος λόγω χαμένης παραγωγής κατά τη διάρκεια αντικατάστασης-παραγωγικής αδράνειας.

Βέβαια,και οι υπόλοιπες καταστάσεις συνδέονται με κόστη λόγω παραγωγής ελλατωματικών προιόντων ακόμα και πριν την κατάσταση 3 ως εξής:

| Κατάσταη | Αναμενόμενο κόστος λογω ελλατωματικών προιόντων/εβδομαδα($) |
|---|---|
| 0 | **0** |
| 1 | **1.000** |
| 2 | **3.000** |

Figure 11: κόστη λόγω ελαττωματικής παραγωγής/εβδομάδα

Τέλος,υπάρχει η δυνατότητα επιδιόρθωσης του μηχανήματος με κόστος 2.000$.Αυτή η ενέργεια μπορεί να πραγματοποιηθεί μόνο στην κατάσταση 2 και έχει ως αποτέλεσμα την κατηγοροποόηση του εξεταζόμενου εξαρτήματος ως λειτουργικό-μικρή αλλοίωση(επιστροφή στην κατάσταση 1).

Βάση αυτών των πληροφοριών πλέον είμαστε ενημερωμένοι για τα κόστη που συνδέονται με τις καταστάσεις του συστήματος.Θεωρώντας ως R μια πολίτικη και $di(R)$ την αντίστοιχη απόφαση έχουμε τους εξής συνδυασμούς:

| ΑΠΟΦΑΣΗ | ΕΝΕΡΓΕΙΑ | ΚΑΤΑΣΤΑΣΗ |
|---|---|---|
| 1 | Δεν επεμβαίνουμε | 0,1,2 |
| 2 | Επιδιόρθωση(επιστροφή στη κατ.1) | 2 |
| 3 | αντικατάσταση | 1,2,3 |

Figure 12: σχέσεις αποφάσεων-καταστάσεων

| ΑΠΟΦΑΣΗ | ΚΑΤΑΣΤΑΣΗ | ΑΝΑΜΕΝΟΜΕΝΟ ΚΟΣΤΟΣ ΛΟΓΩ ΕΛΑΤΤΩΜΑΤΙΚΩΝ ΠΡΟΙΟΝΤΩΝ($) | ΚΟΣΤΟΣ ΣΥΝΤΗΡΗΣΗΣ($) | ΚΟΣΤΟΣ ΛΟΓΩ ΜΗ ΠΑΡΑΓΩΓΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ($) | ΣΥΝΟΛΙΚΟ ΚΟΣΤΟΣ($) |
|---|---|---|---|---|---|
| 1. ΔΕΝ ΕΠΕΜΒΑΙΝΟΥΜΕ | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1.000 | 0 | 0 | 1.000 |
| | 2 | 3.000 | 0 | 0 | 3.000 |
| 2 .ΕΠΙΔΙΟΡΘΩΣΗ | 2 | 0 | 2.000 | 2.000 | 4.000 |
| 3.ΑΝΤΙΚΑΤΑΣΤΑΣΗ | 1,2,3 | 0 | 4.000 | 2.000 | 6.000 |

Figure 13: Συνολικές σχέσεις κόστους

.Έτσι το άτομο το οποίο ερευνεί την βέλτιστη πολιτική(decision maker) του συστήματος θα μπορούμε να ελεγξει την πολιτικες :

- **Ra: Αντικατάσταση στην κατάσταση 3 και καμία άλλη ενέργεια συντήρησης**

  Υπό αυτήν την πολιτική το σύστημα εξακολουθεί να αποτελεί μια μαρκοβιανή διαδικασία της οποίας ο πίνακας μετάβασης είναι :

$$P_a = \begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \\ \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}$$

- **Rb:Αντικατάσταση στην κατάσταση 3 και επιδιόρθωση στην κατάσταση 2** με πίνακα πιθανοτήτων μετάβασης

$$P_b = \begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \\ \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}$$

- **Rc:Αντικατάσταση στην κατάσταση 2 και 3** με πίνακα πιθανοτήτων μετάβασης

$$P_c = \begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \\ \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}$$

- **Rd:Αντικατάσταση στις καταστάσεις 1,2,3** με πίνακα πιθανοτήτων μετάβασης

$$P_d = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cccc} 0 & 1 & 2 & 3 \end{array}$$

$$P_d = \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array}$$

Παρατηρούμε ότι εδώ τόσο οι πιθανότητες μετάβασης, όσο και τα κόστη δεν εξαρτώνται από το χρονικό βήμα,οπότε η εν λόγω αλυσίδα είναι χρονικά ομογενής

## 4.2 Επίλυση με τη μέθοδο exhaustive enumeration

Στη διαδικασία αυτή αρχικά υπολογίζουμε τα ανεμενόμενα μακροπρόθεσμα μέσα κόστη ως προαναφέρθηκαν

$$E(C) = \sum_{i=0}^{M} c_{ik}\pi_i$$

όπου $k = d_i(R)$ για κάθε μία από τις τεσσερίς πολιτικές και τελικά επιλέγουμε την πολιτική εκείνη που συνδέεται με το ελάχιστο αναμενόμενο κόστος ως εξής:
Οσόν αφορά την πολιτική $R_a$ ισχύει ότι
$d_0(Ra) = 1, d_1(Ra) = 1, d_2(Ra) = 1, d_3(Ra) = 3$ και άρα

$$E(C) = C_{01}\pi_0 + C_{11}\pi_1 + C_{21}\pi_2 + C_{33}\pi_3 =$$

$$= 0\pi 0 + 1.000\pi_2 + 3.000\pi_2 + 6.000\pi 3$$

όπου $\pi = (\pi_0, \pi_1, \pi_2, \pi_3)$ αποτελεί την σταθερή κατανομη του συστήματος στις καταστάσεις $i = 0, 1, 2, 3$ για την πολιτική $R_a$ και ικανοποιεί τις εξισώσεις(2.3.4):

$$\begin{cases} \pi_0 + \pi_1 + \pi_2 + \pi_3 = 1 \\ \pi_0 = \pi_3 \\ \pi_1 = \frac{7}{8}\pi_0 + \frac{3}{4}\pi_1 \\ \pi_2 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_3 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_3 \end{cases} . \tag{7}$$

| Πολιτική | Στάσιμη κατανομή(π0,π1,π2,π3) | Υπολογισμός | Αναμενόμενο κόστος($)/ανα εβδομάδα | |
|---|---|---|---|---|
| Ra | 2/13, 7/13, 2/13, 2/13 | 2/13(0)+7/13(1.000)+2/13(3.000)+2/13(6.000) | 1.923 | |
| Rb | 2/21, 5/7, 2/21, 2/21 | 2/21(0)+5/7(1.000)+2/21(4.000)+2/21(6.000) | 1.667 | minimum |
| Rc | 2/11, 7/11, 1/11, 1/11 | 2/11(0)+7/11(1.000)+1/11(6.000)+1/11(6.000) | 1.727 | |
| Rd | 1/2, 7/16, 1/32, 1/32 | 1/2(0)+7/16(6.000)+1/32(6.000)+1/32(6.000) | 3.000 | |

Figure 14: optimal policy by exhaustive enumeration

των οποίων η λύση είναι $\pi_0 = \frac{2}{3}, \pi_1 = \frac{7}{13}, \pi_2 = \frac{2}{13}, \pi_3 = \frac{2}{13}$ και άρα το αναμενομενο κόστος υπολογίζεται σε 1.923,08\$. Συνεχίζοντας την ίδια διαδικασία για τις υπόλοιπες πολιτικές λαμβάνουμε τον παρακάτω πίνακα

από όπου προκύπτει πως η βέλτιστη πολιτική είναι η Rb:Αντικατάσταση στην κατάσταση 3 και επιδιόρθωση στην κατάσταση 2 με μέσο αναμενόμενο κόστος ισο με 1.667\$

Σε αυτό το σημείο είναι σημαντικές δύο παρατηρήσεις. Η πρώτη αφορά ότι η μαρκοβιανή αλυσίδα του παραδείγματος είναι στάσιμη (ανεξάρτητη του χρόνου)και η δεύτερη ότι η πολιτική του παραδείγματος είναι ντετερμινιστική. δηλ. όποτε το σύστημα βρίσκεται στην κατάσταση i, ο κανόνας για την λήψη η απόφαση επιλέγει σίγουρα μια συγκεκριμένη απόφαση. (Λόγω της φύσης του εμπλεκόμενου αλγορίθμου, η επόμενη ενότητα εξετάζει τυχαιοποιημένες πολιτικές, όπου χρησιμοποιείται μια κατανομή πιθανοτήτων για τη λήψη της απόφασης.)

## 4.3 Κατηγοριοποίηση πολιτικών

> Ορίζουμε ως πολιτική (ή στρατηγική) τους κανόνες ή τρόπους λήψης απόφασεων που χρησιμοποιήθηκαν σε όλες τις χρονικές περιόδους απόφασης του συστήματος.

Ένας κανόνας απόφασης περιγράφει τη διαδικάσια για την επιλογή απόφασης σε κάθε κατάσταση σε μια συγκεκριμένη χρονική περίοδο απόφασης.Έχουμε τις εξής κατηγορίες:

- **ιστορικά εξαρτώμενος και τυχαιοπηοιμένος(HR)**:

Στον οποίο όταν $\chi_n = i$ η λήψη της απόφασης $\alpha \in A_i$ εξαρτάται και από την παρελθούσα εξέλιξη $h_n$ της διαδικασίας και των αντίστοιχων αποφάσεων ακολουθώντας την κατανομή $q_n(h_n, a), a \in A(i), i = s_n$

- **ιστορικά εξαρτώμενος και ντετερμινιστικός(HD)**:

  Εδώ ο τρόπος λήψης απόφασης εξαρτάται από την παρελθούσα ιστορία,οπότε αποτελεί μία συνάρτηση της $h_n = (s1, a1, \ldots, s_{n-1}, a_{n-1}, s_n)$, όπου $s_k = x_k, a_k \in A(s_k)$, k =1,2, ... n, η κατάσταση και η ληφθείσα απόφαση την χρονική στιγμή k. Όταν $x_n = s_n = i$, η απόφαση $a_n = d_n$ λαμβάνεται από το σύνολο δυνατών αποφάσεων $A(S_n)$.Ως εκ τούτου έχουμε $d_n : S \times A \times S \times A \times S \to A, h_n \to d_n(h_n) \in A_(s_n)$

- **Μαρκοβιανός και τυχαιοποιημένος (MR)**: Κατά τον τυχαιοποιημένο μαρκοβιανό κανόνα λήψης αποφάσεων όταν $x_n = i$ η λήψη της απόφασης $\alpha \in A(i)$ γίνεται βάση μιας κατανομής πιθανότητας $q_n(i, a)$

- **Μαρκοβιανός και ντετερμινιτστικος(MD)**: Εδώ ο τρόπος λήψης της απόφασης είναι μία συνάρτηση $d_n(i) : S \to A(i)$ ο οποίος καθορίζει την απόφαση του συστήματος όταν βρίσκεται στην κατάσταση ι στην χρονική περίοδο απόφασης n.Είναι μαρκοβιανός καθώς εξαρτάται απο το παρελθόν μόνο μέσω της τρέχουσας κατάστασης του συστήματος και ντετερμινιστικός καθώς αποφασίζει με πιθανότητα 1.$(\alpha_n = d_n(i))$

Στο προηγούμενο παράδειγμα περιγράψαμε το βασικό είδος πολιτικής (στασιμη,ντετερμινιστική).Η πολιτική R μπορεί να περιγραφεί αποδίδοντας τις τιμές $D_{ik}=0$ ή 1 στον πίνακα

$$P=\begin{array}{c} \\ \\ \end{array}\begin{array}{cccc} 0 & 1 & \dots & M \end{array}$$

$$P=\begin{pmatrix} D_{01} & D_{02} & \dots & D_{0k} \\ D_{11} & D_{12} & \dots & D_{1k} \\ \dots & \dots & \dots & \dots \\ D_{M1} & D_{M2} & \dots & D_{MK} \end{pmatrix}\begin{array}{c} 0 \\ 1 \\ \dots \\ M \end{array}$$

όπου κάθε $D_{ik}(i=0,1,...,M)$ και $k=0,1,\dots k)$ ορίζεται ως

$$D_{ik}=\begin{cases} 1, & \text{εαν παρθηκε αποφαση k στην i} \\ 0, & \text{αλλιως} \end{cases}.$$

Έτσι, για παράδειγμα η πολιτική Rb μπορεί να περιγραφείι ως τον πίνακα

$$Pb=\begin{array}{c} \\ \end{array}\begin{array}{ccc} 1 & 2 & 3 \end{array}$$

$$Pb=\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array}$$

Βάση αυτής της επιλογής δίνεται κίνητρο για εισαγωγή της τυχαιοποιημένης μαρκοβιανής πολιτικής η οποία είναι απαραίτητη για τον γραμμικό προγραμματισμό του προβλήματος μας.Πλέον ,κάθε φορά που βρισκόμαστε στην κατάσταση i ορίζουμε μια κατανομή πιθανότητας για την απόφαση που θα παρθεί ως:

$$D_{ik}=P\{αποφαση=k/κατασταση=i\}$$

οπού μπορεί να περιγραφεί με τον ίδιο πίνακα αλλά αυτήν την φορά $0\leq D_{ik}\leq 1$

## 4.4 Ο γραμμικός προγραμματισμός(linear programming formulation)

Ορίζουμε ως $y_{ik}$ τις μεταβλητές απόφασης όπου

$$y_{ik}\text{=P\{απόφαση=k και κατάσταση =i\}}$$

$\forall i=0,1,\dots,M$ και $k=1,2,\dots,k$ για την οποία ισχύει ότι $y_{ik}=\pi_i D_{ik}$ και $\pi_i=\sum_{k=1}^{k}y_{ik}$ και άρα

$D_{ik} = \frac{y_{ik}}{\pi_i} = \frac{y_{ik}}{\sum_{k=1}^{k} y_{ik}}$.(1) Για τις μεταβλητές απόφασης ισχύουν οι περιορισμοί:

- $\sum_{i=1}^{M} \pi_i = 1$ έτσι ώστε $\sum_{i=0}^{M} \sum_{k=1}^{k} y_{ik} = 1$
- $\sum_{k=1}^{k} y_{jk} - \sum_{i=0}^{M} \sum_{k=1}^{k} y_{ik} p_{ij}(k) = 0$ για $j = 0, 1, \ldots, M$
  καθώς γνωρίζουμε πως $\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}$
- $y_{ik} \geq 0$ για i=0,1,...,M και k=0,1,...,k

Στόχος της μεθόδου είναι η ελαχιστοποίηση της αντικειμενικής συνάρτησης

$$minz = \sum_{i=0}^{M} \sum_{k=1}^{k} c_{ik} y_{ik}$$

καθώς αντιπροσωπεύει το(μακροπρόθεσμο) μεσο αναμενόμενο κόστος ανα μονάδα χρόνου υπό τους προαναφερόρθέντες περιορισμούς.

**Λύση του προβλήματος (3.1) με γραμμικό προγραμματισμό**

Από τον πίνακα για τις σχέσεις αποφάσεων καταστάσεων και κόστεων και βρίσκουμε ποιες είναι οι μεταβλητές οι οποίες θα συμμετέχουν στην αντικειμενική συνάρτηση και ποιες μηδενίζονται.Άρα

$$Z = 1.000 y_{11} + 3.000 y_{21} + 4.000_{y22} + 6.000 y_{13} + 6.000 y_{23} + 6.000 y_{33}$$

υπό τους περιορισμούς

$$y_{01} + y_{11} + y_{13} + y_{21} + y_{22} + y_{23} + y_{33} = 1$$

$$y_{01} - (y_{13} + y_{23} + y_{33}) = 0$$

$$y_{11} + y_{13} - (\frac{7}{8} y_{01} + \frac{3}{4} y_{11} + y_{22}) = 0$$

$$y_{21} + y_{22} + y_{23} - (\frac{1}{16} y_{01} + \frac{1}{8} y_{11} + \frac{1}{2} y_{21}) = 0$$

$$y_{33} - (\frac{1}{16} y_{01+} \frac{1}{8} y_{11} + \frac{1}{2} y_{21}) = 0$$

και όλα τα

$$y_{ik} \geq 0$$

εφαρμόζοντας την μέθοδο simplex λαμβάνουμε την βέλτιστη λύση

$$y_{01} = \frac{2}{21}$$

$$(y_{11}, y_{13}) = (\frac{5}{7}, 0)$$

$$(y_{21}, y_{22}, y_{23}) = (0, \frac{2}{21}, 0)$$

$$y_{33} = (\frac{2}{21})$$

και άρα απο (1)έχουμε

- $D_{01} = 1$,δηλ. στην κατάσταση 0 να παρθεί η απόφαση 1(δεν επεμβαίνουμε)
- $(D_{11}, D_{13}) = (1, 0)$, δηλ.στην κατάσταση 1 να παρθεί η απόφαση 1 ,
- $(D_{21}, D_{22}, D_{23}) = (0, 1, 0)$, δηλ.στην κατάσταση 2 να παρθεί η απόφαση 2 και
- $D_{33} = 1$, δηλ.στην κατάσταση 3 να παρθεί η απόφαση 3

Αυτή είναι η ίδια απάντηση που πήραμε εφαρμόζοντας την μέθοδο της exhaustive enumeration.

Το βασικό συμπέρασμα είναι ότι η βέλτιστη πολιτική που βρέθηκε με τη μέθοδο simplex είναι ντετερμινιστική και όχι τυχαίοποιημένη. Επομένως, το να επιτρέπεται η τυχαιοποίηση των πολιτικών δεν βοηθά στη βελτίωση της τελικής πολιτικής όμως στη μετατροπή ακέραιων μεταβλητών (το Dik) σε συνεχείς μεταβλητές, έτσι ώστε να μπορεί να χρησιμοποιηθεί ο γραμμικός προγραμματισμός .

## 4.5 Ο αλγόριθμος βελτίωσης της πολιτικής(policy improvement algorithm)

**Εισαγωγή στον αλγόριθμο και μαθηματικά εργαλεία**

Θεωρώ ως R μια σταθερή ντετερμινιστική πολιτική και $d_i(R) = k$ η απόφαση στην κατάσταση i.Εώς τώρα ορίσαμε ως $c_{ik}$ το αναμενόμενο κόστος το οποίο εξαρτάται μόνο μόνο απο την παρατηρειθήσα κατάσταση και την αντίστοιχη απόφαση που λαμβάνεται.Το κόστος όμως μπορεί να εξαρτάται και απο την κατάσταση στην οποία μεταβαίνει το σύστημα και τότε γράφουμε

$$c_{ik} = \sum_{j=0}^{M} q_{ij}(k) p_{ij}(k)$$

όπου $p_{ij}(k)$ η πιθανότητα μετάβασης και $q_{ij}(k)$ το αναμενόμενο κόστος που προκύπτει όταν το σύστημα βρίσκεται στην κατάσταση i, λαμβάνεται η απόφαση k και το σύστημα μεταβαίνει στην κατάσταση j την επόμενη χρονική περίοδο. Ορίζουμε ως **συνολικό αναμενόμενο κόστος** ενός συστήματος το οποίο ξεκίνησε στην κατάσταση i κατα την πρώτη χρονική περίοδο παρατήρησης και εξελίχθηκε για n χρονικές περιόδους ως $V_i^n(R)$. Αυτό αποτελείται από

- το $c_{ik}$ δηλαδή το κόστος που προέκυψε κατα τη διάρκεια της πρώτης χρονικής περιόδου παρατήρησης και
- $\sum_{j=0}^{M} p_{ij}(k) V_j^{n-1}(R)$ το αναμενόμενο κόστος που προέκυψε κατά την διάρκεια των υπόλοιπων n-1 χρονικών περιόδων.

Άρα

$$V_i^n(R) = c_{ik} + \sum_{j=0}^{M} p_{ij}(k) V_j^{n-1}(R_n) \tag{8}$$

όπου προφανώς $V_i^1(R) = c_{ik}$.

Επίσης όπως έχει προαναφερθεί το **(μακροπρόθεσμο) αναμενόμενο μέσο κόστος ανά μονάδα χρόνου μετά** από οποιαδήποτε πολιτική R εκφράζεται ως:$g(R) = \sum_{i=0}^{M} \pi_\iota c_{ik}$το οποίο είναι ανεξάρτητο της αρχικής κατάστασης i. Σε αυτό το σημείο είναι χρήσιμο να παρατηρήσουμε τη συμπεριφορά του $V_i^n(R)$ καθώς το $n \to \infty$.Το συνολικό κόστος τότε προσεγγίζει το ng(R).Συγκεκριμένα,

$$V_i^n(R) \sim ng(R) + v_i R$$

όπου καθώς η πρώτη συνιστώσα είναι ανεξάρτητη της αρχικής κατάστασης το $v_i R$ ερμηνέυεται ως την επιρροή της αρχικής κατάστασης i. Τότε η ποσότητα $v_i^n - v_j^n \sim v_i - v_j$ εκφράζει την διαφορά στο μέσο συνολικό κόστος,αν η διαδικασία ξεκινήσει από την κατάσταση i σε σύγκριση με το αν ξεκινούσε από την κατάσταση j , όταν εφαρμόζεται η πολιτική R . Ισοδύναμα, η διαφορά αυτή ουσιαστικά αποτελεί το μέγιστο ποσό που συμφέρει κάποιον να διαθέσει,ώστε το σύστημα (η αλυσίδα) να ξεκινήσει από την κατάσταση j παρά από την i κάτω από την πολιτική R .Τέλος,εφόσον $V_j^{n-1}(R) = (n-1)g(R) + v_j(R)$με αντικατάσταση στην επαναληπτική σχέση προκύπτει πως

$$g(R) + v_i(R) = c_{ik} + \sum_{j=0}^{M} p_{ij}(k) v_j(R_n)$$

για $i = 0, 1, 2, \ldots, M$ Παρατηρούμε ότι στο παραπάνω σύστημα υπάρχουν οι τιμές που το ικανοποιούν. Συνεπώς, το σύστημα έχει M + 1 εξισώσεις, με M + 2 άγνωστες τιμές έτσι ώστε μία από αυτές να επιλεχθεί αυθαίρετα, και το $M(R)$ να επιλεχθεί ίσο με το μηδέν. Η εύρεση του (μακροπρόθεσμου) αναμενόμενου μέσου κόστους g(R) προκύπτει από την επίλυση του συστήματος των γραμμικών εξισώσεων, όταν ακολουθείται η πολιτική R. Η εξαγωγή της βέλτιστης πολιτικής που ελαχιστοποιεί το g(R) θα πραγματοποιηθεί αποτελεσματικά με τον αλγόριθμο βελτίωσης των πολιτικών.

Ο αλγόριθμος ξεκινάει επιλέγοντας μία αυθαίρετη πολιτική $R_1$.Τότε λύνει το σύστημα εξισώσεων βρίσκοντας τα $g(R_1), v_0(R_1), v_1 R_1 \ldots, v_{m-1}(R_1)$ όπου θεωρούμε $v_M(R_1) = 0$.Αυτό το βήμα ονομάζεται ορισμός τιμών.Στην συνέχεια και αν η πρώτη πολιτική δεν είναι και η βέλτιστη κατασκευάζεται μια καλύτερη πολιτική $R_2$.Αυτό το βήμα ονομάζεται βελτίωση της πολιτικής.Ουσιαστικά αυτά είναι και τα δυο βήματα του αλγορίθμου τα οποία εναλλάσονται εως ότου 2 διαδοχικές επαναλήψεις να να οδηγούν σε 2 πανομοιότυπες πολιτικές δηλαδή την βέλτιστη.

**Σύνοψη του αλγορίθμου βελτίωσης της πολιτικής**

Αρχικοποίση:Διαλέγω τυχαία αρχική δοκιμαστική πολιτική $R_1$.Θέτω n=1

- Βήμα 1: Αρχικοποίηση τιμών.

  Για την πολιτική $R_n$ χρησιμοποιούμε $p_{ij}(k), c_{ik}$ και $v_M(R_n)$ για να λύσουμε το σύστημα των M+1 εξισώσεων

  $$g(R) = c_{ik} + \sum_{j=0}^{M} p_{ij}(k) v_j(R_n) - v_i(R_n)$$

  για $i = 0, 1, 2, \ldots, M$

- Βήμα 2:

  Χρησιμοποιώντας τις τιμές των $v_i(R_n)$ οι οποίες υπολογίστηκαν στο προηγούμενο βήμα για την πολιτική $R_n$ βρίσκω την εναλλακτική πολιτική $R_{n+1}$όπου για κάθε κατάσταση i,$d_i(R_{n+1})$είναι η απόφαση η οποία ελαχιστοποιεί το $c_{ik} + \sum_{j=0}^{M} p_{ij}(k) v_j(R_n) - v_i(R_n)$

Δηλαδή για κάθε κατάσταση i θέλουμε:

$$Minimize_{k=1,2,\ldots,k}[c_{ik} + \sum_{j=0}^{M} p_{ij}(k)v_j(R_n) - v_i(R_n)]$$

και τότε θέτουμε $d_i(R_{n+1})$ ίσο με την παραπάνω τιμή ελαχιστοποίησης του k.

*Δοκιμή βελτιστοποίσης.*Η τρέχουσα πολιτική $R_{n+1}$ θα είναι η βέλτιστη εάν είναι ίδια με την πολιτική $R_n$. Εάν είναι ίδιες, τότε ο αλγόριθμος σταματά. Διαφορετικά, θέτουμε n=n+1 και εκτελούμε άλλη μία επανάληψη.

**Λύση του παραδείγματος με τον αλγόριθμο βελτίωσης της πολιτικής**

Αρχικοποίηση:Επιλέγουμε μια αυθαίρετη πολιτική $R_1$.Εστω ότι η πολιτική επιλογή μας είναι η **Ra:Αντικαθιστούμε το μηχάνημα στην κατάσταση 3 καμία άλλη ενέργεια συντήρησης** η οποία υπενθυμίζουμε ότι συνδέεται με τον πίνακα πιθανοτήτων μετάβασης

$$P_a = \begin{array}{c} \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}$$

και

| **κατάσταση** | κόστος |
|---|---|
| **0** | 0 |
| **1** | 1.000 |
| **2** | 3.000 |
| **3** | 6.000 |

**Επανάληψη 1η**

- Βήμα 1: Θέτουμε $v_3(R_1) = 0$ και έχουμε τις 4 εξισώσεις:

$$g(R_1) = \frac{7}{8}v_1(R_1) + \frac{1}{16}v_2(R_1) - v_0(R_1) \tag{9}$$

$$g(R_1) = 1.000 + \frac{3}{4}v_1(R_1) + \frac{1}{8}v_2(R_1) - v_1(R_1) \tag{10}$$

$$g(R_1) = 3.000 + \frac{1}{2}v_2(R_1) - v_2(R_1) \tag{11}$$

$$g(R_1) = 6.000 + v_0(R_1) \tag{12}$$

των οποίων η λύση είναι:

$$g(R_1) = \frac{25.000}{13} = 1.923$$

$$v_0(R_1) = -4.077$$

$$v_1(R_1) = -2.615$$

$$v_2(R_1) = 2.154$$

- Βήμα 2:

  Θέλουμε να βρούμε μια βελτιωμένη πολιτική $R_2$ τ.ω η απόφαση k στην κατάσταση i να ελαχιστοποιεί την παρακάτω έκφραση:

  - κατάσταση 0:

  $$c_{0k} + p_{00}V_0(R_1) + p_{01}v_1(R_1) + p_{02}v_2(R_1) + p_{03}v_3(R_1) - v_0 =$$
  $$c_{0k} - p_{00}(4.077) - p_{01}(2.615) + p_{02}(2.154) + 4.077 \tag{13}$$

  - κατάσταση 1:

  $$c_{1k} + p_{10}V_0(R_1) + p_{11}v_1(R_1) + p_{12}v_2(R_1) + p_{13}v_3(R_1) - v_1(R_1) =$$
  $$c_{1k} - p_{10}(4.077) - p_{11}(2.615) + p_{12}(2.154) + 2.615 \tag{14}$$

  - κατάσταση 2 :

  $$c_{2k} + p_{20}V_0(R_1) + p_{21}v_1(R_1) + p_{22}v_2(R_1) + p_{23}v_3(R_1) - v_2(R_1) =$$
  $$c_{2k} - p_{20}(4.077) - p_{21}(2.615) + p_{22}(2.154) - 2.154 \tag{15}$$

  - κατάσταση 3:

  $$c_{3k} + p_{30}V_0(R_1) + p_{31}v_1(R_1) + p_{32}v_2(R_1) + p_{33}v_3(R_1) - v_3(R_1) =$$
  $$c_{3k} - p_{30}(4.077) - p_{31}(2.615) + p_{32}(2.154) \tag{16}$$

Στην πραγματικότητα η διαδικασία ελαχιστοποιήσης του παραπάνω συστήματος μπορεί να συμβεί μόνο για τις καταστάσεις 1 και 2 καθώς η κατάσταση 0 συνδέεται μόνο με την απόφαση 1 και η κατάσταση 3 μόνο με την από απόφαση 3. Για την απόφαση που ελαχιστοποιεί την έκφραση της κατάστασης 1: Αυτή συνδέεται μόνο με τις καταστάσεις 1 και 3 και άρα βάση του παραπάνω πίνακα με τις αντίστοιχες τιμές πιθανοτήτων και κόστων.
Άρα έχουμε τιμή της αρχικής έκφρασης (14) ως

| απόφαση | $C_{1k}$ | $p_{10}(k)$ | $p_{11}(k)$ | $p_{12}(k)$ | $p_{13}(k)$ |
|---|---|---|---|---|---|
| 1 | 1.000 | 0 | 3/4 | 1/8 | 1/8 |
| 3 | 6.000 | 1 | 0 | 0 | 0 |

Figure 15: κατάσταση 1 σε σχέση με αποφάσεις 1 και 3

- κατάσταση 1 και απόφαση 1: 1.923 $\leftarrow$ **Minimum**
- κατάσταση 1 και απόφαση 3: 4.538

Αντίστοιχα για την κατάσταση 2 η οποία συνδέεται με τις καταστάσεις 1,2,3 γνωρίζουμε ότι:

| απόφαση | $C_{2k}$ | $p_{20}(k)$ | $p_{21}(k)$ | $p_{22}(k)$ | $p_{23}(k)$ |
|---|---|---|---|---|---|
| 1 | 3.000 | 0 | 0 | 1/2 | 1/2 |
| 2 | 4.000 | 0 | 1 | 0 | 0 |
| 3 | 6.000 | 1 | 0 | 0 | 0 |

Figure 16: κατάσταση 2 σε σχέση με αποφάσεις 1,2 και 3

και έτσι με αντικατάσταση στη σχέση (15) λαμβάνουμε τις τιμές της έκφρασης αντίστοιχα

- κατάσταση 2 και απόφαση 1: 1.923
- κατάσταση 2 και απόφαση 2: $-769$ $\leftarrow$ **Minimum**
- κατάσταση 2 και απόφαση 3: $-231$

**ΠΟΛΙΤΙΚΗ R2**

| ΚΑΤΑΣΤΑΣΗ | ΑΠΟΦΑΣΗ |
|---|---|
| **0** | 1 |
| **1** | 1 |
| **2** | 2 |
| **3** | 3 |

Το δεύτερο βήμα του αλγορίθμου για την πρώτη επανάληψη έδειξε ότι: για την κατάσταση 1 λαμβάνουμε απόφαση 1 και για την κατάσταση 2 απόφαση 2.Άρα συνοψίζοντας:
με πίνακα πιθανοτήτων μετάβασης :

$$P_b = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \\ \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}$$

και κόστη Επίσης,εφόσον οι δυο πολιτικές $R_1$ και $R_2$ δεν ταυτίζονται προ-

| ΚΑΤΑΣΤΑΣΗ | ΚΟΣΤΟΣ |
|---|---|
| **0** | 0 |
| **1** | 1.000 |
| **2** | 4.000 |
| **3** | 6.000 |

χωράμε ακομα μια επανάληψη.

**Επανάληψη 2η**

Βήμα 1:το αντίστοιχο σύστημα υπολογίζεται ως:

$$g(R_2) = \frac{7}{8}v_1(R_2) + \frac{1}{16}v_2(R_2) - v_0(R_2) \tag{17}$$

$$g(R_2) = 1.000 + \frac{3}{4}v_1(R_2) + \frac{1}{8}v_2(R_2) - v_1(R_2) \tag{18}$$

$$g(R_2) = 4.000 + v_1(R_2) - v_2(R_2) \tag{19}$$

$$g(R_2) = 6.000 + v_0(R_2) \tag{20}$$

π΄που η ταυτόχρονη λύση αυτή τη φορά δίνεται ως:

$$g(R_2) = \frac{5.000}{3} = 1.667$$
$$v_0(R_2) = -4.333$$
$$v_1(R_2) = -3.000$$
$$v_2(R_2) = -667$$

Βήμα 2: Για τις καταστάσεις που συνδέονται με περισσότερες από μια αποφάσεις υπολογίζουμε

- κατάσταση 1:

$$c_{1k} - p_{10}(4.333) - p_{11}(3.000) - p_{12}(667) + 3.000 \tag{21}$$

- κατάσταση 2:

$$c_{2k} - p_{20}(4.333) - p_{21}(3.000) + p_{22}(667) + 667 \tag{22}$$

Έτσι, με ανάλογο τρόπο με αυτόν του βήματος 2 για την πρώτη επανάληψη λαμβάνουμε:

- κατάσταση 1 και απόφαση 1: $1.667 \leftarrow$ **Minimum**
- κατάσταση 1 και απόφαση 3: $4.667$

και

- κατάσταση 2 και απόφαση 1: $3.333$
- κατάσταση 2 και απόφαση 2: $1.667 \leftarrow$ **Minimum**
- κατάσταση 2 και απόφαση 32.334

Δηλαδή η πολιτική $R_3$
Παρατηρούμε πως ο αλγοριθμός σε δυο επαναλήψεις κατέληξε σε δυο πολιτικές που ταυτίζονται και άρα αυτή είναι και η βέλτιστη.

| ΠΟΛΙΤΙΚΗ R3 | |
|---|---|
| **ΚΑΤΑΣΤΑΣΗ** | ΑΠΟΦΑΣΗ |
| **0** | 1 |
| **1** | 1 |
| **2** | 2 |
| **3** | 3 |

Figure 17: πολιτική R3

# 5 Το κριτήριο του κόστους προεξόφλησης και προσαρμογή των αλγορίθμων βάση αυτού

Το κριτήριο το οποίο επιλέξαμε στο προηγούμενο κεφάλαιο για την εύρεση της βέτιστης πολιτικής ήταν η ελαχιστοποίηση του μέσου αναμενόμενου κόστους ανα μονάδα χρόνου .Παρολά αυτά,το σύνολο του χρόνου στα μοντέλα βελτιστοποιήσης μας όπως είδαμε μπορούν να πάρουν και εκτεταμένες τιμές το οποίο είναι δύσκολο να διαχειριστεί.Τότε, το σύνολο θεωρείται άπειρο και βέβαια αυτό έχει ως αποτέλεσμα το συνολικό κόστος να είναι μη φραγμένο. Για την αντιμετώπιση αυτού του προβλήματος ορίζουμε ένας παράγοντα προεξόφλησης $0 \leq \alpha \leq 1$ το οποίο εκφράζει πως το κέρδος( ή κόστος) το οποίο θα πραγματοποιηθεί στο μέλλον αξίζει λιγότερο από την ανταμοιβή που θα πραγματοποιούταν σήμερα.Συγκεκριμένα, θεωρούμε ως M το μέγιστο κέρδος σε όλες τις πιθανές καταστάσεις και αποφάσεις που μποιρεί να ληφθεί σε μια περίοδο.Τοτε, το συνολικό μέγιστο κέρδος σε άπειρο χρονικό ορίζοντα υπολογίζεται ως

$$M + Ma + Ma^2 + \cdots = \frac{M}{1-a} \leq \infty$$

## 5.1 Γραμμικός προγραμματισμός(για το κόστος προεξόφλησης)

Ο γραμμικός προγραμματισμός για το κόστος προεξόφλησης είναι όμοιος με εκείνον για το μέσο αναμενόμενο κόστος ανά μονάδα χρόνου μόνο που σε αυτήν την περίπτωση ο πρώτος περιορισμός δεν είναι απαραίτητος ενώ οι περιορισμοί μας θα πρέπει να περιέχουν τον παράγοντα προεξόφλησης a.Μία ακόμα διαφορά είναι ότι για την εφαρμογή του μοντέλου μας δεν χρειάζεται ο περιορισμός της αλυσίδας να είναι μη υποβιβάσιμη όπως στο μέσο αναμενέμενο κόστος ανα μονάδα χρόνου.

Τέλος, ο γραμμικός προγραμματισμός θα πρέπει να περιέχει σταθερές $\beta_j, j = 0, 1, \ldots, M$ οι οποίες επιλέγονται εν γένει αυθαίρετα αλλά θα πρέπει να ικανοποιούν με τη σειρά τους τους περιορισμούς:

$$\sum_{j=0}^{M} \beta_j = 1$$

και

$$\beta_j \geq 0, j = 0, 1, \ldots, M$$

Το τελικό μοντέλο είναι:

Ορίζουμε ως $y_{ik}$ τις μεταβλητές απόφασης .Στόχος της μεθόδου είναι η ελαχιστοποίηση της αντικειμενικής συνάρτησης

$$minz = \sum_{i=0}^{M} \sum_{k=1}^{k} c_{ik} y_{ik}$$

υπό τους περιορισμούς:

- $\sum_{k=1}^{k} y_{jk} - a \sum_{i=0}^{M} \sum_{k=1}^{k} y_{ik} p_{ij}(k) = \beta_j$ για $j = 0, 1, \ldots, M$
- $y_{ik} \geq 0$ για i=0,1,...,M και k=0,1,...,k

Τα $y_{ik}$ ερμηνεύονται ως ο προεξοφληθείς αναμενόμενος χρόνος για όπου το σύστημα όντας στην κατάσταση i λαμβάνεται η απόφαση k όταν η κατανομή της αρχικής κατάστασης είναι $Px_0 = j = \beta_j για j = 0, 1, 2, \ldots, M$ Συγκεκριμένα, $z_{ik}^n = P$[σε χρόνο n,κατάσταση i και απόφαση k] και άρα οι μεταβήτές αποφάσης είναι

$$y_{ik} = zik^0 + azik^1 + a^2 zik^2 + a^3 zik^3 + \ldots$$

Βάση των παραπάνω το Z μπορεί να ορισθεί ως το αναμενόμενο συνολικό κόστος προεξόφλησης .Η επιλογή των $\beta_j$ επηρεάζουν την τιμή του όμως όχι την τιμή της βέλτιστης πολιτικής.

Στο τέλος της μεθόδου και εφόδον έχει βρεθεί η βέλτιστη λύση του μοντέλου μέσω της μεθόδου Simplex υπολογίζονται οι τιμές $D_{ik}$=P[απόφαση =k και κατάσταση=i] $=\frac{y_{ik}}{\sum_{k=1}^{k} y_{ik}}$ οι οποίες οδηγούν και πάλι σε μια ντετερμινιστική πολιτική με τιμέ $D_{ik} = 0$ ή 1

**Επίλυση του παραδείγματος με γραμμικό προγραμματισμό και παράγοντα προεξόφλησης=0.9**

Η αντικέιμενική συνάρτηση έχει πάρει τη μορφή:

$$Z = 1.000y_{11} + 6.000y_{13} + 3.000y_{21} + 4.000y_{22} + 6.000y_{23} + 6.000y_{33}$$

υπό τους περιορισμούς

$$y_{01} - 0.9(y_{13} + y_{23} + y_{33}) = \frac{1}{4} \tag{23}$$

$$y_{11} + y_{13} - 0.9(\frac{7}{8}y_{01} + \frac{3}{4}y_{11} + y_{22}) = \frac{1}{4} \tag{24}$$

$$y_{21} + y_{22} + y_{23} - 0.9(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}) = \frac{1}{4} \tag{25}$$

$$y_{33} - 0.9(\frac{1}{16}y_{01+}\frac{1}{8}y_{11} + \frac{1}{2}y_{21}) = \frac{1}{4} \tag{26}$$

και όλα τα

$$y_{ik} \geq 0$$

όπου επιλέχθηκαν $\beta_0, \beta_1, \beta_2$και $\beta_3$ ίσα με $\frac{1}{4}$ Από μέθοδο simplex οι μεταβλητές υπολογίζονται ως:

$$y_{01} = 1.210$$

$$(y_{11}, y_{13}) = (6.656, 0)$$

και

$$(y_{21}, y_{22}, y_{23}) = (0, 1.067, 0)$$

και άρα

$$D_{01} = 1$$

$$(D_{11}, D_{13}) = (1, 0)$$

$$(D_{21}, D_{22}, D_{23}) = (0, 1, 0)$$

$$D_{33} = 1$$

Αντικαθιστώντας στην αντικειμενική συνάρτηση λαμβάνουμε Z=17,325

## 5.2 Ο αλγόριθμος βελτίωσης της πολιτικής για προεξοφλημένο κόστος

Για κάθε κατάσταση i ($i = 0, 1, 2, \ldots, M$) μιας μαρκοβιανής αλυσίδας αποφάσεων η οποία λειτουργελι υπό την πολιτική R ορίζουμε ως:$V_i^n(R)$ το ολικό αναμενόμενο προεξοφλημένο κόστος για ένα σύστημα το οποίο ξεκινά στην κατάσταση i και εξελίσσεται για n χρονικές περιόδους το οποίο αποτελείται από δύο στοιχεία:

- $C_{ik}$ το κόστος το οποίο προέκυψε κατά την πρώτη χρονική περίοδο παρατήρησης και
- $\alpha \sum_{j=0}^{M} p_{ij}(k) v_j^{n-1}(R)$ το ολικό προεξοφλημένο κόστος για τις υπόλοιπες n-1 περιόδους

**Σύνοψη του αλγορίθμου βελτίωσης της πολιτικής**

Αρχικοποίηση:Διαλέγω τυχαία αρχική δοκιμαστική πολιτική $R_1$.Θέτω n=1. Επανάληψη n:

- Βήμα 1: Ορισμός τιμών.

  Για την πολιτική $R_n$ χρησιμοποιούμε $p_{ij}(k), c_{ik}$ και $v_M(R_n)$ για να λύσουμε το σύστημα των M+1 εξισώσεων

  $$V_i(R_n) = c_{ik} + a \sum_{j=0}^{M} p_{ij}(k) v_j(R_n)$$

  για $i = 0, 1, 2, \ldots, M$

- Βήμα 2:

  Χρησιμοποιώντας τις τιμές των $v_i(R_n)$ οι οποίες υπολογίστηκαν στο προηγούμενο βήμα για την πολιτική $R_n$ βρίσκω την εναλλακτική πολιτική $R_{n+1}$όπου για κάθε κατάσταση i,$d_i(R_{n+1})$είναι η απόφαση η οποία ελαχιστοποιεί το $c_{ik} + a \sum_{j=0}^{M} p_{ij}(k) v_j(R_n)$

  Δηλαδή για κάθε κατάσταση i θέλουμε:

  $$Minimize_{k=1,2,\ldots,k}[c_{ik} + a \sum_{j=0}^{M} p_{ij}(k) v_j(R_n)]$$

  και τότε θέτουμε $d_i(R_{n+1})$ ίσο με την παραπάνω τιμή ελαχιστοποίησης του k.

*Δοκιμή βελτιστοποίσης.*Η τρέχουσα πολιτική $R_{n+1}$ θα είναι η βέλτιστη εάν είναι ίδια με την πολιτική $R_n$. Εάν είναι ίδιες, τότε ο αλγόριθμος σταματά. Διαφορετικά, θέτουμε n=n+1 και εκτελούμε άλλη μία επανάληψη.

**Λύση του παραδείγματος με τον αλγόριθμο βελτίωσης της πολιτικής και κριτήριο το προεξοφλημένο κόστος**

Με α=0.9 ξεκινάμε με αρχική αυθαίρετα επιλεγμένη πολιτική $R_1$. Δηλαδή, δεν επεμβαίνουμε στις καταστάσεις 0 και 1, επιδιορθώση στην κατάσταση 2 και αντικατάσταση στην κατάσταση 3 με γνωστούς πίνακες :

$$
\begin{array}{cc}
 & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\
P = & \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array}
\end{array}
$$

και κόστη:

| ΚΑΤΑΣΤΑΣΗ | ΚΟΣΤΟΣ |
|---|---|
| **0** | 0 |
| **1** | 1.000 |
| **2** | 4.000 |
| **3** | 6.000 |

- **Βήμα 1**: Ορισμός τιμών

  Με αντικατάσταση λαμβάνουμε το σύστημα των εξισώσεων:

$$V_0(R_1) = 0.9[\frac{7}{8}V_1(R_1) + \frac{1}{6}V_2(R_1) + \frac{1}{16}V_3(R_1)] \tag{27}$$

$$V_1(R_1) = 1.000 + 0.9[\frac{3}{4}V_1(R_1) + \frac{3}{8}V_2(R_1) + \frac{1}{8}V_3(R_1)] \tag{28}$$

$$V_2(R_1) = 4.000 + 0.9[V_1(R_1)] \tag{29}$$

$$V_3(R_1) = 6.000 + 0.9[V_0(R_1)] \tag{30}$$

  των οποίων η ταυτόχρονη λύση είναι:

$$V_0(R_1) = 14.949$$

$$V_1(R_1) = 16.262, V_2(R_1) = 18.636$$
$$V_3(R_1) = 19.454$$

- **Βήμα 2**:Βελτίωση της πολιτικής

$$Minimize_{k=1,2,\ldots,k}[c_{ik} + a\sum_{j=0}^{M} p_{ij}(k)v_j(R_n)]$$

Χρειάζεται να κατασκευάσω τις εκφράσεις μόνο για τις καταστάσεις που σχετίζονται με περισσότερο από μια απόφαση

 – κατάσταση 1:

$$c_{1k} + 0.9[(p_{10}(14.949) + p_{11}(16.262) + p_{12}(18.636) + p_{13}(19.454)]$$

όπου σχετίζεται με τις καταστάσεις 1 και 3 ως εξής:

| απόφαση | $C_{1k}$ | $p_{10}(k)$ | $p_{11}(k)$ | $p_{12}(k)$ | $p_{13}(k)$ | τιμή | |
|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 0 | 3/4 | 1/8 | 1/8 | 16.262 | minimum |
| 3 | 6.000 | 0 | 0 | 0 | 0 | 19.454 | |

 – κατάσταση 2:

$$c_{2k} + 0.9[(p_{20}(14.949) + p_{21}(16.262) + p_{22}(18.636) + p_{23}(19.454)]$$

όπου σχετίζεται με τις καταστάσεις 1 και 3 ως εξής:

| απόφαση | $C_{2k}$ | $p_{20}$ | $p_{21}$ | $p_{22}$ | $p_{23}$ | τιμή | |
|---|---|---|---|---|---|---|---|
| 1 | 3.000 | 0 | 0 | 1/2 | 1/2 | 20.140 | |
| 2 | 4.000 | 0 | 1 | 0 | 0 | 18.636 | minimum |
| 3 | 6.000 | 1 | 0 | 0 | 0 | 19.454 | |

Άρα η απόφαση 1 ελαχιστοποιεί την έκφραση για την κατάσταση 1 και η απόφαση 2 αυτή για την κατάσταση 2 και βλέπουμε πως η πολιτική $R_2$ προκύπτει ως Πόρισμα:Εφόσον η πολιτική ταυτίζεται με την προηγούμενη είναι και η βέλτιστη.Σε αυτό το σημείο είναι σημαντική η παρατήρηση ότι ο αλγόριθμος βελτίωσης της πολιτικής για το μέσο αναμενόμενο κόστος ανά μονάδα χρόνου δίνει συχνά αλλά όχι κατα κανόνα την ίδια απάντηση με αυτόν για το αναμενόμενο προεξοφλημένο κόστος.

| ΠΟΛΙΤΙΚΗ R2 | |
|---|---|
| **ΚΑΤΑΣΤΑΣΗ** | ΑΠΟΦΑΣΗ |
| **0** | 1 |
| **1** | 1 |
| **2** | 2 |
| **3** | 3 |

# 6 Η μέθοδος των διαδοχικών προσεγγίσεων( method of succesive approximations )

Βασικό στοιχείο αυτής της μεθόδου είναι η αναζήτηση μιας προσέγγισης της βέλτιστης πολιτικής για ένα πρόβλημα αορίστου χρόνου το οποίο εν γένει απαιτεί προσεκτική διαχείρηση. Η βέλτιστη πολιτική για το πρόβλημα αορίστου χρόνου θα συγκλίνει στην βέλτιστη πολιτική του αντίστοιχου προλήματος ορισμένου χρόνου καθώς διαχειριζόμαστε το δεύτερο για $n = 1, n = 2, n = 2, \ldots$ και καθώς το n αυξάνει .Ο λόγος για τον οποίο προτιμάται αυτή η μέθοδος είναι ότι έχουμε ήδη στη διάθεση μας μία γρήγορη μέθοδο έυρεσης της βελτιστής πολιτικής για ένα σύστημα που δρα σε n χρονικές περιόδους από την θεωρία του πιθανολογικού δυναμικού προγραμματισμού.

Ορίζουμε , για $i = 0, 1, \ldots, M$ ως
$V_i^n$: το ολικό αναμενόμενο προεξοφλημένο κόστος ενός συστήματος το οποίο ακολουθεί μια βέλτιστη πολιτική, δεδομένου ότι η διαδικασία ξεκινά στην κατάσταση i και έχει n περιόδους λειτουργίας όπου τα $V_i^n$ υπολογίζονται απο την σχέση:

$$V_i^n = min[c_{ik} + a \sum_{j=0}^{M} p_{ij}(k) v_j^{n-1}] \quad \gamma\iota\alpha \quad i = 0, 1, 2, \ldots, M \tag{31}$$

Η τιμή ελαχιστοποίσης k παρέχει την βέλτιστη απόφαση προς λήψη όταν η διαδικασία ξεκινά στην κατάσταση i. Ξεκινώντας για n=1 όλα τα $v_i^0 = 0$ και έτσι

$$v_i^1 = min\{c_{ik}\}$$

για $i = 0, 1, 2, \ldots, M$. Επίσης, αν η μαρκοβιανή διαδικασία έχει όντως n περιόδους λειτουργίας τότε η μέθοδος των διαδοχικών προσεγγίσεων θα καταλήξει με σιγουρια στη βέλτιστη πολιτική.Σε αυτήν την περίπτωση θέτουμε α=1 και τότε στόχος είναι η ελαχιστοποίηση του αναμενόμενου κόστους σε n περιόδους.

.

**Λύση του παραδείγματος με τη μέθοδο των διαδοχικών προσεγγίσεων**

Χρησιμοποιούμε α=0.9

**1η επανάληψη(n=1)**: Οι τιμές που λαμβάνουμε για το $V_i^n$ είναι:

- i=0: $V_0^1 = min_{k=1}\{c_{0k}\} = 0 \quad (k = 1)$
- i=1: $V_1^1 = min_{k=1,3}\{c_{1k}\} = min\{c_{11}, c_{13}\} = 1.000 \quad (k = 1)$
- i=2: $V_2^1 = min_{k=1,2,3}\{c_{2k}\} = min\{c_{21}, c_{22}, c_{23}\} = 3.000 \quad (k = 1)$
- i=3: $V_3^1 = min_{k=3}\{c_{3k}\} = min\{c_{33}\} = 6.000 \quad (k = 3)$

**Συμπέρασμα πρώτης προσέγγισης**:Να παρθεί απόφαση 1 όταν το σύστημα είναι στην κατάσταση 0,1 και 2 ενώ να παρθεί απόφαση 3 όταν το σύστημα βρίσκεται στην κατάσταση 3.

**Επανάληψη 2(n=2)**

- i=0:

$$V_0^2 = min_{k=1}\{c_{0k}+0.9\sum_{j=0}^{M} p_{oj}V_j\} = 0+0.9[\frac{7}{8}(1.000)+\frac{1}{16}(3.000)+\frac{1}{16}(6.000)] = 1.294$$

  $(k = 1)$

- i=1:

$$V_1^2 = min\{1.000+0.9[\frac{3}{4}(1.000)+\frac{1}{8}(3.000)+\frac{1}{8}(6.000)], 6.000+0.9[1(0)]\} = 2,688$$

  $(k = 1)$

- i=2:

$$V_2^2 = min\{3.000+0.9[\frac{1}{2}(3.000)+\frac{1}{2}(6.000)], 4.000+0.9[1(1.000)], 6.000+0.9[1(0)]\} = 4.900$$

  $(k = 2)$

- i=3:

$$V_3^2 = \{6.000 + 0.9[1(0)]\} = 6.000$$

  $(k = 3)$

**Συμπέρασμα δεύτερης προσέγγισης**:Να παρθεί απόφαση 1 όταν το σύστημα είναι στην κατάσταση 0,1 να παρθεί απόφαση 2 όταν το σύστημα είναι στην κατάσταση 2 ενώ να παρθεί απόφαση 3 όταν το σύστημα βρίσκεται στην κατάσταση 3.Σε αυτό το σημείο σημειώνεται πως αυτή είναι και η βέλτιστη πολιτική όπως υποδείξαμε και με τις προηγούμενες μεθόδους.Παρόλα αυτά,το ολικό αναμενόμενο προεξοφλούμενο κόστος στην κατάσταση για το πρόβλημα 2 χρονικών περιόδων δεν έχει προσεγγίσει ακόμα την πραγματική του τιμή.(για το πρόβλημα του αορίστου χρόνου)

.

**Επανάληψη 3η(n=3)**

- i=0:

$$V_0^3 = min\{0 + 0.9[\frac{7}{8}(2.688) + \frac{1}{16}(4.900) + \frac{1}{16}(6.000)] = 2.730$$

$(k = 1)$

- i=1:

$$V_1^3 = min\{1.000+0.9[\frac{3}{4}(2.688)+\frac{1}{8}(4.900)+\frac{1}{8}(6.000)], 6.000+0.9[1(1.294)]\} = 4.041$$

$(k = 1)$

- i=2:

$$V_2^3 = min\{3.000+0.9[\frac{1}{2}(4.900)+\frac{1}{2}(6.000)], 4.000+0.9[1(2.688)], 6.000+0.9[1(1.294)]\} = 6.419$$

$(k = 2)$

- i=3:

$$V_3^3 = min\{6.000 + 0.9[1(1.294)]\} = 7.165$$

$(k = 3)$

**Συμπεράσματα της τρίτης επανάληψης**.Η βέλτιστη λύση έχει και πάλι επιτευχθεί και αυτή τη φορα οι τιμές $V_i^3$ είναι πιο κοντά στις πραγματικές.Βάση αυτού συμπεραίνουμε πως καθώς αυξάνεται το n οι τιμές $v_0^n, v_1^n, v_2^n, v_3^n$ θα είναι 14,949, 16,262, 18,636, και 19,454 αντίστοιχως.
Όπως φαίνεται σημαντικό πλεονέκτημα αυτής της μεθόδου συγκριτικά με αυτήν της βέλτιστης πολιτικής και του γραμμικού προγραμματισμου είναι πως δεν απαιτείται να λυθεί ένα σύστημα εξισώσεων και έτσι κάθε επανάληψη μπορεί να λυθεί γρήγορα και εύκολα.

# 7 Γενικές Εφαρμογές

## 7.1 Εφαρμογή 1

**Εφαρμογή 1**
Μια εταιρεία σαπουνιών ειδικεύεται σε ένα πολυτελές είδος σαπουνιού μπάνιου. Οι πωλήσεις αυτού του σαπουνιού κυμαίνονται μεταξύ δύο επιπέδων—χαμηλές και υψηλή—εξαρτάται από δύο παράγοντες: (1) εάν πραγματοποιείται διαφημίση και (2) από τη διαφήμιση και την εξαγωγή νέων προϊόντων από τους ανταγωνιστές. Ο δεύτερος παράγοντας είναι εκτός ελέγχου της εταιρείας, αλλά προσπαθεί να καθορίσει ποια πρέπει να είναι η δική της διαφημιστική πολιτική. Για παράδειγμα, η πρόταση του διευθυντή μάρκετινγκ είναι η διαφημίση να πραγματοποιείται όταν οι πωλήσεις είναι χαμηλές αλλά όχι όταν οι πωλήσεις είναι υψηλές (μια συγκεκριμένη πολιτική). Η διαφήμιση σε οποιοδήποτε τρίμηνο του έτους έχει πρωταρχικό αντίκτυπο στις πωλήσεις του επόμενο τρίμηνου. Στην αρχή κάθε τριμήνου, οι απαραίτητες πληροφορίες που χρειάζονται για να προβλεφθεί εάν οι πωλήσεις θα είναι χαμηλές ή υψηλές αυτό το τρίμηνο είναι διαθέσιμες και έτσι μπορεί να αποφασιστεί εάν θα γίνει διαφήμιση. Το κόστος της διαφήμισης είναι 1 εκατομμύριο δολάρια για κάθε τρίμηνο του έτους στο οποίο γίνεται. Όταν η διαφήμιση γίνεται κατά τη διάρκεια ενός τριμήνου, η πιθανότητα να έχουμε υψηλές πωλήσεις το επόμενο τρίμηνο είναι $\frac{1}{2}$ και $\frac{3}{4}$ ανάλογα με το αν οι πωλήσεις του τρέχοντος τριμήνου είναι χαμηλές ή υψηλές. Αυτές οι πιθανότητες πέφτουν σε $\frac{1}{4}$και $\frac{1}{2}$ όταν η διαφήμιση δεν γίνεται κατά το τρέχον τρίμηνο. Τα τριμηνιαία κέρδη της εταιρείας (χωρίς το κόστος διαφήμισης) είναι 4 εκατομμύρια δολάρια όταν οι πωλήσεις είναι υψηλές αλλά μόνο 2 εκατομμύρια δολάρια όταν οι πωλήσεις είναι χαμηλές. Η διοίκηση θέλει τώρα να καθορίσει τη διαφημιστική πολιτική που θα μεγιστοποιήσει την πολιτική της εταιρείας (μακροχρόνια) αναμενόμενο μέσο καθαρό κέρδος (κέρδος μείον διαφημίσεις κόστος) ανά τρίμηνο
**Λύση:**
Η κατάσταση του συτήματος είναι το επίπεδο των πωλήσεων στην διάρκεια της περιόδου(τριμήνοτ t) όπου:

$$X_t = \left\{ \begin{array}{ll} 0, & \text{χαμηλές} \\ 1, & \text{υψηλές} \end{array} \right\}$$

Επίσης υπάρχουν δύο δυνατές αποφάσεις σε κάθε κατάσταση i:

$$d_i = \left\{ \begin{array}{ll} 1, & \text{δεν πραγματοποιείται διαφήμιση} \\ 2, & \text{πραγματοποιείται διαφήμιση} \end{array} \right\}$$

Θα ορίσουμε κέρδος(αντί κόστους) ενός βήματος και η αντικειμενική συνάρτηση θα είναι το αναμενόμενο μέσο κέρδος ανά τρίμηνο.Άρα αυτή τη φορά κάνουμε λόγο για μεγιστοποίηση (maximize).
Το κέρδος ενός βήματος στην κατάσταση i και κάτω από την απόφαση k είναι ίσο με $C_{ik}$ όπου:

$$c_{01} = 2 \quad c_{02} = 1$$

$$c_{11} = 4 \quad c_{12} = 3$$

και οι πιθανότητες μετάβασης είναι:

$$p_{00}(1) = \frac{3}{4} \quad p_{00}(2) = \frac{1}{2}$$

$$p_{10}(1) = \frac{1}{2} \quad p_{10}(2) = \frac{1}{4}$$

$$p_{01}(1) = \frac{1}{4} \quad p_{01}(2) = \frac{1}{2}$$

$$p_{11}(1) = \frac{1}{2} \quad p_{11}(2) = \frac{3}{4}$$

Δηλαδή

$$P^1 = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \end{array}$$

$$P^2 = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \end{array}$$

Για όλες τις ντετερμινιστικές στάσιμες πολιτικές η αντίστοιχη μαρκοβιανή αλυσίδα έχει το παρακάτω διάγραμμα Επίσης,οι δυνατές πολιτικές είναι

- **R1**: Όταν οι πωλήσεις είναι χαμηλές ή υψηλές δεν πραγματοποιείται διαφήμιση $d_0 = 1, d_1 = 1$

$$P^1 = \begin{array}{cc} \begin{array}{cc} 0 & 1 \end{array} & \\ \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

με $c_{01} = 2, c_{11} = 4$

- **R2**: Όταν οι πωλήσεις είναι χαμηλές δεν πραγματοποιείται διαφήμιση ενώ όταν είναι υψηλές πραγματοποιείται διαφήμιση $d_0 = 1, d_1 = 2$

$$P^2 = \begin{array}{cc} \begin{array}{cc} 0 & 1 \end{array} & \\ \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

με $c_{01} = 2, c_{12} = 3$

- **R3**: Όταν οι πωλήσεις είναι χαμηλές πραγματοποιείται διαφήμιση ενώ όταν είναι υψηλές δεν πραγματοποιείται διαφήμιση $d_0 = 2, d_1 = 1$

$$P^3 = \begin{array}{cc} \begin{array}{cc} 0 & 1 \end{array} & \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

με $c_{02} = 1, c_{11} = 4$

- **R4**: Όταν οι πωλήσεις είναι χαμηλές ή υψηλές πραγματοποιείται διαφήμιση $d_0 = 2, d_1 = 2$

$$P^4 = \begin{array}{cc} \begin{array}{cc} 0 & 1 \end{array} & \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

με $c_{02} = 1, c_{12} = 3$

Έτσι, πλέον μπορούν να υπολογιστούν οι στάσιμες κατανομές κατάστασης για την κάθε πολοτική και να υπολογιστεί το μέσο αναμενόμενο κέρδος Θα βρεθούν για για την πολιτική $R_1$ καθώς ο υπολογισμός των υπολοίπων είναι όμοιος.

$$\left\{ \begin{array}{l} \pi_0 + \pi_1 = 1 \\ \pi_0 = \frac{3}{4}\pi_0 + \frac{1}{2}\pi_1 \end{array} \right\}$$

από όπου προκύπτει πως $(\pi_0, \pi_1) = (\frac{2}{3}, \frac{1}{3})$

Ομοίως καταλήγουμε στον παρακάτω πίνακα.

| **Πολιτική** | **Στάσιμη κατανομή(π0,π1)** | **Υπολογισμός** | **Αναμενόμενο κέρδος/τρίμηνο ($εκατο/ων)** |
|---|---|---|---|
| **R1** | (2/3,1/3) | 2(2/3)+4(1/3) | 8/3=2.67 |
| **R2** | (1/2,1/2) | 2(1/2)+3(1/2) | 5/2=2.5 |
| **R3** | (1/2,1/2) | 1(1/2)+4(1/2) | 5/2=2.5 |
| **R4** | (1/3,2/3) | 1(1/3)+3(2/3) | 7/3=2.33 |

**Συμπέρασμα:**

Το κέρδος μεγιστοποιείται αν ακολουθήσουμε την πρώτη πολιτική δηλαδή να μη γίνει διαφήμιση είτε οι πωλήσεις είναι υψηλές είτε χαμηλές.

## 7.2 Εφαρμογή 2

**Εφαρμογή 2**

Προσεγγίζουμε ένα πρόβλημα αποθέματος (inventory problem) αορίστου χρόνου το οποίο περιλαμβάνει ένα μόνο προϊόν.Στην αρχή κάθε περιόδου πρέπει να ληφθεί μία απόφαση σχετικά με τη ποσότητα προϊόντος που συνίσταται να παραχθεί την ίδια αυτή χρονική περίοδο.Το κόστος εγκατάστασης (set up cost) ανέρχεται σε στα 10$ και το κόστος κάθε παραγόμενης μονάδας στα 5$.Το κόστος αποθήκευσης για κάθε είδος που δεν πουλήθηκε κατα την διάρκεια της περιόδου είναι 4$(η ποσότητα των ειδών για τα οποία υπάρχει συνατότητα αποθήκευσης είναι εως και 2 είδη).Η ζήτηση κάθε περίοδο ακολουθεί μια γνωστή κατανομή πιθανότητας και συγκριμένα πιθανότητα $\frac{1}{3}$ για ζήτηση κανενός ενός και δύο ειδών αντίστοιχα. Εάν η ζήτηση υπερβεί τη διαθέσιμη ποσότητα προς προσφορά τότε αυτές οι πωλήσεις χάνονται για την επιχείρηση και προκύπτει κόστος 8$ για την έλλειψη ενός είδους και 32$ για την έλλειψη 2 ειδών(συμπεριλαμβανομένων και των χαμένων εσόδων. Ερωτήσεις:

- Εξετάστε την πολιτική σύμφωνα με την οποία παράγονται δύο είδη αν δεν υπάρχουν είδη στο απόθεμα και προσιορίστε το αντίστοιχο (μακροπρόθεσμο) μέσο αναμενόμενο κόστος ανά χρονική περίοδο

- Αναγνωρίστε όλες τις εφικτές ντετερμινιστικές πολιτικές που δεν οδηγούν σε υπέρβαση του διαθέσιμου αποθέματος
- Χρησιμοποιήστε τον αλγόριθμο βελτίωσης της πολιτικής και προσδιορίστε την βέλτιστη πολιτική του προβλήματος

**Λύση**

Θεωρούμε ως :

$$X_t = \{ \text{ ο αριθμός των προϊόντων σε απόθεμα στην αρχή της περιόδου t } \}$$

Επίσης κάθε κατάσταση i συνδέεται με έναν αριθμό αντίστοιχων αποφάσεων.Συγκεκριμένα, στην κατάσταση 0 μπορεί να παρθεί είτε η απόφαση να μην παραχθεί κανένα είδος ή να παραχθούν ένα η δύο είδη .Στην κατάσταση 1 μπορεί να παρθεί η απόφαση να παραχθούν ένα είδος η κανένα ενώ στην κατάσταση 2 μπορεί να παρθεί μόνο η απόφαση της παραγωγικής αδράνειας. Και άρα λαμβάνουμε τον παρακάτω πίνακα.

| **κατάσταση** | **απόφαση d(i)** |
|---|---|
| **0** | 0,1,2 |
| **1** | 1,0 |
| **2** | 0 |

Figure 18: σχέσεις καταστάσεων -αποφάσεων

Αν $\chi_t = i$ τότε η μέγιστη παραγόμενη ποσότητα ισούται με 2-i καθώς η επιχείρηση μπορεί να φυλάξει εώς και δύο είδη και επίσης υποθέτουμε πως η ποσότητα αποθηκέυεται πριν εμφανιστεί η ζήτηση.

Το κόστος παραγωγής k μονάδων το οποίο αποτελείται απο το κόστος της εγκατάστασης και το κόστος παραγωγής κάθε μονάδας υπολογίζεται ως:

$$C_{\pi\alpha\rho\alpha\gamma}(\kappa) = \left\{ \begin{array}{ll} 10 + 5k, & k > 0 \\ 0, & 0 \end{array} \right\}$$

Για το συνολικό κόστος κάθε περιόδου έχουμε ότι

Συνολικό αναμενόμενο κόστος($C_{ik}$)=κόστος παραγωγής +αναμενόμενο κόστος αποθήκευσης των προϊόντων που δεν πουλήθηκαν+αναμενόμενο κόστος των ελλείψεων

Για το κόστος αποθήκευσης των απούλητων προϊόντων ισχύει ότι :
Κόστος αποθήκευσης =4*αριθμό των απούλητων προϊόντων.
Αν $\chi_t = i$ και $d_i = k$ τότε στην αρχή της ημέρας θα υπάρχουν i+k μονάδες.Αν $D_t$ η ζήτηση κατα τη διάρκεια της ημέρας τότε για να τα προσδιορίσουμε τα απούλητα προϊόντα διακρίνουμε τις περιπτώσεις.:

- Εάν $D_t < i + k$ τότε η ποσότητα των προϊοντων που δεν πουλήθηκαν είναι $i + k - D_t$
- Εάν $D_t > i + k$ τότε η ποσότητα των απούλητων προϊόντων είναι 0

και το κόστος αποθήκευσης υπολογίζεται συνολικά απο την σχέση

$$C_{\alpha\pi o\theta}(i, k) = 4E[max(i + k - D_t, 0)]$$

Για το κόστος των ελλείψεων έχουμε ότι το κόστος των ελλείψεων θα είναι ίσο με 8 αν έχουμε έλλειψη ενός είδους και ίσο με 32 αν έχουμε έλλειψη δύο ειδών και άρα:

$$C_{\varepsilon\lambda\lambda} = 8P[D_t = i + k + 1] + 32P[D_t = i + k + 2]$$

Συνολικά το κόστος υπολογίζεται απο τη σχέση

$$C_{ik} = C_{\pi\alpha\rho\alpha\gamma}(k) + C_{\alpha\pi o\theta}(i, k) + C_{\varepsilon\lambda\lambda}(i, k) \tag{32}$$

Έτσι διαδοχικά έχουμε

- i=0
  - k=0

    $$C_{\pi\alpha\rho\alpha\gamma}(0) = 0, C_{\alpha\pi o\theta}(0, 0) = 0, C_{\varepsilon\lambda\lambda}(0, 0) = 8(\frac{1}{3}) + 32(\frac{1}{3}) = \frac{40}{3}$$

    Συνολικά,

    $$C_{00} = 13,33$$

  - k=1

    $$C_{\pi\alpha\rho\alpha\gamma}(1) = 15, C_{\alpha\pi o\theta}(0, 1) = 4E[max\{1-D_t, 0\}] = 4[\frac{1}{3}+0(\frac{1}{3})+0(\frac{1}{3})] = 1.33$$

    $$C_{\varepsilon\lambda\lambda}(0, 1) = 8P[D_t = 2] + 32P[D_t = 3] = 8(\frac{1}{3}) + 0 = 2.67$$

    Συνολικά,

    $$C_{01} = 15 + 1.33 + 2.67$$

- k=2

$$C_{\pi\alpha\rho\alpha\gamma}(2) = 20, C_{\alpha\pi o\theta}(0,2) = 4E[max\{1-D_t, 0\}] = 4[2(\frac{1}{3})+(\frac{1}{3})] = 4$$

$$C_{\varepsilon\lambda\lambda}(0,1) = 8P[D_t = 3] + 32P[D_t = 4] = 0$$

Συνολικά,

$$C_{02} = 24$$

- i=1
  - k=0

$$C_{\pi\alpha\rho\alpha\gamma}(0) = 0, C_{\alpha\pi o\theta}(1,0) = 4E[max\{1 - D_t, 0\}] = 4[1(\frac{1}{3})]$$

$$C_{\varepsilon\lambda\lambda}(1,0) = 8P[D_t = 2] + 32P[D_t = 3] = 8(\frac{1}{3})$$

Συνολικά,

$$C_{10} = 4$$

  - k=1

$$C_{\pi\alpha\rho\alpha\gamma}(1) = 15, C_{\alpha\pi o\theta}(1,1) = 4 = C_{\alpha\pi o\theta}(0,2)$$

$$C_{\varepsilon\lambda\lambda}(1,1) = C_{\varepsilon\lambda\lambda}(0,2) = 0$$

Συνολικά,

$$C_{11} = 19$$

- i=2
  - k=0

$$C_{\pi\alpha\rho\alpha\gamma}(0) = 0, C_{\alpha\pi o\theta}(2,0) = 4 = C_{\alpha\pi o\theta}(0,2)$$

$$C_{\varepsilon\lambda\lambda}(2,0) = C_{\varepsilon\lambda\lambda}(0,2) = 0$$

Συνολικά,

$$C_{20} = 4$$

Θεωρούμε την πολιτική $R$ η οποία παράγει δύο μονάδες προϊόντων για $x_t = 0$ και τίποτα άλλο.Άρα $R : d_0 = 2, d_1 = 0, d_2 = 0$ η οποία αντιστοιχεί σε πίνακα πιθανοτήτων μετάβασης

$$\begin{array}{cc} \begin{array}{ccc} 0 & 1 & 2 \end{array} & \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \end{array}$$

και άρα εχουμε το σύστημα των εξισώσεων

$$\begin{cases} \pi_0 + \pi_1 + \pi_2 = 1 \\ \pi_0 = \frac{1}{3}\pi_0 + \frac{2}{3}\pi_1 + \frac{1}{3}\pi_2 \\ \pi_1 = \frac{1}{3}\pi_0 + \frac{1}{3}\pi_1 \\ \pi_2 = \frac{1}{3}\pi_1 + \frac{1}{3}\pi_2 \end{cases} . \tag{33}$$

απο όπου προκύπτει πως

$$\pi = (\frac{4}{9}, \frac{3}{9}, \frac{2}{9})$$

και το συνολικό αναμενόμενο μακροπρόθεσμο μέσο κόστος υπολογίζεται ως

$$c(R) = \frac{4}{9}24 + \frac{3}{9}4 + \frac{2}{9}4 = 12,88$$

οι δυνατές πολιτικές βρίσκονται με τη βοήθεια του πίνακα 12(figure 12) και είναι οι εξής:

- $R_1 : d_0 = 0, d_1 = 0, d_2 = 0$
- $R_2 : d_0 = 0, d_1 = 1, d_2 = 0$
- $R_3 : d_0 = 1, d_1 = 0, d_2 = 0$
- $R_4 : d_0 = 1, d_1 = 1, d_2 = 0$
- $R_5 : d_0 = 2, d_1 = 0, d_2 = 0$
- $R_6 : d_0 = 2, d_1 = 1, d_2 = 0$

**Επίλυση με τον αλγόριθμο βελτίωσης της πολιτικής**
Ξεκινάμε αυθαίρετα με την πολιτική $R_1 = (0,0,0)$ η οποία αντιστοιχεί στον πίνακα πιθανοτήτων μετάβασης

$$\begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ P_1 = & \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \end{array}$$

και σχετίζεται με τα κόστη

$$C_{00} = 13.333, C_{10} = 4, C20 = 4$$

Βήμα προσδιορισμού τιμών $v_2(R_1) = 0$

$$\begin{cases} g(R_1) = 13.33 + v_0(R_1) - v_0(R_1) \\ g(R_1) = 4 + \frac{2}{3}v_0(R_1) + \frac{1}{3}v_1 R_1 - v_1 R_1 \\ g(R_1) = 4 + \frac{1}{3}v_0(R_1) + \frac{1}{3}v_1(R_1) \end{cases} \tag{34}$$

από όπου βρίσκουμε πως

$$\begin{cases} g(R_1) = 13.33 \\ v_0(R_1) = 21 \\ v_1(R_1) = 7 \end{cases} \tag{35}$$

Βήμα βελτίωσης:

- i=0

$$min\{13.33, 19+\frac{2}{3}v_0(R_1)-\frac{1}{3}v_1(R_1)-v_0(R_1), 24+\frac{1}{3}v_0(R_1)+\frac{1}{3}v_1(R_1)-v_0(R_1)\} = \tag{36}$$

  =min{13.33,14.33,12.33} =12.33(k=2)

- i=1

$$min\{4+\frac{2}{3}v_0(R_1)+\frac{1}{3}v_1(R_1)-v_1(R_1), 19+\frac{1}{3}v_0(R_1)+\frac{1}{3}v_1(R_1)-v_1(R_1)\} = \tag{37}$$

  =min{13.33,21.33}=13.33 (k=0)

- i=2
  έχουμε μόνη δυνατή απόφαση k=0 και άρα αυτό το βήμα είναι περιττό (k=0)

Άρα η νέα πολιτική είναι η $R_2 = (2,0,0)$ για την οποία έχουμε αντίστοιχο πίνακα μεταβάσεων

$$P_2 = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}$$

και αντίστοιχα κόστη

$$C_{02} = 24$$

$$C_{10} = 4$$

$$C_{20} = 4$$

Βήμα προσδιορισμού τιμών($v_2(R_2) = 0$)

$$\begin{cases} g(R_2) = 24 + \frac{1}{3}v_0(R_2) + \frac{1}{3}v_1(R_2) - v_0(R_2) \\ g(R_2) = 4 + \frac{2}{3}v_0(R_2) + \frac{1}{3}v_1(R_2) - v_1(R_2) \\ g(R_2) = 4 + \frac{1}{3}v_0(R_2) + \frac{1}{3}v_1(R_2) \end{cases} \tag{38}$$

από όπου βρίσκουμε πως

$$\begin{cases} g(R_2) = \frac{116}{9} = 12.889 \\ v_0(R_2) = 20 \\ v_1(R_2) = \frac{20}{3} = 6.667 \end{cases} \tag{39}$$

Βήμα βελτίωσης:

- i=0

$$min\{13.33, 19+\frac{2}{3}v_0(R_2)+\frac{1}{3}v_1(R_2)-v_0(R_2), 24+\frac{1}{3}v_0(R_2)+\frac{1}{3}v_1(R_2)-v_0(R_2)\} = \tag{40}$$

=min{13.33,14.556,12.889} =12.889(k=2)

- i=1

$$min\{4+\frac{2}{3}v_0(R_2)+\frac{1}{3}v_1(R_2)-v_1(R_2), 19+\frac{1}{3}v_0(R_2)+\frac{1}{3}v_1(R_2)-v_1(R_2)\} = \tag{41}$$

=min{12,889,21.222}=12.889 (k=0)

- i=2
έχουμε μόνη δυνατή απόφαση k=0 και άρα αυτό το βήμα είναι περιττό (k=0)

Καταλήγουμε λοιπόν,στην πολιτική $R_2 = (2,0,0)$η οποία ταυτίζεται με την προηγούμενη ευρεθείσα πολιτική και άρα είναι και η βέλτιστη $p_{ij}^n > 0$

# 8 Βιβλιογραφία

## References


[1] **1.1-1.4** Κολέτσος, Ι. και Στογιάννης, Δ. (2012). Εισαγωγή στην Επιχειρησιακή Έρευνα. Εκδόσεις Συμεών, Αθήνα.

[2] **2.1,2.2,2.3** Introduction to Operations Research (Frederick S. Hillier and Gerald J. Lieberman), McGraw-Hill, 2001(σελ.780-.792),

Κολέτσος, Ι. και Στογιάννης, Δ. (2012). Εισαγωγή στην Επιχειρησιακή Έρευνα. Εκδόσεις Συμεών, Αθήνα

[3] **3.1,3.2** Introduction to Operations Research (Frederick S. Hillier and Gerald J. Lieberman), McGraw-Hill, 2001(σελ.827.828),
Λουλάκης Μ. (2015). Στοχαστικές Διαδικασίες, σελ. 1-2.
Propability,markov chains,queues and simulation by William J.Stewart(σελ 202,207)

[4] Introduction to Operations Research (Frederick S. Hillier and Gerald J. Lieberman), McGraw-Hill, 2001(σελ.830)
Λουλάκης Μ. (2015). Στοχαστικές Διαδικασίες, σελ. 9-10.
Operational research Applications and Algorithms by Wayne L.Winston(σελ.930-944)

Κολέτσος, Ι. και Στογιάννης, Δ. (2012). Εισαγωγή στην Επιχειρησιακή Έρευνα. Εκδόσεις Συμεών, Αθήνα.

[5] **4** Introduction to Operations Research (Frederick S. Hillier and Gerald J. Lieberman), McGraw-Hill, 2001

[6] **4,5** Puterman, M. L. (2005). Markov Decision Processes, Discrete Stochastic Dynamic Programming. John Wiley and Sons(σελ.17-20) New Jersey,
Introduction to Operations Research (Frederick S. Hillier and Gerald J. Lieberman), McGraw-Hill, 2001(σελ.1078-1082).