



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ
ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
Δ.Π.Μ.Σ. «ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ:

**«ΘΕΩΡΙΑ ΠΑΙΓΝΙΩΝ, ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ
ΚΑΙ ΣΤΡΑΤΗΓΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ»**

Μαρία Ε. Σοφούλη

Επιβλέπων:

Ι. Κολέτσος, Επίκουρος Καθηγητής Ε.Μ.Π.

Επιτροπή:

Ι. Κολέτσος, Επίκουρος Καθηγητής Ε.Μ.Π.

Ε. Τυχόπουλος, Επίκουρος Καθηγητής Ε.Μ.Π.

Δ. Φουσκάκης, Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Απρίλιος 2014

Οφείλεις να μάθεις τους κανόνες του παιχνιδιού.

Και έπειτα να παίζεις καλύτερα από τον καθένα.

Albert Einstein

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1 ^ο	4
ΕΙΣΑΓΩΓΗ	4
1.1 ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ	4
1.2 ΘΕΩΡΙΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ	6
1.3 ΘΕΩΡΙΑ ΠΑΙΓΝΙΩΝ	7
ΚΕΦΑΛΑΙΟ 2 ^ο	11
ΑΠΛΑ ΜΟΝΤΕΛΑ ΣΤΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ	11
2.1.ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ	11
2.2.ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ	12
2.3 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΟΡΘΟΛΟΓΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ	14
2.4.ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΦΥΣΙΚΗΣ ΕΠΙΛΟΓΗΣ	20
2.5 ΒΕΛΤΙΣΤΗ ΣΥΜΠΕΡΙΦΟΡΑ	23
2.6 ΔΕΝΤΡΑ ΑΠΟΦΑΣΕΩΝ	24
2.7 ΣΤΡΑΤΗΓΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ -	26
ΑΜΙΓΕΙΣ ΚΑΙ ΜΙΚΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ	26
2.8 ΒΕΛΤΙΣΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ	29
ΚΕΦΑΛΑΙΟ 3 ^ο	33
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ	33
3.1 ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΗΣ ΕΞΑΡΤΩΜΕΝΕΣ ΑΠΟ ΤΗΝ ΚΑΤΑΣΤΑΣΗ....	33
3.2 ΜΑΡΚΟΒΙΑΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΕΩΝ	35
3.3 ΣΤΟΧΑΣΤΙΚΕΣ ΜΑΡΚΟΒΙΑΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΕΩΝ	38
3.4 ΒΕΛΤΙΣΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ ΓΙΑ ΠΕΠΕΡΑΣΜΕΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ.....	42
3.5 ΜΑΡΚΟΒΙΑΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΕΩΝ ΜΗ ΠΕΠΕΡΑΣΜΕΝΟΥ ΟΡΙΖΟΝΤΑ – ΒΕΛΤΙΣΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ	44
3.6 ΒΕΛΤΙΩΣΗ ΠΟΛΙΤΙΚΗΣ	50
ΚΕΦΑΛΑΙΟ 4 ^ο	54
ΠΑΙΓΝΙΑ ΠΛΗΡΟΥΣ ΠΛΗΡΟΦΟΡΗΣΗΣ	54
4.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΓΙΑ ΤΑ ΠΑΙΓΝΙΑ ΜΕ ΠΛΗΡΗ ΠΛΗΡΟΦΟΡΗΣΗ	54
4.2 ΣΤΑΤΙΚΑ ΠΑΙΓΝΙΑ 2 ΠΑΙΚΤΩΝ ΜΗΔΕΝΙΚΟΥ ΑΘΡΟΙΣΜΑΤΟΣ	56
4.3 ΕΠΙΛΥΣΗ ΠΑΙΓΝΙΩΝ ΜΕ ΠΛΗΡΗ ΠΛΗΡΟΦΟΡΗΣΗ	57
4.3.1 ΜΕΘΟΔΟΣ Minimax	57
4.3.2 ΔΙΑΔΟΧΙΚΗ ΑΠΑΛΟΙΦΗ ΑΥΣΤΗΡΑ ΚΥΡΙΑΡΧΟΥΜΕΝΩΝ	

ΣΤΡΑΤΗΓΙΚΩΝ.....	58
4.4 ΜΙΚΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ.....	62
ΚΕΦΑΛΑΙΟ 5 ^ο	66
ΠΕΠΕΡΑΣΜΕΝΑ ΔΥΝΑΜΙΚΑ ΠΑΙΓΝΙΑ ΠΛΗΡΟΥΣ ΠΛΗΡΟΦΟΡΗΣΗΣ.....	66
5.1 ΔΕΝΤΡΑ ΠΑΙΓΝΙΩΝ	66
5.2 ΙΣΟΡΡΟΠΙΑ ΚΑΤΑ NASH.....	70
5.3 ΣΥΝΟΛΑ ΠΛΗΡΟΦΟΡΙΑΣ.....	72
5.4 ΤΕΛΕΙΟΤΗΤΑ ΑΝΑ ΥΠΟΠΑΙΓΝΙΟ.....	74
5.5 ΙΣΟΡΡΟΠΙΑ Nash – ΒΕΛΤΙΩΣΕΙΣ.....	75
ΚΕΦΑΛΑΙΟ 6 ^ο	78
ΠΑΙΓΝΙΑ ΜΕ ΣΥΝΕΧΗ ΣΥΝΟΛΑ ΣΤΡΑΤΗΓΙΚΩΝ.....	78
6.1 ΑΠΕΙΡΑ ΣΥΝΟΛΑ ΣΤΡΑΤΗΓΙΚΩΝ	78
6.2 ΜΟΝΤΕΛΟ ΔΥΟΠΩΛΙΟΥ Cournot.....	78
6.3 ΜΟΝΤΕΛΟ ΔΥΟΠΩΛΙΟΥ Stackelberg.....	82
6.4 ΤΟ ΜΟΝΤΕΛΟ «War of Attrition»	85
ΚΕΦΑΛΑΙΟ 7 ^ο	89
ΜΗ ΠΕΠΕΡΑΣΜΕΝΑ ΔΥΝΑΜΙΚΑ ΠΑΙΓΝΙΑ	89
7.1 ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΑ ΠΑΙΓΝΙΑ.....	89
7.2 ΤΕΛΕΙΟΤΗΤΑ ΑΝΑ ΥΠΟΠΑΙΓΝΙΟ.....	95
7.3 FOLK THEOREM.....	99
7.4 ΣΤΟΧΑΣΤΙΚΑ ΠΑΙΓΝΙΑ	102
Βιβλιογραφία	107

ΚΕΦΑΛΑΙΟ 1^ο

ΕΙΣΑΓΩΓΗ

1.1 ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ

Ο τομέας της Επιχειρησιακής Έρευνας (Ε.Ε.) ασχολείται με τη μελέτη, την ανάπτυξη και την εφαρμογή προηγμένων αναλυτικών μεθόδων οι οποίες έχουν ως στόχο τον συντονισμό των δραστηριοτήτων ενός οργανισμού. Το εύρος των εφαρμογών της είναι αρκετά μεγάλο καθώς μπορεί να χρησιμοποιηθεί από δημόσιες υπηρεσίες, το στρατό, κατασκευαστικές εταιρίες και τη βιομηχανία.

Η Ε.Ε. ακολουθεί για την έρευνα της μία σύνθετη διαδικασία η οποία ξεκινά με την προσεκτική παρατήρηση και τη συλλογή δεδομένων. Έπειτα κατασκευάζεται ένα επιστημονικό μαθηματικό μοντέλο το οποίο πρέπει να εντοπίζει την ουσία του προβλήματος, και να είναι καλώς ορισμένο και επαρκώς τεκμηριωμένο με τα βασικά χαρακτηριστικά του προβλήματος που καλείται να αντιμετωπίσει. Στη συνέχεια διενεργούνται τα κατάλληλα πειράματα για να δοκιμαστεί το μοντέλο (*tests*) κι αν χρειάζεται γίνονται οι απαραίτητες τροποποιήσεις ώστε να εξάγονται έγκυρα αποτελέσματα.

Παρ' όλα αυτά η έρευνα δε σταματά εκεί. Η διοίκηση του οργανισμού ανήκει στους τομείς που μελετά η επιχειρησιακή έρευνα και γι' αυτό πρέπει να παρέχει αποτελέσματα κατανοητά και εφαρμόσιμα στους αρμόδιους για τη λήψη των αποφάσεων. Για το λόγο αυτό, ο όρος «*management*» ταυτίζεται συχνά με την Επιχειρησιακή Έρευνα.

Αξίζει να αναφέρουμε ότι πολλές φορές οι ερευνητές καλούνται να δώσουν λύσεις οι οποίες είναι αντικρουόμενες για τα επιμέρους τμήματα του οργανισμού αλλά σε αυτές τις περιπτώσεις μελετάται το πρόβλημα με γνώμονα τους στόχους της επιχείρησης ως σύνολο. Επίσης, όταν αναφερόμαστε στη λύση ενός προβλήματος στην Επιχειρησιακή Έρευνα, εννοούμε τη «βέλτιστη» (*optimal*) λύση και όχι απαραίτητα την καλύτερη. Αυτό συμβαίνει διότι η καλύτερη λύση μπορεί να μην είναι πρακτικά εφικτή ή να μην είναι μοναδική.

Ιστορικά, η ανάπτυξη των επιχειρήσεων ως προς το μέγεθος και την πολυπλοκότητα των λειτουργιών τους συνδέεται άρρηκτα με την περίοδο της βιομηχανικής επανάστασης. Τα μικρά μαγαζιά που άνηκαν σε τεχνίτες της

προηγούμενης εποχής εξελίχθηκαν σε επιχειρήσεις εκατομμυρίων. Εκείνη την περίοδο σημειώθηκε δραματική αύξηση στην κατάκτηση των διοικητικών ευθυνών και υποχρεώσεων στους οργανισμούς και αυτό άρχισε να δημιουργεί προβλήματα συντονισμού και εποπτείας.

Η τάση των επιμέρους τμημάτων να αναπτύσσονται σχεδόν αυτόνομα με τους δικούς τους στόχους και τα δικά τους συστήματα αξιών προκάλεσε πρόβλημα στην εποπτεία όλων των δραστηριοτήτων του οργανισμού και στη σύνδεση των στόχων των τμημάτων προς όφελος της επιχείρησης ως σύνολο. Συχνά τα συμφέροντα των τμημάτων ήταν, όπως και σήμερα άλλωστε, αντικρουόμενα. Για παράδειγμα η ανάγκη στο τμήμα παραγωγής ενός προϊόντος μιας επιχείρησης για χρήση ποιοτικών υλικών που κοστίζουν ακριβά πιθανόν να έρχεται σε αντίθεση με τις απαιτήσεις του οικονομικού τμήματος για χαμηλό κόστος παραγωγής. Ένα ακόμη πρόβλημα είναι ότι η διανομή των διαθέσιμων πόρων μίας επιχείρησης με αποτελεσματικό τρόπο καθίσταται όλο και πιο δύσκολη όσο αυξάνεται η εξειδίκευση και η πολυπλοκότητα των λειτουργιών της.

Αυτού του είδους προβλήματα καθώς και η ανάγκη να βρεθεί ο βέλτιστος τρόπος επίλυσης τους, οδήγησαν στην ανάπτυξη της Επιχειρησιακής Έρευνας, οι ρίζες της οποίας ανιχνεύονται την περίοδο του Β' παγκοσμίου πολέμου. Η προσπάθεια να χρησιμοποιηθεί μία επιστημονική προσέγγιση η οποία θα συνέβαλλε στη σωστή διοίκηση των επιχειρήσεων αποδίδεται στον Βρετανικό και Αμερικανικό στρατό. Η διοίκηση του στρατού θέλοντας να οργανώσει με αποτελεσματικό τρόπο τους λιγοστούς διαθέσιμους πόρους κάλεσε μία ομάδα επιστημόνων να εκπονήσουν μελέτες σε αυτό το ερευνητικό πεδίο. Αυτή η ομάδα έθεσε τα θεμέλια για τη νίκη στις μάχες αέρος της Βρετανίας με την ανάπτυξη αποτελεσματικών μεθόδων στη χρήση του , τότε νέου, εργαλείου radar.

Με το πέρας του πολέμου, η επιτυχία της έρευνας αυτής κέρδισε το ενδιαφέρον και άλλων οργανισμών εκτός του στρατού. Εφόσον άρχισε και πάλι να αναπτύσσεται η βιομηχανία, εμφανίστηκαν παρόμοια προβλήματα πολυπλοκότητας και εξειδίκευσης, τα οποία ήταν τελικά σχεδόν ίδια με αυτά που αντιμετώπισε ο στρατός όταν κάλεσε την ομάδα ερευνητών, αλλά σε διαφορετικό αντικείμενο.

Ξεκίνησε τότε, στις αρχές της δεκαετίας του 1950, η εξάπλωση της Ε.Ε., η οποία συνεχώς βελτιώνει τις τεχνικές της ενώ ολοένα και περισσότεροι επιστήμονες στρεφόντουσαν σε αυτήν τη μελέτη. Αξίζει να αναφερθεί ως παράδειγμα η μέθοδος Simplex επίλυσης γραμμικού προγραμματισμού, την οποία θεμελίωσε ο George

Dantzig το 1947, ενώ παράλληλα αναπτύσσονταν και άλλοι τομείς, όπως η θεωρία ουρών αναμονής, ο δυναμικός προγραμματισμός και ο έλεγχος αποθεμάτων.

Οι ηλεκτρονικοί υπολογιστές συνέβαλλαν επίσης στη ραγδαία ανάπτυξη και εξάπλωση της Ε.Ε.. Αφενός, γιατί αποτέλεσαν (και αποτελούν) βασικό εργαλείο των ερευνητών, καθώς τους δίνεται η δυνατότητα να προβούν σε αριθμητικούς υπολογισμούς σύνθετων προβλημάτων ίσως και χιλιάδες φορές ταχύτερα σε σύγκριση με τον υπολογισμό χωρίς τους Η/Υ. Αφετέρου, γιατί δημιουργήθηκαν λογισμικά πακέτα επιχειρησιακής έρευνας που έκαναν πιο προσιτή αυτή τη μελέτη στο ευρύ κοινό που τη χρειαζόταν.

1.2 ΘΕΩΡΙΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ

Η Θεωρία Αποφάσεων ασχολείται με την εφαρμογή μεθόδων για τη λήψη σημαντικών αποφάσεων για τις οποίες δεν είναι προφανές τι ενέργεια πρέπει να γίνει, δεδομένου ότι οι αποφάσεις πρέπει να ληφθούν κάτω από συνθήκες αβεβαιότητας.

Η μελέτη αυτής της θεωρίας αφορά διάφορους επιστημονικούς κλάδους όπως τα μαθηματικά, την κοινωνιολογία, την ψυχολογία, τα οικονομικά καθώς και τις πολιτικές επιστήμες. Οι φιλόσοφοι μελετούσαν τις αποφάσεις των ανθρώπων διότι πίστευαν πως αυτές δείχνουν την ηθική και τις αξίες τους. Οι ιστορικοί αντίστοιχα μελετούν τις αποφάσεις μεγάλων ηγετών για να εξετάσουν τις στρατηγικές τους σε κρίσιμες στιγμές. Η έρευνα σχετικά με τον κίνδυνο και την οργανωτική συμπεριφορά πηγάζει από μια πιο πρακτική επιθυμία: να βοηθήσει τους διοικητές (για παράδειγμα μίας εταιρίας ή μίας χώρας) να πετύχουν καλύτερα αποτελέσματα.

Η αυξανόμενη μελέτη της διαχείρισης του κινδύνου, η προσπάθεια ανάλυσης και κατανόησης της ανθρώπινης συμπεριφοράς και η πρόοδος της τεχνολογίας που στηρίζουν τις γνωστικές διαδικασίες έχουν βελτιώσει τη λήψη αποφάσεων σε διάφορες καταστάσεις. Βέβαια, η θεωρία λήψης αποφάσεων και στρατηγικής συμπεριφοράς δεν εξελίχθηκε άμεσα και ξεκάθαρα προς την ορθολογική συμπεριφορά.

Μέσα από τις ανάγκες για βέλτιστη απόφαση άρχιζαν οι περιορισμοί, οι διευκρινίσεις και οι πιο αυστηροί ορισμοί. Οι δύσκολες συνθήκες, ο περιορισμένος χρόνος και η ανεπαρκής νοητική ικανότητα καθιστούν τον υπεύθυνο αποφάσεων περιορισμένο ορθολογικά. Αυτές όμως οι δυσκολίες ανάγκασαν τους επιστήμονες να

δώσουν μοντέλα που να τους δίνουν τουλάχιστον αποδεκτές ή βέλτιστες λύσεις και επιλογές.

Στις μέρες μας οι εταιρίες μελετούν συμπεριφορικές και ανθρωπιστικές επιστήμες καθώς και επιστήμες της τεχνολογίας όσο το δυνατό περισσότερο εξαιτίας του ανταγωνισμού και της ανάγκης για μεγιστοποίηση κέρδους. Όταν σύγχρονες διαδικασίες και οικονομικά μοντέλα όπως τα παράγωγα, τα ετήσια οικονομικά πλάνα, και η πρόβλεψη επιβάλλονται τότε η θεωρία λήψης αποφάσεων πρέπει να γίνει εργαλείο προς αρωγή.

Η βασική διαφορά με τη θεωρία παιγνίων όπως θα δούμε και παρακάτω είναι πως μελετάται το σύνολο των αποφάσεων για ένα άτομο, δεδομένων συνθηκών ανεξάρτητα από τον αντίπαλο. Στη θεωρία αυτή δεν μελετάμε την κίνηση του αντιπάλου γιατί δεν αφορά παίγνιο αλλά στρατηγική συμπεριφορά.

Αποφάσεις υπό βεβαιότητα: Έχουμε όλη την πληροφορία και μπορούμε να υπολογίσουμε το εξαγόμενο από κάθε απόφαση π.χ. ένα πρόβλημα γραμμικού προγραμματισμού.

Αποφάσεις υπό σύγκρουση: Λαμβάνουμε υπόψη τις αποφάσεις του ανταγωνιστή π.χ. παίγνια.

Αποφάσεις υπό αβεβαιότητα: Παίρνουμε υπ' όψη μας την τύχη (στοχαστικότητα) π.χ. θα βρέξει ή δεν θα βρέξει; Το ιστορικό (γνώση παρελθόντος) βοηθά στη λήψη απόφασης.

1.3 ΘΕΩΡΙΑ ΠΑΙΓΝΙΩΝ

Η μαθηματική θεωρία παιγνίων έχει ως σκοπό την αυστηρή, φορμαλιστική ανάλυση ανταγωνιστικών καταστάσεων. Αυτές οι καταστάσεις περιλαμβάνουν παιχνίδια όπως το σκάκι, το πόκερ, το baseball, αλλά και αντιπαλότητες μεταξύ εταιριών, στρατιωτικών δυνάμεων ή ακόμα και εθνών τα οποία στη θεωρία αυτή καλούνται παίγνια.

Η ανάλυση των παιχνιδιών έχει δύο στόχους. Σύμφωνα με τον πρώτο στόχο απαιτείται να κατανοηθούν οι κανόνες και ο τρόπος με τον οποίο συμπεριφέρονται οι παίκτες σε αντίπαλες θέσεις και αποδεικνύεται σημαντικός σε παίγνια με πολλούς συμμετέχοντες και πολύπλοκους κανόνες όπως στην οικονομία ή την διεθνή πολιτική.

Ο δεύτερος στόχος είναι πιο πρακτικός αλλά και «φιλόδοξος» διότι συνιστά δυνατότητα να συμβουλευσεις τους παίκτες για την βέλτιστη τακτική που πρέπει να ακολουθήσουν στο παιχνίδι.

Οι λογικοί συλλογισμοί είναι το απαιτούμενο εργαλείο για την θεωρία αυτή, ενώ σε πρακτικό και ερευνητικό επίπεδο μαθηματικές γνώσεις άλγεβρας και θεωρίας πιθανοτήτων είναι απαραίτητες. Η ακρίβεια και η σαφήνεια που απαιτούνται για την μαθηματική θεμελίωση της θεωρίας παιγνίων δεν συνεπάγεται πάντα η μαθηματική αυστηρότητα. Αυτό είναι προφανές στις φυσικές επιστήμες στις οποίες συχνά η μαθηματική αυστηρότητα στέκεται εμπόδιο στην ανάπτυξη μίας θεωρίας. Αν για παράδειγμα, μία πρόταση αποδειχθεί για πεπερασμένο πλήθος είτε παραγόντων, είτε επαναλήψεων ή παραμέτρων δεν σημαίνει πως ισχύει και για άπειρο πλήθος. Συχνά τα αποτελέσματα βασίζονται σε προσεγγίσεις και όταν μελετάμε πραγματικά παιχνίδια σε ρεαλιστικό χρόνο, το άπειρο δεν μας απασχολεί.

Αξίζει να σημειωθεί πως όταν μιλάμε για λύση ενός παιχνιδιού εννοούμε την βέλτιστη στρατηγική όλων των παικτών. Η στρατηγική αποτελεί ένα προκαθορισμένο κανόνα που καθορίζει πλήρως πώς κάποιος σκοπεύει να ανταποκριθεί σε κάθε πιθανή περίπτωση σε κάθε στάδιο του παιχνιδιού. Για παράδειγμα, η στρατηγική ενός παίκτη σκάκι υποδεικνύει την επομένη κίνηση του για κάθε πιθανή κίνηση του αντιπάλου του. Αυτό σημαίνει πως το πλήθος των δυνατών στρατηγικών είναι τεράστιο. Οι εφαρμογές τη θεωρίας παιγνίων περιλαμβάνουν συνήθως λιγότερο περίπλοκες καταστάσεις από το σκάκι αλλά οι στρατηγικές που εμπλέκονται μπορεί να είναι αρκετά σύνθετες. Παρ' όλα αυτά ακόμα και στις περιπτώσεις που το παιχνίδι δεν μπορεί να λυθεί, η θεωρία παιγνίων συχνά δύναται να προσφέρει μικρές συμβουλές στους παίκτες για το πώς θα παίξουν καλύτερα.

Σε αντίθεση με τη θεωρία λήψης αποφάσεων, της οποίας η βασική υπόθεση είναι ότι ο υπεύθυνος για τις αποφάσεις συμμετέχει σε ένα παιχνίδι με παθητικό αντίπαλο, η θεωρία παιγνίων χρησιμοποιεί τη γνώση για τις πιθανές «επιθετικές» κινήσεις του αντιπάλου, για να λύσει το παίγνιο. Επίσης, για την εύρεση λύσης συχνά ανάγεται το παιχνίδι σε ένα άλλο απλούστερο για να μελετηθεί ευκολότερα και δοθεί η δυνατότητα στον παίκτη να δει μπροστά μερικές κινήσεις.

Η κύρια ανάπτυξη της θεωρίας παιγνίων οφείλεται στους μαθηματικούς John Von Neumann (1903 – 1957) και Émile Borel (1871 -1956), ενώ η δημοσίευση του

βιβλίου «Theory of games and economic behavior»¹ των Von Neumann και του οικονομολόγου Oskar Morgenstern έθεσε τις βάσεις σε επιστημονικό επίπεδο.

- Ο John Von Neumann γεννήθηκε στην Βουδαπέστη (Ουγγαρία) το 1903 και από μικρή ηλικία φάνηκε η μαθηματική του ευφυΐα. Σπούδασε μαθηματικά στο Πανεπιστήμιο της Βουδαπέστης, χημεία στο Πανεπιστήμιο του Βερολίνου και αποφοίτησε από το Ελβετικό Ομοσπονδιακό Ινστιτούτο Τεχνολογίας της Ζυρίχης ως χημικός μηχανικός, ενώ στη συνέχεια εργάστηκε στο Institute of Advanced Study στο Princeton μαζί με τον Albert Einstein. Επίσης ήταν σύμβουλος στην C.I.A., τον Αμερικανικό Στρατό, τη Standard Oil, την General Electric, την IBM και πολλές άλλες εταιρίες. Αξίζει να αναφέρουμε πως γνώριζε λατινικά, αρχαία ελληνικά και είχε εξαιρετικές γνώσεις στην Βυζαντινή Ιστορία. Ο πολυπράγμων αυτός άνθρωπος έσβησε από την επάρατη νόσο το 1957.

Στη συνέχεια, σημαντική μορφή για τη θεωρία Παιγνίων αποτέλεσε ο John F. Nash ο οποίος εισήγαγε την έννοια της Ισορροπίας (Nash Equilibrium). Σε εκείνο το σημείο άρχισε η οικονομική θεωρία, οι πολιτικές και συμπεριφορικές επιστήμες να μελετούν και να χρησιμοποιούν τα παιγνιοθεωρητικά μοντέλα, ενώ το 1970 χρησιμοποιήθηκε και από την εξελικτική βιολογία. Το 1994, βραβείο Νόμπελ οικονομίας απέσπασαν οι επιστήμονες της θεωρίας Παιγνίων John Nash, John Harsanyi και Reinhard Selten.

- Ο John (Forbes) Nash γεννήθηκε το 1928 στο Bluefield της Δυτικής Βιρτζίνια. (Η.Π.Α.). Σπούδασε μαθηματικά στο Τεχνολογικό Ινστιτούτο του Carnegie (Carnegie Mellon University), ενώ την ακαδημαϊκή του καριέρα συνέχισε με υποτροφία στο Πανεπιστήμιο του Πρίνστον, αργότερα στο Harvard ενώ δούλεψε και ως καθηγητής στο Τεχνολογικό Ινστιτούτο Μασαχουσέτης (M.I.T.). Το 2011 η NASA εμφάνισε επιστολές του Nash από το έτος του 1950 στις οποίες πρότεινε ένα νέο σύστημα κρυπτογράφησης παρόμοιο με το σύγχρονο.

Πολλοί από τους πρώτους ερευνητές του τομέα αυτού υποστηρίχθηκαν από το Αμερικανικό Υπουργείο Άμυνας και εργάστηκαν πάνω στην ατομική βόμβα ως όπλο

¹ Von Neumann J., Morgenstern Oskar (1953), Theory of games and Economic Behavior, Princeton University Press

για τον πόλεμο ενάντια στη Σοβιετική Ένωση. Για το λόγο αυτό, στο ευρύ κοινό η θεωρία παιγνίων ήταν ταυτόσημη με τη μελέτη στα «παιχνίδια πολέμου» και θεωρούνταν σκληρή και απάνθρωπη. Γενικότερα όμως η απήχηση της πέρασε από στάδια ύφεσης και ανάπτυξης μέχρι να μπορούμε πια να ξεχωρίσουμε πως δεν είναι ούτε η σωτήρια θεωρία αλλά ούτε κι ένα απλό μαθηματικό παιχνίδι.

Ορισμοί:

- **Παίγνιο:** Μία ανταγωνιστική κατάσταση κατά την οποία οι παίκτες (δύο ή και περισσότεροι) επιλέγουν τους τρόπους με τους οποίους θα ενεργήσουν.
- **Παίκτης:** Η μονάδα που αποφασίζει για τις κινήσεις μέσα στο παιχνίδι –για παράδειγμα άτομο, επιχείρηση, πολιτική παράταξη- και προσπαθεί να βελτιστοποιήσει προς όφελος της το αποτέλεσμα σύμφωνα με τους κανόνες και με πιθανές πληροφορίες που μπορεί να διαθέτει..
- **Στρατηγική:** Οι προκαθορισμένοι κανόνες του παιγνίου που ορίζουν τις δυνατές κινήσεις που μπορεί να ακολουθήσει κάθε παίκτης μέχρι το τέλος.
 - **Αμιγής στρατηγική:** Η επιλογή μίας από τις εφικτές στρατηγικές του παίκτη ως βέβαιο ενδεχόμενο.
 - **Μικτή στρατηγική:** Η επιλογή ενός συνδυασμού εφικτών στρατηγικών.
- **Πίνακας Αποδόσεων:** Δείχνει τις αμοιβές (αρνητικές ή θετικές) για τον παίκτη που μελετάμε για κάθε συνδυασμό στρατηγικών.
- **Λύση:** Η βέλτιστη στρατηγική όλων των παικτών.

Βασικές Υποθέσεις:

- Όλοι οι παίκτες σκέφτονται και λειτουργούν με λογική (ορθολογικοί παίκτες).
- Όλοι οι παίκτες αποφασίζουν να ακολουθήσουν τη στρατηγική τους αποκλειστικά για δικό τους όφελος, χωρίς συμπάθεια για τον αντίπαλο.

ΚΕΦΑΛΑΙΟ 2^ο

ΑΠΛΑ ΜΟΝΤΕΛΑ ΣΤΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ

2.1.ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Έστω ότι αντιμετωπίζουμε ένα πρόβλημα στο οποίο πρέπει να λάβουμε μία απόφαση. Μία προσέγγιση σε αυτό θα μπορούσε να είναι ο προσδιορισμός του επιθυμητού αποτελέσματος και μία συμπεριφορά ανάλογη η οποία να οδηγεί σε αυτό. Δεν είναι όμως πάντοτε εφικτό το επιθυμητό αποτέλεσμα. Έτσι αναζητούμε μία εναλλακτική προσέγγιση, κατά την οποία προσδιορίζεται μία σειρά δυνατών δράσεων και το αντίστοιχο αποτέλεσμα της καθεμιάς. Τότε επιλέγεται ένα από αυτά τα αποτελέσματα -και η αντίστοιχη συμπεριφορά από την οποία προήλθε- διότι αυτό (το επιλεγθέν αποτέλεσμα) είναι που μεγιστοποιεί την τιμή του στόχου, ή αντίστοιχα ελαχιστοποιεί όταν αναφερόμαστε σε ζημία. Αυτή λοιπόν, η δεύτερη προσέγγιση, καλείται «λήψη της βέλτιστης απόφασης» και σκοπός μας είναι να μελετήσουμε το μαθηματικό πλαίσιο για την ανάπτυξη της σε ένα πλήθος συνθηκών και καταστάσεων.

Η εύρεση μέγιστης τιμής είναι μία γνωστή διαδικασία στα βασικά μαθηματικά (αριθμητική ανάλυση). Αν για παράδειγμα αναζητούμε το μέγιστο μίας συνεχούς και παραγωγίσιμης συνάρτησης $f(x)$, τότε αρκεί να την παραγωγίσουμε, να εντοπίσουμε τη ρίζα ή τις ρίζες της παραγώγου (έστω x^*) και το πρόσημο στην γύρω περιοχή. Με αυτό τον τρόπο βρίσκουμε που επιτυγχάνεται η μέγιστη τιμή της συνάρτησης, δηλαδή την τιμή $f(x^*)$. Παρ' όλα αυτά θα αντιμετωπίσουμε προβλήματα στα οποία οι συναρτήσεις μας δεν είναι απαραίτητα συνεχείς ή/και παραγωγίσιμες, όπως οι συναρτήσεις σε διακριτό χώρο.

Ας δούμε λοιπόν μερικά παραδείγματα:

Παράδειγμα: Έστω η συνάρτηση $f(x) = \sqrt{x}$ ορισμένη στο $[0, +\infty)$ η οποία αυξάνεται συνεχώς όσο μεγαλώνει η τιμή του x , συνεπώς δεν παρουσιάζει μέγιστο. Αν όμως θεωρήσουμε την ίδια συνάρτηση αλλά ορισμένη στο διάστημα $[0, 9]$, τότε έχουμε μέγιστο. Δεδομένου ότι η f είναι αύξουσα συνάρτηση στο πεδίο ορισμού της, τότε για $x^* = 9$ έχουμε μέγιστο την τιμή $f(9)=3$.

Παράδειγμα : Έστω α, β και γ θετικές σταθερές και οι συναρτήσεις:

$$f(x) = \begin{cases} 1 - \frac{x}{\beta} & \text{αν } 0 \leq x \leq \beta \\ 0 & \text{αλλού} \end{cases}$$

$$g(x) = \begin{cases} 0 & \text{αν } x \leq 0 \\ 1 & \text{αν } x > 0 \end{cases}$$

Βλέπουμε ότι για $x=0$ η $f(x)$ μεγιστοποιείται: $f(0) = 1$. Επίσης η $-g(x)$ μεγιστοποιείται όταν η $g(x)$ παίρνει την ελάχιστη τιμή της, δηλαδή για $x=0$, $g(0) = 0$

Συνεπώς η συνάρτηση $h(x) = \alpha x f(x) - \gamma g(x)$ μεγιστοποιείται όταν $x=0$, με $h(0) = 0$.

Ορισμός: Έστω ένα τυχαίο στοιχείο x ενός συνόλου X και $f(x)$ μία συνάρτηση ορισμένη στο X . Τότε

$$x^* \in \operatorname{argmax}_{x \in X} f(x) \Leftrightarrow f(x^*) = \max_{x \in X} f(x)$$

όπου x^* είναι η τιμή που μεγιστοποιεί την $f(x)$. Ας σημειωθεί βεβαίως ότι δεν ισχύει η ισότητα $x^* = \operatorname{argmax}_{x \in X} f(x)$ διότι μία συνάρτηση μπορεί να λαμβάνει περισσότερες από μία τιμή στο X που να δίνουν μέγιστο. Για παράδειγμα οι συναρτήσεις $f_n(x) = x^n$ με $x \in [-a, a]$, όπου a θετικός ακέραιος και n άρτιος φυσικός αριθμός: Κάθε μία έχει δύο μέγιστες τιμές $x^{*1} = -a$ και $x^{*2} = +a$. Δεδομένου λοιπόν, ότι το argmax επιστρέφει ένα σύνολο τιμών, κι όχι μόνο μία, δεν θεωρείται συνάρτηση αλλά «ανταπόκριση».

2.2.ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ

Την πιο απλή συνθήκη προς συλλογισμό αποτελεί η *μη τυχειότητα στο περιβάλλον*: κάθε επιλογή που κάνουμε επιφέρει ένα βέβαιο αποτέλεσμα. Παρακάτω θα ορίσουμε κάποιες έννοιες απαραίτητες για τη θεωρία της βέλτιστης απόφασης.

Ορισμός: Μία επιλογή συμπεριφοράς σε ένα πρόβλημα που απαιτείται η λήψη μίας μοναδικής απόφασης, καλείται κίνηση (ή δράση). Το σύνολο των διαθέσιμων εναλλακτικών κινήσεων συμβολίζεται A .

Ορισμός: Η συνάρτηση απόδοσης $\pi: A \rightarrow \mathbb{R}$, είναι η συνάρτηση που συνδέει μία αριθμητική τιμή με κάθε κίνηση του συνόλου A .

Ορισμός: Μία κίνηση α^* στο A καλείται βέλτιστη εάν

$$\pi(\alpha^*) \geq \pi(\alpha) \quad \text{για κάθε } \alpha \text{ στο } A \text{ ή ισοδύναμα } \alpha^* \in \operatorname{argmax}_{\alpha \in A} \pi(\alpha)$$

Αυτό σημαίνει πως η βέλτιστη απόφαση είναι να επιλεγθεί ένα $\alpha^* \in A$ το οποίο να μεγιστοποιεί τη συνάρτηση απόδοσης $\pi(x)$. Προφανώς, όπως προαναφέραμε, η τιμή του α^* δεν είναι πάντα μοναδική, αρκεί, όποια και να επιλέξουμε, να οδηγεί στο μέγιστο της $\pi(x)$.

Παράδειγμα: Σε ένα νέο άνεργο προσφέρονται δύο θέσεις για δουλειά η W_1 και η W_2 . Οι πιθανές κινήσεις του άνεργου είναι α_1 : Να δεχτεί να εργασθεί στην W_1 και α_2 : Να δεχτεί να εργασθεί στην W_2 . Οι αποδόσεις αντιστοιχούν στο μισθό που προσφέρει κάθε εταιρία: Η W_1 ανταμείβει με 21.000 ευρώ/έτος, ενώ η δεύτερη W_2 με 24.000 ευρώ/έτος

Έτσι έχουμε $\pi(\alpha_1)=21000 < \pi(\alpha_2)=24.000$. Συνεπώς ο νέος άνεργος θα επιλέξει την W_2 .

Ορισμός: Ορίζεται ο μετασχηματισμός της $\pi(x)$ σε $\pi^*(x)$ ως εξής:

$$\pi^*(x) = c\pi(x) + d, \text{ όπου } c, d \text{ πραγματικές σταθερές ανεξάρτητες του } x, \text{ με } c > 0.$$

Θεώρημα: Η βέλτιστη κίνηση παραμένει αμετάβλητη αν οι αποδόσεις μετασχηματιστούν σύμφωνα με τον παραπάνω ορισμό.

$$\begin{aligned} \text{Απόδειξη: Εφόσον } c > 0, \text{ έχουμε: } \operatorname{argmax} \pi^*(x) &= \operatorname{argmax}[c\pi(x)+d] \\ &= \operatorname{argmax} \pi(x). \end{aligned}$$

■

Συχνά, στα προβλήματα απόφασης, όπως προαναφέραμε, οι κινήσεις δεν ανήκουν απαραίτητα σε συνεχή υποσύνολα του $\mathbb{R}$. Πολλές φορές, όμως μπορούμε να παραβλέπουμε αυτό τον περιορισμό για να επιλύσουμε το πρόβλημα και να τον λαμβάνουμε υπόψη στη συνέχεια. Για παράδειγμα, αν ένας έμπορος πουλά ένα προϊόν και σκέφτεται να κάνει μία έκπτωση στην τιμή πώλησης από 0.50 ευρώ έως 4.00 ευρώ, τότε το διάστημα $[0.5, 4]$ είναι συνεχές αλλά δεν υπάρχουν νομίσματα για κάθε τιμή αυτού του διαστήματος, οπότε μιλάμε για διακριτό σύνολο. Προτιμάται όμως να αντιμετωπιστεί ως συνεχές διάστημα και να κάνουμε ίσως μία στρογγυλοποίηση στο αποτέλεσμα εάν χρειαστεί.

Παράδειγμα:² Μία ελληνική εταιρία παραγωγής πορτοκαλάδας (Orange Co.) πουλά το χυμό p ευρώ/lt. Ένα κατάστημα super market σκοπεύει να αγοράσει $Q(p)$ λίτρα. Αν Q_0 είναι η μέγιστη ποσότητα που θα μπορούσε να αγοράσει σε οποιαδήποτε τιμή πώλησης του χυμού και p_0 μία ακόμη σταθερά που δηλώνει τη μέγιστη τιμή ανά λίτρο που θα μπορούσε να πληρώσει για να αγοράσει το χυμό, τότε η συνάρτηση $Q(p)$ διαμορφώνεται ως εξής:

$$Q(p) = Q_0 \left(1 - \frac{p}{p_0}\right) \quad \text{αν } p < p_0$$

$$\text{ή } Q(p) = 0 \quad \text{αν } p \geq p_0$$

Οι δυνατές δράσεις της Orange Co. είναι να διατηρήσει την τιμή p στο διάστημα $[0, p_0]$, αλλιώς το super market δεν θα αγοράσει καθόλου χυμό. Αν θεωρήσουμε το καθαρό κέρδος ως απόδοση για την Orange Co. με δεδομένο κόστος παραγωγής χυμού c , έχουμε:

$$\pi(p) = (p - c)Q(p)$$

Αναζητούμε τώρα την τιμή p^* που μεγιστοποιεί την απόδοση, δηλαδή το καθαρό κέρδος. Αρκεί λοιπόν να βρούμε τη ρίζα της εξίσωσης:

$$\begin{aligned} \frac{d\pi(p)}{dp} &= 0 \Rightarrow \\ Q_0 \left(1 - \frac{2p - c}{p_0}\right) &= 0 \Rightarrow \\ p &= \frac{p_0 + c}{2} \end{aligned}$$

Άρα η τιμή πώλησης που μεγιστοποιεί το κέρδος για την Orange Co. είναι $p^* = \frac{p_0 + c}{2}$.

2.3 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΟΡΘΟΛΟΓΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ

Ένας πλούσιος φιλάνθρωπος προσεγγίζει ένα άτομο για να του προσφέρει χρήματα. Του δίνει τις εξής δύο επιλογές: είτε να δεχτεί 1 ευρώ με πιθανότητα 1, είτε 5 ευρώ με πιθανότητα 0,5 (προφανώς μπορεί να μην λάβει καθόλου χρήματα με πιθανότητα επίσης 0,5). Τι πρέπει να κάνει το άτομο σε αυτή την περίπτωση; Τι θα συνέβαινε αν τα ποσά ήταν μεγαλύτερα; Ποια είναι η ορθολογική συμπεριφορά; Σύμφωνα με όσα έχουμε πει μέχρι τώρα θα έπρεπε να δεχτεί τη δεύτερη επιλογή γιατί

² Webb James N. (2006). Game Theory: Decisions, Interaction and Evolution, Springer

το πιθανό κέρδος είναι μεγαλύτερο. Αν όμως τα ποσά ήταν 1 εκατομμύριο ευρώ και 5 εκατομμύρια αντίστοιχα, οι περισσότεροι θα επέλεγαν την ασφαλή οδό (1 εκατ. με πιθανότητα 1). Είναι αυτή ορθολογική συμπεριφορά;³

Πράγματι, η μεγιστοποίηση δεν είναι πάντα η σοφή επιλογή και μάλιστα υπάρχουν αρκετά επιχειρήματα που το επιβεβαιώνουν. Αρχικά πρέπει να λάβουμε υπόψη πως υπάρχουν αγαθά τέτοια όπως οι καλοκαιρινές διακοπές, η φιλία και η υγεία τα οποία δεν μπορούμε να τα «κοστολογήσουμε» και είναι πολύ σημαντικά για τον άνθρωπο, ειδικά σε σύγκριση με τα ίδια τα χρήματα. Έπειτα, τα χρήματα για πολλούς είναι απλά το μέσο για την επίτευξη ενός στόχου κι όχι ο στόχος καθαυτός. Όπως αναφέραμε και παραπάνω, κάποιος στην περίπτωση που είχε να επιλέξει ανάμεσα σε 1.000.000 ευρώ ως βέβαιο ενδεχόμενο και σε 5.000.000 ευρώ με πιθανότητα 0,5 θα επέλεγε το πρώτο γιατί του δίνει τη δυνατότητα να πραγματοποιήσει όνειρα ή στόχους που δεν μπορούσε πριν, ενώ στην δεύτερη επιλογή ή εξασφαλίζει τους ίδιους πάλι στόχους αν κερδίσει ή κανέναν. Τα όνειρα δεν αυξάνονται απαραίτητα όταν αυξάνεται το χρηματικό ποσό που διαθέτει κάποιος. Για παράδειγμα αν είχε την επιθυμία για την αγορά ενός πολυτελούς αυτοκινήτου που τα έξοδα αγοράς, συντήρησης και μίας άνετης ζωής για τον ίδιο καλύπτονται τόσο από το 1 εκατομμύριο όσο και από τα 5 δεν θα μπει στη διαδικασία τζόγου.

Ένας ακόμα σημαντικός παράγοντας, ο οποίος είναι αντίθετος στην ιδέα πως η μεγιστοποίηση της *προσδοκώμενης χρηματικής αξίας* (expected monetary value-EMV) είναι πάντα ορθολογική, είναι πως αλλάζει η αιτία μίας συμπεριφοράς με τις συνέπειες αυτής. Η σχέση αιτίας-αιτιατού αντιστρέφεται. Θέλουμε λοιπόν να ορίσουμε τι είναι ορθολογικό και να προσδιορίσουμε με βάση αυτό τον ορισμό οποιαδήποτε ποσότητα που μεγιστοποιείται. Αν μπορούμε να το κάνουμε αυτό τότε μπορούμε να συνεχίσουμε βρίσκοντας τις τιμές που μεγιστοποιούν κι να τις αντικαταστήσουμε στην προσδοκώμενη χρηματική αξία (EMV). Πώς όμως ορίζουμε τι είναι ορθολογικό;

Αρχικά πρέπει να αναφέρουμε πως ο ορθολογισμός δεν ισοδυναμεί απαραίτητα με αμερόληπτους συλλογισμούς του νου. Συχνά, οι επιθυμίες και οι προτιμήσεις ενός ατόμου είναι αυτές που τον οδηγούν στο αποτέλεσμα, χωρίς αυτό να σημαίνει πως πρέπει αυτές οι προτιμήσεις να συμβαδίζουν με κάποιου άλλου στην

³ Axzo Press LLC, Project Management: Intermediate (2002), ILT

αντίστοιχη θέση. Όμως, πρέπει αυτές οι ενδόμυχες επιθυμίες και προτιμήσεις του ατόμου να διακρίνονται από σταθερότητα. Με αυτό τον τρόπο ορίζουμε το *ορθολογικό* άτομο, ως εκείνον που είναι σταθερός στις προτιμήσεις του και προσπαθεί να πετύχει τα επιθυμητά αποτελέσματα με γνώμονα αυτές.

Έστω ότι έχουμε ένα σύνολο δυνατών αποτελεσμάτων, δίνουμε για παράδειγμα ένα δωρεάν γεύμα σε μία ομάδα ατόμων με το δικαίωμα να διαλέξουν ανάμεσα σε πίτσα, ή σαλάτα και κάθε ένας μπορεί να επιλέξει ο,τι από τα δύο προτιμά, χωρίς να δοθεί ένα κοινό πιάτο σε όλους. Γνωρίζουμε επίσης, πως από τη στιγμή που έχουμε βέβαια ενδεχόμενα, η επιλογή του καθενός οδηγεί στο επιθυμητό αποτέλεσμα. Έτσι λοιπόν, ένα ορθολογικό άτομο θα επιλέξει αυτό που θέλει να φάει. Άρα αν κάποιος επιλέξει σαλάτα, ενώ επιθυμούσε να φάει πίτσα δεν θεωρείται ορθολογικός, ενώ κάποιος που είπε πως θα ήθελε να φάει πίτσα αλλά λόγω δίαιτας επιλέγει σαλάτα απλώς απέτυχε να συμπεριλάβει τις επιθυμίες του στην προσπάθεια περάτωσης του στόχου (να κάνει δίαιτα).

Αν θεωρήσουμε το σύνολο όλων των δυνατών ενδεχομένων $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$, λέμε ότι κάποιος επιλέγει «αυστηρά» το αποτέλεσμα ω_1 έναντι του ω_2 γράφοντας $\omega_1 > \omega_2$, ενώ αν κάποιος απλά επιλέγει χωρίς την αυστηρή έννοια το αποτέλεσμα ω_1 έναντι του ω_2 γράφουμε $\omega_1 \geq \omega_2$ (αν δηλαδή προτιμά εξίσου και τα δύο) και αν κάποιος είναι αδιάφορος για τα δύο αυτά αποτελέσματα γράφουμε $\omega_1 \sim \omega_2$.

Ορισμός: Ένα άτομο καλείται *ορθολογικό υπό βεβαιότητα* αν οι επιλογές του για το αποτέλεσμα ικανοποιούν τις ακόλουθες ιδιότητες:

- i. Είτε $\omega_i > \omega_j$, ή $\omega_j > \omega_i$, για κάθε $i, j = 1, 2, 3, \dots$ και $i \neq j$ (πληρότητα)
- ii. Αν $\omega_i > \omega_k$ και $\omega_k > \omega_j$, τότε $\omega_i > \omega_j$ για κάθε $i, j, k = 1, 2, 3, \dots$ και $i \neq j \neq k$ (μεταβατική)

Με βάση αυτό τον ορισμό εισάγεται τώρα η έννοια της συνάρτησης ωφέλειας για κάθε ενδεχόμενο ω . Κάθε άτομο θεωρείται ότι έχει τη δική του συνάρτηση ωφέλειας $u(\omega)$ η οποία μπορεί να αντιστοιχεί σε αριθμό (π.χ. χρήματα) ή όχι (π.χ. κατάσταση υγείας)⁴. Όποια κι αν είναι λοιπόν η ανταμοιβή, η συνάρτηση ωφέλειας περικλείει κάθε σημαντική πληροφορία που πρέπει να λάβει υπόψιν το άτομο.

⁴ Όπως και στη στατιστική, θεωρούμε κατηγορικές μεταβλητές όσες δεν έχουν να κάνουν με νούμερα, αλλά τυποποιούμε με κάποιον αριθμό για λόγους ευκολίας στην επίλυση του προβλήματος.

Ορισμός: Η συνάρτηση $u: \Omega \rightarrow \mathbb{R}$ καλείται συνάρτηση ωφέλειας εάν:

$$u(\omega_i) > u(\omega_j) \Leftrightarrow \omega_i \succ \omega_j,$$

$$u(\omega_i) = u(\omega_j) \Leftrightarrow \omega_i \sim \omega_j,$$

για $i, j = 1, 2, 3, \dots$ και $i \neq j$

Μία άμεση συνέπεια αυτού του ορισμού είναι ότι ένα *ορθολογικό υπό βεβαιότητα* άτομο πρέπει να αναζητά μεθόδους μεγιστοποίησης της συνάρτησης ωφέλειας. Η σχέση μεταξύ της συνάρτησης ωφέλειας $u(\omega)$ και της συνάρτησης απόδοσης $\pi(\alpha)$ (μίας κίνησης α που αντιστοιχίζεται σε ένα αποτέλεσμα $\omega(\alpha)$) είναι άμεση: $\pi(\alpha) = u(\omega(\alpha))$.

Ας δούμε όμως στη συνέχεια τι γίνεται όταν δεν έχουμε βέβαια ενδεχόμενα, αλλά όταν μία πράξη α «προκαλεί» ένα αποτέλεσμα ω με θετική πιθανότητα μικρότερη της μονάδας. Αυτό που αποκαλούμε «λοταρία».

Ορισμός: Μία απλή λοταρία λ , είναι ένα σύνολο πιθανοτήτων για κάθε ενδεχόμενο $\omega \in \Omega$. Κάθε πιθανότητα σε αυτό το σύνολο γράφεται $p(\omega|\lambda)$.

Ορισμός: Μία σύνθετη λοταρία είναι ένας γραμμικός συνδυασμός απλών.

Για παράδειγμα, αν $0 < p^* < 1$ και λ_1, λ_2 δύο απλές λοταρίες, τότε η

$$\lambda = (1 - p^*) \lambda_1 + p^* \lambda_2 \quad \text{είναι σύνθετη λοταρία.}$$

Ορισμός: Ένα άτομο καλείται ορθολογικό αν οι προτιμήσεις του στη λοταρία ικανοποιούν τις εξής συνθήκες:

- i. Είτε $\lambda_i \succcurlyeq \lambda_j$ ή $\lambda_j \succcurlyeq \lambda_i$ (πληρότητα)
- ii. Αν $\lambda_i \succcurlyeq \lambda_k$ και $\lambda_k \succcurlyeq \lambda_j$, τότε $\lambda_i \succcurlyeq \lambda_j$ (μεταβατικότητα)
- iii. Αν $\lambda_i \succ \lambda_j$ και $c_i > c_j$, τότε $c_i \lambda_i + (1 - c_i) \lambda_j \succ c_j \lambda_i + (1 - c_j) \lambda_j$ (μονοτονία)
- iv. Αν $\lambda_i \succcurlyeq \lambda_k$ και $\lambda_k \succcurlyeq \lambda_j$, τότε υπάρχει μία πιθανότητα q τέτοια ώστε $\lambda_k \sim q\lambda_i + (1 - q) \lambda_j$ (συνέχεια)
- v. Αν $\lambda_i \succ \lambda_k$ τότε $q\lambda_i + (1 - q) \lambda_j \succ q\lambda_k + (1 - q) \lambda_j$ (ανεξαρτησία)

Όπου $i, j, k = 1, 2, 3, \dots$ με $i \neq j \neq k$ και $0 \leq c_i, q \leq 1$

Αξίζει εδώ να αναφέρουμε πως η πληρότητα επιβεβαιώνει ότι όλες οι λοταρίες μπορούν να συγκριθούν μεταξύ τους και η μεταβατικότητα δείχνει πως

μπορούν να ταξινομηθούν βάση των προτιμήσεων. Η μονοτονία και η συνέχεια δείχνουν ότι η λοταρία γίνεται καλύτερη καθώς η πιθανότητα του προτιμώμενου αποτελέσματος αυξάνεται. Η ανεξαρτησία δείχνει πως οι προτιμήσεις εξαρτώνται μόνο από τις διαφορές ανάμεσα στις λοταρίες, ενώ ίδια στοιχεία μπορούν να παραληφθούν.

Υποθέτουμε τώρα πως επιλέγοντας μία κίνηση α προκύπτει μία λοταρία $\lambda(\alpha)$. Ποιά είναι τότε η συνάρτηση απόδοσης που πρέπει ένα ορθολογικό άτομο να προσπαθήσει να μεγιστοποιήσει; Εδώ εισάγεται η έννοια της συνάρτησης ωφέλειας για το αναμενόμενο ενδεχόμενο $E(\omega)$. Αντιμετωπίζουμε όμως δύο προβλήματα με αυτό. Αρχικά δεν είναι ξεκάθαρη αυτή η έννοια όταν τα ενδεχόμενα δεν είναι αριθμητικά. Έπειτα, ακόμα κι αν ήταν αριθμητικά δεν θα ήταν απαραίτητο να υπολογίζουμε τη μέγιστη τιμή οποιασδήποτε συνάρτησης αναμενόμενου ενδεχομένου.

Θεώρημα: (Θεώρημα Αναμενόμενης Ωφέλειας)⁵

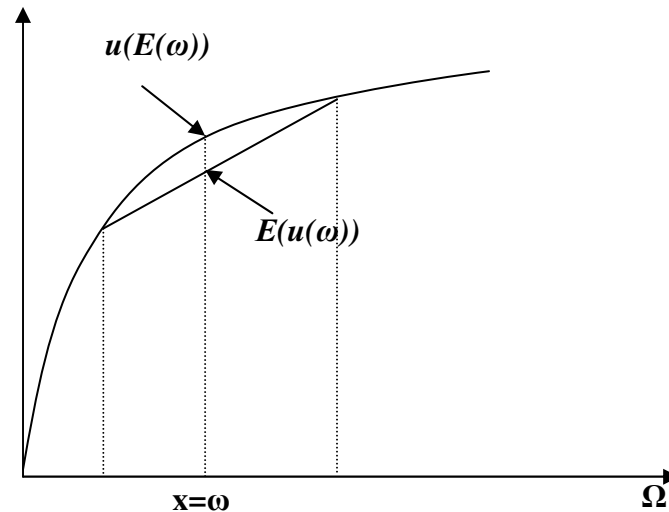
Για κάθε ορθολογικό άτομο (ορσ. 9^{ος}) μπορούμε να ορίσουμε μια συνάρτηση ωφέλειας $u: \Omega \rightarrow \mathbb{R}$. Τότε, δεδομένου οι κινήσεις γίνονται ορθολογικά ώστε να μεγιστοποιούν τη συνάρτηση απόδοσης $\pi(\alpha)$, η ίδια καλείται αναμενόμενη ωφέλεια και δίνεται από τον τύπο :

$$\pi(\alpha) = \sum_{\omega \in \Omega} p(\omega|\lambda(\alpha))u(\omega)$$

Ορισμός 10^{ος}: Έστω μία συνάρτηση ωφέλειας $u: \Omega \rightarrow \mathbb{R}$ για ένα άτομο. Αν ικανοποιείται η σχέση:

- i. $E(u(\omega)) < u(E(\omega))$ τότε λέμε ότι είναι σε αντίθεση ρίσκου
- ii. $E(u(\omega)) > u(E(\omega))$ τότε λέμε ότι είναι σε τάση ρίσκου
- iii. $E(u(\omega)) = u(E(\omega))$ τότε λέμε ότι είναι ουδέτερος ρίσκου

⁵ Myerson Roger (1991), Game Theory: Analysis of Conflict



Σχήμα 2.1 Σύμφωνα με τον ορισμό φαίνεται πως σύμφωνα με την καμπύλη της συνάρτησης ωφέλειας το άτομο είναι σε θέση κατά του ρίσκου (αντίθετος ρίσκου).

Παράδειγμα: (Κλασικό πρόβλημα επιλογής χαρτοφυλακίου) Δύο περιουσιακά στοιχεία είναι διαθέσιμα σε έναν επενδυτή. Η πρώτη επιλογή δεν έχει καθόλου ρίσκο (ένας λογαριασμός ταμειευτηρίου) και του επιστρέφει ένα σταθερό ποσό έναντι του αρχικού, έστω r (σταθερός τόκος). Η άλλη επιλογή έχει ρίσκο (χρηματιστήριο) και η απόδοση είναι μεταβλητή με μέσο μ και σταθερή απόκλιση σ . Αν ο επενδυτής ανήκει στην κατηγορία αυτών που τείνουν να μεγιστοποιούν τη Προσδοκώμενη Χρηματική Αξία (EMV), τότε θα έπρεπε να επενδύσει όλα του χρήματα στη δεύτερη επιλογή αν ο μέσος $\mu > r$. Παρ' όλα αυτά, υπό ορισμένες συνθήκες μπορεί ένας επενδυτής αντίθετος ρίσκου να επιλέξει με ένα γραμμικό τρόπο την μεταβλητή απόδοση περιορίζοντας με κάποιο τρόπο τη μεταβλητότητα της απόδοσης συνολικά. Αυτό επιτυγχάνεται δημιουργώντας ένα χαρτοφυλάκιο σύμφωνα με το οποίο θα τοποθετούσαμε ένα μέρος των χρημάτων στην τράπεζα (με σταθερή απόδοση r) και τα υπόλοιπα στο χρηματιστήριο. Αν θεωρήσουμε α το μέρος των χρημάτων που επενδύθηκαν στο χρηματιστήριο, τότε η αναμενόμενη απόδοση από αυτό το χαρτοφυλάκιο είναι $\alpha\mu + (1-\alpha)r$ με διασπορά $\alpha^2\sigma^2$. Συνεπώς η αναμενόμενη ωφέλεια για τον επενδυτή είναι:

$$\pi(\alpha) = \alpha\mu + (1 - \alpha)r - \frac{c}{2}\alpha^2\sigma^2$$

Όπου c είναι η παράγοντας που ορίζει ο επενδυτής στη διασπορά ανάλογα με τις προσδοκίες. Πότε λοιπόν μεγιστοποιείται η αναμενόμενη ωφέλεια;

Έχουμε:

$$\begin{aligned}\frac{d\pi(\alpha)}{d\alpha} &= -c\sigma^2\alpha + \mu - r \\ \frac{d\pi(\alpha)}{d\alpha} = 0 &\Leftrightarrow -c\sigma^2\alpha + \mu - r = 0 \\ &\Leftrightarrow \alpha = \frac{\mu - r}{c\sigma^2} \\ \frac{d^2\pi(\alpha)}{d\alpha^2} &= -c\sigma^2 < 0\end{aligned}$$

Οπότε συνοψίζοντας έχουμε ότι η αναμενόμενη ωφέλεια μεγιστοποιείται όταν:

$$\alpha^* = \begin{cases} 0 & \text{αν } \mu < r \\ \frac{\mu - r}{c\sigma^2} & \text{αν } 0 < \mu - r < c\sigma^2 \\ 1 & \text{αν } \mu - r > c\sigma^2 \end{cases}$$

2.4.ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΦΥΣΙΚΗΣ ΕΠΙΛΟΓΗΣ

Σε αυτή την παράγραφο θα δούμε πως εφαρμόζεται στη συμπεριφορά των ζώων αυτό που τα μαθηματικά περιγράφουν ως βέλτιστη απόφαση από ορθολογικό άτομο. Η έννοια της βέλτιστης συμπεριφοράς των ζώων βασίζεται στην ακόλουθη ερμηνεία που δίνει η *φυσική επιλογή*. Κάθε ζώο υιοθετεί μία συμπεριφορά κοινή για όλο τον πληθυσμό (π.χ. αγέλες, σμήνη) στον οποίο ανήκει λόγω γενετικού υλικού. Ασφαλώς η σχέση συμπεριφοράς και γενετικής είναι αρκετά περίπλοκη κι ελάχιστα κατανοητή, αλλά παρατηρείται ότι συμπεριφορές κληροδοτούνται από γενιά σε γενιά, σύμφωνα με την *φαινοτυπική πλαστικότητα*.⁶

Ανάλογα λοιπόν με τη συμπεριφορά που έχει υιοθετήσει ένα ζώο έχει στην πορεία λίγους ή πολλούς απογόνους. Οι γονείς μεταβιβάζουν τη γενετική πληροφορία συμπεριφοράς στους απογόνους κι έτσι ο πληθυσμός του αυξάνεται ή όχι, οπότε το είδος του ζώου που αφήνει περισσότερους απογόνους επικρατεί αριθμητικά.

⁶ http://en.wikipedia.org/wiki/Phenotypic_plasticity

Παράδειγμα:⁷ Έστω ένας πληθυσμός ζώων στον οποίο έχουν περάσει δύο είδη γενετικού υλικού για τη συμπεριφορά τους. Λαμβάνουμε υπόψη πως τα ζώα ζουν ένα έτος, γεννούν μία φορά και έπειτα πεθαίνουν. Ανάλογα λοιπόν το γενετικό υλικό (τύπος $i = 1$ ή 2) αποκτούν s_i απογόνους. Δεδομένου ότι $s_1 < s_2$, υποθέτουμε ότι τη χρονική στιγμή t θα υπάρχουν $n_i(t)$ το πλήθος απόγονοι από το είδος i . Ποιό είναι το επικρατέστερο είδος;

Την χρονική στιγμή $t + 1$ θα είναι σε πλήθος $n_i(t + 1) = s_i n_i(t)$ για το είδος i . Έτσι εξελίσσονται γεωμετρικά ως εξής:

$$\text{Για } t = 1, \quad n_i(1) = s_i n_i(0)$$

$$\text{Για } t = 2, \quad n_i(2) = s_i n_i(1) = s_i^2 n_i(0),$$

$$\text{Για } t = 3, \quad n_i(3) = s_i n_i(2) = s_i s_i^2 n_i(0) = s_i^3 n_i(0),$$

...

$$\text{Για } t = k, \quad n_i(k) = s_i n_i(k - 1) = s_i s_i^{k-1} n_i(0) = s_i^k n_i(0)$$

Για να δούμε τον επικρατέστερο πληθυσμό αρκεί να εξετάσουμε το λόγο:

$$\frac{n_1(k)}{n_2(k)} = \left(\frac{s_1}{s_2} \right)^k \frac{n_1(0)}{n_2(0)}$$

όπου, καθώς $k \rightarrow \infty$ ο λόγος τείνει στο μηδέν. Συνεπώς το επικρατέστερο είδος για τον πληθυσμό είναι το $i=2$ και σύμφωνα με την αρχή της φυσικής επιλογής, αυτή η συμπεριφορά θα επιλεγεί από τα ζώα.

Ορισμός: Η ασυμπτωτική ανάπτυξη του ρυθμού ενός υποπληθυσμού ζώων που χρησιμοποιούν μία συμπεριφορά ορίζεται ως η καταλληλότητα αυτής. Για ζώα με συμπεριφορικό τύπο i , πλήθους $n_i(t)$ τη χρονική στιγμή t , ορίζεται ένας ετήσιος ρυθμός ανάπτυξης:

$$s_i = \frac{n_i(t + 1)}{n_i(t)}$$

Ενώ η καταλληλότητα του τύπου προκύπτει από την απόδοση του ρυθμού:

$$\pi(i) = \lim_{t \rightarrow \infty} s_i(t)$$

Συνεπώς η φυσική επιλογή δείχνει να δρα με τέτοιο τρόπο ώστε να μεγιστοποιεί την καταλληλότητα της συμπεριφοράς που υιοθετεί ένας υποπληθυσμός ζώων και να επικρατεί αυτή. Με άλλα λόγια θεωρούμε ένα σύνολο εύλογων εναλλακτικών συμπεριφορών από το οποίο επιλέγεται μία συμπεριφορά που έχει τη

⁷ Grafen, A. (1991), Modelling in behavioural ecology, in 'Behavioural Ecology: An Evolutionary Approach', 3 edn, Blackwell, Oxford, pp. 5–31.

μέγιστη καταλληλότητα. Βεβαίως, το πλήθος των απογόνων δεν είναι πάντα το σωστό κριτήριο για τη μέτρηση της καταλληλότητας όπως θα δούμε στο επόμενο παράδειγμα.

Παράδειγμα: Έστω ένα ζώο που διαθέτει τις εξής επιλογές για τη ζωή και την αναπαραγωγή.

α_1 : Να γεννήσει 7 απογόνους και έπειτα να πεθάνει.

α_2 : Να γεννήσει το πρώτο έτος της ζωής του 4 απογόνους, το επόμενο έτος 5 απογόνους κι έπειτα να πεθάνει.

Η απόδοση (καταλληλότητα) της $1^{\text{ης}}$ επιλογής είναι προφανώς:

$$\pi(\alpha_1) = \frac{n_1(t+1)}{n_1(t)} = 7$$

Η δεύτερη επιλογή είναι λίγο πιο σύνθετη. Αν $g_1(t)$ το πλήθος των ζώων-γονέων ενός έτους τη χρονική στιγμή t , τότε οι απόγονοι τους θα είναι $4g_1(t)$. Και αν $g_2(t)$ το πλήθος των ζώων-γονέων ηλικίας δύο ετών, τότε το πλήθος των απογόνων τους είναι: $5g_2(t)$. Συνεπώς για τη χρονική στιγμή $t+1$ θα έχουμε:

$$g_1(t+1) = 4g_1(t) + 5g_2(t)$$

και

$$g_2(t+1) = g_1(t)$$

(εξ'ορισμού διότι τα ζώα δύο ετών έχουν πεθάνει άρα μένει το πλήθος των πρωτοετών)

Κι έχουμε: $g_1(t+1) + g_2(t+1) = 4g_1(t) + g_1(t) + 5g_2(t) \Leftrightarrow$

$$g_1(t+1) + g_2(t+1) = 5g_1(t) + 5g_2(t) \Leftrightarrow$$

$$g_1(t+1) + g_2(t+1) = 5[g_1(t) + g_2(t)] \Leftrightarrow$$

$$\frac{[g_1(t+1) + g_2(t+1)]}{[g_1(t) + g_2(t)]} = 5 \Leftrightarrow$$

$$\pi(\alpha_2) = \frac{n_2(t+1)}{n_2(t)} = 5$$

Οπότε: $\pi(\alpha_1) > \pi(\alpha_2)$, δηλαδή η καταλληλότερη συμπεριφορά είναι η πρώτη.⁸

Στην πράξη το αναμενόμενο πλήθος των απογόνων δεν μπορεί να υπολογισθεί ακριβώς. Αυτό που είναι όμως δυνατό, είναι να ξεχωρίσουμε λογικά, μέσα από ένα σύνολο συμπεριφορών, έναν παράγοντα που να επηρεάζει την καταλληλότητα.

⁸ Η θεωρία του «ζήσε γρήγορα, πέθανε νέος/ live fast and die young»

2.5 ΒΕΛΤΙΣΤΗ ΣΥΜΠΕΡΙΦΟΡΑ

Μέχρι τώρα σε προβλήματα απόφασης αναζητούσαμε τη βέλτιστη κίνηση που μεγιστοποιεί την απόδοση. Σε αυτό όμως το σημείο πρέπει να συμπεριλάβουμε μία ακόμα δυνατή επιλογή για το άτομο, την δυνατότητα να επιλέξει *τυχαία* μία κίνηση από το σύνολο κινήσεων A και να εξετάσουμε αν αυτό μπορεί να επιφέρει τα επιθυμητά αποτελέσματα.

Ορισμός: Προσδιορίζουμε μία γενική συμπεριφορά β μέσα από μία λίστα πιθανοτήτων για κάθε δυνατή κίνηση. Η πιθανότητα μία κίνηση α να επιλεγεί είναι $p(\alpha)$ και ισχύει ότι: $\sum_{\alpha \in A} p(\alpha) = 1$.

Το σύνολο όλων των εφικτών τυχαίων συμπεριφορών γράφεται ως B .

Η απόδοση για την υιοθέτηση μίας συμπεριφοράς β είναι άμεσα συνδεδεμένη με τις αποδόσεις των κινήσεων α που έχουν επιλεγεί κι έτσι έχουμε ότι:

$$\pi(\beta) = \sum_{\alpha \in A} p(\alpha)\pi(\alpha)$$

Ορίζουμε επίσης και τις αποδόσεις μίας γενικής συμπεριφοράς β δεδομένων κινήσεων x : $\pi(\beta|x) = \sum_{\alpha \in A} p(\alpha)\pi(\alpha|x)$,

Και προκύπτει ότι:

$$\pi(\beta) = \sum_{\alpha \in A} p(X = x)\pi(\beta|x)$$

Ορισμός: Ορίζουμε ως βέλτιστη συμπεριφορά β^* εκείνη για την οποία ισχύει ένα από τα παρακάτω:

- i. $\pi(\beta^*) \geq \pi(\beta)$ για κάθε $\beta \in B$
- ii. $\beta^* \in \operatorname{argmax}_{\beta \in B} \pi(\beta)$

Ορισμός: Στήριξη (support) μίας συμπεριφοράς β καλείται το υποσύνολο του A , $A(\beta)$, όλων των κινήσεων για τις οποίες η β ορίζει θετική πιθανότητα. Δηλαδή $p(\alpha) > 0$.

Θεώρημα: Αν β^* μία βέλτιστη συμπεριφορά με στήριξη A^* , τότε $\pi(\alpha) = \pi(\beta^*)$ για κάθε $\alpha \in A^*$.

Απόδειξη: Προφανώς το θεώρημα ισχύει για την τετριμμένη περίπτωση που το A^* είναι μονοσύνολο.

Έστω λοιπόν ότι το A^* περιέχει περισσότερα από ένα στοιχεία. Έστω ακόμα ότι το θεώρημα δεν είναι αληθές. Τότε θα υπάρχει στοιχείο $\alpha^* \in A^*$ τέτοιο ώστε να μην δίνει την ίδια απόδοση με τη συμπεριφορά β^* αλλά μεγαλύτερη. Έτσι θα ισχύει ότι:

$$\begin{aligned}
 \pi(\beta^*) &= \sum_{\alpha \in A^*} p(\alpha)\pi(\alpha) \\
 &= \sum_{\alpha \in A^* \setminus \{\alpha^*\}} p(\alpha)\pi(\alpha) + p(\alpha^*)\pi(\alpha^*) \\
 &< \sum_{\alpha \in A^* \setminus \{\alpha^*\}} p(\alpha)\pi(\alpha^*) + p(\alpha^*)\pi(\alpha^*) \quad (\text{όπου } \pi(\alpha^*) > \pi(\alpha)) \\
 &= \left(\sum_{\alpha \in A^* \setminus \{\alpha^*\}} p(\alpha) + p(\alpha^*) \right) \pi(\alpha^*) \\
 &= \left(\sum_{\alpha \in A^*} p(\alpha) \right) \pi(\alpha^*) \\
 &= \pi(\alpha^*)
 \end{aligned}$$

Άτοπο. Διότι η συμπεριφορά β^* είναι βέλτιστη.

■

Συνεπώς, αν μία τυχαία συμπεριφορά είναι βέλτιστη τότε υπάρχουν τουλάχιστον δύο κινήσεις (που ανήκουν στη στήριξη) και επιτυγχάνουν εξίσου τη μέγιστη απόδοση. Άρα προτιμάται αυτή η στρατηγική της τυχαίας συμπεριφοράς (randomising) όταν όλες οι δυνατές επιλογές είναι ισοδύναμες.

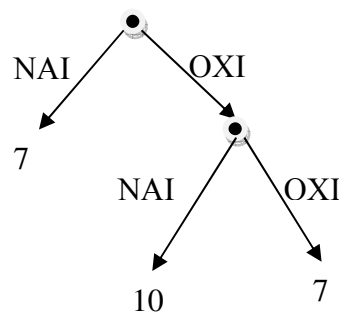
2.6 ΔΕΝΤΡΑ ΑΠΟΦΑΣΕΩΝ

Ένας υπάλληλος εργάζεται σε μία πολυεθνική και κάθε φορά που ολοκληρώνει επιτυχώς και έγκαιρα ένα project ο προϊστάμενος του προσφέρει ως bonus για επιβράβευση, την επιλογή να διαλέξει ανάμεσα σε ένα μικρό χρηματικό ποσό ή να καλύψει τα έξοδα του υπαλλήλου για ένα ταξίδι αναψυχής. Προς έκπληξη της συζύγου του ο υπάλληλος επιλέγει κάθε φορά το πρώτο. Για να της λύσει την

απορία, δικαιολογεί την απόφαση του με την εξής απάντηση: «Αν επέλεγα το ταξίδι, δεν θα είχα ξανά προσφορά bonus από τον προϊστάμενο μου!».

Με το παραπάνω παράδειγμα παρατηρούμε ότι υπάρχουν προβλήματα απόφασης τα οποία μπορεί να τερματιστούν πριν ληφθούν όλες οι πιθανές αποφάσεις και έτσι έχουμε δύο συνιστώσες που καθορίζουν την στρατηγική συμπεριφορά. Σύμφωνα με την πρώτη συνιστώσα, η προσδοκία για καλή πληρωμή στο μέλλον έχει μεγαλύτερη βαρύτητα από την άμεση αμοιβή, ενώ σύμφωνα με τη δεύτερη οφείλουμε να λάβουμε υπόψη και τη συμπεριφορά των άλλων.⁹ Στο σημείο αυτό για να αναπαραστήσουμε το πρόβλημα σχεδιάζουμε ένα δένδρο απόφασης. Σε κάθε δένδρο, οι δυνατές αποφάσεις ή αλλιώς κινήσεις αντιστοιχούν σε ένα μικρό κύκλο, ο οποίος αποτελεί ένα κόμβο και καλείται «φύλλο» του δέντρου. Αφήνοντας έναν κόμβο, είτε τερματίζουμε, είτε με μία γραμμή που καλείται «κλαδί» οδηγούμαστε στον επόμενο κόμβο. Όταν όλες οι αποφάσεις έχουν ληφθεί οδηγούμαστε μέσω του δέντρου στο τέλος ενός μονοπατιού, στο οποίο αναγράφονται οι αποδόσεις. Φυσικά η κορυφή του δέντρου είναι ψηλά και τα κλαδιά «ανοίγουν» κατεβαίνοντας προς τα κάτω.

Ας δούμε ένα παράδειγμα δέντρου.



Σχήμα 2.2 : Δέντρο απόφασης το οποίο τερματίζει στο 1^ο βήμα αν αποφασίσουμε «ναι».

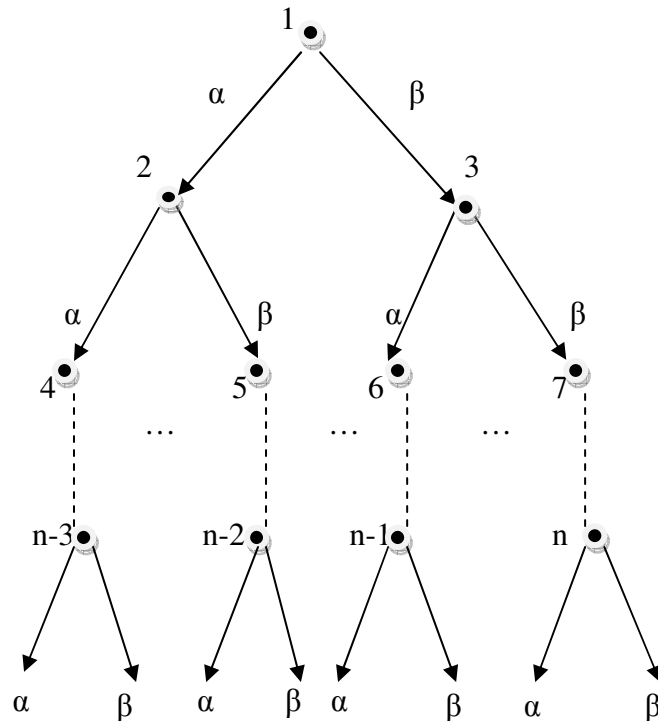
Σ' αυτό το σημείο πρέπει να επιλέξουμε την κατάλληλη στρατηγική.

Ορισμός: Μία *αμιγής στρατηγική* σημαίνει ότι μελετάμε τις επιλογές μας και δεν εμφανίζεται σε αυτήν το στοιχείο της τυχαιότητας. Το σύνολο όλων των δυνατών

⁹ Αναλυτικά προβλήματα στα οποία έχουμε «κοινή γνώση» μελετώνται σε επόμενα κεφάλαια διότι ανήκουν στο φάσμα της θεωρίας παιγνίων.

καθαρών στρατηγικών το συμβολίζουμε με S .

Έστω ότι έχουμε ένα δένδρο με n κόμβους και σε κάθε κόμβο έχουμε 2 δυνατές επιλογές (Σχήμα 3). Τότε το πλήθος των στοιχείων του S είναι 2^n .



Σχήμα 2.3: Δένδρο απόφασης n κόμβων

2.7 ΣΤΡΑΤΗΓΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ - ΑΜΙΓΕΙΣ ΚΑΙ ΜΙΚΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ

Μία *στρατηγική* είναι ένας κανόνας επιλογής της ενέργειας ή του βήματος που πρέπει να πραγματοποιηθεί σε κάθε σημείο που αυτό απαιτείται. Αμιγής στρατηγική είναι εκείνη στην οποία δεν συναντάται η έννοια της τυχειότητας. Σε αυτό το σημείο ορίζουμε το σύνολο S , το οποίο περιέχει όλες τις δυνατές αμιγείς στρατηγικές.

Μία αμιγής στρατηγική επιλέγει ένα μονοπάτι από το σημείο εκκίνησης μέχρι το σημείο περάτωσης, μέσα στο δέντρο αποφάσεων. Όμως, μία αμιγής στρατηγική δεν αποτελεί μόνο ένα μονοπάτι μέσα στο δέντρο αποφάσεων αλλά καθορίζει και τα

βήματα που πρέπει να γίνουν σε κάθε κόμβο “απόφασης” περιλαμβάνοντας κι εκείνα που δεν είναι απαραίτητα για να επιτευχθεί το αποτέλεσμα.

Παράδειγμα: Θεωρούμε το παιχνίδι κορώνα (K) ή γράμματα (Γ) (με αμερόληπτο κέρμα) το οποίο πραγματοποιείται σε δύο ρίψεις του κέρματος. Νικητής του παιχνιδιού είναι εκείνος που θα φέρει για πρώτη φορά κορώνα. Το σύνολο όλων των αμιγών στρατηγικών είναι $S = \{ \Gamma\Gamma, \Gamma K, K\Gamma, KK \}$. Παρατηρούμε εδώ πως οι δύο τελευταίες στρατηγικές, KΓ και KK επιφέρουν το ίδιο αποτέλεσμα (νίκη) στην πρώτη ρίψη, οπότε δεν χρειάζεται επιπλέον απόφαση. Παρόλα αυτά περιλαμβάνεται στο σύνολο S .

Το παραπάνω παράδειγμα μας δίνει την αφορμή να ορίσουμε ένα νέο σύνολο στρατηγικών, S_R , το οποίο είναι το αρχικό σύνολο στρατηγικών διαμορφωμένο έτσι ώστε να περιλαμβάνει μόνο τα απαιτούμενα βήματα σε κάθε στρατηγική.

Έτσι στο παράδειγμα 1 έχουμε και το $S_R = \{ \Gamma\Gamma, \Gamma K, KX \}$ όπου KX σημαίνει ότι φέραμε (K) κορώνα στην πρώτη ρίψη και δεν χρειάζεται άλλη απόφαση.

Στη συνέχεια εισάγουμε την τυχαιότητα στη στρατηγική.

Ορισμός: Η μικτή στρατηγική σ καθορίζει την πιθανότητα $p(s)$ με την οποία κάθε αμιγή στρατηγική $s \in S$ χρησιμοποιείται. Θεωρούμε το σύνολο των στρατηγικών $S = \{s_\alpha, s_\beta, s_\gamma, \dots\}$. Τότε μία μικτή στρατηγική σ δίνεται ως μία κατανομή πιθανοτήτων ως εξής:

$$\sigma = (p(s_\alpha), p(s_\beta), p(s_\gamma), \dots)$$

Συνεπάγεται πως οι μικτές στρατηγικές μπορούν να γραφούν ως γραμμικός συνδυασμός αμιγών στρατηγικών:

$$\sigma = \sum_{s \in S} p(s)s$$

Στήριγμα μικτής στρατηγικής σ καλείται το σύνολο $S(\sigma) \subseteq S$ όλων των αμιγών στρατηγικών για τις οποίες η σ ορίζει θετική πιθανότητα, δηλαδή $p(s) > 0$.

Ορισμός: Σε κάθε κόμβο i με $i \in \{1, 2, 3, \dots, n\}$ θεωρούμε το σύνολο κινήσεων

$A_i = \{a_1^i, a_2^i, a_3^i, \dots, a_{k_i}^i\}$ όπου το πλήθος των δυνατών κινήσεων k_i πρέπει να είναι

διαφορετικό σε κάθε κόμβο απόφασης i . Η συμπεριφορά ενός ατόμου στον κόμβο i καθορίζεται από το διάνυσμα πιθανοτήτων:

$$p_i = [p(a_1^i), p(a_2^i), p(a_3^i), \dots, p(a_{k_i}^i)].$$

Κάθε $p(a_j^i)$ είναι η πιθανότητα με την οποία επιλέγεται η κίνηση $a_j^i \in A_i$.

Μία συμπεριφοριστική στρατηγική β είναι το σύνολο πιθανοτήτων

$$\beta = \{p_1, p_2, p_3, \dots, p_n\}$$

Σε αντίθεση με τη μικτή στρατηγική, στη συμπεριφοριστική συναντάται η τυχαιότητα *αρκετές φορές* καθώς διασχίζεται ένα δένδρο απόφασης.

Ορισμός: Μία συμπεριφοριστική στρατηγική και μία μικτή στρατηγική είναι ισοδύναμες αν δίνουν τις ίδιες πιθανότητες σε κάθε δυνατή αμιγή στρατηγική που είναι διαθέσιμη.

Θεώρημα:

- i) Κάθε συμπεριφοριστική στρατηγική διαθέτει μία μικτή αναπαράσταση και
- ii) κάθε μικτή στρατηγική έχει μία συμπεριφοριστική αναπαράσταση.

Απόδειξη:

i) Έστω $s \in S$ μία αμιγής στρατηγική με ένα σύνολο από κόμβους απόφασης $I(s) \subseteq I = \{1, 2, 3, \dots, n\}$ και $a^i(s) \in A_i$ η επιλογή μίας κίνησης για κάθε κόμβο i στο $I(s)$. Για κάθε κόμβο απόφασης i μία δοσμένη συμπεριφοριστική στρατηγική β θα καθορίσει την επιλογή της κίνησης με πιθανότητα $p(a^i(s))$. Τότε η πιθανότητα ένα άτομο να διασχίσει το δέντρο αποφάσεων χρησιμοποιώντας την συμπεριφοριστική στρατηγική β μέσω των κόμβων $I(s)$ είναι:

$$p(s) = \prod_{i \in I(s)} p(a^i(s))$$

Συνεπώς μία μικτή στρατηγική αναπαράσταση του β είναι:

$$\sigma_\beta = \sum_{s \in S} p(s) \cdot s$$

διότι:

$$\sum_{s \in S} \prod_{i \in I(s)} p(a^i(s)) = 1$$

ii) Έστω $\sigma = \sum_{s \in S} p(s)s$ μία μικτή στρατηγική. Για κάθε αμιγή στρατηγική s με $I(s)$

$\subseteq I$ το σύνολο των κόμβων απόφασης όταν το υποκείμενο ακολουθεί τη στρατηγική s . Για κάθε κόμβο απόφασης $i \in I$, θεωρούμε $S(i) \subseteq S$ το σύνολο των αμιγών στρατηγικών που οδηγούν σε αυτό τον κόμβο i . Τότε η πιθανότητα της μικτής στρατηγικής σ να φτάσει στον κόμβο i είναι:

$$p_\sigma(i) = \sum_{s \in S(i)} p(s).$$

Αν τώρα θεωρήσουμε $S(a^i, i) \subseteq S(i)$ να είναι το σύνολο των αμιγών στρατηγικών που οδηγούν στον κόμβο i και σε αυτό το σημείο επιλέξουμε τη κίνηση $a^i \in A_i$, τότε η πιθανότητα το άτομο ακολουθώντας τη μικτή στρατηγική σ να φτάσει στον κόμβο i και να επιλέξει την απόφαση a^i είναι:

$$p_\sigma(a^i, i) = \sum_{s \in S(a^i, i)} p(s).$$

Με τον περιορισμό ότι $p_\sigma(i) \neq 0$ μπορούμε να ορίσουμε την πιθανότητα να επιλεγεί η a^i :

$$p(a^i) = \frac{p_\sigma(a^i, i)}{p_\sigma(i)}.$$

Ακόμα κι αν $p_\sigma(i) = 0$ για κάποιο κόμβο i τότε κάθε σύνολο πιθανοτήτων $p(a^i)$ με $\sum_{a^i \in A_i} p(a^i) = 1$ θα ικανοποιείται.

Έτσι το σύνολο αυτών των πιθανοτήτων για όλες τις δράσεις για κάθε κόμβο καθιστά μία συμπεριφοριστική αναπαράσταση της μικτής στρατηγικής σ .

■

2.8 ΒΕΛΤΙΣΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ

Όπως είδαμε σε προηγούμενη παράγραφο, σε προβλήματα μίας και μόνο απόφασης η έννοια της τυχαίας συμπεριφοράς δεν έχει νόημα δεδομένου ότι η βέλτιστη απόφαση σε ένα βήμα μπορεί πάντα να βρεθεί. Ένα παρόμοιο αποτέλεσμα ισχύει και για ανεπίλυτες αποφάσεων.

Λήμμα: Έστω σ^* μία βέλτιστη μικτή στρατηγική με στήριγμα S^* . Τότε $\pi(s) = \pi(\sigma^*)$ για κάθε $s \in S^*$.

Απόδειξη: Για την τετριμμένη περίπτωση που το S^* περιέχει ένα μόνο στοιχείο ισχύει προφανώς το Λήμμα.

Έστω ότι το S^* περιέχει τουλάχιστον δύο στοιχεία. Αν δεν ισχύει η υπόθεση μας, τότε μία τουλάχιστον στρατηγική θα δίνει υψηλότερη απόδοση από $\pi(\sigma^*)$. Θεωρούμε s' τη στρατηγική με την μεγαλύτερη απόδοση από τις αναφερθείσες.

Έχουμε:

$$\begin{aligned}\pi(\sigma^*) &= \sum_{s \in S^*} p^*(s)\pi(s) \\ &= \sum_{s \neq s'} p^*(s)\pi(s) + p^*(s')\pi(s') \\ &< \sum_{s \neq s'} p^*(s)\pi(s') + p^*(s')\pi(s') \\ &= \pi(s') \cdot \left(\sum_{s \in S^*} p^*(s) \right) \\ &= \pi(s') \\ &\Rightarrow \pi(\sigma^*) < \pi(s')\end{aligned}$$

Δεδομένου, όμως, ότι σ^* είναι βέλτιστη, δηλαδή $\pi(\sigma^*) \geq \pi(s)$ για κάθε $s \in S^*$ οδηγούμαστε σε άτοπο.

■

Θεώρημα: Για κάθε διαδικασία λήψης απόφασης υπάρχει πάντα μία βέλτιστη αμιγή στρατηγική.

Απόδειξη: Σε προηγούμενο θεώρημα είδαμε ότι κάθε συμπεριφοριστική στρατηγική έχει τουλάχιστον μία ισοδύναμη μικτή στρατηγική. Συνεπώς καμία συμπεριφοριστική στρατηγική δεν μπορεί να έχει μεγαλύτερη απόδοση από αυτή που επιτυγχάνεται μέσω μίας μικτής στρατηγικής. Έπεται λοιπόν πως αν υπάρχει μία βέλτιστη μικτή στρατηγική, θα υπάρχει και μία βέλτιστη αμιγής.

■

Ορισμός: Μερικό ιστορικό h καλείται μία σειρά αποφάσεων που έχουν παρθεί από ένα άτομο σε δεδομένο χρόνο. Στο αρχικό σημείο, όπου καμία απόφαση δεν έχει παρθεί, έχουμε $h = \emptyset$, ενώ ένα πλήρες ιστορικό μία στρατηγικής s είναι μία πλήρης ακολουθία αποφάσεων ενός ατόμου που ακολουθεί την στρατηγική s και σημειώνεται ως $H(s)$.

Αξίζει εδώ να σημειωθεί πως αν ένα άτομο έχει τέλεια μνήμη τότε κάθε κόμβος

απόφασης έχει μοναδικό ιστορικό και κάθε ιστορικό προσδιορίζει ένα μοναδικό κόμβο.

Ορίζουμε στη συνέχεια ένα υποσύνολο αμιγών στρατηγικών $S(h) \in S$, το οποίο περιέχει όλες τις στρατηγικές με ιστορικό h , το οποίο διαφέρει στις κινήσεις που θα λάβουν χώρα στο μέλλον. Τότε η βέλτιστη απόδοση ενός ατόμου που μπορεί να επιτευχθεί, δεδομένου του ιστορικού h είναι:

$$\pi^*(s|h) = \max_{s \in S(h)} \pi(s)$$

Θεωρούμε τώρα $A(h)$ το σύνολο των δυνατών κινήσεων ενός ατόμου. Από τη στιγμή που μία απόφαση θα ληφθεί, άμεσα προσαρτάται στο ιστορικό h .

Θεώρημα: (Αρχή βέλτιστου)

Για ένα άτομο με τέλεια μνήμη ισχύουν τα παρακάτω:

- i. $\pi^*(s|H(s)) = \pi(s)$
- ii. $\pi^*(s|h) = \max_{a \in A(h)} \pi^*(s|h, a)$
- iii. $\pi^* = \max_{s \in S(\emptyset)} \pi^*(s|\emptyset)$

Απόδειξη:

i. Προκύπτει άμεσα διότι από τον ορισμό του $H(s)$ το άτομο έχει ολοκληρώσει τη διαδικασία λήψης αποφάσεων μέσω της στρατηγικής s , η οποία αντιστοιχεί σε απόδοση $\pi(s)$.

ii. Μία γνήσια στρατηγική είναι μία ακολουθία δράσεων $\{\alpha_0, \alpha_1, \dots, \alpha_h, a_{h+1}, a_{h+2}, \dots, \alpha_H\}$ με αποδόσεις:

$$\pi(s) = \pi(\alpha_0, \alpha_1, \dots, \alpha_h, a_{h+1}, a_{h+2}, \dots, \alpha_H)$$

Αν h το μερικό ιστορικό, $h = \{\alpha_0, \alpha_1, \dots, \alpha_h\}$, τότε

$$\begin{aligned} \pi^*(s|h) &= \max_{a_{h+1}} \max_{a_{h+2}} \dots \max_{a_H} \pi(\alpha_0, \alpha_1, \dots, \alpha_h, a_{h+1}, a_{h+2}, \dots, \alpha_H) \\ &= \max_{a_{h+1}} \pi^*(s|h, a_{h+1}) \end{aligned}$$

iii) Για $h = \emptyset$ έχουμε πρόβλημα βελτιστοποίησης από την αρχή, οπότε για $S(\emptyset) = S$ έχουμε:

$$\max_{s \in S(\emptyset)} \pi^*(s|\emptyset) = \max_{s \in S} \pi(s) = \pi^*$$

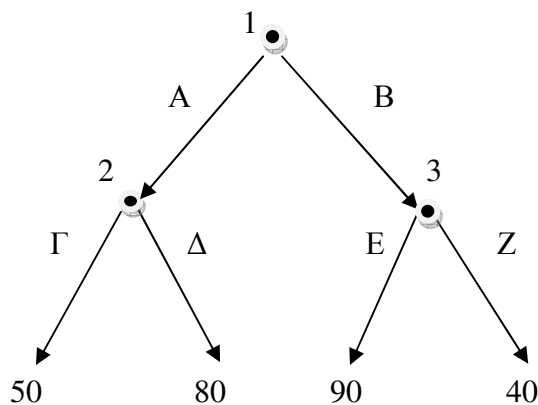
■

Η αρχή του βέλτιστου οδηγεί απευθείας σε μία βολική μέθοδο επίλυσης δυναμικών

προβλημάτων απόφασης. Αν λοιπόν ελπίζουμε σε μία βέλτιστη απόφαση τώρα, θεωρώντας ότι θα έχουμε τη βέλτιστη συμπεριφορά στο μέλλον, αρκεί να κάνουμε ξεκάθαρες τις μελλοντικές μας επιλογές δουλεύοντας οπισθοδρομικά (αντίστροφη επαγωγή).

Παράδειγμα:

Δίνεται στο σχήμα 4 ένα δένδρο απόφασης με 3 κόμβους και αναζητούμε τη βέλτιστη στρατηγική. Θα εργασθούμε με την οπισθοδρομική επαγωγή.



Σχήμα 2.4

Επιλέγουμε έναν από τους τερματικούς κόμβους απόφασης 2 ή 3 (ισοδύναμοι κόμβοι).

Έστω λοιπόν ότι βρισκόμαστε στον κόμβο 3. Τότε ανάμεσα στις επιλογές E και Z βέλτιστη είναι η E γιατί μας δίνει την υψηλότερη απόδοση. Γυρνώντας προς τα πίσω βλέπουμε πως οδηγηθήκαμε στον κόμβο 3 μέσω της δράσης B. Άρα μία βέλτιστη στρατηγική είναι η $s_1 = \{B, E\}$ με απόδοση $\pi(s_1) = 90$

Στην συνέχεια, μελετάμε τις επιλογές απόφασης στον κόμβο 2. Τότε παρατηρούμε πως καλύτερη απόφαση είναι η Δ γιατί δίνει μεγαλύτερη απόδοση. Επιστρέφοντας, βλέπουμε πως οδηγηθήκαμε στον κόμβο 2 μέσω της δράσης A. Άρα μία ακόμα βέλτιστη στρατηγική είναι η $s_2 = \{A, \Delta\}$ με $\pi(s_2) = 80$.

Τελικά, συγκρίνοντας τις αποδόσεις των δύο στρατηγικών προκύπτει ως βέλτιστη η στρατηγική s_1 .

ΚΕΦΑΛΑΙΟ 3^ο

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ

3.1 ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΗΣ ΕΞΑΡΤΩΜΕΝΕΣ ΑΠΟ ΤΗΝ ΚΑΤΑΣΤΑΣΗ

Οι ανελίξεις που θα μελετηθούν σε αυτό το κεφάλαιο είναι είτε ντετερμινιστικές είτε στοχαστικές. Αρχικά θεωρούμε πως η διαδικασία τερματίζεται σε δοσμένο χρόνο T (μοντέλα μας έχει πεπερασμένου ορίζοντα), ενώ οι αποφάσεις μπορούν να ληφθούν τις χρονικές στιγμές $t = 0, 1, 2, \dots, T - 1$, παρόλο που το πλήθος των αποφάσεων τελικά, μπορεί να μην υπερβαίνει την τιμή T αν η διαδικασία τερματιστεί νωρίτερα λόγω των επιλεγμένων κινήσεων. Μοντέλα τα οποία δεν έχουν εκ των προτέρων περιορισμό στο πλήθος των αποφάσεων που μπορούν να ληφθούν, θα μελετηθούν αργότερα (μοντέλα άπειρου ορίζοντα).

Θεωρούμε ότι τα άτομα προσδιορίζονται από μια μεταβλητή κατάστασης x που ανήκει στο χώρο καταστάσεων X , ο οποίος μπορεί να είναι είτε διακριτός είτε συνεχής. Η κατάσταση κάθε φορά μπορεί να αντιπροσωπεύει κάτι διαφορετικό· για παράδειγμα το πλούτο ενός ατόμου, το πλήθος των απογόνων, την ανάγκη για φαγητό. Τα άτομα καθορίζουν τη συμπεριφορά στην κατάσταση στην οποία βρίσκονται κάθε χρονική στιγμή και εμείς προσδιορίζουμε ένα σύνολο κινήσεων για την κατάσταση x τη χρονική στιγμή t . Μία κίνηση a_t προκαλεί μετάβαση σε μία νέα κατάσταση και συμβολίζουμε: $x_t \rightarrow x_{t+1}$. Με αυτό τον τρόπο ορίζεται η διαδικασία (x_t, a_t) για $t = 0, 1, 2, \dots$.

Ορισμός: Ιστορικό μίας ανελίξης καλείται μία ακολουθία καταστάσεων και κινήσεων $x_0, a_0, x_1, a_1, x_2, a_2, \dots$. Συμβολίζουμε ως $h_t = (x_0, a_0, x_1, a_1, \dots, x_t)$ το ιστορικό μίας ανελίξης μέχρι τη χρονική στιγμή t .

Πώς όμως λαμβάνονται οι αποφάσεις; Στην πιο απλή μορφή, μία αμιγής στρατηγική s θα μπορούσε να προκύπτει από το σύνολο των κινήσεων $a_0, a_1, a_2, \dots$. Γενικότερα όμως, οι στρατηγικές προσδιορίζουν μία επιλογή κινήσεων $a(x, t)$ για κάθε κατάσταση και κάθε στιγμή απόφασης. Ξεκινώντας από την κατάσταση x_0 ,

αυτή η στρατηγική παράγει ένα ιστορικό $x_0, a_0, x_1, a_1, x_2, a_2, \dots$ και $a_t = a(x_t, t)$ με αντίστοιχες αμοιβές $r_t(x_t, a_t)$ για $t = 1, 2, \dots, T-1$. Δηλαδή, ένα άτομο στην κατάσταση x_t τη χρονική στιγμή t , που χρησιμοποιεί την κίνηση a_t θα λάβει ως άμεση αμοιβή $r_t(x_t, a_t)$. Για πεπερασμένες ανελίξεις, περιλαμβάνεται και μία προαιρετική αμοιβή $r_T(x_T)$ η οποία λαμβάνεται στο τέλος της διαδικασίας. Η συνολική αμοιβή που αποκτήθηκε από ένα άτομο που ξεκίνησε από την κατάσταση x και ακολούθησε τη στρατηγική s , είναι:

$$\pi(x|s) = \sum_{t=0}^{T-1} r_t(x_t, a_t) + r_T(x_T).$$

Όλα όσα προαναφέραμε αφορούν διαδικασίες στις οποίες οι καταστάσεις ήταν συγκεκριμένες. Τι συμβαίνει όμως όταν έχουμε στοχαστικές διαδικασίες απόφασης; Έστω λοιπόν η τυχαία μεταβλητή X_t που δηλώνει την κατάσταση σε χρόνο t . Οι πιθανότητες μετάβασης $x_t \rightarrow x_{t+1}$ μπορούν γενικά να εξαρτώνται από το ιστορικό της ανελίξης αλλά και από την επιλεγμένη κίνηση a_t :

$$P(X_{t+1} = x_{t+1}) = p(x_{t+1}|h_t, a_t)$$

με

$$\sum_{x_{t+1} \in X} p(x_{t+1}|h_t, a_t) = 1.$$

Για όλες τις χρονικές στιγμές, τις κινήσεις και όλο το ιστορικό. Αν τώρα έχουμε τυχαίες στρατηγικές σ τέτοιες ώστε η κίνηση που επιλέχθηκε να είναι κι αυτή μία τυχαία μεταβλητή, έστω A_t , η συνολική αμοιβή που αποκτήθηκε από ένα άτομο, ξεκινώντας από μία κατάσταση x και επιλέγοντας τη στρατηγική σ είναι:

$$\begin{aligned} \pi(x|\sigma) &= \mathbb{E} \left[\sum_{t=0}^{T-1} r_t(X_t, A_t) + r_T(X_T) \right] \\ &= \sum_{t=0}^{T-1} \mathbb{E}[r_t(X_t, A_t)] + \mathbb{E}[r_T(X_T)]. \end{aligned}$$

Στη συνέχεια, για την απόδοση που αντιστοιχεί στη βέλτιστη στρατηγική ξεκινώντας από την κατάσταση x θα χρησιμοποιούμε το συμβολισμό $\pi^*(x)$, όπου:

$$\pi^*(x) = \pi^*(x|\sigma^*).$$

Προφανώς στόχος κάθε ατόμου που καλείται να λάβει αποφάσεις είναι να μεγιστοποιήσει αυτή την αναμενόμενη συνολική αμοιβή¹⁰.

¹⁰Ή αντίστοιχα να ελαχιστοποιήσει τη ζημία, ανάλογα με το πρόβλημα.

3.2 ΜΑΡΚΟΒΙΑΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΕΩΝ

Οι Μαρκοβιανές διαδικασίες αποφάσεων είναι μια ειδική κατηγορία, διότι οι πιθανότητες μετάβασης από μία κατάσταση σε μία άλλη βασίζονται μόνο στην τρέχουσα κατάσταση και όχι στο ιστορικό. Δεν μας ενδιαφέρει λοιπόν πως φτάσαμε στην τρέχουσα κατάσταση.

Ορισμός: Μια διαδικασία αποφάσεων έχει την Μαρκοβιανή¹¹ ιδιότητα αν:

$$p(x_{t+1}|h_t, a_t) = p(x_{t+1}|x_t, a_t)$$

Ορισμός: Μία διαδικασία αποφάσεων που έχει την Μαρκοβιανή ιδιότητα λέγεται Μαρκοβιανή Διαδικασία Αποφάσεων (ΜΔΑ).

Στη συνέχεια θα ασχοληθούμε αποκλειστικά με ΜΔΑ. Ο υπολογισμός των αποδόσεων για γενικευμένες στρατηγικές είναι αρκετά πολύπλοκος με συνέπεια να γίνεται ακόμη πιο δύσκολο το έργο της εύρεσης της βέλτιστης στρατηγικής. Όμως, για πεπερασμένες ΜΔΑ η μέθοδος της οπισθοδρομικής επαγωγής είναι σαφώς χρήσιμη.

Ακολουθεί ένα παράδειγμα Μαρκοβιανής Διαδικασίας Αποφάσεων.

Παράδειγμα: Ένας επενδυτής πρέπει να λάβει ορισμένες αποφάσεις για τη διαχείριση του κεφαλαίου του. Πρέπει να βρει τη βέλτιστη λύση που αφορά πόσα από τα χρήματα του τρέχοντος κεφαλαίου θα καταναλωθούν (π.χ. για αγορά αγαθών) και πόσα από αυτά θα επενδύσει. Τη χρονική στιγμή t το κεφάλαιο είναι x_t με κόστος (δηλαδή το ποσό του κεφαλαίου προς κατανάλωση) c_t , ενώ το υπόλοιπο ποσό επενδύεται με επιτόκιο $r - 1$ για να προκύψει ένα ποσό κεφαλαίου $r(x_t - c_t)$ τη στιγμή της επόμενης απόφασης. Για λόγους ευκολίας, περιορίζουμε το πρόβλημα σε δύο περιόδους $t = 0$ και $t = 1$, όπου σε κάθε μία από αυτές πρέπει ο επενδυτής να αποφασίσει για το ποσό που θα καταναλώσει. Υποθέτουμε ότι ο επενδυτής έχει άμεσο όφελος από την απόφαση και ότι η συνάρτηση ωφέλειας είναι λογαριθμική, έστω $\pi(c_0, c_1) = \ln(c_0) + \ln(c_1)$.

Α' τρόπος: Μέθοδος Lagrange¹². Η αρχική συνθήκη μας είναι: $x_1 = r(x_0 - c_0)$ με

¹¹ Το όνομα οφείλεται στο Ρώσο μαθηματικό Andrei Andreyevich Markov (1856 -1922)

περιορισμό $c_1 \leq x_1 \Rightarrow c_1 \leq r(x_0 - c_0)$.

Έχουμε λοιπόν ότι:

$$L(c_0, c_1) = \ln(c_0) + \ln(c_1) - \lambda(c_1 + rc_0 - rx_0)$$

Ενώ ταυτόχρονα πρέπει να ικανοποιούνται οι ακόλουθες εξισώσεις:

$$\begin{cases} \frac{\partial L(c_0, c_1)}{\partial c_0} = 0 \\ \frac{\partial L(c_0, c_1)}{\partial c_1} = 0 \\ c_1 + rc_0 - rx_0 = 0 \end{cases} \Rightarrow \begin{cases} \frac{1}{c_0^*} - \lambda r = 0 \\ \frac{1}{c_1^*} - \lambda = 0 \\ c_1^* + rc_0^* - rx_0 = 0 \end{cases}$$

Το πρώτο ζεύγος εξισώσεων μας οδηγεί στην παρακάτω σχέση για τα κόστη στις δύο περιόδους:

$$c_1^* = rc_0^*.$$

Αντικαθιστώντας στην τρίτη ισότητα προκύπτει η βέλτιστη στρατηγική:

$$c_0^* - \frac{1}{2}x_0, \quad x_1^* = \frac{r}{2}x_0$$

Β' τρόπος: (Δυναμικός προγραμματισμός) Με τη βοήθεια της οπισθοδρομικής επαγωγής, έχουμε:

Για $t = 1$, η απόδοση είναι $\ln(c_1)$ λόγω του περιορισμού $c_1 \leq x_1$ (όπου x_1 το κεφάλαιο που διαθέτει ο επενδυτής εκείνη τη στιγμή. Έτσι λοιπόν η βέλτιστη απόφαση είναι $c_1^*(x_1) = x_1$. Προφανώς δεν γνωρίζουμε την τιμή του x_1 διότι εξαρτάται από τη συμπεριφορά στην αρχική φάση (δηλαδή για $t = 0$) όπου $x_1 = r(x_0 - c_0)$. Έτσι αυτό που έχουμε μέχρι στιγμής είναι μία γενική λύση που ικανοποιείται για κάθε τιμή του x_1 .

Για $t = 0$, το πρόβλημα του επενδυτή έγκειται στη μεγιστοποίηση της συνολικής απόδοσης, με υπόθεση ότι έχουμε τη βέλτιστη συμπεριφορά για $t = 1$. Έτσι, έχουμε:

$$c_0 \in \operatorname{argmax}_{c_0 \in [0, x_0]} (\ln(c_0) + \ln(x_1(c_0))).$$

Παραγωγίζοντας την απόδοση και εξισώνοντας με το μηδέν προκύπτει:

¹²Παλός μαθηματικός, φυσικός και αστρονόμος Joseph –Louis Lagrange (1736- 1813), Τσίτσας Λ., Εφαρμοσμένος Διανυσματικός Απειροστικός Λογισμός (2003), Εκδόσεις Συμμετρία

$$x_1 - rc_0^* = 0 \Rightarrow c_0^* = \frac{x_1}{r}.$$

Αντικαθιστώντας με τη σχέση $x_1 = r(x_0 - c_0)$ με x_0 γνωστό, έχουμε:

$$\begin{aligned} c_0^* = \frac{x_1}{r} &\Rightarrow c_0^* = \frac{r(x_0 - c_0^*)}{r} = x_0 - c_0^* \\ &\Rightarrow 2c_0^* = x_0 \Rightarrow c_0^* = \frac{1}{2}x_0 \end{aligned}$$

Συνεπώς η βέλτιστη στρατηγική $s^*(x, t) = (c_0^*, c_1^*)$ είναι:

$$\begin{aligned} (c_0^*, c_1^*) &= \left(\frac{1}{2}x_0, x_1\right) \Rightarrow \\ (c_0^*, c_1^*) &= \left(\frac{1}{2}x_0, rc_0^*\right) \Rightarrow \\ (c_0^*, c_1^*) &= \left(\frac{1}{2}x_0, \frac{r}{2}x_0\right). \end{aligned}$$

Στη συνέχεια, ως πρόλογο για την εισαγωγή στο στοχαστικό δυναμικό προγραμματισμό θα αναπτύξουμε μία γενική μορφή για το δυναμικό προγραμματισμό στη ντετερμινιστική του μορφή.

Η ιδέα του δυναμικού προγραμματισμού είναι να απαντά στο ερώτημα «ποια είναι η καλύτερη κίνηση τώρα, δεδομένου ότι θα έχουμε βέλτιστη συμπεριφορά σε κάθε πιθανό σημείο απόφασης στο μέλλον;» Η λέξη «πιθανό» αναφέρθηκε για να δηλώσει ότι πρέπει να γνωρίζουμε τι θα γίνει σε όλα τα μελλοντικά στάδια συμπεριλαμβανομένων και εκείνων που μπορεί να μην φτάσουμε αν βρεθεί και παρθεί η βέλτιστη απόφαση. Χωρίς αυτή την πληροφορία ως δεδομένο, δεν μπορούμε να αποφασίσουμε για το βέλτιστο. Συνεπώς, αν $\pi_t^*(x)$ η απόδοση στο μέλλον, με αφετηρία την κατάσταση x τη χρονική στιγμή t , τότε θα έχουμε βέλτιστη συμπεριφορά τις χρονικές στιγμές $t, t+1, t+2, \dots, T-1$. Έτσι

$$\pi_t^*(x) = \sum_{n=t}^{T-1} r_t(x_n^*, a_n^*) + r_T(x_T^*)$$

Όπου $x_t = x$ και η ακολουθία καταστάσεων $(x_n^*)_{n=t+1, \dots, T-1}$ παράγεται από την ακολουθία των βέλτιστων κινήσεων $(a_n^*)_{n=t+1, \dots, T-1}$. Οπότε η γενική μορφή της εξίσωσης για το δυναμικό προγραμματισμό στην ντετερμινιστική περίπτωση θα είναι:

$$\pi_t^*(x) = \max_{a \in A(x)} [r_t(x, a) + \pi_{t+1}^*(x_{t+1}(a))]$$

Όπου $x_{t+1}(a)$ είναι η επόμενη κατάσταση μετά την x αν επιλέξουμε και

χρησιμοποιήσουμε την κίνηση a από το διαθέσιμο σύνολο κινήσεων της κατάστασης x , το $A(x)$. Η διαδικασία της οπισθοδρομικής επαγωγής αρχίζει αν θέσουμε:

$$\pi_T^*(x) = r_T(x_T) \quad \text{για κάθε } x \in X$$

Και η απόδοση που επιτυγχάνεται από τη βέλτιστη στρατηγική s^* με σημείο εκκίνησης την κατάσταση x δίνεται από $\pi(x|s^*) = \pi_0^*(x)$.

3.3 ΣΤΟΧΑΣΤΙΚΕΣ ΜΑΡΚΟΒΙΑΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΕΩΝ

Στις ντετερμινιστικές Μαρκοβιανές διαδικασίες αποφάσεων (ΜΔΑ), η επιλογή κίνησης σε μία συγκεκριμένη κατάσταση προσδιορίζει μοναδικά την κατάσταση στο επόμενο στάδιο της διαδικασίας. Στο υπόλοιπο μέρος του κεφαλαίου στις διαδικασίες απόφασης που θα μας απασχολήσουν, οι μεταβάσεις από τη μία κατάσταση στην άλλη δεν είναι βέβαιες. Θα υποθέσουμε ότι οι πιθανότητες μετάβασης από μία κατάσταση σε μία άλλη δεν εξαρτώνται από το χρόνο, δεδομένου ότι ένα άτομο στη κατάσταση x τη χρονική στιγμή t , επιλέγει μία κίνηση a η πιθανότητα να βρεθεί στην κατάσταση x' τη χρονική στιγμή $t + 1$ είναι $p(x'|x, a) \leq 1$ για κάθε $x' \in X$. Παρόλο που έχουμε προβεί σε μία υπόθεση περισσότερο στάσιμη από άποψη χρόνου, η βέλτιστη στρατηγική μπορεί να είναι χρονοεξαρτώμενη, όπως είδαμε στην προηγούμενη παράγραφο. Άλλωστε και οι ντετερμινιστικές ΜΔΑ, είναι στοχαστικές ΜΔΑ στις οποίες οι πιθανότητες μετάβασης είναι 0 ή 1. Έτσι γίνεται φανερό πως η εξάρτηση από το χρόνο μπορεί να προκύψει εξαιτίας του γεγονότος ότι τα προβλήματα που μελετάμε (μέχρι στιγμής) είναι πεπερασμένα, διότι σε κάθε κατάσταση έχουμε μηδενική άμεση αμοιβή.

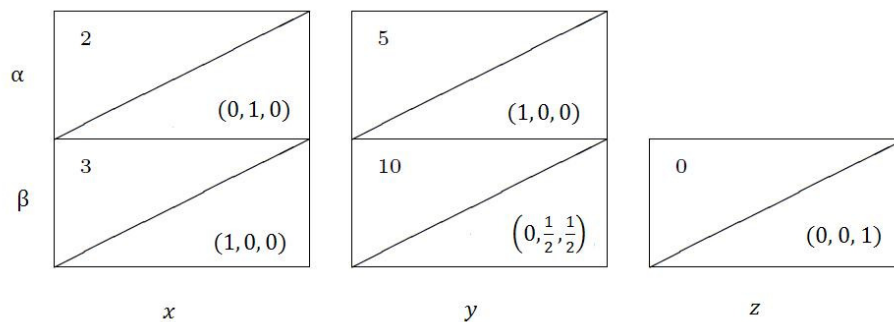
Υπό ορισμένες συνθήκες μία στοχαστική Μαρκοβιανή διαδικασία αποφάσεων (ΜΔΑ) διαθέτει μία απλή διαγραμματική αναπαράσταση. Οι συνθήκες είναι οι ακόλουθες:

- i. Το πλήθος των καταστάσεων και των διαθέσιμων κινήσεων σε κάθε μία από αυτές είναι μικρό
- ii. Οι διαθέσιμες κινήσεις σε κάθε κατάσταση είναι ανεξάρτητες από το χρόνο (τουλάχιστον για $t < T$).
- iii. Οι αμοιβές που αντιστοιχούν σε κάθε ζεύγος κατάστασης-κίνησης είναι

ανεξάρτητες από το χρόνο ($r_t(x, a) = r(x, a)$ για $t < T$)

Παράδειγμα: Έστω το σύνολο καταστάσεων $X = \{x, y, z\}$. Στις καταστάσεις x και y μπορούμε να επιλέξουμε μία κίνηση από το σύνολο $A(x) = A(y) = \{\alpha, \beta\}$, ενώ στην κατάσταση z έχουμε ως μοναδική επιλογή την κίνηση $A(z) = \{\beta\}$. Αν επιλέξουμε την κίνηση α στην κατάσταση x τότε προκύπτει αμοιβή $r(x, \alpha) = 2$ και μετακινούμαστε στην κατάσταση y με πιθανότητα 1. Αν επιλέξουμε την κίνηση β , έχουμε αμοιβή $r(x, \beta) = 3$ και παραμένουμε στην κατάσταση αυτή με πιθανότητα 1. Στην κατάσταση y έχουμε αντίστοιχα για την επιλογή της κίνησης α αμοιβή $r(y, \alpha) = 5$ και μετάβαση στην κατάσταση x με πιθανότητα 1, ενώ για την επιλογή της κίνησης β ¹³ η αμοιβή είναι $r(y, \beta) = 10$ και με πιθανότητα $1/2$ γίνεται η μετάβαση στην κατάσταση z , αλλιώς με πιθανότητα επίσης $1/2$ παραμένουμε στην κατάσταση y . Αν όμως είμαστε στην κατάσταση z , για την κίνηση β έχουμε αμοιβή $r(z, \beta) = 0$ και με πιθανότητα 1 παραμένουμε σε αυτή την κατάσταση.

Αυτή η αναλυτική περιγραφή μπορεί να παρασταθεί διαγραμματικά όπως φαίνεται στο ακόλουθο σχήμα. Στην περίπτωση όμως που είχαμε ένα πρόβλημα πεπερασμένου ορίζοντα αυτή η διαγραμματική απεικόνιση του προβλήματος θα απαιτούσε να προσδιοριστεί ο ορίζοντας T και οι τελικές αποδόσεις $r_T(x)$ για κάθε $x \in X$.



Σχήμα 3.1

¹³ Η επιλογή μίας κίνησης χωρίς εναλλακτική είναι επίσης γνωστή ως «επιλογή Hobson» από τον Thomas Hobson (Cambridge, 1544 -1631), ο οποίος ήταν ιδιοκτήτης υποστασίου και είχε ως όρο σε υποψήφιους αγοραστές των αλόγων του να πάρουν εκείνο που θα ήταν πλησιέστερο στην πόρτα αλλιώς να μην αγοράσουν τίποτα.

Παρατηρούμε ότι η κατάσταση z είναι απορροφητική κατάσταση και έχει μηδενική αμοιβή. Η ύπαρξη απορροφητικών καταστάσεων με μηδενική αμοιβή είναι αρκετά συχνή στα μοντέλα ΜΔΑ. Για παράδειγμα, στη βιολογία, μία τέτοια κατάσταση μπορεί να σήμαινε τον θάνατο ενός οργανισμού.

Για να λυθούν προβλήματα όπως αυτό στο προηγούμενο παράδειγμα θα χρειαστούμε τη στοχαστική εκδοχή της εξίσωσης των αποδόσεων που είδαμε προηγουμένως στο ντετερμινιστικό δυναμικό προγραμματισμό¹⁴. Αν $\pi_t(x|a)$ είναι η απόδοση που θα έχω στο μέλλον αν επιλέξω την κίνηση a ενώ βρίσκομαι στην κατάσταση x τη χρονική στιγμή t και συμπεριφέρομαι βέλτιστα τις χρονικές στιγμές $t + 1, t + 2, \dots, T - 1$.

Οπότε:

$$\pi_t(x|a) = r(x, a) + \sum_{x' \in X} p(x'|x, a)\pi_{t+1}^*(x').$$

Όπου $\pi_t^*(x)$ η αμοιβή στο μέλλον αν ξεκινήσουμε από την κατάσταση x τη χρονική στιγμή t και με βέλτιστη συμπεριφορά τις χρονικές στιγμές $t + 1, t + 2, \dots, T - 1$.

Τότε η ισότητα του στοχαστικού δυναμικού προγραμματισμού είναι:

$$\begin{aligned} \pi_t^*(x) &= \max_{a \in A} \pi_t(a|x) \\ &= \max_{a \in A} \left[r_t(x, a) + \sum_{x' \in X} p(x'|x, a)\pi_{t+1}^*(x') \right]. \end{aligned}$$

Αν δύο κινήσεις, έστω α και β , οδηγούν και οι δύο στη μέγιστη απόδοση $\pi^*(x)$, τότε μπορεί να επιλεγεί οποιαδήποτε από αυτές. Στην πραγματικότητα ο συνδυασμός « α με πιθανότητα p και β με πιθανότητα $1 - p$ » δίνει επίσης την ίδια απόδοση στο μέλλον και είναι μέγιστη, αλλά στην περίπτωση που έχουμε τυχαιότητα θεωρείται καλύτερα ως ένας τρόπος να «σπάσει» την ισοπαλία παρά ως αναγκαιότητα.

Εκτός από μερικές πολύ σπάνιες περιπτώσεις, οι στοχαστικές ΜΔΑ δεν επιλύονται για αυθαίρετες παραμετρικές τιμές. Για παράδειγμα, στο πρόβλημα του σχήματος 3.1, αν όλες οι αμοιβές, οι πιθανότητες μετάβασης και ο ορίζοντας T αφήνονταν απροσδιόριστοι θα είχαμε να αντιμετωπίσουμε πάρα πολλές παραμέτρους με αποτέλεσμα να δυσχεραίνεται η επίλυση και να προκύπτει υψηλό υπολογιστικό κόστος. Ακόμα και αν μπορούσε να βρεθεί μία αναλυτική λύση, δεν θα ήταν εύκολο να διακρίνουμε πως μεταβάλλεται αυτή η λύση για τις διάφορες τιμές αυτών των

¹⁴ Bellman, Michael. Dynamic Programming. Dover Publications, 2003

παραμέτρων. Αντιθέτως, η συνήθης προσέγγιση είναι να σταθεροποιήσουμε τις παραμέτρους και να βρούμε μία αριθμητική λύση. Ένας ηλεκτρονικός υπολογιστής είναι ασφαλώς χρήσιμος όταν οι καταστάσεις, το πλήθος των επιλογών και ο ορίζοντας T δεν είναι μικρός αριθμός.

Παράδειγμα: Θεωρούμε το πρόβλημα ΜΔΑ που αντιστοιχεί στο Σχήμα 3.1. Επιπροσθέτως, υποθέτουμε ότι οι αποφάσεις πρέπει να ληφθούν στις χρονικές στιγμές $t = 0, 1$ και 2 (δηλαδή έχουμε ορίζοντα $T = 3$) και αμοιβές $r_3(x) = r_3(y) = r_3(z) = 0$. Αναφέρουμε επίσης ότι στην κατάσταση z δεν υπάρχουν διαθέσιμες επιλογές. $a^*(z, t) = \beta$ και $\pi_t^*(z) = 0$, για κάθε t . Η απουσία τερματικών αμοιβών μας οδηγεί στις αποδόσεις: $\pi_3^*(x) = \pi_3^*(y) = 0$. Οπότε:

Για $t = 2$, στην κατάσταση x , έχουμε:

$$\pi_2(x|a) = 2 + \pi_3^*(y) = 2$$

$$\pi_2(x|\beta) = 3 + \pi_3^*(x) = 3.$$

Άρα επιλέγουμε κίνηση $a^*(x, 2) = \beta$ με απόδοση $\pi_2^*(x) = 3$

Για $t = 2$, στην κατάσταση y , έχουμε:

$$\pi_2(y|a) = 5 + \pi_3^*(x) = 5$$

$$\pi_2(y|\beta) = 10 + \frac{1}{2}\pi_3^*(y) + \frac{1}{2}\pi_3^*(z) = 10$$

Άρα επιλέγουμε κίνηση $a^*(y, 2) = \beta$ με απόδοση $\pi_2^*(y) = 10$

Για $t = 1$, στην κατάσταση x , έχουμε:

$$\pi_1(x|a) = 2 + \pi_2^*(y) = 12$$

$$\pi_1(x|\beta) = 3 + \pi_2^*(x) = 6.$$

Άρα επιλέγουμε κίνηση $a^*(x, 1) = a$ με απόδοση $\pi_1^*(x) = 12$

Για $t = 1$, στην κατάσταση y , έχουμε:

$$\pi_1(y|a) = 5 + \pi_2^*(x) = 8$$

$$\pi_1(y|\beta) = 10 + \frac{1}{2}\pi_2^*(y) + \frac{1}{2}\pi_2^*(z) = 15.$$

Άρα επιλέγουμε κίνηση $a^*(y, 1) = \beta$ με απόδοση $\pi_1^*(y) = 15$.

Για $t = 0$, στην κατάσταση x , έχουμε:

$$\pi_0(x|a) = 2 + \pi_1^*(y) = 17$$

$$\pi_0(x|\beta) = 3 + \pi_1^*(x) = 15$$

Άρα επιλέγουμε κίνηση $a^*(x, 0) = a$ με απόδοση $\pi_0^*(x) = 17$

Για $t = 0$, στην κατάσταση y , έχουμε:

$$\pi_0(y|a) = 5 + \pi_1^*(x) = 17$$

$$\pi_0(y|\beta) = 10 + \frac{1}{2}\pi_1^*(y) + \frac{1}{2}\pi_1^*(z) = 17,5$$

Άρα επιλέγουμε κίνηση $a^*(y, 0) = a$ με απόδοση $\pi_0^*(y) = 17,5$

Έτσι οδηγούμαστε στην ακόλουθη βέλτιστη s^* :

$$s^* = \begin{matrix} & t = 0 & 1 & 2 \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{pmatrix} \alpha & \alpha & \beta \\ \beta & \beta & \beta \\ \beta & \beta & \beta \end{pmatrix} \end{matrix}$$

με απόδοση 17 αν η κατάσταση εκκίνησης είναι η x και 17,5 αν η κατάσταση εκκίνησης είναι η y .

3.4 ΒΕΛΤΙΣΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ ΓΙΑ ΠΕΠΕΡΑΣΜΕΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ

Σε αυτή την παράγραφο θα μελετήσουμε αν μία αμιγής στρατηγική που υπολογίστηκε με τη βοήθεια του δυναμικού προγραμματισμού είναι πράγματι βέλτιστη. Αποδεικνύεται, όπως θα φούμε παρακάτω πως είναι όντως βέλτιστη ακόμα και όταν το πλήθος των ομάδων των πιθανών στρατηγικών μεγαλώνει περιλαμβάνοντας περισσότερους τύπους γενικών στρατηγικών. Αυτό φυσικά δεν αποκλείει το ενδεχόμενο να υπάρχουν στρατηγικές που είναι το ίδιο αποδοτικές, αλλά σίγουρα δεν θα είναι καλύτερες. Ας θεωρήσουμε αρχικά πως η πιθανότητα των τυχαίων στρατηγικών εξαρτάται μόνο από την τρέχουσα κατάσταση κι ας συνεχίσουμε με τον παρακάτω ορισμό.

Ορισμός: Έστω $A(x, t)$ το σύνολο των διαθέσιμων κινήσεων στην κατάσταση x τη χρονική στιγμή t . Μία γενική Μαρκοβιανή στρατηγική β ορίζει μία κίνηση $a \in A(x, t)$ με πιθανότητα $f(a|x, t)$, όπου:

$$\sum_{a \in A(x, t)} f(a|x, t) = 1$$

Το σύνολο όλων των Μαρκοβιανών στρατηγικών σημειώνεται ως $\mathcal{B}$.

Θεώρημα: Έστω μία Μαρκοβιανή Διαδικασία Αποφάσεων πεπερασμένου ορίζοντα και s^* η στρατηγική που υπολογίστηκε μέσω του δυναμικού προγραμματισμού. Τότε η s^* είναι μία βέλτιστη Μαρκοβιανή στρατηγική για αυτή τη διαδικασία.

Απόδειξη: Η απόδειξη θα γίνει επαγωγικά στη χρονική στιγμή t , Έστω $\pi_t(x|\beta)$ η αναμενόμενη απόδοση τη χρονική στιγμή t για τη στρατηγική β που ακολουθήθηκε, δεδομένου ότι η διαδικασία βρίσκεται εκείνη τη στιγμή στην κατάσταση x . Υποθέτουμε τώρα πως η στρατηγική s^* (από την επίλυση με δυναμικό προγραμματισμό) είναι βέλτιστη τη χρονική στιγμή $t + 1$, δηλαδή:

$$\pi_{t+1}(x|\beta) \leq \pi_{t+1}(x|s^*) \equiv \pi_{t+1}^*(x) \quad \text{για κάθε } x \in X \text{ και } \beta \in \mathcal{B}.$$

Τότε, για τυχαίο β , έχουμε:

$$\begin{aligned} \pi_t(x|\beta) &= \sum_{a \in A(x,t)} f(a|x,t) \left[r(x,a) + \sum_{x' \in X} p(x'|x,a) \pi_{t+1}(x'|\beta) \right] \\ &\leq \sum_{a \in A(x,t)} f(a|x,t) \left[r(x,a) + \sum_{x' \in X} p(x'|x,a) \pi_{t+1}^*(x') \right] \\ &\leq \left(\sum_{a \in A(x,t)} f(a|x,t) \right) \pi_t^*(x) \\ &= \pi_t^*(x) \end{aligned}$$

Όπου η πρώτη ανισότητα προκύπτει από την αρχική υπόθεση μας, ενώ στη συνέχεια χρησιμοποιείται η ισότητα του στοχαστικού δυναμικού προγραμματισμού¹⁵. Διότι:

$$\pi_T(x|\beta) = r_T(x) = \pi_T^*(x) \quad \text{για κάθε } x \in X.$$

Η ανισότητα $\pi_t(x|\beta) \leq \pi_t^*(x)$ ισχύει λοιπόν για κάθε κατάσταση x και χρονική στιγμή t . Συνεπώς η συνθήκη (βελτίστου) $\pi_0(x|\beta) \leq \pi_0^*(x)$ για κάθε $x \in X$ ισχύει.

■

Έχοντας τώρα αποδείξει ότι η στρατηγική που βρέθηκε από το δυναμικό προγραμματισμό αποτελεί μία βέλτιστη Μαρκοβιανή στρατηγική, θα δούμε τι συμβαίνει αν αυξηθεί το πλήθος των κλάσεων των διαθέσιμων στρατηγικών που οδηγούν σε μία καλύτερη στρατηγική.

Ορισμός: Έστω $A(x,t)$ το σύνολο των διαθέσιμων κινήσεων στην κατάσταση x τη χρονική στιγμή t . Για κάθε ιστορικό $h_t(x)$ που οδηγεί στην κατάσταση x τη χρονική

¹⁵ Ross S. M. (1995), Introduction to Stochastic Dynamic Programming, Academic Press

στιγμή t , μία συμπεριφοριστική στρατηγική ορίζει μία κίνηση $a \in A(x, t)$ με πιθανότητα $\varphi(a|h_t(x), t)$, όπου:

$$\sum_{a \in A(x, t)} \varphi(a|h_t(x), t) = 1$$

Προφανώς οι Μαρκοβιανές στρατηγικές είναι υποσύνολο των συμπεριφοριστικών στρατηγικών· το υποσύνολο όπου οι αποφάσεις καθορίζονται όχι από το σύνολο του ιστορικού αλλά από το μέρος αυτού που αφορά την τρέχουσα κατάσταση.

Θεώρημα: Καμία συμπεριφοριστική στρατηγική δεν δίνει υψηλότερη απόδοση από τη στρατηγική που προκύπτει μέσω του δυναμικού προγραμματισμού.

Απόδειξη: Για κάθε συμπεριφοριστική στρατηγική, υπάρχει μία Μαρκοβιανή στρατηγική που δίνει την ίδια αναμενόμενη απόδοση. Συνεπώς καμία συμπεριφοριστική στρατηγική δεν είναι καλύτερη από καμία Μαρκοβιανή στρατηγική. Έχουμε όμως δείξει ότι οι στρατηγικές που προέρχονται από το δυναμικό προγραμματισμό είναι βέλτιστες μαρκοβιανές στρατηγικές, και έτσι προκύπτει το ζητούμενο.¹⁶

■

3.5 ΜΑΡΚΟΒΙΑΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟΦΑΣΕΩΝ ΜΗ ΠΕΠΕΡΑΣΕΝΟΥ ΟΡΙΖΟΝΤΑ – ΒΕΛΤΙΣΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ

Θα θεωρήσουμε τώρα διαδικασίες, στις οποίες δεν προσδιορίζεται το σημείο τερματισμού εκ των προτέρων. Συνεπώς η διαδικασία μπορεί να εξελίσσεται επ' άπειρον εκτός εάν παρθεί κάποια απόφαση που προκαλεί τον τερματισμό αυτής.

Στην περίπτωση λοιπόν, μίας διαδικασίας μη πεπερασμένου ορίζοντα, θα θεωρήσουμε πως η απόδοση, με σημείο εκκίνησης την κατάσταση x , θα προκύπτει από το όριο όταν $T \rightarrow \infty$ της αντίστοιχης απόδοσης πεπερασμένου ορίζοντα:

$$\pi(x|\sigma) = \mathbb{E} \left[\sum_{t=0}^{\infty} r_t(X_t, A_t) \right]$$

¹⁶ Competitive Markov Decision Processes, J. Filar, K. Vrieze (Springer, 1997)

$$= \sum_{t=0}^{\infty} \mathbb{E}[r_t(X_t, A_t)].$$

Προφανώς, εφόσον δεν υπάρχει εκ των προτέρων σημείο τερματισμού, δεν υπάρχει και τερματική αμοιβή. Αρχικά όμως, τίθεται το ερώτημα εάν η παραπάνω σειρά συγκλίνει για όλες τις πιθανές στρατηγικές.

Παράδειγμα: Θεωρούμε μία πολύ απλή διαδικασία με μία μόνο κατάσταση, την x . Σε αυτή την κατάσταση, το σύνολο των διαθέσιμων κινήσεων είναι $A = \{a_1, a_2\}$ και $r(x, a_1) = 1$, $r(x, a_2) = 2$, ανεξάρτητα από το χρόνο t . Προφανώς οποιαδήποτε κίνηση και αν επιλεγεί η διαδικασία επιστρέφει αναγκαστικά στην κατάσταση x . Υποθέτουμε τώρα πως έχουμε δύο στρατηγικές: $\sigma_1 = \text{"να επιλέγεις μόνο } a_1\text{"}$ και $\sigma_2 = \text{"να επιλέγεις μόνο } a_2\text{"}$. Οπότε με την παραπάνω σχέση για την απόδοση, έχουμε:

$$\begin{aligned} \pi(x|\sigma_1) &= \sum_{t=0}^{\infty} \mathbb{E}[r_t(X_t, A_t)] \\ &= \sum_{t=0}^{\infty} \mathbb{E}[1] = \sum_{t=0}^{\infty} 1 = \infty \end{aligned}$$

Και

$$\begin{aligned} \pi(x|\sigma_2) &= \sum_{t=0}^{\infty} \mathbb{E}[r_t(X_t, A_t)] \\ &= \sum_{t=0}^{\infty} \mathbb{E}[2] = \sum_{t=0}^{\infty} 2 = \infty \end{aligned}$$

Οπότε δεν μπορούμε να επιλέξουμε την καλύτερη στρατηγική.

Για προβλήματα τέτοιου είδους, όπου έχουμε σταθερή αμοιβή ($r_t = c$, $c \in \mathbb{R}$), ανεξάρτητη με το χρόνο, διακρίνουμε τις περιπτώσεις όπου

- i. $|c| < 1$, η απόδοση θα είναι $\pi(x|\sigma) = 0$ καθώς $T \rightarrow \infty$
- ii. $|c| > 1$, η απόδοση θα είναι $\pi(x|\sigma) = \infty$ καθώς $T \rightarrow \infty$

Για τη δεύτερη περίπτωση θα χρησιμοποιήσουμε τον συντελεστή δ , με $\delta \in [0, 1)$ στην συνάρτηση αμοιβής r_t , δηλαδή: $r_t(x, a) = \delta^t r(x, a)$. Τότε η απόδοση είναι:

$$\pi(x|\sigma) = \mathbb{E} \left[\sum_{t=0}^{\infty} \delta^t r(X_t, A_t) \right]$$

Και είναι πεπερασμένη για κάθε στρατηγική σ και κάθε αρχική κατάσταση x .

Έτσι καταφέρνουμε να εξασφαλίσουμε πεπερασμένη τιμή για την απόδοση. Συχνά αυτός ο παράγοντας δ είναι η πιθανότητα να συνεχίσει η διαδικασία για τουλάχιστον ένα ακόμη βήμα. Συνεπώς για πιθανότητα $1 - \delta$ έχουμε πρόβλημα στη διαδικασία (ανεξάρτητα από το ποια στρατηγική υιοθετήθηκε). Για παράδειγμα στη βιολογία, η πιθανότητα $1 - \delta$ θα μπορούσε να δηλώνει ότι ένας οργανισμός πεθαίνει ως αποτέλεσμα παραγόντων που δεν έχουν γίνει σαφείς στο πρόβλημα.

Στη συνέχεια εισάγεται ο συντελεστής δ στην απόδοση που προκύπτει από το δυναμικό προγραμματισμό για πρόβλημα πεπερασμένου ορίζοντα:

$$\pi_t^*(x) = \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right].$$

Το μοντέλο μη πεπερασμένου ορίζοντα έχει την ιδιότητα μετά από τη λήψη μίας απόφασης από ένα άτομο, το ίδιο θα βρεθεί στην ίδια θέση να αντιμετωπίσει το πρόβλημα όπως και πριν, παρόλο που μπορεί να ξεκινά από διαφορετική κατάσταση. Για το λόγο αυτό μπορούμε να θέσουμε $\pi_t^*(x) = \pi_{t+1}^*(x) = \pi^*(x)$ για κάθε $x \in X$ στην παραπάνω ισότητα.

Θεώρημα: Οι βέλτιστες αποδόσεις ικανοποιούν την ισότητα του δυναμικού προγραμματισμού:

$$\pi^*(x) = \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right].$$

Απόδειξη: Έστω σ μία τυχαία στρατηγική που επιλέγει την κίνηση a την χρονική στιγμή $t = 0$ με πιθανότητα $f(a)$. Τότε

$$\pi(x|\sigma) = \sum_{a \in A(x)} f(a) \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi_1(x'|\sigma) \right]$$

Όπου $\pi_1(x'|\sigma)$ είναι η απόδοση που επιτυγχάνεται από τη στρατηγική σ , ξεκινώντας από την κατάσταση x' τη χρονική στιγμή $t = 1$. Εφόσον $\pi_1(x'|\sigma) \leq \pi^*(x')$, έχουμε:

$$\begin{aligned} \pi(x|\sigma) &\leq \sum_{a \in A(x)} f(a) \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right] \\ &\leq \sum_{a \in A(x)} f(a) \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right] \end{aligned}$$

$$= \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right].$$

Δεδομένου ότι η παραπάνω ανισότητα αληθεύει για την τυχαία σ , θα ισχύει και για τη βέλτιστη στρατηγική.

Έτσι έχουμε:

$$\begin{aligned} \pi^*(x) &\equiv \pi(x|\sigma^*) \\ &\leq \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right] \end{aligned}$$

Για να δείξουμε τώρα ότι ισχύει και η αντίθετη φορά της ανισότητας, θεωρούμε:

$$\hat{a} \in \arg \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right]$$

Και έστω $\sigma(\hat{a})$ να είναι η στρατηγική που επιλέγει την κίνηση $\hat{a}$ τη χρονική στιγμή $t = 0$ η οποία είναι βέλτιστη για τη χρονική στιγμή $t = 1$. Τότε:

$$\begin{aligned} \pi^*(x) &\geq \pi(x|\sigma(\hat{a})) \\ &= r(x, \hat{a}) + \delta \sum_{x' \in X} p(x'|x, \hat{a}) \pi^*(x') \\ &= \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right] \end{aligned}$$

Συνδυάζοντας τελικά τις δύο ανισότητες προκύπτει η ζητούμενη ισότητα και ολοκληρώνεται η απόδειξη.

■

Θεώρημα: (Μοναδικότητα)

Η απόδοση $\pi^*(x)$ αποτελεί μοναδική λύση για την εξίσωση

$$\pi^*(x) = \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi^*(x') \right]. \quad (*)$$

Απόδειξη: Έστω ότι υπάρχουν οι αποδόσεις $\pi_1(x)$ και $\pi_2(x)$ οι οποίες ικανοποιούν την (*). Θέτουμε

$$\hat{a}(x) \in \arg \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi_1(x') \right]$$

Για κάθε κατάσταση x έχουμε:

$$\begin{aligned}
\pi_1(x) - \pi_2(x) &= \delta \max_{a \in A(x)} \sum_{x' \in X} p(x'|x, a) [\pi_1(x') - \pi_2(x')] \\
&= \delta \sum_{x' \in X} p(x'|x, \hat{a}(x)) [\pi_1(x') - \pi_2(x')] \\
&\leq \delta \sum_{x' \in X} p(x'|x, \hat{a}(x)) |\pi_1(x') - \pi_2(x')| \\
&\leq \delta \sum_{x' \in X} p(x'|x, \hat{a}(x)) \max_{x' \in X} |\pi_1(x') - \pi_2(x')| \\
&\leq \delta \max_{x' \in X} |\pi_1(x') - \pi_2(x')|
\end{aligned}$$

Έστω τώρα $x_m \in \arg \max_{x' \in X} |\pi_1(x') - \pi_2(x')|$, τότε έχουμε:

$$\pi_1(x_m) - \pi_2(x_m) \leq \delta |\pi_1(x_m) - \pi_2(x_m)|$$

Επίσης, ισχύει ότι:

$$\pi_2(x_m) - \pi_1(x_m) \leq \delta |\pi_2(x_m) - \pi_1(x_m)|.$$

Οπότε:

$$|\pi_1(x_m) - \pi_2(x_m)| \leq \delta |\pi_1(x_m) - \pi_2(x_m)|$$

Έχουμε όμως ότι $\delta < 1$, συνεπώς πρέπει να είναι:

$$|\pi_1(x_m) - \pi_2(x_m)| = \max_{x' \in X} |\pi_1(x') - \pi_2(x')| = 0.$$

$$\Rightarrow |\pi_1(x_m) - \pi_2(x_m)| = 0 \text{ για κάθε } x \in X.$$

■

Ορισμός: Έστω σ μία σταθερή και όχι τυχαία στρατηγική που επιλέγει την κίνηση $a(x) \in A(x)$ κάθε φορά που η διαδικασία διέρχεται από την κατάσταση x . Αν $g(x): X \rightarrow \mathbb{R}$ είναι μία φραγμένη συνάρτηση, τότε ορίζουμε τον τελεστή T_σ ως εξής:

$$(T_\sigma g)(x) = r(x, a(x)) + \delta \sum_{x' \in X} p(x'|x, a(x)) g(x').$$

Στη συνέχεια υποθέτουμε ότι ο τελεστής T_σ δρα πάνω στη $g(x)$, και αν έπειτα αφήσουμε τον T_σ στο αποτέλεσμα της πρώτης διεργασίας προκύπτει ο τελεστής $(T_\sigma^2 g)(x)$. Ομοίως στο n -οστό βήμα θα έχουμε τον T_σ^n .

Λήμμα: Για κάθε φραγμένη συνάρτηση $g(x): X \rightarrow \mathbb{R}$ ισχύει: $\lim_{n \rightarrow \infty} (T_\sigma^n g)(x) = \pi(x|\sigma)$.

Απόδειξη:

$$\begin{aligned}
 (T_\sigma^n g)(x) &= r(x, a(x)) + \delta \sum_{x' \in X} p(x'|x, a(x))(T_\sigma^{n-1} g)(x') \\
 &= r(x, a(x)) + \delta \sum_{x' \in X} p(x'|x, a(x))r(x', a(x')) \\
 &\quad + \delta^2 \sum_{x'' \in X} p(x''|x', a(x'))(T_\sigma^{n-2} g)(x'')
 \end{aligned}$$

Οπότε για $\delta < 1$, έχουμε:

$$\lim_{n \rightarrow \infty} (T_\sigma^n g)(x) = \sum_{t=0}^{\infty} \mathbb{E}[r(X_t, a_t)] \equiv \pi(x|\sigma).$$

■

Θεώρημα: Έστω

$$a^*(x) \in \arg \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a)\pi^*(x') \right] \quad \text{για κάθε } x \in X$$

και σ^* μία σταθερή και όχι τυχαία στρατηγική που επιλέγει την κίνηση $a^*(x) \in A(x)$ κάθε φορά που η διαδικασία είναι στην κατάσταση x . Τότε η σ^* είναι βέλτιστη.

Απόδειξη: Από τον ορισμό της σ^* και την ισότητα του δυναμικού προγραμματισμού έχουμε:

$$(T_{\sigma^*} \pi^*)(x) = \pi^*(x)$$

το οποίο σημαίνει ότι:

$$(T_{\sigma^*}^n \pi^*)(x) = \pi^*(x) \quad \text{για κάθε } n.$$

Καθώς όμως $n \rightarrow \infty$, και λόγω του λήμματος που αποδείξαμε παραπάνω, έχουμε:

$$\pi(x|\sigma^*) = \lim_{n \rightarrow \infty} (T_{\sigma^*}^n \pi^*)(x) = \pi^*(x).$$

■

3.6 ΒΕΛΤΙΩΣΗ ΠΟΛΙΤΙΚΗΣ

Η ισότητα του δυναμικού προγραμματισμού προτείνει την ακόλουθη επαναληπτική διαδικασία για την εύρεση βέλτιστης στρατηγικής και τις αντίστοιχες αποδόσεις σε μία Μαρκοβιανή Διαδικασία Αποφάσεων μη πεπερασμένου ορίζοντα. Η διαδικασία καλείται «Βελτίωση Πολιτικής» διότι οι στρατηγικές καλούνται πολιτικές από τα άτομα που μελετούν ΜΔΑ αλλά όχι παίγνια. Η διαδικασία εξελίσσεται σύμφωνα με τα ακόλουθα αλγόριθμο:

1. Ξεκίνα επιλέγοντας μία τυχαία στρατηγική σ που επιλέγει την κίνηση $a(x)$ στην κατάσταση x .

2. Λύσε την εξίσωση:

$$\pi(x|\sigma) = r(x, a) + \delta \sum_{x' \in X} p(x'|x, a)\pi(x'|\sigma)$$

για να βρεις τις αποδόσεις που αντιστοιχούν στη στρατηγική του βήματος (1).

3. Βρες μία βελτιωμένη στρατηγική υπολογίζοντας:

$$\hat{a}(x) \in \arg \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a)\pi(x'|\sigma) \right].$$

4. Επανάλαβε τα βήματα (2) και (3) μέχρι η στρατηγική σου να μην αλλάζει.

Ακολουθεί ένα θεώρημα που επαληθεύει τον αλγόριθμο:

Θεώρημα: Υποθέτουμε πως έχουμε μία σταθερή αμιγή στρατηγική με απόδοση $\pi(x|\sigma)$. Έστω

$$\hat{a}(x) \in \arg \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a)\pi(x'|\sigma) \right]$$

για κάθε $x \in X$ και $\hat{\sigma}$ μία (σταθερή και όχι τυχαία) στρατηγική σ που επιλέγει την κίνηση $\hat{a}(x)$ κάθε φορά που η διαδικασία είναι στην κατάσταση x . Τότε η $\hat{\sigma}$ είναι είτε μία καλύτερη στρατηγική από την σ , είτε είναι και οι δύο βέλτιστες.

Απόδειξη: Θεωρούμε τον τελεστή $T_{\hat{\sigma}}$ που συνδέεται με μία νέα στρατηγική $\hat{\sigma}$ πάνω σε μία φραγμένη και πραγματική συνάρτηση $g(x)$:

$$(T_{\hat{\sigma}}g)(x) = r(x, \hat{a}(x)) + \delta \sum_{x' \in X} p(x'|x, \hat{a}(x))g(x').$$

Από τον ορισμό όμως του $\hat{a}(x)$ έχουμε:

$$T_{\hat{\sigma}}\pi(x|\sigma) \geq \pi(x|\sigma) \quad \text{για κάθε } x \in X.$$

Εφαρμόζοντας επαναλαμβανόμενα αυτή την ανίσωση με τον τελεστή $T_{\hat{\sigma}}$, έχουμε:

$$T_{\hat{\sigma}}^n \pi(x|\sigma) \geq T_{\hat{\sigma}}^{n-1} \pi(x|\sigma) \geq \dots \geq \pi(x|\sigma).$$

Στη συνέχεια, για $n \rightarrow \infty$ και λόγω του λήμματος παραπάνω, έχουμε:

$$\pi(x|\hat{\sigma}) \geq \pi(x|\sigma) \quad \text{για κάθε } x \in X.$$

Δείξαμε με αυτό τον τρόπο πως η στρατηγική $\hat{\sigma}$ είναι τουλάχιστον όσο καλή είναι και η προηγούμενη. Θα δείξουμε έπειτα πως αν η στρατηγική $\hat{\sigma}$ δεν είναι αυστηρά καλύτερη από τη σ σε τουλάχιστον μια κατάσταση, τότε θα είναι και οι δύο βέλτιστες.

Έχουμε ήδη αποδείξει ότι:

$$\pi(x|\hat{\sigma}) \geq T_{\hat{\sigma}}\pi(x|\sigma) \geq \pi(x|\sigma).$$

Υποθέτουμε τώρα ότι $\pi(x|\hat{\sigma}) = \pi(x|\sigma)$ για κάθε $x \in X$. Τότε:

$$\begin{aligned} \pi(x|\hat{\sigma}) &= T_{\hat{\sigma}}\pi(x|\sigma) = T_{\hat{\sigma}}\pi(x|\hat{\sigma}) \\ &= \max_{a \in A(x)} \left[r(x, a) + \delta \sum_{x' \in X} p(x'|x, a) \pi(x'|\hat{\sigma}) \right]. \end{aligned}$$

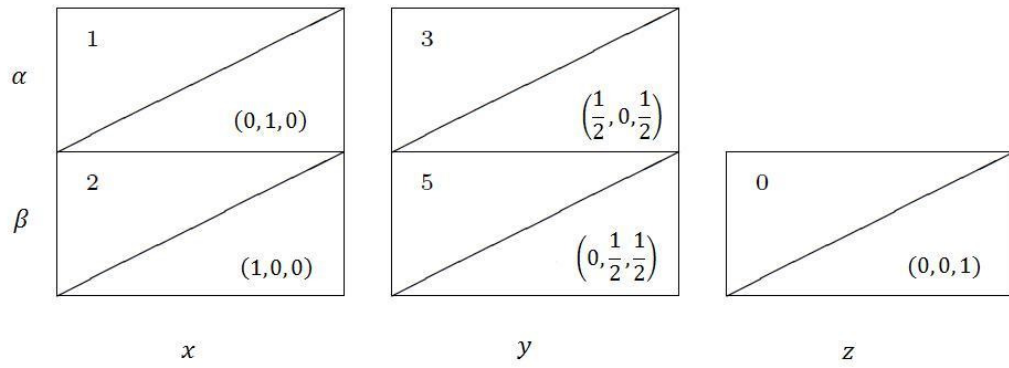
Οπότε η $\pi(x|\hat{\sigma})$ ικανοποιεί την ισότητα για τις βέλτιστες στρατηγικές του δυναμικού προγραμματισμού. Από τη μοναδικότητα της λύσης της εξίσωσης, έχουμε ότι:

$$\pi(x|\hat{\sigma}) = \pi^*(x) = \pi(x|\sigma)$$

το οποίο αποδεικνύει ότι η $\hat{\sigma}$ και η σ είναι και οι δύο βέλτιστες στρατηγικές.

■

Παράδειγμα: Θεωρούμε το πρόβλημα που φαίνεται στο διάγραμμα του σχήματος 3.2 παρακάτω με συντελεστή $\delta = \frac{9}{10}$. Έστω η αρχική στρατηγική $\sigma_0 = \{a(x) = \beta, a(y) = \beta\}$.



Σχήμα 3.2

Θα εφαρμόσουμε τον αλγόριθμο «βελτίωση πολιτικής»:

Επανάληψη 1^η: Οι αποδόσεις για την σ_0 είναι:

$$\pi(x|\sigma_0) = 2 + \delta\pi(x|\sigma_0) = 20 \quad \text{και}$$

$$\pi(y|\sigma_0) = 5 + \frac{1}{2}\delta\pi(y|\sigma_0) = \frac{100}{11}.$$

Στη συνέχεια, οι αποδόσεις για την αλλαγή της κίνησης που είχαμε σε κάθε κατάσταση σε a είναι:

$$r(x, a) + \delta\pi(y|\sigma_0) = 1 + \frac{9}{10} \cdot \frac{100}{11} = \frac{101}{11} \quad \text{και}$$

$$r(y, a) + \frac{1}{2}\delta\pi(x|\sigma_0) = 3 + \frac{9}{20} \cdot 20 = 12$$

Εφόσον $\frac{101}{11} < 20$ και $12 > \frac{100}{11}$, μπορούμε να καταλήξουμε στις κινήσεις

$$\hat{a}(x) = \beta \quad \text{και} \quad \hat{a}(y) = a.$$

Θεωρούμε τώρα μία νέα στρατηγική σ_1 .

Επανάληψη 2^η: Οι αποδόσεις για την σ_1 είναι:

$$\pi(x|\sigma_1) = 2 + \delta\pi(x|\sigma_1) = 20 \quad \text{και}$$

$$\pi(y|\sigma_1) = 3 + \frac{1}{2}\delta\pi(y|\sigma_1) = 12.$$

Τότε οι αποδόσεις για την αλλαγή της κίνησης που είχαμε σε κάθε κατάσταση, σε a είναι:

$$r(x, a) + \delta\pi(y|\sigma_1) = 1 + \frac{9}{10} \cdot 12 = \frac{118}{10} \quad \text{και}$$

$$r(y, a) + \frac{1}{2} \delta \pi(x|\sigma_1) = 5 + \frac{9}{20} \cdot 20 = 14$$

Από τις τελευταίες τιμές συμπεραίνουμε ότι μια αλλαγή στρατηγικής δεν επιφέρει καλύτερες αποδόσεις, οπότε η βέλτιστη στρατηγική είναι η:

$$\sigma^* = \{a(x) = \beta, a(y) = \beta\}$$

Η βέλτιστη στρατηγική που βρέθηκε από τον αλγόριθμο δείχνει λογική για $\delta \rightarrow 1$ διότι μειώνει την πιθανότητα η διαδικασία να τελειώσει στην κατάσταση z δημιουργώντας άπειρες μηδενικές αμοιβές.

ΚΕΦΑΛΑΙΟ 4^ο

ΠΑΙΓΝΙΑ ΠΛΗΡΟΥΣ ΠΛΗΡΟΦΟΡΗΣΗΣ

4.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΓΙΑ ΤΑ ΠΑΙΓΝΙΑ ΜΕ ΠΛΗΡΗ ΠΛΗΡΟΦΟΡΗΣΗ

Ένα διαδραστικό πρόβλημα απόφασης περιλαμβάνει τουλάχιστον δύο άτομα, τα οποία καλούνται να λάβουν μία απόφαση η οποία καθορίζει τις αποδόσεις για κάθε άτομο που συμμετέχει. Αυτό το πρόβλημα καλούμε παιχνίδι, και τα άτομα καλούνται παίκτες. Σε αντίθεση με το προηγούμενο κεφάλαιο στο οποίο κάθε άτομο αποφάσιζε για τις κινήσεις του με σκοπό την βέλτιστη απόδοση, τώρα πια κάθε παίκτης στις αποφάσεις του πρέπει να λαμβάνει υπόψη ότι επηρεάζονται οι αποδόσεις και από τις κινήσεις των άλλων παικτών. Αφού λοιπόν δώσουμε βασικούς ορισμούς, θα μελετήσουμε παρακάτω τύπους παιγνίων μέσα από σημαντικά παραδείγματα.

Ορισμός: Παίγνιο πλήρους πληροφόρησης είναι εκείνο στο οποίο οι παίκτες γνωρίζουν την εκτεταμένη μορφή του παιχνιδιού και τις αποδόσεις για κάθε στρατηγική.¹⁷

Συνεπώς, η συνάρτηση απόδοσης κάθε παίκτη με βάση το συνδυασμό κινήσεων που επέλεξαν αποτελεί κοινή γνώση.

Σε κάθε παίγνιο απαιτείται ο προσδιορισμός:

- α) των παικτών,
 - β) των διαθέσιμων στρατηγικών του καθενός,
 - γ) των αποδόσεων (payoffs) για κάθε πιθανή στρατηγική ή συνδυασμό αυτών.
- Οπότε αν $S_i = \{s_{i1}, s_{i2}, \dots, s_{im}\}$ το σύνολο των στρατηγικών του i παίκτη, με συνάρτηση απόδοσης a_i , έχουμε το παίγνιο $G = \{S_1, S_2, \dots, S_n | a_1, a_2, \dots, a_n\}$

Ορισμός: Στατικό καλείται ένα παίγνιο στο οποίο κάθε παίκτης καλείται να λάβει μία απόφαση ταυτόχρονα με τον αντίπαλο (ή τους αντιπάλους) χωρίς να γνωρίζει την απόφαση του (ή τους).

¹⁷Κ. Μηλολιδάκης, Θεωρία Παιγνίων-Μαθηματικά Μοντέλα Σύγκρουσης και Συνεργασίας, εκδόσεις Σοφία

Ορισμός: Το σύνολο των κινήσεων που καθορίζουν πλήρως τη συμπεριφορά του παίκτη στο παιχνίδι καλούμε αμιγή στρατηγική.

Παράδειγμα: Το δίλημμα του φυλακισμένου

Δύο κρατούμενοι (Α και Β) από τις αρχές θεωρούνται ύποπτοι για ένα σοβαρό έγκλημα. Παρόλα αυτά ο κατηγορος δεν νιώθει ότι έχει αρκετές αποδείξεις για την καταδίκη τους. Οι ύποπτοι κρατούνται σε διαφορετικά κελιά και τους προτείνεται η εξής συμφωνία:

- Αν ο κρατούμενος Α καταθέσει και ο κρατούμενος Β δεν μιλήσει, ο Α αφήνεται ελεύθερος και ο Β θα υποστεί κάθειρξη 10 ετών.
- Αν ο κρατούμενος Β καταθέσει και ο κρατούμενος Α δεν μιλήσει, ο Β αφήνεται ελεύθερος και ο Α θα υποστεί κάθειρξη 10 ετών.
- Αν και οι δύο κρατούμενοι παραμείνουν σιωπηλοί και δεν ομολογήσουν τότε θα καταδικαστούν σε κάθειρξη 6 μηνών. (0,5 έτη)
- Αν καταθέσουν και οι δύο θα καταδικαστούν σε 2 έτη φυλάκισης.

Κάθε κρατούμενος πρέπει να αποφασίσει αν θα προδώσει τον άλλο ομολογώντας ή θα σωπάσει. Κανένας από τους δύο κρατούμενους δεν γνωρίζει την απόφαση του άλλου (αν θα ομολογήσει ή θα σωπάσει).

Με βάση τα δεδομένα μας, διαμορφώνεται ο παρακάτω πίνακας αποδόσεων με. Ας αναφέρουμε εδώ πως στόχος κάθε κρατούμενου (παίκτη) είναι η ελαχιστοποίηση της διάρκειας της φυλάκισης. Αν λοιπόν Ο σημαίνει ομολογία και Σ σιωπή, έχουμε:

Πίνακας 4.1

Στρατηγική		Κρατούμ.Β	
		Ο	Σ
Κρατούμ. Α	Ο	(-2, -2)	(0,- 10)
	Σ	(-10, 0)	(-0.5 , -0.5)

Τι πρέπει λοιπόν να κάνει ένας κρατούμενος; Ας δούμε το πρόβλημα από την πλευρά του κρατούμενου Α. Αν ο Β σωπάσει πρέπει να ομολογήσει γιατί έχει καλύτερη απόδοση (0 έναντι -0.5). Αν ο Β ομολογήσει επίσης πρέπει να ομολογήσει για τον ίδιο λόγο (απόδοση -2 έναντι -10). Άρα ανεξάρτητα από την απόφαση του κρατουμένου

B, ο A πρέπει να ομολογήσει. Με τον ίδιο τρόπο καταλήγουμε πως και ο παίκτης B θα αναγκαστεί να ομολογήσει αφού αντιστοιχεί στην καλύτερη δυνατή απόδοση και για εκείνον. Έτσι η λύση του προβλήματος είναι να ομολογήσουν και οι δύο (O, O).

Ορισμός: Μία λύση καλείται βέλτιστη κατά Pareto¹⁸ όταν η απόδοση ενός παίκτη δεν γίνεται να αυξηθεί χωρίς να μειωθεί η απόδοση του αντιπάλου.

Ορισμός: Σημείο Ισορροπίας Nash (για παίγνιο με 2 παίκτες) καλείται ένα ζεύγος αμιγών στρατηγικών που αντιστοιχεί σε κάθε παίκτη και κάθε μία είναι η καλύτερη δυνατή ως απάντηση στην άλλη.

Ας δώσουμε τώρα και έναν πιο αυστηρό μαθηματικό ορισμό:¹⁹

Έστω σύνολο A και 2^A το δυναμοσύνολο του. Η συνάρτηση $F: X \rightarrow 2^A \subset A$ με $x \in X$ και $F(x) \in 2^A$ καλείται ανταπόκριση από το x στο A . Ένα σημείο $x^* \in X$ καλείται σταθερό σημείο της ανταπόκρισης αν και μόνο αν $x^* \in F(x^*)$.

Έστω μία στρατηγική κατάσταση s στο σύνολο S των στρατηγικών καταστάσεων. Ορίζουμε το σύνολο των βέλτιστων αποκρίσεων (bestreaction) του παίκτη i ($i=1,2, \dots, n$) στη στρατηγική s : $BR^i(s)$ και $BR(s) = \prod_{i=1}^n BR^i(s)$. Η $BR(s)$ είναι μία ανταπόκριση, άρα αν το διάνυσμα $u = (u_1, u_2, \dots, u_n) \in BR(s)$ η u_1 είναι η βέλτιστη απάντηση του 1^{ου} παίκτη στην s , η u_2 βέλτιστη απάντηση του 2^{ου} παίκτη στην s , ..., η u_n η βέλτιστη απάντηση του n -οστού παίκτη στην s . Άρα αν $s^* \in S$, τέτοιο ώστε $s^* \in BR(s)$, τότε για κάθε παίκτη i η s^{*i} είναι η βέλτιστη απάντηση στην s^* και καλείται Σημείο Ισορροπίας Nash.

4.2 ΣΤΑΤΙΚΑ ΠΑΙΓΝΙΑ 2 ΠΑΙΚΤΩΝ ΜΗΔΕΝΙΚΟΥ ΑΘΡΟΙΣΜΑΤΟΣ

Τα παιχνίδια μηδενικού αθροίσματος (με δύο παίκτες) έχουν λύση τέτοια ώστε το συνολικό άθροισμα των αμοιβών να είναι μηδενικό. Για να το αντιληφθούμε καλύτερα αρκεί να σκεφτούμε το σκάκι όπου όταν ο ένας παίκτης κερδίζει το παιχνίδι

¹⁸ Από τον Ιταλό οικονομολόγο Vilfredo Pareto, Neisser Hans (1940), A Note on Pareto's Theory of Production, *Econometrica*, 8 July, 253-262

¹⁹Κ. Μηλολιδάκης, Θεωρία Παιγνίων-Μαθηματικά Μοντέλα Σύγκρουσης και Συνεργασίας, εκδόσεις Σοφία

ο άλλος έχει αναγκαστικά χάσει, ή το πόκερ στο οποίο όταν ο ένας παίκτης κερδίσει 100 ευρώ ο άλλος έχει χάσει 100 ευρώ. Ας δούμε όμως ένα απλό πρόβλημα μηδενικού αθροίσματος δύο παικτών:

Παίγνιο “Morra”²⁰: Έχουμε το παιχνίδι «μονά-ζυγά» σύμφωνα με το οποίο δύο παίκτες κάνοντας γροθιά το χέρι τους, μετρούν μέχρι το 3 και δείχνουν 1 ή 2 δακτύλα. Αν το άθροισμα των δακτύλων και των δύο παικτών είναι άρτιο (ζυγός αριθμός) κερδίζει ο παίκτης A 1 χρηματική μονάδα, ενώ αν το άθροισμα είναι περιττό (μονός αριθμός) κερδίζει ο παίκτης B 1 χρηματική μονάδα. Οι παίκτες συνεπώς γνωρίζουν τις στρατηγικές που έχουν και τον πίνακα αποδόσεων όπως απεικονίζεται παρακάτω:

Πίνακας 4.2

Στρατηγική		Παίκτης B	
		1	2
Παίκτης A	1	(1, -1)	(-1, 1)
	2	(-1, 1)	(1, -1)

Παρατηρούμε εδώ αρχικά πως δεν υπάρχει αυστηρά κυριαρχούμενη στρατηγική και πως δεν υπάρχει ένα σημείο ισορροπίας. Συμπεραίνουμε πως η ιδιότητα ύπαρξης Σημείου Ισορροπίας Nash δεν υφίσταται.

4.3 ΕΠΙΛΥΣΗ ΠΑΙΓΝΙΩΝ ΜΕ ΠΛΗΡΗ ΠΛΗΡΟΦΟΡΗΣΗ

4.3.1 ΜΕΘΟΔΟΣ Minimax

Μέθοδος Minimax: Για κάθε παίγνιο με γνωστό πίνακα αποδόσεων $A = (a_{ij})^{n \times n}$ και αμιγείς στρατηγικές, το σημείο (i_k, j_k) που αντιστοιχεί στο μεγιστοελάχιστο της

²⁰ Αναφορά στο παιχνίδι odds and evens: [http://en.wikipedia.org/wiki/Morra_\(game\)](http://en.wikipedia.org/wiki/Morra_(game))

γραμμής και ελαχιστομέγιστο της στήλης του πίνακα αποδόσεων είναι σημείο ισορροπίας Nash (Σ.Ι.). Συγκεκριμένα, αν η κάτω τιμή του παιχνιδιού $\underline{w} = \max_i(\min_j(a_{ij}))$ και η άνω τιμή $\bar{w} = \min_j(\max_i(a_{ij}))$ συμπίπτουν ($\underline{w} = \bar{w}$) τότε έχουμε σημείο ισορροπίας κατά Nash.

Ας δούμε αν επιβεβαιώνεται η μέθοδος στο προαναφερθέν παίγνιο Morra:

Σύμφωνα με τον πίνακα 1.1 έχουμε ότι $\min_j(a_{1j}) = -1$ και $\min_j(a_{2j}) = -1$ (ελάχιστα των στηλών), οπότε και το μέγιστο από τα ελάχιστα είναι:

$$\underline{w} = \max_i(\min_j(a_{ij})) = -1 ,$$

Ενώ, $\max_i(a_{i1}) = 1$ και $\max_i(a_{i2}) = 1$, οπότε η ελάχιστη τιμή από τα δύο μέγιστα είναι $\bar{w} = \min_j(\max_i(a_{ij})) = 1$. Συνεπώς δεν έχουμε σημείο ισορροπίας (Σ.Ι.) εφ' όσον $\underline{w} \neq \bar{w}$.

4.3.2 ΔΙΑΔΟΧΙΚΗ ΑΠΑΛΟΙΦΗ ΑΥΣΤΗΡΑ ΚΥΡΙΑΡΧΟΥΜΕΝΩΝ ΣΤΡΑΤΗΓΙΚΩΝ

Ας μελετήσουμε το παρακάτω παίγνιο:

Δύο εταιρίες κινητής τηλεφωνίας Α και Β δημιούργησαν μία καινούρια υπηρεσία για το internet και σκοπεύουν να τη διαφημίσουν την ίδια χρονική περίοδο. Δεχόμαστε ότι το κόστος διαφήμισης είναι το ίδιο σε κάθε περίπτωση κι έτσι έχουν τις επιλογές:

1. Να διαφημιστεί η υπηρεσία σε τηλεοπτικό κανάλι
2. Να διαφημιστεί στο ραδιόφωνο
3. Να διαφημιστεί σε εφημερίδες και περιοδικά

Δίνεται ο πίνακας (1.2) διπλής εισόδου με τις αποδόσεις. Οι ακέραιοι αριθμοί που εμφανίζονται δηλώνουν πόσοι, σε χιλιάδες, υποψήφιοι πελάτες θα ενδιαφερθούν για το προϊόν της εταιρίας.

Πίνακας 4.3

Στρατηγική		Παίκτης B		
		1	2	3
Παίκτης A	1	(4,4)	(7,3)	(9,0)
	2	(3,7)	(3,3)	(5,1)
	3	(0,9)	(1,5)	(1,1)

Σε κάθε ζεύγος αποδόσεων το στοιχείο στα αριστερά είναι για τον παίκτη A (εταιρία A), ενώ το στοιχείο στα δεξιά αντιστοιχεί στον παίκτη B (εταιρία B).

Δεδομένου ότι κάθε εταιρία σκέφτεται ορθολογικά, η στρατηγική 3 θα απορριφθεί από την εταιρία B γιατί οι αποδόσεις της είναι μικρότερες από αυτές της στρατηγικής 2 ($0 < 3$, $1 < 3$, $1 < 5$). Επομένως, λέμε ότι η στρατηγική 3 κυριαρχείται από την στρατηγική 2. Προφανώς αυτό το γνωρίζει και η εταιρία A. Έτσι προκύπτει ο ισοδύναμος πίνακας (1.3)

Πίνακας 4.4

Στρατηγική		Παίκτης B	
		1	2
Παίκτης A	1	(4,4)	(7,3)
	2	(3,7)	(3,3)
	3	(0,9)	(1,5)

Στη συνέχεια, παρατηρούμε πως και η στρατηγική 2 κυριαρχείται από την 1, οπότε οδηγούμαστε στον πίνακα 1.4 ισοδύναμο:

Πίνακας 4.5

Στρατηγική	Παίκτης B	
	1	
Παίκτης A	1	(4,4)
	2	(3,7)
	3	(0,9)

Τώρα ο παίκτης A γνωρίζοντας ότι ο παίκτης B έχει επιλέξει την στρατηγική 1, ως βέλτιστη απόκριση θα επιλέξει κι αυτός τη στρατηγική 1, γιατί συγκρίνοντας τις αποδόσεις του ανά γραμμή έχουμε $0 < 3 < 4$.

Αρα οδηγούμαστε στη λύση (4,4) που προκύπτει αν και οι δύο εταιρίες επιλέξουν τη στρατηγική 1, δηλαδή να διαφημίσουν τη νέα υπηρεσία που προσφέρουν σε τηλεοπτικό κανάλι.

Υπάρχουν όμως και παίγνια τα οποία δεν έχουν αυστηρά κυριαρχούμενες στρατηγικές όπως στο παρακάτω παράδειγμα:

Παράδειγμα: Δύο πολιτικοί κατά την προεκλογική τους εκστρατεία διαθέτουν τις εξής πόλεις ως επιλογές για να δώσουν μία ομιλία:

Ο πρώτος πολιτικός μπορεί να πάει είτε στην Άρτα, είτε στην Τρίπολη, είτε στη Λάρισα, ενώ ο δεύτερος (B) μπορεί να επισκεφθεί είτε τα Γιάννενα, είτε τη Μεγαλόπολη, είτε τη Λαμία. Ποια είναι η βέλτιστη απόφαση για τον καθένα σύμφωνα με τον πίνακα αποδόσεων (πίνακας 1.4) που δίνεται;

Αρχικά μοντελοποιώντας το πρόβλημα παρατηρούμε πως οι στρατηγικές που διαθέτουν οι δύο πολιτικοί είναι οι εξής:

$$S_A = \{A, T, Λμ\} \text{ και } S_B = \{Γ, Μ, Λρ\}$$

Πίνακας 4.6

Στρατηγική		Παίκτης B		
		Γ	Μ	Δρ
Παίκτης A	A	(0,3)	(3,0)	(4,2)
	T	(3,0)	(0,3)	(4,2)
	Δμ	(2,4)	(2,4)	(5,5)

Παρατηρούμε λοιπόν πώς καμία στήλη δεν έχει αυστηρά κυριαρχούμενες στρατηγικές. Συγκρίνοντας την 1^η (Γ) με την 2^η (Μ) έχουμε: $3 > 0$ αλλά $0 < 3$ και $4 = 4$.

Πίνακας 4.7

Στρατηγική		Παίκτης B	
		Γ	Μ
Παίκτης A	A	(0, <u>3</u>)	(3, <u>0</u>)
	T	(3, <u>0</u>)	(0, <u>3</u>)
	Δμ	(2, <u>4</u>)	(2, <u>4</u>)

Ή την 1^η (Μ) με την 3^η (Δρ) όπου $3 > 2$, αλλά $0 < 2$ και $4 < 5$.

Πίνακας 4.8

Στρατηγική		Παίκτης B	
		Γ	Δρ
Παίκτης A	A	(0, <u>3</u>)	(4, <u>2</u>)
	T	(3, <u>0</u>)	(4, <u>2</u>)
	Δμ	(2, <u>4</u>)	(5, <u>5</u>)

4.4 ΜΙΚΤΕΣ ΣΤΡΑΤΗΓΙΚΕΣ

Στο παίγνιο *Morra* σε προηγούμενη παράγραφο είδαμε πως δεν έχουμε ισορροπία Nash. Αξίζει να παρατηρήσουμε πως κάθε παίκτης σε αυτό το παιχνίδι θα ήθελε να είχε ξεγελάσει τον αντίπαλο του. Παρόμοιες καταστάσεις συναντάμε σε διάφορα γνωστά παιχνίδια όπως το πόκερ, στο οποίο το μπλοφάρισμα είναι επιτρεπτό αλλά η συχνότητα με την οποία γίνεται από ένα παίκτη μπορεί να λειτουργήσει προς όφελος του αλλά και το ανάποδο. Στην περίπτωση που ο παίκτης δεν μπλοφάρει ποτέ οι αντίπαλοι θα πηγαίνουν πάσο, ενώ αν η μπλόφα γίνεται αρκετά συχνά το ρίσκο είναι μεγάλο και συνήθως οδηγεί στην ήττα. Βλέπουμε λοιπόν πως σε τέτοιου είδους παίγνια η ισορροπία Nash δεν υφίσταται εξ' αιτίας της αβεβαιότητας της δράσης του αντιπάλου. Με τη βοήθεια της *μικτής στρατηγικής*²¹ γίνεται η προσπάθεια να ομαλοποιηθεί το πρόβλημα της αβεβαιότητας και της τυχαιότητας των δράσεων των παικτών.

²¹ Η έννοια εμφανίζεται και στη θεωρία λήψης αποφάσεων, κεφάλαιο 2^ο.

Ορισμός: Σε παίγνιο της μορφής $G = \{S_1, S_2, \dots, S_n | a_1, a_2, \dots, a_n\}$, μικτή στρατηγική για τον παίκτη i με αντίστοιχο σύνολο δυνατών στρατηγικών $S_i = \{s_{i1}, s_{i2}, \dots, s_{im}\}$, καλείται μία κατανομή πιθανότητας $p_i = (p_{i1}, p_{i2}, \dots, p_{im})$ με $p_{ij} \in [0,1]$ ($j = 1, 2, \dots, m$) και $\sum_{j=1}^m p_{ij} = 1$.

Παράδειγμα: Δίνεται ο παρακάτω πίνακας διπλής εισόδου ο οποίος αφορά δύο ποδοσφαιριστές –έναν επιθετικό και έναν τερματοφύλακα- οι οποίοι βρίσκονται σε φάση πέναλτι και έχουν αντίστοιχα τις επιλογές να κινηθούν προς τη δεξιά κατεύθυνση ή προς την αριστερή του τέρματος, μόνο που ο επιθετικός κλοτσά την μπάλα λίγο χειρότερα προς τα δεξιά απ’ ότι στα αριστερά. Και το ερώτημα είναι αν πρέπει ο επιθετικός δεδομένων των συνθηκών να κλοτσά συχνότερα προς τα αριστερά.

Πίνακας 4.9

Στρατηγική	Τερματοφύλακας	
	Α	Δ
Επιθετικός	Α	(0, 1)
	Δ	(3/4, 1/4)

Αν τώρα θεωρήσουμε q την πιθανότητα ο επιθετικός να πάει αριστερά, τότε $1-q$ η πιθανότητα να πάει δεξιά.

Άρα οι αντίστοιχες αποδόσεις θα είναι:

Αριστερά: $q + \frac{1}{4}(1 - q)$

Δεξιά: $0q + 1 - q$

Ας υπολογίσουμε την πιθανότητα q , βάση των αποδόσεων. Πρέπει:

$$\alpha_1(\text{αριστ.}) = \alpha_1(\text{δεξιά})$$

$$\Rightarrow q + \frac{1}{4}(1 - q) = 1 - q$$

$$\Rightarrow \frac{3}{4}q + q = 1 - \frac{1}{4}$$

$$\Rightarrow \frac{7}{4}q = \frac{3}{4} \Rightarrow q = \frac{3}{7}$$

Αρα έχουμε τη μικτή στρατηγική: $p_{\text{επιθ}} = (\frac{3}{7}, \frac{4}{7})$

Ενώ για τον τερματοφύλακα, με πιθανότητα p να κινηθεί αριστερά και $1-p$ δεξιά και αποδόσεις:

Αριστερά: $0p + 1 - p$

Δεξιά: $\frac{3}{4}p + 0(1 - p)$

Κι έχουμε: $\alpha_2(\text{αριστ.}) = \alpha_2(\text{δεξιά})$

$$\Rightarrow 1 - p = \frac{3}{4}p \Rightarrow \frac{7}{4}p = 1 \Rightarrow p = \frac{4}{7}$$

Έτσι φαίνεται ο τερματοφύλακας να πηγαίνει συχνότερα στα αριστερά, κι έτσι έχει ένα μικρό προβάδισμα.

Αν ο τερματοφύλακας όμως κινούνταν ισοπίθانا δεξιά κι αριστερά, τότε το προβάδισμα θα ήταν του επιθετικού, ο οποίος θα έπρεπε να παίζει μόνο στα αριστερά κι έτσι θα κέρδιζε με πιθανότητα $\frac{1}{2}$ αντί για $\frac{3}{7}$ που είναι τώρα.

Όσον αφορά την ύπαρξη ισορροπίας (Nash), σύμφωνα με τον ορισμό που δώσαμε σε προηγούμενη παράγραφο, κάθε παίγνιο δύο παικτών με αμιγείς στρατηγικές έχει μία ισορροπία Nash σύμφωνα με την οποία η αμιγής στρατηγική του ενός είναι η καλύτερη δυνατή ως απόκριση στην αμιγή στρατηγική του αντιπάλου.

Για να διευρύνουμε αυτό τον ορισμό στις μικτές στρατηγικές αρκεί να αναζητήσουμε την μικτή στρατηγική που αποτελεί άριστη απάντηση στην μικτή στρατηγική του αντιπάλου (ή για $n > 2$ των άλλων παικτών). Κι αυτό αποτελεί τη γενικευμένη μορφή, περιλαμβάνοντας τα προηγούμενα διότι κάθε αμιγής στρατηγική μπορεί να παρασταθεί ως μικτή με πιθανότητα 1, δίνοντας στις υπόλοιπες αμιγείς στρατηγικές του παίκτη πιθανότητα μηδέν.

Οδηγούμαστε έτσι στο Θεώρημα ισορροπίας:

Θεώρημα (Nash):²² Στο κανονικής μορφής παίγνιο n παικτών $G = \{S_1, \dots, S_n; a_1, \dots, a_n\}$ αν S_i πεπερασμένο για κάθε i , τότε υπάρχει μία τουλάχιστον ισορροπία κατά Nash που μπορεί να περιλαμβάνει μικτές στρατηγικές.

²² Morris Peter, Introduction to Game Theory (1994), Springer

ΚΕΦΑΛΑΙΟ 5^ο

ΠΕΠΕΡΑΣΜΕΝΑ ΔΥΝΑΜΙΚΑ ΠΑΙΓΝΙΑ ΠΛΗΡΟΥΣ ΠΛΗΡΟΦΟΡΗΣΗΣ

5.1 ΔΕΝΤΡΑ ΠΑΙΓΝΙΩΝ

Μέχρι τώρα είδαμε στατικά παίγνια σύμφωνα με τα οποία οι κινήσεις πραγματοποιούνταν ταυτόχρονα. Υπάρχουν όμως αρκετές ενδιαφέρουσες καταστάσεις στις οποίες οι αποφάσεις λαμβάνονται σε διάφορες χρονικές στιγμές, ενώ συχνά κάποιες από τις προηγούμενες επιλογές αποτελούν κοινή γνώση τη στιγμή της επερχόμενης απόφασης. Αυτού του είδους τα παιχνίδια καλούνται δυναμικά, διότι υπάρχει ένα ξεκάθαρο χρονοδιάγραμμα που περιγράφει πότε παίρνει απόφαση κάθε παίκτης.

Τα δυναμικά παίγνια μπορούν να παρασταθούν από ένα δέντρο παιγνίου κι αυτό αποτελεί την εκτεταμένη μορφή του δέντρου απόφασης που αφορά ένα και μόνο άτομο όπως είδαμε σε προηγούμενο κεφάλαιο. Τη στιγμή που λαμβάνεται μία απόφαση θα τη δούμε σε ένα μικρό κύκλο και καλείται (όπως έχουμε ξαναδεί στο κεφάλαιο 2) κόμβος απόφασης. Τον κόμβο απόφασης ακολουθεί ένα κλαδί που αντιστοιχεί σε μία ενέργεια/δράση. Τη στιγμή που ένα άτομο λάβει όλες τις αποφάσεις και πραγματοποιήσει τις αντίστοιχες ενέργειες, λέμε πως το άτομο έφτασε στο τέλος του μονοπατιού μέσα από το δέντρο. Σε εκείνο το σημείο αναγράφονται οι αποδόσεις για κάθε μονοπάτι και η πρώτη απόδοση σε κάθε ζεύγος αντιστοιχεί στον παίκτη που παίζει πρώτος. Επίσης, ο σχεδιασμός του δέντρου γίνεται από πάνω προς τα κάτω εκφράζοντας τις χρονικές στιγμές οι οποίες περνούν με αύξουσα πορεία όσο κατεβαίνουμε στο δέντρο.²³

Παράδειγμα: Δύο μαϊμούδες, διαφορετικής ηλικίας και ανάπτυξης, τρέφονται με μπανάνες από ένα δέντρο το οποίο παράγει ένα φρούτο ανά ημέρα και δίνει ενέργεια 100 θερμίδες. Αν η (μεγάλη) μαϊμού Α τρέξει να σκαρφαλώσει στο δέντρο για να κουνήσει το κλαδί και να ρίξει το φρούτο καταναλώνει 20 θερμίδες, ενώ η

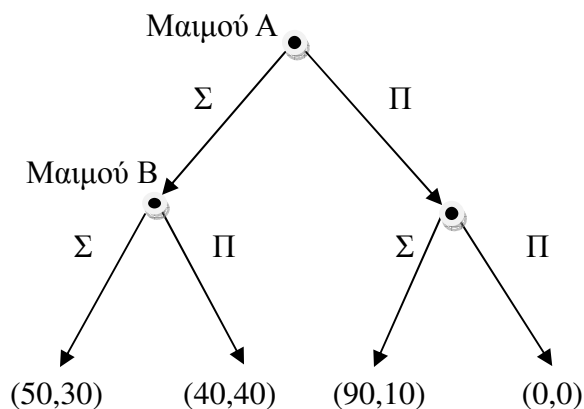
²³ Gintis Helbert, Game Theory Evolving, Princeton University Press 2nd Edition

(μικρόσωμη) μαϊμού B μπορεί να κάνει την ίδια διαδικασία χωρίς ενεργειακό κόστος. Στην περίπτωση που και οι δύο μαϊμούδες ξεκινήσουν να τρώνε ταυτόχρονα θα λάβουν 70 θερμίδες η μαϊμού A και 30 θερμίδες η μαϊμού B.

Αν όμως δεν ξεκινήσουν ταυτόχρονα αλλά ξεκινήσει πρώτα η μαϊμού B να τρώει (διότι δεν έχει προλάβει ακόμα να κατέβει η μαϊμού A από το δέντρο) τότε λαμβάνει 40 θερμίδες από τη μπανάνα ενώ οι υπόλοιπες 60 μένουν για τη μαϊμού A. Στην αντίθετη περίπτωση που ανέβει στο δέντρο η μαϊμού B, το μεγαλύτερο μέρος της μπανάνας καταναλώνεται από τη μαϊμού A η οποία θα πάρει 90 θερμίδες, και για την μαϊμού B αντιστοιχούν 10 θερμίδες από την μπανάνα.

Ποια είναι λοιπόν η τακτική που πρέπει να ακολουθήσει κάθε μαϊμού;

Αντιμετωπίζουμε λοιπόν το πρόβλημα σαν ένα παίγνιο με παίκτες τις μαϊμούδες A και B. Σε πρώτη φάση παρατηρούμε ποιες είναι οι επιλογές που έχει κάθε παίκτη: να σκαρφαλώσει στο δέντρο (Σ) ή να περιμένει την μπανάνα (Π). Έπειτα επιλέγουμε την μαϊμού A να αποφασίζει πρώτη αν θα σκαρφαλώσει στο δέντρο (Σ) ή όχι (Π). Αυτές οι δύο κινήσεις αποτελούν τις δύο στρατηγικές του πρώτου παίκτη. Φτιάχνοντας στη συνέχεια ένα δέντρο θα δούμε γραφικά τις δυνατές στρατηγικές του δεύτερου παίκτη.



Σχήμα 5.1

Παρατηρούμε λοιπόν πως οι διαθέσιμες στρατηγικές για την μαϊμού B είναι τέσσερις:

- i. Να σκαρφαλώσει σε κάθε περίπτωση ($\Sigma\Sigma$)
- ii. Να περιμένει το φρούτο σε κάθε περίπτωση ($\Pi\Pi$)
- iii. Να κάνει το ίδιο με τη μαϊμού A ($\Sigma\Pi$)

iv. Να κάνει το αντίθετο με τη μαϊμού A(ΠΣ)

Στους τερματικούς κόμβους (Σχήμα 4.1) αναφέρονται οι αποδόσεις για κάθε παίκτη. Η πρώτη τιμή αντιστοιχεί στη μαϊμού A, ενώ η δεύτερη στη μαϊμού B. Ας δούμε την περίπτωση που σκαρφαλώσει η μαϊμού A και περιμένει η μαϊμού B αναλυτικά. Τότε έχουμε σύμφωνα με τα δεδομένα μας:

Για τη μαϊμού A:

$$\left. \begin{array}{l} 60 \text{ θερμίδες που έλαβε από τη μπανάνα} \\ 20 \text{ θερμίδες που κατανάλωσε} \end{array} \right\} = 40 \text{ θερμίδες}$$

Για τη μαϊμού B:

$$\left. \begin{array}{l} 40 \text{ θερμίδες που έλαβε από τη μπανάνα} \\ 0 \text{ θερμίδες που κατανάλωσε} \end{array} \right\} = 40 \text{ θερμίδες}$$

Συνεπώς αν η μαϊμού A, η οποία αποφασίζει πρώτη, επιλέξει να σκαρφαλώσει, η μαϊμού B θα περιμένει, οπότε κερδίζει 40 θερμίδες, ενώ αν επιλέξει να περιμένει η μαϊμού B θα σκαρφαλώσει, άρα κερδίζει 90 θερμίδες. Η βέλτιστη τακτική για τη μαϊμού A είναι να περιμένει. Όσον αφορά τη μαϊμού B, δεδομένου ότι η μαϊμού A θα επιλέξει να περιμένει, η καλύτερη δυνατή στρατηγική που πρέπει να ακολουθηθεί είναι να σκαρφαλώσει για να κερδίσει 10 θερμίδες γιατί διαφορετικά δεν θα λάβει καθόλου ενέργεια.

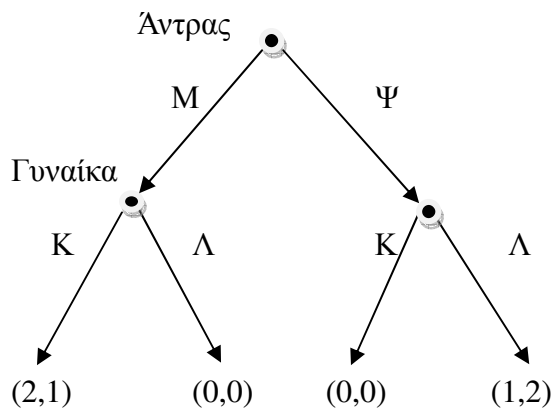
Άρα δεδομένων των συνόλων των στρατηγικών $S_A = \{\Sigma, \Pi\}$ και $S_B = \{\Sigma\Sigma, \Pi\Pi, \Sigma\Pi, \Pi\Sigma\}$, η λύση προκύπτει όταν πραγματοποιηθούν οι στρατηγικές $(\Pi, \Sigma\Pi)$. Το ίδιο αποτέλεσμα θα είχαμε όμως και με την ισοδύναμη λύση $(\Pi, \Sigma\Sigma)$ όμως η πρώτη εξασφαλίζει στη μαϊμού B ότι σε περίπτωση λανθασμένης απόφασης της A, θα έχει καλύτερη αμοιβή: $(\Sigma, \Sigma\Sigma) \leftrightarrow (50, 30)$ έναντι $(\Sigma, \Sigma\Pi) \leftrightarrow (40, 40)$. Τελικά η λύση του παιγνίου είναι η μαϊμού A να περιμένει και η μαϊμού B να κάνει το αντίθετο με απόδοση $(90, 10)$ δηλαδή το ζεύγος $(\Pi, \Sigma\Pi)$.

Αυτή η λύση αποτελεί το σημείο ισορροπίας κατά Nash.

Παράδειγμα(το δείπνο): Ένα παντρεμένο ζευγάρι θέλει να οργανώσει ένα δείπνο. Ο άντρας έχει αναλάβει να αγοράσει για το κυρίως πιάτο μοσχαρίσιο κρέας ή ψάρι, ενώ η σύζυγος του θα αγοράσει κρασί λευκό ή κόκκινο. Και οι δύο προτιμούν τους

συνδυασμούς μοσχαρίσιο κρέας-κόκκινο κρασί (M – K), ψάρι- λευκό κρασί (Ψ - Λ) σε σύγκριση με τα αντίθετα (M-Λ) ή (Ψ-K). Το πρόβλημα όμως είναι ότι ο άντρας προτιμά το μοσχαρίσιο κρέας για το κύριο πιάτο ενώ η γυναίκα του το ψάρι. Θεωρούμε παίκτη 1 τον άντρα και παίκτη 2 τη γυναίκα. Δίνονται παρακάτω οι αποδόσεις και το αντίστοιχο δέντρο του παιγνίου μας με τη βασική υπόθεση πως ο άντρας ενημερώνει τη σύζυγο για την αγορά του.

$$\begin{array}{lll} \alpha_1(M, K) = 2 & \alpha_1(\Psi, \Lambda) = 1 & \alpha_1(M, \Lambda) = \alpha_1(\Psi, K) = 0 \\ \alpha_2(M, K) = 1 & \alpha_2(\Psi, \Lambda) = 2 & \alpha_2(M, \Lambda) = \alpha_2(\Psi, K) = 0 \end{array}$$



Σχήμα 5.2

Ο προφανής τρόπος επίλυσης του παιγνίου μας είναι η οπισθοδρομική επαγωγή (backward induction). Στο σημείο αυτό όμως πρέπει να ανακαλέσουμε μία σημαντική πληροφορία: η γυναίκα (παίκτης 2) τη στιγμή της απόφασης της γνωρίζει τι έχει αποφασίσει ο άντρας της, αν αγόρασε δηλαδή μοσχαρίσιο κρέας ή ψάρι. Έτσι λοιπόν, ξέρει πως αν ο σύζυγος αγοράσει ψάρι εκείνη θα αγοράσει λευκό κρασί (απόδοση 2 έναντι 0 που δίνει η απόφαση Ψ-K) ενώ αν αγοράσει M, η καλύτερη απόκριση είναι να αγοράσει κόκκινο κρασί (απόδοση 1, έναντι 0 για το συνδυασμό M-Λ).

Φτάνοντας τώρα στην απόφαση του συζύγου, αυτός θα προτιμήσει μοσχαρίσιο κρέας, εφόσον η γυναίκα του, σε αυτή την επιλογή του, θα αγοράσει κόκκινο κρασί με απόδοση 2 έναντι της επιλογής ψάρι, σύμφωνα με την οποία η σύζυγος θα αγοράσει λευκό κρασί και δίνει απόδοση 1 μονάδα.

Το μενού λοιπόν θα περιλαμβάνει μοσχαρίσιο κρέας και κόκκινο κρασί.

5.2 ΙΣΟΡΡΟΠΙΑ ΚΑΤΑ NASH

Όπως είδαμε στο προηγούμενο παράδειγμα καταλήξαμε σε μία λύση (M-K).

Αποτελεί όμως αυτή η λύση ισορροπία κατά Nash;

Το σύνολο των δυνατών δράσεων για τον άντρα είναι M ή Ψ και το σύνολο των αμιγών στρατηγικών είναι επίσης $S_1 = \{M, \Psi\}$. Ενώ η σύζυγος για τις επιλογές της K ή Λ διαθέτει το εξής σύνολο αμιγών στρατηγικών: $S_2 = \{KK, K\Lambda, \Lambda\Lambda, \Lambda K\}$ όπου:

KK: αγορά K σε κάθε απόφαση του συζύγου

KΛ: αγορά K αν ο άντρας επιλέξει M και Λ αν ο άντρας επιλέξει Ψ

ΛΛ: αγορά Λ σε κάθε απόφαση του συζύγου

ΛK: αγορά Λ αν ο άντρας επιλέξει M και K αν ο άντρας επιλέξει Ψ.

Έτσι διαμορφώνεται σύμφωνα με την κανονική μορφή παιγνίου ο παρακάτω πίνακας αποδόσεων:

		Γυναίκα			
		KK	KΛ	ΛΛ	ΛK
Στρατηγική	Αντρας M	<u>2,1</u>	<u>2,1</u>	0,0	0,0
	Ψ	0,0	1, <u>2</u>	<u>1,2</u>	0,0

Πίνακας 5.1

Είναι προφανές ότι έχουμε τις εξής τρεις ισορροπίες κατά Nash: (M,KK) (M, KΛ) και (Ψ,ΛΛ). Αν σ_2^* η μικτή στρατηγική που αντιστοιχεί σε πιθανότητα p για την KK και 1-p για την KΛ, τότε παρατηρούμε πως το (M, σ_2^*) αποτελεί ισορροπία Nash κι έτσι φαίνεται πως η λύση που δώσαμε προηγουμένως με την οπισθοδρομική επαγωγή είναι ισορροπία Nash απλώς δεν είναι μοναδική.

Η στρατηγική ΛΛ της συζύγου, έχει νόημα αν ενημερώσει τον άντρα της γι' αυτή, διότι αν αυτός γνωρίζει πως σε κάθε περίπτωση θα έχει στο τραπέζι λευκό κρασί, δεδομένου ότι ο συνδυασμός Μ-Λ του προκαλεί τη λιγότερη ευχαρίστηση, θα «αναγκαστεί» να αγοράσει ψάρι. Αν όμως αυτός επιλέξει Μ εξ' αρχής τότε η γυναίκα θα αναγκαστεί να επιλέξει Κ γιατί ούτε σε εκείνη είναι ευχάριστος ο συνδυασμός Μ-Λ. Έπεται λοιπόν πως ο άντρας δεν θα ενδώσει στην «απειλή». Το συμπέρασμα από αυτά είναι πως η ισορροπία (Ψ, ΛΛ) βασίζεται στην «απειλή» πως η γυναίκα θα πάρει μία απόφαση που κινδυνεύει να μην την ευχαριστήσει καθόλου.

Αντίστοιχα, αν η γυναίκα ενημερώσει τον σύζυγο για την απόφαση της να αγοράσει κόκκινο κρασί σε κάθε περίπτωση (ΚΚ) δεν αποτελεί απειλή και διευκολύνει τη θέση του άντρα να επιλέξει Μ αφού ο ίδιος προτιμά περισσότερο τον συνδυασμό Μ-Κ από κάθε άλλο. Η κίνηση αυτή δεν έχει όμως νόημα εφόσον γνωρίζουμε πως με αυτή την επιλογή η γυναίκα αποκλείει μόνη της τον συνδυασμό Ψ-Λ που εκείνη προτιμά περισσότερο από κάθε άλλο.

Καταλήγουμε λοιπόν στο συμπέρασμα πως οι ισορροπίες Nash δεν ταιριάζουν απόλυτα στα δυναμικά παίγνια διότι δεν λαμβάνεται υπόψη η χρονική σειρά που παίρνονται η αποφάσεις καθώς οι παίκτες δεν αποφασίζουν ταυτόχρονα. Με την οπισθοδρομική επαγωγή βρήκαμε λύση που σέβεται τη χρονική σειρά και τη δομή του παιγνίου και ταυτόχρονα αποτελεί ισορροπία.

5.3 ΣΥΝΟΛΑ ΠΛΗΡΟΦΟΡΙΑΣ

Η πραγματική διαφορά ανάμεσα στα στατικά παίγνια και στα δυναμικά, προφανώς δεν είναι παράσταση σε δέντρα, αλλά έγκειται στις πληροφορίες που διαθέτουν οι παίκτες τη στιγμή που πρέπει να λάβουν μία απόφαση. Απαιτείται λοιπόν να μοντελοποιήσουμε αυτή τη διαφορά στα παίγνια και να ορίσουμε τα σύνολα πληροφορίας.

Ορισμός: Ένα σύνολο πληροφορίας ενός παίκτη είναι το σύνολο των κόμβων απόφασης ενός δέντρου που περιγράφει το παίγνιο με τις εξής ιδιότητες:

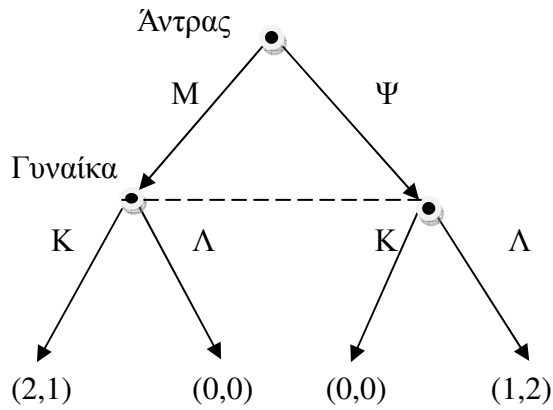
- i. Η απόφαση λαμβάνεται αποκλειστικά από τον παίκτη στον οποίο αναφερόμαστε.
- ii. Ο παίκτης δεν γνωρίζει σε ποιον κόμβο έχει φτάσει παρά μόνο ότι βρίσκεται σε έναν κόμβο του συνόλου. Έπεται πως ο παίκτης σε κάθε κόμβο έχει τις ίδιες διαθέσιμες επιλογές για να κινηθεί.

Το πρόβλημα του δείπνου που μελετήσαμε προηγουμένως αποτελούσε ένα δυναμικό παίγνιο. Αν όμως το εξετάζαμε σαν στατικό θα έπρεπε να υποθέσουμε πως κάθε παίκτης αποφασίζει ταυτόχρονα με τον άλλο και χωρίς πληροφορία για την επιλογή του άλλου. Συγκεκριμένα, η σύζυγος θα έπρεπε να επιλέξει κρασί μη γνωρίζοντας την επιλογή πιάτου που έχει κάνει ο άντρας της και έτσι διαμορφώνεται ο παρακάτω πίνακας για τη στατική μορφή:

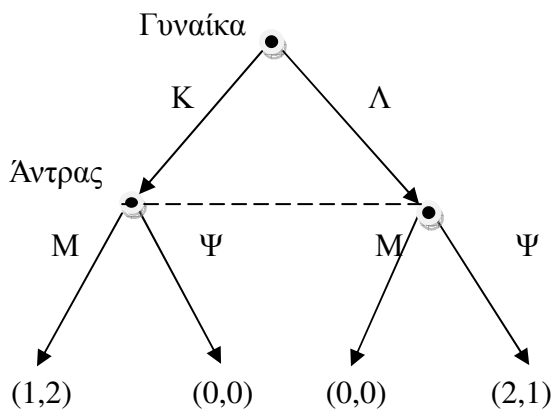
		Γυναίκα	
		K	Λ
Στρατηγική	M	(2,1)	(0,0)
	2	(0,0)	(1,2)
Άντρας			

Πίνακας 5.2

Φαίνεται λοιπόν πως σε αυτή την περίπτωση η γυναίκα έχει μόνο δύο στρατηγικές K, Λ οι οποίες είναι αμιγείς κι έτσι οδηγούμαστε σε δύο άλλα πιθανά δέντρα (σχήμα 4.2 και 4.3). Οι διακεκομμένες γραμμές υπάρχουν για να διαχωρίσουν τα σύνολα πληροφορίας κάθε παίκτη.



Σχήμα 5.3



Σχήμα 5.4

5.4 ΤΕΛΕΙΟΤΗΤΑ ΑΝΑ ΥΠΟΠΑΙΓΝΙΟ

Όπως είδαμε σε προηγούμενη παράγραφο η οπισθοδρομική επαγωγή μας έδωσε λύση σε δυναμικό παίγνιο που αποτελούσε και ισορροπία κατά Nash αλλά δεν ισχύει το αντίστροφο. Ενώ βρέθηκαν περισσότερες από μία ισορροπίες δεν ήταν λογικές λύσεις και για το λόγο αυτό θα προχωρήσουμε στον παρακάτω ορισμό:

Ορισμός: Υποπαίγνιο καλούμε ένα μέρος του δέντρου ενός παιγνίου που ικανοποιεί τις ακόλουθες συνθήκες:

- i. Ξεκινά σε έναν κόμβο απόφασης
- ii. Το σύνολο πληροφορίας που περιλαμβάνει τον αρχικό κόμβο δεν περιέχει κανέναν άλλο κόμβο. Ο παίκτης γνωρίζει όλες τις αποφάσεις που έχουν ληφθεί μέχρι εκείνη τη στιγμή.
- iii. Το μέρος του δέντρου που αντιστοιχεί στο υποπαίγνιο περιέχει όλους τους κόμβους που ακολουθούν τον αρχικό κόμβο και τίποτα άλλο.

Στο παράδειγμα δυναμικού παιγνίου με το δείπνο, με βάση το Σχήμα 4.1 υπάρχουν τα εξής υποπαίγνια: α) τα μέρη του δέντρου στα οποία ξεκινά να αποφασίζει η γυναίκα και β) ολόκληρο το δέντρο. Επίσης, στις περιπτώσεις των παιγνίων που οι παίκτες αποφασίζουν να δράσουν ταυτόχρονα το υποπαίγνιο ταυτίζεται με ολόκληρο το δέντρο παιγνίου.

Ορισμός: Τέλεια ισορροπία Nash ανά υποπαίγνιο είναι μία ισορροπία Nash στην οποία η καθορισμένη συμπεριφορά αποτελεί επίσης ισορροπία Nash για το υποπαίγνιο.

Συμπεραίνουμε λοιπόν πως η λύση (M, ΚΛ) που βρήκαμε στο παράδειγμα παιγνίου με το δείπνο αποτελεί μία τέλεια ισορροπία Nash διότι στα υποπαίγνια που αποφασίζει η γυναίκα η επιλογή Κ είναι η καλύτερη απόκριση στο Μ, και η Λ η καλύτερη απόκριση στο Ψ, ενώ η επιλογή Μ του άντρα είναι η βέλτιστη για όλο το παίγνιο. Αντιθέτως, η ισορροπία (Ψ, ΛΛ) δεν είναι τέλεια ισορροπία ανά υποπαίγνιο διότι η συμπεριφορά της γυναίκας στο υποπαίγνιο που ορίζεται από την απόφαση της γυναίκας όταν ο άντρας επιλέξει Μ δεν αποτελεί ισορροπία.

Έπεται λοιπόν από τον ορισμό πως κάθε ισορροπία κατά Nash που προκύπτει από οπισθοδρομική επαγωγή είναι τέλεια.

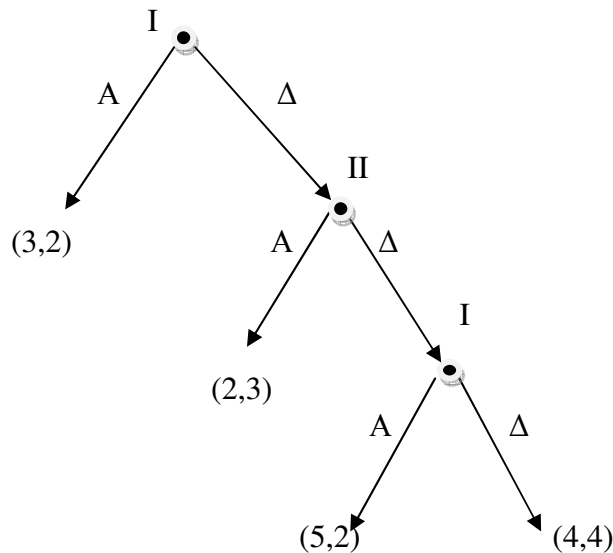
Θεώρημα: Κάθε πεπερασμένο δυναμικό παίγνιο έχει μία τέλεια ισορροπία κατά Nash ανά υποπαίγνιο.

Απόδειξη: Από το θεώρημα Nash γνωρίζουμε πως κάθε πεπερασμένο παίγνιο κανονικής μορφής έχει τουλάχιστον μία ισορροπία κατά Nash. Συνεπώς και κάθε υποπαίγνιο θα έχει μία ισορροπία. Από τον ορισμό που δώσαμε προηγουμένως προκύπτει το ζητούμενο.

5.5 ΙΣΟΡΡΟΠΙΑ Nash – ΒΕΛΤΙΩΣΕΙΣ

Η τελειότητα υποπαιγνίων αποτελεί μία βελτίωση της ισορροπίας κατά Nash. Όλες οι βελτιώσεις έχουν ως στόχο να εμπλουτίσουν με συνθήκες τον ορισμό της ισορροπίας έτσι ώστε να μειώνεται το πλήθος των ισορροπιών και να οδηγούμαστε σε πιο λογικά αποτελέσματα. Σε αυτή την προσέγγιση αντιμετωπίζουμε δύο προβλήματα. Αρχικά η έννοια του «λογικού» διαφέρει ανάλογα την κατάσταση που εξετάζουμε. Έπειτα, το πλήθος των ισορροπιών που ικανοποιούν τις συνθήκες βελτίωσης είναι σπανίως μόνο μία· συχνά βλέπουμε να παραμένουν αρκετές. Επιπλέον, υπάρχουν παιχνίδια που με κάποιες βελτιώσεις (πέραν της τέλειας ισορροπίας ανά υποπαίγνιο, η οποία υπάρχει πάντα) δεν οδηγούν απαραίτητα σε ικανοποιητικό αποτέλεσμα, διότι είναι πιθανό η ισορροπία Nash του παιγνίου να μην ικανοποιεί τις πρόσθετες συνθήκες βελτίωσης. Η τελειότητα ανά υποπαίγνιο προσπαθεί να συλλέξει συγκεκριμένες ισορροπίες που να είναι πιο λογικές καθώς κινούμαστε οπισθοδρομικά στο δέντρο. Μία άλλη προσέγγιση είναι η *μπρος επαγωγή* σύμφωνα με την οποία κινούμαστε κατακόρυφα εμπρός στο δέντρο. Ας δούμε τώρα ένα παράδειγμα:

Δίνεται παρακάτω ένα δέντρο παιγνίου το οποίο παρουσιάζει ένα πρόβλημα με την τελειότητα ανά υποπαίγνιο. Η μοναδική τέλεια ισορροπία κατά Nash ανά υποπαίγνιο οδηγεί τον παίκτη 1 να πρέπει να κινηθεί αριστερά (Α) από το πρώτο βήμα όπου και τερματίζεται το παίγνιο.



Σχήμα 5.5

Παρ' όλα αυτά τι θα συμβεί και ποια στρατηγική πρέπει να ακολουθήσει ο παίκτης II αν ο παίκτης I δεν κινηθεί ορθολογικά και το παίγνιο δεν ολοκληρωθεί στο 1^ο βήμα αλλά του δώσει τη δυνατότητα να λάβει απόφαση;

Σύμφωνα με την οπισθοδρομική επαγωγή η οποία βασίζεται στην υπόθεση ότι ο παίκτης θα συμπεριφερθεί ορθολογικά στο μέλλον, ο παίκτης II θα αναγκαστεί να παίξει A και να τερματίσει εκείνος το παιχνίδι με απόδοση 3 (και 2 για τον παίκτη I). Δεδομένου όμως ότι ο παίκτης I έχει συμπεριφερθεί επιπόλαια ήδη μία φορά, μπορεί να το επαναλάβει. Με αυτό σαν επιχείρημα ο παίκτης II πρέπει να κινηθεί Δ (δεξιά) δίνοντας πάλι την ευκαιρία στον αντίπαλο να κατευθυνθεί Δ ξανά.

Το πρόβλημα εντοπίζεται στην προσπάθειά μας να αναλύσουμε μία μη ορθολογική συμπεριφορά με βάση ότι οι παίκτες μας είναι ορθολογικοί με αποτέλεσμα να μην οδηγούμαστε πουθενά. Ο τρόπος για να λυθεί αυτό το πρόβλημα είναι να υποθέσουμε όντως ότι οι παίκτες μας είναι ορθολογικοί κι ότι διεπράχθη λάθος κατά την εφαρμογή της στρατηγικής τους και έτσι η επιπόλαιη συμπεριφορά είναι μόνο φαινομενικά.

Ένα καλό παράδειγμα είναι στο σκάκι, τη στιγμή που ο παίκτης I πρόκειται να κινήσει ένα πιόνι και έχει μόνο δύο διαθέσιμες κινήσεις, εκ των οποίων η μία είναι καλύτερη από την άλλη. Έχοντας ο παίκτης I διαλέξει την καλύτερη δυνατή που τον οδηγεί σε θέση ματ, αγχώνεται και το πιόνι γλιστρά από τα χέρια του και πέφτει σε λάθος τετράγωνο χωρίς φυσικά τη δυνατότητα αναίρεσης της κίνησης. Έτσι η

ισορροπία που υπολείπεται καλείται «ισορροπία Nash τρεμάμενου χειρός»²⁴. Αν θεωρήσουμε πιθανότητα λάθους ε παίρνουμε οριακά $\varepsilon \rightarrow 0$.

Ο χαρακτηρισμός ως απροσδόκητης συμπεριφοράς υπονοεί ότι το λάθος ενός παίκτη δεν είναι απαραίτητο να επαναληφθεί λόγω επιπόλαιης συμπεριφοράς και δεν υπάρχει λόγος να το συσχετίζουμε με τη συμπεριφορά του στο σύνολο της. Αυτό σημαίνει πως στο προηγούμενο παράδειγμα (σχήμα) ο παίκτης II πρέπει να επιλέξει A όσο η πιθανότητα λάθους του αντιπάλου διατηρείται αμελητέα.

²⁴ Ο όρος προκύπτει από το “trembling hand Nash equilibria”

ΚΕΦΑΛΑΙΟ 6^ο

ΠΑΙΓΝΙΑ ΜΕ ΣΥΝΕΧΗ ΣΥΝΟΛΑ ΣΤΡΑΤΗΓΙΚΩΝ

6.1 ΑΠΕΙΡΑ ΣΥΝΟΛΑ ΣΤΡΑΤΗΓΙΚΩΝ

Μετά από τη μελέτη παιγνίων στα οποία οι παίκτες διαθέτουν σύνολα στρατηγικής διακριτά και πεπερασμένα κρίνεται σημαντικό να εξετασθούν παίγνια στα οποία τα σύνολα στρατηγικής είναι συνεχή κι όχι απαραίτητα πεπερασμένα. Στην πραγματική ζωή άλλωστε συναντώνται κυρίως μοντέλα παιγνίων που περιγράφουν καταστάσεις στις οποίες οι παίκτες πρέπει να επιλέξουν στρατηγική από μη διακριτό σύνολο, αλλά από υποσύνολο των πραγματικών αριθμών. Για το λόγο αυτό θα μελετηθούν παρακάτω παραδείγματα τέτοιων περιπτώσεων όπου φαίνεται καθαρά πως η θεωρία παιγνίων απαντά σε προβλήματα της οικονομικής επιστήμης.

Υποθέτουμε αρχικά πως τα σύνολα αμιγών στρατηγικών αντιστοιχούν σε κλειστά διαστήματα της μορφής $[\alpha, \beta] \subseteq \mathbb{R}$. Τότε μία αμιγής στρατηγική είναι μία επιλογή $x \in [\alpha, \beta]$ και μία μικτή στρατηγική ορίζεται από μία συνάρτηση κατανομής $p(x)$ τέτοια ώστε η πιθανότητα η επιλογή του παίκτη να είναι ανάμεσα στο x και στο $x+dx$ να είναι $p(x)dx$.

6.2 ΜΟΝΤΕΛΟ ΔΥΟΠΩΛΙΟΥ Cournot

Δυοπώλιο είναι μία αγορά στην οποία δύο επιχειρήσεις ανταγωνίζονται να προμηθεύσουν στο ίδιο αγοραστικό κοινό το ίδιο προϊόν. Υπάρχουν τρία κλασικά μοντέλα δυοπωλίου: Cournot, Bertrand και το μοντέλο του Stackelberg. Στο μοντέλο του Bertrand, η ποσότητα προϊόντος που καταναλώνεται καθορίζεται από την τιμή πώλησης και οι δύο εταιρίες πρέπει να αποφασίσουν για την τιμή που θα προσπαθήσουν να πουλήσουν το απόθεμά τους ταυτόχρονα. Έτσι, η εταιρία που κατοχυρώνει τη χαμηλότερη τιμή πώλησης κερδίζει την αγορά, ενώ η αντίπαλη εταιρία δεν πουλά τίποτα. Μία απλή ανάλυση της κατάστασης στην οποία χρησιμοποιούμε τη μέθοδο της απαλοιφής κυριαρχούμενων στρατηγικών δείχνει πως οι εταιρίες θα έπρεπε να ορίσουν μία τιμή ίση με αυτή του κόστους παραγωγής του

προϊόντος. Στα μοντέλα Cournot και Stackelberg οι δύο εταιρίες πρέπει να αποφασίσουν πόσο προϊόν θα παράγουν και η τιμή πώλησης θα καθοριστεί από την ποσότητα αυτή που παρήγαγαν. Στο μοντέλο Cournot οι αποφάσεις λαμβάνονται ταυτόχρονα, ενώ στο μοντέλο του Stackelberg οι αποφάσεις λαμβάνονται σταδιακά, με την απόφαση της πρώτης επιχείρησης να γίνεται ευρέως γνωστή. Πολλές φορές οι οικονομολόγοι ονομάζουν αυτές τις ισορροπίες είτε ως «Cournot», είτε «Bertrand» ή «Stackelberg», αλλά όλες είναι ισορροπίες κατά Nash για κάθε μοντέλο δυοπωλίου που περιγράφουν, με την ισορροπία του Stackelberg να είναι και τέλεια ανά υποπαίγνιο.

Θεωρούμε μία αγορά με δύο αντίπαλες επιχειρήσεις οι οποίες εμπορεύονται ένα προϊόν που πωλείται σε οποιαδήποτε ποσότητα καθώς διαιρείται χωρίς πρόβλημα, όπως για παράδειγμα το πετρέλαιο. Το μοντέλο Cournot βασίζεται στην εξής υπόθεση:

Επιτρέπεται στις εταιρίες να διαλέξουν όση ποσότητα προϊόντος θα παράγουν (ή θα εμπορευθούν) με συνέπεια το σύνολο των ενεργειών τους να προσδιορίζεται από ποσότητες q_i . Δεδομένου ότι το προϊόν μπορεί να διαιρείται απείρως, θεωρούμε το σύνολο των κινήσεων συνεχές.

Αν η εταιρία i παράγει μία ποσότητα q_i του προϊόντος, τότε η συνολική ποσότητα που θα παραχθεί θα είναι $Q = q_1 + q_2$. Η τιμή πώλησης του προϊόντος υποτίθεται πως εξαρτάται από το συνολικό απόθεμα ως εξής:

$$P(Q) = \begin{cases} P_0 \left(1 - \frac{Q}{Q_0}\right) & \text{αν } Q < Q_0 \\ 0 & \text{αν } Q \geq Q_0 \end{cases}$$

Έτσι λοιπόν η τιμή πώλησης πέφτει από τη μέγιστη τιμή P_0 όταν το προϊόν είναι πολύ σπάνιο, στο μηδέν όταν παράγεται η ποσότητα Q_0 . Τα κόστη παραγωγής θα είναι $C(q_i) = cq_i$ όπου δεν υπάρχει σταθερό κόστος παραγωγής (ανεξάρτητο της ποσότητας) και το κόστος παραγωγής ανά μονάδα προϊόντος (c) είναι ίδιο για κάθε επιχείρηση. Οι αποδόσεις για κάθε εταιρία δίνονται από το καθαρό κέρδος στην αγορά και από τη συμπεριφορά των δύο εταιριών. Συνεπώς η απόδοση για την εταιρία i είναι:

$$a_i(q_1, q_2) = q_i P(Q) - cq_i.$$

Παρατηρούμε ότι σίγουρα δεν έχει νόημα οι εταιρίες να παράγουν ποσότητα μεγαλύτερη από Q_0 διότι αυτό θα οδηγούσε σε βέβαιη ζημία. Κατά συνέπεια

περιορίζουμε το σύνολο δράσεων να είναι: $[0, Q_0]$.

Θα αναζητήσουμε τώρα την καλύτερη απόκριση της εταιρίας 1, δεδομένης της ποσότητας q_2 . Και έτσι έχουμε:

$$\frac{\partial a_1}{\partial q_1}(\hat{q}_1, q_2) = 0$$

Άρα:

$$\hat{q}_1 = \frac{Q_0}{2} \left(1 - \frac{q_2}{Q_0} - \frac{c}{P_0} \right).$$

Για να ελέγξουμε εάν είναι όντως η καλύτερη απόκριση (κι όχι η χειρότερη) υπολογίζουμε:

$$\frac{\partial^2 a_1}{\partial q_1^2}(\hat{q}_1, q_2) = -\left(\frac{P_0}{Q_0}\right) < 0.$$

Χρειάζεται να επιβεβαιώσουμε και τη συνθήκη: $\hat{q}_1 + q_2 \leq Q_0$ ώστε να εξασφαλίσουμε μη αρνητικό κέρδος:

$$\begin{aligned} \hat{q}_1 + q_2 &= \frac{Q_0}{2} \left(1 - \frac{q_2}{Q_0} - \frac{c}{P_0} \right) + q_2 \\ &= \frac{Q_0}{2} + \frac{q_2}{2} - \frac{cQ_0}{2P_0} \\ &\leq \frac{Q_0}{2} + \frac{Q_0}{2} - \frac{cQ_0}{2P_0} \\ &= Q_0 \left(1 - \frac{c}{2P_0} \right) < Q_0 \end{aligned}$$

Ομοίως βρίσκουμε την καλύτερη απόκριση στην δράση q_1 να είναι για την εταιρία 2 να παράγει ποσότητα:

$$\hat{q}_2 = \frac{Q_0}{2} \left(1 - \frac{q_1}{Q_0} - \frac{c}{P_0} \right).$$

Μία αμιγής στρατηγική που ισορροπεί κατά Nash είναι το ζεύγος (q_1^*, q_2^*) όπου η κάθε τιμή αποτελεί την καλύτερη απόκριση στην άλλη. Από τις ισότητες:

$$q_1^* = \frac{Q_0}{2} \left(1 - \frac{q_2^*}{Q_0} - \frac{c}{P_0} \right) \quad q_2^* = \frac{Q_0}{2} \left(1 - \frac{q_1^*}{Q_0} - \frac{c}{P_0} \right)$$

Προκύπτει ότι:

$$q_1^* = q_2^* = \frac{Q_0}{3} \left(1 - \frac{c}{P_0} \right) \equiv q_c^*$$

Όπου q_c^* έχουμε ορίσει την τιμή της ποσότητας που έχει επιλεγεί από κάθε εταιρία σε

κατάσταση ισορροπίας. Η απόδοση για κάθε εταιρία σε αυτή την περίπτωση είναι:

$$\begin{aligned} a_1(q_c^*, q_c^*) &= a_2(q_c^*, q_c^*) \\ &= q_c^* P(2q_c^*) - cq_c^* \\ &= \frac{Q_0 P_0}{9} \left(1 - \frac{c}{P_0}\right)^2 \end{aligned}$$

Ας συγκρίνουμε τώρα αυτή την ανταγωνιστική ισορροπία με την κατάσταση που επικρατεί υπό συνθήκες μονοπωλίου.

Η απόδοση στο μονοπώλιο προφανώς είναι: $a_\mu(q) = qP(q) - cq$

Τότε η βέλτιστη στρατηγική για την εταιρία είναι:

$$q_\mu^* = \frac{Q_0}{2} \left(1 - \frac{c}{P_0}\right).$$

Εφόσον $q_\mu^* < 2q_c^*$ η τιμή πώλησης του αγαθού είναι υψηλότερη συγκριτικά με αυτή στο δυοπώλιο. Έτσι το μοντέλο μας δείχνει πως ο ανταγωνισμός λειτουργεί προς όφελος του καταναλωτή.

Ας υποθέσουμε τώρα πως οι δύο ανταγωνιστικές εταιρίες του δυοπωλίου σχηματίσουν μία σύμπραξη²⁵ και συμφωνήσουν να χρησιμοποιήσουν τις εξής στρατηγικές:

$$q_1 = q_2 = \frac{1}{2} q_\mu^* .$$

Δηλαδή συμφώνησαν, να παράγει η καθεμία εταιρία, τη μισή ποσότητα από αυτή στην περίπτωση του μονοπωλίου. Τότε οι απολαβές θα είναι:

$$\begin{aligned} a_i &= \left(\frac{1}{2} q_\mu^*, \frac{1}{2} q_\mu^*\right) = \frac{1}{2} q_\mu^* P(q_\mu^*) - \frac{1}{2} c q_\mu^* \\ &= \frac{Q_0 P_0}{8} \left(1 - \frac{c}{P_0}\right)^2 \end{aligned}$$

Όπως φαίνεται λοιπόν, οι απολαβές είναι καλύτερες για τις δύο εταιρίες σε σύγκριση με αυτές που υπολογίσαμε μέσω του δυοπωλίου Cournot, και η τιμή πώλησης του αγαθού θα είναι η ίδια με αυτή του μονοπωλίου. Παρ' όλα αυτά μία τέτοια συμφωνία είναι ασταθής διότι η καλύτερη απόκριση σε μία εταιρία που παράγει ποσότητα προϊόντος $\frac{1}{2} q_\mu^*$ (αυτή της σύμπραξης), είναι να παράγει:

$$\hat{q} = \frac{Q_0}{2} \left(1 - \frac{q_\mu^*}{2Q_0} - \frac{c}{P_0}\right)$$

²⁵ Τα λεγόμενα «cartel»

<http://www.oecd.org/competition/cartels/>

$$= \frac{3}{4} q_{\mu}^* > \frac{1}{2} q_{\mu}^* .$$

Αξίζει να αναφέρουμε σε αυτό το σημείο πως δεν έχουμε αποδείξει ότι μία σύμπραξη είναι αδύνατη, παρά μόνο ότι δεν μπορεί να πραγματοποιηθεί σε καταστάσεις που περιγράφονται από το μοντέλο του Cournot.

6.3 ΜΟΝΤΕΛΟ ΔΥΟΠΩΛΙΟΥ Stackelberg

Σύμφωνα με το μοντέλο του Stackelberg, οι δύο εταιρίες ανταγωνίζονται με σκοπό να πουλήσουν ένα προϊόν που δεν αποτελεί μία αυτόνομη μονάδα αλλά διαιρείται και μετράται η ποσότητα παραγωγής συνολικά. Έτσι πρέπει να αποφασίσουν πόση ποσότητα προϊόντος (q_1 για την 1^η και q_2 για τη 2^η εταιρία) θα παράγουν. Όπως και στο μοντέλο του Cournot, υποθέτουμε πως η τιμή του αγαθού δίνεται από τη σχέση:

$$P(Q) = P_0 \left(1 - \frac{Q}{Q_0} \right)$$

Όπου $Q = q_1 + q_2$ και c το κόστος παραγωγής ανά μονάδα προϊόντος. Όπως προαναφέραμε, αντίθετα με το μοντέλο του Cournot, στο μοντέλο του Stackelberg οι αποφάσεις δεν λαμβάνονται ταυτόχρονα αλλά διαδοχικά. Η πρώτη εταιρία (η οποία καλείται και «αρχηγός της αγοράς»²⁶) αποφασίζει την ποσότητα που θα παράγει και δεύτερη εταιρία (αλλιώς «ακόλουθος της αγοράς»²⁷) αφού παρατηρήσει την απόφαση της 1^{ης}, επιλέγει κι αυτή την ποσότητα του αγαθού που θα παράγει. Βασική υπόθεση και σε αυτό το μοντέλο είναι η επιθυμία κάθε εταιρίας να μεγιστοποιήσει το κέρδος της και η μέγιστη τιμή πώλησης να είναι μεγαλύτερη από το κόστος παραγωγής ανά μονάδα προϊόντος ($P_0 > c$).²⁸

Λύνουμε το παίγνιο με οπισθοδρομική επαγωγή για να βρούμε μία τέλεια ανά υποπαίγνιο ισορροπία κατά Nash. Ξεκινούμε βρίσκοντας την καλύτερη απόκριση της 2^{ης} εταιρίας $\hat{q}_2(q_1)$ για κάθε πιθανή απόφαση της 1^{ης} εταιρίας. Δεδομένου ότι η 1^η εταιρία γνωρίζει την καλύτερη απόκριση της 2^{ης} σε κάθε επιλογή q_1 αυτής,

²⁶ Στα αγγλικά ο όρος είναι “market leader”.

²⁷ Στα αγγλικά ο όρος είναι “market follower”.

²⁸ Gabszewicz J. J. (1999), Strategic Interaction and Markets, Oxford University Press.

μπορούμε να βρούμε μία ισορροπία Nash γι' αυτό το παίγνιο καθορίζοντας τη μέγιστη απόδοση που μπορεί να πετύχει η 1^η θεωρώντας πως η 2 θα ανταποκριθεί βέλτιστα.

Το κέρδος της 2^{ης} εταιρίας δίνεται από της συνάρτηση απόδοσης:

$$a_2(q_1, q_2) = q_2[P(Q) - c]$$

Αναζητούμε λοιπόν, τη μέγιστη τιμή $\hat{q}_2$ αυτής της συνάρτησης λύνοντας την εξίσωση:

$$\frac{\partial a_2}{\partial q_2}(q_1, \hat{q}_2) = 0$$

η οποία δίνει

$$\hat{q}_2(q_1) = \frac{Q_0}{2} \left(1 - \frac{q_1}{Q_0} - \frac{c}{P_0} \right)$$

Τότε για την επιλογή $\hat{q}_2(q_1)$, το κέρδος της 1^{ης} θα είναι:

$$\begin{aligned} a_1(q_1, \hat{q}_2(q_1)) &= q_1 \left[P_0 \left(1 - \frac{q_1 + \hat{q}_2(q_1)}{Q_0} \right) - c \right] \\ &= \frac{q_1 P_0}{2} \left(1 - \frac{q_1}{Q_0} - \frac{c}{P_0} \right). \end{aligned}$$

Αν $\hat{q}_1$ η ποσότητα που μεγιστοποιεί το κέρδος της πρώτης εταιρίας, τότε αυτό θα είναι ίσο με:

$$\hat{q}_1 = \frac{Q_0}{2} \left(1 - \frac{c}{P_0} \right).$$

Τότε η ισορροπία κατά Nash επιτυγχάνεται για

$$q_1^* = \frac{Q_0}{2} \left(1 - \frac{c}{P_0} \right)$$

Και για

$$q_2^* = \hat{q}_2(q_1),$$

Δηλαδή για

$$q_2^* = \frac{Q_0}{4} \left(1 - \frac{c}{P_0} \right)$$

Ενδιαφέρον παρουσιάζει το γεγονός ότι, παρόλο που η δεύτερη εταιρία έχει περισσότερες πληροφορίες πριν λάβει απόφαση για την ποσότητα που θα παράγει δεδομένου ότι η απόφαση της θα ληφθεί μεταγενέστερα από της πρώτης, τελικά

μεγαλύτερο κέρδος έχει η πρώτη εταιρία (διότι $q_1^* > q_2^*$ ²⁹).

Συχνά, η τέλεια ανά υποπαίγνιο ισορροπία κατά Nash που εξήχθη (q_1^*, q_2^*) καλείται και «ισορροπία Stackelberg». Δεν είναι όμως η μοναδική ισορροπία κατά Nash στο συγκεκριμένο μοντέλο. Μία ακόμη ισορροπία είναι η 1^η εταιρία να παράγει την ποσότητα που αντιστοιχεί στο μοντέλο του Cournot και η 2^η εταιρία να κάνει το ίδιο, χωρίς να την απασχολεί τι θα κάνει η πρώτη. Έτσι, αν $q_1 = q_c^*$, τότε και

$$\hat{q}_2 = \frac{Q_0}{2} \left(1 - \frac{q_c^*}{Q_0} - \frac{c}{P_0} \right) = q_c^*$$

Δηλαδή σε αυτή την περίπτωση η 2^η εταιρία επιλέγει πάντα $q_2 = q_c^*$.

Τότε, από τη συνάρτηση απόδοσης της πρώτης εταιρίας:

$$a_1(q_1, q_2^*) = q_1 \left[P_0 \left(1 - \frac{q_1 + q_2^*}{Q_0} \right) - c \right]$$

Αναζητούμε τη μέγιστη τιμή $\hat{q}_1$, λύνοντας την εξίσωση:

$$\frac{\partial a_1}{\partial q_1}(\hat{q}_1, q_c^*) = 0$$

Κι έχουμε ότι:

$$\hat{q}_1 = \frac{Q_0}{2} \left(1 - \frac{q_c^*}{Q_0} - \frac{c}{P_0} \right) = q_c^*$$

Άρα η καλύτερη απόκριση για κάθε εταιρία ως απάντηση στην άλλη είναι να παράγουν ποσότητα q_c^* , οπότε το ζεύγος (q_c^*, q_c^*) αποτελεί ισορροπία κατά Nash. Όμως, η απάντηση $q_2 = q_c^*$ της 2^{ης} εταιρίας είναι η καλύτερη δυνατή μόνο στην περίπτωση που η πρώτη αποφασίσει $q_1 = q_c^*$. Σε αντίθετη περίπτωση ($q_1 \neq q_c^*$) δεν είναι η καλύτερη απόκριση. Έπεται λοιπόν πως αυτή η ισορροπία δεν είναι τέλεια ανά υποπαίγνιο.

²⁹ Εφόσον το κέρδος a_i είναι ευθέως ανάλογο της ποσότητας q_i

6.4 TO MONTELO «War of Attrition»

Στο μοντέλο αυτό, δύο παίκτες διεκδικούν τον ίδιο πόρο ή αγαθό αξίας v . Για παράδειγμα, δύο ζώα-αρχηγοί παλεύουν για μία περιοχή στην οποία θέλουν να εγκατασταθούν για να αναπαραχθούν ή δύο ανταγωνιστικά super market τα οποία βρίσκονται σε πόλεμο για τις τιμές μπορούν να περιγραφούν ικανοποιητικά και να μελετηθούν από το μοντέλο «War of Attrition». Η στρατηγική για κάθε παίκτη είναι η επιλογή του χρόνου t_i που θα επιμένει ενάντια στον αντίπαλο. Έτσι προκύπτουν οι παρακάτω τρεις υποθέσεις:

- i. Το κόστος της διαμάχης εξαρτάται μόνο από τη διάρκεια του.
- ii. Ο παίκτης που επιμένει το μεγαλύτερο χρονικό διάστημα κερδίζει όλο το αγαθό. Αν όμως οι παίκτες παραιτηθούν ταυτόχρονα, τότε δεν κερδίζει κανείς τίποτα.
- iii. Το κόστος που καταβάλλεται από κάθε παίκτη είναι ανάλογο του μικρότερου χρόνου που επέλεξε να επιμένει. (Προφανώς όταν ένας παίκτης παραιτείται και το παιχνίδι τελειώσει δεν υπάρχουν περαιτέρω κόστη.)

Με αυτές τις υποθέσεις, οι συναρτήσεις απόδοσης για κάθε παίκτη i ($i=1, 2$) είναι:

$$a_1(t_1, t_2) = \begin{cases} v - ct_2 & \text{αν } t_1 > t_2 \\ -ct_1 & \text{αν } t_1 \leq t_2 \end{cases}$$

και

$$a_2(t_1, t_2) = \begin{cases} v - ct_1 & \text{αν } t_2 > t_1 \\ -ct_2 & \text{αν } t_2 \leq t_1 \end{cases}$$

Εφόσον για τον παίκτη 1 έχουμε ότι:

$$a_1(t_1, 0) = v \quad \text{για κάθε } t_1 > 0 \\ \text{και } a_1(0, 0) = 0$$

για τον παίκτη 2:

$$a_2(v/c, t_2) = -ct_2 < 0 \quad \text{για κάθε } t_2 \leq v/c$$

και

$$a_2(v/c, t_2) = 0 \quad \text{για κάθε } t_2 > v/c.$$

Προκύπτει ότι οι αμιγείς στρατηγικές $t_1^* = v/c$ και $t_2^* = 0$ αποτελούν ισορροπία Nash διότι: $a_1(t_1, t_2^*) \leq a_1(t_1^*, t_2^*)$ για κάθε t_1 και $a_2(t_1^*, t_2) \leq a_2(t_1^*, t_2^*)$ για κάθε

t_2 .

Ένα ακόμη ζεύγος αμιγών στρατηγικών που αποτελούν ισορροπία Nash είναι:

$$t_1^* = 0 \quad \text{και} \quad t_2^* = v/c$$

με

$$a_1(0, v/c) = 0 \quad \text{και} \quad a_2(0, v/c) = v.$$

Για να βρούμε ισορροπία με μικτές στρατηγικές, βολεύει να αναζητήσουμε στρατηγικές βασισμένες στα κόστη της διαμάχης, δηλαδή $x \equiv ct_1$ και $y \equiv ct_2$.

Οπότε, οι αντίστοιχες αποδόσεις θα είναι:

$$a_1(x, y) = \begin{cases} v - y & \text{αν } x > y \\ -x & \text{αν } x \leq y \end{cases}$$

$$a_2(x, y) = \begin{cases} v - x & \text{αν } y > x \\ -y & \text{αν } y \leq x \end{cases}$$

Μία μικτή στρατηγική σ_1 ορίζει μία επιλογή κόστους με εύρος από x έως $x+dx$ με πιθανότητα $p(x)dx$ και η σ_2 ορίζει μια συνάρτηση πυκνότητας πιθανότητας $q(y)$. Η αναμενόμενη απόδοση για τον παίκτη 1 αν επιλέξει ένα κόστος x ενάντια στη μικτή στρατηγική σ_2^* είναι:

$$a_1(x, \sigma_2^*) = \int_0^x (v - y)q(y)dy + \int_x^{+\infty} (-x)q(y)dy$$

Όπου ο πρώτος όρος προκύπτει από την πιθανότητα ο παίκτης 2 να επιλέξει κόστος $y \leq x$ και ο δεύτερος όρος από την πιθανότητα ο παίκτης 2 να επιλέξει $y > x$.

Πρέπει επίσης, να έχουμε ότι η απόδοση $a_1(x, \sigma_2^*)$ είναι σταθερή. Οπότε για σταθερό σ_2^* η συνάρτηση $a_1(x, \sigma_2^*)$ είναι ανεξάρτητη του x , με

$$\frac{\partial a_1}{\partial x} = 0$$

Οπότε

$$\frac{\partial a_1}{\partial x} = \frac{d}{dx} \int_0^x (v - y)q(y)dy - \int_x^{+\infty} q(y)dy - x \frac{d}{dx} \int_x^{+\infty} q(y)dy$$

Όπου:

$$\frac{d}{dx} \int_0^x (v - y)q(y)dy = (v - x)q(x)$$

Και

$$\frac{d}{dx} \int_x^{+\infty} q(y)dy = \frac{d}{dx} \left[1 - \int_0^x q(y)dy \right] = -q(x)$$

Έτσι έχουμε ότι:

$$\begin{aligned}
\frac{\partial a_1}{\partial x} &= (v-x)q(x) - \int_x^{+\infty} q(y)dy + xq(x) \\
&= vq(x) - \int_x^{+\infty} q(y)dy \\
&= 0.
\end{aligned}$$

Από την τελευταία σχέση μπορούμε να αναγνωρίσουμε πως η $q(y)$ είναι η συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής με σταθερά $\lambda=1/v$:

$$\begin{aligned}
vq(x) &= \int_x^{+\infty} q(y)dy \\
q(x) &= \frac{1}{v} \int_x^{+\infty} q(y)dy
\end{aligned}$$

και

$$\int_x^{+\infty} q(y)dy = \frac{1}{v} \int_x^{+\infty} e^{-\frac{y}{v}} dy = e^{-\frac{x}{v}}$$

Άρα

$$q(y) = \frac{1}{v} e^{-\frac{y}{v}}$$

Με άλλα λόγια η κατανομή κόστους που επιλέχθηκε για την μικτή κατανομή είναι η εκθετική με μέσο κόστος v . Και για τον αντίπαλο παίκτη, με τον ίδιο τρόπο οδηγούμαστε στην ίδια κατανομή κόστους (συμμετρική ισορροπία):

$$p(x) = \frac{1}{v} e^{-\frac{x}{v}}$$

Τώρα που βρήκαμε την κατανομή από πλευράς κόστους, μπορούμε εύκολα να βρούμε την ισορροπία κατά Nash με τις συνθήκες που ορίζει η κατανομή του χρόνου επιμονής. Από τη σχέση:

$$p(t) = p(x) \frac{dx}{dt}$$

έχουμε:

$$p(t) = \frac{c}{v} e^{-\frac{ct}{v}}$$

Έτσι η κατανομή του χρόνου επιμονής είναι εκθετική με μέσο $\left(\frac{c}{v}\right)^{-1} = \frac{v}{c}$

Όμως, παρόλο που η $p(t)$ είναι η κατανομή του χρόνου επιμονής για κάθε παίκτη, δεν αποτελεί την κατανομή της διάρκειας του παιχνιδιού. Υπολογίζουμε λοιπόν ως εξής:

$$P(\text{διάρκεια} \leq t) = 1 - P(\text{δεν έχει σταματήσει κανείς πριν τη χρον. στιγμή } t)$$

Όπου ,

$P(\text{ο παίκτης } i \text{ δεν παραιτείται πριν τη χρ. στιγμή } t)$

$$= \int_t^{+\infty} p(r) dr$$

$$= e^{-\frac{ct}{v}}$$

Άρα,

$$P(\text{διάρκεια} \leq t) = 1 - \left(e^{-\frac{ct}{v}}\right) \left(e^{-\frac{ct}{v}}\right)$$

$$= 1 - \left(e^{-\frac{ct}{v}}\right)^2$$

$$= 1 - \left(e^{-\frac{2ct}{v}}\right)$$

▪

ΚΕΦΑΛΑΙΟ 7^ο

ΜΗ ΠΕΠΕΡΑΣΜΕΝΑ ΔΥΝΑΜΙΚΑ ΠΑΙΓΝΙΑ

7.1 ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΑ ΠΑΙΓΝΙΑ

Μελετώντας το δίλημμα του φυλακισμένου οδηγηθήκαμε σε μία λύση του παιγνίου που προσφέρει την καλύτερη δυνατή απόδοση αν οι κρατούμενοι παίζουν (λαμβάνοντας απόφαση ταυτόχρονα) χωρίς συνεργασία. Αν όμως μπορούσαν να συνεργαστούν, θα είχαμε άλλη λύση στο παίγνιο που θα συνέφερε περισσότερο τους παίκτες μας. Τίθεται σε αυτό το σημείο το ερώτημα: Απαιτείται μία εξωτερική παρέμβαση για να διατηρηθεί μία συνεργασία ή μπορεί μία συνεργασία να εξελιχθεί από μόνη της ελεύθερα, ανεξάρτητα και λογικά; Στο δυοπώλιο του Cournot, η συνεργασία (cartel) δεν ευσταθούσε. Σε πολλές χώρες, έχουν γίνει ουσιώδεις προσπάθειες να χαρακτηρισθούν τα καρτέλ παράνομα και να ψηφισθούν νόμοι κατά αυτών. Και αυτό συμβαίνει γιατί υπάρχει κίνδυνος στο μοντέλο των καρτέλ να μη διατηρηθούν οι συμφωνίες και επιπλέον, όπως είδαμε δεν συμφέρει τους καταναλωτές και έτσι ψηφίζονται νόμοι προστασίας του καταναλωτή. Πώς θα μπορούσε όμως να γίνει ευσταθές το μοντέλο των καρτέλ;³⁰

Μία ένδειξη για πιθανή λύση στα προβλήματα που εμφανίζονται βασίζεται στην απόκριση πολλών ανθρώπων στην κανονική μορφή του διλήμματος του φυλακισμένου: είναι ο φόβος για τις συνέπειες των αποφάσεων τους στο μέλλον. Μέσα σε μία κοινωνία, τα άτομα αλληλεπιδρούν συχνά και βλέπουμε πολλάκις να αποφασίζουν με γνώμονα την επίδραση των ενεργειών τους στο μέλλον. Φυσικά και στον χώρο των επιχειρήσεων, οι εταιρίες συχνά αναγκάζονται να πάρουν περισσότερες από μία αποφάσεις για την παραγωγή των αγαθών. Ίσως λοιπόν να μην υπήρχαν ζητήματα ευστάθειας του μοντέλου των καρτέλ αν δίνονταν οι κατάλληλες υποσχέσεις για το τι θα συμβεί στο μέλλον ή ακόμα και να επιβάλλονταν ποινές για τη μη τήρηση της συμφωνίας, ελαχιστοποιώντας με αυτό τον τρόπο το άγχος και το φόβο για τα επερχόμενα των συνιστωσών της συμφωνίας.

Δεδομένων των ανωτέρω παρατηρήσεων θα μελετήσουμε καταστάσεις στις οποίες οι παίκτες αλληλεπιδρούν επαναλαμβανόμενα. Οι αποδόσεις του παιγνίου

³⁰ Gintis Helbert, Game Theory Evolving (2008), Princeton University Press

εξαρτώνται από προηγούμενες επιλογές, αφενός διότι οι παίκτες μπορεί να καθορίσουν τις στρατηγικές τους πάνω στο ιστορικό των αλληλεπιδράσεων και αφετέρου διότι οι προηγούμενες επιλογές τους οδήγησαν σε διαφορετική κατάσταση.

Θεωρούμε αρχικά την περίπτωση του παιγνίου που απαιτείται μία μόνο απόφαση³¹ και προκύπτει μία μόνο κατάσταση ως αποτέλεσμα. Στη συνέχεια, υποθέτουμε πως οι παίκτες βρίσκονται ξανά στην αρχική θέση, δηλαδή το παίγνιο επαναλαμβάνεται. Σε αυτής της μορφής το παίγνιο (μίας απόφασης) η πιο λογική επιλογή είναι να ακολουθήσουν οι παίκτες τη λύση που αντιστοιχεί σε ισορροπία κατά Nash κάθε φορά που αυτό παίζεται. Παρόλα αυτά, αν δούμε ως ένα παίγνιο όλα τα παίγνια μίας απόφασης που επαναλήφθηκαν, τότε το σύνολο των στρατηγικών που διαθέτει κάθε παίκτης εμπλουτίζεται. Συνεπώς οι παίκτες μπορούν να καθορίσουν τη συμπεριφορά τους ανάλογα με τις περασμένες κινήσεις των αντιπάλων τους ή να προκαλέσουν «απειλή» για τις μελλοντικές τους πράξεις στην περίπτωση που δεν είναι ικανοποιητική η πορεία του παιχνιδιού μέχρι εκείνη τη στιγμή.

Για τη μελέτη μας θα περιοριστούμε σε διακριτό και πεπερασμένο σύνολο στρατηγικών.

Στο σημείο αυτό θα μοντελοποιήσουμε την απλή μορφή του παιγνίου του διλήμματος του φυλακισμένου κι έπειτα θα το μελετήσουμε ως επαναλαμβανόμενο παίγνιο.

Θεωρούμε το σύνολο των στρατηγικών $S_i = \{\Sigma, A\}$ για κάθε παίκτη $i=1, 2$ όπου Σ δηλώνει συνεργασία και A δηλώνει αντίσταση. Τότε ο πίνακας αποδόσεων είναι:

Στρατηγική		Κρατούμενος B	
		Σ	A
Κρατούμενος A	Σ	(x, x)	(z, w)
	A	(w, z)	(y, y)

³¹ Αναφερόμαστε στα «stage games»

<http://www.stanford.edu/~jdlevin/Econ%20203/RepeatedGames.pdf> (15/3/2014)

Πίνακας 7.1

Όπου x, y, w, z θετικοί ακέραιοι με $w > x > y > z = 0$. Και λύνοντας το παίγνιο, προκύπτει πως έχουμε ισορροπία στην περίπτωση όπου και οι δύο παίκτες επιλέξουν Α με αποδόσεις (y, y) .

Στη συνέχεια, ας δούμε το επαναλαμβανόμενο δίλλημα του φυλακισμένου στη γενική μορφή. Αρχικά υποθέτουμε πως το πρόβλημα επαναλαμβάνεται σε δύο στάδια. Θα λυθεί ως δυναμικό παίγνιο με την οπισθοδρομική επαγωγή. Στο τελικό στάδιο, δεδομένου ότι δεν θα υπάρχει μελλοντική διάδραση, η τελική απόφαση αντιστοιχεί στην απόδοση του μονοπατιού που ακολουθήθηκε. Όπως είδαμε στην απλή μορφή του παιγνίου, η καλύτερη δυνατή απόκριση σε κάθε παίκτη είναι να παίξει Α ανεξάρτητα από την κίνηση του αντιπάλου. Έτσι προκύπτει ισορροπία Nash (A, A) . Στο παιχνίδι δύο σταδίων, η λύση (A, A) είναι τέλεια ισορροπία στο υποπαίγνιο.

Τώρα ας μελετήσουμε το πρώτο στάδιο. Εφόσον οι στρατηγικές έχουν δημιουργηθεί για το τελικό στάδιο που οδηγεί στις τελικές αποδόσεις, οι αποδόσεις του πρώτου υποπαιγνίου μπορούν να υπολογισθούν προσθέτοντας τις αποδόσεις του σημείου ισορροπίας του τελικού σταδίου στο πρώτο στάδιο, δημιουργώντας έτσι ένα νέο πίνακα αποδόσεων συνολικά. Το σύνολο των αμιγών στρατηγικών κάθε παίκτη είναι το $S = \{\Sigma, \Sigma A, A \Sigma, AA\}$ αλλά επειδή μας ενδιαφέρει μία ισορροπία τέλεια ανά υποπαίγνιο χρειαζόμαστε ένα μόνο υποσύνολο του πίνακα αποδόσεων. Έτσι έχουμε:

Στρατηγική	Κρατούμενος Β	
	ΣΑ	ΑΑ
Κρατούμενος Α	ΣΑ	$(x+y, x+y)$ $(z+y, w+y)$
	ΑΑ	$(w+y, z+y)$ $(2y, 2y)$

Πίνακας 7.2

Η ισορροπία Nash του παιγνίου είναι (AA, AA) . Έτσι η τέλεια ανά υποπαίγνιο

ισορροπία είναι κάθε παίκτης να παίζει A σε κάθε στάδιο. Ας σημειωθεί βέβαια πως δεν μπορεί να υπάρξει συνεργασία μεταξύ των παικτών, ούτε ενδιάμεσα από τα στάδια. Ας το δούμε επαγωγικά, για παράδειγμα έστω ότι συνεννοούνται οι παίκτες να συνεργαστούν στο δεύτερο στάδιο, αν ο ένας τηρούσε τη συμφωνία, τότε τον παίκτη B τον συμφέρει να εξαπατήσει και να παίζει A διότι $x > z \Rightarrow x + y > z + y$. Ή ακόμα κι αν η «συνεργασία» μεταφραστεί ως απειλή του ενός να παίζει A και να μην συνεργαστούν τότε πάλι οδηγούμαστε στη ισορροπία A,A, δηλαδή δεν τίθεται θέμα απειλής. Αυτό λοιπόν το επιχείρημα για μη σύμπραξη στο παίγνιο δύο σταδίων στέκει και για περισσότερες από δύο επαναλήψεις του παιγνίου. Καταλήγουμε στο συμπέρασμα πως και οι δύο παίκτες θα επιλέξουν A σε κάθε στάδιο, όσα κι αν είναι αυτά. Με άλλα λόγια η αμφίπλευρη συνεργασία (αλλιώς cartel) δεν ευσταθεί στο πεπερασμένα επαναλαμβανόμενο παίγνιο του διλήμματος του φυλακισμένου.

Ας θεωρήσουμε τώρα πως έχουμε άπειρο αριθμό επαναλήψεων, συμβολίζοντας τις επαναλήψεις με $t = 0, 1, 2, \dots$. Αν δεν υπάρχει (τουλάχιστον γνωστό) πέρας του παιγνίου, τότε δεν υπάρχει και τελικό στάδιο για να εργασθούμε με την οπισθοδρομική επαγωγή. Έτσι, σε κάθε στάδιο θα υπάρχει πάντα και ένα επόμενο. Θα πρέπει τότε να αναζητήσουμε μία στάσιμη στρατηγική εφόσον όλα τα υποπαίγνια είναι ίδια μεταξύ τους.

Ορισμός: Στάσιμη στρατηγική καλείται εκείνη η στρατηγική που ακολουθείται όταν κάθε υποπαίγνιο διέπεται από τους ίδιους κανόνες για την επιλογή της επόμενης κίνησης. Ας σημειωθεί ότι δεν είναι απαραίτητο οι επιλεγμένες κινήσεις να ταυτίζονται σε κάθε στάδιο.

Παράδειγμα: Οι στρατηγικές «επιλέγω Σ σε κάθε στάδιο», «επιλέγω Α σε κάθε στάδιο» είναι προφανώς στάσιμες στρατηγικές στο επαναλαμβανόμενο παίγνιο του διλήμματος του φυλακισμένου. Η υπό περιορισμούς στρατηγική «Επιλέγω Σ αν ο άλλος παίκτης δεν έχει επιλέξει ποτέ Α, αλλιώς επιλέγω Α» είναι επίσης στάσιμη.

Η απόδοση μίας στάσιμης στρατηγικής είναι το άπειρο άθροισμα των αποδόσεων που επιτυγχάνονται σε κάθε στάδιο. Υποθέτουμε πως ένας παίκτης i λαμβάνει απόδοση $r_i(t)$ για κάθε στάδιο t . Τότε η συνολική απόδοση είναι:

$$\sum_{r=0}^{\infty} r_i(t) .$$

Δυστυχώς η προσέγγιση με ευθεία επαγωγή οδηγεί σε πρόβλημα. Θεωρούμε τη στρατηγική s_{Σ} = «επιλέγω Σ σε κάθε στάδιο». Αν και οι δύο παίκτες ακολουθήσουν αυτή τη στρατηγική με $r_i(t) = x$, x θετικός ακέραιος, άρα $x \geq 1$, τότε η συνολική απόδοση για κάθε παίκτη θα είναι:

$$a_i(s_{\Sigma}, s_{\Sigma}) = \sum_{t=0}^{\infty} x = \infty$$

Ενώ αν επιλεγεί από τον ένα παίκτη η στρατηγική «Α σε κάθε στάδιο», τότε η συνολική απόδοση θα είναι ίδια για κάθε παίκτη και ίση με:

$$a_1(s_{\Sigma}, s_A) = a_2(s_A, s_{\Sigma}) = \sum_{t=0}^{\infty} w = \infty$$

Εφόσον w ακέραιος με $w > x > y > z = 0$, δηλαδή $w > 1$.

Καταλήγουμε στο συμπέρασμα πως δεν μπορούμε να εξάγουμε λύση, αν δηλαδή η στρατηγική s_{Σ} είναι καλύτερη από την s_A εφόσον οι αποδόσεις και των δύο τείνουν στο άπειρο.

Αναγκάζομαστε να εισάγουμε τον παράγοντα δ^{32} , με $\delta \in (0,1)$ με σκοπό να επιτευχθούν αποδόσεις που να αντιστοιχούν σε πραγματικό αριθμό και να μην απειρίζονται. Έτσι η συνολική απόδοση για τον παίκτη i θα είναι:

$$a_i = \sum_{t=0}^{\infty} \delta^t r_i(t)$$

Ανάλογα με το μοντέλο με το οποίο εργαζόμαστε αυτός ο παράγοντας θα μπορούσε να δηλώνει για παράδειγμα τον πληθωρισμό, ή κάποιον άλλο -ποσοστιαίο πιθανότητα- δείκτη που να ανήκει προφανώς στο διάστημα $(0, 1)$.

Συνεπώς, με την εισαγωγή του νέου αυτού παράγοντα, η απόδοση του παίκτη i αν επιλεγεί και από τους δύο η στρατηγική s_{Σ} = «επιλέγω Σ σε κάθε στάδιο» θα είναι:

$$\begin{aligned} a_i(s_{\Sigma}, s_{\Sigma}) &= \sum_{t=0}^{\infty} x \delta^t \\ &= x \sum_{t=0}^{\infty} \delta^t = x \cdot \frac{1}{1-\delta} = \frac{x}{1-\delta} \end{aligned}$$

Ενώ για την περίπτωση που ο ένας δεν ακολουθήσει αυτή τη στρατηγική αλλά

³² Discount factor

αποφασίσει μονομερώς να επιλέγει Α σε κάθε στάδιο (s_A) , θα έχουμε:

$$a_1(s_\Sigma, s_A) = a_2(s_A, s_\Sigma) \\ = \sum_{t=0}^{\infty} w \delta^t = \frac{w}{1-\delta}$$

Άρα οι αποδόσεις πια είναι πεπερασμένος αριθμός κι έτσι μπορούμε να τις συγκρίνουμε και να επιλέξουμε τη βέλτιστη στρατηγική.

Δεδομένου ότι μπορούμε να συγκρίνουμε αποδόσεις που επιτυγχάνονται από διαφορετικές στρατηγικές, μπορεί να ευσταθεί μία μόνιμη σύμπραξη (cartel); Ας δούμε το παρακάτω παράδειγμα.

Παράδειγμα: Θεωρούμε τη στρατηγική $s_k =$ «επιλέγω Σ (συνεργασία) όσο διατηρείται η συμφωνία και από τα δύο μέρη. Αν αλλάξει το άλλο μέρος, επιλέγω από εκείνο το σημείο κι έπειτα σταθερά Α (αντίσταση)». Αν και οι δύο παίκτες υιοθετήσουν αυτή τη στρατηγική τότε θα παρατηρούνταν μόνιμη συνεργασία και κάθε παίκτης θα μπορούσε να πετύχει απόδοση:

$$a_i(s_k, s_k) = \sum_{t=0}^{\infty} x \delta^t = \frac{x}{1-\delta}$$

Όπου x θετικός ακέραιος μεγαλύτερος του μηδενός.

Αποτελεί όμως η (s_k, s_k) ισορροπία κατά Nash; Χάριν ευκολίας υποθέτουμε πως και οι δύο παίκτες είναι περιορισμένοι να διαλέξουν ανάμεσα στο σύνολο S των αμιγών στρατηγικών: $S = \{s_k, s_\Sigma, s_A\}$. Υποθέτουμε επίσης πως ο παίκτης 1 επιλέγει πάντα να συνεργάζεται (Σ) και ο άλλος ακολουθεί τη στρατηγική s_k που προαναφέραμε. Τότε θα έχουμε μόνιμη συνεργασία με αποδόσεις:

$$a_1(s_\Sigma, s_k) = a_2(s_\Sigma, s_k) \\ \sum_{t=0}^{\infty} x \delta^t = \frac{x}{1-\delta}$$

Το ίδιο αποτέλεσμα προκύπτει κι αν ο παίκτης 2 αποφασίσει να αλλάξει γνώμη, διότι η καλύτερη δυνατή είναι η s_Σ . Αν όμως ο παίκτης 1 αποφασίζει να ακολουθήσει τη στρατηγική s_A , δηλαδή να αντιτίθεται και να μη συνεργάζεται, και ο παίκτης 2 να διατηρεί τη στρατηγική s_k , τότε θα έχουμε την παρακάτω ακολουθία κινήσεων των παικτών:

$t=$	0	1	2	3	4	5	...
Παίκτης 1 (s_A)	A	A	A	A	A	A	...
Παίκτης 2 (s_k)	Σ	A	A	A	A	A	...

Τότε η απόδοση για τον παίκτη 1 είναι:

$$\begin{aligned}
 a_1(s_A, s_k) &= w + \sum_{t=1}^{\infty} \delta^t y \\
 &= w + y(\delta + \delta^2 + \delta^3 + \delta^4 + \dots) \\
 &= w + y \cdot \frac{\delta}{1 - \delta}
 \end{aligned}$$

Άρα η στρατηγική s_k είναι καλύτερη έναντι της s_A αν:

$$\begin{aligned}
 a_1(s_{\Sigma}, s_k) &\geq a_1(s_A, s_k) \\
 \Rightarrow \frac{x}{1 - \delta} &\geq w + y \cdot \frac{\delta}{1 - \delta} \\
 \Rightarrow x - w &\geq \delta(y - w)
 \end{aligned}$$

Όμως έχουμε ορίσει: $w > x > y > 0$, δηλαδή $x - w < 0$ και $y - w < 0$, οπότε:

$$\delta \geq \frac{x - w}{y - w}$$

Άρα το ζεύγος στρατηγικών (s_k, s_k) αποτελεί ισορροπία Nash αν ο δείκτης δ τείνει να πλησιάσει τη μονάδα.

7.2 ΤΕΛΕΙΟΤΗΤΑ ΑΝΑ ΥΠΟΠΑΙΓΝΙΟ

Στην προηγούμενη παράγραφο είδαμε ότι το ζεύγος (υπό όρους) στρατηγικών (s_k, s_k) αποτελεί ισορροπία Nash. Όμως δεν είναι μία ισορροπία τέλεια ανά υποπαίγνιο. Ας δούμε τώρα γιατί συμβαίνει αυτό. Σε κάθε σημείο του παιγνίου το μέλλον του υποπαιγνίου είναι τυπικά ισοδύναμο με το σύνολο του παιγνίου. Τα πιθανά υποπαίγνια διαχωρίζονται σε τέσσερις ομάδες:

- i. κανένας παίκτης δεν έχει παίξει A (αντίσταση/μη συνεργασία)
- ii. και οι δύο παίκτες έχουν παίξει A
- iii. ο παίκτης 1 έπαιξε A στο πιο πρόσφατο στάδιο αλλά ο παίκτης 2 έπαιξε Σ
- iv. ο παίκτης 2 έπαιξε A στο πιο πρόσφατο στάδιο αλλά ο παίκτης 1 έπαιξε Σ

Τι δηλώνει λοιπόν το σημείο ισορροπίας (s_k, s_k) σε κάθε ομάδα υποπαιγνίων;

Στις ομάδες i και ii το σημείο ισορροπίας δεν έρχεται σε αντίθεση με την ιδέα της κάθε ομάδας. Στην ομάδα i ο αντίπαλος κάθε παίκτη δεν έχει παίξει A, άρα η στρατηγική s_k καθορίζει τη συνέχεια τη συνεργασίας μέχρις ότου ο άλλος παίκτης να έρθει σε αντίθεση (να παίξει A). Αυτό λοιπόν είναι το ζεύγος στρατηγικών (s_k, s_k) που προσδιορίζεται από την ομάδα i υποπαιγνίων το οποίο αποτελεί σημείο ισορροπίας κατά Nash στο υποπαίγνιο εφόσον είναι ισορροπία ολόκληρου του παιγνίου. Στην ομάδα ii ο αντίπαλος κάθε παίκτη (δηλαδή και οι δύο) έχουν παίξει A, άρα η ισορροπία (s_k, s_k) εδώ δηλώνει πως οι παίκτες θα παίζουν σταθερά A για πάντα. Έτσι η στρατηγική που υιοθετείται σε αυτή την ομάδα υποπαιγνίων είναι η (s_A, s_A) η οποία αποτελεί σημείο ισορροπίας αφού είναι και αυτή ισορροπία για όλο το παίγνιο.

Στην ομάδα iii, η στρατηγική s_k δηλώνει πως ο παίκτης 1 θα έπρεπε να αλλάξει και να παίξει A για πάντα, εφόσον ο αντίπαλος –μόλις– έπαιξε A. Παρόλα αυτά ο παίκτης 1 δεν έπαιξε A, με αποτέλεσμα ο παίκτης 2 να συνεχίζει τη στρατηγική s_k (το οποίο πράγματι οδηγεί στην επιλογή A στον επόμενο γύρο του παιχνιδιού). Έτσι, η ισορροπία Nash για όλο το παιχνίδι ορίζει το ζεύγος στρατηγικών (s_A, s_k) που θα πρέπει να υιοθετηθεί στην τρίτη ομάδα υποπαιγνίων. Δεν αποτελεί όμως ισορροπία Nash στο υποπαίγνιο, γιατί ο παίκτης 2 θα μπορούσε να εξασφαλίσει καλύτερη απόδοση μέσω της στρατηγικής s_A έναντι της s_k . Παρόμοιο είναι το επιχείρημα για την τέταρτη ομάδα υποπαιγνίων. Ως εκ τούτου η ισορροπία Nash για ολόκληρο το παίγνιο δεν αποτελεί ισορροπία σε κάθε υποπαίγνιο, άρα δεν είναι μία τέλεια ισορροπία.

Όμως, παρόλο που η (s_k, s_k) δεν αποτελεί τέλεια ανά υποπαίγνιο ισορροπία, μία παρόμοια στρατηγική τελικά μπορεί να οδηγήσει σε τέλεια ισορροπία, αν υιοθετηθεί κι από τους δύο παίκτες. Θεωρούμε λοιπόν τη στρατηγική $s_K = \ll \text{Ξεκινώ επιλέγοντας συνεργασία (}\Sigma\text{) και διατηρώ την επιλογή αυτή μέχρις ότου, είτε ο ένας παίκτης είτε ο άλλος επιλέξει A όπου τότε θα παίζω A για πάντα} \gg$. Τότε το ζεύγος (s_K, s_K) θα αποτελεί τέλεια ανά υποπαίγνιο ισορροπία διότι καθορίζει τη βέλτιστη στρατηγική στις ομάδες iii και iv υποπαιγνίων, η οποία είναι (s_A, s_A) .

Μέχρι τώρα τα επαναλαμβανόμενα παίγνια που είδαμε είχαν ένα πεπερασμένο σύνολο στρατηγικών. Τι γίνεται όμως αν έχουμε ένα πιο γενικό σύνολο με στρατηγικές; Μπορούμε να γνωρίζουμε αν κάποιο άλλο ζεύγος εκτός της ισορροπίας που έχει βρεθεί μπορεί να δίνει καλύτερη λύση; Για παράδειγμα στο

επαναλαμβανόμενο παίγνιο του διλήματος του φυλακισμένου, αν επιτραπεί η εισαγωγή κι άλλων στρατηγικών στο σύνολο, τότε θα συνεχίζει να είναι η καλύτερη δυνατή λύση η στρατηγική s_k που μας δίνει η ισορροπία Nash; Μία απάντηση στα ερωτήματα αυτά έρχεται να δώσει ο ακόλουθος ορισμός της αρχής της αποκλίσεως σε ένα στάδιο για τα επαναλαμβανόμενα παίγνια.

Ορισμός: Ένα ζεύγος στρατηγικών (σ_1, σ_2) ικανοποιεί τη συνθήκη της αποκλίσεως σε ένα στάδιο αν κανένας από τους παίκτες δεν μπορεί να αυξήσει την απόδοση του αποκλίνοντας (μονομερώς) από τη στρατηγική του σε οποιοδήποτε στάδιο και επιστρέφοντας έπειτα σε αυτήν.

Παράδειγμα: Θεωρούμε το επαναλαμβανόμενο παίγνιο του διλήματος του φυλακισμένου και την τέλεια ανά υποπαίγνιο ισορροπία s_K , όπου $s_K = \langle \text{Ξεκινώ επιλέγοντας συνεργασία (Σ) και διατηρώ την επιλογή αυτή μέχρις ότου, είτε ο ένας παίκτης είτε ο άλλος επιλέξει Α όπου τότε θα παίζω Α για πάντα} \rangle$. Ικανοποιεί αυτή η στρατηγική τη συνθήκη αποκλίσεως ενός σταδίου;

Σε οποιοδήποτε δοσμένο στάδιο το παιχνίδι θα είναι σε μία από τις εξής δύο ομάδες: είτε και οι δύο παίκτες συνεργάζονται πάντα ή τουλάχιστον ένας παίκτης έχει επιλέξει να μην συνεργαστεί (Α) σε προηγούμενο γύρο. Αν και οι δύο παίκτες συνεργάζονται πάντα τότε η s_K δηλώνει συνεργασία σε αυτό το στάδιο. Αν από την άλλη κάποιος παίκτης αλλάξει συμπεριφορά επιλέγοντας (Α) σε εκείνο το στάδιο, τότε η s_K δηλώνει την επιλογή Α ως μόνιμη επιλογή από εκείνο το σημείο. Η εκτιμώμενη μελλοντική απόδοση για αυτή την αλλαγή, θα είναι:

$$w + y \frac{\delta}{1 - \delta}$$

Η οποία είναι μικρότερη από την τιμή της απόδοσης σε διαρκή συνεργασία αν $\delta > \frac{x-w}{y-w}$ (και φυσικά αυτή είναι η αναγκαία συνθήκη για να είναι η s_K ισορροπία κατά Nash). Αν τώρα κάθε παίκτης έχει αρνηθεί συνεργασία στο παρελθόν (έχει δηλαδή επιλέξει Α), τότε η s_K δηλώνει άρνηση συνεργασίας και σε αυτό το επίπεδο. Ενώ αν κάθε παίκτης αλλάξει επιλέγοντας συνεργασία (Σ) σε αυτό το επίπεδο τότε η s_K δηλώνει και πάλι την επιλογή άρνησης συνεργασίας (Α) από εδώ και στο εξής. Η εκτιμώμενη απόδοση για τον παίκτη αν κάνει την αλλαγή είναι:

$$z + y \frac{\delta}{1 - \delta}$$

η οποία είναι μικρότερη από την απόδοση που αντιστοιχεί στη συμπεριφορά που περιγράφει η s_K με την προϋπόθεση $\delta < 1$. Έτσι το ζεύγος s_K ικανοποιεί τη συνθήκη απόκλισης ενός σταδίου υπό τον όρο $\frac{x-w}{y-w} < \delta < 1$ (όπου $0 \leq z < y < x < w$).

Θεώρημα: Ένα ζεύγος στρατηγικών αποτελεί τέλεια ανά υποπαίγνιο ισορροπία Nash για ένα προκαθορισμένο, επαναλαμβανόμενο παίγνιο αν και μόνο αν αυτό ικανοποιεί τη συνθήκη απόκλισης ενός σταδίου.

Απόδειξη: Για παίγνια με πεπερασμένες το πλήθος επαναλήψεις, προκύπτει άμεσα από την μέθοδο της οπισθοδρομικής επαγωγής πως αν έχουμε ισορροπία τέλεια ανά υποπαίγνιο τότε θα ικανοποιείται και η συνθήκη απόκλισης ενός σταδίου. Κι αυτό διότι η απόδοση μίας τέτοιας ισορροπίας δεν επιδέχεται βελτίωση. Εφόσον η τελειότητα ανά υποπαίγνιο εξ' ορισμού συνεπάγεται τη συνθήκη απόκλισης του ενός σταδίου για παίγνια με πεπερασμένες αλλά και άπειρες το πλήθος επαναλήψεις, έτσι απομένει μόνο να αποδειχθεί η αντίστροφη κατεύθυνση.

Έστω λοιπόν ότι δεν ισχύει η υπόθεση του θεωρήματος, δηλαδή αν ένα ζεύγος στρατηγικών (σ_1, σ_2) που ικανοποιεί την συνθήκη απόκλισης σε ένα στάδιο δεν αποτελεί ισορροπία τέλεια ανά υποπαίγνιο. Αυτό συνεπάγεται πως υπάρχει κάποιο στάδιο t στο οποίο θα ήταν προτιμότερο για κάποιον παίκτη –έστω ότι αναφερόμαστε στον παίκτη 1 χωρίς βλάβη της γενικότητας– να υιοθετήσει μια διαφορετική στρατηγική $\hat{\sigma}_1$. Τότε θα υπάρχει ένα $\varepsilon > 0$ τέτοιο ώστε:

$$a_1(\hat{\sigma}_1, \sigma_2) - a_1(\sigma_1, \sigma_2) > 2\varepsilon. \quad (I)$$

Θεωρούμε τώρα μια στρατηγική σ_1' που να ταυτίζεται με την $\hat{\sigma}_1$ από το στάδιο t έως το στάδιο T και με την σ_1 από το στάδιο T και έπειτα.

Εφόσον θα μελετήσουμε τις αποδόσεις στο σύνολο των σταδίων (δηλαδή για $t=0$ έως ∞), εισάγεται ο παράγοντας δ με $0 < \delta < 1$. Οπότε η $|a_1(\hat{\sigma}_1, \sigma_2) - a_1(\sigma_1', \sigma_2)|$ είναι ανάλογη της ποσότητας δ^{T-t} .

Αν $T \rightarrow \infty$ τότε $\delta^{T-t} \rightarrow 0$, εφόσον $\delta \in (0,1)$, άρα υπάρχει ένα $\varepsilon > 0$, πολύ μικρό τέτοιο ώστε:

$$|a_1(\hat{\sigma}_1, \sigma_2) - a_1(\sigma_1', \sigma_2)| < \varepsilon \quad (II)$$

Συνδυάζοντας τις σχέσεις (I) και (II) προκύπτει ότι:

$$a_1(\sigma_1', \sigma_2) - a_1(\sigma_1, \sigma_2) < \varepsilon.$$

Όμως, σ_1 και σ_1' διαφέρουν μόνο σε ένα πεπερασμένο πλήθος σταδίων, οπότε η ανίσωση έρχεται σε αντίθεση με την αρχική υπόθεση πως ικανοποιείται η συνθήκη απόκλισης σε ένα στάδιο, δηλαδή οδηγηθήκαμε σε άτοπο. Έπεται λοιπόν πως ένα ζεύγος στρατηγικών δεν μπορεί να ικανοποιεί τη συνθήκη αποκλίσεως σε ένα στάδιο χωρίς να είναι τέλεια ανά υπό παίγνιο ισορροπία κατά Nash.

■

7.3 FOLK THEOREM

Στην παράγραφο αυτή αφού εισάγουμε τις απαραίτητες έννοιες, θα αποδείξουμε το *Folk Theorem*, το οποίο πήρε το όνομα αυτό³³ διότι έγινε ευρέως γνωστό πριν δημοσιευτεί η επίσημη απόδειξη του. Από τα βασικά αποτελέσματα αυτού, υπήρξαν αρκετές παραλλαγές του θεωρήματος, όπου η κάθε μία αποδείκνυε ένα ελάχιστα διαφορετικό αποτέλεσμα βασισμένο σε ελαφρώς διαφορετικές υποθέσεις. Παρόλα αυτά, η γενική ιδέα παρέμενε πάντα ίδια. Δεδομένου ότι η ισορροπία Nash σε ένα στατικό παίγνιο είναι κοινωνικά κάτω από το βέλτιστο, οι παίκτες μπορούν πάντα να παίξουν και καλύτερα αν το παίγνιο επαναλαμβάνεται και ο παράγοντας δ είναι αρκετά υψηλός και φυσικά μικρότερος της μονάδας. Παρά την ονομασία του θεωρήματος πρέπει να σημειωθεί πως είναι καλώς αποδεδειγμένο.

Ορισμός: Εφικτά ζεύγη αποδόσεων είναι τα ζεύγη αποδόσεων (a_i, a_j) που μπορούν να παραχθούν από τις διαθέσιμες στρατηγικές στους παίκτες.

Ορισμός: Έστω ότι έχουμε ένα επαναλαμβανόμενο παίγνιο με συντελεστή δ . Αν ο δ αντιστοιχεί στην πιθανότητα το παιχνίδι να συνεχίσει, τότε το αναμενόμενο πλήθος των σταδίων που εξελίσσεται το παιχνίδι είναι:

$$T = \frac{1}{1 - \delta}$$

³³ Folk:Παραδοσιακός, λαϊκός

Rasmusen Eric (2006), Games and Information: An Introduction to Game Theory

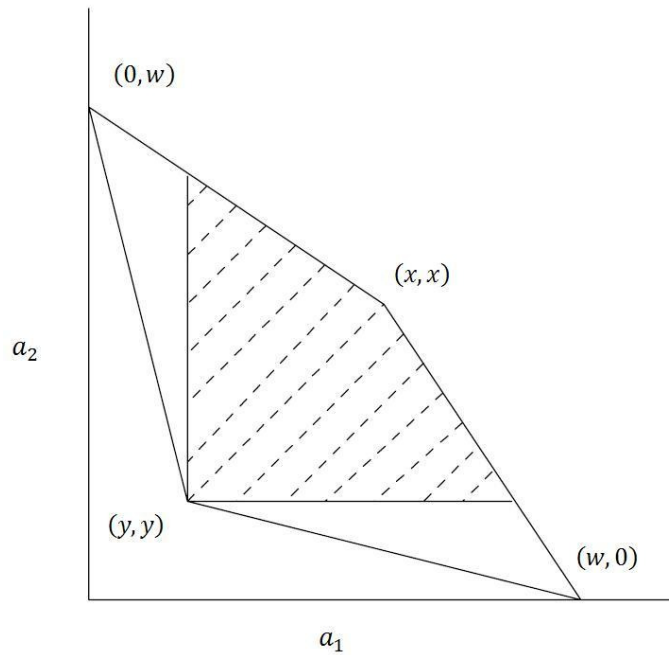
Υποθέτουμε ότι οι δύο παίκτες μας ακολουθούν τις στρατηγικές σ_1 και σ_2 , οι οποίες μπορεί και να μην αποτελούν ισορροπία, τότε η συνολική εκτιμώμενη απόδοση για τον παίκτη i είναι $a_i(\sigma_1, \sigma_2)$ και οι μέσες αποδόσεις ανά στάδιο δίνονται από:

$$\frac{1}{T} a_i(\sigma_1, \sigma_2) = (1 - \delta) a_i(\sigma_1, \sigma_2).$$

Αν το παίγνιο επαναλαμβάνεται τότε το εύρος των εφικτών αποδόσεων σε ζεύγη σε στατικό παίγνιο και το εύρος των εφικτών μέσων αποδόσεων σε ζεύγη είναι ίσα.

Ορισμός: Ζεύγος ορθολογικών αποδόσεων είναι οι μέσες αποδόσεις που υπερβαίνουν την απόδοση των στρατηγικών (των δύο παικτών) που αντιστοιχεί στην ισορροπία κατά Nash σε παίγνιο ενός σταδίου.

Παράδειγμα: Στο στατικό παίγνιο του γενικευμένου διλήμματος του φυλακισμένου, τα ζεύγη των αποδόσεων (a_1, a_2) είναι (x, x) , $(0, w)$, $(w, 0)$ και (y, y) (με $w > x > y > 0$) και είναι προφανώς εφικτά από τη στιγμή που παράγονται από συνδυασμούς αμιγών στρατηγικών. Όμως, παρόλο που κάθε παίκτης θα μπορούσε να έχει απόδοση 0, που είναι η ιδανική για αυτόν, το ζεύγος αποδόσεων $(0, 0)$ δεν είναι εφικτό εφόσον δεν υπάρχει ζεύγος στρατηγικών που να παράγει αυτές τις αποδόσεις και για τους δύο παίκτες. Αν ο παίκτης 1 και ο παίκτης 2 χρησιμοποιήσουν τη στρατηγική (Σ) με πιθανότητες p και $q = 1 - p$ με αποδόσεις a_1 και a_2 αντίστοιχα, τότε τα εφικτά ζεύγη αποδόσεων μπορούν να υπολογισθούν, δίνοντας στα p και q όλες τις δυνατές τιμές ανάμεσα στο 0 και στο 1. Επίσης τα ζεύγη ορθολογικών αποδόσεων είναι εκείνα των οποίων οι τιμές είναι τουλάχιστον ίσες με αυτές που δίνει η ισορροπία Nash. Στο Σχήμα 7.1 βλέπουμε ότι οι εφικτές αποδόσεις, βρίσκονται εντός του τετραπλεύρου με κορυφές τα σημεία (x, x) , $(0, w)$, $(w, 0)$ και (y, y) , ενώ οι ορθολογικές αποδόσεις βρίσκονται εντός της γραμμοσκιασμένης επιφάνειας.



Σχήμα 7.2

Θεώρημα: (Folk Theorem)

Αν (a_1^*, a_2^*) ένα ζεύγος αποδόσεων που αποτελεί ισορροπία Nash για ένα παίγνιο ενός σταδίου και (u_1, u_2) ένα ζεύγος εφικτών αποδόσεων στην περίπτωση που το παίγνιο ενός σταδίου επαναλαμβάνεται. Για κάθε ζεύγος ορθολογικής απόδοσης (u_1, u_2) , δηλαδή με τιμές τέτοιες ώστε $u_1 > a_1^*$ και $u_2 > a_2^*$, υπάρχει ένας αριθμός δ τέτοιος ώστε, για κάθε παράγοντα $\delta > \delta$ να υπάρχει μια ισορροπία τέλεια ανά υποπαίγνιο με αποδόσεις (u_1, u_2) .

Απόδειξη: Έστω (σ_1^*, σ_2^*) η ισορροπία Nash που αντιστοιχεί σε ζεύγος αποδόσεων (a_1^*, a_2^*) καθώς και το ζεύγος αποδόσεων (u_1, u_2) που παράγεται από τις κινήσεις v_1 και v_2 των παικτών 1 και 2 αντίστοιχα σε κάθε στάδιο. Θεωρούμε την ακόλουθη στρατηγική -trigger³⁴:

«Ξεκίνα με συμφωνία να παιχθεί η κίνηση v_i και συνέχισε με αυτή όσο οι συμφωνημένες κινήσεις ακολουθούνται κι από τους δύο παίκτες. Αν ένας από τους δύο δεν ακολουθήσει την κίνηση v_i , τότε χρησιμοποίησε τη στρατηγική σ_i^* για το

³⁴ Trigger strategy (ή Grim Trigger) είναι μία στρατηγική που χρησιμοποιείται στα επαναλαμβανόμενα παίγνια σύμφωνα με την οποία ο παίκτης συνεργάζεται όσο διατηρεί τη συμφωνία κι αντίπαλος, ενώ στην περίπτωση αθέτησης από αυτόν επέρχεται «τιμωρία». Η ιδέα χρησιμοποιήθηκε πρώτη φορά από τον James W. Friedman σε άρθρο του 1971.

υπόλοιπο του παιχνιδιού.»

Από κατασκευής, οποιαδήποτε ισορροπία Nash περιέχει αυτές τις (trigger) στρατηγικές θα είναι και τέλεια ανά υποπαίγνιο. Συνεπώς αρκεί μόνο να βρεθούν οι συνθήκες για μία ισορροπία. Έστω μία κίνηση v_1' τέτοια ώστε η απόδοση αυτής για τον παίκτη 1 να είναι $a_1(v_1', v_2) > u_1$. Τότε η συνολική απόδοση για την εναλλαγή από την κίνηση v_1' στην στρατηγική trigger που περιγράψαμε, εναντίον του αντιπάλου δεν θα υπερβαίνει την παρακάτω τιμή:

$$a_1(v_1', v_2) + \delta \frac{a_1^*}{1 - \delta}.$$

Δεν είναι λοιπόν επωφελές να αλλάξουμε την κίνηση v_1' αν $\delta \geq \delta_1$, όπου

$$\delta_1 = \frac{a_1(v_1', v_2) - u_1}{a_1(v_1', v_2) - a_1^*}.$$

Προφανώς θα ισχύει ότι $0 < \delta_1 < 1$, δεδομένου ότι $a_1^* < u_1 < a_1(v_1', v_2)$.

Παρόμοια οδηγούμαστε στον ελάχιστο δυνατό παράγοντα δ_2 και για τον παίκτη 2.

Συνεπώς για $\delta = \max(\delta_1, \delta_2)$ ολοκληρώθηκε το πρώτο μέρος της απόδειξης.

Έστω τώρα ότι έχουμε τις αποδόσεις u_i , οι οποίες προέκυψαν από τυχαίες στρατηγικές σ_i . Θεωρούμε πως υπάρχει μία συσκευή η οποία παράγει τυχαία αποτελέσματα (για παράδειγμα ένα αμερόληπτο ζάρι) τα οποία παρατηρούνται κι από τους δύο παίκτες. Υποθέτουμε επίσης πως υπάρχει ένας συμφωνημένος κανόνας με τον οποίο γίνεται αντιστοίχιση των αποτελεσμάτων με την επιλογή της κίνησης κάθε παίκτη. Αυτές οι υποθέσεις γίνονται για να γίνει φανερό ότι οι στρατηγικές αποτελούν αντικείμενο παρατήρησης από τους παίκτες. Με αυτό τον τρόπο, το προηγούμενο επιχείρημα μπορεί να επαναληφθεί αντικαθιστώντας τις κινήσεις u_i και u_i' με τις στρατηγικές σ_i και σ_i' .

■

7.4 ΣΤΟΧΑΣΤΙΚΑ ΠΑΙΓΝΙΑ

Ένα στοχαστικό παίγνιο ορίζεται από ένα χώρο καταστάσεων X , με το παίγνιο ενός σταδίου να αντιστοιχεί σε μία κατάσταση. Σε κάθε κατάσταση x ο παίκτης i μπορεί να διαλέξει κινήσεις από ένα σύνολο $\mathcal{A}_i(x)$. Κάθε στάδιο του παιχνιδιού παίζεται σε διακριτό χρόνο $t = 0, 1, 2, \dots$. Η επιλογή κίνησης από τους παίκτες σε μία

συγκεκριμένη κατάσταση προσδιορίζει τόσο τις άμεσες αμοιβές αλλά και την πιθανότητα άφιξης σε κάθε άλλη γνωστή κατάσταση στο επόμενο σημείο που πρέπει να ληφθεί απόφαση. Διότι, δεδομένου ότι οι παίκτες βρίσκονται σε μία κατάσταση x και επιλέγουν κινήσεις $v_1 \in \mathcal{A}_1(x)$ και $v_2 \in \mathcal{A}_2(x)$, οι παίκτες λαμβάνουν άμεσες αμοιβές $r_1(x, v_1, v_2)$ και $r_2(x, v_1, v_2)$ και η πιθανότητα να βρεθούν στην κατάσταση x' στην επόμενη απόφαση τους είναι $p(x'|x, v_1, v_2)$.

Ορισμός: Μία στρατηγική καλείται Μαρκοβιανή στρατηγική αν η συμπεριφορά του παίκτη τη χρονική στιγμή t εξαρτάται μόνο από την κατάσταση x . Μία αμιγής Μαρκοβιανή στρατηγική καθορίζει την κίνηση $v(x)$ για κάθε κατάσταση $x \in X$.

Τώρα θα προχωρήσουμε κάνοντας τις παρακάτω υποθέσεις:

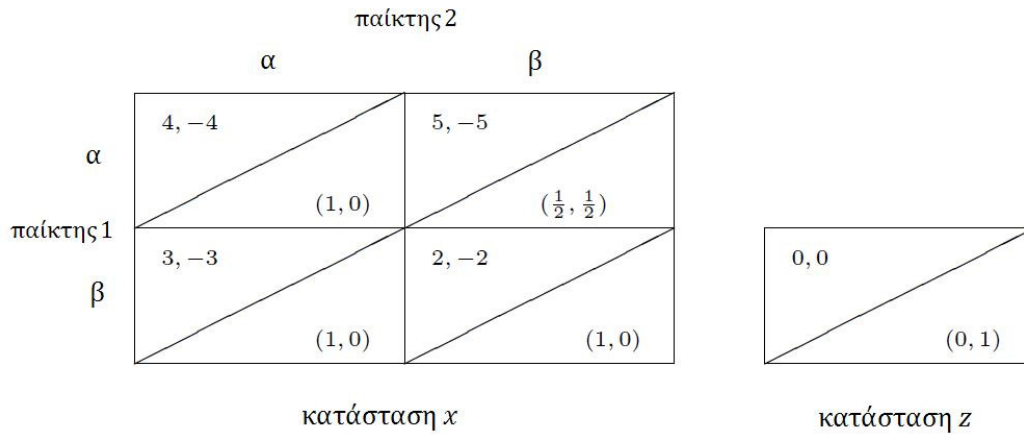
- i. Η διάρκεια του παιχνιδιού δεν είναι γνωστή στους παίκτες.
- ii. Οι αμοιβές και οι πιθανότητες μετάβασης είναι ανεξάρτητες από το χρόνο.
- iii. Οι στρατηγικές είναι Μαρκοβιανές.

Αν το πλήθος των καταστάσεων και το πλήθος των διαθέσιμων κινήσεων σε κάθε κατάσταση είναι μικρά, τότε ένα στοχαστικό παίγνιο μπορεί να έχει μία απλή διαγραμματική αναπαράσταση. Ας δούμε το ακόλουθο παράδειγμα.

Παράδειγμα: Έστω ο χώρος καταστάσεων $X = \{x, z\}$. Στην κατάσταση x , και οι δύο παίκτες μπορούν να επιλέξουν ανάμεσα σε κινήσεις από τα σύνολα $\mathcal{A}(x) = \mathcal{A}(z) = \{\alpha, \beta\}$. Οι άμεσες αμοιβές για τον παίκτη 1 στην κατάσταση x είναι: $r_1(x, \alpha, \alpha) = 4$, $r_1(x, \alpha, \beta) = 5$, $r_1(x, \beta, \alpha) = 3$ και $r_1(x, \beta, \beta) = 2$. Αυτό είναι ένα παίγνιο μηδενικού αθροίσματος³⁵, οπότε $r_2(x, v_1, v_2) = -r_1(x, v_1, v_2)$ για κάθε ζεύγος κινήσεων ($v_i = \alpha$ ή β , με $i = 1, 2$). Αν οι παίκτες επιλέξουν το ζεύγος κινήσεων $[\alpha, \beta]$ στην κατάσταση x , τότε θα κινηθούν στην κατάσταση z με πιθανότητα $\frac{1}{2}$. Αν επιλεγεί οποιοδήποτε άλλο ζεύγος κινήσεων, τότε οι παίκτες παραμένουν στην κατάσταση x με πιθανότητα 1. Αν οι παίκτες είναι στην κατάσταση z , τότε έχουν ως μοναδική επιλογή την κίνηση $\mathcal{A}(z) = \{\beta\}$ και οι άμεσες αμοιβές είναι: $r_1(z, \beta, \beta) = r_2(z, \beta, \beta) = 0$. Από τη στιγμή που οι παίκτες φτάσουν στην κατάσταση z , παραμένουν σε αυτή με πιθανότητα 1. Άρα η κατάσταση z είναι μία απορροφητική

³⁵ Zero sum game

κατάσταση μηδενικής απόδοσης. Δίνεται το ακόλουθο διάγραμμα (σχήμα 7.2) συνοπτικής αναπαράστασης του παιγνίου:



Σχήμα 7.3

Έστω ένα παίγνιο στην κατάσταση x τη χρονική στιγμή t . Αν γνωρίζαμε τις στρατηγικές που αντιστοιχούν στην ισορροπία Nash για κάθε παίκτη από τη χρονική στιγμή $t + 1$ κι έπειτα, θα μπορούσαμε να εκτιμήσουμε τις μελλοντικές αποδόσεις που θα είχε κάθε παίκτης από $t + 1$ κι έπειτα, δεδομένου ότι ξεκινούν και οι δύο από μία συγκεκριμένη κατάσταση. Συμβολίζουμε τις αναμενόμενες μελλοντικές αποδόσεις που προέρχονται από τις στρατηγικές τις ισορροπίας Nash (για κάθε παίκτη) με $a_i^*(x)$. Τη χρονική στιγμή t , οι παίκτες, οι παίκτες παίζουν σε παίγνιο μίας απόφασης με αποδόσεις:

$$a_i(v_1, v_2) = \left(r_i(x, v_1, v_2) + \delta \sum_{x' \in X} p(x'|x, v_1, v_2) a_i^*(x') \right)$$

Όπου οι μελλοντικές αποδόσεις έχουν συντελεστή δ όπως τον είχαμε ορίσει σε προηγούμενη παράγραφο³⁶ έτσι ώστε να πετύχουμε ρεαλιστικό αποτέλεσμα. Ένα τέτοιο παίγνιο καλείται *αποτελεσματικό παίγνιο* στην κατάσταση x .

Για μία Μαρκοβιανή στρατηγική οι εκτιμώμενες αποδόσεις στο μέλλον, σε κατάσταση x είναι ανεξάρτητες του χρόνου t . Για το λόγο αυτό οι αποδόσεις για μία ισορροπία από Μαρκοβιανές στρατηγικές δίνονται από το συνδυασμό των λύσεων

³⁶ Discount factor

των παρακάτω εξισώσεων (για κάθε $x \in X$).

$$a_1^*(x) = \max_{v_1 \in \mathcal{A}_1(x)} \left(r_1(x, v_1, v_2^*) + \delta \sum_{x' \in X} p(x'|x, v_1, v_2^*) a_1^*(x') \right)$$

και

$$a_2^*(x) = \max_{v_2 \in \mathcal{A}_2(x)} \left(r_2(x, v_1^*, v_2) + \delta \sum_{x' \in X} p(x'|x, v_1^*, v_2) a_2^*(x') \right)$$

Πρέπει όμως να αναφέρουμε ότι δυστυχώς η επίλυση αυτών των εξισώσεων έχει υψηλό υπολογιστικό κόστος.

Παράδειγμα: Θεωρούμε το στοχαστικό παίγνιο με τις μεταβάσεις κατάστασης και τις αποδόσεις να δίνονται από το Σχήμα 7.2 και παράγοντα (discount factor) $\delta = \frac{2}{3}$. Η αμοιβή στην κατάσταση z είναι 0 και για τους δύο παίκτες. Αν u είναι η τρέχουσα αμοιβή για τον παίκτη 1 στην κατάσταση x , τότε η αντίστοιχη τιμή για τον παίκτη 2 θα είναι $-u$ εφόσον είναι παιχνίδι μηδενικού αθροίσματος. Αυτό σημαίνει πως στην κατάσταση x , οι παίκτες αντιμετωπίζουν το εξής αποτελεσματικό παίγνιο:

Πίνακας 7.3

		Παίκτης 2	
		α	β
Παίκτης 1	α	$(4 + \frac{2}{3}u, -4 - \frac{2}{3}u)$	$(5 + \frac{1}{3}u, -5 - \frac{1}{3}u)$
	β	$(3 + \frac{2}{3}u, -3 - \frac{2}{3}u)$	$(2 + \frac{2}{3}u, -2 - \frac{2}{3}u)$

Προφανώς το ζεύγος (β, α) δεν αποτελεί ισορροπία κατά Nash για καμία τιμή του u . Αγνοώντας τις οριακές περιπτώσεις, η ισορροπία κατά Nash για αυτό το αποτελεσματικό παίγνιο στην κατάσταση x θα είναι:

$$Ισορ. Nash: \begin{cases} (\alpha, \alpha) & αν \ u < 3 \\ (\beta, \beta) & αν \ u > 9 \\ (\alpha, \beta) & αν \ 3 < u < 9 \end{cases}$$

Υποθέτουμε τώρα πως οι παίκτες επιλέγουν να παίζουν (α, α) . Τότε $u = 4 + \frac{2}{3}u \Rightarrow u = 12$. Όμως αυτή η τιμή δεν ικανοποιεί τη σχέση $u < 3$ άρα δεν είναι ισορροπία. Αν οι παίκτες επιλέξουν να παίζουν (β, β) , θα έχουμε: $u = 2 + \frac{2}{3}u \Rightarrow u = 6 < 9$, άρα ούτε το ζεύγος (β, β) είναι ισορροπία αφού δεν ικανοποιείται η συνθήκη $u > 9$. Αν όμως επιλεγθεί από τους παίκτες το ζεύγος (α, β) τότε $u = 5 + \frac{1}{3}u \Rightarrow u = 7,5$, δηλαδή $u \in [3, 9]$ οπότε ικανοποιείται η αναγκαία συνθήκη και έτσι έχουμε μοναδική μαρκοβιανή στρατηγική (α, β) στην κατάσταση x που αποτελεί ισορροπία Nash.

■

Βιβλιογραφία

1. Axzo Press LLC, Project Management: Intermediate (2002), ILT
2. Bather J., (2000), Decision Theory: An introduction to dynamic programming and sequential decisions
3. Davis Morton D.,(foreword by Oskar Morgenstern), Game Theory-A nontechnical Introduction, (1970) Basic Books N.Y.
4. Filar J., Vrieze K. (1997), Competitive Markov Decision Processes, Springer
5. Gabszewicz J. J. (1999), Strategic Interaction and Markets, Oxford University Press.
6. Gibbons Robert, A Primer in Game Theory(1992)
7. Gintis Helbert, Game Theory Evolving (2008), Princeton University Press
8. Grafen, A. (1991), Modelling in behavioural ecology, in 'Behavioural Ecology: An Evolutionary Approach', 3 edn, Blackwell, Oxford, p. 5–31
9. Hargreaves Heap S. P., Varoufakis Y., (1995), Game Theory: A Critical Introduction, Routledge.
10. Hillier Frederick S., Lieberman Gerald J.(2001), Introduction to Operations Research
11. Kubilay Ilhan Geckil, Patrick L. Anderson, Applied Game Theory and Strategic Behavior (2009), CRC Press
12. Markus Mobius (2008), Game Theory and Economics, Harvard University
13. Meisel Stephan, Anticipatory Optimization for Dynamic Decision Making (2011), Springer
14. Morris Peter, Introduction to Game Theory (1994), Springer
15. Myerson Roger (1991), Game Theory: Analysis of Conflict
16. Neisser Hans (1940), A Note on Pareto's Theory of Production, Econometrica, 8 July, 253-262
17. Neogy S. K.,Bapat R.B., Das A. K., Parthasarathy T., Mathematical Programming and Game Theory for Decision Making (2008), World Scientific Publishing Co.
18. Notes for the CFA Exam, Level I, Book 2: Economics (2013), Kaplan Inc.
19. Osborne Martin J., Εισαγωγή στη Θεωρία Παιγνίων (1995), Εκδόσεις Κλειδάριθμος

20. Puterman M. L., (1994), Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley.
21. Rama Murthy P. (2005),, Operations Research
22. Rapoport Anatol ,Two-Person Game Theory (the essential ideas) (1999), Ann Arbor Science Paperbacks
23. Rasmusen Eric (2006), Games and Information: An Introduction to Game Theory,
24. Rodica Branzei, Dinco Dimitrov, Stef Tijs (2005), Models in Cooperative Game Theory, Springer
25. Ross S. M. (1995), Introduction to Stochastic Dynamic Programming, Academic Press
26. Schmidt Christian (2002), Game Theory and Economic Analysis: a quiet revolution in economics, Dalloz Editions, Routledge
27. Vajda S. (1963), An Introduction to Linear Programming and the Theory of Games, Science Paperbacks
28. Von Neumann J., Morgenstern Oskar (1953), Theory of games and Economic Behavior, Princeton University Press
29. Webb James N. (2006). Game Theory: Decisions, Interaction and Evolution , Springer
30. Μηλολιδάκης Κωστής, Θεωρία Παιγνίων-Μαθηματικά Μοντέλα Σύγκρουσης και Συνεργασίας (2009), εκδόσεις Σοφία
31. Οικονόμου Αν., Φακίνος Δ. (2003) , Εισαγωγή στην Επιχειρησιακή Έρευνα, Εκδόσεις Συμμετρία
32. Τσίτσας Λ., Εφαρμοσμένος Διανυσματικός Απειροστικός Λογισμός (2003), Εκδόσεις Συμμετρία
33. http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-254-game-theory-with-engineering-applications-spring-2010/lecture-notes/MIT6_254S10_lec15.pdf (12/3/2014)
34. <http://www.oxfordreference.com/view/10.1093/oi/authority.20110803105713705> (21/11/2013)
35. <http://www.stanford.edu/~jdlevin/Econ%20203/RepeatedGames.pdf> (15/3/2014)
36. https://en.wikipedia.org/wiki/Operations_research (04/06/2013)
37. <http://www.oecd.org/competition/cartels/> (02/02/2014)