Στοχαστικά Παιχνίδια

Επειδή στην τάξη υπήρχαν αρκετές ερωτήσεις πάνω στον τρόπο που αντιλαμβανόμαστε τις πληρωμές, γιατί αυτές φράσσονται, κ.ο.κ. , θέλω εδώ να προχωρήσω σε ορισμένες διευκρινίσεις σχετικά με τα μέτρα πιθανότητας και τις αναμενόμενες τιμές σε ένα Στοχαστικό Παιχνίδι 2-παικτών 0-αθροίσματος. Αν αισθάνεστε ότι το μάθημα στην τάξη σας έχει «καλύψει», τότε δεν υπάρχει λόγος να ασχοληθείτε με όσα ακολουθούν, αλλά παρόλ' αυτά, ζητείται από όλους να κάνουν την Άσκηση 1 που περιέχεται σ' αυτό το σημείωμα. Ειδικότερα τώρα όσο αφορά το μπέρδεμα στο τέλος του μαθήματος της Τετάρτης, το θέμα είναι τετριμμένο και αναφέρομαι σε αυτό στο τέλος.

Υπενθυμίζω ότι σε ένα στοχαστικό παιχνίδι, οι τρέχουσες πληρωμές των παικτών δίνονται από τους πίνακες $\{C^1,\ C^2, \dots, C^l\}$, με στοιχεία (c_{ij}^k), $i = 1, \dots, m_k$, $j = 1, \dots, n_k$, $k = 1, \dots , l$ και ότι επίσης για κάθε τριάδα (k, i, j) έχουν δοθεί πιθανότητες μεταπήδησης p_{ij}^{kr}, $r = 1, \dots , l$. Λέμε ότι το στοχαστικό παιχνίδι είναι παιχνίδι με θετικές πιθανότητες τερματισμού εάν κάνουμε την κρίσιμη υπόθεση ότι $\sum_{r=1}^{l} p_{ij}^{kr} < 1$. Τότε, η πιθανότητα σταματήματος (όταν η κατάσταση είναι k και οι παίκτες αποφασίσουν (i, j)) συμβολίζεται με s_{ij}^k, δηλαδή $s_{ij}^k = 1 - \sum_{r=1}^{l} p_{ij}^{kr}$ και επίσης ορίζουμε $s \coloneqq \min_{i,j,k}\ s_{ij}^k$. Πάντα θα υποθέτουμε ότι το παιχνίδι είναι τέλειας ανάμνησης.

Ορίζουμε ως σύνολο των δυνατών ιστοριών του παιχνιδιού μέχρι το στάδιο t το σύνολο όλως των ακολουθιών $h_t = (k_1, i_1, j_1, k_2, i_2, j_2, \dots , k_{t-1}, i_{t-1}, j_{t-1}, k_t)$ που θα μπορούσαν να παρατηρηθούν ως διαδρομές (τροχιές) του παιχνιδιού μέχρι το στάδιο t σε κάποιο παίξιμό του. Εδώ k_τ είναι η κατάσταση και i_τ, j_τ είναι οι αποφάσεις των παικτών στο στάδιο $\tau, \tau = 1, \dots , t$. Η ιστορία του παιχνιδιού h_t εκφράζει τη πληροφορία που έχουν οι παίκτες στη διάθεσή τους όταν καλούνται να κάνουν τις (ταυτόχρονες και ανεξάρτητες) κινήσεις τους στο στάδιο t. Επιτρέπουμε ιστορίες να παρατηρούνται με πιθανότητα 0, αλλά δεν επιτρέπουμε ιστορίες που δεν αντιστοιχούν σε συνεκτικές τροχιές (δηλαδή ιστορίες με αποφάσεις μη συμβατές με την κατάσταση του συστήματος). Με άλλα λόγια, έστω $H \coloneqq \{(k, i, j): k = 1, \dots, l,\ i = 1, \dots, m_k,\ j = 1, \dots, n_k\}$. Τότε το σύνολο όλων των δυνατών ιστοριών κατά το στάδιο t είναι το πεπερασμένο Καρτεσιανό γινόμενο H^t.

Μια συμπεριφορική στρατηγική x για τον παίκτη *I* (αντίστοιχα y για τον *II*) συνίσταται από μια ακολουθία κατανομών πιθανότητας $(x_1(\cdot\,|h_1), x_2(\cdot\,|h_2), \dots, x_t(\cdot\,|h_t), \dots)$ η οποία δίνεται για κάθε ιστορία του παιχνιδιού. Δηλαδή, η κατανομή $x_t(\cdot\,|h_t)$ είναι **συνάρτηση** της h_t και προσδιορίζει με πόση πιθανότητα ο *I* θα επιλέξει κάθε γραμμή του πινακοπαιχνιδιού στο οποίο θα βρεθεί κατά το στάδιο t (και ανάλογα για την $y_t(\cdot\,|h_t)$ του παίκτη *II*). Αυτό συμβαδίζει με τη γενική ιδέα που έχουμε για «ένα πλήρες σχέδιο δράσης», πλην όμως προσέξτε ότι οι στρατηγικές αυτές ΔΕΝ είναι οι συνήθεις μεικτές, οι οποίες αναμειγνύουν διαδρομές απείρου μήκους. Ο Aumann (1974)—η παραπομπή υπάρχει στη βιβλιογραφία—επεξέτεινε για όχι πεπερασμένα παιχνίδια το θεώρημα του Kuhn που επιτρέπει να περιοριζόμαστε σε συμπεριφορικές στρατηγικές στα παιχνίδια τέλειας ανάμνησης και το θεώρημά του εφαρμόζεται στα στοχαστικά παιχνίδια. Επομένως, χωρίς βλάβη της γενικότητας, μπορούμε να περιοριστούμε σε συμπεριφορικές στρατηγικές.

Ακολουθεί ορολογία που αφορά επιμέρους περιπτώσεις στρατηγικών.

Εάν για κάθε t η $x_t(\cdot\,|h_t)$ είναι της μορφής $x_t(\cdot\,|k_1, k_t)$, δηλαδή εξαρτάται μόνο από την αρχική και από την παρούσα κατάσταση, τότε η στρατηγική αυτή ονομάζεται *ημι-Μαρκοβιανή*. Αν επιπλέον, για κάθε t η $x_t(\cdot\,|h_t)$ είναι της μορφής $x_t(\cdot\,|k_t)$, δηλαδή εξαρτάται μόνο από την παρούσα κατάσταση, τότε η ημι-Μαρκοβιανή στρατηγική πλέον λέγεται *Μαρκοβιανή*. Αν επιπλέον, για κάθε t η Μαρκοβιανή στρατηγική είναι της μορφής $x(\cdot\,|k_t)$, δηλαδή αυτή δεν εξαρτάται από το χρονικό στάδιο, αλλά μόνο από την παρούσα κατάσταση, τότε αυτή ονομάζεται *στάσιμη*. Μια στρατηγική ονομάζεται *ντετερμινιστική* όταν οι κατανομές πιθανότητας πάντα δίνουν μάζα πιθανότητας 1 σε κάποια απόφαση. Οι στάσιμες ντετερμινιστικές στρατηγικές συχνά αναφέρονται και ως «καθαρές» στάσιμες στρατηγικές.

Δεδομένων των πιθανοτήτων μεταπήδησης p_{ij}^{kr}, $r = 1, \dots, l$, η αρχική κατάσταση του παιχνιδιού k_1 μαζί με την επιλογή συμπεριφορικών στρατηγικών (x, y) επάγουν ένα μέτρο πιθανότητας $\mathbf{P}_{k_1,x,y}^{\mathrm{t}}$ πάνω στο πεπερασμένο H^t. Συγκεκριμένα,

Άσκηση 1: Να δειχτεί ότι $\mathbf{P}_{k_1,x,y}^{\mathrm{t}}(k_1, i_1, j_1, k_2, i_2, j_2, \dots, k_{t-1}, i_{t-1}, j_{t-1}, k_t) =$

$$= x_1(i_1|k_1)y_1(j_1|k_1)\, p_{i_1j_1}^{k_1k_2} x_2(i_2|k_1, i_1, j_1, k_2)y_2(j_2|k_1, i_1, j_1, k_2)p_{i_2j_2}^{k_2k_3} \dots\dots\, p_{i_{t-1}j_{t-1}}^{k_{t-1}k_t}$$

[Υπόδειξη: Ο τύπος προκύπτει από διαδοχικές δεσμεύσεις, την παρατήρηση ότι η μετάβαση στην επόμενη κατάσταση γίνεται σύμφωνα με κατανομή που εξαρτάται αποκλειστικά από την τιμή της παρούσας κατάστασης και αποφάσεων των παικτών, και τυπικά ολοκληρώνεται με επαγωγή. Μπορείτε να δείτε και τις σημειώσεις πάνω στις ελεγχόμενες διαδικασίες Markov («Στοχαστικές Μέθοδοι στην Επιχειρησιακή Έρευνα ΙΙ»)].

Ο N. Kolmogorov έδειξε ότι η ακολουθία μέτρων $\mathbf{P}_{k_1,x,y}^{\mathrm{t}}$, $t = 1,2, \dots$, επεκτείνεται κατά «φυσικό» τρόπο σε ένα μοναδικό μέτρο πιθανότητας $\mathbf{P}_{k_1,x,y}$ πάνω στο σύνολο των δυνατών ιστοριών H^∞ του παιχνιδιού άπειρου μήκους (βλ. Billingsley (1968)).

Δεδομένων στρατηγικών (x, y) των παικτών, έστω ότι οι τυχαίες μεταβλητές $K_{x,y}^t$, $I_{x,y}^t$, $J_{x,y}^t$, $t = 1,2, \dots$, αντιπροσωπεύουν την κατάσταση του παιχνιδιού και τις αποφάσεις των παικτών στο στάδιο t. Προφανώς οι περιθώριες κατανομές των $K_{x,y}^t$, $I_{x,y}^t$, $J_{x,y}^t$ καθορίζονται από το $\mathbf{P}_{k_1,x,y}$. Έτσι, η αναμενόμενη πληρωμή του *I* κατά το στάδιο t θα δίνεται από την

$$h_t(k_1, x, y) = \sum_{(k,i,j)\in H} c_{ij}^k\, \mathbf{P}_{k_1,x,y}(K_{x,y}^t = k,\ I_{x,y}^t = i,\ J_{x,y}^t = j)$$

και η συνολική πληρωμή του στο παιχνίδι θα δίνεται από την

$$h(k_1, x, y) = \sum_{t=1}^{\infty} h_t(k_1, x, y)$$

(Προσωπικά βρίσκω τα παραπάνω ενδιαφέροντα μόνο στο βαθμό που ξεκαθαρίζουν το θεωρητικό πλαίσιο. Όμως, αλλοίμονο αν αρχίσουμε να σκεφτόμαστε έτσι κατά την αντιμετώπιση προβλημάτων, αφού τότε θα κολλάμε στις γενικότητες διυλίζοντες τον κώνωπα.)

Ας ονομάσουμε τώρα $M = \max_{(k,i,j)\in H} |c_{ij}^k|$. Τότε,

$$|h_t(k_1,x,y)| \leq M \sum_{(k,i,j)\in H} \mathbf{P}_{k_1,x,y}(K_{x,y}^t = k,\ I_{x,y}^t = i,\ J_{x,y}^t = j)$$

Όμως,

$$\sum_{(k,i,j)\in H} \mathbf{P}_{k_1,x,y}(K_{x,y}^t = k,\ I_{x,y}^t = i,\ J_{x,y}^t = j) = \sum_k \mathbf{P}_{k_1,x,y}(K_{x,y}^t = k)$$

Το τελευταίο άθροισμα εκφράζει την πιθανότητα να έχει φτάσει το παιχνίδι μέχρι το στάδιο t. Άρα,

$$\sum_k \mathbf{P}_{k_1,x,y}(K_{x,y}^t = k) \leq (1-s)^{t-1}$$

Συμπεραίνουμε ότι

$$|h_t(k_1,x,y)| \leq M(1-s)^{t-1}$$

Δηλαδή, η σειρά που δίνει την αναμενόμενη πληρωμή κυριαρχείται από τη γεωμετρική σειρά και επομένως η πληρωμή είναι καλά ορισμένη. Αλλιώς,

$$|h(k_1,x,y)| \leq \sum_{t=1}^{\infty} M(1-s)^{t-1} = \frac{M}{s}$$

Ας έρθουμε τώρα στο πρώτο στάδιο.

Στο βιβλίο υπάρχει ο ισχυρισμός (σελ. 164) ότι αν ο *I* διαλέξει i και ο *II* διαλέξει j, τότε ο *I* εξασφαλίζει τουλάχιστον

$$c_{ij}^k + \sum_{r=1}^{l} p_{ij}^{kr}\,(v^r - \varepsilon)$$

Στην τάξη, από μπέρδεμα δικό μου, μπλεχτήκαμε (όχι όλοι πάντως). Ο παραπάνω ισχυρισμός είναι προφανής δεδομένου ότι με πιθανότητα p_{ij}^{kr} το παιχνίδι θα βρεθεί αύριο στην κατάσταση r, από όπου και μετά ο *I* εξασφαλίζει $v^r - \varepsilon$ παίζοντας την ε-βέλτιστη στρατηγική του που ήδη γνωρίζουμε ότι υπάρχει.