ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
Τμήμα Επιστήμης Υπολογιστών

ΗΥ-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες - Εαρινό Εξάμηνο 2021
Διδάσκων: Π. Τσακαλίδης

Φροντιστήριο 6: Μαρκοβιανές αλυσίδες – Εξισώσεις ισορροπίας

Άσκηση 1.
Σε μια πολύ μεγάλη ακολουθία DNA έχουν καταγραφεί οι εμπειρικές συχνότητες σύμφωνα με τις οποίες κάθε μια από τις βάσεις A, C, G, T ακολουθείται από μια άλλη. Αυτές φαίνονται στν παρακάτω πίνακα.

$$P = \begin{pmatrix} 0.200 & 0.255 & 0.285 & 0.260 \\ 0.302 & 0.322 & 0.178 & 0.198 \\ 0.392 & 0.139 & 0.292 & 0.177 \\ 0.148 & 0.198 & 0.346 & 0.308 \end{pmatrix}$$

Αν θεωρήσουμε (προσεγγιστικά) ότι η εναλλαγή των βάσεων ακολουθεί το μοντέλο μιας Μαρκοβιανής αλυσίδας με χώρο καταστάσεων $S = \{A,\ C,\ G,\ T\}$. Κάποιος συνάδελφός σας ισχυρίζεται ότι η οριακή κατανομή αυτής της αλυσίδας είναι η $[\pi_A,\ \pi_C,\ \pi_G,\ \pi_T] = [0.2639,\ 0.2245,\ 0.2772,\ 0.2344]$.

α΄) Είναι σωστός ο ισχυρισμός του συναδέλφου σας; Βρείτε και ερμηνεύστε την οριακή κατανομή.

β΄) Βρείτε την πιθανότητα σε ένα τυχαίο απομακρυσμένο σημείο της ακολουθίας να υπάρχουν δύο διαδοχικές βάσεις A.

Λύση.

α΄) Έστω X_n ο τύπος της n-οστής βάσης στην ακολουθία, με $X_n \in S = \{A, C, G, T\}$.

Εφόσον ο πίνακας μετάβασης $P > 0$ (όλα τα στοιχεία του είναι θετικά), η αλυσίδα έχει μια μοναδική επαναληπτική κλάση, η οποία είναι μοναδική. Υπό αυτές τις συνθήκες, η οριακή κατανομή $\underline{\pi} = [\pi_A,\ \pi_C,\ \pi_G,\ \pi_T]$ υπάρχει και είναι η μοναδική λύση των εξισώσεων ισορροπίας:

$$\begin{cases} \underline{\pi} = \underline{\pi} \cdot P \\ \underline{\pi} \cdot \underline{e} = 1 \end{cases}$$

Αντικαθιστώντας την κατανομή που μας δίνει ο συνάδελφός μας, διαπιστώνουμε ότι ικανοποιείται από τις εξισώσεις ισορροπίας, άρα είναι σωστός ο ισχυρισμός του.

Σημείωση: Θα μπορούσαμε να υπολογίσουμε και μόνοι μας την στατική κατανομή από τις εξισώσεις ισορροπίας, αλλά οι πιθανότητες στον παραπάνω στοχαστικό πίνακα δεν βοηθάνε για πράξεις στο χαρτί.

Ερμηνεία 1. (Οριακή Κατανομή) Αν πάρουμε μια βάση της αλυσίδα σε ένα σημείο πολύ μακριά από την αρχή, τότε ανεξάρτητα από την αρχική βάση, η πιθανότητα να βρούμε A τείνει στο 0.2639 κ.ο.κ.

Ερμηνεία 2. (Συχνότητες) Σε μια μεγάλη αλυσίδα, περίπου 26.39% των βάσεων είναι τύπου A, 22.45% είναι τύπου G κ.ο.κ.

β΄) Ζητείται η πιθανότητα $P(X_n = A, X_{n+1} = A)$, καθώς το $n \to \infty$.

$$\begin{aligned}
\lim_{n\to\infty} P(X_n = A, X_{n+1} = A) &= \lim_{n\to\infty} P(X_{n+1} = A | X_n = A) P(X_n = A) \\
&= P_{AA} \cdot \pi_A = 0.200 \cdot 0.2639 \\
&= 0.0528
\end{aligned}$$

Άσκηση 2.
Στο γκαράζ του Πανεπιστημίου έχει εγκατασταθεί μία μπάρα η οποία λειτουργεί με κάρτες και δυστυχώς κινδυνεύει από το προσωπικό και τους καθηγητές. Συγκεκριμένα, κάθε μέρα ένα αυτοκίνητο προσκρούει στην μπάρα με πιθανότητα p και στην περίπτωση αυτή μία καινούρια μπάρα πρέπει να εγκατασταθεί. Επίσης, μία μπάρα η οποία λειτουργεί για m μέρες αντικαθίσταται από μία νέα για λόγους καλής λειτουργίας.

α΄) Ορίστε τη Μαρκοβιανή αλυσίδα που περιγράφει πλήρως τη λειτουργία του συστήματος της μπάρας του γκαράζ. Ορίστε επακριβώς τις καταστάσεις, δώστε το γράφημα της αλυσίδας με τις πιθανότητες μετάβασης καθαρά γραμμένες στις ακμές και γράψτε τον αντίστοιχο πίνακα πιθανοτήτων μετάβασης.

β΄) Γράψτε τις εξισώσεις ισορροπίας και υπολογίστε τη στατική κατανομή της αλυσίδας

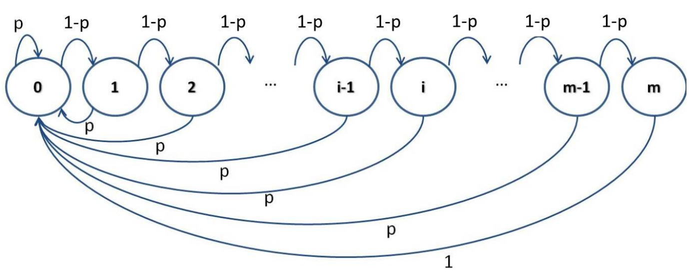
γ΄) Σε βάθος χρόνου (μακροπρόθεσμα), ποια είναι η συχνότητα με την οποία αντικαθίσταται η μπάρα; Με τι ισούται αυτή η συχνότητα όταν η διάρκεια m της φυσικής ζωής της μπάρας είναι πολύ μεγάλη;

Λύση.

α΄) Ορίζουμε $m+1$ καταστάσεις $\{0, 1, 2, \ldots, m-1, m\}$ ως εξής:

Κατάσταση i: 'Η μπάρα βρίσκεται σε λειτουργία για i μέρες.' $\quad i = 0, 1, \ldots, m.$

Το γράφημα της αλυσίδας δίδεται στο Σχήμα 1:



Σχήμα 1: Το διάγραμμα καταστάσεων για την άσκηση 2.

Ο πίνακας πιθανοτήτων μετάβασης είναι:

$$P = \begin{bmatrix} p & 1-p & 0 & 0 & \ldots & 0 \\ p & 0 & 1-p & 0 & \ldots & 0 \\ p & 0 & 0 & 1-p & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p & 0 & 0 & 0 & \ldots & 1-p \\ 1 & 0 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

β΄) Συμβολίζοντας με $\underline{\pi} = [\pi_0, \pi_1, \ldots, \pi_m]$ το διάνυσμα της στατικής κατανομής των καταστάσεων, οι εξισώσεις ισορροπίας παίρνουν τη μορφή:

$$\begin{cases} \underline{\pi} = \underline{\pi} \cdot P \\ \underline{\pi} \cdot \underline{e} = 1 \end{cases}$$

Επομένως:

$$\begin{aligned}
&\pi_0 = \pi_0 p + \pi_1 p + \ldots + \pi_{m-1} p + \pi_m \\
&\pi_1 = \pi_0(1-p) \\
&\pi_2 = \pi_1(1-p) = \pi_0(1-p)^2 \\
&\pi_3 = \pi_2(1-p) = \pi_0(1-p)^3 \\
&\ \vdots \\
&\pi_i = \pi_{i-1}(1-p) = \pi_0(1-p)^i, \quad i = 1, 2, \ldots, m
\end{aligned} \tag{1}$$

και

$$\begin{aligned}
1 &= \pi_0 + \pi_1 + \cdots + \pi_m = \pi_0 + \pi_0(1-p) + \cdots + \pi_0(1-p)^m \\
&= \pi_0 \sum_{i=0}^{m} (1-p)^i = \pi_0 \frac{1-(1-p)^{m+1}}{1-(1-p)} = \pi_0 \frac{1-(1-p)^{m+1}}{p}
\end{aligned} \tag{2}$$

Άρα

$$\pi_0 = \frac{p}{1-(1-p)^{m+1}}$$

και αντικαθιστώντας στην (1) έχουμε τελικά:

$$\pi_i = \frac{p(1-p)^i}{1-(1-p)^{m+1}}, \quad i = 1, 2, \ldots, m.$$

γ') Μακροπρόθεσμα, η συχνότητα με την οποία αντικαθίσταται η μπάρα ισούται με τη συχνότητα επισκέψεων της κατάστασης 0 (όπου η μπάρα λειτουργεί για 0 μέρες, δηλαδή μόλις έχει αντικατασταθεί), που είναι

$$\pi_0 = \frac{p}{1-(1-p)^{m+1}}.$$

Καθώς $m \to \infty$, $\lim_{m\to\infty} \pi_0 = p$.

Με άλλα λόγια, για πολύ μεγάλη διάρκεια φυσικής ζωής της μπάρας ($m \to \infty$) (δηλαδή για καλή ποιότητα του εξαρτήματος), η συχνότητα αντικατάστασής της ισούται με την πιθανότητα p με την οποία ο χρήστης της προκαλεί ζημιά.

Άσκηση 3.
Ένα μηχάνημα στο καζίνο έχει ρυθμιστεί έτσι ώστε κάθε φορά που παίζει ένας παίκτης είτε κερδίζει είτε χάνει με μια πιθανότητα. Πιο συγκεκριμένα, αν σε ένα παίξιμο κερδίσει, τότε στο επόμενο παίξιμο έχει πιθανότητα να κερδίσει ξανά ίση με 0.25. Αντίστοιχα, αν σε ένα παίξιμο χάσει, τότε στο επόμενο έχει πιθανότητα να κερδίσει ίση με 0.75. Ο παίκτης πληρώνει 1€ για κάθε παίξιμο και εισπράττει 2.5€ κάθε φορά που κερδίζει, αλλιώς δεν παίρνει τίποτα. Βρείτε αν το παιχνίδι συμφέρει τον παίκτη μακροπρόθεσμα.

Λύση.
Έστω X_n το αποτέλεσμα του n-οστού παιχνιδιού, οπού $X_n = 1$ αν κερδίσει και $X_n = 0$ αν χάσει.

Εφόσον η πιθανότητα μετάβασης σε μια κατάσταση μπορεί να περιγραφεί πλήρως από την τελευταία κατάσταση, το πρόβλημα μπορεί να μοντελοποιηθεί ως μια Μαρκοβιανή αλυσίδα, με πίνακα μετάβασης

$$P = \begin{bmatrix} 0.25 & 0.75 \\ 0.75 & 0.25 \end{bmatrix}$$

Ο πίνακας μετάβασης έχει όλα του τα στοιχειά θετικά, επομένως η αλυσίδα περιέχει μια μοναδική κλάση, η οποία είναι απεριοδική. Επομένως, υπάρχει η στατική κατανομή της αλυσίδας και δίνεται από τις εξισώσεις ισορροπίας:

$$\left.\begin{array}{l} \underline{\pi} = \underline{\pi} \cdot P \\ \underline{\pi} \cdot \underline{e} = 1 \end{array}\right\} \Rightarrow \left.\begin{array}{l} \pi_0 = 0.25\pi_0 + 0.75\pi_1 \\ \pi_1 = 0.75\pi_0 + 0.25\pi_1 \\ \pi_0 + \pi_1 = 1 \end{array}\right\} \Rightarrow \pi_0 = \pi_1 = 0.5$$

Το καθαρό κέρδος στην κατάσταση 0 ισούται με -1€ και στην κατάσταση 1 ισούται με $+1.5$€

Επομένως, το αναμενόμενο κέρδος ανά παιχνίδι μετά από πολλά παιξίματα, ανεξάρτητα από το αποτέλεσμα του πρώτου παιξίματος, είναι $\pi_0 \cdot (-1) + \pi_1 \cdot (1.5) = 0.25 > 0$. Επομένως, μακροπρόθεσμα το παιχνίδι αποφέρει κέρδος.

Άσκηση 4.
Μια εταιρία κατασκευής κλιματιστικών υποβάλλει τα κλιματιστικά με τη σειρά που παράγονται σε τεχνικό έλεγχο ποιότητας και τα χαρακτηρίζει είτε ως 'τέλεια' είτε ως 'ελαττωματικά'. Έχει παρατηρηθεί ότι η ποιότητα κάθε κλιματιστικού εξαρτάται από την ποιότητα του αμέσως προηγούμενού του στη σειρά παραγωγής και ότι ένα 'τέλειο' κλιματιστικοί ακολουθείται από ένα επίσης 'τέλειο' κλιματιστικό με πιθανότητα $3/4$, ενώ ένα 'ελαττωματικό' ακολουθείται από ένα άλλο 'ελαττωματικό' με πιθανότητα $1/3$. Αν το τρίτο κλιματιστικό που εξετάζεται είναι 'τέλειο', ποια είναι η πιθανότητα να είναι 'τέλειο' το εικοστό;

Λύση.
Έστω X_n η ποιότητα του n-οστού κλιματιστικού, οπού $X_n = 0$ αν το κλιματιστικό είναι 'τέλειο' και $X_n = 1$ είναι 'ελαττωματικό'. Εφόσον οι πιθανότητες μετάβασης περιγράφονται πλήρως από την αμέσως προηγούμενη κατάσταση, το πρόβλημα μπορεί να μοντελοποιηθεί ως μια Μαρκοβιανή αλυσίδα.

Ο πίνακας μετάβασης είναι:

$$P = \begin{bmatrix} 3/4 & 1/4 \\ 2/3 & 1/3 \end{bmatrix} \doteq \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

Έχουμε δείξει ότι για τον πίνακα μετάβασης του n-οστού βήματος $P^{(n)}$ ισούται με τη n-οστή δύναμη του αρχικού πίνακα μετάβασης:

$$P^{(n)} = P^n \ , \quad n \geq 1 \ .$$

Μας ζητείται η πιθανότητα

$$P(X_{20} = 0|X_3 = 0) = P(X_{17} = 0|X_0 = 0) = (P^{(17)})_{00} = (P^{17})_{00} \ ,$$

όπου η δεύτερη ισότητα προκύπτει από τον ορισμό της Μαρκοβιανής αλυσίδας (υποθέτουμε ότι η αρχική μας κατάσταση είναι το τρίτο βήμα).

Θα υπολογίσουμε τον πίνακα P^n. Ξέρουμε ότι $P^n = U\Lambda^n U^{-1}$, όπου U είναι ο $2x2$ πίνακας με στήλες τα δεξιά ιδιοδιανύσματα του P και Λ είναι ο διαγώνιος πίνακας των αντίστοιχων ιδιοτιμών. Οι μη-μηδενικές ιδιοτιμές υπολογίζονται μηδενίζοντας την ορίζουσα:

$$\begin{aligned} &|P - \lambda I| = 0 \\ \Rightarrow &\begin{vmatrix} 1-a-\lambda & a \\ b & 1-b-\lambda \end{vmatrix} = 0 \\ \Rightarrow &(1-a-\lambda)(1-b-\lambda) - a \cdot b = 0 \\ \Rightarrow &\lambda^2 - (2-a-b)\lambda - a \cdot b = 0 \ . \end{aligned}$$

$$\Delta = (2-a-b)^2 - 4(1-a-b) = a^2 + b^2 + 2ab = (a+b)^2 \ .$$

$$\lambda_{1,2} = \frac{(2-a-b) \pm (a+b)}{2} \Rightarrow \begin{cases} \lambda_1 = 1 \\ \lambda_2 = 1-a-b \end{cases}$$

Ας συμβολίσουμε με u_1, u_2 τις δύο στήλες του U (τα ιδιοδιανύσματα του P). Για το u_1 με ιδιοτιμή λ_1 ξέρουμε ότι είναι ίσο με $u_1 = \mathbf{e} = [1,\ 1]^T$, αφού κάθε στοχαστικός πίνακας έχει ένα τέτοιο ζεύγος ιδιοτιμών-ιδιοδιανυσμάτων.

Για το u_2 έχουμε:

$$\begin{aligned}
Pu_2 = \lambda_2 u_2 &\Rightarrow \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \begin{bmatrix} u_{2(1)} \\ u_{2(2)} \end{bmatrix} = (1-a-b) \begin{bmatrix} u_{2(1)} \\ u_{2(2)} \end{bmatrix} \\
&\Rightarrow \begin{cases} (1-a)u_{2(1)} + au_{2(2)} = (1-a-b)u_{2(1)} \\ bu_{2(1)} + (1-b)u_{2(2)} = (1-a-b)u_{2(2)} \end{cases} \\
&\Rightarrow u_{2(1)} = -\frac{b}{a} u_{2(2)}
\end{aligned}$$

Έστω $u_{2(1)} = a \Rightarrow u_{2(2)} = -b$. Επομένως, το δεύτερο ιδιοδιάνυσμα είναι οποιοδήποτε πολλαπλάσιο του

$$u_2 = \begin{bmatrix} a \\ -b \end{bmatrix}$$

Άρα:

$$\begin{aligned}
U &= \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix} \\
\Rightarrow P^n &= \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-a-b)^n \end{bmatrix} \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix}^{-1} \\
\Rightarrow P^n &= \begin{bmatrix} 1 & a(1-a-b)^n \\ 1 & -b(1-a-b)^n \end{bmatrix} \begin{bmatrix} \frac{-b}{-a-b} & \frac{-a}{-a-b} \\ \frac{-1}{-a-b} & \frac{1}{-a-b} \end{bmatrix} \\
\Rightarrow P^n &= \begin{bmatrix} 1 & a(1-a-b)^n \\ 1 & -b(1-a-b)^n \end{bmatrix} \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{1}{a+b} & \frac{-1}{a+b} \end{bmatrix} \\
\Rightarrow P^n &= \begin{bmatrix} \frac{b+a(1-a-b)^n}{a+b} & \frac{a-a(1-a-b)^n}{a+b} \\ \frac{b-b(1-a-b)^n}{a+b} & \frac{a+b(1-a-b)^n}{a+b} \end{bmatrix}
\end{aligned}$$

Άρα

$$(P^{17})_{00} = \frac{b + a(1-a-b)^{17}}{a+b}$$

Αντικαθιστώντας τις τιμές $a = 1/4,\ b = 2/3$, προκύπτει ότι $(P^{17})_{00} \approx 0.7273$.