Ανάλυση Απόδοσης Πληροφοριακών Συστημάτων

Διάλεξεις 2-4: Διαδικασίες Markov Διακριτού Χρόνου

Δρ. Αθανάσιος Ν. Νικολακόπουλος

ΜΔΕ Επιστήμης και Τεχνολογίας Υπολογιστών

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

4 Νοεμβρίου 2016

Μοντέλα Markov Διακριτού Χρόνου I

Ένα μοντέλο αλυσίδας Markov (Markov Chain Model) ορίζεται από:

1. Ένα σύνολο καταστάσεων S
2. Ένα σύνολο των πιθανών μεταβάσεων, μεταξύ των καταστάσεων αυτών, δηλαδή τα (i, j) για τα οποία $p_{ij} > 0$ και
3. Τις τιμές των θετικών πιθανοτήτων μετάβασης, $p_{ij} > 0$.

Η ***Αλυσίδα Markov Διακριτού Χρόνου*** (στο εξής **DTMC**) που αντιστοιχεί στο μοντέλο είναι μία στοχαστική ακολουθία $\{X(n) \in \mathbf{S}, n \in \mathbb{Z}\}$ η οποία ικανοποιεί:

$$\Pr\{X(n+1) = j \mid X(n) = i, X(n-1) = i_{n-1}, \dots, X(0) = i_0\}$$

$$\xrightarrow[\text{Markov}]{\text{Ιδιότητα}} = \Pr\{X(n+1) = j \mid X(n) = i\} = p_{ij} \qquad (1)$$

Μία πεπερασμένη DTMC μπορεί να περιγραφεί πλήρως από το στοχαστικό μητρώο $\mathbf{P} \triangleq [p_{ij}]$ το οποίο καλείται **μητρώο πιθανοτήτων μετάβασης**.

Μοντέλα Markov Διακριτού Χρόνου II

Example (**’Ενας Απλός Server**)

- Αν είναι **Busy**, παραμένει **Busy** με πιθανότητα $1 - \alpha$.
- Αν είναι **Idle**, παραμένει **Idle** με πιθανότητα $1 - \beta$.

Ερωτήσεις

- Χώρος Κατάστασης?
- Μητρώο Πιθανοτήτων Μετάβασης?

Example (**Απλή Ουρά Διακριτού χρόνου**)

- Αφίξεις : $BP(p)$
- Εξυπηρέτηση : $Geometric(q)$
- Χωρητικότητα : K πελάτες
- Πολιτική Εξυπηρέτησης : $FCFS$

Ερωτήσεις

- Χώρος Κατάστασης?
- Διάγραμμα πιθανοτήτων μετάβασης?
- Μητρώο Πιθανοτήτων Μετάβασης?

Πιθανότητες n βημάτων I

Πρόβλημα: Θέλουμε να υπολογίσουμε: $p_{ij}^{(n)} \triangleq \Pr\{i \to j \text{ σε } n \text{ βήματα}\}$

Εξισώσεις Chapman-Kolmogorov

Η Λύση μπορεί να βρεθεί αναδρομικά:

$$p_{ij}^{(n)} = \sum_{k \in \mathbf{S}} p_{ik}^{(n-1)} p_{kj}, \forall i, j \tag{2}$$

η οποία μπορεί να γραφεί πιο συνοπτικά σε matrix form:

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)} \mathbf{P} \tag{3}$$

οπότε λύνοντας την αναδρομή:

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)} \mathbf{P} = \mathbf{P}^{(n-2)} \mathbf{P}^2 = \ldots = \mathbf{P}^{(1)} \mathbf{P}^{n-1} = \mathbf{P}^n \tag{4}$$

Πιθανότητες n βημάτων II

Έστω λ_i η i-οστή ιδιοτιμή και που αντιστοιχεί στο δεξιό ιδιοδιάνυσμα u_i, του Μητρώου Πιθανοτήτων Μετάβασης:

$$\mathbf{P}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \tag{5}$$

Θα εξετάσουμε για απλότητα την περίπτωση που το μητρώο πιθανοτήτων μετάβασης είναι διαγωνιοποιήσιμο. Τότε από τη γραμμική άλγεβρα γνωρίζουμε πως ισχύει:

$$\mathbf{P} = \mathbf{U}\Lambda\mathbf{U}^{-1} \equiv \mathbf{U}\Lambda\mathbf{V} \text{ όπου } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{bmatrix} \text{ και } \mathbf{V} \triangleq \mathbf{U}^{-1} \tag{6}$$

Οπότε, υψώνοντας το $\mathbf{P}$ στη n-οστή δύναμη θα έχουμε:

$$\mathbf{P}^n = \mathbf{U}\Lambda^n\mathbf{V} = \sum_{i \in S} \lambda_i^n \mathbf{u}_i \mathbf{v}_i^\top, \tag{7}$$

Πιθανότητες n βημάτων III

Συνεπώς η κατανομή n βημάτων μπορεί να εκφραστεί:

$$\mathbf{p}^\top(n) = \mathbf{p}^\top(0)\mathbf{P}^n = \sum_{k \in S} \lambda_k^n \mathbf{p}^\top(0)\mathbf{u_k}\mathbf{v}_k^\top \tag{8}$$

και η πιθανότητα η αλυσίδα Markov να είναι στην κατάσταση i τη χρονική στιγμή n δίνεται από:

$$p_i(n) = \sum_{k \in S} \lambda_k^n (\mathbf{p}^\top(0)\mathbf{u_k}) v_{ki}, i \in S. \tag{9}$$

Πιθανότητες n βημάτων IV

Example (Ένας Απλός **Server** συνέχεια...)

- Αν είναι **Busy**, παραμένει **Busy** με πιθανότητα $1 - \alpha$.
- Αν είναι **Idle**, παραμένει **Idle** με πιθανότητα $1 - \beta$.

$$\mathbf{P}^n_{server} = \begin{pmatrix} \frac{\beta+\alpha(1-\alpha-\beta)^n}{\alpha+\beta} & \frac{\alpha-\alpha(1-\alpha-\beta)^n}{\alpha+\beta} \\ \frac{\beta-\beta(1-\alpha-\beta)^n}{\alpha+\beta} & \frac{\alpha+\beta(1-\alpha-\beta)^n}{\alpha+\beta} \end{pmatrix}$$

Παρατήρηση:

$$\lim_{n\to\infty} \mathbf{P}^n_{server} = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$$

1. Υπάρχει μία **οριακή κατανομή** με την οποία βρισκόμαστε σε κάθε κατάσταση!
2. Η κατανομή αυτή είναι **ανεξάρτητη** της αρχικής κατάστασης!

Ερώτηση: Γιατί είναι σημαντικό αυτό?

Πότε Υπάρχει Οριακή Κατανομή?

$\$10^9$ Questions:

- Ποιές Ιδιότητες της Αλυσίδας εξασφαλίζουν ότι:
  1. Το $\lim_{n\to\infty} p_{ij}^{(n)}$ **υπάρχει** $\forall i, j$
  2. Το $\lim_{n\to\infty} p_{ij}^{(n)}$ είναι **ανεξάρτητο** του i
  3. Το $\lim_{n\to\infty} p_{ij}^{(n)}$ είναι **μη μηδενικό** $\forall j$
- Πώς σχετίζεται η "time-average" κατανομή: $\varpi_j \triangleq \lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t} \mathbb{I}_{[X(n)=j]}$ με την οριακή κατανομή.

Πριν επιχειρήσουμε να απαντήσουμε θα πρέπει να ορίσουμε κάποιες χρήσιμες ιδιότητες των καταστάσεων και των αλυσίδων.

Κατηγοριοποίηση Αλυσίδων Markov I

Ορισμός **Προσβασιμότητα και Επικοινωνία**

Μια κατάσταση $j \in S$ είναι **προσβάσιμη** από μια κατάσταση $i \in S$ αν υπάρχει ένας ακέραιος $n \geqslant 1$ τέτοιος ώστε $P_{ij}^{(n)} > 0$. Αν η κατάσταση i είναι προσβάσιμη από την κατάσταση j και η κατάσταση j είναι προσβάσιμη από την i, τότε οι καταστάσεις i και j λέγεται ότι **επικοινωνούν** το οποίο συμβολίζεται $i \leftrightarrow j$.

Πρόταση **Η Επικοινωνία είναι μια σχέση ισοδυναμίας**

- Ανακλαστική ιδιότητα: $i \leftrightarrow i$ για όλα τα $i \in S$.
- Συμμετρική ιδιότητα: αν $i \leftrightarrow j$, τότε $j \leftrightarrow i$.
- Μεταβατική ιδιότητα: αν $i \leftrightarrow j$ και $j \leftrightarrow k$, τότε $i \leftrightarrow k$.

Κατηγοριοποίηση Αλυσίδων Markov II

Ορισμοί **Κλειστά, Ανοιχτά και Αμείωτα σύνολα καταστάσεων**

Ένα σύνολο καταστάσεων C ονομάζεται **κλειστό** αν

$$P_{ij} = 0 \text{ για όλα τα } i \in C, j \notin C \tag{10}$$

αλλιώς, ονομάζεται **ανοιχτό**.

Ένα σύνολο C ονομάζεται **αμείωτο** αν

$$i \leftrightarrow j \text{ για όλα τα } i, j \in C \tag{11}$$

Αν όλες οι καταστάσεις S μιας αλυσίδας Markov είναι αμείωτες, τότε η αλυσίδα ονομάζεται **αμείωτη**. Αν ένα κλειστό σύνολο περιέχει ακριβώς μια κατάσταση, η κατάσταση αυτή ονομάζεται **απορροφητική**.

Κατηγοριοποίηση Αλυσίδων Markov III

Ορίζουμε $f_{ij}^{(n)}$ την πιθανότητα, ξεκινώντας από την i, το βήμα **πρώτης διάβασης** της κατάστασης j να είναι ίσο με n. Αυστηρότερα:

$$f_{ij}^{(n)} \triangleq \Pr\{X(0) = i, X(1) \neq j, \dots, X(n-1) \neq j, X(n) = j\}$$

Τότε η πιθανότητα επίσκεψης της κατάστασης j (**κάποια στιγμή**), ξεκινώντας από την i, θα είναι:

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \tag{12}$$

Ορισμός Επαναληπτική και Μεταβατική κατάσταση

Αν $f_{ii} = 1$, η κατάσταση i ονομάζεται **επαναληπτική**. Αν $f_{ii} < 1$ η κατάσταση ονομάζεται **μεταβατική**.

Κατηγοριοποίηση Αλυσίδων Markov IV

Μέσος Χρόνος Πρώτης Διέλευσης της i ξεκινώντας από την j: Συμβολίζεται με μ_{ij} και ισχύει:

$$\mu_{ij} = \begin{cases} \sum_{n=1}^{\infty} n f_{ij}^{(n)}, & \text{αν } f_{ij} = 1, \\ \infty, & \text{αν } f_{ij} < 1, \end{cases} \tag{13}$$

Όταν $f_{ij} = 1$, για ένα ζευγάρι καταστάσεων i και j, ο μέσος χρόνος πρώτης διάβασης μ_{ij} ικανοποιεί την αναδρομική εξίσωση

$$\mu_{ij} = p_{ij} + \sum_{k \neq j} p_{ik}(\mu_{kj} + 1), \tag{14}$$

Κατηγοριοποίηση Αλυσίδων Markov V

Όταν $j = i$, ο μέσος χρόνος πρώτης διάβασης ονομάζεται **Μέσος Χρόνος Επιστροφής στην Κατάσταση** i.

Ορισμός **Μηδενικά και Βέβαια επαναληπτικές καταστάσεις**

Μια επαναληπτική κατάσταση $i \in S$ ονομάζεται μια **μηδενικά επαναληπτική** αν $\mu_{ii} = \infty$. Αν $\mu_{ii} < \infty$, ονομάζεται μια **βέβαια επαναληπτική**

Παρατήρηση: Σε μια πεπερασμένη αλυσίδα Markov υπάρχουν μόνο βέβαια επαναληπτικές και μεταβατικές καταστάσεις.

Ορισμός **Περιοδική και Απεριοδική κατάσταση**

Μια κατάσταση $i \in S$ ονομάζεται **περιοδική** με περίοδο d_i αν $d_i > 1$ και **μη-περιοδική** αν $d_i = 1$, όπου

$$d_i = \gcd\{n : P_{ii}^{(n)} > 0\}. \tag{15}$$

Κατηγοριοποίηση Αλυσίδων Markov VI

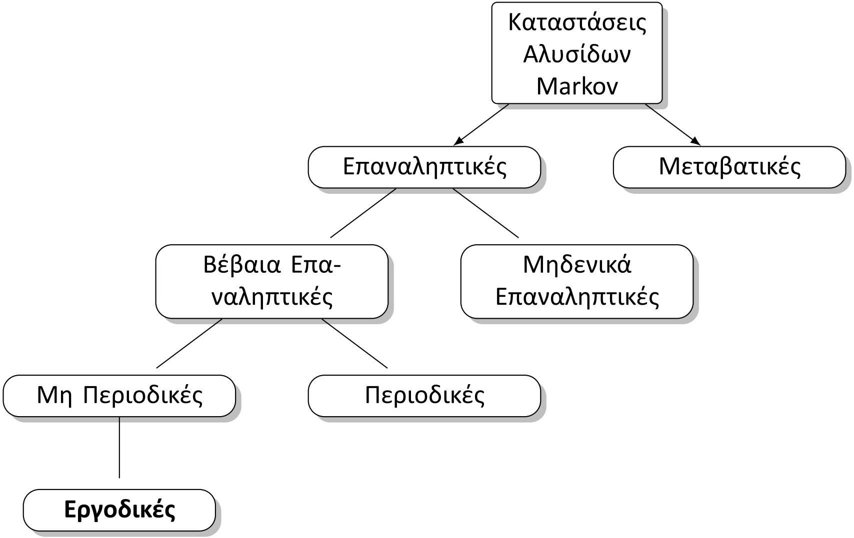
Ορισμός **Εργοδικές καταστασεις**

Μια κατάσταση $i \in S$ ονομάζεται **εργοδική** αν είναι βέβαια επαναληπτική και μη-περιοδική, δηλαδή, $f_{ii} = 1, \mu_{ii} < \infty$, και $d_i = 1$. Αν όλες οι καταστάσεις μίας αλυσίδας είναι εργοδικές, η αλυσίδα λέγετε, εργοδική.

Theorem (**Ιδιότητα μιας αμείωτης αλυσίδας Markov**)

Σε μια αμείωτη αλυσίδα Markov όλες οι καταστάσεις ανήκουν στην ίδια κλάση: είναι όλες μεταβατικές, όλες μηδενικά επαναληπτικές, ή όλες βέβαια επαναληπτικές. Επιπλέον, είναι είτε όλες μη-περιοδικές είτε όλες περιοδικές με ίδια περίοδο.

Κατηγοριοποίηση Αλυσίδων Markov VII



Κατανομές αλυσίδων Markov I

Ορισμός **Στάσιμη κατανομή**

Μία κατανομή π, καλείται στάσιμη κατανομή αν και μόνο αν, $\pi^\top = \pi^\top \mathbf{P}, \quad \sum \pi_i = 1.$

Θεώρημα **Στάσιμη κατανομή αμείωτης απεριοδικής αλυσίδας Markov**

Για μια αμείωτη μη-περιοδική αλυσίδα Markov η **οριακή κατανομή υπάρχει πάντα** και είναι ανεξάρτητη της αρχικής. Επίσης, ισχύει ένα από τα παρακάτω:

1. Οι καταστάσεις της είναι όλες μεταβατικές ή όλες μηδενικά επαναληπτικές. Στην περίπτωση αυτή, **δεν υπάρχει καμία στάσιμη κατανομή** και $\lim_{n\to\infty} p_{ij}^{(n)} = 0, \forall i, j$
2. Όλες οι καταστάσεις είναι εργοδικές. Στην περίπτωση αυτή, υπάρχει **μια μοναδική στάσιμη κατανομή** και $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$, για όλα τα $i, j \in S$

Επιπλέον, ισχύει $\pi_i = \frac{1}{\mu_{ii}}$

Υπολογισμός Στάσιμης Κατανομής Εργοδικών Αλυσίδων I

Η στάσιμη κατανομή μπορεί να υπολογιστεί:

- Λύνοντας το σύστημα $\pi^{\mathsf{T}} = \pi^{\mathsf{T}}\mathbf{P} \quad \sum \pi_i = 1$
- Λύνοντας τις "**εξισώσεις ισορροπίας**" $\sum_{j \neq i} \pi_i p_{ij} = \sum_{j \neq i} \pi_j p_{ji}$

Example (**Απλός Server: επίλυση**)

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \quad \text{και} \quad \pi_1 + \pi_2 = 1.$$

Για να βρούμε τη στάσιμη κατανομή λύνουμε το σύστημα και παίρνουμε $\pi_1 = \frac{\beta}{\alpha+\beta}$ και $\pi_2 = \frac{\alpha}{\alpha+\beta}$

ΠΡΟΣΟΧΗ! Όταν $\alpha = \beta = 1$ η αλυσίδα γίνεται περιοδική. Η Οριακή κατανομή δεν υπάρχει, αλλά η στάσιμη κατανομή υπάρχει και ταυτίζεται με την time-average κατανομή.

Υπολογισμός Στάσιμης Κατανομής Εργοδικών Αλυσίδων ΙΙ

Example (**Ουρά διακριτού χρόνου άπειρου μήκους: επίλυση**)

$$\mathbf{P} = \begin{pmatrix} 1-r & r & & \\ s & 1-r-s & r & \\ & s & 1-r-s & \ddots \\ \vdots & & \ddots & \ddots \end{pmatrix}$$

Ερώτηση Πότε υπάρχει οριακή κατανομή, και ποιά είναι?

Ερώτηση Ποιος είναι ο μέσος αριθμών εργασιών στο server?

Χρησιμοποιώντας τις εξισώσεις ισορροπίας έχουμε:

$$\begin{aligned} \pi_0 r &= \pi_1 s \implies \pi_1 = (r/s)\pi_0 \\ \pi_1(r+s) &= \pi_2 s + \pi_0 r \implies \pi_2 = (r/s)^2\pi_0 \\ &\vdots \\ \pi_k(r+s) &= \pi_{k+1}s + \pi_{k+1}r \implies \pi_k = (r/s)^k\pi_0 \\ &\vdots \end{aligned}$$

Υπολογισμός Στάσιμης Κατανομής Εργοδικών Αλυσίδων III

Example (**Ουρά διακριτού χρόνου άπειρου μήκους: επίλυση** συνέχεια...)

Συνεπώς, λαμβάνοντας υπόψη πως οι πιθανότητες αθροίζουν στο 1 και συμβολίζοντας με ρ το r/s:

$$\sum_{k=0}^{\infty} \rho^k \pi_0 = \pi_0 \frac{1}{1-\rho} = 1 \quad \Longrightarrow \quad \boxed{\pi_0 = 1-\rho, \quad \pi_k = (1-\rho)\rho^k}$$

και για τον μέσο αριθμό εργασιών στον server έχουμε:

$$\begin{aligned} \mathbb{E}[N] &= \sum_{k=0}^{\infty} k\pi_k = (1-\rho)\sum_{k=0}^{\infty} k\rho^k = (1-\rho)\rho\sum_{k=0}^{\infty} k\rho^{k-1} = (1-\rho)\rho\sum_{k=0}^{\infty} \frac{d}{d\rho}(\rho^k) \\ &= (1-\rho)\rho\frac{d}{d\rho}\sum_{k=0}^{\infty}(\rho^k) = (1-\rho)\rho\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho} \quad \Longrightarrow \quad \boxed{\mathbb{E}[N] = \frac{\rho}{1-\rho}} \end{aligned}$$

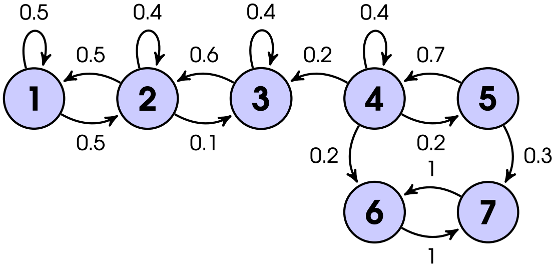
Άσκηση 1

Ένα Λειτουργικό Σύστημα ΛΣ ελέγχει μία CPU μοιράζοντας τον υπολογιστικό χρόνο της μεταξύ εργασιών χρηστών και εσωτερικών διεργασιών του ΛΣ. Ο υπολογιστικός χρόνος της CPU διαιρείται σε μικρά τμήματα σταθερού μήκους, τα slots.
Στο τέλος κάθε slot, η CPU θα συνεχίσει να εξυπηρετεί το ίδιο είδος εργασίας (χρηστών ή ΛΣ) για ένα ακόμα slot, με σταθερή πιθανότητα p_X αν εξυπηρετεί χρήστη, και σταθερή πιθανότητα p_Λ αν εξυπηρετεί εσωτερική διεργασία του ΛΣ.

1. Σχεδιάστε το διάγραμμα καταστάσεων – πιθανοτήτων μεταβάσεων της αλυσίδας Markov διακριτού χρόνου που μοντελοποιεί τη λειτουργία του παραπάνω συστήματος. Σε τι αντιστοιχούν οι καταστάσεις της αλυσίδας;
2. Παρουσιάστε τη λύση του συστήματος στη μόνιμη κατάσταση.
3. Μας δίνεται η απαίτηση ο μέσος χρόνος που η CPU εξυπηρετεί εργασίες χρηστών, να είναι το 90% του συνολικού υπολογιστικού χρόνου της, ενώ γνωρίζουμε ότι το p_Λ είναι ίσο με 0.2 . Ποια τιμή του p_X ικανοποιεί την απαίτηση αυτή;

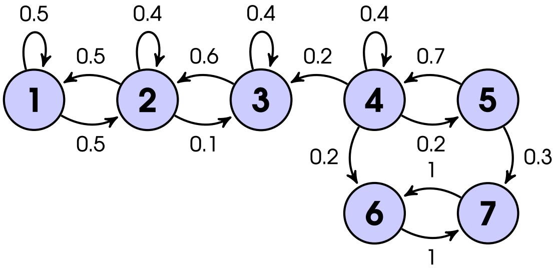
Άσκηση 2

Θεωρείστε την παρακάτω αλυσίδα.



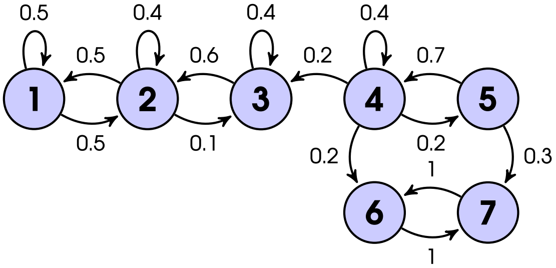
1. Να χαρακτηριστούν οι καταστάσεις της αλυσίδας.
2. Υπάρχουν πιθανότητες μόνιμης κατάστασης δεδομένου πως η αλυσίδα ξεκινάει από την κατάσταση 1. Αν ναι, ποιες είναι?
3. Υπάρχουν πιθανότητες μόνιμης κατάστασης δεδομένου πως η αλυσίδα ξεκινάει από την κατάσταση 6. Αν ναι, ποιες είναι?

Άσκηση 2 (ΙΙ)



4. Θεωρείστε ότι η αλυσίδα ξεκινάει στην κατάσταση 1, αλλά εμείς αρχίζουμε να την παρατηρούμε εφόσον αυτή έχει φθάσει σε ισορροπία.
    α. Να βρεθεί η υπό συνθήκη πιθανότητα η διαδικασία να ήταν στην κατάσταση 2 όταν αρχίσαμε να την παρατηρούμε, δεδομένου πως η κατάσταση αυξήθηκε κατά ένα κατά την πρώτη μετάβαση που παρατηρήσαμε.
    β. Να βρεθεί η πιθανότητα η κατάσταση να αυξήθηκε κατά ένα κατά τη διάρκεια της πρώτης αλλαγής κατάστασης που παρατηρήσαμε.

Άσκηση 2 (III)



5. Θεωρείστε ότι η αλυσίδα ξεκινάει στην κατάσταση 4.
    α. Για κάθε επαναληπτική κλάση, καθορίστε την πιθανότητα να φτάσουμε τελικά στη συγκεκριμένη κλάση.
    β. Ποιος είναι ο μέσος αριθμός μεταβάσεων μέχρι και την μετάβαση με την οποία φθάνουμε μία επαναληπτική κλάση για πρώτη φορά?