

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ  
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



## Σύζευξη και εφαρμογές

Δημήτρης Πετρόγιαννος

Επιβλέπων καθηγητής: κ. Λουλάκης Μιχαήλ

08/10/2018

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγικές έννοιες</b>	<b>6</b>
<b>2</b>	<b>Coupling-σύζευξη</b>	<b>7</b>
2.1	Σύζευξη και απόσταση ολικής κύμανσης . . . . .	10
2.2	Coupling και ρυθμός σύγκλισης στη στάσιμη κατανομή . . . . .	14
<b>3</b>	<b>Ο αλγόριθμος Metropolis</b>	<b>19</b>
<b>4</b>	<b>Coupling from the past</b>	<b>20</b>
4.1	Γιατί δεν επιλέγουμε coupling to the future . . . . .	33
4.2	Γιατί χρησιμοποιούμε τις ίδιες συνατήσεις που κλήθηκαν σε προηγούμενες επαναλήψεις . . . . .	34
4.3	Η μονότονη περίπτωση coupling from the past . . . . .	35
<b>5</b>	<b>Εφαρμογή στο μοντέλο του Ising</b>	<b>38</b>
<b>6</b>	<b>Χρόνος μίξης για το μοντέλο του Ising με τη προσομοίωση με δυναμική Glauber</b>	<b>42</b>
6.1	Η μετρική μεταφοράς . . . . .	43
6.2	Σύζευξη μονοπατιών . . . . .	45
6.3	Άνω όριο για το χρόνο μίξης του μοντέλου του Ising σε υψηλές θερμοκρασίες . . . . .	48
<b>7</b>	<b>Χρόνος μίξης για το τυχαίο περίπατο</b>	<b>51</b>
<b>8</b>	<b>Αριθμητικά πειράματα</b>	<b>57</b>
8.1	Μοντέλο Ising . . . . .	57
8.2	Τυχαίος περίπατος στο σύνολο $\{0, \dots, n\}$ . . . . .	57
8.3	Μοντέλο Ehrenfest . . . . .	59
<b>9</b>	<b>Κώδικες</b>	<b>61</b>

## Ευχαριστίες

Θα ήθελα να ευχαριστήσω πρώτα από όλα την οικογένεια μου για την καθημερινή στήριξη που μου παρείχε. Ακόμη, ευχαριστώ τον καθηγητή μου κ. Λουλάκη Μιχάλη που με εισήγαγε στον κλάδο του χρόνου μείξης των μαρκοβιανών αλυσίδων και για τη συνολική βοήθεια στην εκπόνηση της διπλωματικής. Τέλος, θα ήθελα να αφιερώσω την παρούσα εργασία στον παππού μου Γιώργο για όλα αυτά που πρόλαβε να μου μάθει.

## Εισαγωγή

Η στάσιμη κατανομή αποτελεί χαρακτηριστικό γνώρισμα της μαρκοβιανής αλυσίδας. Όταν η αλυσίδα μας φτάσει στην στάσιμη κατανομή τότε η πιθανότητα να βρεθεί σε οποιαδήποτε θέση κάθε στιγμή ταυτίζεται με τη στάσιμη κατανομή. Ένα πολύ χρήσιμο εργαλείο ώστε να εκτιμήσουμε πόσο μακριά είναι από τη στάσιμη κατανομή η μαρκοβιανή αλυσίδα καθώς και το ρυθμό με τον οποίο συγκλίνει σε αυτή είναι η σύζευξη. Η μέθοδος της σύζευξης είναι χρήσιμη και στην απόδειξη θεωρημάτων και προτάσεων εκτός από τον κλάδο της θεωρίας πιθανοτήτων.

Το 1996 οι James Propp και David Wilson δημιούργησαν έναν αλγόριθμο για δειγματοληψία από τη στάσιμη κατανομή. Σε αντίθεση με τον αλγόριθμο Metropolis-Hastings, η έξοδος του αλγορίθμου δίνει αποτέλεσμα με μηδενικό σφάλμα, όταν μπορεί να εφαρμοστεί.

Σκοπός της διπλωματικής είναι να μελετήσουμε τις ιδιότητες του αλγορίθμου καθώς και τη μονότονη περίπτωση του. Εάν η αλυσίδα μας φτάνει σχετικά γρήγορα στη στάσιμη κατανομή τότε ο αλγορίθμος μας μπορεί να εφαρμοστεί εύκολα. Τέλος, εφαρμόζουμε την ιδέα σε διάφορα μοντέλα και υπολογιστικά με τη γλώσσα Python.

## Abstract

The stationary distribution is a characteristic of a Markov chain, if it exists. When Markov chains reach stationarity then they move like fixed chains. One powerful technique in order to estimate how far is our Markov chain from stationarity and the rate of convergence is the coupling technique. Coupling arguments can be useful in proofing a wide variety of theorems except from probability theory.

Coupling from the past is a method of sampling from the stationary distribution of a Markov chain developed by James Propp and David Wilson in 1996. Unlike the Metropolis-Hastings algorithm, the Propp-Wilson algorithm, when it is applicable, outputs samples from stationarity with zero bias.

In this thesis we study the Propp-Wilson algorithm and its monotone case. If our Markov chain is rapidly mixing then we can apply our algorithm. We apply it to various models and run some simulations in Python.

# 1 Εισαγωγικές έννοιες

Θέτουμε  $P$  τον πίνακα μετάβασης μιας μαρκοβιανής αλυσίδας με πεπερασμένο σύνολο καταστάσεων  $\Omega$  και με την ακολουθία χαρακτήρων  $Pr$  θα συμβολίζουμε τη πιθανότητα.

Θα εισάγουμε κάποιες έννοιες από τη θεωρία των πεπερασμένων στοχαστικών ανελίξεων.

**Ορισμός 1** Μια μαρκοβιανή αλυσίδα είναι μη υποβιβάσιμη εάν  $\forall x, y \in \Omega$

$$\exists t > 0 \ Pr[X_t = y | X_0 = x] > 0$$

ή ισοδύναμα  $P^t[x, y] > 0$

Υπάρχει λοιπόν θετική πιθανότητα η αλυσίδα μας να βρεθεί απο οποιαδήποτε κατάσταση σε μια άλλη σε πεπερασμένο χρόνο.

**Ορισμός 2** Για κάθε κατάσταση  $x$  ορίζουμε το σύνολο δυνατών χρόνων επιστροφής ως

$$I(x) = \{t \geq 0 : P^t(x, x) > 0\}$$

Ως περίοδος της κατάστασης ορίζεται ως ο μέγιστος κοινός διαυρέτης του συνόλου δυνατών χρόνων επιστροφής.

Η κατάσταση λέγεται απεριοδική αν η περίοδος της είναι 1.

Σε μια μη υποβιβάσιμη αλυσίδα η περίοδος είναι κοινή για όλες τις καταστάσεις.

**Ορισμός 3** Στάσιμη κατανομή  $\pi$  μιας μαρκοβιανής αλυσίδας ορίζουμε τη κατανομή που ικανοποιεί την εξής σχέση

$$\pi = \pi P$$

Είναι εύκολο να δούμε ότι αν η αλυσίδα επιλέξει την αρχική της θέση σύμφωνα με την στάσιμη κατανομή τότε η πιθανότητα να βρεθεί σε οποιαδήποτε κατάσταση ακολουθεί πάλι την κατανομή  $\pi$ .

Αυτό συμβαίνει γιατί αν  $\mu_t$  η κατανομή κάθε χρονική στιγμή  $t$  τότε  $\mu_t = \mu_{t-1}P$  και επαγωγικά  $\mu_t = \mu_0 P^t$  και το προηγούμενο συμπέρασμα προκύπτει εύκολα χρησιμοποιώντας τον ορισμό της στάσιμης κατανομής.

Ένας εύκολος τρόπος να ελέγξουμε αν μια κατανομή είναι στάσιμη είναι μέσω των συνθηκών ακριβούς ισορροπίας

**Ορισμός 4** Η κατανομή  $\pi$  ικανοποιεί τις συνθήκες ακριβούς ισορροπίας σύμφωνα με το πίνακα  $P$  εάν

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

Αθροίζοντας ως προς όλα τα  $y \in \Omega$  προκύπτει

$$\pi(x) = \sum_{y \in \Omega} \pi(y)P(y, x) \iff \pi = \pi P$$

Ένα πολύ χρήσιμο λήμμα το οποίο ισχύει για μαρκοβιανές αλυσίδες θα χρησιμοποιήσουμε και αργότερα

**Λήμμα 1.1** Αν η  $X_n$  είναι μη υποβιβάσιμη και απεριοδική τότε  $\exists n_0 \in \mathbb{N}$  τέτοιο ώστε

$$P^n[x, y] > 0 \quad \forall n \geq n_0$$

**Πρόταση 1.1** Αν μια μαρκοβιανή αλυσίδα είναι μη υποβιβάσιμη με πεπερασμένο χώρο καταστάσεων τότε έχει μοναδική στάσιμη κατανομή.

**Πρόταση 1.2** Αν μια μαρκοβιανή αλυσίδα είναι μη υποβιβάσιμη και απεριοδική τότε συγκλίνει στη μοναδική στάσιμη κατανομή.

Αποδείξεις σε όσα αφήσαμε αναπόδεικτα μπορούν να βρεθούν σε κάθε βιβλίο στοχαστικών ανελίξεων. (Ενδεικτικά το βιβλίο [2])

## 2 Coupling-σύζευξη

**Ορισμός 5** Coupling-σύζευξη δυο κατανομών πιθανότητας  $\mu, \nu$  ορισμένων στον ίδιο χώρο πιθανότητας είναι ένα ζευγάρι τυχαίων μεταβλητών  $(X, Y)$  τέτοιων ώστε

$$Pr[X = x] = \mu(x) \text{ και } Pr[Y = y] = \nu(y)$$

Στην περίπτωση της σύζευξης δύο κατανομών πιθανότητας σκοπός μας είναι να κατασκευάσουμε μια δισδιάστατη τυχαία μεταβλητή ορισμένη στο χώρο γινόμενο που οι περιθώριες κατανομές τους είναι οι αρχικές.

Αντίστοιχα, υπάρχει ορισμός για τη σύζευξη μιας μαρκοβιανής αλυσίδας να είναι:

**Ορισμός 6** Coupling-Σύζευξη δύο μαρκοβιανών αλυσίδων που περιγράφονται από τον ίδιο πίνακα μετάβασης  $P$  είναι ένα ζευγάρι διαδικασιών  $(X_t, Y_t)_{t=0}^{\infty}$  οι οποίες  $X_t, Y_t$  έχουν ως πίνακα μετάβασης τον πίνακα  $P$ , και διαφορετικές

αρχικές κατανομές.

Πολλές φορές όταν θα δημιουργούμε συζεύξεις μαρκοβιανών αλυσίδων που περιγράφονται από τον πίνακα  $P$  θα απαιτούμε οι αλυσίδες να εκτελούν τις ίδιες κινήσεις από το χρόνο συνάντησής τους και μετά.

Αν συμβολίσουμε με  $\mathbb{1}_x, \mathbb{1}_y$  τις αρχικές κατανομές των τυχαίων μεταβλητών, δηλαδή  $X_0 = x$  και  $Y_0 = y$  και θέσουμε

$$t_{couple} := \min\{t : X_t = Y_t\}$$

την πρώτη φορά που οι αλυσίδες συναντιούνται τότε η διαδικασία σύζευξης της μαρκοβιανής αλυσίδας περιγράφεται:

$$\begin{array}{lll} t = 0 & X_0 = x & Y_0 = y \\ t < t_{couple} & X_t = X_{t-1}P & Y_t = Y_{t-1}P \\ t \geq t_{couple} & X_s = Y_s = X_{s-1}P & \end{array}$$

**Παράδειγμα 2.1** Θα μελετήσουμε την περίπτωση σύζευξης δύο κατανομών  $\mu, \nu$  που περιγράφουν το βάρος που δίνει ένα τίμιο νόμισμα στα δυο πιθανά αποτελέσματα, που τα συμβολίζουμε  $[0, 1]$ .

Εμείς θέλουμε να κατασκευάσουμε ένα μέτρο  $q(x, y) = Pr[X = x, Y = y]$  στο χώρο  $\Omega \times \Omega$  με την παρακάτω ιδιότητα:

$$\begin{aligned} \sum_{y \in \Omega} q(x, y) &= \sum_{y \in \Omega} Pr[X = x, Y = y] = Pr[X = x] = \mu(x) \\ \sum_{x \in \Omega} q(x, y) &= \sum_{x \in \Omega} Pr[X = x, Y = y] = Pr[Y = y] = \nu(y) \end{aligned}$$

Μια λύση των παραπάνω σχέσεων στη περίπτωση μας είναι

$$q(x, y) = 1/4 \quad \forall (x, y) \in [0, 1]^2$$

Μια δεύτερη επιλογή είναι η:

$$q(x, y) = \begin{cases} \frac{1}{2} & \text{εάν } (x, y) = (0, 0), (x, y) = (1, 1) \\ 0 & \text{εάν } (x, y) = (0, 1), (x, y) = (1, 0) \end{cases}$$

Μία πρώτη παρατήρηση είναι ότι για τη σύζευξη δυο κατανομών δεν υπάρχει μοναδική επιλογή και εύκολα μπορούμε να δείξουμε ότι υπάρχει πάντα σύζευξη όπου για κάθε  $x, y \in \Omega$  επιλέγουμε στη διδιάστατη κατανομή ανεξάρτητα από τις περιθώριες

$$q(x, y) = \mu(x)\nu(y) \quad \forall x, y \in \Omega$$

Στη συνέχεια, θα παρουσιάσουμε δύο προτάσεις που η αποδείξη τους προκύπτει εύκολα χρησιμοποιώντας τη μέθοδο της σύζευξης



**Πρόταση 2.1** Έστω ο συμμετρικός τυχαίος περίπατος στο σύνολο  $\{1, \dots, N\}$ .  
Για  $x \leq y$  ισχύει

$$P^t(x, n) \leq P^t(y, n)$$

### Απόδειξη

Έστω  $\{X_t\}, \{Y_t\}$  δύο αλυσίδες που ξεκινούν από τις καταστάσεις  $x, y$  αντίστοιχα, οι οποίες εκτελούν και αυτές το τυχαίο περίπατο στο σύνολο  $\{1, \dots, N\}$ .

Καλούμε  $\Delta_t$  μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών που παίρνουν τις τιμές  $\{-1, 1\}$  με πιθανότητα  $1/2$  τη κάθε μία.

Επιλέγουμε κάθε χρονική στιγμή  $t$  οι αλυσίδες μας να εκτελούν το ίδιο βήμα ανάλογα με τη τιμή της μεταβλητής  $\Delta_t$ . Αν  $\Delta_t = +1$  τότε οι αλυσίδες κινούνται δεξιότερα ενώ στην αντίθετη περίπτωση κινούνται προς τα αριστερά.

Εάν κάποια χρονική στιγμή συναντηθούν τότε οι πορείες τους θα ταυτίζονται από το χρόνο αυτό και μετά.

Είναι προφανές πως η κατανομή της αλυσίδας  $X_t$  είναι η  $P^t(x, \cdot)$  ενώ της  $Y_t$  είναι η  $P^t(y, \cdot)$ .

Αφού  $x \leq y$  και το επόμενο τους βήμα εξαρτάται από τον ίδιο παράγοντα, συνεπάγεται ότι  $X_t \leq Y_t$ . Επιπλέον, αν  $X_t = n$  τότε  $Y_t = n$

$$P^t(x, n) = Pr\{X_t = n\} \leq Pr\{Y_t = n\} = P^t(y, n)$$



**Πρόταση 2.2** Έστω τυχαία μεταβλητή με τιμές στο  $\mathbb{R}$ . Αν οι συναρτήσεις  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  είναι και οι δύο αύξουσες ή και οι δύο φθίνουσες, τότε

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)]$$

### Απόδειξη

Καλούμαι  $Y$  μια τυχαία μεταβλητή ανεξάρτητη και ισόνομη με τη τυχαία μεταβλητή  $X$ .

Παρατηρούμε ότι αν οι  $f, g$  έχουν το ίδιο είδος μονοτονίας τότε για κάθε  $x, y \in \mathbb{R}$

$$(f(x) - f(y))(g(x) - g(y)) \geq 0$$

Συνεπώς,

$$\begin{aligned} 0 &\leq E(f(X) - f(Y))(g(X) - g(Y)) \\ &= E[f(X)g(X)] + E[f(Y)g(Y)] - E[f(X)g(Y)] - E[f(Y)g(X)] \\ &\quad (\text{λόγω ανεξαρτησίας των τυχαίων μεταβλητών } X, Y) \\ &= E[f(X)]E[g(X)] + E[f(Y)g(Y)] - E[f(X)]E[g(Y)] - E[f(Y)]E[g(X)] \\ &\quad (X, Y \text{ ισόνομες} \rightarrow E[h(X)] = E[h(Y)]) \\ &= 2(E[f(X)g(X)]) - E[f(X)]E[g(X)] \end{aligned}$$

## 2.1 Σύζευξη και απόσταση ολικής κύμανσης

Μια πάρα πολύ χρήσιμη ιδιότητα της σύζευξης μαρκοβιανών αλυσίδων είναι ότι μπορούμε να τη χρησιμοποιήσουμε ώστε να εκτιμήσουμε το ρυθμό με τον οποίο συγκλίνει η αρχική μαρκοβιανή αλυσίδα στη στάσιμη κατανομή. Για να μετρήσουμε το ρυθμό που συγκλίνουν οι αλυσίδες στη στάσιμη κατανομή χρειάζεται να ορίσουμε πρώτα μια έννοια απόστασης.

**Ορισμός 7** Ορίζουμε απόσταση ολικής κύμανσης (*total variation distance*) δύο κατανομών που ορίζονται στον ίδιο χώρο πιθανότητας  $\Omega$

$$\|\mu - \nu\| = \max_{A \subset \Omega} |\mu(A) - \nu(A)|$$

Υπάρχει σύνδεση μεταξύ της σύζευξης δύο κατανομών στον ίδιο χώρο πιθανότητας και της απόστασης ολικής κύμανσης.

### Πρόταση 2.3

$$\begin{aligned}\|\mu - \nu\| &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| \\ &= \sum_{x \in \Omega, \mu(x) \geq \nu(x)} [\mu(x) - \nu(x)] \\ &= \sum_{y \in \Omega, \nu(y) \geq \mu(y)} [\nu(y) - \mu(y)]\end{aligned}$$

#### Απόδειξη

Θέτουμε  $B = \{x : \mu(x) > \nu(x)\}$ . Έστω  $A \subset \Omega$

$$\begin{aligned}\mu(A) - \nu(A) &= \mu(A \cap B) + \mu(A \cap B^c) - \nu(A \cap B) - \nu(A \cap B^c) \\ &\leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B)\end{aligned}$$

Ομοίως,

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c)$$

Όμως

$$\mu(B) + \mu(B^c) = \nu(B) + \nu(B^c)$$

Άρα επιλέγοντας  $A = B$  ή  $A = B^c$  η διαφορά των δύο μέτρων λαμβάνει τη μέγιστη τιμή της

$$\|\mu - \nu\| = \sum_{x \in B} |\mu(x) - \nu(x)| = \sum_{y \in B^c} |\mu(y) - \nu(y)|$$

και επειδή  $B \cup B^c = \Omega$

$$\|\mu - \nu\| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$$

### Θεώρημα 2.1

$$\|\mu - \nu\| = \inf[Pr[X \neq Y] : (X, Y) \text{ είναι σύζευξη των } \mu, \nu]$$

#### Απόδειξη

Έστω  $x \in \Omega$  και  $X, Y$  σύζευξη των  $\mu, \nu$

$$\begin{aligned}\mu(x) - \nu(x) &= Pr[X = x] - Pr[Y \neq x] \\ &= Pr[X = x, Y = x] + Pr[X = x, Y \neq x] - Pr[Y = x, X \neq x] - Pr[Y = x, X = x] \\ &\leq Pr[X = x, Y \neq x] - Pr[Y = x, X \neq x]\end{aligned}$$

$$|\mu(x) - \nu(x)| \leq Pr[X = x, Y \neq x] + Pr[Y = x, X \neq x]$$

$$\|\mu - \nu\| \leq Pr[X \neq Y]$$

$$\|\mu - \nu\| \leq \inf[Pr[X \neq Y] : (X, Y) \text{ είναι σύζευξη των } \mu, \nu]$$

Αρκεί, πλέον, να βρούμε μια σύζευξη που η πιθανότητα οι δυο τυχαίες μεταβλητές να διαφέρουν να είναι ίση με την απόσταση των κατανομών. Θέτουμε

$$p = \sum_{x \in \Omega} \mu(x) \wedge \nu(x)$$

και παρατηρούμε ότι

$$\sum_{x \in \Omega} \mu(x) \wedge \nu(x) = \sum_{x \in \Omega, \mu(x) \leq \nu(x)} \mu(x) + \sum_{x \in \Omega, \mu(x) > \nu(x)} \nu(x) \Rightarrow$$

$$\sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - \sum_{x \in \Omega, \mu(x) > \nu(x)} \mu(x) - \nu(x)$$

$$\sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\| = p$$

Τώρα αποφασίζουμε να στρίψουμε ένα νόμισμα με πιθανότητα εμφάνισης κεφαλής ίσης με  $p$  και ανάλογα με το αποτέλεσμα του στριψίματος θα ενεργούμε διαφορετικά

1. Εάν το αποτέλεσμα του στριψίματος είναι κορώνα τότε διαλέγουμε μια τιμή  $Z$  με πιθανότητα

$$\gamma_{III}(x) = \frac{\mu(x) \wedge \nu(x)}{p}$$

και θέτουμε  $X = Y = Z$

2. Εάν το αποτέλεσμα του στριψίματος είναι γράμματα τότε διαλέγουμε τη τυχαία μεταβλητή  $X$  να παίρνει τη τιμή  $x$  με πιθανότητα

$$\gamma_I(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|} & \text{εάν } \mu(x) > \nu(x) \\ 0 & \text{αλλιώς} \end{cases}$$

και για τη τυχαία μεταβλητή  $Y$  να παίρνει τη τιμή  $x$  με πιθανότητα

$$\gamma_{II}(x) = \begin{cases} \frac{\nu(x) - \mu(x)}{\|\mu - \nu\|} & \text{εάν } \nu(x) > \mu(x), \\ 0 & \text{αλλιώς} \end{cases}$$

Σε κάθε περίπτωση

$$\begin{aligned}p\gamma_{III}(x) + (1 - p)\gamma_I(x) &= \mu \\p\gamma_{III}(x) + (1 - p)\gamma_{II}(x) &= \nu\end{aligned}$$

και η κατανομή των τυχαίων μεταβλητών  $X, Y$  που ορίσαμε είναι  $\mu, \nu$  αντίστοιχα.

Η μόνη περίπτωση να διαφέρουν οι δύο τυχαίες μεταβλητές  $X \neq Y$  από την ανάλυση που κάναμε είναι όταν το αποτέλεσμα του στριψίματος είναι γράμματα και πραγματοποιείται με πιθανότητα:

$$Pr[X \neq Y] = 1 - p = \|\mu - \nu\|$$

## 2.2 Coupling και ρυθμός σύγκλισης στη στάσιμη κατανομή

Αφού δώσαμε συνθήκες ύπαρξης μοναδικότητας και σύγκλισης στη στάσιμη κατανομή μιας μαρκοβιανής αλυσίδας, θα ήταν πολύ χρήσιμο να μπορούσαμε να εκτιμήσουμε το ρυθμό με τον οποίο οι αλυσίδες τείνουν να ισορροπήσουν. Το βασικό θεώρημα είναι το εξής.

**Ορισμός 8** Καλούμε χρόνο σύζευξης- $t_{couple}$  τη χρονική στιγμή που δύο μαρκοβιανές αλυσίδες που είναι σύζευξη του ίδιου μαρκοβιανού πίνακα  $P$  και είναι κατασκευασμένες ώστε όταν συναντηθούν και μετά να εξελίσσονται μαζί.

Όταν ασχολούμαστε με το ρυθμό σύγκλισης στη στάσιμη κατανομή στις συζεύξεις που καλούμε απαιτούμε να ταυτίζονται από το χρόνο σύζευξης και μετά.

**Θεώρημα 2.2** Έστω  $\{X_t, Y_t\}$  δύο αλυσίδες οι οποίες αποτελούν τη σύζευξη μαρκοβιανών αλυσίδων που περιγράφονται από τον πίνακα μετάβασης  $P$  και αρχικές συνθήκες  $X_0 = x, Y_0 = y$ . Τότε

$$\|P^t(x, \cdot) - P^t(y, \cdot)\| \leq Pr_{x,y}\{X_t \neq Y_t\}$$

### Απόδειξη

Παρατηρούμε ότι  $P^t(x, z) = Pr_x\{X_t = z\}$  και  $P^t(y, z) = Pr_y\{Y_t = z\}$ .

Από το θεώρημα 2.1, που αποδεικνύει ότι η απόσταση ολικής κύμανσης δυο κατανομών είναι το ελάχιστο άνω φράγμα πάνω στη ποσότητα  $Pr[X_t \neq Y_t]$

$$\|P^t(x, \cdot) - P^t(y, \cdot)\| \leq Pr_{x,y}\{X_t \neq Y_t\}$$

Όμως από τις ιδιότητες της σύζευξης

$$Pr_{x,y}\{X_t \neq Y_t\} = Pr_{x,y}\{t_{couple} > t\}$$

Για να μελετήσουμε διεξοδικότερα το θέμα ορίζουμε δύο ειδικές περιπτώσεις της απόστασης ολικής κύμανσης.

Η πρώτη αφορά την απόσταση της αλυσίδας από τη στάσιμη κατανομή τη χρονική στιγμή  $t$

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|$$

Αν  $d(t) = 0.25$  σημαίνει πως από όπου και να ξεκινήσει η αλυσίδα μας τη χρονική στιγμή  $t$  θα απέχει στη χειρότερη περίπτωση 25% από τη στάσιμη κατανομή. Πολλές φορές είναι δυνατό να φράξουμε την ποσότητα  $\|P^t(x, \cdot) - P^t(y, \cdot)\|$  ομοιόμορφα πάνω σε όλα τα ζεύγη  $(x, y)$ . Ορίζουμε

$$\bar{d}(t) := \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|$$

Με τις έννοιες που εισηγάγαμε γίνεται προφανές

#### Πρόταση 2.4

$$d(t) \leq \max_{x,y \in \Omega} Pr_{x,y} \{t_{couple} > t\}$$

**Παράδειγμα 2.2** Αν θεωρήσουμε τη μαρκοβιανή αλυσίδα  $X_t$  που καταγράφει το ελάχιστο μεταξύ ενός φυσικού αριθμού  $n$  και των συνεχόμενων επιτυχημένων στριψιμάτων με αποτέλεσμα κορώνα ενός τίμιου νομίσματος, θα δώσουμε ένα παράδειγμα πως μπορούμε να εκτιμήσουμε την απόσταση από τη στάσιμη κατανομή.

Για  $n = 6$

$$\begin{aligned} t : 001001 & X_t = 1 \\ t + 1 : 010011 & X_{t+1} = 2 \\ t + 2 : 100110 & X_{t+2} = 0 \end{aligned}$$

Ξεκινώντας δύο αλυσίδες που ξεκινούν από  $a \neq b \in \{0, \dots, n\}$

$$Pr[t_{couple} > t] \leq 2^{-t} \Rightarrow d(t) \leq 2^{-t}$$

Ισχύει η εξής ανισότητα για τις δυο ποσότητες που ορίσαμε προηγουμένως:

#### Λήμμα 2.1

$$d(t) \leq \bar{d}(t) \leq 2d(t)$$

**Απόδειξη** Για το δεύτερο μέρος της ανισότητας

$$\begin{aligned} \|P^t(x, \cdot) - P^t(y, \cdot)\| &= \frac{1}{2} \sum_{z \in \Omega} |P^t(x, z) - P^t(y, z)| \\ &= \frac{1}{2} \sum_{z \in \Omega} |P^t(x, z) - P^t(y, z) \pm \pi(z)| \leq \\ &\frac{1}{2} \sum_{z \in \Omega} |P^t(x, z) - \pi(z)| + \frac{1}{2} \sum_{z \in \Omega} |P^t(y, z) - \pi(z)| \\ \Rightarrow \|P^t(x, \cdot) - P^t(y, \cdot)\| &\leq \|P^t(x, \cdot) - \pi\| + \|P^t(y, \cdot) - \pi\| \\ &\Rightarrow \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\| \leq d(t) + d(t) \\ &\Rightarrow \bar{d}(t) \leq 2d(t) \end{aligned}$$

Για να δείξουμε την πρώτη ανισότητα χρειάζεται να εκμεταλλευτούμε ότι η  $\pi$  είναι στάσιμη κατανομή  $\pi = \pi P$ .

$$\pi(A) = \sum_{y \in \Omega} \pi(y) P^t(y, A)$$

Συνεπώς,

$$\begin{aligned} \|P^t(x, \cdot) - \pi\| &= \max_{A \subset \Omega} |P^t(x, A) - \pi(A)| \\ &= \max_{A \subset \Omega} \left| \sum_{y \in \Omega} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \end{aligned}$$

Εφαρμόζοντας τριγωνική ανισότητα

$$\|P^t(x, \cdot) - \pi\| \leq \sum_{y \in \Omega} \pi(y) \max_{A \subset \Omega} |P^t(x, A) - P^t(y, A)| = \bar{d}(t) \Rightarrow$$

$$d(t) \leq \bar{d}(t)$$

Μια πολύ σημαντική ιδιότητα για την απόστασης ολικής κύμανσης  $\bar{d}$  είναι η εξής.

**Λήμμα 2.2**  $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$

**Απόδειξη**

Σταθεροποιούμε δύο καταστάσεις  $x, y$  και θέτουμε  $(X_s, Y_s)$  τη βέλτιστη σύζευξη των κατανομών  $P^s(x, \cdot), P^s(y, \cdot)$ .

Δηλαδή,  $\|P^s(x, \cdot) - P^s(y, \cdot)\| = Pr[X_s \neq Y_s]$

$$P^{s+t}(x, w) = \sum_z P^s(x, z)P^t(z, w) = \sum_z Pr[X_s = z]P^t(z, w) = E(P^t(X_s, w))$$

Ομοίως,  $P^{s+t}(y, w) = E(P^t(Y_s, w))$

$$\begin{aligned} P^{s+t}(x, w) - P^{s+t}(y, w) &= E(P^t(X_s, w)) - E(P^t(Y_s, w)) \\ &= E(P^t(X_s, w) - P^t(Y_s, w)) \end{aligned}$$

$$\begin{aligned} \|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\| &= \frac{1}{2} \sum_w |P^{s+t}(x, w) - P^{s+t}(y, w)| \\ &= \frac{1}{2} \sum_w |E(P^t(X_s, w) - P^t(Y_s, w))| \\ &\leq \frac{1}{2} \sum_w E|P^t(X_s, w) - P^t(Y_s, w)| \\ &= E\left(\frac{1}{2} \sum_w |P^t(X_s, w) - P^t(Y_s, w)|\right) \\ &= E(\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|) \end{aligned}$$



Η ποσότητα  $\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|$  είναι μηδέν όταν  $X_s = Y_s$  και είναι φραγμένη από  $\bar{d}(t)$

$$\|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\| \leq \bar{d}(t)E(\mathbb{1}_{X_s \neq Y_s}) = \bar{d}(t)Pr[X_s \neq Y_s]$$

και επειδή η σύζευξη  $(X_s, Y_s)$  είναι βέλτιστη

$$\begin{aligned} \|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\| &\leq \|P^s(x, \cdot) - P^s(y, \cdot)\| \bar{d}(t) \\ &\Rightarrow \bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t) \end{aligned}$$

Συνέπεια του προηγούμενου λήμματος είναι

$$d(ct) \leq \bar{d}(ct) \leq \bar{d}(t)^c \quad \forall c \in \mathbb{N}$$

που προκύπτει εύκολα παρατηρώντας επαγωγικά

$$\bar{d}(ct) = \bar{d}(t + (c-1)t) \leq \bar{d}(t)\bar{d}((c-1)t) \leq \bar{d}(t)^c$$

και από τη σχέση  $d(t) \leq \bar{d}(t)$ .

Μια άλλη σημαντική ιδιότητα της απεικόνισης  $d(t)$  είναι πως σε κάθε βήμα μειώνεται

**Θεώρημα 2.3** Η απεικόνιση  $d(t)$  είναι φθίνουσα

**Απόδειξη**

Θα δείξουμε ότι  $d(t+1) \leq d(t)$

$$\begin{aligned} \|P^{t+1}(x, \cdot) - \pi\| &= \frac{1}{2} \sum_{z \in \Omega} |P^{t+1}(x, z) - \pi(z)| = \frac{1}{2} \sum_{z \in \Omega} |P^{t+1}(x, z) - \pi P(z)| \\ &= \frac{1}{2} \sum_{z \in \Omega} \left| \sum_{y \in \Omega} P^t(x, y)P(y, z) - \sum_{y \in \Omega} \pi(y)P(y, z) \right| \\ &= \frac{1}{2} \sum_{z \in \Omega} \left| \sum_{y \in \Omega} P(y, z)(P^t(x, y) - \pi(y)) \right| \\ &\leq \frac{1}{2} \sum_{y \in \Omega} \sum_{z \in \Omega} P(y, z) |P^t(x, y) - \pi(y)| \\ &= \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| = \|P^t(x, \cdot) - \pi\| \end{aligned}$$

Παίρνοντας το *maximum* και στις δύο μέλη της ανισότητας πάνω σε όλα τις δυνατές καταστάσεις προκύπτει το ζητούμενο. Παρατηρούμε πως σε κάθε βήμα η αλυσίδα ολοένα και πλησιάζει στη στάσιμη κατανομή.

**Ορισμός 9** Καλούμε  $\varepsilon$ -χρόνο μείξης

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}$$

και χρόνο μείξης

$$t_{mix} := t_{mix}(1/4)$$

Επιλέγουμε στο χρόνο μείξης η παράμετρος  $\varepsilon$  να είναι ίση με  $1/4$  γιατί για τυχαίο ακέραιο  $l$

$$d(lt_{mix}(\varepsilon)) \leq \bar{d}(lt_{mix}(\varepsilon)) \leq \bar{d}(t_{mix}(\varepsilon))^l \leq (2\varepsilon)^l$$

$$\text{Για } \varepsilon = \frac{1}{4} \quad d(lt_{mix}) \leq 2^{-l}$$

$$d(lt_{mix}) \leq 2^{-l}$$

$$t_{mix}(\varepsilon) \leq \log_2 \varepsilon^{-1} t_{mix}$$

Υπολογίζοντας το  $\varepsilon$ -χρόνο μείξης γνωρίζουμε τον ελάχιστο χρόνο που χρειάζεται η αλυσίδα μας ξεκινώντας από οποιαδήποτε κατάσταση ώστε να απέχει στη χειρότερη περίπτωση 100% από τη στάσιμη κατανομή.

Μία δεύτερη παρατήρηση είναι πως εάν τρέξουμε την αλυσίδα για πολλαπλάσιο χρόνο του  $t_{mix}$  τότε η απόσταση ολοένα και πέφτει με γεωμετρικό ρυθμό.

**Παράδειγμα 2.3** Στην αλυσίδα του παραδείγματος 2.2 δείξαμε πως  $d(t) \leq 2^{-t}$ . Αντικαθιστώντας στο τύπο για την εκτίμηση για το χρόνο μείξης προκύπτει

$$t_{mix}(\varepsilon) \leq \log_2(\varepsilon^{-1})$$

### 3 Ο αλγόριθμος Metropolis

Ο αλγόριθμος Metropolis είναι πιο γνωστός αλγόριθμος για να προσομειώσουμε μια αλυσίδα με δεδομένη στάσιμη κατανομή. Γνωρίζοντας τη πιθανότητα να μεταβούμε από την κατάσταση  $x$  στη κατάσταση  $y$ , τη συμβολίζουμε με  $r(x, y)$ , δημιουργούμε μια καινούργια μαρκοβιανή αλυσίδα με πίνακα μετάβασης  $Q(x, y)$ . Εάν ο αρχικός πίνακας  $P$  είναι συμμετρικός τότε ορίζεται ο πίνακας  $Q$ .

$$Q(x, y) = \begin{cases} r(x, y)a(x, y) & \text{εάν } x \neq y \\ 1 - \sum_{z: z \neq x} r(x, z)a(x, z) & \text{εάν } x = y \end{cases}$$

όπου  $a(x, y)$  είναι η πιθανότητα αποδοχής της μετάβασης από την κατάσταση  $x$  στη κατάσταση  $y$ .

Ο πίνακας  $Q$  έχει στάσιμη κατανομή την  $\pi$  εάν ικανοποιεί τις εξισώσεις ακριβούς ισορροπίας, δηλαδή, εάν

$$\pi(x)r(x, y)a(x, y) = \pi(y)r(y, x)a(y, x)$$

και γιατί ο πίνακας είναι συμμετρικός

$$\pi(x)a(x, y) = \pi(y)a(y, x)$$

Αφού εκφράζουν πιθανότητες αρκεί το σύστημα να ικανοποιεί τις δύο παρακάτω σχέσεις.

$$\begin{aligned} \pi(x)a(x, y) &\leq \pi(x) \\ \pi(x)a(x, y) &= \pi(y)a(y, x) \leq \pi(y) \end{aligned}$$

Και επειδή δεν είναι ωφέλιμο να απορρίπτουμε συχνά ενδεχόμενες κινήσεις θα προσπαθήσουμε να μεγιστοποιήσουμε την πιθανότητα  $a(x, y)$ . Έτσι επιλέγουμε

$$a(x, y) = \min(1, \pi(y)/\pi(x))$$

Στη περίπτωση που ο πίνακας μετάβασης δεν είναι συμμετρικός η πρώτη σχέση μας οδηγεί στην επιλογή

$$a(x, y) = \min(1, \frac{\pi(y)r(y, x)}{\pi(x)r(x, y)})$$

Και στις δύο περιπτώσεις η επιλογή του πίνακα μετάβασης ικανοποιεί τις εξισώσεις ακριβούς ισορροπίας σύμφωνα με τη κατανομή  $\pi$  και είναι προφανώς απεριοδική και μη υποβιβάσιμη εφόσον ο αρχικός πίνακας  $P$  που περιγράφει τη

μαρκοβιανή αλυσίδα είναι και αυτός απεριοδκός και μη υποβιβάσιμος. Εμείς για να πάρουμε δείγματα απο τη κατανομή  $\pi$  τρέχουμε τον παραπάνω αλγόριθμο για ενα μεγάλο αριθμό επαναλήψεων ελπίζοντας οτι η αλυσίδα μας έχει φτάσει στην στάσιμη κατανομή.

## 4 Coupling from the past

Πως θα μας φαινόταν η ιδέα να δημιουργήσουμε τόσες αλυσίδες όσες το πλήθος καταστάσεων και να της αφήσουμε να εξελιχθούν μέχρι να συναντηθούν-να έρθουν σε σύζευξη. Το αποτέλεσμα που θα λάβουμε φαίνεται να προκύπτει απο τη στάσιμη κατανομή.

Πόσο μακρυνά είναι όμως, στην πραγματικότητα ο παραπάνω ισχυρισμός ώστε να μπορούμε να πάρουμε δείγματα σύμφωνα με τη στάσιμη κατανομή χωρίς μεροληψία.

Μια λύση στο πρόβλημα μας να λάβουμε δείγματα απο τη στάσιμη κατανομή με πιθανότητα 1 είναι μια ιδέα που υλοποιήθηκε απο τους James Propp και David Wilson το 1996 και ονομάζεται *coupling from the past* ή αλγόριθμος Propp-Wilson. Η ιδέα βασίζεται στο να τρέξουμε  $|\Omega|$  το πλήθος αλυσίδες που ξεκινά η κάθε μία απο μία κατάσταση μέχρι να έρθουν σε σύζευξη. Η μόνη διαφορά είναι πως ο χρόνος ξεκινά από το παρελθόν. Δηλαδή ξεκινάμε για μεγάλο χρόνο  $t < 0$  και τη χρονική στιγμή 0 ελέγχουμε αν έχει επιτευχθεί η σύζευξη. Αν όχι τότε ξεκινάμε πάλι τη διαδικασία για  $t' < t < 0$ .

Θα πρέπει να δείξουμε ότι ο αλγόριθμος μας τελειώνει σε πεπερασμένο χρόνο καθώς και να βρούμε μια εύκολη συνθήκη τερματισμού που θα μας βοηθάει να λαμβάνουμε δείγματα από τη στάσιμη κατανομή.

Θα εισάγουμε κάποιες έννοιες για να μπορέσουμε να αποδείξουμε όλες τις ιδιότητες του αλγορίθμου.

Θα προσπαθήσουμε να δημιουργήσουμε έναν τρόπο να αναπαραστήσουμε την εξέλιξη  $|\Omega|$  το πλήθος αλυσίδων που περιγράφονται από το πίνακα  $P$  και κάθε μία ξεκινά από διαφορετική κατάσταση.

**Ορισμός 10** Θεώρημα αναπαράστασης μέσω τυχαίας συνάρτησης (*Random function representation*) για τον πίνακα  $P$

Συμβολίζουμε  $\Omega^\Omega$  το σύνολο των συναρτήσεων από το  $\Omega^{|\Omega|} \rightarrow \Omega^{|\Omega|}$  και  $F$  τη τυχαία συνάρτηση με την ιδιότητα

$$Pr_F[F(x) = y] = P(x, y) \quad \forall x, y \in \Omega$$

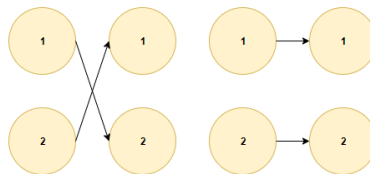
Στην τυχαία συνάρτηση  $F$  υπάρχουν συναρτήσεις  $f$  της μορφής

$$f = \begin{cases} f(x_1) = x_{j_1} \\ f(x_2) = x_{j_2} \\ \dots\dots \\ f(x_{|\Omega|}) = x_{j_{|\Omega|}} \end{cases}$$

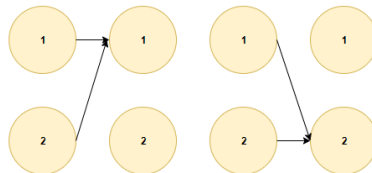
με  $j_i \in \{1, 2, \dots, |\Omega|\}$ . Η πιθανότητα να διαλέξουμε μια από όλες τις συναρτήσεις που απεικονίζει την κατάσταση  $x$  στην κατάσταση  $y$  είναι ίση με  $P(x, y)$ .

**Παράδειγμα 4.1** Έστω ότι έχουμε τον διπλανό πίνακα  $P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$

**1η δυνατή περίπτωση**

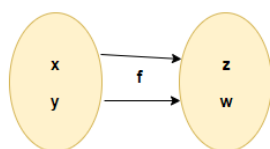


**2η δυνατή περίπτωση**



Σε κάθε περίπτωση επιλέγουμε ένα από τα γραφήματα με πιθανότητα  $1/2$ . Οι απεικονίσεις που δημιουργήσαμε ικανοποιούν τις απαιτήσεις του θεωρήματος αναπαράστασης μέσω τυχαίας συνάρτησης. Μια πρώτη παρατήρηση είναι πως δεν υπάρχει μοναδική επιλογή που να ικανοποιεί τις απαιτήσεις του θεωρήματος. Με το παραπάνω ορισμό, στο προηγούμενο παράδειγμα, παρατηρούμε πως αν έχουμε ένα διάνυσμα  $(x, y)$  με στοιχεία από το χώρο καταστάσεων, τότε επιλέγοντας για κάθε κατάσταση ξεχωριστά μια συνάρτηση από το θεωρήματος αναπαράστασης μέσω τυχαίας συνάρτησης και οι επιλογές αυτές είναι ανεξάρτητες τότε μπορούμε να προσομειώσουμε την εξέλιξη τους για ένα βήμα.

$$(x, y) \xrightarrow{f=(f_x, f_y)} (f_x(x) = z, f_y(y) = w)$$



Ένας τρόπος να δημιουργούμε τέτοια γραφήματα που ικανοποιούν τις απαιτήσεις του *random map representation* είναι με τα δέντρα διάγραμματα.

Στο πρώτο επίπεδο, στις κορυφές που συνδέονται με τη ρίζα, δημιουργούμε  $|\Omega|$  το πλήθος διαφορετικές κορυφές. Κάθε μία κορυφή περιέχει τις συναρτήσεις που  $f(1) = j, j = 1, \dots, |\Omega|$  και πιθανότητα μετάβασης ίση με  $P(1, j)$ .

Στο δεύτερο επίπεδο, για κάθε προηγούμενη κορυφή που υπάρχει τη συνδέουμε με άλλες  $|\Omega|$  νέες κορυφές. Κάθε νέα κορυφή κληρονομεί τις ιδιότητες των κορυφών με τις οποίες συνδέεται και έχει την επιπλέον ιδιότητα  $f(2) = k, k \in \{1, \dots, |\Omega|\}$  ανάλογα με τη θέση-ύψος της κορυφής. Μπορούμε να φανταστούμε κάθε κορυφή σαν έναν κύκλο που περιλαμβάνει τις ιδιότητες που έχουν οι συναρτήσεις που περιγράφει. Για παράδειγμα, αν βρίσκεται στην τρίτη θέση από τον πρώτο κύκλο-κορυφή που έχουν κοινό πατέρα, όπως βλέπουμε το γράφημα από πάνω προς τα κάτω, τότε η κορυφή που μελετάμε κληρονομεί όλες τις ιδιότητες από τον πατέρα της και έχει την επιπλέον ιδιότητα  $f(2) = 3$   $k \in \{1, \dots, |\Omega|\}$  με πιθανότητα σύνδεσης με το πατέρα της ίση με  $P(2, 3)$ . Συνεπώς, στο δεύτερο επίπεδο βρίσκονται όλες οι συναρτήσεις με την ιδιότητα  $f : f(1) = j, f(2) = k, j, k \in \{1, \dots, |\Omega|\}$  και πιθανότητα μετάβασης από τον πατέρα στον εαυτό τους ίση με  $P(2, k)$

Η διαδικασία συνεχίζεται μέχρι να έχουμε φτιάξει  $|\Omega|$  το πλήθος επίπεδα.

Πρέπει να δείξουμε ότι η πιθανότητα να διαλέξουμε μία από τις  $|\Omega|^{|\Omega|}$  συναρτήσεις που δημιουργήσαμε ικανοποιεί τις απαιτήσεις του *random map representation*. Αν παρατηρήσουμε σωστά στο δέντρο διάγραμμα σε κάθε επίπεδο  $i$  καθορίζεται που απεικονίζει η συνάρτησή μας το στοιχείο  $i$ .

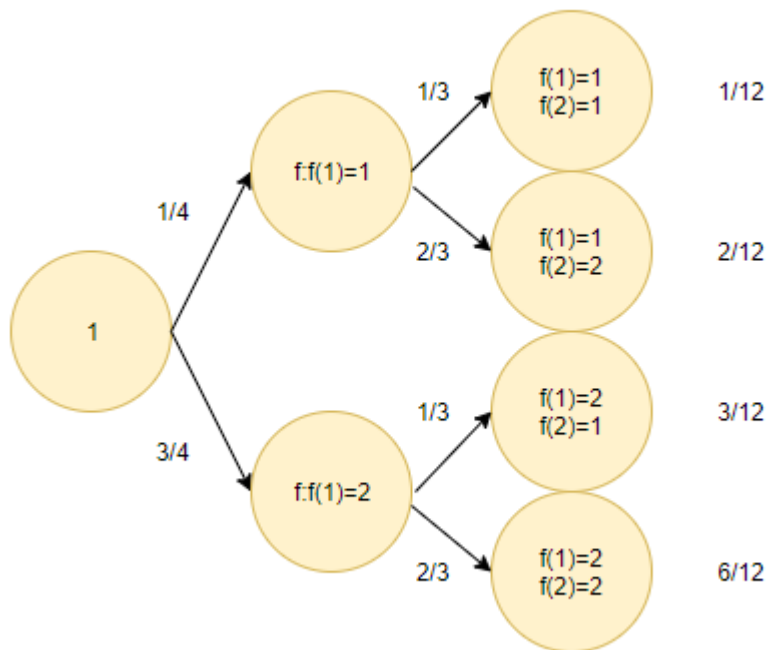
Σε κάθε επίπεδο ξεκινάει μια καινούργια μαρκοβιανή αλυσίδα που εξελίσσεται όπως μια μαρκοβιανή αλυσίδα που περιγράφεται σύμφωνα με τον πίνακα και αρχική κατανομή  $\mathbb{1}_i$ .

Για να υπολογίσουμε την πιθανότητα να παρατηρήσουμε μια συνάρτηση που απεικονίζεται στα τελικά φύλλα του δέντρου διαγράμματος είναι η πιθανότητα σε κάθε επίπεδο  $i$  να ξεκινάει μια μαρκοβιανή αλυσίδα που

- Περιγράφεται από τον πίνακα  $P$ .
- Έχει αρχική κατανομή  $\mathbb{1}_i$ .

και είναι ανεξάρτητη από τις προηγούμενες μαρκοβιανές αλυσίδες.  
 Συνεπώς  $Pr[f(x) = y] = P(x, y)$

**Παράδειγμα 4.2** Έστω  $P = \begin{bmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{bmatrix}$



Κάθε τελικός κόμβος-συνάρτηση έχει πιθανότητα να επιλεγεί ίση με εκείνη που βλέπουμε δεξιά από κάθε κόμβο.

Ορίζουμε

$$F_{t_i}^{t_j} = f_{t_j-1} \circ f_{t_j-2} \dots \circ f_{t_i}$$

τη σύνθεση συναρτήσεων  $f_i$  από τη χρονική στιγμή  $t_i$  μέχρι  $t_j$  όπου  $f_t$  συμβολίζει την επιλογή μία συνάρτησης όπως ορίζει το θεώρημα αναπαράστασης μέσω τυχαίας συνάρτησης για τον πίνακα  $P$ .

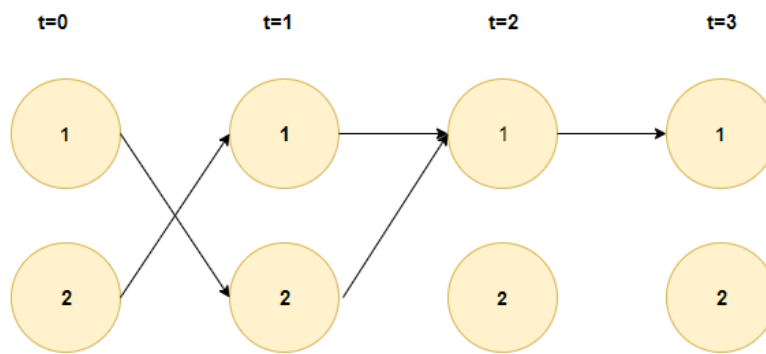
**Παράδειγμα 4.3** Χρησιμοποιώντας τα δεδομένα από το παράδειγμα 4.2

$$f_0 = \begin{cases} f(1) = 2 \\ f(2) = 1 \end{cases}$$

$$f_1 = \begin{cases} f(1) = 1 \\ f(2) = 1 \end{cases}$$

$$f_2 = \begin{cases} f(1) = 1 \\ f(2) = 2 \end{cases}$$

Το αποτέλεσμα της  $F_0^3 = f_2 \circ f_1 \circ f_0$  είναι  $X_3 = (1, 1)$



Στις προσομιώσεις θα χρησιμοποιήσουμε την έννοια της συνάρτησης ανανέωσης.

**Ορισμός 11** Συνάρτηση ανανέωσης

Συνάρτηση ανανέωσης ενός πίνακα μετάβασης  $P$  είναι μια απεικόνιση  $\phi : \Omega \times \mathbb{R} \rightarrow \Omega$

$$X^{t+1} = \phi(X^t, U^{t+1}) \text{ εάν } P(x, y) = Pr(\phi(x, U^{t+1}) = y)$$

όπου  $U^{t+1}$  αριθμός με κατανομή ομοιόμορφη στο  $[0,1]$ . Γνωρίζοντας την κατάσταση της αλυσίδας τη χρονική στιγμή  $t$  και επιλέγοντας έναν αριθμό, ομοιόμορφα καταμεμημένο στο  $[0,1]$ , μεταβαίνουμε στην επόμενη κατάσταση.

**Παράδειγμα 4.4** Τυχαίος περίπατος στο κανονικό  $n$ -γωνο

$$\phi(i, U^t) = \begin{cases} i+1 & \text{εάν } U^t \leq 1/2 \\ i-1 & \text{εάν } U^t > 1/2 \end{cases}$$



Μια πολύ χρήσιμη τεχνική για να προσομοιώνουμε την χρονική εξέλιξη των μαρκοβιανών αλυσίδων, συνήθως υπολογιστικά, είναι γνωρίζοντας τον πίνακα μετάβασης  $P$  να δημιουργούμε την παρακάτω συνάρτηση ανανέωσης:

$$\phi(x_i, r) = \begin{cases} x_1 & \text{εάν } x \in [0, P_{i,1}) \\ x_2 & \text{εάν } x \in [P_{i,1}, P_{i,1} + P_{i+2}) \\ \cdot & \cdot \\ x_j & \text{εάν } x \in \left[ \sum_{l=1}^{j-1} P_{i,l}, \sum_{l=1}^j P_{i,l} \right) \\ \cdot & \cdot \\ x_n & \text{εάν } x \in \left[ \sum_{l=1}^{n-1} P_{i,l}, 1 \right] \end{cases}$$

Η συνάρτηση που ορίσαμε είναι μια έγκυρη συνάρτηση ανανέωσης και περιγράφει το πίνακα μετάβασης  $P$  γιατί για κάθε  $x_i, x_j$  με  $i, j \in \{1, \dots, n\}$

$$Pr[\phi(x_i, r) = x_j] = \sum_{l=1}^j P_{i,l} - \sum_{l=1}^{j-1} P_{i,l} = P_{i,j}$$

Δηλαδή αν ξέρουμε την αρχική θέση, έστω  $x_i$ , δημιουργούμε το μερικό άθροισμα της γραμμής του πίνακα μετάβασης που περιγράφει τη πιθανότητα μετάβασης από τη θέση  $x_i$  στις υπόλοιπες θέσεις και ανάλογα με τον αριθμό  $r$  επιλέγουμε το επόμενο βήμα σύμφωνα με το πίνακα.

Θα αναφέρουμε πρώτα τον αλγόριθμο παρουσιάζοντας δυο ισοδύναμες ανα-  
 παρστάσεις και στη συνέχεια θα τον εξηγήσουμε αναλυτικά.

**Result:** Δειγματοληψία από τη στάσιμη κατανομή  
 $t=0;$   
 $F_t^0$  = vector with all the possible starting states  
**while**  $F_t^0 \neq \text{constant}$  **do**  
 | sample random  $f_t \sim F$   
 |  $F_{t-1}^0 = F_t^0 \circ f_t$   
 |  $t=t-1$   
**end**  
 return the unique state in the range of  $F_t^0$

**Algorithm 1:** Coupling from the past with random map representation

**Result:** Δειγματοληψία από τη στάσιμη κατανομή  
 $t=0;$   
 $F_t^0$  = vector with all the possible starting states  
**while**  $F_t^0 \neq \text{constant}$  **do**  
 | pick a random number  $U^{t+1} \sim U[0,1]$   
 |  $X^{t+1} = \phi(X^t, U^{t+1})$   
 |  $F_{t-1}^0 = F_t^0 \circ f_t$   
 |  $t=t-1$   
**end**  
 return the unique state in the range of  $F_t^0$

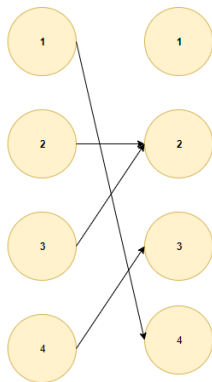
**Algorithm 2:** Coupling from the past with the update function

Αρχικά, δημιουργούμε τόσες αλυσίδες όσες και το πλήθος των στοιχείων του συνόλου των καταστάσεων της αλυσίδας και ξεκινούν από την κάθε κατάσταση ξεχωριστά. Η πρώτη αλυσίδα ξεκινά από τη  $x_1$ , η δεύτερη από τη  $x_2$  κ.ο.κ μέχρι να εξαντληθεί το πλήθος των δυνατών καταστάσεων.



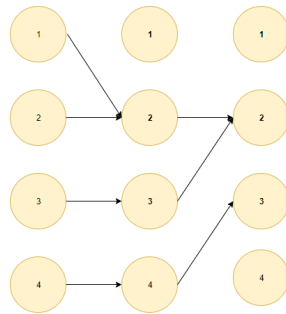
Σχήμα 1: Παράδειγμα με 4 καταστάσεις

Για  $t=-1$  αφήνουμε κάθε αλυσίδα να εξελιχθεί σύμφωνα με την αρχική αλυσίδα και ελέγχουμε αν όλες έχουν συναντηθεί στην ίδια κατάσταση.



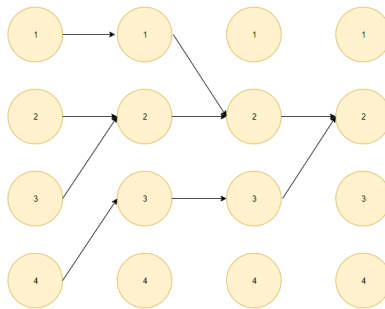
Σχήμα 2:  $t=-1$

Αν δεν έχουν συναντηθεί τότε ξεκινάμε από τη χρονική στιγμή  $t = -2$  να εξελιχθούν σύμφωνα με το πίνακα μετάβασης με τη μόνη υποχρέωση, για χρόνους που έχουμε τρέξει τον αλγόριθμο να διατηρούμε τις επιλογές από τις προηγούμενες επαναλήψεις. Αυτό επιτυγχάνεται αποθηκεύοντας με κάποιο τρόπο τις επιλογές των  $f$  ή των αριθμών  $u$ . Η διαδικασία συνεχίζεται πηγαίνοντας πίσω στο χρόνο μέχρι να επιτευχθεί η



Σχήμα 3:  $t=-2$

ταύτιση όλων των αλυσίδων διατηρώντας τα βήματα της αλυσίδας που έχουν πραγματοποιηθεί σε χρόνους που έχουμε τρέξει τον αλγόριθμο. Στο παράδειγμα



Σχήμα 4:  $t=-3$

που βλέπουμε στα σχήμα για  $t=-3$  ο αλγόριθμος τερματίζει και στην έξοδο βλέπουμε την κατάσταση 2.

Χρειάζεται να εξασφαλίσουμε ότι ο αλγόριθμος τερματίζει σε πεπερασμένο χρόνο και το αποτέλεσμα που προκύπτει ανήκει στη στάσιμη κατανομή. Ακόμη πρέπει να δικαιολογήσουμε την επιλογή να ξεκινάμε στο χρόνο από το παρελθόν αλλά και την επιλογή να μην επιτέπουμε στην αλυσίδα να εξελιχθεί ανεξάρτητα σε κάθε αποτυχημένη εκτέλεση του αλγορίθμου.

**Θεώρημα 4.1** Ο αλγόριθμος CFTP τερματίζει με πιθανότητα 1

**Απόδειξη**

Αφού η μαρκοβιανή αλυσίδα είναι μη υποβιβάσιμη και απεριοδική με πεπερασμένο χώρο καταστάσεων υπάρχει ένας ακέραιος αριθμός που θα το συμβολίζουμε με  $L > 0$  τέτοιος ώστε  $P^L(x, y) > 0 \quad \forall x, y \in \Omega$ . Δηλαδή, για την μας αλυσίδα υπάρχει ένας ακέραιος για τον οποίο ξεκινώντας από οποιαδήποτε κατάσταση μπορεί να φτάσει σε οποιαδήποτε άλλη με θετική πιθανότητα.

Άρα εάν έχουμε επιλέξει κατάλληλες συναρτήσεις για να περιγράψουμε την εξέλιξη των αλυσίδων ώστε να είναι συνεπής με τον πίνακα μετάβασης αλλά και να δίνουν θετικό βάρος στο να συναντηθούν σε πεπερασμένο χρόνο, έστω  $T$ , τότε

$$Pr[F_{-T}^0 \text{ είναι σταθερή}] = Pr[F_{-2L}^{-L} \text{ είναι σταθερή}] = \dots = q$$

Η πιθανότητα να μην τελειώσει ο αλγόριθμος είναι μετά από  $kL$  βήματα είναι να μην είναι σταθερή η απεικόνιση  $F$ .

$$(1 - q)^k$$

γιατί είναι ανεξάρτητες οι επιλογές των απεικονίσεων που επιλέγουμε.

Η πιθανότητα να μην τελειώσει σε πεπερασμένο χρόνο είναι  $\lim_{k \rightarrow \infty} (1 - q)^k = 0$ .

Συνεπώς, η ζητούμενη πιθανότητα να είναι ίση με 1.

Θα δικαιολογήσουμε τι εννοούμε κατάλληλη συνάρτηση.

**Παράδειγμα 4.5** Έστω  $P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$

Αν επιλέξουμε για συνάρτηση ανανέωσης την

$$\phi(1, r) = \begin{cases} 1 & \text{εάν } r \in [0, 0.5) \\ 2 & \text{εάν } r \in [0.5, 1] \end{cases}$$

$$\phi(2, r) = \begin{cases} 2 & \text{εάν } r \in [0, 0.5) \\ 1 & \text{εάν } r \in [0.5, 1] \end{cases}$$

και όπως χρειάζεται ο αλγόριθμος χρησιμοποιούμε τους ίδιους αριθμούς για να ανανεώνουμε τις δύο αλυσίδες, που η μία ξεκινά από το 1 και η άλλη από το 2, με τον παραπάνω τρόπο οι αλυσίδες δε θα συναντηθούν ποτέ και ο αλγόριθμος δε θα τελειώσει ποτέ.

Αν όμως χρησιμοποιήσουμε

$$\phi(x, r) = \begin{cases} 1 & \text{εάν } r \in [0, 0.5) \\ 2 & \text{εάν } r \in [0.5, 1] \end{cases}$$

όπου  $x = 1$  ή  $2$  τότε ο αλγόριθμος τελειώνει με ένα βήμα εκτέλεσης.  
 Με τη μέθοδο του δέντρου διαγράμματος και του μερικού συσσωρευτικού α-  
 θροίσματος υπάρχει πάντα θετική πιθανότητα να συναντηθούν οι πορείες των  
 αλυσίδων.

Χρειάζεται να βρούμε ένα εύκολο τρόπο

**Πρόταση 4.1** Έστω  $t_1 < t_2$ . Εάν  $F_{t_1}^{t_2}$  είναι σταθερή απεικόνιση τότε για  
 $t_1 < t_2 < t_3 < t_4$  και η  $F_{t_1}^{t_4}$  είναι σταθερή.

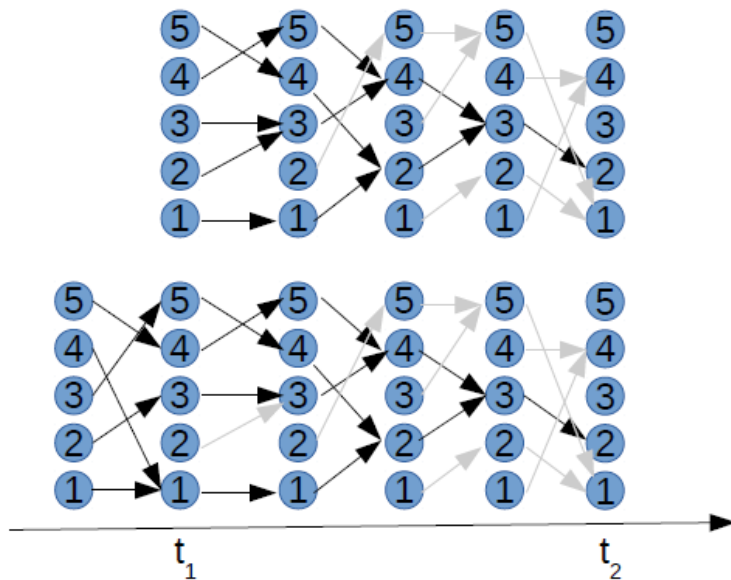
**Απόδειξη**

Από το τρόπο που έχουμε ορίσει την  $F$

$$F_{t_1}^{t_4} = F_{t_3}^{t_4} \circ F_{t_2}^{t_3} \circ F_{t_1}^{t_2}$$

Αφού το πρώτο όρισμα της σύνθεσης είναι σταθερό το αποτέλεσμα της σύνθεσης  
 όλων των απεικονίσεων είναι και αυτό κάποια σταθερή απεικόνιση όχι απαραίτητα  
 ίδια με τη πρώτη που προκύπτει

Εάν ,όμως , κινούμαστε στο χρόνο όπως περιγράφει ο αλγόριθμος τότε το α-  
 ποτέλεσμα της σύνθεσης παραμένει το ίδιο όσο πιο μακριά απο το παρελθόν  
 ξεκινήσουμε.



**Πρόταση 4.2** Εάν  $F_{-M}^0 = c$  τότε  $\forall -t < -M$  ισχύει  $F_{-t}^0 = c$

**Απόδειξη**  $F_{-t}^0 = F_{-M}^0 \circ F_{-t}^{-M} = c$

**Πρόταση 4.3**  $\forall t \in \mathbb{N}$  και  $\forall y \in \Omega$   $Pr[F_0^t = y] = Pr[F_{-t}^0 = y]$

**Απόδειξη** Η απεικόνιση  $F_{t_1}^{t_2}$  δεν εξαρτάται από το χρόνο αφού σε κάθε βήμα επιλέγουμε ανεξάρτητα μία από τις  $f$ .

Αυτο είναι και το κλειδί για να δείξουμε ότι το αποτέλεσμα του αλγορίθμου μας είναι από τη στάσιμη κατανομή της αλυσίδας.

**Ορισμός 12** Ορίζουμε  $M = \min\{t > 0 : F_{-M}^0 \text{ είναι σταθερή}\}$

**Θεώρημα 4.2**  $F_{-M}^0 \sim \pi$

**Απόδειξη**

Συνδυάζοντας τα δύο παραπάνω θεωρήματα θα προκύψει εύκολα το ζητούμενο.

$$\begin{aligned} Pr[F_{-M}^0 = y] &= \lim_{t \rightarrow \infty} Pr[F_{-t}^0 = y] \\ &= \lim_{t \rightarrow \infty} Pr[F_0^t = y] \\ &= \pi(y) \end{aligned}$$

Ακόμη, μπορούμε να δείξουμε και κάτι πιο όμορφο.

Το αποτέλεσμα του αλγορίθμου μας κατανέμεται ακριβώς όπως η στάσιμη κατανομή.

**Θεώρημα 4.3** Έστω  $Y$  η έξοδος του αλγορίθμου. Τότε

$$Pr[Y = s] = \pi(s)$$

**Απόδειξη**

Έστω  $\varepsilon > 0$  και  $s \in \Omega$ .

Αρκεί να δείξουμε ότι  $|Pr[Y = s] - \pi(s)| \leq \varepsilon$

Έστω  $X$  μια αλυσίδα που επιλέγει την αρχική της θέση σύμφωνα με τη στάσιμη κατανομή και  $M > 0$  :

$$Pr[\text{ο αλγόριθμος τελειώνει σε λιγότερο από } M \text{ βήματα}] \geq 1 - \varepsilon$$

Αυτός ο αριθμός που καλέσαμε υπάρχει αφού ο αλγόριθμος μας τελειώνει σε πεπερασμένο αριθμό επαναλήψεων με πιθανότητα 1.

Έστω  $\tilde{Y} = F_{-M}^0$  η έξοδος του αλγορίθμου τη χρονική στιγμή  $M$ .

Αρχίζουμε να τρέχουμε τον αλγόριθμο και για την αλυσίδα  $X$  εφαρμόζοντας τις ίδιες συναρτήσεις ανανέωσης.

Εάν ο αλγόριθμος σταματήσει σε λιγότερα από  $M$  βήματα τότε  $\tilde{Y} = X$ .

Άρα για κάθε  $\varepsilon > 0$   $Pr[\tilde{Y} \neq X] \leq \varepsilon$ .

$$\begin{aligned} Pr[X = s] - \pi(s) &= Pr[X = s] - Pr[\tilde{Y} = s] \\ &\leq Pr[X = s, \tilde{Y} \neq s] \leq Pr[X \neq \tilde{Y}] \leq \varepsilon \end{aligned}$$

Ομοίως

$$\begin{aligned}\pi(s) - Pr[X = s] &= Pr[\tilde{Y} = s] - Pr[X = s] \\ &\leq Pr[X = s, \tilde{Y} \neq s] \leq Pr[X \neq \tilde{Y}] \leq \varepsilon\end{aligned}$$

Άρα  $|Pr[Y = s] - \pi(s)| \leq \varepsilon$  και αφού ισχύει η σχέση για κάθε  $\varepsilon$  θετικό προκύπτει το ζητούμενο

Τώρα μένει να δικαιολογήσουμε τις επιλογές στον αλγόριθμο να προχωράμε από το παρελθόν και να διατηρούμε τις κλήσεις μας από προηγούμενες επαναλήψεις.



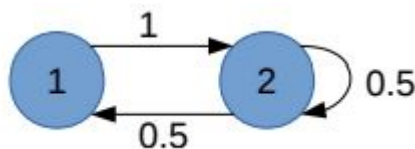
#### 4.1 Γιατί δεν επιλέγουμε coupling to the future

Εάν επιθυμούμε να απλοποιήσουμε τον αλγόριθμο ώστε να ξεκινάμε μία αλυσίδα για κάθε κατάσταση και να τις αφήσουμε να εξελιχθούν σύμφωνα με τον πίνακα  $P$ . Το αποτέλεσμα του αλγορίθμου συμφωνεί:

Έστω ότι θέλουμε να τρέξουμε τον αλγορίθμο μας για το παράδειγμα με δυο καταστάσεις και πίνακα μετάβασης

$$\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

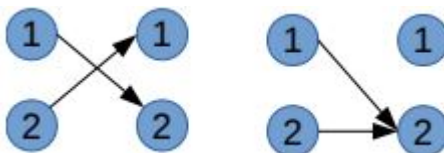
Η αλυσίδα είναι εργοδική. Και λύνοντας τις εξισώσεις  $\pi = \pi P$  προκύπτει ότι



η στάσιμη κατανομή είναι

$$\pi(1) = 1/3, \pi(2) = 2/3$$

Σε αυτή την περίπτωση, η συναρτήση ανανέωσης είναι αυτή που βλέπουμε



στην εικόνα με την πιθανότητα να επιλέξουμε το κάθε γράφημα είναι  $1/2$ . Υπάρχει μοναδική δυνατή επιλογή από το θεώρημα αναπαράστασης μέσω τυχαίας συνάρτησης. Αυτό αποδεικνύεται συμβολίζοντας

$$\alpha : f(1) = 1, f(2) = 1$$

$$\beta : f(1) = 1, f(2) = 2$$

$$\gamma : f(1) = 2, f(2) = 1$$

$$\delta : f(1) = 2, f(2) = 2$$

$$\alpha + \beta + \gamma + \delta = 1$$

$$\alpha + \beta = 0$$

$$\gamma + \delta = 1$$

$$\alpha + \gamma = 1/2$$

$$\beta + \delta = 1/2$$

Λύνοντας το παραπάνω σύστημα προκύπτει η μοναδική λύση  $\gamma = \delta = 1/2$ .  
 Αν επιλέξουμε να τρέξουμε τον αλγόριθμο *coupling from the past* με τη μόνη διαφορά να ξεκινάμε από τη χρονική στιγμή  $t = 0$  μέχρι να συναντηθούν οι αλυσίδες τότε παρατηρούμε ότι στην έξοδο του αλγορίθμου θα παρατηρούμε με πιθανότητα 1 την κατάσταση 2.

## 4.2 Γιατί χρησιμοποιούμε τις ίδιες συνατήσεις που κλήθηκαν σε προηγούμενες επαναλήψεις

Έστω ότι σε κάθε αποτυχημένη επανάληψη του αλγορίθμου ξεκινάμε από την αρχή και δεν αποθηκεύουμε τίποτα. Υποθέτουμε ότι έχουμε το προηγούμενο πίνακα μετάβασης.

Θέτουμε  $Y$  την έξοδο του τροποποιημένου CFTP και  $M$  τον ακέραιο για τον οποίο σταματά ο τροποποιημένος αλγόριθμος.

$$\begin{aligned} Pr[Y = 2] &= \sum_{m=1}^{\infty} Pr[Y = 2, M = m] \\ &\geq Pr[Y = 2, M = 1] + Pr[Y = 2, M = 2] \\ &= Pr[M = 1]Pr[Y = 2|M = 1] + Pr[M = 2]Pr[Y = 2|M = 2] \\ &= \frac{1}{2} \cdot 1 + \frac{3}{8} \cdot \frac{2}{3} \\ &= \frac{3}{4} > \frac{2}{3} \end{aligned}$$

Αφού

$$\begin{aligned} Pr[M = 2] &= Pr[M \neq 1] \cdot Pr[M = 2|M \neq 1] \\ &= \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8} \end{aligned}$$

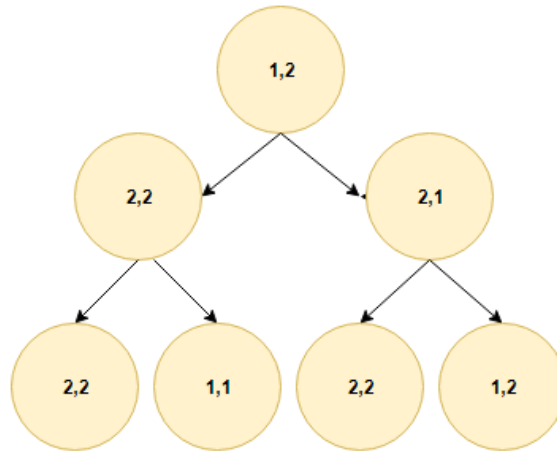
Πολλές φορές στη βιβλιογραφία επιλέγεται ο χρόνος σε κάθε αποτυχημένη εφαρμογή του αλγορίθμου να διπλασιάζεται αντί να μειώνεται κατά 1.

Η επιλογή αυτή δικαιολογήθηκε από τους συγγραφείς Propp&Wilson ως εξής

$$\begin{aligned} T_{total} &= |\Omega| \cdot T_0 + |\Omega| \cdot T_0 \cdot r + \dots + |\Omega| \cdot T_0 \cdot r^k \\ &= |\Omega| \cdot T_0 \frac{r^{k+1} - 1}{r - 1} < |\Omega| \cdot T_0 \frac{r^2 - 1}{r - 1} \leq |\Omega| \cdot T_{couple} \frac{r^2 - 1}{r - 1} \end{aligned}$$

Με τη παραπάνω παράσταση να ελαχιστοποιείται για  $r = 2$

Η μέθοδος δειγματοληψίας CFTP μας επιστρέφει ακριβή δείγματα από τη στάσιμη κατανομή μιας



### 4.3 Η μονότονη περίπτωση coupling from the past

Όπως αναφέραμε στη γενική περίπτωση του *coupling from the past* χρειάζεται να δημιουργήσουμε και να εξετάσουμε στο χρόνο τη συνάντηση τόσων αλυσίδων όσο και του πλήθους του χώρου καταστάσεων. Ο χώρος καταστάσεων μπορεί να είναι πολύ μεγάλος με αποτέλεσμα η εκτέλεση του αλγορίθμου να είναι αργή ή αδύνατη υπολογιστικά. Για αυτό το λόγο αν παρατηρήσουμε κάποια ειδική δομή στο χώρο καταστάσεων μπορεί να απλοποιηθεί η διαδικασία.

Συγκεκριμένα, αν ο χώρος είναι εφοδιασμένος με μια μερική διάταξη και αφού είναι πεπερασμένος μπορούμε να βρούμε ένα μέγιστο και ένα ελάχιστο στοιχείο. Αν αυτά τα δύο είναι μοναδικά, τα ονομάζουμε  $X_{max}$  &  $X_{min}$  αντίστοιχα.

Πλέον αρκεί να ελέγξουμε αν οι τροχιές των δύο αυτών αλυσίδων που ξεκινούν από τα δύο στοιχεία που ξεχωρίσαμε ταυτίζονται τη χρονική στιγμή  $t = 0$ . Χρειάζεται, βέβαια, και οι απεικονίσεις από το *random function representation* ή οι συναρτήσεις ανανέωσης να διατηρούν τη διάταξη.

**Ορισμός 13** Υποθέτουμε ότι ο χώρος καταστάσεων  $\Omega$  είναι εφοδιασμένος με μία διάταξη  $\preceq$ . Η συλλογή των απεικονίσεων  $f_n$  από την απεικόνιση  $F$  του *random map representation* καλείται μονότονη εάν

$$x \preceq y \Rightarrow Pr[f(x) \preceq f(y)]$$

Η μονότονη περίπτωση του αλγορίθμου *coupling from the past* μειώνει σημαντικά τη πολυπλοκότητα εφαρμογής της γενικής μεθόδου αφού ο χώρος καταστάσεων μπορεί να είναι τεράστιος.

**Πρόταση 4.4** Στη περίπτωση του μονότονου *coupling from the past* χρόνος να έρθουν δυο οποιεσδήποτε καταστάσεις σε σύζευξη κυριαρχείται απο τον χρόνο να έρθουν σε σύζευξη οι αλυσίδες που ξεκινούν η μία απο το μέγιστο στοιχείο και η άλλη από το ελάχιστο ,δηλαδή

$$Pr[T_{x,y} > t] \leq Pr[T_{x_{max},x_{min}} > t] \forall t \ \& \ \forall x, y \in \Omega$$

### Απόδειξη

$$F_t^0 = f_t \circ f_{t-1} \dots \dots \dots f_1$$

Κάθε απεικόνιση  $f$  διατηρεί τη διάταξη. Από την μονότονη ιδιότητα και με επαγωγή στο χρόνο

$$X_{min} \preceq x \Rightarrow F_{-t}^0(X_{min}) \preceq F_{-t}^0(x)$$

και

$$y \preceq X_{max} \Rightarrow F_{-t}^0(y) \preceq F_{-t}^0(X_{min})$$

Άρα αν  $F_{-t}^0(X_{max}) = F_{-t}^0(X_{min})$  συνεπάγεται ότι  $F_{-t}^0(x) = F_{-t}^0(y)$  με πιθανότητα 1

Στη μονότονη περίπτωση του *coupling from the past* μπορεί να δώσουμε μία εκτίμηση για το αναμενόμενο χρόνο βημάτων ώστε οι αλυσίδες να έρθουν σε σύζευξη.

**Λήμμα 4.1** Έστω  $l$  το μήκος της μεγαλύτερης αλυσίδας στο διατεταγμένο χώρο. Τότε

$$\frac{Pr[T^{up} > t]}{l} \leq \tilde{d}(t) \leq Pr[T^{up} > t]$$

$$\text{όπου } T^{up} = \min\{t | F_0^t(X_{max}) = F_0^t(X_{min})\}$$

Η απόδειξη του προηγούμενο λήμματος μπορεί να βρεθεί στο [2].

Θα δείξουμε οτι η συνάρτηση  $F(k) = Pr[T^{up} > k]$  είναι υποπολλαπλασιασική

**Λήμμα 4.2** Έστω  $K_1, K_2$  μη αρνητικοί ακέραιοι

$$Pr[T^{up} > K_1 + K_2] \leq Pr[T^{up} > K_1] \cdot Pr[T^{up} > K_2]$$

### Απόδειξη

Τα γεγονότα

$$\begin{cases} F_0^{K_1} & \text{είναι σταθερή} \\ F_{K_1}^{K_1+K_2} & \text{είναι σταθερή} \end{cases}$$

είναι ανεξάρτητα και αν ένα από τα δύο ισχύει τότε ισχύει αυτόματα πως  $F_0^{K_1+K_2}$  είναι σταθερό. Είναι εύκολο απο ορισμό

$$Pr[F_0^{K_1} \text{ είναι σταθερή} \cup F_{K_1}^{K_1+K_2} \text{ είναι σταθερή}] \leq Pr[T^{up} \leq K_1 + K_2]$$

Παίρνοντας τα συμπληρωματικά ενδεχόμενα σε κάθε περίπτωση και από την ανεξαρτησία των ενδεχομένων

$$\begin{aligned} Pr[T^{up} > K_1 + K_2] &\leq Pr[F_0^{K_1} \text{ δεν είναι σταθερή} \cup F_{K_1}^{K_1+K_2} \text{ δεν είναι σταθερή}] \\ &= Pr[T^{up} > K_1] \cdot Pr[F_{K_1}^{K_1+K_2} \text{ δεν είναι σταθερή}] \end{aligned}$$

Και επειδή ισχύει

$$Pr[F_{K_1}^{K_2+K_1} \text{ είναι σταθερή}] = Pr[F_0^{K_2} \text{ είναι σταθερή}]$$

**Λήμμα 4.3**  $t \cdot Pr[T^{up} > t] \leq E[T^{up}] \leq t/Pr[T^{up} \leq t]$

### Απόδειξη

Το πρώτο προκύπτει απο την ανισότητα Markov.

Για την απόδειξη της δεύτερης ανισότητας θέτουμε  $\varepsilon = Pr[T^{up} > t]$ .

Από την υποπολλαπλασιαστικότητα  $Pr[T^{up} > kt] \leq \varepsilon^k$ .

Για μία θετική διακριτή τυχαία μεταβλητή  $Y$  ισχύει  $E[Y] = \sum_{t \geq 0} Pr[Y > t]$ .

$$tE[T^{up}/t] = t \sum_{i \geq 0} Pr[T^{up}/t > i] = t \sum_{i \geq 0} Pr[T^{up} > it] \leq t \sum_{i \geq 0} \varepsilon^i = \frac{t}{1 - \varepsilon}$$

Συνεπώς,  $E[T^{up}] \leq \frac{t}{Pr[T^{up} \leq t]}$

Θέτοντας  $T = t_{mix}(1 + \log_2 l)$  από την υποπολλαπλασιαστικότητα της συνάρτησης  $\bar{d}(t)$

$$\bar{d}(T) = \bar{d}(t_{mix}(1 + \log_2 l)) \leq \bar{d}(t_{mix} \cdot \log_2 l) \bar{d}(t_{mix}) \leq \frac{1}{2} \frac{1}{2}^{\log_2 l} 2 = \frac{1}{2l}$$

Από το λήμμα 4.1

$$Pr[T^{up} > T] \leq l \cdot \frac{1}{2l} = \frac{1}{2}$$

και από το λήμμα 4.3

$$E[T^{up}] \leq \frac{T}{1 - \frac{1}{2}} \leq 2T \leq 2t_{mix}(1 + \log_2(l))$$

Ένα γενικό συμπέρασμα είναι πως εάν η αλυσίδα μας φτάνει σχετικά γρήγορα αρκετά κοντά στη στάσιμη κατανομή, δηλαδή έχει σχετικά μικρό χρόνο μίξης, τότε ο αναμενόμενος αριθμός βημάτων είναι ικανοποιητικός ενώ σε αντίθετη περίπτωση ο αλγόριθμος καθυστερεί να τερματίσει.

## 5 Εφαρμογή στο μοντέλο του Ising

Το μοντέλο του Ising προέρχεται από το χώρο της στατιστικής φυσικής και χρησιμοποιείται συχνά στη μελέτη μεθόδων δειγματοληψίας. Υποθέτουμε ότι έχουμε ένα μη κατευθυνόμενο πεπερασμένο γράφημα, συνήθως πλέγμα διάστασης  $N \times N$ , που οι κορυφές τους δέχονται τις τιμές  $+1$  ή  $-1$  που συμβολίζουν το σπιν του ατόμου.

Ο χώρος καταστάσεων είναι  $S^V$  όπου  $S = \{-1, +1\}$  και  $V$  το σύνολο των κορυφών το γράφου.

Κάθε στοιχείο του συνόλου  $S^V$  θα το καλούμε κατάσταση και θα το συμβολίζουμε με  $x$ .

Ως ενέργεια μιας κατάστασης ορίζεται να είναι η ποσότητα

$$H(x) = - \sum_{i \sim j} x(i)x(j)$$

όπου  $x(i)$  είναι η τιμή του σπιν στην κατάσταση  $x$  στη θέση  $j$ .

Η πιθανότητα να βρεθεί το σύστημα στην κατάσταση  $x$  είναι

$$\pi_{G,\beta}(x) = \frac{1}{Z_{G,\beta}} e^{-\beta H(x)}$$

όπου η παράμετρος  $\beta$  ονομάζεται αντίστροφη θερμοκρασία και ισούται με  $\frac{1}{kT}$ .

Η ποσότητα  $Z(\beta)$  ισούται με

$$\sum_{x \in \Omega} e^{-\beta H(x)}$$

και έχει το ρόλο της κανονικοποίησης.

Η σταθερά  $Z(\beta)$  είναι πολύ δύσκολο να υπολογιστεί και ο χώρος καταστάσεων είναι μεγέθους  $2^V$ .

Δημιουργείται το ερώτημα μήπως μπορούμε να χρησιμοποιήσουμε τον αλγόριθμο Propp-Wilson και ακόμα καλύτερα τη μονότονη περίπτωση.

Χρειάζεται να βρούμε ένα τρόπο να ανανεώνουμε την κατάσταση των αλυσίδων του αλγορίθμου και εάν είναι να δυνατό να διατηρεί μια μερική διάταξη. Θα αποδείξουμε μια σειρά από λήμματα και προτάσεις που θα μας οδηγήσουν στο να μπορέσουμε να ορίσουμε με επιτυχία τη μέθοδο *monotone coupling from the past* για το μοντέλο του Ising

Αρχικά, χρειάζεται να δημιουργήσουμε μια διαδικασία που θα μας βοηθήσει να

προσομειώσουμε την εξέλιξη της αλυσίδα μας. Δηλαδή, επιθυμούμε να κατασκευάσουμε μια μαρκοβιανή αλυσίδα η οποία σε κάθε επανάληψη θα επιλέγει μια κορυφή και θα δοκιμάζει να αλλάξει το υπάρχον σπιν και θα θέλαμε να ικανοποιεί τις εξισώσεις ακριβούς ισοροπίας.

Τη λύση θα μας τη δώσει μια γνωστή κατηγορία μαρκοβιανών αλυσίδων, που χρησιμοποιούνται για να προσομειώσουμε αλυσίδες που το σύνολο καταστάσεων είναι της μορφής  $S^V$ , γνωστές ως δυναμική Glauber. Τα βήματα της προσομείωσης για κάθε  $n$  είναι τα εξής:

1. Διαλέγουμε μια κορυφή ομοιόμορφα, έστω  $x$ .
2. Δεδομένης της κατάστασης τη χρονική στιγμή  $n$ ,  $X_n$ , επιλέγουμε να μεταβούμε στη κατάσταση  $X_{n+1}$  αλλάζοντας το σπιν στη θέση  $x$  σύμφωνα με τη δεσμευμένη πιθανότητα  $\pi_{G,\beta}$  δεδομένης της κατάστασης  $X_n(V \setminus \{x\})$ .

Συμβολίζουμε  $\xi \in \{-1, +1\}^{V \setminus \{x\}}$

Θέτουμε  $k_+(x, \xi)$  το πλήθος των κορυφών που έχουν σπιν  $+1$  στο και αλληλεπιδρούν με τη κορυφή  $x$

και αντίστοιχα  $k_-(x, \xi)$  το πλήθος των κορυφών που έχουν  $-1$  και γειτνιάζουν με τη κορυφή  $x$ .

Ακόμη, θέτουμε  $\xi^+$  την εκτεταμένη κατάσταση όπου  $\xi^+(x) = +1$  και  $\xi^-$  όπου  $\xi^-(x) = -1$ .

**Λήμμα 5.1**  $\frac{\pi_{G,\beta}(\xi^+)}{\pi_{G,\beta}(\xi^-)} = e^{2\beta(k_+(x,\xi) - k_-(x,\xi))}$

**Απόδειξη**

Από ορισμό

$$\pi_{G,\beta}(\xi) = \frac{1}{Z_{G,\beta}} e^{-\beta H(\xi)} = \frac{1}{Z_{G,\beta}} e^{\beta \sum_{(y,z)} \xi(y)\xi(z)}$$

Έτσι αν διατηρήσουμε τις κορυφές που γειτνιάζουν με τη κορυφή που είναι υποψήφια να αλλάξει τιμή

$$\frac{\pi_{G,\beta}(\xi^+)}{\pi_{G,\beta}(\xi^-)} = \frac{e^{\beta(k_+(x,\xi) - k_-(x,\xi))}}{e^{\beta(k_-(x,\xi) - k_+(x,\xi))}} = e^{2\beta(k_+(x,\xi) - k_-(x,\xi))}$$

**Πρόταση 5.1**

$$\pi_{G,\beta}(X(x) = +1 | X(V \setminus x) = \xi) = \frac{e^{2\beta(k_+(x,\xi) - k_-(x,\xi))}}{e^{2\beta(k_-(x,\xi) - k_+(x,\xi))} + 1}$$

## Απόδειξη

$$\begin{aligned}
\pi_{G,\beta}(X(x) = +1 | X(V \setminus x) = \xi) &= \\
\frac{\pi_{G,\beta}(X(x) = +1 \cap X(V \setminus x) = \xi)}{\pi_{G,\beta}(X(V \setminus x) = \xi)} &= \\
\frac{\pi_{G,\beta}(\xi_+)}{\pi_{G,\beta}(\xi_+) + \pi_{G,\beta}(\xi_-)} &= \\
\frac{\frac{\pi_{G,\beta}(\xi^+)}{\pi_{G,\beta}(\xi^-)}}{\frac{\pi_{G,\beta}(\xi^+)}{\pi_{G,\beta}(\xi^-)} + 1} &= \\
\frac{e^{2\beta(k_+(x,\xi) - k_-(x,\xi))}}{e^{2\beta(k_-(x,\xi) - k_+(x,\xi))} + 1} &=
\end{aligned}$$

Επειδή κάθε κορυφή μπορεί να έχει σπιν είτε +1 είτε -1, εύκολα, προκύπτει

$$k_+(x, \xi) - k_-(x, \xi) = \sum_{u \sim x} \sigma(u) := S(\xi, u)$$

Δηλαδή, η πιθανότητα να μεταβούμε από την κατάσταση  $\sigma$  στην κατάσταση  $w$ , που έχει στην κορυφή σπιν +1 και οι υπόλοιπες ταυτίζονται με εκείνες της  $\sigma$  είναι:

$$Pr(\sigma, w) := \frac{e^{\beta S(\sigma, w)}}{e^{\beta S(\sigma, w)} + e^{-\beta S(\sigma, w)}}$$

και αντίστοιχα η πιθανότητα να μεταβούμε από τη κατάσταση  $\sigma$  στη κατάσταση  $\delta$  είναι:

$$Pr(\sigma, \delta) = \frac{1}{|V|} \sum_{u \in V} \frac{e^{\beta \delta(u) S(\sigma, w)}}{e^{\beta \delta(u) S(\sigma, w)} + e^{-\beta \delta(u) S(\sigma, w)}} \mathbf{1}_{\sigma(u) = \delta(u)} \text{ για } u \neq w$$

Τώρα, είναι εύκολο να δείξουμε ότι η αλυσίδα από τη δυναμική Glauber ικανοποιεί τις εξισώσεις ακριβούς ισορροπίας σύμφωνα με τη κατανομή  $\pi_{G,\beta}$ . Ενδιαφέρον έχουν οι περιπτώσεις κατά τις οποίες η κορυφή  $x$  αλλάζει σπιν αφού για τις υπόλοιπες περιπτώσεις ισχύουν με τετριμμένο τρόπο οι εξισώσεις.

**1η περίπτωση** Η κορυφή  $\xi$  έχει σπιν +1 και αλλάζει το σπιν σε -1  
Από το λήμμα

$$\begin{aligned}
\frac{\pi_{G,\beta}(\xi^{-1})}{\pi_{G,\beta}(\xi)} &= e^{2\beta(k_-(x,\xi) - k_+(x,\xi))} = e^{-2\beta S(\xi, \xi^+)} \\
\frac{Pr(\xi, \xi^{-1})}{Pr(\xi^{-1}, \xi)} &= \frac{\frac{e^{-\beta S(\xi, \xi^{-1})}}{e^{-\beta S(\xi, \xi^{-1})} + e^{\beta S(\xi, \xi^{-1})}}}{\frac{e^{\beta S(\xi, \xi^{-1})}}{e^{\beta S(\xi, \xi^{-1})} + e^{-\beta S(\xi, \xi^{-1})}}} = e^{-2\beta S(\xi, w)}
\end{aligned}$$



**2η περίπτωση** Η κορυφή  $\xi$  έχει σπιν +1 στη θέση  $x$  και αλλάζει το σπιν σε -1. Κάνουμε αντίστοιχη ανάλυση και οι εξισώσεις ακριβούς ισορροπίας ισχύουν.

Η μέθοδος που ανανεώνουμε την κατάσταση της αλυσίδας είναι

$$X_{n+1}(x, \xi) = \begin{cases} +1 & \text{εάν } U_{n+1} \leq \frac{e^{2\beta(k_+(x,\xi)-k_-(x,\xi))}}{e^{2\beta(k_+(x,\xi)-k_-(x,\xi))+1}}, \\ -1 & \text{αλλιώς} \end{cases}$$

Χρειάζεται να δείξουμε ότι η παραπάνω μέθοδος ορίζει μια μονότονη περίπτωση.

### Απόδειξη

Σε κάθε επανάληψη του αλγορίθμου εμείς επιλέγουμε την ίδια κορυφή ως υποψήφια και τον ίδιο τυχαίο αριθμό. Έστω  $x$  η κορυφή που είναι υποψήφια για αλλαγή, οποιαδήποτε κορυφή του γραφήματος μπορεί να είναι αυτή και  $\eta \preceq \xi$ . Οι υπόλοιπες κορυφές παραμένουν απαράλλαχτες.

$$\forall y \neq x \quad \xi(y) \geq \eta(y)$$

Μένει να δείξουμε ότι

$$\xi(x) \geq \eta(x)$$

Αφού η κατάσταση  $\xi$  έχει συνολικά περισσότερες κορυφές με σπιν με θετικό πρόσημο σε σχέση με την κατάσταση  $\eta$ , αυτό θα ισχύει και τοπικά. Δηλαδή,

$$k^+(\xi, x) \geq k^+(\eta, x)$$

Αντίστοιχα, λόγω μονοτονίας

$$k^-(\eta, x) \leq k^-(\xi, x)$$

και

$$k^+(\xi, x) - k^-(\xi, x) \leq k^-(\eta, x) - k^-(\eta, x)$$

Συνεπώς,

$$\frac{e^{2\beta(k_+(x,\xi)-k_-(x,\xi))}}{e^{2\beta(k_+(x,\xi)-k_-(x,\xi))+1}} \geq \frac{e^{2\beta(k_+(x,\eta)-k_-(x,\eta))}}{e^{2\beta(k_+(x,\eta)-k_-(x,\eta))+1}}$$

γιατί η συνάρτηση  $\zeta(x) = \frac{e^{2\beta x}}{e^{2\beta x} + 1}$  είναι αύξουσα συνάρτηση.

## 6 Χρόνος μίξης για το μοντέλο του Ising με τη προσομείωση με δυναμική Glauber

Σε αυτή την ενότητα θα προσπαθήσουμε να βρούμε κάποια άνω φράγματα για το χρόνο μίξης του μοντέλου Ising, όπως το προσομειώσαμε στο προηγούμενο κεφάλαιο, ώστε να έχουμε μια εκτίμηση για τον αναμενόμενο αριθμό επαναλήψεων του αλγορίθμου.

Κατ' αρχάς, θα εισάγουμε τη τεχνική εύρεσης άνω φραγμάτων του χρόνου μίξης με τη σύζευξη μονοπατιών (path coupling).

Υποθέτουμε ότι βρισκόμαστε σε ένα μετρικό χώρο  $(\Omega, \rho)$  όπου  $\rho$  είναι η μετρική με την οποία είναι εφοδιασμένος ο χώρος μας και ικανοποιεί την εξής σχέση:

$$\rho(x, y) \geq \mathbf{1}\{x, y\} \quad \text{για κάθε } x, y \in \Omega$$

Επιπλέον, υποθέτουμε ότι για οποιοδήποτε ζεύγος καταστάσεων υπάρχει σύζευξη  $(X_1, Y_1)$  των κατανομών  $P(x, \cdot)$ ,  $P(y, \cdot)$  τέτοια ώστε:

$$E_{x,y}\rho(X_1, Y_1) \leq e^{-\alpha}\rho(x, y)$$

**Ορισμός 14** Διαμέτρος του  $\Omega$  ορίζεται η ποσότητα

$$diam(\Omega) := \max_{x,y \in \Omega} \rho(x, y)$$

Επαγωγικά,

$$E_{x,y}\rho(X_t, Y_t) \leq e^{-at} diam(\Omega)$$

Μπορούμε να συμπεράνουμε, εύκολα, ότι

$$\begin{aligned} \|P^t(x, \cdot) - P^t(y, \cdot)\| &\leq Pr_{x,y}[X_t, Y_t] \\ &= Pr_{x,y}[\rho(X_t, Y_t) \geq 1] \\ &\leq E_{x,y}\rho(X_t, Y_t) \\ &\leq diam(\Omega)e^{-at} \end{aligned}$$

και συνεπώς

$$t_{mix}(\varepsilon) \leq \frac{1}{\alpha} [\log(diam(\Omega)) + \log(\varepsilon^{-1})]$$

## 6.1 Η μετρική μεταφοράς

Η μετρική δυο κατανομών πιθανοτήτων  $\mu, \nu$  που ορίζονται στον ίδιο χώρο είναι:

**Ορισμός 15**  $\rho_k(\mu, \nu) := \inf\{E(\rho(X, Y)) : (X, Y) \text{ είναι σύζευξη των } \mu, \nu\}$

Είναι προφανές πως αν επιλέξουμε ως  $\rho(x, y) = \mathbf{1}_{\{x \neq y\}}$  τότε  $\rho_k(\mu, \nu) = \|\mu - \nu\|$  γιατί

$$E(\mathbf{1}_{\{A\}}) = P[A] \text{ για κάθε } A \subset \Omega$$

Πολλές φορές είναι πιο απλό να εργαζόμαστε στο χώρο γινόμενο  $\Omega \times \Omega$  και να χρησιμοποιούμε τις κατανομές που δημιουργούνται από τις προβολές. Η προβολή στην πρώτη και στη δεύτερη συντεταγμένη ορίζεται να είναι αντίστοιχα

$$q(\cdot \times \Omega) = \sum_{y \in \Omega} q(\cdot, y)$$

$$q(\Omega \times \cdot) = \sum_{y \in \Omega} q(y, \cdot)$$

Δοθέντος ενός μέτρου πιθανότητας  $q$  στο χώρο γινόμενο με  $\mu, \nu$  τις προβολές του στη πρώτη και στη δεύτερη συντεταγμένη αντίστοιχα τότε η ταυτοτική απεικόνιση στο χώρο  $(\Omega \times \Omega, q)$  είναι σύζευξη των  $\mu, \nu$ .

Αντίστροφα, δοθέντος μίας σύζευξης  $(X, Y)$  δύο μέτρων πιθανότητας  $\mu, \nu$ , τότε η διδιάστατη τυχαία μεταβλητή  $(X, Y)$  έχει προβολές τις κατανομές  $\mu, \nu$  στην πρώτη και στη δεύτερη μεταβλητή αντίστοιχα.

Παρατηρώντας πως

$$\mathbf{E}(\rho((X, Y))) = \sum_{(x, y) \in \Omega \times \Omega} \rho(x, y) q(x, y)$$

όταν το ζεύγος τυχαίων μεταβλητών  $(X, Y)$  έχει κατανομή  $q$ , τότε η μετρική μπορεί να γραφτεί ως:

$$\rho_K(\mu, \nu) = \inf \left[ \sum_{(x, y) \in \Omega \times \Omega} \rho(x, y) q(x, y) : q(\cdot \times \Omega) = \mu, q(\Omega \times \cdot) = \nu \right]$$

**Ορισμός 16** Καλούμε βέλτιστη σύζευξη δύο μέτρων πιθανότητας  $\mu, \nu$  το ζεύγος τυχαίων μεταβλητών  $(X_*, Y_*)$

$$E(\rho(X_*, Y_*)) = \rho_K(\mu, \nu)$$

**Λήμμα 6.1** Η συνάρτηση  $\rho_K$  που ορίσαμε παραπάνω ικανοποιεί τις ιδιότητες της μετρικής

### Απόδειξη

Η απεικόνιση  $\rho_K(\mu, \nu)$  είναι μη αρνητική ως το ελάχιστο άνω φράγμα μη αρνητικών ποσοτήτων και είναι συμμετρική αφού  $\rho_K(\mu, \nu) = \rho_K(\nu, \mu)$ . Μένει να δείξουμε ότι ικανοποιεί την ιδιότητα της τριγωνικής ανισότητας. Δηλαδή, να δείξουμε ότι

$$\rho_K(\mu, \eta) \leq \rho_K(\mu, \nu) + \rho_K(\nu, \eta)$$

Έστω  $\mu, \nu, \eta$  τρία μέτρα πιθανότητας στο χώρο  $\Omega$ . Καλούμε  $p$  την πιθανότητα κατανομής στο χώρο  $\Omega \times \Omega$  που είναι σύζευξη των κατανομών  $\mu$  και  $\nu$  και  $q$  τη σύζευξη των κατανομών  $\nu$  και  $\eta$ .

Ορίζουμε τη κατανομή  $r$  στο χώρο  $\Omega \times \Omega \times \Omega$  ως εξής

$$r(x, y, z) := \frac{p(x, y)q(y, z)}{\nu(y)}$$

Η προβολή της κατανομής  $r$  στις δυο πρώτες συντεταγμένες είναι

$$\sum_{z \in \Omega} r(x, y, z) = \sum_{z \in \Omega} \frac{p(x, y)q(y, z)}{\nu(y)} = p(x, y)$$

είναι η κατανομή  $p$  ενώ η προβολή στις δυο τελευταίες είναι

$$\sum_{x \in \Omega} r(x, y, z) = \sum_{x \in \Omega} \frac{p(x, y)q(y, z)}{\nu(y)} = q(y, z)$$

η κατανομή  $q$  ενώ η προβολή στην πρώτη και στην τελευταία συντεταγμένη

$$\sum_{y \in \Omega} r(x, y, z) = \sum_{y \in \Omega} \frac{p(x, y)q(y, z)}{\nu(y)}$$

είναι σύζευξη των  $\mu, \eta$  γιατί

$$\sum_{y \in \Omega} \sum_{x \in \Omega} r(x, y, z) = \sum_{y \in \Omega} q(y, z) = \eta$$

$$\sum_{y \in \Omega} \sum_{z \in \Omega} r(x, y, z) = \sum_{y \in \Omega} p(x, y) = \mu$$

Εάν υποθέσουμε, επιπλέον, ότι οι κατανομές  $p$  και  $q$  είναι βέλτιστες συζεύξεις τότε για ένα τυχόν διάνυσμα  $(X, Y, Z)$  με κατανομή  $r$  ισχύει

$$\begin{aligned} \rho(X, Z) &\leq \rho(X, Y) + \rho(Y, Z) \Rightarrow \\ E(\rho(X, Z)) &\leq E(\rho(X, Y)) + E(\rho(Y, Z)) = \\ &\rho_K(\mu, \nu) + \rho_K(\nu, \eta) \end{aligned}$$

και επειδή η παραπάνω σχέση ισχύει για κάθε σύζευξη των κατανομών  $\mu, \nu$  προκύπτει το ζητούμενο

## 6.2 Σύζευξη μονοπατιών

Υποθέτουμε πως ο χώρος καταστάσεων είναι ένας συνεκτικός γράφος που θα το συμβολίζουμε ως  $G = (\Omega, E_0)$  έχει μια συνάρτηση βάρους  $l$  που ικανοποιεί τη σχέση  $l \geq 1$  για κάθε ακμή.

Ορίζουμε ένα μονοπάτι από τη κατάσταση  $x$  στην κατάσταση  $y$  να είναι μια ακολουθία καταστάσεων την οποία θα συμβολίζουμε :

$$\xi = (x_0, x_1, \dots, x_r) \text{ με } x_0 = x \text{ και } x_r = y$$

Το μήκος του μονοπατιού είναι  $\sum_{i=1}^r l(x_{i-1}, x_i)$

**Πρόταση 6.1** Μετρική μονοπατιού στο χώρο  $\Omega$  είναι η ποσότητα:

$$\rho(x, y) = \min\{\text{μήκος του μονοπατιού } \xi : \xi \text{ είναι ένα μονοπάτι από το } x \text{ στο } y\}$$

Αφού  $l(x, y) \geq 1$ , από ορισμό του μήκους και της μετρικής ισχύει  $\rho(x, y) \geq \mathbf{1}\{x \neq y\}$ . Συνεπώς,

$$P\{X \neq Y\} = E(\mathbf{1}\{X \neq Y\}) \leq E_p(X, Y)$$

Ελαχιστοποιώντας ως προς όλα τις δυνατές συζεύξεις των κατανομών  $\mu, \nu$  προκύπτει:

$$\|\mu - \nu\| \leq \rho_K(\mu, \nu)$$

**Θεώρημα 6.1** Υποθέτουμε ότι ο χώρος καταστάσεων είναι το σύνολο κορυφών ενός γράφου με συνάρτηση βάρους  $l$ . Η συνάρτηση  $\rho$  ικανοποιεί τις ιδιότητες της μετρικής μονοπατιού. Εάν για κάθε ακμή  $\{x, y\}$  υπάρχει σύζευξη  $(X_1, Y_1)$  των κατανομών  $P(x, \cdot), P(y, \cdot)$  που ικανοποιεί τη σχέση:

$$\mathbf{E}_{x,y}(\rho(X_1, Y_1)) \leq \rho(x, y)e^{-a} = l(x, y)e^{-a}$$

Τότε για κάθε μέτρα πιθανότητας  $\mu, \nu$  στο  $\Omega$

$$\rho_K(\mu P, \nu P) \leq e^{-a} \rho_K(\mu, \nu)$$

Πρώτα θα δείξουμε ένα χρήσιμο λήμμα

**Λήμμα 6.2** Υποθέτουμε ότι ισχύουν όλες οι υποθέσεις του θεωρήματος καθώς και το συμπέρασμα. Τότε,

$$d(t) \leq e^{-\alpha t} \text{diam}(\Omega)$$

και συνεπώς

$$t_{\text{mix}}(\varepsilon) \leq \frac{-\log(\varepsilon) + \log(\text{diam}(\Omega))}{\alpha}$$

### Απόδειξη λήμματος 6.2

Χρησιμοποιώντας τη σχέση  $\rho_K(\mu P, \nu P) \leq e^{-\alpha} \rho_K(\mu, \nu)$  επαγωγικά προκύπτει

$$\rho_K(\mu P^t, \nu P^t) \leq e^{-\alpha t} \rho_K(\mu, \nu) \leq e^{-\alpha t} \max_{x,y} \rho(x, y)$$

Από τη σχέση  $\|\mu - \nu\| \leq \rho_K(\mu, \nu)$  και θέτοντας όπου  $\mu = \delta_x$  και  $\nu = \pi$  προκύπτει

$$\|P^t(x, \cdot) - \pi\|_{TV} \leq e^{-\alpha t} \text{diam}(\Omega)$$

### Απόδειξη θεωρήματος 6.1

Έστω  $x, y$  δύο καταστάσεις και  $(x_0, x_1, \dots, x_r)$  ένα μονοπάτι που ικανοποιεί τη μετρική του μονοπατιού από το  $x$  στο  $y$ . Από τη τριγωνική ιδιότητα της  $\rho_K$

$$\rho_K(P(x, \cdot), P(y, \cdot)) \leq \sum_{k=1}^r \rho_K(P(x_{k-1}, \cdot), P(x_k, \cdot))$$

Επειδή  $\rho_K$  είναι η ελάχιστη πάνω σε όλα τις δυνατές συζεύξεις συνεπάγεται ότι για κάθε ακμή

$$\rho_K(P(a, \cdot), P(b, \cdot)) \leq e^{-\alpha} l(a, b)$$

Αντικαθιστώντας στη σχέση προκύπτει

$$\rho_K(P(x, \cdot), P(y, \cdot)) \leq e^{-\alpha} \sum_{k=1}^r l(x_{k-1}, x_k)$$

Αφού το μονοπάτι που επιλέξαμε είναι ελαχίστου μήκους

$$\rho(x, y) = \sum_{k=1}^r l(x_{k-1}, x_k)$$

και συνεπώς

$$\rho_K(P(x, \cdot), P(y, \cdot)) \leq e^{-\alpha} \rho(x, y)$$

που είναι και το ζητούμενο.

Έστω  $\eta$  μια βέλτιστη σύζευξη των κατανομών  $\mu, \nu$  τέτοια ώστε

$$\rho_K(\mu, \nu) = \sum_{x,y \in \Omega} \rho(x, y) \eta(x, y)$$

Από τη σχέση που μόλις δείξαμε αντιλαμβανόμαστε ότι για κάθε ζεύγος καταστάσεων  $x, y$  υπάρχει μια βέλτιστη σύζευξη  $\theta_{x,y}$  των κατανομών  $P(x, \cdot), P(y, \cdot)$  τέτοια ώστε:

$$\sum_{u,w \in \Omega} \rho(u, w) \theta_{x,y}(u, w) \leq e^{-\alpha} \rho(x, y)$$

Η ποσότητα  $\theta := \sum_{x,y \in \Omega} \eta(x, y) \theta_{x,y}$  είναι ένα μέτρο πιθανότητας στο χώρο γινόμενο  $\Omega \times \Omega$  και αποτελεί σύζευξη των κατανομών  $\mu P$  και  $\nu P$ .

$$\begin{aligned} \sum_{w \in \Omega} \theta &= \sum_{w \in \Omega} \sum_{x,y \in \Omega} \eta(x, y) \theta_{x,y}(u, w) \\ &= \sum_{x,y \in \Omega} \eta(x, y) \sum_{w \in \Omega} \theta_{x,y}(u, w) \\ &= \sum_{x,y \in \Omega} \eta(x, y) P(x, u) \\ &= \sum_{x \in \Omega} P(x, u) \sum_{y \in \Omega} \eta(x, y) \\ &= \sum_{x \in \Omega} P(x, u) \mu(x) = \mu P \end{aligned}$$

Ομοίως δείχνουμε ότι

$$\sum_{u \in \Omega} \theta = \nu P$$

Συνδυάζοντας τις σχέσεις έχουμε

$$\begin{aligned} \sum_{u,w \in \Omega} \rho(u, w) \theta(u, w) &= \sum_{x,y \in \Omega} \sum_{u,w \in \Omega} \rho(u, w) \theta_{x,y}(u, w) \eta(x, y) \\ &\leq e^{-\alpha} \sum_{x,y \in \Omega} \rho(x, y) \eta(x, y) = e^{-\alpha} \rho_K(\mu, \nu) \end{aligned}$$

και επειδή ισχύει  $\rho_K(\mu P, \nu P) \leq \sum_{u,w \in \Omega} \rho(u, w) \theta(u, w)$  συνεπάγεται

$$\rho_K(\mu P, \nu P) \leq e^{-\alpha} \rho_K(\mu, \nu)$$

Συνοψίζοντας τι κάναμε μέχρι τώρα:

1. Δημιουργώντας το γράφημα με κορυφές το σύνολο του χώρου καταστάσεων και μία συνάρτηση βάρους  $l$  μπορούμε να ορίσουμε μία μετρική μεταξύ των καταστάσεων  $r$ .
2. Μπορούμε να ορίσουμε μια μετρική μεταξύ κατανομών.
3. Εάν για κάθε ακμή  $\{x, y\}$  υπάρχει σύζευξη  $(X, Y)$  των κατανομών  $P(x, \cdot), P(y, \cdot)$  τέτοια ώστε

$$E_{x, y}(\rho(X, Y)) \leq \rho(x, y)e^{-\alpha}$$

τότε μπορώ να έχω μια εκτίμηση για το  $t_{mix}$

### 6.3 Άνω όριο για το χρόνο μίξης του μοντέλου του Ising σε υψηλές θερμοκρασίες

**Θεώρημα 6.2** Για το μοντέλο *Ising* που το προσομειώνουμε με τη δυναμική *Glauber* σε ένα γράφημα με  $n$  κορυφές και μέγιστο βαθμό  $\Delta$  ισχύει

$$t_{mix}(\varepsilon) \leq \frac{n(\log(n) + \log(1/\varepsilon))}{c(\beta)}$$

εάν  $\Delta \cdot \tanh(\beta) \leq 1$  όπου  $c(\beta) := 1 - (\Delta) \tanh(\beta)$

Ο μέγιστος βαθμός ενός γραφήματος είναι το μέγιστο όλων των κορυφών πάνω στο πλήθος των κορυφών με τις οποίες συνδέεται.

Σε ένα πλέγμα  $n \times n$  είναι 4 ενώ σε ένα γράφημα με κορυφές τις κορυφές ενός ισοπλεύρου τριγώνου είναι 2.

Θα αποδείξουμε ένα χρήσιμο λήμμα

**Λήμμα 6.3** Η συνάρτηση  $\phi(x) := \tanh(\beta(x+1)) - \tanh(\beta(x-1))$  είναι περιττή και φθίνουσα στο  $[0, \infty)$

#### Απόδειξη λήμματος 6.3

Η συνάρτηση  $\psi(x) := \tanh(\beta x)$  έχει πρώτη παράγωγο  $\psi'(x) = \frac{\beta}{\cosh^2(\beta x)}$  και δεύτερη παράγωγο  $\psi''(x) = -\frac{2 \sinh \beta x}{\cosh^3(\beta x)}$  που είναι αρνητική για  $x > 0$

Άρα

$$\phi'(x) = \frac{\beta}{\cosh^2(\beta(x+1))} - \frac{\beta}{\cosh^2(\beta(x-1))} < 0$$

Η συνάρτηση  $\tanh(x)$  είναι περιττή. Συνεπώς,

$$\phi(-x) = \tanh(\beta(-x+1)) - \tanh(\beta(-x-1)) = \phi(x)$$



## Απόδειξη Θεωρήματος 6.2

Αρχικά, ορίζουμε την μετρική  $\rho'$  στο χώρο  $\Omega$

$$\rho'(\sigma, \tau) = \frac{1}{2} \sum_{u \in V} |\sigma(u) - \tau(u)|$$

Η συνάρτηση που ορίσαμε είναι μετρική μονοπατιού.

Έστω  $\sigma, \tau$  δύο διατάξεις τέτοιες ώστε  $\rho'(\sigma, \tau) = 1$ , δηλαδή, δύο διατάξεις που διαφέρουν μόνο σε μια κορυφή που θα τη καλούμε πλέον  $v$ .

Υποθέτουμε χωρίς βλάβη της γενικότητας ότι  $\sigma(v) = -1$  και  $\tau(v) = 1$

Ορίζουμε ακόμη το σύνολο  $N(u) := \{v : u \sim v\}$  που περιέχει τις κορυφές που γειτνιάζουν με τη κορυφή  $u$ .

Θα περιγράψουμε μια σύζευξη  $(X, Y)$  που η μία ξεκινά από τη διάταξη  $\sigma$  και η άλλη ξεκινά από τη διάταξη  $\tau$ .

Διαλέγουμε μια κορυφή  $w$  τυχαία από το σύνολο των κορυφών  $V$ .

**1η περίπτωση:** Η κορυφή  $w$  δεν ανήκει στο σύνολο  $N(u)$

Τότε οι γειτονικές κορυφές της  $w$  συμφωνούν και στις δύο διατάξεις και οι αλυσίδες πραγματοποιούν την ίδια αλλαγή.

**2η περίπτωση:** Η κορυφή  $w$  ανήκει στο σύνολο  $N(u)$

Είναι προφανές από την ανάλυση που έχουμε κάνει προηγουμένως ότι η πιθανότητα να μεταβούν στην επόμενη κατάσταση δεν είναι πλέον η ίδια για τις δύο αλυσίδες.

Θα χρησιμοποιήσουμε μόνο μια μορφή τυχαιότητας για τη μετάβαση στη νέα κατάσταση και των δυο αλυσίδων ώστε να υπάρχει υψηλότερη πιθανότητα να συναντηθούν οι τροχιές τους.

Έστω  $U$  μια τυχαία μεταβλητή με κατανομή ομοιόμορφη στο  $[0,1]$  και θέτουμε

$$X(w) = \begin{cases} +1 & \text{εάν } U \leq \rho(\sigma, w) \\ -1 & \text{εάν } U > \rho(\sigma, w) \end{cases}$$

$$Y(w) = \begin{cases} +1 & \text{εάν } U \leq \rho(\tau, w) \\ -1 & \text{εάν } U > \rho(\tau, w) \end{cases}$$

με τις άλλες κορυφές να μένουν ίδιες.

Συνεπώς, εάν  $w = u$  τότε  $\rho'(X, Y) = 0$  και αν  $w \notin N(u) \cup \{u\}$  τότε  $\rho'(X, Y) = 1$ .

Εάν  $w \in N(u)$  και  $\rho(\sigma, w) < U \leq \rho(\tau, w)$  τότε  $\rho'(X, Y) = 2$ .

$$E_{\sigma, \tau} \rho'(X, Y) \leq 1 - \frac{1}{n} + \frac{1}{n} \sum_{w \in N(u)} |\rho(\tau, w) - \rho(\sigma, w)|$$

Αν παρατηρήσουμε πως  $S(w, \tau) = S(w, \sigma) + 2 = S + 2$  προκύπτει

$$\begin{aligned} \rho(\tau, w) - \rho(\sigma, w) &= \frac{e^{\beta(S+2)}}{e^{\beta(S+2)} + e^{-\beta(S+2)}} - \frac{e^{\beta S}}{e^{\beta S} + e^{-\beta S}} \\ &= \frac{1}{2} [\tanh(\beta(S+2)) - \tanh(\beta S)] \end{aligned}$$

Θέτοντας  $\tilde{S} = S + 1$  και εφαρμόζοντας τη σχέση

$$\rho(\tau, w) - \rho(\sigma, w) = \frac{1}{2} [\tanh(\beta(\tilde{S} + 1)) - \tanh(\beta(\tilde{S} - 1))] \leq \tanh(\beta)$$

Αντικαθιστώντας στη μέση τιμή

$$E_{\sigma, \tau} \rho(X, Y) \leq 1 - \left[ \frac{1 - \Delta \tanh(\beta)}{n} \right] \leq \exp\left(-\frac{1 - \Delta \tanh(\beta)}{n}\right) = e^{-c(\beta)/n}$$

Εάν  $\Delta \tanh(\beta) < 1$  τότε  $c(\beta) > 0$ .

Η διάμετρος του συνόλου είναι  $\text{diam}(\Omega) = n$

Εφαρμόζοντας στη σχέση όπου  $\alpha = \frac{c(\beta)}{n}$  προκύπτει η ζητούμενη σχέση τεξτ

Εάν παρατηρήσουμε πως  $\tan(x) \leq x$  τότε αν  $\beta < \Delta^{-1}$  συνεπάγεται πως  $\Delta \tanh(\beta) < 1$  και μπορούμε να ισχυριστούμε ότι το θεώρημα που μόλις αποδείξαμε ισχύει και για το εύρος τιμών που είναι μικρότερο από  $\Delta^{-1}$ .

Αποδεικνύεται πως η σχέση που δείξαμε ισχύει για κάθε  $\beta < \beta_c = \frac{1}{2} \ln 1 + \sqrt{2}$

## 7 Χρόνος μίξης για το τυχαίο περίπατο

Έστω  $\phi$  μια συνάρτηση  $\Omega \rightarrow \mathbb{R}$ .

Μέση τιμή μιας πραγματικής τυχαίας μεταβλητής σύμφωνα με τη στάσιμη κατανομή  $E_\pi \phi$  είναι η ποσότητα

$$E_\pi \phi = \sum_{x \in \Omega} \pi(x) \phi(x)$$

Αντίστοιχα, διασπορά μιας πραγματικής τυχαίας μεταβλητής σύμφωνα με τη στάσιμη κατανομή

$$\begin{aligned} \text{Var}_\pi \phi &:= \sum_{x \in \Omega} \pi(x) (\phi(x) - E_\pi \phi)^2 \\ &= \sum_{x \in \Omega} \pi(x) \phi(x)^2 - (E_\pi \phi)^2 \\ &= \sum_{x \in \Omega} \pi(x) \phi(x)^2 \sum_{y \in \Omega} \pi(y) - \sum_{x \in \Omega} \pi(x) \phi(x) \sum_{y \in \Omega} \pi(y) \phi(y) \\ &= \sum_{x, y} (\pi(x) \pi(y) \phi(x)^2 - \pi(x) \pi(y) \phi(x) \phi(y)) \\ &= \frac{1}{2} \sum_{x, y} \pi(x) \pi(y) (\phi(x) - \phi(y))^2 \end{aligned}$$

**Ορισμός 17** Καλούμε τοπική διασπορά τη ποσότητα

$$\mathcal{E}_\pi(\phi, \phi) = \frac{1}{2} \sum_{x, y} \pi(x) P(x, y) (\phi(x) - \phi(y))^2$$

**Ορισμός 18** Ορίζουμε σταθερά του Poincaré

$$\alpha = \inf_{\phi \text{ δεν είναι σταθερή}} \frac{\mathcal{E}_\pi(\phi, \phi)}{\text{Var}_\pi \phi}$$

**Θεώρημα 7.1** Για κάθε πίνακα μετάβασης  $P$  που αναπαριστά μια απεριοδική και μη υποβιβασίμη μαρκοβιανή αλυσίδα ισχύει

$$\tau_x(\varepsilon) \leq \frac{1}{\alpha} (2 \ln \varepsilon^{-1} + \ln \pi(x)^{-1})$$

Ράθυμη καλούμε τη μαρκοβιανή αλυσίδα που ο πίνακας που τη περιγράφει  $P$  μπορεί να διασπαστεί σε ένα άθροισμα της μορφής

$$P = \frac{\hat{P} + I}{2}$$

όπου ο πίνακας  $\hat{P}$  είναι και αυτός στοχαστικός.

Εύκολα μπορούμε από την ιδιότητα  $\pi = \pi P$  να ελέγξουμε ότι οι δύο πίνακες έχουν την ίδια στάσιμη κατανομή.

Πολλές φορές καταφεύγουμε να μελετάμε την ράθυμη εκδοχή της μαρκοβιανής αλυσίδας αφού έχει την ίδια στάσιμη κατανομή με την αρχική εκδοχή της αλυσίδας και επιπλέον με τη θετική πιθανότητα παραμονής στην ίδια θέση αποτρέπουμε τυχόν προβλήματα που δημιουργούνται με την απεριοδικότητα που υπήρχε προηγουμένως

### Απόδειξη

Ορίζουμε  $[P_\phi](x) := \sum_y P(x, y)\phi(y)$ . Αφού η αλυσίδα μας είναι απεριοδική και μη υποβιβάσιμη συνεπάγεται

$$\lim_{t \rightarrow \infty} P^t \phi = E_\pi \phi$$

και κρατάμε την σχέση ότι  $E_\pi[P^t \phi] = E_\pi[\phi]$

$$\begin{aligned} E_\pi[P^t \phi] &= \sum_x \pi(x) P^t \phi(x) \\ &= \sum_x \pi(x) \sum_y P^t(x, y) \phi(y) \\ &= \sum_y \phi(y) \sum_x \pi(x) P^t(x, y) \\ &= \sum_y \phi(y) \pi(y) = E_\pi[\phi] \end{aligned}$$

**1ο βήμα:** Θα δείξουμε ότι για κάθε  $\phi : \Omega \rightarrow R$

$$\text{Var}_\pi[P_\phi] \leq \text{Var}_\pi \phi - \mathcal{E}_\pi(\phi, \phi)$$

Αφού ασχολούμαστε με τη ράθυμη εκδοχή της αλυσίδας

$$\begin{aligned} [P_\phi](x) &= \sum_{y \in \Omega} P(x, y) \phi(y) \\ &= \frac{1}{2} \phi(x) + \frac{1}{2} \sum_{y \in \Omega} \hat{P}(x, y) \phi(y) \\ &= \frac{1}{2} \sum_{y \in \Omega} \hat{P}(x, y) (\phi(x) + \phi(y)) \end{aligned}$$

Μπορούμε να ισχυριστούμε χωρίς βλάβη της γενικότητας ότι  $E_\pi \phi = 0$  γιατί αν θέσουμε  $\phi' = \phi - E_\pi \phi$  τότε  $E_\pi \phi' = 0$  και οι διασπορές μένουν ίδιες αφού δεν επηρεάζονται από μετατοπίσεις.

$$\begin{aligned} \text{Var}_\pi[P\phi] &= \sum_{x \in \Omega} \pi(x) ([P\phi(x)])^2 \\ &= \frac{1}{4} \sum_{x \in \Omega} \pi(x) \sum_{y \in \Omega} \hat{P}(x, y) (\phi(x) + \phi(y))^2 \\ &\leq \frac{1}{4} \sum_{x, y} \pi(x) \hat{P}(x, y) (\phi(x) + \phi(y))^2 \end{aligned}$$

με τη τελευταία ανισότητα να προέκυψε από την ανισότητα *Cauchy – Schwarz*

$$\begin{aligned} \text{Var}_\pi(\phi) &= \frac{1}{2} \sum_{x \in \Omega} \pi(x) \phi(x)^2 + \frac{1}{2} \sum_{y \in \Omega} \pi(y) \phi(y)^2 \\ &= \frac{1}{2} \sum_{x, y} \pi(x) \phi(x)^2 \hat{P}(x, y) + \frac{1}{2} \sum_{x, y} \pi(x) \phi(y)^2 \hat{P}(x, y) \\ &= \frac{1}{2} \sum_{x, y \in \Omega} \pi(x) \hat{P}(x, y) (\phi(x)^2 + \phi(y)^2) \end{aligned}$$

Αφαιρώντας κατά μέλη

$$\text{Var}_\pi \phi - \text{Var}_\pi[P_\phi] \geq \frac{1}{4} \sum_{x, y} \pi(x) \hat{P}(x, y) (\phi(x) - \phi(y))^2$$

Το δεξί μέλος ισούται με

$$\frac{1}{2} \sum_{x, y \in \Omega} \pi(x) P(x, y) (\phi(x) - \phi(y))^2 = \mathcal{E}_\pi(\phi, \phi)$$

για το στο άθροισμα δεν έχει καμία συνεισφορά τα διαγώνια στοιχεία του πίνακα

**2ο βήμα:** Θα δείξουμε ότι  $Var_{\pi}[P_{\phi}^t] \leq (1 - \alpha)^t Var_{\pi}\phi$

Δείξαμε πως  $Var_{\pi}\phi - Var_{\pi}[P_{\phi}] \geq \mathcal{E}_{\pi}(\phi, \phi)$

Από ορισμό της σταθεράς του Poincaré  $Var_{\pi}[P_{\phi}] \leq Var_{\pi}\phi(1 - \alpha)$  και επαγωγικά προκύπτει το ζητούμενο

Τώρα είμαστε σε θέση να αποδείξουμε το θεώρημα

Έστω  $A \subset \Omega$ . Ορίζουμε

$$\phi_A(x) = \begin{cases} 1 & \text{εάν } x \in A, \\ 0 & \text{αλλιώς} \end{cases}$$

Για τη συνάρτηση που ορίσαμε ισχύει

$$Var_{\pi}\phi_A = \sum_{x \in A} \pi(x) \cdot 1 \leq 1$$

και αντικαθιστώντας στο βήμα 2

$$Var_{\pi}[P_{\phi_A}^t] \leq (1 - \alpha)^t \leq e^{-\alpha t}$$

Θέτοντας  $t = \frac{1}{\alpha}(\ln \pi(x)^{-1} + 2 \ln \varepsilon^{-1})$  στη παραπάνω σχέση προκύπτει

$$Var_{\pi}[P_{\phi_A}^t] \leq \varepsilon^2 \cdot \pi(x)$$

Από την άλλη πλευρά, διαλέγοντας κάποιο  $x \in A$

$$Var_{\pi}[P_{\phi_A}^t] \geq \pi(x)(P_{\phi_A}^t - E_{\pi}[P_{\phi_A}^t])^2 \\ \pi(x)(P_x^t(A) - \pi(A))^2$$

όπου η τελευταία σχέση προκύπτει από τον ορισμό της συνάρτησης  $\phi_A(x)$  και από το γεγονός πως  $E_{\pi}[P_{\phi}] = E_{\pi}\phi$ .

Από την επιλογή του χρόνου  $t$  έχουμε

$$(P_x^t(A) - \pi(A))^2 \leq \varepsilon^2$$

και επειδή το σύνολο είναι αυθαίρετο έχουμε  $\|P_x^t - \pi\| \leq \varepsilon$

Για να βρούμε ένα άνω φράγμα στο χρόνο μίξης θα εισάγουμε κάποιες έννοιες από τα δίκτυα που συνδέονται με τις αντιστρέψιμες μαρκοβιανές αλυσίδες.

**Ορισμός 19** Ορίζουμε αγωγιμότητα μίας ακμής  $v = (x, y)$  τη ποσότητα

$$C(v) := \pi(x)Pr(x, y) = Pr_{\pi}(x, y)$$

Πολλές φορές η αγωγιμότητα καλείται και εργοδική ροή και έχει τη φυσική σημασία της πιθανότητας η αλυσίδα μας να μεταβεί από τη κατάσταση  $x$  στη κατάσταση  $y$  ξεκινώντας από τη στάσιμη κατανομή.

**Ορισμός 20** Καλούμε ζήτηση  $D(x, y) := \pi(x)\pi(y) \forall x, y \in \Omega$

Επιπλέον, θέτοντας  $P = \cup_{xy} P_{xy}$  όπου  $P_{xy}$  το σύνολο όλων των απλών μονοπατιών από την κατάσταση  $x$  στην κατάσταση  $y$  είμαστε σε θέση να ορίσουμε και τη ροή.

**Ορισμός 21** Ροή είναι μια συνάρτηση  $f : P \rightarrow \mathbb{R}^+ \cup \{0\}$  τέτοια ώστε

$$\sum_{p \in P_{xy}} f(p) = D(x, y)$$

Τέλος, ορίζουμε ροή μίας ακμής τη ποσότητα  $f(e) = \sum_{e \in p} f(p)$ , κόστος μιας ροής  $\rho(f) := \max_e \frac{f(e)}{C(e)}$  και μήκος μιας ροής αντίστοιχα  $l(f) := \max_{p: f(p) > 0} |p|$ . Χρειάζεται να βρούμε κάποια σχέση μεταξύ της σταθεράς του Poincaré και κάποιων από τους όρους που ορίσαμε ώστε να μπορούμε να βρούμε κάποια σχέση για να βρούμε κάποιο φράγμα για το χρόνο μίξης. Στην πραγματικότητα ισχύει

**Θεώρημα 7.2** Για κάθε εργοδική αλυσίδα και κάθε ροή  $f$

$$\alpha \geq \frac{1}{\rho(f)l(f)}$$

**Απόδειξη**

$$\begin{aligned} 2Var_{\pi}\phi &= \sum_{xy} \pi(x)\pi(y)(\phi(x) - \phi(y))^2 \\ &= \sum_{xy} \sum_{p \in P_{xy}} f(p)(\phi(x) - \phi(y))^2 \end{aligned}$$

Παρατηρούμε ότι για κάθε μονοπάτι  $p \in P_{xy}$

$$\phi(x) - \phi(y) = \sum_{(u,v) \in p} (\phi(u) - \phi(v))$$

Έτσι η πρώτη σχέση σύμφωνα με την προηγούμενη μπορεί να γραφτεί ως:

$$\begin{aligned}
& \sum_{xy} \sum_{p \in P_{xy}} f(p) \left[ \sum_{(u,v) \in p} (\phi(x) - \phi(y)) \right]^2 \\
& \leq \sum_{xy} \sum_{p \in P_{xy}} f(p) |p| \sum_{(u,v) \in p} (\phi(u) - \phi(v))^2 \\
& = \sum_{e=(u,v)} (\phi(u) - \phi(v))^2 \sum_{e \in p} f(p) |p| \\
& \leq l(f) \sum_{e \in (u,v)} (\phi(u) - \phi(v))^2 \sum_{p \in e} f(p)
\end{aligned}$$

Επειδή, για μία ακμή ισχύει  $\sum_{e \in p} f(p) = f(e)$  και  $\rho(f) := \max_e \frac{f(e)}{C(e)}$

$$\begin{aligned}
2\text{Var}_\pi \phi & \leq l(f) \rho(f) \sum_{e \in (u,v)} (\phi(u) - \phi(v))^2 C(e) \\
& = 2l(f) \rho(f) (E)_\pi(\phi, \phi)
\end{aligned}$$

Άρα

$$\frac{E_\pi(\phi, \phi)}{\text{Var}_\pi \phi} \geq \frac{1}{l(f) \rho(f)}$$

και προκύπτει το ζτούμενο Το άμεσο συμπέρασμα του προηγούμενο θεωρήματος είναι

**Θεώρημα 7.3** Για κάθε ράθυμη και εργοδική μαρκοβιανή αλυσίδα ισχύει

$$\tau_x(\varepsilon) \leq \rho(f) l(f) (2 \ln \varepsilon^{-1} + \ln \pi(x)^{-1}) \quad \forall x \in \Omega \quad \forall \varepsilon > 0$$

Στη δική μας περίπτωση  $C(e) = \pi(x)P(x, y) = \frac{1}{4N}$  για κάθε ακμή και η ζήτηση είναι  $D(x, y) = \frac{1}{N^2} \quad \forall x, y \in \Omega$ .

Επειδή για κάθε ακμή υπάρχει μόνο ένα μονοπάτι και η συνάρτηση  $f = 1/N^2 \cdot |P|$  είναι μια αποδεκτή ροή

$$f((i, i+1)) = i(N-i) \frac{1}{N^2} \leq \frac{1}{4}$$

Συνεπώς,

$$\rho(f) = \max_e \frac{f(e)}{C(e)} \leq \frac{\frac{1}{4}}{\frac{1}{4N}} = N$$

και το μήκος είναι

$$l(f) = N$$

Γίνεται προφανώς πως η σταθερά του Poincaré είναι μεγαλύτερη από  $\frac{1}{N^2}$  και ο χρόνος μίξης είναι

$$t_{mix} = O(N^2 \log N)$$



## 8 Αριθμητικά πειράματα

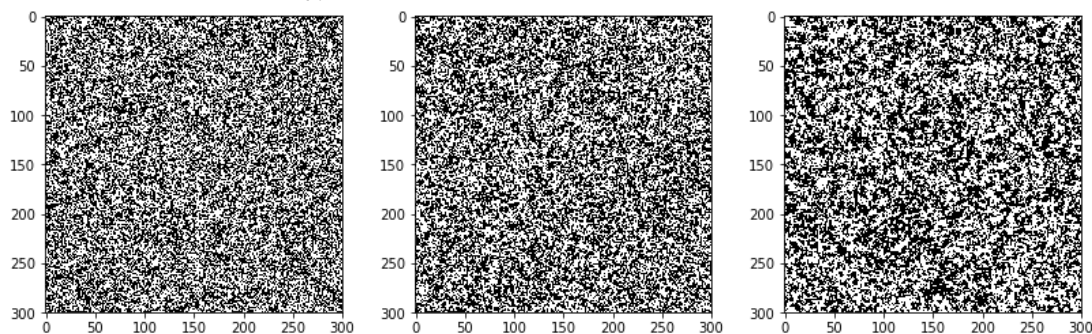
### 8.1 Μοντέλο Ising

Όπως αναφέραμε στην ενότητα στην οποία μελετούσαμε το χρόνο μίξης για το μοντέλο του Ising υπάρχει μια αλλαγή φάση για τη θερμοκρασία  $\beta_c = \frac{1}{2} \ln(1 + \sqrt{2})$  με αποτέλεσμα ο αλγόριθμος να μην μπορεί να μας δώσει αποτέλεσμα σε εύλογο χρονικό διάστημα για θερμοκρασίες μεγαλύτερες από  $\beta_c$  όπου ο χρόνος μίξης είναι  $O(e^n)$ .

Το μήκος της μεγαλύτερης αλυσίδας με τη διάταξη που έχουμε ορίσει είναι  $n^2$ . Συνεπώς,

$$E[T^{up}] = \begin{cases} O(n^2 \log(n^2)(\log_2(n^2))) & \text{εάν } \beta < \beta_c, \\ O(e^{\Omega n} \ln n^2) & \text{αλλιώς} \end{cases}$$

Θα προσομειώσουμε τον αλγόριθμο για το πλέγμα  $300 \times 300$  για θερμοκρασίες 0.10, 0.20 και 0.30 αντίστοιχα

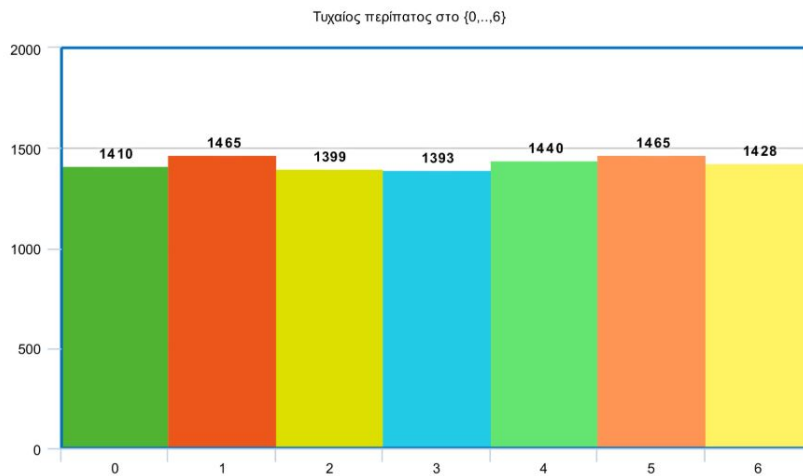


### 8.2 Τυχαίος περίπατος στο σύνολο $\{0, \dots, n\}$

Σύμφωνα με προηγούμενη ανάλυση

$$E[T^{up}] \leq O(n^2 \log n \log n)$$

Δοκιμάζουμε για  $n = 7$  και τρέχουμε τον αλγόριθμο για 10.000 φορές και προκύπτει το εξής ιστόγραμμα



Ενώ ο έλεγχος  $X^2$  μας δίνει το εξής αποτέλεσμα  $p - value = 0.7187$ .

Εάν παρατηρήσουμε καλύτερα μπορούμε να ορίσουμε μια μονότονη περίπτωση του αλγορίθμου με το πίνακα που μελετήσαμε. Ξεκινάμε δύο αλυσίδες από τις τιμές 0 και  $n$ . Σε κάθε βήμα επιλέγουμε τον ίδιο αριθμό για να καθορίσει το επόμενο βήμα μας με μία διαφορά. Αν βρισκόμαστε σε οποιαδήποτε κατάσταση εκτός της τελευταίας,  $x = n$ , επιλέγουμε το τύπο με τα μερικά αθροίσματα. Ενώ αν βρισκόμαστε στην τελευταία κατάσταση επιλέγουμε να μείνουμε στην ίδια θέση εάν  $r \geq 0.75$  και να μεταβούμε στη γειτονική κατάσταση σε αντίθετη περίπτωση. Σχηματικά,

$$\begin{aligned} \text{Για } x_n = 0 \text{ Εάν } r < 0.75 \quad x_{n+1} = 0 \\ r \geq 0.75 \quad x_{n+1} = 1 \end{aligned}$$

$$\begin{aligned} \text{Για } x_n = 1, \dots, n-1 \text{ Εάν } r < 0.25 \quad x_{n+1} = x_n - 1 \\ r < 0.75 \quad x_{n+1} = x_n \\ r \geq 0.75 \quad x_{n+1} = x_n + 1 \end{aligned}$$

$$\begin{aligned} \text{Για } x_n = n \text{ Εάν } r < 0.25 \quad x_{n+1} = n - 1 \\ r \geq 0.25 \quad x_{n+1} = n \end{aligned}$$

Τρέχοντας τον αλγόριθμο για 100.000 επαναλήψεις προκύπτει τα παρακάτω αποτελέσματα [1396,1432,1398,1443,1471,1429,1431] και  $p - value = 0.8317$ . Ο αλγόριθμος στη μονότονη περίπτωση είναι πιο γρήγορος και τα αποτελέσματα έχουν πιο μεγάλη ισχύ αλλά πάλι μπορούμε να απορρίψουμε την υπόθεση ότι τα δεδομένα προέρχονται από την διακριτή ομοιόμορφη κατανομή με πιθανότητα  $1/7$ .

### 8.3 Μοντέλο Ehrenfest

Στο μοντέλο του Ehrenfest υποθέτουμε πώς έχουμε  $N$  σωματίδια και δύο κουτιά, που θα τα αποκαλούμε δεξί και αριστερό κουτί αντίστοιχα, τα οποία επικοινωνούν με τον εξής τρόπο:

Κάθε χρονική στιγμή διαλέγουμε να στείλουμε ένα και μόνο σωματίδιο από το ένα κουτί στο άλλο με πιθανότητα ανάλογη του βάρους κάθε κουτιού. Δηλαδή, αν το δεξί κουτί έχει  $x$  μπάλες τότε το άλλο έχει  $N - x$  και οι πιθανότητες μετάβασης για το πλήθος των σωματιδίων είναι

$$p(x, x + 1) = \frac{N - x}{N} \quad \text{και} \quad p(x, x - 1) = \frac{x}{N}$$

Είναι προφανές πως η αλυσίδα μας είναι μη υποβιβάσιμη και από τις εξισώσεις ακριβούς ισορροπίας μπορούμε να ελέγξουμε πως η μοναδική στάσιμη κατανομή της είναι η διωνύμικη με πιθανότητα επιτυχία  $p = \frac{1}{2}$

$$\pi(n) = \frac{1}{2^N} \binom{N}{n}$$

Πλέον, έχουμε δύο προβλήματα να αντιμετωπίσουμε.

1. Πρέπει να κάνουμε την αλυσίδα μας να είναι αperiοδική.
2. Εάν είναι εφικτό να δημιουργήσουμε μία μονότονη περίπτωση δειγματοληψίας από τη στάσιμη κατανομή της αλυσίδας.

Στη περίπτωση που είναι αδύνατη η μονότονη περίπτωση περίπτωση, όταν το πλήθος των σωματιδίων είναι 100 χρειάζεται να ελέγχουμε για τη μετάβαση 100 αλυσίδων.

Η απάντηση τελικά είναι θετική και για τα δύο παραπάνω προβλήματα.

Για να κάνουμε την αλυσίδα μας αperiοδική επιλέγουμε σε κάθε βήμα να διαλέγουμε ένα από τα δύο κουτιά με πιθανότητα  $\frac{1}{2}$ . Αφού επιλέξουμε το κουτί που επιθυμούμε να δοκιμάσουμε να αλλάξουμε τη θέση ενός σωματιδίου ανβάλογα με το βάρος κάθε κουτιού.

Σχηματικά

$$X_{n+1} = \begin{cases} X_n + 1 & \text{εάν } U_{n+1} \geq X_n/N \text{ και } V_{n+1} = 0 \\ X_n - 1 & \text{εάν } U_{n+1} \leq X_n/N \text{ και } V_{n+1} = 1 \\ X_n & \text{αλλιώς} \end{cases}$$

όπου 0 κωδικοποιεί την επιλογή του δεξιού κουτιού και 1 του αριστερού και οι τυχαίες μεταβλητές  $U, V$  είναι ανεξάρτητες.

Εδώ οι πιθανότητες μετάβασης τη χρονική στιγμή  $n$  είναι

$$Pr[x, x - 1] = Pr[U_n \leq \frac{x}{N} \& V_n = 1] = \frac{x}{2N}$$

$$Pr[x, x + 1] = Pr[U_n \geq \frac{x}{N} \& V_n = 0] = \frac{N - x}{2N}$$

$$Pr[x, x] = \frac{1}{2}$$

Τώρα μπορούμε να δώσουμε απάντηση και στο δεύτερο πρόβλημα. Είναι εφικτή η μονότονη περίπτωση του αλγορίθμου εάν ξεκινήσουμε από δύο μοντέλα κάθε φορά επιλέγουμε να αλλάξουμε σωματίδια από τα ίδια κουτιά.

Αφού το κουτί 1 είναι πιο ισχυρό ως προς τη διάταξη που ορίσαμε,  $box1 \succeq box2$ , συνάγεται ότι

$$\begin{aligned} r_{box1} &\geq r_{box2} \\ l_{box2} &\geq l_{box1} \end{aligned}$$

Μπορούμε να ισχυριστούμε ότι αν έχουμε 5,3 και 2,6 στα κουτιά θα το συμβολίζουμε ως (5,-3) και (2,-6).

Εάν τα κουτιά είναι ίδια τότε θα 'κινούνται' μαζί.

Εάν διαφέρουν έστω κατά ένα σωματίδιο:

**1η περίπτωση:**  $r < r_{box1}/N$  και αλλάζουμε μόνο σωματίδια από το δεξί κουτί

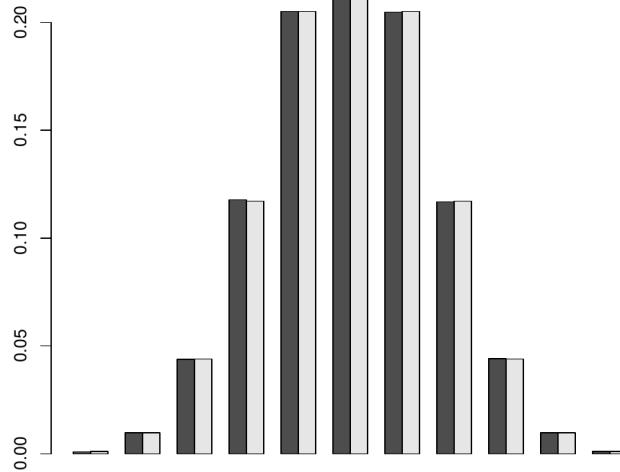
Το κουτί 2 μπορεί να μην πραγματοποιήσει αλλαγή και να ταυτίζεται πλέον με το πρώτο κουτί ή πραγματοποιεί και η διάταξη διατηρείται.

**2η περίπτωση:**  $r > l_{box1}/N$  και αλλάζουμε μόνο σωματίδια από το αριστερό κουτί

Το κουτί 1 μπορεί να μην πραγματοποιήσει αλλαγή και να ταυτίζεται πλέον με το πρώτο κουτί ή πραγματοποιεί και η διάταξη διατηρείται.

Επαναλαμβάνοντας τον αλγόριθμο 1.000.000 φορές προκύπτει το παρακάτω αποτέλεσμα.

Με γκρι φαίνονται οι θεωρητικές τιμές και με μάυρο χρώμα οι τιμές εξόδου του αλγορίθμου.



## 9 Κώδικες

**Algorithm for random walk on  $\{0, \dots, n\}$**

```
import numpy as np
import random as random
import math

markov_table=np.matrix('0.75 0.25 0.0 0.0 0 0 0; 0.25 0.5 0.25 0.0 0 0 0 ;
0.0 0.25 0.5 0.25 0 0 0;0 0 0.25 0.5 0.25 0 0 ;0 0 0 0.25 0.5 0.5 0 ;0 0 0 0 0.25
0.5 0.25;0.0 0.0 0 0 0 0.25 0.75')
n=math.sqrt(markov_table.size)
print(n)
coalescence=False
random_numbers=[0]
current_position=[]
random.uniform(0,1)
T=-1
total_pos=[]

def initial_list(alist):
```

```

    alist=[]
    alist=np.arange(n)
    return alist
def initial_list0(alist):
    alist=[]
    alist=np.zeros(int(n))
    return alist
def check_coalescence(a_list):
    return all(x == a_list[0] for x in a_list)
def last_move(init_pos,time,random_numbers,markov_table):
    temp=init_pos
    for i in range(time,0,+1):
        r=random_numbers[-i]
        temp=np.argmin(np.cumsum(markov_table[temp])<r)
    return temp

current_position =initial_list(current_position)
total_pos=initial_list0(total_pos)
for k in range(1000):
    coalescence=False
    current_position =initial_list(current_position)
    random_numbers=[0]
    T=-1
    while(coalescence==False):
        current_position =initial_list(current_position)
        random_numbers.append(random.uniform(0,1))
        for i in range(int(n)):
            temp1=last_move(i,T,random_numbers,markov_table)
            current_position[i]=temp1
            coalescence=check_coalescence(current_position)
        T=T-1
    total_pos[int(current_position[1])] +=1
l=total_pos
print(total_pos)
print('end')

```

### Algorithm for Ising model

```
import numpy as np
import random as random
import math

grid_size=10
n=grid_size
nNodes=np.power(grid_size,2)
print('The spin system has ',nNodes ,'nodes')
coalescence=False
T=-1
max_state=[]
min_state=[]
random_numbers=[-1]
position=[-1]
beta=2
sum=0
def initial_list(list1,list2):
    list1=np.empty(nNodes)
    list1.fill(1)
    list2=np.empty(nNodes)
    list2.fill(-1)
    return list1 , list2

def check_coalescence(list1,list2):
    return np.array_equal(list1,list2)

def get_neighbors(pos,list1):
    neighbors=[]
    sum=0
    if ((pos%n)!=0):
        neighbors.append(pos-1)
    if ((pos
        neighbors.append(pos+1)
    if ((pos+n)<=nNodes-1):
        neighbors.append(pos+n)
    if ((pos-n)>=0):
        neighbors.append(pos-n)
    for i in neighbors:
        sum=sum+1.0*list1[i]
```

```

return sum

def update(time,random_numbers,position,alist1,alist2):
    for i in range(time,0,+1):
        r=random_numbers[-i]
        thesi=position[-i]
        sum1=get_neighbors(thesi,alist1)
        sum2=get_neighbors(thesi,alist2)
        ratio1=math.exp(2.0*beta*(sum1))/(math.exp(2.0*beta*(sum1))+1)
        ratio2=math.exp(2.0*beta*(sum2))/(math.exp(2.0*beta*(sum2))+1)
        if (r<ratio1):
            alist1[thesi]=+1
        else:
            alist1[thesi]=-1
        if (r<ratio2):
            alist2[thesi]=+1
        else:
            alist2[thesi]=-1
    return alist1,alist2

while(coalescence==False):
    max_state,min_state=initial_list(max_state,min_state)
    for i in range(T,10*T,-1):
        random_numbers.append(random.uniform(0,1))
        position.append(random.randint(0,nNodes-1))
    max_state,min_state=update(T,random_numbers,position,max_state,min_state)
    coalescence=check_coalescence(max_state,min_state)
    T=10*T

```



### Algorithm for Ehrenfest model

```
import numpy as np
import random

coalescence=False
T=-1
random_numbers=[-1]
max_state=[]
min_state=[]
box=[-1]
coalescence=False
n=222
T=-1
def check_coalescence(right1,left1,right2,left2):
    if ((right1==right2) and (left1==left2)):
        return True
    else:
        return False
def initial(n):
    right1=n
    left1=0
    right2=0
    left2=n
    return right1,left1,right2,left2
    def update_boxes(right,left,random,w_box):
        ratio=(right*1.0)/n
        if(random<ratio)and(w_box==0):
            right-=1
            left+=1
        elif(random>(ratio))and(w_box==1):
            left-=1
            right+=1
    return right,left

while(coalescence==False):
    right1,left1,right2,left2=initial(n)
    for i in range(T,2*T,-1):
        random_numbers.append(random.uniform(0,1))
        box.append(int(np.random.choice(2, 1)))
    for i in range(T,0,+1):
```

```
w_box=box[-i]
right1,left1=update_boxes(right1,left1,random_numbers[-i],w_box)
right2,left2=update_boxes(right2,left2,random_numbers[-i],w_box)
coalescence=check_coalescence(right1,left1,right2,left2)
T=2*T
```

## Βιβλιογραφία

- [1] David A. Levin, Yuval Peres, Elizabeth L. Wilmer (2009). "Markov Chains and Mixing Times" . American Mathematical Society,
- [2] Λουλάκης Μιχαήλ (2015). "Στοχαστικές Διαδικασίες " . Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών
- [3] James Gary Propp and David Bruce Wilson (1996). "Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics " , Random Structure and Algorithms 9
- [4] Xeni K. Dimakos (2001). "A guide to Exact Simulation" International Statistical Review / Revue Internationale de Statistique Vol. 69, No. 1 , pp. 27-48
- [5] Constantinos Daskalakis Notes in Probability and Computation <http://people.csail.mit.edu/costis/6896sp11/>
- [6] Haim Kaplan. Seminar on Markov Chains and Mixing Times [http://www.cs.tau.ac.il/~haimk/mcmc-seminar/cftp\\_presentation.pdf](http://www.cs.tau.ac.il/~haimk/mcmc-seminar/cftp_presentation.pdf)
- [7] Häggström Olle (2002). Finite Markov Chains and Algorithmic Applications. London Mathematical Society Student texts vol. 52, Cambridge University Press