



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

**Προσομοίωση Monte Carlo Αλυσίδων Markov:
Αλγόριθμοι Metropolis & Metropolis-Hastings
Προσομοιωμένη Ανόπτηση – Simulated Annealing
Markov Random Fields, Ising Model**

καθ. Βασίλης Μάγκλαρης
maglaris@netmode.ntua.gr
www.netmode.ntua.gr

Video Conference μέσω Cisco Webex

Πέμπτη 2/4/2020

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στατιστική Μηχανική και Μηχανική Μάθηση (Επανάληψη)

- Εκτίμηση (inference) στατιστικών ιδιοτήτων δειγματικών στοιχείων εισόδου $\mathbf{x}(i)$ σε συστήματα **Μηχανικής Μάθησης** που αυτό-οργανώνονται χωρίς επίβλεψη (**Unsupervised Learning**) για ταξινόμηση, συμπλήρωση ατελειών... από δείγμα μάθησης μέσω **γενίκευσης**
- Επιλογή μοντέλων **Στατιστικής Μηχανικής** για κωδικοποίηση στατιστικών ιδιοτήτων m χαρακτηριστικών (**features**) **μεγάλου** πλήθους N δειγματικών στοιχείων μάθησης που προβάλλονται σε τυχαίες μεταβλητές του δείγματος εισόδου. Τα χαρακτηριστικά των N στοιχείων του δείγματος αντιστοιχίζονται στις τιμές των συντεταγμένων των διανυσμάτων εισόδου διαστάσεως ($m \times 1$) του περιβάλλοντος δειγματικού χώρου:
$$\mathbf{x}(i) = [x_1(i), x_2(i), \dots, x_m(i)]^T, i = 1, 2, \dots, N$$
- Οριακή προσέγγιση στατιστικής κατανομής χαρακτηριστικών των δειγματικών στοιχείων $\mathbf{x}(i)$ με αντιστοίχιση μακροσκοπικών μοντέλων συστημάτων **φυσικής μηχανικής σε δυναμική ισορροπία** κάτω από ορισμένη θερμοκρασία. Αναλογία εννοιών θερμοκρασίας και εντροπίας (αταξίας) με καταστάσεις και παραμέτρους ελέγχου συστημάτων **Μηχανικής Μάθησης**
- Πρωτοποριακή εφαρμογή: **Μηχανή Boltzmann** (**Hinton – Sejnowski**, 1983) για επεξεργασία και ταξινόμηση εικόνων μέσω στατιστικής **γενίκευσης** δειγμάτων μάθησης

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στατιστική Μηχανική: Κατανομή Gibbs, Partition Function, Εντροπία (Επανάληψη)

Θερμική Ισορροπία Φυσικού Συστήματος με πολλούς Βαθμούς Ελευθερίας

Φυσικό σύστημα με πολλούς βαθμούς ελευθερίας ισορροπεί σε T βαθμούς Kelvin σε καταστάσεις i , ενέργειας E_i

Οι πιθανότητες ισορροπίας p_i (σχετική συχνότητα εμφάνισης της i) είναι αντιστρόφως ανάλογες των E_i :

$$p_i \propto \exp\left(-\frac{E_i}{T}\right), \quad \frac{p_i}{p_j} = \exp\left(-\frac{E_i - E_j}{T}\right)$$

Οι $p_i \geq 0$, $\sum_i p_i = 1$ ακολουθούν **κατανομή Gibbs** (1902) ή **Boltzmann** (1868) με Z τη Σταθερά Κανονικοποίησης (*Zustadsumme*) που αποκαλείται **Συνάρτηση Κερματισμού (Partition Function)**

$$p_i = \frac{1}{Z} \exp\left(-\frac{E_i}{T}\right), \quad Z = \sum_i \exp\left(-\frac{E_i}{T}\right)$$

Καταστάσεις i χαμηλής ενέργειας E_i έχουν μεγαλύτερη πιθανότητα να συμβαίνουν από καταστάσεις υψηλής ενέργειας. Η ενέργεια της κατάστασης i είναι $E_i = -T \log(Zp_i)$, η μέση ενέργεια $\langle E \rangle = \sum_i p_i E_i$ και η συνολική Ελεύθερη Ενέργεια F (**Helmholtz Free Energy**):

$$F = -T \log Z \Rightarrow \langle E \rangle - F = -T \sum_i p_i \log p_i$$

Η **εντροπία** του συστήματος είναι $H \triangleq \sum_i p_i \log p_i \Rightarrow \langle E \rangle - F = TH$ ή $F = \langle E \rangle - TH$

Αρχή της Ελάχιστης Ελεύθερης Ενέργειας (**Landau & Lifshitz, 1980**)

Σε θερμική ισορροπία η εντροπία τείνει στη μέγιστη τιμή, η F παίρνει την ελάχιστη τιμή της και οι καταστάσεις ακολουθούν την κατανομή **Gibbs**

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αλυσίδες Markov Διακριτής Κατάστασης (Markov Chains) (Επανάληψη)

Ορισμοί Αλυσίδων Markov, Μεταβάσεις Καταστάσεων

Θεωρούμε **Στοχαστικές Ανεξίξεις Διακριτού Χρόνου και Διακριτής Κατάστασης** $X_n = i$ με μεταβάσεις $(X_n = i) \rightarrow (X_{n+1} = j)$ ανεξάρτητες του παρελθόντος σε διακριτά βήματα

Πιθανότητες μετάβασης σε ένα βήμα είναι σταθερές και ανεξάρτητες της χρονικής στιγμής n :

$$p_{ij} = P(X_{n+1} = j | X_n = i) \geq 0, \sum_j p_{ij} = 1 \quad \forall i$$

Αν $i \leq K$ οι πιθανότητες μετάβασης δίνονται από την μήτρα \mathbf{P} ($K \times K$) με άθροισμα στοιχείων γραμμών $\sum_j p_{ij} = 1$ (**στοχαστική μήτρα**)

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1K} \\ \vdots & \ddots & \vdots \\ p_{K1} & \cdots & p_{KK} \end{bmatrix}$$

Οι πιθανότητες μεταβάσεων σε m βήματα είναι $p_{ij}^{(m)} = P(X_{n+m} = j | X_n = i), m = 1, 2, \dots$

$$p_{ij}^{(m+1)} = \sum_k p_{ik}^{(m)} p_{kj} \quad \text{και} \quad p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}, \quad m, n = 1, 2, \dots$$

(**Ταυτότητα Chapman-Kolmogorov**)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ιδιότητες Καταστάσεων Αλυσίδων Markov (3/4) (Επανάληψη)

- Οι καταστάσεις i των οποίων οι πιθανότητες σταθερής κατάστασης π_i μπορούν να υπολογισθούν σαν αναλογία χρόνου στην i σε άπειρο χρονικό ορίζοντα εξέλιξης ενός δείγματος της αλυσίδας Markov X_i ορίζονται σαν **εργοδικές καταστάσεις**
- Οι πιθανότητες π_i των εργοδικών καταστάσεων είναι **εργοδικές πιθανότητες** και η X_i ορίζεται σαν **εργοδική αλυσίδα Markov**. Μια **irreducible** μη περιοδική Markov Chain είναι πάντα εργοδική
- **Σύγκλιση πιθανοτήτων καταστάσεων σε αναλλοίωτη κατανομή εργοδικών πιθανοτήτων**
Οι πιθανότητες καταστάσεων εργοδικής αλυσίδας Markov $X_i, i = 1, 2, \dots, K$ στο βήμα μετάβασης $n = 0, 1, 2, \dots$ ορίζουν διάνυσμα $(1 \times K)$ $\boldsymbol{\pi}^{(n)}$ με εξισώσεις μετάβασης από τη στοχαστική μήτρα \mathbf{P} ($K \times K$) και αρχικές συνθήκες $\boldsymbol{\pi}^{(0)}$

$$\boldsymbol{\pi}^{(n)} = \left[\pi_1^{(n)} \quad \pi_2^{(n)} \quad \dots \quad \pi_K^{(n)} \right], \quad \boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(n-1)} \mathbf{P} = \boldsymbol{\pi}^{(n-2)} \mathbf{P}^2 = \dots = \boldsymbol{\pi}^{(0)} \mathbf{P}^n$$

Στο όριο της εξέλιξης έχουμε σύγκλιση στις εργοδικές πιθανότητες $\boldsymbol{\pi} = [\pi_1 \quad \pi_2 \quad \dots \quad \pi_K]$

$$\lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)} \times \lim_{n \rightarrow \infty} \mathbf{P}^n = \boldsymbol{\pi}^{(0)} \begin{bmatrix} \pi_1 & \dots & \pi_K \\ \vdots & \ddots & \vdots \\ \pi_1 & \dots & \pi_K \end{bmatrix} = \boldsymbol{\pi}^{(0)} \begin{bmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{bmatrix} = \sum_{j=1}^K \pi_j^{(0)} \times \boldsymbol{\pi} = 1 \times \boldsymbol{\pi} = \boldsymbol{\pi}$$

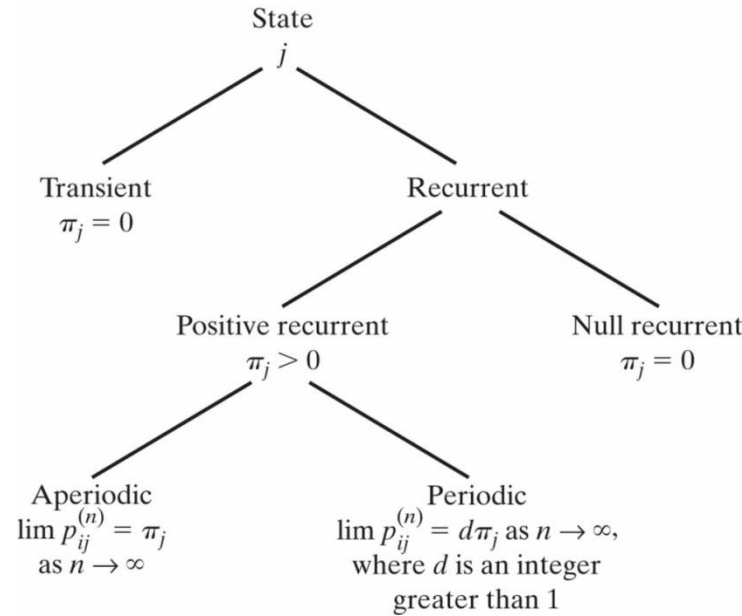
Άρα οι $\boldsymbol{\pi} = \lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(n)}$ είναι **ανεξάρτητες** της αρχικής συνθήκης $\boldsymbol{\pi}^{(0)}$ και υπολογίζονται μέσω του γραμμικού συστήματος **αναλλοίωτης κατανομής εργοδικών πιθανοτήτων**:

$$\pi_j = \sum_{i=1}^K \pi_i p_{ij} \quad j = 1, 2, \dots, K \quad \text{ή} \quad \boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P} \quad \text{και} \quad \sum_{j=1}^K \pi_j = 1$$

(για γραμμική ανεξαρτησία απαιτείται και η εξίσωση **κανονικοποίησης** των πιθανοτήτων)

Ιδιότητες Καταστάσεων Αλυσίδων Markov (4/4) (Επανάληψη)

Σύνοψη Ταξινόμησης Καταστάσεων Αλυσίδων Markov



Χρονικά Αναστρέψιμες Διαδικασίες - Εξισώσεις Ακριβούς Ισορροπίας

- Στη θερμική ισορροπία σύστημα με πιθανότητες καταστάσεων **Gibbs** $\pi_i = \frac{1}{Z} \exp\left(-\frac{E_i}{T}\right)$ όπου Z η **Partition Function**, κάθε δυνατή μετάβαση $i \rightarrow j$ πραγματοποιείται με σχετική συχνότητα ίση με την αντίστροφή της $j \rightarrow i$. Τότε ισχύουν οι εξισώσεις **Ακριβούς Ισορροπίας (Detailed Balance Equations)** $\pi_i p_{ij} = \pi_j p_{ji}$ και οι π_i είναι συμβατές με τις εξισώσεις **Αναλλοίωτων**

Πιθανοτήτων Αλυσίδων Markov: $\sum_{i=1}^K \pi_i p_{ij} = \sum_{i=1}^K \left(\frac{\pi_i}{\pi_j} p_{ij}\right) \pi_j = \sum_{i=1}^K p_{ji} \pi_i = \pi_j$

- Αλυσίδα Markov με ισχύουσες τις εξισώσεις ακριβούς ισορροπίας έχει την ιδιότητα της **χρονικής αντιστρεψιμότητας (Time Reversibility)** με ίδιες εργοδικές πιθανότητες είτε σε μεταβάσεις p_{ij} προς το μέλλον ή προς στο παρελθόν με $\hat{p}_{ji} = \frac{\pi_i}{\pi_j} p_{ij}$.

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (1/5)

Στόχος

Εκτίμηση της κατανομής $\pi(x)$ (ιστογράμματος) τυχαίας μεταβλητής X ή ροπών της μέσω γεννήτριας ακολουθίας n τυχαίων αριθμών (pseudo-random numbers) $X_n = x, n = 1, 2, \dots$

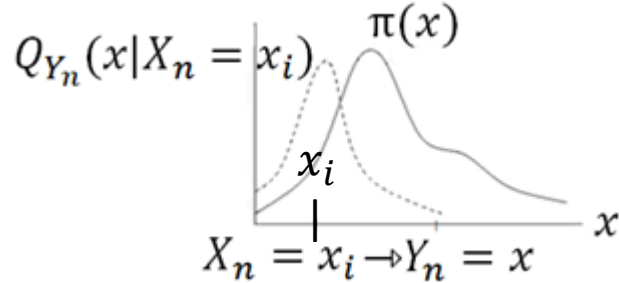
Για μεγάλο πλήθος $n \rightarrow \infty$, $\lim_{n \rightarrow \infty} \frac{f_n(x)}{n} = \lim_{n \rightarrow \infty} P(X_n = x) = \pi(x)$ όπου $f_n(x)$ ο αριθμός εμφάνισης της τιμής x σε n διαδοχικά βήματα προσομοίωσης δειγμάτων

Παραδοχές και Προσέγγιση

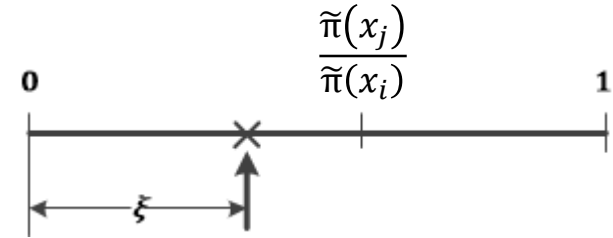
- Η $\pi(x)$ είναι μια κατανομή στόχος (**Target**) που αντανακλά σχετικές συχνότητες εμφάνισης της τιμής x στο δειγματικό χώρο της τυχαίας μεταβλητής X : $\sum_x P(X_n = x) = 1$
- Η $\pi(x)$ έχει γνωστή μορφή $\pi(x) \propto \tilde{\pi}(x)$, π.χ. κατανομή **Gibbs** $\pi(x) \propto \exp\left(-\frac{E(x)}{T}\right)$ αλλά δεν μπορεί να δημιουργήσει τυχαία δείγματα π.χ. λόγω αδυναμίας πλήρους καταγραφής του δειγματικού χώρου και υπολογισμού της σταθεράς κανονικοποίησης (**partition function**)
- Ζητείται γεννήτρια (**αλγόριθμος δημιουργίας**) ακολουθίας $X_n = x, n = 1, 2, \dots$ με κατανομή $\pi(x) \propto \tilde{\pi}(x)$ όπου $\tilde{\pi}(x)$ ανάλογη της συχνότητας εμφάνισης μιας τιμής x
- Η εκτίμηση των σχετικών συχνοτήτων (ιστόγραμμα) $P(X_n = x) \rightarrow \pi(x)$ γίνεται με την καταμέτρηση επισκέψεων **random walk** εργοδικής αλυσίδας **Markov** σε καταστάσεις $\{X_n\}$ που αντιστοιχούν στις τιμές x . Η αλυσίδα προσδιορίζεται ώστε οι εργοδικές της πιθανότητες να είναι ίσες (κατά προσέγγιση) με τις ζητούμενες $\pi(x)$ εξού η μέθοδος εντάσσεται στη κατηγορία προσομοιώσεων **Markov Chain Monte Carlo (MCMC)**
- Η προσομοίωση **MCMC** προσεγγίζει ζητούμενο ιστόγραμμα αλλά με **συσχετισμένα** δείγματα

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (2/5)



$$\pi(x) \propto \tilde{\pi}(x)$$



- Γνωρίζοντας τη μορφή $\tilde{\pi}(x)$ των **εργοδικών πιθανοτήτων στόχου** $\pi(x) \propto \tilde{\pi}(x)$ για παράμετρο $x \in \{0, 1, 2, \dots\}$ ζητείται η δημιουργία (**generation**) χρονικά αναστρέψιμης ακολουθίας **Markov** καταστάσεων $X_n = x$ με κατανομή που να προσεγγίζει τις $\pi(x)$
- Βήμα n : Έστω $X_n = x_i$. Επιλέγω τυχαία μεταβλητή $Y_n = x_j$ με βάση αυθαίρετη **προτεινόμενη** κατανομή (**Proposal Conditional Density**) $Q_{Y_n}(x|X_n = x_i)$. Οι πιθανότητες μετάβασης $(X_n = x_i) \rightarrow (Y_n = x_j)$ επιλέγονται ώστε να ισχύει **συμμετρία**

$$P(Y_n = x_j | X_n = x_i) = P(Y_n = x_i | X_n = x_j)$$

- Αν $\tilde{\pi}(x_j) \geq \tilde{\pi}(x_i)$ η επιλογή $X_n \rightarrow Y_n$ γίνεται αποδεκτή και $X_{n+1} = x_j$
- Αν $\tilde{\pi}(x_j) < \tilde{\pi}(x_i)$ δημιουργώ τυχαίο αριθμό ξ ομοιόμορφα κατανομημένο μεταξύ $(0, 1)$
Αν $\xi < \frac{\tilde{\pi}(x_j)}{\tilde{\pi}(x_i)} = \frac{\pi(x_j)}{\pi(x_i)}$ η επιλογή $X_n \rightarrow Y_n$ γίνεται αποδεκτή $\Rightarrow X_{n+1} = x_j$. Αλλιώς $X_{n+1} = x_i$
- Προσοχή στην επιλογή αρχικής τιμής $X_0 = x$ και προτεινόμενης κατανομής $Q_{Y_n}(x|X_n = x_i)$: Πρέπει να είναι συμμετρική και να καλύπτει ικανό εύρος τιμών x ώστε η αλυσίδα Markov να μην εγκλωβίζεται σε υποσύνολα καταστάσεων. Συνήθεις επιλογές: (1) **Gauss** με μέση τιμή $\mu = x_i$ και $\sigma^2 = 1$ και (2) **Ομοιόμορφη Κατανομή** $x \in (x_i - a, x_i + a)$

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (3/5)

Δημιουργία Τυχαίων Δειγμάτων με Εργοδικές Πιθανότητες Θερμικής Ισορροπίας

Ο αλγόριθμος **Metropolis** δημιουργεί μέσω προσομοίωσης **χρονικά αναστρέψιμης διαδικασίας Markov** ακολουθία καταστάσεων $X_n = x_i$ με σχετική συχνότητα εμφάνισης θερμικής ισορροπίας **Gibbs**: $\pi_i = \frac{1}{Z} \exp\left(-\frac{E_i}{T}\right)$. Η αναλογία των π_i είναι γνωστή αλλά η ακριβής τιμή τους απαιτεί γνώση της **Partition Function** Z μέσω κανονικοποίησης $\sum_i \pi_i = 1$ σε περίπλοκο δειγματικό χώρο

Έστω χρονικά αναστρέψιμη αλυσίδα **Markov** X_n (ακολουθία αριθμών) με **συμμετρικές πιθανότητες μετάβασης ακριβούς ισορροπίας** που μετά από n μεταβάσεις (βήματα) παράγει την κατάσταση x_i . Με τυχαίο τρόπο δημιουργούμε νέα κατάσταση x_j άλλης διαδικασίας Y_n θεωρώντας τη **συνθήκη συμμετρίας** στις μεταβάσεις $X_n \rightarrow Y_n$:

$$P(Y_n = x_j | X_n = x_i) = P(Y_n = x_i | X_n = x_j)$$

Η μετάβαση δημιουργεί διαφορετικό ενέργειας $\Delta E = E_j - E_i$

- Αν $\Delta E < 0$ η μετάβαση οδηγεί σε κατάσταση ($Y_n = x_j$) μικρότερης ενέργειας και γίνεται αποδεκτή: $X_{n+1} := Y_n$
- Αν $\Delta E > 0$ η μετάβαση σε ($Y_n = x_j$) γίνεται αποδεκτή και $X_{n+1} := Y_n$ με πιθανότητα $\exp\left(-\frac{\Delta E}{T}\right)$ όπου T η θερμοκρασία του συστήματος. Αλλιώς $X_{n+1} := X_n$
- Η εξέλιξη της κατάστασης για $\Delta E > 0$ οδηγείται από **προσομοίωση Monte Carlo** μέσω δοκιμών **Bernoulli** δημιουργίας (ψευδο)τυχαίου αριθμού ξ ομοιόμορφα κατανομημένου μεταξύ $(0,1)$

$$\text{Αν } \xi < \exp\left(-\frac{\Delta E}{T}\right), X_{n+1} := Y_n. \text{ Αλλιώς } X_{n+1} := X_n$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (4/5)

Επιλογή Προτεινόμενων Πιθανοτήτων Μετάβασης

Θεωρούμε αλυσίδα Markov με **συμμετρικές** πιθανότητες μετάβασης τ_{ij} :

$$\tau_{ij} \geq 0, \forall i, j \text{ και } \sum_j \tau_{ij} = 1, \forall i$$
$$\tau_{ij} = \tau_{ji}, \forall i, j$$

Ζητείται αλυσίδα Markov X_n καταστάσεων $x_i = 1, 2, \dots, K$ και π_i εργοδικές πιθανότητες **Gibbs** :

$$\pi_i = \frac{1}{Z} \exp\left(-\frac{E_i}{T}\right)$$

Η X_n έχει πιθανότητες μετάβασης p_{ij} που ορίζονται από τις συμμετρικές τ_{ij} ως εξής:

$$p_{ij} = \begin{cases} \tau_{ij} \frac{\pi_j}{\pi_i} & \text{για } \frac{\pi_j}{\pi_i} < 1 \\ \tau_{ij} & \text{για } \frac{\pi_j}{\pi_i} \geq 1 \end{cases}$$

και με τον ορισμό των p_{ii} για τη συνθήκη κανονικοποίησης $\sum_j p_{ij} = 1, \forall i$:

$$p_{ii} = \tau_{ii} + \sum_{j \neq i} \tau_{ij} \left(1 - \frac{\pi_j}{\pi_i}\right)$$

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (5/5)

Επιλογή Πιθανοτήτων Μετάβασης

Οι p_{ij} ορίζονται από τον λόγο $\frac{\pi_j}{\pi_i} = \exp\left(-\frac{E_j - E_i}{T}\right) = \exp\left(-\frac{\Delta E}{T}\right)$, χωρίς υπολογισμό της **partition function** Z , και είναι συμβατές με τις **detailed balance equations**:

- $\Delta E < 0: \left(\frac{\pi_j}{\pi_i}\right) > 1$

$$\pi_i p_{ij} = \pi_i \tau_{ij} = \pi_i \tau_{ji} \text{ και } \pi_j p_{ji} = \pi_j \tau_{ji} \frac{\pi_i}{\pi_j} = \pi_i \tau_{ji} \Rightarrow \pi_j p_{ji} = \pi_i p_{ij}$$

- $\Delta E > 0: \left(\frac{\pi_j}{\pi_i}\right) < 1$

$$\pi_i p_{ij} = \pi_i \tau_{ij} \frac{\pi_j}{\pi_i} = \pi_j \tau_{ij} = \pi_j \tau_{ji} \Rightarrow \pi_j p_{ji} = \pi_i p_{ij}$$

Άρα έχουμε προσομοιώσει διαδικασία X_n που οδηγεί μονοσήμαντα σε **θερμική ισορροπία Gibbs** με βάση τις ενέργειες E_i των καταστάσεων x_i παρακάμπτοντας τον υπολογισμό της Z

Οι μεταβάσεις τ_{ij} που οδηγούν τις p_{ij} προς εργοδικές πιθανότητες $\pi_j = \frac{1}{Z} \exp\left(-\frac{E_j}{T}\right)$ προκύπτουν από την προσομοίωση **Monte Carlo** όπου η διαδικασία εξελίσσεται $i \rightarrow j$ για $E_j - E_i = \Delta E < 0$ ή αν $\Delta E > 0$ εξελίσσεται με πιθανότητα $\exp\left(-\frac{\Delta E}{T}\right)$, αλλιώς με πιθανότητα $1 - \exp\left(-\frac{\Delta E}{T}\right)$ παραμένει στην κατάσταση i (βάση ανεξάρτητων τυχαίων δοκιμών **Bernoulli**)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Γενίκευση του Αλγορίθμου Metropolis: Ο Αλγόριθμος Metropolis-Hastings (1/2)

- Ο αλγόριθμος προσομοίωσης του *Nicholas Metropolis et.al., 1953* υποθέτει συμμετρικές μεταβάσεις και **time reversible** αλυσίδες Markov που συνεπάγονται **detailed balance equations**. Συγκλίνει προς τη κατανομή **Gibbs (Boltzmann)** $\pi_i = \frac{1}{Z} \exp\left(-\frac{E_i}{T}\right)$ με κατάλληλο ορισμό ενεργειών καταστάσεων E_i
- Γενικεύτηκε από τον *W.K. Hastings* το 1970 στον αλγόριθμο **Metropolis-Hastings**

Γενίκευση Προτεινόμενης Κατανομής Μετάβασης: Δεν απαιτείται συμμετρία της

$$Q_{Y_n}(x|X_n = x_i): P(Y_n = x_j|X_n = x_i) \neq P(Y_n = x_i|X_n = x_j)$$

Τροποποίηση Πιθανότητας Αποδοχής $X_{n+1} = x_j$

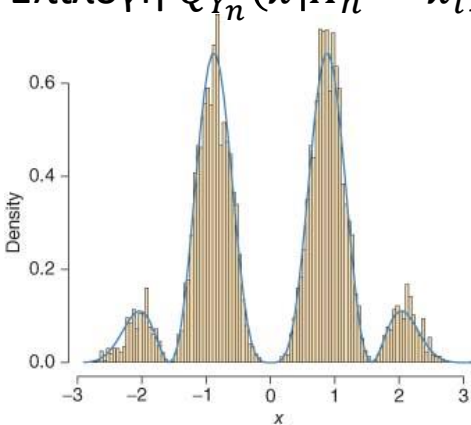
- Αν $\tilde{\pi}(x_j) \times P(Y_n = x_i|X_n = x_j) \geq \tilde{\pi}(x_i) \times P(Y_n = x_j|X_n = x_i)$ η επιλογή $X_n \rightarrow Y_n$ γίνεται αποδεκτή και $X_{n+1} = x_j$
- Αλλιώς δημιουργώ τυχαίο αριθμό ξ ομοιόμορφα κατανεμημένο μεταξύ $(0,1)$
Αν $\xi < \frac{\tilde{\pi}(x_j)}{\tilde{\pi}(x_i)} \times \frac{P(Y_n=x_i|X_n=x_j)}{P(Y_n=x_j|X_n=x_i)}$ η επιλογή $X_n \rightarrow Y_n$ γίνεται αποδεκτή και $X_{n+1} = x_j$. Αν όχι η X_n παραμένει στη τιμή της και $X_{n+1} = x_i$
- Ο αλγόριθμος **Metropolis** είναι ειδική περίπτωση με συμμετρική $Q_{Y_n}(x|X_n = x_i)$

Προβλήματα Αλγορίθμου

- Ο αλγόριθμος **Metropolis-Hastings** δημιουργεί αλυσίδα Markov με εργοδικές πιθανότητες συμβατές με ζητούμενη κατανομή αλλά προσομοιώνει **συσχετισμένες** τυχαίες μεταβλητές
- Η επιλογή κατάλληλης προτεινόμενης κατανομής μεταβάσεων $Q_{Y_n}(x|X_n = x_i)$ και αρχικής κατάστασης $X_0 = x_i$ έχουν ιδιαίτερη σημασία για την ορθή και ταχεία σύγκλιση του αλγορίθμου <https://onlinelibrary.wiley.com/doi/full/10.1002/9781118445112.stat07834>
- Η επιρροή της μεταβατικής κατάστασης σβήνει αν ξεχάσουμε τις πρώτες μεταβάσεις π.χ. $n \leq 1000$

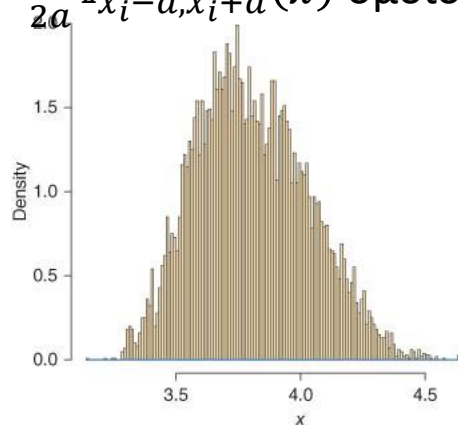
Παράδειγμα: Προσομοίωση κατανομής στόχου (**target**) $\pi(x)$ ανάλογης των $\tilde{\pi}(x) = \sin^2(x) \times \sin^2(2x) \times \varphi(x)$ όπου $\varphi(x)$ κανονική κατανομή Gauss $\mathcal{N}(0,1)$

Επιλογή $Q_{Y_n}(x|X_n = x_i) = \frac{1}{2a} \mathbb{1}_{x_i-a, x_i+a}(x)$ ομοιόμορφη με μέσο όρο x_i και εύρος $2a$



$a = 1, X_0 = 3.14$

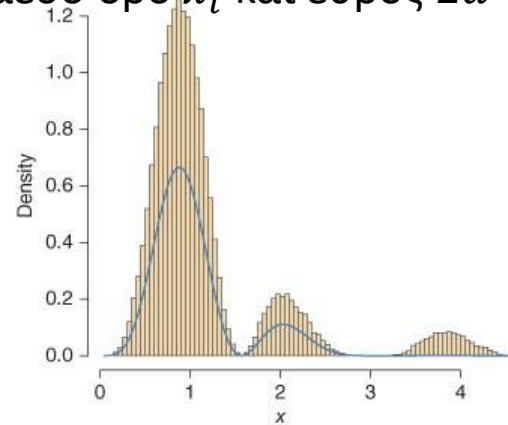
Ορθή σύγκλιση σε 10^4 βήματα



$a = 0.1, X_0 = 3.14$

Αστοχία σε 10^4 βήματα

(η διαδικασία παγιδεύτηκε σε 1 λοβό)



$a = 0.2, X_0 = 3.14$

Μερική αστοχία σε 10^4 βήματα

(η διαδικασία παγιδεύτηκε σε θετικές τιμές)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ενδεικτικές Εφαρμογές & Επεκτάσεις Αλγόριθμου Metropolis-Hastings (1/2)

Ακολουθία Τυχαίου Δείγματος προς Θερμική Ισορροπία

Ο αλγόριθμος **Metropolis** παράγει συσχετισμένες καταστάσεις **time reversible** αλυσίδας Markov X_n με εργοδικές πιθανότητες συμβατές με κατανομή **Gibbs (Boltzmann)** χωρίς γνώση της **partition function** Z . Οι καταστάσεις είναι σχεδόν αδύνατο να προσδιορισθούν με πληρότητα, ιδίως αν είναι **πολυδιάστατες**, και άρα η κανονικοποίησή τους είναι δυσεπίλυτη

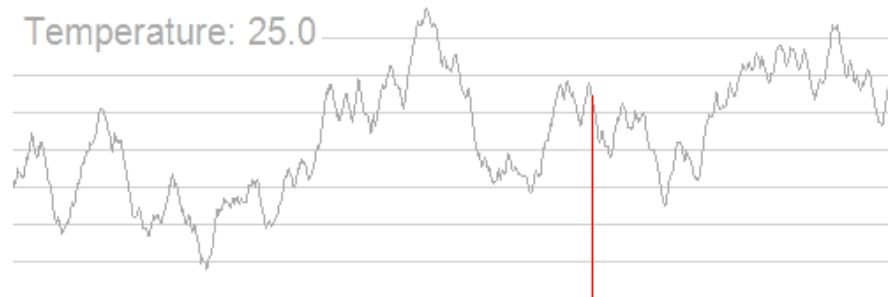
Υπολογισμοί Ολοκληρωμάτων και Στατιστικών Παραμέτρων

Η δημιουργία ακολουθίας τιμών που προσομοιώνουν **Markov Chain Random Walks** μπορεί να αποτελέσει εργαλείο για τον αριθμητικό υπολογισμό ολοκληρωμάτων και ροπών κατανομών **πολυδιάστατων** τυχαίων μεταβλητών χωρίς πλήρη στοιχεία

Εντοπισμός Ακραίων Τιμών, Simulated Annealing

Οι μέθοδοι MCMC παρέχουν τη δυνατότητα να μην εγκλωβισθεί μια επαναληπτική διαδικασία σε περιοχές με τοπικά άκρα αλλά επιτρέπει να διερευνηθούν με κάποια πιθανότητα και εναλλακτικές προς μη ελκυστικές κατευθύνσεις που ένας αλγόριθμος τύπου deepest descent δεν θα εντόπιζε. Αυτή είναι η αρχή αλγορίθμων τύπου **Προσομοιωμένης Ανόπτωσης (Simulated Annealing)** με εφαρμογή σε περίπλοκα προβλήματα συνδυαστικής βελτιστοποίησης, π.χ. **Travelling Salesman Problem**

ΕΥΡΕΣΗ ΣΦΑΙΡΙΚΟΥ ΑΚΡΟΥ: Ο αλγόριθμος ακολουθεί κατευθύνσεις προς το ζητούμενο σφαιρικό (**global**) άκρο (ψηλότερη κορυφή), αλλά επιτρέπει φαινομενικά **λάθος** βήματα με σταδιακά μειούμενη πιθανότητα όσο μειώνεται η «θερμοκρασία»



https://en.wikipedia.org/wiki/File:Hill_Climbing_with_Simulated_Annealing.gif

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

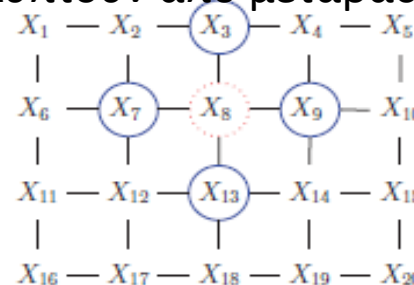
Ενδεικτικές Εφαρμογές & Επεκτάσεις Αλγόριθμου Metropolis-Hastings (2/2)

Μοντέλο Ising – Markov Random Fields, Προσομοίωση Μεταβάσεων Φάσεων

Αφορά σε υλικά με μαγνητικά δίπολα s κατανεμημένα στις κορυφές ενός γράφου G με ακμές μεταξύ τους αν υπάρχει αμφίδρομη αλληλοεπίδραση. Κάθε δίπολο μπορεί να βρίσκεται σε δύο καταστάσεις (**spins**) $y_s \in \{-1, +1\}$. Αν $s \leftrightarrow t$ τα δίπολα s, t είναι γειτονικά και αλληλοεπιδρούν με δύναμη $J_{s,t}$ θετική (π.χ. $J_{s,t} = 1$) αν τείνει να ευθυγραμμίσει τις καταστάσεις y_s και y_t ή αρνητική (π.χ. $J_{s,t} = -1$) αν τις ωθεί προς αντίθετη κατεύθυνση.

Σε ένα **Markov Random Field** οι καταστάσεις ενός διπόλου s μπορεί να επηρεάζονται μόνο από τα αμέσως γειτονικά του $t \in N(s) \subset G$. Αν τα δίπολα είναι κατανεμημένα σε επίπεδο δύο διαστάσεων με τοπολογία πλέγματος (**lattice**) η συνολική κατάσταση σε ισορροπία του συστήματος $\mathbf{y}(G) = (y_1 y_2 \dots y_s y_t \dots)^T$ αφορά σε νόμιμους σχηματισμούς (**configurations**) των επιμέρους καταστάσεων y_s των διπόλων που προκύπτουν από μεταβάσεις με πιθανότητες $P(y_s | \mathbf{y}(G)) = P(y_s | \mathbf{y}(N(s)))$

Η κατάσταση του διπόλου X_8 εξαρτάται μόνο από τα X_3, X_7, X_9 και X_{13}



Το **Markov Random Field** του μοντέλου **Ising** ισορροπεί σε καταστάσεις $\mathbf{y}(G)$ με πιθανότητες κατανομής **Gibbs (Boltzmann)** με ή χωρίς εξωτερική μαγνητική επιρροή h_t στα δίπολα t :

$$P(\mathbf{y}(G)) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{y}(G))}{T}\right) \text{ με } E(\mathbf{y}(G)) \approx -\sum_{s \leftrightarrow t} J_{s,t} y_s y_t - \mu \sum_t h_t y_t$$

Αν $J_{s,t} = J$ για όλα τα γειτονικά ζεύγη $s \leftrightarrow t$ τότε $E(\mathbf{y}(G)) \approx -J \sum_{s \leftrightarrow t} y_s y_t - \mu \sum_t h_t y_t$

Η σύγκλιση σε καταστάσεις ισορροπίας και η σταθερά Z μπορούν να προσεγγισθούν με αλγόριθμο random walk **Metropolis-Hastings**