ε.δε.μ^2

ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσομοίωση Monte Carlo Αλυσίδων Markov:

Αλγόριθμοι Metropolis & Metropolis-Hastings

Markov Random Fields, Ising Model

Προσομοιωμένη Ανόπτηση - Simulated Annealing

Gibbs Sampling

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

Μέσω Πλατφόρμας Webex

Τρίτη 19/3/2024

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στατιστική Μηχανική και Μηχανική Μάθηση (*Επανάληψη*)

- Οριακή προσέγγιση στατιστικής κατανομής χαρακτηριστικών των δειγματικών στοιχείων $\mathbf{x}(n)$ με αντιστοίχηση 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**

- Εκτίμηση (**inference**) στατιστικών ιδιοτήτων δειγματικών στοιχείων εισόδου $\mathbf{x}(n)$ σε συστήματα **Μηχανικής Μάθησης** που αυτό-οργανώνονται χωρίς επίβλεψη (**Unsupervised Learning**) για κωδικοποίηση με εξαγωγή κύριων χαρακτηριστικών, συμπίεση, ταξινόμηση, συμπλήρωση ατελειών, παραγωγή (**generation**) δειγματικών στοιχείων συμβατών με υποθέσεις κατανομής δειγματικού χώρου (**statistical sampling** - **Generative AI**)

- 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 μάθησης αντιστοιχίζονται στις τιμές των συντεταγμένων των διανυσμάτων εισόδου διαστάσεως $(m \text{ x } 1)$ του περιβάλλοντος δειγματικού χώρου:

$$\mathbf{x}(n) = [x_1(n)\ x_2(i)\ \ \dots\ \ x_m(i)]^{\mathrm{T}}, n = 1,2,\dots,N$$

- Πρωτοποριακή εφαρμογή: **Μηχανή Boltzmann** (***Hinton – Sejnowski***, 1983) για επεξεργασία και ταξινόμηση εικόνων μέσω στατιστικής **γενίκευσης** δειγμάτων μάθησης

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (1/6)

Μέθοδοι Προσομοίωσης Markov Chain Monte Carlo (MCMC)

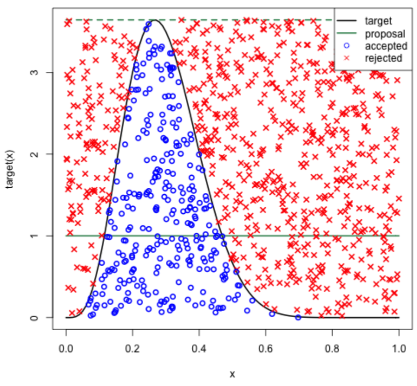
Στόχος η καταγραφή δειγματικών στοιχείων (***sample elements***) $X_n = x, n = 1,2, \ldots$ μέσω προσομοίωσης της κατάστασης x σαν ***random walk*** αλυσίδας ***Markov*** με τυχαίες αλλά ελεγχόμενες μεταβάσεις. Το προκύπτον δείγμα (***sample***) αποτελείται από στοιχεία με ιστόγραμμα συμβατό με κατανομή δειγματικού χώρου διακριτών τυχαίων μεταβλητών

Μία γεννήτρια ***MCMC*** παρακολουθεί την κατάσταση x προσομοιώνοντας την εξέλιξη φυσικού συστήματος προς θερμική ισορροπία με εργοδικές πιθανότητες $\pi(x)$ που προκύπτουν στο όριο σχετικών συχνοτήτων εμφάνισης $f_n(x)$ της x σε n μεταβάσεις:

$$\pi(x) = \lim_{n\to\infty} P(X_n = x) = \lim_{n\to\infty} \frac{f_n(x)}{n}$$

Μια κύρια εφαρμογή: Η εκτίμηση ροπών διακριτών τυχαίων μεταβλητών όταν δεν είναι δυνατή η πλήρης καταγραφή του δειγματικού χώρου και ο υπολογισμός της σταθεράς κανονικοποίησης (***partition function***)

- Οι παραγόμενες τυχαίες μεταβλητές κατανέμονται σύμφωνα με επιθυμητό ιστόγραμμα (κατανομή ισορροπίας) αλλά λόγω παραγωγής τους μέσω ***Markov random walk*** είναι συσχετισμένες (***correlated***)
- Εναλλακτικά για δημιουργία ***sample*** με γνωστή κατανομή και χωρίς συσχέτιση μπορούν να χρησιμοποιηθούν απλές μέθοδοι, π.χ. παραγωγή τυχαίων σημείων και καταγραφή όσων είναι κάτω από την επιφάνεια του ιστογράμματος (***Rejection Sampling***)



https://en.wikipedia.org/wiki/Rejection_sampling#Adaptive_rejection_sampling

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (2/6)

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: $\sum_x P(X_n = x) = 1$

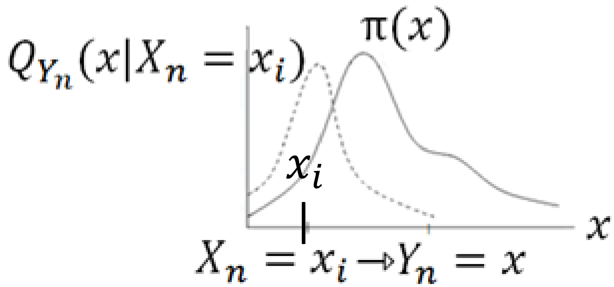
- Η $\pi(x)$ έχει γνωστή μορφή $\pi(x) \propto \tilde{\pi}(x)$, π.χ. κατανομή ***Gibbs*** $\pi(x) \propto \exp\left(-\frac{E(x)}{T}\right)$ αλλά δεν μπορεί να δημιουργήσει τυχαία δειγματικά στοιχεία λόγω αδυναμίας πλήρους καταγραφής του δειγματικού χώρου και υπολογισμού της ***partition function***

- Ζητείται γεννήτρια (***sampling***) ακολουθίας $X_n = x, n = 1,2, \ldots$ με κατανομή $\pi(x) \propto \tilde{\pi}(x)$ όπου $\tilde{\pi}(x)$ ανάλογη της συχνότητας εμφάνισης μιας τιμής x

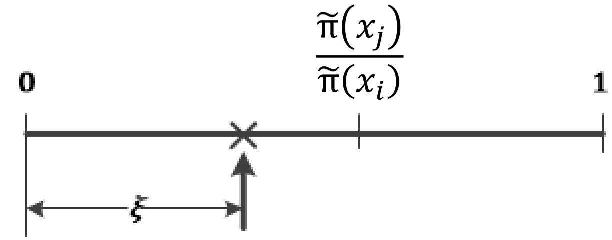
- Η εκτίμηση των σχετικών συχνοτήτων (ιστόγραμμα) $P(X_n = x) \to \pi(x)$ γίνεται με την καταμέτρηση επισκέψεων ***random walk*** εργοδικής αλυσίδας ***Markov*** σε καταστάσεις $\{X_n\}$ που αντιστοιχούν στις τιμές x. Η αλυσίδα προσδιορίζεται ώστε οι εργοδικές της πιθανότητες να είναι ίσες (κατά προσέγγιση) με τις ζητούμενες $\pi(x)$ εξού η μέθοδος εντάσσεται στη κατηγορία προσομοιώσεων ***Markov Chain Monte Carlo*** (**MCMC**)

- Η προσομοίωση **MCMC** προσεγγίζει ζητούμενο ιστόγραμμα αλλά παράγει **συσχετισμένα** δειγματικά στοιχεία (τυχαίες μεταβλητές)

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (3/6)



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



- 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 $(X_n = x_i) \to (Y_n = x_j)$ 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 $X_n \to Y_n$ γίνεται αποδεκτή και $X_{n+1} = x_j$
- Αν $\tilde{\pi}(x_j) < \tilde{\pi}(x_i)$ δημιουργώ τυχαίο αριθμό ξ ομοιόμορφα κατανεμημένο μεταξύ $(0,1)$

Αν $\xi < \frac{\tilde{\pi}(x_j)}{\tilde{\pi}(x_i)} = \frac{\pi(x_j)}{\pi(x_i)}$ η επιλογή $X_n \to Y_n$ 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 (4/6)

Παραγωγή 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 $X_n \to Y_n$:

$$P\left(Y_n = x_j \middle| X_n = x_i\right) = P\left(Y_n = x_i \middle| X_n = x_j\right)$$

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

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

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εξέλιξη Συστήματος προς Θερμική Ισορροπία: Ο Αλγόριθμος Προσομοίωσης Markov Chain Monte Carlo (MCMC) του Metropolis (5/6)

Παραγωγή μέσω Προτεινόμενων Πιθανοτήτων Μετάβασης

Ζητείται η δημιουργία χρονικά αναστρέψιμης αλυσίδας Markov $X_n = x_i, \quad i = 1,2,\dots,K$ που να συγκλίνει σε εργοδικές πιθανότητες π_i κατανομής ***Gibbs*** με παραμέτρους E_i και T :

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

Θεωρούμε νέα αλυσίδα Markov με ***αυθαίρετα*** ορισμένες ***συμμετρικές προτεινόμενες*** πιθανότητες μετάβασης τ_{ij} από την $X_n = x_i$ σε νέα κατάσταση $Y_n = x_j$:

$$\tau_{ij} = P\left(Y_n = x_j \middle| X_n = x_i\right) = P\left(Y_n = x_\iota \middle| X_n = x_j\right) = \tau_{ji}$$

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

$$\tau_{ij} = \tau_{ji}, \forall i,j$$

Η ζητούμενη αλυσίδα X_n έχει πιθανότητες μετάβασης p_{ij} οριζόμενες από τις τ_{ij} ως εξής:

$$p_{ij} = P\left(X_{n+1} = x_j \middle| X_n = x_i\right) = \begin{cases} \tau_{ij}\frac{\pi_j}{\pi_i} = \tau_{ij}\exp(-\frac{E_j - E_i}{T}) & \text{για } \frac{\pi_j}{\pi_i} < 1 \\ \tau_{ij} & \text{για } \frac{\pi_j}{\pi_i} \geq 1 \end{cases}, \quad i \neq j$$

και με p_{ii} όπως προκύπτει από 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 (6/6)

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***:

- $\Delta E < 0$: $\left({}^{\pi_j}/_{\pi_i}\right) > 1$
  $\pi_i p_{ij} = \pi_i \tau_{ij} = \pi_i \tau_{ji}$ και $\pi_j p_{ji} = \pi_j \tau_{ji} \frac{\pi_i}{\pi_j} = \pi_i \tau_{ji} \Rightarrow \pi_j p_{ji} = \pi_i p_{ij}$
- $\Delta E > 0$: $\left({}^{\pi_j}/_{\pi_i}\right) < 1$
  $\pi_i p_{ij} = \pi_i \tau_{ij} \frac{\pi_j}{\pi_i} = \pi_j \tau_{ij} = \pi_j \tau_{ji} \Rightarrow \pi_j p_{ji} = \pi_i p_{ij}$

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 $i \to j$ για $E_j - E_i = \Delta E < 0$ ή αν $\Delta E > 0$ εξελίσσεται με πιθανότητα $\exp(-\frac{\Delta E}{T})$, αλλιώς με πιθανότητα $1 - \exp(-\frac{\Delta E}{T})$ παραμένει στην κατάσταση i (βάση ανεξάρτητων τυχαίων δοκιμών ***Bernoulli***)

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

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

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

- Αν $\tilde{\pi}(x_j) \times P\left(Y_n = x_i \middle| X_n = x_j\right) \geq \tilde{\pi}(x_i) \times P\left(Y_n = x_j \middle| X_n = x_i\right)$ 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 $\xi < \frac{\tilde{\pi}(x_j)}{\tilde{\pi}(x_i)} \times \frac{P\left(Y_n = x_i \middle| X_n = x_j\right)}{P\left(Y_n = x_j \middle| X_n = x_i\right)}$ 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 $Q_{Y_n}(x|X_n = x_i)$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

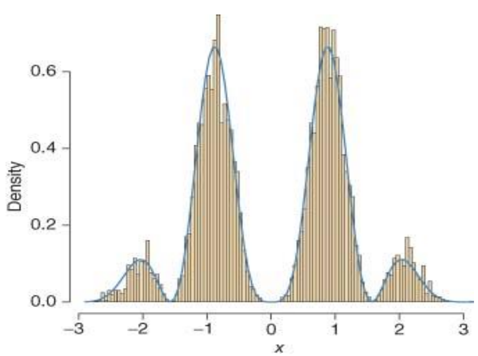
Γενίκευση του Αλγορίθμου Metropolis: Ο Αλγόριθμος Metropolis-Hastings (2/2)

Προβλήματα Αλγορίθμου

- Ο αλγόριθμος ***Metropolis-Hastings*** δημιουργεί αλυσίδα Markov με εργοδικές πιθανότητες συμβατές με ζητούμενη κατανομή αλλά παράγει **συσχετισμένες** τυχαίες μεταβλητές
- Αν απαιτείται η παραγωγή (***sampling***) τυχαίων διανυσμάτων $\mathbf{X}_n$ με πολλές διαστάσεις, ο αλγόριθμος υποφέρει από το ***curse of dimensionality***
- Η επιλογή κατάλληλης προτεινόμενης κατανομής μεταβάσεων $Q_{Y_n}(x|X_n = x_i)$ και αρχικής κατάστασης $X_0 = x_i$ έχουν ιδιαίτερη σημασίας για την ορθή και ταχεία σύγκλιση του αλγορίθμου https://onlinelibrary.wiley.com/doi/full/10.1002/9781118445112.stat07834
- Η επιρροή της μεταβατικής κατάστασης σβήνει χωρίς αρχικές μεταβάσεις π.χ. $n \leq 1000$

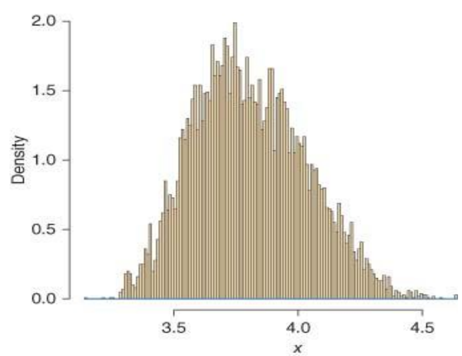
Παράδειγμα: Προσομοίωση κατανομής στόχου (***target***) $\pi(x)$ ανάλογης των $\widetilde{\pi}(x) = \sin^2(x) \times \sin^2(2x) \times \varphi(x)$ όπου $\varphi(x)$ κανονική κατανομή Gauss $\mathbb{N}(0,1)$

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



$a = 1, X_0 = 3.14$

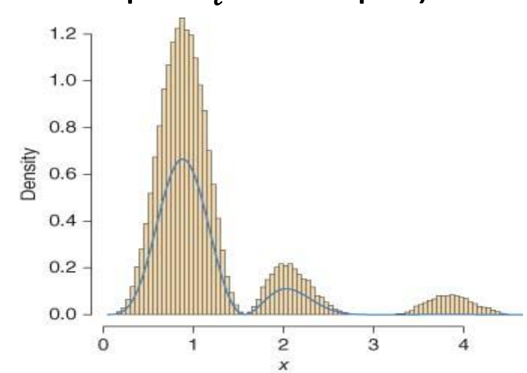
Ορθή σύγκλιση σε 10^4 βηματα



$a = 0.1, X_0 = 3.14$

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)

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

Παραγωγή Ακολουθίας 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

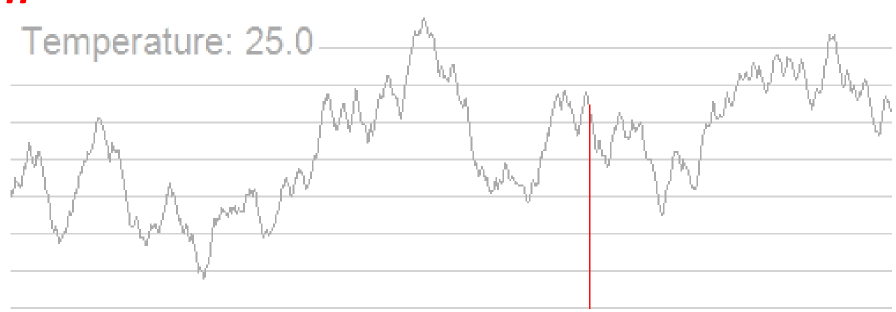
Υπολογισμοί Ολοκληρωμάτων και Στατιστικών Παραμέτρων

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 και με λογικό αριθμό διαστάσεων

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 Ανόπτησης*** (***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

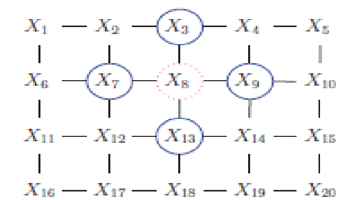
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

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

Μοντέλο *Ising* – Markov Random Fields, Προσομοίωση Μεταβάσεων Φάσεων

https://en.wikipedia.org/wiki/Ising_model

- 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 (***ferromagnetic***) ή αρνητική (π.χ. $J_{s,t} = -1$) αν τις ωθεί προς αντίθετη κατεύθυνση (***antiferromagnetic***)
- 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 $\boldsymbol{y}(G) = [y_1 \; y_2 \ldots y_s \; y_t \ldots]^{\mathrm{T}}$ αφορά σε νόμιμους σχηματισμούς (***configurations***) των επιμέρους καταστάσεων y_s των διπόλων που προκύπτουν από μεταβάσεις με πιθανότητες $P(y_s|\boldsymbol{y}(G)) = P(y_s|\boldsymbol{y}(N(s)))$



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 $\boldsymbol{y}(G)$ με πιθανότητες κατανομής ***Gibbs*** (***Boltzmann***) με ή χωρίς εξωτερική μαγνητική επιρροή h_t στα δίπολα t:

$$P\big(\boldsymbol{y}(G)\big) = \frac{1}{Z}\exp\left(-\frac{E(\boldsymbol{y}(G))}{T}\right) \text{ με } E\big(\boldsymbol{y}(G)\big) \approx -\sum\nolimits_{s\leftrightarrow t} J_{s,t} y_s y_t - \mu \sum\nolimits_t h_t y_t$$

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

Η σύγκλιση σε καταστάσεις ισορροπίας και η σταθερά Z μπορούν να προσεγγισθούν με αλγόριθμο ***Metropolis-Hastings*** (αν $J_{s,t} = J > 0$ η σύγκλιση οδηγεί στη ***ferromagnetic*** φάση)

Αντιστοίχιση Εννοιών Στατιστικής Φυσικής – Συνδυαστικής Βελτιστοποίησης

Οι διαδικασίες ***Simulated Annealing*** και τα ***Random Markov Fields*** – ***Ising Model*** έχουν εμπνεύσει αλγορίθμους Συνδυαστικής Βελτιστοποίησης κατ' αναλογία εννοιών Στατιστικής Φυσικής (π.χ. μεταβάσεις καταστάσεων σε θερμική ισορροπία ***Gibbs/Boltzmann*** όπως αυτές εξελίσσονται σε προσομοιώσεις Monte Carlo αλυσίδων Markov τύπου ***Metropolis-Hastings***)

Κοινά Χαρακτηριστικά

- Ζητείται ο εντοπισμός δύστροπων πολυδιάστατων καταστάσεων ισορροπίας ελαχίστου ***Κόστους*** ή ***Ενέργειας***, προς τις οποίες συγκλίνει ένα μοντέλο στοχαστικών μεταβάσεων. Η σύγκλιση μπορεί να υποβοηθείται από παραμέτρους ελέγχου (***Bias*** σε συστήματα Μηχανικής Μάθησης - Νευρωνικά Δίκτυα ή ***Θερμοκρασία*** σε Φυσικά Συστήματα)
- Οι αλγόριθμοι βελτιστοποίησης δεν προχωρούν μονοσήμαντα προς βήματα ελάττωσης του κόστους (ή αύξησης της ανταμοιβής - reward) αλλά δημιουργούν τυχαίες ακολουθίες Markov (***Random Walk Samples***) που επισκέπτονται μεγάλο εύρος πιθανών καταστάσεων ώστε να μην εγκλωβίζεται η αναζήτηση σε κάποιο ***τοπικό ελάχιστο κόστος*** (ή μέγιστη ανταμοιβή)

TABLE 11.1 Correspondence between Statistical Physics and Combinatorial Optimization

Statistical physics	Combinatorial optimization
Sample	Problem instance
State (configuration)	Configuration
Energy	Cost function
Temperature	Control parameter
Ground-state energy	Minimal cost
Ground-state configuration	Optimal configuration

Προσομοιωμένη Ανόπτηση - Simulated Annealing (1/2)

Αρχική Θερμοδυναμική Προσέγγιση

Μελέτη φυσικού συστήματος πολλών σωματιδίων: Σύγκλιση αλυσίδας Markov σε κατάσταση ισορροπίας χαμηλής θερμοκρασίας T

- Το σύστημα συγκλίνει στις πιθανότητες ***Gibbs*** $p_i = \frac{1}{Z}\exp\left(-\frac{E_i}{T}\right)$ όταν $T \to 0$ με τις καταστάσεις χαμηλής ενέργειας να έχουν τις μεγαλύτερες πιθανότητες

Ο αλγόριθμος ***Metropolis*** εντοπίζει καταστάσεις ισορροπίας με γρήγορη σύγκλιση για υψηλές θερμοκρασίες T αλλά με εξαιρετική βραδύτητα σε χαμηλές T εξού και η τροποποίηση του με την τεχνική ***Simulated Annealing*** που περιλαμβάνει δύο συστατικά:

- Πρόγραμμα μείωσης της θερμοκρασίας (***cooling schedule***) προς την ζητούμενη χαμηλή τιμή της $T \to 0$ μέσω επαναλήψεων $T_{0,} T_1, \dots, T_k, \dots, T$ με συνεχώς μειούμενα άλματα (π.χ. $T_k = \alpha T_{k-1}, 0.8 < \alpha < 0.99$)
- Προσδιορισμός καταστάσεων ισορροπίας με τον αλγόριθμος ***Metropolis*** αρχικά για υψηλή T_0 που να επιτρέπει μεγάλο εύρος μεταβάσεων και λογικό αριθμό διαδοχικών επαναλήψεων (π.χ. 10 μεταβάσεις καταστάσεως) σε χαμηλότερες $T_k \to T_{k+1}$ με αρχικοποίηση στην κατάσταση που οδηγήθηκε το σύστημα στην προηγούμενη θερμοκρασία T_k

Ο αλγόριθμος ***Simulated Annealing*** μπορεί να εγκλωβισθεί σε **πολλαπλά τοπικά ελάχιστα**. Η σύγκλιση στο σφαιρικό ελάχιστο (***global minimum***) δεν είναι εγγυημένη για ρεαλιστικά προγράμματα μείωσης της θερμοκρασίας

Προσομοιωμένη Ανόπτηση - Simulated Annealing (2/2)

Εφαρμογή σε Συνδυαστική Βελτιστοποίηση (Combinatorial Optimization)

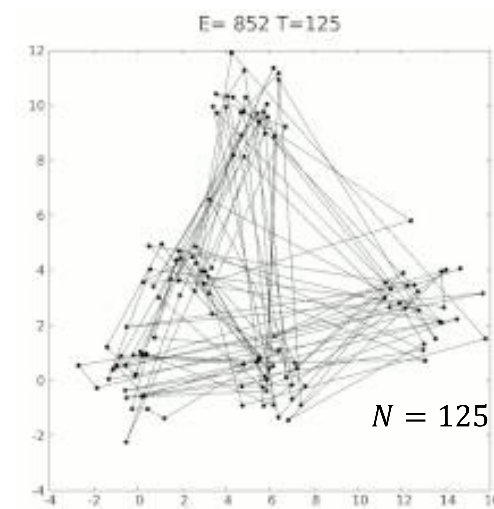
Η ανίχνευση θερμοδυναμικής ισορροπίας σε χαμηλές θερμοκρασίες μεταφράζεται σε ανεύρεση κατάστασης ελάχιστου κόστους σε προβλήματα **Συνδυαστικής Βελτιστοποίησης** περίπλοκων προβλημάτων με πολλαπλά τοπικά ελάχιστα και μεγάλο αριθμό καταστάσεων

- Οι ενέργειες των καταστάσεων E_i αντιστοιχούν με το αριθμητικό κόστος των διακριτών καταστάσεων. Η θερμοκρασίας T είναι παράμετρος εξωτερικού ελέγχου μεταβάσεων του συστήματος και ξεκινά με υψηλή αρχική τιμή και μειώνεται $T \rightarrow \alpha T$ με βήμα (υπερπαράμετρο - ***hyperparameter***) $\alpha,\ 0 < \alpha < 1$ μέχρι $T \cong 0$
- Σε κάθε διαδοχική μείωση της T ο αλγόριθμος ***Metropolis*** εκκινεί από την προηγούμενη τελική κατάσταση προς νέα ισορροπία με πιθανότητες ***Gibbs*** $p_i = \frac{1}{Z}\exp\left(-\frac{E_i}{T}\right)$ για E_i και T. Η αποδοχή μεταβάσεων $i \rightarrow j$ προς γειτονικές καταστάσεις υψηλότερου κόστους $E_j > E_i$ με μη μηδενική πιθανότητα $\exp(-\frac{E_j - E_i}{T})$, αποτρέπει εγκλωβισμό σε **τοπικά ελάχιστα**. Η πιθανότητα μειώνεται για $T \rightarrow 0$ και ο αλγόριθμος ***θεωρεί*** πως βρήκε το **σφαιρικό ελάχιστο**

Traveling Salesman Problem: Διάσημο *NP-Complete* Πρόβλημα Συνδυαστικής Βελτιστοποίησης

Δίνεται γράφος με N σημεία (πόλεις) κόστους c_{ij} μεταξύ σημείων (κόμβων) i και j. Ζητείται διαδρομή ελαχίστου κόστους περιπλανώμενου πωλητή (***Traveling Salesman***) που να επισκέπτεται όλα τα σημεία μία φορά

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



Δειγματοληψία Gibbs (1/2)

- Παράγει μια ακολουθία τυχαίων στοιχείων δείγματος **πολλών διαστάσεων** μέσω προσομοίωσης ***Monte Carlo*** αλυσίδας ***Markov*** **διανυσματικών** καταστάσεων πολλών μεταβλητών με εργοδικές πιθανότητες κατανομής ***Gibbs***. Τα παραγόμενα τυχαία στοιχεία δεν είναι ανεξάρτητα (συσχετίσεις λόγω μεταβάσεων ***Markov***)
- Αποτελεί παραλλαγή του αλγόριθμου ***Metropolis*** με επαναλήψεις (βήματα) που οδηγούν σε καταστάσεις χαμηλότερης ενέργειας, αλλά επιτρέπονται και αντίθετες μεταβάσεις με κάποια φθίνουσα πιθανότητα
- Κάθε επανάληψη περιλαμβάνει διαδοχικές μεταβάσεις για κάθε συνιστώσα του διανυσματικού δείγματος. Οι πιθανότητες μετάβασης ανά συνιστώσα εξαρτώνται από τις παρούσες τιμές όλων των συνιστωσών **πλην της παρούσας τιμής της ίδιας συνιστώσας** και δεν είναι χρονοσταθερές. Οι παρούσες τιμές περιλαμβάνουν νέες τιμές άλλων συνιστωσών όπως ανανεώθηκαν στη παρούσα επανάληψη πριν από την υπό μετάβαση συνιστώσα
- Συγκλίνει προς καταστάσεις ισορροπίας ***Gibbs***, επιτρέποντας εκτιμήσεις συνδυασμένων (***joint***) ή οριακών (***marginal***) εργοδικών πιθανοτήτων καταστάσεων και κατανομών σαν σχετικό πλήθος επισκέψεων σε καταστάσεις στη διάρκεια της προσομοίωσης
- **Εφαρμογές**
  - Δημιουργία με προσομοίωση ***Monte Carlo*** διανυσματικού δείγματος κατανομής ***Gibbs***
  - Συμπλήρωση διανυσματικών δεδομένων με μη παρατηρήσιμες ή παραμορφωμένες συνιστώσες (***hidden variables*** σε Νευρωνικά Δίκτυα με κρυφούς νευρώνες, ***Deep Neural Networks***), υποθέτοντας δείγμα κατανομής ***Gibbs***
  - Προσέγγιση συνάρτησης εξαρτώμενης από μια μόνο μεταβλητή, π.χ. μέση τιμή μιας συνιστώσας διανυσματικού δείγματος κατανομής ***Gibbs***

Δειγματοληψία Gibbs (2/2)

- Ορισμός τυχαίου διανύσματος με K συνιστώσες στο βήμα (επανάληψη) n:
$$\mathbf{X}(n) = [X_1(n)\, X_2(n)\ \dots\ X_K(n)]^{\mathrm{T}}$$
- Οι διανυσματικές τιμές των τυχαίων μεταβλητών $\mathbf{X}(n)$ ορίζουν καταστάσεις :
$$\mathbf{x}(n) = [x_1(n)\, x_2(n)\ \dots\ x_K(n)]^{\mathrm{T}}$$

Αλγόριθμος (*Stuart* & *Donald Geman*, 1984)

$n = 0$: Αυθαίρετη αρχικοποίηση $\mathbf{x}(0) = [x_1(0)\, x_2(0)\ \dots\ x_K(0)]^{\mathrm{T}}$

$n \to n+1$: Κατά σειρά για $i = 1, \dots, K$ δημιουργείται νέα τιμή $x_i(n+1)$ της συνιστώσας i της $X_i(n+1)$ με ψευτο-τυχαίο τρόπο βάση της υπό συνθήκη πιθανότητας
$$P[X_i(n+1)|\{x_1(n+1)\ \dots\ x_{i-1}(n+1)\ x_{i+1}(n)\ \dots\ x_K(n)\}]$$
Η συνθήκη περιλαμβάνει τις τιμές των άλλων συνιστωσών όπως έχουν ορισθεί μέχρι το στάδιο αυτό και **δεν** περιλαμβάνει την τιμή της συνιστώσας $X_i(n)$

Η ακολουθία των δειγμάτων $x_i(n)$, $i = 1, \dots, K$ σηματοδοτεί αλυσίδα ***Markov*** που συγκλίνει (γεωμετρικά) προς καταστάσεις θερμικής ισορροπίας $\mathbf{x} = [x_1\, x_2\ \dots\ x_K]$ με οριακές (**marginal**) πιθανότητες
$$\lim_{n\to\infty} P(X_i(n) = x_i(n)|x_i(0)) = P(X_i = x_i)$$
και με συνδυασμένες (**joint**) πιθανότητες ***Gibbs***
$$\lim_{n\to\infty} P(\mathbf{X}_n = \mathbf{x}_n) = P(X_1 = x_1, \dots, X_K = x_K) = P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right)$$

Εφαρμογή: Παραγωγικά Μοντέλα Μάθησης - ***Boltzmann Machine*** (**BM**)***, Restrictive*** **BM**