

*How and why
the Monte Carlo method
works*

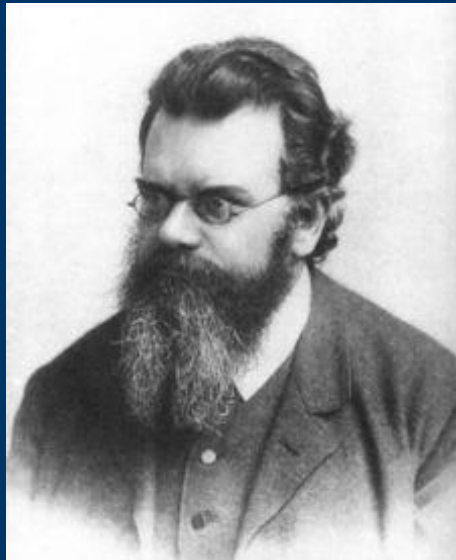


Julien Michel
March 2007

Outline: Monte Carlo

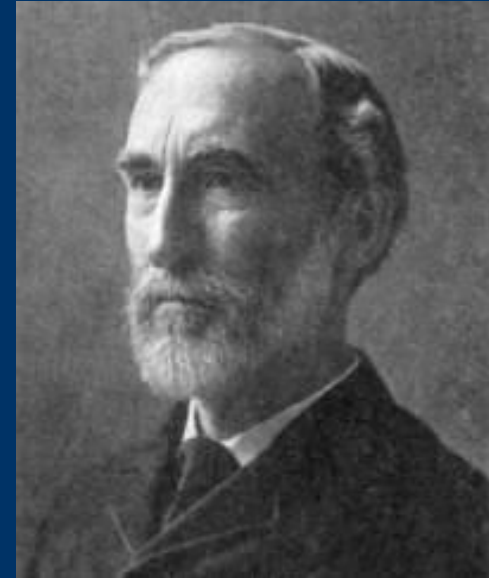
- Origin of statistical mechanics
 - The partition function.
 - Calculation of thermodynamic properties as an ensemble average
 - Numerical techniques to estimate integrals
 - Direct quadrature, basic Monte Carlo integration, importance sampling
 - Markov Chains, to draw from Boltzmann without knowing it
 - The Metropolis algorithm
 - What is meant by detailed balance. Example of moves that satisfy/break it and the consequences.
 - Beyond detailed balance. Markov chains that still work and the Perron Frobenius theorem.
 - References for further reading
-
-

Statistical mechanics



Ludwig Eduard Boltzmann

1844-1906



Josiah Willard Gibbs

1839-1903

Partition function and ensemble averages

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int \int dp^N dr^N \exp(-\beta E(p^N, r^N))$$

$$A_{obs} = \langle A_{ens} \rangle = \frac{1}{N! h^{3N}} \frac{1}{Q_{NVT}} \int \int dp^N dr^N A(p^N, r^N) \exp(-\beta E(p^N, r^N))$$

$$\langle A_{ens} \rangle = \frac{\int dr^N A(r^N) \exp(-\beta U(r^N))}{Z_{N,NVT}}$$

$$Z_{N,NVT} = \int dr^N \exp(-\beta U(r^N))$$

Some thermodynamic properties

$$U = - \left(\frac{\partial \ln Z}{\partial \beta} \right)_{N,V} = \langle E \rangle$$

$$P = \frac{1}{\beta} \left(\frac{\partial \ln Z}{\partial V} \right)_{N,T} = \frac{NkT}{V} + \frac{1}{3V} \left\langle \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} f_{ij} \right\rangle$$

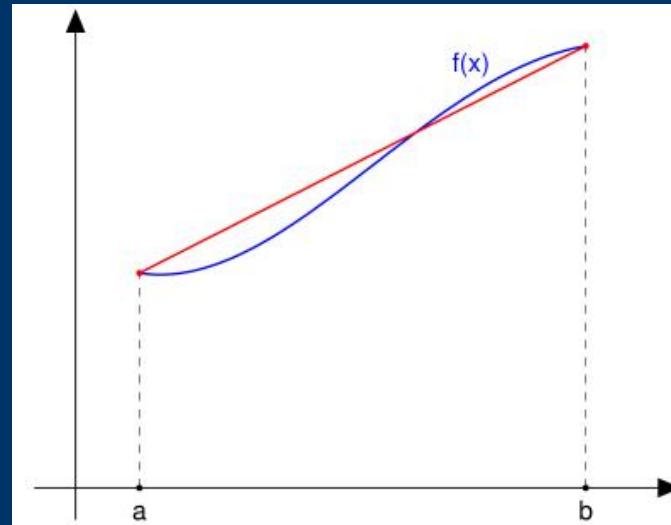
$$C_v = \left(\frac{\partial U}{\partial T} \right)_V = \frac{(\langle E^2 \rangle - \langle E \rangle^2)}{kT^2}$$

So how do I integrate this?

$$\langle A_{ens} \rangle = \frac{\int dr^N A(r^N) \exp(-\beta U(r^N))}{Z_{N,\mu,T}}$$

Given the very complicated nature of U in molecular simulation, it is very unlikely analytical solutions will ever be devised

Numerical integration: quadrature techniques



$$\int_a^b f(x) dx \approx \frac{b-a}{2n} (f(x_0) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(x_n))$$

10 pts / dimension in 10 dimensions = 10 billion function evaluations

$$\langle A_{ens} \rangle = \frac{\int dr^N A(r^N) \exp(-\beta U(r^N))}{Z_{N,NVT}}$$

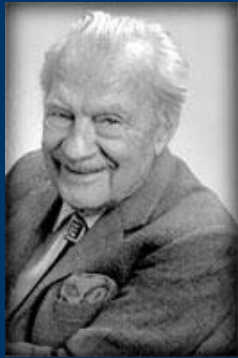
Too many
dimensions !

MANIAC, the birth of computing and Monte Carlo



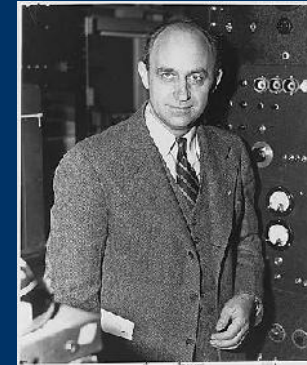
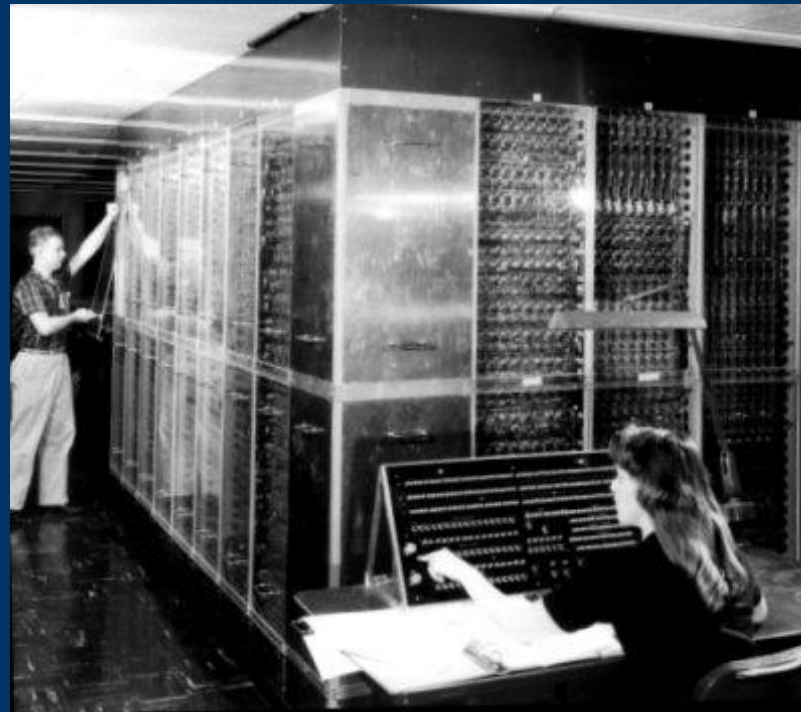
John von Neumann

1903-1957



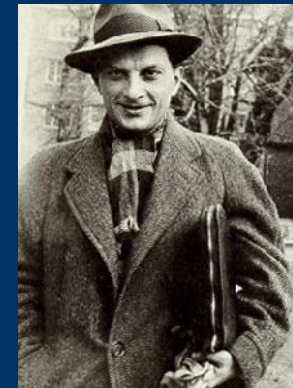
**Nicholas Constantine
Metropolis**

1915-1999



Enrico Fermi

1901-1954



Stanislaw Ulam

1909-1984

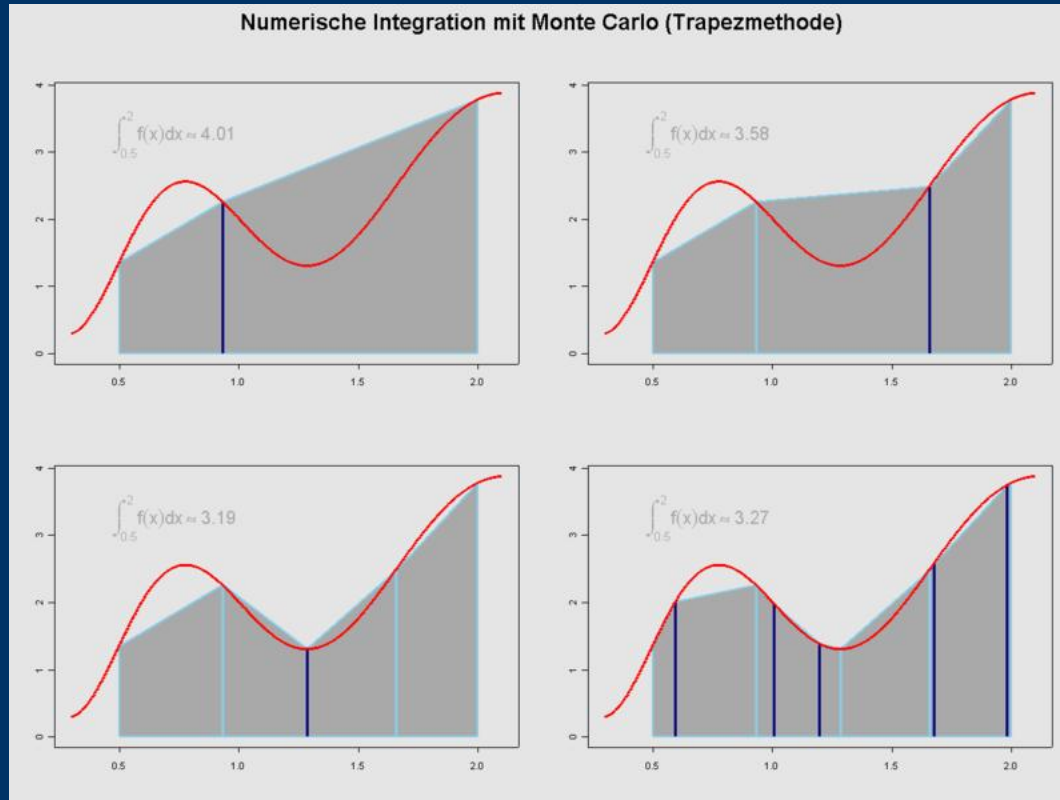
Scientific triumphs of MANIAC

“The nice thing about having the first computing machine is that almost anything you do on it is new and important”

Herbert L. Anderson

- Pion-Proton phase-shift analysis (Fermi, Metropolis; 1952)
 - Phase-shift analysis (Bethe, deHoffman, Metropolis; 1954)
 - Non linear coupled oscillators (Fermi, Pasta, Ulam; 1953)
 - Genetic code (Gamow, Metropolis; 1954)
 - **Equation of state: Importance Sampling (Metropolis, Teller, 1953)**
 - Two-dimensional hydrodynamics (Metropolis, von Neumann; 1954)
 - Universalities of iterative functions (Metropolis, Stein, Stein; 1973)
 - Nuclear cascades using Monte Carlo (Metropolis, Turkevich; 1954)
 - Anti-clerical chess (Wells; 1956)
 - The lucky numbers (Metropolis, Ulam; 1956)
-
-

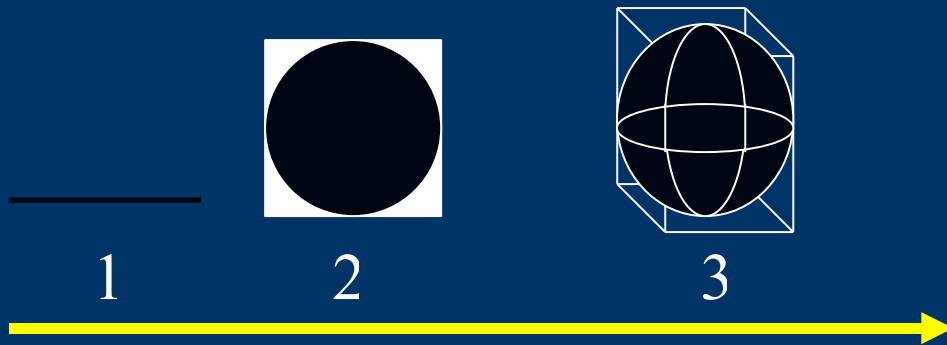
The Monte Carlo Method



$$I_{est} = \frac{1}{N} \sum_{i=1}^N I_S(X_i)$$

Monte Carlo rapidly outperforms quadrature techniques in higher dimensions

What is the volume of a "sphere" ?



$$\frac{V}{V_R} = \frac{\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2} + 1\right) 2^k}$$

Number of
dimensions k

k	V/V_R
1	1.00E+00
2	7.85E-01
3	5.24E-01
5	1.64E-01
10	2.49E-03
50	1.54E-27
100	1.87E-69

As k increases, the vast majority of the points in the k -dimension space lies outside of the sphere

Random selection of points not efficient

Strong analogy with statistical mechanical integrals. There are **few low energy states** that contributes meaningfully to the integral and **many high energy states** (e.g atomic overlaps) that do not contribute.

Importance sampling

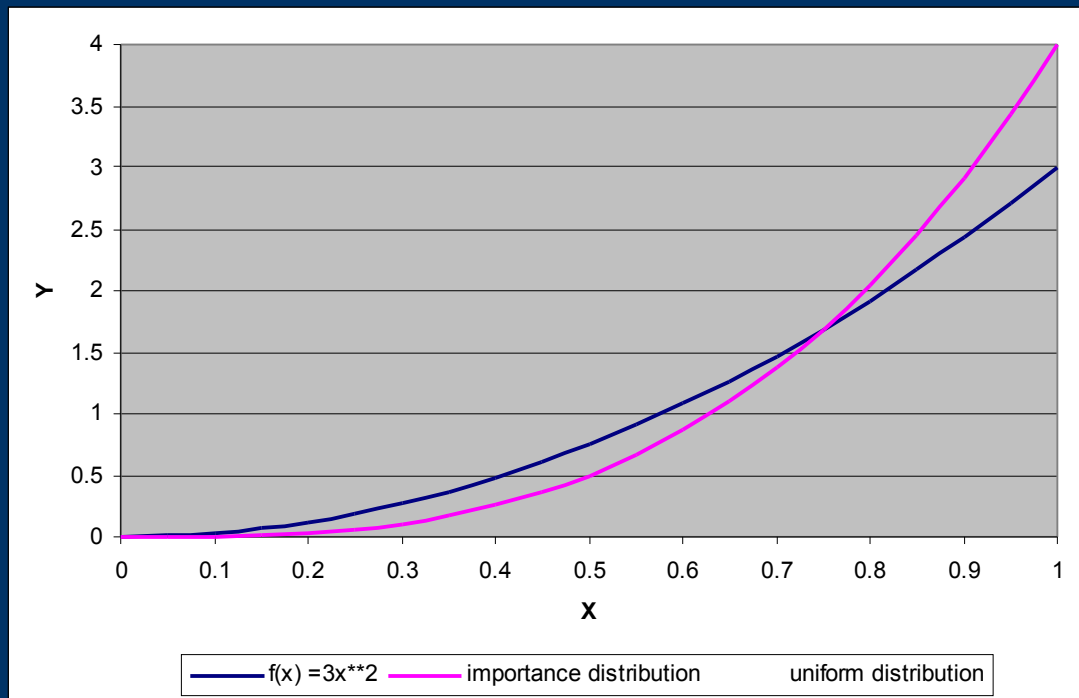
- Instead of drawing random points from an uniform distribution, draw points from a distribution π . The Monte Carlo integration equation becomes

$$I_{est} = \frac{1}{N} \sum_{i=1}^N \frac{I(X_i)}{\pi(X_i)}$$

- π is selected so that points are in the region of space which contributes the most to the integrand (e.g, in the sphere)
 - The bias on the selection of X_i is removed when the contribution to the integrand is estimated
-
-

Importance sampling: example

$$f(x) = 3x^2$$
$$I = \int_0^1 \frac{f(x)}{\pi_k(x)} \pi_k(x) dx$$
$$\pi_0 = 1$$
$$\pi_1 = 4x^3$$



Estimate of I after drawing 100 samples with two different importance sampler

Function	Average	Std. Dev
π_0	1.027	0.111
π_1	0.999	0.036

Importance sampling: implications

$$\langle A_{ens} \rangle = \frac{\int dr^N A(r^N) \exp(-\beta U(r^N))}{Z_{N,NVT}}$$

If $A(r^N)$ does not dominate the product in the numerator, then an ideal importance sampling function to estimate our integral is :

$$\pi(r^i) = \frac{\exp(-\beta U(r^i))}{\int \exp(-\beta U(r^N)) dr^N}$$

Problem : Impossible to draw samples from $\pi(r_i)$ without knowing the denominator, which we can't (as it involves solving directly a very difficult integral)

Markov Chains

- A Markov Chain is a set of **probabilistic rules** which governs **transitions between states** and is often represented as a **transition matrix Π**



Andrei Markov
1856-1922

$$\Pi = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

← Probability of moving from state 1 to 3

- Assuming Π obeys a number of mathematical properties, then some interesting properties arise...

Properties of Markov Chains

$$\rho^{(2)} = \rho^{(1)}\Pi$$

$$\rho^{(3)} = \rho^{(2)}\Pi = \rho^{(1)}\Pi^{(2)}$$

$$\pi = \lim_{n \rightarrow \infty} \rho^{(1)}\Pi^{(n)}$$

$$\pi = \lim_{n \rightarrow \infty} \left[\rho^{(1)}\Pi^{(n-1)} \right] \Pi$$

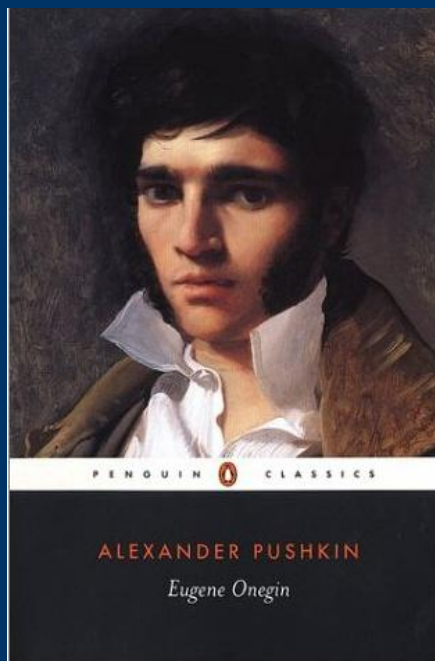
$$\pi = \pi\Pi$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(X_i) = \int_{R_k} f(x)\pi(x)dx$$

Markov Chains : Take home message

- Repeated application of a transition matrix Π converges any arbitrary initial distribution to a unique limiting distribution π
 - Further applications let us draw samples from this distribution π
 - This is achieved without ever having to know the whole distribution π !
-
-

The first application of a Markov Chain



Не мысля гордый свет забавить,
Вниманье дружбы возлюбя,
Хотел бы я тебе представить
Залог достойнее тебя,
Достойнее души прекрасной,
Святой исполненной мечты,
Поэзии живой и ясной,
Высоких дум и простоты;

$$(0.128, 0.872) = (1.0, 0.0) \begin{pmatrix} 0.128 & 0.872 \\ 0.663 & 0.337 \end{pmatrix}$$

$$(0.594, 0.405) = (0.128, 0.872) \begin{pmatrix} 0.128 & 0.872 \\ 0.663 & 0.337 \end{pmatrix}$$

(...)

$$(0.432, 0.568) = (0.432, 0.568) \begin{pmatrix} 0.128 & 0.872 \\ 0.663 & 0.337 \end{pmatrix}$$

20,000 letters !

The catch....

Given a transition matrix Π we can draw samples from its limiting distribution π

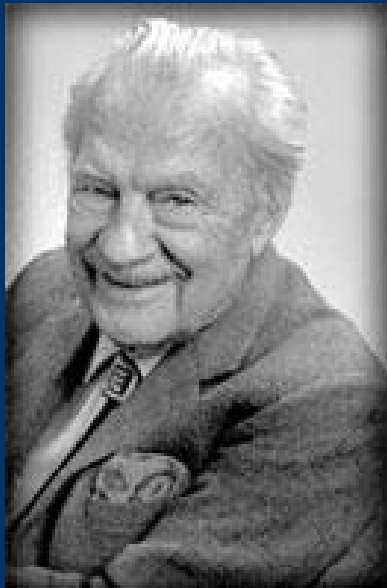
$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(X_i) = \int_{R_k} f(x) \pi(x) dx$$

$$\langle A_{ens} \rangle = \frac{\int dr^N A(r^N) \exp(-\beta U(r^N))}{Z_{N,NVT}} \quad \pi(r^i) = \frac{\exp(-\beta U(r^i))}{\int \exp(-\beta U(r^N)) dr^N}$$

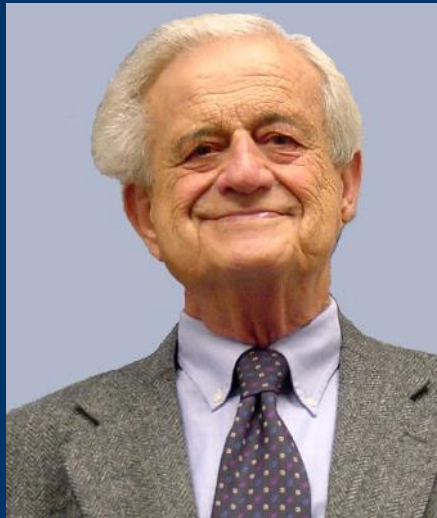
Problem 1: How do we build a transition matrix Π whose limiting distribution is the Boltzmann distribution ?

Problem 2: How do we manipulate Π if there are a large number of states ?

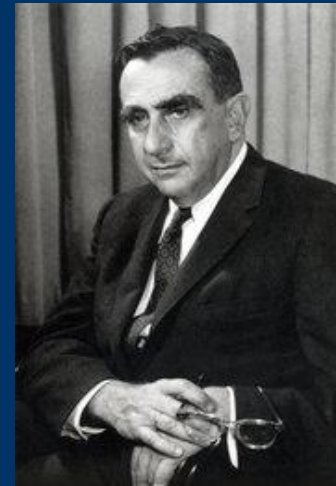
Markov Chain Monte Carlo: Metropolis Sampling



**Nicholas
Constantine
Metropolis
1915-1999**



**Marhsall
Rosenbluth
1927-2003
and Arianna
Rosenbluth**



**Edward Teller
1908-2003
and Augusta
Teller**



The Metropolis algorithm

1. Start in state I
2. Attempt a move to state j with probability a_{ij}
3. Accept this move with probability

$$\alpha_{ij} = \frac{\pi_j}{\pi_i} = \frac{\exp(-\beta U_j) / Z_N}{\exp(-\beta U_i) / Z_N} = \exp(-\beta(U_j - U_i))$$

6. If the move is accepted, set $i = j$, otherwise $i = i$
7. Accumulate any property of interest $A(i)$
8. Return to 1 or terminate after N iterations

Detailed Balance

$$(\pi_1, \pi_2) = (\pi_1, \pi_2) \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

$$(\pi_1, \pi_2) = (\pi_1 p_{11} + \pi_2 p_{21}, \pi_1 p_{12} + \pi_2 p_{22})$$

$$\pi_1 p_{12} = \pi_2 p_{21}$$

$$(\pi_1, \pi_2) = (\pi_1 (p_{11} + p_{12}), \pi_2 (p_{21} + p_{22}))$$

$$\pi_i p_{ij} = \pi_j p_{ji}$$

Detailed Balance in Metropolis

Let p_{ij} be the probability that the move i to j is accepted and note

$$p_{ij} = a_{ij}\alpha_{ij}$$

where

$$\alpha_{ij} = \min\left(1, \frac{\pi_j}{\pi_i}\right)$$

if π is the Boltzmann distribution

$$\frac{\pi_j}{\pi_i} = \frac{\frac{\exp(-\beta U_j)}{Z_{N,NVT}}}{\frac{\exp(-\beta U_i)}{Z_{N,NVT}}} = \frac{\exp(-\beta U_j)}{\exp(-\beta U_i)}$$

assume $\pi_j < \pi_i$. Then

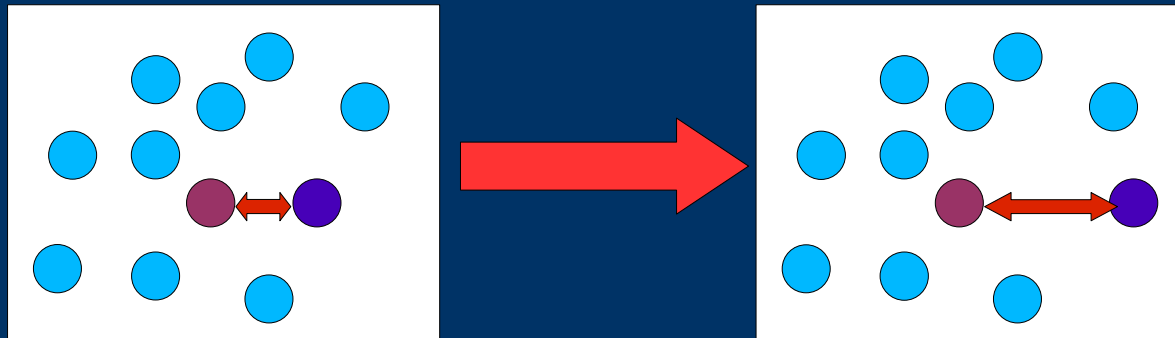
$$\pi_i p_{ij} = \pi_j p_{ji}$$

$$\pi_i a_{ij} \alpha_{ij} = \pi_j a_{ji} \alpha_{ji}$$

$$\pi_i a_{ij} \frac{\pi_j}{\pi_i} = \pi_j a_{ji}$$

$$a_{ij} = a_{ji}$$

A simple Monte Carlo move



$$\pi_j < \pi_i$$

$$\begin{aligned}\pi_i p_{ij} &= \pi_j p_{ji} \\ \pi_i \frac{1}{N} \frac{dV}{V_m} \alpha_{ij} &= \pi_j \frac{1}{N} \frac{dV}{V_m} \alpha_{ji} \\ \pi_i \frac{1}{N} \frac{dV}{V_m} \frac{\pi_j}{\pi_i} &= \pi_j \frac{1}{N} \frac{dV}{V_m} 1\end{aligned}$$

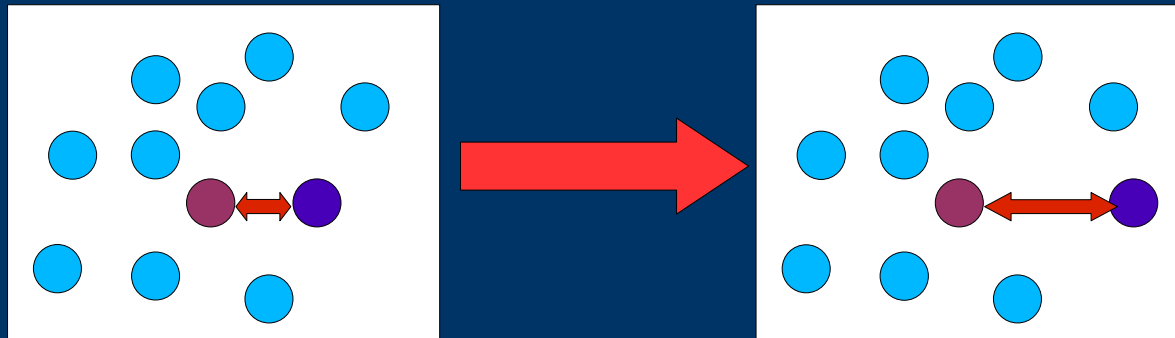
Biased Monte Carlo Moves

$$\pi_i p_{ij} = \pi_j p_{ji}$$

$$\pi_i a_{ij} \alpha_{ij} = \pi_j a_{ji} \alpha_{ji}$$

$$\pi_i a_{ij} \frac{\pi_j a_{ji}}{\pi_i a_{ij}} = \pi_j a_{ji}$$

A correct biased move: Preferential Sampling



$$\pi_j < \pi_i$$

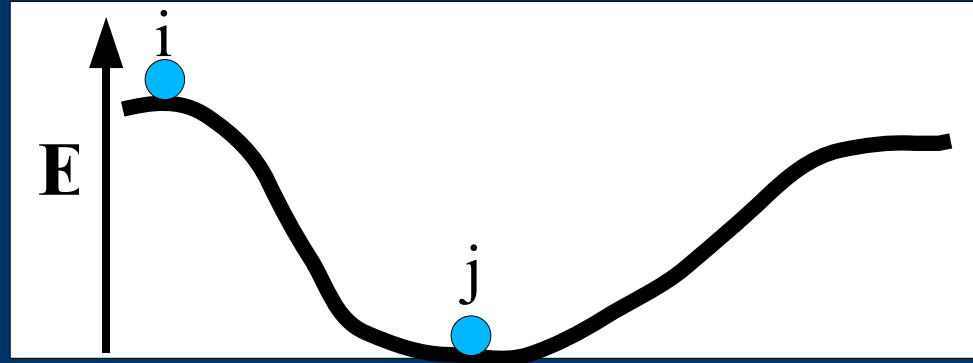
$$W_i = \frac{1}{r_{is}} \quad W'_i = \frac{W_i}{\sum_{j=1}^N W_j}$$

$$\pi_i p_{ij} = \pi_j p_{ji}$$

$$\pi_i W'_i \frac{dV}{V_m} \alpha_{ij} = \pi_j W'_j \frac{dV}{V_m} \alpha_{ji}$$

$$\pi_i W'_i \frac{dV}{V_m} \frac{\pi_j W'_j}{\pi_i W'_i} = \pi_j W'_j \frac{dV}{V_m} 1$$

An impossible biased Move: Minimisation



$$\pi_j > \pi_i$$

$$\pi_i p_{ij} = \pi_j p_{ji}$$
$$\pi_i 1 \alpha_{ij} = \pi_j 0 \alpha_{ji}$$

This can only be true if α_{ij} is 0. e.g, if the forward move is certain and the reverse move impossible, the forward move cannot be accepted.(*)

(*). However, $p_{ij} = 0$ for $j \neq i$ would not satisfy the conditions required to ensure that the Markov Chain is ergodic

Beyond Metropolis sampling

$$\lambda_d \mu_d = \mu_d \Pi$$

An approximate phrasing of the Perron-Frobenius theorem: Π has one dominant eigenvalue $\lambda = 1$ and the associated eigenvector μ is the only eigenvector whose individual values are all positive or null. The magnitude of the other eigenvalues are lower than unity.”

$$\Pi = PDP^{-1}$$

$$\begin{aligned}\Pi^{(n)} &= (PDP^{-1}) \dots (PDP^{-1}) \\ &= PD^{(n)}P^{-1}\end{aligned}$$

$$D^{(1)} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \lambda_n \end{pmatrix}$$

$$D_{n \rightarrow \infty}^{(n)} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 \end{pmatrix}$$

Beyond Metropolis sampling

$\pi (0.33, 0.33, 0.33)$

Transition matrices

Eigenvalues

**After 100
applications(*)**

$$\begin{pmatrix} 0.96 & 0.02 & 0.02 \\ 0.01 & 0.97 & 0.02 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}$$

1 ; 0.95 ; 0.94

$\pi_{\text{est}}(0.62, 0.20, 0.18)$

$$\begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{pmatrix}$$

1 ; -0.50 ; -0.50

$\pi_{\text{est}}(0.25, 0.38, 0.37)$

$$\begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.2 & 0.7 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}$$

1 ; -0.25 ; -0.25

$\pi_{\text{est}}(0.37, 0.31, 0.32)$

(*) starting from $\pi_{\text{est}} = (1, 0, 0)$

References

- Appendix of my thesis
 - Pr. Kofke lecture's notes <http://www.eng.buffalo.edu/~kofke/ce530>
 - Evans, A.; Swartz. T “Approximating Integrals via Monte Carlo and deterministic methods” Oxford University Press, 2000, Oxford, UK
 - Gammelman, D. “Markov Chain Monte Carlo”, Chapman and Hall, 1997, London UK
 - Metropolis, N. and Ulam, S. “The Monte Carlo Method” Journal of the American Statistical Association 1949, 44, 335-341
 - Metropolis; N. Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. “Equation of State Calculations by Fast Computing Machines” Journal of Chemical Physics 1953, 21, 1087-1092
-
-