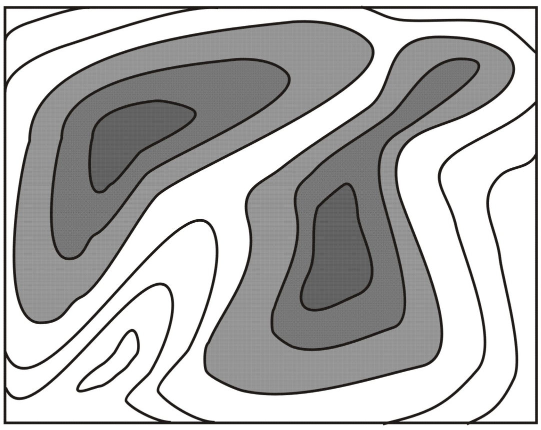
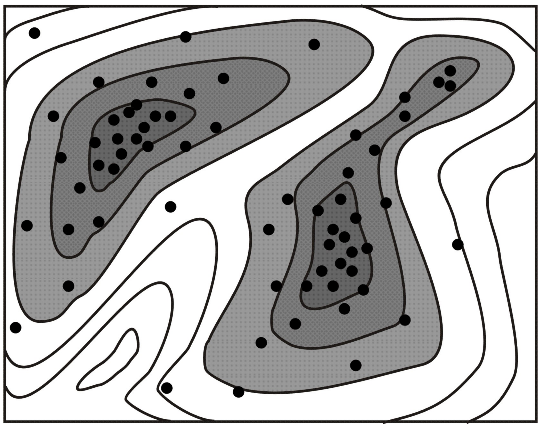
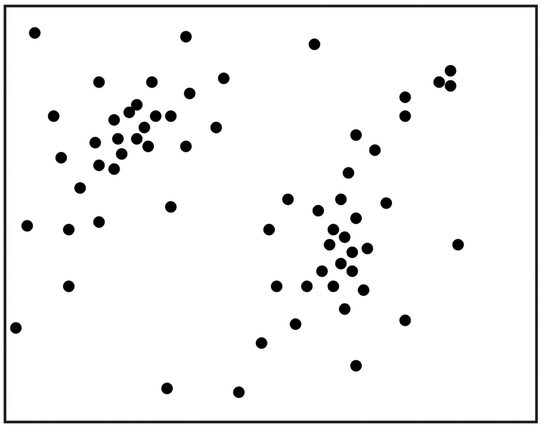
Basic Monte Carlo Theory

Klaus Mosegaard
Niels Bohr Institute
University of Copenhagen

Sampling of solutions to inverse problems



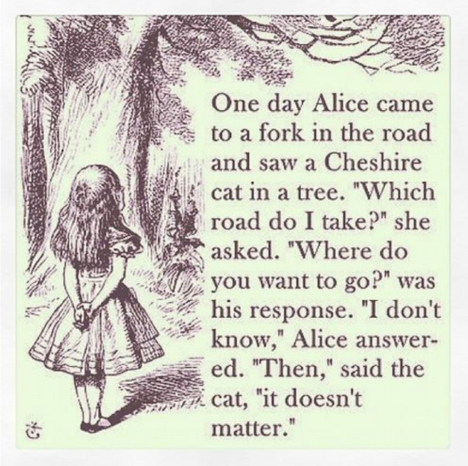




Why Random Sampling?

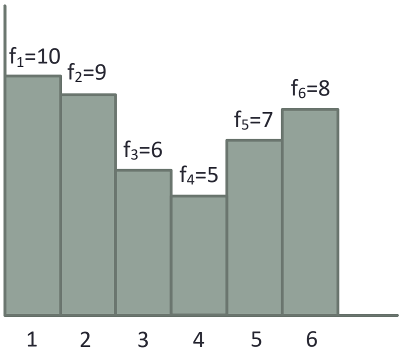
“If you don't know where you want to go, then it doesn't matter which path you take.”

— Lewis Carroll, Alice in Wonderland



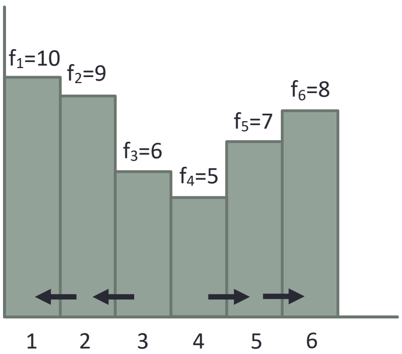
The Old Controversy: Random or Deterministic?

A Simple 2-peak Model

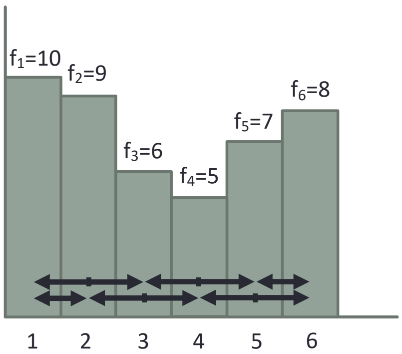


A Simple 2-peak Model

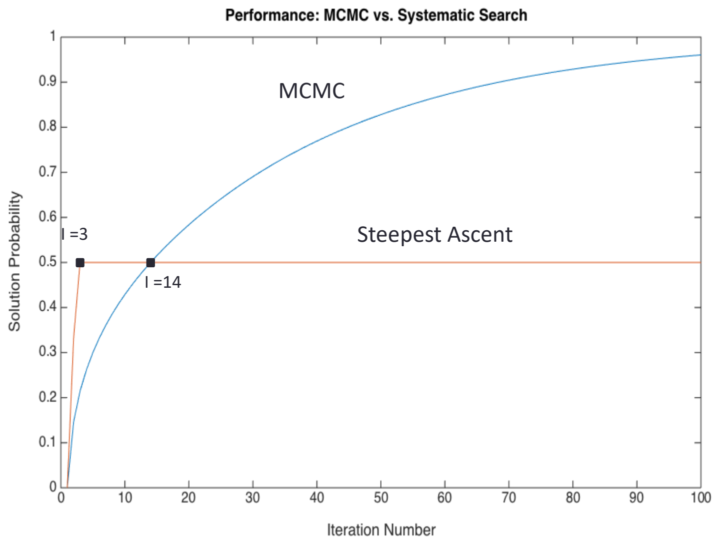
Steepest Ascent:
Always uphill



MCMC:
Random steps left/right



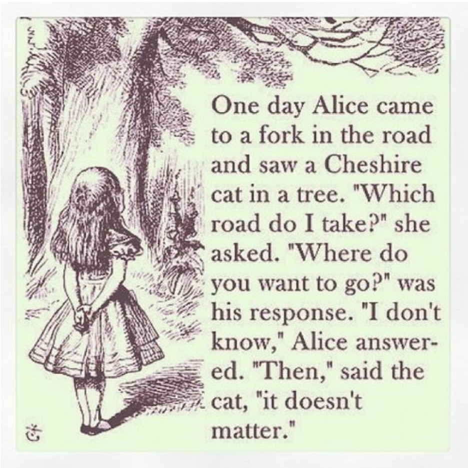
Evolution of Solution Probability



Why Random Sampling?

"If you don't know where you want to go, then it doesn't matter which path you take."

— Lewis Carroll, Alice in Wonderland



Monte Carlo Sampling: Estimation of averages

Estimation of integrals of the form

$$I = \int_X h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

is an important application of Monte Carlo methods:

- Mean values: $\langle \mathbf{x} \rangle = \int_X \mathbf{x}\, f(\mathbf{x}) d\mathbf{x}$
- Covariances: $\mathbf{C}_{ij}(\mathbf{x}) = \int_X (x_i - \langle x \rangle_i)(x_j - \langle x \rangle_j)\, f(\mathbf{x}) d\mathbf{x}$
- Probability of events A: $P(A) = \int_A f(\mathbf{x}) d\mathbf{x}$

Approximation of averages

Given N **statistically independent** samples $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N$ from the pdf $f(\mathbf{x})$, we have the following approximation:

$$I = \int_X h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{x}_n)$$

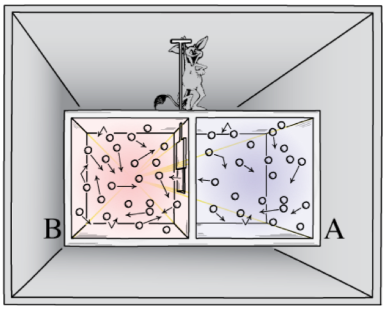
Efficiency

- All other methods using N points $\mathbf{x}_1, \mathbf{x}_2, \ldots\ \mathbf{x}_N$ in an M-dimensional space to approximate I have an absolute error that decreases not faster than $N^{-1/M}$.
- The absolute error of a (perfect) Monte Carlo method decreases as $N^{-1/2}$, that is, **independent of** M (!)
- But the number of iterations between roughly independent samples increases strongly with M (!)

Basic Monte Carlo Methodology

But First: Introduction to ”Demonology”

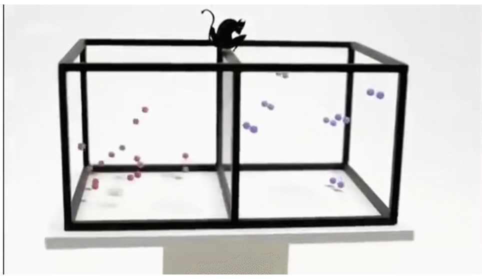
Maxwell's Demon



The Gate-keeper Demon apparently violates the second law of thermodynamics!
The demon only allows ”slow” atoms to enter the compartment to the right.
In this way it increases the temperature difference between the two compartments!

But First: Introduction to ”Demonology”

Maxwell's Demon



The Gate-keeper Demon apparently violates the second law of thermodynamics!
The demon only allows ”slow” atoms to enter the compartment to the right.
In this way it increases the temperature difference between the two compartments!

The Two Demons of Monte Carlo sampling

Proposal Demon
Randomly suggests reasonable points. This demon may ”know” something about the target distribution (perhaps not much!)

Gate-keeper Demon
Accepts or rejects the suggestions from the proposal demon. Knows nothing about the target distribution!

Rejection Sampling (simple)

In the n'th step of the algorithm:

1. Propose a candidate sampling point $\mathbf{x}_{cand}$ with a **constant** proposal distribution $q(\mathbf{x})$
2. Accept $\mathbf{x}_{cand}$ only with probability

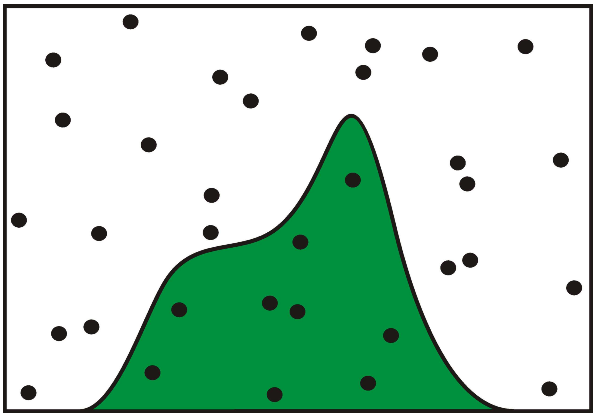
$$p_{accept} = \frac{p(\mathbf{x}_{cand})}{M}$$

where

$$M \geq \max_{\mathbf{x}}\big(p(\mathbf{x})\big).$$

The set of thus accepted candidate points are samples from the probability distribution $p(\mathbf{x})$.

Rejection Sampling: Why does it work?

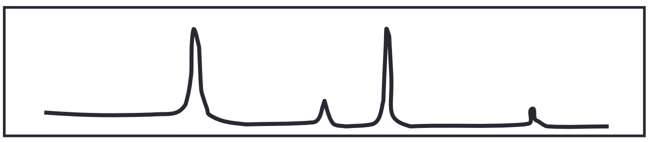


Why not always use Rejection Sampling?

The acceptance probability

$$p_{accept} = \frac{p(\mathbf{x}_{cand})}{M}$$

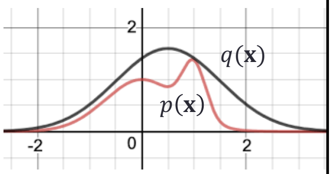
is negligible almost everywhere when p has "narrow maxima".



Rejection Sampling (more sophisticated)

In the n'th step of the algorithm:

1. Propose a candidate sampling point $\mathbf{x}_{cand}$ with a **non-constant** (global) proposal distribution $q(\mathbf{x})$
2. Accept $\mathbf{x}_{cand}$ only with probability



$$p_{accept} = \frac{p(\mathbf{x}_{cand})}{M\ q(\mathbf{x}_{cand})}$$

where

$$M \geq \max_{\mathbf{x}}\big(p(\mathbf{x})/g(\mathbf{x})\big)\,.$$

The set of thus accepted candidate points are samples from the probability distribution $p(\mathbf{x})$.

Markov Chain Monte Carlo (MCMC)

MCMC (simple)

In each step (starting in the current point $\mathbf{x}_j$) :

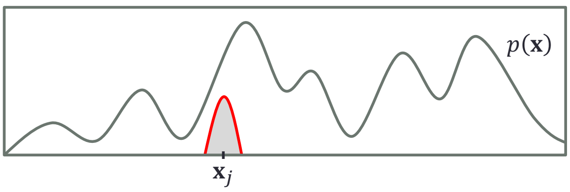
- Propose a jump $\mathbf{x}_j \longrightarrow \mathbf{x}_i$ using a (local or global) proposal distribution $U(\mathbf{x}_i|\mathbf{x}_j)$ with symmetry: $U(\mathbf{x}_i|\mathbf{x}_j) = U(\mathbf{x}_j|\mathbf{x}_i)$
- Accept $\mathbf{x}_i$ only with probability

$$p_{accept} = \min\left(1, \frac{p(\mathbf{x}_i)}{p(\mathbf{x}_j)}\right)$$

  Otherwise repeat $\mathbf{x}_j$

Advantages of MCMC

1. In MCMC, it is possible to use a **local** proposal distribution $U(\mathbf{x}_i|\mathbf{x}_j)$ if our only knowledge of the target distribution $p(\mathbf{x})$ is its ”smoothness”.



2. In MCMC we don’t need to normalize the target distribution $p(\mathbf{x})$. This is a **huge** advantage!

MCMC (more sophisticated)

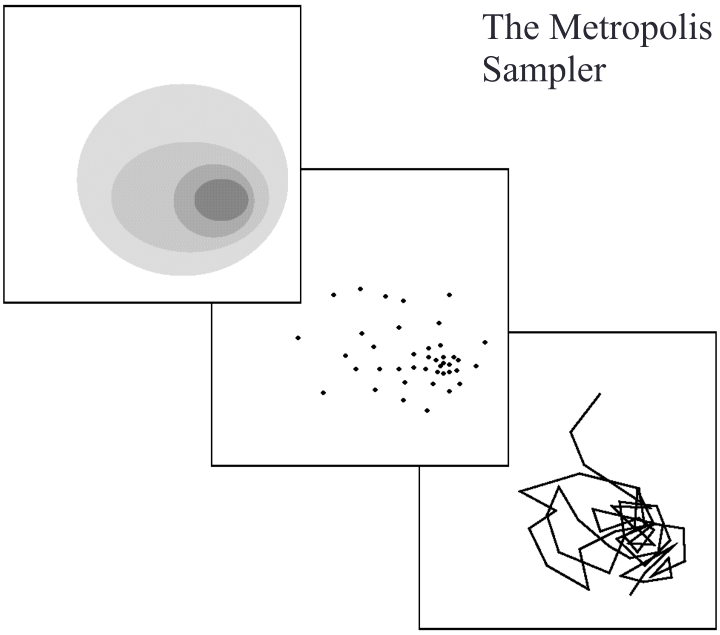
In each step (starting in the current point $\mathbf{x}_j$) :

- Propose a jump $\mathbf{x}_j \longrightarrow \mathbf{x}_i$ using a (possibly assymmetric, local or global) proposal distribution $U(\mathbf{x}_i|\mathbf{x}_j)$
- Accept $\mathbf{x}_i$ only with probability

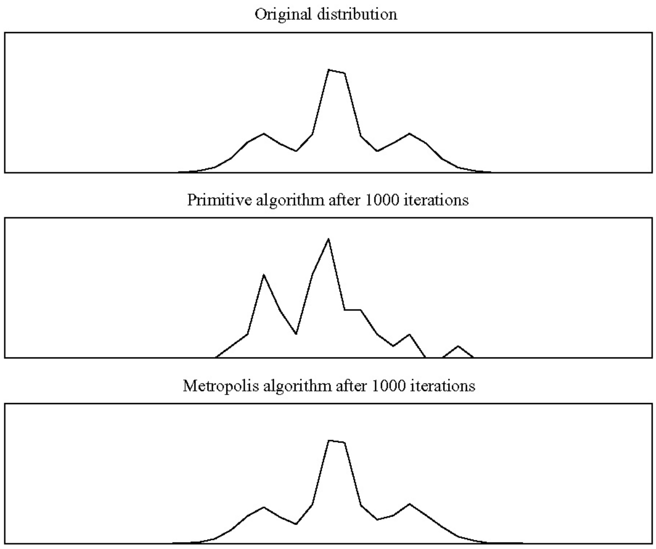
$$p_{accept} = \min\left(1, \frac{p(\mathbf{x}_i)U(\mathbf{x}_j|\mathbf{x}_i)}{p(\mathbf{x}_j)U(\mathbf{x}_i|\mathbf{x}_j)}\right)$$

Otherwise repeat $\mathbf{x}_j$

The Metropolis Sampler



MCMC vs. Rejection Sampling



MCMC: Why does it work?

Random walks in equilibrium

One step of a random walk in X is given by a set of **transition probability densities**

$$P(\mathbf{x}_i|\mathbf{x}_j)$$

The location distribution $p(\mathbf{x})$ for the random walk is an **equilibrium pdf** if it is left unchanged after one step of the walk, i.e., if

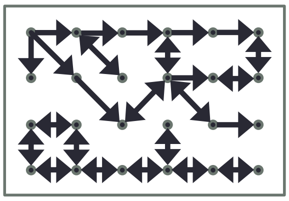
$$p(\mathbf{x}_i) = \int_X P(\mathbf{x}_i|\mathbf{x}_j)\ p(\mathbf{x}_j) d\mathbf{x}$$

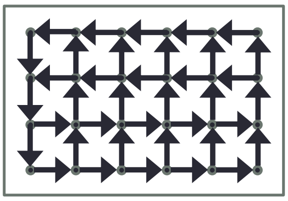
Random walks in equilibrium

It can be shown that, under certain mild assumptions, the sampling of a random walk will reach a stable (equilibrium) distribution.

The assumptions are that the walk is **not**:

Reducible or **Periodic**





Microscopic Reversibility: A way of maintaining equilibrium

Microscopic Reversibility means that in each iteration,

$$P(\mathbf{x}_i|\mathbf{x}_j)\ p(\mathbf{x}_j)d\mathbf{x}_i d\mathbf{x}_j = P(\mathbf{x}_j|\mathbf{x}_i)\ p(\mathbf{x}_i)d\mathbf{x}_i d\mathbf{x}_j$$

or,

$$\frac{P(\mathbf{x}_i|\mathbf{x}_j)}{P(\mathbf{x}_j|\mathbf{x}_i)} = \frac{p(\mathbf{x}_i)}{p(\mathbf{x}_j)}$$

and this is satisfied by the MCMC rule:

$$p_{accept} \equiv P(\mathbf{x}_i|\mathbf{x}_j) = \min\left(1, \frac{p(\mathbf{x}_i)}{p(\mathbf{x}_j)}\right)$$