technische universität
dortmund

Statistical Methods of Data Analysis

Monte Carlo Methods

Prof. Dr. Dr. Wolfgang Rhode   Dr. Maximilian Linhoff

2023

Overview

Overview

Previously Learned Methods for Random Number Generation

Inversion Sampling

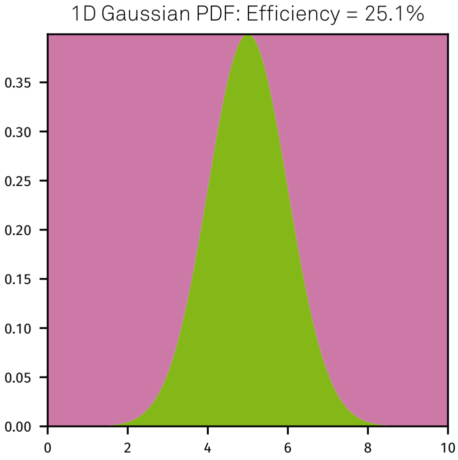
- \+ No rejection of random numbers
- \+ Efficient
- \- Integration of PDF
- \- Inversion of CDF
- \- Not applicable to every distribution

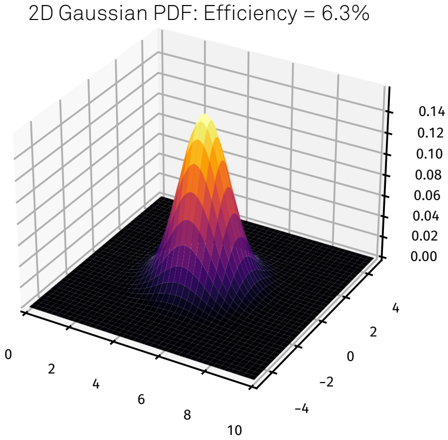
Rejection Sampling

- \+ No integration or inversion
- \+ Applicable to every function
- \- Rejection of random numbers
- \- Increasingly inefficient with increasing dimension
- \- Problematic for infinite distributions

“The Curse of Dimensionality”

The number of rejected random numbers increases with the dimension





Importance Sampling

Wanted distribution $f(x)$

1. Find appropriate function $g(x)$ with the properties
   - $g(x) \geq f(x) \, \forall x$
   - $G(x) = \int_{-\infty}^{x} g(x')\mathrm{d}x'$ is invertible
2. Modified *rejection sampling*
   - Sampling of two uniformly distributed random numbers $\xi_1, \xi_2 \in [0, 1)$
   - Transformation from $\xi_1 \rightarrow \xi_1'$ according to the distribution $g(x)$ with *inversion sampling*
   - Acceptance of ξ_1' if $g(\xi_1') \cdot \xi_2 \leq f(\xi_1')$

$\Longrightarrow$ The fraction $\frac{g(x)-f(x)}{f(x)}$ of random numbers is discarded

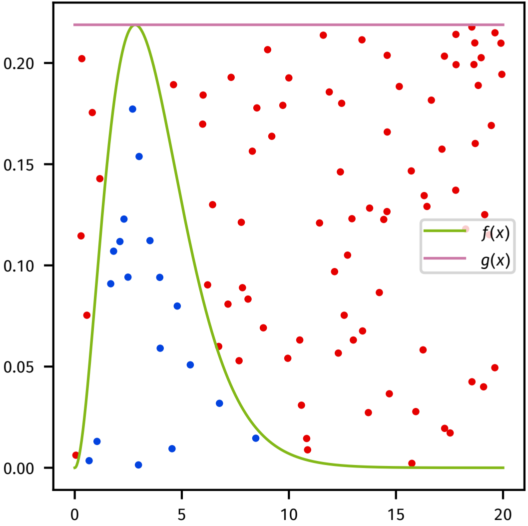
Importance Sampling (example)

Planck distribution $f(x) = N\frac{x^3}{e^x - 1}$

Difficulties:

- exponentially decreasing
- infinite distribution
- example: 82 out of 100 random numbers are discarded

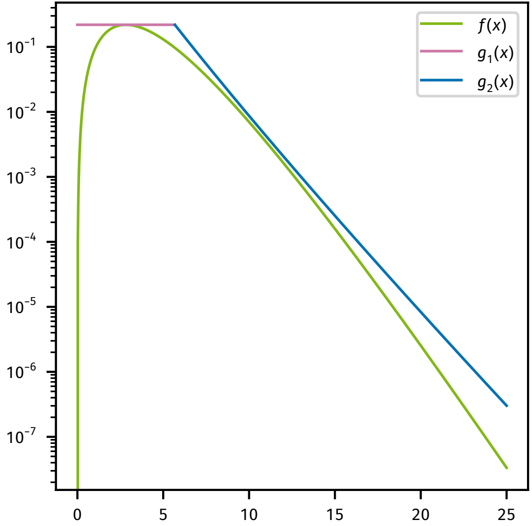
$\Longrightarrow$ see exercise sheet



Importance Sampling (example)

Majorant function

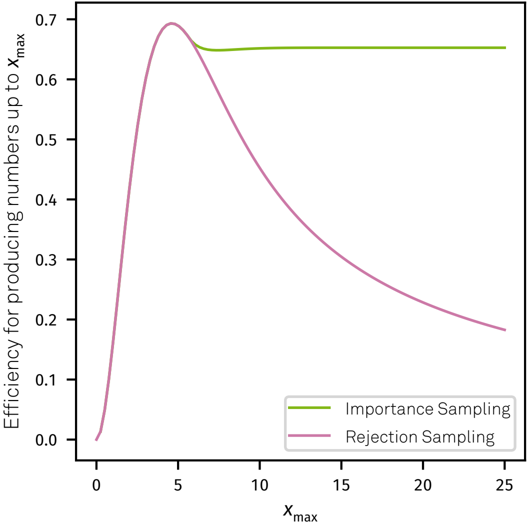
$$g(x) = \begin{cases} f(x_{max}), & \text{if } x \leq x_1 \\ 200N \cdot x^{-0.1} \exp(-x^{0.9}), & \text{if } x > x_1 \end{cases}$$



Importance Sampling (example)

Majorant function

$$g(x) = \begin{cases} f(x_{\text{max}}), & \text{if } x \leq x_1 \\ 200N \cdot x^{-0.1} \exp(-x^{0.9}), & \text{if } x > x_1 \end{cases}$$



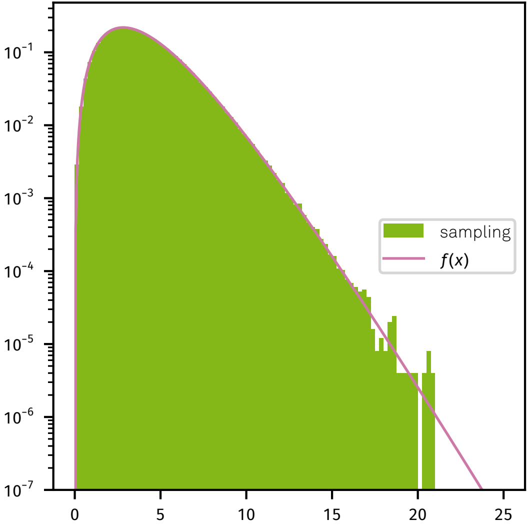
Importance Sampling (example)

Comparison: Generation of 10^6 random numbers with cutoff at 25

	rejected samples	run time
rejection	4 463 097	466 ms
importance	347 175	144 ms

Rejection sampling

- → uses 12 times more rejected random numbers
- → but only 3 times more runtime
- → Inefficiency gets larger with larger cutoff



Importance Sampling: Summary

Advantages:

+ Sampling to infinity is possible
+ More efficient than simple rejection sampling
+ No problem of dimensionality

Disadvantages:

- Finding a suitable majorant function can be difficult or impossible
- Simple rejection sampling might be faster

Overview

Importance Sampling

Markov Chain Monte Carlo

Monte Carlo Simulations

Markov
Chain
Monte
Carlo

In putting together this issue of *Computing in Science & Engineering*, we knew three things: it would be difficult to list just 10 algorithms; it would be fun to assemble the authors and read their papers; and, whatever we came up with in the end, it would be controversial. We tried to assemble the 10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century. Following is our list (here, the list is in chronological order; however, the articles appear in no particular order):

- Metropolis Algorithm for Monte Carlo
- Simplex Method for Linear Programming
- Krylov Subspace Iteration Methods
- The Decompositional Approach to Matrix Computations
- The Fortran Optimizing Compiler
- QR Algorithm for Computing Eigenvalues
- Quicksort Algorithm for Sorting
- Fast Fourier Transform
- Integer Relation Detection
- Fast Multipole Method

hand in developing the algorithm, and in other cases, the author is a leading authority.

In this issue

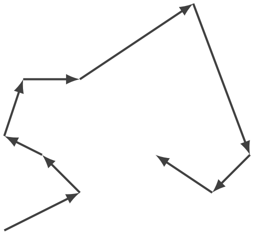
Monte Carlo methods are powerful tools for evaluating the properties of complex, many-body systems, as well as nondeterministic processes. Isabel Beichl and Francis Sullivan describe the Metropolis Algorithm. We are often confronted with problems that have an enormous number of dimensions or a process that involves a path with many possible branch points, each of which is governed by some fundamental probability of occurence. The solutions are not exact in a rigorous way, because we randomly sample the problem. However, it is possible to achieve nearly exact results using a relatively small number of samples compared to the problem's dimensions. Indeed, Monte Carlo methods are the only practical choice for evaluating problems of high dimensions.

John Nash describes the Simplex method for solving linear programming problems. (The use of the word *programming* here really refers to scheduling or

Comput. Sci. Eng. 2 (2000), 22

Markov Chain

- Stochastic transitions
- Transition probability is only dependent on initial and final state
- Dynamic Sampling



Detailed Balance (Reversible Processes)

- Stationary distribution $f(x)$
- Transition probability $M_{i\to j}$

$$M_{i\to j} f(x_i) = M_{j\to i} f(x_j)$$

$$\frac{M_{i\to j}}{M_{j\to i}} = \frac{f(x_j)}{f(x_i)}$$

The transition probability is proportional to the ratio of states.

Metropolis Algorithm

Transition probability

$$M_{i\to j} = \min\left(1, \frac{f(x_j)}{f(x_i)}\right)$$

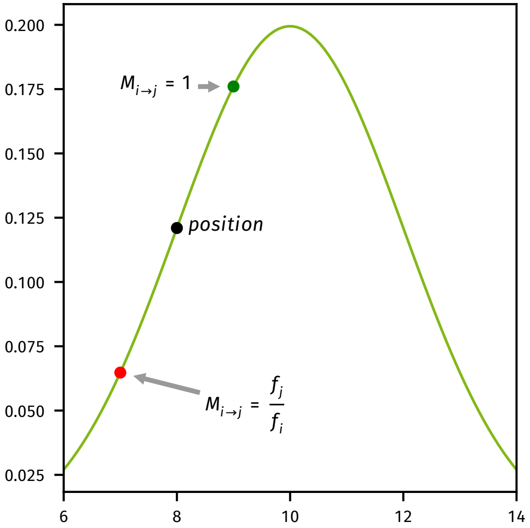
fulfills the "Detailed Balance" condition.

Generate random number $\xi \in [0, 1]$

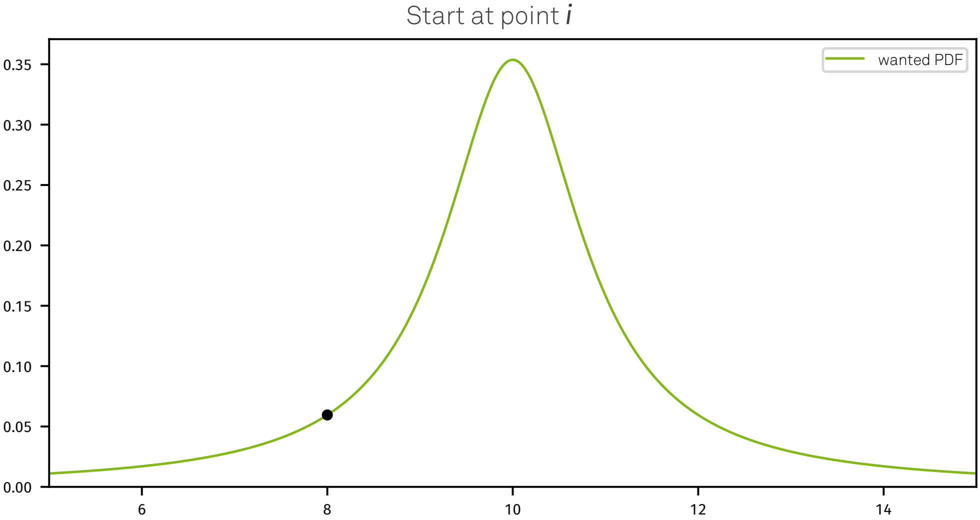
$$x_{i+1} = \begin{cases} x_j, & \text{if } \xi \leq M_{i\to j} \\ x_i, & \text{if } \xi > M_{i\to j} \end{cases}$$

How to determine x_j?

⇒ Random sampling from fixed step size distribution

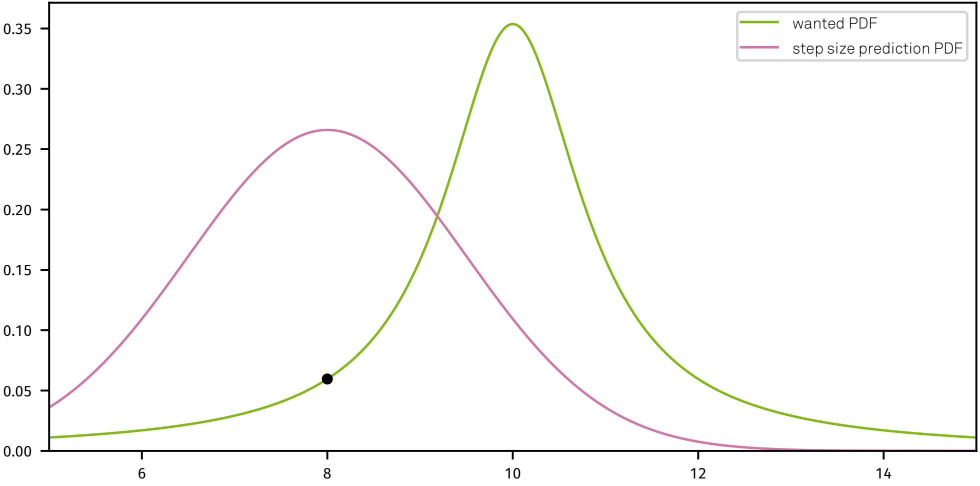


Determination of the Step Size

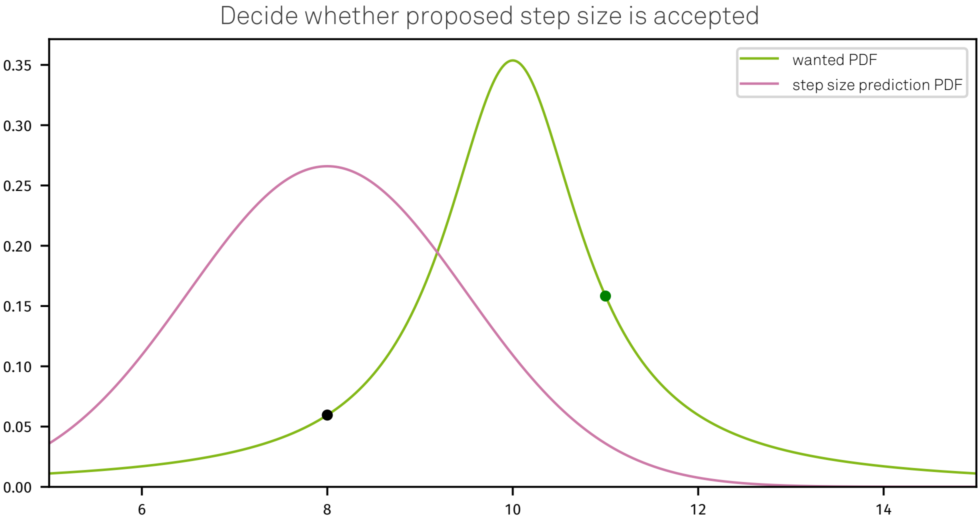


Determination of the Step Size

Sample next step from proposed step size PDF $g(x)$

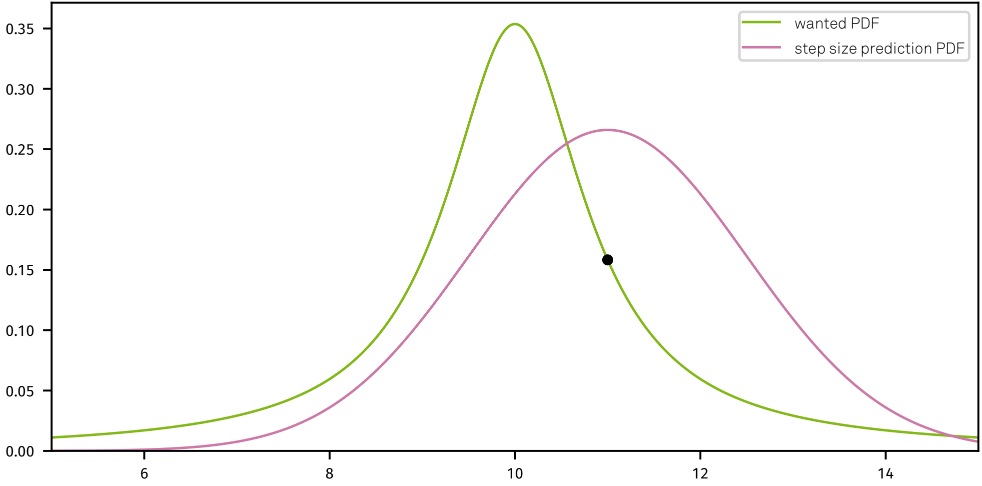


Determination of the Step Size



Determination of the Step Size

Sample next step from distribution $g(x)$ (different mean)



Metropolis-Hastings Algorithm

Proposed step for antisymmetric distribution

$$g_{i\to j} > g_{j\to i}$$

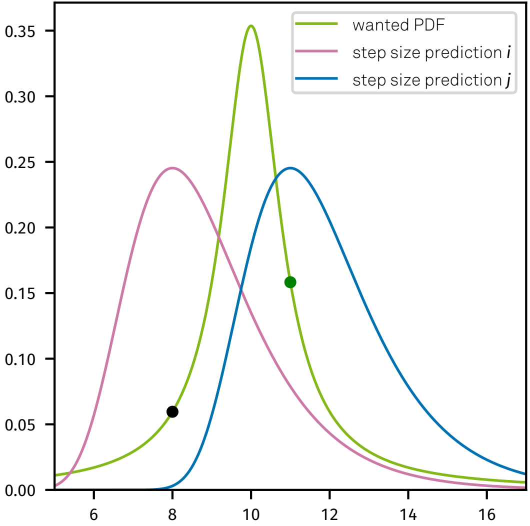
does not fulfill the "detailed balance" condition, but

$$M_{i\to j} f(x_i) g_{i\to j} = M_{j\to i} f(x_j) g_{j\to i}$$

does.

The transition probability is

$$M_{i\to j} = \min\left(1, \frac{f(x_j)}{f(x_i)} \frac{g(x_i|x_j)}{g(x_j|x_i)}\right)$$



MCMC Algorithm

- Initialize: $t = 0$, starting point x_0
- Repeat:
  1. Propose step: sample x' from $g(x|x_t)$
  2. Calculate acceptance probability: $p = \min\left(1, \frac{p(x')}{p(x_t)} \frac{g(x_t|x')}{g(x'|x_t)}\right)$
  3. Acceptance step:
     - 3.1 Sample uniformly distributed random number $\xi \in [0, 1)$
     - 3.2 If $\xi \leq p$, accept $x' \Rightarrow x_{t+1} = x'$
     - 3.3 Else: Reject $x' \Rightarrow x_{t+1} = x_t$
  4. Raise $t = t + 1$

Autocorrelation vs Sampling Efficiency

- Acceptance rate too small (< 30 %)
  - too small area of distribution is explored
  - few points represent distribution

$\Longrightarrow$ Step size too large

- Acceptance rate too large (> 70%)
  - Many points close to one another
  - Flat area of distribution
  - Ratio close to 1

$\Longrightarrow$ Step size too small

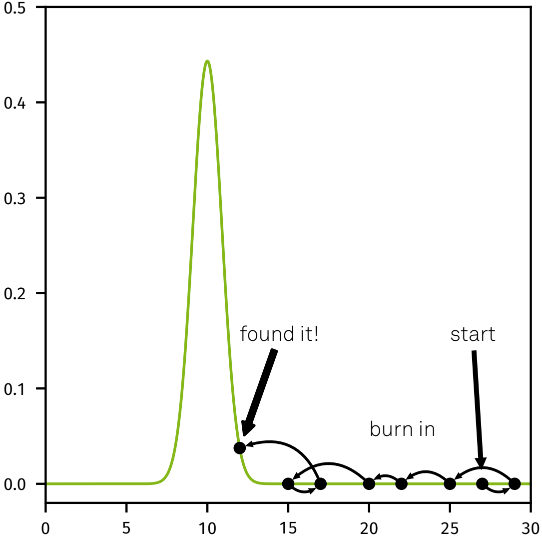
Burn-in Phase

Scenario when starting point is far from significant area:

- Flat distribution
- Majority of points not in significant area
- Significant area not found

Ansatz:

- Rejection of burn-in phase
- Ensemble of *walkers*



Problems and Outlook

Problems:

- No independent sampling
- Dependence on starting point
- Dependence on step size

Further modifications $\Longrightarrow$ Hamilton Monte-Carlo

- Considers state and gradient
- Analogous to potential and kinetic energy
- Step size distribution depends on gradient

Overview

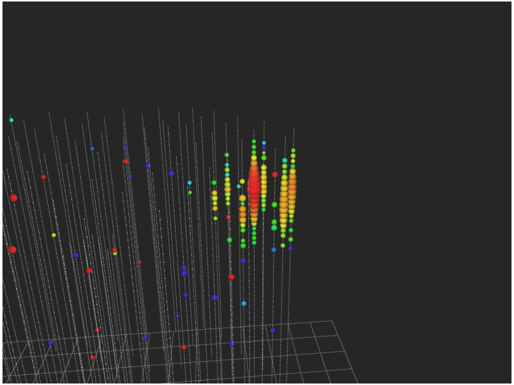
Importance Sampling

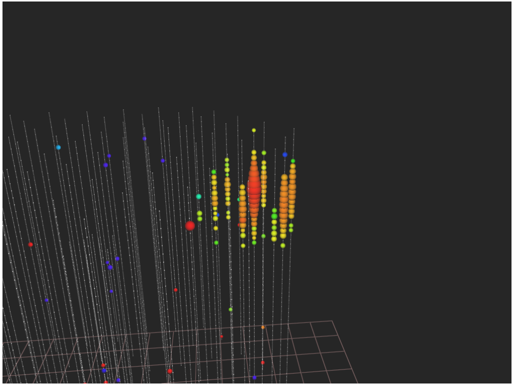
Markov Chain Monte Carlo

Monte Carlo Simulations

Monte Carlo Simulations

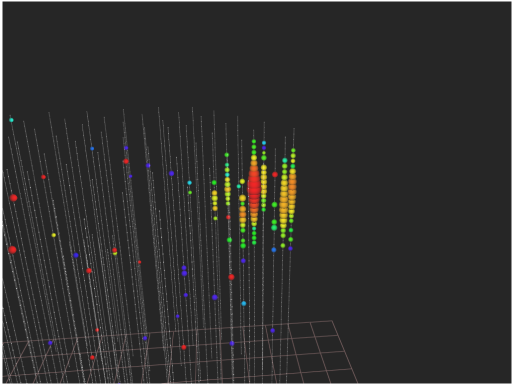
Which event is simulated? Which one is measured?



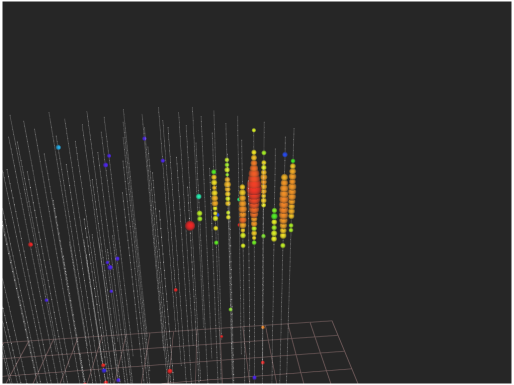


Monte Carlo Simulations

Which event is simulated? Which one is measured?



measured

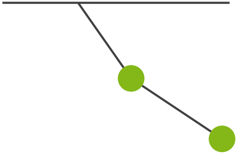


simulated

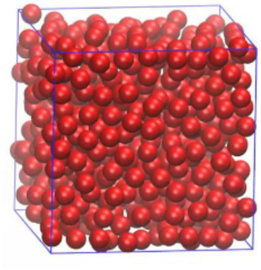
Analytical Simulations (without Monte Carlo)

- Analytical solutions of differential equations
- Final state is determinable from initial state $\Longrightarrow$ deterministic

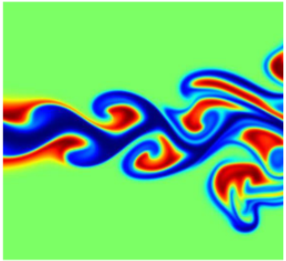
Double pendulum (Euler-Lagrange)



Molecule dynamics (Newton)



Currents (Navier-Stokes)



Monte Carlo Simulations

Stochastic Processes

Same starting conditions ⇒ different outcomes

Solid state physics:

- Simulation of ensembles
- Calculation of integrals / mean values by averaging over random samples
- Comp. phys. : mean energy of 2D Ising model

Particle physics (our focus):

- Simulation of individual particles
- Random generation, propagation, detection and reconstruction of individual events
- Averaging over statistical processes not meaningful here

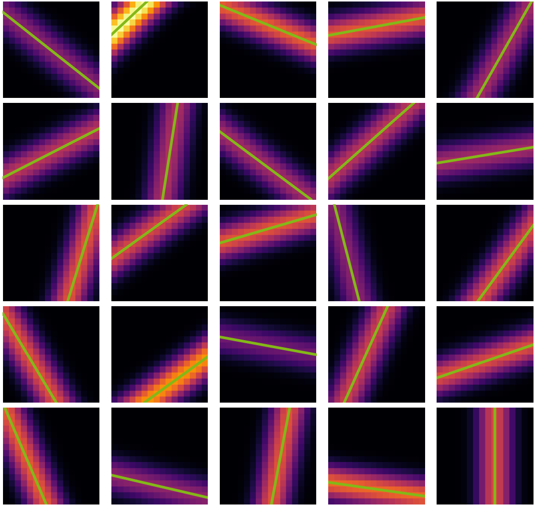
Monte Carlo Simulation Chain

- An event is a superposition of multiple complex processes
- Analytical solution of entire process is impossible
- Separate treatment of individual processes
- At every step: Sample random number from distribution and determine result
- Coverage of parameter space by frequent repetition

Generation → Propagation → Detection

Why We Need Monte Carlo Simulations

No Monte Carlo



Example: Muon track inside detector (IceCube)

Assumptions

- Constant muon energy
- Gaussian signal attenuation

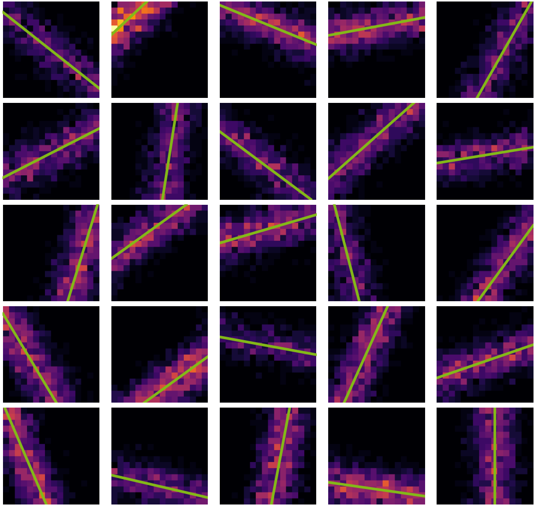
Why We Need Monte Carlo Simulations

Example: Muon track inside detector (IceCube)

Assumptions

- Constant muon energy
- Gaussian signal attenuation
- Poisson distributed photon hits

With Poisson-Noise



Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
    I. Sampling of scattering length
    II. Propagation to next scattering point
    III. Sampling of scattering angle
    IV. Rotation around scattering angle

Initialize

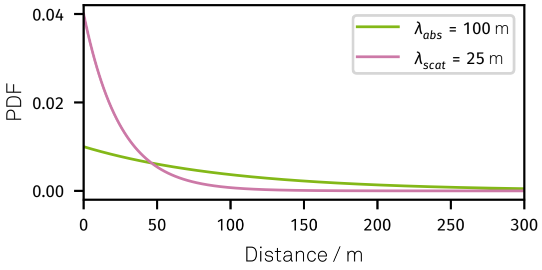
- Starting position
- Direction
- Average absorption length λ_{abs}
- Average scattering length λ_{scat}
- Average scattering angle $\cos(\Theta)$

Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
    I. Sampling of scattering length
    II. Propagation to next scattering point
    III. Sampling of scattering angle
    IV. Rotation around scattering angle

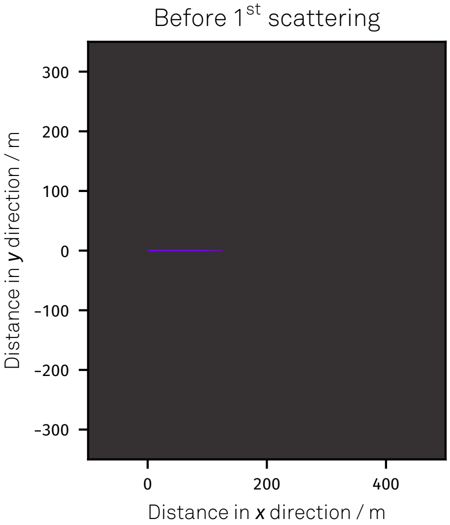
Sample exponentially distributed absorption and scattering length with inversion sampling

$$p(x) = \frac{1}{\lambda} e^{\frac{-x}{\lambda}}$$



Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
    I. Sampling of scattering length
    II. Propagation to next scattering point
    III. Sampling of scattering angle
    IV. Rotation around scattering angle

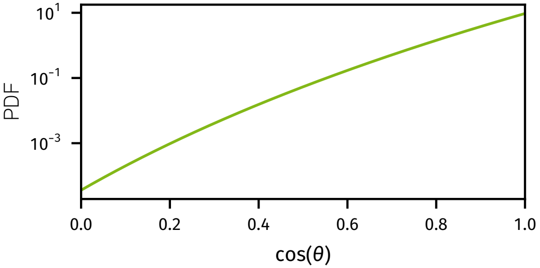


Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
   I. Sampling of scattering length
   II. Propagation to next scattering point
   III. Sampling of scattering angle
   IV. Rotation around scattering angle

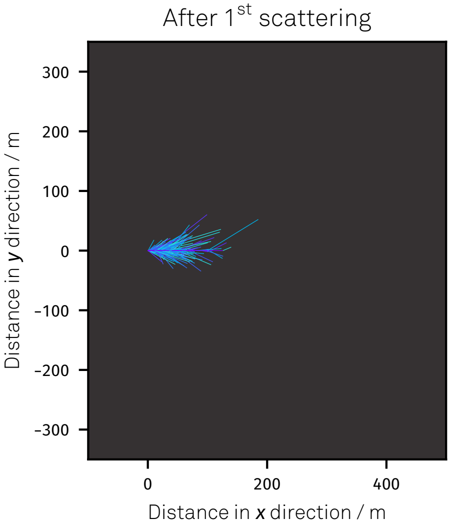
The scattering angle is distributed as

$$f(\cos\Theta) = \frac{1+\alpha}{2}\left(\frac{1+\cos\Theta}{2}\right)^{\alpha}, \alpha = \frac{2\langle\cos\Theta\rangle}{1-\langle\cos\Theta\rangle}$$



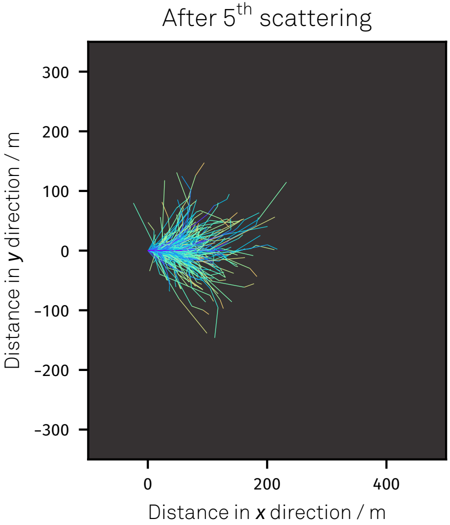
Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
   I. Sampling of scattering length
   II. Propagation to next scattering point
   III. Sampling of scattering angle
   IV. Rotation around scattering angle



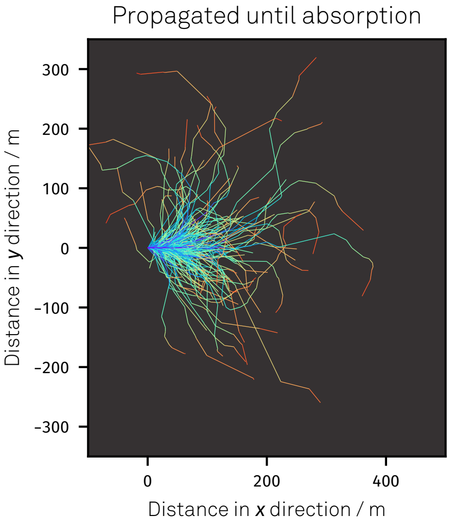
Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
    I. Sampling of scattering length
    II. Propagation to next scattering point
    III. Sampling of scattering angle
    IV. Rotation around scattering angle

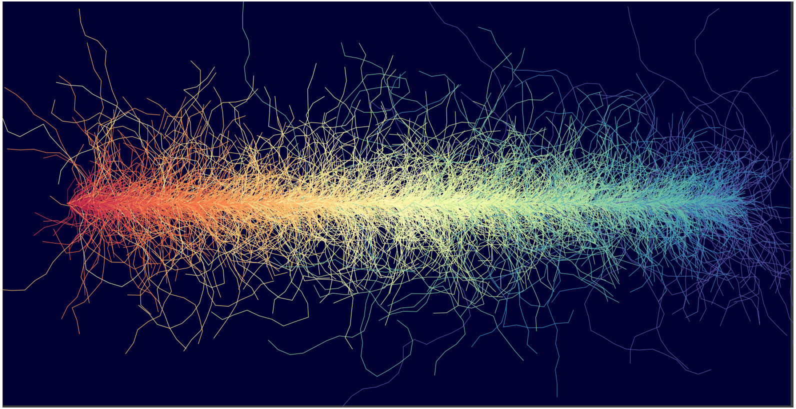


Example: Photon Propagation

1. Initialization
2. Sampling of absorption length
3. Propagation and scattering until absorption length is reached
    I. Sampling of scattering length
    II. Propagation to next scattering point
    III. Sampling of scattering angle
    IV. Rotation around scattering angle



Photon Propagation along Muon Path



here no stochastic muon losses yet

Particle Propagation

Analogous to photon propagation

- Absorption $\Longrightarrow$ decay
- Scattering $\Longrightarrow$ stochastic energy losses + scattering
- Continuous losses

Recap: Sampling of Random Events From Distribution

Photon propagation

- PDF is invertible (exponential distribution) $\Longrightarrow$ Inversion sampling

Muon propagation

- Complex distributions (cross sections) are often only numerically available or they must be integrated first
- How are these sampled then?

Recap: Sampling of Random Events From Distribution

Photon propagation

- PDF is invertible (exponential distribution) $\Longrightarrow$ Inversion sampling

Muon propagation

- Complex distributions (cross sections) are often only numerically available or they must be integrated first
- How are these sampled then? $\Longrightarrow$ inversion sampling (because more efficient):
  - MC simulations are resource-intensive
  - Large MC are necessary for analyses
  - Interpolation instead of integration
  - Look-up tables instead of new calculations

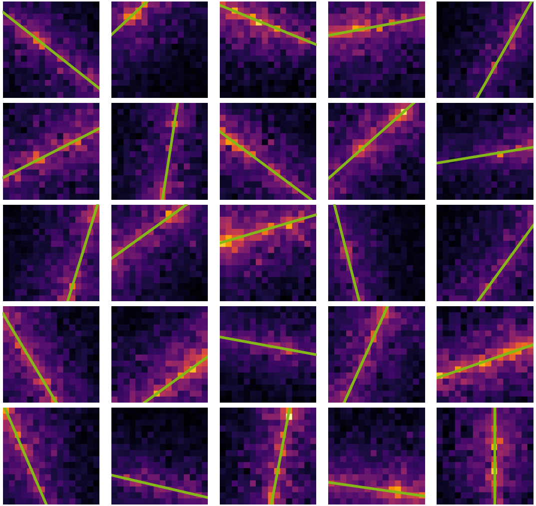
Why We Need Monte Carlo Simulations

Example: Muon track inside detector (IceCube)

Assumptions

- Constant muon energy
- Gaussian signal attenuation
- Poisson distributed photon hits
- Stochastic energy losses

With Stocachistic Losses



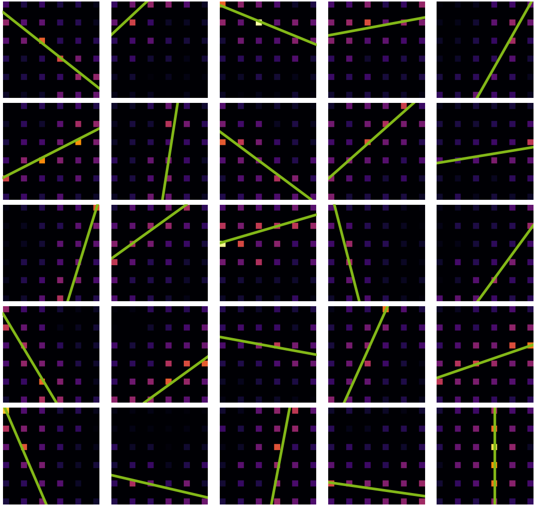
Why We Need Monte Carlo Simulations

Example: Muon track inside detector (IceCube)

Assumptions

- Constant muon energy
- Gaussian signal attenuation
- Poisson distributed photon hits
- Stochastic energy losses
- Poor detector resolution

Detector with Low Resolution



Complete Monte Carlo Simulation Chain

- Up to now: only Monte Carlo propagation
- Generator
  - Particle energy spectrum
  - Particle flux
- Detector
  - Acceptance probability
  - Trigger
  - Electronics (e. g. white noise)

$\Longrightarrow$ fully simulated events are comparable to measurements

Reconstruction of a single event