1. Introduction to Monte Carlo techniques

Torbjörn Sjöstrand

Theoretical High Energy Physics
Department of Astronomy and Theoretical Physics
Lund University
Sölvegatan 14A, SE-223 62 Lund, Sweden

Lund, 5 November 2012
Course overview

Coordinator: Anders Irbäck, 046–222 34 93, anders@thep.lu.se

- P0: Introduction to Monte Carlo Techniques (TS, A Irbäck)
- P1: Write a parton shower (TS)
- P2: Markov chain Monte Carlo simulation of protein fibril formation (A Irbäck)
- P3: Stellar populations with clusters (M Davies)
- P4 Monte Carlo simulation of photon interactions with matter (M Ljungberg)
- P5: Molecular simulations (P Linse, M Lund)
- Px: longer project continuing on either of above

My lectures and other course material can be found at http://home.thep.lu.se/~torbjorn/compute2012.html
My involvement: next two+ weeks

- **Today**: introduction, integration of 1-dimensional functions, and random selection according to them
- **Tomorrow**: random number generators, special tricks, more (but still few) dimensions
- **(Wednesday, Thursday)**: Anders lectures
- **Friday 17.00**: deadline for hand-in of two warmup exercises
- **Next Monday**: problems with time evolution/ordering, and the technology of the main project of the week
- **Next Tuesday**: introduction to particle physics, and the context of the main project of the week
- **Next Wednesday (– Friday)**: check progress of project
- **Next Friday 17.00**: deadline for hand-in of project report
- **Following Tuesday**: my feedback
Buffon’s needle (proposed 1733): probability for needle to cross line is related to $\pi$

…but gambling & odds are older
“Spatial” problems: no memory/ordering (this week)

1. Integrate a function
2. Pick a point at random according to a probability distribution

“Temporal” problems: has memory (next week)

1. Radioactive decay: probability for a radioactive nucleus to decay at time \( t \), given that it was created at time 0

In reality combined into multidimensional problems:

1. Random walk (variable step length and direction)
2. Charged particle propagation through matter (stepwise loss of energy by a set of processes)
3. **Parton showers** (cascade of successive branchings)
Random numbers

For now assume algorithm that returns “random numbers” $R$, uniformly distributed in range $0 < R < 1$ and uncorrelated. More explanation/examples tomorrow (but not my expertise).

Task 1: find and learn how to use a random number generator on your platform of preference.

Fallback: an implementation of the Marsaglia–Zaman–Tsang algorithm is available on my course page. It allows for $\sim 900\,000\,000$ different sequences, seq numbered from 1 upwards.

- C++: `Rndm rndm(seq);` creates+initializes; `rndm.flat();` returns next random number

- Fortran: `RNDM(seq)` initializes first time called; always returns next random number

Adapt and check that it works (rewritten for standalone).
Pick among discrete possibilities

Assume \( n \) possible outcomes with (unnormalized) probabilities \( P_i, 1 \leq i \leq n \). Pick one of them according to

1. \( i = 0 \)
   \[ P_R = R \sum_{i=0}^{n} P_i \]
2. \( i = i + 1 \)
   \[ P_R = P_R - P_i \]
3. if \( P_R > 0 \) cycle to 2

Example 1:
Poissonian \( P_i = \left( \frac{\langle n \rangle^i}{i!} \right) e^{-\langle n \rangle}, i \geq 0 \).
Note that \( P_i = \left( \frac{\langle n \rangle}{i} \right) P_{i-1} \):

1. \( i = -1; \quad P_R = R; \quad P_{\text{now}} = e^{-\langle n \rangle} \)
2. \( i = i + 1; \quad \text{if} \ (i > 0) \ P_{\text{now}} = P_{\text{now}} \left( \frac{\langle n \rangle}{i} \right); \)
   \[ P_R = P_R - P_{\text{now}} \]
3. if \( P_R > 0 \) cycle to 2
Assume function $f(x)$, studied range $x_{\text{min}} < x < x_{\text{max}}$, where $f(x) \geq 0$ everywhere.

Two connected standard tasks:

1. Calculate (approximatively)
   $$\int_{x_{\text{min}}}^{x_{\text{max}}} f(x') \, dx'$$

2. Select $x$ at random according to $f(x)$

In step 2 $f(x)$ is viewed as “probability distribution” with implicit normalization to unit area, and then step 1 provides overall correct normalization.
Integral as an area/volume

Theorem

An $n$-dimensional integration $\equiv$ an $n + 1$-dimensional volume

$$
\int f(x_1, \ldots, x_n) \, dx_1 \ldots dx_n \equiv \int \int_0^{f(x_1, \ldots, x_n)} 1 \, dx_1 \ldots dx_n \, dx_{n+1}
$$

since $\int_0^{f(x)} 1 \, dy = f(x)$. 
Integral as an area/volume

Theorem

An \( n \)-dimensional integration \( \equiv \) an \( n + 1 \)-dimensional volume

\[
\int f(x_1, \ldots, x_n) \, dx_1 \ldots dx_n \equiv \int \int_0^{f(x_1, \ldots, x_n)} 1 \, dx_1 \ldots dx_n \, dx_{n+1}
\]

since \( \int_0^{f(x)} 1 \, dy = f(x) \).

So, for 1 + 1 dimension, selection of \( x \) according to \( f(x) \) is equivalent to uniform selection of \((x, y)\) in the area \( x_{\text{min}} < x < x_{\text{max}}, 0 < y < f(x) \).

Therefore

\[
\int_{x_{\text{min}}}^{x} f(x') \, dx' = R \int_{x_{\text{min}}}^{x_{\text{max}}} f(x') \, dx'
\]

(area to left of selected \( x \) is uniformly distributed fraction of whole area)
Basic method 1: analytical solution

If know primitive function $F(x)$ and know inverse $F^{-1}(y)$ then

$$F(x) - F(x_{\text{min}}) = R (F(x_{\text{max}}) - F(x_{\text{min}})) = R A_{\text{tot}}$$

$$\implies x = F^{-1}(F(x_{\text{min}}) + R A_{\text{tot}})$$

Proof: introduce $z = F(x_{\text{min}}) + R A_{\text{tot}}$. Then

$$\frac{dP}{dx} = \frac{dP}{dR} \frac{dR}{dx} = 1 \frac{1}{\frac{dx}{dR}} = \frac{1}{\frac{dx}{dz} \frac{dz}{dR}} = \frac{1}{\frac{dF^{-1}(z)}{dz} \frac{dz}{dR}} = \frac{\frac{dF(x)}{dx}}{A_{\text{tot}}} = \frac{f(x)}{A_{\text{tot}}}$$

Example 2:

$f(x) = 2x, \ 0 < x < 1, \implies F(x) = x^2$

$$F(x) - F(0) = R (F(1) - F(0)) \implies x^2 = R \implies x = \sqrt{R}$$

Example 3:

$f(x) = e^{-x}, \ x > 0, \ F(x) = 1 - e^{-x}$

$$1 - e^{-x} = R \implies e^{-x} = 1 - R = R \implies x = -\ln R$$
Basic method 2: hit-and-miss

If \( f(x) \leq f_{\text{max}} \) in \( x_{\text{min}} < x < x_{\text{max}} \)
use interpretation as an area

1. select \[ x = x_{\text{min}} + R (x_{\text{max}} - x_{\text{min}}) \]
2. select \( y = R f_{\text{max}} \) (new \( R! \))
3. while \( y > f(x) \) cycle to 1

Integral as by-product:

\[
l = \int_{x_{\text{min}}}^{x_{\text{max}}} f(x) \, dx = f_{\text{max}} (x_{\text{max}} - x_{\text{min}}) \frac{N_{\text{acc}}}{N_{\text{try}}} = A_{\text{tot}} \frac{N_{\text{acc}}}{N_{\text{try}}}
\]

Binomial distribution with \( p = N_{\text{acc}}/N_{\text{try}} \) and \( q = N_{\text{fail}}/N_{\text{try}} \),
so error

\[
\frac{\delta l}{l} = \frac{A_{\text{tot}} \sqrt{p \, q / N_{\text{try}}}}{A_{\text{tot}} \, p} = \sqrt{\frac{q}{p \, N_{\text{try}}}} = \sqrt{\frac{q}{N_{\text{acc}}}} < \frac{1}{\sqrt{N_{\text{acc}}}}
\]
Example 4:

\[ f(x) = x^\alpha, \quad \alpha > 0, \]

\[ 0 \leq x \leq 1 \Rightarrow 0 \leq f(x) \leq 1 \]

\[ F(x) = x^{\alpha+1}/(\alpha + 1) \]

\[ p = l = \int_0^1 f(x) \, dx = 1/(\alpha + 1) \]

\[ q = 1 - l = \alpha/(\alpha + 1) \]

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( l )</th>
<th>( \sqrt{N}_{\text{try}} \delta l )</th>
<th>( \sqrt{N}_{\text{try}}(\delta l/l) )</th>
<th>( \sqrt{N}_{\text{acc}}(\delta l/l) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>0.707</td>
</tr>
<tr>
<td>2</td>
<td>0.333</td>
<td>0.471</td>
<td>1.413</td>
<td>0.816</td>
</tr>
<tr>
<td>1/2</td>
<td>0.667</td>
<td>0.471</td>
<td>0.706</td>
<td>0.577</td>
</tr>
<tr>
<td>9</td>
<td>0.1</td>
<td>0.3</td>
<td>3.0</td>
<td>0.949</td>
</tr>
<tr>
<td>1/9</td>
<td>0.9</td>
<td>0.3</td>
<td>0.333</td>
<td>0.316</td>
</tr>
</tbody>
</table>
Crude Monte Carlo integration

Hit-and-miss not most efficient integration: for each $x$ picked, and $f(x)$ evaluated, only accept/reject statistics is used.

Better use full $f(x)$ information:

\[
I = \int_{x_{\min}}^{x_{\max}} f(x) \, dx = (x_{\max} - x_{\min}) \frac{1}{N_{\text{try}}} \sum_{i=1}^{N_{\text{try}}} f(x_i) = \Delta x \langle f(x) \rangle
\]

\[
\delta I = \frac{1}{\sqrt{N_{\text{try}}}} \Delta x \sqrt{\langle f^2(x) \rangle - \langle f(x) \rangle^2}
\]

with \[
\langle f^2(x) \rangle = \frac{1}{N_{\text{try}}} \sum_{i=1}^{N_{\text{try}}} f^2(x_i)
\]
Example 4 (continued):

\[ f(x) = x^\alpha, \quad \alpha > 0, \quad 0 \leq x \leq 1 \]

\[ \langle f(x) \rangle = \int_0^1 x^\alpha \, dx = \frac{1}{\alpha + 1} \]

\[ \langle f^2(x) \rangle = \int_0^1 x^{2\alpha} \, dx = \frac{1}{2\alpha + 1} \]

\[ \delta f = \sqrt{\langle f^2(x) \rangle - \langle f(x) \rangle^2} = \frac{\alpha}{(\alpha + 1)\sqrt{2\alpha + 1}} \]

<table>
<thead>
<tr>
<th>\alpha</th>
<th>\text{I}</th>
<th>\sqrt{N_{\text{try}}} \delta I_{\text{HaM}}</th>
<th>\sqrt{N_{\text{try}}} \delta I_{\text{CMCi}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.289</td>
</tr>
<tr>
<td>2</td>
<td>0.333</td>
<td>0.471</td>
<td>0.298</td>
</tr>
<tr>
<td>1/2</td>
<td>0.667</td>
<td>0.471</td>
<td>0.236</td>
</tr>
<tr>
<td>9</td>
<td>0.1</td>
<td>0.3</td>
<td>0.206</td>
</tr>
<tr>
<td>1/9</td>
<td>0.9</td>
<td>0.3</td>
<td>0.090</td>
</tr>
</tbody>
</table>
Conventional integration

Plethora of “deterministic” integration formulae:
Simpson, Newton-Cotes, Gauss, Richardson, Romberg, . . .

For \( I = \int_{0}^{1} f(x) \, dx \)

\[
\begin{align*}
\text{Trapezoid} & \quad I = \frac{1}{2} f(0) + \frac{1}{2} f(1) + O(f'') \\
\text{Simpson} & \quad I = \frac{1}{6} f(0) + \frac{4}{6} f \left( \frac{1}{2} \right) + \frac{1}{6} f(1) + O(f^{(4)})
\end{align*}
\]

extended by split into subranges.

For comparable number of points \( N \) error then scales like

\[
\begin{align*}
\text{Monte Carlo} & \quad 1/\sqrt{N} \\
\text{Trapezoid} & \quad 1/N^2 \\
\text{Simpson} & \quad 1/N^4
\end{align*}
\]

Monte Carlo will not win for 1-dimensional integration.
Conventional integration (2)

Game changes for $d$ dimensions:

- Monte Carlo: $1/\sqrt{N}$
- Trapezoid: $1/N^{2/d}$
- Simpson: $1/N^{4/d}$

Also: 20 dimensions, Simpson $\Rightarrow 3^{20} \approx 3 \cdot 10^9$ points, all except one on border.

Generally, advantages of simple Monte Carlo integration include

- proportionately faster convergence in many dimensions
- discontinuous functions no problem
- arbitrarily complex integration regions
- few points needed to get first estimate
- easy error estimate
- by-product of Monte Carlo selection of $x$
Detour: stratified sampling

Split integration range into subranges (adjoint, non-overlapping).
Assume \( n \) subranges, \( 1 \leq i \leq n \), \( \Delta x_i = x_{i,\text{max}} - x_{i,\text{min}} \),
and \( N_i \) points in respective subrange:

\[
I = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \Delta x_i \langle f(x) \rangle_i
\]

\[
(\delta I)^2 = \sum_{i=1}^{n} \frac{(\Delta x_i \delta f_i)^2}{N_i} = \sum_{i=1}^{n} \frac{(\Delta x_i)^2}{N_i} \left( \langle f^2(x) \rangle_i - \langle f(x) \rangle_i^2 \right)
\]

Uniform stratification: all \( \Delta x_i \) and \( N_i \) the same, does reduce \( \delta I \).
Ultimately \( N_i = 1 \), \( n \) large, \( \approx \) conventional integration.
Variance reduction: pick smaller ranges wherever \( f'(x) \) is large,
rather than \( f(x) \) itself.
Wrong way to go for selection according to a distribution!
From now on only study techniques that allow (unbiased) selection.
Importance sampling

Improved version of hit-and-miss:
If \( f(x) \leq g(x) \) in
\( x_{\min} < x < x_{\max} \)
and \( G(x) = \int g(x') \, dx' \) is simple
and \( G^{-1}(y) \) is simple

1. select \( x \) according to \( g(x) \) distribution
2. select \( y = R \, g(x) \) (new \( R! \))
3. while \( y > f(x) \) cycle to 1

Example 5:
\( f(x) = x \, e^{-x}, \ x > 0 \)
Attempt 1: \( F(x) = 1 - (1 + x) \, e^{-x} \) not invertible
Attempt 2: \( f(x) \leq f(1) = e^{-1} \) but \( 0 < x < \infty \)
Importance sampling (2)

**Attempt 3:** $g(x) = N e^{-x/2}$

$$\frac{f(x)}{g(x)} = \frac{x e^{-x}}{N e^{-x/2}} = \frac{x e^{-x/2}}{N} \leq 1$$

for rejection to work, so find maximum:

$$\frac{d}{dx} \left( \frac{f(x)}{g(x)} \right) = \frac{1}{N} \left( 1 - \frac{x}{2} \right) e^{-x/2} = 0 \implies x = 2$$

Normalize so $g(2) = f(2) \Rightarrow N = \frac{2}{e}$

$G(x) \propto 1 - e^{-x/2} = R \Rightarrow x = -2 \ln R$

1. select $x = -2 \ln R$
2. select $y = R g(x) = R 2e^{-(1+x/2)}$
3. while $y > f(x) = x e^{-x}$ cycle to 1

efficiency $= \frac{\int_{0}^{\infty} f(x) \, dx}{\int_{0}^{\infty} g(x) \, dx} = \frac{e}{4}$
Variable transformation

Importance sampling can be reinterpreted as variable transformation

\[ \int f(x) \, dx = \int \frac{f(x)}{g(x)} g(x) \, dx = \int \frac{f(x)}{g(x)} \, dG(x) \]

- map to finite \( x \) range
- map away singular/peaked regions

Example 6:
\( f(x) = \exp(-x^2), \ 1 \leq x < \infty \) (or \( \exp(-x^\alpha) \) with \( \alpha > 1 \))
def. \( t = \exp(-x) \Rightarrow x = -\ln t, \ 0 \leq t \leq 1/e \)

\[ \int_1^\infty e^{-x^2} \, dx = \int_1^\infty \frac{e^{-x^2}}{e^{-x}} e^{-x} \, dx = \int_0^{1/e} \frac{e^{-\ln^2 t}}{t} \, dt = \int_0^{1/e} t^{-1-\ln t} \, dt \]

Pick \( t \) uniformly in \( 0 < t \leq 1/e \), repeatedly until \( t^{-1-\ln t} > R \), and then obtain \( x = -\ln t \)
If \( f(x) \leq g(x) = \sum_i g_i(x) \), where all \( g_i \) “nice” (\( G_i(x) \) invertible) but \( g(x) \) not

1. select \( i \) with relative probability
   \[
   A_i = \int_{x_{\text{min}}}^{x_{\text{max}}} g_i(x') \, dx'
   \]

2. select \( x \) according to \( g_i(x) \)
3. select \( y = R \, g(x) = R \sum_i g_i(x) \)
4. while \( y > f(x) \) cycle to 1

Works since

\[
\int f(x) \, dx = \int \frac{f(x)}{g(x)} \sum_i g_i(x) \, dx = \sum_i A_i \int \frac{g_i(x) \, dx}{A_i} \frac{f(x)}{g(x)}
\]
Example 7:

\[ f(x) = \frac{1}{\sqrt{x(1-x)}}, \quad 0 < x < 1 \]

\[ g(x) = \frac{1}{\sqrt{x}} + \frac{1}{\sqrt{1-x}} = \frac{\sqrt{x} + \sqrt{1-x}}{\sqrt{x(1-x)}}, \quad \frac{1}{\sqrt{2}} \leq \frac{f(x)}{g(x)} \leq 1 \]

1. if \( R < 1/2 \) then \( g_1(x) \) else \( g_2(x) \)

2. \( g_1: \ G_1(x) = 2\sqrt{x} = 2R \implies x = R^2 \)

\[ g_2: \ G_2(x) = 2(1 - \sqrt{1-x}) = 2R \implies x = 1 - R^2 \]

3, 4 as previous page
Recall previous formula

\[ \int f(x) \, dx = \int \frac{f(x)}{g(x)} \sum_i g_i(x) \, dx = \sum_i A_i \int \frac{g_i(x) \, dx}{A_i} \frac{f(x)}{g(x)} \]

Now assume split \( f(x) = \sum_i f_i(x) \), with \( f_i(x) < g_i(x) \).

(\text{Brute-force, always possible with } f_i(x) = \left(\frac{g_i(x)}{g(x)}\right) f(x).)

Then

\[ \int f(x) \, dx = \int \sum_i \frac{f_i(x)}{g_i(x)} g_i(x) \, dx = \sum_i A_i \int \frac{g_i(x) \, dx}{A_i} \frac{f_i(x)}{g_i(x)} \]

1. select \( i \) with relative probability \( A_i \)
2. select \( x \) according to \( g_i(x) \)
3. select \( y = R g_i(x) \)
4. while \( y > f_i(x) \) cycle to 1
Discussed today:
- Pick among discrete possibilities
- Integral as an area/volume
- Analytical solution
- Hit-and-miss
- Crude Monte Carlo integration
- Conventional integration
- Stratified sampling
- Importance sampling
- Variable transformations
- Multichannel

To come:
- Random number generators
- Special tricks
- Several dimensions
- Time evolution/ordering
- Combining it: a particle physics example

Material:
- Exercises, week 1 and 2
- these lecture notes (need more? on what?)