To estimate
\[ \langle f \rangle_\pi = \int dx \, \pi(x) f(x) \quad \text{\(\pi(x)\) probability density} \]
we use the algorithm we described before.

We perform \(N\) iterations and compute
\[ \bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(X_i) \quad \text{[SAMPLE MEAN]} \]

\(\bar{f}\) is an estimator of \(\langle f \rangle_\pi\).

Now we wish to define the error. We wish to define the error for any value of \(N\) not necessarily large. We can even take \(N=2,3\).

To define the error we imagine the following procedure:

1. We perform a simulation \((N_{\text{iter}})\) and obtain \(\bar{f}^{(1)}\).
2. We perform a second different simulation, again with \(N_{\text{iter}}\) iterations and obtain a second, different estimate \(\bar{f}^{(2)}\).
3. We perform a third different simulation \(\rightarrow \bar{f}^{(3)}\).
   \[
   \vdots
   \]
4. We perform a \(N_{\text{MC}}\)-th different simulation \(\rightarrow \bar{f}^{(N_{\text{MC}})}\).
Then, we plot the distribution of the results.

In each bin $i$, the height that is equal to the number of times $\bar{f}$ is in the corresponding interval.

The number of times $\bar{f}$ is in $a < \bar{f} \leq b$ is 1 $[\text{# = number}]$.
The number of times $\bar{f}$ is in $b < \bar{f} \leq c$ is 8.
The number of times $\bar{f}$ is in $c < \bar{f} \leq d$ is 17 ...

For $N_{MC} \rightarrow \infty$, the distribution (once normalized so that the area is 1) converges to the probability of obtaining a given value for $\bar{f}$ in a simulation of $N$ iterations.

\[
P(\bar{f})
\]

Two quantities of interest:

\[
\mu = \langle \bar{f} \rangle_{MC} \quad \text{average of } \bar{f} \text{ with respect to } P(\bar{f})
\]

\[
\sigma^2 = \langle (\bar{f} - \mu)^2 \rangle_{MC} = \langle \bar{f}^2 \rangle_{MC} - \mu^2 \quad \text{variance of } \bar{f} \text{ with respect to } P(\bar{f})
\]
NOTE:
Do not confuse $\langle \cdot \rangle_{MC}$ and $\langle \cdot \rangle_{\pi}$
They are two different averages, although it is customary to use the same symbol $\langle \cdot \rangle$ (with no suffix) for both of them.

ERROR: by definition is the standard deviation $\sigma$
It gives the width of the distribution of the values of $\bar{f}$ obtained in the simulations.

BIAS: It is defined as
\[ \text{bias} = \mu - \langle f \rangle_{\pi} \]

COMMENT: $N$ is fixed and arbitrary
Instead we assume $N_{MC} \to \infty$ in the computation of error and bias
(we assume that we are repeating the same simulation of length $N$ an infinite number of times).
SOME BASIC CALCULATIONS.

Consider again the simulation of \(N\) iterations. We wish to compute \(\langle f(X_5) \rangle_{MC}\).

The meaning of the average is the following:

Simul. 1: \(X_1^{(1)} \ldots X_N^{(1)} \rightarrow \) compute \(f(X_5^{(1)})\) and the value of \(f\) using the result at \(i = 5\).

Simul. 2: \(X_1^{(2)} \ldots X_N^{(2)} \rightarrow \) compute \(f(X_5^{(2)})\).

Simul. 3: \(X_1^{(3)} \ldots X_N^{(3)} \rightarrow \) compute \(f(X_5^{(3)})\).

and so on.

\[
\langle f(X_5) \rangle_{MC} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} f(X_5^{(i)}) \quad \text{for } N_{MC} \to \infty
\]

Now \(X_5^{(i)}\) for each \(i\) is extracted with probability \(\pi(X)\). We can thus use the sample-mean theorem:

\[
\langle f(X_5) \rangle_{MC} = \langle f \rangle_{\pi}
\]

There is nothing special about "5":

\[
\langle f(X_i) \rangle_{MC} = \langle f \rangle_{\pi} \quad \text{for any } 1 \leq i \leq N
\]

Analogously:

\[
\langle f(X_i^2) \rangle = \langle f^2 \rangle_{\pi}
\]
Now we wish to compute \( \langle f(X_5) f(X_7) \rangle_{HC} \).

We reason as before.

**Simulation:** we compute \( f(X_5^{(i)}) f(X_7^{(i)}) \), i.e. we use the values of for \( X_5^{(i)} \), the 5th-extracted number and \( X_7^{(i)} \), the 7th-extracted number.

We repeat \( N_{HC} \) times

\[
\langle f(X_5) f(X_7) \rangle_{HC} = \frac{1}{N_{HC}} \sum_{i=1}^{N_{HC}} f(X_5^{(i)}) f(X_7^{(i)})
\]

for \( N_{HC} \to \infty \)

To understand how to apply the sample-mean theorem note that \((X_5, X_7)\) are a pair \((x, y)\) of random numbers distributed according to

\[
P(x, y) = \pi(x) \pi(y)
\]

Here we use the hypothesis of no correlations among different random numbers.

The joint probability is simply the product of the individual probabilities

\[
\frac{1}{N_{HC}} \sum_{i=1}^{N_{HC}} f(X_5^{(i)}) f(X_7^{(i)}) \to \int dx \, dy \, P(x, y) f(x) f(y)
\]

\[
= \int dx \, dy \, \pi(x) \pi(y) f(x) f(y) =\]

\[
= \int dx \pi(x) f(x) \cdot \int dy \pi(y) f(y) = \langle f \rangle^2 \frac{1}{\pi}
\]
SUMMARY

\[ \langle f(x_i) \rangle_{\text{MC}} = \langle f \rangle_{\pi} \]

\[ \langle f(x_i) f(x_j) \rangle_{\text{MC}} = \begin{cases} 
\langle f^2 \rangle_{\pi} & i = j \\
\langle f \rangle_{\pi}^2 & i \neq j
\end{cases} \]

We can now compute \( \langle f \rangle_{\text{MC}} \) and \( \langle f^2 \rangle_{\text{MC}} \).

\[ \langle f \rangle_{\text{MC}} = \frac{1}{N} \sum_{i=1}^{N} \langle f(x_i) \rangle = \frac{f}{N} \sum_{i=1}^{N} \langle f \rangle_{\pi} = \langle f \rangle_{\pi} \]

\[ \langle f^2 \rangle_{\text{MC}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle f(x_i) f(x_j) \rangle_{\text{MC}} \]

\[ = \frac{1}{N^2} \sum_{i=1}^{N} \langle f(x_i)^2 \rangle_{\text{MC}} + \frac{1}{N^2} \sum_{i \neq j} \langle f(x_i) f(x_j) \rangle_{\text{MC}} \]

\[ = \frac{1}{N^2} \sum_{i=1}^{N} \langle f^2 \rangle_{\pi} + \frac{1}{N^2} \sum_{i \neq j} \langle f \rangle_{\pi}^2 \]

\[ = \frac{1}{N^2} N \langle f^2 \rangle_{\pi} + \frac{1}{N^2} N(N-1) \langle f \rangle_{\pi}^2 \]

\[ = \frac{1}{N} \langle f^2 \rangle_{\pi} + \frac{N-1}{N} \langle f \rangle_{\pi}^2 \]

THERE IS NO BIAS.
\[ \sigma^2 = \langle \bar{f}^2 \rangle_{MC} - \langle \bar{f} \rangle_{MC}^2 \]
\[ = \frac{1}{N} \langle f^2 \rangle_\pi + \left(1 - \frac{1}{N} \right) \langle f \rangle_\pi^2 - \langle f \rangle_\pi^2 \]
\[ = \frac{1}{N} \left[ \langle f^2 \rangle_\pi - \langle f \rangle_\pi^2 \right] \text{ cancel} \]
\[ = \frac{1}{N} \text{Var}_\pi f \]

\[ \text{Var}_\pi f = \int dx \pi(x) f(x)^2 - \left( \int dx \pi(x) f(x) \right)^2 \]
\[ = \text{a number independent of } N \]

The error scales as \( \frac{1}{\sqrt{N}} \).

The MC method converges slowly, compared to deterministic integration algorithms, but it is the only method we have to perform integrals in \( D \) dimensions with \( D \) large.
Efficiency of Monte Carlo algorithms

**The basic Monte Carlo algorithm:**
1) Repeat \(N\) times the basic iteration step: generate \(X_n\) uniformly distributed in \([a, b]\) and compute \(g_n = g(X_n)\).
2) An estimate of \(I\) is simply
\[
I \approx \frac{(b-a)}{N} \sum_{n=1}^{N} g_n.
\]

The Monte Carlo algorithm is not efficient in one dimension, since errors vanish as \(1/\sqrt{N}\). Deterministic algorithms have much faster convergence rates. For instance, compute the integral as
\[
I \approx \frac{h}{2} [g(a) + g(b)] + h \sum_{n=1}^{N-1} g(x_n)
\]
where \(x_n\) are equally spaced points such that \(x_0 = a, x_N = b\), and \(h = x_n - x_{n-1}\). The convergence rate is \(1/N^2\) (Simpson’s rule gives \(1/N^4\) convergence).

**Example:**
\[
I = \int_0^1 x^2 \, dx
\]
Using the two methods (\(N\) is the number of points in the trapezoidal rule, and the number of iterations in the MC calculation), we obtain

<table>
<thead>
<tr>
<th>(N)</th>
<th>(I_{\text{trap}})</th>
<th>(I_{\text{MC}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.33335</td>
<td>0.261</td>
</tr>
<tr>
<td>1000</td>
<td>0.3333335</td>
<td>0.327</td>
</tr>
<tr>
<td>10000</td>
<td>0.33333335</td>
<td>0.335</td>
</tr>
</tbody>
</table>

**The main problem of deterministic algorithms.** They become inefficient in large dimensions \(D\). Since they use essentially a regular grid, to obtain reliable results one needs at least 10 points in each direction, hence at least \(10^D\) points. But, if \(D \geq 10\), the number of points is far too large. Moreover, the convergence rate is slower.

**An example:** suppose we wish to compute
\[
I = 3^5 \int_0^1 x^2 y^2 z^2 t^2 u^2 \, dx dy dz dt du = 1
\]
We use the trivial multidimensional generalization of the trapezoidal rule (\(\text{trap}\)) and Monte Carlo. We obtain (\(\Delta_{\text{MC}}\) is the Monte Carlo error, \(\Delta_{\text{trap}} = I_{\text{trap}} - 1\))

<table>
<thead>
<tr>
<th>(n) points</th>
<th>(I_{\text{trap}})</th>
<th>(I_{\text{MC}})</th>
<th>(\Delta_{\text{MC}})</th>
<th>(\Delta_{\text{trap}}/\Delta_{\text{MC}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3^5 = 2.43 \cdot 10^2)</td>
<td>1.802</td>
<td>1.174</td>
<td>0.257</td>
<td>3.1</td>
</tr>
<tr>
<td>(4^5 = 1.02 \cdot 10^3)</td>
<td>1.310</td>
<td>1.001</td>
<td>0.125</td>
<td>2.5</td>
</tr>
<tr>
<td>(5^5 = 3.13 \cdot 10^3)</td>
<td>1.166</td>
<td>1.091</td>
<td>0.072</td>
<td>2.3</td>
</tr>
<tr>
<td>(6^5 = 7.78 \cdot 10^3)</td>
<td>1.104</td>
<td>0.956</td>
<td>0.045</td>
<td>2.3</td>
</tr>
<tr>
<td>(7^5 = 16.8 \cdot 10^4)</td>
<td>1.071</td>
<td>0.942</td>
<td>0.031</td>
<td>2.3</td>
</tr>
<tr>
<td>(10^5 = 1 \cdot 10^5)</td>
<td>1.031</td>
<td>1.004</td>
<td>0.013</td>
<td>2.4</td>
</tr>
</tbody>
</table>