ERRORS AND BIAS

To estimate

$$\langle f \rangle_\pi = \int dx \, \pi(x) f(x)$$

$\pi(x)$ probability density

We use the algorithm we described before

We perform N iterations and compute

$$\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(X_i) \qquad \text{[SAMPLE MEAN]}$$

$\bar{f}$ IS AN <u>ESTIMATOR</u> 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 $N=2,3$.

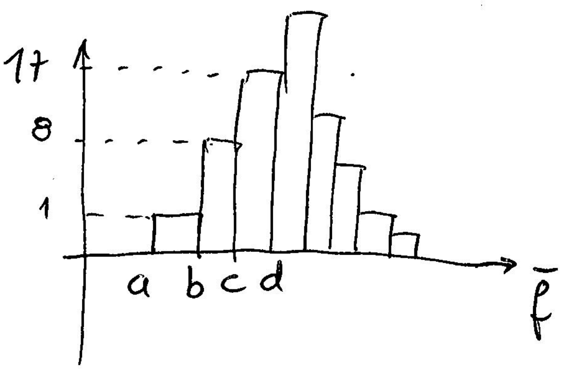
To define the error we imagine the following procedure.

- We perform a simulation (N iter) and obtain $\bar{f}^{(1)}$
- We perform a second different simulation, again with N iterations and obtain a second, different estimate $\bar{f}^{(2)}$
- We perform a third different simulation $\longrightarrow \bar{f}^{(3)}$

⋮

- We perform a N_{MC}-th different simulation $\longrightarrow \bar{f}^{(N_{MC})}$

Then, we plot the distribution of the results



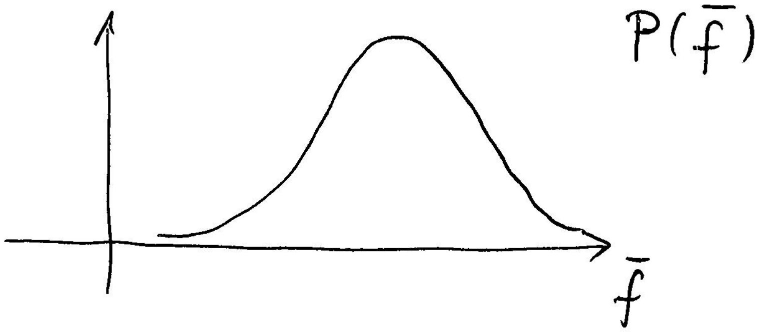
~~In~~ Each bin ~~was~~ has a height that is equal to the the number of times $\bar{f}$ is in the corresponding interval.

The # of times $\bar{f}$ is in $a < \bar{f} \leq b$ is 1 [# = number]

The # of times $\bar{f}$ is in $b < \bar{f} \leq c$ is 8

The # of times $\bar{f}$ is in $c < \bar{f} \leq d$ is 17 ....

For $N_{MC} \to \infty$ the distributions (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



Two quantities of interest

$\mu = \langle \bar{f} \rangle_{MC}$ average of $\bar{f}$ with respect to $P(\bar{f})$

$\sigma^2 = \langle (\bar{f} - \mu)^2 \rangle_{MC} = \langle \bar{f}^2 \rangle_{MC} - \mu^2$ variance of $\bar{f}$ 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 $\langle \ \rangle$ (with no suffix) for both of them.

ERROR : by definition is the standard deviation σ

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 : $X_1^{(1)} \dots X_N^{(1)} \longrightarrow$ compute $f(X_5^{(1)})$
the value of f using
the result at $i=5$

simul. 2 : $X_1^{(2)} \dots X_N^{(2)} \longrightarrow$ compute $f(X_5^{(2)})$

simul. 3 : $X_1^{(3)} \dots X_N^{(3)} \longrightarrow$ compute $f(X_5^{(3)})$

and so on

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

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":

$$\langle f(X_i) \rangle_M = \langle f \rangle_\pi \qquad \text{for any } 1 \le i \le N$$

Analogously

$$\langle f(X_i)^2 \rangle = \langle f^2 \rangle_\pi$$

Now we wish to compute $\langle f(X_5) f(X_7) \rangle_{MC}$

We reason as before

Simul 1: we compute $f(X_5^{(1)}) f(X_7^{(1)})$, i.e. we use the values of for $X_5^{(1)}$, the 5th-extracted number and $X_7^{(1)}$, the 7th-extracted number.

We repeat N_{MC} times

$$\langle f(X_5) f(X_7) \rangle_{MC} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} f(X_5^{(i)}) f(X_7^{(i)})$$

for $N_{MC} \to \infty$

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

$$\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} f(X_5^{(i)}) f(X_7^{(i)}) \longrightarrow \int dx\, dy\, P(x,y) f(x) f(y)$$

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

$$= \int dx\, \pi(x) f(x) \cdot \int dy\, \pi(y) f(y) = \langle f \rangle_\pi^2$$

SUMMARY

$$\langle f(x_i) \rangle_{MC} = \langle f \rangle_\pi$$

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

We can now compute $\langle \bar{f} \rangle_{MC}$ and $\langle \bar{f}^2 \rangle_{MC}$

$$\langle \bar{f} \rangle_{MC} = \frac{1}{N} \left\langle \sum_{i=1}^{N} f(x_i) \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \langle f \rangle_\pi = \langle f \rangle_\pi$$

$\boxed{\text{THERE IS NO BIAS}}$

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

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

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

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

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

$$\sigma^2 = \langle \bar{f}^2 \rangle_{MC} - \langle \bar{f} \rangle_{MC}^2$$

$$= \frac{1}{N} \langle f^2 \rangle_\pi + \underbrace{\left(1 - \frac{1}{N}\right) \langle f \rangle_\pi^2 - \langle f \rangle_\pi^2}_{\text{cancel}}$$

$$= \frac{1}{N} \Big[\underbrace{\langle f^2 \rangle_\pi - \langle f \rangle_\pi^2}_{\text{variance of } f(x) \text{ with respect to } \pi(x)} \Big]$$

$$= \frac{1}{N} \mathrm{Var}_\pi f$$

$$\mathrm{Var}_\pi f = \int dx\, \pi(x) f(x)^2 - \left[\int dx\, \pi(x) f(x) \right]^2$$

$$= \text{a number independent of } N$$

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

$N = 100$	$I_{\text{trap}} = 0.33335$	$I_{\text{MC}} = 0.261$
$N = 1000$	$I_{\text{trap}} = 0.3333335$	$I_{\text{MC}} = 0.327$
$N = 10000$	$I_{\text{trap}} = 0.333333335$	$I_{\text{MC}} = 0.335$

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 $D \gtrsim 10$, the number of points is far too large. Moreover, the convergence rate is slower.

An example: suppose we wish to compute

$$I = 3^5 \int_0^1 x^2 y^2 z^2 t^2 u^2 \, dxdydzdtdu = 1$$

We use the trivial multidimensional generalization of the trapezoidal rule (trap) and Monte Carlo. We obtain (Δ_{MC} is the Monte Carlo error, $\Delta_{\text{trap}} = I_{\text{trap}} - 1$)

n.points	I_{trap}	I_{MC}	Δ_{MC}	$\Delta_{\text{trap}}/\Delta_{MC}$
$3^5 = 2.43 \cdot 10^2$	1.802	1.174	0.257	3.1
$4^5 = 1.02 \cdot 10^3$	1.310	1.001	0.125	2.5
$5^5 = 3.13 \cdot 10^3$	1.166	1.091	0.072	2.3
$6^5 = 7.78 \cdot 10^3$	1.104	0.956	0.045	2.3
$7^5 = 16.8 \cdot 10^4$	1.071	0.942	0.031	2.3
$10^5 = 1 \cdot 10^5$	1.031	1.004	0.013	2.4