# Monte Carlo Methods in Statistical Physics: I. Error analysis for uncorrelated data

Andrea Pelissetto

Andrea.Pelissetto@roma1.infn.it

Dipartimento di Fisica
Università di Roma “La Sapienza”

Thermodynamics is a macroscopic theory of thermal phenomena, like the melting of a solid, the freezing of a liquid, the heating of water, …, and it is relevant for a wide range of phenomena in physics, chemistry, biology, ….
Statistical mechanics provides the microscopic foundations for thermodynamics. The basic ingredient is the Gibbs measure.

**Example: A gas**. Consider a gas (fluid) of $N$ molecules in a box of volume $V$. The configurations of the gas are specified by the set of positions $\mathbf{r_1}, \mathbf{r_2}, \ldots$ (each vector defines a position inside the box) and of momenta $\mathbf{p_1}, \mathbf{p_2}, \ldots$ The configuration space is the phase space available to the system. On this space we define the Gibbs measure

$$\exp\left[-\frac{1}{k_B T} H(\{\mathbf{r}_i\}, \{\mathbf{p}_i\})\right] \prod_i [d\mathbf{r}_i d\mathbf{p}_i]$$

where $H$ is the Hamiltonian, $T$ the absolute temperature, and $k_B$ the Boltzmann constant.

For the ideal gas we have simply

$$H = \sum_i \frac{p_i^2}{2m_i}.$$

For a real gas we should also consider a pair potential (for instance a Lennard-Jones potential).

When considering a gas, a fluid, …, we are not interested in the detailed behavior of each single molecule, but rather on average properties of the system, which, as long as the system is macroscopic, are time-independent (the system is in equilibrium) and independent of the microscopic chaotic motion of the molecules.
These average quantities can be computed as averages over the Gibbs probability distribution.

For instance the average energy at constant volume can be computed as

$$E = \frac{1}{Z} \int [dp][dr] H e^{-\beta H},$$

where the integral is over the whole configuration space (all values of the momenta and all positions inside the box), $\beta = 1/(k_B T)$ and $Z$ is the partition function.

From $E$ one can compute the constant-volume specific heat and reobtain from statistical mechanics all thermodynamic quantities.

The integrals over the Gibbs measure are integrals over a large number of variables (of the order of the number of molecules, atoms, spins present in the system, which is of the order of the Avogadro number for a macroscopic system). Except in a very few cases, they cannot be performed analytically.

Thus, in order to predict the thermodynamic behavior of nontrivial systems, we need to use numerical techniques. There are essentially two methods:

**MONTE CARLO METHOD**. It is very general, easy to implement, in several instances quite fast. Uses a nonphysical dynamics to compute the average values over the Gibbs distribution.

**MOLECULAR DYNAMICS**. It is appropriate for molecular systems. It is not competitive with the Monte Carlo method for the understanding of the static behavior (but in some instances, it may be useful to combine molecular dynamics and Monte Carlo simulations). It is however the only available method to study time-dependent (dynamic) phenomena, like relaxation, equilibration, …, since the classical dynamics correctly describes the physical time behavior of molecular systems (as long as electrons are not involved).

The Monte Carlo method is essentially a technique to compute high-dimensional integrals via the generation of elements of a given configuration space with a given probability distribution.

We begin by considering the simplest possible case:

The configuration space is the interval $[a,b]$ of the real line;
The Gibbs distribution is a positive, normalized function $f(x)$.

**The algorithmic problem**:

> We wish to generate numbers $X_n \in [a,b]$, with probability density $f(x)$:

$$\text{Prob}(x \leq X_n \leq x+dx) = \begin{cases} f(x)dx & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

Let us **assume** that we know how to generate independent data $X_n$. Can we use them to compute averages with respect to $f(x)$?

The basic result.

If $\{X_n\}$, $n : 1, \ldots N$, are distributed with probability density $f(x)$, then the sample mean

$$\overline{g(X)} = \frac{1}{N} \sum_n g(X_n)$$

converges, as $N \to \infty$, to

$$\langle g(x) \rangle_f = \int g(x) f(x) dx \equiv G$$

Rate of convergence.

Let us compute the average of the square deviations

$$\Delta^2 = \left\langle \left[ G - \overline{g(X)} \right]^2 \right\rangle_f .$$

Expanding the square we obtain

$$\Delta^2 = \left\langle G^2 - 2G\overline{g(X)} + \frac{1}{N^2} \sum_{nm} g(X_n) g(X_m) \right\rangle_f .$$

Now, if $n \neq m$ we have

$$\langle g(X_n) g(X_m) \rangle_f = \langle g(X_n) \rangle_f \langle g(X_m) \rangle_f = G^2$$

since $X_n$ and $X_m$ are independent variables. If $n = m$

$$\langle g(X_n)^2 \rangle_f = \int g(x)^2 f(x) dx = \langle g^2 \rangle_f$$

Collecting everythig together we have

$$\Delta^2 = G^2 - 2G \times G + \frac{1}{N^2}\left[N(N-1)G^2 + N\left\langle g^2 \right\rangle_f\right]$$

$$\Delta^2 = G^2 - 2G^2 + G^2 - G^2/N + \left\langle g^2 \right\rangle_f /N = \frac{1}{N}\mathrm{Var}\,(g)$$

where the variance of $g$ is

$$\mathrm{Var}\,(g) = \left\langle g^2 \right\rangle_f - G^2 = \int (g(x) - G)^2 f(x) dx.$$

Thus, **deviations scale as $1/\sqrt{N}$. This is the basic feature of any Monte Carlo calculation**.

## imple integration method

Suppose we wish to compute

$$I = \int_a^b g(x)dx,$$

where $g(x)$ is an arbitrary function.

**A simple 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 result follows immediately from the previous results, if we take $f(x) = 1/(b-a)$, i.e. the uniform distribution in $[a,b]$.

This Monte Carlo algorithm is terrible, since corrections 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:**

$$I = \int_0^1 x^2 dx$$

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

## 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.

| n.points | $I_{\text{trap}}$ | $I_{MC}$ | $\Delta_{MC}$ |
|---|---|---|---|
| $3^5$ | 1.802 | 1.174 | 0.257 |
| $4^5$ | 1.310 | 1.001 | 0.125 |
| $5^5$ | 1.166 | 1.091 | 0.072 |
| $6^5$ | 1.104 | 0.956 | 0.045 |
| $7^5$ | 1.071 | 0.942 | 0.031 |
| $10^5$ | 1.031 | 1.004 | 0.013 |

Here $\Delta_{MC}$ is the expected standard deviation on the MC result.

**Rate of convergence**: if the number of points is $p^5$, the deterministic algorithm converges as $1/p^2$. Monte Carlo converges as $1/\sqrt{p^5} = 1/p^{2.5}$. The Monte Carlo converges faster.

The convergence rate of the Monte Carlo calculation is

$$\Delta = \frac{1}{\sqrt{N}} \left[ \int_a^b [(b-a)g(x) - I]^2 \frac{dx}{b-a} \right]^{1/2}$$

We wish $\Delta$ to be as small as possible (of course it vanishes for $g(x) = I/(b-a)$).
It is large if $g(x)$ varies significantly in the interval $[a,b]$.

Importance sampling provides a way out.

We have computed the integral $I$ by using the flat distribution.
But we could use any probability distribution $f(x)$.

$$I = \int_a^b g(x)dx = \int_a^b \frac{g(x)}{f(x)} f(x)dx = \langle g/f \rangle_f .$$

Hence,if $\{X_n\}$, $n : 1, \ldots N$ are distributed with probability density $f(x)$, we have

$$\frac{1}{N}\sum_{n=1}^{N}\frac{g(X_n)}{f(X_n)} \to I$$

for $N \to \infty$.

Convergence rate:

$$\Delta = \frac{1}{\sqrt{N}}\left[\int_a^b \left(\frac{g(x)}{f(x)} - I\right)^2 f(x)dx\right]^2$$

We can improve the convergence rate by choosing $f(x)$ in such a way that $g(x)/f(x)$ is flatter than $g(x)$.
**NOTE**: $f(x)$ is a probability density, hence positive. Thus, the method can work only if $g(x)$ is (almost) always positive or negative. If it changes sign, importance sampling does not work (**the so-called sign problem**).

## rcise:

Suppose $g(x) = e^{-x^2}$. We wish to compute

$$I = \int_0^2 g(x)dx$$

Verify the following results:
a) If $f(x) = 1/2$, the convergence rate is $0.69/\sqrt{N}$;
b) If $f(x) = (3-x)/4$ (write the algorithm for such a distribution), the convergence rate is $0.45/\sqrt{N}$;
c) If $f(x) = Ke^{-x}$ the convergence rate is $0.27/\sqrt{N}$;
d) If $f(x) = Ke^{-3x/2}$ the convergence rate is $0.19/\sqrt{N}$.
The last method is 13 times faster than the first, but it is more CPU time-consuming to generate numbers with such an $f(x)$. Compare the convergence in CPU time!

## imating the variance

In MC calculations the convergence rate is controlled by the variance of $g(x)$.
If

$$G = \langle g \rangle_f = \int_a^b g(x) f(x) dx$$

the **sample mean**

$$\overline{g(X)} = \frac{1}{N} \sum_n g(X_n)$$

is an **estimator** of $G$.

We wish to find an estimator of

$$\mathrm{Var}\,(g) = \left\langle (g-G)^2 \right\rangle_f = \left\langle g^2 \right\rangle_f - G^2 = \int (g(x)-G)^2 f(x) dx.$$

It is natural to consider the following quantity

$$V = \frac{1}{N}\sum_n g(X_n)^2 - \left[\overline{g(X)}\right]^2$$

Of course, on average

$$\left\langle \frac{1}{N}\sum_n g(X_n)^2 \right\rangle_f = \frac{1}{N}\sum_n \left\langle g(X_n)^2 \right\rangle_f = \left\langle g^2 \right\rangle_f .$$

However, for finite $N$ it is not true that

$$\left\langle \left[\overline{g(X)}\right]^2 \right\rangle_f = G^2$$

Let's compute it explicitly. Using the definition of sample mean we obtain

$$\left\langle \left[\overline{g(X)}\right]^2 \right\rangle_f = \frac{1}{N^2}\sum_{nm}\langle g(X_n)g(X_m)\rangle_f$$

Now we should distinguish two cases: if $n = m$ we have $\langle g(X_n)g(X_m)\rangle_f = \langle g^2\rangle_f$ (this occurs $N$ times); if $n \neq m$ (it occurs $N^2 - N$ times) we have $\langle g(X_n)g(X_m)\rangle_f = G^2$. Hence we obtain

$$\frac{1}{N^2}\left[(N^2-N)G^2 + N\langle g^2\rangle_f\right] = \frac{N-1}{N}G^2 + \frac{1}{N}\langle g^2\rangle_f.$$

Therefore

$$V = \langle g^2\rangle_f - \frac{N-1}{N}G^2 - \frac{1}{N}\langle g^2\rangle_f = \frac{N-1}{N}(\langle g^2\rangle_f - G^2)$$

1) $V$ is a biased estimator of the variance: it corresponds to the variance only for $N \to \infty$.
2) The unbiased estimator of the variance is

$$\frac{N}{N-1}V = \frac{N}{N-1}\left(\frac{1}{N}\sum_n g(X_n)^2 - \left[\overline{g(X)}\right]^2\right).$$

The correction factor is irrelevant if $N$ is large.
3) The error on the sample mean can thus be estimated as

$$\sigma_{\overline{g}} = \sqrt{\frac{1}{N-1}\left(\frac{1}{N}\sum_n g(X_n)^2 - \left[\overline{g(X)}\right]^2\right)}$$

# rcise

Compute $M = \left\langle \left( \overline{g(X)} - G \right)^3 \right\rangle_f$ .

1) Prove first that

$$\left\langle \left( \overline{g(X)} \right)^3 \right\rangle_f = G^3 + \frac{3G}{N} \mathrm{Var}\, g + \frac{1}{N^2} \langle (g-G)^3 \rangle_f$$

2) It follows

$$M = \frac{1}{N^2} \langle (g-G)^3 \rangle_f .$$

A general formula can be obtained by considering

$$M(\alpha) = \left\langle \exp\left[\alpha\left(\overline{g(X)} - G\right)\right]\right\rangle_f .$$

Prove that this quantity can be written as

$$M(\alpha) = \left[\int_a^b f(x)dx\, e^{\alpha(g(x)-G)/N}\right]^N$$

Expanding in powers of $\alpha$ and comparing the coefficients of $\alpha^n$ we obtain all previous relations.
Use this expression to show that:

1) $\left\langle \left(\overline{g(X)} - G\right)^4 \right\rangle_f$ is of order $1/N^2$.

2) $\left\langle \left(\overline{g(X)} - G\right)^5 \right\rangle_f$ and $\left\langle \left(\overline{g(X)} - G\right)^6 \right\rangle_f$ are of order $1/N^3$.

Let us consider again a function $g(x)$, its average $G = \langle g \rangle_f$ over a distribution $f$. We are interested in computing

$$K = H(G)$$

where $H(x)$ is some function.

A MC determination of $K$ would estimate it as

$$K_{\text{est}} = H\left(\overline{g(X)}\right)$$

We wish to show that $K_{\text{est}}$ is a **biased** estimator of $K$.

Let us compute

$$\langle K_{\text{est}} \rangle_f = \left\langle H\left(\overline{g(X)}\right) \right\rangle_f .$$

We assume that $N$ is large so that the sample mean is close to the correct value.

Write

$$\overline{g(X)} = G + \Delta G, \qquad \Delta G = \overline{g(X)} - G$$

and assume that $\Delta G$ is small, so that we can Taylor expand all quantities. Hence

$$\langle K_{\text{est}} \rangle = H(G) + H'(G) \langle \Delta G \rangle_f + \frac{1}{2} H''(G) \left\langle (\Delta G)^2 \right\rangle_f + \dots$$

Now, we have $\langle \Delta G \rangle_f = 0$ and $\left\langle (\Delta G)^2 \right\rangle_f = (\text{Var}\, g)/N$. Therefore,

$$\langle K_{\text{est}} \rangle = H(G) + \frac{1}{2N} H''(G)(\text{Var}\, g) + \dots$$

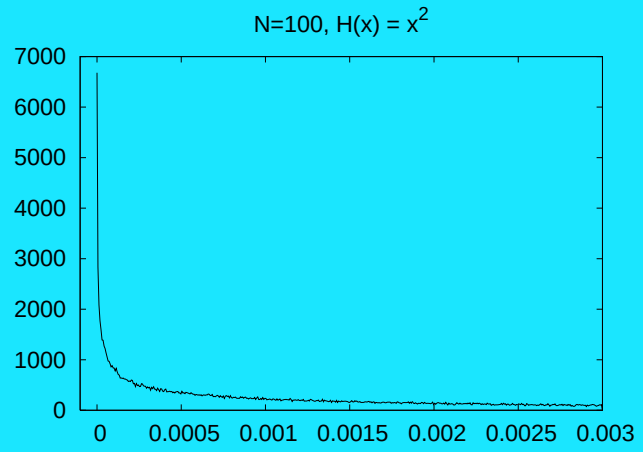
For finite $N$ $\langle K_{\text{est}} \rangle$ differs from $H(G)$: the corrections, of order $1/N$, are called the **bias**.
For instance: if $H''(G) > 0$, the previous formula indicates that it is more probable to obtain an estimate $K_{\text{est}}$ that is larger than $H(G)$ than an estimate that is smaller than $H(G)$.

# example

Let $f(x) = 1$ in $[0,1]$, $g(x) = 2x - 1$, $H(x) = x^2$,

$$G = \langle g \rangle_f = \int_0^1 g(x)dx$$

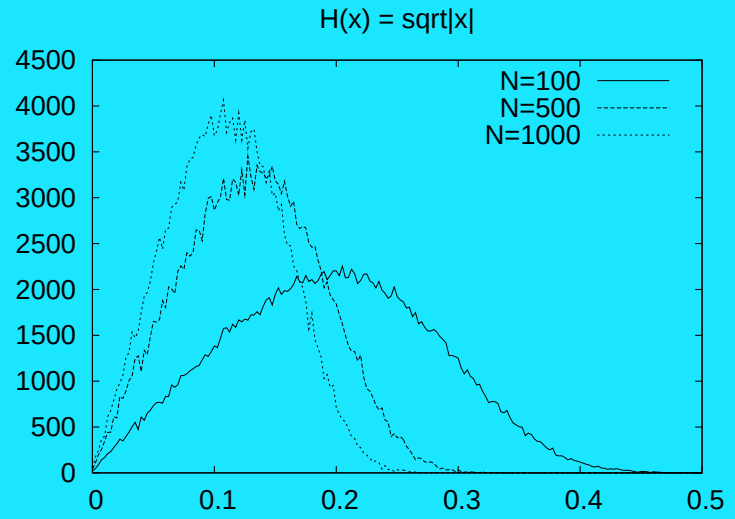
Of course $G = 0$, $H(G) = 0$.
We now fix $N = 100$ and perform several MC determination of $K_{\text{est}}$. The distribution of the results is reported in the figure




The distribution has support on $x \geq 0$, is peaked at $x = 0$, and has a very long tail. The mean of $x$ over this distribution is $0.00333$

The mean is exactly the bias.
Indeed: $1/(2N)H''(G)\text{Var}\, g = 1/300$, since $H''(G) = 2$,
$\text{Var}\, g = \int dx\,(2x-1)^2 = 1/3$.

**WARNING: the bias decreases as $1/N$ if $H(x)$ is analytic at $x = G$. Otherwise, the bias can decrease slower.**




Here $H(x) = \sqrt{|x|}$. The bias decreases as $N^{-1/4}$.

**COMMENT**: In MC calculations, biased estimates are usually obtained when considering functions of several different averages. For instance, one often wishes to estimate

$$\frac{\langle g \rangle_f}{\langle h \rangle_f}$$

The estimator

$$\frac{\overline{g(X)}}{\overline{h(X)}}$$

is biased.

Relevance of the bias:

The bias is relevant in short simulations. In this case the bias (of order $1/N$) may be comparable with the statistical error (of order $1/\sqrt{N}$). It is then crucial to take it into account [this is the case of quenched random systems]: a bias correction must be performed.

If $N$ is large, the bias can be neglected: statistical fluctuations are much more important.

The error on the results can be computed analogously. We consider the average (square) deviation:

$$\Delta^2 = \left\langle (K_{\text{est}} - H(G))^2 \right\rangle_f .$$

Expanding at first order in $\Delta G$ we obtain

$$\Delta^2 = \left\langle (H'(G)\Delta G)^2 \right\rangle_f = \frac{[H'(G)]^2}{N} \text{Var}\, g$$

This is nothing but the standard **error propagation formula**.

# rcise

Let $f(x) = 1$ in $[0,1]$, $g(x) = x^4$, $H(x) = 1/x^4$.

1) Suppose we estimate $K$ as $1/(\overline{g(X)})^4$. Show that the relative bias (bias/$G$) decreases as $17.8/N$, while the relative error decreases as $5.3/\sqrt{N}$. The bias is larger than the error for $N \lesssim 12$.

2) Show that a better estimator for short simulations is

$$\frac{1}{(\overline{g(X)})^4} - \frac{10}{N-1}\frac{\overline{(g(X)^2)} - \left(\overline{g(X)}\right)^2}{\left(\overline{g(X)}\right)^6}$$

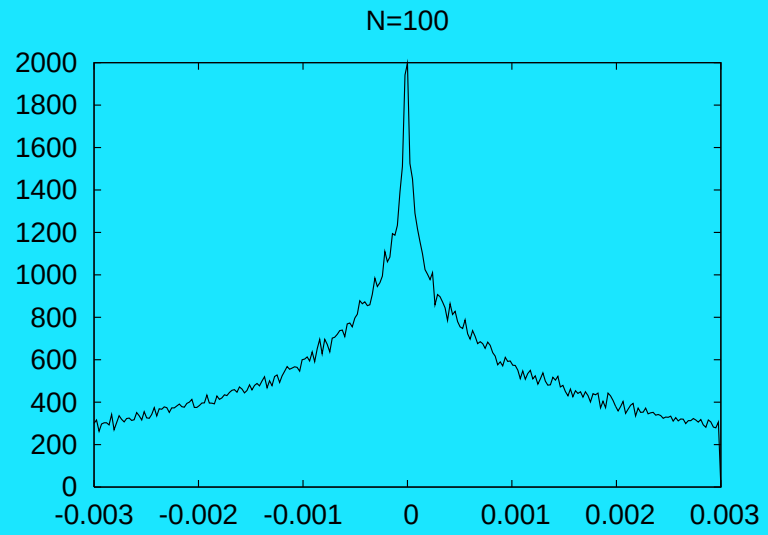
Prove it has no bias to order $1/N$.

# rcise

Consider the case $H(x) = x^2$. Show that an unbiased estimator is (this is a particular case of the jackknife estimator)

$$\frac{4}{N^2}\left[\sum_{n=1}^{N/2} g(X_n)\right]\left[\sum_{n=N/2+1}^{N} g(X_n)\right]$$

Compute the error on this estimator and compare it with that on the naive estimator [Ans.: no difference].




Distribution of the MC results for the choice (as before): $f(x) = 1$ in $[0,1]$, $g(x) = 2x - 1$, $N = 100$. the distribution is clearly symmetric (no bias).

A common error.

In order to estimate $K$, suppose you perform a run of $N$ iterations. The sample mean of $g(x)$ gives an estimate $G_1$ of $G$, and $K_{\text{est},1} = H(G_1)$.
To improve the results, a second run (again of $N$ iterations) is performed. The sample mean of $g(x)$ gives an estimate $G_2$ of $G$, and $K_{\text{est},2} = H(G_2)$.

**What is the final estimate of $K$?**

Because of the bias, it is wrong to quote $K = (K_{\text{est},1} + K_{\text{est},2})/2$. Instead, compute $G = (G_1 + G_2)/2$. The final estimate of $K$ is $H(G)$.

Consider a probability distribution $f(x)$ in $[a,b]$ and two functions $g(x)$, $h(x)$. We define

$$G = \langle g \rangle_f = \int_a^b g(x) f(x) dx \qquad H = \langle h \rangle_f = \int_a^b h(x) f(x) dx$$

and

$$K = \frac{G}{H}$$

The quantity $K$ can be estimated by considering

$$K_{\text{est}} = \frac{\overline{g(X)}}{\overline{h(X)}}.$$

This is a biased estimator of $K$.

**How do we compute the error on $K_{\text{est}}$?**

We already discussed the error on the sample mean. Hence the error on $\overline{g(X)}$ and $\overline{h(X)}$ is simply given by

$$\sigma_{\overline{g}} = \frac{1}{N}\text{Var}\, g \qquad\qquad \sigma_{\overline{h}} = \frac{1}{N}\text{Var}\, h$$

To obtain the final error, we might use expressions that use only $\sigma_{\overline{g}}$ and $\sigma_{\overline{h}}$.

If $K = K(\langle g\rangle_f, \langle h\rangle_f)$ two commonly used expressions are the following. Define

$$K_1(x,y) = \frac{\partial K(x,y)}{\partial x}, \qquad\qquad K_2(x,y) = \frac{\partial K(x,y)}{\partial y},$$

1) **Worst-error formula**:

$$\sigma_K = |K_1(\overline{g},\overline{h})|\sigma_{\overline{g}} + |K_2(\overline{g},\overline{h})|\sigma_{\overline{h}}$$

2) **Independent-error formula**:

$$\sigma_K = \sqrt{|K_1(\overline{g},\overline{h})|^2\sigma_{\overline{g}}^2 + |K_2(\overline{g},\overline{h})|^2\sigma_{\overline{h}}^2}$$

**Comments**:
Formula 1 always overestimates the error. It is a (usually very) conservative expression.
Formula 2 is unreliable: it may overestimate as well as underestimate the true error.

Let us now compute the correct error. The squared deviation is given by

$$\Delta^2 = \left\langle \left( \frac{\overline{g(X)}}{\overline{h(X)}} - \frac{G}{H} \right)^2 \right\rangle_f = \left\langle \left( \frac{G+\Delta G}{H+\Delta H} - \frac{G}{H} \right)^2 \right\rangle_f$$

Expanding in $\Delta G$ and $\Delta H$ to first order, we obtain

$$\left\langle \left( \frac{\Delta G}{H} - \frac{G\Delta H}{H^2} \right)^2 \right\rangle_f = \left( \frac{G}{H} \right)^2 \left[ \frac{\langle \Delta G^2 \rangle_f}{G^2} + \frac{\langle \Delta H^2 \rangle_f}{H^2} - \frac{2\langle \Delta G \Delta H \rangle_f}{GH} \right]$$

The first two terms in brackets are $\sigma_{\overline{g}}^2/G^2$ and $\sigma_{\overline{h}}^2/H^2$

Let us investigate the third term

$$\langle \Delta G \Delta H \rangle_f = -GH + \frac{1}{N^2} \sum_{nm} \langle g(X_n) h(X_m) \rangle_f =$$

$$-GH + \frac{1}{N^2} [(N^2 - N) GH + N \langle g(X) h(X) \rangle_f] = \frac{1}{N} \mathrm{Cov}(g,h)$$

where the Covariance of $g$ and $h$ is defined as

$$\mathrm{Cov}(g,h) = \langle g(X) h(X) \rangle_f - GH = \int_a^b (g(x) - G)(h(x) - H) f(x) dx$$

$$= \int_a^b (g(x) h(x) - GH) f(x) dx$$

The different expressions for the variance:

1) **correct**

$$\Delta^2 = \frac{1}{N}\left(\frac{G}{H}\right)^2\left(\frac{\mathrm{Var}\, g}{G^2} + \frac{\mathrm{Var}\, h}{H^2} - \frac{2\mathrm{Cov}(g,h)}{GH}\right)$$

2) **independent-error formula**

$$\Delta^2 = \frac{1}{N}\left(\frac{G}{H}\right)^2\left(\frac{\mathrm{Var}\, g}{G^2} + \frac{\mathrm{Var}\, h}{H^2}\right)$$

3) **worst-error formula**

$$\Delta^2 = \frac{1}{N}\left(\frac{G}{H}\right)^2\left(\frac{\sqrt{\mathrm{Var}\, g}}{G} + \frac{\sqrt{\mathrm{Var}\, h}}{H}\right)^2$$

The independent-error formula amounts to neglecting the covariance. If $h$ and $g$ are strongly correlated [$\mathrm{Cov}(g,h) > 0$] the formula overestimates the error. If they are anticorrelated, it underestimates the error.
The worst-error formula provides an upper bound on the correct error.

# per bound on the error

Define

$$\delta g(x) = g(x) - G, \qquad \delta h(x) = h(x) - H.$$

Obviously we have

$$[\delta g(x)\delta h(y) - \delta g(y)\delta h(x)]^2 \geq 0 \Rightarrow \delta g(x)^2\delta h(y)^2 + \delta h(x)^2\delta g(y)^2 \geq 2\delta g(x)\delta g(y)\delta h(x)\delta h(y)$$

Multiply by $f(x)f(y)$ and integrate over $x$ and $y$:

$$2\mathrm{Var}\, g\, \mathrm{Var}\, h \geq 2\mathrm{Cov}(g,h)^2$$

Hence (this relation is known as **triangular inequality**)

$$|\mathrm{Cov}(g,h)| \leq \sqrt{\mathrm{Var}\, g\, \mathrm{Var}\, h}$$

Using this relation it is immediate to prove that the worst-error formula provides an upper bound on the error.

## estimator for the error

Exercise: Prove that an unbiased estimator of the covariance is

$$\frac{N}{N-1}\left(\frac{1}{N}\sum_n g(X_n)h(X_n) - \overline{g(X)}\,\overline{h(X)}\right)$$

This formula is the analogous of that obtained for the variance.

For the computation of the error it is however much more convenient to rewrite the error formula as

$$\Delta^2 = \frac{1}{N}\left(\frac{G}{H}\right)^2 \mathrm{Var}\left(\frac{g}{G} - \frac{h}{H}\right).$$

## rcise

Take $f(x) = 1$ in $[0,1]$, $g(x) = x^4$, $h(x) = x^3$, $K = G/H$.
Show that the correct error on the estimate of $K$ scales as $0.45/\sqrt{N}$. The independent-error formula would give an error which scales as $1.2/\sqrt{N}$ (it overestimates the error by a factor of 2.7) while the worst-error formula would give an error which scales as $2.0/\sqrt{N}$ (error overestimated by more than 4).

Take $f(x) = 1$ in $[0,1]$, $g(x) = x^4$, $h(x) = (1-x)^4$, $K = G/H$.
Show that the correct error on the estimate of $K$ scales as $1.53/\sqrt{N}$. The independent-error formula would give an error which scales as $1.37/\sqrt{N}$ (now it slightly underestimates the error) while the worst-error formula would give an error which scales as $2.7/\sqrt{N}$.

The jackknife method is very powerful.
It enjoys three very good properties:

1. It takes care of the bias: the jackknife estimate has a bias of order $1/N^2$.
2. Provides a correct estimate of the errors: correlations are taken into account.
3. It is numerically stable and accurate for small values of $N$ (for instante, it takes into account the nonlinearities of the function $H(G)$).

The basic ingredient in the jackknife method is the partial jackknife average.

In our setting, suppose we generate $N$ numbers $X_n$ with probability distribution $f(x)$. Let $g(x)$ and $h(x)$ be two functions and $g_n = g(X_n)$, $h_n = h(X_n)$.

In order to apply the jackknife method, we first choose a number $n$ which exactly divides $N$.
$n$ should not be too small, neither too large (for practical reasons): in practice $n \approx 100$ is a good choice.
If $N$ is small, we can take $n = N$.

If $M = N/n$, we define

$$\hat{g}_1 = \frac{1}{M}(g_1 + \ldots g_M)$$
$$\hat{g}_2 = \frac{1}{M}(g_{M+1} + \ldots g_{2M})$$
$$\ldots\ldots$$
$$\hat{g}_n = \frac{1}{M}(g_{N-M+1} + \ldots g_N)$$

and the partial jackknife averages:

$$\hat{g}_1^J = \frac{1}{n-1}(\hat{g}_2 + \hat{g}_3 + \ldots \hat{g}_n) = \frac{1}{n-1}\sum_{i\neq 1}\hat{g}_i$$
$$\hat{g}_2^J = \frac{1}{n-1}(\hat{g}_1 + \hat{g}_3 + \ldots \hat{g}_n) = \frac{1}{n-1}\sum_{i\neq 2}\hat{g}_i$$
$$\ldots\ldots$$
$$\hat{g}_n^J = \frac{1}{n-1}(\hat{g}_1 + \hat{g}_2 + \ldots \hat{g}_{n-1}) = \frac{1}{n-1}\sum_{i\neq n}\hat{g}_i$$

Analogously, we can define $\hat{h}_i^J$. Of course

$$\langle \hat{g}_i^J \rangle_f = G \qquad\qquad \langle \hat{h}_i^J \rangle_f = H$$

Using the partial jackknife averages we can compute the jackknife estimators of any quantity we are interested in. For instance, suppose we wish to compute

$$K_1 = H(G), \qquad\qquad K_2 = \frac{G}{H}$$

The jackknife partial estimators are the following:

$$K_{1,i}^J = H(\hat{g}_i^J), \qquad\qquad K_{2,i}^J = \frac{\hat{g}_i^J}{\hat{h}_i^J}$$

Let moreover be $K_{1,\mathrm{est}}$ and $K_{2,\mathrm{est}}$ the standard estimates obtained by using all data.

The **jackknife estimate** is

$$K^{JK} = nK_{\text{est}} - \frac{n-1}{n}\sum_{i=1}^{n} K_i^J$$

The bias on this estimator is of order $1/N^2$ (same formula for both $K_1$ and $K_2$, or any other interesting observable).

An estimate of the error $\sigma_K$ on $K^{JK}$ is

$$\sigma_K^2 = (n-1)\left[\frac{1}{n}\sum_i (K_i^J - K^{JK})^2\right]$$

This error takes into account correlations.

# nputation of the bias on $K_{JK}$

As we discussed earlier, given a sample of $m$ numbers, we have

$$\langle H(\overline{g})\rangle_f = H(G) + a/m + b/m^2 + \dots$$

where $a/m + b/m^2 + \dots$ is the bias.
For all $i$, $K_i^J$ corresponds to an estimate obtained by using a sample of $N-M$ numbers, while $K_{\text{est}}$ is obtained using the complete sample of $N$ numbers.
We have (we use $N-M = N(1-1/n) = N(n-1)/n$)

$$\begin{aligned}
\langle K_{JK}\rangle_f &= n\left(H(G) + \frac{a}{N} + \frac{b}{N^2} + \dots\right) - \frac{n-1}{n} n\left(H(G) + \frac{a}{N-M} + \frac{b}{(N-M)^2} + \dots\right) \\
&= H(G) + a\left(\frac{n}{N} - (n-1)\frac{n}{N(n-1)}\right) + \dots \\
&= H(G) - \frac{n}{n-1}\frac{b}{N^2} + \dots
\end{aligned}$$

**NOTE**: The unbiased estimator we discussed before for $H(G) = G^2$ is the jackknife estimator with $n = 2$. Prove it.

# eck of the error formula for $K_{JK}$

Assume $K = H(G)$. Derive the following relations:

$$\left\langle [K_i^J - H(G)][K_j^J - H(G)] \right\rangle_f = \frac{a^2}{M} \frac{n-2}{(n-1)^2} \mathrm{Var}\, g \qquad i \neq j$$

$$\left\langle [K_i^J - H(G)]^2 \right\rangle_f = \frac{a^2}{M} \frac{1}{(n-1)} \mathrm{Var}\, g$$

$$\left\langle [K_i^J - H(G)][K_{\mathrm{est}} - H(G)] \right\rangle_f = \frac{a^2}{M} \frac{1}{n} \mathrm{Var}\, g$$

$$\left\langle [K_{\mathrm{est}} - H(G)]^2 \right\rangle_f = \frac{a^2}{M} \frac{1}{n} \mathrm{Var}\, g$$

where $a = H'(G)$.
Using these expressions prove that

$$\Delta^2 = \left\langle [K^{JK} - H(G)]^2 \right\rangle_f = \frac{a^2}{M} \frac{1}{n} \mathrm{Var}\, g = \frac{a^2}{N} \mathrm{Var}\, g.$$

$$\left\langle \frac{1}{n} \sum_i (K_i^J - K^{JK})^2 \right\rangle_f = \frac{a^2}{M} \frac{1}{n(n-1)} \mathrm{Var}\, g = \frac{\Delta^2}{n-1}$$

To understand the origin of the factor $(n-1)$ that appears in the error formula, it is useful to consider the simple case $H(G) = G$, $n = N$.
Show that:

$$\frac{1}{n}\sum_i (K_i^J)^2 = \frac{n(n-2)}{(n-1)^2}\left(\frac{1}{n}\sum_i \hat{g}_i\right)^2 + \frac{1}{n(n-1)^2}\sum_i \hat{g}_i^2 \qquad K^{JK} = \frac{1}{n}\sum_i K_i^J = \frac{1}{n}\sum_i \hat{g}_i.$$

It follows

$$\frac{1}{n}\sum_i (K_i^J - K^{JK})^2 = \frac{1}{(n-1)^2}\left[\frac{1}{n}\sum_i \hat{g}_i^2 - \left(\frac{1}{n}\sum_i \hat{g}_i\right)^2\right]$$

The variance is smaller by a factor $(n-1)^2$. Hence the prefactor $1/(n-1)$ which usually appears in the error formula is replaced by $(n-1)^2 \times 1/(n-1) = (n-1)$.