Course 12/2

Errors in correlated sampling

- Errors in uncorrelated sampling
- Correlation function
- Correlation time
- Expression for error
- Error from « time » averages
- Blocking analysis

Errors in uncorrelated sampling

$$I = \int \underbrace{d\vec{X}\;\omega(\vec{X})}_{\substack{\text{with configurations}\\ \text{sampled according}\\ \text{to normalized } \omega(\vec{X})}} A(\vec{X})$$

The integral is obtained as a « time » average: $I = \overline{A} = \frac{1}{N} \sum_{n=1}^{N} A_n$

where A_n is a configuration-dependent physical quantity.

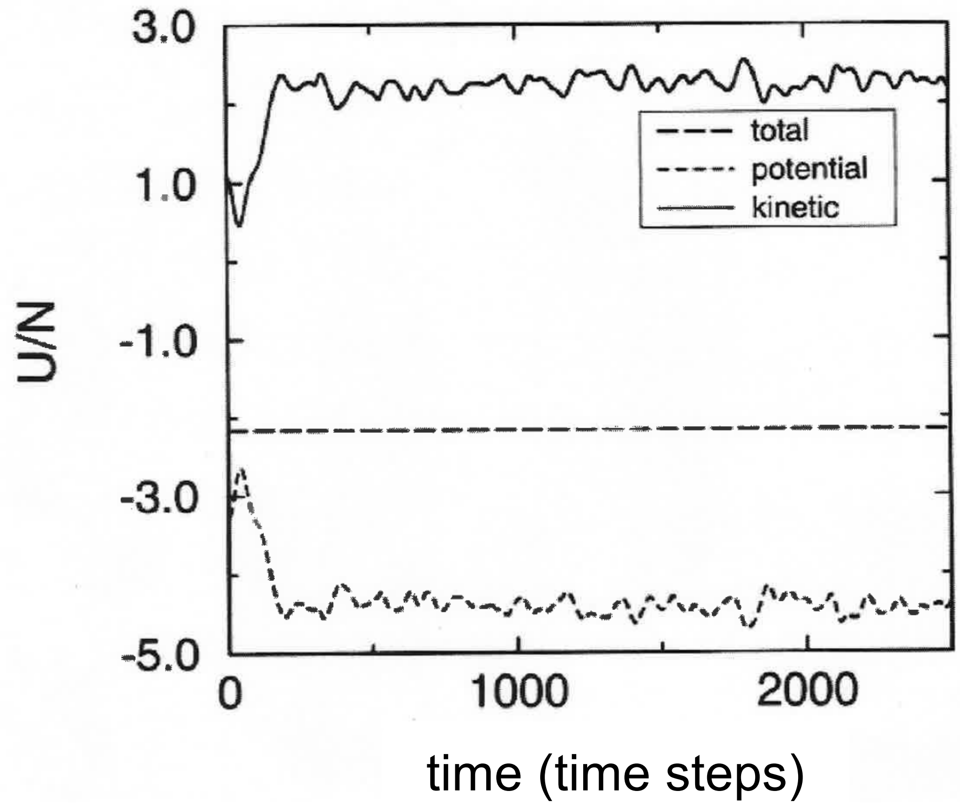
When the A_n are independent, we have seen that $\sigma_I = \frac{1}{\sqrt{N}} \sigma_A$

with $\sigma_A^2 = \langle A^2 \rangle - \langle A \rangle^2$, where $\langle \ldots \rangle$ indicates the average over many independent simulations.

For sufficiently long simulations, σ_A can be obtained from $\sigma_A^2 = \overline{A^2} - (\overline{A})^2$

Correlated sampling

$$I = \int \underbrace{d\vec{X}\ \omega(\vec{X})}_{\text{with configurations sampled according to normalized } \omega(\vec{X})} A(\vec{X})$$



What is σ_I for N correlated data ?

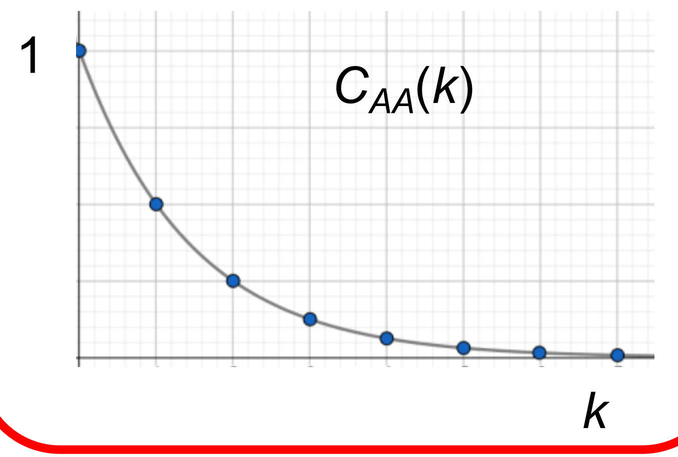
Correlation function

$$C_{AA}(k) = \frac{\langle\ (A_n - \langle A_n \rangle)(A_{n+k} - \langle A_{n+k} \rangle)\ \rangle}{\langle A_n^2 \rangle - \langle A_n \rangle^2} = \frac{\langle A_n A_{n+k} \rangle - \langle A_n \rangle^2}{\sigma_A^2}$$

where $\langle \ldots \rangle$ indicates the average over many independent simulations.

- Time translation invariance at equilibrium

- $C_{AA}(k = 0) = 1$

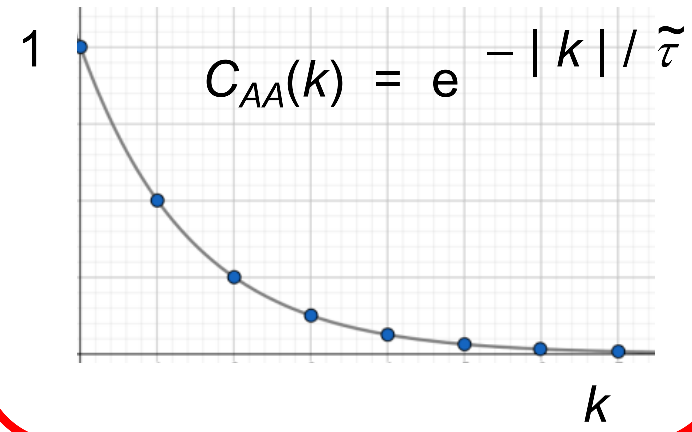


- $\lim\limits_{k \to \infty} \langle A_n A_{n+k} \rangle \underset{\uparrow}{=} \langle A_n \rangle \langle A_{n+k} \rangle = \langle A_n \rangle^2 \Rightarrow \lim\limits_{k \to \infty} C_{AA}(k) = 0$

independence at large k

Correlation time

Definition $$\tau = \frac{1}{2}\sum_{k=-\infty}^{+\infty} C_{AA}(k)$$



For an exponentially decaying correlation function, one finds $\tau = \tilde{\tau}$ for large τ.

Proof

We use for this: $$\sum_{k=0}^{\infty} q^n = \frac{1}{1-q} \qquad \text{for } q < 1$$

by definition

$$\tau = \frac{1}{2}\sum_{k=-\infty}^{+\infty} e^{-|k|/\tilde{\tau}} = \frac{1}{2}\left(\sum_{k=0}^{\infty} e^{-|k|/\tilde{\tau}} + \sum_{k=0}^{\infty} e^{-|k|/\tilde{\tau}} - 1\right)$$

$$= \frac{1}{1-e^{-1/\tilde{\tau}}} - \frac{1}{2} \underset{\tilde{\tau} \gg 1}{=} \tilde{\tau}$$

Expression for error

$$\sigma_I{}^2 = \left\langle \left[\frac{1}{N} \sum_{n=1}^{N} A_n - \left\langle \frac{1}{N} \sum_{n=1}^{N} A_n \right\rangle \right]^2 \right\rangle$$

$\langle \ldots \rangle$ indicates the average over many independent simulations.

$$= \left\langle \frac{1}{N^2} \sum_{n,m=1}^{N} A_n A_m \right\rangle - \left\langle \frac{1}{N} \sum_{n=1}^{N} A_n \right\rangle^2$$

$$= \frac{1}{N^2} \left\{ \sum_{n,m=1}^{N} \langle A_n A_m \rangle - \sum_{n,m=1}^{N} \langle A_n \rangle \langle A_m \rangle \right\}$$

$$= \frac{\sigma_A{}^2}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} C_{AA}(m-n)$$

The error can be expressed in terms of the correlation function!

Expression for error

frequency of term ℓ

$$\sum_{n=1}^{N}\sum_{m=1}^{N} C_{AA}(m-n) \quad \longrightarrow \quad \sum_{\ell=\ell_{min}}^{\ell_{max}} C_{AA}(\ell)\cdot F(\ell)$$

double sum but C_{AA} only depends on $\ell = m - n$,

with $\ell_{min} = -(N-1)$ and $\ell_{max} = N-1$

Frequency

Example: $N = 4$

$\ell = m - n$

$n \backslash m$	1	2	3	4
1	0	–1	–2	–3
2	1	0	–1	–2
3	2	1	0	–1
4	3	2	1	0

frequency of $\ell = 0$: N

frequency of $\ell = +1$: $N-1$

frequency of $\ell = -1$: $N-1$

$\vdots$

frequency of ℓ : $N - |\ell|$

Expression for error

$$\sigma_I^{\,2} = \frac{\sigma_A^{\,2}}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} C_{AA}(m-n) = \frac{\sigma_A^{\,2}}{N} \sum_{\ell=\ell_{\min}}^{\ell_{\max}} C_{AA}(\ell) \cdot \left(1 - \frac{|\ell|}{N}\right)$$

$$\underset{N \to \infty}{=} \frac{\sigma_A^{\,2}}{N} \sum_{\ell=-\infty}^{+\infty} C_{AA}(\ell) = \frac{\sigma_A^{\,2}}{N} \cdot 2\tau$$

The error can be expressed in terms of the correlation time!

To sum up:

$$\sigma_I^{\text{UNCORR.}} = \frac{\sigma_A}{\sqrt{N}}$$

$$\sigma_I^{\text{CORR.}} = \frac{\sigma_A}{\sqrt{N}} \cdot \sqrt{2\tau} = \sigma_I^{\text{UNCORR.}} \cdot \sqrt{2\tau}$$

Error from « time » averages

$$\sigma_A{}^2 = \overline{A^2} - (\overline{A})^2$$

estimated from « time » averages

$$\tau = \frac{1}{2} \sum_{n=-\infty}^{+\infty} \overline{C_{AA}(n)}$$

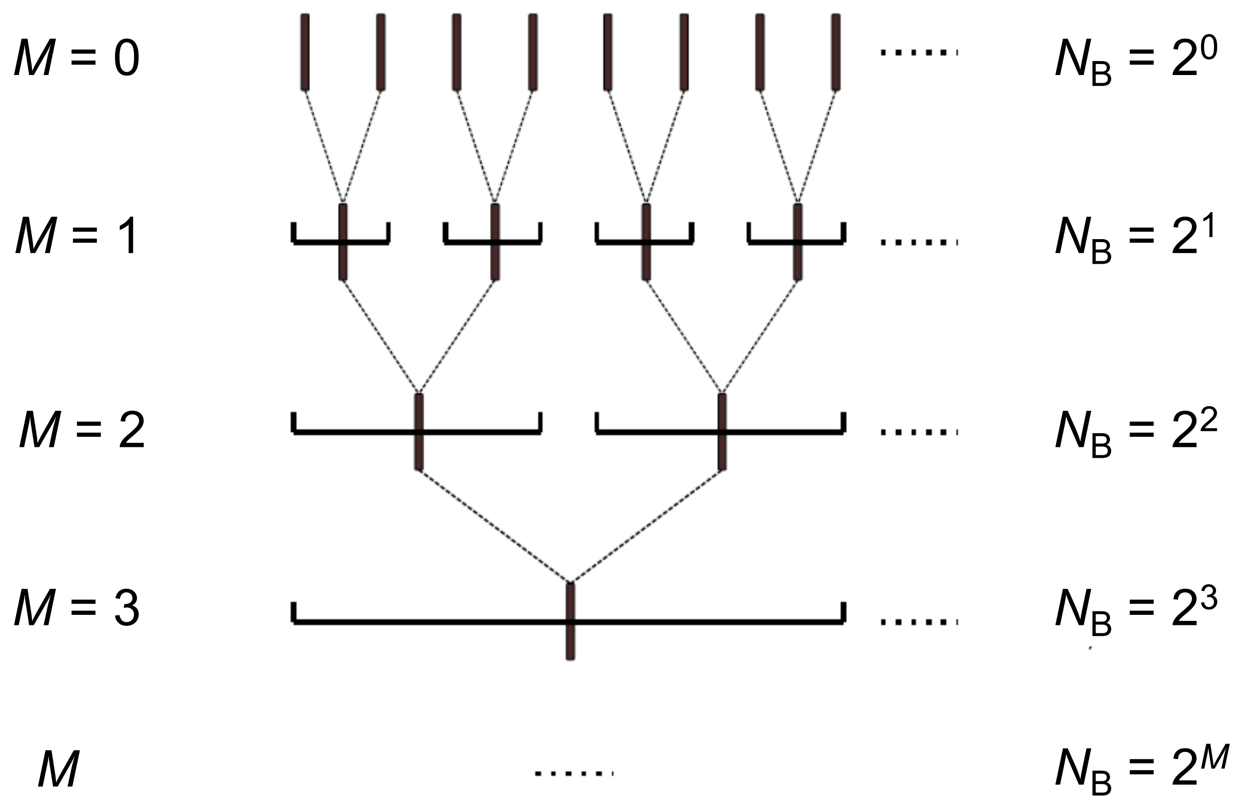
estimated from « time » averages

$$\Rightarrow \quad \boxed{\sigma_I{}^2 = \frac{\sigma_A{}^2}{N} \cdot 2\tau}$$

Blocking analysis

H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys*. **91**, 461 (1989).

The data are chopped into blocks of size N_B and are averaged over the blocks.
The averages obtained in this way are treated as if they were independent.
When N_B is sufficiently large, this will eventually become true.



Blocking analysis

For a given block size N_B, we obtain N / N_B block averages.

When N_B is larger than the correlation time, these block averages can be considered as independent and the associated variance $\sigma_B{}^2$ can be calculated.

From our error analysis, for a block of N_B data, the error can be written as

correlation time in the limit of large N_B

$$\sigma_B{}^2 = \sigma_A{}^2 \cdot \frac{2\tau_B}{N_B}$$

variance of block averages considered as independent

from «time» average

block size

$$2\tau_B = \frac{N_B \sigma_B{}^2}{\sigma_A{}^2} \xrightarrow[N_B \to \infty]{} 2\tau_c$$

constant correlation time !

Blocking analysis

Full simulation

$$\sigma_I^{\,2} = \frac{\sigma_A^{\,2}}{N} \cdot 2\tau$$

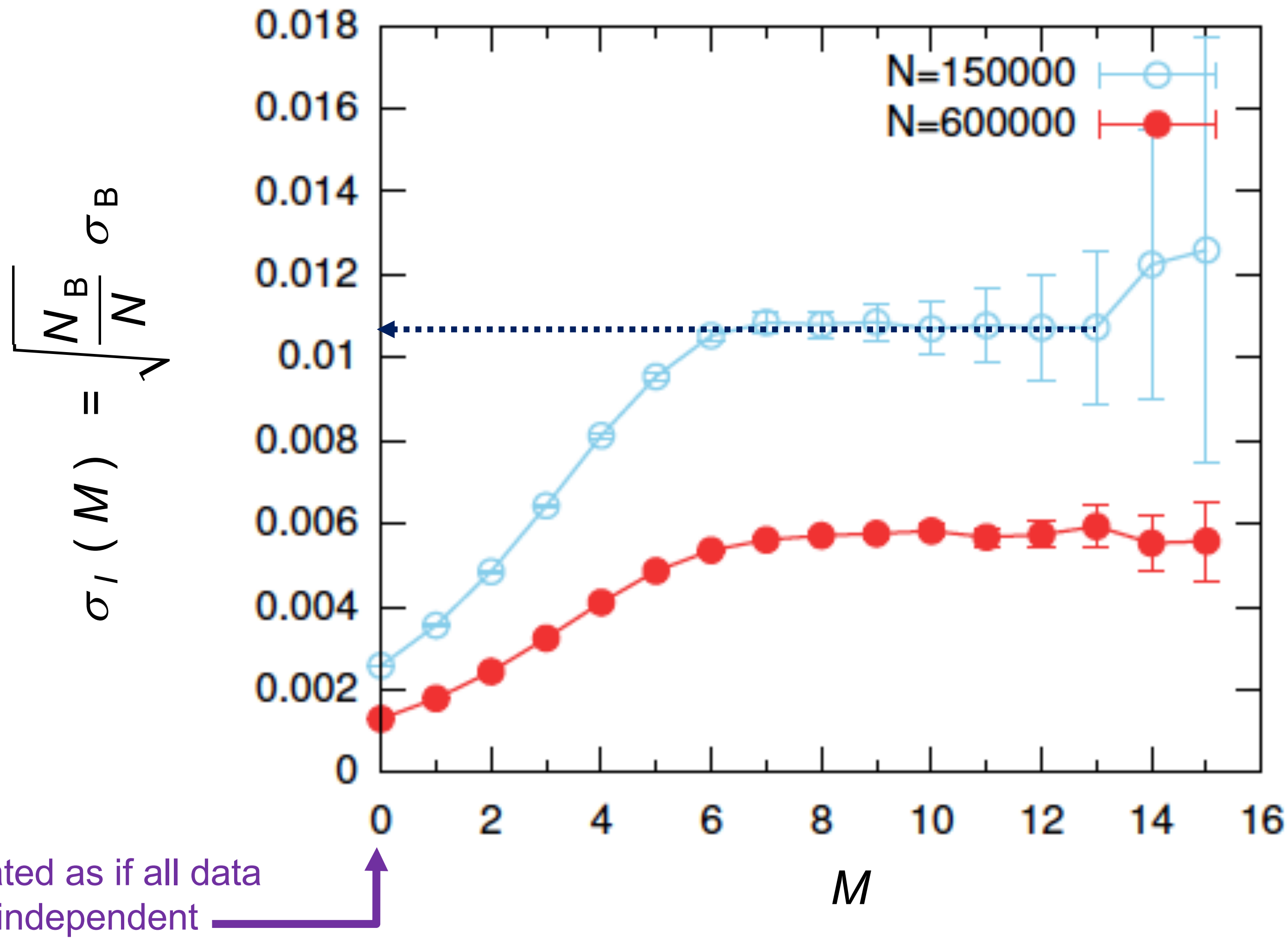
For a given transformation with block size $N_B = 2^M$:

$$\sigma_I^{\,2}(M) = \frac{\sigma_A^{\,2}}{N} \cdot 2\tau_B = \frac{N_B\,\sigma_B^{\,2}}{N} \xrightarrow[N_B \to \infty\,,\ M \to \infty]{} \sigma_I^{\,2}$$

constant N

$$\boxed{2\tau_B = \frac{N_B\,\sigma_B^{\,2}}{\sigma_A^{\,2}}}$$

Blocking analysis



Error calculated as if all data points were independent

Note that fluctuations set in for large M : few blocks, large errors.

If the plateau is not seen, more data points need to be collected (increase N).

Course 12/2

Errors in correlated sampling

- Errors in uncorrelated sampling
- Correlation function
- Correlation time
- Expression for error
- Error from « time » averages
- Blocking analysis