Reliable error estimation

Autocorrelation time

In deriving the error estimate $\delta_{N_{meas}}$, we assumed that the different "measurements" were independent:

$$p.33 \Rightarrow \int d\bar{\phi}_i \int d\bar{\phi}_j \, f(\bar{\phi}_i)\, f(\bar{\phi}_j) = \langle f \rangle^2 .$$

In a Markov chain, we change the variable $\bar{\phi}_i$ very little at a time, however. Is it justified to assume subsequent configurations statistically independent?

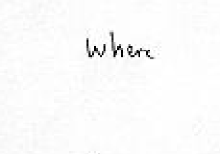
A way to estimate this is to consider the autocorrelation function

$$C(t) = \frac{\frac{1}{N_{meas}-t} \sum_{i=1}^{N_{meas}-t} f_i f_{i+t} - \langle f \rangle^2}{\frac{1}{N_{meas}} \sum_{i=1}^{N_{meas}} f_i^2 - \langle f \rangle^2} ,$$

where $\langle f \rangle \equiv \frac{1}{N_{meas}} \sum_{i=1}^{N_{meas}} f_i$ and $f_i \equiv f(\bar{\phi}_i)$.

Properties:

- $C(0) = 1$
- for $N_{meas} \to \infty$ and $t \gg 1$ such that f_i, f_{i+t} are independent, $\langle f_i f_{i+t} \rangle = \langle f_i \rangle \langle f_{i+t} \rangle = \langle f \rangle^2$, and $C(t) \approx 0$.
- it may be assumed (and it is observed in practice) that $C(t)$ extrapolates between these limits as



$$C(t) \approx \exp\left(-\frac{t}{\tau_{exp}}\right),$$

where τ_{exp} is the exponential autocorrelation time.

- the integrated autocorrelation time τ_{int} is defined as

$$\tau_{int} = \frac{1}{2} + \sum_{t=1}^{\infty} C(t) \approx \frac{1}{2} - 1 + \frac{1}{1 - \exp\left(-\frac{1}{\tau_{exp}}\right)} \overset{\tau_{exp} \gg 1}{\approx} \tau_{exp} + O\left(\frac{1}{\tau_{exp}}\right).$$

- with typical updates, $\tau_{int} \sim 10 \ldots 500$ sweeps (through the whole lattice).

- the exact value depends on the details of the algorithm — the smaller τ_{int}, the better the algorithm!

Thus, a practical simulation could be organised as follows:

(1) Carry out a test run to estimate τ_{int} for the algorithm used.

(2) Start the actual simulation from some initial configuration — "cold" = totally ordered, "hot" = totally random, or something in between.

(3) Discard $n \gg \tau_{int}$, maybe $10 \cdot \tau_{int}$ sweeps from the beginning, to allow for "thermalisation".

(4) Pick up a configuration for measurement, and wait again $\gg \tau_{int}$, maybe $2 \cdot \tau_{int}$ sweeps, before taking the next one.

(5) Pick up about ~ 1000 such independent configurations, to reach a relative error of $\sim \frac{1}{\sqrt{1000}} \sim 3\%$.
The absolute error is now $\delta_{N_{meas}}$, as defined before.

(6) τ_{int} depends on the parameters, and often grows close to a phase transition — "critical slowing down".

— o —

Jackknife & bootstrap

General methods for error estimation ; have to be used particularly for observables which are non-linear functions of quantities measured directly in the simulation, e.g.

$$f = \frac{\langle E \rangle}{\langle M \rangle^2} \quad , \text{etc.}$$

Jackknife :

- Consider statistically independent configurations ϕ_i.
- Denote the measurement of various operators from ϕ_i by O_i, while
$$O_{(i)} \equiv \frac{1}{N_{meas}-1} \sum_{j \neq i} O_j \quad \text{indicates that } i \text{ was omitted.}$$
- The expectation value is $\langle f \rangle = f\left(\frac{1}{N_{meas}} \sum_i O_i\right)$, or $\langle f \rangle = \frac{1}{N_{meas}} \sum_i f(O_{(i)})$.
- Denote $f_{(i)} = f(O_{(i)})$.
- The jackknife error is $\delta_{N_{meas}} \equiv \sqrt{\frac{(N_{meas}-1)}{N_{meas}} \sum_{i=1}^{N_{meas}} \left[f_{(i)} - \langle f \rangle \right]^2}$
- The factors can be understood by considering the special case of a linear function :

$$\frac{(N_{meas}-1)}{N_{meas}} \sum_{i=1}^{N_{meas}} \left[\frac{1}{N_{meas}-1} \sum_{j \neq i} O_j - \langle O \rangle \right]^2 \qquad \Big| \; \langle O \rangle = \frac{1}{N_{meas}} \sum_i O_i$$

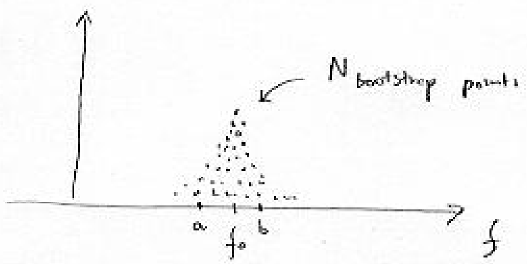
$$= \frac{(N_{meas}-1)}{N_{meas}} \sum_{i=1}^{N_{meas}} \left[\left(\frac{N_{meas}}{N_{meas}-1} - 1 \right) \langle O \rangle - \frac{1}{N_{meas}-1} O_i \right]^2$$

$$= \frac{1}{N_{meas}(N_{meas}-1)} \sum_{i=1}^{N_{meas}} \left[O_i - \langle O \rangle \right]^2 \qquad \%$$

- "Jackknife on blocked configurations" : divide the original N_{meas} configurations into a smaller number N_{block} of blocked, or averaged configurations, and perform the jackknife analysis with the blocks.

Bootstrap:

- Consider again N_{mean} configurations (or N_{block} blocks).
- Pick randomly N_{mean} out of them, without constraints. (Some will appear many times, some no times at all.)
- Compute $\langle O \rangle$, $f(\langle O \rangle)$ over the selected sample.
- Repeat this procedure a large number of times, $N_{bootstrap}$.
- Consider the outcomes as a statistical distribution:



- The value of f is now $f = f_0 {}^{+ (b-f_0)}_{- (f_0-a)}$,

  where f_0 is the average of the bootstrap points, b is chosen so that 16% of points lie above it, and a so that 16% lie below it.

- If the distribution is symmetric, $\delta_{N_{mean}} = \dfrac{b-a}{2}$.

Reweighting

- a method to extend measurements carried out at some $\beta \equiv \beta_0$, to neighbouring β, without additional simulations.

(Ferrenberg, Swendsen, Phys. Rev. Lett. 61 (1988) 2635
63 (1989) 1195)

Denote $S = \beta E$.

Consider $\langle O \rangle = \frac{1}{N_{meas}} \sum_i O_i = \frac{\sum_i O_i}{\sum_i 1}$.

The configurations here have been generated according to probability

$$p_{\beta_0}(\bar\phi_i) = \frac{1}{Z} e^{-S(\bar\phi_i)} = \frac{1}{Z} e^{-\beta_0 E(\bar\phi_i)}.$$

If now $\beta = \beta - \beta_0 + \beta_0 = \beta_0 + \delta\beta$, the correct probability would be

$$p_\beta(\bar\phi_i) \propto e^{-\delta\beta E(\bar\phi_i)} \, p_{\beta_0}(\bar\phi_i).$$

$\Rightarrow$ Simply insert $e^{-\delta\beta E(\bar\phi_i)}$ into the observables!

$$\Rightarrow \quad \langle O \rangle \approx \frac{\sum_i O_i e^{-\delta\beta E_i}}{\sum_i e^{-\delta\beta E_i}}, \quad \text{where } E_i \equiv E(\bar\phi_i).$$

Errors of reweighted $\langle O \rangle$ by jackknife / bootstrap.