Monte Carlo simulations and supercomputers

Piotr Korcyl

March 23, 2025

Summary of lecture 3

Last week we discussed

- Monte Carlo simulation for statistical systems -> Metropolis's paper
- Markov chains

Markov chains

Requirements for the transition probabilities:

- $T(s_i \to s_j) \geq 0$ for all s_i,s_j and

$$\sum_{s_j} T(s_i \to s_j) = 1$$

  for all s_i

- $$\sum_{s_i} P^{eq}(s_i) T(s_i \to s_j) = P^{eq}(s_j)$$

  for all s_j

- $T(s_i \to s_i) > 0$ for all s_i
- if S is a non-empty proper subset of states, there exist two states $s_i \in S$ and $s_j \notin S$ such that $T(s_i \to s_j) > 0$

Markov chains

In practice most algorithms satisfy detailed balance condition which is stronger:

$$T(s_\mu \to s_\nu) P^{eq}(s_\mu) = T(s_\nu \to s_\mu) P^{eq}(s_\nu)$$

for all states μ and ν.

$$\sum_{s_\nu} T(s_\mu \to s_\nu) = 1, \qquad \sum_{s_\nu} P^{eq}(s_\nu) T(s_\nu \to s_\mu) = P^{eq}(s_\mu)$$

Proof:

$$\sum_{s_\nu} T(s_\mu \to s_\nu) P^{eq}(s_\mu) = \sum_{s_\nu} T(s_\nu \to s_\mu) P^{eq}(s_\nu)$$

$$1 P^{eq}(s_\mu) = \sum_{s_\nu} P^{eq}(s_\nu) T(s_\nu \to s_\mu)$$

$$1 P^{eq}(s_\mu) = P^{eq}(s_\mu)$$

Markov chains

It is often practical to factor the transition probability into two parts:

- proposal probability $g(s_\mu \to s_\nu)$ being the probability to propose s_ν when in s_μ
- acceptance probability $A(s_\mu \to s_\nu)$ being the probability to accept the proposed state (when the state is rejected, the current state is repeated)

Hence,

$$T(s_\mu \to s_\nu) = g(s_\mu \to s_\nu) A(s_\mu \to s_\nu)$$

and the detailed balance condition

$$T(s_\mu \to s_\nu) P(s_\mu) = T(s_\nu \to s_\mu) P(s_\nu)$$

becomes

$$\frac{P(s_\nu)}{P(s_\mu)} = \frac{T(s_\mu \to s_\nu)}{T(s_\nu \to s_\mu)} = \frac{g(s_\mu \to s_\nu) A(s_\mu \to s_\nu)}{g(s_\nu \to s_\mu) A(s_\nu \to s_\mu)}$$

Markov chains

$\Rightarrow$ generate states with a simple g and then tune A such that detailed balance is satisfied.

Remarks about g:

- if $g(s_\mu \to s_\nu)$ favors big steps in configuration space
  - configurations/states are far apart
  - small acceptance probability
  - same configuration is repeated
  - long correlation times
- if $g(s_\mu \to s_\nu)$ favors small steps in configuration space
  - configurations/states are close together
  - high acceptance probability
  - configuration space traversed very slowly
  - long correlation times

Markov chains

Metropolis algorithm: a particular proposal for g and A such that configurations from the Boltzmann probability distribution are generated.

1. proposal: g generate a new proposal s_ν in the neighborhood of s_μ according to a probability distribution g which is symmetric

$$g(s_\mu \to s_\nu) = g(s_\nu \to s_\mu)$$

2. compute the energies E_{s_ν} and E_{s_μ}
3. acceptance: A
   - if $E_{s_\nu} \leq E_{s_\mu}$; accept
   - if $E_{s_\nu} > E_{s_\mu}$, accept with probability

$$\rho = \frac{P^{\mathrm{eq}}(s_\nu)}{P^{\mathrm{eq}}(s_\mu)} = e^{-\beta(E_{s_\nu} - E_{s_\mu})}$$

The acceptance probability is thus

$$A(s_\mu \to s_\nu) = \min\Big(1, \rho\Big)$$

Markov chains

Proof:

$$\frac{T(s_\mu \to s_\nu)}{T(s_\nu \to s_\mu)} = \frac{g(s_\mu \to s_\nu)A(s_\mu \to s_\nu)}{g(s_\nu \to s_\mu)A(s_\nu \to s_\mu)} = \frac{g(s_\mu \to s_\nu)\min\left(1, \rho\right)}{g(s_\nu \to s_\mu)\min\left(1, \rho^{-1}\right)} =$$

$$= \rho = \frac{P^{\mathrm{eq}}(s_\nu)}{P^{\mathrm{eq}}(s_\mu)}$$

Markov chains

A generic proposal probability can be incorporated into such an accept/reject step

$$A(s_\mu \to s_\nu) = \min\left(1, \rho \frac{g(s_\nu \to s_\mu)}{g(s_\mu \to s_\nu)}\right)$$

Markov chains

Transition matrix including the accept/reject step. We said that

$$T(s_\mu \to s_\nu) = g(s_\mu \to s_\nu) A(s_\mu \to s_\nu)$$

however special care has to be taken when dealing with the diagonal elements.
In the implementation we have:

- $s_\mu \neq s_\nu$

$$T(s_\mu \to s_\nu) = g(s_\mu \to s_\nu) A(s_\mu \to s_\nu)$$

- $s_\mu = s_\nu$

$$T(s_\mu \to s_\nu) = g(s_\mu \to s_\nu) A(s_\mu \to s_\nu) + \\ + \sum_\kappa g(s_\mu \to s_\kappa) \Big(1 - A(s_\mu \to s_\kappa)\Big)$$

Data analysis techniques

Assume we have constructed a Markov process and have one instance of a Markov chain. We measure some observable O on each configuration.

How do we correctly calculate statistical uncertainties of the results?

How to correctly account for:

- primary/secondary observables
- correlations
- autocorrelations

Primary/secondary observables

Observables can be divided into two classes

- primary:
  we call an observable primary when it can be estimated on a single element of the Markov chain
- secondary:
  all observables that are not primary

Primary/secondary observables

Observables are affected by

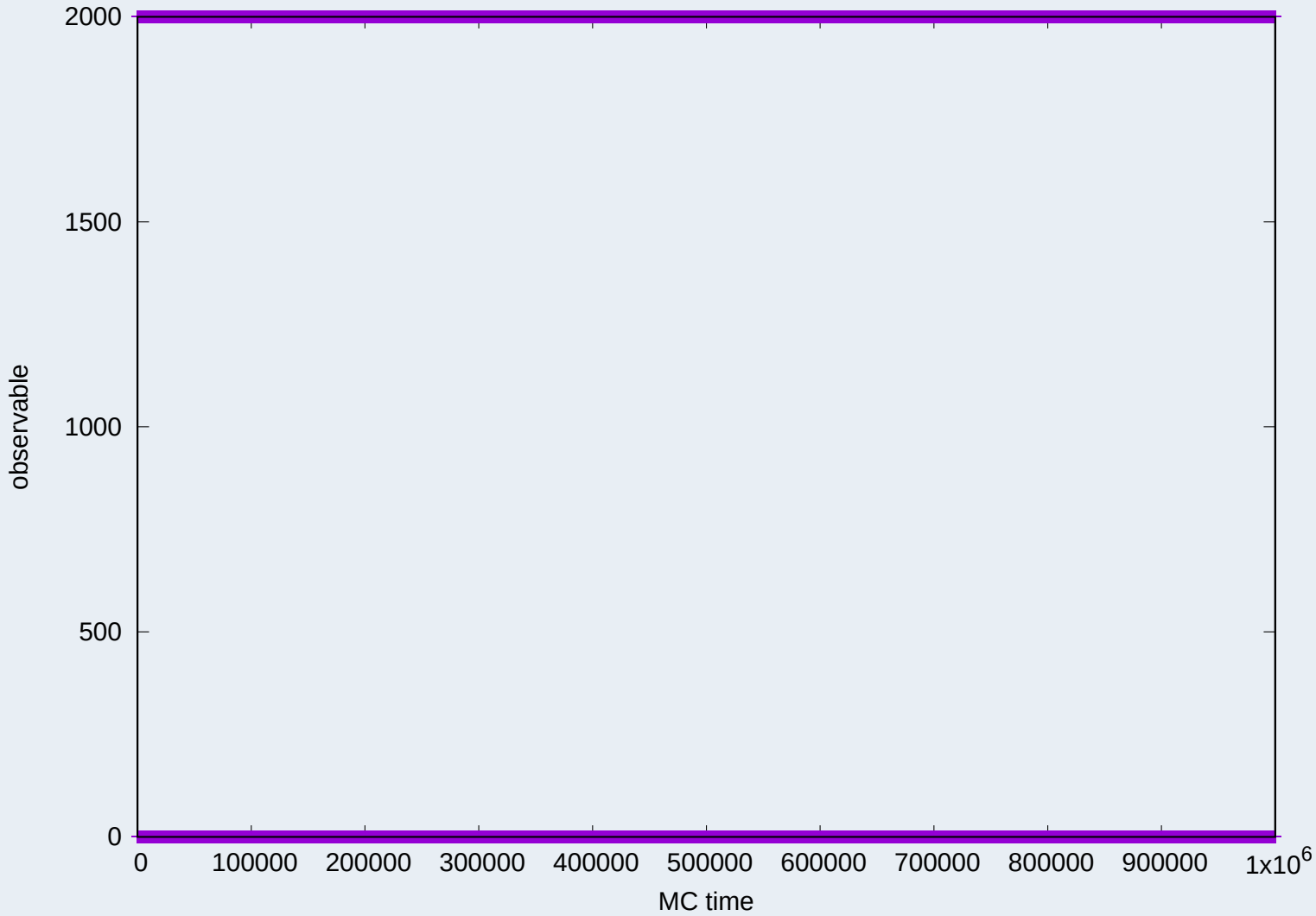
- primary:
  - autocorrelations
- secondary:
  - autocorrelations
  - correlations with other observables

Autocorrelations

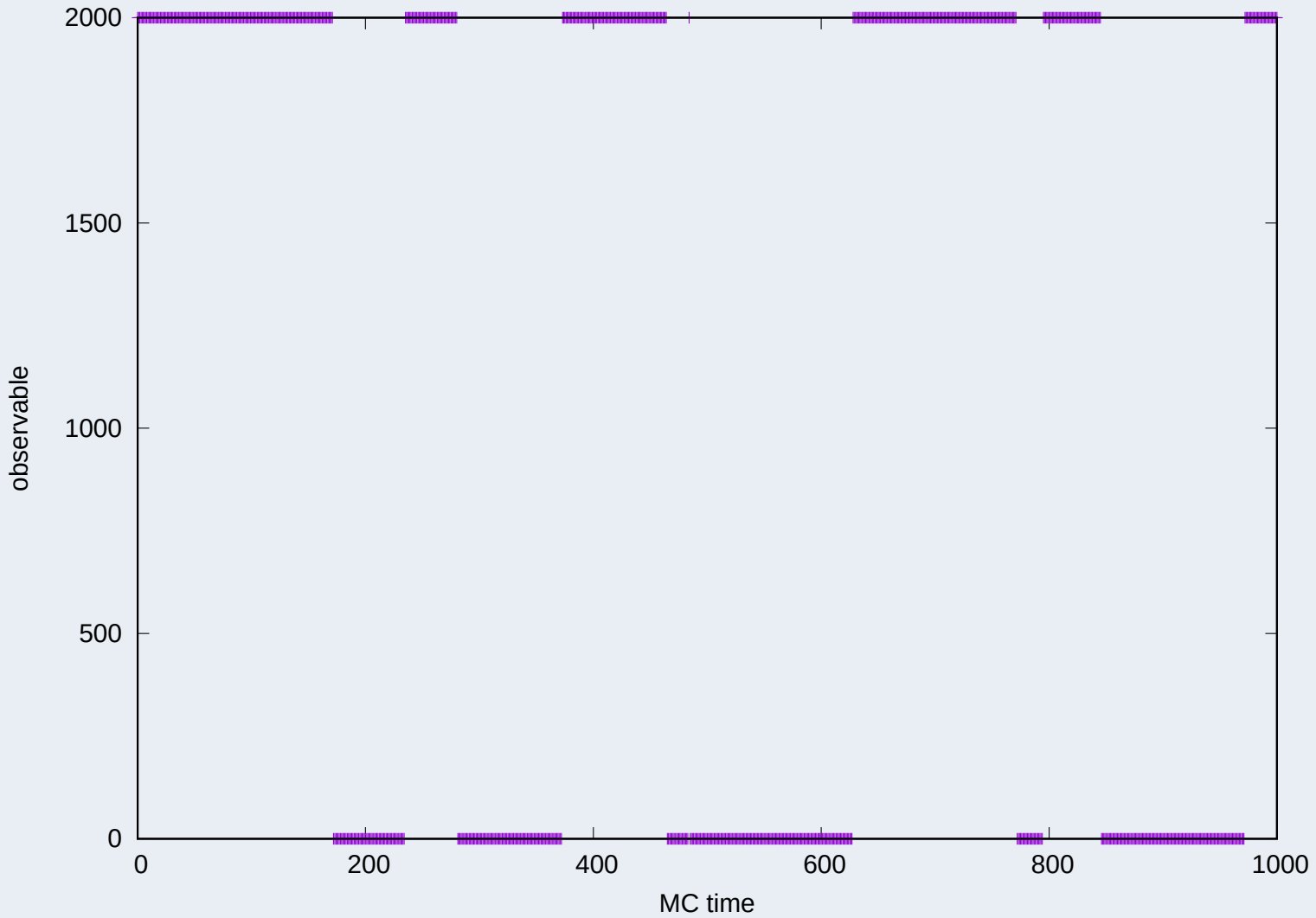
Example:

- system with two states $\{s_0, s_1\}$
- when system in state s_0 we "measure" $\mathcal{O}(s_0) = 2000$
- when system in state s_1 we "measure" $\mathcal{O}(s_1) = 0$
- transition probabilities:
  - $s_0 \rightarrow s_0 = 0.99$
  - $s_0 \rightarrow s_1 = 0.01$
  - $s_1 \rightarrow s_0 = 0.01$
  - $s_1 \rightarrow s_1 = 0.99$
- on the average we expect the system to occupy equally state s_0 and s_1
- on the average we expect to measure $< \mathcal{O} >= 1.0$

Autocorrelations



Autocorrelations



Autocorrelations

- average value from 2000 measurements: 1165 ± 22
- average value from 1000000 measurements: 1002 ± 1

Autocorrelations

The variance of $\bar{y}$ is

$$\sigma_{\bar{y}}^2 = \langle \bar{y}^2 \rangle - \langle \bar{y} \rangle^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \langle y_i y_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^{N} \langle y_i \rangle \langle y_j \rangle \tag{1}$$

We separate the sums into two cases: $i = j$ and $i \neq j$

$$\sigma_{\bar{y}}^2 = \frac{1}{N^2} \sum_{i=1}^{N} \left(\langle y_i^2 \rangle - \langle y_i \rangle^2 \right) + \frac{1}{N^2} \sum_{i \neq j}^{N} \left(\langle y_i y_j \rangle - \langle y_i \rangle \langle y_j \rangle \right) \tag{2}$$

Autocorrelations

$$\sigma_{\bar{y}}^2 = \frac{1}{N}\sigma_y^2 + \frac{1}{N^2}\sum_{i\neq j}^{N}\Big(\langle y_i y_j\rangle - \langle y_i\rangle\langle y_j\rangle\Big) \tag{3}$$

We use the symmetry between $i \leftrightarrow j$ to write

$$\sigma_{\bar{y}}^2 = \frac{1}{N}\sigma_y^2 + \frac{2}{N^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Big(\langle y_i y_j\rangle - \langle y_i\rangle\langle y_j\rangle\Big) \tag{4}$$

and substitute $t = j - 1$ to get

$$\sigma_{\bar{y}}^2 = \frac{1}{N}\sigma_y^2 + \frac{2}{N^2}\sum_{i=1}^{N-1}\sum_{t=1}^{N-i}\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big) \tag{5}$$

Autocorrelations

Finally,

$$\sigma^2_{\bar{y}} = \frac{1}{N}\sigma^2_y + \frac{2}{N^2}\sum_{t=1}^{N-1}\sum_{i=1}^{N-t}\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big) \tag{6}$$

Now we use the fact that the average values are independent of the position in the Markov chain i, so we can write

$$\sigma^2_{\bar{y}} = \frac{1}{N}\sigma^2_y + \frac{2}{N^2}\sum_{t=1}^{N-1}\Big(N-t\Big)\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big) \tag{7}$$

and eventually,

$$\sigma^2_{\bar{y}} = \frac{1}{N}\sigma^2_y\Big[1 + \frac{2}{N\sigma^2_y}\sum_{t=1}^{N-1}\Big(N-t\Big)\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big)\Big] \tag{8}$$

Autocorrelations

We arrived at

$$\sigma_{\bar{y}}^2 = \frac{1}{N}\sigma_y^2\Big[1 + \frac{2}{N\sigma_y^2}\sum_{t=1}^{N-1}\Big(N-t\Big)\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big)\Big] \tag{9}$$

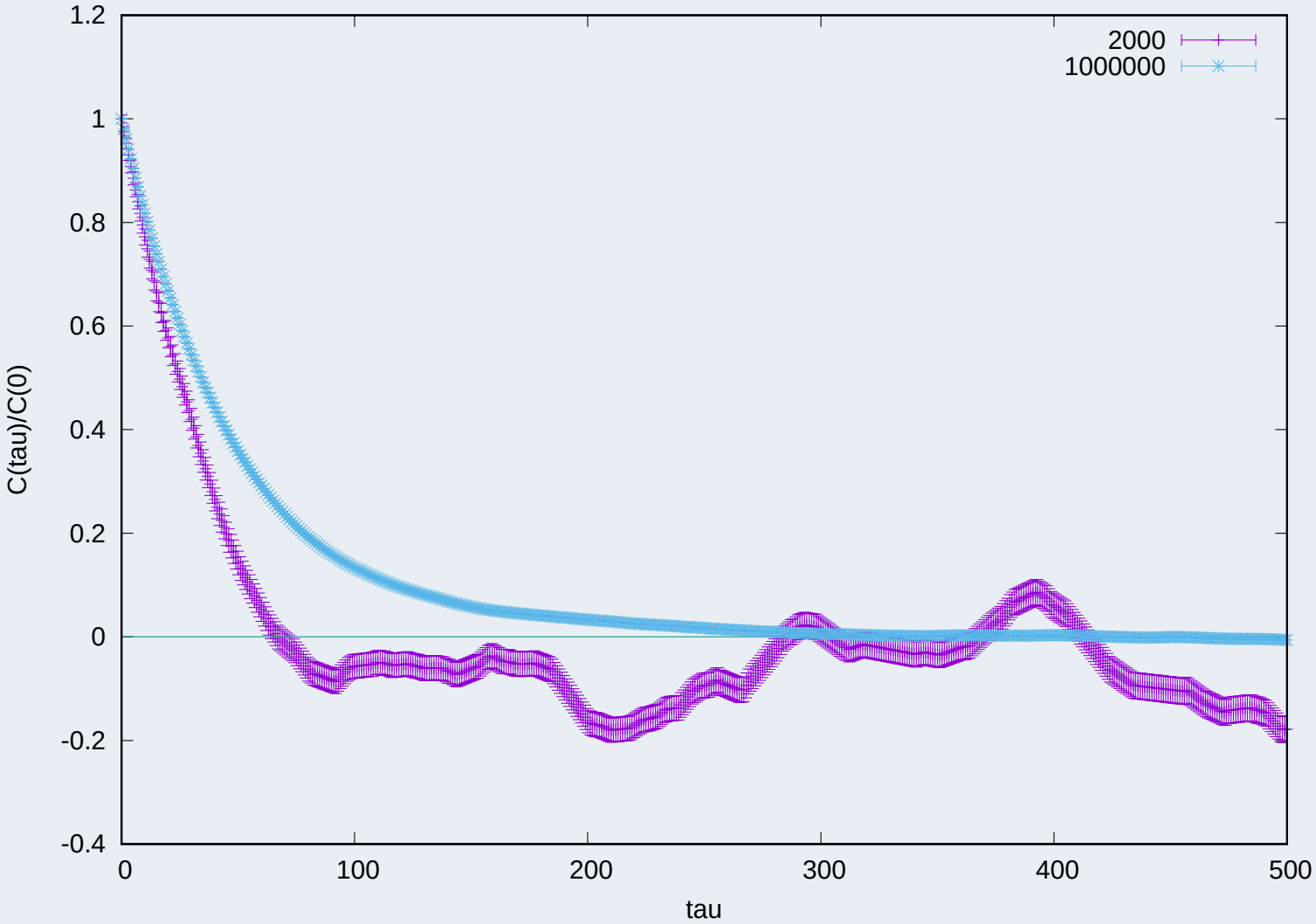
where the quantity in the bracket is the integrated autocorrelation time

$$2\tau_{\rm int} = 1 + \frac{2}{N\sigma_y^2}\sum_{t=1}^{N-1}\Big(N-t\Big)\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big) \tag{10}$$

or

$$\tau_{\rm int} = \frac{1}{2} + \frac{1}{\sigma_y^2}\sum_{t=1}^{N-1}\Big(1-\frac{t}{N}\Big)\Big(\langle y_i y_{i+t}\rangle - \langle y_i\rangle\langle y_{i+t}\rangle\Big) \tag{11}$$

Autocorrelations



Autocorrelations

Summarizing:

$$\sigma_{\bar{y}}^2 = \frac{1}{N}\sigma_y^2 \times 2\tau_{\mathrm{int}} \tag{12}$$

where

$$\tau_{\mathrm{int}} = \frac{1}{2} + \sum_{t=1}^{N-1} \rho(t) \tag{13}$$

with

$$\rho(t) = \frac{C_y(t)}{C_y(0)} \tag{14}$$

Blocking method

- divide the N measurements in n_b blocks, each block with N_b measurements
- measure the observable on each block

$$O_{(i)} = \frac{1}{N_b} \sum_{i=1}^{N_b} O_i \tag{15}$$

- compute the block-average

$$\bar{O}(n_b) = \frac{1}{n_b} \sum_{i=1}^{n_b} O_{(i)} \tag{16}$$

Blocking method

- compute the block-average

$$\bar{O}(n_b) = \frac{1}{n_b} \sum_{i=1}^{n_b} O_{(i)} \tag{17}$$

together with the block-variance

$$\sigma_{\bar{O}(n_b)} = \frac{1}{\sqrt{n_b}} \sqrt{\frac{1}{n_b} \sum_{i=1}^{n_b} O_{(i)}^2 - \frac{1}{n_b^2} \sum_{i=1}^{n_b} O_{(i)} \sum_{i=1}^{n_b} O_{(i)}} \tag{18}$$

- increase n_b and find the value for which the block-variance changes no longer

Blocking method

- increase n_b and find the value for which the block-variance changes no longer
- the method is simple and robust
- it may however fail if the autocorrelation time is of the order or larger than N

Primary/secondary observables

Primary observables are affected by autocorrelations. Example:

- mean magnetization in the Ising model

$$\langle |m| \rangle = \frac{1}{N} \sum_{t=1}^{N} |\frac{1}{V} \sum_{x} s_x| \tag{19}$$

- together with the uncertainty

$$\sigma_{|\bar{m}|} = \sigma_{|m|} \sqrt{\frac{2\tau_{\mathrm{int}}}{N}} \tag{20}$$

What about secondary observables?

- functions of primary observables
- more complicated constructions of primary observables: fits

Secondary observables

Let's take the simplest example: $\langle AB \rangle - \langle A \rangle \langle B \rangle$

- uncertainties of $\langle A \rangle$ and $\langle B \rangle$ correlated
- naively one would calculate

$$\sigma_{\langle A \rangle \langle B \rangle} = \sigma_A |\langle B \rangle| + \sigma_B |\langle A \rangle| \tag{21}$$

- or

$$\sigma_{\langle A \rangle \langle B \rangle} = \sqrt{\sigma_A^2 (\langle B \rangle)^2 + \sigma_B^2 (\langle A \rangle)^2} \tag{22}$$

- but the above assumes that σ_A and σ_B are independent
- when we estimate A and B in the same Markov chain, the uncertainties of $\langle A \rangle$ and $\langle B \rangle$ are related

Secondary observables

Let's take the simplest example: $\langle AB \rangle - \langle A \rangle \langle B \rangle$

- more precisely and generally one should calculate

$$\sigma_f^2 = (\partial_x f)^2 \sigma_{\bar{x}}^2 + (\partial_y f)^2 \sigma_{\bar{y}}^2 + 2(\partial_x f)(\partial_y f)\mathrm{cov}(\bar{x}, \bar{y}) \tag{23}$$

where the covariance matrix

$$\mathrm{cov}(x, y) = \langle (x_i - \mu_x)(y_i - \mu_y) \rangle \tag{24}$$

and is such that

$$\mathrm{cov}(x, x) = \sigma_x^2 \tag{25}$$

Secondary observables

How to correctly take into account the covariance matrix? $\Rightarrow$ resampling methods

- jackknife
- bootstrap

Jackknife method

Consider a Markov chain S with N elements: S_i, $0 \leq i < N$. Let's also take a secondary observable measured on S, $Q(S)$.

- construct N new chains by dropping one element of S. Hence,
  - S^0 will contain $N-1$ elements, $S^0_i = S_{i+1}$, $0 \leq i < N-1$,
  - S^1 will contain $N-1$ elements, $S^1_0 = S_0$ and $S^1_i = S_{i+1}$, $1 \leq i < N-1$,
  - S^2 will contain $N-1$ elements, $S^2_0 = S_0$,$S^2_1 = S_1$ and $S^2_i = S_{i+1}$, $2 \leq i < N-1$,
  - ...
- evaluate Q on each 'jackknife sample', $Q(S^i)$
- the jackknife estimate for the standard error of $Q(S)$ is

$$\sigma_Q = \sqrt{\frac{N-1}{N} \sum_{i=1}^{N} \left(Q(S^i) - \bar{Q}_{\text{jackknife}} \right)^2} \tag{26}$$

Jackknife method

Consider a Markov chain S with N elements: S_i, $0 \leq i < N$. Let's also take a secondary observable measured on S, $Q(S)$.

- the jackknife estimate for the standard error of $Q(S)$ is

$$\sigma_Q = \sqrt{\frac{N-1}{N} \sum_{i=1}^{N} \left(Q(S^i) - \bar{Q}_{\text{jackknife}} \right)^2} \tag{27}$$

where

$$\bar{Q}_{\text{jackknife}} = \frac{1}{N} \sum_{i=1}^{N} Q(S^i) \tag{28}$$

- final result:

$$Q(S) \pm \sigma_Q \tag{29}$$

Bootstrap method

Consider a Markov chain S with N elements: S_i, $0 \leq i < N$. Let's also take a secondary observable measured on S, $Q(S)$.

- construct M new chains by randomly choosing N elements of S allowing for duplicates. Hence,
  - S^0 will contain N elements, $S^0_i = S_{\mathrm{rand}(i)}$, $0 \leq i < N$,
  - S^1 will contain N elements, $S^1_0 = S_{\mathrm{rand}(i)}$, $0 \leq i < N$,
  - S^2 will contain N elements, $S^2_0 = S_{\mathrm{rand}(i)}$, $0 \leq i < N$,
  - ...
- evaluate Q on each 'pseudo-sample', $Q(S^i)$, $0 \leq i < M$
- the bootstrap estimate for the standard error of $Q(S)$ is

$$\sigma_Q = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(Q(S^i) - \bar{Q}_{\mathrm{bootstrap}}\right)^2} \tag{30}$$

Bootstrap method

Consider a Markov chain S with N elements: S_i, $0 \leq i < N$. Let's also take a secondary observable measured on S, $Q(S)$.

- the bootstrap estimate for the standard error of $Q(S)$ is

$$\sigma_Q = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(Q(S^i) - \bar{Q}_{\text{bootstrap}}\right)^2} \tag{31}$$

where

$$\bar{Q}_{\text{bootstrap}} = \frac{1}{M}\sum_{i=1}^{M} Q(S^i) \tag{32}$$

- final result:

$$Q(S) \pm \sigma_Q \tag{33}$$

- the uncertainty decreases with increasing N
- the uncertainty is estimated more precisely with increasing M