3.Expectation values: estimates and standard error

- In statistical physics: interested in expectation values of measurements performed on states of a system that are distributed according to some probability distribution, e.g. the Boltzmann distribution.
- Consider (high - dimension) physical states (vector)x distributed according to a probability distribution p(x)

$$\mu \equiv < y > = \int d\vec{x} p(\vec{x}) y(\vec{x})$$

where \int dx = \int dx_1 \int dx_2 … \int dx_N

for x = (x_1, x_2, …, x_N)

And variance

$$\sigma_y^2 = < (y - \mu_y) > = \int d\vec{x} p(\vec{x}) \left(y(\vec{x}) - \mu_y \right)^2 = < y >^2 - < y^2 >$$

The exact solution of μ_y and σ_y are unknown and are measured in experiment; simulations.

Note: the states $\vec{x}$ of the system one typically high-dimensional (e.g. spins on a grid) and $\int d\vec{x}$ represents a high-dimensional integral.

Theorem: Given a sample to $S = \left\{ \vec{x}_i \,|\, i = 1,\ldots, N \right\}$ of size N with states $\vec{x}_i$ **independently** distributed according to $p(\vec{x})$

a) The sample mean $\bar{y}$ yields an estimate for $< y >$ with;

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \quad \text{where} \quad y_i = y(\vec{x}_i)$$

b) The accuracy of the estimate $\bar{y}$ is given by the **standard error** $\sigma_{\bar{y}}$, defined as the standard deviation of the distribution of sample means $\bar{y}$ measured over all possible independent samples S of some size N (repeated sampling) and is given by:

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}} \quad \sigma_y = \sqrt{< (y - \mu_y)^2 >} \qquad < (y - \mu_y)^2 > = \int d\vec{x} p(\vec{x}) (y(\vec{x}) - \mu_y)^2$$

Proof:

a) In repeated sampling, the sample means $\bar{y}$ are distributed probabilistically with expectation value $\langle \bar{y} \rangle$

$$\langle \bar{y} \rangle = \langle \frac{1}{N} \sum_{i=0}^{N} y_i \rangle = \frac{1}{N} \sum_{i=0}^{N} \langle y_i \rangle = \mu_y = \langle y \rangle$$

—> we can use the mean $\bar{y}$ of one sample as estimated for $\langle y \rangle$

b) Variance of sample means $\bar{y}$ are repeated samplings;

$$\begin{aligned}
\sigma_{\bar{y}}^2 &= \left\langle \left(\bar{y} - \langle \bar{y} \rangle\right)^2 \right\rangle, \quad \langle \bar{y} \rangle = \mu_y \\
&= \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} y_i - \mu_y \right)^2 \right\rangle \\
&= \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} (y_i - \mu_y) \right)^2 \right\rangle \\
&= \left\langle \frac{1}{N^2} \sum_{i,j=1}^{N} (y_i - \mu_y)(y_j - \mu_y) \right\rangle \\
&= \frac{1}{N^2} \sum_{i,j=1}^{N} \left\langle (y_i - \mu_y)(y_j - \mu_y) \right\rangle \\
&= \frac{1}{N^2} \sum_{i,j=1}^{N} \left\langle (y_i - \mu_y)(y_j - \mu_y) \right\rangle \\
&= \ldots \\
&= \frac{\sigma_y^2}{N} \\
&\Rightarrow \sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}} \quad \text{stadard error}
\end{aligned}$$

- A measured estimate only makes sense together with its standard error

$$\langle y \rangle = \bar{y} \pm \frac{\sigma_y}{\sqrt{N}}$$

The standard error σ_y gives the uncertainty on $\bar{y}$ as estimate for $\langle y \rangle$

- In practice: exact σ_y is unknown

  —> approximate σ_y by the sample standard deviation of the measured sample

$$\sigma_y \approx S_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2} \quad \Rightarrow \quad \text{using the information of a single sample}$$

- Central limit Theorem: For sample size $N \to \infty$, the sample means $\bar{y}$ are <u>normally distributed</u> with mean μ_y and standard deviation $\frac{\sigma_y}{\sqrt{N}}$, independently of the shape of $p(\vec{x})$

—> Hence, for large N

$$P\left(|\bar{y}-\mu_y| < \sigma_{\bar{y}}\right) = 68.3\,\%\,, \quad P\left(|\bar{y}-\mu_y| < 2\sigma_{\bar{y}}\right) = 95.4\,\%\,, \quad P\left(|\bar{y}-\mu_y| < 3\sigma_{\bar{y}}\right) = 99.7\,\%$$

4 Classical spin systems in statistical physics

4.1 Ising model:

- Simplest spin model in statistical physics
- Discrete, classical spin model

More sophisticated models:
- Continuous spin models
- Quantum spin models

- Ising model: idealized model for ferromagnet. Spins represent the magnetic dipole moments of atom.
- Has many properties in common with more complicated systems —> ideal test lab
- Consider d - dimensional grid, where grid points $\vec{x}$ have integer coordinates

$$\vec{x} = (x_1, \ldots, x_d) \quad x_\nu \in \{1, \ldots, L_\nu\} \quad and \quad \nu \in \{1, \ldots, d\}$$

e.g. $d = 2$, $L_1 = L_2 = 4$

Volume: $$V = \prod_{\nu=1}^{d} L_\nu$$

- Ising model: each grid point has a spin $s \in \{+1, -1\} \quad or \quad \{\uparrow, \downarrow\}$
- Simplify notation: replace $\vec{x}$ by a linear indexation of the grid points: $i = 1, \ldots, V$. Each grid point i has a spin S_i
- A configuration of state is given by a tuple of $\mu = (S_1, S_2, \ldots, S_V)$
- Hamiltonian to the ferromagnetic Ising model ($J > 0$):

$$H(\mu) = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i$$

Where $\langle ij \rangle$ denotes nearest neighbors i and j and so $\sum_{\langle ij \rangle}$: „sum over all nearer neighbors“

- Nearest neighbor interaction:

$$-JS_iS_j = \begin{cases} -J & for \ \uparrow\uparrow \ or \ \downarrow\downarrow \\ +J & for \ \uparrow\downarrow \ or \ \downarrow\uparrow \end{cases}$$

—> Spins prefer to be aligned, anti - aligned neighbors

- Extend magnetic field B: contribution of S_i is

$$-BS_i = \begin{cases} -B & for \ \uparrow \\ +B & for \ \downarrow \end{cases}$$

—> Spin orientated against B costs energy $2B$

Ground state:
- for $B > 0$: all spin $\uparrow$ —> minimal energy
- for $B = 0$: two ground stats: all spins $\uparrow$ or all spins $\downarrow$
- Simulations in finite volume —> choose boundary conditions (b.c.)

<u>E.g.</u> open b.c. : no neighbors on edge

fixed b.c. : set all neighbors outside the grid to $\uparrow$ or $\downarrow$

periodic b.c. : wrap around the lattice <— best

helical b.c. : follows linear indexing

4.2 Canonical ensamble and observables

- State of the system is determined by the spin values on all grid points —> spin state $\mu = (S_i)$

E.g. $d = 2,\ L = 2$

Enumerate the states:

$$\mu_1 : (\uparrow\uparrow\uparrow\uparrow) \quad \mu_2 : (\uparrow\uparrow\uparrow\downarrow) \quad \mu_3 : (\uparrow\uparrow\downarrow\uparrow) \quad \dots \quad \mu_{16} : (\downarrow\downarrow\downarrow\downarrow)$$

Number of states: $2^4 = 16$
In General for Ising model for V grid points : 2^V states

- <u>Ensemble:</u> set of all states $\{\mu\}$ where each state μ occurs with a probability P_μ

Canonical ensemble: temperature of the system is kept constant —> system is in thermal equilibrium at temperature T

- the statistical distribution of states is described by the <u>partition function:</u>

$$Z = \sum_{\mu \in \Omega} e^{-\beta E_\mu} \quad \text{with } \beta = \frac{1}{kT}, \text{ temperature } T, \text{ Boltzmann constant } k$$

- Probability that the system (in thermal equilibrium) is in state μ with energy $E_\mu = H(\mu)$, is:

$$P_\mu = \frac{e^{-\beta E_\mu}}{Z} \text{ Boltzmann distribution}$$

- P_μ is a probability:

$$P_\mu \geq 0 \ for\ \mu = (S_i), \quad \sum_{\mu \in \Omega} P_\mu = \frac{1}{Z} \sum_{\mu \in \Omega} e^{-\beta E_\mu} = \frac{1}{Z} Z = 1$$

in SI units: $[E] = J$, $[T] = K$, $k_B = 1{,}3806488 \cdot 10^{-23} \frac{J}{K}$

in simulations: choose natural units where $k_B = 1$ $\rightarrow [E] = [T]$, e.g. J or eV

$1J \rightarrow 7{,}243 \cdot 10^{22}\,K$ $\qquad$ $1eV \,\hat{=}\, 1{,}16 \cdot 10^{4}\,K$

Observables

- Expectation values: defined as ensemble averages
  - Energy: $\langle E \rangle = \sum_{\mu \in \Omega} p_\mu E_\mu \overset{\text{can. Ens.}}{=} \frac{1}{Z} \sum_{\mu \in \Omega} e^{-\beta E_\mu} E_\mu$

    This can also be computed as derivative of $\ln Z$ $\quad \boxed{\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = \frac{1}{Z}\sum_{\mu \in \Omega} e^{-\beta E_\mu} E_\mu}$

  - Magnetization M

    A state μ has magnetization $M_\mu = \sum_{i=1}^{V} s_i$

    $\langle M \rangle = \sum_{\mu \in \Omega} p_\mu M_\mu = \frac{1}{Z} \sum_{\mu \in \Omega} e^{-\beta E_\mu} M_\mu$

    This again a derivative of $\ln Z$, because $M_\mu = -\frac{\partial E_\mu}{\partial B}$

    $\Rightarrow \boxed{\langle M \rangle = \frac{1}{\beta} \frac{\partial \ln Z}{\partial B}}$

Susceptibilities

* specific heat: change of the average energy under variation of the temperature

$$C = \frac{\partial \langle E \rangle}{\partial T} = \frac{\partial \beta}{\partial T} \cdot \frac{\partial}{\partial \beta}\left(\frac{1}{Z}\sum_{\mu \in \Omega} e^{-\beta E_\mu} E_\mu\right) = -k\beta^2 \left(-\frac{1}{Z}\left(\sum_\mu e^{-\beta E_\mu} E_\mu^2\right) - \frac{1}{Z^2}\frac{\partial Z}{\partial \beta}\left(\sum_\mu e^{-\beta E_\mu} E_\mu\right)\right)$$

$$= k\beta^2 \left(\langle E^2 \rangle - \langle E \rangle^2\right) = k\beta^2 \sigma_E^2$$

$\rightarrow$ C is proportional to the variance of the energy in the ensemble

* magnetic susceptibility: change of the average magnetization under a variance of B

$$\chi_V = \frac{\partial \langle M \rangle}{\partial B} = \beta\left(\langle M^2 \rangle - \langle M \rangle^2\right) = \beta\sigma_M^2$$

- If Z is known analytically, the derivative formular are useful, but typically ensemble observables are estimated numerically using simulations
- E, M, C, χ_V are extensive qualities; i.e. proportional to volume V

  C, χ_V are also proportional to V (not V^2) as they are derivatives of $\langle E \rangle$ and $\langle M \rangle$ with respect to parameters
- Extensive quantities are often computed per spin; i.e. as densities (e.g. $e = \frac{E}{V}$, $m = \frac{M}{V}$, $c = \frac{C}{V}$, $\chi = \frac{\chi_V}{V}$

  → intensive quantities → allows for comparison of observables measured for different V and to keep the result finite as $V \to \infty$
- Partition function is related to the free energy density $f = -\lim \frac{1}{\beta V} \ln Z$
- in 1d the Ising model can be solved analytically:

$$F = -\frac{1}{\beta} \ln\left(e^{\beta J}\cosh\beta B + \sqrt{e^{2\beta J}(\sinh\beta B)^2 + e^{-2\beta J}}\right)$$

for $B=0$: $\langle m \rangle = 0$ for all T: no spontaneous magnetization in 1d Ising model → disordered phase

- in 2d the Ising model can be solved analytically, but it's extremly hard

  The magnetization density (for $V \to \infty$) is: $\lim\limits_{B \to 0} \lim\limits_{V \to \infty} \langle m \rangle = \begin{cases} \pm\left(1 - \sinh^{-4}(2\beta J)\right)^{\frac{1}{8}} & , \beta > \beta_c, \; T < T_c \\ 0 & , \beta < \beta_c, \; T > T_c \end{cases}$

  with $2J\beta_c = \ln(1 + \sqrt{2})$

  - T dependence of magnetization:
    - → high T: system is disordered → paramagnetic phase where $\langle m \rangle = 0$ (for $B=0$)
    - → T lowered: magnetization remains zero till the critical or Curie temperature T_c is reached

      ⇒ second order phase transition
    - → when T is lowered further: magnetization becom non zero, $\langle m \rangle \neq 0$ ($B=0$) → system is in ferromagnetic phase

      ↑ order parameter
    - → phenomen is called spontaneous magnetization: system has a preferred spin direction even without external B-field
- In finite volume: phase transition is weakened to crossover

Note: In infinite volume: when system has choosen a direction it cannot flip back → orientation of spins is fixed

In finite volume: perferred direction can flip during simulation → overall $\langle m \rangle_V = 0$. Flips happen more often for small volume

Practical hint: measure magnetization using $\langle |m| \rangle$ instead $\langle m \rangle$ → insensitive to flips

→ useful information about infinite volume system from simulations at finite V

4.3 Potts model

Classical spin model with q states per spin: each spin s_i can take integer values from $1, \dots, q$

$$H = -J \sum_{\langle ij \rangle} \delta_{s_i s_j}$$

- 2 neighboring spins having same spin-value: energy contribution $-J$
- " " different spin-value: " 0

- The $q=2$ potts model is equivalent to Ising model (with rescaled J and shifted 0-point energy)
- The $q=3$ Potts model in 3d can be used to approximate QCD at finite T with heavy quarks
- in 2d: Potts model has first order phase transition when $q>4$: discontinuity in the $\langle E \rangle$, $\langle m \rangle$
  → jumping of $\langle m \rangle$ at T_c

4.4 Continous spin models (spins have continous values)

Example:

- XY model: spins are 2-dim vectors with length = 1 → represented by the angle θ of point on unit circle
  O(2) model

→ phase transition for $d \geq 3$, Kosterlitz-Thouless transition for $d=2$
dim

- Heisenberg model: spins are 3-dim vectors with Length = 1 → represented by polar and azimuthal angles of point on unit sphere O(3) model

- Generalization: O(n) model → spins are n-dim vectors with length one
- remark: Ising model is the O(1) model, as $s^2 = 1$
- There are many other continous spin models, with other symmetries
- The dim of the spins is not directly related to dim of space: for example the XY model can be studied in 3-d
- Hamiltonian: nearest-neighbor interaction $H = -J \sum_{\langle ij \rangle} \vec{s}_i \cdot \vec{s}_j$

Note: system has a continuous energy spectrum between E_{min} and E_{max}
system has an infinite number of ground states, where all spins are aligned in the same direction

- The partition is a sum over all possible states of system, which are distributed according to some probability distribution
  - Recall: Ising model $Z = \sum_{\mu \in \Omega} e^{-\beta E_\mu}$ sum over all possible states

    Generate all states of Ising model: $\Omega = \prod_{i=1}^{V} \{\uparrow, \downarrow\}_i = \{\uparrow,\downarrow\}_1 \times \{\uparrow,\downarrow\}_2 \times \ldots \times \{\uparrow,\downarrow\}_V \Rightarrow 2^V$ configurations

    The sum over all states for Ising model: $Z = \sum_{S_1 \in \{-1,1\}} \ldots \sum_{S_V \in \{-1,1\}} e^{-\beta H(S_1,\ldots,S_V)}$

    Generalization to continuous space of $O(n)$:

    space of states: $\Omega = \prod_{i=1}^{V} S^{n-1} = S_1^{n-1} \times S_2^{n-1} \times \ldots S_V^{n-1}$

    where S^{n-1} is the $n-1$ dim. unit sphere, and V is number of grid points

    For each spin the sum over all possible values becomes an integral $\underbrace{\sum_{S \in \{-1,1\}}}_{\text{Ising}} \rightarrow \underbrace{\int_{S^{n-1}} d\vec{S}}_{O(n)}$

    The $O(n)$ partition function is: $Z = \int_{S^{n-1}} d\vec{S}_1 \ldots \int_{S^{n-1}} d\vec{S}_V \, e^{-\beta H(\vec{S}_1,\ldots,\vec{S}_V)}$

    Expectation values for $O(n)$ model $\langle M \rangle = \frac{1}{Z} \int_{S^{n-1}} d\vec{S}_1 \ldots \int_{S^{n-1}} d\vec{S}_V \, e^{-\beta H(\vec{S}_1,\ldots,\vec{S}_V)} M(\vec{S}_1,\ldots,\vec{S}_V)$

5 Monte Carlo Methods in statistical physics

5.1 The Monte Carlo method

- Consider the canonical ensemble for Ising model. The partition function represents the probability distribution of all states: $Z = \sum_{\mu \in \Omega} e^{-\beta E_\mu}$

  $\uparrow$ set of all states of Ising system

  An expectation value $\langle Y \rangle = \frac{1}{Z} \sum_{\mu \in \Omega} e^{-\beta E_\mu} Y_\mu$

- Can we compute expectation values explicitly for realistically large system

  e.g. $d=2$, $L=100$, $V=100 \times 100 = 10^4$

  $\rightarrow$ number of configurations: $2^V = 2^{10^4} = 2^{10 \cdot 10^3} = (2^{10})^{1000} \approx (10^3)^{1000} = 10^{3000}$
- Ex? calculation of Z on $\langle Y \rangle$ is prohibitively expensive $\rightarrow$ use stochastic sampling: Monte Carlo Methods
- Fundamental idea of the MC method: replace the set of all states by a much smaller sample of randomly chosen states $\rightarrow$ sample of state: $S_N = \{\mu_t \mid t = 1, 2, \ldots, N\}$ t: Monte Carlo time
- Naively: "uniform sampling" $\rightarrow$ each state μ has equal probability $\frac{1}{2^V}$ (choose every spin on grid to be 1 or -1 with 50/50 probability)

  An expectation value can be estimated as: $\langle Y \rangle \approx \bar{Y} = \dfrac{\frac{1}{N} \sum_{\mu \in S_N} e^{-\beta E_\mu} Y_\mu}{\frac{1}{N} \sum_{\mu \in S_N} e^{-\beta E_\mu}}$

  In general, this is a very bad choice.

Example: 2d - Ising model, $L=100 \rightarrow 10^{3000}$ states

If the sample S_N contains 10^9 states $\rightarrow$ only one state in 10^{2992} is sampled $\rightarrow$ Most sampled states are irrelevant in Z as $e^{-\beta E_\mu}$ will be exponentially suppressed!

- Importance sampling: assume that the states in sample S_N are choosen to some propability p_μ
  $\rightarrow$ the sample estimate is $\langle Y \rangle \approx \bar{Y} = \dfrac{\frac{1}{N}\sum_{\mu \in S_N} \frac{e^{-\beta E_\mu}}{p_\mu} Y_\mu}{\frac{1}{N}\sum_{\mu \in S_N} \frac{e^{-\beta E_\mu}}{p_\mu}}$ (divide by p_μ as states in S_N are chosen according to p_μ
  $\hookrightarrow \sum_\mu p_\mu \frac{e^{-\beta E_\mu}}{p_\mu}$

- Aim: chose p such that $\bar{Y}$ is a good approximation to $\langle Y \rangle$ for $N \ll 2^V$
- Preferred choice of p_μ? Choose the states μ according to the probability with which they enter the partition function
  i.e. for the canonical ensemble $\boxed{p_\mu = \frac{e^{-\beta E_\mu}}{Z}}$
  $\Rightarrow \langle Y \rangle \approx \bar{Y} = \frac{1}{N} \sum_{\mu \in S_N} Y_\mu$ $\Rightarrow$ standard. error: $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}} \approx \frac{S_Y}{\sqrt{N}}$ (for N independent states in sample)

Problem: How do we sample states μ according to a certain p_μ?
$\rightarrow$ construct a Markov process with equilibrium p

5.2 Markov processes and Markov chains:

$$\mu(t) \rightarrow \mu(t+1)$$

Transition probabilities $P(\mu \rightarrow \nu)$, $\forall_{\mu,\nu\in\Omega}$

$$\sum_{\nu\in\Omega} P(\mu \rightarrow \nu) = 1 \ , \quad \forall_{\mu,\nu\in\Omega} \qquad (1)$$

Note:

- $P(\mu \rightarrow \nu)$ can be $\neq 0$: state at time t + 1 can be same as state at time t
- $P(\mu \rightarrow \nu)$ can be $= 0$: for some (or many) states ν —> not all states ν can be reached from all states μ in one Markov step
- $P(\mu \rightarrow \nu)$ only depends on the states μ and ν
- $P(\mu \rightarrow \nu)$ is independent of the Markov time t : i.e., independent of position in Markov chain
- Consecutive states in Markov chain are correlated : autocorrelations
- Define $\omega_\mu(t)$ as the probability to find the system in state μ after a time t
- The probabilities for all states $\mu \in \Omega$ at time t can be represented by the state vector $\vec{\omega}(t)$ (length 2^V for Ising model) with $\sum_{\mu\in\Omega} \omega_\mu(t) = 1$, $\forall_t$ (follows from eq.1)
- Start from some initial configuration μ_0 ; e.g. an initial state vector $\vec{\omega}(0)$ with $\omega_\mu(0) = \delta_{\mu,\mu_0}$
- Probability to find the system in a state μ at time $t+1$

$$\omega_\mu(t+1) = \sum_{\nu\in\Omega} \omega_\nu(t) P(\nu \rightarrow \mu)$$

Or in matrix notation with $P_{\mu\nu} \equiv P(\nu \rightarrow \mu)$

$$\vec{\omega}(t+1) = P \,.\, \vec{\omega}(t)$$

i.e. the state vector $\vec{\omega}(t)$ is transformed in a new state vector $\vec{\omega}(t+1)$, and in general $\vec{\omega}(t) \neq \vec{\omega}(t+1)$, unless in equilibrium

- Relation between Markov process and a Markov chain: a Markov chain is a particular realization of the Markov process
- Repeated Markov chains would reproduce the distribution $\vec{\omega}(t)$ at each t

5.3 Markov process and its equilibrium

Equilibrium: transition from t to $t+1$ leaves all ω_μ unchanged and equals to the equilibrium distribution p_μ

$$\Rightarrow \omega(t+1) \overset{!}{=} \omega_\mu(t) \overset{!}{=} p_\mu \quad \forall_{\mu\in\Omega}$$

Thus $\forall_{\mu\in\Omega}$: $\omega_\mu(t+1) = \sum_{\nu\in\Omega} \omega_\nu(t) P(\nu \to \mu)$

⇒ Equilibrium condition: $\sum_{\nu\in\Omega} p_\nu P(\nu \to \mu) = p_\nu$

Matrix notation—> $P\,.\vec{p} = \vec{p}$

i.e. equilibrium distribution of a Markov process is the eigenvector of the transition matrix P with eigenvalue 1 when the distribution is $\vec{p}$ at time t, it stays in $\vec{p}$ at all times!

- For Monte Carlo simulations we want the Boltzmann distribution $\vec{p}$ to be the equilibrium of the Markov process
- Ergodicity algorithm can each any state of the system starting form any other state in a finite number of Markov steps
- This is necessary to reach the Boltzmann distribution: If a configuration μ cannot be reached then $p_\mu = 0$, but it should be $p_\mu = \dfrac{e^{-\beta E_\mu}}{Z}$
- We want more: $\vec{p}$ should not only be equilibrium of Markov process, but should also be its fixed point: i.e. Markov process should converge to $\vec{p}$ starting from any initial state vector $\vec{\omega}(0)$

$$\lim_{t\to\infty} \vec{\omega}(t) = \vec{p}$$

- Lemma: If the algorithm is ergodic and satisfies equilibrium condition $\left(P\,.\vec{p} = \vec{p}\right)$, then the Boltzmann distribution will be obtained as fixed - point of the process

Proof:

Define distance between states vectors as:

$$\|\vec{\omega} - \vec{p}\| = \sum_{\mu\in\Omega} |\,\omega_\mu - p_\mu\,| \quad \hat{L}\text{-norm}$$

- At Markov time $t+1$ the distance form the equilibrium distribution is:

$$\begin{aligned}
\left\| \vec{\omega}(t+1) - \vec{p} \right\| &= \sum_{\mu\in\Omega} \left|\omega_\mu(t+1) - p_\mu\right| = \sum_{\mu\in\Omega} \left| \sum_{\nu\in\Omega} P(\nu\to\mu)\left(\omega_\nu - p_\nu\right) \right| \qquad (\vec{p}\text{ is equilibrium}) \\
&\leq \sum_{\mu\in\Omega}\sum_{\nu\in\Omega} P(\nu\to\mu)\left|\omega_\nu(t) - p_\nu\right| \\
&= \sum_{\nu} \left|\omega_\nu(t) - p\nu\right| = \left\| \vec{\omega} - \vec{p} \right\| \\
\Rightarrow \left\| \vec{\omega}(t+1) - \vec{p} \right\| &\leq \left\| \vec{\omega}(t) - \vec{p} \right\|
\end{aligned}$$

—> The Markov process brings the state vector closer and closer to the equilibrium distribution:
—> $\vec{p}$ is ficht point of the Markov process

—> starting from any state μ_0 the distribution $\vec{p}$ re reached <u>often</u> in the same <u>equilibration time</u> (thermalization)

- Equilibration happens exponentially fast: When equilibrium is reached, the state vector remains unchanged in further Markov steps. The set of configurations generated <u>after equilibrium</u> reproduce the equilibrium ensemble and can be used to compute observables of the system in thermal equilibrium
- Markov chain: after equilibration the state are sampled according to the equilibrium distribution
  —> MC time average after equilibration $\longrightarrow$ <u>ensemble average</u>
- <u>Problem:</u> Construct a transition matrix $P(\mu, \nu) \quad (\forall_{\mu,\nu\in\Omega})$ which satisfies $P \,.\, \vec{p} = \vec{p}$ and ergodicity
- In practice: most algorithms satisfies <u>detailed balance</u>, which is a stronger condition then $P \,.\, \vec{p} = \vec{p}$

$$P(\mu \to \nu)p_\mu = P(\nu \to \mu)p_\nu \ , \qquad \forall_{\mu,\nu\in\Omega}$$

Equilibrium follows from detailed balance (DB): sum detailed balance equation over ν:

$$\begin{aligned}\sum_{\nu\in\Omega} P(\mu \to \nu)p_\mu &= \sum_{\nu\in\Omega} P(\nu \to \mu)p_\nu \\ \Rightarrow p_\mu &= \sum_{\nu\in\Omega} P(\nu \to \mu)p_\nu\end{aligned}$$

- DB is a sufficient but not necessary condition to each the correct equilibrium distribution $\vec{p}$
- DB avoids „limit cycles“ —> fixed point is always reached from any initial state μ_0
- The transition probabilities $P(\mu \to \nu)$ are not uniquely determined by DB but have to satisfy

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{p_\nu}{p_\mu} \overset{canoncial}{=} e^{-\beta(E_\nu - E_\mu)} \qquad \text{(DB)}$$

- Different MC algorithms differ in there choice of $P(\mu \to \nu)$ satisfying DB

8.2 Error propagation for functions of primary qualities

Consider a secondary quality $f(\langle y_1 \rangle, \ldots, \langle y_n \rangle)$ estimated by $f(\bar{y}_1, \ldots, \bar{y}_n)$. What is the standard error on f, i.e. what is the standard deviation of $f(\bar{y}_1, \ldots, \bar{y}_n)$ in repeated samplings?

Notation: set $\langle y_k \rangle \equiv \mu_k$

Standard error on f: standard deviation σ_f of $f(\bar{y}_1, \ldots, \bar{y}_n)$

- Te variance of $f(\bar{y}_1, \ldots, \bar{y}_n)$ under repeated sampling is:

$$\sigma_f^2 = \left\langle \left(f(\bar{y}_1, \ldots, \bar{y}_n - \langle f(\bar{y}_1, \ldots, \bar{y}_n \rangle \right)^2 \right\rangle$$

- Taylor expansion of $f(\bar{y}_1, \ldots, \bar{y}_n)$ around $(\mu_1, \ldots, \mu_n)$:

$$f(\bar{y}_1, \ldots, \bar{y}_n) = f(\mu_1, \ldots, \mu_n) + \sum_{k=1}^{n} f_k \Delta \bar{y}_k + \Theta(\Delta)$$

with $\Delta \bar{y}_k \equiv \bar{y}_k - \mu_k$

and $f_k \equiv \frac{\partial f}{\partial \mu_k} \Big|_{(\mu_1, \ldots, \mu_n)}$

- First order Taylor expansion is justified if we assume small standard errors $\sigma_{\bar{y}_k}$ on the primary qualities $\bar{y}_k$
- Compute expectation value of $f(\bar{y}_1, \ldots, \bar{y}_n)$ over the repeated sampling.

$$\left\langle f(\bar{y}_1, \ldots, \bar{y}_n) \right\rangle \cong f(\mu_1, \ldots, \mu_n) + \mathcal{O}(\langle \Delta^2 \rangle) \rightarrow \text{ negligeable}$$

- To leading order of variance σ_f^2 is:

$$\sigma_f^2 = \left\langle \left(\sum_{k=1}^{n} f_k \Delta \bar{y}_k \right)^2 \right\rangle$$ —> depends on variance $\langle (\Delta y_k) \rangle$

and on covariance $\langle \Delta \bar{y}_k \Delta y_l \rangle) = \langle \Delta(\bar{y}_k \Delta - \mu_k)(\Delta \bar{y}_l - \mu_l) \rangle$

- How can this and propagation be used in the case of autocorrelated measurements in a Markov chain?
- Define an <u>effective observable</u> g, which can be measured on each configuration of the Markov chain:

$$g(y_1, \ldots, y_n) = \sum_{k=1}^{n} f_k y_k \qquad \text{with constants} \qquad f_k \equiv \frac{\partial f}{\partial \mu_k} \Big|_{(\mu_1, \ldots, \mu_n)}$$

Rewrite σ_f^2 as:

$$\sigma_f^2 = \left\langle \left(g(\bar{y}_1, \ldots, \bar{y}_n) - g(\mu_1, \ldots, \mu_n) \right)^2 \right\rangle$$

- Because g is a <u>linear</u> function of measurements $g(\langle y_1 \rangle, \ldots, \langle y_n \rangle) = \langle g(y_1, \ldots, y_n) \rangle$

—> a linear function of expectation values is an expectation value of function applied to sample measurements and is a primary quantity!

Proof: $$g(\langle y_1\rangle, \ldots, \langle y_n\rangle) = \sum_{k=1}^{n} f_k \langle y_k\rangle = \left\langle \sum_{k=1}^{n} f_k y_k \right\rangle = \langle g(y_1, \ldots, y_n)\rangle$$

The same hods for sample means:

$$g(\bar{y}_1, \ldots, \bar{y}_n) = \overline{g(y_1, \ldots, y_n)}$$

⇒ We find:

$$\sigma_f^2 = \left\langle \left(\overline{g(y_1, \ldots, y_n)} - \langle g(y_1, \ldots, y_n)\rangle\right)^2 \right\rangle$$

—> This is exactly the definition of standard error on the primary quantity $g(y_1, \ldots, y_n) = \sum_{k=1}^{n} f_k y_k$

—> to compute the standard error σ_f on $f(\bar{y}_1, \ldots, \bar{y}_n) \longrightarrow$ compute the standard error $\sigma_{\bar{g}}$ on the primary quantities $\bar{g}$ taking into account the integrated autocorrelation time on g:

$$\sigma_f \simeq \sigma_{\bar{g}} = \sqrt{\frac{2\tau_{int,g}}{N}}\sigma_g \qquad \text{with} \qquad \tau_{int,g} = \frac{1}{2} + \sum_{t=1}^{N} \rho_g(t)\left(1 - \frac{t}{N}\right)$$

- To construct g we need estimates for f_k from a single chain —> replace expectation values μ_k by sample means $\bar{y}_k$

$$f_k \equiv \frac{\partial f}{\partial \mu_k}\bigg|_{(\mu_1, \ldots, \mu_n)} \approx \frac{\partial f}{\partial \mu_k}\bigg|_{(\bar{y}_1, \ldots, \bar{y}_n)}$$

- This error propagation automatically takes into account both the correlations between different observables AND the autocorrelations in the Markov chain
- Apply this to specific heat in Ising model

$$c = \frac{k\beta^2}{V}\left(\langle E^2\rangle - \langle E\rangle^2\right) \qquad \Rightarrow \; y_1 = E^2 \quad y_2 = E \quad | \quad \mu_1 = \langle E^2\rangle \quad \mu_2 = \langle E\rangle$$

8.3 Blocking method

- Consider an equilibrated sample S of N configurations (autocorrelated)

  $S = \{\mu_t \,|\, t = 1, \ldots, N\}$ (t Markov time, μ_t configuration at time t after equilibration)

  and a secondary quantity $Q(S)$, e.g. the specific heat

- Divide the N configurations in n_b blocks S_i , $i = 1, \ldots, n_b$ each of block size N_i and measure $Q(S_i) = Q_i$ on each block
- If the blocks are large enough ($N_i \gg \tau_{int,i}$) the measurements Q_i on successive blocks are uncorrelated

- Apply the standard error for the block measurements Q_i to get an estimate for standard error on $Q(S)$

$$\sigma_Q = \frac{\sigma_{block}}{\sqrt{n_b}}$$

with variance $\sigma^2_{block} = \frac{1}{n_b}\sum_{i=1}^{n_b}(Q_i - \overline{Q}_{block})^2$ with $\overline{Q}_{block} = \frac{1}{n_b}\sum_{i=1}^{n_b} Q_i$

Result for Q: $Q = Q(S) \pm \sigma_Q$

^ (the estimate for Q is $Q(S)$ <u>and not</u> $\overline{Q}_{block}$)

- Typically n_b is between 10 and 20
- Advantage: no need to compute autocorrelation times
- Disadvantage: error estimate is not accurate and depends on n_b
- Danger: if autocorrelation is large —> method is wrong
- Method works: because it kind of mimics the definition of standard error: standard deviation over repeated samples (=blocks)

$$\longrightarrow \frac{1}{\sqrt{n_b N_i}}$$

<u>8.4 The Jackknife method</u>

- Consider a sample S of N <u>independent</u> configurations

$$S = \{\mu_t \,|\, t = 1,\ldots,N\}$$

- Consider a secondary quantity $Q(S)$
- Construct N <u>jackknife samples</u> of $N-1$ configurations by removing each configuration of the original sample

$$S_i = \{\mu_1\ldots,\mu_{i-1},\mu_{i+1},\ldots,\mu_N\} \quad \forall_{i=1,\ldots,N}$$

Computes N jackknife measurements of the secondary quantity:

$$Q_i \equiv Q(S_i)$$

- Jackknife estimate of standard error on $Q(S)$

$$\sigma_Q = \sqrt{\sum_{i=1}^{N}(Q_i - \overline{Q}_{Jack})^2} \quad \text{with} \quad \overline{Q}_{Jack} = \frac{1}{N}\sum_{i=1}^{N} Q_i$$

Result for Q; $Q = Q(S) \pm \sigma_Q$

- σ_Q : has no explicit $\frac{1}{\sqrt{N}}$, but σ_Q still decreases with $\frac{1}{\sqrt{N}}$ if N increases

- Example: take a primary quantity a secondary quantity

  ↳ assume $Q(S) = \bar{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_i$

  Jackknife measurements: $Q_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} Y_j$

  Jackknife average: $\bar{Q}_{Jack} = \frac{1}{N} \sum_{i=1}^{N} Q_i = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} Y_j \overset{+Y_i - Y_i}{=} \frac{1}{N-1} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} Y_j - Y_i \right)$

  The jackknife standard error²:
  $$\sigma_Q^2 = \sum_{i=1}^{N} (Q_i - \bar{Q}_{Jack})^2 = \sum_{i=1}^{N} \left(\frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} Y_i - \bar{Y} \right)^2 \overset{+Y_i - Y_i}{=} \sum_{i=1}^{N} \left(\frac{1}{N-1} \sum_{j=1}^{N} Y_j - \frac{Y_i}{N-1} - \bar{Y} \right)^2$$
  $$= \sum_{i=1}^{N} \left(\frac{N}{N-1} \bar{Y} - \frac{Y_i}{N-1} - \bar{Y} \right)^2 = \sum_{i=1}^{N} \left(\frac{\bar{Y}}{N-1} - \frac{Y_i}{N-1} \right)^2 = \frac{1}{(N-1)^2} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$$
  $$= \frac{S_Y^2}{N-1} \approx \frac{\sigma_Y^2}{N}$$
  standard error for primary quantities

- Jackknife method assumes uncorrelated configurations in sample
  - → Autocorrelations from Markov process are removed as following: compute τ_{int} using some "appropriate" observable and then: starting from S_i make a smaller sample with $N_{indep} = \frac{N}{2\tau}$ independent configurations

    compute the jackknife error on this smaller sample

8.5 The bootstrap method

- Assume again that the secondary quantity Q is estimated by $Q(S)$ computed on a sample S of independent configurations, with

$$S = \{\mu_t \mid t = 1, \dots, N\}$$

?

Generate M pseudosamples S_i, $i = 1, \dots, M$, where a pseudosample S_i is generated by randomly choosing N configurations out of the N configurations of S, with replacement, i.e. allowing for duplicates
In a pseudosample some configurations will occur more than once, while others will be absent
- Compute the M pseudo-measurements on the M pseudosamples:

$$Q_i = Q(S_i)$$

The bootstrap standard error on $Q(S)$ is the standard deviation of the pseudo measurements:

$$\sigma_Q = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(Q_i - \bar{Q}_B\right)^2} \qquad \text{with bootstrap average } \bar{Q}_B = \frac{1}{M} \sum_{i=1}^{M} Q_i$$

Result for Q: $Q = Q(S) \pm \sigma_Q$
- Bootstrap formula for σ_Q is different from Jackknife formula
- In bootstrap: N is sample size, which determines the size of the standard error, while M determines the accuracy of the standard error
- Standard error decreases with increasing N (as $\frac{1}{\sqrt{N}}$), but not with M
- in practice: $M = 1000 \rightarrow \sigma_Q$ is determined with $\approx 3\%$ accuracy
- For large N ($N \gg M$): bootstrap faster than Jackknife
- Bootstrap is based on assumption that the equilibrated Markov chain is a coarse rendering of the full Boltzmann distribution and the pseudo-samples are repeated samples on this coarse distribution $\rightarrow$ bootstrap error follows the definition of standard error on measurements made on coarse distribution
- What about autocorrelations? $\rightarrow$ compute τ_{int} from some appropiate observable
  - starting from S, make a smaller sample with $N_{indep} = \frac{N}{2\tau}$ independent configurations compute bootstrap error on smaller sample

    or easier:

    $\rightarrow$ use all N config. of S, but only draw $\frac{N}{2\tau}$ random config. to construct each pseudosample
    $\rightarrow$ pseudosamples contain $\frac{N}{2\tau}$ independent config. $\rightarrow$ compute bootstrap error on the smaller pseudosamples

9. Heatbath algorithm

Principle: consider a reduced system of single spin in thermal equilibrium in a fixed background —> local heatbath

Recipe: choose a lattice i with spin s_i , and choose new values s_i' according to a local Boltzmann distribution at temperature T in a fixed background of all other spins s_j , $i \neq j$

- Assume that the current configuration is $\mu = (\bar{\mu}, s_i)$ where $\bar{\mu}$ consists of all spins s_j where $i \neq j$. Choose a new spin values s_i' at site i according to the probability:

$$P_{local}(s_i \rightarrow s_i' | \bar{\mu}) = p_{hb}(s_i' | \bar{\mu}) = \frac{e^{-\beta E(\bar{\mu} s_i')}}{\sum_{s_i'} e^{-\beta E(\bar{\mu} s_i')}}$$

With: $\sum\limits_{s_i'} P_{local}(s_i \rightarrow s_i' | \bar{\mu}) = 1$ and $P_{local} > 0 \longrightarrow$ local ergodicity

- Global transition probability from any $\mu = (\bar{\mu}, s_i)$ to any $\nu = (\bar{\nu}, s_i')$:

$$P(\mu \rightarrow \nu) = P_{local}(s_i \rightarrow s_i' | \bar{\mu}) \delta_{\bar{\mu}\bar{\nu}} = P_{hb}(s_i' | \bar{\mu}) \delta_{\bar{\mu}\bar{\nu}}$$

- Rewrite the Boltzmann probability as follows:

$$P(\mu) = \frac{e^{-\beta E(\bar{\mu} s_i)}}{\sum\limits_{\mu} e^{-\beta E(\mu)}} \times \frac{\sum\limits_{s_i} e^{-\beta E(\bar{\mu} s_i)}}{\sum\limits_{s_i} e^{-\beta E(\bar{\mu} s_i)}} = P_{hb}(s_i | \bar{\mu}) P_{bg}(\bar{\mu}),$$

where $P_{bg} = \frac{\sum\limits_{s_i} e^{-\beta E(\bar{\mu} s_i)}}{\sum\limits_{\mu} e^{-\beta E(\mu)}} = \sum\limits_{s_i} p(\bar{\mu}, s_i)$ —> probability of the background $\bar{\mu}$ in the ensemble

Detailed balance?

$$\begin{aligned} P(\mu \rightarrow \nu) p(\mu) &= P_{hb}(s_i' | \bar{\mu}) \delta_{\bar{\mu}\bar{\nu}} P_{hb}(s_i | \bar{\mu}) P_{bg}(\bar{\mu}) \\ &= P_{hb}(s_i | \bar{\nu}) \delta_{\bar{\mu}\bar{\nu}} P_{hb}(s_i' | \bar{\nu}) P_{bg}(\bar{\nu}) \\ &= P(\nu \rightarrow \mu) p(\nu) \longrightarrow \underline{\text{DB is staisfied!}} \end{aligned}$$

- Ergodicity: each lattice site i is chosen with equal probability and at this site any spin value can be reached —> any state of system can b e reached from any other state of the system in a finite time
- Repeat the local heatbath V times to perform one ˝sweep˝ and then perform many sweeps to construct a Markov chain
- Just like for Metropolis one has to equilibrate the chain before making final measurements, and autocorrelation have to be taken into account, when computing standard errors on primary and secondary quantities
- Example: Ising model
  - Generate configurations by successive single - spin updates

- For one spin —> two possible spin values ↑ or ↓ after the update with local Boltzmann probabilities:

$$P(\uparrow) = \frac{e^{-\beta E(\bar{\mu},\uparrow)}}{e^{-\beta E(\bar{\mu},\uparrow)} + e^{-\beta E(\bar{\mu},\downarrow)}}, \quad P(\downarrow) = \frac{e^{-\beta E(\bar{\mu},\downarrow)}}{e^{-\beta E(\bar{\mu},\uparrow)} + e^{-\beta E(\bar{\mu},\downarrow)}}$$

- $P(\uparrow) + P(\downarrow) = 1$

- Probabilities for the new Spin values are independent of the current spin value, it only depends on the background spins and on the new value
- Choose the new spin value according to $P(\uparrow)$ and $P(\downarrow)$
  $\longrightarrow$ generate a uniform random number $r \in [0,1)$. If $r < P(\uparrow)$ than the new spin is ↑, else it is ↓

- H is local → to compute $P(\uparrow)$ one only needs to consider the spin s_i and its nearest neighbors The energy contribution inside the background cancels between the numerator and denominator in $P(\uparrow)$ and $P(\downarrow)$ $\longrightarrow$ fast computation
- Repeat the single spin update until V spin updates habe been considered —> one sweep —> new config. in in Markov chain
- Ising model; Metropolis is somewhat more efficient than heatbath

Compute the probability that the current spin is flipped.

Assume $\Delta E = E_{new} - E_{old}$

Metropolis $\quad P_{new} = min(1, e^{-\beta \Delta E})$

Heatbath $\quad P_{new} = \dfrac{e^{-\beta E_{new}}}{e^{-\beta E_{new}} + e^{-\beta E_{old}}} = \dfrac{1}{1 + \frac{e^{-\beta E_{old}}}{e^{-\beta E_{new}}}} = \dfrac{1}{1 + e^{\beta \Delta E}}$

- For Ising model: the acceptance of spin flip:

$A_{HB} < A_{Metropolis} \rightarrow$ HB has longer autocorrelation —> HB somewhat less efficient

But: For Potts model (e.g. with q = 10): HB is way more efficient than Metropolis.

- In general it can be quite expansive to generate s_i' according to the local Boltzmann distribution $P_{hb}(s_i' | \bar{\mu})$ if at all possible
- Metropolis is simpler and always applicable
  - In general slower convergence to equilibrium and lower autocorrelation times
  - Updates are very cheep (for local Hamiltonians)

10. Monte Carlo simulations at phase transitions

- Consider Ising model the Ising model in 2 dimensions:
  at the critical temperature $kT_c = \dfrac{2J}{\ln(1+\sqrt{2})} \approx 2.269J \quad (\text{for } V \to \infty)$
  => phase transition from the disordered phase (paramagnetic $\langle M \rangle = 0$) to the ordered phase (ferromagnetic $\langle M \rangle \neq 0$) → spontaneous magnetization

- What happens when we approach T_c from above: $T \to T_c^+$?
- At high T the spins are random and uncorrelated
- At somewhat lower temperatures, but still $T > T_c$, neighboring spins tend to align → domains of aligned spin are forming
- Consider the spatial correlation function, i.e. the two-point connected correlation function of the spin s between two sites i and j :

$$G(i,j) \equiv \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle = \left\langle \left(s_i - \langle s_i \rangle \right) \left(s_j - \langle s_j \rangle \right) \right\rangle$$

- In practice: $G(i,j) = \langle s_i s_j \rangle - \langle |m| \rangle^2$ at finite volume
- $G(i,j)$ is defined for any pair of lattice sites (i,j)
- Translational invariance and invariance over $\dfrac{\pi}{2}$ rotation (on square lattice)

  $\to G(r) \equiv G(i,j)$ where i and j are on same horizontal or vertical line

  with $r = r_{i,j} = |\vec{r}_i - \vec{r}_j|$

- If there are m_r pairs of lattice sites (i,j) with same vertical or horizontal separation r, we can write:

$$G(r) = \sum_{\substack{(i,j) \\ r_{i,j} = r}} G(i,j)$$

- Note: $r_{i,j}$ is the shortest distance between i and j → can wrap over the boundary because of the periodic lattice
- For large separations r :

  $G(r) \overset{large\ r}{\sim} \exp(-\dfrac{r}{\xi})$ where the correlation length ξ is typical size of the domains, which depends on T

- For $T \to T_c$ the correlation length $\xi \to \infty$ for $L \to \infty$
- For finite size L → maximal correlation length $\xi \propto L$
- At $T = T_c$: the system spontaneously chooses a preferred direction into which most spins point, such that $\langle M \rangle \neq 0$
- For $B = 0$ this is random and depends on random thermal fluctuations during the phase transition

- Critical fluctuations:
  - For $T \approx T_c^+$ the domains are very large and regularly flip as $\langle M \rangle = 0$ but ξ is large
  - When they flip → large fluctuations in M and E → „critical fluctuations"

  ⟶ large standard errors on measurements ($\propto \sigma_E, \sigma_M$) → need more measurements → longer run times

- In MC algorithms (e.g. Metropolis, heatbath) an additional effect occurs:
  - Inside domains, the probability to flipp spins is extremely low → algorithm only nibbles at the domain borders
  - Autocorrelation time τ increases dramatically as $T \to T_c$ and diverges as $L \to \infty$
  - Independent measurements require a time 2τ in Markov chain

    ⟹ „Critical slowing down" of the algorithm close to T_c

- Two negative effects for simulations close to T_c . As $\sigma_{\bar{y}} = \sqrt{\frac{2\tau}{N}}\sigma_y$:
  1. Critical fluctuations $\propto \sigma_y$: property of the system which cannot be changed
  2. Critical slowing down: property <u>of</u> the local update algorithm → find better algorithm!

- At T_c :
  - Autocorrelation time: $\tau \propto \xi^z$ where z is dynamical exponent at algorithm
  - For finite L: $\tau \propto L^z$
  - Large z → strong critical slowing down

- For Ising model in 2d: Metropolis algorithm has $z = 2.17$

  ⟹ CPU time for z sweeps $\sim V\tau \sim L^{2+z} \sim L^{4.17}$
- <u>Solution:</u> use <u>cluster algorithms</u> instead of local updates

  <u>Idea:</u> flip all the spins in a cluster at once, instead of flipping single spins

11. Cluster algorithms

11.1 Wolf algorithm for the Ising model

1. Consider the current configuration in the Markov chain. Randomly choose a site i (sometimes called the seed site). The spin s_i of this site determines the cluster orientation
2. Consider in turn each nearest neighbor j of site i :
   - If $s_j \neq s_i \rightarrow$ do not add j to the cluster
   - If $s_j = s_i \rightarrow$ add site j to the cluster with probability $P_{add} = 1 - e^{-2\beta\tau}$

   ⟹ Do this by choosing a uniformly distributed random number r between 0 and 1 and add the site j to the cluster if $r \leq P_{add}$
3. Repeat step **2.** for each new site in the cluster until no new sites are added
4. Flip the complete cluster → new configuration in the Markov chain (beware: this is not a sweep!)
5. Repeat from step **1.** to construct the next cluster with a new seed site

Remarks about step **2.** :

- If a neighbor already belongs to the cluster, it is not considered again (never kick a spin out of the cluster). Implement this by immediately flipping the spin when it is added to the cluster!
- If a neighbor was already considered before but failed the P_{add} test, it gets another chance to get added to the cluster

Proof of detailed balance:

- Consider two configurations μ and ν, which differ exactly by one cluster flip (otherwise $P(\mu \rightarrow \nu) = P(\nu \rightarrow \mu) = 0$)

(a) Consider the transition $\mu \rightarrow \nu$
- Choose the seed site
- The other sites of the clusters are added according to their orientation and to P_{add}
- Some neighboring sites with right orientation have been rejected

(b) Reverse transition $\nu \rightarrow \mu$
- Choose the seed site with same probability as in (a)
- Add the other chosen sites with the same probability as in (a)
- Again some sites have the right orientation (along the outer border of the cluster) but have been rejected → these sites are different from those in (a)

- Probability to construct this specific cluster in μ and ν

The difference between the constructions (a) and (b) is only given by the sites along the outer edge of the cluster that have the same orientation as the cluster but failed the accept - reject test once or more times.
When the cluster is flipped attractive interactions along the border become repulsive after cluster flipp and vice-versa.

(a) Assume that there are m attractive interactions along the outer cluster edge in configuration μ (in the example $m = 5$).

⟹ Probability to form the cluster in μ and get ν after the flip:

$$P(\mu \rightarrow \nu) \propto (1 - P_{add})^m$$

(b) Similarly, there are n attractive interactions along the outer cluster edge in ν (in the example: $n = 13$)

$$P(\nu \to \mu) \propto (1 - P_{add})^n$$

- When flipping the cluster from μ to ν, the energy change is:

  $+2J$ for each of the m attractive interactions that turn repulsive
  $-2J$ for each of the n repulsive interaction that turn attractive

  $$\Rightarrow \Delta E = E_\nu - E_\mu = 2J(m - n)$$

- Detailed balance requires that: $(1 - P_{add})^m e^{-\beta E_\mu} \stackrel{!}{=} (1 - P_{add})^n e^{-\beta E_\nu}$
  Note: all other factors in P are identical and cancel

  $$\Rightarrow (1 - P_{add})^{m-n} \stackrel{!}{=} e^{-\beta(E_\nu - E_\mu)} \stackrel{!}{=} e^{-2\beta J(m-n)}$$
  $$\Rightarrow 1 - P_{add} \stackrel{!}{=} e^{-2\beta J}$$
  $$\Rightarrow \underline{P_{add} \stackrel{!}{=} 1 - e^{-2\beta J}} \longrightarrow$$ explains the choice of P_{add} in Wolf algorithm

Ergodicity:

- There is a nonzero probability for any site to be seed and a non zero probability that the cluster has size 1
  ⟹ A single spin is flipped
  ⟹ Any configuration can be reached from any configuration in finite time

Properties of Wolf algorithm

- Slide of cluster density versus T : the lower the temperature, the larger the clusters
- High temperature: $\left(e^{-2\beta J} \to 1 \ \Rightarrow \underline{P_{add} \to 0}\right)$
  ⟹ Cluster size $\to 1$ $\longrightarrow$ single-spin-flips
  ⟹ Wolf alg. similar behavior as Metropolis at high T
  (Wolf alg. is somewhat slower as cluster construction has to be generated)

- Low temperature ($T \ll T_c$) compare Metropolis and Wolf:

1. Metropolis:
- Most spins pint in same direction (assume $\downarrow$)
  a few excited spins point in appropriate direction ($\uparrow$)
- during a sweep:
  a) excited spins are flipped form $\uparrow$ to $\downarrow$ as $\Delta E < 0$
  b) Occasionally $\downarrow$ - Spin gets exited by thermal excitation
  ($\Delta E > 0 \to$ Metropolis accept/reject test)

  ⟹ mimics the dynamical behavior of physical systems

⇒ on average the exited spins are relaxed after 1 sweep, new independent excitations are generated

→ at very low T: one independent conic per sweep:

2. <u>Wolf algorithm at low T:</u>

When $T \to 0 \Rightarrow P_{add} \to 1$

- Assume thermalized configuration where most spins point ↓ , then seed spin is probably ↓
- Most spins are ↓ and are added to cluster with high probability P_{add}