The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization
IAEA

1929-1

Advanced School on Quantum Monte Carlo Methods in Physics and Chemistry

21 January - 1 February, 2008

Intro on random walks and error analysis.

D. Ceperley
University of Illinois at Urbana-Champaign

Basics of the Monte Carlo Method

- General properties of simulations and Monte Carlo
- Markov Chains
- Error analysis

Lab today will use do some variational Monte Carlo and error analysis.

Simulation is advantageous for high dimensional integrals -the best general method

Consider an integral in the unit *D-dimensional* hypercube:

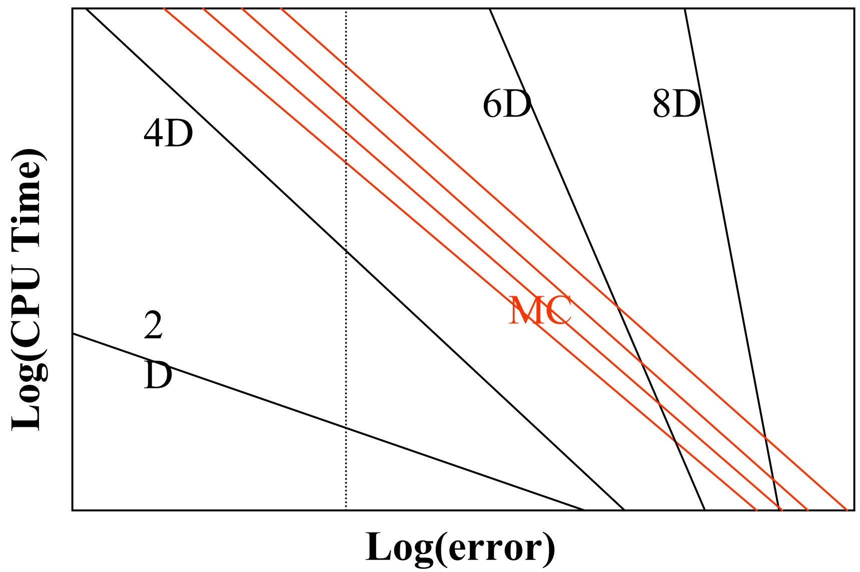
$$I = \int dx_1 \ldots dx_D \quad f(x_1, \ldots, x_D)$$

By conventional deterministic methods:

- Lay out grid with L points in each direction: $h=1/L$
- Number of points is $N = L^D = h^{-D} \propto$ CPU time.

How does error scale with CPU time?

- Error in trapezoidal rule goes as $\varepsilon = f''(x)\, h^2$ since
- Hence CPU time $\propto \varepsilon^{-D/2}$ since $\varepsilon \sim h^2$ and $CPU \sim h^{-D}$
- But by MC: ε^{-2} since $\varepsilon \sim M^{-1/2}$ and $CPU \sim M$



MC is advantageous for D>4!

Quantum Monte Carlo Integrals

What integrals do we need to do?
NB They are all un-normalized

$$R \equiv \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$$

$$\hat{P}R \equiv \{\mathbf{r}_{P1}, \mathbf{r}_{P2}, \ldots, \mathbf{r}_{PN}\}$$

$$\langle \hat{O} \rangle = \frac{\int dR \pi(R) O(R)}{\int dR \pi(R)}$$

METHOD	$\pi(\mathrm{R})$
Classical MC	$\exp(-\beta V(R))$
Variational MC	$\lvert \Psi_T(R) \rvert^2 = \lvert \det(R) \rvert^2 e^{-2U(R)}$
Diffusion MC	$\Psi_T(R) \langle R \lvert e^{-\tau H} \rvert R' \rangle \Psi_T(R')^{-1}$
Reptation MC	$\Psi_T(R_0) \langle R_0 \lvert e^{-\tau H} \rvert R_1 \rangle \langle R_1 \lvert e^{-\tau H} \rvert R_2 \rangle \ldots \langle R_{p-1} \lvert e^{-\tau H} \rvert R_p \rangle \Psi_T(R_p)$
Path Integral MC	$\langle R_0 \lvert e^{-\tau H} \rvert R_1 \rangle \langle R_1 \lvert e^{-\tau H} \rvert R_2 \rangle \ldots \langle R_{p-1} \lvert e^{-\tau H} \rvert R_p \rangle \langle R_{p-1} \lvert e^{-\tau H} \rvert \hat{P} R_0 \rangle$

Simulations

- What is a simulation?
  - It has an internal state "S"
    - In classical mechanics, the state = positions $\{q_i\}$ and velocities $\{p_i\}$ of the particles.
    - In Ising model, they are the spins (up or down $\{\sigma_i\}$) of the particles.
  - A rule for changing the state $S_{n+1} = T\ (S_n)$
    - In a random case, the new state is sampled from a distribution $T(S_{n+1}|S_n)$.
  - From initial state S_0, we repeat the iteration many times: $n \Rightarrow \infty$

$$S_0 \Rightarrow S_1 \Rightarrow S_2 \Rightarrow S_3 \Rightarrow S_4 \Rightarrow S_5 \Rightarrow \ldots\ldots S_n \Rightarrow S_{n+1} \Rightarrow$$

  - The *iteration index* "n" is called "time." It could be either "real time" or an iteration count, a pseudo-time, sometimes called *Monte Carlo time*.

- Simulations can be:
  - Deterministic (e.g. Newton's equations via Molecular Dynamics)
  - Stochastic (Monte Carlo, Brownian motion,…)

  Nonetheless, you analyze in a similar way.

- **Why do a simulation? It is the only exact method for general many-body problems! As with experiment: the rules of the simulation can be simple but output can be unpredictable.**

Ergodicity

- Typically simulations are assumed to be ergodic:
  - after a certain time the system loses memory of its initial state, S_0, except possibly for certain conserved quantities such as the energy, momentum.
  - The *correlation time* κ (which we will define soon) is the number of iterations it takes to forget.
  - If you look at (non-conserved) properties for times much longer κ, they are unpredictable as if randomly sampled from some distribution.
  - Ergodicity is often easy to prove for the random transition but usually difficult for the deterministic simulation.

The assumption of egodicity is used for:

- Warm up period at the beginning (or equilibration)
- To get independent samples for computing errors.

Equilibrium distribution

- Let $F_t(S|S_0)$ be the distribution of state after time t.
- If the system is ergodic, no matter what the initial state, one can characterize the state of the system for t >> κ by a unique *probability distribution: the equilibrium state* $F^*(S)$.

$$\boxed{\lim_{t\to\infty} F(S \mid S_0) = F^*(S)}$$

  - In classical statistical systems, this is the canonical Boltzmann distribution: $F^*(S)=\exp(-V(S)/kT)/Z$
  - In VMC it is the square of the wavefunction
  - In PIMC, it is the path distribution.

- One goal is to compute averages to get static properties in equilibrium. e.g., **the energy:**

$$U = \int dS F^*(S) V(S) \equiv \langle V(S) \rangle_{F^*}$$

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

I. INTRODUCTION

THE purpose of this paper is to describe a general method, suitable for fast electronic computing machines, of calculating the properties of any substance which may be considered as composed of interacting individual molecules. Classical statistics is assumed, only two-body forces are considered, and the potential field of a molecule is assumed spherically symmetric. These are the usual assumptions made in theories of liquids. Subject to the above assumptions, the method is not restricted to any range of temperature or density. This paper will also present results of a preliminary two-dimensional calculation for the rigid-sphere system. Work on the two-dimensional case with a Lennard-Jones potential is in progress and will be reported in a later paper. Also, the problem in three dimensions is being investigated.

* Now at the Radiation Laboratory of the University of California, Livermore, California.

II. THE GENERAL METHOD FOR AN ARBITRARY POTENTIAL BETWEEN THE PARTICLES

In order to reduce the problem to a feasible size for numerical work, we can, of course, consider only a finite number of particles. This number N may be as high as several hundred. Our system consists of a square† containing N particles. In order to minimize the surface effects we suppose the complete substance to be periodic, consisting of many such squares, each square containing N particles in the same configuration. Thus we define d_{AB}, the minimum distance between particles A and B, as the shortest distance between A and any of the particles B, of which there is one in each of the squares which comprise the complete substance. If we have a potential which falls off rapidly with distance, there will be at most one of the distances AB which can make a substantial contribution; hence we need consider only the minimum distance d_{AB}.

† We will use the two-dimensional nomenclature here since it is easier to visualize. The extension to three dimensions is obvious.

Markov chain MC or Random Walk

- Markov chain is a random walk through phase space:

$$s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4 \rightarrow \ldots$$

  Here "*s*" is the state of the system.
- ALL QMC is some type of Markov process. VMC is the simplest.
- The transition probability is $P(s_n \rightarrow s_{n+1})$ a *stochastic matrix*

$$P(s \rightarrow s') \geq 0 \qquad \sum_{s'} P(s \rightarrow s') = 1$$

- In a Markov chain, the distribution of s_{n+1} depends only on s_n (by definition). *A drunkard has no memory!*
- Let $f_n(s)$ be the probability after "n" steps. It evolves according to a "master equation."

$$f_{n+1}(s') = \Sigma_s f_n(s) P(s \rightarrow s')$$

OR: $$f_{n+1} = P f_n$$

- The stationary states are eigenfunctions of *P:* $P \phi = \varepsilon \phi$

- Because P is positive, the eigenvalues have $\varepsilon \leq 1$.

  An equilibrium state must have $\varepsilon = 1$.

- How many equilibrium states are there?
  - If it is *ergodic*, then it will converge to a unique stationary distribution (only one eigenfunction with eigenvalue whose eigenvalue =1)
- ergodicity can be proven if:
  - One can move everywhere in a finite number of steps with non-zero probability. *No barriers!*
  - Non-periodic transition rules. (e.g. not hopping on a bi-partite lattice)
  - Average return time is finite. (No expanding universe.) Not a problem in a finite system.
- If ergodic, convergence is *geometrical and monotonic* to a unique state: $\pi(s)$

$$f_n(s) = \pi(s) + \sum_\lambda \varepsilon^n_\lambda c_\lambda \phi_\lambda(s)$$

Metropolis algorithm

Three key concepts:

1. Sample by using an ergodic random walk.
2. Determine equilibrium state by using detailed balance.
3. Achieve detailed balance by using rejections.

Detailed balance: $\pi(s)\, P(s \rightarrow s') = \pi(s')P(s' \rightarrow s)$.

Rate balance from s to s' implies $\pi(s)$ ***equilibrium state.***

PROOF:Put $\pi(s)$ into the master equation.

$$\Sigma_s\, \pi(s)\, P(s \rightarrow s') = \pi(s')\, \Sigma_s\, P(s' \rightarrow s) = \pi(s')$$

- Hence, $\pi(s)$ is an eigenfunction of P.
- If $P(s \rightarrow s')$ is ergodic, $\pi(s)$ is unique steady state solution.

Rejection Method

Metropolis achieves detailed balance by *rejecting* moves.

General Approach:

1. Choose distribution to sample, e.g., $\pi(s) = \exp[-\beta H(s)]/Z$
2. Impose detailed balance on transition: $K(s \to s') = K(s' \to s)$

   where $K(s \to s') = \pi(s)\, P(s \to s')$

   (*probability of being at s)* * (*probability of going to s'*).
3. Break up transition probability into sampling and acceptance:

   $P(s \to s') = T(s \to s')\, A(s \to s')$

(*probability of generating s'* from s) * (*probability of accepting move)*

The optimal acceptance probability that gives detailed balance is:

$$A(s \to s') = \min[1, \frac{T(s' \to s)\,\pi(s')}{T(s \to s')\,\pi(s)}] = \min[1, \frac{\pi(s')}{\pi(s)}]$$

Normalization of $\pi(s)$ is not needed or used!

If T is *constant*!

The Metropolis method for classical systems

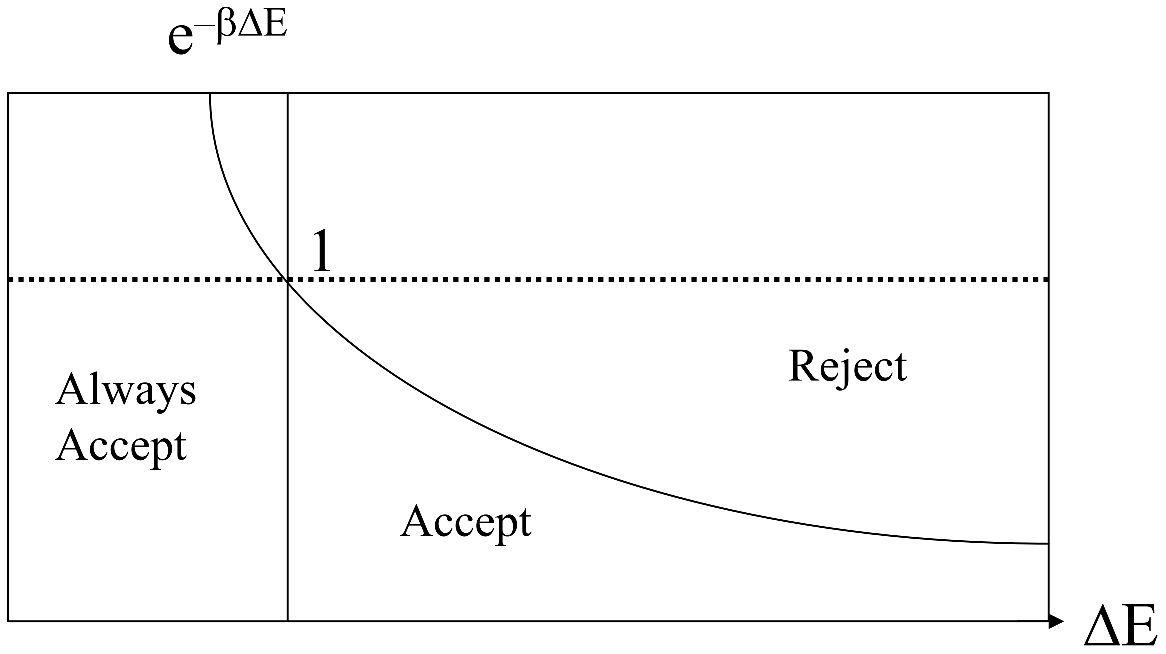
Metropolis-Rosenbluth² -Teller² (1953) method is:

- Move from s to s' with probability $T(s \rightarrow s')$= constant
- Accept with move with probability:

$$A(s \rightarrow s') = \min\left[1 , \exp\left(- (E(s')-E(s))/k_B T \right) \right]$$

- Repeat many times
- Given ergodicity, the distribution of s will be the canonical distribution: $\pi(s) = \exp(-E(s)/k_B T)/Z$
- **Convergence is guaranteed but the rate is not!**

Metropolis Acceptance Probability



- If $\Delta E < 0$, it lowers the system energy → accept.

Otherwise

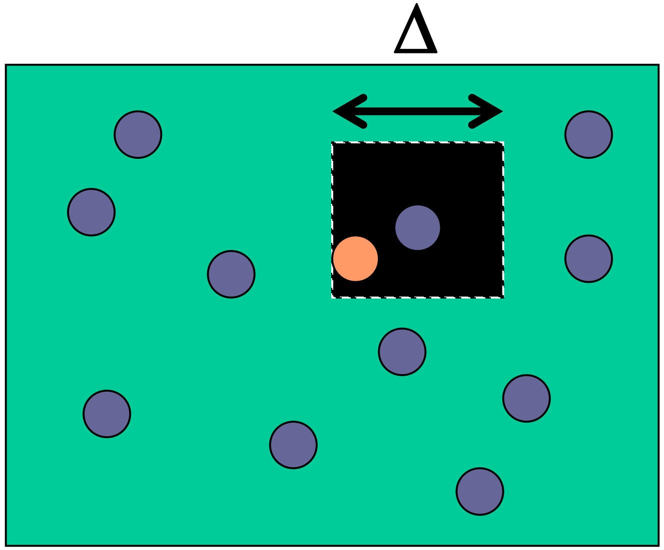
- Generate UDRN u_n on (0,1)
- Compare u_n to $e^{-\beta\Delta E}$: **If $u_n < e^{-\beta\Delta E}$, accept.**

If $u_n > e^{-\beta\Delta E}$, reject.

How to sample

$$S\_new = S\_old + \Delta \cdot (sprng\ - 0.5)$$

Uniform distribution in a cube of side "Δ".



Note: It is more efficient to *move one particle at a time* because only the energy of that particle comes in and the *acceptance ratio will be larger*.

$$A(s \to s') = \exp[-\beta(V(s') - V(s))]$$
$$= \exp[-\beta \sum_{j \neq i} (v(r_i' - r_j) - v(r_i - r_j))]$$

For V with cut-off range, difference is local.

MONTE CARLO CODE

```
call initstate(s_old)
E_old = action(s_old)
LOOP{
    call  sample(s_old,s_new,T_new,1)
    E_new = action(s_new)
    call sample(s_new,s_old,T_old,0)
    A=exp(-E_new+E_old) T_old/T_new
    if(A.gt.sprng()) {
            s_old=s_new
          E_old=E_new
          naccept=naccept+1}
  call averages(s_old)          }
```

Initialize the state

Sample snew

Trial action

Find prob. of going backward

Acceptance prob.

Accept the move

Collect statistics

Overview of MCMC

- Decide how to move from state to state.
- Initialize the state
- Throw away first *k* states as being out of equilibrium.
- Then collect statistics but be careful about correlations.

Common errors:

1. **If you can move from *s* to *s'*, the reverse move must also be possible.**
2. **Accepted and rejected states count the same!**

• *Always measure acceptance ratio.* Adjust ratio to *roughly* 0.5 by varying the "step size". RULE: 0.1<a.r.<0.9

• A 20% acceptance ratio actually achieves better diffusion than a 50% acceptance ratio *in this example*.

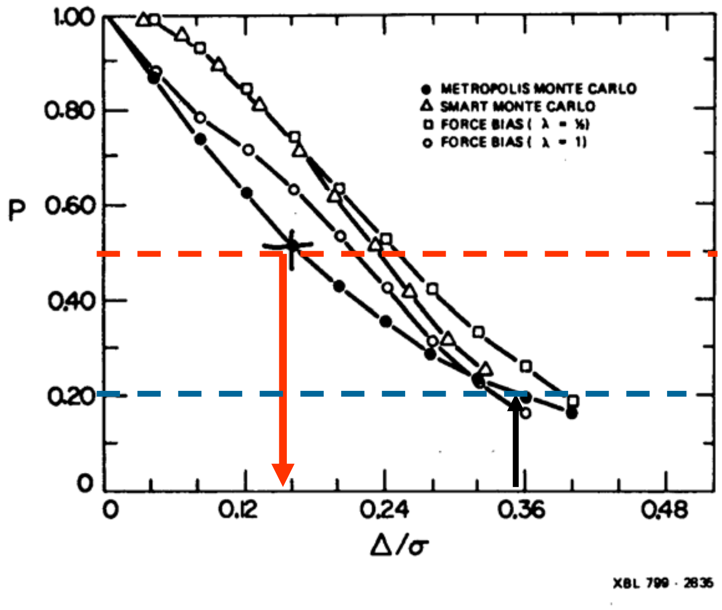


Fig. 1. Average acceptance probability.

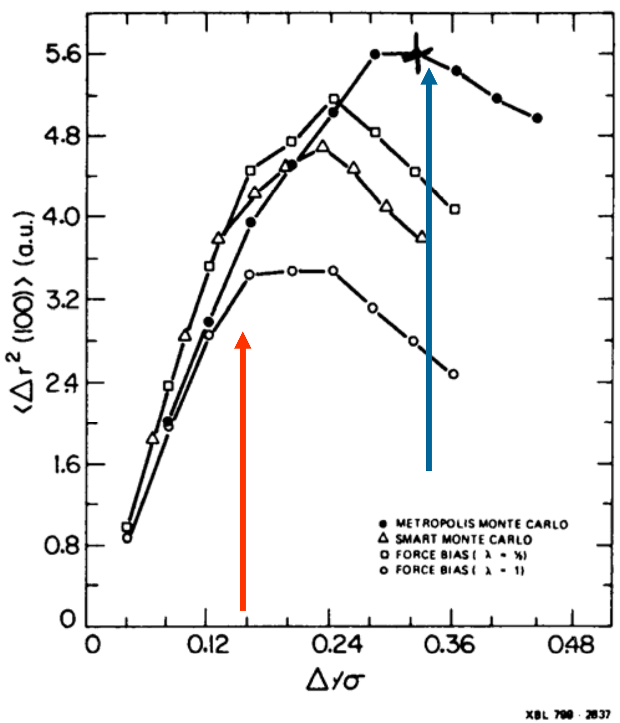


Fig. 3. $\langle(\vec{r}(i) - \vec{r}(i + 100))^2\rangle$

$\vec{r}(i)$ = 3n vector of argon positions at step i.

Variance of energy (local quantity) is not as sensitive to step size.

MC is a robust method! You don't need to fine tune things!

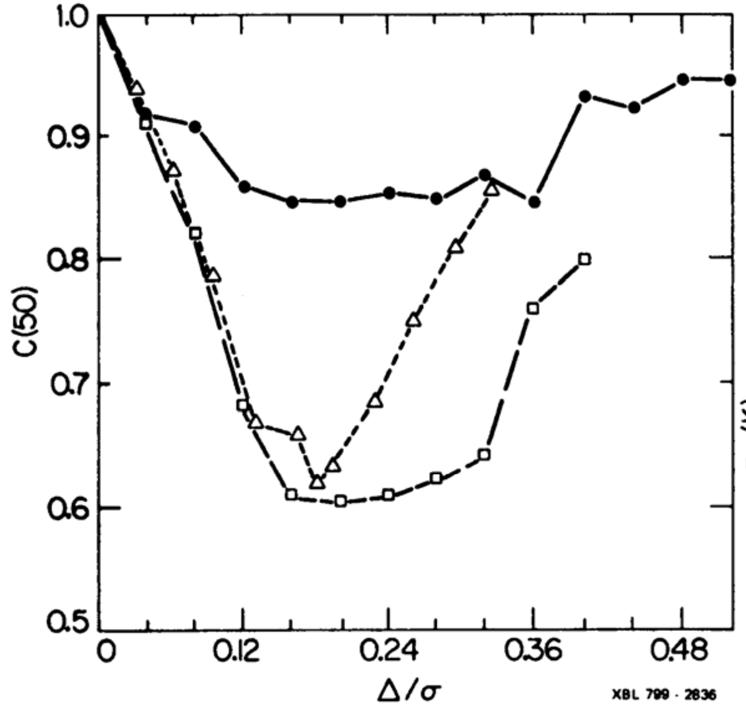


Fig. 2. $\frac{\langle \Delta V(i)\, \Delta V(i+50)\rangle}{\langle \Delta V(i)^2\rangle}$

where i = step number and
Δ V is the deviation of potential
energy from the mean.

Acceptable range of efficiency.

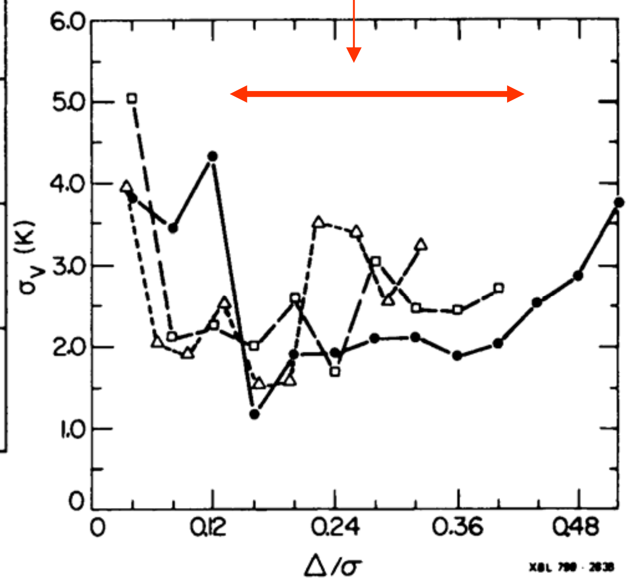


Fig. 4. The variance of the total potential energy for calculations with the same number of steps.

Optimizing the moves

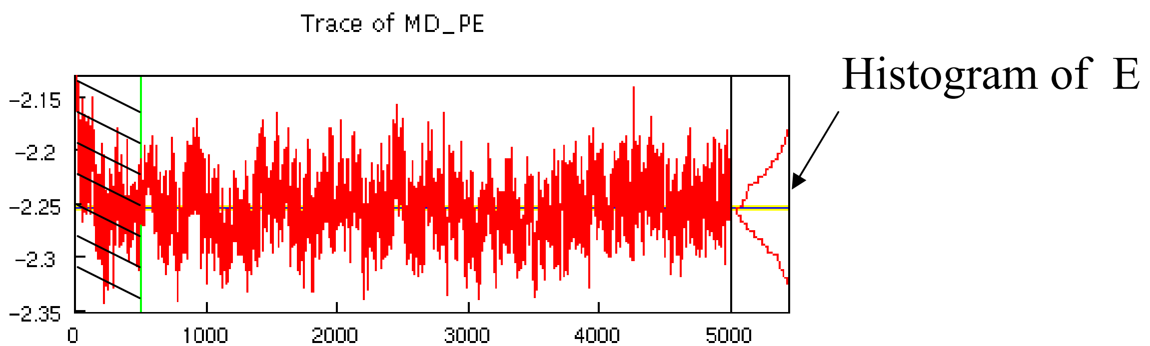
- *Any transition rule is allowed* as long as you can go *anywhere in phase space with a finite number of steps*. (Ergodicity)

- Try to find a $T(s \rightarrow s') \approx \pi(s')/C$.
  - If you can, the acceptance ratio will be 1.

- Can use the forces to push the walk in the right direction.
  - Taylor expand about current point: $V(\mathbf{r})=V(\mathbf{r}_0)-F(\mathbf{r})(\mathbf{r}-\mathbf{r}_o)$
  - **Then set $T(s \rightarrow s') \approx \exp[-\beta(V(r_0)-F(r_0)(r-r_o))]$**
  - Leads to ***Force-Bias Monte Carlo.***
  - Related to Brownian motion (Smoluchowski Eq.) and to diffusion Monte Carlo

Estimated Errors

- In what sense do we calculate exact properties? Answer: if we average long enough the error goes to zero, the errors of the simulation are controlled.
- Next: how accurate is the estimate of the exact value?
  - Simulation results without error bars are only suggestive.
    - Without error bars one has no idea of its significance.
    - You should **understand formulas and be able to make an "*eye-ball*" *estimate***.
- **Error bar**: the ***estimated error*** in the ***estimated mean***.
  - Error estimates based on Gauss' ***Central Limit Theorem***.
  - **Average** of statistical processes has *normal* (Gaussian) distribution.
  - **Error bars**: square root of the variance of the distribution divided by the number of ***uncorrelated*** steps.



Central Limit Theorem (Gauss)

Sample N independent values from $F^*(x)dx$, i.e. $(x_1, x_2, x_3, \ldots, x_N)$.

Calculate mean as $y = (1/N)\sum x_i$.

What is the pdf of mean? *Solve by fourier transforms*

Characteristic function: $c_x(k) = \langle e^{ikx} \rangle = \int_{-\infty}^{\infty} dx\, F^*(x) e^{ikx} \quad c_y(k) = c_x(k/N)^N$

$$\lim_{N\to\infty} \quad c_y(k) = e^{ik\kappa_1 - k^2\kappa_2/2N - ik^3\kappa_3/6N^2 \ldots}$$

Cumulants: Mean = κ_1 Variance= κ_2 Skewness = κ_3 Kurtosis= κ_4

The n=1 moment remains invariant but the rest get reduced by higher powers of N.

CLT: Given enough averaging almost anything becomes a Gaussian distribution.

$$\boxed{P(y) = (N/2\pi\kappa_2)^{1/2} \exp\left[-\frac{N(y-\kappa_1)^2}{2\kappa_2}\right] \quad \text{standard error(y)} = \sigma = \sqrt{\frac{\kappa_2}{N}}}$$

Approach to normality

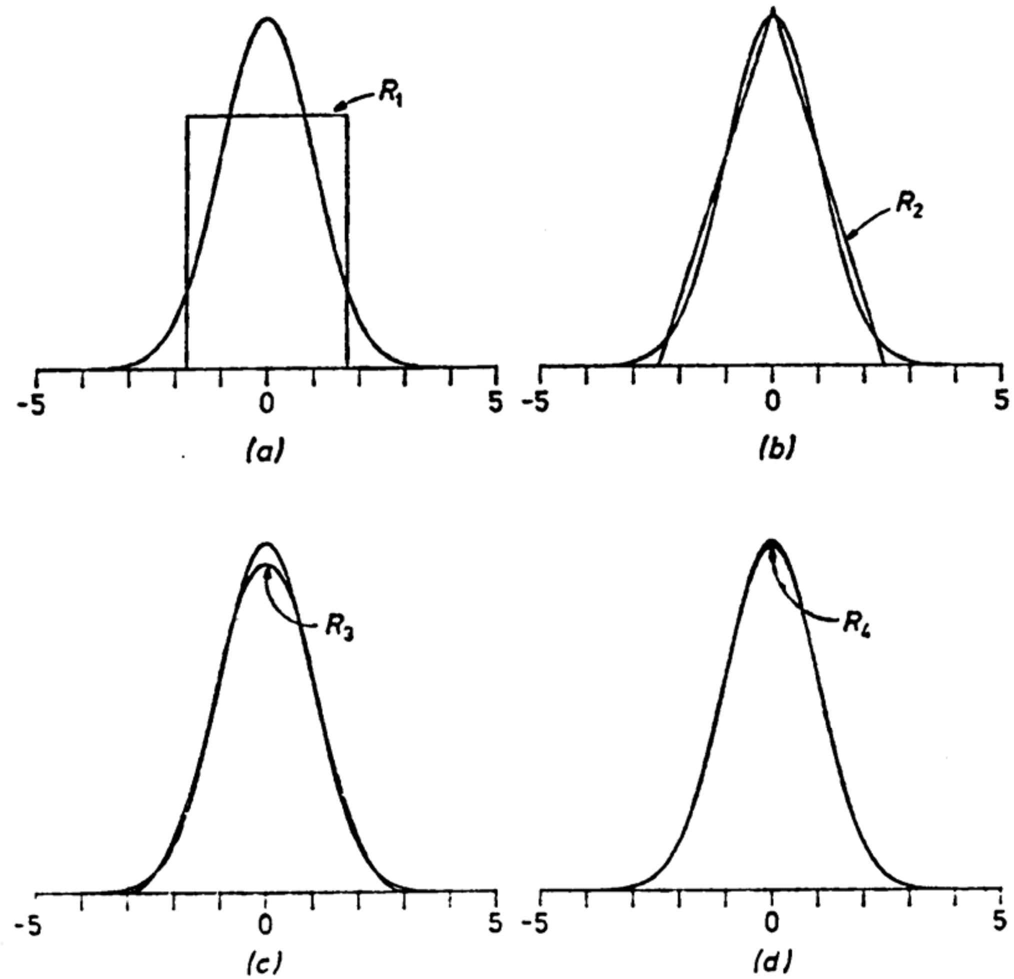


Figure 1. Distributions of sums of uniform random numbers, each compared with the normal distribution. (*a*) R_1, the uniform distribution. (*b*) R_2, the sum of two uniformly distributed numbers. (*c*) R_3, the sum of three uniformly distributed numbers. (*d*) R_{12}, the sum of twelve uniformly distributed numbers.

Conditions on Central Limit Theorem

$$I_n = \langle x^n \rangle = \int_{-\infty}^{\infty} dx\, F^*(x) x^n$$

- We need the first three moments to exist.
  - If I_0 is not defined $\Rightarrow$ not a pdf
  - If I_1 does not exist $\Rightarrow$ not mathematically well-posed.
  - If I_2 does not exist $\Rightarrow$ infinite variance. Important to know if variance is finite for simulations.
- Divergence could happen because of tails of distribution

$$I_2 = \langle x^2 \rangle = \int_{-\infty}^{\infty} dx\, F^*(x) x^2$$

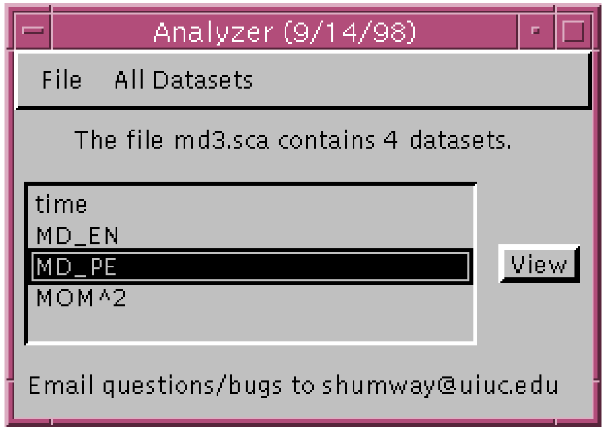
We need: $$\lim_{x \to \pm\infty} \quad x^3 F^*(x) \to 0$$

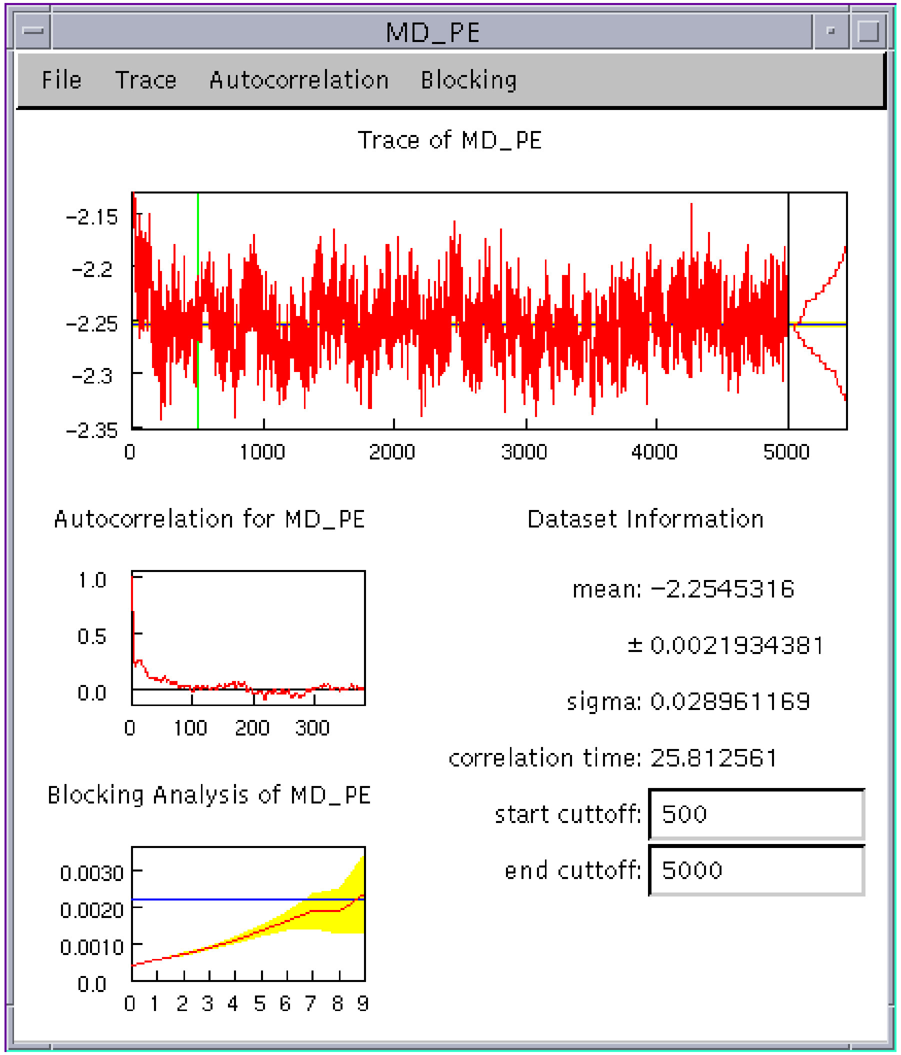
- OR Divergence because of singular behavior of F^* at finite x:

We need: $$\lim_{x \to 0} \quad x F^*(x) \to 0$$

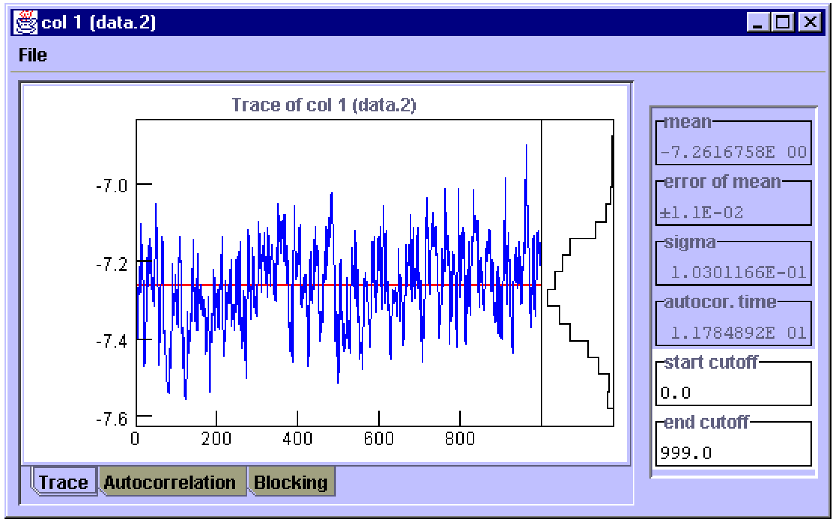
DataSpork

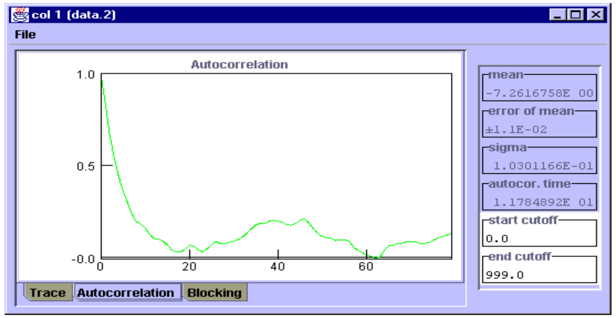
Interactive code to perform statistical analysis of data



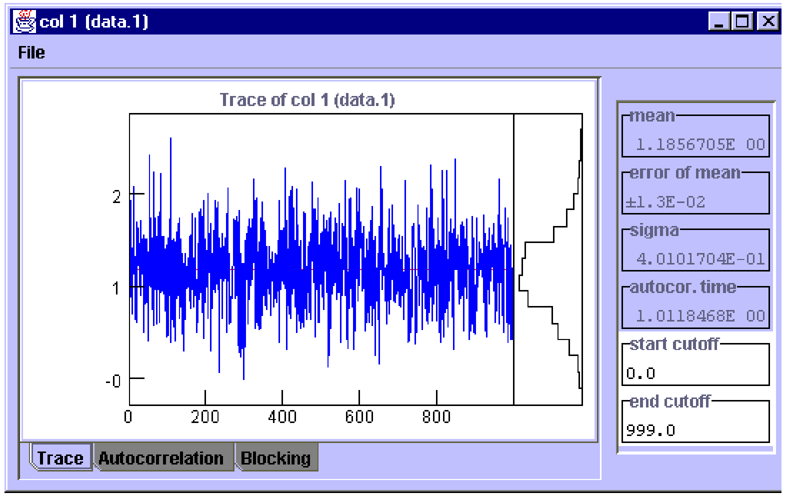


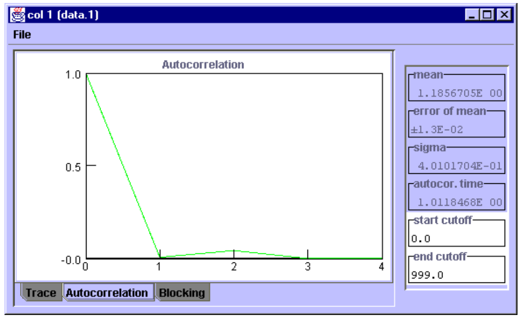
Correlated data





Uncorrelated data





Statistical vs. Systematic Errors

- What are statistical errors?
  - Statistical error measures the distribution of the averages about their avg.
  - ***Statistical error can be reduced by extending or repeating runs***, increase N.

$$\text{standard error(y)}=\sigma=\sqrt{\frac{\kappa_2}{N}}$$

- The efficiency is how we measure the rate of convergence of the statistical errors.

$$\boxed{\zeta = \frac{1}{T\sigma^2}}$$

  - It depends on the computer, the algorithm, the property etc. But not on the length of the run.
- What are systematic errors ?
  - Systematic error measures the others errors. Even if you sample forever you do not get rid of systematic errors.
  - Systematic error is caused by round-off error, non-linearities, bugs, non-equilibrium, etc.

Estimating Errors

- Uncorrelated data

$$\{a_t\} \qquad 0 < t \le N$$

$$\langle a_t \rangle \approx \bar{a} = \frac{1}{N}\sum_t a_t$$

$$error(\bar{a}) = \left\langle \left(\bar{a} - \langle a \rangle\right)^2 \right\rangle^{1/2} \approx \left[\frac{\sum_t \delta a_t^{\,2}}{N(N-1)}\right]^{1/2}$$

$$\delta a_t \equiv a_t - \bar{a}$$

- Correlated data

$$error(\bar{a}) = \left\langle \left(\bar{a} - \langle a \rangle\right)^2 \right\rangle^{1/2} \approx \left\langle \frac{\kappa \sum_t \left(a_t - \bar{a}\right)^2}{N(N-1)} \right\rangle^{1/2}$$

$$\kappa = 1 + 2\sum_{t=1}^{\infty} \frac{\langle \delta a_t \delta a_0 \rangle}{\langle \delta a^2 \rangle} = \text{correlation time}$$

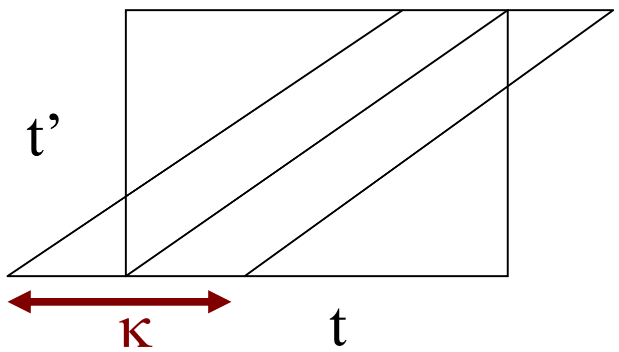
- Problem: how to cut off the summation for κ.
- Blocking method: average together data in blocks longer than the correlation time until it is uncorrelated.

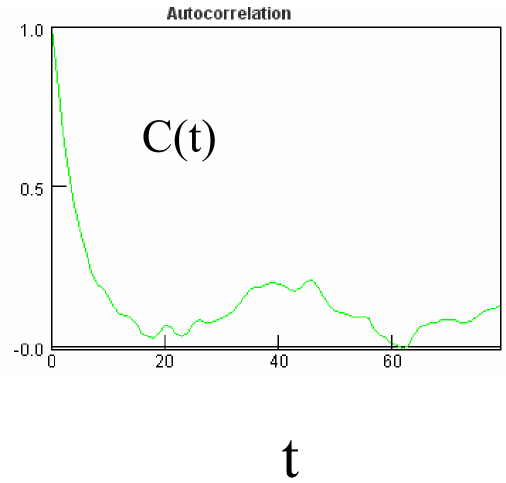
Estimate of errors for correlated data

$$\boxed{\text{estimate of sample mean is } \bar{a} = \frac{1}{N}\sum_t a_t}$$

$$error(\bar{a}) = \left\langle \left(\bar{a} - \langle a \rangle\right)^2 \right\rangle^{1/2}$$

$$\left\langle \left(\bar{a} - \langle a \rangle\right)^2 \right\rangle = \left\langle \frac{1}{N^2}\sum_{t,t'}^{N} \delta a_t \delta a_{t'} \right\rangle = \frac{\left\langle \delta a^2 \right\rangle}{N^2}\sum_{t,t'}^{N} C_{|t-t'|} \le \frac{\left\langle \delta a^2 \right\rangle}{N^2}\sum_{t'=1}^{N}\sum_{t=-\infty}^{\infty} C_t = \left\langle \delta a^2 \right\rangle \frac{\kappa}{N}$$





$$\kappa = 1 + 2\sum_{t=1}^{\infty} C(t) = \text{correlation time} \approx 2\int_0^{\infty} \frac{dt}{\delta t}\,\mathrm{C(t)}$$

$$C(t,t') \equiv \frac{\left\langle \delta a_t \delta a_{t'} \right\rangle}{\left\langle \delta a^2 \right\rangle} = C\left(|t-t'|\right) = \text{autocorrelation function}$$

Statistical Vocabulary

- **Trace of A(t)**:
- **Equilibration time**.
- **Histogram** of values of A (P(A)).
- **Mean** of A (a).
- **Variance** of A (v).
- **estimate of the mean:** $\Sigma A(t)/N$
- **estimate of the variance**
- **Autocorrelation** of A (C(t)).
- **Correlation time** κ .
- The (estimated) **error** of the (estimated) **mean** (σ).
- **Efficiency** [= 1/(CPU time * error ^{2})]

Statistical thinking is slippery: be careful

- "Shouldn't the energy settle down to a constant"
  - NO. It fluctuates forever. It is the overall mean which converges.
- Because data is correlated, the central limit theorem is invalid
- "The cumulative energy has converged".
  - BEWARE. Even pathological cases have smooth cumulative energy curves.
- "Data set A differs from B by 2 error bars. Therefore it must be different".
  - This is normal in 1 out of 10 cases. If things agree too well, something is wrong!
- "My procedure is too complicated to compute errors"
  - NO! Run your whole code 10 times and compute the mean and variance from the different runs. If a quantity is important, you MUST estimate its errors.

Recap: problems with estimating errors

- Any good simulation quotes *systematic and statistical* errors for anything important.
- The *error and mean* are simultaneously determined from the same data. HOW?
- Central limit theorem: the distribution of an average approaches a normal distribution (*if the variance is finite*).
  - One *standard deviation* means ~2/3 of the time the correct answer is within σ of the sample average.
- Problem in simulations is that *data is correlated in time*.
  - It takes a "correlation" time κ to be "ergodic"
  - Correction errors for autocorrelation.
  - throw away the initial transient.
- We need about 25 *independent* data points to estimate errors. (so that the error of the error is only 1/sqrt(N)= 20%)