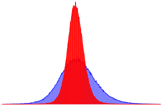


Probability and Statistics in Quantum Monte Carlo

Pablo López Ríos, John Trail

QMC school at S.N.Bose Centre

24 March 2015

The need for statistical analysis

- A QMC calculation produces **millions** of data values
- We want a single number (with its **error bar**) as a result:

$$\boxed{E \pm \sigma_E}$$

- Serial correlation needs to be removed
- How to **manipulate** quantities with error bars

Basic statistics

- The configurations $\{\mathbf{R}_i\}_{i=1}^{i=M}$ distributed according to $|\Psi(\mathbf{R})|^2$
- The local energy $E_i = E_L(\mathbf{R}_i) = \Psi^{-1}(\mathbf{R}_i)\hat{H}\Psi(\mathbf{R}_i)$
- $E_L(\mathbf{R})$ forms a distribution with:

Mean

$$E_V = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \approx \bar{E} = \frac{\sum_{i=1}^{M} E_i}{M}$$

Variance

$$\sigma_{E_L}^2 = \frac{\langle\Psi|\hat{H}^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} - \left[\frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}\right]^2 \approx \tilde{\sigma}_{E_L}^2 = \frac{\sum_{i=1}^{M}\left(E_i - \bar{E}\right)^2}{M-1}$$

Basic statistics

- $\bar{E}$ can be determined to a given degree of certainty
- Different calculations yield different $\bar{E}$ values
- $\bar{E}$ is itself a random number distributed according to

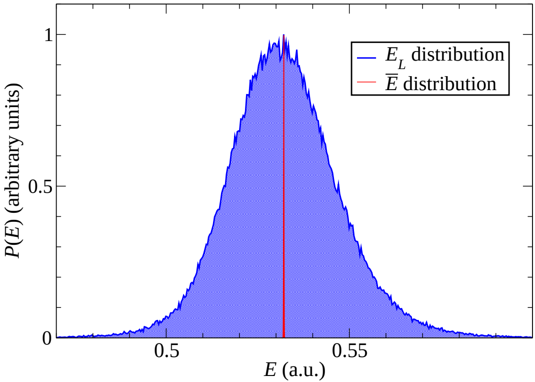
Mean

$$\bar{E} \approx \frac{\sum_{i=1}^{M} E_i}{M}$$

Variance

$$\sigma_{\bar{E}}^2 \approx \tilde{\sigma}_{\bar{E}}^2 = \frac{\sum_{i=1}^{M} \left(E_i - \bar{E}\right)^2}{M(M-1)}$$

Local energy and mean energy



The local energy distribution is what we sample.
The mean energy distribution is what we obtain.

Sampling of configuration space

$\{\mathbf{R}_i\}_{i=1}^{i=M}$ must be distributed according to $|\Psi(\mathbf{R})|^2$.

Sampling algorithm at i-th step

- Start at config $\mathbf{R}_i$
- **Propose** a new config $\mathbf{R}_i'$
- Compute the **wave function ratio** $q_i = \left|\frac{\Psi(\mathbf{R}_i')}{\Psi(\mathbf{R}_i)}\right|^2$
- Generate uniform **random number** $\xi \in [0,1)$
- Accept/reject step:
    - if $\xi < q_i \rightarrow$ set $\mathbf{R}_{i+1} = \mathbf{R}_i'$ (accept new config)
    - if $\xi > q_i \rightarrow$ set $\mathbf{R}_{i+1} = \mathbf{R}_i$ (reject new config)

Proposing $\mathbf{R}_i \to \mathbf{R}'_i$

- If $\mathbf{R}'_i$ proposed at random:
  → Small chance of landing in a **reasonable** region of configuration space
  → q_i will be small
  → most moves are rejected
  → poor sampling
- If $\mathbf{R}'_i$ is $\mathbf{R}_i$ plus a small displacement:
  → $\mathbf{R}'_i$ similar to $\mathbf{R}_i$
  → $E_L(\mathbf{R}'_i)$ similar to $E_L(\mathbf{R}_i)$
  → **Serial correlation**

Effect of serial correlation

- Consider an uncorrelated set of energies $\{E_1, E_2, E_3, \ldots, E_M\}$
- Generate a new set with artificial serial correlation:

$$\{\underbrace{E_1, \ldots, E_1}_{\tau}, \underbrace{E_2, \ldots, E_2}_{\tau}, \underbrace{E_3, \ldots, E_3}_{\tau}, \ldots, \underbrace{E_M, \ldots, E_M}_{\tau}\}$$

- No new information → mean and error bar should be unchanged
- Computed mean of new set is $\bar{E}' = \bar{E}$
- Computed error bar of new set is $\tilde{\sigma}'_{\bar{E}} = \tilde{\sigma}_{\bar{E}}/\sqrt{\tau}$ → **error bar underestimated**

Removing serial correlation

- In this example we can remove serial correlation by ignoring $\tau - 1$ of every τ consecutive energies
- For real data the correlation time τ varies during the run
  $\rightarrow$ would need to ignore $\tau_{\max} - 1$ of each $\tau_{\max}$ data
  $\rightarrow$ lots of relevant data discarded
  $\rightarrow$ inefficiency
- However the formula

$$\tilde{\sigma}_{\bar{E}} = \sqrt{\tau}\tilde{\sigma}'_{\bar{E}}$$

  still holds, where τ is the **average** correlation time
- This is an alternative approach to the **reblocking algorithm**

The reblocking algorithm

- Consider the following operation on data, where the item under each brace is the average of the two numbers above:

$$\underbrace{\underbrace{E_1^{(0)} \quad E_2^{(0)}}_{E_1^{(1)}} \quad \underbrace{E_3^{(0)} \quad E_4^{(0)}}_{E_2^{(1)}}}_{\ldots} \quad \underbrace{\underbrace{E_5^{(0)} \quad E_6^{(0)}}_{E_3^{(1)}} \quad \underbrace{E_7^{(0)} \quad E_8^{(0)}}_{E_4^{(1)}}}_{\ldots}$$

- Succesively apply transformations until $\tau_{\max}$ original data are averaged together → **resulting data are uncorrelated**
- Cannot compute $\tau_{\max}$ directly — need another way to determine how many reblocking transformations to apply

Error estimator after reblocking

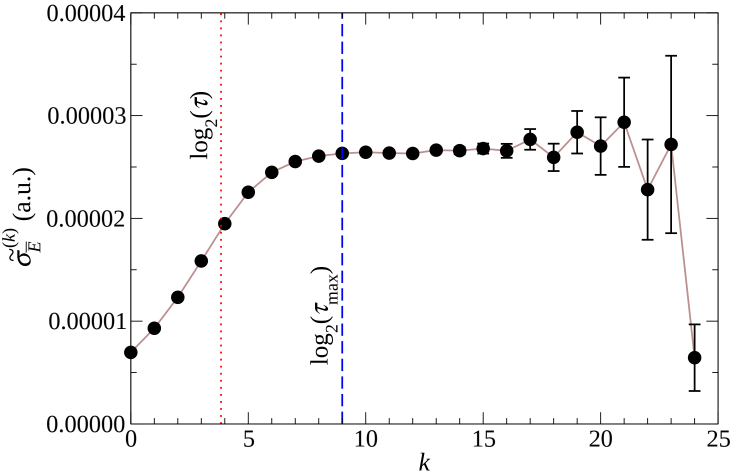
- At the k-th iteration in this procedure:

$$\tilde{\sigma}_{\bar{E}}^{(k+1)2} \approx \tilde{\sigma}_{\bar{E}}^{(k)2} + \frac{2\sum_{i=1}^{M^{(k)}/2}\left(E_{2i-1}^{(k)} - \bar{E}\right)\left(E_{2i}^{(k)} - \bar{E}\right)}{M^{(k)}(M^{(k)}-2)}$$

- If there is no serial correlation, the last term tends to zero
- If there is serial correlation, the last term is positive
- Hence $\tilde{\sigma}_{\bar{E}}^{(k)}$ will increase until it reaches the true error bar at $k \approx \log_2(\tau_{\max})$

Plateau in $\tilde{\sigma}_{\bar{E}}^{(k)}$ signals convergence of reblocking algorithm

Reblock plot



How to run efficient VMC calculations

- Reducing serial correlation, by
  - Choosing an appropriate timestep
  - Using electron-by-electron sampling
  - Skipping the right number of steps between every two calculations of expectation values
- Reducing the intrinsic variance/expense, by
  - Using appropriate trial wave functions

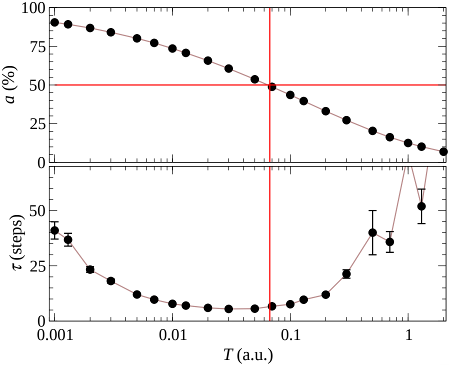
The VMC timestep

- The "timestep" T is the variance of the distribution used to generate the random displacements when proposing moves
- It is actually a squared length, but can be regarded a time if considering a diffusion process
- T does **not** enter the VMC formalism
  → can be chosen so as to improve run statistics
  - T small → $\mathbf{R}_i'$ very similar to $\mathbf{R}_i$
    → serial correlation increased
  - T large → $\mathbf{R}_i'$ very dissimilar from $\mathbf{R}_i$
    → most moves are rejected
    → serial correlation increased

The 50% rule

The 50% rule

Choose T such that the acceptance ratio $a = 50\%$



Electron-by-electron sampling

- QMC sampling usually described using configuration moves → Configuration-by-configuration sampling (CBCS)
- In practice, one-electron moves proposed and accepted or rejected individually → Electron-by-electron sampling (EBES)
- Two case comparisons:
  - Set T to the same value in CBCS and EBES
    $\rightarrow a_{\rm C} = a_{\rm E}^N$ (very small)
  - Set a to the same value in CBCS and EBES
    → the chance of $\mathbf{R}_{i+1} = \mathbf{R}_i$ in CBCS is $1 - a$
    → the chance of $\mathbf{R}_{i+1} = \mathbf{R}_i$ in EBES is $(1-a)^N$ (very small)

EBES is more efficient

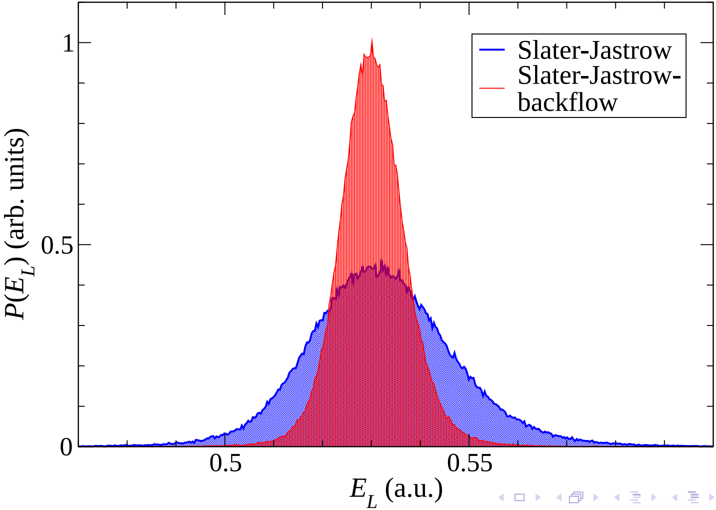
Choosing the right wave function

- With a more sophisticated wave function (e.g., adding backflow, 3-body Jastrow terms, etc):
    - Lower energy
    - **Lower** variance → fewer steps for target error bar
    - Higher cost of evaluation
    - Harder optimization
    - Diminishing returns
    - Similar energy differences (cancellation of errors)

Important!

The **best** trial wave function for a problem need **not** be the **most sophisticated**

Wave functions and the local energy distribution



The DMC algorithm

- **Start** from P walkers $\{\mathbf{R}_{0,\alpha}\}_{\alpha=1}^{P}$ distributed according to $|\Psi(\mathbf{R})|^2$ (from VMC)
- DMC evolution of the walkers:
    - **Drift-diffusion**: move $\mathbf{R}_{i,\alpha} \to \mathbf{R}'_{i,\alpha}$
    - **Branching**: define weight $w_{i,\alpha}$
      $\to$ configurations breed/die according to **branching factor** $w'_{i,\alpha}/w_{i,\alpha}$
      $\to$ variable number of walkers P_i
- **Equilibrate** the walkers until we reach infinite-time limit
  $\to$ look at $E_i = \sum_{\alpha=1}^{P_i} w_{\alpha,i} E_{\alpha,i} / \sum_{\alpha=1}^{P_i} w_{\alpha,i}$

The DMC algorithm

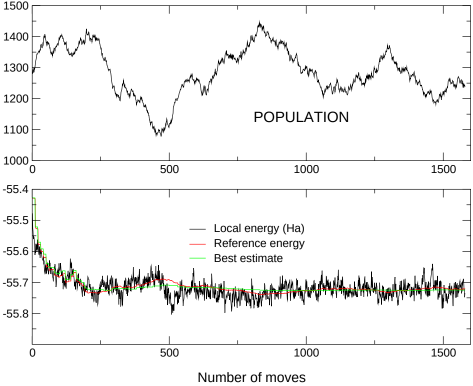
- **Accumulate** data after equilibration to improve statistics of result

DMC mixed estimator

$$\langle A\rangle_{\mathrm{DMC}} = \lim_{t\to\infty}\langle\Psi|\hat{A}|\Phi(t)\rangle/\langle\Psi|\Phi(t)\rangle$$

$$E_D \approx \bar{E} = \frac{\sum_{i=1}^{M} W_i E_i}{\sum_{i=1}^{M} W_i} \quad ; \quad \sigma_{\bar{E}}^2 \approx \tilde{\sigma}_{\bar{E}}^2 = \frac{\sum_{i=1}^{M} W_i \left(E_i - \bar{E}\right)^2}{M\left(\sum_{i=1}^{M} W_i - \frac{\sum_{i=1}^{M} W_i^2}{\sum_{i=1}^{M} W_i}\right)}$$

Calculation of the energy in DMC



Sources of error in DMC

- **Timestep**: we have assumed that T is small
  $\rightarrow$ must extrapolate to zero timestep to obtain a reliable result
  $\rightarrow$ cannot use timestep to improve statistics
- **Population**: Φ is represented by set of configurations
  $\rightarrow$ must use sufficient configurations to represent it accurately
  $\rightarrow$ possible to extrapolate to infinite population
- **Fixed-node error**: only limitation of DMC
  $\rightarrow$ E_D is still variational (very important!)
  $\rightarrow$ can be reduced by using Ψ with better nodes
- **Locality approximation**: from pseudopotentials
  $\rightarrow$ E_D non-variational
  $\rightarrow$ goes away with good Ψ

Central Limit Theorem (CLT)

Derivation of the CLT:

- Let $P_1(x)$ be a probabily distribution of Fourier Transform

$$\mathscr{F}\left[P_1(x)\right] = \exp\left[ia_1k - a_2k^2 + \mathscr{O}(k^3)\right]$$

- Let $P_2(x)$ be the probability that the sum of two numbers drawn from $P_1(x)$ is x:

$$\begin{aligned} P_2(x) &= \int\int P_1(x_1)P_1(x_2)\delta(x_1+x_2-x)dx_1dx_2 \\ &= \int P_1(x_1)P_1(x-x_1)dx_1 \end{aligned}$$

- The Fourier transform of $P_2(x)$ is

$$\mathscr{F}\left[P_2(x)\right] = \mathscr{F}\left[P_1(x)\right]^2 = \exp\left(i2a_1k - 2a_2k^2 + \ldots\right)$$

Central Limit Theorem (CLT)

- Let $P_M(x)$ be the probability that the sum of M numbers drawn from $P_1(x)$ is x:

$$\mathscr{F}\left[P_M(x)\right] = \mathscr{F}\left[P_1(x)\right]^M = \exp(iMa_1k - Ma_2k^2 + \ldots)$$

- $P_M(Mx)$ is the probability that the mean of M numbers drawn from $P_1(x)$ is x, and at large M:

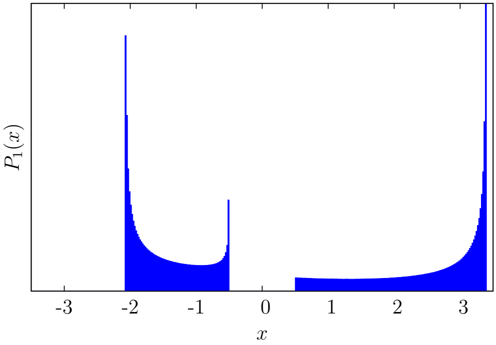
$$\mathscr{F}\left[P_M(Mx)\right] \approx \exp(ia_1k - \frac{a_2}{M}k^2)$$

- Invert $\mathscr{F}$, redefine in terms of $\mu = \mathrm{Mean}[P_1]$, $\sigma^2 = \mathrm{Var}[P_1]/M$:

CLT

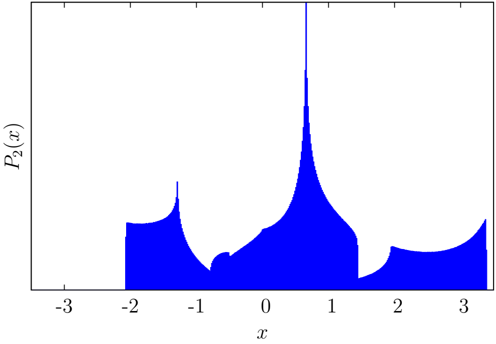
$$\lim_{M\to\infty} P_M(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

CLT: example with peculiar-looking distribution



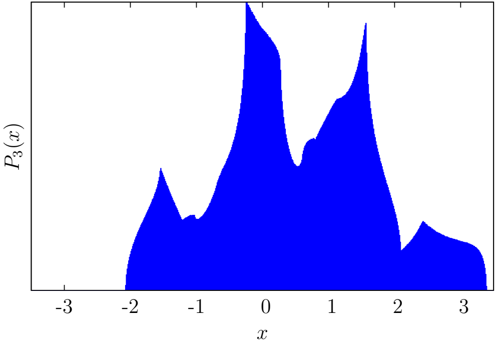
- Average of 1 random variable
- $P_1(x)$ is PDF of $x = x_1$

CLT: example with peculiar-looking distribution



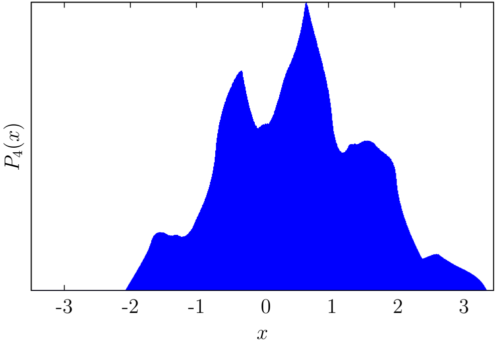
- Average of 2 random variables
- $P_2(x)$ is PDF of $x = \frac{1}{2}(x_1 + x_2)$

CLT: example with peculiar-looking distribution



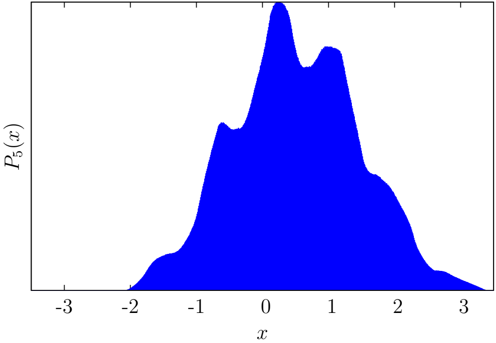
- Average of 3 random variables
- $P_3(x)$ is PDF of $x = \frac{1}{3}(x_1 + x_2 + x_3)$

CLT: example with peculiar-looking distribution



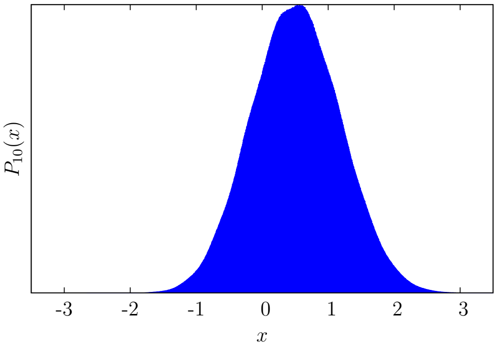
- Average of 4 random variables
- $P_4(x)$ is PDF of $x = \frac{1}{4}(x_1 + x_2 + x_3 + x_4)$

CLT: example with peculiar-looking distribution



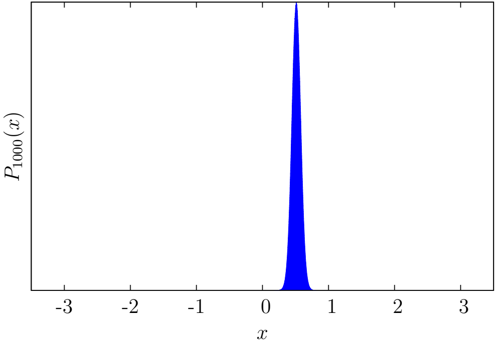
- Average of 5 random variables
- $P_5(x)$ is PDF of $x = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5)$

Central Limit Theorem



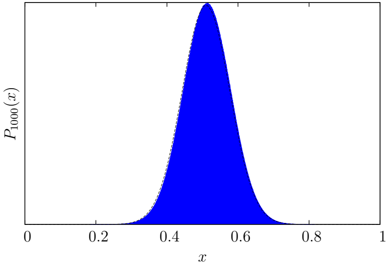
- Average of 10 random variables
- $P_{10}(x)$ is PDF of $x = \frac{1}{10}\sum_{n=1}^{10} x_n$

Central Limit Theorem



- Average of 1000 random variables
- $P_{1000}(x)$ is PDF of $x = \frac{1}{1000}\sum_{n=1}^{1000} x_n$

Central Limit Theorem



- Average of M random variables → *Normal distribution*
- Defined by 2 numbers, the mean and standard deviation
- Centred at mean, width of $\sigma \propto 1/\sqrt{N}$
- Probability is all close to the mean

Is the CLT always true?

- Usually CLT is true iff the mean and variance of P_1 are finite
- Counterexample: $P_1(x)$ with x^{-2} tails
- $\mathscr{F}\left[P_1(x)\right] = \exp(ia_1k - a_2|k| + \ldots)$:
- $\mathscr{F}\left[P_M(Mx)\right] \approx \exp(ia_1k - a_2|k|)$

Limit theorem for x^{-2} tails

$$\lim_{M\to\infty} P_M(Mx) = \frac{\beta}{\pi}\frac{1}{\beta^2+(x-\alpha)^2} \quad (1)$$

$\alpha \neq$ mean, and $\beta \neq$ standard error

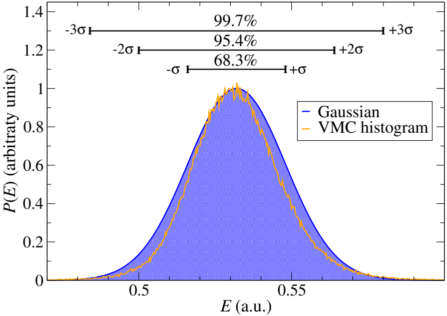
For total energy in QMC we can prove that the CLT is true
(Not so for certain other expectation values)

The normal distribution

- The normal distribution is $D(E;\bar{E},\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(E-\bar{E})^2}{2\sigma^2}\right]$
- The probability of the E being in an interval (A,B) is
    - $P(A<E<B)=f\left(\frac{B-\bar{E}}{\sigma}\right)-f\left(\frac{A-\bar{E}}{\sigma}\right)$
    - $f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-y^2/2\right)dy$
- One-sigma interval $(\bar{E}-\sigma,\bar{E}+\sigma)\rightarrow 68.3\%\rightarrow$ unreliable
- Two-sigma interval $(\bar{E}-2\sigma,\bar{E}+2\sigma)\rightarrow 95.4\%\rightarrow$ reliable
- Three-sigma interval $(\bar{E}-3\sigma,\bar{E}+3\sigma)\rightarrow 99.7\%\rightarrow$ very reliable

The normal distribution

Comparison of a Gaussian and the local energy distribution



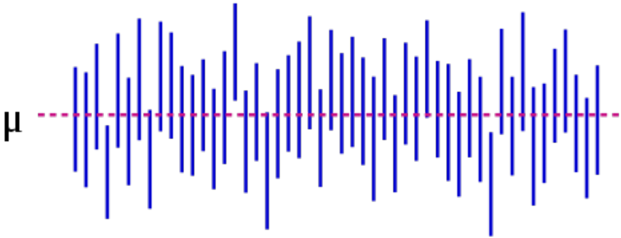
From Central Limit Theorem:

The mean energy is exactly normal

How to compare quantities with errorbars

- Want to find distribution of difference, denoted $(\bar{E}_- \pm \sigma_-) = (\bar{E}_1 \pm \sigma_1) - (\bar{E}_2 \pm \sigma_2)$
- Results in
  - $\bar{E}_- = \bar{E}_1 - \bar{E}_2$
  - $\sigma_-^2 = \sigma_1^2 + \sigma_2^2$
- Example:
  - Ψ_1 gives $E_1 = -14.66728(2)$ a.u.
  - Ψ_2 gives $E_2 = -14.66733(7)$ a.u.
  - Comparison: $E_- = 0.00005(7)$ a.u. → 76% chance of $E_2 < E_1$ → unreliable!
  - If $E_2 = -14.66733(2)$ a.u. instead → $E_- = 0.00005(3)$ a.u. → 95% chance of $E_2 < E_1$ → reliable

What are error bars?



"x% of error bars will include exact mean" is the definition of a confidence inteval

E.g., "68.3% of error bars will include exact mean"

Is random error an “extra” error?

- The presence of an error bar often creates the first impression that QMC has an “extra” error that other methods do not
- However, computers cannot do integration exactly:
  - **Finite basis sets** → basis set error (unkown, controlled)
  - **Quadrature on grid** → quadrature error (unkown, controlled)
  - **Monte Carlo** → random error (known, controlled)
- QMC has a different **type** of integration error

Summary

- **Reblocking algorithm** applied using the REBLOCK utility
- **Average correlation time** τ given in VMC runs and REBLOCK utility
- **VMC timestep** automatically optimized to give $a = 50\%$ (do not apply on HEG)
- **EBEA** is the default in both VMC and DMC
- **DMC statistics** monitored using GRAPHIT utility
- **Timestep extrapolation** carried out using the EXTRAPOLATE_TAU utility