Random Estimates in QMC

J.R. Trail and R.J. Needs

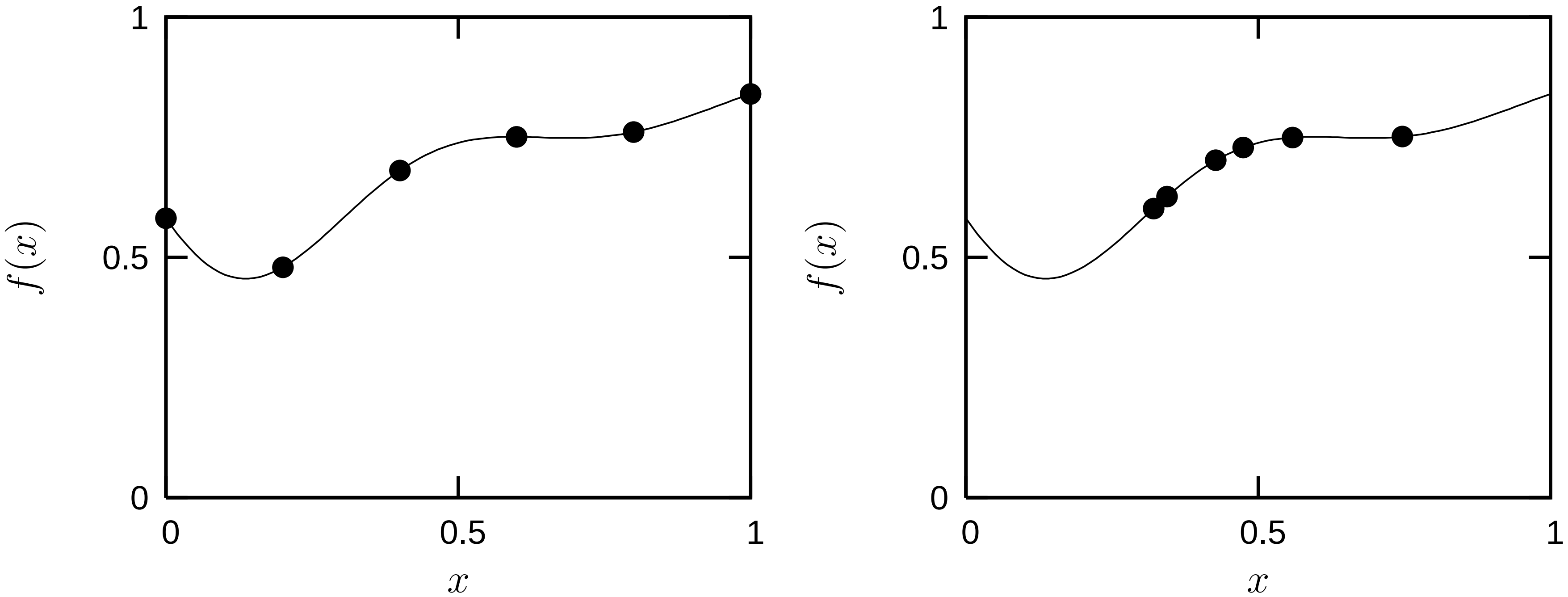
Theory of Condensed Matter Group, Cavendish Laboratory,

University of Cambridge, UK

ESDG February 2007

Numerical Integration

Unkown function, r sample points



Evenly sampled grid:

$$\int_0^1 f(x)dx \approx \sum a_n f(x_n) \quad , \quad \epsilon \propto \Delta^p \approx \left(\frac{1}{r}\right)^{p/D}$$

Points sampled randomly, with PDF $P(x)$:

$$\int_0^1 f(x)dx \approx \frac{1}{r}\sum f(x_n)/P(x_n) \quad , \quad \text{e} \propto \frac{1}{\sqrt{r}}$$

Quantum Monte Carlo (VMC)

Sample 3N dimensional space with PDF $P(\mathbf{R})$

$$\text{Est}\left[E_{tot}\right] = \frac{\sum \psi^2 E_L/P}{\sum \psi^2/P} \quad = \quad \frac{\langle\psi|\hat{H}|\psi\rangle + \text{Y}}{\langle\psi|\psi\rangle + \text{X}}$$

where $E_L = \psi^{-1}\hat{H}\psi$.

Simplest case is 'Standard Sampling': Choose $P(\mathbf{R}) = A\psi(\mathbf{R})^2$, then

$$\text{Est}\left[E_{tot}\right] = \frac{1}{r}\sum E_L \quad = \quad \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} + \text{W}$$

- W is the random error in a sum of random variables, so what is its distribution?
- **IF** the CLT is valid then it is Gaussian with mean 0, and variance $\sigma/r^{1/2}$.

3N $\longrightarrow$ 1 dimension

Why?: Easier to deal with the general case analytically

A change of the random variable from spatial to energy:

$$\begin{aligned} E_{tot} &= \int_V \psi^2 E_L d^{3N}\mathbf{R} / \int_V \psi^2 d^{3N}\mathbf{R} \\ &= \int_{-\infty}^{\infty} P_{\psi^2}(E) E dE \end{aligned}$$

with

$$P_{\psi^2}(E) = \int_{E=E_L} \frac{P(\mathbf{R})}{|\nabla_{\mathbf{R}} E_L|} d^{3N-1}\mathbf{R}$$

- A histogram of E_L approximates the 'seed' PDF P_{ψ^2}
- $|\nabla_{\mathbf{R}} E_L|$ results from curvilinear co-ordinates and change of variables.
- Useless numerically, but useful analytically.

What can we say about P_{ψ^2}?

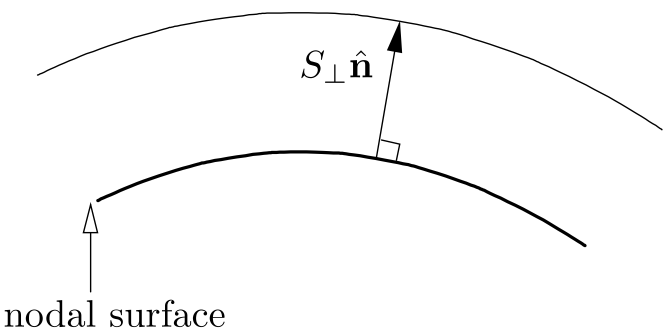
Singularities in the local energy:

$$E_L(\mathbf{R}) = -\frac{1}{2}\frac{\nabla^2_{\mathbf{R}}\psi}{\psi} + \sum_{i<j}\frac{1}{r_{ij}} - \sum_i \frac{Z}{r_i}$$

- $E_L(\mathbf{R}) = E_{tot}$ if the trial wavefunction, ψ, is exact
- Enforce Kato cusp conditions $\longrightarrow$ no Coulomb singularities
- Nodal surface is $\psi = 0$, and is $3N-1$ dimensional
- Kinetic energy part gives singularity on a $3N-1$ dimensional surface
- Singularities provide information about P_{ψ^2} for large $|E|$

What can we say about P_{ψ^2}?

$$P_{\psi^2}(E) = \int_{E=E_L} \frac{P(\mathbf{R})}{|\nabla_{\mathbf{R}} E_L|} d^{3N-1}\mathbf{R}$$



$$\begin{aligned} \psi &= a_1 S_\perp + \ldots \\ E_L &= b_{-1} S_\perp^{-1} + \ldots \\ P(\mathbf{R})/|\nabla E_L| &= c_4 S_\perp^4 + \ldots \\ P_{\psi^2}(E) &= d_{-4} E^{-4} + \ldots \end{aligned}$$

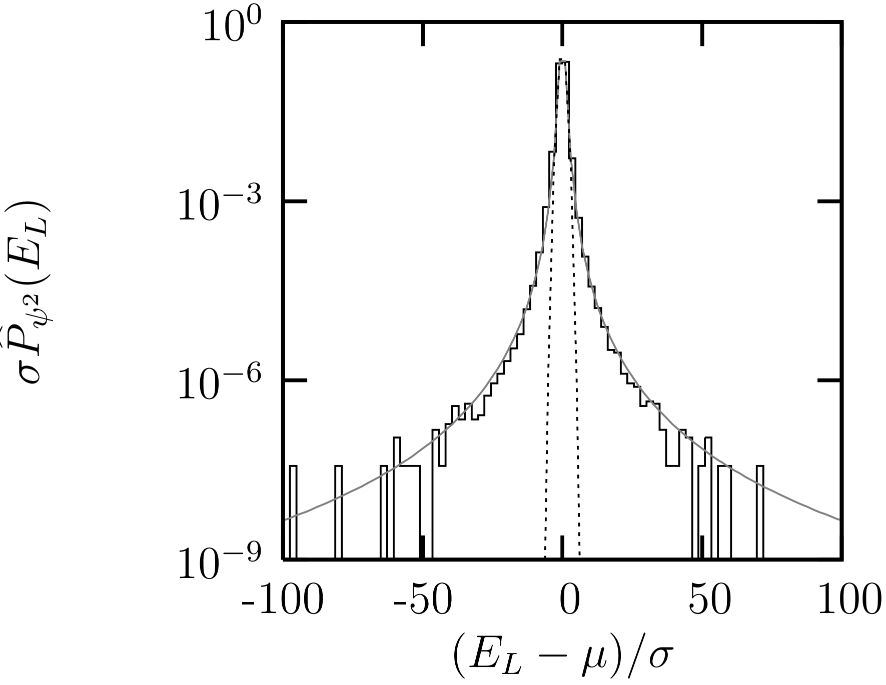
or more completely

$$P_{\psi^2}(E) = (E - E_0)^{-4} \left(e_0 + \frac{e_1}{(E - E_0)} + \ldots \right) \quad |E| \gg E_0$$

Example: All-electron isolated Carbon atom

- Jastrow + 48 determinants + backflow:

$$\psi = e^{J(\mathbf{R})} \sum_m a_m D_m^{\uparrow}(\mathbf{R}') D_m^{\downarrow}(\mathbf{R}') \quad with \quad \mathbf{R}' = \mathbf{R}'(\mathbf{R})$$



Estimated seed probability density function

- 93% correlation energy at VMC level

Also shown is $\frac{\sqrt{2}}{\pi} \frac{\sigma^3}{\sigma^4 + (E-\mu)^4}$, and a Normal distribution

Random error in total energy estimate

$$\text{Est}\left[E_{tot}\right] = \frac{1}{r}\left(E_1 + \ldots + E_r\right)$$

Product of probability of r samples energies that add up to $rE_{tot} \longrightarrow$ convolution integrals

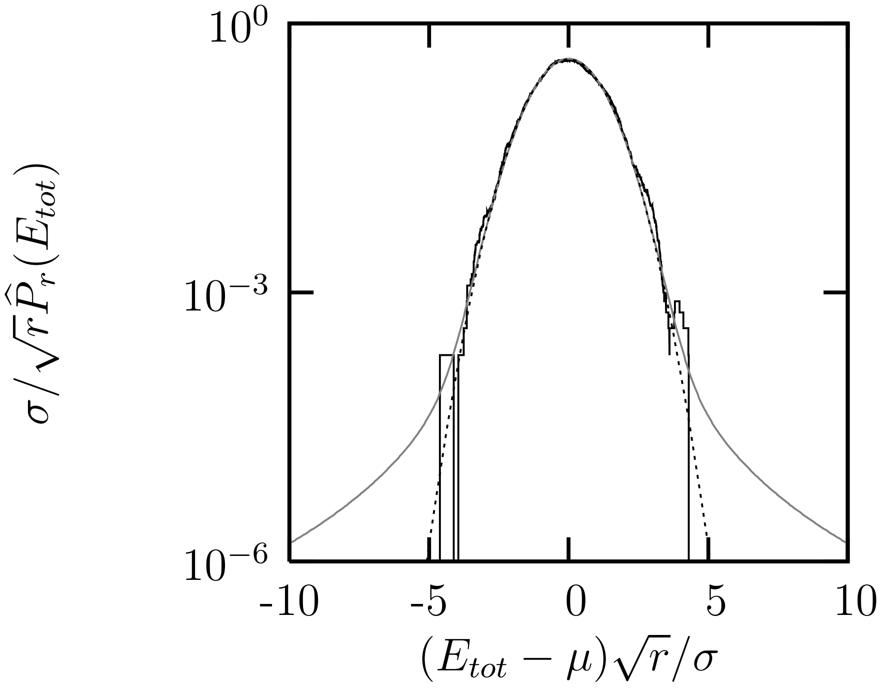
$$\begin{aligned} P_{r=2}(2E_{tot}) &= \int P_{\psi^2}(E_1)P_{\psi^2}(E_2)\delta(E_1+E_2-2E_{tot})dE_1dE_2 = P_{\psi^2}\star P_{\psi^2} \\ P_r(rE_{tot}) &= P_{\psi^2}\star P_{\psi^2}\star\ldots\star P_{\psi^2} \end{aligned}$$

- Take Fourier transform of $P_{\psi^2}(E)$
- Take the r^{th} power
- Take the inverse Fourier transform
- Rescale some variables to get the PDF of averages instead of sum

$$P_r(y) = \frac{1}{\sqrt{2\pi}}\left[1+\frac{\eta}{\sqrt{r}}\frac{d^3}{dy^3}+\mathcal{O}\left(\frac{1}{r}\right)\right]e^{-y^2/2}+\left[\frac{\lambda}{3\pi}\frac{1}{\sqrt{r}}\frac{d^3}{dy^3}D\left(\frac{y}{\sqrt{2}}\right)+\mathcal{O}\left(\frac{1}{r}\right)\right]$$

$$y = (E_{tot}-\mu)/\sigma$$

PDF of estimate of Total energy



- Approximate PDF from 10^4 estimates of total energy, with $r = 10^3$
- For small $|E|$, PDF is dominated by $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- For large $|E|$, PDF is dominated by $\frac{\sqrt{2}}{\pi} \frac{\lambda}{\sqrt{r}} 1/x^4$ ($\lambda \approx 1$ for Carbon trial wavefunction)
- CLT is true in its weakest form

PDF of estimate of the 'Residual Variance', v

Optimisation of wavefunctions using the 'residual variance', v

$$\left(\hat{H} - E_{tot}\right)\psi = \delta = \left(E_L - E_{tot}\right)\psi$$

$$v = \int \delta^2 d\mathbf{R} \geq 0, \text{ and zero for exact } \psi$$

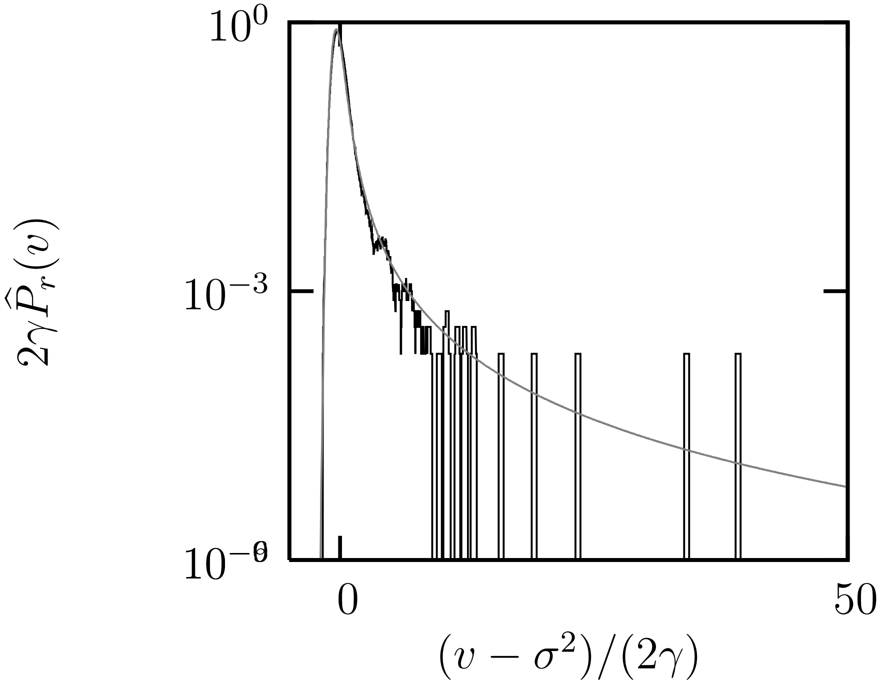
- To optimise the wavefunction v is usually minimised
- Analyse effect of tails, as before:

$$\begin{aligned} P_r(\overline{v}) &= \frac{\sqrt{3}}{\pi}\frac{1}{2\gamma}\left[\frac{\overline{v}-\sigma^2}{2\gamma}\right]^2 \exp\left(\left[\frac{\overline{v}-\sigma^2}{2\gamma}\right]^3\right) \\ &\times \left[-\text{sgn}\left[\overline{v}-\sigma^2\right] K_{1/3}\left(\left|\frac{\overline{v}-\sigma^2}{2\gamma}\right|^3\right) + K_{2/3}\left(\left|\frac{\overline{v}-\sigma^2}{2\gamma}\right|^3\right)\right] \end{aligned}$$

with the 'width' of the PDF decided by the magnitude of the tails

$$\gamma = \left[\frac{6\lambda^2}{\pi r}\right]^{1/3} \sigma^2 \tag{1}$$

PDF of estimate of the 'Residual Variance', v



- Approximate PDF from 10^4 estimates of residual variance, with $r = 10^3$
- Small v, $\propto e^{x^3}$. Large $v \propto 1/x^{5/2}$
- PDF has no variance, γ has no vigorous statistical estimate and is $\propto r^{-1/3}$

- CLT is valid in its weakest form for the total energy
- CLT not valid for residual variance
- CLT is likely to be invalid for estimates of other physical quantites
- **Because**: ψ^2 samples E_L rarely where it is largest, at the nodal surface

Can the CLT be reinstated ?

Residual sampling

Instead of sampling with $P = A\psi^2$, sample with $P = A\psi^2/w$, then

$$\text{Est}\left[E_{tot}\right] = \frac{\sum wE_L}{\sum w} = \frac{\langle\psi|\hat{H}|\psi\rangle + \text{Y}}{\langle\psi|\psi\rangle + \text{X}}$$

and the residual variance,

$$\text{Est}\left[\int \delta^2 d\mathbf{R}\right] \quad = \quad \frac{\sum w(E_L - E_{tot})^2}{\sum w} = \frac{\int \psi^2(E_L - E_{tot})^2 d\mathbf{R} + \text{Y}}{\int \psi^2 d\mathbf{R} + \text{X}}$$

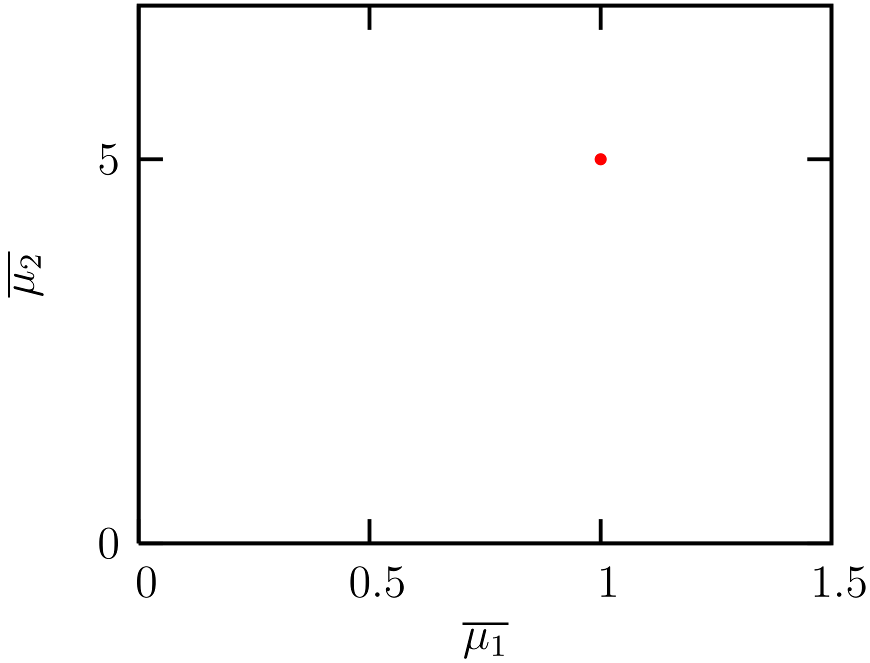
Choose the weighting function

$$w(E_L) = \frac{\epsilon^2}{(E_L - E_0)^2 + \epsilon^2}$$

to ‘interpolate’ beween sampling the numerator and denominator perfectly.

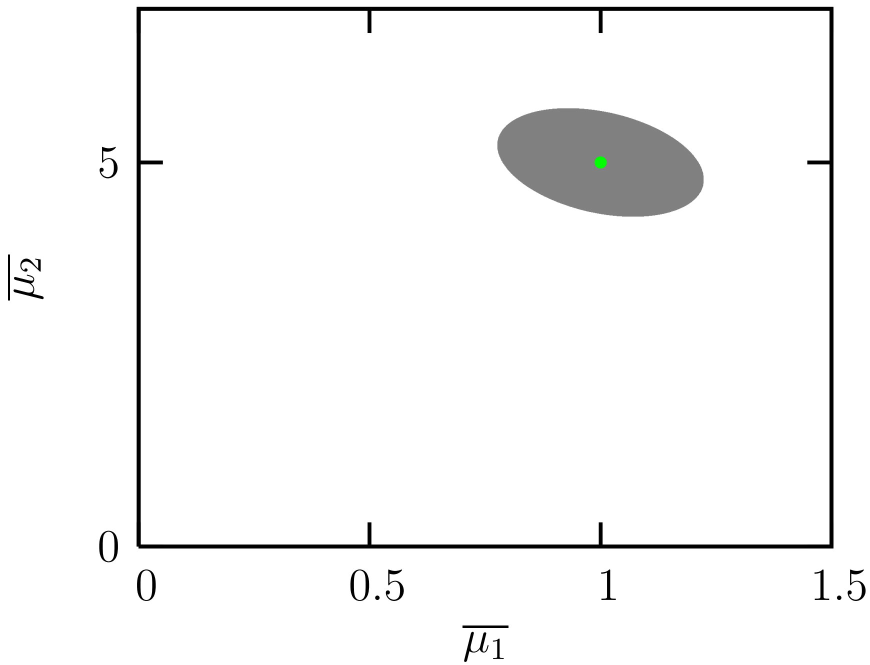
- No singularities, and no power law tails
- Quotient of two correlated random variables, each a sum of random variables

Fieller's Theorem



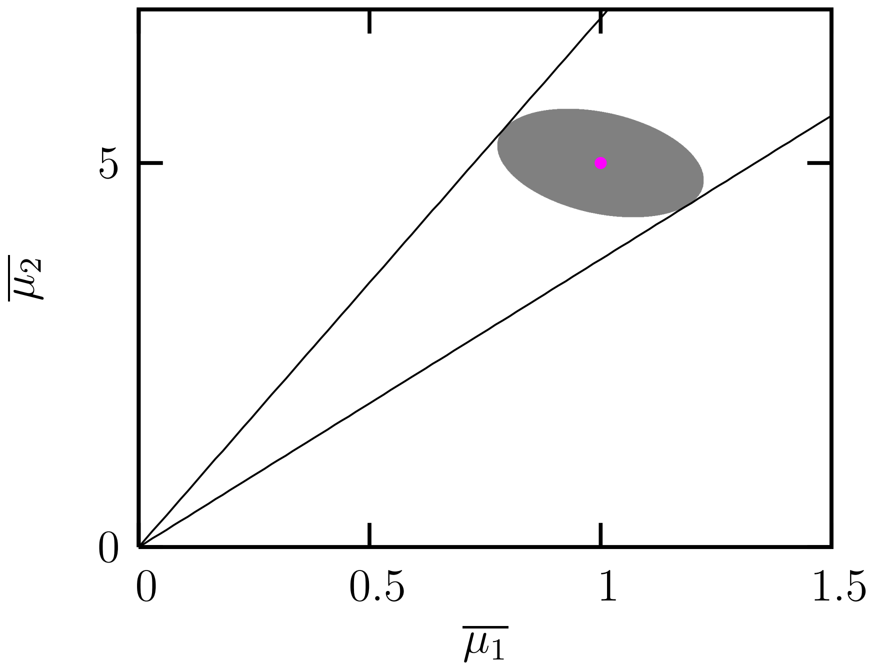
- (μ_2, μ_1) that give Est $= \mu_2/\mu_1$

Fieller's Theorem



- (μ_2, μ_1) that give Est $= \mu_2/\mu_1$
- Ellipse containing 39% of probability from covariance matrix and bivariate CLT

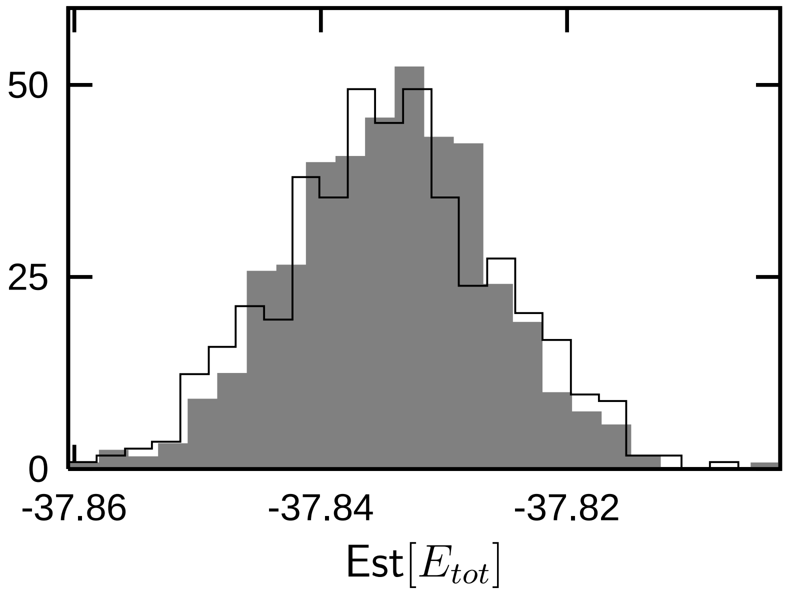
Fieller's Theorem



- (μ_2, μ_1) that give Est $= \mu_2/\mu_1$
- Ellipse containing 39% of probability from covariance matrix
- Wedge that contains 68.3% of probability

$\Rightarrow m_1 < \mu_2/\mu_1 < m_2$ with confidence 68.3%

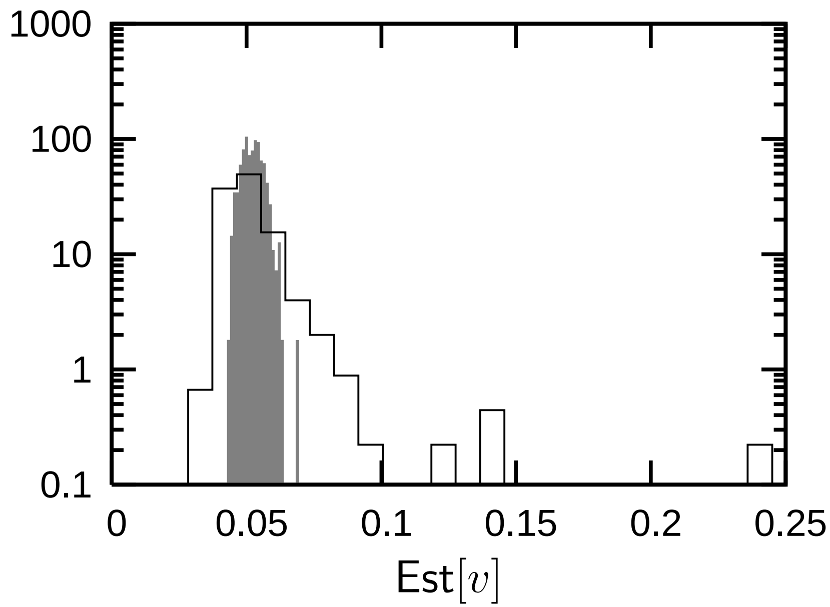
Estimate of total energy



Histogram of 10^3 total energy estimates, each total energy estimate from 10^3 configurations.

- Residual sampling (filled) and standard sampling (unfilled) are not significantly different
- Residual sampling reduces error by $\sim 30\%$
- For other systems standard sampling may give ‘power law’ outliers (depending on λ)
- For all systems residual sampling does not give ‘power law’ outliers

Estimate of residual variance



Histogram of 10^3 residual variance estimates, each estimated from 10^3 configurations.

- Residual sampling and standard sampling are very different
- Standard sampling shows the $v^{-5/2}$ tails and outliers expected
- Residual sampling gives well defined confidence limits for estimate via the bivariate CLT
- Standard sampling does not

Optimisation

$1)$ Take samples using wavefunction with parameter α_0, $\{\mathbf{R}\}_r$

$2)$ These define random sample from a distribution of Optimisation functions, $\mathsf{O}(\alpha)$

$3)$ Find the minimum of $\mathsf{O}(\alpha)$, at $\alpha = \alpha_{min}$

$4)$ set $\alpha_0 = \alpha_{min}$, and return to $1)$

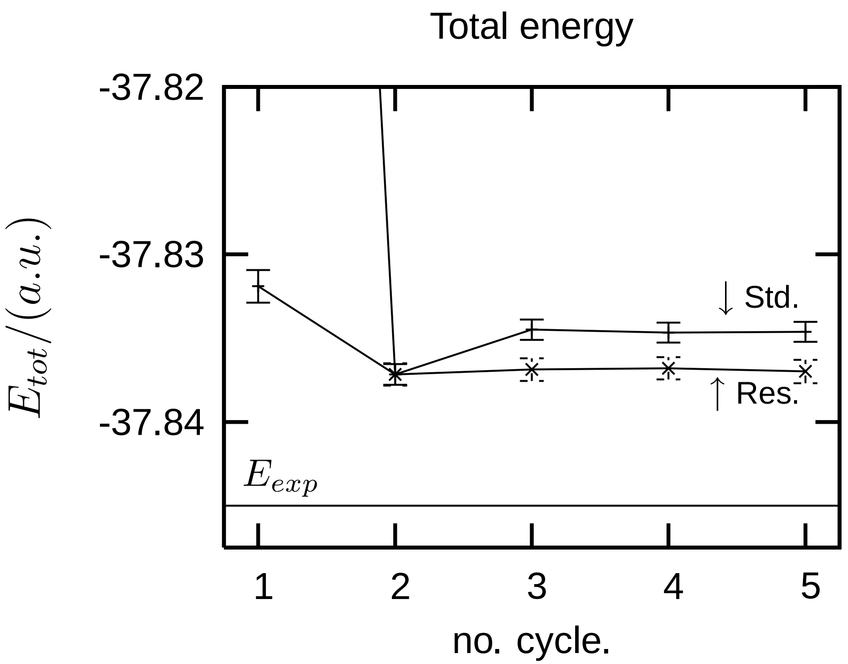
What is the distribution of $\mathsf{O}(\alpha)$?

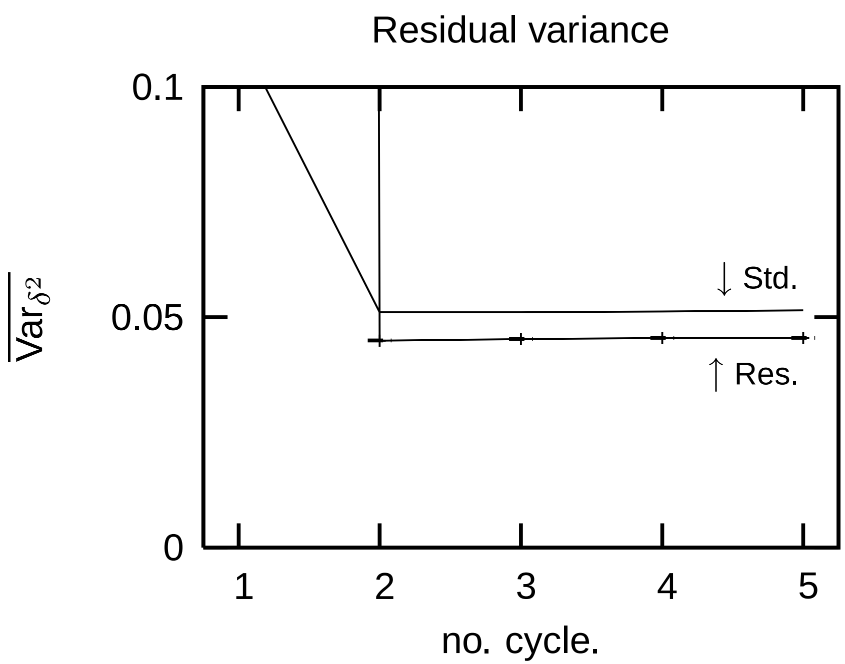
$$\mathsf{O}(\alpha) = \frac{\mathsf{a}_0 + \mathsf{a}_1(\alpha - \alpha_0) + \ldots}{\mathsf{b}_0 + \mathsf{b}_1(\alpha - \alpha_0) + \ldots}$$

- For each type of sampling/optimisation this expansion gives statistics of random error
- For 'Standard Sampling' error is not normal unless the nodes are suppressed by introducing a weight function into $\mathsf{O}(\alpha)$ solely for this purpose (eg $\int g(E_L)\delta^2 d^{3N}\mathbf{R}$).
- For Residual sampling the error is always normal

Optimisation

$r = 10^5$ configurations.





Std. - standard sampling with nodal surface 'suppressed' ($93\%\ E_{corr}$)

Res. - residual sampling of residual variance ($95\%\ E_{corr}$)

- New sampling provides lower total energy and a lower residual variance than standard sampling with nodal surface 'suppressed'

Conclusions

- We cannot assume the CLT is true for estimates in ‘standard sampling QMC’
- ‘r large’ enough must be shown to be true for each estimate in ‘standard sampling QMC’
- The CLT can be reinstated by using an alternative sampling strategy
- Random functions whose minimum gives ‘optimum’ wavefunctions are not generally normally distributed. Some do not converge as $r \longrightarrow \infty$
- The residual sampling strategy can guarantee that the CLT is valid for estimates and optimisation functions, as long as they exist
- With residual sampling optimisation functions can be chosen on physical grounds - to give a good wavefunction at the nodal surface and small *fixed node error* in DMC

Acknowledgements

Financial support was provided by the Engineering and Physical Sciences Research Council (EPSRC), UK, and computational resources were provided by the Cambridge-Cranfield High Performance Computing Facility.