QMC: What are the odds of that?

J.R. Trail

Theory of Condensed Matter Group, Cavendish Laboratory,

University of Cambridge, UK

TTI July 2006

What are the questions?

- What are the statistics of estimates in QMC?
- Is the statistical error kept under control?
- Can better estimates be made?
- What influence does the nodal surface have on all this?

Here VMC and variance minimisation is examined analytically, and numerically for an isolated C atom.

Answered in three sections:

1 - Statistical analysis of ‘standard sampling’ VMC

2 - A new ‘residual sampling’ strategy, and an analysis of its statistics

3 - Statistical analysis of variance minimisation for both standard sampling and residual sampling

1 - Standard VMC

Basic equation of MC:

$$\int_V f d\mathbf{R} \approx V\overline{f} \pm V\epsilon[f], \quad P(\mathbf{R}) = 1/V \tag{1}$$

For estimate of operator $\hat{f}$ ($f = \frac{\hat{f}\psi}{\psi}$) using unormalised many-body trial wavefunction $\psi^2(\mathbf{R})$

$$\langle f \rangle \approx \frac{\overline{\psi^2 f}}{\overline{\psi^2}} \pm \epsilon\left[\psi^2 f, \psi^2\right], \quad P(\mathbf{R}) = 1/V \tag{2}$$

Using importance sampling and *assuming* the CLT is valid:

$$\approx \quad \overline{f} \pm \epsilon\left[f\right], \quad P(\mathbf{R}) = \lambda\psi^2 \tag{3}$$

$$\approx \quad \overline{f} \pm \sqrt{\frac{\overline{\text{Var}[f]}}{r}} \tag{4}$$

- Importance sampling with ψ^2 makes the equations simple. Is it the best choice?
- Does the CLT hold? For r finite samples what replaces it?
- At the nodal surface $\psi^2 \to 0$ and $E_L \to \pm\infty$. This may be bad sampling for $f = f(E_L)$

3N-d distribution → 1-d distribution

Why?: Easier to deal with the general case analytically.

A change of the random variable from spatial to energy:

$$\begin{aligned} \langle E_L \rangle &= \int_V \psi^2 E_L d\mathbf{R} && (5) \\ &= \int_{-\infty}^{\infty} P_{\psi^2}(E) E dE && (6) \end{aligned}$$

with

$$P_{\psi^2}(E) = \int_{E=E_L} \frac{P(\mathbf{R})}{|\nabla_{\mathbf{R}} E_L|} d^{3N-1}\mathbf{R} \tag{7}$$

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

Form of P_{ψ^2} and singularities in $E_L = T_L + V_L$

3 types for electron+atomic nuclei problems:

1 - singularity in nuclear potential part of V_L not cancelled by singularity in T_L

2 - singularity in e-e potential not cancelled by singularity in T_L

3 - singularity in T_L due to zeroes in $\psi(\mathbf{R})$

1&2 can be prevented by enforcing correct cusp conditions on ψ^2, 3 cannot.

Type 3 only

Introduce new co-ordinates $\mathbf{R} = \mathbf{X} + S_\perp \hat{n}$ for expansion about nodal surface:

- $\mathbf{X}$ vector to nodal surface, $S_\perp$ distance $\perp$ to nodal surface

$$\psi^2(\mathbf{R}) = a_2(\mathbf{X}) S_\perp^2 + a_3(\mathbf{X}) S_\perp^3 + \ldots \quad (8)$$

$$E_L(\mathbf{R} + S_\perp \hat{n}) - E_0 = b_{-1}(\mathbf{X}) S_\perp^{-1} + b_0(\mathbf{X}) + b_1(\mathbf{X}) S_\perp + \ldots \quad (9)$$

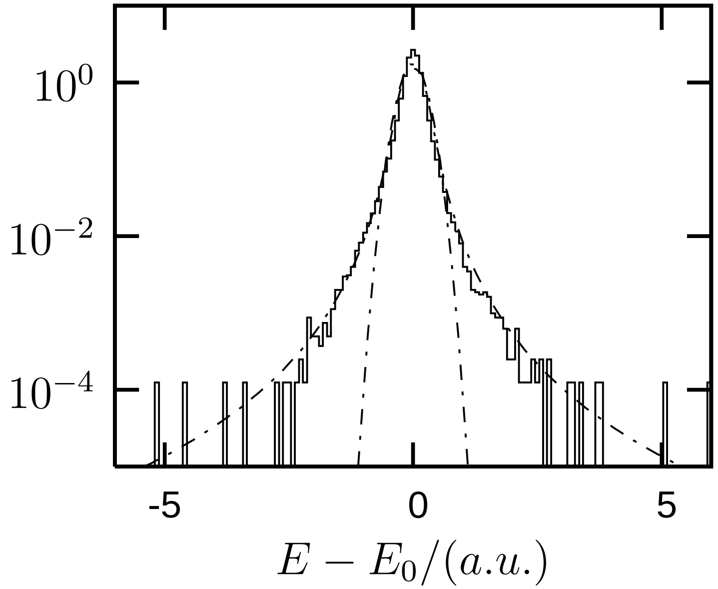
$\Rightarrow$

$$P_{\psi^2}(E) = \frac{1}{(E - E_0)^4} \left(e_0 + \frac{e_1}{(E - E_0)} + \ldots \right) \quad (10)$$

E^{-4} (‘leptokurtotic’ or ‘fat’) tails are general to any trial wavefunction with Type 3 singularities only.

Type 3 singularities only

All-electron Carbon. Trial wavefunction is multideterminant+jastrow+backflow.



Estimated seed probability distribution

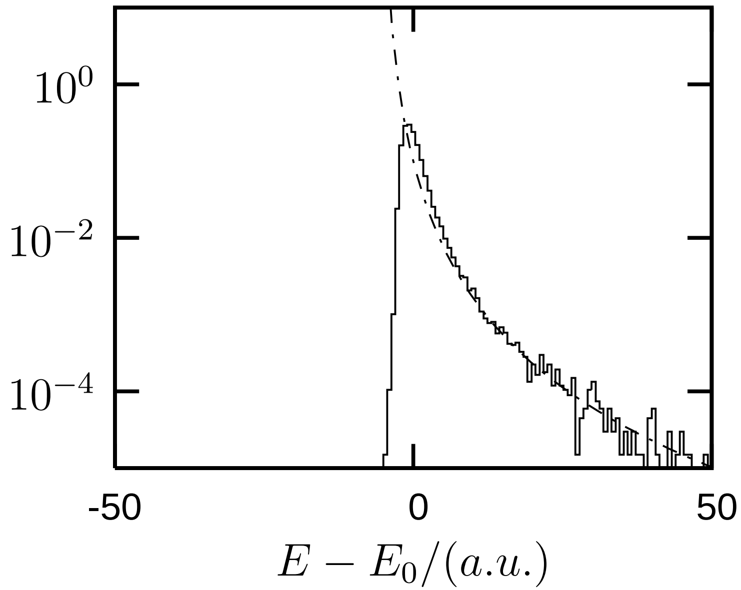
General asymptotic form is:

$$\lim_{|E|\to\infty} P_{\psi^2}(E) = c_3 E^{-4} \quad E \to \pm\infty \tag{11}$$

Also shown are $P_{\psi^2} = \frac{\sqrt{2}}{\pi}\frac{\sigma^3}{\sigma^4+(E-E_0)^4}$, and Gaussian with E_0 and σ the mean and standard deviation of sampled E_L.

Type 2 singularities only

All-electron C. Trial wavefunction is HF determinant.



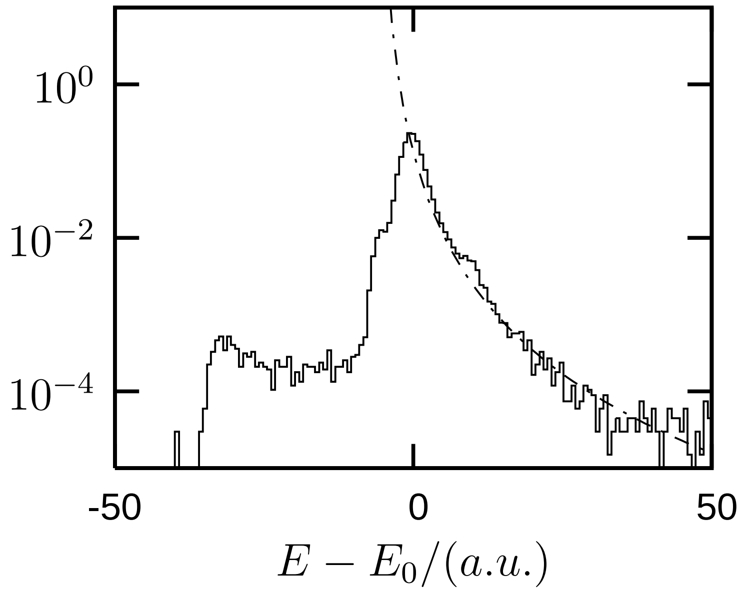
Estimated seed probability distribution

General asymptotic form is:

$$\lim_{|E|\to\infty} P_{\psi^2}(E) = \begin{cases} c_2 E^{-4} & E \to +\infty \\ 0 & E \to -\infty \end{cases} \tag{12}$$

Type 1 & Type 2 singularities

All-electron C. Trial wavefunction is HF determinant with Gaussian basis.



Estimated seed probability distribution

General asymptotic form is:

$$\lim_{|E|\to\infty} P_{\psi^2}(E) = \begin{cases} c_2 E^{-4} & E \to +\infty \\ c_1 E^{-4} & E \to -\infty \end{cases} \tag{13}$$

The Central Limit theorem - summary

Consider a distribution, $p(x)$, mean 0

CLT is derived by finding the distribution of the sum of r x's sampled from $p(x)$:

$$s_r = x_1 + \ldots + x_r \tag{14}$$

The distribution of s_r is given by the convolution relations

$$P_r(s_r) = p(x) \star P_{r-1}(s_{r-1}) \tag{15}$$

Taking the fourier transform of this gives

$$P_r(k) = p(k)^r = e^{r \ln p(k)} \tag{16}$$

IF $p(k)$ is continuous at $k = 0$ **THEN**

- Taylor expansion of $\ln p(k)$ (cumulant expansion)
- Factor out largest term in $P_r(k)$
- Expand the smaller factor as series, and FT back:

$$P_r(\rho) = \frac{1}{\sqrt{2\pi}} e^{-\rho^2/2} \left[1 + \frac{p_1(\rho)}{\sqrt{r}} + \dots \right] \tag{17}$$

with each $p_n(\rho)$ a polynomial in ρ - a Gram-Charlier expansion.*

- As $r \longrightarrow \infty$ $P_r(\rho)$ approaches a Normal distribution.
- Deviations from the normal distribution for finite r decay away exponentially in ρ
- Deviations from the normal distribution for finite r decay away as $1/r^{1/2}$

$* \rho = \frac{\sqrt{r}}{\sigma} \left(\overline{E} - \mu \right)$

BUT, a general property of fourier transforms is

$$FT \tag{18}$$

$$p(x)|_{x\to\pm\infty} \sim 1/|x|^q \quad \longrightarrow \quad p(k)|_{k\to\pm 0} \sim |k|^{q-1} \tag{19}$$

For our trial wavefunctions the seed distribution $P_{\psi^2}(E) \sim 1/E^4$

This means there is $|k|^3$ discontinuity in the FT of $P_{\psi^2}(E)$, so no cumulant or Gram-Charlier expansion is possible.

CLT for total energy estimate

Rescale energy variables so ‘seed’ distribution has mean 0 and variance 1, $P_{\psi^2}(E) \rightarrow p(x)$.

$$s_r = x_1 + \ldots + x_r \tag{20}$$

distribution given by convolution

$$P_r(s_r) = p(x) \star P_{r-1}(s_{r-1}) \quad , \quad P_r(k) = p(k)^r = e^{r \ln p(k)} \tag{21}$$

$p(k)$ can be expanded about $k = 0$ as

$$P_r(k) = \exp\left[-r\frac{1}{2}k^2 + r\frac{\lambda}{3\sqrt{2}}|k|^3 + \ldots\right] \tag{22}$$

with λ a measure of the magnitude of the E^{-4} tails, and not related to the mean or average of $P_{\psi^2}(E)$.

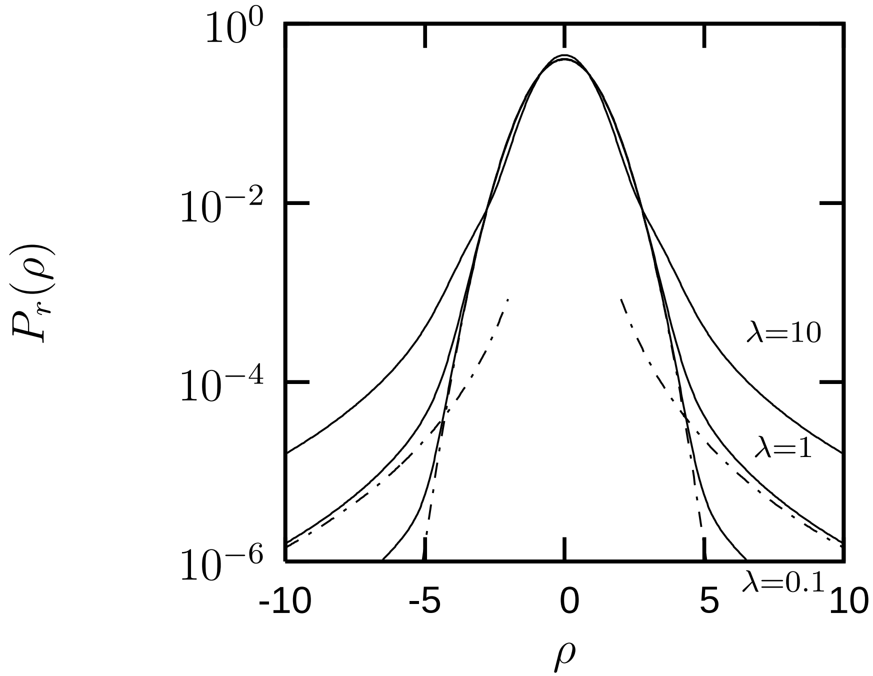
Factoring, series expansion of smaller factor, and inverse transformation gives*

$$P_r(\rho) = \phi_0(\rho) + \frac{\lambda}{3\sqrt{2r}}\phi_1(\rho) + \dots \tag{23}$$

- $\phi_0(\rho) = \frac{1}{\sqrt{2\pi}} e^{-\rho^2/2}$, with mean and variance as before
- $\lim_{\rho \to \pm\infty} P_r(\rho) = \sqrt{\frac{2}{r}} \frac{1}{\pi} \frac{\lambda}{\rho^4}$
- CLT is valid.
- Deviations from the normal distribution for finite r decay away as $1/\rho^4$.

$* \; \rho = \frac{\sqrt{r}}{\sigma}\left(\overline{E} - E_0\right)$

Total energy estimate for finite r ?



Distribution of errors in the total energy estimate - $r = 10^5$

- Crossover between Gaussian and $1/\rho^4$ occurs at $\rho_c^2 \approx \ln \frac{\pi r}{4\lambda^2}$
- For $\lambda = 1, r > 10^3$ then confidence of $< 99.99\%$ is CLT
- For $\lambda > 10, r > 10^3$ then finite r effects lower confidence
- Depends weakly on r/λ^2, with λ an unknown property of the trial wavefunction.
- For all cases probability of an outlier does not decrease exponentially, but much slower.

CLT for variance of the local energy

Same strategy as before, but sum of $x^2 - 1$:

Rescale energy variables to $u = x^2 - 1$ and $p(u) \to 1/u^{5/2}$ as $u \to \infty$

Find the distribution of the sum of r u's sampled from $p(u)$:

$$s_r = x_1^2 + \ldots + x_r^2 - r = u_1 + \ldots + u_r \tag{24}$$

distribution given by convolution

$$P_r(s_r) = p(u) \star P_{r-1}(s_{r-1}) \quad , \quad P_r(k) = p(k)^r = e^{r \ln p(k)} \tag{25}$$

and expansion about $k = 0$

$$P_r(k) = \exp\left[-r \frac{4\lambda}{3\pi^{1/2}} (1 \mp i)|k|^{3/2} + rk^2 + \ldots \right] \tag{26}$$

Factoring, series expansion of smaller factor, and inverse transformation gives*

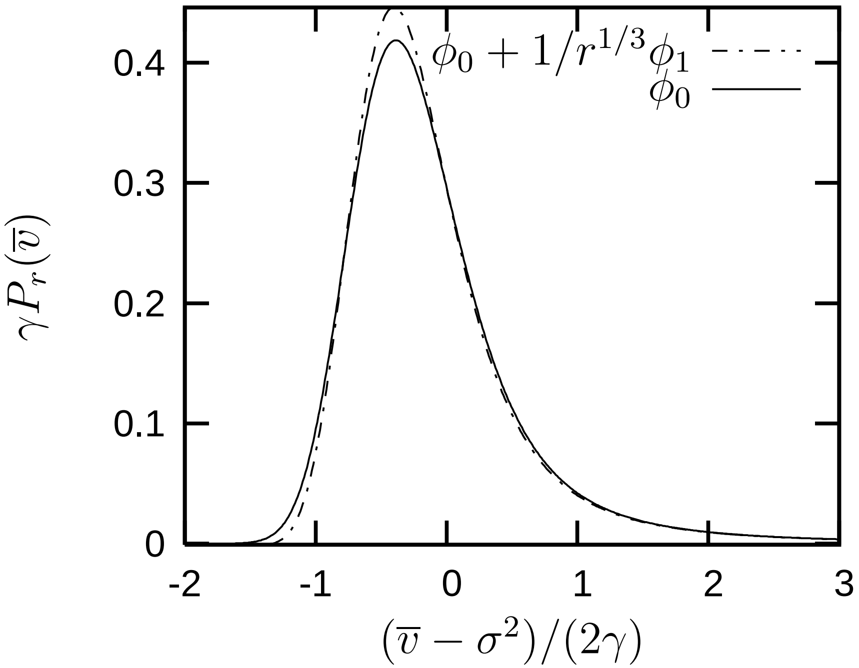
$$P_r(\overline{v}) = \frac{\sqrt{3}}{\pi}\frac{1}{2\gamma}\left[\frac{\overline{v}-\sigma^2}{2\gamma}\right]^2 \exp\left(\left[\frac{\overline{v}-\sigma^2}{2\gamma}\right]^3\right) \times \left[-\text{sgn}\left[\overline{v}-\sigma^2\right] K_{1/3}\left(\left|\frac{\overline{v}-\sigma^2}{2\gamma}\right|^3\right) + K_{2/3}\left(\left|\frac{\overline{v}-\sigma^2}{2\gamma}\right|^3\right)\right] \quad (27)$$

with the ‘width’ of the distribution decided by the magnitude of the tails

$$\gamma = \left[\frac{6\lambda^2}{\pi r}\right]^{1/3} \sigma^2 \quad (28)$$

- Not a normal distribution in the limit $r \to \infty$
- γ is not related to moments of seed distribution

$* \; \overline{v} = \overline{\text{Var}[E_L]}$



Distribution of errors in the variance estimate - $r = 10^3$

- CLT is not valid (variance is infinite). Law of large number (LLN).
- A sample is most likely to be below mean, and outliers are very likely.
- Outlier probablility falls of as $1/\overline{v}^{5/2}$, and not exponentially.
- Confidence limits defined via CLT are not valid. A new definition needs λ, and will scale as $r^{-1/3}$

4^{th} moment, μ_4?

Same strategy as before, but sum of x^4

Obtain distribution of $u = x^4 - 1$

- $P_r(k) \sim \exp[-ark^{3/4} + \ldots]$
- $P_r(\mu_4) \asymp r^{1/4}/\mu_4^{7/4}$
- $P_r(\mu_4)$ gets wider as r increases
- $P_r(\mu_4)$ has infinite mean and variance
- neither CLT or LLN are valid $\longrightarrow$ no statistical convergence

Conclusion

- CLT applies to energy estimate for large enough r.
- Outliers are not exponentially unlikely for $r < \infty$.
- CLT does not apply to variance estimates as r increases. LLN does.
- Neither LLN or CLT apply to higher moments than the variance.
- Error in the variance estimate are unknown (unless we stop being rigorous), but does go to zero.

2. ‘Residual Sampling’ - can the CLT be reinstated?

Use importance sampling with a different sampling distribution - not ψ^2

$$\langle f(E_L)\rangle \approx \frac{\overline{w(E_L)f(E_L)}}{\overline{w(E_L)}} \pm \epsilon\,[wf, w]\,, \quad P(\mathbf{R}) = \lambda\psi^2/w(E_L) \tag{29}$$

choose

$$w = \frac{\epsilon^2}{(E_L - E_0)^2 + \epsilon^2} \tag{30}$$

Why?:

- $P(\mathbf{R})$ is now non-zero and smooth over the nodal surface.
- $\epsilon \longrightarrow \infty$ gives standard sampling.
- Estimate of error is different - ratio of two random variables.
- Sample from $P(\mathbf{R})$ with Metropolis

Error from the Bivariate CLT

Define $\overline{\mu_2} = \overline{wf}$ and $\overline{\mu_1} = \overline{w}$

The pair $\overline{\mu_2}, \overline{\mu_1}$ from r samples is a 2d random vector sampled from the distribution

$$P_r(\overline{\mu_2}, \overline{\mu_1}) = \frac{1}{2\pi} \frac{1}{\sqrt{c_{11}c_{22} - c_{12}^2}} e^{-q^2/2} \tag{31}$$

$$q^2 = \frac{1}{c_{11}c_{22} - c_{12}^2} \left[c_{22} \left(\overline{\mu_1} - \mu_1 \right)^2 - 2c_{12} \left(\overline{\mu_1} - \mu_1 \right) \left(\overline{\mu_2} - \mu_2 \right) + c_{11} \left(\overline{\mu_2} - \mu_2 \right)^2 \right] \tag{32}$$

and

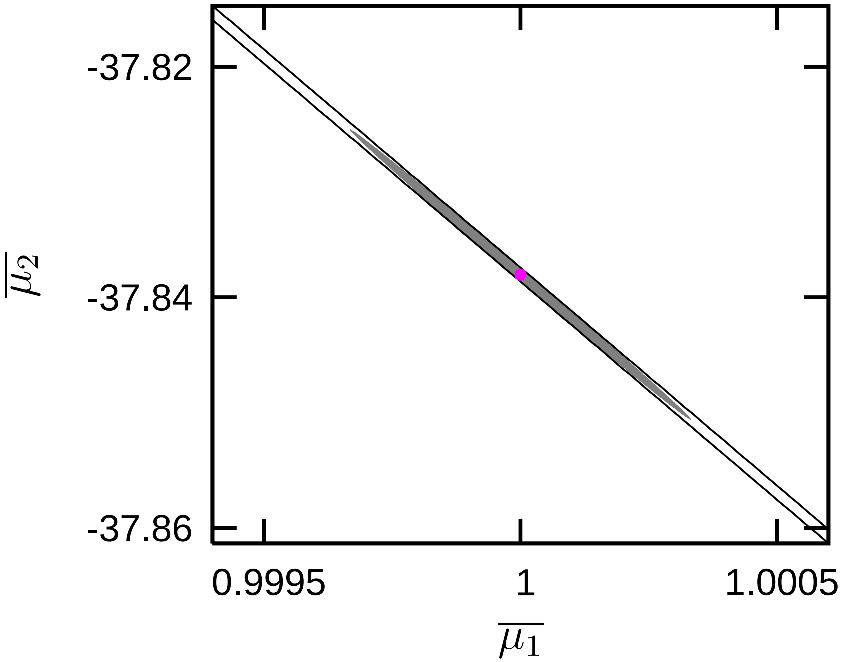
$$\begin{aligned} c_{22} &= \frac{1}{r}\overline{(wf - \mu_2)^2} \\ c_{12} &= \frac{1}{r}\overline{(wf - \mu_2)(w - \mu_1)} \\ c_{11} &= \frac{1}{r}\overline{(w - \mu_1)^2} \end{aligned} \tag{33}$$

$f = E_L$ gives distribution of numerator/denominator for total energy estmate $\overline{wE_L}/\overline{w}$

$f = (E_L - \mu_2/\mu_1)^2$ gives distribution of numerator/denominator for residual variance estimate.

All co-moments exist $\longrightarrow$ CLT is valid, and tails are exponential

Confidence limits



Confidence ellipse and confidence wedge

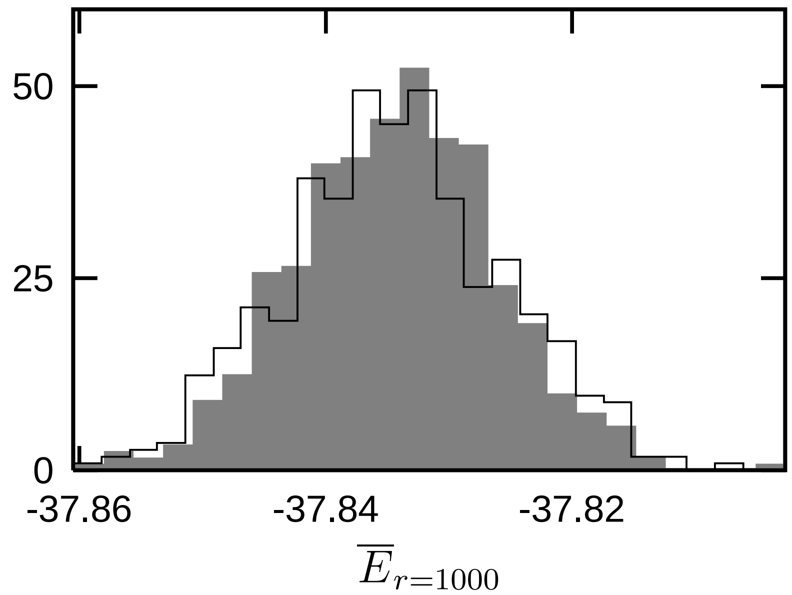
Get confidence limits using Fieller's theorem. Confidence range of $\overline{\mu_2}/\overline{\mu_1}$ is (l_l, l_u) with

$$l_{u/l} = \frac{(\overline{\mu_1}.\overline{\mu_2} - q_0^2\overline{c_{12}}) \pm \sqrt{(\overline{\mu_1}.\overline{\mu_2} - q_0^2\overline{c_{12}})^2 - (\overline{\mu_1}^2 - q_0^2\overline{c_{11}})(\overline{\mu_2}^2 - q_0^2\overline{c_{22}})}}{\overline{\mu_1}^2 - q_0^2\overline{c_{11}}} \quad (34)$$

and $q_0 = \sqrt{2}\text{erf}^{-1}(c)$ defining confidence of c in the estimate of μ_2/μ_1.

- The CLT is now valid for any $f(E_L)$

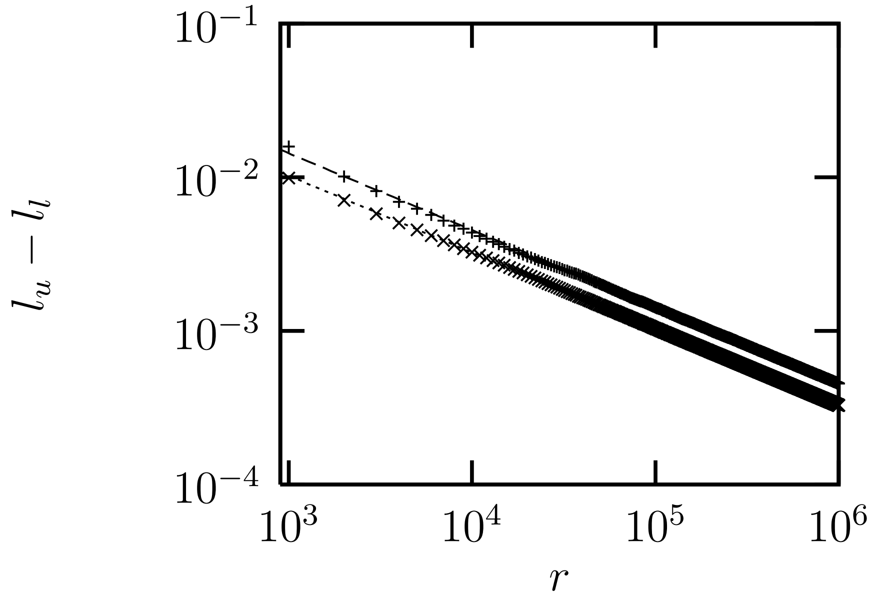
Estimate of total energy



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

- Residual sampling and standard sampling are not significantly different

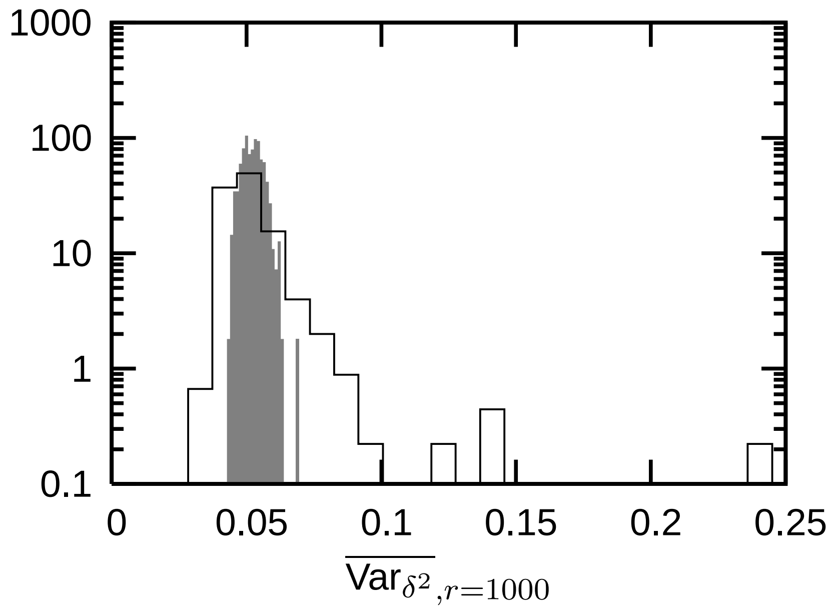
Estimate of error in total energy



Size of confidence interval estimated using CLT for standard, Fieller's theorm for residual sampling.

- Residual sampling and standard sampling are not significantly different
- For both error $\sim 1/r^{1/2}$

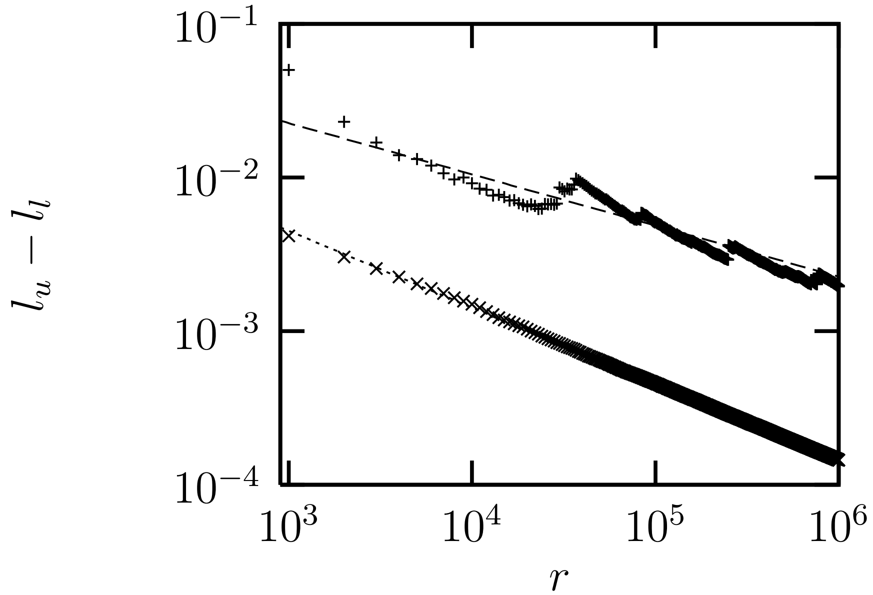
Estimate of variance of local energy



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

- Residual sampling and standard sampling are very different
- Standard sampling shows the $[\text{Var}]^{-5/2}$ tails and outliers expected
- Residual sampling gives well defined confidence limits from the co-moments and bivariate CLT.
- Standard sampling does not.

Estimate error in variance of local energy



Size of confidence interval estimated using CLT expression for standard, and Fieller's theorm for residual sampling.

- Residual sampling and standard sampling are very different
- Standard sampling error $\sim 1/r^{1/3}$ and random noise. Error estimate is not valid.
- Residual sampling error $\sim 1/r^{1/2}$. Error is valid.
- Residual sampling gives well defined confidence limits from the co-moments and bivrariate CLT.
- Standard sampling does not.

The difference is near the nodal hypersurface

Conclusions

- If we want to reintroduce the CLT, and remove the persistent x^{-q} tails in the distribution of estimates, then we can, using residual sampling.

- For the variance this interpolates between sampling the numerator perfectly, and sampling the denominator perfectly.

- Residual sampling gives us well defined confidence limits for the variance in terms of the moments, while standard sampling does not.

- Residual sampling adds 2 new parameters (E_0 and ϵ) but is not sensitive to them. They can be optimised.

3. Variance minimisation and Correlated sampling

- Sample using distribution $P(\alpha_0)$, with α_0 a parameters of the trial wavefunction
- Choose a quantity whose minimum we wish to find, eg total energy:

$$O(\alpha) = \left\langle \frac{P(\alpha)}{P(\alpha_0)} E_L(\alpha) \right\rangle_{P_{\alpha_0}} / \left\langle \frac{P(\alpha)}{P(\alpha_0)} \right\rangle_{P_{\alpha_0}} = \langle f_2(\alpha,\alpha_0)\rangle / \langle f_1(\alpha,\alpha_0)\rangle \tag{35}$$

Expand the averaged quantity in the numerator and denominator as a taylor series, and taking numerical averages gives

$$\overline{O(\alpha)} = \frac{\overline{f_2(\alpha,\alpha_0)}}{\overline{f_1(\alpha,\alpha_0)}} = \frac{\overline{f_2(\alpha_0)} + \overline{f_2'(\alpha_0)}(\alpha-\alpha_0) + \ldots}{\overline{f_1(\alpha_0)} + \overline{f_1'(\alpha_0)}(\alpha-\alpha_0) + \ldots} \tag{36}$$

- What is the statistical error in this estimate of $O(\alpha)$?

  Analyse statistics of each coefficient seperately:

  - Does it converge to a constant as $r \longrightarrow \infty$?
  - Does it obey the CLT?

Example: $O(\alpha) =$ total energy, standard sampling

$\mathbf{X}$ = vector to nodal surface, $\hat{n}$ = vector $\perp$ nodal surface at $\mathbf{X}$, $S_\perp$ = distance $\perp$ to nodal surface

$$
\begin{aligned}
P(\mathbf{R};\alpha) &= a_2(\mathbf{X};\alpha)\left(S_\perp - S_0(\mathbf{X};\alpha)\right)^2 + \ldots \\
E_L(\mathbf{R};\alpha) - E_0(\alpha) &= b_{-1}(\mathbf{X};\alpha)\left(S_\perp - S_0(\mathbf{X};\alpha)\right)^{-1} + \ldots \\
f_2^{(n)}(\mathbf{R}) &= \frac{1}{P(\mathbf{R};\alpha_0)}\frac{d^n}{d\alpha^n}\left[P(\mathbf{R};\alpha)E_L(\mathbf{R};\alpha)\right]_{\alpha_0} \\
f_1^{(n)}(\mathbf{R}) &= \frac{1}{P(\mathbf{R};\alpha_0)}\frac{d^n}{d\alpha^n}\left[P(\mathbf{R};\alpha)\right]_{\alpha_0}
\end{aligned}
\tag{37}
$$

- For each coefficient $\overline{f_{1/2}^{(n)}}$ transform to a 1-D distribution, with the new random variable $x = f_{1/2}^{(n)}(\mathbf{R})$
- This is done by integrating over $f_{1/2}^{(n)}(\mathbf{R}) = x$ hypersurface, as for VMC analysis.
- We get the asymptotic tails of the distribution $p(x)$ whose average is $\overline{f_{1/2}^{(n)}}$

Limit theorems for sample average of $p(x) \asymp \|x\|^{-q}$	
q	Limit theorem
$3 < q$	CLT
$2 < q \leq 3$	LLN
$1 < q \leq 2$	No statistical limit
$q \leq 1$	Not a PDF

- The distribution of the numerator or denominator is the fattest distribution of all the coefficents (for $\alpha \neq \alpha_0$)
- The distribution of the num./den. is bivariate CLT if all coefficients are CLT.
- The distribution of the num./den. is bivariate LLN if all coefficients are CLT or LLN
- The distribution of the num./den. does not converge if any coefficient is not CLT or LLN.

Standard sampling - $P = \lambda \psi_{\alpha_0}^2$

		Numerator			Denominator			Stat. of $O(\alpha)$
Optimate		$n=0$	$n=1$	$n>1$	$n=0$	$n=1$	$n>1$	
Energy	reweighted	CLT	LLN	LLN	CLT	CLT	LLN	bivariate LLN
Variance	reweighted	LLN	LLN	LLN	CLT	CLT	LLN	bivariate LLN
	unweighted	LLN	$\times$	$\times$		exact		$\times$
	limited reweight	LLN	$\times$	$\times$	CLT	CLT	CLT	$\times$
	artificial weight	CLT	CLT	CLT	CLT	CLT	CLT	bivariate CLT

reweighted : $O(\alpha) = \left\langle \frac{\psi^2}{\psi_{\alpha_0}^2} (E_L - \langle E_L \rangle)^2 \right\rangle / \left\langle \frac{\psi^2}{\psi_{\alpha_0}^2} \right\rangle$

unweighted : $O(\alpha, \alpha_0) = \langle (E_L - \langle E_L \rangle)^2 \rangle$

limited reweight : As reweighting, with a maximum $P/P(\alpha_0)$ enforced

artificial weight : $O(\alpha, \alpha_0) = \langle h(E_L)(E_L - \langle E_L \rangle)^2 \rangle \,/\, \langle h(E_L) \rangle$ with $h(E_L) \asymp$ Gaussian in E_L

Residual sampling - $P = \lambda\psi^2_{\alpha_0}/w(\alpha_0)$

		Numerator			Denominator			Stat. of $O(\alpha)$
Optimate		$n=0$	$n=1$	$n>1$	$n=0$	$n=1$	$n>1$	
Energy	reweighted	CLT	CLT	CLT	CLT	CLT	CLT	bivariate CLT
Res. Variance	reweighted	CLT	CLT	CLT	CLT	CLT	CLT	bivariate CLT
	unweighted	$\times$	$\times$	$\times$		exact		$\times$
	limited reweight	$\times$	$\times$	$\times$	CLT	CLT	CLT	$\times$
	artificial weight	CLT	CLT	CLT	CLT	CLT	CLT	bivariate CLT

reweighted : $O(\alpha) = \left\langle \frac{\psi^2}{\psi^2_{\alpha_0}} w(\alpha_0)(E_L - \langle E_L \rangle)^2 \right\rangle / \left\langle \frac{\psi^2}{\psi^2_{\alpha_0}} w(\alpha_0) \right\rangle$

unweighted : $O(\alpha, \alpha_0) = \langle (E_L - \langle E_L \rangle)^2 \rangle$

limited reweight : As reweighting, with a maximum $P/P(\alpha_0)$ enforced

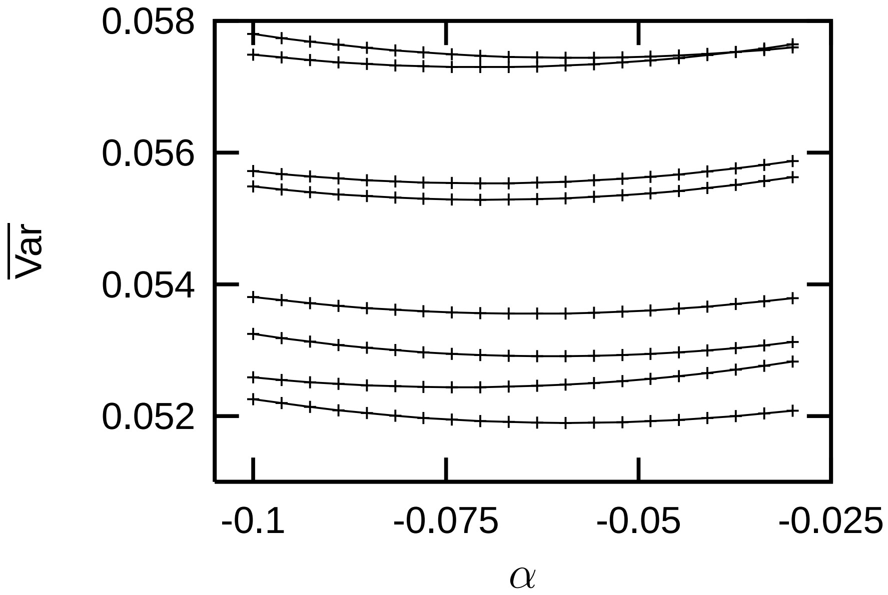
artificial weight : $O(\alpha, \alpha_0) = \langle h(E_L)(E_L - \langle E_L \rangle)^2 \rangle \,/\, \langle h(E_L) \rangle$ with $h(E_L) \asymp$ Gaussian in E_L

Estimated $O(\alpha)$

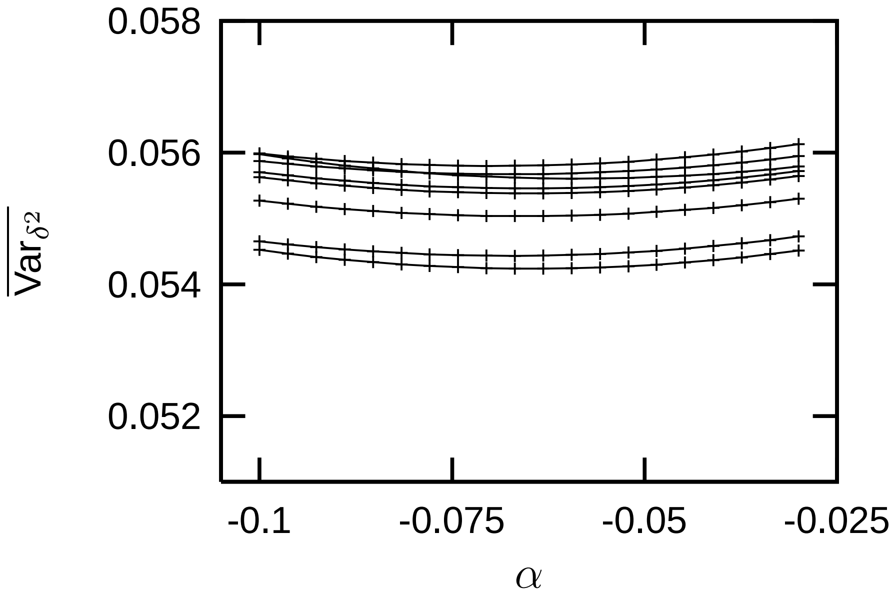
$r = 10^5$ configurations for each of $8\ \overline{O(\alpha)}$'s

Estimate of variance using reweighting

Variance estimated with Standard sampling



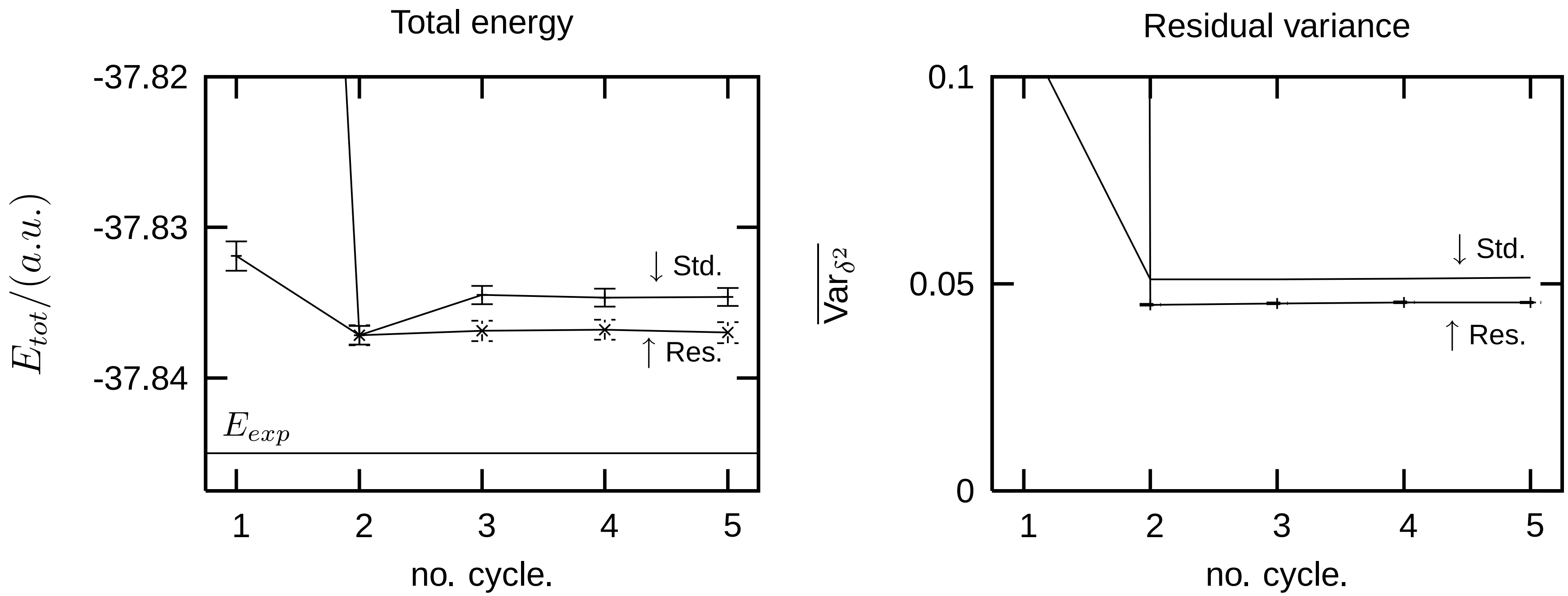
Variance estimated with Residual sampling



- Standard sampling to generate $\overline{O(\alpha)}$ is distributed via LLN
- Residual sampling to generate $\overline{O(\alpha)}$ is distributed via CLT
- Residual sampling provides the best estimate to $\overline{O(\alpha)}$

Optimisation

$r = 10^5$ configurations.



Std. - artificial weights and samples using $\psi(\alpha_0)^2$

Res. - reweighting and samples using $\psi(\alpha_0)^2/w(\alpha_0)$

- The standard method starts with jastrow/multidet. optimised, backflow parameters set to zero
- The residual method starts with jastrow/backflow/multdeterminant parameters set to zero
- Optimisation using reweighting and residual sampling provides a lower energy and variance than standard sampling with artifical weights.

Conclusions

- For standard VMC we cannot assume that CLT and 'r is large enough' apply. Many of the estimates are not distributed as CLT for $r \longrightarrow \infty$.

- A new sampling 'Residual Sampling' with a distribution that is non-zero at the nodal hypersurface reintroduces the CLT for all estimates.

- Optimisation for standard sampling finds the minimum of $\overline{O(\alpha)}$. This is not distributed as CLT, unless the nodal surface is removed from sampling (using artificial weights).

- Optimisation for residual sampling finds the minimum of $\overline{O(\alpha)}$. This is distributed as CLT, with sampling taking place at the nodal surface.

- Optimisation with residual sampling gives the lowest total energy and variance of the local energy, and the lowest statistical error.

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.