Commun. Comput. Phys.
doi: 10.4208/cicp.190313.171013a

Vol. **15**, No. 5, pp. 1352-1367
May 2014

The Lognormal Distribution and Quantum Monte Carlo Data

Mervlyn Moodley*

School of Chemistry and Physics, Quantum Research Group, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban, 4000, South Africa.

Received 19 March 2013; Accepted (in revised version) 17 October 2013

Communicated by Michel A. Van Hove

Available online 26 February 2014

Abstract. Quantum Monte Carlo data are often afflicted with distributions that resemble lognormal probability distributions and consequently their statistical analysis cannot be based on simple Gaussian assumptions. To this extent a method is introduced to estimate these distributions and thus give better estimates to errors associated with them. This method entails reconstructing the probability distribution of a set of data, with given mean and variance, that has been assumed to be lognormal prior to undergoing a blocking or renormalization transformation. In doing so, we perform a numerical evaluation of the renormalized sum of lognormal random variables. This technique is applied to a simple quantum model utilizing the single-thread Monte Carlo algorithm to estimate the ground state energy or dominant eigenvalue of a Hamiltonian matrix.

PACS: 02.70.Ss, 02.50.Cw, 02.50.Ng, 02.50.-r

Key words: Lognormal distribution, renormalization transformation, quantum Monte Carlo.

1 Introduction

Quantum Monte Carlo simulations utilizing the technique of multiplying weights together often give spurious results when one calculates expectation values of operators. Often, one is faced with a dilemma when having to choose a final estimate together with its corresponding error estimate from a set of estimators converging to the exact result. This may arise as a consequence of the estimators developing a distribution that is somewhat different from the Gaussian distribution. Correct statistical inference is based on the assumption that data under consideration adheres to a specific distribution. If this distribution is incorrect, the results obtained after statistical analysis may be invalid [1].

*Corresponding author. *Email address:* moodleym2@ukzn.ac.za (M. Moodley)

With respect to quantum Monte Carlo applications where one is interested in the statistical iteration of some operator, Hetherington [2] observed that the probability distribution of the estimators depends on the number of Monte Carlo iterations. In fact, as shown in this paper, the estimators exhibit a lognormal distribution that has been block-transformed a number of times. This attribute is inherited from the distribution of the product of weights associated with importance sampling. By the central limit theorem, the lognormal distribution should approach the Gaussian limit for a sufficiently large number of block transformations. The estimators however are sometimes not blocked sufficiently often to have reached the Gaussian limit but they do resemble the Gaussian distribution with slight deviations. It would therefore be incorrect to assume that standard statistical analysis, giving the average plus or minus one standard error to be within a 68% confidence interval, is appropriate here.

In the context of statistical and probability theory, a block or renormalization transformation as described in this work, corresponds to the renormalized sum of identically independent random variables. In this paper we consider the sum of lognormal random variables. Sums of lognormal random variables appear in many branches of science [3,4] and finance [5] but most prominently in the field of communications [6,7]. For a historic perspective of finding the distribution of sums of lognormally distributed random variables, see [8]. The difficulty in evaluating these sums of distributions analytically is due to the fact that the characteristic function of the lognormal distribution is not know in closed form and as a result approximation methods are used. Many of these approximations are based on approximating the sum of lognormal variables by another lognormal variable [6,7,9–11]. Other methods of approximation have also been introduced by Beaulieu et al. [12,13]. In the results presented in this work, the sums of lognormal distributions were evaluated simply by using the trapezoidal rule which produced excellent results without resorting to more elaborate numerical integration techniques [14].

The paper is organized as follows. In Section 2 we recall the standard statistical methods applied to a set of data if normality is assumed. In Section 3, we describe a method of calculating the number of times a set of data has undergone a block transformation by relating the cumulants of the blocked data to that of the original data. From this, a recursion relation results which relates successive blocked cumulants. Section 4 focuses on the lognormal distribution and here we construct the block transformed lognormal distribution numerically by first calculating the characteristic function and then Fourier transforming to obtain the probability distribution. A recipe is given in Section 5 to construct the probability distribution of a set of data, with given mean and variance, that has been assumed to be lognormal prior to blocking. By constructing this distribution we are able to give better estimates of the standard errors corresponding to a desired confidence interval. As an application of these methods we consider data obtained from using the single-thread Monte Carlo technique to estimate the groundstate energy (dominant eigenvalue) of a 3×3 symmetric Hamiltonian matrix. This is described in Section 6. Here we consider data with small ensemble sizes that do not ideally converge to an expected value. By constructing the probability distributions of these data, we show that the errorbars are actually asymmetric as compared to those obtained by standard statistical methods.

2 Standard statistical methods

Let $x_1,x_2,\cdots,x_N$ be possible realizations of the stochastic variable X. The most common method of estimating the mean of this set of independent identically distributed data, $\{x_i\}$, is

$$\langle x\rangle=\frac{1}{N}\sum_{i=1}^{N}x_i. \tag{2.1}$$

The spread or uncertainty of the data points from the mean is the standard error, given by the estimate of the standard deviation of the mean,

$$\sigma_{\langle x\rangle}=\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i-\langle x\rangle)^2}. \tag{2.2}$$

Under the assumption that for a large enough value of N, the distribution of the $\{x_i\}$ approach the normal distribution (by the central limit theorem), the previous definitions have precise meanings. Here one can write down an estimate of the data together with an uncertainty or error bar given by $x=\langle x\rangle\pm k\sigma_{\langle x\rangle}$, such that

$$\begin{aligned}&\mathrm{Prob}\big(x\in[\langle x\rangle-k\sigma_{\langle x\rangle},\langle x\rangle+k\sigma_{\langle x\rangle}]\big)\\&=\int_{\langle x\rangle-k\sigma_{\langle x\rangle}}^{\langle x\rangle+k\sigma_{\langle x\rangle}}\frac{1}{\sqrt{2\pi}\sigma_{\langle x\rangle}}e^{-\frac{(x-\langle x\rangle)^2}{2\sigma_{\langle x\rangle}}}dx\\&=\mathrm{erf}\Big(\frac{k}{\sqrt{2}}\Big),\end{aligned} \tag{2.3}$$

where for $k=1$ and $k=2$, the corresponding probabilities are 68% and 95% respectively.

If the data do not conform to the central limit theorem in the sense that they are not numerous enough, are correlated or lack normality, the previous probabilities for the uncertainties cannot be assumed. In this case, one needs to know explicitly what form the distributions take on in order to make any intelligent guess in estimating the uncertainties. This problem is addressed in the following sections where a method is developed to obtain error estimates corresponding to the probabilities mentioned above.

3 The blocking coefficient

Let $\{x_1,x_2,\cdots,x_{2^N}\}$ be a set of independent, identically distributed random data of a stochastic variable X with probability distribution $P_X(x)$. We block transform this set of data into a new set $\{x_1', x_2', \cdots, x_{2^{N-1}}'\}$, corresponding to the stochastic variable X', such that

$$x_i' = \frac{1}{2^{1/\alpha}}(x_{2i-1} + x_{2i}), \tag{3.1}$$

where the characteristic exponent α is chosen to be 2 so that the $\{x_i'\}$ will have the same variance as the $\{x_i\}$. The characteristic function for the transformed data are related to the one for the original data by

$$f_{X'}(k) = \langle e^{\imath k x'} \rangle = \langle e^{\frac{\imath k x}{\sqrt{2}}} \rangle^2 = \left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]^2. \tag{3.2}$$

Now expanding in terms of the cumulants, we have,

$$f_{X'}(k) = \exp\left\{ \sum_{n=1}^{\infty} \frac{(\imath k)^n}{n!} C_n(X') \right\} = \exp\left\{ 2 \sum_{n=1}^{\infty} \frac{(\frac{\imath k}{\sqrt{2}})^n}{n!} C_n(X) \right\}, \tag{3.3}$$

by Eq. (3.2), and by comparing terms one can relate the cumulants of the original data to the cumulants of the blocked data:

$$C_n(X') = \frac{C_n(X)}{2^{(n/2-1)}}. \tag{3.4}$$

Now if the $\{x_i'\}$ are blocked further, say, b times from the $\{x_i\}$, then the cumulants of the b^{th} blocked data, $\{x^{(b)}\}$, are related to the cumulants of the original data by

$$C_n(X^{(b)}) = \frac{C_n(X)}{2^{b(n/2-1)}}. \tag{3.5}$$

So, if the cumulants are known, one can calculate the number of times the $\{x_i\}$ have been block transformed from

$$b = \frac{\log\left|\frac{C_n(X)}{C_n(X^{(b)})}\right|}{\log 2^{(n/2-1)}}. \tag{3.6}$$

It can also be shown that there exists a recursion relation between successive blocks given by

$$C_n(X^{(b)}) = \frac{C_n(X^{(b-1)})}{2^{(n/2-1)}}. \tag{3.7}$$

It should be noted that by choosing $\alpha = 2$, the mean, given by the first cumulant, grows by a factor of $\sqrt{2}$ for successive blocks. It is therefore more convenient to initially transform the data such that $C_1(X) = 0$ and $C_2(X) = 1$. This transformation also leaves b invariant.

4 Constructing the block transformed lognormal distributions

If $\{y_1,y_2,\cdots,y_{2^N}\}$ is normally distributed with $\mu=0$ and $\sigma^2=1$, then $\{e^{y_1},e^{y_2},\cdots,e^{y_{2^N}}\}=\{x_1,x_2,\cdots,x_{2^N}\}$ is lognormally distributed with distribution given by

$$P_X(x)=\frac{1}{x\sqrt{2\pi}}e^{-\frac{1}{2}(\ln x)^2}. \tag{4.1}$$

If the $\{x_i\}$ are block transformed into a new set $\{x_i^{(b')}\}$ such that

$$x_i^{(b')}=\frac{1}{\sqrt{b'}}(x_{b'(i-1)+1}+x_{b'(i-1)+2}+\cdots+x_{b'i}), \tag{4.2}$$

then the characteristic function for the new set is given by

$$f_{X^{(b')}}(k)=\langle e^{\frac{\imath kx}{\sqrt{b'}}}\rangle^{b'}=\Big[f_X\Big(\frac{k}{\sqrt{b'}}\Big)\Big]^{b'}, \tag{4.3}$$

where $b'=2^b$. Now the probability distribution for $x^{(b')}$ can be reconstructed from Eq. (4.3) by taking the Fourier transform such that,

$$P_{X^{(b')}}(x^{(b')})=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{\imath kx}{\sqrt{b'}}}f_{X^{(b')}}(k)dk=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{\imath kx}{\sqrt{b'}}}\Big[f_X\Big(\frac{k}{\sqrt{b'}}\Big)\Big]^{b'}dk, \tag{4.4}$$

which is a function of the original data $\{x_i\}$. There is however no closed form expression for this probability distribution since the characteristic function of a lognormal variable is not known in closed form. It should be noted, that the lognormal distribution is not stable, in the sense that a sum of lognormal random variables is not lognormally distributed. The lognormal distribution under a transformation, as given above, with increasing b' approaches the stable normal distribution. The lognormal distribution is guaranteed to belong to the domain of attraction of the normal distribution since the former distribution has a finite variance given by the blocking transformation having a characteristic exponent of 2 [15]. This corresponds to the renormalization group proof of the central limit theorem [16, 17]. Another important property of the lognormal distribution is that the product and ratio of independent lognormal variables are also lognormally distributed.

We now proceed to show how the blocked probabilities are calculated numerically. Without loss of generality we consider the case of $b'=2$. From Eq. (4.4) above we have

$$P_{X^{(2)}}=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{\imath kx}{\sqrt{2}}}\Big[f_X\Big(\frac{k}{\sqrt{2}}\Big)\Big]^2dk=\frac{1}{\pi}\int_{0}^{\infty}e^{-\frac{\imath kx}{\sqrt{2}}}\Big[f_X\Big(\frac{k}{\sqrt{2}}\Big)\Big]^2dk, \tag{4.5}$$

since the characteristic function has the property that $f_X(-k)=f_X^*(k)$. Our only consideration here is the real part of the probability distribution since the imaginary part vanishes exactly. $P_{X^{(2)}}$ is calculated as follows:

$$\begin{aligned}
P_{X^{(2)}} &= \frac{1}{\pi}\int_0^\infty \Re\left[e^{-\frac{\imath kx}{\sqrt{2}}}\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]^2\right]dk \\
&= \frac{1}{\pi}\int_0^\infty \left\{\Re[e^{-\frac{\imath kx}{\sqrt{2}}}]\Re\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]^2 - \Im[e^{-\frac{\imath kx}{\sqrt{2}}}]\Im\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]^2\right\}dk \\
&= \frac{1}{\pi}\int_0^\infty \left\{\cos(kx/\sqrt{2})\left[\left[\Re\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]\right]^2 - \left[\Im\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]\right]^2\right]\right. \\
&\qquad\left. + 2\sin(kx/\sqrt{2})\Re\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]\Im\left[f_X\left(\frac{k}{\sqrt{2}}\right)\right]\right\}dk,
\end{aligned} \tag{4.6}$$

where

$$\Re f_X\left(\frac{k}{\sqrt{2}}\right) = \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{\cos(kx/\sqrt{2})}{x}e^{-\frac{1}{2}(\ln x)^2}dx \tag{4.7}$$

and

$$\Im f_X\left(\frac{k}{\sqrt{2}}\right) = \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{\sin(kx/\sqrt{2})}{x}e^{-\frac{1}{2}(\ln x)^2}dx. \tag{4.8}$$

$P_{X^{(2)}}$ therefore gives the probability distribution of the lognormal distribution blocked twice, which is the distribution of a sum of two lognormal variables. The above integrals were estimated by using the extended trapezoidal rule. Fig. 1 shows a plot of different blocked probability distributions with the means of the blocked variables, $x^{(b')}$, centered at zero for $b=2$ to $b=12$. Here it is clearly seen that the blocked probability distributions with increasing values of b converge to the normal distribution.

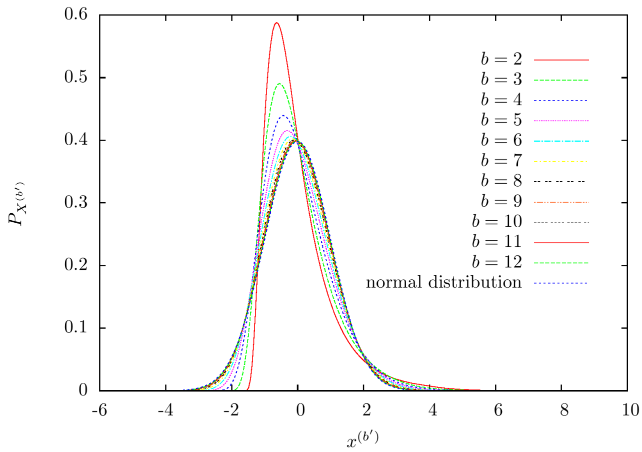

Figure 1: (Color online) Plots of the blocked probabilities for $b'=2^b$ with $b=2,3,4,5,6,7,8,9,10,11,12$ (with means centered at zero) and the normal distribution.

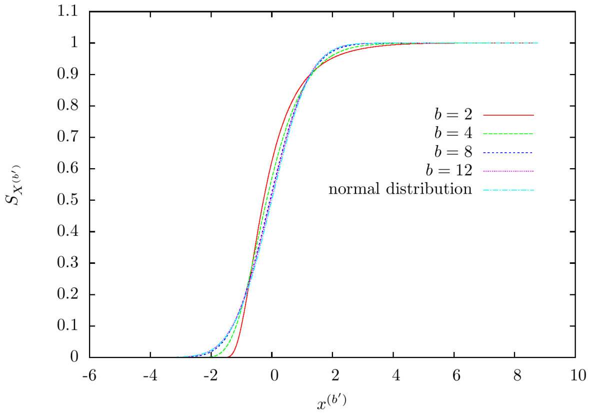

Figure 2: (Color online) Plots of the blocked cumulative distribution functions for $b'=2^b$ with $b=2,4,8,12$ and the normal distribution. Note that the $b=12$ cumulative distribution function coincides almost exactly with the normal cumulative distribution.

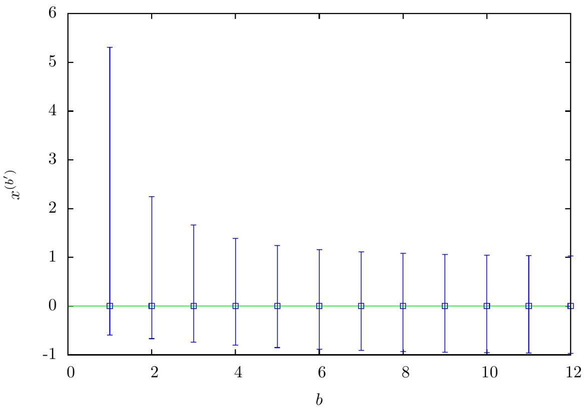

Figure 3: (Color online) Errorbars for the blocked probabilities for the different blocking coefficients b, corresponding to a confidence interval of 68%.

Once these distributions are established, one can construct the blocked cumulative distribution functions, $S_{X^{(b')}}$, as shown in Fig. 2. By appropriately summing the probabilities in $S_{X^{(b')}}$ to the left and right of the mean, one can obtain the standard error corresponding to a desired confidence interval. The errorbars corresponding to a confidence interval of 68% are shown in Fig. 3 for the probability distributions corresponding to different values of b. Note that these errorbars start off asymmetric and converge to the symmetric standard deviation of $\sigma=1$ for the normal distribution.

The cumulative distributions can also be used as the theoretical or reference distribution in the Kolmogorov-Smirnov test. Here one calculates the Kolmogorov-Smirnov statistic,

$$D=\max|S_N(x)-S(x)|, \tag{4.9}$$

which is the maximum absolute difference between a theoretical cumulative distribution

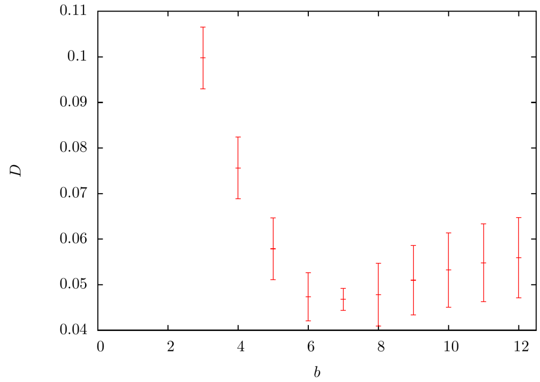

Figure 4: (Color online) The blocking coefficient b versus the Kolmogorov-Smirnov statistic D for lognormal distributions blocked 7 times.

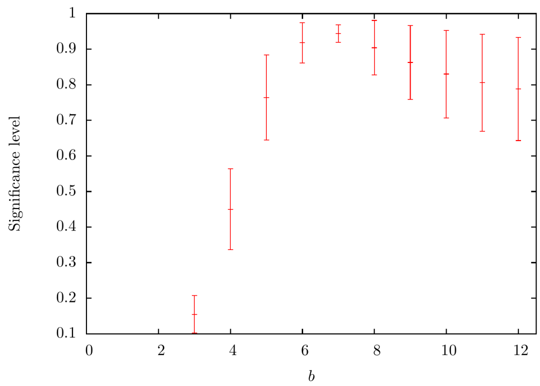

Figure 5: (Color online) The blocking coefficient b versus the significance level of D for lognormal distributions blocked 7 times.

$S(x)$ and an estimator $S_N(x)$ of the cumulative distribution of a set of N data points sampled from the same probability distribution. For a sample size of N, the N data points are arranged in ascending order $x_1,x_2,\cdots,x_N$ and one calculates the cumulative proportions $S_N(x_1)=1/N$, $S_N(x_2)=2/N$ and so on. D is a random variable and as such has some probability distribution. The significance level of D is given by [18],

$$\text{Prob}(D>\text{observed value})=2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2\lambda^2}, \tag{4.10}$$

where $\lambda=D[\sqrt{N}+0.12+0.11/\sqrt{N}]$. To see how accurately the previously developed blocked probability distributions would represent a blocked sample, we considered a sample size of 2^{14} random lognormal variables, blocked them and then applied the Kolmogorov-Smirnov test with the appropriate $S_{X(b')}$ as the reference distribution. For all values of the blocking coefficient we obtained Kolmogorov-Smirnov statistics corresponding to significance levels of about 95%. As an illustration, Fig. 4 and Fig. 5 show respectively a plot of the blocking parameter b versus the Kolmogorov-Smirnov statistic and significance level for lognormal samples blocked 7 times. As can be seen from both figures, their is no doubt that the data do indeed correspond to data with blocking coefficient $b=7$. The Kolmogorov-Smirnov statistic is one of many statistics that can be used to measure the difference between two distributions. See [18] for other statistics and references.

The previous discussion can be generalized to the case where μ is finite and σ^2 is not necessarily unity. Here the probability distribution of the lognormal data, in its most general form, is given by

$$P_X(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \tag{4.11}$$

with moments

$$\langle x^n \rangle = \int_0^\infty x^n P_X(x)dx = e^{n\mu+\frac{1}{2}n^2\sigma^2}, \tag{4.12}$$

where n labels the order of the moments. In this form the numerical integration for the characteristic function is more difficult since the integrands become more rapidly oscillating and have slowly converging envelopes. These integrals however need not be calculated, since one can always transform a given set of data to another with zero mean and variance equal to one. Note also that, given a set of data with distribution $P_X(x)$ one can obtain an approximation for μ and σ^2 from the first and second cumulants, $C_1(X)=m=e^{\mu+\sigma^2/2}$ and $C_2(X)=s^2=e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$. Solving, one obtains,

$$\mu = \ln m^2 - \frac{1}{2}\ln[m^2+s^2] \tag{4.13}$$

and

$$\sigma^2 = \ln\left[\frac{m^2+s^2}{m^2}\right]. \tag{4.14}$$

From these values one can also obtain the true errorbars corresponding to a 68% confidence interval from noting that,

$$\begin{aligned}\mathrm{Prob}\Big(X\in\Big[\frac{e^\mu}{e^\sigma},e^\mu e^\sigma\Big]\Big) &= \int_{e^{\mu-\sigma}}^{e^{\mu+\sigma}} \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{2}\Big[\mathrm{erf}\Big(\frac{1}{\sqrt{2}}\Big) - \mathrm{erf}\Big(\frac{-1}{\sqrt{2}}\Big)\Big] = 68\%. \end{aligned} \tag{4.15}$$

5 Constructing the probability distribution of blocked data that were originally lognormal

In the previous section, we showed how the lognormal distribution was block transformed to give new distributions which represented the sum of lognormal variables. In this section we address the problem on how to construct the probability distribution of a set of data $\{x_i\}$, with mean μ_x and variance σ_x^2, that we assume was originally lognormally distributed prior to undergoing some blocking transformation. Using the methods developed previously we write down the following recipe:

1. Transform the data into new data $\{\tilde{x}_i\}$ with mean equal to zero and variance equal to one via the transformation: $\tilde{x}=(x-\mu_x)/\sigma_x$.
2. Apply the Kolmogorov-Smirnov test to this data with each of the blocked cumulative distributions, $S_{X^{(b')}}$, as the reference distributions. The $S_{X^{(b')}}$ which gives the largest significance level probability, infers the number of times the data has been blocked, i.e., the blocking coefficient b. This also infers the probability distribution $P_{X^{(b')}}$, as constructed in the previous section.
3. Obtain the probability distribution of the data $\{x_i\}$ by the following transformation: $x=\mu_x+\tilde{x}\sigma_x$. This gives the probability distribution of $\{x_i\}$, with mean μ_x and variance σ_x^2, to be

$$P_X(x)=P_{X^{(b')}}\left(\frac{x-\mu_x}{\sigma_x}\right)\frac{1}{|\sigma_x|}. \tag{5.1}$$

One can also apply the recursion relation given by Eq. (3.7), with the above value of b, to the data set $\{x_i\}$ to get the cumulants of the original lognormal distribution. From this, Eq. (4.13) and Eq. (4.14) can be used to obtain the mean (μ) and variance (σ^2) of the original Gaussian distribution. One can then write down an expression for the original lognormal distribution as given by Eq. (4.11). Block transforming this distribution according to b also gives the probability distribution of the data $\{x_i\}$.

6 Application: Single-thread Monte Carlo

As an application of the above methods we consider data, in the form of ground state energies, as obtained from the implementation of the single-thread [19, 20] algorithm to a system defined by a 3×3 symmetric Hamiltonian matrix, $\mathcal{H}$, with elements distributed uniformly in the interval $(-1,0)$. For more details on the single-thread Monte Carlo algorithm, see [20].

Defining the evolution matrix operator, $G=e^{-\tau\mathcal{H}}$, and since $[\mathcal{H},G]=0$ we can estimate the ground state energy (dominant eigenvalue) by

$$\mathcal{E}_{TT}^{(p)}=\frac{\langle\psi_T|G^p\mathcal{H}|\psi_T\rangle}{\langle\psi_T|G^p|\psi_T\rangle}, \tag{6.1}$$

where p is the power of the G matrix or the projection time and ψ_T is a trial wave function. Our estimation in Eq. (6.1) is based on the fact that for sufficiently large p, $\mathcal{E}_{TT}$ approaches the ground state, $\mathcal{E}_0$. Define the trial wave function in a state S as $\psi_T(S)=\phi^\alpha(S)$, where

$\phi(S)$ are the eigenvectors of $\mathcal{H}$ and $0\leq\alpha\leq 1$. The choice of $\alpha=1$ corresponds to ideal importance sampling while that of $\alpha=0$ corresponds to total ignorance of the trial wave function. Performing an importance sampling transformation on G,

$$\hat{G}(S'|S)=\phi^{\alpha}(S')G(S'|S)\phi^{-\alpha}(S), \tag{6.2}$$

we can write $\hat{G}(S'|S)$ in a factorizable form given by,

$$\hat{G}(S'|S)=\hat{g}(S)\hat{P}(S'|S). \tag{6.3}$$

Here $\hat{g}(S)=\sum_{S'}\hat{G}(S'|S)$ is the weight matrix and from this the transition matrix is defined by

$$\hat{P}(S'|S)=\frac{\hat{G}(S'|S)}{\sum_{S'}\hat{G}(S'|S)}. \tag{6.4}$$

Since $G(S'|S)$ is symmetric in our system, it can be shown that $\hat{P}(S'|S)$ has a known stationary distribution: $\sum_S\psi_T(S)G(S'|S)\psi_T(S')\equiv\psi_G^2$.

Now by defining the configuration energy as,

$$\mathcal{E}_T(S)=\frac{\sum_{S'}\langle S|\mathcal{H}|S'\rangle\psi_T(S')}{\psi_T(S)} \tag{6.5}$$

and by repeated insertion of the resolution of the identity in Eq. (6.1), we obtain,

$$\mathcal{E}_{TT}^{(p)}=\frac{\sum_{S_p,\cdots,S_0}\psi_T(S_p)\mathcal{E}_T(S_p)\left[\prod_{i=0}^{p-1}G(S_{i+1}|S_i)\right]\psi_T(S_0)}{\sum_{S_p,\cdots,S_0}\psi_T(S_p)\left[\prod_{i=0}^{p-1}G(S_{i+1}|S_i)\right]\psi_T(S_0)}. \tag{6.6}$$

Eq. (6.6) can be converted into a time average by defining the hatted trial wave function $\hat{\psi}_T(S)=\psi_T(S)/\psi_G(S)$ and noting that

$$\text{Prob}(S_t,S_{t+1},\cdots,S_{t+p})\propto\left[\prod_{i=0}^{p-1}\hat{P}(S_{t+i+1}|S_{t+i})\right]\psi_G(S_t)^2, \tag{6.7}$$

we have

$$\mathcal{E}_{TT}^{(p)}=\lim_{M\to\infty}\frac{\sum_{t=1}^{M}\hat{\psi}_T(S_{t+p})\mathcal{E}_T(S_{t+p})\hat{W}_t(p)\hat{\psi}_T(S_t)}{\sum_{t=1}^{M}\hat{\psi}_T(S_{t+p})\hat{W}_t(p)\hat{\psi}_T(S_t)}, \tag{6.8}$$

where

$$\hat{W}_t(p)=\prod_{i=0}^{p-1}\hat{g}(S_{t+i+1}|S_{t+i}). \tag{6.9}$$

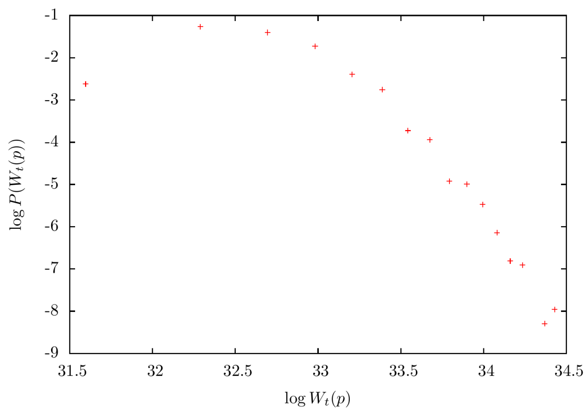

Figure 6: (Color online) A log-log histogram plot of the distribution of $\hat{W}_t(p)$ for $p=20$.

Eq. (6.8) gives a Monte Carlo estimate of the ground state energy. In order to obtain a statistical estimate, one calculates an ensemble of $\mathcal{E}_{TT}^{(p)}$'s for different seeds of the random number generator and then performs standard statistical analysis on this ensemble. These standard statistical techniques are based on the assumption that the $\mathcal{E}_{TT}^{(p)}$'s are Gaussian in nature. From Eq. (6.8) it can be inferred that the distribution of $\mathcal{E}_{TT}^{(p)}$ must depend on the distribution of the product of weights, $\hat{W}_t(p)$. Fig. 6 shows a log-log histogram plot of the distribution of $\hat{W}_t(p)$ for $p=20$ and it is clear that this distribution is lognormal in nature. Due to the time average, the denominator and numerator of Eq. (6.8) are sums of the lognormally distributed variable $\hat{W}_t(p)$ which corresponds to a block transformation. Now, since the distribution of the ratio of two lognormally distributed variables retains the lognormality it can be assumed that the $\mathcal{E}_{TT}^{(p)}$'s are block transformed lognormal variables. Therefore the methods described in the previous sections can be appropriately applied here.

Fig. 7 shows $\mathcal{E}_{TT}^{(p)}$ for different projections for 2^{13} and 2^8 time steps, where the dashed line represents the exact ground state. These values and their corresponding errorbars were calculated using standard statistical methods. With 2^{13} time steps one observes a rapid convergence to the exact value and as expected the corresponding blocking coefficients for all projections were found to be very close to that which gives a normal distribution. With 2^8 time steps on the other hand, due to the number of time steps being small, $\mathcal{E}_{TT}^{(p)}$ does not converge to the exact ground state. In this case the corresponding blocking coefficients for each projection varies. For small values of projections ranging from $p=1$ to $p=3$ the distribution does not vary significantly from the normal distribution. For larger values of p, the blocking coefficients range from $b=7$ to $b=9$.

To illustrate the techniques developed previously we now show the construction of the probability distribution together with its errorbars for the case of $p=5$ for which $b=7$. Fig. 8 shows the transformed probability distributions (mean zero and variance one) for

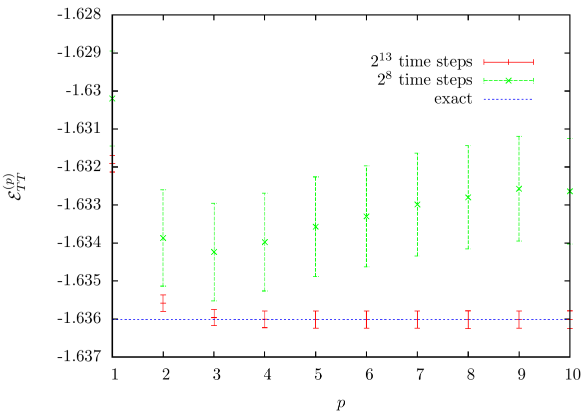

Figure 7: (Color online) Plot of the ground state energy estimate $\mathcal{E}_{TT}^{(p)}$ for different projections for 2^{13} and 2^8 time steps with an ensemble size of 2^8. The horizontal dashed line indicates the exact ground state energy.

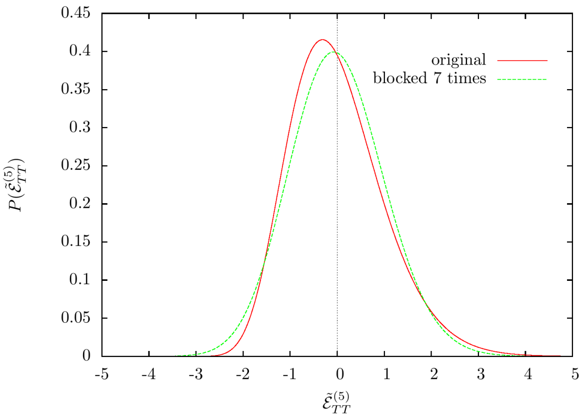

Figure 8: (Color online) The transformed probability distributions for $p=5$. The solid curve shows the original lognormal distribution and the dashed curve shows this distribution blocked 7 times.

the original lognormal distribution together with this distribution blocked seven times. Here, as expected, the blocked distribution approaches the normal distribution but does not exactly overlap the latter (cf. Fig. 1). We can now transform this blocked distribution into the true distribution described by the set of original data points obtained from the ensemble of $\mathcal{E}_{TT}^{(p)}$ for $p=7$. This distribution as obtained by using Eq. (5.1) is shown in Fig. 9. From this distribution one can easily obtain the errorbars corresponding to a confidence interval of 68%. This method was also applied to data corresponding to the other projections. Fig. 10 shows a comparison between the errorbars obtained from the standard method to the ones obtained from the constructed probability distributions. Note that, the errorbars obtained from the constructed probability distributions are asymmetric, which is due to the fact that the probability distributions are not Gaussian.

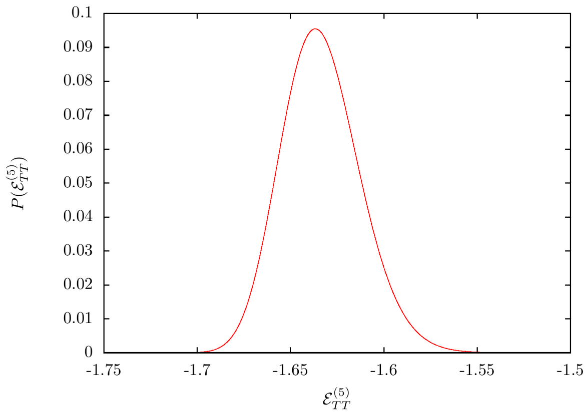

Figure 9: (Color online) Plot of the probability distribution corresponding to data obtained for $p=5$.

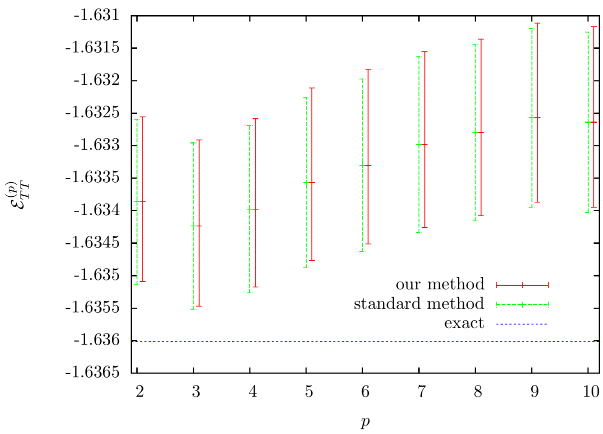

Figure 10: (Color online) Plot of the ground state energy estimate $\mathcal{E}_{TT}^{(p)}$ for different projections with 2^8 time steps and an ensemble size of 2^8. The left errorbars were obtained by using standard statistical methods while the right ones were obtained from the constructed probability distributions. The errorbars shown correspond to a 68% confidence interval.

7 Conclusions

In this paper, we demonstrated a method of obtaining the probability distribution together with estimates of the errorbars for a given set of data. The probability distribution was obtained by assuming that the given data was block transformed from the lognormal distribution. For simplicity, we only considered blocking transformations done in powers of 2. More accurate results could be obtained if a continuous spectrum of blocking coefficients were considered. In this case one could ideally obtain Kolmogorov-Smirnov statistics with significance levels of close to 100%.

The errorbars which were computed from the reconstructed probability distributions, to give a 68% confidence interval, indicated that the ones obtained by standard statistical methods incorrectly represented the relative uncertainties. Unlike the latter, the errorbars obtained using the above method were asymmetric, indicating that the probability distributions were not Gaussian. It was evident from this study that our model employing a 3×3 symmetric Hamiltonian matrix did not give data that were significantly similar to the lognormal distribution. That is, large contrasts in the errorbars obtained by the two different methods would be evident if our data corresponded to data having small blocking coefficients. Even though this is the case, the results have some credibility since many quantum Monte Carlo techniques are becoming more and more refined and the number of significant figures quoted for estimates are ever increasing. As a consequence, more accurate statistical methods need to be employed in order to account for the non-Gaussian nature of these estimates.

Acknowledgments

The author would like to acknowledge initial insightful discussions with Peter Nightingale and thank Darryl Fleming for verifying many of the numerical calculations. This work is based partially upon research supported by the University of KwaZulu-Natal Competitive Grant.

References

[1] H. A. Noughabi and N. R. Arghami, Monte Carlo comparison of seven normality tests, J. Stat. Comput. Sim., 81 (2011), 965–972.
[2] J. H. Hetherington, Observation on the statistical iteration of matrices, Phys. Rev. A, 30 (1984), 2713–2719.
[3] M. Romeo, V. Da Costa and F. Bardou, Broad distribution effects in sums of lognormal random variables, Euro. Phys. J. B, 32 (2003), 513–525.
[4] E. Limpert, W. A. Stahel, and M. Abbt, Log-normal distributions across the sciences: keys and clues, Bioscience, 51(5) (2001), 341–352.
[5] D. Dufresne, The log-normal approximation in financial and other computations, Adv. Appl. Prob., 36 (2004), 747–773.
[6] L. Fenton, The sum of lognormal probability distributions in scatter transmission systems, IRE Trans. Commun. Sys., 8 (1960), 57–67.
[7] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, Estimating the distribution of a sum of independent lognormal random variables, IEEE Trans. Commun., 43 (1995), 2869–2873.
[8] D. Dufresne, Sums of lognormals, in Proceedings of the 43rd Actuarial Research Conference, University of Regina, Regina, Canada, August 2008.
[9] D. Schleher, Generalized Gram-Charlier series with application to the sum of log-normal variates, IEEE Trans. Inform. Theory, 23 (1977), 275–280.
[10] S. C. Schwartz and Y. S. Yeh, On the distribution function and moments of power sums with log-normal components, Bell Systems Tech. J., 61(7) (1982), 1441–1462.
[11] Neelesh B. Mehta et al., Approximating a sum of random variables with a lognormal, IEEE Trans. Wirel. Commun., 6 (2007), 2690–2699.

[12] N. C. Beaulieu and Q. Xie, On optimal lognormal approximation to lognormal sum distributions, IEEE Trans. Veh. Technol., 53 (2004), 479–489.

[13] N. C. Beaulieu and F. Rajwani, Highly accurate simple closed-form approximations to lognormal sum distributions and densities, IEEE Commun. Lett., 8 (2004), 709–711.

[14] D. Senaratne and C. Tellambura, Numerical computation of the lognormal sum distribution, in Proceedings of the 28th IEEE conference on Global telecommunications (GLOBECOM'09), Mehmet Ulema (Ed.), IEEE Press, Piscataway, NJ, USA, (2009), 3966–3971.

[15] J. Honerkamp, Statistical Physics, Springer-Verlag, Berlin, 1998.

[16] G. Jona-Lasinio, The renormalization group: a probabilistic view, Nuovo Cimento B, 26 (1975), 99–119.

[17] G. Jona-Lasinio, Renormalization group and probability theory, Phys. Reports, 352 (2001), 439–458.

[18] W. H Press, B. P. Flannery, S. Teukolsky and W. T. Vetterling, Numerical Recipes, Cambridge University Press, Cambridge, 1992.

[19] M. P. Nightingale and C. J. Umrigar, Monte Carlo Eigenvalue Methods in Quantum Mechanics and Statistical Mechanics, Advances in Chemical Physics, Vol. 105, Monte Carlo Methods in Chemistry, edited by David M. Ferguson, J. Ilja Siepmann, and Donald G. Truhlar, series editors I. Prigogine and Stuart A. Rice, John Wiley & Sons, New York, 1999.

[20] M. P. Nightingale, Basics, Quantum Monte Carlo and Statistical Mechanics, in Quantum Monte Carlo Methods in Physics and Chemistry, edited by M. P. Nightingale and C. J. Umrigar, NATO Science Series, Series C; Mathematical and Physical Sciences-Vol. 525, Kluwer, 1999.