Neutron monitor generated data distributions in quantum variational Monte Carlo

A S Kussainov\textsuperscript{1,2} and N Pya\textsuperscript{3}

\textsuperscript{1} Physics and Technology Department, al-Farabi Kazakh National University, Almaty, Kazakhstan
\textsuperscript{2} National Nanotechnology Laboratory of Open Type, al-Farabi Kazakh National University, Almaty, Kazakhstan
\textsuperscript{3} Department of Mathematics, School of Science and Technology, Nazarbayev University, Astana, Kazakhstan

E-mail: arman.kussainov@gmail.com

Abstract. We have assessed the potential applications of the neutron monitor hardware as random number generator for normal and uniform distributions. The data tables from the acquisition channels with no extreme changes in the signal level were chosen as the retrospective model. The stochastic component was extracted by fitting the raw data with splines and then subtracting the fit. Scaling the extracted data to zero mean and variance of one is sufficient to obtain a stable standard normal random variate. Distributions under consideration pass all available normality tests. Inverse transform sampling is suggested to use as a source of the uniform random numbers. Variational Monte Carlo method for quantum harmonic oscillator was used to test the quality of our random numbers. If the data delivery rate is of importance and the conventional one minute resolution neutron count is insufficient, we could always settle for an efficient seed generator to feed into the faster algorithmic random number generator or create a buffer.

1. Introduction

Though hardly conceivable today, one may run into a situation when a random number could not be generated sequentially on a conventional computer. Specific statistical or computational tasks may lead to decision to use the random number look-up tables instead of pseudorandom software generators [1]. These tables have a long and successful history of applications. Tippett [2] is credited as the inventor of the random number table in 1927.

The sources for these tables are fairly different, and while ones came from the census data, the others are produced from the previously computed tables, like logarithmic, see Fisher and Yates [3]. Regardless of the order or location in the tables, these numbers should have the desired properties of the local or global randomness [4]. The mechanisms of sampling these numbers from the table are parallel topics of discussion.

One may also come across a situation when the iterative processes behind the regular pseudorandom software generator are not available at the very low level of data acquisition and processing, or inaccessible because of the rate, slow or fast, these data are sampled. In these cases we may want to take an advantage of the prerecorded data pattern, or equivalently, to create a buffer of the preregistered numbers [5], [6].
This approach is still intrinsically related to hardware random number generators or so called true random number generator [7]. These are devices that generate random numbers from the real physical processes. It is assumed though that the data are being registered in a stable state of a system, away from any transient phenomena and completely unpredictable. The last condition should be a subject of an experimental test.

In our study, we have selected Markov chain Monte Carlo (MCMC) algorithm [11] as a fine example of the quantum Monte Carlo method to test the random number distributions generated from an experiment. MCMC algorithm is used to approximate the ground state of a quantum system using variational method by calculating an integral related to the average value of energy of a system.

MCMC algorithm proceeds by randomly moving about the space spanned by the integral. It is restricted in the choice of the next step by the certain set of rules called importance sampling. In basic case, we need two individual random number distributions to move a single walker around, where the walker is an object following the single chain of successive positions selected in accordance with the previously mentioned rules. The first one is the normal distribution driving our walker toward the region of maximum probability defined by the trial set of wavefunctions and another one is a uniform random number distribution used to sample the regions with low-density probability along the path.

We chose to use the neutron counts data from the ground–based neutron detectors. These are designed to measure the secondary products of collisions of the high-energy charged particles from the outer space with the nuclei of atmospheric gases [8].

The registered data have a rich spectral composition modulated by the Earth rotation, solar cycles and multiple other events. It also has an intensive random component we intend to use.

2. Methods
We have considered one and three dimensional quantum harmonic oscillators as a quantum system to model by the MCMC method [9].

One dimensional, time independent, Schrödinger equation is given by the following expression

$$\frac{-\hbar}{2m} \frac{d^2 \varphi(x)}{dx^2} + \frac{1}{2} mk^2 x^2 \varphi(x) = E \varphi(x)$$

where \( k \) is the force constant, \( \hbar \) is the reduced Planck constant and \( m \) is the mass of the oscillating particle. If we know the complete set of \( N \) eigenfunctions from the following superposition notation

$$\Phi(x) = \sum_{i=0}^{N-1} c_i \varphi_i(x)$$

then, assuming the orthonormality of the eigenfunctions, the average value of energy \( \langle E \rangle \) of the oscillator could be calculated from the following expressions

$$\langle E \rangle = \frac{\int_{-\infty}^{+\infty} dx \Phi(x)^* H \Phi(x)}{\int_{-\infty}^{+\infty} dx \Phi(x)^* \Phi(x)} = E_0 + \frac{\sum_{i=0}^{N-1} |c_i|^2 (E_i - E_0)}{\sum_{i=0}^{N-1} |c_i|^2}$$

where \( H \) is the Hamiltonian given by the left part of equation (1) and \( E_0 \) is the exact value of the ground state energy.

Variational Monte Carlo method uses equation (3) as a starting point. Replacing \( \Phi(x) \), the unknown superposition of the eigenfunctions, by the trial wavefunction \( \Phi_{T,\alpha}(x) \), we are starting to vary the parameter \( \alpha \). If we were right in our choice of the trial wave function, the calculated averaged energy, as a function of \( \alpha \), will have a minimum for the ground state value and we could reconstruct the corresponding eigenfunction.
One of the distinct features of this method is that we are not sampling the whole configuration space from negative to positive infinity indiscriminately but rather traversing it in the manner prescribed by the Metropolis–Hastings algorithm, which is a subcase of MCMC. The targeted wave function’s tails, though stretching from negative to positive infinity, are vanishing to zero pretty fast in just a few steps away from the origin, and its overall shape is similar to the normal distribution. The Metropolis–Hastings algorithm samples the integral’s space, guided by the weight function $\omega(x)$, as given in the equation (4)

$$\langle E \rangle = \int_{-\infty}^{+\infty} dx \omega(x)E_L(x), \quad \text{where} \quad \omega(x) = \frac{|\Phi_{T,a}(x)|^2}{\int_{-\infty}^{+\infty} dx' |\Phi_{T,a}(x')|^2} \text{and} \quad E_L(x) = \frac{H\Phi_{T,a}(x)}{\Phi_{T,a}(x)} \quad (4)$$

where $\Phi_{T,a}(x)$ is the trial wavefunction of our choice.

That is we place the $n$ walkers randomly and uniformly across the selected region, not far away from the origin, at coordinates $\{x_n\}_t$, where index $t$ is used for the current set of coordinates. The next set of the walkers’ coordinates $\{x_n\}_{t+1}$ is determined by the following rules for each $n$

$$x_{t+1} = x_t + \delta, \quad \text{if} \quad \frac{\omega(x_{t+1})}{\omega(x_t)} \geq 1, \text{the step is accepted,} \quad (5)$$

$$\frac{\omega(x_{t+1})}{\omega(x_t)} < 1, \text{accepted only if it is larger than a random number from [0, 1]}$$

here $\delta$ is the step’s size which is determined based on the problems conditions and in general is subject to changes depending on the acceptance ratio. If this step is accepted the local energy $E_L(x)$ is calculated at this new walker’s coordinates and contributes to the integral in equation (4). All the walkers are sampling the integral in equation (4) independently and may cross their paths.

If, for one dimensional case, the trial function is chosen to be $\Phi_{T,a}(x) = \exp(-\alpha x^2)$, then the value of $\langle E \rangle$ could be calculated according to the formula

$$\langle E \rangle = \frac{1}{\text{counts}} \sum_{i=1}^{\text{counts}} E_L(x_i), \text{where} \quad E_L(x) = \alpha + x^2(0.5 - 2\alpha^2) \quad (6)$$

Here, the counts are the number of accepted steps across the all walkers’ trajectories. Expression for $E_L$ in the equation (6) was derived from the equation (4) and explicit notation for the trial wave function $\Phi_{T,a}(x)$. We also assume that $\hbar = m = k = 1$.

Three dimensional case is slightly different by the form of the Schrodinger equation we need to solve. Assuming the radial symmetry, it states that

$$-\frac{\hbar}{2m} \left[ \frac{d^2\varphi(r)}{dr^2} + \frac{2}{r} \frac{d\varphi(r)}{dr} \right] + \frac{e^2}{r} \varphi(r) = E\varphi(r) \quad (7)$$

and

$$E_L(r) = -0.5 \left[ \alpha^2 - 2\alpha/r \right] - 1/r \quad (8)$$

In fact, the computational part of the problem may be directly constructed from the one dimensional case. All we need is to represent a single three dimensional step as a sequence of three one dimensional steps in each of the three dimensions.
3. Results
For our purposes, we have selected the 38000 minutes long data piece. It was simultaneously registered by the 6 channels of unit A of the neutron monitor, at the Tien-Shan mountain high-elevation research station. This monitor is located 3340 meters above the sea level. The data were registered for about 633 hours, starting at 00.00 UTC, January 23rd, 2016 [13].

As we said before, we need two individual random numbers distributions to move a single walker around. The first one is the normal distribution and another one is a uniform random number distribution used to sample the regions with low-density probability, see equations (4)-(5).

A piecewise cubic spline fit with breaks (knots) every 250 points implemented by the `splinefit` function in Octave package on Debian, has been used to detect characteristic variations in the intensities of the cosmic rays and to subtract them as a baseline from the raw data thus producing the Gaussian white noise, see Fig. 1 (a)-(c).

![Figure 1](image)

**Figure 1.** Original data (a) fitted with the `splinefit` function, extracted white noise (b), histogram (c) of its spectral composition and cumulative distribution function (d) to produce a uniform random distribution.

The standardization procedure, see equation (9), further transforms the filtered noise component to a new one with zero mean and unit variance, which is our final normal distribution, see Fig. 1 (b)-(c)

\[ x' = \frac{(x - \bar{x})}{s}, \]  

where \( s \) and \( \bar{x} \) are the standard deviation and mean value of the extracted noise before standardization procedure.

It is known from the probability integral transform that data values that are generated from any given continuous distribution, normal in our case, can be transformed to random deviates with a uniform distribution on the interval \([0,1]\) as \( U = F(x) \), where \( F(x) \) is the standard normal cumulative distribution function. We used this to produce a uniform random distribution plotted on Fig. 1 (d).

As a reference simulations we have used a pseudo–random generator provided by the g++ compiler on Debian Linux OS. In this reference case, Box-Muller transform [12] was used to go from a uniform random number distribution to a normal random distribution. In order to compare these data with our experimentally generated random numbers distributions we need the number of walkers to be the same in both cases.

If the walkers are placed not far away from the origin, we need no more than 1000 steps, including thermalization stage, to reach and amply sample the area of maximum probability.

All channels are highly correlated in general. The sources of extracted noise in each individual channel may be different from just a local background radiation and other independent processes in the registration hardware and could correlate as well. Fitting the data with splines may be not good enough to get rid of the possible statistical imprints of the common registered events.
Nevertheless, we obtained a nearly perfect match, see Fig. 2, when we pulled the new values of the random normal and random uniform deviates from all available channels, in a successive order, one by one from each channel. We have got this match from as many as 17 walkers.

As one can see, the quality of experimental data, after few simple transformations, is reasonably good to bring the walker even from the remote location to the maximum probability point, see Fig. 2 (a). The data were good enough not to introduce a special procedure to treat the outliers in the raw data.

![Figure 2](image)

**Figure 2.** The 3D trajectories (a) of individual walkers driven to the maximum probability region by a regular pseudorandom numbers (red), and by a random numbers from our experimental data (blue). Computed average energy values (b) and computed values of the variance (c), where blue open circles stand for a regular pseudorandom numbers and red pentagrams stand for an experimental data.

The next two plots to the right from trajectories’ plot, represent the average energy and its variance versus variation parameter $\alpha = 0$. As expected, we can clearly see the minimum at $\alpha = 0.5$ because the $E_L$ loses its $\alpha$ dependence at this point, see equations (6) and (8). This value of $\alpha$ corresponds to the zero variance value on the right plot which pinpoints the ground state in our problem. The slight differences observed on the picture are related to the quality of the random numbers used in our simulations.

Supporting our MCMC simulations, the one-sample Kolmogorov-Smirnov test $kstest(x)$ returns a positive test decision for the null hypothesis that our data indeed come from a standard normal distribution. The formal application of the full range of normality tests, as well as the tests of global and local randomness, has been left for a further study.

4. Conclusions
We have successfully used the neutron monitor data as a source for the normal and uniform random number generators to drive a MCMC method. The desired stochastic component was extracted by fitting the data with splines and subtracting this fit from the raw data. Further scaling of our data to zero mean and variance of one was sufficient to obtain a stable standard normal random variate. Cumulative distribution function was used as a source of the uniform random numbers.

We are able to obtained the exact values of the variational parameter $\alpha$ and local energy value $E_L$ required to pinpoint the ground state of the three dimensional quantum harmonic oscillator. The negligible difference in the calculated values of the local energy for different $\alpha$ values as compared to a reference case may be attributed to the multiple factors related not only to the quality of the noise extraction but to a thorough analysis of this noise sources in different channels. For some problems, as ours, this quality is good enough or may be improved by additional treatment of the raw data. Distributions under consideration pass additional goodness of fit test for normality.
Acknowledgments
This study was supported by grant #3824/GF4 provided by the Science Committee at the Ministry of Science and Education of Republic of Kazakhstan to the principal investigator at the National Nanotechnology Open Laboratory. We also would like to thank Prof. Yelena White, from East Georgia State College, Swainsboro, GA, USA, for editing the manuscript.

References
[1] Mitchell R L and Stone C R 1977 IEEE Transactions on Computers C-26 1006-08
[2] Tippett L H C 1927 Random sampling numbers. no. 15 Tracts for Computers (Cambridge: Cambridge Univ. Press) reprinted in 1952
[3] Fisher R A, Yates F 1948 Statistical Tables for Biological, Agricultural and Medical Research 3rd ed. (London: Oliver & Boyd)
[4] Kendall M G, Smith B B 1938 Journal of the Royal Statistical Society 101 147-66
[5] Kravtsov K S, Radchenko I V, Kulik S P, Molotkov S N 2015 Journal of the Optical Society of America B 32 1743
[6] Self J, Macke M C 2015 ArXiv e-prints arXiv:1512.07303
[7] Dynes J F, Yuan Z L, Sharpe A W and Shields A J 2008 Appl. Phys. Lett. 93 031109
[8] Simpson J A 2000 Space Science Reviews 93 11-32
[9] Foulkes W M C 2001 Rev. Mod. Phys. 73 33
[10] Dodge Y, Cox D, Commenges D 2003 The Oxford Dictionary of Statistical Terms (Oxford: The Oxford University Press)
[11] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, Teller E 1953 Journal of Chemical Physics 21 1087-92
[12] Box G E P, Muller M E A 1958 The Annals of Mathematical Statistics 29 610-11
[13] Shepetov Yu V 2014 The Tien-Shan Mountain Station’s Database (Moscow: Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation named after Nikolay Pushkov of the Russian Academy of Sciences) http://cr29.izmiran.ru/vardbaccess/frames-vari.html