PROCEEDINGS OF SCIENCE

A method to estimate observables with infinite variance in fermionic systems

Cagin Yunus[a,*] **and William Detmold**[a,b]

[a] *Massachusetts Institute of Technology,*
77 Massachusetts Avenue, Cambridge, MA, USA 02139
[b] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*

E-mail: cyunus@mit.edu, wdetmold@mit.edu

Numerical estimation of fermionic observables often requires introduction of auxiliary bosonic fields that have no direct physical relevance through a Hubbard-Stratonovich transformation. The variance of some fermionic observables (for example, appropriately constructed local four-fermion operators in 2d Gross-Neveu model) when they are estimated using such auxiliary fields may not correspond to any physical observable and in particular may diverge. In such cases it is not clear how one can reliably estimate the corresponding observable. An important example may be found in the context of exceptional configurations in quenched QCD. We demonstrate simpler examples in toy models by constructing particular four-fermion operators. We propose two methods to overcome this issue. The first method is a discrete Hubbard-Stratonovich transformation that does not suffer from such divergences for these classes of observables. While this method works in principle, it is not practically feasible. The second method we propose is a reweighting method.

MIT-CTP/5385

PoS(LATTICE2021)145

**** The 38th International Symposium on Lattice Field Theory, ****
**** 26-30 Jul, 2021 ****
**** MIT, Cambridge, MA, USA ****

[*]Speaker

1. Introduction

In a Monte Carlo simulation of a Lattice QFT model, depending on the parameters of the Lagrangian, one may sample field configurations with probability weights arbitrarily close to 0. Then, there will be random variables that have an infinite variance. In particular, such large variances are expected for correlation functions constructed from large number of fermionic propagators. A similar problem in the context of the condensed matter physics was investigated in [5].

In order to construct confidence intervals for the mean of the random variable from the sample standard deviation one typically employs the Central Limit Theorem. However, for a random variable with infinite variance, the Central Limit Theorem does not apply[1].

We will propose two methods to overcome this issue. The first method we propose is applicable to fermionic theories where Hubbard-Stratonovich transformation is used to construct auxiliary bosonic variables. We introduce a discrete version of the Hubbard-Stratonovich transformation which generates discrete auxiliary bosonic variables. In this approach, the variance will be finite although can be very large for specific model parameter choices. We will see that while there are instances where this method is useful, it is not practical for large lattice sizes. The second method we propose is a reweighting method which is applicable to semi positive random variables. In this method, the mean of a random variable is expressed as the product of the means of the several random variables each having finite variance. We will test the first method for a toy model and the second method for the 2D Gross-Neveu model.

1.1 An example in Euclidean Field Theory

The partition function of a Euclidean theory involving bosonic fields U and biliniear in fermionic fields Ψ can be written as:

$$\begin{aligned} Z &= \int \mathcal{D}[\Psi\bar{\Psi}]\mathcal{D}[U]e^{-S[U]-\bar{\Psi}D[U]\Psi} \\ &= \int \mathcal{D}[U]e^{-S[U]} \det D[U] \\ &= \int \mathcal{D}[U]e^{-S[U]} \prod_{a=1}^{N_D} \lambda_a[U] \end{aligned} \tag{1}$$

in the third line the fermions are integrated exactly and where N_D is the number of eigenvalues λ_a of the Dirac operator $D[U]$ which is finite in a discretized theory in a finite volume. We will assume that $D[U]$ is diagonalizable for all U. Then, $D[U]$ can be written as $D[U] = Q[U]\Lambda[U]Q^{-1}[U]$. If $V_k = \prod_{n=1}^{k} \bar{\Psi}_{i_n}\Psi_{j_n}$ is any product of k fermion bilinears, then after integrating over the fermionic fields its expectation value can be written as a sum of products of inverses of k eigenvalues of the

[1] Sample variance for a random variable with infinite variance is also not meaningful, in the sense that it doesn't converge to a particular value as the sample size is increased. Furthermore, it must be stressed that the Central Limit Theorem is valid only asymptotically, that is, as the sample size goes to infinity. Therefore, similar issues will arise when the variance of a random variable is finite but very large compared to its mean squared.

Dirac operator. Precisely[2],

$$\langle V_k \cdots \rangle = \frac{1}{Z} \int \mathcal{D}[U] e^{-S[U]} \prod_{b=1}^{N_D} \lambda_b[U] \sum_{\pi \in S_n} \text{sgn}\,(\pi) \prod_{n=1}^{k} \sum_{a=1}^{N_D} \frac{1}{\lambda_a[U]} v_{i_n}^{(a)}[U] w_{j_{\pi(n)}}^{(a)}[U] (\cdots) \tag{2}$$

where $\cdots$ is any operator insertion that does not coincide with any of the fermions in V_k. One can project products of inverses of eigenvalues λ_a by considering (U dependent) linear combinations of V_k as follows. Let us define a new operator $O_{a_1,\cdots,a_k}$ by $O_{a_1,\cdots,a_k} = \prod_{n=1}^{k} \sum_{i_n,j_n} w_{i_n}^{(a_n)}[U] v_{j_n}^{(a_n)}[U] \bar{\Psi}_{i_n} \Psi_{j_n}$ where it is assumed that $a_m \neq a_n$ for $m \neq n$. A natural estimator for $O_{a_1,\cdots,a_k}$ is given by:

$$\hat{O}_{a_1,\cdots,a_n} = \frac{1}{N_S} \sum_{i=1}^{N_S} \frac{1}{\prod_{n=1}^{k} \lambda_{a_n}[U_i]} \tag{3}$$

where N_S is the sample size. Difficulties arise when one of the eigenvalues, say $\lambda_{a_1}[U]$, vanishes at $U = U^*$. In such a case, ($\hat{\theta}$ and hence the observable V_k associated with it) becomes infinite. We will call such confiigurations **exceptional configurations**. We note that the probability of sampling U^* is 0 since the probability distribution at U is proportional to $\det D[U] = \prod_{a=1}^{N_D} \lambda_a[U]$ and thus observables with a single inverse power of λ_a, are finite even in this case. However, the variance of such a quantity will contain λ_a^{-2} and so will be strictly infinite. It is easy to see that $\hat{O}_{a_1}$ has infinite variance. The above explains how infinite variance can arise LQFT calculations of complicated theories such as QCD. In a quenced version of the theory the determinant is absent and even the observable will have an infinite contribution. In partially-quenched or mixed-action QCD, where the fermion actionis different in the measure and in defining observables, the observable $O_{a_1,\cdots,a_k}$ is similarly ill-defined.

2. Simple examples with infinite variance

To investigate infinite variance in Monte Carlo sampling, two simple models are introduced and the particulars of infinite variance correlation functions in each model are presented.

2.1 A Toy Model

We will consider a toy model that exhibits most of the issues we are concerned with. It is a zero dimensional (Euclidean) QFT model of interacting fermions[3]. The Lagrangian is given as $\mathcal{L} = m\bar{\Psi}\Psi - \frac{g}{2}\left(\bar{\Psi}\Psi\right)^2$. The standard way of getting rid of the quadratic term is to introduce an auxiliary field through the Hubbard-Stratonovich transformation. After fermions are integrated, one obtains $Z[\eta] = \int d\phi\, e^{-\frac{1}{2}\phi^2 + \bar{\eta}\frac{1}{m+\sqrt{g}\phi}\eta} (m + \sqrt{g}\phi)^{2N_f}$. In a Monte Carlo simulation, one will use

[2]The columns $v^{(a)}[U]$ of $Q[U]$ are right eigenvectors of $D[U]$ and the rows $(w^{(a)})^T[U]$ of $Q^{-1}[U]$ are left eigenvectors of $D[U]$. They satisfy $(w^{(a)})^T[U] v^{(b)}[U] = \delta^{ab}$.

[3]The fermions will be denoted by $\Psi = \begin{pmatrix} \Psi_1 \\ \cdots \\ \Psi_{N_f} \end{pmatrix}$ and $\bar{\Psi} = \begin{pmatrix} \bar{\Psi}_1 \cdots \bar{\Psi}_{N_f} \end{pmatrix}$ where $\Psi_i = \begin{pmatrix} \Psi_i^{\uparrow} \\ \Psi_i^{\downarrow} \end{pmatrix}$, $\bar{\Psi}_i = \begin{pmatrix} \bar{\Psi}_i^{\uparrow} \\ \bar{\Psi}_i^{\downarrow} \end{pmatrix}$ are two component Grassmannian variables.

$P(\phi) \propto e^{-\frac{1}{2}\phi^2}\left(m+\sqrt{g}\phi\right)^{2N_f}$ as the probability weight. Suppose we are interested in the observable $O = \prod_{i=1}^{2N_f}\prod_{s=\uparrow,\downarrow}\bar{\Psi}_i^s\Psi_i^s$. In terms of the auxiliary field, expectation value of the O is given as:

$$\langle O\rangle = \frac{\int d\phi\, e^{-\frac{1}{2}\phi^2}}{\int d\phi e^{-\frac{1}{2}\phi^2}\left(m+\sqrt{g}\phi\right)^{2N_f}} \tag{4}$$

It is clear from the above expression that $\langle O\rangle$ is finite. However, one will run into trouble while calculating this quantity in a Monte Carlo simulation. This happens because one estimates the expectation value of the observable by the following estimator in a Monte Carlo simulation $\langle O\rangle_{MC} = \frac{1}{N_s}\sum_{n=1}^{N_s} O(\phi_n)$ where N_s is the sample size and $O(\phi) = \left(m+\sqrt{g}\phi\right)^{-N_f}$ is the representation of the observable in terms of the auxiliary field. We see that, this quantity has a singularity at $\phi^* = -\frac{m}{\sqrt{g}}$. In fact, the variance of this estimator is divergent as the second moment diverges:

$$\left\langle O^2(\phi)\right\rangle = \frac{\int d\phi\, e^{-\frac{1}{2}\phi^2}\left(m+\sqrt{g}\phi\right)^{-2N_f}}{\int d\phi e^{-\frac{1}{2}\phi^2}\left(m+\sqrt{g}\phi\right)^{2N_f}} = \infty \tag{5}$$

2.2 Gross-Neveu Model

The Gross-Neveu (GN) model [2] is a simple model of fermions in $d = 2$ dimensions (where it is renormalizable) interacting via four-fermion couplings that shares some important properties with QCD. Notably it is asymptotically free and demonstrates chiral[4] symmetry breaking. These properties make the Gross-Neveu model useful as a simple testing ground for new ideas that may eventually be applied to QCD. The Lagrangian of the Gross-Neveu model in the continuum is given by: $\mathcal{L} = \bar{\Psi}\left(\partial\!\!\!/ + m\right)\Psi - \frac{g}{2}\left(\bar{\Psi}\Psi\right)^2$. In this work, we will consider the Gross-Neveu model on the lattice with Wilson fermions and $N_f = 2$ species.

3. Discrete Hubbard-Stratonovich Transformation

The continuous Hubbard-Stratonovich transformation is valid for all commuting variables Φ and is given by:

$$e^{\frac{1}{2}\Phi^2} = \frac{1}{\sqrt{2\pi}}\int du\, e^{-\frac{1}{2}u^2+u\Phi} \tag{6}$$

However, in problems where Φ is constructed out of fermions, the above equation is only required to satisfied up to an even power of Φ which we will denote by[5] $2N_f$ since $\Phi^{2N_f+1} = 0$ because of the anticommuting nature of the fermion fields. This observation brings the possibility of solving this equation by introducing an auxiliary variable that takes values in a discrete finite set. We aim to solve the following equation (here w_a should be positive to have a probabilistic representation and t_a should be real to avoid a sign problem):

$$e^{\frac{1}{2}\Phi^2} = \sum_a w_a e^{t_a\Phi} \tag{7}$$

[4]For the version of the model we are discussing, it is a discrete version of the chiral symmetry. But it is easy to modify the action to obtain a theory with the continuous chiral symmetry.

[5]In theories that have spinor dimension 2, N_f is the number of fermions.

where Φ is assumed to satisfy $\Phi^{2N_f+1} = 0$. Now we consider Φ as a real variable and interpret the above equation (we make the transformation $\Phi \to i\Phi$) as equality of the two real power series in Φ up to the $2N_f$th order in Φ:

$$e^{-\frac{1}{2}\Phi^2} = \sum_a w_a e^{it_a\Phi} + O\left(\Phi^{2N_f+1}\right) \qquad \text{where } \Phi \in \mathbb{R} \tag{8}$$

We may also think of the series on the left and the right side as the characteristic functions[6] of the two probability densities in the conjugate variable t. These densities are given as $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$ and $\sum_a w_a\delta(t-t_a)$, respectively. Then, the condition (7) is equivalent to the following: if $f(t)$ is a polynomial of degree at most $2N_f$, then it needs to satisfy the following equation:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dt\, e^{-\frac{1}{2}t^2} f(t) = \sum_a w_a f(t_a) \tag{9}$$

This equation can be solved using the method of Gaussian quadrature. Explicitly, if $He_n(t)$ are the (probabilist's) Hermite polynomial[7] then t_a are the roots of $He_{N_f+1}(t)$ and w_a are given by $w_a = \frac{(N_f!)^2}{He'_{N_f+1}(t_a)He_{N_f}(t_a)}$.

A less general discrete version of the Hubbard-Stratonovich transformation was first proposed in [3]. This transformation has proven to be useful for Quantum Monte Carlo simulations, see for example [1] and [6].

3.1 Discrete Sampling vs. Continuous Sampling for the Toy Model

In the context of the toy model, we compare the discrete Hubbard-Stratonovich transformations to the continuous Hubbard-Stratonovich transformation and discrete transformations to each other by inspecting the mean and the logarithm of the standard deviation of the observable $O = \prod_{i=1}^{2N_f}\prod_{s=\uparrow,\downarrow}\bar{\Psi}_i^s\Psi_i^s$ as a function of sample size. In Figure 1a, we show the logarithm of the standard deviation of O against the logarithm of the standard deviation of the sample size. In Figure 1b, we show the mean of O against the sample size for the parameters $m = \sqrt{3}$ and $g = 1.0$. One expects the slope of $log(std(O))$ to converge to -0.5. While this is observed to be correct for discrete sampling, one observes that jumps continue over the sample sizes that could be investigated for the continuous sampling. We further note that it is possible to choose m and g such that the exceptional configuration is one of the roots of the $He_{N_f+1}(t)$. In this case, the discrete sampling will obey the correct scaling law however the mean will be biased.

We have chosen the parameters such that the exceptional configuration ($\phi^* = -\sqrt{3}$) is extremely close to one of the roots of He_3. There are several observations to make:

- In Figure 1a the plot of He_8 exhibits jumps because it has a root that is very close to the exceptional configuration $\phi^* = -\sqrt{3}$. The probability of sampling this root is $p \simeq 3\times 10^{-7}$ and one expects to sample this root about 30 times within a sample size of $N_S = 10^8$. Indeed we observe that the first jump appears around $N_S \sim \frac{1}{p}$ with later jumps that are less noticeable.

[6] Characteristic function of a random variable X is defined as $\phi(t) = \left\langle e^{itX}\right\rangle$.
[7] $He_n(t) = (-1)^n e^{\frac{1}{2}t^2}\frac{d^n}{dt^n}e^{-\frac{1}{2}t^2}$.

- In Figure 1a the plot of He_3 doesn't exhibit any jumps although it has a root that is extremely close the exceptional configuration. This is because the sample size $N_S = 10^8$ is not large enough to hit the root that is close to the exceptional configuration which has probability $p \simeq 10^{-13}$. However, since the net contribution of this root is large, we observe a bias for the mean of O. This bias should start to disappear for $N_S \gtrsim 10^{13}$ and one would start to see jumps in the plot of standard deviation. For $N_S \gg 10^{13}$ both bias and jumps will disappear.

- We further investigate the last point by changing the mass in Figure 1d. We don't see any jumps for $m = 1.03$ because the roots are not close to the exceptional configuration. We observe that as we approach $m = 1.73$ jumps starts to appear before disappearing again.

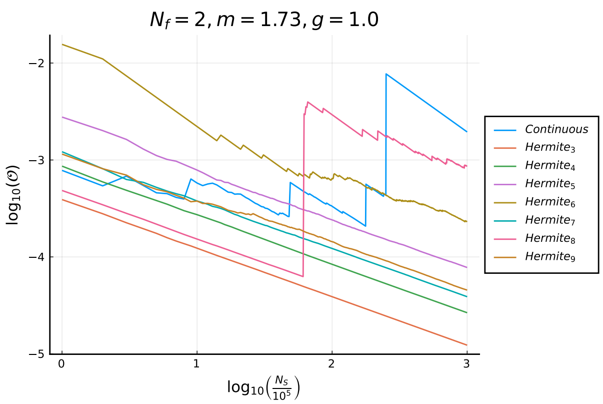


Figure 1(a): Logarithm of the standard deviation of O vs. logarithm of the sample size for $m = 1.73$, $g = 1.0$ and $N_f = 2$ with various sampling schemes for the toy model.

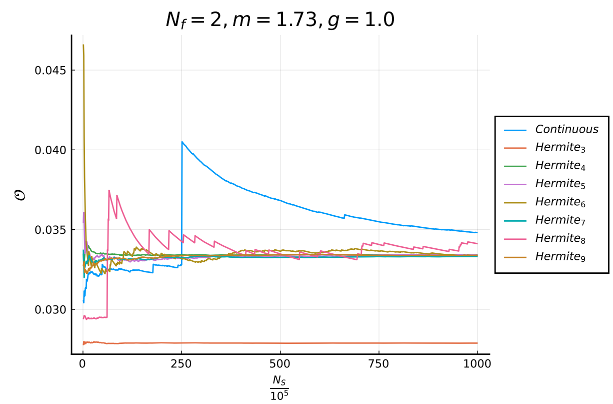


Figure 1(b): Ratio of the sample mean of O to mean of O vs. sample size for $m = 1.73$, $g = 1.0$ and $N_f = 2$ with various sampling schemes for the toy model.

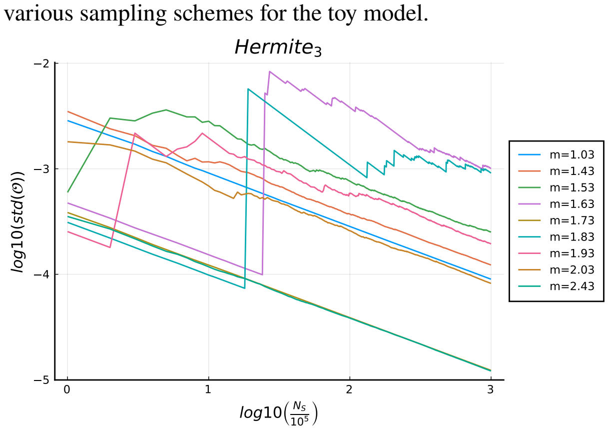


Figure 1(c): Logarithm of the standard deviation of O vs. logarithm of the sample size for various m, $g = 1.0$ and $N_f = 2$. He_3 is chosen for the discrete sampling scheme. Total sample size is $N_S = 10^8$.

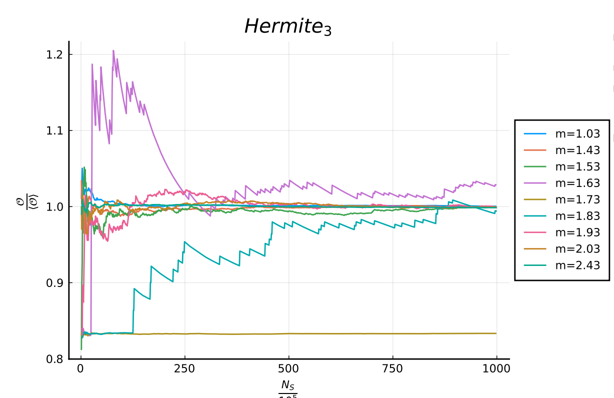


Figure 1(d): Ratio of the sample mean of O to mean of O vs. sample size for various m, $g = 1.0$ and $N_f = 2$. He_3 is chosen for the discrete sampling scheme. Total sample size is $N_S = 10^8$.

3.2 Summary

The discrete sampling schemes we have proposed has finite variance by construction. Furthermore, it is useful for calculating exact quantities for small lattices and elucidating the underlying issues of infinite variance. However, as this variance can be very large this scheme may not be practically feasible. In fact ongoing work shows that the discrete sampling scheme does not solve issues we have explored for the $2d$ Gross-Neveu model with a lattice size as small as $L = 8 \times 8$.

4. Reweighting

In this section, we suggest a more general method to sample observables with infinite variance by dismantling the difference between the probability measure and the observable gradually.

Assume that we have a (unnormalized) probability weight $P(x)$ and a positive observable we want to estimate $O(x)$. One wants to calculate $\langle O \rangle = \frac{\sum_x P(x)O(x)}{\sum_x P(x)}$. A standard estimator for this quantity is given by $\hat{O} = \frac{1}{N_s}\sum_{i=1}^{N_s} O(x_i)$ where x_i has sampled with respect to the probability weight $P(x)$. If the second moment of O with respect to $P(x)$ is infinite then the expected variance of this estimator is not defined. To overcome this issue, we will introduce new (unnormalized) probability weights $P_\mu(x) \equiv P(x)O(x)^\mu$. Note that $P_0(x) = P(x)$. We will denote the expectation value of an observable with respect to $P_\mu(x)$ by $\langle \cdot \rangle_\mu$. Now it is easy to verify that:

$$\langle O \rangle = \prod_{r=0}^{N-1} \left\langle O(x)^{1/N} \right\rangle_{\frac{r}{N}}, \tag{10}$$

where N is a positive integer. One expects that for a suitable N, the estimators $\frac{1}{N_r}\sum_{i=1}^{N_r} O(x_i)^{\frac{1}{N}}$ where x_i are sampled with respect to $P_{\frac{r}{N}}(x)$ will have finite variance for $r = 0, \cdots, N-1$. If this is the case, then a new estimator $\tilde{O}$ of $\langle O \rangle$ is given by:

$$\tilde{O} = \prod_{r=0}^{N-1} \left(\frac{1}{N_r} \sum_{i_r=1}^{N_r} O\left(x_{i_r}\right) \right) \tag{11}$$

where x_{i_r} are sampled with respect to $P_{\frac{r}{N}}(x)$.

Figure 2 represents the results of the estimator $\tilde{O}$ for the Gross-Neveu model for $N = 10$ and $L = 2 \times 2$ with varying sample sizes where $O = \prod_{i=1}^{2} \prod_{\sigma=\uparrow,\downarrow} \bar{\Psi}_i^\sigma(1,1)\Psi_i^\sigma(1,1)$. We also compare different step numbers for a total sample size (TSS) of 10^7. We use the median of the means estimator to determine the confidence intervals, see [4] for further details.

PoS(LATTICE2021)145

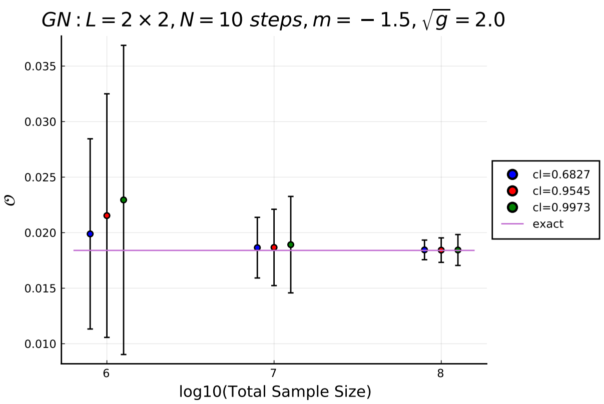
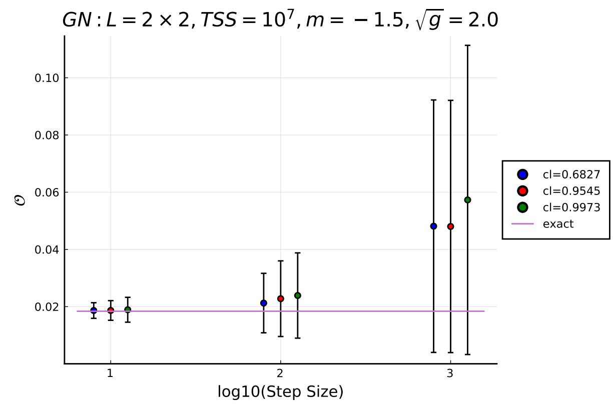


Figure 2(a): Estimators for the mean of O obtained with median of the means estimator for various sample sizes and for $N = 10$ steps for the Gross-Neveu model with $L = 2 \times 2$, $m = -1.5$, $\sqrt{g} = 2.0$. The purple line shows the exact value obtained with discrete sampling scheme. The abbreviation "cl" refers to "confidence level".



Estimators for the mean of O obtained with median of the means estimator for various step numbers and for total sample size $TSS = 10^7$ for the Gross-Neveu model with $L = 2\times2$, $m = -1.5$, $\sqrt{g} = 2.0$. The purple line shows the exact value obtained with the discrete sampling scheme. The abbreviation "cl" refers to "confidence level".

5. Conclusions

We have developed a discrete sampling scheme to overcome the issue of infinite variance. It is observed that while the resulting estimators have finite variances, the variances can be very large and therefore these estimators may be impractical. We further developed another sampling scheme which may be applied in any context where it is possible to generate configurations in a Monte Carlo setting and the observable one is interested in is non-negative. We stress that while the method is constructed to estimate observables with infinite variances, it is also expected to be useful in situations with finite but very large noise to signal ratios.

Acknowledgments

WD and CY are supported in part by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under grant Contract Number DE-SC0011090. WD is also supported by the SciDAC4 award DE-SC0018121, and by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, http://iaifi.org/).

References

[1] A. Bulgac, J. E. Drut, and P. Magierski. Quantum monte carlo simulations of the bcs-bec crossover at finite temperature. *Physical Review A*, 78(2), Aug 2008. ISSN 1094-1622. doi: 10.1103/physreva.78.023625. URL http://dx.doi.org/10.1103/PhysRevA.78.023625.

[2] D. J. Gross and A. Neveu. Dynamical symmetry breaking in asymptotically free field theories. *Physical Review D*, 10(10):3235–3253, Nov. 1974. doi: 10.1103/PhysRevD.10.3235.

[3] J. E. Hirsch. Discrete hubbard-stratonovich transformation for fermion lattice models. *Phys. Rev. B*, 28:4059–4061, Oct 1983. doi: 10.1103/PhysRevB.28.4059. URL https://link.aps.org/doi/10.1103/PhysRevB.28.4059.

[4] M. Lerasle. Lecture notes: Selected topics on robust statistical learning theory, 2019.

[5] H. Shi and S. Zhang. Infinite variance in fermion quantum monte carlo calculations. *Physical Review E*, 93(3), Mar 2016. ISSN 2470-0053. doi: 10.1103/physreve.93.033303. URL http://dx.doi.org/10.1103/PhysRevE.93.033303.

[6] C. Wu and S.-C. Zhang. A Sufficient condition for the absence of the sign problem in the fermionic quantum Monte-Carlo algorithm. *Phys. Rev. B*, 71:155115, 2005. doi: 10.1103/PhysRevB.71.155115.