PROCEEDINGS
OF SCIENCE

Listening to Noise

Michael G. Endres
Theoretical Research Division, RIKEN Nishina Center, Wako, Saitama 351-0198, Japan
E-mail: endres@riken.jp

David B. Kaplan[*]
Institute for Nuclear Theory, University of Washington, Seattle, WA 98195-1550, USA
E-mail: dbkaplan@uw.edu

Jong-Wan Lee
Institute of Particle and Nuclear Studies (IPNS), KEK, 1-1 Oho, Tsukuba, Ibaraki, Japan
E-mail: jongwan@post.kek.jp

Amy N. Nicholson
Deptartment of Physics, University of Maryland, College Park MD 20742-4111, USA
E-mail: amynn@umd.edu

We show how sign problems in simulations of many-body systems can manifest themselves in the form of heavy-tailed correlator distributions, similar to what is seen in electron propagation through disordered media. We propose an alternative statistical approach for extracting ground state energies in such systems, illustrating the method with a toy model and with lattice data for unitary fermions.

---

[*]Speaker.

1. Introduction

While QCD is accepted as the correct theory of the strong interactions, it is frustrating that we still are unable to derive from it the basic properties of matter or map out a phase diagram that is more than a cartoon. The problem is that the only tool we have for analyzing QCD nonperturbatively outside of special kinematic regions is the lattice, and lattice QCD faces daunting barriers to investigating systems with significant baryon number. In the grand canonical formulation, the problem encountered is that the fermion determinant has a rapidly varying phase leading to strong cancelations in the path integral and a suppression of the partition function that is exponential in the volume. This is often called the "fermion sign problem", although we will argue it is not specific to fermions. In a canonical formulation, where one computes correlation functions of fixed baryon number, one encounters a couple of problems. The first is that the measurements may become increasingly noisy as the baryon number is increased, which we will call a "noise problem". The second is that the value obtained for the correlator may be relatively quiet, yet drift with time, giving no evidence for a plateau in the effective mass from which one can infer a ground state energy. The effective mass computed from a correlator in a system with discrete nonzero energy levels must exhibit a plateau at late time; so if a plateau is absent it follows that the path integral is not being estimated correctly, with little overlap between the field configurations being sampled and the ones that actually dominate in a correct estimation of the correlator. We will refer to this second problem as an "overlap problem". In an actual simulation both problems may be found simultaneously: a drifting and noisy value for the correlator.

After arguing that the problems encountered in the canonical and grand canonical approaches arise from the same physics, we focus on the canonical method and the statistical distribution of correlators. We give examples where the overlap problem is a manifestation of an underlying statistical distribution for the correlator which is very heavy-tailed, and not too far off a log-normal distribution. We show that the appearance of nearly log-normal distributions is indicative of a mean field theory expansion, such as that used to study electron properties in disordered media. We then develop a technique to effectively overcome the overlap problem by defining mass plots using an estimator which is not minimum bias, but where the log-normal result is the first term in a well-defined expansion. This method has been used with great success in studies of trapped "unitary fermions" — a strongly coupled nonrelativistic conformal theory relevant to trapped atoms at a Feshbach resonance — allowing a measurement of the ground state energy of up to 70 fermions in a harmonic trap. We also show QCD data from the NPLQCD collaboration that indicates similar techniques might aid in the study of QCD; we expect this to be especially so at larger baryon number. This work has been published in Ref. [1].

As a possible direction for future research, our approach suggests a renormalization group approach to the statistical distributions of lattice observables, and a way to analyze such distributions in a manner analogous to effective field theory, where the nth cumulants of the log of the observable play the role of higher dimension operators in the effective Lagrangian.

2. Noise and the physical spectrum

We first examine the connections between the noise problem encountered in canonical sim-
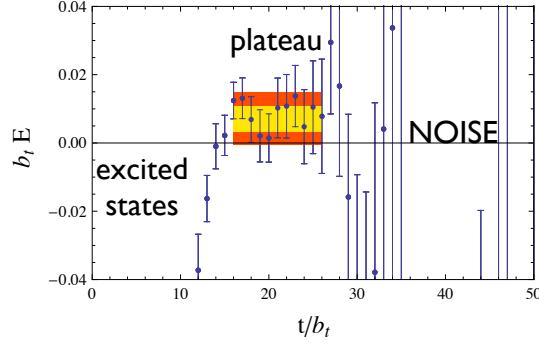
Figure 1: Effective mass plot for nine quarks in the triton channel from NPLQCD [7]; annotation by D.K.

ulations and the phase problem due to the fermion determinant in grand canonical computations, developing a heuristic picture for why a sign/noise/overlap problem arises in the first place.

Grand canonical simulations of lattice QCD at nonzero quark chemical potential involve a path integral measure that includes a fermion determinant which is neither real nor positive, leading to an exponentially hard computation [2]. It was noticed long ago [3] that the determinant phase problem sets in for quark chemical potential $\mu \geq m_\pi/2$, even at zero temperature, which seems peculiar since nothing physical happens until $\mu > m_N/3$, m_N being the nucleon mass. A beautiful explanation for this precocious annoyance was given by Splittorff and Verbaarschott [5] (see also [6] for earlier related ideas): they pointed out that in two flavor QCD with degenerate quarks, the quark determinant phase is exactly what sets apart a simulation at finite isospin chemical potential ($\mu_u = -\mu_d = \mu$) from a simulation at finite quark number ($\mu_u = \mu_d = \mu$). In the former system, pion condensation occurs at $\mu = m_\pi/2$ leading to a drop in the free energy density, and therefore growth of the partition function Z which is exponential in the volume. Therefore a simulation at finite baryon number with $\mu \geq m_\pi/2$ must have the quark determinant phase oscillate wildly to cancel this exponential growth, since there is no pion condensate in this system. The sign problem in QCD is therefore characterized by $\Delta \mu = \zeta \equiv (M_N/3 - m_\pi/2)$, the difference between the values for μ where the ground states rearrange themselves in the isospin and baryon number systems.

In a canonical simulation of the A-nucleon state one looks instead at the $3A$ quark correlator over Euclidian time τ, $C_A(\tau, U)$, which is a function of the link fields U for the gluons. For short times, the correlator gets contributions from the many excited states with $3A$ quarks, but at later time it should be dominated by the ground state for A nucleons. One typically creates an effective mass plot of

$$m_{\text{eff}}(\tau) = -\frac{1}{\Delta\tau} \ln\langle C_A(\tau + \Delta\tau, U)\rangle / \langle C_A(\tau, U)\rangle \tag{2.1}$$

where the average is over an ensemble of link fields and $\Delta\tau = O(1)$, chosen to optimize the result. When the correlator is dominated by the ground state, the effective mass should be a constant, the ground state energy. Fig. 1 shows an effective mass plot for the case for $A = 3$ provided by the NPLQCD collaboration clearly exhibiting three types of behavior: strong τ-dependence at short times when excited states contribute, a plateau dominated by the mass of the triton, and degradation into noise at later time.

The growth of noise in the correlator makes it difficult to extract a signal without a lot of statistics, and the problem is expected to get worse with increasing baryon number. It would have been strange if properties of matter at finite baryon density, which are extremely difficult to compute in the grand canonical ensemble, were easy to compute in a canonical simulation — and indeed this noise problem exhibited by the correlator at late time is presumably an avatar of the sign problem. This sign/noise problem does not arise simply because of Fermi statistics; for example, constructing correlators C_A as $3A \times 3A$ Slater determinants of quark propagators accounts for Fermi statistics and leads to a computational cost from the determinant only scaling as $(3A)^3$, not the exponential difficulty seen in the grand canonical computation. Instead, the noise is closely related to the physical spectrum, as has been quantified by Lepage [4]. For example, in QCD the expectation of a $3A$ quark correlator for a nucleus of atomic number A and mass M_A is $\langle C_A \rangle \sim e^{-M_A \tau}$, while the variance in the sample mean $\overline{C}_A$ can be estimated as

$$\sigma^2 = \frac{1}{\mathcal{N}} \left(\langle C_A^\dagger C_A \rangle - \langle C_A^\dagger \rangle \langle C_A \rangle \right) \sim \frac{1}{\mathcal{N}} e^{-3Am_\pi \tau} \tag{2.2}$$

for sample size $\mathcal{N}$. Since C_A corresponds to $3A$ quark propagators and $C_A^\dagger$ to $3A$ anti-quark propagators, the variance is dominated by the state with $3A$ pions. Thus the signal to noise ratio scales as $\sim \sqrt{\mathcal{N}} \exp(-3A\zeta\tau)$, where $\zeta = (M_N/3 - m_\pi/2)$ is the same parameter we saw characterizing the sign problem in the grand canonical case. This reasoning is rather simplistic, as the overlap between the operators and the nucleon or pion states — which will typically contain volume factors – has not been included. However, Fig. 2 shows evidence from QCD simulations by NPLQCD that the Lepage argument is qualitatively correct. The noise and sign problems are therefore presumably closely related and determined by the physical spectrum of the theory and should not be thought of as solely a "fermion sign problem"; similar issues can also arise in interacting boson theories.

In either case the problem arises from the existence of multiparticle states for which the energy/constituent is lower than for the states one wants to study: in QCD, quarks are in some sense lighter when they are in a pion than when they are in a nucleon, by the amount ζ. This is a problem
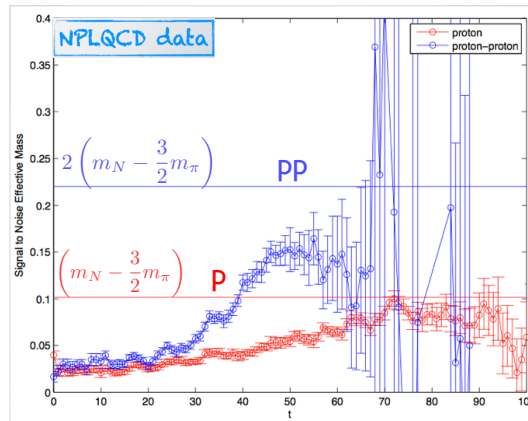

Figure 2: A plot of $-1/t \ln$(signal/noise ratio) versus time for p (red) and pp (purple) correlators computed in lattice QCD; horizontal lines give the expected values for these quantities using the Lepage argument. Figure supplied by K. Orginos (NPLQCD), annotation by D.K.
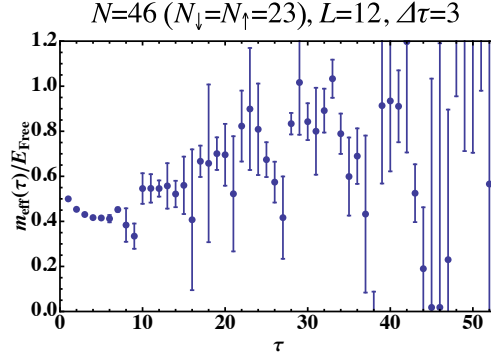
Figure 3: The effective mass plot for 46 unitary fermions in a box of size $L = 12$.

that can occur in bosonic systems as well, and so one should not expect the sign problem to simply be a fermion problem. Simple bosonic path integrals in the grand canonical approach, such as gauged theories with a ϕ^4 interaction, manifestly have complex action in Euclidian spacetime and a sign problem.

Applying the Lepage argument for $C_A(\tau, U)$ to higher moments one sees that all odd moments are proportional to $\exp(-M_A\tau)$ while even moments are all dominated by pion states[1]. Thus at very late time τ, the probability distribution for $C_A(\tau, U)$ will be almost symmetric, implying a noisy sample with small mean. At earlier times, however, we will see that the distribution appears to be very asymmetric and heavy tailed in both QCD and other theories, and it is this regime we wish to analyze.

3. Unitary fermions and a Mean Field Description

Nonrelativistic spin half fermions with strong short-range interactions tuned to a conformal fixed point where the phase shift satisfies $\delta(k) = \pi/2$ for all k are called "unitary fermions". One can think of the inter-particle interaction as being described by a square well potential tuned to have a single bound state with zero binding energy, as one takes the width of the potential to zero. The field theory describing this system is conformal, and interesting to study both for its simplicity and universality, its challenges for many-body theory, and because it can be realized and studied experimentally using trapped atoms tuned to a Feshbach resonance [8]. It is also an ideal nonperturbative theory for studying fermion sign problems on the lattice, being simpler and faster to simulate than QCD. At its most basic, the lattice action involves the simplest discretization of the Euclidean Lagrangian [9]

$$\psi^\dagger(\partial_\tau - \nabla^2/2M)\psi - \tfrac{1}{2}m^2\phi^2 + \phi\,\psi^\dagger\psi \tag{3.1}$$

where ϕ is a non-propagating auxiliary field with m^2 tuned to a critical value, and ψ is a spin $\frac{1}{2}$ fermion with mass M; a more sophisticated action tuned to reduce discretization errors was recently presented in [10, 11].

---

[1]Thanks to Martin Savage for first pointing this out to us.

5

The ground state for $N = (N_\uparrow + N_\downarrow)$ unitary fermions can be determined by computing the correlator for each background ϕ field, and then averaging over the ensemble of such fields. One finds that when N is large, the correlators rapidly become extremely noisy and drift away from a plateau; see for example Fig. 3 for an effective mass plot for $N = 46$ unpolarized fermions in a box with periodic boundary conditions of size $L = 12$. In order to better understand the problems apparent for $\tau \gtrsim 10$, it is instructive to plot a histogram of the correlator before averaging over the ϕ field. This is shown in Fig. 4, which reveals a distribution for N-body correlators $C_N(\tau, \phi)$ which is increasingly non-Gaussian at late τ; in fact, Fig. 4 shows that it is $\ln C_N$ which appears to be roughly normally distributed, so that $C_N(\tau, \phi)$ is approximately *log-normal* distributed with an increasingly large σ and long tail at late time.

A log-normal distribution for the correlator is difficult to reconcile with the Lepage argument. First of all, Lepage assumes that the central limit theorem applies and that the distribution of correlators is close to Gaussian and well described by the mean and variance, which is evidently not the case here. As we will see in a toy model below, the Central Limit Theorem may be irrelevant in such cases, requiring an exponentially large number of configurations to be applicable. Secondly, the size of the variance for this log-normal distribution cannot be related to the physical spectrum of the theory using the Lepage argument. The Lepage argument relates the variance in measurements of C_N to the energy of the lightest state coupling to $(C_N)^2$; there are no anti-particles in this theory, and and so $(C_N)^2$ couples to the ground state of a system N particles each of two different species of spin $1/2$ unitary fermions [2]. However, such a theory in the continuum does not have a finite energy ground state, being unstable against condensing into a point-like clump. On the lattice, a ground state exists for such a system since there is a maximum density due to Fermi statistics and the cutoff; this ground state is a lattice artifact, however, and even then may have a vanishingly small probability to be created by $(C_N)^2$, in which case the variance might be determined by some metastable or excited state.

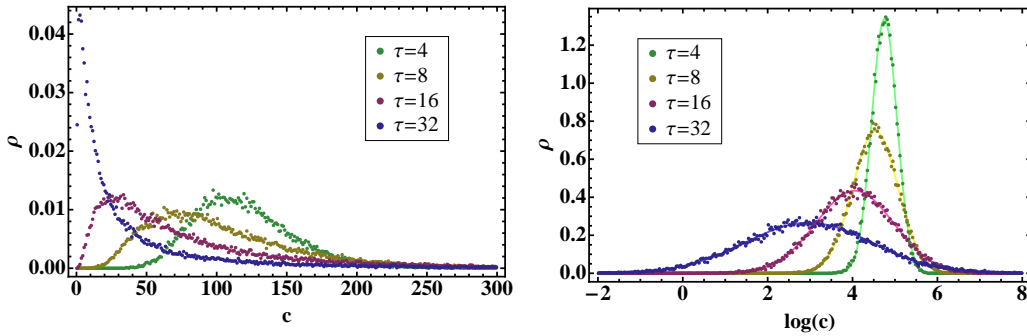The appearance of a heavy-tailed distribution should not be surprising, however, since having

Figure 4: Distribution histograms for $c = C_N(\tau, \phi)$ and $\ln(c)$ for $N = 4$ unitary fermions at several times τ, taken from Ref. [11]. Curves fitting $\ln(c)$ are Gaussian, implying that c is approximately log-normal distributed, with σ^2 increasing with time.

---

[2]If C_N is the product of $N_\downarrow$ antisymmetrized correlators times $N_\uparrow$ antisymmetrized correlators, then $(C_N)^2$ corresponds to a product of four groups of antisymmetrized correlators, and is equivalent to the propagator for two distinguishable species of spin half fermions.

fermions wander through a random ϕ background is similar to the problem of electrons propagating in disordered media, where heavy-tailed distributions are ubiquitous in the vicinity of the Anderson localization transition. For example, it is found that for physical quantities such as the current relaxation time or normalized local density of states, the distribution function $P(z)$ scales as $\exp(-C_d \ln^d z)$. A particularly simple way to derive these results is to use the optimal fluctuation method of Ref. [12], which is a mean field approach. We can adapt these methods to the current problem, defining the variable $Y = \ln C_N(\tau, \phi)$ and computing its probability distribution $P(y)$ as

$$P(y) \propto \int D\phi \, e^{-S_\phi} \, \delta(Y(\tau, \phi) - y) = \int D\phi \, \frac{dt}{2\pi} e^{-S} \tag{3.2}$$

where $S_\phi = \int d^4x \frac{m^2}{2} \phi^2$ and $S = S_\phi - it(\ln C_N(\tau, \phi) - y)$. Using the PDS subtraction scheme [13] we have $m^2 = M\lambda/4\pi$, where the renormalization scale λ is taken to be the physical momentum scale in the problem — in this case $\lambda = k_F \equiv (3\pi^2 N/V)^{1/3}$, $N/2$ being the number of fermions with a single spin orientation. We proceed now to evaluate this integral using a mean field expansion; it is not evident that there is a small parameter to justify this expansion, but the leading order result is illuminating and fits the numerical data well. We expand about $\phi(x) = \phi_0$, $t = t_0$, and use the fact that for large τ the nth functional derivative of $\ln C_N(\tau, \phi)$ with respect to $\phi(x)$ equals the the 1-loop Feynman diagram with n insertions of $\psi^\dagger \psi$ in the presence of a chemical potential $\mu = k_F^2/(2M)$. The equations for ϕ_0 and t_0 are given by

$$t_0 = -i\frac{m^2\phi_0}{\langle n(x) \rangle_c} = -i\frac{Vm^2\phi_0}{N}$$
$$\phi_0 = -\frac{y - \ln Z + \tau E_0(N)}{N\tau} \tag{3.3}$$

where $E_0(N) = 3NE_F/5$ is the total energy of N free degenerate fermions ($N/2$ of each spin), and Z is the overlap of the source and sink with the free fermion state. The leading term in the mean field expansion for $P(y)$ can therefore be expressed as $P(y) \propto \exp\left[-\frac{(y-\bar{y})^2}{2\sigma^2}\right]$ with

$$\bar{y} = \ln Z - \tau E_0(N) , \quad \sigma^2 = \frac{40}{9\pi} E_0(N)\tau . \tag{3.4}$$

This describes a log-normal distribution for the N-fermion propagator $C_N(\tau, \phi)$, with both mean and variance growing in magnitude with time in units of the energy of N free degenerate fermions. In Fig. 5 we plot the quantities $-\frac{1}{E_0}\frac{\partial \bar{y}}{\partial \tau}$ and $\frac{1}{E_0}\frac{\partial \sigma^2}{\partial \tau}$ as a function of N obtained from correlator distribution data for unitary fermions at late τ, and find that the gross features of the results are compatible with the mean field estimates of unity and $40/9\pi$ obtained from eq. (3.4).

4. A toy model

It would be useful to devise an algorithm to reliably estimate energies without having to exhaustively sample the long tail of the correlator distribution, yet without making incorrect assumptions about the exact functional form of that tail. An approach we suggest here is to exploit the general relationship between stochastic variables X and $Y = \ln X$:

$$\ln\langle X \rangle = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} \tag{4.1}$$

where κ_n is the nth cumulant of Y. This relation can be proved by noting that the generating function for the κ_n is $\ln \phi_Y(t)$ where $\phi_Y(t) = \langle e^{Yt} \rangle = \langle X^t \rangle$ is the moment generating function for Y, and evaluating at $t = 1$, assumed to be within the radius of convergence. The motivation for investigating eq. (4.1) is that if the distribution $P(X)$ were exactly log-normal, the above sum would end after the second term, as $\kappa_{n>2}$ would all vanish; therefore by replacing the κ_n by sample cumulants $\bar{\kappa}_n$ and truncating the sum at $n = n_{max}$, one might hope to have a reliable estimator for $\ln \langle X \rangle$, provided that the systematic error from truncating eq. (4.1) and the statistical error from sampling $\bar{\kappa}_n$ can be simultaneously minimized.

Distributions with log-normal-like tails arise naturally in products of stochastic variables. The propagator $C_N(\tau, \phi)$ for unitary fermions can be expressed in a transfer matrix formalism as the product of τ matrices — one per time hop — each of which is the direct product of N $V \times V$ matrices of the form $e^{-K/2}(1 + g\varphi)e^{-K/2}$, where K is a constant matrix (the spatial kinetic operator), φ is a random diagonal matrix with $O(1)$ entries corresponding to stochastic ϕ fields living on the time links, and g is a coupling constant (identified with $1/m^2$ from eq. (3.1)) that has been tuned to a particular critical value that is $O(1)$. Unfortunately, little seems to be known about products of random matrices beyond dimension two [16]. Therefore we analyze instead a toy model where we define a "correlator" C_τ as a product of random numbers, and an "energy" $\mathscr{E} = \lim_{\tau \to \infty} \mathscr{E}_\tau$ where:

$$C_\tau = \prod_{i=1}^{\tau}(1 + g\varphi_i), \quad \mathscr{E}_\tau = -\frac{1}{\tau}\ln \langle C_\tau \rangle \tag{4.2}$$

where $0 \leq g \leq 1$ and the φ_i are independent and identically distributed random numbers with a uniform distribution on the interval $[-1, 1]$. The exact value for the energy is obviously $\mathscr{E}_\tau = 0$ since the statistical average of the correlator is $\langle C_\tau \rangle = 1$. The cumulants of the variable $Y = \ln(C_\tau)$ are given by

$$\kappa_1 = \tau \left[\tfrac{1}{2} \log\left(1 - g^2\right) + \frac{\tanh^{-1}(g)}{g} - 1 \right],$$
$$\frac{\kappa_n}{n!} = \tau \left(\frac{(-1)^n}{n} - Li_{1-n}\left(\frac{1+g}{1-g}\right) \frac{\left(2\tanh^{-1}(g)\right)^n}{n!} \right)$$

for $n \geq 2$; for $g < 1$ one finds that the $\kappa_n/n!$ rapidly decrease as n increases.

In Fig. 6 we show the results of a simulation where we compute $\mathscr{E}_\tau$ for $g = \frac{1}{2}$ and $\tau = 1, \ldots, 1000$. At each value of τ we independently generated an ensemble of values for C_τ of
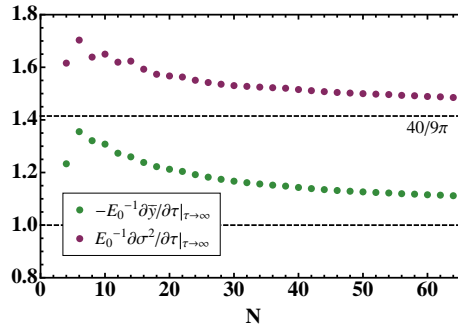

Figure 5: The quantities $-\frac{1}{E_0}\frac{\partial \bar{y}}{\partial \tau}$ and $\frac{1}{E_0}\frac{\partial \sigma^2}{\partial \tau}$ as a function of N for unitary fermions at late times on a lattice of size $L = 10$, compared to mean field prediction (dashed lines).
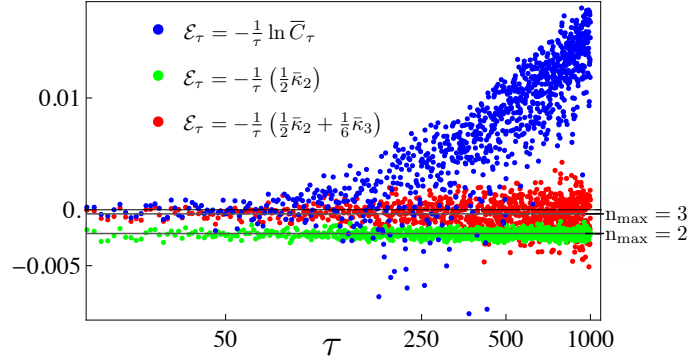
Figure 6: Simulation of the energy $\mathcal{E}_\tau$ for the toy model eq. (4.2) with $g = \frac{1}{2}$. The exact answer is $\mathcal{E}_\tau = 0$ (black line); exact values of the expansion eq. (4.1) truncated at order $n_{\max} = 2, 3$ are indicated.

Method	$\mathcal{E}$	stat. error	syst. error
conventional	0.014932	0.002485	–
$\kappa_{n\leq 2}$	-0.002159	0.000304	-0.002165
$\kappa_{n\leq 3}$	-0.000412	0.001618	-0.000324
$\kappa_{n\leq 4}$	-0.000647	0.008379	0.000050
$\kappa_{n\leq 5}$	-0.001794	0.037561	3.34×10^{-6}
$\kappa_{n\leq 6}$	0.010943	0.147739	-1.22×10^{-6}

Table 1: $\mathcal{E}$ determined from 250 blocks of 50,000 configurations each for the model with $\tau = 1000$, $g = 1/2$.

size $N = 50,000$. From that ensemble we computed $\mathcal{E}_\tau$ by (i) using the conventional estimator $\mathcal{E}_\tau = -\frac{1}{\tau} \ln \overline{C}_\tau$ (blue), which shows a striking systematic error for $\tau \gtrsim 50$, and statistical noise increasing up to $\tau \simeq 500$ but decreasing beyond that; (ii) using eq. (4.1) truncated at $n_{max} = 2$ using sample cumulants $\overline{\kappa}_n$ (green), showing a τ-independent systematic error with smaller but slowly growing statistical error; (iii) eq. (4.1) truncated at $n_{max} = 3$ (red) with a negligible constant systematic error but a larger statistical error. Evidently, one trades systematic error for statistical error by truncating eq. (4.1) at increasingly large n_{max}.

Table 1 displays results of a simulation of 1.25×10^7 ϕ configurations blocked into 250 blocks of 50,000 each, for the model eq. (4.2) at $\tau = 1000$ and $g = 1/2$. We give the conventional estimate $\mathcal{E}_\tau = -1/\tau \ln \overline{C}_\tau$ and estimates based on the cumulant expansion eq. (4.1) truncated at various n_{max}, where the exact value is $\mathcal{E} = 0$. For each method we give the computed value for $\mathcal{E}$ and the statistical error; for the cumulant expansion we also give the exact systematic error from truncating eq. (4.1) at $n = n_{max}$ using our analytic expressions for κ_n. These numbers show how the conventional method gives a wrong answer with deceptively small statistical error. For the cumulant expansion one sees again the trade of systematic error for statistical error with increasing n_{max}. Table 1 shows the combined error is minimized for $n_{max} = 3$, justified by noting that the $n_{max} = 4$ result with statistical errors encompasses the $n_{max} = 3$ result; we suggest this as a practical algorithm for determining where to truncate the cumulant expansion in general.

Leaving the toy model and returning to real simulations of unitary fermions, Fig. 7 shows how the cumulant expansion works for 50 trapped unitary fermions [11], where truncating the expansion
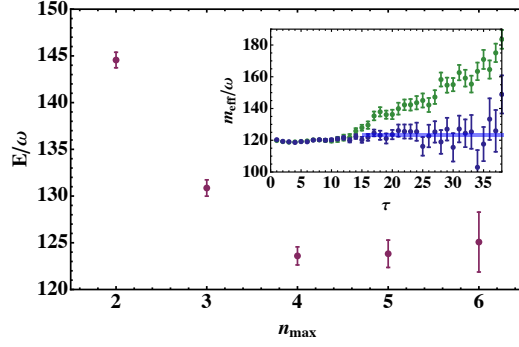
Figure 7: Energy for 50 unitary fermions in a harmonic trap of frequency ω, 10^6 configurations; fits performed using truncated at order n_{max}. Inset: conventional effective mass $m_{eff}(\tau) = \log \bar{C}(\tau)/\bar{C}(\tau+1)$ (green) and corresponding fitted cumulant effective mass with $n_{max} = 4$ (blue).
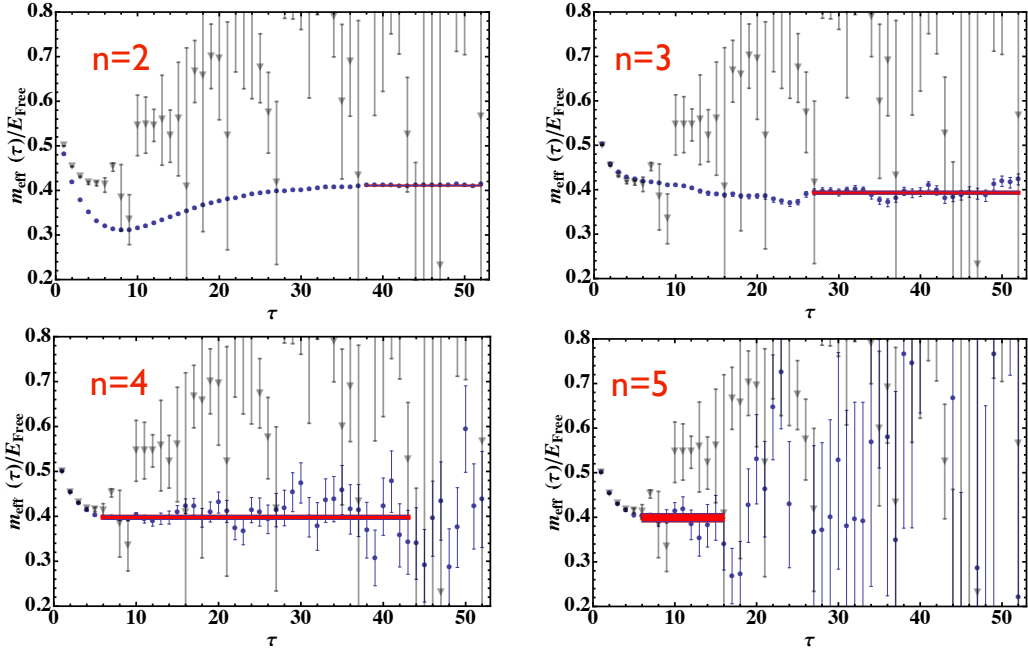

Figure 8: Effective mass plot for the same data displayed in Fig. 3, including $n = 2,3,4,5$ terms in the expansion eq. (4.1). The conventional analysis is given by gray triangles; the cumulant expansion results are blue circles. In each case we have marked the plateaus we extract, with bands whose widths signify the computed error. If the correlator distribution was exactly log-normal, the $n = 2$ analysis would be exact.

at $n_{max} = 4$ is supported by the data. In Fig. 8 we show the improvement to the same data displayed above in Fig. 3 when we use the expansion eq. (4.1) to order $n = 2,3,4,5$. Again, as in the toy model, it is evident from this plot how increasing the number of cumulants included in the truncated version of eq. (4.1) decreases the systematic error, but at the cost of increased statistical error.

5. Discussion

Heavy-tail distributions are likely to be ubiquitous in *N*-body simulations, including multi-
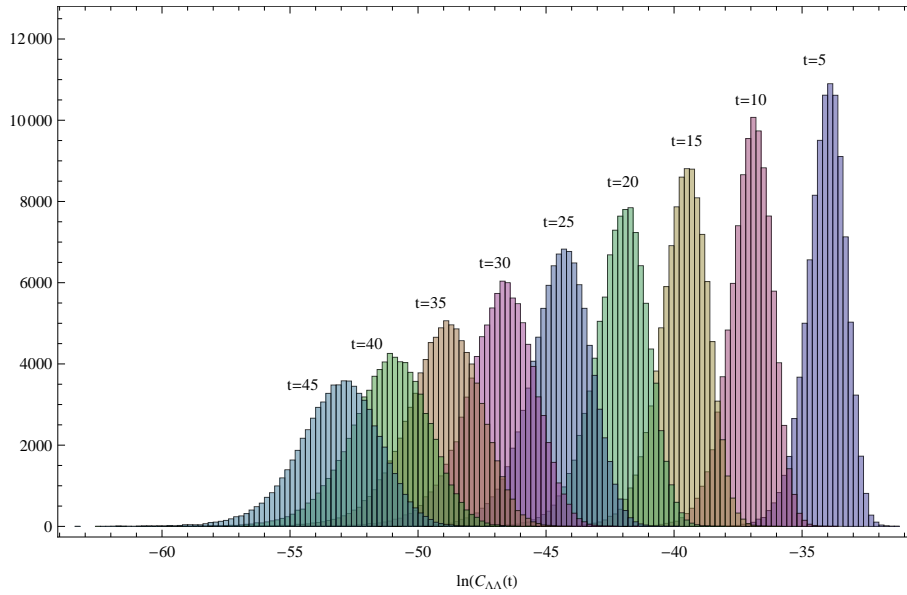
Figure 9: Statistical distribution of the log of the real part of $\Lambda\Lambda$ correlators at different temporal extent t suggesting a nearly log-normal distribution for the correlators. Each curve represents 100,000 samples; a small number of correlators with negative real part were discarded. Figure from W. Detmold (NPLQCD).

baryon computations in lattice QCD. An example of the latter is given in Fig. 9, consisting of histograms of the log of the $\Lambda\Lambda$ correlator, as computed by the NPLQCD collaboration from lattice QCD simulations. This plot shows evidence that the correlator must have a nearly log-normal distribution during the times considered. Presumably at very late times the distribution of the correlator will go over to a Gaussian with large variance and small mean, following the Lepage argument; however, at these earlier times that argument evidently fails, and the types of statistical techniques described here and in Ref. [1] might improve the analysis of such QCD data. It is interesting to speculate whether evidence for log-normal behavior in QCD implies that a mean field description could be valid for QCD correlators at moderate times, as discussed in §3; note the similarity between the QCD example Fig. 9, and the analogous plot for unitary fermions, Fig. 4 — as well as the different way the two distributions scale with time. One might expect that if such a mean field description is approximately valid for multibaryon correlators in QCD, as baryon number is increased that description would become even more accurate, and log-normal behavior would become ubiquitous.

It is also interesting to wonder whether the expansion in eq. (4.1) could be reformulated in an effective field theory language, where higher cumulants play the role of irrelevant operators. The reason the correlator distribution is flowing toward a simple universal form (log-normal) is presumably because fermions propagating through a single background field configuration are sampling many values of that field which are nearly statistically independent. What one measures is a sum of products of weakly correlated random matrices. The log of the correlator is in some sense averaging over short range fluctuations in the field, and it must be this averaging process — similar to RG blocking — which accounts for the flow of the distribution toward a Gaussian infrared attractor, with deviations characterized by the cumulants of the log of the correlator. Analogies between the

RG and statistics are well known — an example being the Central Limit Theorem (see, for example, [14]), or extreme statistics [15]. If the physical basis for attractors for the statistical distribution of correlators was better understood (not much is known about the distributions of products of random matrices, let alone sums of such products, but see [16]), including a systematic parameterization of deviations from the fixed point — analogous to the systematic inclusion of the effects of irrelevant operators in an effective field theory — then it is conceivable that one could greatly enhance one's ability to find a signal hiding in what superficially looks like useless noise.

Acknowledgments

References

[1] M. G. Endres, D. B. Kaplan, J. -W. Lee and A. N. Nicholson, *Noise, sign problems, and statistics*, Phys. Rev. Lett. **107**, 201601 (2011) [arXiv:1106.0073 [hep-lat]].

[2] M. Troyer and U. -J. Wiese, *Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations*, Phys. Rev. Lett. **94**, 170201 (2005) [cond-mat/0408370].

[3] P. E. Gibbs, *Understanding finite baryonic density simulations In lattice QCD*, PRINT-86-0389 (GLASGOW).

[4] G. P Lepage, *The analysis of algorithms for lattice field theory*, invited lectures given at TASI'89 Summer School, Boulder, CO, Jun 4-30, 1989.

[5] K. Splittorff and J. J. M. Verbaarschot, *Phase of the Fermion Determinant at Nonzero Chemical Potential*, Phys. Rev. Lett. **98**, 031601 (2007) [hep-lat/0609076].

[6] T. D. Cohen, *QCD functional integrals for systems with nonzero chemical potential*, In *Shifman, M. (ed.) et al.: From fields to strings, vol. 1* 101-120 [hep-ph/0405043].

[7] S. R. Beane et al. (NPLQCD collaboration), *High Statistics Analysis using Anisotropic Clover Lattices. II. Three-Baryon Systems*, Phys. Rev. D **80**, 074501 (2009) [arXiv:0905.0466 [hep-lat]].

[8] T.-L Ho, *Universal thermodynamics of degenerate quantum gases in the unitarity limit*, Phys. Rev. Lett. **92**, 090402 (2004) [cond-mat/0309109].

[9] J. -W. Chen and D. B. Kaplan, *A Lattice theory for low-energy fermions at finite chemical potential*, Phys. Rev. Lett. **92**, 257002 (2004) [hep-lat/0308016].

[10] M. G. Endres, D. B. Kaplan, J. -W. Lee and A. N. Nicholson, *A new approach for studying large numbers of fermions in the unitary regime*, PoSLATTICE **2010**, 182 (2010) [arXiv:1011.3089 [hep-lat]].

[11] M. G. Endres, D. B. Kaplan, J. -W. Lee and A. N. Nicholson, *Lattice Monte Carlo calculations for unitary fermions in a harmonic trap*, Phys. Rev. A **84**, 043644 (2011) [arXiv:1106.5725 [hep-lat]].

[12] I. Smolyarenko and B. Altshuler, *Statistics of rare events in disordered conductors*, Physical Review B **55**, 10451 (1997) [cond-mat/9606181].

[13] D. B. Kaplan, M. J. Savage and M. B. Wise, *A new expansion for nucleon-nucleon interactions*, Phys. Lett. B **424**, 390 (1998) [nucl-th/9801034].

[14] D. Sornette (2004). *Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder: concepts and tools*, Berlin: Springer.

[15] Györgyi, G. et al., *Renormalization-group theory for finite-size scaling in extreme statistics*, Phys. Rev. E **81**, 041135, (2010)

[16] A. D Jackson et al., *Products of random matrices*, Phys. Rev. E **66**, 066124 (2002) [physics/0202037].