An exact formula for the variance of linear statistics in the one-dimensional jellium model

Ana Flack\textsuperscript{1}, Satya N Majumdar\textsuperscript{1} and Grégory Schehr\textsuperscript{2,\*} \\textsuperscript{1} LPTMS, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France \textsuperscript{2} Sorbonne Université, Laboratoire de Physique Théorique et Hautes Energies,CNRS UMR 7589, 4 Place Jussieu, 75252 Paris Cedex 05, France \\E-mail: schehr@lpthe.jussieu.fr

Received 28 November 2022; revised 24 January 2023 \Accepted for publication 1 February 2023 \Published 21 February 2023

Abstract

We consider the jellium model of $N$ particles on a line confined in an external harmonic potential and with a pairwise one-dimensional Coulomb repulsion of strength $\alpha>0$. Using a Coulomb gas method, we study the statistics of $s = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$ where $f(x)$, in principle, is an arbitrary smooth function. While the mean of $s$ is easy to compute, the variance is nontrivial due to the long-range Coulomb interactions. In this paper we demonstrate that the fluctuations around this mean are Gaussian with a variance $\text{Var}(s) \approx b/N^3$ for large $N$. In this paper, we provide an exact compact formula for the constant $b = 1/(4\alpha) \int_{-\infty}^{\infty} [f'(x)]^2 \, dx$. In addition, we also calculate the full large deviation function characterizing the tails of the full distribution $P(s,N)$ for several different examples of $f(x)$. Our analytical predictions are confirmed by numerical simulations.

Keywords: Coulomb gas, random matrices, long-range interactions

(Some figures may appear in colour only in the online journal)

1. Introduction

Understanding the probability distribution of the sum of random variables is a well-known problem in probability theory with multiple applications in physics, chemistry, statistics and biology. Consider for instance $N$ random variables $\{x_1, x_2, \ldots, x_N\}$ drawn from a joint probability distribution function (jPDF) $P_{\text{joint}}(x_1, x_2, \ldots, x_N)$. The general question is: how is the sum $S_N = \sum_{i=1}^{N} x_i$ distributed? In the simplest case where the variables are independent and...
identically distributed, each drawn from \( p(x) \) (with zero mean for simplicity), then the jPDF factorizes, i.e., \( P(x_1, x_2, \ldots, x_N) = p(x_1)p(x_2) \ldots p(x_N) \). In this case, the central limit theorem (CLT) guarantees that, as long as the variance \( \sigma^2 = \int x^2 p(x) \, dx \) is finite, the sum \( S_N \) converges for large \( N \) to a Gaussian random variable, i.e., \( S_N/(\sigma \sqrt{N}) \to N(0, 1) \) where \( N(0, 1) \) is a standard normal variable with zero mean and unit variance.

However, for correlated random variables, in particular when the correlations are strong, the CLT does not hold in general and determining the PDF of \( S_N \) for large \( N \) is a challenging problem and appears in many different contexts. A celebrated example is a long-ranged gas with repulsive pairwise interactions and confined in a harmonic potential in one-dimension. This is the so-called Riesz gas where the energy of a configuration of charges with positions \( x_i \)'s on a line is given by \([1, 2]\)

\[
E_{\{x_i\}} = \frac{A}{2} \sum_{i=1}^{N} x_i^2 + \alpha \sum_{i \neq j} |x_i - x_j|^{-k},
\]

where \( \alpha > 0 \) and \( A > 0 \) are fixed coupling constants of order \( O(1) \). While the first term represents the potential energy due to the external harmonic potential, the second term represents the repulsive interaction between the \( i \)th and \( j \)th particle. In equation (1), we assume that \( k > 2 \) (otherwise the harmonic potential is not sufficient to confine the gas). The jPDF of the positions of the particles is given by the Gibbs–Boltzmann distribution

\[
P(x_1, x_2, \ldots, x_N) = \frac{1}{Z_N} e^{-E_{\{x_i\}}/(k_B T)},
\]

where \( T \) is the temperature, \( k_B \) is the Boltzmann constant and \( Z_N \) is the normalizing partition function. In this case, the jPDF \( P(x_1, x_2, \ldots, x_N) \) does not factorize because of the interaction term in equation (1), which is long-ranged. Hence, it represents an example of a strongly correlated gas which has been extensively studied \([3–8]\). There are three limiting cases of this Riesz gas that have been studied widely in the literature, namely (a) \( k = -1 \); this is the famous jellium model also known as the one-dimensional one-component plasma (1d-OCP) \([9–17]\), (b) \( k \to 0^+ \); this is the well-known Dyson’s log-gas \([18–23]\) where the positions \( x_i \)'s correspond to the eigenvalues of Gaussian random matrices and (c) \( k = 2 \), corresponding to the Calogero–Moser model which represents an important integrable classical model \([24–28]\). In these examples of the Riesz gas, \( S_N/N \) represents the position of the center of mass of the particles, a natural and important physical observable.

More generally, one can also study the probability distribution of the so-called linear statistics

\[
S_N = \sum_{i=1}^{N} f(x_i),
\]

where \( f(x) \) is an arbitrary function (not necessarily linear). In the case when \( f(x) = x \), it reduces to the center of mass (multiplied by \( N \)) of the gas. However, the case of a general \( f(x) \) is also of interest and has been widely studied in several special cases, in particular in the context of the random matrix theory \([29–44]\). Some examples are as follows

- Full counting statistics in the Dyson’s log-gas \([36, 45–50]\) or in the jellium model \([14, 16]\), where one is interested in the number of particles \( N_L = \sum_{i=1}^{N} \mathbb{1}_{[-L, L]}(x_i) \) in an interval \([-L, L]\). Here, \( f(x) = \mathbb{1}_{[-L, L]}(x) \) is an indicator function, i.e. \( \mathbb{1}_{[-L, L]}(x) = 1 \) if \( -L \leq x \leq L \) and zero otherwise.
Conductance (with \( f(x) = x \) and shot noise \( f(x) = x(1 - x) \)) in chaotic transport through a cavity where \( 0 \leq x_i \leq 1 \)'s represent the eigenvalues of an \( N \times N \) Jacobi random matrix [30, 38, 40, 51–58].

Rényi entropy in a random pure state of a bipartite system where \( f(x) = x^q \), with \( q > 0 \) represents the Rényi index [59–61].

In the case of Dyson’s log-gas, i.e. \( k \to 0^+ \) limit of the Riesz gas, the linear statistics \( S_N \) in equation (3) with general \( f(x) \) have been studied extensively and several exact results are known. In order to study the linear statistics in the log-gas, it is convenient to first rescale \( x_i \to x_i/\sqrt{N} \)'s such that the average density of the gas is supported, for large \( N \), over an interval of size of order \( O(1) \). Then one defines \( s = (1/N) \sum f(x_i) \) where \( x_i \)'s now represent the rescaled coordinates such that \( s \sim O(1) \) in the large \( N \) limit. It is known that the distribution of the typical value of \( s \) for a general \( f(x) \) is Gaussian around its mean \( \bar{s} \) with a variance given, for large \( N \), by an explicit formula [30, 35, 40]

\[
\text{Var}(s) \approx \frac{1}{\beta^2 N^2} \int_{0}^{\infty} k |\hat{f}(k)|^2 dk, \tag{4}
\]

where \( \hat{f}(k) = \int_{-\infty}^{\infty} e^{ikx} f(x) dx \) is the Fourier transform of \( f(x) \) and the parameter \( \beta = 1, 2, 4 \) represents the three standard symmetry classes of Gaussian random matrices. Note that this formula assumes that the integral in (4) is convergent. For certain choices of \( f(x) \) one has to augment this formula with some regularization. For example for the center of mass where \( f(x) = x \), its Fourier transform \( \hat{f}(k) \) is not well defined and one cannot use the formula in equation (4) directly and use alternative formulas that are more complicated [30].

A natural question is whether one can obtain an explicit formula for general \( f(x) \) for other values of \( k \) in the Riesz model (1). The purpose of this paper is to derive an explicit formula for \( k = -1 \), i.e. for the 1d jellium model, in the limit of large \( N \). As in the case of the log-gas discussed above, it is useful to first rescale \( x_i \to L_N y_i \) where \( y_i = O(1) \) and \( L_N \) is an \( N \)-dependent length scale to be chosen as follows. Under this rescaling, the energy in equation (1) for \( k = -1 \) reads

\[
E[\{y_i\}] = \frac{A}{2} \sum_{i=1}^{N} y_i^2 - \alpha L_N \sum_{i \neq j} |y_i - y_j|. \tag{5}
\]

For the system to exhibit any interesting physical behavior, the two terms in equation (5) must scale in the same way for large \( N \). The first term scales as \( AL_N^2 N \) since the sum \( \sum_i y_i^2 \sim N \) as \( y_i = O(1) \). The second term scales as \( \alpha L_N N^2 \) since there are \( N(N - 1) \) terms of order \( O(1) \) each in the double sum. Equating these two, one gets, for large \( N \),

\[
AL_N \sim \alpha N. \tag{6}
\]

Since both the coupling constants \( A \) and \( \alpha \) are of order \( O(1) \), this tells us that we must choose \( L_N = O(N) \). Choosing \( L_N = N \) and setting \( A = 1 \) for convenience the energy in equation (5) can be written as

\[
E[\{y_i\}] = \frac{N^2}{2} \sum_{i=1}^{N} y_i^2 - \alpha \sum_{i \neq j} |y_i - y_j|. \tag{7}
\]

For convenience, we henceforth will denote \( y_i \to x_i \) and write the energy as

\[
E[\{x_i\}] = \frac{N^2}{2} \sum_{i=1}^{N} x_i^2 - \alpha N \sum_{i \neq j} |x_i - x_j|. \tag{8}
\]
Note that, since $x_i$'s are typically of order $O(1)$, both terms in the energy in (8) scale as $N^3$. Hence the total energy scales as $N^3$. In this case, the average density is defined by

$$\langle \rho_N(x) \rangle = \frac{1}{N} \left( \sum_{i=1}^{N} \delta(x-x_i) \right),$$

where $\langle \ldots \rangle$ denotes an average over the Gibbs–Boltzmann measure in equation (2). It is known to converge in the large $N$ limit to a flat profile supported over $[-2\alpha, 2\alpha]$ [10, 13]

$$\lim_{N \to \infty} \langle \rho_N(x) \rangle = \frac{1}{4\alpha}, \quad -2\alpha \leq x \leq 2\alpha.$$

As in the log-gas case, we now define the linear statistics in the rescaled coordinates as

$$s = \frac{1}{N} \sum_{i=1}^{N} f(x_i),$$

such that $s \sim O(1)$ in the large $N$ limit. In this case, using the average density in (10), it is easy to see that the average value of $s$ converges to

$$\bar{s} = \frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} f(x) \, dx.$$

Our main result in this paper is to obtain, using a Coulomb gas and large deviation method, an explicit formula for the variance $\text{Var}(s) = \langle (s-\bar{s})^2 \rangle$ which simply reads for large $N$

$$\text{Var}(s) \approx \frac{1}{4\alpha N^2} \int_{-2\alpha}^{2\alpha} [f'(x)]^2 \, dx.$$

This formula is valid only when $[f'(x)]^2$ exists and is integrable. We will discuss precisely the assumptions implicit in deriving this explicit formula and also verify this prediction by numerical simulations in several specific examples of $f(x)$. Using the same Coulomb gas method, we show how to access the full large deviation function that characterizes the tails of the full distribution $\mathcal{P}(s,N)$. For some examples of $f(x)$, we compute explicitly the rate function describing these large atypical fluctuations. In some cases, such as $f(x) = |x|$, we show from explicit computation that the rate function has a singular point where the third derivative is discontinuous, signaling a third order phase transition in the underlying Coulomb gas.

We will also provide an alternative derivation of this formula (13), which in addition demonstrates that, up to an overall constant, the formula in equation (13) holds for a general external potential $V(x_i)$, not necessarily harmonic.

Let us remark that the variance of linear statistics has also been studied in another well known ensemble of random matrix theory (RMT), namely the Ginibre ensemble where the entries are complex Gaussian but there is no Hermitian symmetry. In this case, the eigenvalues are complex and the eigenvalues behave like charged particles in two-dimensions repelling each other via a logarithmic Coulomb interaction and also confined in a harmonic potential. For this 2d-case, a formula for the variance of $s = (1/N) \sum_{i=1}^{N} f(\vec{x}_i)$ in the large $N$ limit was derived by Forrester [62], which looks formally very similar to our one-dimensional formula in (13), up to overall constants and $N$-dependent factors, i.e. $\text{Var}(s) \propto \int (\nabla f)^2 d\vec{r}$ where the integral is over the support of the Coulomb gas. This result in $d = 2$ was later proved rigorously in the mathematics literature [63, 64] and the formula was also extended to confined Coulomb gases in $d \geq 2$ [65, 66]. However, these rigorous methods do not seem to easily extend to the jellium model in $d = 1$. On the other hand, our exact result for $d = 1$ in equation (13) seems to suggest that the formula for the variance $\text{Var}(s) \propto \int (\nabla f)^2 d\vec{r}$ holds even in $d = 1$ up to an
overall factor. Our method, in addition, gives access to the full large deviation function for any linear statistics.

The rest of the paper is organized as follows. In section 2, we discuss the Coulomb gas method leading to the exact asymptotic formula in equation (13) for general \( f(x) \). We also study few examples where we provide explicit results for the large deviation function associated with the full distribution of \( s \). In section 3, we provide an alternative derivation of the formula for the variance. In section 4, we discuss the criteria for the validity of this general formula. Finally, in section 5, we conclude with a summary and open problems.

2. The computation of the variance via a Coulomb gas method

We consider the linear statistics \( s \) defined in equation (11) with an arbitrary function \( f(x) \). Clearly \( s \) is a random variable since the \( x_i \)'s are also random variables distributed via the Gibbs–Boltzmann weight (2) with the energy given in equation (8). Our goal is to calculate the variance \( \text{Var}(s) \) of \( s \) in the large \( N \) limit. To compute it, we use the following strategy. We first express the PDF \( \mathcal{P}(s,N) \) as follows

\[
\mathcal{P}(s,N) = \int dx_1 \int dx_2 \ldots \int dx_N P\{\{x_i\}\} \delta \left( s - \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right),
\]

with \( P\{\{x_i\}\} \) given by the Boltzmann weight (using equations (2) and (8))

\[
P\{\{x_i\}\} = \frac{e^{-E[\{x_i\}]}}{\int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_N e^{-E[\{x_i\}]}}
\]

where

\[
E[\{x_i\}] = \frac{N^2}{2} \sum_{i=1}^{N} x_i^2 - \alpha N \sum_{i \neq j} |x_i - x_j|.
\]

The idea is to evaluate this distribution in the large \( N \) limit using a Coulomb gas method detailed below. For large \( N \), we will see that \( \mathcal{P}(s,N) \) admits a large deviation form

\[
\mathcal{P}(s,N) \sim e^{-N \Psi(s)},
\]

where \( \Psi(s) \) is a rate function that implicitly depends on \( f(x) \). The factor \( N^3 \) in the exponent arises from the fact that the energy scales as \( N^3 \). Typically, one would expect that \( \Psi(s) \) has a minimum around \( s = \bar{s} \) (where \( \bar{s} \) denotes the average value of \( s \) in equation (12)) and behaves quadratically around its mean

\[
\Psi(s) \approx \frac{1}{2b} (s - \bar{s})^2.
\]

Substituting this behavior in the large deviation form (17) one gets

\[
\mathcal{P}(s,N) \sim e^{-\frac{N^3}{2b} (s - \bar{s})^2},
\]

indicating that the distribution, near its peak, has a Gaussian form with mean \( \bar{s} \) and variance

\[
\text{Var}(s) \approx \frac{b}{N^3}.
\]

Therefore the idea would be to compute first the large deviation function \( \Psi(s) \) and read off the number \( b \) by expanding \( \Psi(s) \) up to quadratic order around its minimum at \( s = \bar{s} \). In the next subsection, we outline the Coulomb gas method to compute \( \Psi(s) \).
2.1. The Coulomb gas method: general set up

We start from the multiple integral in (14). The first step is to get rid of the delta-function by replacing it with its integral representation \( \delta(x) = \frac{1}{2\pi i} \int_{\Gamma} e^{\mu x} \frac{dx}{x} \) where \( \Gamma \) is a Bromwich contour going along the imaginary axis in the complex \( \mu \)-plane. This gives

\[
P(s,N) \propto \int_{\Gamma} d\mu \int dx_1 \int dx_2 \ldots \int dx_N e^{-E_\mu([x_i])}
\]

where \( E_\mu([x_i]) = \frac{N^2}{2} \sum_{i=1}^{N} x_i^2 - N\alpha \sum_{i\neq j} |x_i - x_j| + \mu \left[ \frac{1}{N} \sum_{i=1}^{N} f(x_i) - s \right] \). (22)

Since the \( x_i \)'s are of order \( O(1) \), we see that both the first two terms scale as \( O(N^3) \) for large \( N \), while the third term multiplying \( \mu \) scales as \( O(\mu N^2) \). In order that all the three terms scale in the same way with increasing \( N \) (needed to compute the distribution of \( s \)), we need to rescale \( \mu = N^3 \mu \) where \( \mu = O(1) \). This then gives

\[
P(s,N) = \int_{\Gamma} d\mu \frac{\int dx_1 \int dx_2 \ldots \int dx_N e^{-E_\mu([x_i])}}{\int dx_1 \ldots dx_N e^{-E_\mu([x_i])}} = \frac{\int_{\Gamma} d\mu Z_N(\mu)}{Z_N}.
\] (23)

In the denominator, the energy function \( E_\mu([x_i]) \) is given in equation (8)—note that for convenience we have set \( k_BT = 1 \), without any loss of generality. In the numerator in equation (23), the modified energy function \( E_\mu([x_i]) \) reads

\[
E_\mu([x_i]) = \frac{N^2}{2} \sum_{i=1}^{N} x_i^2 - N\alpha \sum_{i\neq j} |x_i - x_j| + N^3 \mu \left[ \frac{1}{N} \sum_{i=1}^{N} f(x_i) - s \right]
= \frac{N^2}{2} \sum_{i=1}^{N} (x_i^2 + \mu f(x_i)) - N\alpha \sum_{i\neq j} |x_i - x_j| - \mu s N^3.
\] (24)

Up to a constant shift of energy \( -\mu s N^3 \), the energy in equation (24) can be interpreted as the energy of the same jellium gas, but in the presence of an effective potential

\[
V_{\text{eff}}(x) = \frac{1}{2} x^2 + \mu f(x).
\] (25)

Thus the ‘chemical potential’ \( \mu \) can be interpreted as the amplitude of the perturbation of the original potential \( \lambda^2/2 \). Consequently, \( Z_N(\mu = 0) = Z_N \) in equation (23).

Evaluating the multiple integrals in the numerator and denominator of equation (23) for all \( N \) is hard in general. However, for large \( N \), one can make a ‘continuum/hydrodynamic’ approximation. This proceeds in two steps. In the first step, we fix a macroscopic density profile \( \rho(x) \) (normalized to unity) and sum over all microscopic configurations of \( x_i \)'s that correspond to this macroscopic density profile. In the second step, we integrate (functional integration) over all possible (normalized to unity) macroscopic density profiles. Following these two steps, one can write \( Z_N(\mu) \) in equation (23) as (up to an overall \( N \)-dependent constant factor) [67]

\[
Z_N(\mu) \approx \int D\rho(x) e^{-N^2 E_\mu[\rho(x)] - N \int dx \rho(x) \ln \rho(x)} \delta \left( \int dx \rho(x) - 1 \right),
\] (26)

where \( E_\mu[\rho(x)] \) is given by

\[
E_\mu[\rho(x)] = \int \left( \frac{x^2}{2} + \mu f(x) \right) \rho(x) dx - \alpha \int \int \rho(x) \rho(y) |x-y| dx dy - \mu s.
\] (27)
The second term $N \int dx \rho(x) \ln \rho(x)$ inside the exponent in the integrand in equation (26) is an entropy term that comes from the first step of coarse-graining mentioned above, i.e. from the sum over all possible microscopic configurations corresponding to a given macroscopic density profile $\rho(x)$. Note that this entropy term scales as $O(N)$, while the energy $\sim N^3$ is much bigger. Hence in the large $N$ limit, we will henceforth neglect the entropy term. The delta-function in equation (26) represents the fact that only the normalized (to unity) macroscopic density profiles are allowed. In fact, it is again convenient to replace this delta-function by its integral representation $\delta(x) = N^3 \int_\mathbb{R} d\mu_0 e^{-\mu_0 N^3 x}$, as was done before in equation (23). Finally the distribution $P(s, N)$ can then be written as

$$P(s, N) \approx \frac{\int D\rho(x) \int d\mu \int d\mu_0 e^{-N^3 S[\rho(x), \mu, \mu_0]} \int D\rho(x) \int d\mu_0 e^{-N^3 S[\rho(x), \mu=0, \mu_0]}}{\int D\rho(x) \int d\mu_0 e^{-N^3 S[\rho(x), \mu=0, \mu_0]}},$$

where the effective action is given by

$$S[\rho(x), \mu, \mu_0] = \int \frac{x^2}{2} \rho(x) dx - \alpha \int \rho(x) \rho(y) |x-y| dx dy + \mu \left( \int f(x) \rho(x) dx - s \right) + \mu_0 \left( \int \rho(x) dx - 1 \right).$$

The next step is to evaluate both the numerator and the denominator in equation (28) by the saddle point method valid for large $N$. We consider them separately, starting with the denominator.

### 2.1.1. Denominator: In this case, the saddle point equations read

$$\frac{\delta S[\rho(x), \mu=0, \mu_0]}{\delta \rho(x)} = 0 \quad \text{and} \quad \frac{\partial S[\rho(x), \mu=0, \mu_0]}{\partial \mu_0} = 0.$$  

The second equation actually gives the normalization condition $\int \rho(x) dx = 1$, while the first equation gives the saddle point density [10, 13]

$$\rho_0^\ast(x) = \frac{1}{4\alpha}, \quad -2\alpha \leq x \leq 2\alpha,$$

which is flat over the finite support $[-2\alpha, +2\alpha]$. Inserting the saddle-point density into (29) with $\mu = 0$, one gets the leading large $N$ behavior of the unconstrained partition function of the jellium model [10, 13]

$$Z_N = Z_N(\mu = 0) \approx e^{4\alpha^2 N^3}.$$  

### 2.1.2. Numerator: In the case of the numerator, there are three saddle-point equations

$$\frac{\delta S[\rho(x), \mu, \mu_0]}{\delta \rho(x)} = 0, \quad \frac{\delta S[\rho(x), \mu, \mu_0]}{\delta \mu} = 0 \quad \text{and} \quad \frac{\delta S[\rho(x), \mu, \mu_0]}{\delta \mu_0} = 0.$$  

The last two equations give the two constraints: $\int f(x) \rho(x) dx = s$ and $\int \rho(x) dx = 1$. We therefore need to find the solution $\rho^\ast_0(x)$ of the first equation that satisfies these two constraints. The first equation in (33), using the action $S[\rho(x), \mu, \mu_0]$ from (29), reads

$$\frac{x^2}{2} - 2\alpha \int \rho^\ast_0(y) |x-y| dy + \mu f(x) + \mu_0 = 0.$$  

$$\int dx \rho(x) = 1.$$  

Finally, the distribution $P(s, N)$ can be written as

$$P(s, N) \approx \frac{\int D\rho(x) \int d\mu \int d\mu_0 e^{-N^3 S[\rho(x), \mu, \mu_0]} \int D\rho(x) \int d\mu_0 e^{-N^3 S[\rho(x), \mu=0, \mu_0]}}{\int D\rho(x) \int d\mu_0 e^{-N^3 S[\rho(x), \mu=0, \mu_0]}},$$

where the effective action is given by

$$S[\rho(x), \mu, \mu_0] = \int \frac{x^2}{2} \rho(x) dx - \alpha \int \rho(x) \rho(y) |x-y| dx dy + \mu \left( \int f(x) \rho(x) dx - s \right) + \mu_0 \left( \int \rho(x) dx - 1 \right).$$

The next step is to evaluate both the numerator and the denominator in equation (28) by the saddle point method valid for large $N$. We consider them separately, starting with the denominator.

### 2.1.1. Denominator: In this case, the saddle point equations read

$$\frac{\delta S[\rho(x), \mu=0, \mu_0]}{\delta \rho(x)} = 0 \quad \text{and} \quad \frac{\delta S[\rho(x), \mu=0, \mu_0]}{\delta \mu_0} = 0.$$  

The second equation actually gives the normalization condition $\int \rho(x) dx = 1$, while the first equation gives the saddle point density [10, 13]

$$\rho_0^\ast(x) = \frac{1}{4\alpha}, \quad -2\alpha \leq x \leq 2\alpha,$$

which is flat over the finite support $[-2\alpha, +2\alpha]$. Inserting the saddle-point density into (29) with $\mu = 0$, one gets the leading large $N$ behavior of the unconstrained partition function of the jellium model [10, 13]

$$Z_N = Z_N(\mu = 0) \approx e^{4\alpha^2 N^3}.$$  

### 2.1.2. Numerator: In the case of the numerator, there are three saddle-point equations

$$\frac{\delta S[\rho(x), \mu, \mu_0]}{\delta \rho(x)} = 0, \quad \frac{\delta S[\rho(x), \mu, \mu_0]}{\delta \mu} = 0 \quad \text{and} \quad \frac{\delta S[\rho(x), \mu, \mu_0]}{\delta \mu_0} = 0.$$  

The last two equations give the two constraints: $\int f(x) \rho(x) dx = s$ and $\int \rho(x) dx = 1$. We therefore need to find the solution $\rho^\ast_0(x)$ of the first equation that satisfies these two constraints. The first equation in (33), using the action $S[\rho(x), \mu, \mu_0]$ from (29), reads

$$\frac{x^2}{2} - 2\alpha \int \rho^\ast_0(y) |x-y| dy + \mu f(x) + \mu_0 = 0.$$  

$$\int dx \rho(x) = 1.$$
Figure 1. A schematic plot of the density $\rho^*_\mu(x)$ in equation (36) for $\alpha = 1, \mu = 1$ and $f(x) = x^4$. The left and the right edges of the support are denoted respectively by $L_1$ and $L_2$.

We assume that the saddle-point density $\rho^*_\mu(x)$ has a single support $x \in [L_1, L_2]$ (to be determined \textit{a posteriori}). This equation holds only for $x$ belonging to the support. Taking a derivative of equation (34) with respect to $x$ gives

$$x - 2\alpha \int \rho^*_\mu(y) \text{sgn}(x - y) dy + \mu f'(x) = 0.$$  \hfill (35)

Taking further one more derivative and using $\frac{d}{dx} \text{sgn}(x) = 2\delta(x)$ one finds

$$1 + \mu f''(x) - 4\alpha \rho^*_\mu(x) = 0 \quad \text{implying} \quad \rho^*_\mu(x) = \frac{1}{4\alpha} (1 + \mu f''(x)) , \quad L_1 \leq x \leq L_2 .$$  \hfill (36)

For a schematic plot of this density in equation (36)—with $f(x) = x^4$ as an example—see figure 1. At this stage, we have four unknown parameters: $L_1$, $L_2$, $\mu_0$ and $\mu$. They can be determined as follows. Setting $x = L_2$ in equation (35) and using $\text{sgn}(L_2 - y) = 1$ for all $y \leq L_2$ and using the normalization $\int_{L_1}^{L_2} \rho^*_\mu(x) dx = 1$, one gets

$$L_2 - 2\alpha + \mu f'(L_2) = 0 .$$  \hfill (37)

Similarly, by setting $x = L_1$ in equation (35) and using $\text{sgn}(L_1 - y) = -1$ for all $y \geq L_1$ gives

$$L_1 + 2\alpha + \mu f'(L_1) = 0 .$$  \hfill (38)

Setting $x = L_2$ in equation (34) and using $|L_2 - y| = L_2 - y$ for all $y \leq L_2$, gives a third independent relation

$$\mu_0 = -\frac{L_2^2}{2} + 2\alpha L_2 - \frac{L_2^2}{4} (L_2^2 - L_1^2) - \frac{\mu}{2} \int_{L_1}^{L_2} y f''(y) dy - \mu f(L_2) .$$  \hfill (39)
Note that we have already used the normalization condition in arriving at the first two equations (37) and (38). But we are still left with one condition

$$\int_{L_1}^{L_2} f(x)\rho_\mu^*(x) \, dx = s.$$  \hfill (40)

Inserting the saddle-point density from equation (36) in this condition and performing the integral gives the desired fourth relation

$$\frac{\mu^2}{4\alpha} \int_{L_1}^{L_2} f(x)\rho''(x) \, dx = s - \frac{1}{4\alpha} \int_{L_1}^{L_2} f(x) \, dx.$$  \hfill (41)

Thus the four unknown parameters \((L_1, L_2, \mu_0, \mu)\) are determined from the four independent non-linear relations (37), (38), (39) and (41). Once they are determined (for a given \(s\)), they characterize the saddle-point density \(\rho^*_\mu(x)\) in equation (36) fully. Let us remark that, when \(\mu = 0\), one recovers the unconstrained density \(\rho^*_0(x) = 1/(4\alpha)\) for \(L_1 \leq x \leq L_2\) with \(L_1 = -2\alpha\) and \(L_2 = +2\alpha\). In this case, we get \(s = \bar{s}\) given in equation (12).

Substituting the saddle-point density \(\rho^*_\mu(x)\) back in equations (28) and (29) gives the leading large \(N\) behavior of the numerator

$$\int D\rho(x) \int d\mu \int d\mu_0 \ e^{-N S[\rho(x), \mu, \mu_0]} \sim e^{-N S[\rho^*_\mu(x), \mu, \mu_0]},$$

where the saddle-point action \(S[\rho^*_\mu(x), \mu, \mu_0]\) depends implicitly on \(s\). Using the behavior of the denominator in equation (32), we then get

$$\mathcal{P}(s, N) \sim e^{-N\Psi(s)} \quad \text{where} \quad \Psi(s) = S[\rho^*_\mu(x), \mu, \mu_0] + \frac{2}{3} \alpha^2.$$  \hfill (43)

By substituting the explicit solution for the saddle-point density from equation (36), we find after straightforward algebra,

$$\Psi(s) = \frac{L_1^3 - L_2^3}{48\alpha} + \frac{L_1^2 - L_2^2}{8} + \frac{L_1^3}{2} - \alpha L_2 + \frac{2}{3} \alpha^2 \mu^2 + \frac{L_1}{4} \int_{L_1}^{L_2} x^2 f''(x) \, dx - 2s + 2f(L_2) + \int_{L_1}^{L_2} xf''(x) \, dx,$$  \hfill (44)

where \(L_1, L_2\) and \(\mu\) are functions of \(s\) and are determined from the relations (37), (38), (39) and (41). This expression gives the exact rate function for the linear statistics \(s = (1/N) \sum f(x_i)\) with arbitrary \(f(x)\). This is one of the main results of this paper. Let us end this subsection with a remark. When \(s\) approaches \(\bar{s}\) given in equation (12), we have seen that \(\mu \to 0\) (see the discussion below equation (41)) and the density \(\rho^*_\mu(x) \to \rho^*_0(x) = 1/(4\alpha)\) for \(L_1 \leq x \leq L_2\) with \(L_1 \to -2\alpha\) and \(L_2 \to 2\alpha\). Consequently, one can easily verify from equation (43) that \(\Psi(s) \to 0\) as \(s \to \bar{s}\). Since \(\Psi(s)\) is a positive convex function, clearly \(s = \bar{s}\) is a minimum of \(\Psi(s)\).

2.2. Extraction of the variance in the large \(N\) limit

In the previous section, we have determined the large deviation form of the PDF \(\mathcal{P}(s, N)\) in equation (43) with the rate function \(\Psi(s)\) given in equation (44). The goal of this subsection is to extract the variance by expanding the rate function \(\Psi(s)\) around its minimum at \(s = \bar{s}\), as explained in the beginning of this section. We have shown at the end of the previous subsection
that, indeed, $\Psi(s)$ has a minimum at $s = \bar{s}$ with $\bar{s}$ given in equation (12). For easy reading, we recall it here

$$
\bar{s} = \frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} f(x) \, dx.
$$

We now expand $\Psi(s)$ around this minimum at $s = \bar{s}$. We have seen that when $s = \bar{s}$, the saddle-point solution is the unconstrained flat density over $[L_1, L_2]$ with $L_1 = -2\alpha$ and $L_2 = +2\alpha$. When $s$ changes slightly from $\bar{s}$, say $s = \bar{s} + \epsilon$, we expect

$$
s = \bar{s} + \epsilon, \quad L_1 = -2\alpha + \delta_1(\epsilon), \quad L_2 = 2\alpha + \delta_2(\epsilon),
$$

where $\delta_1(\epsilon)$ and $\delta_2(\epsilon)$ are small. The idea is first to evaluate $\mu, L_1$ and $L_2$ using the independent relations (37), (38) and (41) for small $\epsilon$ and then substitute this result in equation (44) and evaluate $\Psi(s)$ up to order $O(\epsilon^2)$, where we recall that $\epsilon = s - \bar{s}$. We also recall that $\Psi(s)$ is just the saddle point action up to a constant—see equation (43). It turns out that, in order to extract the expansion of $\Psi(s)$ around $\bar{s}$, it is enough to compute $\mu$ only up to order $O(\epsilon)$. This follows from a thermodynamic identity [57, 58, 68] (for a simple proof, see equations (40) and (41) in [15]) which states that

$$
\frac{\partial \Psi}{\partial s} = \frac{\partial S[\rho_{\mu_0}(x) \mu, \mu_0]}{\partial \mu} = -\mu(s),
$$

where the first equality follows from equation (43). From this relation (47) it is clear that to expand $\Psi(s)$ around $\bar{s}$ up to quadratic order, it is enough to compute $\mu(s)$ only up to linear order in $\epsilon = s - \bar{s}$. We then substitute the expansions (46) in the four independent relations (37), (38), (39) and (41) and expand up to order $\epsilon$. This straightforward expansion leads to

$$
\mu = -\frac{4\alpha \epsilon}{\int_{-2\alpha}^{2\alpha} \left| f'(x) \right|^2 \, dx} + O(\epsilon^2).
$$

In addition, $\delta_1(\epsilon)$ and $\delta_2(\epsilon)$ are given by

$$
\delta_1(\epsilon) = \frac{4\alpha f'(-2\alpha) \epsilon}{\int_{-2\alpha}^{2\alpha} \left| f'(x) \right|^2 \, dx} + O(\epsilon^2)
$$

and

$$
\delta_2(\epsilon) = \frac{4\alpha f'(2\alpha) \epsilon}{\int_{-2\alpha}^{2\alpha} \left| f'(x) \right|^2 \, dx} + O(\epsilon^2).
$$

We then substitute this relation (48) in (47) and integrate over $s$ to obtain

$$
\psi(s) = \frac{1}{2b} (s - \bar{s})^2 + O((s - \bar{s})^3), \quad \text{with} \quad b = \frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} \left| f'(x) \right|^2 \, dx.
$$

Hence, from equation (20), we get

$$
\text{Var}(s) \approx \frac{b}{N^3} \approx \frac{1}{4\alpha N^3} \int_{-2\alpha}^{2\alpha} \left| f'(x) \right|^2 \, dx.
$$

### 2.3. A few examples

Here we work out two simple examples of $f(x)$, namely $f(x) = x$ and $f(x) = x^2$.

- $f(x) = x$: in this case the effective potential in equation (25) reads $V_{\text{eff}}(x) = x^2/2 + \mu x$, which is always confining for all $\mu$ and hence we expect a single support solution for the density,
Figure 2. Left: a plot of the saddle-point density for $f(x) = x$ and two different values of $s = \bar{s} = 0$ and $s = 1$. For $f(x) = x$, the average unperturbed value of $s$ is $\bar{s} = 0$, where the density is flat over $[-2\alpha, 2\alpha]$ (this is shown by the red curve). When $s$ changes from $s = 0$ to $s = 1$, the new density (shown by the green curve) gets shifted by $s = 1$. Right: a plot of the saddle-point density for $f(x) = x^2$ and two different values of $s = \bar{s} = 4\alpha^2/3$ and $s = 1$. For $f(x) = x^2$, the average unperturbed value of $s$ is $\bar{s} = 4\alpha^2/3$, where the density is flat over $[-2\alpha, 2\alpha]$ (this is shown by the red curve). When $s$ changes from $s = \bar{s}$ to $s = 1$, the new density (shown by the green curve), while remaining flat over $[-\sqrt{3}, +\sqrt{3}]$, gets rescaled (unlike in the case $f(x) = x$ where it is just shifted without changing the shape from the unperturbed case). In the figures, we chose $\alpha = 1/10$.

an assumption which is crucial for the derivation of the general formula for the variance of $s$ in equation (52). In this case, the saddle-point density in equation (36) reads

$$\rho^*_\mu(x) = \frac{1}{4\alpha} \mathbb{I}_{[L_1, L_2]}(x),$$

where $\mathbb{I}_{[L_1, L_2]}(x)$ is an indicator function, which is 1 if $x \in [L_1, L_2]$ and 0 otherwise. The parameters $L_1$, $L_2$ and $\mu$ are determined respectively from equations (38), (37) and (41). They simply read

$$L_1 = s - 2\alpha, \quad L_2 = s + 2\alpha, \quad \mu = -s \quad \text{and} \quad \bar{s} = 0.$$  

Thus the original density $\rho_0^*(x) = \frac{1}{4\alpha} \mathbb{I}_{[-2\alpha, 2\alpha]}(x)$ just gets shifted for a finite $s$, or equivalently for finite $\mu$, as seen from equation (53) (see the left panel of figure 2). Since $\mu(s) = -s$ for all $s$, we can integrate the exact relation (47) with the condition that $\Psi(s = \bar{s}) = 0$. This gives the full rate function

$$\Psi(s) = \frac{s^2}{2}.$$  

Hence, clearly, from equation (17), it follows that the variance is $\text{Var}(s) \approx b/N^3$ with $b = 1$, in agreement with our general formula in equation (52).

- $f(x) = x^2$: in this case the effective potential in equation (25) reads $V_{\text{eff}}(x) = (\mu + 1/2)x^2$. Hence, for all $\mu > -1/2$, the potential is confining and we expect to have a single support around $x = 0$. The saddle-point density from equation (36) reads

$$\rho^*_\mu(x) = \frac{1}{2\sqrt{3}s} \mathbb{I}_{[\mu, L_2]}(x),$$

where $L_2 = s + 2\alpha$, $\mu = -s$, and $\bar{s} = 0$. The density is flat over $[-\sqrt{3}, +\sqrt{3}]$, getting rescaled (unlike in the case $f(x) = x$ where it is just shifted without changing the shape from the unperturbed case). In the figures, we chose $\alpha = 1/10$. 


with the parameters $L_1$, $L_2$ and $\mu$ again determined respectively from equations (38), (37) and (41). We get

$$ L_1 = -\sqrt{3}s, \quad L_2 = \sqrt{3}s, \quad \mu = -\frac{1}{2} + \frac{\alpha}{\sqrt{3}s} \quad \text{and} \quad \bar{s} = \frac{4\alpha^2}{3}. \quad (57) $$

Note that, for any $s > 0$, we have $\mu > -1/2$ from equation (57) and hence the effective potential $V_{\text{eff}}(x) = (\mu + 1/2)x^2$ will be always confining for all $s$, leading to a single support solution for any $s$. Thus, as in the $f(x) = x$ case discussed before, the saddle-point density is uniform but unlike the case $f(x) = x$, it is not just a shift of $\rho_0^s(x)$ but also the height and the width gets modified (see the right panel of figure 2). Using the exact $\mu(s)$ from equation (57) in the exact relation (47), and integrating with respect to $s$ using $\Psi(s = \bar{s}) = 0$, one gets the full rate function

$$ \Psi(s) = \frac{s^2}{2} - \frac{2\alpha}{\sqrt{3}}s + \frac{2\alpha^2}{3}. \quad (58) $$

A plot of this rate function is shown in figure 3. By expanding around $s = \bar{s} = 4\alpha^2/3$ up to quadratic order, one gets $\text{Var}(s) = b/N^3$ with $b = 16\alpha^2/3$, in agreement with our general formula in equation (52).

Note that in the derivation of the general formula (52), we have assumed that the effective potential $V_{\text{eff}}(x) = x^2/2 + \mu(x)$ is confining, so that one has a single support for the saddle-point density and also the fact that $f(x)$ is a smooth function such that its first derivative $f'(x)$ exists and the integral in equation (52) is finite. In the two examples discussed above, namely $f(x) = x$ and $f(x) = x^2$, both conditions are met. In section 4, we will demonstrate examples where either one of the two conditions breaks down and yet we will show that the formula (52) will still be valid.
3. An alternative derivation of the formula for the variance

In this section, we provide an alternative derivation of the formula for the variance given in equation (13). Following [35] for the log-gas case, we focus on the generating function of $s$, namely

$$G(\lambda) = \left< e^{-\lambda s} \right> = \left< e^{-\frac{\lambda}{1} \sum_{i=1}^{N} f(x_i)} \right> = \int_{-\infty}^{\infty} ds \, P(s, N) e^{-\lambda s}, \quad (59)$$

where we recall that $\left< \ldots \right>$ denotes an average over the Gibbs–Boltzmann measure in equation (2) and $P(s, N)$ is the PDF of $s$ given in equation (14). The starting point of the analysis presented here is the following identity

$$\frac{\partial}{\partial \lambda} \ln G(\lambda) = -\frac{1}{N} \left< \sum_{i=1}^{N} f(x_i) \right> \lambda, \quad (60)$$

where $\left< \ldots \right>_{\lambda}$ denotes an average with respect to the modified weight

$$\tilde{P}_\lambda(x_1, x_2, \ldots, x_N) = \frac{1}{Z_N(\lambda)} e^{-E[\{x_i\}] - \frac{\lambda}{1} \sum_{i=1}^{N} f(x_i)}, \quad (61)$$

where $Z_N(\lambda)$ is a normalization constant. Hence, in the large $N$ limit the relation (59) becomes, to leading order for large $N$

$$\frac{\partial}{\partial \lambda} \ln G(\lambda) \approx -\int_{-\infty}^{\infty} \tilde{\rho}_\lambda(x) f(x) \, dx, \quad (62)$$

where $\tilde{\rho}_\lambda(x)$ is the equilibrium density associated to the joint PDF in (61). By comparing equations (61) and (24), it is easy to see that, to leading order for large $N$,

$$\tilde{\rho}_\lambda(x) \approx \rho^{\ast}_{\mu=\lambda/N^d}(x), \quad (63)$$

where $\rho^{\ast}_{\mu}(x)$ is given in equation (36). Hence $\tilde{\rho}_\lambda(x)$ reads, to leading order for large $N$,

$$\tilde{\rho}_\lambda(x) \approx \frac{1}{4\alpha} \left( 1 + \frac{\lambda}{N^d} f''(x) \right), \quad \tilde{L}_1(\lambda) \leq x \leq \tilde{L}_2(\lambda), \quad (64)$$

where $\tilde{L}_1(\lambda)$ and $\tilde{L}_2(\lambda)$ are given respectively by equations (38) and (37) with the substitution $\mu = \lambda/N^d$, i.e.

$$\tilde{L}_1(\lambda) + 2\alpha + \frac{\lambda}{N^d} f''(\tilde{L}_1(\lambda)) = 0 \quad (65)$$

$$\tilde{L}_2(\lambda) - 2\alpha + \frac{\lambda}{N^d} f''(\tilde{L}_2(\lambda)) = 0. \quad (66)$$

To compute the variance, we need to compute the small $\lambda$ expansion of $\ln G(\lambda)$ in (62) up to order $O(\lambda^2)$. For this purpose, it turns out that we need to expand $\tilde{L}_1(\lambda)$ and $\tilde{L}_2(\lambda)$ up to order $O(\lambda)$ only. Expanding equations (65) and (66) for small $\lambda$ one finds

$$\tilde{L}_1(\lambda) = -2\alpha - \frac{\lambda}{N^d} f''(2\alpha) + O(\lambda^2), \quad \tilde{L}_2(\lambda) = 2\alpha - \frac{\lambda}{N^d} f''(2\alpha) + O(\lambda^2). \quad (67)$$

Inserting equation (64) in equation (62) one obtains

$$\frac{\partial}{\partial \lambda} \ln G(\lambda) \approx -\frac{1}{4\alpha} \int_{-\tilde{L}_1}^{\tilde{L}_2} f(x) \, dx - \frac{\lambda}{4\alpha N^d} \int_{\tilde{L}_1}^{\tilde{L}_2} f(x) f''(x) \, dx. \quad (68)$$
Performing an integration by parts in the second integral in (68) and using the small $\lambda$ expansion of $L_1(\lambda)$ and $L_2(\lambda)$ in (67), one obtains after straightforward algebra

$$\frac{\partial}{\partial \lambda} \text{ln} G(\lambda) = -\frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} f(x) \, dx + \frac{\lambda}{4\alpha N^2} \int_{-2\alpha}^{2\alpha} dx |f'(x)|^2 + O(\lambda^2).$$

(69)

Therefore, integrating over $\lambda$, using $G(\lambda = 0) = 1$ one finds

$$\text{ln} G(\lambda) = -\frac{\lambda}{4\alpha} \int_{-2\alpha}^{2\alpha} f(x) \, dx + \frac{\lambda^2}{8\alpha N^2} \int_{-2\alpha}^{2\alpha} dx |f'(x)|^2 + O(\lambda^3).$$

(70)

From this expression, one can immediately read off the mean $\bar{s}$ and the variance $\text{Var}(s)$ of $s$ as

$$\bar{s} = \frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} f(x) \, dx, \quad \text{Var}(s) = \frac{1}{4\alpha N^2} \int_{-2\alpha}^{2\alpha} dx |f'(x)|^2,$$

(71)

which indeed coincide with the results in equations (12) and (13) obtained by a different method.

The exercise above can actually be repeated for arbitrary confining potential $V(x)$, not necessarily harmonic. In this case, the equilibrium density in equation (36) gets replaced by

$$\tilde{\rho}_s(x) = \frac{1}{4\alpha} [V(x) + \mu f'(x)],$$

(72)

which is supported over the interval $[L_1, L_2]$. Note that $L_1$ and $L_2$ do depend on $V(x)$. One can then repeat the steps above and finds that the formula for the variance (71) actually holds for a general $V(x)$ and reads, up to an overall $N$-dependent constant,

$$\text{Var}(s) \propto \int_{L_1}^{L_2} dx |f'(x)|^2.$$ (73)

Thus the dependence of the variance on $V(x)$ enters only through the support edges $L_1$ and $L_2$, but not explicitly.

4. Validity of the formula for the variance

As mentioned above, one of the crucial assumptions leading to the derivation of the formula (52) is the following. Once the chemical potential is switched on, it changes the effective potential of the jellium to $V_{\text{eff}}(x) = x^2/2 + \mu f(x)$ (see equation (25)). We assume $V_{\text{eff}}(x)$ is still confining and the saddle-point density still has a single support. In addition $f'(x)$ must exist and the integral in equation (52) should be finite. Below, we discuss two examples where one of the two conditions breaks down. For instance, when $f(x) = x^3$, the effective potential is always non confining, except of course for $\mu = 0$. In some other examples, $f'(x)$ may be singular, such as $f(x) = |x|$, which also leads to two disjoint supports for the density for $\mu < 0$. However, in both cases, we will show that the general formula (52) still gives the correct answer for the variance of $s$.

4.1. The special case $f(x) = x^3$

In this case, the effective potential $V_{\text{eff}}(x) = x^2/2 + \mu x^3$ is non-confining for any $\mu$. For example, for $\mu > 0$, it diverges negatively as $x \to -\infty$. A plot of $V_{\text{eff}}(x)$ in this case is given in figure 4 for $\mu > 0$. 


Figure 4. Plot of the effective potential $V_{\text{eff}}(x) = x^2/2 + \mu x^3$ with $\mu = 1/10$. The maximum on the left of the origin occurs at $x = -1/(3\mu)$, while the height of the maximum is $1/(54\mu^2)$.

For $\mu > 0$, the effective potential $V_{\text{eff}}(x)$ has always a minimum at $x = 0$ and a maximum at $x = -1/(3\mu)$. The height of the maximum is given by

$$V_{\text{eff}} \left( x = -\frac{1}{3\mu} \right) = \frac{1}{54\mu^2}.$$  \hfill (74)

For small $\mu$, the barrier height is thus very large and, hence, the minimum at $x = 0$ is deep. Hence, for small $\mu$, we expect to have a single support. For large positive $\mu$, the barrier height will decrease and the minimum at $x = 0$ will be shallow. In this case, some charges may split from the support around $x = 0$ and go over the barrier to very large negative $x$ and our assumption about the single support will no longer be valid. However, for the computation of the variance of $s$, we only need $\mu$ very small. Hence, the single support assumption will still be correct and hence we expect that our formula in equation (52) is still valid for large $N$. Indeed, assuming a single support solution as in equation (36), one can in principle compute the parameters $L_1, L_2$ and $\mu$ and check that the rate function $\Psi(s)$ in equation (44) is again quadratic around $s = \bar{s}$ with the variance given by equation (52) with $b$ given by

$$b = \frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} (3x^2)^2 dx = \frac{144}{5} \alpha^4.$$  \hfill (75)

Thus, while we are able to compute the variance, from the expansion of $\Psi(s)$ in equation (44) around $s = \bar{s}$, the expression of the rate function $\Psi(s)$—obtained by assuming a single support—is not expected to be valid for all $s$. Computing the full rate function $\Psi(s)$ for all $s$ is an interesting challenge left for future investigations. Note that, here, we just discussed the specific example of $f(x) = x^3$, but a similar discussion will hold for any odd function $f(x)$ that diverges faster than $x^2$, for instance $f(x) = x^5, x^7, \ldots$. In all such cases, we expect the formula for the variance in equation (52) to be still valid, since it arises only from the small $\mu$ expansion. We have verified the validity of our prediction for the variance in (75) for $f(x) = x^3$ by Monte-Carlo simulations, as shown in figure 5 where we plot the difference...
Figure 5. A plot of $|b - b_{MC}|$ vs $N$ for $\alpha = 0.1$ (left panel) and $\alpha = 0.01$ (middle panel) for $f(x) = x^3$. The right panel shows the same data as in the left and middle panels but in a log-log plot. Here $b = \frac{244}{5} \alpha^4$ is the theoretical prediction for $N^3 \text{Var}(s)$ for $f(x) = x^3$ and $b_{MC}$ is the Monte-Carlo value of $N^3 \text{Var}(s)$.

Figure 6. Effective potentials $V_{\text{eff}}(x) = x^2/2 + \mu |x|$ vs $x$ for three values of $\mu$. From left to right the values of $\mu$ are $-1$, $0$, and $1$. If $\mu < 0$ the potential has two minima and for $\mu > 0$ just one. The transition point is at $\mu = 0$.

between the theoretical value of $b = \frac{244}{5} \alpha^4$ (see equation (75)) and the Monte-Carlo value $b_{MC} = N^3 \text{Var}(s) \bigg|_{MC}$ as a function of increasing $N$ and for two different values of $\alpha$. We find that it decreases to zero as $N \to \infty$, thus verifying the theoretical prediction.

4.2. The special case $f(x) = |x|

In this case, the effective potential in equation (25) felt by the charges in the jellium model reads

$$V_{\text{eff}}(x) = \frac{1}{2} x^2 + \mu |x|.$$  \hfill (76)

Clearly, for $\mu > 0$, $V_{\text{eff}}(x)$ has a single minimum at $x = 0$. In this case, one would expect that the saddle-point density $\rho_{s\mu}^*(x)$ will be supported over a single interval around this minimum. In contrast, when $\mu < 0$, the effective potential has two minima located at $\pm |\mu|$ (see figure 6).

Hence, in this case, one expects that the saddle-point density will have two disjoint supports, one around each minimum. Below, we calculate the saddle-point density explicitly, confirming
this scenario. Before proceeding, we note that the average value of \( s \) (corresponding to \( \mu = 0 \)) is given by (using \( f(x) = |x| \) in equation (12))

\[
\bar{s} = \alpha . \quad (77)
\]

Our starting point is the action \( S[\rho(x), \mu, \mu_0] \) in equation (29) with \( f(x) = |x| \). The saddle point equation for the density then reads, from equation (35)

\[
x - 2\alpha \int_0^\infty \rho_\mu^*(y) \text{sgn}(x - y) dy + \mu \text{sgn}(x) = 0 . \quad (78)
\]

Note that this equation holds only at points inside the support of \( \rho_\mu^*(x) \). We now consider the two cases \( \mu > 0 \) and \( \mu < 0 \) separately.

**4.2.1. The case \( \mu > 0 \):** In this case, we expect that there is a single support over \([-\ell, +\ell]\) where \( \ell \) remains to be determined. Note that, since \( V_{\text{eff}}(x) \) is symmetric around \( x = 0 \), we expect the support to be also symmetric around \( x = 0 \), and hence we chose it to be \([-\ell, +\ell]\).

Taking one more derivative of equation (78) with respect to \( x \), and using \( \frac{d}{dx} \text{sgn}(x) = 2\delta(x) \), we get

\[
\rho_\mu^*(x) = \frac{1}{4\alpha} + \frac{\mu}{2\alpha} \delta(x) . \quad (79)
\]

Thus the density has a flat profile with a spike (delta-function) at its center. This is confirmed in our Monte-Carlo simulations (see the middle panel figure 7). The unknown parameters are \( \ell \) and \( \mu \). The normalization condition \( \int_{-\ell}^{+\ell} \rho_\mu^*(x) dx = 1 \) gives the relation

\[
\ell + \mu = 2\alpha , \quad (80)
\]

and the other condition \( s = \int_{-\ell}^{+\ell} \rho_\mu^*(x)|x| dx \) gives

\[
s = \frac{1}{4\alpha}(2\alpha - \mu)^2 . \quad (81)
\]

By inverting this relation, one gets \( \mu = 2\alpha \pm \sqrt{4\alpha s} \). From equation (80), it is clear that \( \mu < 2\alpha \) in order that \( \ell > 0 \). Hence we choose the negative root of the quadratic equation for \( \mu \) and set

\[
\mu = 2\alpha - \sqrt{4\alpha s} \quad \text{and} \quad \ell = \sqrt{4\alpha s} . \quad (82)
\]

Since \( \mu > 0 \), it follows that this single support solution in equation (78) is valid only for \( s < \alpha = \bar{s} \), where we used equation (77) for the last equality.

**4.2.2. The case \( \mu < 0 \):** In this case, as mentioned before, we expect a two-support solution. We need to parametrize the solution and we expect the two supports to be placed symmetrically around the origin. Let us assume that the supports are \([-L, -a] \cup [a, L]\) with \( 0 \leq a \leq L\)—see figure 7. Taking a derivative with respect to \( x \) in the general saddle-point equation (78), we see that the density in the bulk (i.e. away from the edges) is always flat and is given by

\[
\rho_\mu^*(x) = \frac{1}{4\alpha} \quad \text{for} \quad -L < x < -a \quad \text{or} \quad a < x < L . \quad (83)
\]

Note that equation (78) holds for all points belonging to both supports. Choosing \( x \in [a, L] \) (right support), and using (83), one gets, from equation (78) the relation

\[
\mu = -a . \quad (84)
\]
Figure 7. Left: Plot of the Monte-Carlo simulations for the average density (red dots) for $\alpha = 1/10$ and $N = 1000$ for $s = 0.6$, compared with the saddle-point prediction in equation (83). In this case, since $s = 0.6 > \bar{s} = \alpha = 1/10$, the density has two disjoint supports with parameters $L = s + \alpha = 0.7$ and $a = s - \alpha = 0.5$. Right: Plot of the Monte-Carlo simulations for the average density (red dots) for $\alpha = 1/10$ and $N = 1000$ for $s = 0.025$, compared with the saddle-point prediction in equation (79). In this case since $s = 0.025 < \bar{s} = 1/10$, the density has a single support with an additional delta peak at $x = 0$.

The normalization condition $2 \int_{-a}^{a} \rho_\mu^*(x) \, dx = 1$ gives the relation

$$L - a = 2\alpha. \quad (85)$$

Furthermore, the condition $2 \int_{-a}^{a} \mid x \mid \rho_\mu^*(x) \, dx = s$ gives another relation

$$L^2 - a^2 = 4\alpha s. \quad (86)$$

Thus we have three equations (84)–(86) for three unknowns $\mu$, $a$ and $L$ for a fixed $s$. Solving them, we get the three parameters

$$L = s + \alpha, \quad a = s - \alpha \quad \text{and} \quad \mu = \alpha - s. \quad (87)$$

Since $\mu < 0$, this two-support solution holds for $s > \alpha = \bar{s}$.

Thus summarizing the two cases $\mu > 0$ (or equivalently $s < \bar{s}$) and $\mu < 0$ (or equivalently $s > \bar{s}$), we find that, as $s$ approaches $\bar{s}$ from above, the gap $2a = 2(s - \alpha) = 2(s - \bar{s})$ between the two supports shrinks linearly and vanishes exactly at $s = \bar{s}$, where the two supports merge with each other. When $s$ reduces further below $\bar{s}$, the delta-spike at $s = 0$ appears in the single support solution. Thus there is indeed a phase transition that occurs at $s = \bar{s}$ due to the vanishing of the gap between the two supports. This actually shows up in a singularity of the rate function $\Psi(s)$ at $s = \bar{s}$, as demonstrated below.

In order to compute the rate function $\Psi(s)$ from the identity (47), we need to first determine $\mu(s)$ as a function of $s$ for all $s$. Indeed, this is given from the analysis above in equations (82) and (87). We thus get

$$\mu(s) = \begin{cases} 
2\alpha - \sqrt{4\alpha s}, & s < \bar{s} = \alpha \\
\alpha - s, & s > \bar{s} = \alpha 
\end{cases} \quad (88)$$
Figure 8. A plot of $|b - b_{MC}|$ vs $N$ for $\alpha = 0.1$ (left panel) and $\alpha = 0.01$ (middle panel) for $f(x) = |x|$. The right panel shows the same data as in the left and middle panels but in a log-log plot. Here $b = 1$ is the theoretical prediction for $N^{3}\text{Var}(s)$ for $f(x) = |x|$ and $b_{MC}$ is the Monte-Carlo value of $N^{3}\text{Var}(s)$.

Substituting this in the identity (47), integrating with respect to $s$, and using $\Psi(s = \bar{s}) = 0$, we get

$$
\Psi(s) = \begin{cases} 
-2\alpha s + \frac{4}{3}\sqrt{\alpha s^3/2} + \frac{2}{3}\alpha^2, & s < \alpha = \bar{s}, \\
\frac{1}{2}(s - \alpha)^2, & s > \alpha = \bar{s}.
\end{cases}
$$

One can easily check that $\Psi(s), \Psi'(s)$ as well as $\Psi''(s)$ are all continuous at $s = \bar{s} = \alpha$. However, the third derivative is discontinuous and is given by

$$
\Psi'''(s) = \begin{cases} 
\frac{1}{2\alpha}, & s \to \bar{s}^-,
0, & s \to \bar{s}^+.
\end{cases}
$$

Thus there is a third order phase transition in $\Psi(s)$ at $s = \bar{s}$. Such phase transitions have shown up in many other examples involving Coulomb and log-gases (for a review see [23]). Since $\Psi''(s)$ is continuous at $s = \bar{s}$, the expansion of $\Psi(s)$ around $s = \bar{s}$ is quadratic in leading order $\Psi(s) \approx \frac{1}{2}(s - \bar{s})^2$, indicating from equation (17) that

$$
\text{Var}(s) \approx \frac{b}{N^3} \text{ with } b = 1.
$$

Thus this is completely consistent with our general formula in equation (52) which also predicts $b = 1$ for $f(x) = |x|$. We have verified the validity of our prediction for the variance in (91) for $f(x) = |x|$ by Monte-Carlo simulations, as shown in figure 8 where we plot the difference between the theoretical value of $b = 1$ (see equation (91)) and the Monte-Carlo value $b_{MC} = N^{3}\text{Var}(s)^{MC}$ as a function of increasing $N$ and for two different values of $\alpha$. We find that it decreases to zero as $N \to \infty$, thus verifying the theoretical prediction.

5. Conclusion

To summarize, we have considered the jellium model of $N$ particles in one-dimension with energy

$$
E([x_i]) = \frac{N^2}{2} \sum_{i=1}^{N} x_i^2 - \alpha N \sum_{i \neq j} |x_i - x_j|,
$$

$$
91.
$$
and studied the statistics of \( s = \left( \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right) \) in the Gibbs–Boltzmann state at any temperature of order \( \Theta(1) \). The main result in this paper is to derive a nice and compact general formula for the variance of \( s \) in the limit of a large \( N \)

\[
\text{Var}(s) = \frac{1}{N^{3\alpha}} \int_{-2\alpha}^{2\alpha} (f'(x))^2 \, dx. \tag{93}
\]

One expects this formula to be valid for a wide class of \( f(x) \)'s for which the integral in equation (93) is convergent. We have discussed with several examples, both analytically and numerically, the precise criteria behind the validity of this formula. We have provided two different derivations of this formula in this paper.

In addition to computing the formula for the variance for general \( f(x) \), we have also shown that, for large \( N \), the full PDF of \( s \) exhibits a large deviation form \( P(s,N) \sim e^{-N^\Psi(s)} \) and we have computed the rate function \( \Psi(s) \) for general \( f(x) \) in equation (44) and provided more explicit forms in several examples such as \( f(x) = x, x^2 \) and \( f(x) = |x| \). In all cases, \( \Psi(s) \) exhibits a quadratic behavior around its minimum at \( s = \bar{s} = \left( \frac{1}{4\alpha} \int_{-2\alpha}^{2\alpha} f(x) \, dx \right) \). This quadratic form reads \( \Psi(s) \approx (s - \bar{s})^2/(2\text{Var}(s)) \), where \( \text{Var}(s) \) is given in equation (93).

It would be interesting to extend the technique presented in this paper to compute the covariance of two different linear statistics \( s_1 = \left( \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right) \) and \( s_2 = \left( \frac{1}{N} \sum_{i=1}^{N} g(x_i) \right) \) where \( f(x) \) and \( g(x) \) are two different but arbitrary functions. In the log-gas case, this was computed in [40] and it is natural to compute this covariance of two linear statistics for the jellium model.

**Data availability statement**

No new data were created or analysed in this study.

**Acknowledgment**

We thank P Bourgade for useful discussions.

**ORCID iD**

Grégory Schehr https://orcid.org/0000-0001-6648-5213

**References**

[1] Riesz M 1938 Riemann Liouville integrals and potentials Acta Sci. Math. 9 1

[2] For a recent review, see Lewin M 2022 Coulomb and Riesz gases: the known and the unknown J. Math. Phys. 63 061101

[3] Leblé T and Serfaty S 2017 Large deviation principle for empirical fields of Log and Riesz gases Invent. Math. 210 645

[4] Hardin D P, Leblé T, Saff E B and Serfaty S 2018 Large deviation principles for hypersingular Riesz gases Constr. Approx. 48 61

[5] Agarwal S, Dhar A, Kulkarni M, Kundu A, Majumdar S N, Mukamel D and Schehr G 2019 Harmonically confined particles with long-range repulsive interactions Phys. Rev. Lett. 123 100603

[6] Ketepalli J, Kulkarni M, Kundu A, Majumdar S N, Mukamel D and Schehr G 2021 Harmonically confined long-ranged interacting gas in the presence of a hard wall J. Stat. Mech. Theory Exp. 2021 103209

[7] Ketepalli J, Kulkarni M, Kundu A, Majumdar S N, Mukamel D and Schehr G 2022 Edge fluctuations and third-order phase transition in harmonically confined long-range systems J. Stat. Mech. Theory Exp. 2022 033203
[8] Santra S, Ketepalli J, Agrawal S, Dhar A, Kulkarni M and Kundu A 2022 Gap statistics for confined particles with power-law interactions Phys. Rev. Lett. 128 170603
[9] Lenard A 1961 Exact statistical mechanics of a one-dimensional system with Coulomb forces J. Math. Phys. 2 682
[10] Prager S 1962 The one-dimensional plasma Adv. Chem. Phys. 4 201
[11] Baxter R J 1963 Statistical mechanics of a one-dimensional Coulomb system with a uniform charge background Math. Proc. Camb. Philos. Soc. 59 779
[12] Aizenman M and Martin P A 1980 Structure of Gibbs states of one dimensional Coulomb systems Commun. Math. Phys. 78 99
[13] Dhar A, Kundu A, Majumdar S N, Sabhapandit S and Schehr G 2017 Exact extremal statistics in the classical 1D Coulomb gas Phys. Rev. Lett. 119 060601
[14] Dhar A, Kundu A, Majumdar S N, Sabhapandit S and Schehr G 2018 Extreme statistics and index distribution in the classical 1d Coulomb gas J. Phys. A: Math. Theor. 51 295001
[15] Flack A, Majumdar S N and Schehr G 2021 Truncated linear statistics in the one dimensional one-component plasma J. Phys. A: Math. Theor. 54 435002
[16] Flack A, Majumdar S N and Schehr G 2022 Gap probability and full counting statistics in the one dimensional one-component plasma J. Stat. Mech. Theory Exp. 5 053211
[17] Chafai D, Garcia-Zelada D and Jung P 2022 At the edge of a one-dimensional jellium Bernoulli 28 1784–809
[18] Dyson F J 1962 Statistical theory of the energy levels of complex systems. I. J. Math. Phys. 3 140
[19] Mehta M L and Dyson F J 1963 Statistical Theory of the Energy Levels of Complex Systems J. Math. Phys. 4 713
[20] Mehta M L 1991 Random Matrices and the Statistical Theory of Spectra (New York: Academic Press)
[21] Forrester P J 2010 Log-Gases and Random Matrices (Princeton, NJ: Princeton University Press)
[22] Saff E B and Totik V 2013 Logarithmic Potentials with External Fields vol 316 (Berlin: Springer Science & Business Media)
[23] Majumdar S N and Schehr G 2014 Top eigenvalue of a random matrix: large deviations and third order phase transition J. Stat. Mech. Theory Exp. 2014 01012
[24] Calogero F 1971 Solution of the one-dimensional N-body problems with quadratic and/or inversely quadratic pair potentials J. Math. Phys. 12 419
[25] Calogero F 1975 Exactly solvable one-dimensional many-body problems Lett. Nuovo Cimento 13 411
[26] Sutherland B 1971 Exact results for a quantum many-body problem in one dimension Phys. Rev. A 4 2019
[27] Moser J 1975 Three integrable Hamiltonian systems connected with isospectral deformations Adv. Math. 16 197
[28] Agarwal S, Kulkarni M and Dhar A 2019 Some connections between the classical Calogero–Moser model and the log-gas J. Stat. Phys. 176 1463
[29] Politzer H D 1989 Random-matrix description of the distribution of mesoscopic conductance Phys. Rev. B 40 11917
[30] Beenakker C W J 1993 Universality in the random-matrix theory of quantum transport Phys. Rev. Lett. 70 1155
[31] Beenakker C W J 1993 Random-matrix theory of mesoscopic fluctuations in conductors and superconductors Phys. Rev. B 47 15763
[32] Basor E L and Tracy C A 1993 Variance calculations and the Bessel kernel J. Stat. Phys. 73 415
[33] Chen Y and Manning S M 1994 Distribution of linear statistics in random matrix models (metallic conductance fluctuations) J. Phys.: Condens. Matter 6 3039
[34] Baker T H and Forrester P J 1997 Finite-N fluctuation formulas for random matrices J. Stat. Phys. 88 1371
[35] Johansson K 1998 On fluctuations of eigenvalues of random Hermitian matrices Duk. Math. J. 91 151
[36] Soshnikov A 2000 The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities Ann. Probab. 28 1353
[37] Pastur L 2006 Limiting laws of linear eigenvalue statistics for Hermitian matrix models J. Math. Phys. 47 103303
[38] Sommers H J, Wieczorek W and Savin D V 2007 Statistics of conductance and shot noise power for chaotic cavities Acta Phys. Pol. A 112 691
[39] Lytova A and Pastur L 2009 Central limit theorem for linear eigenvalue statistics of random matrices with independent entries Ann. Probab. 37 1778
[40] Cunden F D and Vivo P 2014 Universal covariance formula for linear statistics on random matrices Phys. Rev. Lett. 113 070202
[41] Grabsch A, Majumdar S N and Texier C 2017 Truncated linear statistics associated with the top eigenvalues of random matrices J. Stat. Phys. 167 234
[42] Grabsch A, Majumdar S N and Texier C 2017 Truncated linear statistics associated with the eigenvalues of random matrices II. Partial sums over proper time delays for chaotic quantum dots J. Stat. Phys. 167 1452
[43] Grabsch A 2021 General truncated linear statistics for the top eigenvalues of random matrices J. Phys. A: Math. Theor. 55 124001
[44] Texier C and Majumdar S N 2013 Wigner time-delay distribution in chaotic cavities and freezing transition Phys. Rev. Lett. 110 250602
[45] Costin O and Lebowitz J L 1995 Gaussian fluctuation in random matrices Phys. Rev. Lett. 75 69
[46] Fogler M M and Shklovskii B I 1995 Probability of an eigenvalue number fluctuation in an interval of a random matrix spectrum Phys. Rev. Lett. 74 3312
[47] Majumdar S N, Nadal C, Scardicchio A and Vivo P 2009 Index distribution of Gaussian random matrices Phys. Rev. Lett. 103 220603
[48] Majumdar S N, Nadal C, Scardicchio A and Vivo P 2011 How many eigenvalues of a Gaussian random matrix are positive? Phys. Rev. E 83 041105
[49] Marino R, Majumdar S N, Schehr G and Vivo P 2014 Phase transitions and edge scaling of number variance in Gaussian random matrices Phys. Rev. Lett. 112 254101
[50] Smith N R, Le Doussal P, Majumdar S N and Schehr G 2021 Full counting statistics for interacting trapped fermions SciPost Phys. 11 110
[51] Osipov V A and Kanzieper E 2008 Integrable theory of quantum transport in chaotic cavities Phys. Rev. Lett. 101 176804
[52] Vivo P, Majumdar S N and Bohigas O 2008 Distributions of conductance and shot noise and associated phase transitions Phys. Rev. Lett. 101 216809
[53] Khoruzhenko B A, Savin D V and Sommers H J 2009 Systematic approach to statistics of conductance and shot-noise in chaotic cavities Phys. Rev. B 80 125301
[54] Osipov V A and Kanzieper E 2009 Statistics of thermal to shot noise crossover in chaotic cavities J. Phys. A: Math. Theor. 42 475103
[55] Vivo P, Majumdar S N and Bohigas O 2010 Probability distributions of linear statistics in chaotic cavities and associated phase transitions Phys. Rev. B 81 104202
[56] Damle K, Majumdar S N, Tripathi V and Vivo P 2011 Phase transitions in the distribution of the Andreev conductance of superconductor-metal junctions with multiple transverse modes Phys. Rev. Lett. 107 177206
[57] Grabsch A and Texier C 2015 Capacitance and charge relaxation resistance of chaotic cavities—joint distribution of two linear statistics in the Laguerre ensemble of random matrices Europhys. Lett. 109 50004
[58] Grabsch A and Texier C 2016 Distribution of spectral linear statistics on random matrices beyond the large deviation function—Wigner time delay in multichannel disordered wires J. Phys. A Math. Theor. 49 465002
[59] Facchi P, Marzolino U, Parisi G, Pascazio S and Scardicchio A 2008 Phase transitions of bipartite entanglement Phys. Rev. Lett. 101 050502
[60] Nadal C, Majumdar S N and Vergassola M 2011 Statistical distribution of quantum entanglement for a random bipartite state J. Stat. Phys. 142 403
[61] Nadal C, Majumdar S N and Vergassola M 2010 Phase transitions in the distribution of bipartite entanglement of a random pure state Phys. Rev. Lett. 104 110501
[62] Forrester P J 1999 Fluctuation formula for complex random matrices J. Phys. A: Math. Gen. 32 159
[63] Rider B and Virag B 2007 The noise in the circular law and the Gaussian free field Int. Math. Res. Not. 2007 rnm006
[64] Ameur Y, Hedenmalm H and Makarov N 2011 Fluctuations of eigenvalues of random normal matrices Duke Math. J. 159 31
[65] Leblé T and Serfati S 2018 Fluctuations of two dimensional Coulomb gases Geom. Funct. Anal. 28 443
[66] Armstrong S and Serfaty S 2021 Local laws and rigidity for Coulomb gases at any temperature Ann. Probab. 49 46

[67] Dean D S and Majumdar S N 2008 Extreme value statistics of eigenvalues of Gaussian random matrices Phys. Rev. E 77 041108

[68] Cunden F D, Facchi P and Vivo P 2016 A shortcut through the Coulomb gas method for spectral linear statistics on random matrices J. Phys. A Math. Theor. 49 135202