ON THE REGULARITY OF THE CONDITIONAL DISTRIBUTION OF THE SAMPLE MEAN

VICTOR CHULAEVSKY

Abstract. We show that the hypothesis of regularity of the conditional distribution of the empiric average of a finite sample of IID random variables, given all the sample “fluctuations”, which appeared in our earlier manuscript [4] in the context of the eigenvalue concentration analysis for multi-particle random operators, is satisfied for a class of probability distributions with sufficiently smooth probability density. It extends the well-known property of Gaussian IID samples.

1. Introduction

In a few talks given at workshops on disordered quantum systems, I have mentioned a simple result of the elementary probability theory which has an interesting application to the multi-particle Anderson localization theory. It is difficult to say if the result itself is original; personally, I would be glad to learn that it is not, and to provide some bibliographical reference, for it is indeed hard to believe that the elementary probabilistic problem in question was never addressed, for example, in statistics. However, I am unaware of any such published (or folkloric) result.

The goal of this short note is to fill this gap and provide an elementary proof of the regularity (with high probability) of the conditional sample mean of a finite sample of uniformly distributed IID random variables, given the sigma-algebra of “fluctuations”.

This text is an improvement of the previous version (25.04.2013) in two ways:

- we consider a larger class of probability distributions, including those with piecewise-constant probability density, on the intervals of arbitrary length \( \ell \); while such a generalization is quite straightforward, it renders more convenient references to the main result of this paper; moreover, we extend the main result to a class of smooth probability densities;
- the probabilistic estimates are made slightly stronger; again, this is a minor improvement, but it may prove useful in the applications.

2. Prelude: Gaussian IID samples

Consider a sample of \( N \) IID (independent and identically distributed) random variables with Gaussian distribution \( \mathcal{N}(0, 1) \), and introduce the sample mean \( \xi = \xi_N \) and the “fluctuations” \( \eta_i \) around the mean:

\[
\xi_N = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad \eta_i = X_i - \xi_N, \quad i = 1, \ldots, N.
\]
It is well-known from elementary courses of the probability theory that $\xi_N$ is independent from the sigma-algebra $\mathcal{F}_N$ generated by $\{\eta_1, \ldots, \eta_n\}$ (the latter are linearly dependent, and have rank $N - 1$). To see this, it suffices to note that $\eta_i$ are all orthogonal to $\xi_N$ with respect to the standard scalar product in the linear space formed by $X_1, \ldots, X_N$ given by

$$\langle Y, Z \rangle := \mathbb{E}[Y Z]$$

where $Y$ and $Z$ are real linear combinations of $X_1, \ldots, X_N$ (recall: $\mathbb{E}[X_i] = 0$).

Therefore, the conditional probability distribution of $\xi_N$ given $\mathcal{F}_N$ coincides with the unconditional one, so $\xi_N \sim N(0, N^{-1})$, thus $\xi_N$ has bounded density

$$p_\xi(t) = \frac{\exp(-\frac{1}{2}t^2)}{\sqrt{2\pi N^{-1}}} \leq \frac{\pi^{1/2}}{\sqrt{2\pi}}$$

Moreover, for any interval $I \subset \mathbb{R}$ of length $|I|$, we have

$$\text{ess sup} \mathbb{P}\{\xi_N(\omega) \in I \mid \mathcal{F}\} = \mathbb{P}\{\xi_N(\omega) \in I\} \leq \frac{\pi^{1/2}}{\sqrt{2\pi}} |I|. \tag{2.1}$$

The essential supremum in the above LHS is a bureaucratic tribute to the formal rule saying that $\mathbb{P}\{\cdot \mid \mathcal{F}\}$ is a random variable (which is $\mathcal{F}$-measurable), and as such is defined, generally speaking, only up to subsets of measure zero.

In this particular case – for Gaussian samples – the conditional regularity of the sample mean $\xi_N$ given the fluctuations $\mathcal{F}$ is granted, but is not always so, as shows the following elementary example where the common probability distribution of the sample $X_1, X_2$ is just excellent: $X_1 \sim \text{Unif}([0, 1])$, so $X_1$ admit a compactly supported probability density bounded by 1. Indeed, in this simple example, set

$$\xi = \xi_2 = \frac{X_1 + X_2}{2}, \quad \eta = \eta_1 = \frac{X_1 - X_2}{2}.$$  

The random vector $(X_1, X_2)$ is uniformly distributed in the unit square $[0, 1]^2$, and the condition $\eta = c$ selects a straight line in the two-dimensional plane with coordinates $(X_1, X_2)$, parallel to the main diagonal $\{X_1 = X_2\}$. The conditional distribution of $\xi$ given $\{\eta = c\}$ is the uniform distribution on the segment

$$J_c := \{(x_1, x_2) : x_1 - x_2 = 2c, 0 \leq x_1, x_2 \leq 1\}$$

of length vanishing at $2c = \pm 1$. For $|2c| = 1$, the conditional distribution of $\xi$ on $J_c$ is concentrated on a single point, which is the ultimate form of singularity.

Yet, the good news in this example is that the conditions of singularity are quite explicit, and it is simple to assess the probability of the event that the conditional probability density of $\xi$ given $\mathcal{F}$ is bigger than a given threshold. In the next Section, we exploit this elementary observation in a more general case of $N \geq 2$ IID random variables uniformly distributed in $[0, 1]$. The applications of the main result of Section 3 are discussed in Section 4.

3. The principal applications

3.1. The conditional empirical mean in EVC bounds. Let $\Lambda$ be a finite graph, with $|\Lambda| = N \geq 1$, and $H_\Lambda(\omega)$ be a random DSO acting in the finite-dimensional Hilbert space $\mathcal{H} = \mathcal{H}_\Lambda = \ell^2(\Lambda)$, with IID random potential potential $V : \Lambda \times \Omega \to \mathbb{R}$, relative to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Decomposing the random field $V$ on $\Lambda$, $V(x; \omega) = \xi_N(\omega) + \eta_x(\omega), \xi_N(\omega)$,
we can represent $H(\omega)$ as follows:

$$H(\omega) = \xi_N(\omega) \mathbf{1} + A(\omega),$$

where the operator $A(\omega)$ is $\mathfrak{F}_\eta$-measurable, and so are its eigenvalues $\tilde{\mu}_j(\omega)$, $j = 1, \ldots, N$. Since $A(\omega)$ commutes with the scalar operator $\xi_N(\omega) \mathbf{1}$, the eigenvalues $\lambda_j(\omega)$ of $H(\omega)$ have the form

$$\lambda_j(\omega) = \xi_N(\omega) + \mu_j(\omega). \tag{3.1}$$

The numeration of the eigenvalues $\lambda_j(\omega)$, $\mu_j(\omega)$ is, of course, not canonical, but they can be consistently defined as random variables on $\Omega$.

The representation (3.1) implies immediately the following EVC bound: for any interval $I = [t, t + s],

$$\mathbb{P}\{ \text{tr} \ P_I(H(\omega)) \geq 1 \} \leq \sum_{j=1}^N \mathbb{P}\{ \lambda_j(\omega) \in I \} = \sum_{j=1}^N \mathbb{P}\{ \xi_N(\omega) + \mu_j(\omega) \in I \}

= \sum_{j=1}^N \mathbb{E}\left[ \mathbb{P}\{ \xi_N(\omega) + \mu_j(\omega) \in I \mid \mathfrak{F}_\eta \} \right]

= \sum_{j=1}^N \mathbb{E}\left[ \mathbb{P}\{ \xi_N(\omega) \in [-\mu_j(\omega) + t, -\mu_j(\omega) + t + s] \mid \mathfrak{F}_\eta \} \right]. \tag{3.2}$$

Further, omitting the argument $\omega$ for notational brevity, we have

$$\mathbb{P}\{ \xi_N + \tilde{\mu}_j \in I \mid \mathfrak{F}_\eta \} = \mathbb{P}\{ \xi_N \in [\mu_j + t, \mu_j + t + s] \mid \mathfrak{F}_\eta \}

= \mathbb{P}\{ \xi_N \in [\tilde{\mu}_j, \tilde{\mu}_j + s] \mid \mathfrak{F}_\eta \}$$

where $\tilde{\mu}_j(\omega) := -\mu_j(\omega) + t$ are $\mathfrak{F}_\eta$-measurable, i.e., fixed under the conditioning.

Now introduce the conditional continuity modulus of $\xi_N$, given $\mathfrak{F}_\eta$:

$$\nu_N(s) := \sup_{t \in \mathbb{R}} \text{ess sup} \ \mathbb{P}\{ \xi_N \in [t, t + s] \mid \mathfrak{F}_\eta \}, \ s > 0.$$ 

Obviously,

$$\mathbb{P}\{ \lambda_j \in I \mid \mathfrak{F}_\eta \} \leq \nu_N(s),$$

thus

$$\mathbb{P}\{ \text{tr} \ P_I(H(\omega)) \geq 1 \} \leq |A| \nu_N(s). \tag{3.3}$$

In this section, we discuss by way of example the Wegner-type bounds for a conventional, single-particle DSO, but in applications to the multi-particle EVC bounds, similar objects turn out to be of interest:

$$s \mapsto \mathbb{P}\{ \xi_N(\omega) \in [\tilde{\mu}(\omega), \tilde{\mu}(\omega) + s] \}, \tag{3.4}$$

and

$$s \mapsto \mathbb{P}\{ \xi_N(\omega) \in [\tilde{\mu}(\omega), \tilde{\mu}(\omega) + s] \mid \mathfrak{F}_\eta \}, \tag{3.5}$$

with an $\mathfrak{F}_\eta$-measurable random variable $\tilde{\mu}$. 
3.2. The Gaussian case. In the particular case where $X_i \sim \mathcal{N}(0, 1)$, we can apply the estimate (2.1) and infer from (3.3) that
\[
P\{ \text{tr} P_I(H(\omega)) \geq 1 \} \leq N \cdot \frac{N^{1/2}}{\sqrt{2\pi}} |I| = \frac{|\Lambda|^{3/2}}{\sqrt{2\pi}} |I|.
\] (3.6)

The above RHS gives the correct (linear) dependence upon the length of the interval $|I|$, but the volume factor is has wrong exponent $(3/2)$, compared to the Wegner estimate (with $|\Lambda|^1$). This is not surprising: we have actually exploited only one of the degrees of freedom in the random potential, related to the normalized empirical mean $\tilde{\xi}_N$, while the well-known proof, due to Wegner [8], as well as its more recent variants, make use of all $N = |\Lambda|$ degrees of freedom. The bound (3.6) is certainly insufficient for the proof of absolute continuity of the limiting eigenvalue distribution for the random operator $H(\omega)$ in an infinite graph (e.g., in the lattice $\mathbb{Z}^d$), and this is not an intended application of our method, as was explained in the introduction. On the other hand, it is more than sufficient for applications to the localization analysis, especially for the MSA. It would not be easy to find an even more elementary derivation of a Wegner-like EVC bound suitable for the analysis of resonances in disordered systems, particularly for the Gaussian potentials.

Another drawback of the described approach to the EVC estimates is that the ”abstract” probabilistic component of the proof, viz. the estimate on $v_N(s)$, becomes more complicated for IID random potentials with low regularity of their common probability distribution function (PDF) $F_V$. The existing methods, used in the single-particle Anderson localization theory, provide a large choice of bounds applicable, formally, to arbitrary continuous PDF $F_V$ (i.e., continuous marginal probability distributions); in practice, the MSA for the DSO on lattices and more general countable graphs requires at least log-Hölder continuity of the marginal distribution. The Fractional Moments Method (FMM), which usually provides stronger (exponential) probabilistic localization bounds, when applicable, is even more exigent: it requires Hölder continuity of the marginal measure.

With these considerations in mind, we have to stress again that we aim here mainly at localization analysis for multi-particle Hamiltonians, where the traditional approaches have been unable so far to obtain efficient localization bounds in arbitrarily large finite volumes.

3.3. Reduction to the local analysis in the sample space. Assume that the support $S \subset \mathbb{R}$ of the common continuous marginal probability measure $P_V$ of the IID random variables $X_j$, $1 \leq j \leq N$, is covered by a finite or countable union of intervals:
\[
S \subset \bigcup_{k \in K} J_k, \quad K \subset \mathbb{Z}, \quad J_k = [a_k, b_k], \quad a_{k+1} \geq b_k.
\]
Let $K = K^N$, and for each $k = (k_1, \ldots, k_N) \in K$, denote
\[
J_k = \times_{i=1}^N J_{k_i}.
\]
Owing to the continuity of the marginal measure, $J_k$ are ”essentially” disjoint: for all $k \neq l$, $P_V(J_k \cap J_l) = 0$. Respectively, the family of the parallelepipeds $\{J_k, \ k \in K\}$

---

1As it is well-known by now, owing to deep works by Bourgain–Klein [3], Aizenman et al. [2], and Germinet–Klein [7], Anderson localization in $\mathbb{R}^d$, $d \geq 1$, can be proven for any nontrivial marginal probability distribution, but for the discrete Schrödinger operators this remains a challenging open problem.
Further, consider the Euclidean space $\sim$ variables $X$ random variables and so are their differences also need a rescaled empirical mean so for the purposes of orthogonal transformation ($X$ then the space $R$ probability space $\Omega$. Further, let $\mathcal{F}_K$ be the sub-sigma-algebra of $\mathcal{F}$ generated by the partition $\mathcal{K}$. Now the quantities of the general form (3.3) can be assessed as follows:

$$P\{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \} = E \left[ P \{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \mid \mathcal{F}_K \} \right] = \sum_{k \in K} P \{ J_k \} \left[ P \{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \mid J_k \} \right].$$

Let $P_k \{ \cdot \}$ be the conditional probability measure, given $\{ X \in J_k \}$, $E_k[\cdot]$ the respective expectation, and $p_k = P \{ J_k \}$. Then we have

$$P \{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \} = \sum_{k \in K} p_k E_k \left[ P \{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \mid \mathcal{F}_k \} \right] \leq \sup_{k \in K} E_k \left[ P \{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \mid \mathcal{F}_k \} \right]. \quad (3.7)$$

This simple formula shows that one may seek a satisfactory upper bound on the LHS of (3.7) by assessing the “local” conditional probabilities $P_k \{ \xi_N \in [\bar{\mu}, \tilde{\mu} + s] \mid \mathcal{F}_k \}$, where each random variable $X_j$ is restricted to a subinterval $J_{k_j}$ of its global support, so the entire sample $X = (X_1, \ldots, X_N)$ is restricted to a parallelepiped $J \subset R^N$.

In the next section, we perform such analysis first in the case of a uniform marginal distribution of the IID variables $X_i$.

### 4. Uniform marginal distributions

Let be given a real number $\ell > 0$ and an integer $N \geq 2$. Consider a sample of $N$ IID random variables with uniform distribution $\text{Unif}([0, \ell])$, and introduce again the sample mean $\xi = \xi_N$ and the “fluctuations” $\eta_i$ around the mean:

$$\xi_N = \frac{1}{N} \sum_{i=1}^N X_i, \quad \eta_i = X_i - \xi_N.$$

For the purposes of orthogonal transformation $(X_1, \ldots, X_n) \mapsto (\tilde{\xi}_N, \tilde{\eta}_2, \ldots, \tilde{\eta}_N)$, we also need a rescaled empirical mean

$$\tilde{\xi}_N = N^{1/2} \xi_N,$$

so

$$X_i = \eta_i + N^{-1/2} \tilde{\xi}_N, \quad i = 1, \ldots, N. \quad (4.1)$$

Further, consider the Euclidean space $\sim R^N$ of real linear combinations of the random variables $X_i$ with the scalar product $\langle X', X'' \rangle = E[ X' X'' ]$. Clearly, the variables $\eta_i : R^N \rightarrow R$ are invariant under the group of translations

$$(X_1, \ldots, X_N) \mapsto (X_1 + t, \ldots, X_N + t), \quad t \in R,$$

and so are their differences $\eta_i - \eta_j \equiv X_i - X_j, \ 1 \leq i < j \leq N$. Introduce the variables

$$Y_i = \eta_i - \eta_N, \quad 1 \leq i \leq N - 1. \quad (4.2)$$

Then the space $R^N$ is fibered into a union of affine lines of the form

$$\tilde{X}(Y) := \{ X \in R^N : \eta_i - \eta_N = Y_i, \ i \leq N - 1 \}$$

and

$$:= \{ X \in R^N : X_i - X_N = Y_i, \ i \leq N - 1 \}, \quad (4.3)$$
labeled by the elements \( Y = (Y_1, \ldots, Y_{N-1}) \) of the \((N-1)\)-dimensional real vector space \( \mathbb{R}^{N-1} \cong \mathbb{R}^{N-1} \). Set
\[
\mathcal{X}(Y) = \tilde{\mathcal{X}}(Y) \cap C_1 = \{ X \in C_1 : X_i = X_N = Y_i, \ i \leq N-1 \}
\]
and endow each nonempty interval \( \mathcal{X}(Y) \subset \mathbb{R}^N \) with the natural structure of a probability space inherited from \( \mathbb{R}^N \):
- if \( |\mathcal{X}(Y)| = 0 \) (an interval reduced to a single point), then we introduce the trivial sigma-algebra and trivial counting measure;
- if \( |\mathcal{X}(Y)| = r > 0 \), then we use the inherited structure of an interval of a one-dimensional affine line and the normalized measure with constant density \( r^{-1} \) with respect to the inherited Lebesgue measure on \( \mathcal{X}(Y) \).

The transformation \( X \mapsto (\xi_N, \eta_1, \ldots, \eta_{N-1}) \) is non-degenerate, but not orthogonal. We will have to work with the metric on \( \mathcal{X}(Y) \), induced by the standard Riemannian metric in the ambient space \( \mathbb{R}^N \); to this end, introduce an orthogonal coordinate transformation in \( \mathbb{R}^N \) \( X \mapsto (\xi_N, \tilde{\eta}_1, \ldots, \tilde{\eta}_{N-1}) \) such that
\[
\tilde{\xi}_N = N^{-1/2} \sum_{i=1}^{N} X_i = N^{1/2} \xi_N;
\] (4.4)
the exact form of \( \tilde{\eta}_j, \ j = 1, \ldots, N-1 \) is of no importance, provided that the transformation is orthogonal.

**Remark 4.1.** For later use, note that, owing to (4.4), each of the re-scaled variables \( N^{1/2}X_i \) can serve as the (normalized) length parameter on the elements \( \mathcal{X}(Y) \). Along an element \( \mathcal{X}(Y) \), one can simultaneously parameterize \( \tilde{\xi} \) and the variables \( X_i \), by setting \( \xi(t) = c_0 + t, \ X_j(t) = c_j + N^{-1/2}t, \) with arbitrarily chosen constants \( c_j \). Here, \( \tilde{\xi}_N \) is a natural length parameter on \( \mathcal{X}(Y) \), since the transformation \( X \mapsto (\xi_N, \tilde{\eta}_1, \ldots, \tilde{\eta}_{N-1}) \) is orthogonal.

It follows from (4.4) that for any given \( a \in \mathbb{R}, \ s > 0, \) and some \( a' \in \mathbb{R}, \)
\[
\xi_N \in [a, a + s] \iff \tilde{\xi}_N \in [a', a' + N^{1/2}s];
\] (4.5)
Next, denote \( J^{(t)} = [0, \ell]^N \) and introduce the random variable
\[
\nu_N(s; J^{(t)}) = \nu_N(s; J^{(t)}; X) := \text{ess sup}_{t \in \mathbb{R}} \sup_{\mathcal{F}_0} \{ \xi_N \in [t, t + s] \mid \mathcal{F}_0 \};
\] (4.6)
Here the presence of ess sup is the tribute to the fact that the conditional probabilities are random variables, usually defined up to subsets of zero measure; \( \ell > 0 \) is the width of the common uniform distribution of \( X_j \). Equivalently, one may write \( \nu_N(s; J^{(t)}; \omega) \) instead of \( \nu_N(s; J^{(t)}; X) \), since the sample space \( \mathbb{R}^N \) is identified with the underlying probability space \( \Omega \).

Since \( \{X_i\} \) are IID with uniform distribution on \([0, \ell]\), the distribution of the random vector \( X(\omega) \) is uniform in the cube \( J^{(t)} = [0, \ell]^N \), inducing a uniform conditional distribution on each element \( \mathcal{X}(Y) \). Therefore, by (4.5) and (4.6),
\[
\nu_N(s; J^{(t)}) = \frac{N^{1/2}s}{|\mathcal{X}(Y)|}.
\] (4.7)
It is to be stressed that both sides of the above equality are random variables: \( \nu_N(s; \ell) = \nu_N(s; \ell; \omega) \) by its definition in (4.4), and \( \mathcal{X}(Y) = \mathcal{X}(X(\omega)) \).
Lemma 1. Consider the IID random variables \(X_1, \ldots, X_N\) with \(X_i \sim \text{Unif}(J_{i,i})\), where \(J_{i,i} = [a_i, a_i + \ell] \subseteq \mathbb{R}, \ell > 0\). For any \(0 < \delta \leq \ell\),
\[
\mathbb{P} \left\{ |\mathcal{X}(X)| \leq \delta \right\} \leq \sum_{i=1}^{N} \mathbb{P} \left\{ X_i - a_i < \delta \right\}. \tag{4.8}
\]

Proof. Without loss of generality, we can consider the case where \(a_i = 0, 1 \leq i \leq N\), so \(X_i \sim \text{Unif}([0, \ell])\). Otherwise, we make change of variables \(X_i \mapsto X_i - a_i\).

Let \(X = X(X) = \min_i X_i\). \tag{4.9}

According to Remark 4.1, each \(N^{1/2}X_i, i = 1, \ldots, N\), restricted to \(\mathcal{X}(Y)\), provides a normalized length parameter on \(\mathcal{X}(Y)\); thus the range of each \(N^{1/2}X_i|_{\mathcal{X}(Y)}\) is an interval of length \(|\mathcal{X}(Y)|\). One can decrease, e.g., the value of \(X_1\), as long as all \(\{X_i, 1 \leq i \leq N\}\) are strictly strictly positive. Therefore, the maximum decrement of \(X_1\) (indeed, of any \(X_i\) along \(\mathcal{X}(Y)\) is given by \(\mathcal{Y}(X)\), so the range of the normalized length parameter \(N^{1/2}X_1\) along \(\mathcal{X}(Y(X))\) is an interval of length \(\geq N^{1/2}\mathcal{X}(X)\):
\[
|\mathcal{X}(Y(X))| \geq N^{1/2}\overline{X}(X). \tag{4.10}
\]

Let \(A_i(t) := \{X_i < t\}, A(t) := \cup_{i=1}^{N} A_i(t), A^c(t) = \Omega \setminus A(t), \tag{4.11}\)
and note that, by (4.10),
\[
\min_{X \in A^c(t)} |\mathcal{X}(X)| \geq N^{1/2}\min_{X \in A^c(t)} X(X) \geq N^{1/2}t. \tag{4.12}
\]

Equivalently, setting \(u = N^{-1/2}t\), so \(t = N^{-1/2}u\), we have
\[
|\mathcal{X}(X)| < u \implies X \in A(N^{-1/2}u). \tag{4.13}
\]

With \(u = \delta\), we infer from (4.10)
\[
\mathbb{P}\left\{ A \left( N^{1/2}N^{-1/2}\delta \right) \right\} = \mathbb{P}\left\{ A(\delta) \right\} \leq \sum_{i=1}^{N} \mathbb{P} \left\{ X_i < \delta \right\}. \tag{4.14}
\]
proving the assertion (4.8). \qed

Theorem 1. Consider IID random variables \(X_1, \ldots, X_N\) with \(X_i \sim \text{Unif}(J_{i,i})\), where \(J_{i,i} = [a_i, a_i + \ell] \subseteq \mathbb{R}, \ell > 0\). For any \(0 < \delta \leq \ell\),
\[
\mathbb{P}\left\{ \nu_N(s; J^{(\ell)}) > \delta^{-1} s \right\} \leq \frac{N\delta}{\ell}. \tag{4.15}
\]

In particular, with \(\delta = s^\alpha\),
\[
\mathbb{P}\left\{ \nu_N(s; J^{(\ell)}) > s^{1-\alpha} \right\} < N\ell^{-1}s^\alpha \tag{4.16}
\]

Proof. The random variable \(X = (X_1, \ldots, X_N) \mapsto |\mathcal{X}(Y(X))|\) is \(\mathcal{F}_{\eta}\)-measurable and takes constant value \(|\mathcal{X}(Y)|\) on each element \(\mathcal{X}(Y)\). By (4.7), for any \(\delta > 0\),
\[
\mathbb{P}\left\{ \nu_N(s; J^{(\ell)}) \geq \delta^{-1} s \right\} \leq \mathbb{P}\left\{ \frac{N^{1/2}\delta}{|\mathcal{X}(Y)|} \geq \delta^{-1} s \right\} = \mathbb{P}\left\{ |\mathcal{X}(Y)| \leq N^{1/2}\delta \right\}. \tag{4.17}
\]

Now (4.14) follows from (4.16) and Lemma 1 since for \(X_i \sim \text{Unif}([0, \ell])\)
\[
\mathbb{P}\left\{ X_i < \delta \right\} = \ell^{-1}\delta. \tag{4.18}
\]

\qed
5. More accurate bounds

A direct inspection shows that the bounds of Lemma 1 (and, consequently, those of Theorem 1) are not optimal, since they are based on the inequality
\[ |X(Y(X))| \geq N^{1/2} \overline{X}(X) \] (5.1)
(cf. (5.15)) which can be easily improved; we do so in Theorem 2 below. However, the method of proof of Lemma 1 is simpler and quite sufficient for our main application to the multi-particle MSA.

**Lemma 2.** Assume that the IID random variables \(X_1, \ldots, X_N\), \(N \geq 2\), admit (common) probability density \(p_V\) with \(\|p_V\|_{\infty} \leq \rho < \infty\). Then
\[ P\{ |X(Y)| < r \} \leq \frac{1}{4} \rho^2 r^2 N. \] (5.2)

In particular, for \(X_j \sim \text{Unif}([0, \ell])\), one has
\[ P\{ |X(Y)| < r \} \leq \frac{r^2 N}{4\ell^2}. \] (5.3)

**Proof.** Let
\[ \underline{X} = \underline{X}(X) = \min_i X_i, \quad \overline{X} = \overline{X}(X) = \max_i X_i. \] (5.4)

While \(\overline{X}(X)\) and \(\underline{X}(X)\) vary along the elements \(X(Y)\), their difference \(\overline{X}(X) - \underline{X}(X)\) does not; it is uniquely determined by \(X(Y)\).

According to Remark 4.1 each \(N^{1/2}X_i\), \(i = 1, \ldots, N\), restricted to \(X(Y)\), provides a normalized length parameter on \(X(Y)\); thus the range of each \(N^{1/2}X_i|_{X(Y)}\) is an interval of length \(|X(Y)|\). One can increase (resp., decrease), e.g., the value of \(X_1\), as long as all \(\{X_i, 1 \leq i \leq N\}\) are strictly smaller than \(\ell\) (resp., strictly positive). Therefore, the maximum increment of \(X_1\) (indeed, of any \(X_i\)) along \(X(Y)\) is given by \(\ell - \overline{X}(X)\), and its maximum decrement equals \(\underline{X}(X)\), so the range of the normalized length parameter \(N^{1/2}X_1\) along \(X(Y(X))\) is an interval of length \(N^{1/2}(\ell - \overline{X}(X) + \underline{X}(X))\):
\[ |X(Y(X))| = N^{1/2}(\ell - \overline{X}(X) + \underline{X}(X)). \] (5.5)

Since both \(\underline{X}(X)\) and \(\ell - \overline{X}(X)\) are non-negative,
\[ \underline{X} + (\ell - \overline{X}) < t \implies \max\{\underline{X}, \ell - \overline{X}\} < t/2. \] (5.6)

With \(0 \leq t \leq \ell\), \((\ell - X_i < t/2)\) implies \((X_i > t/2)\), thus denoting
\[ A_{ij}(t) := \{X_i < t/2\} \cap \{\ell - X_j < t/2\}, \] (5.7)

we have, for any \(i\),
\[ A_{ii}(t) = \{X_i < t/2\} \cap \{\ell - X_i < t/2\} = \emptyset. \] (5.8)

Therefore,
\[ \left\{ \max\{\underline{X}(X), \ell - \overline{X}(X)\} < \frac{t}{2} \right\} \subset \bigcup_{i \neq j} \left\{X_i < \frac{t}{2}, \ell - X_j < \frac{t}{2}\right\}. \] (5.9)

Thus the union \(\bigcup_{i \neq j} A_{ij}(t)\) contains all samples \(X\) with \(|X(Y)| < t/2\).
The sample \( \{X_k\} \) is IID, with common probability density uniformly bounded by \( \overline{p} < \infty \), so for any \( i \neq j \)
\[
\Pr \{ A_{ij}(t) \} = \Pr \left\{ X_i < \frac{t}{2} \right\} \cdot \Pr \left\{ \ell - X_j < \frac{t}{2} \right\} = \frac{1}{4} \overline{p}^2 t^2.
\]
Therefore,
\[
\Pr \{ |\mathcal{X}(Y)| < r \} = \Pr \left\{ N^{1/2} ((\ell - \overline{X}(X)) + \overline{X}(X)) < r \right\} = \Pr \left\{ ((\ell - \overline{X}(X)) + \overline{X}(X)) < r N^{-1/2} \right\} 
\leq \sum_{i \neq j} \Pr \left\{ A_{ij}(r N^{-1/2}) \right\} \leq N(N - 1) \frac{(\overline{p}rN^{-1/2})^2}{4} \tag{5.10}
\leq \frac{1}{4} \overline{p}^2 r^2 N.
\]
\[\square\]

**Theorem 2.** Consider the IID random variables \( X_1, \ldots, X_N \) with \( X_i \sim \text{Unif}([0, \ell]) \). For any \( 0 < \delta \leq s \leq \ell \),
\[
\Pr \left\{ \nu_N(s; J^{(i)}) > \delta^{-1} s \right\} \leq \frac{N^2 \delta^2}{4 \ell^2}. \tag{5.11}
\]
In particular, with \( \delta = s^\alpha \), \( \alpha \in (0, 1) \),
\[
\Pr \left\{ \nu_N(s; J^{(i)}) > s^{1-\alpha} \right\} \leq \frac{N^2 s^{2\alpha}}{4 \ell^2} \tag{5.12}
\]
**Proof.** As before, we associate with each point \( X \in \mathbb{R}^N \) the straight line \( \mathcal{L}(Y(X)) \supset X \) parallel to the vector \( v = (1, \ldots, 1) \), and consider their intersections \( \mathcal{X}(Y(X)) = \mathcal{L}(Y(X)) \cap J^{(i)} \). Owing to Eqn (4.9), for any \( \delta > 0 \),
\[
\Pr \left\{ \nu_N(s) \geq \delta \right\} \leq \Pr \left\{ \frac{N^{1/2} s}{|\mathcal{X}(Y)|} \geq \delta \right\} = \Pr \left\{ |\mathcal{X}(Y)| \leq \frac{N^{1/2} s \delta^{-1}}{\delta} \right\}. \tag{5.13}
\]
Let \( \overline{X} = \overline{X}(X) = \min_i X_i, \underbar{X} = \underbar{X}(X) = \max_i X_i \). While \( \overline{X}(X) \) and \( \underbar{X}(X) \) vary along the elements \( \mathcal{X}(Y) \), their difference \( \overline{X}(X) - \underbar{X}(X) \) does not; it is uniquely determined by \( \mathcal{X}(Y) \).

According to Remark 4.1, each \( N^{1/2} X_i, i = 1, \ldots, N \), restricted to \( \mathcal{X}(Y) \), provides a normalized length parameter on \( \mathcal{X}(Y) \); thus the range of each \( N^{1/2} X_i |_{\mathcal{X}(Y)} \) is an interval of length \( |\mathcal{X}(Y)| \). One can increase (resp., decrease), e.g., the value of \( X_1 \), as long as all \( \{X_i, 1 \leq i \leq N\} \) are strictly smaller than \( \ell \) (resp., strictly positive). Therefore, the maximum increment of \( X_1 \) (indeed, of any \( X_i \)) along \( \mathcal{X}(Y) \) is given by \( \ell - \overline{X}(X) \), and its maximum decrement equals \( \underbar{X}(X) \), so the range of the normalized length parameter \( N^{1/2} X_1 \) along \( \mathcal{X}(Y(X)) \) is an interval of length \( N^{1/2} (\ell - \overline{X}(X) + \underbar{X}(X)) \):
\[
|\mathcal{X}(Y(X))| = N^{1/2} (\ell - \overline{X}(X) + \underbar{X}(X)), \tag{5.15}
\]
Since both \( \overline{X}(X) \) and \( \ell - \overline{X}(X) \) are non-negative,
\[
\overline{X}(X) + (\ell - \overline{X}(X)) < t \implies \max \{ \overline{X}(X), \ell - \overline{X}(X) \} < t/2. \tag{5.16}
\]
With $0 \leq t \leq \ell$, $(\ell - X_i < t/2)$ implies $(X_i > t/2)$, thus denoting
\[ A_{ij}(t) := \{X_i < t/2\} \cap \{\ell - X_j < t/2\}, \]
we have, for any $i$,
\[ A_{ii}(t) = \{X_i < t/2\} \cap \{\ell - X_i < t/2\} = \emptyset. \] (5.18)
Therefore,
\[ \left\{ \max \{X(X), \ell - X(X)\} < \frac{t}{2} \right\} \subset \bigcup_{i \neq j} \left\{ X_i < \frac{t}{2}, \ell - X_j < \frac{t}{2} \right\}. \] (5.19)
Thus the union $\bigcup_{i \neq j} A_{ij}(t)$ contains all samples $X$ with $|X(Y)| < t/2$.
The sample $\{X_k\}$ is IID, with $X_k \sim \text{Unif}([0, \ell])$, so for any $i \neq j$
\[ P \{ A_{ij}(t) \} = P \left\{ X_i < \frac{t}{2} \right\} \cdot P \left\{ \ell - X_j < \frac{t}{2} \right\} = \frac{t^2}{4\ell^2}. \]
Owing to (4.10),
\[ P \{ |X(Y)| < r \} = P \left\{ N^{1/2} ((\ell - X(X)) + X(X)) < r \right\} \]
\[ = P \left\{ ((\ell - X(X)) + X(X)) < rN^{-1/2} \right\} \]
\[ \leq \sum_{i \neq j} P \left\{ A_{ij}(rN^{-1/2}) \right\} \leq N(N - 1) \left( \frac{rN^{-1/2}}{4\ell^2} \right)^2 \]
\[ \leq \frac{r^2N}{4\ell^2}. \] (5.20)
Setting $r = N^{1/2}\delta$, we infer from (5.13)
\[ P \{ \nu(s; \ell) > \delta \} \leq \frac{N^2s^2}{4\ell^2}. \] (5.21)
proving (5.11). The estimate (5.12) is a particular case of (5.11). \qed

In Ref. [4], we introduced the following more general condition, which actually
do not assume the independence of the random variables $X_j$. Let us reformulate
it now in a more general way so as to adapt it to locally finite connected graphs $Z$
with polynomially bounded growth of balls $B_L(u) := \{x \in Z : d_Z(x, u) \leq L\}$ (in
[4], we had $Z = \mathbb{Z}^d$);
\[ \text{card } B_L(u) \leq C_dL^d, \quad l \geq 1. \] (5.22)
(We also adapt the notation of [4] to match the one used in this paper.)
Let $Q \subset B_R(x) \subset Z$ be a subset of a ball of radius $R$. Consider the sample of IID
random variables $\{V(y; \omega), y \in Q\}$; introduce as in (11) the sample mean $\xi_Q$
and the conditional continuity modulus $\nu_Q(s)$ given the sigma-algebra of fluctuations.
Since $Q \subset B_R(x) \subset Z$, where $Z$ satisfies (5.22), we have $|Q| \leq C_dR^d$.
The hypothesis used in [4], reformulated for general index sets $Q$, takes the following form: for some $C', C''$, $A', A'', B', B'' \in (0, +\infty)$
\[ P \left\{ \nu_Q(s) \geq C'|Q|^{A'}s^{B'} \right\} \leq C''|Q|^{A''}s^{B''}. \] (5.23)
To keep track of the length $\ell$ of the interval $[0, \ell]$, re-write it as follows:
\[ P \left\{ \nu_Q(s; \ell) \geq C'|Q|^{A'}s^{B'} \right\} \leq C''|Q|^{A''}s^{B''}. \] (5.24)
ON THE CONDITIONAL DISTRIBUTION OF THE SAMPLE MEAN

We will say that a random field \( V : \mathbb{Z} \times \Omega \rightarrow \mathbb{R} \) on a countable set \( \mathbb{Z} \) (not necessarily IID) is of class (RCM) (here RCM stands for "Regularity of the Conditional Mean") if it satisfies the condition (5.24) for some values \( C', C'', A', A'', B', B'' \in (0, +\infty) \). Naturally, it can be made less cumbersome, since some of these constants can be eliminated by a proper scaling of the variable \( s \), but it might be convenient in some applications to keep all these parameters.

If the random field \( V \) is assumed IID, then (5.24) is merely a condition on the common marginal probability distribution; in this particular (but important) case, one can speak of the class (RCM) of the probability distributions.

We see that, for an IID sample with distribution \( \text{Unif}([0, \ell]) \), \( \ell > 0 \), Theorem 2 can be reformulated in the following way:

**Theorem 3.** Let an IID random field \( V : \mathbb{Z} \times \Omega \) on a finite or countable graph \( \mathbb{Z} \), satisfying the growth condition (5.22), have marginal distribution \( \text{Unif}([c, c+\ell]) \), \( c \in \mathbb{R} \). Then \( V \) satisfies the condition (RCM) of the form (5.24) with the parameters which can be chosen as follows:

\[
C' = 1, \quad A' = 0, \quad b' = 1 - \alpha,
\]
\[
C'' = \frac{1}{4\ell^2}, \quad A'' = 2, \quad b'' = 2\alpha.
\]

(5.25)

For example, one can set

\[
b' = b'' = \frac{2}{3}.
\]

(5.26)

Explicitly,

\[
P\{ \nu_Q(s; \ell) > s^{1-\alpha} \} < \frac{|Q|^2}{4\ell^2} s^{2\alpha}.
\]

(5.27)

6. Smooth positive densities

Now we consider a richer class of probability distributions. While the conditions which we will assume are certainly very restrictive (uniform positivity and smoothness of the probability density on a compact interval), they are quite sufficient for applications to physically realistic Anderson models.

A direct inspection of the proof of Theorem 2 evidences that the hypothesis of strict positivity of the probability density (\( \rho \geq \rho_0 > 0 \), cf. (6.2) below) can be easily replaced by a more general condition of mild decay at the endpoints of \( \text{supp} \rho \), e.g.,

\[
\rho(t) \geq C \left( \min\{t, \ell - t\} \right)^{a}, \quad C, a \in (0, +\infty).
\]

This extends our result to a large class of popular a.c. probability distributions, including the convolution powers of the uniform distribution. Further, the distributions with unbounded support can be treated as well, provided that the probability density decays sufficiently fast at infinity (e.g., the exponential distribution and, more generally, gamma-distributions). We plan to address such probability measures in a forthcoming paper.

**Theorem 4.** Assume that the common probability distribution of the IID random variables \( V_j, j = 1, \ldots, N \), with PDF \( F_V \), satisfies the following conditions:

(i) the probability distribution is absolutely continuous:

\[
dF_V(v) = \rho(v) \, dv, \quad \text{supp} \, \rho = [0, \ell];
\]

(6.1)
(ii) there exist $\rho_*, \overline{\rho} \in (0, +\infty)$ such that
\[ \forall t \in [0, \ell], \quad \rho_* \leq \rho(t) \leq \overline{\rho}; \]  
(6.2)

(iii) $\rho$ has bounded derivative on $(0, \ell)$:
\[ \|\rho'(\cdot)\|_{\infty} \leq C_{\rho'} < +\infty. \]  
(6.3)

Then there exists $c_* = c_*(F_Y) > 0$ such that for any $\delta \in (0, c_* N^{-3/2}]$,
\[ \mathbb{P}\{ \nu_N(s) > \delta^{-1} s \} < \frac{477 N^2 \delta^2}{\ell^2}. \]  
(6.4)

In particular, with $\delta = s^\alpha \leq c_*^{1/\alpha} N^{-3/(2\alpha)}$, $\alpha \in (0, 1)$, one has
\[ \mathbb{P}\{ \nu_N(s) > s^{1-\alpha} \} < \frac{477}{\ell^2} N^2 s^{2\alpha}. \]  
(6.5)

Consequently, the IID random fields satisfying (i)–(iii) belong to the class $\text{RCM}$.

Proof. Step 1. Smoothness of the conditional measure. Unlike the model considered in Section 4, the conditional probability distribution induced on a given interval $\mathcal{X}(Y)$ is no longer constant. However, owing to the smoothness assumption (iii), the product probability measure with density
\[ \mathbf{p}(x_1, \ldots, x_n) = \prod_{j=1}^n \rho(x_j) = e^{\sum_{j=1}^n \ln \rho(x_j)} \]
induces on the interval $\mathcal{X}(Y) \subset \mathcal{L}(Y)$ a measure with smooth density with respect to the Lebesgue measure on the line $\mathcal{L}(Y) \subset \mathbb{R}^N$. Let $t = \xi_N$ be the normalized length parameter along $\mathcal{L}(Y)$, then (cf. (4.1))
\[ \mathcal{L}(Y) = \{ (\eta_1 + t N^{-1/2}, \ldots, \eta_N + t N^{-1/2}), t \in \mathbb{R} \}, \]
so the density at the point $t$ has the form
\[ \rho(t) = Z^{-1}(Y) \prod_{j=1}^n \rho(\eta_j + t) = e^{\sum_{j=1}^n \ln \rho(\eta_j + t)} \]
where $Z^{-1}(Y)$ is the normalization factor. In particular,
\[ \frac{d}{dt} \rho(t) = N^{-1/2} \rho(t) \sum_{j=1}^N \frac{\rho'(\eta_j + t N^{-1/2})}{\rho(\eta_j + t N^{-1/2})}. \]  
(6.6)

Step 2. From $\nu$ to $|\mathcal{X}(Y)|$. By (6.6) combined with assumption (6.2),
\[ \left\| \frac{\rho'}{\rho} \right\|_{\infty} \leq N \cdot N^{-1/2} C_{\rho'} \rho_*^{-1} \leq C_1 N^{1/2}, \]
In particular,
\[ \|\nu'\|_{\infty} \leq C_1 N^{1/2} \left\| \mathbf{p} \right\|_{\infty}. \]  
(6.7)

For notational convenience, identify $\mathcal{L}(Y)$ with the real line $\mathbb{R}$, equipped with the normalized coordinate $t = \xi_N$, and let $t^* = t^*(Y)$ be any point of maximum of the density $\rho$ restricted to $\mathcal{X}(Y)$, and $\rho^*(Y) = \rho(t^*)$; the existence of $t^*(Y)$ follows from the continuity of $\rho$. Assume that
\[ |\mathcal{X}(Y)| > 2 \ell_N, \quad \ell_N \leq \ell_* N^{-1}, \]  
(6.8)
where $\ell_* = \ell_* (F_Y) > 0$ is small enough, viz.

$$
\ell_* (F_Y) = (C_l (F_Y))^{-1},
$$

and depends upon the minimum of the density $p(\cdot)$ and the sup-norm of its derivative; both of these quantities are determined by the PDF $F_Y$. Since $|X(Y)| > 2\ell_N$, at least one of the intervals $[t^* - \ell_N, t^*], [t^*, t^* + \ell_N]$ (perhaps, both of them) is inside the interval $X(Y)$. Denote by $J_*$ such an interval (for definiteness, the first one, if both are inside $X(Y)$).

Then for any $t \in X(Y)$, owing to (6.7),

$$
|\rho(t) - \rho(t^*)| \leq \ell_* N^{-1} \cdot \max_{s \in J_*} \rho'(s) \leq \frac{(C_l \ell_*) N^{1/2} \cdot N^{-1} \cdot \rho^*(Y)}{1}
$$

so that $\forall t \in X(Y)$ and, e.g., $N \geq 4$,

$$
\frac{1}{2} \rho^*(Y) \leq \rho^*(Y) \left(1 - N^{-1/2}\right) \leq \rho(t) \leq \rho^*(Y) \left(1 + N^{-1/2}\right) \leq 2 \rho^*(Y)
$$

The conditional measure induced on $X(Y)$ has the form $dP_Y(t) = Z(Y) \rho(t) dt$, with $Z(Y) = \int_{X(Y)} \rho(t) dt$, and we have

$$
Z(Y) \geq \int_{J_*} \frac{1}{2} \rho(t^*) dt = \frac{1}{2} \rho(t^*) \ell_N.
$$

Therefore, under the assumption $|X(Y)| > 2\ell_N$, we have for any $t' \in \mathbb{R}$:

$$
P \left\{ \xi_N \in [t', t' + s] \mid Y \right\} = P \left\{ \xi_N \in [t'', t'' + N^{1/2} s] \mid Y \right\} = Z^{-1}(Y) \int_{t''}^{t'' + N^{1/2} s} \rho(t) dt = \frac{\rho(t'') N^{1/2} s}{\rho(t^*) \ell_N} = \frac{2 N^{1/2} s}{\ell_N}
$$

(here $t'' = N^{1/2} t'$), yielding, for such $X(Y)$,

$$
\nu_N(s \mid Y) \leq 2 N^{1/2} \ell_N^{-1} s.
$$

Therefore,

$$
\left\{ \nu_N(s) > 2 N^{1/2} \ell_N^{-1} s \right\} \subset \left\{ |X(Y)| < 2\ell_N \right\}
$$

Set $\delta := \frac{1}{2} N^{-1/2} \ell_N$, $c_* = c_*(F_Y) := \frac{1}{2} \ell_* (F_Y)$. Then for any $0 < \delta \leq c_* N^{-3/2},$

$$
\left\{ \nu_N(s) > s \delta^{-1} \right\} \subset \left\{ |X(Y)| < 4 N^{1/2} \delta \right\}. \quad (6.9)
$$

**Step 3. Conclusion.** Now we apply Lemma 2 (cf. (5.2)),

$$
P \left\{ |X(Y)| < r \right\} \leq \frac{1}{4\pi^2 r^2} N,
$$

and obtain, with $r = 4 N^{1/2} \delta,$

$$
P \left\{ |X(Y)| < 4 N^{1/2} \delta \right\} \leq \frac{4\pi^2 N^2}{\delta^2}. \quad (6.10)
$$

Now the main assertion follows from (6.10) and (6.9): for any $\delta \in (0, c_* N^{-3/2})$

$$
P \left\{ \nu_N(s) > s \delta^{-1} \right\} \leq \frac{4\pi^2 N^2}{\delta^2}.
$$

$\square$
REFERENCES

[1] M. Aizenman and S. Warzel, *Localization bounds for multi-particle systems*, Comm. Math. Phys. **290** (2009), no. 3, 903–934.

[2] M. Aizenman, F. Germinet, A. Klein, and S. Warzel, *On Bernoulli decompositions for random variables, concentration bounds and spectral localization*, Probab. Theory Related Fields **143** (2009), 219–238.

[3] J. Bourgain and C.E. Kenig, *On localization in the continuous Anderson-Bernoulli model in higher dimension*, Invent. Math. **161** (2005), 389–426.

[4] V. Chulaevsky, *A remark on charge transfer processes in multi-particle systems*, 2010, arXiv:math-ph/1005.3387.

[5] V. Chulaevsky, *On resonances in disordered multi-particle systems*, C.R. Acad. Sci. Paris, Ser. I, **350** (2011), 81–85.

[6] V. Chulaevsky and Y. Suhov, *Multi-particle Anderson localisation: induction on the number of particles*, Math. Phys. Anal. Geom. **12** (2009), no. 2, 117–139.

[7] F. Germinet and A. Klein, *A comprehensive proof of localization for continuous Anderson models with singular random potentials*, J. Eur. Math. Soc. (JEMS) **15** (2013), 55143.

[8] F. Wegner, *Bounds on the density of states in disordered systems*, Z. Phys. B. Condensed Matter **44** (1981), 9–15.