

On Conditional Distribution of the Sample Mean for Densities with Singular Logarithmic Derivative

Victor Chulaevsky

Département de Mathématiques, Université de Reims, Moulin de la Housse, B.P. 1039, 51687 Reims Cedex 2, France

Copyright ©2015 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract We study the regularity of the conditional distribution of the empiric mean of a finite sample of IID random variables with a bounded common probability density, conditional on the sample "fluctuations", and extend a prior result, proved for strictly positive smooth densities, to a larger class of smooth densities vanishing at one or more points of their support.

Keywords Conditional Sample Mean, Eigenvalue Concentration Estimates, Multi-particle Anderson Localization

1 Introduction

1.1 Formulation of the problem

Consider a sample of $n > 1$ IID (independent and identically distributed) random variables (r.v., in short) X_1, \dots, X_n with Gaussian distribution $\mathcal{N}(0, 1)$, and introduce the sample mean $\xi = \xi_n$ and the "fluctuations" η_i relative to the mean:

$$\xi_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad \eta_i = X_i - \xi_n, \quad i = 1, \dots, n. \quad (1.1)$$

It is well-known from standard courses of the probability theory and statistics (cf., e.g., [6]) that ξ_n is independent of the sigma-algebra \mathfrak{F}_η generated by $\{\eta_1, \dots, \eta_n\}$. Therefore, the conditional probability distribution of ξ_n given \mathfrak{F}_η coincides with the unconditional one, $\mathcal{N}(0, n^{-1})$, thus ξ_n has bounded conditional density

$$p_\xi(t) = \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi n^{-1}}} \leq \sqrt{\frac{n}{2\pi}}, \quad (1.2)$$

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

$$\mathbb{P} \{ \xi_n(\omega) \in I \mid \mathfrak{F}_\eta \} \stackrel{\text{a.s.}}{=} \mathbb{P} \{ \xi_n(\omega) \in I \} \leq \sqrt{\frac{n}{2\pi}} |I|. \quad (1.3)$$

In other words, the conditional probability distribution function (PDF) of the sample mean, given the fluctuations $\{\eta_i\}$, is

Lipschitz continuous. (1.4) implies also the following bound: for any \mathfrak{F}_η -measurable r.v. λ , and any $s \geq 0$

$$\mathbb{P} \{ \xi_n(\omega) \in [\lambda(\omega), \lambda(\omega) + s] \} \leq \sqrt{\frac{n}{2\pi}} s. \quad (1.4)$$

For the proof, it suffices to use conditioning on \mathfrak{F}_η : setting $I_s(\omega) = [\lambda(\omega), \lambda(\omega) + s]$, we have

$$\begin{aligned} \mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \} &= \mathbb{E} [\mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \mid \mathfrak{F}_\eta \}] \\ &\leq \mathbb{E} \left[\sup_{a \in \mathbb{R}} \mathbb{P} \{ \xi_n(\omega) \in [a, a + s] \mid \mathfrak{F}_\eta \} \right] \leq \sqrt{\frac{n}{2\pi}} s. \end{aligned} \quad (1.5)$$

It is to be emphasized that the estimate on the probability $\mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \}$ with a *non-constant, random* interval $I_s(\omega) = [\lambda(\omega), \lambda(\omega) + s]$, is more informative, and more difficult to obtain, than its counterpart for a fixed interval $I_s = [a, a + s]$, $a \in \mathbb{R}$, $s > 0$. With a fixed interval I_s , a number of classical results of the probability theory are available. The regularity of the probability distribution of the sample mean $\xi_n(\omega)$, or simply of the sum of IID r.v. $X_1 + \dots + X_n$, is at least as good as that of each term in the sum, and the smoothing effect of convolution makes it even better with $n \rightarrow +\infty$. Standard textbook examples show that already the sum $X_1 + X_2$ of IID r.v. with purely singular continuous distribution can have absolutely continuous probability distribution; see, e.g., [6, v.II, Section V.4] where it is shown that the uniform distribution in $[0, 1]$ is a convolution of two singular probability measures of Cantor type.

When I_s is random, the situation becomes more complicated.

In the present paper, as in our prior works [3, 4], we handle the problem by using the conditional probability distribution of the sample mean, given the sigma-algebra generated by the fluctuations $\eta_1, \dots, \eta_{n-1}$. The main technical difficulty arising here is that of n degrees of freedom, initially present in the sample mean $\xi_n(\omega) = (X_1(\omega) + \dots + X_n(\omega))/n$, only one remains after fixing $n - 1$ parameters $\eta_1, \dots, \eta_{n-1}$, which can be considered as coordinates in the n -dimensional sample space. When the r.v. X_i have a density $\rho(\cdot)$ (which we always assume in the present paper), their joint probability distribution also has density, viz. $\mathbf{p}(x_1, \dots, x_n) = \rho(x_1) \dots \rho(x_n)$, and the aforementioned conditional distribution of ξ_n is proportional (up to a normalization factor) to

the restriction of $\mathbf{p}(\cdot)$ to the straight line in the n -dimensional sample space selected by the $n - 1$ conditions $\eta_1 = c_1, \dots, \eta_{n-1} = c_{n-1}$. The normalization factor brings up a technical problem, for it can be very large (we address this problem in detail in Sections 4 and 5). Also, if ρ is non-constant, i.e., the marginal distribution of X_i is not a uniform one in some bounded interval, then it takes at least two different values, say, $a < b$, thus the tensor product \mathbf{p} takes values a^n and b^n , with the ratio $(b/a)^n \gg 1$ for $n \gg 1$. In the intended applications, this might result in inadequate regularity bounds, so we have to address this issue, too.

A natural question is, to what extent the above mentioned remarkable property of Gaussian IID samples can be generalized for other types of marginal probability measures. Surprisingly, this problem appears to be unexplored in a reasonably general setup (cf. Section 6). The author of these lines would greatly appreciate any constructive feedback from experts in the field that would shed more light on the regularity problem at hand. A particularly challenging case is where the IID r.v. X_i have a purely singular probability distribution, but the (cumulative) probability distribution function (PDF), $t \mapsto F(t) := \mathbb{P}\{X_1(\omega) \leq t\}$, $t \in \mathbb{R}$, is, for example, Hölder continuous, or more generally, has an explicitly known continuity modulus

$$\mathfrak{s}(s) := \sup_{a \in \mathbb{R}} (F(a + s) - F(a)), \quad s > 0. \quad (1.6)$$

In a prior work [4], we studied this problem under the following condition:

(V1): *The probability measure μ has bounded support, $\text{supp } \mu = [a, b]$, and admits on (a, b) smooth probability density ρ satisfying the following conditions:*

$$\forall t \in (a, b) \quad \begin{cases} 0 < \rho_* \leq \rho(t) \leq \bar{\rho} < \infty, \\ |\rho'(t)/\rho(t)| \leq C < \infty. \end{cases} \quad (1.7)$$

The prototypical example is the uniform distribution on an interval $[a, b]$. Informally, one can say that the probability distributions satisfying (V1) are "comparable" to the uniform distributions. Under this assumption, we proved Hölder continuity of the conditional distribution of the sample mean for typical conditions.

More precisely, we introduced in [4] a property of the probability measures on \mathbb{R} resembling (1.4) and called Regularity of the Conditional Mean:

(RCM) For all $n \geq 2$ and some $C, a \in (0, +\infty)$, $b \in (0, 1]$, for any \mathfrak{F}_η -measurable random variable λ , one has

$$\mathbb{P}\{\xi_n(\omega) \in [\lambda(\eta), \lambda(\eta) + s]\} \leq Cn^a s^b. \quad (1.8)$$

The above form of conditional regularity of the sample mean for typical conditions is well-adapted to the applications (briefly discussed in Section 1.2 below) which served as the principal motivation for our project.

In the present paper, we further develop our approach from [3, 4] and extend (RCM) to a much larger class of smooth marginal densities. Specifically, we allow the logarithmic derivative of the common density ρ to have power-law singularities due to vanishing of ρ at a finite number of points of its support.

The principal technical result of this paper, Theorem 1, is proved under the following hypothesis:

(V2): *The probability measure μ has bounded support, $\text{supp } \mu = [a, b]$, and admits on the interior (a, b) a strictly*

positive smooth probability density ρ satisfying the following condition:

$$\forall t \in (a, b) \quad |\rho'(t)/\rho(t)| \leq C \max\left[\frac{1}{t-a}, \frac{1}{b-t}\right]. \quad (1.9)$$

Moreover, the method of proof of Theorem 1 provides some explicit estimates for the exponents a and b figuring in (1.8).

Note that the condition (1.10) on the logarithmic derivative covers the all the cases where the density ρ vanishes at one of the edges of its support at power-law rate; the upper bound on $(\ln \rho)'$ actually provides some regularity of the decay at the respective edge, but the latter is a substantially milder condition than, for example, an exact decay asymptotics $\rho(t) \sim C(t-a)^\gamma$ as $t \searrow a$.

We shall see that Theorem 1 naturally extends from the measures satisfying (V2) to a much larger class of measures, obtained from those satisfying (V2) by the operations of (i) shifts, (ii) convolution, and (iii) randomization.

The role of shifts is clarified in Corollary 1.

Corollary 2 evidences that, starting from measures satisfying (V2), one can construct a rich class of measures featuring (RCM), under the condition

(V3): *The probability measure μ on \mathbb{R} is a convolution of the form*

$$\mu = \mu_1 * \mu_2 \quad (1.10)$$

where μ_1 fulfills the condition (V2), and μ_2 is an arbitrary probability measure on \mathbb{R} .

Corollary 3 derives from Theorem 1 property (RCM) for measures which fulfill the following condition:

(V4): *The probability measure μ on \mathbb{R} is obtained by randomization of measures μ_1, \dots, μ_K satisfying (V2), i.e., a random variable X with probability distribution μ admits a decomposition of the form*

$$X(\omega) = \sum_{k=1}^K \mathbf{1}_{\mathcal{A}_k}(\omega) X_k(\omega), \quad (1.11)$$

where the r.v. $\{X_k, 1 \leq k \leq K\}$ are independent and fulfill the condition (V2), and $\mathcal{A}_k = \{\omega : \zeta(\omega) = k\}$, $k = 1, \dots, K$, are the events generated by some integer-valued r.v. $\zeta : \Omega \rightarrow \{1, \dots, K\}$ independent of the family $\{X_k, 1 \leq k \leq K\}$.

1.2 Motivation: random N -particle Hamiltonians

The regularity properties of the conditional distribution of the sample mean, given the fluctuations (η_i) , are interesting in itself, and may prove useful in a fairly broad context of mathematical statistics, but the main motivation for this work actually comes from the spectral theory of multi-particle random Hamiltonians studied in the Anderson [1] localization theory; for an introduction to this relatively new area of mathematical physics of disordered quantum systems one can recommend a recent monograph [5].

It has been shown in [3] that optimal decay bounds on the decay of bound (i.e., square-summable) eigenstates of multi-particle quantum systems with a nontrivial interaction, subject to a common random potential field, require concentration estimates for the eigenvalues of random Hamiltonians

which cannot be derived from the standard, Wegner-type estimates (cf. [7]) used in the single-particle models. More to the point, one needs eigenvalue comparison estimates for pairs of random operators $H_1(\omega)$, $H_2(\omega)$ which are stochastically correlated in a very strong way: the same family of random variables which affects $H_1(\omega)$ also affects $H_2(\omega)$, and vice versa. As a result, no stricto sensu stochastic decoupling is possible in such models. However, the eigenvalue comparison analysis for $H_1(\omega)$ and $H_2(\omega)$ becomes much simpler, once the random field generating the potential in a finite domain (e.g., a finite lattice subset $Q \subset \mathbb{Z}^d$) is decomposed into a sum

$$V(x, \omega) = \xi_Q(\omega) \mathbf{1}_Q(x) + \eta_{Q,x}(\omega), \quad (1.12)$$

where

$$\xi_Q(\omega) = \frac{1}{\text{card } Q} \sum_{x \in Q} V(x, \omega) \quad (1.13)$$

and

$$\eta_{Q,x}(\omega) = V(x, \omega) - \xi_Q(\omega). \quad (1.14)$$

The reason is that after conditioning on the random fluctuation field $\eta_{Q,x}(\omega)$, the random potential $V(x, \omega)$ becomes a sum of a nonrandom background potential $\xi_Q(\omega)$, ignored in the principal estimates, and of a random but spatially constant potential $\xi_Q(\omega) \mathbf{1}_Q(x)$. As an operator of multiplication by a constant, the latter commutes with all other operators involved, and this simplifies considerably the spectral analysis. See the details in [3]. Naturally, the crucial issue is the regularity of the conditional distribution of the sample mean $\xi_Q(\omega)$ given the fluctuations $\eta_{Q,x}(\omega)$. Our results show that this conditional distribution is Hölder continuous, and such a regularity suffices for the methods of spectral theory of random operators; cf., e.g., [5].

2 Main results

Our goal is to analyze the case where the common probability distribution of the IID random variables X_j , $1 \leq j \leq n$, is absolutely continuous, with probability density ρ , and the support $\mathcal{S} = \text{supp } \rho$ of the density is a finite union of intervals:

$$\mathcal{S} \subset \cup_{k=1}^K J_k, \quad J_k = [a_k, b_k], \quad a_{k+1} \geq b_k. \quad (2.1)$$

With the help of the well-known randomization procedure (cf. [6]), namely, by making use of Corollary 3, one can reduce the analysis of densities supported by a union of intervals to the case of a single supporting interval, and this is what we shall do first.

The class of densities ρ supported on an interval $[a, b]$ which do not vanish on its interior (a, b) and have there bounded logarithmic derivative,

$$0 < \rho_* \leq \rho(t) \leq \bar{\rho} < \infty, \quad (2.2)$$

$$|\rho'(t)/\rho(t)| \leq C,$$

was studied in [4], so we focus on the case where ρ vanishes at one or both endpoints of the support. To cover in one argument all possible situations, one can always decompose each interval J where $\rho(a) = \rho(b) = 0$ into the union $[a, c] \cup [c, b]$, $c = (a + b)/2$, so on each of the resulting sub-intervals ρ vanishes at exactly one endpoint. The two cases

are equivalent, since we can make on $[c, b]$ the change of variable $t \mapsto b + c - t$, thus getting a density vanishing at the left edge c . Again, randomization allows one to study separately densities with the supports $[a, c]$ and $[c, b]$.

In the next section, we perform such analysis for a density vanishing at exactly one of the endpoints of the supporting interval. The reader will see, however, that the probabilistic bounds stemming from our analysis become slightly better for densities vanishing at both edges of the support (cf. Remark 2).

Theorem 1. *Assume that the common probability distribution of IID random variables X_1, \dots, X_n admits a probability density ρ , with $\|\rho\|_\infty = \bar{\rho} < +\infty$, satisfying the following conditions:*

(i) *supp ρ is a bounded interval:*

$$\text{supp } \rho = [a, a + \ell]; \quad (2.3)$$

(ii) *ρ vanishes at a and admits the upper bound*

$$\rho(t) \leq C(t - a)^\gamma, \quad \gamma > 0; \quad (2.4)$$

(iii) *the logarithmic derivative of ρ admits the upper bound*

$$\forall t \in (a, a + \ell) \quad |\rho'(t)/\rho(t)| \leq C(t - a)^{-1}. \quad (2.5)$$

Then for any $A > 1$ and $\alpha \in (0, 1)$ there exist constants $C', C'' \in (0, +\infty)$, depending upon A , α , ℓ and the density ρ , such that for any $0 \leq s \leq C'' n^{-(\frac{1}{A-1}\alpha)}$ and any \mathfrak{F}_η -measurable random variable λ , one has, with $I_s(\omega) := [\lambda(\omega), \lambda(\omega) + s]$:

$$\mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \} \leq \min [C \bar{\rho} n s^{\alpha(2+\gamma)} + C' \sqrt{n} s^{1-A\alpha}, 1]. \quad (2.6)$$

The explicit form of the RHS in (2.6) shows that (V2) gives rise to the property (RCM).

2.1 Optimization of the Hölder exponents

In applications to the eigenvalue concentration estimates for random Hamiltonians, one often has $s \leq e^{-n^\beta}$, $\beta > 0$, so that s decays as $n \rightarrow \infty$ much faster than any power-law function $n \mapsto n^{-B}$, $0 < B < \infty$. From this perspective, a pre-factor polynomial in n is essentially negligible compared to s , and the exponent figuring in (RCM) is of much greater importance. To balance the contributions from the two terms in the RHS of (2.6), let us find the optimal value (if it exists) $\alpha = \alpha_\gamma$, as a solution of the equation

$$1 - A\alpha = \alpha(\gamma + 2) \Leftrightarrow \alpha_\gamma = \frac{1}{A + 2 + \gamma}, \quad (2.7)$$

resulting in a simpler probability bound

$$\mathbb{P} \{ \xi \in I_s(\eta) \} \leq C'''(n) s^{1 - \frac{A}{A+2+\gamma}} = C'''(n) s^{\frac{2+\gamma}{A+2+\gamma}} \quad (2.8)$$

with $C'(n), C''(n)$ polynomial in n . Comparing to the optimal bound for the probability densities with regular (bounded) logarithmic derivative from [4],

$$\mathbb{P} \{ \xi \in I_s(\eta) \} \leq C''''(n) s^{2/3}, \quad (2.9)$$

we see that the above exponent $2/3$ can be improved by a judicious choice of A , viz.

$$A > A_\gamma = 1 + \frac{1}{2}\gamma, \tag{2.10}$$

and the closer is A to 1, the better. For example, with $\gamma = 1$ (linear decay of the density at the edges), and $A = 4/3$, we obtain

$$\mathbb{P} \{ \xi \in I_s(\eta) \} \leq C(n)s^{9/13} \ll C(n)s^{2/3}, \tag{2.11}$$

for s small enough.

Furthermore, the optimal value $\alpha = \frac{2+\gamma}{A+2+\gamma}$ for the exponent α figuring in (RCM) approaches 1 as $\gamma \rightarrow +\infty$. In view of Remark 2, the final bound becomes even stronger whenever ρ vanishes at both edges of its support.

This corroborates the intuitive idea that vanishing of the density at some edges of its support should enhance the regularity bound for the conditional distribution of the sample mean.

2.2 Extension to the case of random variables with non-identical expectations

By a simple change of variables, the assertion of Theorem 1 can be adapted to independent random variables with different expectations.

Corollary 1. *Suppose that the random variables X_1, \dots, X_n fulfill the hypotheses of Theorem 1 (hence, they have the property (RCM)). Pick any real numbers a_1, \dots, a_n . Then the random variables $\tilde{X}_i := X_i + a_i, i = 1, \dots, n$, also have the property (RCM).*

Indeed, arbitrary shifts $X_i \mapsto X_i + a_i$ result only in a translation of the sample mean ξ_n , even with conditioning on \mathfrak{F}_η , but the upper bounds on the logarithmic derivative of the conditional density of ξ_n given \mathfrak{F}_η remain unaffected. A direct inspection of the proof of Theorem 1 in Section 5 evidences that this suffices for the main assertion to hold true.

2.3 Sufficiency of the condition (V3)

Corollary 2 ((V3)). *The condition (V3) implies (RCM).*

Proof. Let $\mathfrak{F}^{(2)}$ be the sigma-algebra generated by $X^{(2)}$. The conditional distribution of X given $X^{(2)}$ differs from that of $X^{(1)}$ (i.e., from the measure μ_1) only by a shift (nonrandom, conditional on $\mathfrak{F}^{(2)}$). Thus we have

$$\begin{aligned} \mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \} &= \mathbb{E} [\mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \mid \mathfrak{F}_2 \}] \\ &\leq \text{ess sup} \mathbb{P} \{ \xi_n(\omega) \in I_s(\omega) \mid \mathfrak{F}_2 \} \\ &\leq Cn^\alpha s^b, \end{aligned} \tag{2.12}$$

where the last inequality follows from Theorem 1, since μ_1 fulfills (V2). \square

2.4 Sufficiency of the condition (V4)

Corollary 3 ((V4)). *The condition (V4) implies (RCM).*

Proof. Introduce the sigma-algebra \mathfrak{F}_ζ generated by the random vector $\zeta = \{ \zeta_i, 1 \leq i \leq N \}$. The probability distribution of $\{ X_i, 1 \leq i \leq N \}$ conditional on \mathfrak{F}_ζ is concentrated

on parallelepipeds

$$\mathbf{J}_\mathbf{k} = \times_{i=1}^n J_{k_i}, \quad J_{k_i} = [a_{k_i}, b_{k_i}], \tag{2.13}$$

with random multi-indices $\mathbf{k}(\omega) = (k_1(\omega), \dots, k_n(\omega))$ determined by $\{ \zeta_i, 1 \leq i \leq n \}$.

Let $\mathbb{P}_\mathbf{k} \{ \cdot \}$ be the conditional probability measure, given the event $\{ \zeta(\omega) = \mathbf{k} \}$, $\mathbb{E}_\mathbf{k} [\cdot]$ the respective expectation, and $p_\mathbf{k} = \mathbb{P} \{ \zeta(\omega) = \mathbf{k} \}$. Then we have

$$\begin{aligned} \mathbb{P} \{ \xi_n \in I_s(\eta) \} &= \sum_{\mathbf{k} \in \mathbf{K}} p_\mathbf{k} \mathbb{P}_\mathbf{k} \{ \xi_n \in I_s(\eta) \mid \mathfrak{F}_\eta \} \\ &= \sum_{\mathbf{k} \in \mathbf{K}} p_\mathbf{k} \mathbb{E}_\mathbf{k} [\mathbb{P}_\mathbf{k} \{ \xi_n \in I_s(\eta) \mid \mathfrak{F}_\eta \}] \\ &\leq \max_{\mathbf{k} \in \mathbf{K}} \mathbb{E}_\mathbf{k} [\mathbb{P}_\mathbf{k} \{ \xi_n \in I_s(\eta) \mid \mathfrak{F}_\eta \}], \\ &\leq Cn^\alpha s^b, \end{aligned} \tag{2.14}$$

since $\sum_{\mathbf{k}} p_\mathbf{k} = 1$, each term in the last RHS sum is bounded by $Cn^\alpha s^b$, by virtue of Theorem 1 combined with Corollary 1, and the number of terms is finite. \square

3 Basic geometrical objects and notations

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" η_i relative to the mean:

$$\xi_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad \eta_i = X_i - \xi_n. \tag{3.1}$$

Further, consider the n -dimensional Euclidean space of real linear combinations of the random variables X_i . Clearly, the variables $\eta_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are invariant under the group of translations

$$(X_1, \dots, X_n) \mapsto (X_1 + t, \dots, X_n + t), \quad t \in \mathbb{R}, \tag{3.2}$$

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

$$Y_i = \eta_i - \eta_n, \quad 1 \leq i \leq n-1. \tag{3.3}$$

Then the space \mathbb{R}^n is stratified into a union of affine lines of the form

$$\begin{aligned} \mathcal{L}(Y) &:= \{ X \in \mathbb{R}^n : \eta_i - \eta_n = Y_i, i \leq n-1 \} \\ &= \{ X \in \mathbb{R}^n : X_i - X_n = Y_i, i \leq n-1 \}, \end{aligned} \tag{3.4}$$

labeled by the elements $Y = (Y_1, \dots, Y_{n-1})$ of the $(n-1)$ -dimensional real vector space $\mathbb{Y}^{n-1} \cong \mathbb{R}^{n-1}$ orthogonal to the vector $(1, \dots, 1)$. Denote

$$\begin{aligned} \mathcal{X}(Y) &= \mathcal{L}(Y) \cap [0, \ell]^n \\ &= \{ X \in [0, \ell]^n : X_i - X_n = Y_i, i \leq n-1 \} \end{aligned} \tag{3.5}$$

and equip each nonempty interval $\mathcal{X}(Y) \subset \mathbb{R}^n$ with the 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 the 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 $X \mapsto (\xi_n, \eta_1, \dots, \eta_{n-1})$ 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 , $X \mapsto (\tilde{\xi}_n, \tilde{\eta}_1, \dots, \tilde{\eta}_{n-1})$, such that

$$\tilde{\xi}_n = n^{-1/2} \sum_{i=1}^n X_i = n^{1/2} \xi_n; \quad (3.6)$$

the exact form of $\tilde{\eta}_j$, $j = 1, \dots, n-1$ is of no importance for our analysis, provided that the transformation is orthogonal.

Remark 1. For later use, note that, due to (3.6), 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 $\tilde{\xi}(t) = c_0 + t$, $X_j(t) = c_j + n^{-1/2}t$, with arbitrarily chosen constants c_j . Here, $\tilde{\xi}_n$ is a natural length parameter on $\mathcal{X}(Y)$, since the transformation $X \mapsto (\tilde{\xi}_n, \tilde{\eta}_1, \dots, \tilde{\eta}_{n-1})$ is orthogonal.

4 Probability of short intervals

In this section, we assume that $\text{supp } \rho = [0, \ell]$ and the logarithmic derivative ρ'/ρ is well-defined on the open interval $(0, \ell)$. In addition, we assume a power-law decay of ρ at the edge 0:

$$\rho(t) \leq C_\gamma t^\gamma, \quad \gamma > 0. \quad (4.1)$$

Later, we will complete this upper bound by a certain regularity of the edge decay (cf. (2.5)).

By a change of variable $t \mapsto \ell - t$, the obtained bounds will apply to the densities vanishing at the right edge of the support.

In the following preparation lemma, we use only boundedness of ρ and the decay condition (4.1); smoothness of the density is irrelevant for this intermediate result.

Lemma 1. *Assume that the random variables X_1, \dots, X_n , $n \geq 2$, are IID and admit a bounded density ρ supported by an interval $[0, \ell]$, $\ell > 0$, with $\|\rho\|_\infty = \bar{\rho} < \infty$. Furthermore, assume that ρ vanishes at 0 and satisfies (4.1). Then for all $t \in (0, \ell/2]$ one has*

$$\mathbb{P}\{|\mathcal{X}(Y)| < r\} \leq C_\gamma \bar{\rho} n r^{2+\gamma}. \quad (4.2)$$

Proof. Let

$$\underline{X} = \underline{X}(X) = \min_i X_i, \quad \bar{X} = \bar{X}(X) = \max_i X_i. \quad (4.3)$$

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

According to Remark 1, each variable $n^{1/2}X_i$, $i = 1, \dots, n$, restricted to $\mathcal{X}(Y)$, can serve as a length parameter on $\mathcal{X}(Y)$, compatible with the metric induced by the Euclidean distance in the ambient space \mathbb{R}^n . 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 ℓ (resp.,

strictly positive). Therefore, the maximum increment of X_1 (indeed, of any X_i) along $\mathcal{X}(Y)$ is given by $\ell - \bar{X}(X)$, and its maximum decrement equals $\underline{X}(X)$, so the range of the normalized length parameter $n^{1/2}X_1$ along $\mathcal{X}(Y(X))$ is an interval of length

$$|\mathcal{X}(Y(X))| = n^{1/2}(\ell - \bar{X}(X) + \underline{X}(X)). \quad (4.4)$$

Both $\underline{X}(X)$ and $\ell - \bar{X}(X)$ are non-negative, so one has the implication

$$|\mathcal{X}(Y)| = \underline{X} + (\ell - \bar{X}) < t \implies \max\{\underline{X}, \ell - \bar{X}\} < t. \quad (4.5)$$

With $0 \leq t \leq \ell/2$, $(\ell - X_i < t)$ implies $(X_i > t)$, thus denoting

$$A_{ij}(t) := \{X_i < t\} \cap \{\ell - X_j < t\}, \quad (4.6)$$

we have, for any i ,

$$A_{ii}(t) = \{X_i < t\} \cap \{\ell - X_i < t\} = \emptyset. \quad (4.7)$$

Therefore,

$$\left\{ \max\{\underline{X}, \ell - \bar{X}\} < t \right\} \subset \bigcup_{i \neq j} \{X_i < t, \ell - X_j < t\}. \quad (4.8)$$

Thus the union $\cup_{i \neq j} A_{ij}(t)$ contains all the samples X with $|\mathcal{X}(Y)| < t$.

By hypothesis, for any $i \neq j$, the random variables X_i, X_j are independent, so

$$\mathbb{P}\{A_{ij}(t)\} = \mathbb{P}\{X_i < t\} \cdot \mathbb{P}\{\ell - X_j < t\}. \quad (4.9)$$

Since $\|\rho\|_\infty = \bar{\rho}$, we have

$$\mathbb{P}\{\ell - X_j < t\} \leq \bar{\rho} t, \quad (4.10)$$

and by (4.1),

$$\mathbb{P}\{X_i < t\} \leq \int_0^t C_\gamma s^\gamma ds = \frac{C_\gamma t^{\gamma+1}}{\gamma+1} \leq C_\gamma t^{\gamma+1}, \quad (4.11)$$

hence

$$\mathbb{P}\{A_{ij}(t)\} \leq \bar{\rho} t \cdot C_\gamma t^{1+\gamma} = C_\gamma \bar{\rho} t^{2+\gamma}. \quad (4.12)$$

Consequently,

$$\begin{aligned} & \mathbb{P}\{|\mathcal{X}(Y)| < r\} \\ &= \mathbb{P}\left\{n^{1/2}((\ell - \bar{X}(X)) + \underline{X}(X)) < r\right\} \\ &= \mathbb{P}\left\{((\ell - \bar{X}(X)) + \underline{X}(X)) < rn^{-1/2}\right\} \\ &\leq \sum_{i \neq j} \mathbb{P}\left\{A_{ij}(rn^{-1/2})\right\} \leq n(n-1) \frac{C_\gamma \bar{\rho} t^{2+\gamma}}{(n^{1/2})^2} \\ &\leq C_\gamma \bar{\rho} n r^{2+\gamma}. \end{aligned} \quad (4.13)$$

This completes the proof. \square

Remark 2. As was said in Section 2, in the case where ρ vanishes at both edges of its support, 0 and ℓ , the probabilistic bounds for short intervals become stronger. For example, if $\rho(t) \leq Ct^\gamma$ and $\rho(t) \leq C'(\ell - t)^\gamma$, we can replace (4.10) by a stronger bound $\mathbb{P}\{\ell - X_j < t\} \leq C't^{1+\gamma'}$, and the resulting bound is

$$\mathbb{P}\{|\mathcal{X}(Y)| < r\} \leq C'' n r^{2+\gamma'}. \quad (4.14)$$

5 Regularity bounds for densities with singular logarithmic derivative

In order to assess the conditional regularity of the sample mean on typical (viz., not too short) linear elements $\mathcal{X}(Y)$, we need to complement the upper bound (4.1) with some assumption regarding the regularity of the decay of $\rho(t)$ as $t \searrow 0$. Having in mind first of all applications to realistic physical models of disorder, we could restrict ourselves to the case where $\rho(t) \sim Ct^\gamma$, but it actually turns out that one can treat a more general situation where the logarithmic derivative fulfills the condition (2.5), or even a condition of the form

$$\|(\ln \rho)'\mathbf{1}_{(0,\ell)}\|_\infty \leq \frac{C}{t^B} \quad (5.1)$$

for some $B \in (0, +\infty)$. For brevity, we consider only the case $B = 1$ covering all possible rates $\gamma > 0$ of power-law decay at the edge of the support.

Proof of Theorem 1. Without loss of generality, it suffices to prove the claim in the particular case where $\text{supp } \rho = [0, \ell]$, which we assume below.

Fix some $A > 1$ and $\alpha \in (0, 1)$.

Further, fix an element $\mathcal{X}(Y)$ with $|\mathcal{X}(Y)| \geq s^\alpha$; the probability to have an element $\mathcal{X}(Y)$ with $|\mathcal{X}(Y)| < s^\alpha$ is upper-bounded in (5.15), with the help of Lemma 1.

Next, introduce a length parameter t on $\mathcal{X}(Y)$, so we can identify $\mathcal{X}(Y) \cong \tilde{J} = [0, |\mathcal{X}(Y)|]$. As t runs through \tilde{J} , the i -th coordinate of the point $\mathbf{x}(t) \in \mathcal{X}(Y)$ runs through an interval $\tilde{J}_i = [\tilde{a}_i, \tilde{b}_i]$, with $\tilde{b}_i - \tilde{a}_i = |\mathcal{X}(Y)|n^{-1/2}$.

Recall that it is not the sample mean ξ_n but its rescaled counterpart $\tilde{\xi}_n = n^{1/2}\xi_n$ which gives a normalized length parameter on $\mathcal{X}(Y)$, so we re-write the key probability as follows:

$$\mathbb{P}\{\xi_n(\omega) \in [\lambda(\omega), \lambda(\omega) + s]\} = \mathbb{P}\{\tilde{\xi}_n(\omega) \in \tilde{I}_s(\omega)\} \quad (5.2)$$

where

$$\tilde{I}_s(\omega) = [\tilde{\lambda}(\omega), \tilde{\lambda}(\omega) + s\sqrt{n}], \quad \tilde{\lambda}(\omega) = \sqrt{n}\lambda(\omega), \quad (5.3)$$

so that $|\tilde{I}_s(\omega)| = s\sqrt{n}$.

Denote by $\mathbf{p}(t)$ the product density $\rho(x_1(t)) \cdots \rho(x_n(t))$ at $\mathbf{x}(t) = (x_1(\tilde{b}_1 - t), \dots, x_n(\tilde{b}_n - t)) \in \mathcal{X}(Y)$. We have

$$\mathbf{p}(t) = \prod_{j=1}^n \rho(\tilde{b}_j - t), \quad (5.4)$$

and therefore,

$$\frac{\mathbf{p}(t)}{\mathbf{p}(0)} = \prod_{j=1}^n \frac{\rho(\tilde{b}_j - t)}{\rho(\tilde{b}_j)}. \quad (5.5)$$

The function $t \mapsto \mathbf{p}(t)$ gives the conditional density induced on $\mathcal{X}(Y)$, up to a normalizing factor $1/Z_Y$, with

$$Z_Y = \int_0^{|\mathcal{X}(Y)|} \mathbf{p}(r) dr. \quad (5.6)$$

To upper-bound the maximum of the conditional density, we have to lower-bound the integral Z_Y .

Note that for t small enough, the ratio in the RHS of (5.5) is close to 1; to see this, use the hypothesis

$$|(\ln \rho(x))'| \leq Cx^{-1}. \quad (5.7)$$

Recall that we assumed $|\mathcal{X}(Y)| \geq s^\alpha$, and by hypothesis of the Theorem,

$$s \leq n^{-\frac{1}{(A-1)\alpha}}. \quad (5.8)$$

For $x_j = b_j - t$ with $t \in [0, \frac{1}{2}s^\alpha]$, we have $x_j - a_j \geq \frac{1}{2}s^\alpha$, so $|(\ln \rho(x_j))'| \leq Cs^{-\alpha}$. In particular, for $t \in [0, \epsilon_s]$, $\epsilon_s = s^{A\alpha}$, $A > 1$,

$$\begin{aligned} |\ln \rho(b_j) - \ln \rho(b_j - t)| &= |t| \frac{|\rho'(c_j)|}{\rho(c_j)} \leq \epsilon_s \cdot \frac{C}{\frac{1}{2}s^\alpha} \\ &\leq C_1 s^{\alpha(A-1)} \leq C_2 n^{-1}, \end{aligned} \quad (5.9)$$

Further,

$$\begin{aligned} |\ln \mathbf{p}(\mathbf{b}) - \ln \mathbf{p}(\mathbf{x}(t))| &\leq n \max_j |\ln \rho(b_j) - \ln \rho(b_j - t)| \\ &\leq C_3 n s^{\alpha(A-1)} = O(1), \end{aligned} \quad (5.10)$$

hence for such t ,

$$\left| \frac{\mathbf{p}(t)}{\mathbf{p}(0)} \right|^{\pm 1} = e^{\pm(\ln \mathbf{p}(0) - \ln \mathbf{p}(t))} = e^{O(1)} \geq C_4 > 0. \quad (5.11)$$

Therefore,

$$Z_{\mathbf{k}} \geq C_4 \mathbf{p}(0) \epsilon_s \geq \frac{1}{2} \mathbf{p}(0) s^{A\alpha}, \quad (5.12)$$

so the conditional probability density on the segment $\mathcal{X}(Y)$ is uniformly bounded by an s -dependent constant:

$$\frac{\mathbf{p}(t)}{Z_{\mathbf{k}}} \leq \frac{C_5 \mathbf{p}(0)}{s^{A\alpha} \mathbf{p}(0)} \leq C_5 s^{-A\alpha}, \quad (5.13)$$

yielding

$$\begin{aligned} \int_{I_s(\eta)} \frac{\mathbf{p}(\mathbf{x}(t))}{Z_{\mathbf{k}}} dt &\leq C_5 s^{-A\alpha} \cdot |I_s(\eta)| \\ &= C_5 n^{1/2} s^{1-A\alpha}. \end{aligned} \quad (5.14)$$

Owing to Lemma 1, we know that

$$\mathbb{P}\{|\mathcal{X}(Y)| < s^\alpha\} \leq C_6 \bar{\rho} n s^{\alpha(2+\gamma)}. \quad (5.15)$$

Collecting (5.14) and (5.15), the claim follows. \square

6 Open problems

It seems to be a very natural conjecture that the property (RCM) holds true for probability measures with bounded compactly supported probability density. However, operating without any additional assumption on the regularity of the bounded probability density (apart from its measurability and boundedness) would require new analytic ideas.

A more challenging problem concerns the validity of (RCM) for more general probability measures with Hölder continuous PDF. The author's discussions with a number of experts in probability and functional analysis seem to indicate that the question on regularity of conditional distribution of the sample mean on typical fibers

$$\{(x_1, \dots, x_n) : \eta_i = c_i, i = 1, \dots, n-1\} \subset \mathbb{R}^n \quad (6.1)$$

is far from obvious, when \mathbb{R}^n is endowed with the product measure $\mu^{\otimes n}$ with a Hölder continuous, but possibly purely singular continuous measure μ on \mathbb{R} , and it rises in various problems in mathematical physics of disordered media.

7 Conclusion

We have proved that the regularity properties of the sample mean ξ_n , conditional on the sample "fluctuations" relative to ξ_n , generalizing the well-known property of Gaussian samples and proved earlier for a class of marginal densities with bounded logarithmic derivative, hold true for a much larger class of smooth densities which can vanish at some points in a sufficiently regular way. Moreover, we have shown that vanishing of the marginal density can only result in stronger probabilistic concentration bounds for the values of the sample mean. Intuitively, this seems quite natural, but technically, this was not quite clear from prior works based on global regularity of the logarithmic derivative of the density.

From the point of view of applications, the new result covers a large number of models where the probability distribution comes from physical laws or, more generally, from explicit calculations with elementary functions which are smooth in their domains of definition and usually feature a power-law decay at certain points. In particular, this covers many models of disorder in mathematical physics of multi-particle quantum systems, where it provides a crucial ingredient for the analysis of eigenvalue concentration in presence of disorder and of a nontrivial interaction between particles (cf. [5] and references therein).

REFERENCES

- [1] P. W. Anderson (1958). *Absence of diffusion in certain random lattices*. Phys. Rev., 109, 1492–1505.
- [2] V. Chulaevsky & Y. Suhov (2008). *Wegner bounds for a two-particle tight binding model*. Commun. Math. Phys. 283, 479–489.
- [3] V. Chulaevsky (2011). *On resonances in disordered multi-particle systems*. C. R. Acad. Sci. Paris, Ser. I, 350, 81–85.
- [4] V. Chulaevsky (2015). *On the regularity of the conditional distribution of the sample mean*. Markov Proc. Rel. Fields, 21:3.
- [5] V. Chulaevsky & Y. Suhov (2013). *Multi-scale Analysis for Random Quantum Systems with Interaction*. Boston: Birkhäuser.
- [6] W. Feller (1966). *An Introduction to Probability Theory and its Applications*. New York: Wiley.
- [7] F. Wegner (1981). *Bounds on the density of states in disordered systems*. Z. Phys. B. Condensed Matter, 44, 9–15.