

LARGE DEVIATIONS FOR RANDOM MATRICES

SOURAV CHATTERJEE* AND S. R. S. VARADHAN†

Dedicated to Professor K. R. Parthasarathy on the occasion of his 75th birthday

ABSTRACT. We prove a large deviation result for a random symmetric $n \times n$ matrix with independent identically distributed entries to have a few eigenvalues of size n . If the spectrum \mathcal{S} survives when the matrix is rescaled by a factor of n , it can only be the eigenvalues of a Hilbert-Schmidt kernel $k(x, y)$ on $[0, 1] \times [0, 1]$. The rate function for k is $I(k) = \frac{1}{2} \int h(k(x, y)) dx dy$ where h is the Cramér rate function for the common distribution of the entries that is assumed to have a tail decaying faster than any Gaussian. The large deviation for \mathcal{S} is then obtained by contraction.

1. Introduction

We are interested in the large deviation behavior of random $n \times n$ symmetric matrices $X(\omega) = \{x_{i,j}(\omega)\}$ where the entries for $j \geq i$ are independent identically distributed real random variables having a common distribution μ . Let $\Lambda_n(\omega) = \{\lambda_j^n(\omega); 1 \leq j \leq n\}$ be the set of n real eigenvalues of this random matrix. If we assume that the mean is 0 and the variance is equal to σ^2 , according to a theorem of Wigner [6], when divided by \sqrt{n} , the empirical distribution of these eigenvalues converges in probability to the semi-circle law, i.e. for any bounded continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f\left(\frac{\lambda_j(\omega)}{\sqrt{n}}\right) = \int f(y) \phi_\sigma(y) dy$$

in probability, where

$$\phi_\sigma(y) = \begin{cases} 0 & \text{if } |x| \geq 2\sigma, \\ \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2} & \text{if } |x| \leq 2\sigma. \end{cases}$$

In fact it is known [1] that under suitable additional assumptions, for any $\epsilon > 0$,

$$\mathbb{P}\left[\max_{1 \leq j \leq n} |\lambda_j^n| \geq (2\sigma + \epsilon)\sqrt{n}\right] \rightarrow 0.$$

Received 2011-6-23; Communicated by K. B. Sinha.

2000 *Mathematics Subject Classification.* 60B20; 15B52.

Key words and phrases. Large deviations, eigenvalues, random matrices.

* Sourav Chatterjee's research was partially supported by NSF grants DMS-0707054 and DMS-1005312, and a Sloan Research Fellowship.

† S. R. S. Varadhan's research was partially supported by NSF grants DMS-0904701 and OISE-0730136.

Large deviation results for the distribution of eigenvalues to be different from the Wigner distribution has been considered by Ben Arous and Guionnet [2]. If we decide to divide by n rather than \sqrt{n} , and denote the resulting spectrum by $S_n(\omega) = \frac{1}{n}\Lambda_n(\omega)$, then

$$\mathbb{P}[\sup_{\lambda \in S_n} |\lambda| \geq \epsilon] \rightarrow 0.$$

If we drop the assumption that the mean is 0, then there will be one large eigenvalue λ_{max} of size n in Λ_n , with $\frac{\lambda_{max}}{n} \simeq \int x d\mu$ and the remaining eigenvalues will follow the semi-circle law as before when divided by \sqrt{n} . It is also possible to consider scaling by n^α for $\frac{1}{2} < \alpha < 1$, but they would likely behave like moderate deviations and exhibit Gaussian behavior with rate $n^{2\alpha}$.

Let \mathcal{S} denote the set of all finite or countable collections $S = \{\lambda_j\}$ of positive as well as negative real numbers, with repetitions allowed, that have the property

$$\sum_{\lambda \in S} |\lambda|^2 = \sum_j |\lambda_j|^2 < \infty.$$

They are all possible spectra of self adjoint Hilbert-Schmidt operators. The topology on \mathcal{S} will be ordinary convergence as real numbers of the corresponding eigenvalues outside any arbitrarily small interval around 0. Equivalently, it is the minimal topology such that for any bounded continuous function f that is 0 in some interval around 0, the sum $\sum_{\lambda \in S} f(\lambda)$ is continuous as a map of $\mathcal{S} \rightarrow \mathbb{R}$. One can easily construct a metric for this topology. We enumerate separately the positive and negative values in S in decreasing order of their absolute values as $\{u_j^+\}, \{u_j^-\}$. If one or both of them is only a finite set or empty we augment it by adding 0's. If two points S, S' in \mathcal{S} are enumerated as $\{u_j^+\}, \{u_j^-\}$ and $\{v_j^+\}, \{v_j^-\}$, we define the distance

$$d(S, S') := \sum_{j=1}^{\infty} \frac{|u_j^+ - v_j^+|}{2^{j+1}} + \sum_{j=1}^{\infty} \frac{|u_j^- - v_j^-|}{2^{j+1}}.$$

The random matrix yields a random spectrum and after normalization by n , we obtain a random element of \mathcal{S} . Its distribution yields a sequence $\{P_n\}$ of probability measures on \mathcal{S} . We will prove a large deviation result for P_n with rate n^2 . We assume that for all $\theta > 0$,

$$\int \exp[\theta x^2] \mu(dx) < \infty. \quad (1.1)$$

In particular this implies that the moment generating function $M(\theta) = \int e^{\theta x} \mu(dx)$ is finite for all θ and satisfies

$$\limsup_{|\theta| \rightarrow \infty} \frac{1}{\theta^2} \log M(\theta) = 0$$

and the conjugate rate function of Cramér

$$h(x) = \sup_{\theta} [\theta x - \log M(\theta)] \quad (1.2)$$

satisfies

$$\liminf_{|x| \rightarrow \infty} \frac{h(x)}{x^2} = +\infty.$$

The condition (1.1) is important, because an eigenvalue of size n can be produced by a single entry of size n in the random matrix and we would like this to have probability that is super-exponentially small in the scale n^2 .

The random symmetric $n \times n$ matrix $X(\omega) = \{x_{i,j}(\omega)\}$ is first mapped into a symmetric kernel $k(x, y, \omega)$

$$k(x, y, \omega) = \sum_{i,j=1}^n x_{i,j}(\omega) \mathbf{1}_{J_i^n}(x) \mathbf{1}_{J_j^n}(y) \quad (1.3)$$

where J_i^n is the interval $[\frac{i-1}{n}, \frac{i}{n}]$. This induces a family of probability measures Q_n on the space of symmetric kernels $k(x, y)$ on $D = [0, 1] \times [0, 1]$. We will restrict ourselves to $\mathcal{K} = \{k : \int \int_D |k(x, y)|^2 dx dy < \infty\}$. Then $k(x, y)$ defines on $L^2[0, 1]$ a Hilbert-Schmidt operator which has a countable spectrum with 0 as the only limit point. Actually for each fixed n the range of the map $X \rightarrow k$ is a finite dimensional subspace of simple functions $\mathcal{K}_n \subset \mathcal{K}$. The nonzero spectrum of $k(x, y, \omega)$ is the same as that of $\{\frac{x_{i,j}}{n}\}$ and we can obtain P_n from Q_n through the natural map $\mathcal{K} \rightarrow \mathcal{S}$ that takes any k to its set of eigenvalues. We define on \mathcal{K} the following rate function

$$I(k(\cdot, \cdot)) = \frac{1}{2} \int_0^1 \int_0^1 h(k(x, y)) dx dy \quad (1.4)$$

with h given by (1.2).

Any permutation $\sigma \in \Pi(n)$ of the rows and columns of X , mapping $\{x_{i,j}\} \rightarrow \{x_{\sigma(i), \sigma(j)}\}$ leaves the set of eigen-values of X invariant and the group G of measure preserving transformations σ of $[0, 1]$ onto itself lifts to an action on \mathcal{K} mapping $k \rightarrow \sigma k$ where $\sigma k(x, y) := k(\sigma x, \sigma y)$. The map $k \rightarrow \sigma k$ leaves $I(k(\cdot, \cdot))$ as well as the spectrum $S(k)$ of k invariant. For establishing the large deviations of P_n on \mathcal{S} it is therefore enough to prove a large deviation principle for the images \tilde{Q}_n of Q_n on $\tilde{\mathcal{K}} = \mathcal{K}/G$.

We will be working with the space \mathcal{K} of symmetric kernels $k(x, y)$ on D . If $I(k) < \infty$, then the operator defined by $k(x, y)$ on $L^2[0, 1]$ is Hilbert-Schmidt and has a countable spectrum with 0 as the only limit point. We need to show some sort of continuity of the map $\tilde{\mathcal{K}} \rightarrow \mathcal{S}$ mapping the G -orbit \tilde{k} of k to its spectrum $S(k)$, in order to transfer the large deviation result from $\tilde{\mathcal{K}}$ to \mathcal{S} . This requires a topology on $\tilde{\mathcal{K}}$ that will be inherited from \mathcal{K} . The weak topology on \mathcal{K} turns out to be too weak and the strong or L^1 topology too strong. What works is the topology induced by the cut metric

$$d_{\square}(k_1, k_2) = \sup_{\substack{|\phi| \leq 1 \\ |\psi| \leq 1}} \left| \int (k_1(x, y) - k_2(x, y)) \phi(x) \psi(y) dx dy \right|, \quad (1.5)$$

ϕ, ψ being Borel measurable functions on $[0, 1]$. Equivalently,

$$d_{\square}(k_1, k_2) = \sup_{A \times B} \left| \int_{A \times B} (k_1(x, y) - k_2(x, y)) dx dy \right|, \quad (1.6)$$

where the supremum is taken over all Borel subsets A, B of $[0, 1]$. The induced metric on $\tilde{\mathcal{K}}$ is

$$d_{\square}(k_1, k_2) = \inf_{\sigma} d_{\square}(\sigma k_1, k_2) = \inf_{\sigma} d_{\square}(k_1, \sigma k_2) = \inf_{\sigma_1, \sigma_2} d_{\square}(\sigma_1 k_1, \sigma_2 k_2)$$

where $\sigma k(x, y) = k(\sigma x, \sigma y)$. We can define on \mathcal{S} the rate function

$$J(S) = \inf_{k: S(k)=S} I(k).$$

Our main result is the following.

Theorem 1.1. *Under assumption (1.1) the sequence of measures P_n on \mathcal{S} satisfies a large deviation property with rate function $J(S)$, i.e for closed $C \subset \mathcal{S}$*

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(C) \leq - \inf_{S \in C} J(S)$$

and for $U \subset \mathcal{S}$ that are open

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(U) \geq - \inf_{S \in U} J(S).$$

This is based on a large deviation principle for \tilde{Q}_n on $\tilde{\mathcal{K}}$ in the cut topology with rate function $I(\tilde{k})$.

Theorem 1.2. *Under assumption (1.1) the sequence of measures \tilde{Q}_n on $\tilde{\mathcal{K}}$ satisfies a large deviation property with rate function $I(\tilde{k})$, i.e for closed $C \subset \tilde{\mathcal{K}}$*

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{Q}_n(C) \leq - \inf_{\tilde{k} \in C} I(\tilde{k})$$

and for $U \subset \tilde{\mathcal{K}}$ that are open

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{Q}_n(U) \geq - \inf_{\tilde{k} \in U} I(\tilde{k}).$$

To make the connection we need the map $k \rightarrow S(k)$ to be continuous in the cut topology. It is valid, provided we restrict it to sets \mathcal{K}^{ℓ} of the form $\mathcal{K}^{\ell} = \{k : |k(x, y)| \leq \ell\}$ for some $\ell < \infty$. This requires truncation at level ℓ and then removing the cut-off. Condition (1.1) provides super-exponential bounds in the Hilbert-Schmidt norm for the error due to cutoff and that is used to complete the proof.

Similar methods were used in [3] to study the large deviation behavior of the number of triangles or other finite subgraphs in random graphs as the number of vertices goes to ∞ but the probability of an edge being connected remains fixed at some $p > 0$. This will correspond to each $x_{i,j}$ taking the values 0 or 1.

2. Some Useful Lemmas.

We will be working with partitions of the unit interval $[0, 1]$ into a finite disjoint union of subintervals. We will not worry about the end points. We can adopt any convention that makes it a true partition. For each partition \mathcal{P} of the unit interval into m subintervals $\{J_i\}$ there is a corresponding partition of D into m^2 sub-squares $\{J_i \times J_j\}$.

For each integer m we have the special partition \mathcal{P}_m of the unit interval into m equal subintervals and they will be denoted by $J_i^m = [\frac{i-1}{m}, \frac{i}{m}]$. We denote by $\mathcal{K}_m \subset \mathcal{K}$ the space of symmetric kernels of the form

$$f(x, y) = \sum_{i,j} f_{i,j} \mathbf{1}_{J_i^m}(x) \mathbf{1}_{J_j^m}(y).$$

This provides a faithful representation of symmetric matrices of size $m \times m$ as elements of \mathcal{K}_m . If $|f_{i,j}| \leq \ell$, then the corresponding $f \in \mathcal{K}_m^\ell = \mathcal{K}^\ell \cap \mathcal{K}_m$. The following is a simple, but useful lemma.

Lemma 2.1. *If $f(x, y) = \sum f_{i,j} \mathbf{1}_{J_i}(x) \mathbf{1}_{J_j}(y)$ and $g(x, y) = \sum g_{i,j} \mathbf{1}_{J_i}(x) \mathbf{1}_{J_j}(y)$ are both simple functions with respect to the same partition J_1, \dots, J_m of $[0, 1]$, then in calculating the distance*

$$d_{\square}(f, g) = \sup_{A \times B} \left| \int_{A \times B} (f(x, y) - g(x, y)) dx dy \right|$$

the sets A, B can be restricted to sets of the form $\bigcup_{i \in \mathcal{N}} J_i$ where $\mathcal{N} \subset \{1, 2, \dots, m\}$.

Proof. First note that

$$\begin{aligned} \int_{A \times B} (f(x, y) - g(x, y)) dx dy &= \sum_{i,j} \int_{A \cap J_i \times B \cap J_j} (f(x, y) - g(x, y)) dx dy \\ &= \sum_{i,j} |A \cap J_i| |B \cap J_j| (f_{i,j} - g_{i,j}) \\ &= \sum_{i,j} a_i b_j |J_i| |J_j| (f_{i,j} - g_{i,j}), \end{aligned}$$

where $a_i = \frac{|A \cap J_i|}{|J_i|}$ and $b_i = \frac{|B \cap J_i|}{|J_i|}$. It is now clear that the supremum of the absolute value is achieved when each a_i and b_j is either 0 or 1. \square

Remark 2.2. If f and g are defined in terms of two different partitions of $[0, 1]$, they can both be viewed as defined with respect to the finite partition which is their common refinement.

Since our large deviation result on \mathcal{S} is deduced from a large deviation result on \mathcal{K} , we need some continuity property of the map $k \rightarrow S(k)$ of $\mathcal{K} \rightarrow \mathcal{S}$.

Lemma 2.3. *For any $\ell < \infty$, the map $k \rightarrow S(k)$ from $\mathcal{K} \rightarrow \mathcal{S}$ is continuous in the cut topology when restricted to $\mathcal{K}^\ell = \{k : |k(x, y)| \leq \ell\}$.*

Proof. If $f(\lambda) = 0$ near 0 then $\lambda^{-3} f(\lambda)$ can be approximated uniformly on $[-\ell, \ell]$ by a polynomial in λ . Therefore $f(\lambda)$ is approximated by a polynomial involving only powers λ^m for $m \geq 3$. Since we have a bound on $\sum_{\lambda \in \mathcal{S}} |\lambda|^2$ it is enough to show that

$$k \rightarrow \sum_{\lambda \in \mathcal{S}(k)} \lambda^m$$

are continuous maps of $\mathcal{K}^\ell \rightarrow \mathcal{S}$ for each $m \geq 3$. It is elementary to check that since $|k(x, y)| \leq \ell$, the maps

$$k \rightarrow \int_{[0,1]^m} k(x_1, x_2) \cdots k(x_{m-1}, x_m) k(x_m, x_1) dx_1 \cdots dx_m = \sum_{\lambda \in \mathcal{S}(k)} \lambda^m$$

are continuous maps in the cut topology from $\mathcal{K}^\ell \rightarrow \mathcal{S}$ provided $m \geq 3$. We start with

$$\int_{[0,1]^m} k_n(x_1, x_2) \cdots k_n(x_{m-1}, x_m) k_n(x_m, x_1) dx_1 \dots dx_m$$

which can be written as

$$\int_{[0,1]^{m-2}} dx_3 \dots dx_m \int_{[0,1]^2} k_n(x_1, x_2) \phi_n(x_1, x_3, \dots, x_n) \psi_n(x_2, x_3, \dots, x_n) dx_1 dx_2$$

where ϕ_n and ψ_n are uniformly bounded. If $d_\square(k_n, k) \rightarrow 0$, we can then replace $k_n(x_1, x_2)$ by $k(x_1, x_2)$. This is repeated for each factor. \square

We will also need the following lemmas: a multicolor version of Szemerédi's regularity lemma for graphs that can be found in [4] and its consequence.

Lemma 2.4. *Given any $\epsilon > 0$ and integers r and m , there exists M and n_0 such that if the edges of a graph \mathcal{G}_n of size $n \geq n_0$ are colored with any one of r colors, then the vertex set can be partitioned into sets V_0, \dots, V_p for some p in the range $m \leq p \leq M$ so that $|V_0| \leq \epsilon n$, $|V_1| = |V_2| = \dots = |V_p| = K$ and all but at most ϵp^2 pairs (V_i, V_j) satisfy the following regularity condition. For any $X \subset V_i, Y \subset V_j$ with $i, j \geq 1$ and $|X|, |Y| \geq \epsilon K$ we have*

$$|d_\nu(X, Y) - d_\nu(V_i, V_j)| < \epsilon,$$

where $d_\nu(X, Y)$ is the proportion of edges between X and Y that are colored with color ν .

Lemma 2.5. *For any $\ell < \infty$ and $\epsilon > 0$, there is a compact set of simple functions $W_{\ell, \epsilon} \subset \mathcal{K}$ and $n_0(\epsilon, \ell)$ such that for $n \geq n_0$ and any $k \in \mathcal{K}_n^\ell$, there exists $f \in W_{\ell, \epsilon}$ and $\sigma \in \Pi(n)$ such that*

$$d_\square(\sigma k, f) < \epsilon.$$

Proof. Let $\epsilon > 0$ be given. Let the integer r be chosen such that $\frac{\ell}{r} < \frac{\epsilon}{2}$. Let $\epsilon' = \frac{\epsilon}{12r\ell}$ and $m = \frac{1}{\epsilon'}$. Let $x_{i,j}$ be the value of k in the rectangle $J_i^n \times J_j^n$. The interval $[-\ell, \ell]$ is divided into r disjoint equal intervals of length $\frac{2\ell}{r}$ and the edge (i, j) is colored according to the interval into which $x_{i,j}$ falls. The 'color' is defined as the value of the mid point of the interval. The color of the edge (i, j) is then a real number $y_{i,j}$ which can equal any one from the finite set z_1, z_2, \dots, z_r and $|y_{i,j} - x_{i,j}| \leq \frac{\epsilon}{2}$. We apply the multicolor version of Szemerédi's regularity theorem (Lemma 2.4) with parameters (ϵ', m, r) to obtain a partition V_0, V_1, \dots, V_p of $\{1, 2, \dots, n\}$ with the following properties. $|V_0| = K' \leq \epsilon' n$. $|V_1| = |V_2| = \dots = |V_p| = K$. If $a_{s,i,j}$ be the proportion of edges between V_i and V_j that have color z_s , then for all but $\epsilon' p^2$ pairs (V_i, V_j) , $i, j \geq 1$ $i \neq j$, for any two subsets $X \subset V_i, Y \subset V_j$ with $|X| \geq \epsilon' K, |Y| \geq \epsilon' K$ the proportion $b_{s,i,j}$ of edges of color z_s , between X and Y satisfies

$$\sup_s |b_{s,i,j} - a_{s,i,j}| \leq \epsilon'.$$

Let us divide the unit interval into subintervals J_0, \dots, J_p where $J_0 = [0, \frac{K'}{n}]$ and for $i \geq 1$, $J_i = [\frac{K'+(i-1)K}{n}, \frac{K'+iK}{n}]$. We construct a function $f \in \mathcal{K}^\ell$ as

$$f(x, y) = \sum_{s=1}^r z_s f(s, x, y),$$

where

$$f(s, x, y) = \sum_{i,j=0}^p a_{s,i,j} \mathbf{1}_{J_i}(x) \mathbf{1}_{J_j}(y).$$

For fixed ℓ and ϵ as long as p remains bounded such functions vary over a compact subset of $L^1(D)$.

For the permutation $\sigma \in \Pi(n)$ of the vertices, we rearrange the order of the vertices, so that those in V_0 corresponds to the first K' indices and for $1 \leq i \leq p$, those in V_i correspond respectively to indices in the range $(K' + (i-1)K + 1, K' + iK)$. We will denote by k' the image of $\{x_{\sigma(i), \sigma(j)}\}$ in \mathcal{K}_n^ℓ . We define $c(s, i, j) = 1$ if $x_{\sigma(i), \sigma(j)}$ belongs to the interval with mid point z_s . Otherwise it is 0. With

$$k'(s, x, y) = \sum_{i,j=1}^n c(s, i, j) J_i^n(x) J_j^n(y)$$

and

$$k'(x, y) = \sum_{i,j=1}^n x_{\sigma(i), \sigma(j)} J_i^n(x) J_j^n(y)$$

we have that for each x, y ,

$$\left| k'(x, y) - \sum_s z_s k'(s, x, y) \right| \leq \frac{\epsilon}{2}.$$

We need to estimate

$$\left| \int_{A \times B} [k'(x, y) - f(x, y)] dx dy \right| \leq \frac{\epsilon}{2} + r\ell \sup_{1 \leq s \leq r} \left| \int_{A \times B} [k'(s, x, y) - f(s, x, y)] dx dy \right|.$$

For each value of s , we will estimate

$$\int_{A \times B} [k'(s, x, y) - f(s, x, y)] dx dy = \sum_{i,j=0}^p \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy.$$

The summation over (i, j) will be split into several groups. $F_0 = \{(i, i) : i \geq 1\}$. $F_1 = \{(i, j) : i = 0\} \cup \{(i, j) : j = 0\}$. $F_2 = \{(i, j)\}$ is the collection of at most $\epsilon' p^2$ exceptional pairs from Lemma 2.4. $F_3 = \{(i, j)\}$ for which either $|A \cap J_i| \leq \epsilon' |J_i|$ or $|B \cap J_j| \leq \epsilon' |J_j|$. F_4 will be the rest. Then:

$$\left| \sum_{(i,j) \in F_0} \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy \right| \leq \frac{pK^2}{n^2} \leq \frac{1}{p} \leq \epsilon',$$

$$\left| \sum_{(i,j) \in F_1} \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy \right| \leq 2\epsilon',$$

$$\left| \sum_{(i,j) \in F_2} \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy \right| \leq \epsilon' p^2 |J_i| |J_j| \leq \epsilon' p^2 \frac{1}{p^2} = \epsilon',$$

$$\left| \sum_{(i,j) \in F_3} \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy \right| \leq \epsilon' \sum_{i,j} |J_i| |J_j| \leq \epsilon'.$$

Finally in the remaining set F_4 , since $|A \cap J_i| \geq \epsilon' |J_i|$ and $|B \cap J_j| \geq \epsilon' |J_j|$, it follows that

$$\left| \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy \right| \leq \epsilon' |J_i| |J_j|$$

and hence

$$\left| \sum_{(i,j) \in F_4} \int_{A \cap J_i \times B \cap J_j} [k'(s, x, y) - f(s, x, y)] dx dy \right| \leq \epsilon' \sum_{i,j} |J_i| |J_j| \leq \epsilon'.$$

Adding them up gives

$$\left| \int_{A \times B} [k'(x, y) - f(x, y)] dx dy \right| \leq \frac{\epsilon}{2} + 6r\ell\epsilon' \leq \epsilon.$$

□

3. Lower Bound

The lower bound for \tilde{Q}_n on $\tilde{\mathcal{K}}$ can be proved by proving a lower bound for Q_n on \mathcal{K} and it can be done without truncation and under the (weaker) assumption that $\int e^{\theta x^2} d\mu(x) < \infty$ for some θ . This implies a lower bound $h(x) \geq c|x|^2$ when $|x|$ is large.

Theorem 3.1. *Under the weaker assumption that for some $\theta > 0$,*

$$\int e^{\theta x^2} d\mu(x) < \infty,$$

it follows that for any $f \in \mathcal{K}$ such that

$$I(f) = \frac{1}{2} \int h(f(x, y)) dx dy < \infty$$

and for any $\delta > 0$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[d_{\square}(k, f) < \delta] \geq -I(f).$$

Proof. Since h is a convex function of its argument, for any integer q we can replace f by a simple function g of its averages over $J_r^q \times J_s^q$ so that

$$g(x, y) = \sum_{r,s=1}^q g_{r,s} \mathbf{1}_{J_r^q}(x) \mathbf{1}_{J_s^q}(y),$$

where

$$g_{r,s} = q^2 \int_{J_r^q \times J_s^q} f(x, y) dx dy.$$

For any q , $I(g) \leq I(f)$ and for large q , $d_{\square}(f, g) < \frac{\delta}{2}$. It suffices to show that

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[d_{\square}(g, k) \leq \frac{\delta}{2}] \geq -I(g).$$

We have

$$k(x, y) = \sum_{i,j=1}^n x_{i,j} \mathbf{1}_{J_i^n}(x) \mathbf{1}_{J_j^n}(y)$$

and we can view both k and g as members of \mathcal{K}_{nq} . The distance $d_{\square}(g, k)$ can be computed as

$$d_{\square}(g, k) = \sup_{A,B} \left| \int_{A \times B} (g(x, y) - k(x, y)) dx dy \right|,$$

where A and B are taken to be unions of sub-collections of intervals of the form $\{[\frac{i-1}{nq}, \frac{i}{nq}]\}$. There are exactly $2^{nq} \times 2^{nq}$ such pairs A, B . We now tilt the measure so that $\{x_{i,j}\}$ remains symmetric and $\{x_{i,j}\}$ for $j \geq i$ are still independent but the distribution of $x_{i,j}$ is tilted from μ to $\mu_{i,j}$ given by

$$\mu_{i,j}(dx) = \frac{1}{M(\theta_{r,s})} \exp[\theta_{r,s} x] \mu(dx)$$

for $(r-1)n < iq \leq rn$ and $(s-1)n < jq \leq sn$ where $\theta_{r,s} = h'(g_{r,s})$, or equivalently $g_{r,s} = \frac{M'(\theta_{r,s})}{M(\theta_{r,s})}$ for $1 \leq r, s \leq q$. Let Q_n^g be the law of the new k . The law of large numbers applies to each such pair A, B with uniform (in A and B) exponential error bounds of $e^{-cn^2+o(n^2)}$ for some $c > 0$. Therefore,

$$Q_n^g \left[\left| \int_{A \times B} (g(x, y) - k(x, y)) dx dy \right| \geq \frac{\delta}{2} \right] \leq e^{-c(\delta)n^2+o(n^2)}.$$

Since $2^{2nq} \ll e^{cn^2}$, it follows that

$$Q_n^g[d_{\square}(g, k) \geq \frac{\delta}{2}] \rightarrow 0.$$

The relative entropy of Q_n^g with respect to Q_n is easily computed to be $n^2 I(g)$.

The following entropy lower bound using Jensen's inequality, establishes the large deviation lower bound. Suppose α, β are two probability measures and $\beta \ll \alpha$ with $H = \int \phi \log \phi d\alpha < \infty$ where $\phi = \frac{d\beta}{d\alpha}$. Since $y \log y \geq -e^{-1}$ for all $y \geq 0$, $\int |\phi \log \phi| d\alpha \leq H + 2e^{-1}$. Therefore

$$\begin{aligned} \alpha(A) &\geq \int_A \phi^{-1} d\beta = \int_A \exp[-\log \phi] d\beta \\ &= \beta(A) \frac{1}{\beta(A)} \int_A \exp[-\log \phi] d\beta \\ &\geq \beta(A) \exp \left[-\frac{1}{\beta(A)} \int_A \phi \log \phi d\alpha \right] \\ &\geq \beta(A) \exp \left[-\frac{1}{\beta(A)} \int |\phi \log \phi| d\alpha \right] \\ &\geq \beta(A) \exp \left[-\frac{1}{\beta(A)} [H + 2e^{-1}] \right]. \end{aligned}$$

Taking $\alpha = Q_n$ and $\beta = Q_n^g$ and $A = \{k : d_{\square}(k, g) \leq \frac{\delta}{2}\}$ we have $H = n^2 I(g)$ and $\beta(A) \simeq 1$. \square

4. Upper Bound

We assume that $|x_{i,j}| \leq \ell$. According to a result of Lovász and Szegedy [5], for any finite ℓ , the set $\tilde{\mathcal{K}}^\ell = \{\tilde{f} : \sup_{x,y} |f(x,y)| \leq \ell\}$ is compact in \tilde{K} . Therefore in order to prove the large deviation upper bound for \tilde{Q}_n with rate function $I(f)$ it is sufficient to prove the local version of the upper bound.

Theorem 4.1. *Let $\tilde{f} \in \tilde{\mathcal{K}}$. Then*

$$\limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{Q}_n[\tilde{k} : d_{\square}(\tilde{k}, \tilde{f}) < \epsilon] \leq -I(\tilde{f}).$$

Proof. The theorem is proved in several steps. Closed balls $B(f, \epsilon)$ in \mathcal{K} of the form $\{k : d_{\square}(k, f) \leq \epsilon\}$ are weakly closed. It is not hard to prove that

$$\limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[k : d_{\square}(k, f) \leq \epsilon] \leq -I(f). \quad (4.1)$$

The argument goes as follows. Assume $I(f) < \infty$. Given $\delta > 0$ pick a nice g such that

$$\frac{1}{2} \left[\int f(x, y) g(x, y) dx dy - \int \log \left[\int e^{z g(x, y)} \mu(dz) \right] dx dy \right] \geq I(f) - \delta$$

and apply Cramér type estimate using the moment generating function for the half space

$$H_{f, g, \epsilon} = \left\{ k : \langle k, g \rangle \geq \inf_{k' \in B(f, \epsilon)} \langle k', g \rangle \right\}$$

which contains $B(f, \epsilon)$. This gives

$$Q_n[B(f, \epsilon)] \leq Q_n[H_{f, g, \epsilon}] \leq \exp \left[-\frac{n^2}{2} \inf_{k \in B(f, \epsilon)} \langle k, g \rangle \right] \mathbb{E}^{Q_n} \exp \left[\frac{n^2}{2} \langle k, g \rangle \right].$$

It is easy to see (because $x_{i,j} = x_{j,i}$) that

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}^{Q_n} \exp \left[\frac{n^2}{2} \langle k, g \rangle \right] = \frac{1}{2} \int_D \log \left[\int \exp[zg(x, y)] \mu(dz) \right] dx dy$$

and

$$\lim_{\epsilon \rightarrow 0} \inf_{k \in B(f, \epsilon)} \langle k, g \rangle = \langle f, g \rangle.$$

We can let $\delta \rightarrow 0$ at the end. If $I(f) = \infty$, pick g such that

$$\frac{1}{2} \left[\int f(x, y) g(x, y) dx dy - \int \log \left[\int e^{z g(x, y)} \mu(dz) \right] dx dy \right] \geq L$$

and let $L \rightarrow \infty$ in the end.

This proves (4.1). But to prove it for balls in $\tilde{\mathcal{K}}$ we need to estimate the probability of the pre-image in \mathcal{K} of a ball in $\tilde{\mathcal{K}}$. This amounts to showing that

$$\limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[k : \inf_{\sigma \in G} d_{\square}(k, \sigma f) < \epsilon] \leq -I(\tilde{f}).$$

We start with the set

$$U(\tilde{f}, \epsilon) = \cup_{\sigma \in G} B(\sigma f, \epsilon).$$

We saw that for each $f \in \mathcal{K}$, (4.1) holds. We need to replace the union over $\sigma \in G$ by a union over a finite collection. According to Lemma 2.5 there is a compact set $W_{\ell, \epsilon} \subset \mathcal{K}$ such that its orbit by $\Pi(n)$ nearly covers \mathcal{K}_n^ℓ for all sufficiently large n . More precisely for $n \geq n_0(\epsilon, \ell)$,

$$\sup_{k \in \mathcal{K}_n^\ell} \inf_{\sigma \in \Pi(n)} \inf_{g \in W_{\ell, \epsilon}} d_{\square}(\sigma g, k) < \epsilon.$$

$W_{\ell, \epsilon}$ is compact in the cut topology (and even in L^1) and can be covered by the union of balls $B(g, \epsilon)$ of radius ϵ centered around g from a finite collection $F_{\ell, \epsilon}$. Therefore for sufficiently large n

$$\cup_{\sigma \in \Pi(n)} \cup_{g \in F_{\ell, \epsilon}} B(\sigma g, 2\epsilon) \supset \mathcal{K}_n^\ell.$$

We need to estimate the probability under Q_n of

$$\cup_{\sigma' \in \Pi(n)} \cup_{g \in F_{\ell, \epsilon}} \cup_{\sigma \in G} [B(\sigma' g, 2\epsilon) \cap B(\sigma f, \epsilon)]$$

which is at most $n!$ times

$$\sup_{\sigma' \in \Pi(n)} Q_n[\cup_{g \in F_{\ell, \epsilon}} \cup_{\sigma \in G} [B(\sigma' g, 2\epsilon) \cap B(\sigma f, \epsilon)]].$$

Since $F_{\ell, \epsilon}$ is a finite set independent of n and Q_n is invariant under $\Pi(n) \subset G$, it is enough to show that for each $g \in F_{\ell, \epsilon}$

$$\limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[B(g, 2\epsilon) \cap \cup_{\sigma \in G} B(\sigma f, \epsilon)] \leq -I(f).$$

If the intersection is nonempty, then there is a σ' such that $B(g, 2\epsilon) \subset B(\sigma' f, 5\epsilon)$. Thus, from (4.1), the invariance of $I(\sigma f)$ under $\sigma \in G$ and the lower semicontinuity of $I(\cdot)$ on $\tilde{\mathcal{K}}$ we see that there is a $\sigma' \in G$ such that

$$\begin{aligned} & \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[B(g, 2\epsilon) \cap \cup_{\sigma \in G} B(\sigma f, \epsilon)] \\ & \leq \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[B(\sigma' f, 5\epsilon)] \\ & = -I(\sigma' f) = -I(f). \end{aligned}$$

□

5. Truncation

Given $X = \{x_{i,j}\}$ we truncate it at level ℓ . Let $x = f_\ell(x) + g_\ell(x)$ where

$$f_\ell(x) = \begin{cases} x & \text{if } |x| \leq \ell \\ \ell & \text{if } x \geq \ell \\ -\ell & \text{if } x \leq -\ell \end{cases}$$

and $g_\ell(x) = x - f_\ell(x)$. There is a corresponding decomposition of $k = f_\ell(k) + k - f_\ell(k)$ for $k \in \mathcal{K}$. We have the estimate

$$\limsup_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E} \exp \left[\theta \sum_{i,j} [x_{i,j} - f_\ell(x_{i,j})]^2 \right] = 0$$

for any $\theta > 0$. If $\{f_\ell(x_{i,j})\}$ and $\{x_{i,j}\}$ are mapped respectively into k_ℓ and k , we have super exponential estimates on $\Delta(\ell) = \int_D |f_\ell(k(x,y)) - k(x,y)|^2 dx dy$:

$$Q_n[\Delta(\ell) \geq \epsilon] \leq e^{-n^2 \theta \epsilon^2} \mathbb{E} \exp \left[\theta \sum_{i,j} [x_{i,j} - f_\ell(x_{i,j})]^2 \right].$$

Therefore

$$\limsup_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[\Delta(\ell) \geq \epsilon] \leq -\theta \epsilon^2.$$

Since $\theta > 0$ is arbitrary, for any $\epsilon > 0$,

$$\limsup_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[\Delta(\ell) \geq \epsilon] = -\infty.$$

It is easy to complete the proof of Theorem 1.2 using the above identity and the lower-semicontinuity of I .

The difference between the eigenvalues of two kernels k_1 and k_2 can be easily controlled by the Hilbert-Schmidt norm of the difference $k_1 - k_2$. In fact using the variational formula for successive eigenvalues of a compact self adjoint operator,

$$\lambda_{j+1}(A) = \inf_{\substack{\mathcal{M}: \\ \text{codim}(\mathcal{M})=j}} \sup_{\substack{\|z\|=1 \\ z \in \mathcal{M}}} \langle Az, z \rangle$$

it is easily seen that if $\lambda_j(A)$ and $\lambda_j(B)$ are the j th largest positive eigenvalues of two compact self adjoint operators A and B

$$|\lambda_j(A) - \lambda_j(B)| \leq \|A - B\| \leq \|A - B\|_{HS}$$

and a similar formula holds for negative eigenvalues as well. Clearly if $k_1, k_2 \in \mathcal{K}$, then $d(S(k_1), S(k_2)) \leq \|k_1 - k_2\| \leq \|k_1 - k_2\|_{HS}$.

If $C \in \mathcal{S}$ is a closed set, and we truncate X at level ℓ and denote by k_ℓ and k the two images in \mathcal{K} , then

$$P_n(C) = Q_n[S(k) \in C] \leq Q_n[S(k_\ell) \in C^\epsilon] + Q_n[d(S(k_\ell), S(k)) \geq \epsilon].$$

Applying the upper bound for the truncated version,

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n[C] \leq \max \left\{ - \inf_{S \in C^\epsilon} J_\ell(S), -c(\ell, \epsilon) \right\},$$

where

$$c(\ell, \epsilon) = - \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Q_n[d(S(k_\ell), S(k)) \geq \epsilon]$$

and for any $\epsilon > 0$,

$$\lim_{\ell \rightarrow \infty} c(\ell, \epsilon) = -\infty.$$

Moreover

$$J_\ell(S) = \inf_{k: S(k)=S} I_\ell(k); \quad I_\ell(k) = \int_D h_\ell(k(x,y)) dx dy$$

and

$$h_\ell(z) = \sup \left[z\theta - \log \int \exp[\theta f_\ell(x)] \mu(dx) \right].$$

The upper bound is now easily established by letting $\ell \rightarrow \infty$ and then letting $\epsilon \rightarrow 0$. We observe that the superexponential estimate implies $c(\ell, \epsilon) \rightarrow \infty$ as $\ell \rightarrow \infty$ for any $\epsilon > 0$. This completes the proof of Theorem 1.1.

References

1. Bai, Z. D.: Methodologies in the spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** (1999), no. 3, 611–677.
2. Ben Arous, G. and Guionnet, A.: Large deviations for Wigners law and Voiculescu's non-commutative entropy, *Prob. Th. Rel. Fields* **108** (1997) 517–542.
3. Chatterjee, S. and Varadhan, S. R. S.: The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.* **32** (2011), no. 7, 1000–1017
4. Komlós, J. and Simonovits, M.: Szemerédi's regularity lemma and its applications in graph theory. *Bolyai Soc. Math. Stud.*, **2** (1996) 295–352.
5. Lovász, L. and Szegedy, B.: Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17** (2007), no. 1, 252–270.
6. Wigner, E. P.: On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** (1958) 325–327.

SOURAV CHATTERJEE: COURANT INSTITUTE, NEW YORK UNIVERSITY, NEW YORK, NY, 10012, USA

E-mail address: sourav@cims.nyu.edu

URL: <http://www.cims.nyu.edu/~sourav/>

S. R. S. VARADHAN: COURANT INSTITUTE, NEW YORK UNIVERSITY, NEW YORK, NY, 10012, USA

E-mail address: varadhan@cims.nyu.edu

URL: <http://www.cims.nyu.edu/~varadhan/>