Open Problem Session: Large Deviations for the trace of a Wigner matrix to the power k

Jamal Najim

Ecole Nationale Supérieure des Télécommunications and CNRS

YEP 2006 Workshop
March 2006

Statement of the problem

Let X_n be a $n \times n$ matrix with random entries such that

- the matrix $X_n = (X_{ij})$ is symmetric
- the entries X_{ij} $(i \le j)$ are i.i.d. μ-distributed where μ has a compact support.

Question (suggested by Amir Dembo)

Prove a Large Deviation Principle for

$$T_n = \frac{1}{n^k} \mathrm{Trace}(X_n^k)$$

where k is a fixed integer.

A preliminary analysis

$$\text{Trace}(X_n^k) = \sum_{1 \le i_1, \cdots, i_k \le n} \underbrace{X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}}_{Y(i_1, \cdots, i_k)}$$

Exponential integrability

The variables $Y(i_1, \cdots, i_k)$ are bounded due to the assumption on μ. In particular,

$$\mathbb{E} e^{\lambda |Y(i_1, \cdots, i_k)|} < \infty \qquad \forall \lambda \in \mathbb{R}^+$$

▷ No exponential integrability issues

Trivial cases

The cases $k = 1$ and $k = 2$ immediatly follow from Cramér's theorem:

$$
\begin{aligned}
\frac{1}{n}\mathrm{Trace}(X_n) &= \frac{1}{n}\sum_{i=1}^{n} X_{ii} \\
\frac{1}{n^2}\mathrm{Trace}(X_n^2) &= \frac{1}{n^2}\sum_{1\le i,j\le n} X_{ij}X_{ji} = \frac{1}{n^2}\sum_{i,j} X_{ij}^2
\end{aligned}
$$

The Gaussian case

If the random variables X_{ij} $(i \leq j)$ are gaussian then the joint law of the eigenvalues $(\lambda_1, \cdots, \lambda_n)$ associated to $\frac{X_n}{\sqrt{n}}$ has a well-known density:

$$p(\lambda_1, \cdots \lambda_n) = \frac{1}{Z_n} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j| \exp\left\{ -\frac{n}{4} \sum_{i=1}^{n} \lambda_i^2 \right\}$$

(Z_n normalizing constant). In this case

$$\frac{1}{n^k} \mathrm{Trace}(X_n^k) = \frac{1}{n^{\frac{k}{2}}} \mathrm{Trace} \left(\frac{X_n}{\sqrt{n}} \right)^k = \frac{1}{n^{\frac{k}{2}}} \sum_{i=1}^{n} \lambda_i^k$$

The LDP for the empirical measure $\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}$ has been established (Ben Arous - Guionnet '97) but the function $f : x \mapsto x^k$ is not bounded in this case and one cannot use the contraction principle.

The non-trivial case $k = 3$

The case $k = 3$ is the first non-trivial case

$$T_n = \frac{1}{n^3}\mathrm{Trace}(X_n^3) = \frac{1}{n^3}\sum_{1\le i_1,i_2,i_3\le n} X_{i_1i_2}X_{i_2i_3}X_{i_3i_4}.$$

- ▷ It is an empirical mean.
- ▷ Specific dependence structure among the terms.
- ▷ One cannot apply Cramér's theorem.

The formulation of the problem using empirical measures

Denote by

$$L_n = \frac{1}{n^3} \sum_{1 \le i_1, i_2, i_3 \le n} \delta_{(X_{i_1 i_2}, X_{i_2 i_3}, X_{i_3 i_1})}.$$

Since the support of μ (say S) is bounded

$$\begin{aligned} f : S^3 &\longrightarrow \mathbb{R} \\ (x, y, z) &\longmapsto xyz \end{aligned}$$

is bounded continuous and

$$\text{LDP for } L_n \quad \Longrightarrow \quad \text{LDP for} \quad \langle L_n, f \rangle = \frac{1}{n^3}\text{Trace}(X_n^3).$$

by the contraction principle.

It is therefore sufficient to prove the LDP for L_n with respect to the weak topology.

For the sake of comparison with other LDPs for empirical measures, we shall study

$$\tilde{L}_n = \frac{1}{n(n-1)(n-2)} \sum_{\substack{1 \leq i_1, i_2, i_3 \leq n \\ \text{pairwise differents}}} \delta_{(X_{i_1 i_2}, X_{i_2 i_3}, X_{i_3 i_1})}.$$

For simplicity, we call it the Tracial empirical measure.

Both L_n and $\tilde{L}_n$ are exponentially equivalent.

To sum up the problem (in the case $k = 3$):

- either establish directly the LDP for

$$T_n = \frac{1}{n^3}\mathrm{Trace}(X_n^3)$$

- or establish the LDP for

$$\tilde{L}_n = \frac{1}{n(n-1)(n-2)} \sum_{\substack{1 \le i_1, i_2, i_3 \le n \\ \text{pairwise differents}}} \delta_{(X_{i_1 i_2}, X_{i_2 i_3}, X_{i_3 i_1})}.$$

- what about the rate function? Is it convex?

Any ideas?

Remaining plan

1. **A few words on exchangeability**

2. **Two related LDPs for empirical measures**

3. **A graph interpretation of the model**

4. **A conjecture for the rate function**

A few words on exchangeability

N-exchangeability

A finite sequence $(Z_1, \cdots, Z_N)$ of random variables is N-exchangeable if

$$(Z_1, \cdots, Z_N) \stackrel{\mathcal{D}}{=} (Z_{\sigma(1)}, \cdots, Z_{\sigma(N)}) \qquad \forall \sigma \in S_N.$$

Exchangeability

An infinite sequence $(Z_1, \cdots)$ of random variables is exchangeable if

$$(Z_1, Z_2, \cdots) \stackrel{\mathcal{D}}{=} (Z_{\sigma(1)}, Z_{\sigma(2)}, \cdots)$$

for every permutation σ such that $\#\{i,\ \sigma(i) \neq i\} < \infty$.

Extendibility

A finite N-exchangeable sequence $(Z_1, \cdots, Z_N)$ is extendible if there exists an infinite exchangeable sequence $(\tilde{Z}_1, \cdots)$ such that

$$(Z_1, \cdots, Z_N) \stackrel{\mathcal{D}}{=} (\tilde{Z}_1, \cdots \tilde{Z}_N)$$

In general a N-exchangeable sequence IS NOT extendible.

De Finetti's Theorem

"every infinite sequence of exchangeable random variables $(Z_1, Z_2, \cdots)$ is a mixture of i.i.d. random variables ". Otherwise stated:

$$\mathbb{P}\{(Z_1, Z_2, \cdots) \in A\} = \int \pi^{\otimes\infty}(A)\Theta(d\pi)$$

where

- π is a probability measure over $\mathbb{R}$
- Θ is a probability measure over the set of probability measures.

The sequence $(Z_1, Z_2, \cdots)$ can be described by the following two-stages procedure:

1. pick π at random from distribution Θ
2. then let (Z_i) be i.i.d. with distribution π.

The law of an extendible sequence (Z_1, Z_2, Z_3)

Let ν be defined over S^3 by

$$\nu(A \times B \times C) = \int \pi(A)\pi(B)\pi(C)\Theta(d\pi)$$

then ν is the law of an extendible sequence (Z_1, Z_2, Z_3).

Generalities

The relative entropy

Let μ be a probability measure over some space $\mathcal{X}$ ($\mu \in \mathcal{P}(\mathcal{X})$). The relative entropy with respect to μ is defined by:

$$H(\nu \mid \mu) = \begin{cases} \int \left(\frac{d\nu}{d\mu}\right) \log \left(\frac{d\nu}{d\mu}\right) d\mu & \text{if } \nu << \mu, \\ \infty & \text{otherwise.} \end{cases}$$

The relative entropy has a key-role in describing the rate functions assiciated to LDPs of various empirical measures, as we shall see.

Symmetrization

When studying $\tilde{L}_n$, one can check that if $\delta_{X,Y,Z}$ is one of the terms, all the terms based on permutations of (X, Y, Z) are also present. We will keep this feature in the forthcoming examples.

A Sanov-Like theorem

Consider the following empirical probability measure:

$$S_n = \frac{3!}{n(n-1)(n-2)} \sum_{i=1}^{\frac{n(n-1)(n-2)}{3!}} \frac{1}{3!} \sum_{\sigma \in S_3} \delta_{(X_{i\sigma(1)}, X_{i\sigma(2)}, X_{i\sigma(3)})}.$$

The measure S_n has the following properties:

- ▷ S_n is based on $n(n-1)(n-2)$ terms,
- ▷ there are $\frac{n(n-1)(n-2)}{2}$ independent random variables
- ▷ the “range” of X_{ij} is 6
- ▷ the “symmetrization” feature holds

Denote by $\mathcal{P}_3 = \mathcal{P}(S^3)$ and by $C_3 = C(S^3)$.

Theorem

The empirical measure S_n satisfies the LDP in $(\mathcal{P}_3, \sigma(\mathcal{P}_3, C_3))$ with good rate function I i.e:

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n(n-1)(n-2)} \ln \mathbb{P}(S_n \in C) &\leq -I(C)\\ \liminf_{n\to\infty} \frac{1}{n(n-1)(n-2)} \ln \mathbb{P}(S_n \in O) &\geq -I(O)\end{aligned}$$

where

$$I(\nu) = \begin{cases} \frac{1}{3!} H(\nu \mid \mu^{\otimes 3}) & \text{if } \nu \text{ is 3-exchangeable} \\ \infty & \text{otherwise.} \end{cases}$$

Note that I is convex.

Large Deviations for the U-statistics

We follow Eischelsbacher and Schmock (2002) and look at:

$$U_n^3 = \frac{1}{n(n-1)(n-2)} \sum_{\substack{1 \le i_1, i_2, i_3 \le n \\ \text{pairwise differents}}} \delta_{(X_{i_1}, X_{i_2}, X_{i_3})}$$

The main features of this problem are the following:

- U_n is based on $n(n-1)(n-2)$ terms.
- there are n independent random variables
- the range of the random variable X_i is $6(n-1)(n-2)$.
- the "symmetrization" feature holds

Compared to the previously mentioned Sanov theorem, the dependence between the terms is much more important since there are less random variables.

Theorem

The empirical measure U_n satisfies the LDP in $(\mathcal{P}_3, \sigma(\mathcal{P}_3, C_3))$ with good rate function J i.e:

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n(n-1)(n-2)} \ln \mathbb{P}(U_n \in C) &\leq -J(C)\\ \liminf_{n\to\infty} \frac{1}{n(n-1)(n-2)} \ln \mathbb{P}(U_n \in O) &\geq -J(O)\end{aligned}$$

where

$$J(\nu) = \begin{cases} \frac{1}{3} H(\nu \mid \mu^{\otimes 3}) & \text{if } \nu = \nu_1^{\otimes 3} \\ \infty & \text{otherwise.} \end{cases}$$

In this case J is no longer convex.

Comparison with the tracial empirical measure

model	Sanov-like	Tracial measure	U-statistics
measure	S_n	L_n	U_n
# independent r.v.	$\frac{n(n-1)(n-2)}{2}$	$\frac{n(n-1)}{2}$	n
range of each r.v.	6	$2(n-2)$	$6(n-1)(n-2)$
rate function	$\frac{1}{3!}H(\nu \mid \mu)$	$??$	$\frac{1}{3}H(\nu \mid \mu)$
	if ν exchangeable	$??$	if $\nu = \nu_1^{\otimes 3}$
	convex	$??$	non-convex

The measure $\tilde{L}_n$ as a measure on triangles

Let

- G_n be a complete graph
- E_n, the set of its edges, $\mathrm{card}(E_n) = \binom{n}{2}$;
- T_n, the set of its triangles, $\mathrm{card}(T_n) = \binom{n}{3}$

Assume that the random variable $X_{ij} = X_{ji}$ is associated to the edge (i, j), then $\tilde{L}_n$ can be viewed as a measure on the triangles of the graph:

$$\begin{aligned}\tilde{L}_n &= \frac{1}{n(n-1)(n-2)} \sum_{\substack{1 \le i_1, i_2, i_3 \le n \\ \text{pairwise differents}}} \delta_{(X_{i_1 i_2}, X_{i_2 i_3}, X_{i_3 i_1})} \\ &= \frac{1}{\binom{n}{3}} \sum_{(i_1, i_2, i_3) \in T_n} \frac{1}{3!} \sum_{\sigma \in S_3} \delta_{(X_{i_{\sigma(1)} i_{\sigma(2)}}, X_{i_{\sigma(2)} i_{\sigma(3)}}, X_{i_{\sigma(3)} i_{\sigma(1)}})}.\end{aligned}$$

If one defines by

$$\frac{1}{3!}\sum_{\sigma\in S_3}\delta_{(X_{i_{\sigma(1)}i_{\sigma(2)}},X_{i_{\sigma(2)}i_{\sigma(3)}},X_{i_{\sigma(3)}i_{\sigma(1)}})}$$

the empirical measure related to the triangle (i_1, i_2, i_3), its action over a bounded continuous function is well-defined.

The Gibbs Conditioning Principle

Denote by $\pi_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ and recall the Gibbs Conditioning Principle:

$$\mathcal{L}\left(X_1 \,\middle|\, \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \approx \nu\right) \to \nu$$

In loose terms: in order for the empirical measure to behave close to ν, every "particle" X_i should behave as ν.

Question

What can we infer by mimicking this reasoning on $\tilde{L}_n$?

What if a GCP holds for $\tilde{L}_n$?

A Gibbs conditioning principle should hold and yield:

$$\mathcal{L}\left(\text{Triangle}(X_{12}, X_{23}, X_{31}) \,\middle|\, \tilde{L}_n \approx \nu\right) \to \nu$$

In this case,

- every triangle should behave as ν to insure a deviation behaviour for the empirical measure $\tilde{L}_n$
- a “compatibility” condition between adjacent triangles must hold

Constraints on the rate function

Both constraints can be fulfilled if

- ν is a product measure, i.e. $\nu = \nu_1^{\otimes 3}$:
  $\triangleright$ it suffices to let each random variable associated to a given edge be distributed as ν_1, independently.

- ν is extendible, i.e. $\nu = \int \pi^{\otimes 3} \Theta(d\pi)$:
  $\triangleright$ it suffices to first pick at random a distribution π following Θ then to let the $\binom{n}{2}$ edges be i.i.d with distribution π.

In view of the previous (fully qualitative) analysis, we make the following guess

Conjecture

the empirical measure L_n satisfies a LDP in $(\mathcal{P}_3, \sigma(\mathcal{P}_3, C_3))$ with good rate function Γ i.e:

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n(n-1)(n-2)} \ln \mathbb{P}(L_n \in C) &\leq -\Gamma(C) \\ \liminf_{n\to\infty} \frac{1}{n(n-1)(n-2)} \ln \mathbb{P}(L_n \in O) &\geq -\Gamma(O) \end{aligned}$$

where Γ is given by:

4. A conjecture for the rate function

$$\Gamma(\nu) = \begin{cases} \kappa_3 H(\nu \mid \mu^{\otimes 3}) & \text{if } \nu \text{ is 3-exchangeable and extendible} \\ & \text{i.e. } \nu = \int \pi^{\otimes 3} \Theta(d\pi) \\ \infty & \text{otherwise.} \end{cases}$$

In the previous formula, the exact value of κ_3 has to be found. We also believe that the previous formula extends to the case where $k > 3$.