6. Concentration Phenomena and Stronger Forms of Convergence for the Semicircle Law

6.1. Remarks: 1) Recall that our random matrix ensemble is given by probability measures P_N on sets Ω_N of $N \times N$ matrices and we want to see that μ_{A_N} converges weakly to μ_W; equivalently that, for all $z \in \mathbb{C}^+$, $S_{\mu_{A_N}}(z)$ converges to $S_{\mu_W}(z)$.

There are different levels of this convergence with respect to P_N:

i) convergence in average, i.e.

$$E[S_{\mu_{A_N}}(z)] \longrightarrow S_{\mu_W}(z)$$

ii) convergence in probability, i.e.

$$P_N\{A_N : |S_{\mu_{A_N}}(z) - S_{\mu_W}(z)| \geq \varepsilon\} \xrightarrow{N \to \infty} 0 \qquad \forall \varepsilon > 0$$

iii) almost sure convergence, i.e.

$$P\{(A_N)_N : S_{\mu_{A_N}}(z) \text{ does not converge to } S_{\mu_W}(z)\} = 0$$

This is guaranteed, by Borel-Cantelli Lemma, if the convergence in (ii) to zero is sufficiently fast in N, so that

$$\sum_N P_N\{A_N : |S_{\mu_{A_N}}(z) - S_{\mu_W}(z)| \geq \varepsilon\} < \infty$$

Note again that we have here a convergence of probabilistic quantities to a deterministic limit, thus (ii) and (iii) are saying that for large N the eigenvalue distribution of A_N concentrates in a small neighborhood of μ_W. This is an instance of a quite general "concentration of measure" phenomenon:
"A random variable that depends (in a smooth way) on the influence of many independent variables (but not too much on any of them) is essentially constant."
(M. Talagrand)

Note that many classical results in probability theory (like law of large numbers) can be seen as instances of this, dealing with linear functions. However, this principle also applies to non-linear functions - like in our case, to $\operatorname{tr}[(A_N - z)^{-1}]$ considered as a function of the entries of A_N.

2) Often control of the variance of the involved variables is a good way to get concentration estimates.

6.2. Theorem (Markov Inequality): Let X be a random variable taking non-negative values. Then, for any $t > 0$,

$$P\{\omega : X(\omega) \geq t\} \leq \frac{E[X]}{t}$$

[Recall: P probability measure on Ω; random variable $X : \Omega \to \mathbb{R}$ is a measurable function and

$$E[X] = \int_\Omega X(\omega)\, dP(\omega) = \int_{\mathbb{R}} x \, d\underset{\substack{\uparrow \\ \text{distribution of } X}}{\mu_X(x)}$$

]

Proof: Since $X(\omega) \geq 0 \quad \forall \omega \in \Omega$:

$$E[X] = \int_\Omega X(\omega)\, dP(\omega)$$

$$= \underbrace{\int_{\{X(\omega)<t\}} X(\omega)\, dP(\omega)}_{\geq 0} + \int_{\{X(\omega)\geq t\}} X(\omega)\, dP(\omega)$$

$$\geq \int_{\{X(\omega)\geq t\}} \underbrace{X(\omega)}_{\geq t}\, dP(\omega)$$

$$\geq t \int_{\{X(\omega)\geq t\}} 1 \; dP(\omega)$$

$$= t \cdot P\{\omega : X(\omega) \geq t\}$$ □

6.3 Notation: For a random variable $X : \Omega \to \mathbb{R}$ we put

$$E[X] = \int_\Omega X(\omega)\, dP(\omega)$$ "expectation" "mean"

$$Var[X] = E[(X - E[X])^2]$$ "variance"

$$= E[X^2] - E[X]^2$$

$$= \int (X(\omega) - E[X])^2\, dP(\omega)$$

6.4. Theorem (Chebyshev Inequality):

Let X be a random variable with finite mean μ and variance σ^2. Then, for any $\varepsilon > 0$,

$$P\{\omega : |X(\omega) - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon}$$

Proof: Use Markov Inequality for positive random variable $Y := (X-\mu)^2$

Note that

$$E[Y] = E[(X - E(X))^2] = Var(X) = \sigma^2$$

Thus we have

$$P\{\omega : |X(\omega) - \mu| \geq \varepsilon\} = P\{\omega : (X(\omega) - \mu)^2 \geq \varepsilon^2\}$$

$$\overset{6.2.}{\leq} \frac{E[Y]}{\varepsilon^2}$$

$$= \frac{\sigma^2}{\varepsilon^2} \qquad \square$$

6.5. Remark: Our goal will thus be to control the variance of $X = f(X_1, \dots, X_n)$ for $X_1, \dots, X_n$ independent random variables

[in our case: X_i entries of the GOE matrix A

and $f = tr[(A - z\cdot 1)^{-1}]$].

A main idea in this context is to have estimates, which go over from separate control of each variable to control of all variables together, i.e., which are stable under tensorization. There are two prominent types of such estimates, namely

i) Poincaré inequality

ii) LSI = logarithmic Sobolev inequality

We will focus here on (i) and say a few words on (ii) later

<u>6.6. Definition</u>: A random variable $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ satisfies Poincaré inequality with constant $c > 0$ if for any differentiable function with $E[f(X)^2] < \infty$ we have

$$\operatorname{Var}(f(X)) \leq c\, E[\| \nabla f(X) \|_2^2]$$

where $\| \nabla f \|_2^2 = \sum_{i=1}^{n} \left(\frac{\partial}{\partial x_i} f \right)^2$

Let us write this also "explicitly" in terms of the distribution μ_X of $X : \Omega \to \mathbb{R}^n$

$$E[f(X)] = \int f(x_1, .., x_n)\, d\mu(x_1, .., x_n)$$

and PI means:

$$\int_{\mathbb{R}^n} \left(f(x_1, .., x_n) - E[f(X)]\right)^2 d\mu(x_1, .., x_n)$$

$$\leq c \sum_{i=1}^{n} \int_{\mathbb{R}^n} \left(\frac{\partial f}{\partial x_i}(x_1, .., x_n)\right)^2 d\mu(x_1, .., x_n)$$

6.7. Theorem (Efron-Stein Inequality): (6-

Let $X_1, \dots, X_n$ be independent random variables and let $f(X_1, \dots, X_n)$ be a square-integrable function of $X = (X_1, \dots, X_n)$. Then we have

$$\mathrm{Var}(f(X)) \leq \sum_{i=1}^{n} E[\mathrm{Var}^{(i)} f(X)]$$

integrating over all the other variables

taking variance in i-th variable, keeping all the other fixed

Proof: Put $Z = f(X_1, \dots, X_n)$, then we have

$$E[Z] = \int_{\mathbb{R}^n} f(x_1, \dots, x_n) \, d\mu_1(x_1) \dots d\mu_n(x_n)$$

μ_i = distribution of X_i (prob. measure on $\mathbb{R}$)

$X_1, \dots, X_n$ independent $\Rightarrow$ distr. of $X = (X_1, \dots, X_n)$ is given by product measure $\mu_1 \times \dots \times \mu_n$ on $\mathbb{R}^n$

$$\mathrm{Var}(Z) = \int_{\mathbb{R}^n} \left(f(x_1, \dots, x_n) - E[Z]\right)^2 d\mu_1(x_1) \dots d\mu_n(x_n)$$

We will now do the integration E by integrating one variable at a time and control each step

We write

$$
\begin{aligned}
Z - E[Z] = \; & Z - E_1(Z) && \to \Delta_1 \\
& + E_1(Z) - E_{1,2}(Z) && \to \Delta_2 \\
& + E_{1,2}(Z) - E_{1,2,3}(Z) && \to \Delta_3 \\
& \;\vdots \\
& + E_{1,2,\dots,n-1}(Z) - E(Z) && \to \Delta_n
\end{aligned}
$$

where $E_{1,\dots,k}$ denotes integration over the variables $x_1, \dots, x_k$, leaving a function of the variables $x_{k+1}, \dots, x_n$

Thus, with

$$\Delta_i := E_{1,\dots,i-1}(Z) - E_{1,\dots,i}(Z)$$

(is a fct of the variables $x_{i+1}, \dots, x_n$

we have

$$Z - E[Z] = \sum_{i=1}^{n} \Delta_i$$

and thus

$$
\begin{aligned}
Var(Z) &= E\left[(Z - E(Z))^2\right] \\
&= E\left[\left(\sum_{i=1}^{n} \Delta_i\right)^2\right]
\end{aligned}
$$

$$= \sum_{i=1}^{n} E[\Delta_i^2] + \sum_{i \neq j} E[\Delta_i \Delta_j]$$ (6-

Now observe that for all $i \neq j$ we have

$$E[\Delta_i \Delta_j] = 0$$

For example, consider $i=1, j=2$ (and $n=2$)

$$E[\Delta_1 \Delta_2] = E[(Z - E_1(Z))(E_1(Z) - E_{1,2}(Z))]$$

$$= \int \left[f(x_1,x_2) - \int f(\tilde{x}_1,x_2)\, d\mu_1(\tilde{x}_1)\right] \cdot$$

$$\left[\int f(\tilde{x}_1,x_2)\, d\mu_1(\tilde{x}_1) - \int f(\tilde{x}_1,\tilde{x}_2)\, d\mu_1(\tilde{x}_1) d\mu_2(\tilde{x}_2)\right]$$

$$d\mu_1(x_1)\, d\mu_2(x_2)$$

Integration w.r.t. x_1 only affects first factor and integrating this gives zero.

General $i \neq j$ works in the same way.

Thus we get

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} E[\Delta_i^2]$$

Denote with $E^{(i)}$ integration with respect to only variable x_i, leaving a fct of the variables $x_1, .., x_{i-1}, x_{i+1}, .., x_n$

and

$$\mathrm{Var}^{(i)}(Z) = E^{(i)}[(Z - E^{(i)}(Z))^2]$$

then we have

$$\Delta_i = E_{1,\dots,i-1}(Z) - E_{1,\dots,i}(Z)$$

$$= E_{1,\dots,i-1}[Z - E^{(i)}(Z)]$$

and thus, by Jensen's inequality,

$$\Delta_i^2 \leq E_{1,\dots,i-1}[(Z - E^{(i)}(Z))^2]$$

Thus

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} E(\Delta_i^2)$$

$$\leq \sum_{i=1}^{n} E[\underbrace{E^{(i)}(Z - E^{(i)}(Z))^2}_{\mathrm{Var}^{(i)}(Z)}]$$

<u>6.8 Theorem</u>: Let $X_1, \dots, X_n$ be independent random variables in $\mathbb{R}$, s.th. each X_i satisfies a Poincaré inequality with c_i. Then $X = (X_1, \dots, X_n)$ satisfies a Poincaré inequality in $\mathbb{R}^n$ with $c = \max(c_1, \dots, c_n)$.

Proof: By 6.7 we have

$$\mathrm{Var}(f(X)) \leq \sum_{i=1}^{n} E[\underbrace{\mathrm{Var}^{(i)} f(X)}_{\leq c_i E^{(i)}\left[\left(\frac{\partial f}{\partial x_i}\right)^2\right]}]$$

by Poincaré inequality for X_i and fct

$x_i \mapsto f(x_1, .., x_{i-1}, x_i, .., x_n)$

for each fixed $x_1, .., x_{i-1}, x_{i+1}, .. x_n$

$$\leq c \sum_{i=1}^{n} E\left[\left(\frac{\partial f}{\partial x_i}\right)^2\right]$$

$$= c\, E\, \| \nabla f(X) \|_2^2]$$

□

<u>6.9. Theorem (Gaussian Poincaré Inequality)</u>

Let $X_1, .., X_n$ be independent standard Gaussian random variables, $E(X_i) = 0$ and $E(X_i^2) = 1$. Then $X = (X_1, .., X_n)$ satisfies Poincaré inequality with constant 1: for each continuously differentiable $f: \mathbb{R}^n \to \mathbb{R}$ we have

$$\mathrm{Var}(f(X)) \leq E[\| \nabla f \|^2]$$

[Note the independence of n]

Proof: By 6.8. it suffices to prove the statement for $n=1$.

So let X be a Gaussian random variable and $f:\mathbb{R}\to\mathbb{R}$. Then we have to show

$$\operatorname{Var}(f(X)) \le E[f'(X)^2]$$

We might assume that $E[f(X)]=0$, then this means explicitly:

$$\int f(x)^2 e^{-x^2/2}\,dx \le \int f'(x)^2 e^{-x^2/2}\,dx$$

This is not obvious!

One possible proof is to approximate X via a central limit theorem by independent Bernoulli variables Y_i.

Let $Y_1, Y_2, \ldots$ be independent Bernoulli variables, i.e. $P(Y_i=+1)=\frac{1}{2}=P(Y_i=-1)$ and put

$$S_n = \frac{1}{\sqrt{n}}(Y_1+\ldots+Y_n)$$

Then, by the CLT, the distribution of S_n converges (weakly), for $n\to\infty$, to a standard Gaussian.

So we can approximate $f(X)$ by

$$g(Y_1,\dots,Y_n) = f\left(\frac{1}{\sqrt{n}}(Y_1+\dots+Y_n)\right)$$

By Efron-Stein, we have

$$\mathrm{Var}(g(Y_1,\dots,Y_n)) \le \sum_{i=1}^{n} E[\mathrm{Var}^{(i)}(Y_1,\dots,Y_n)]$$

Put $S_n^{[i]} := S_n - \frac{1}{\sqrt{n}} Y_i$

$$= \frac{1}{\sqrt{n}}(Y_1+\dots+Y_{i-1}+Y_{i+1}+\dots+Y_n)$$

Then

$$E^{(i)}(f(S_n)) = \frac{1}{2}\left(f\left(S_n^{[i]}+\frac{1}{\sqrt{n}}\right)+f\left(S_n^{[i]}-\frac{1}{\sqrt{n}}\right)\right)$$

and

$$\mathrm{Var}^{(i)}(f(S_n)) =$$

$$= \frac{1}{2}\left\{\left(f\left(S_n^{[i]}+\frac{1}{\sqrt{n}}\right) - E^{(i)}(f(S_n))\right)^2 + \right.$$

$$\left. + \left(f\left(S_n^{[i]}-\frac{1}{\sqrt{n}}\right) - E^{(i)} f(S_n)\right)^2\right\}$$

$$= \frac{1}{4}\left(f\left(S_n^{[i]}+\frac{1}{\sqrt{n}}\right) - f\left(S_n^{[i]}-\frac{1}{\sqrt{n}}\right)\right)^2$$

and thus by Efron - Stein

$$Var(f(S_n)) \leq \frac{1}{4} \sum_{i=1}^{n} E\left[\left(f(S_n^{[i]}+\tfrac{1}{\sqrt{n}}) - f(S_n^{[i]}-\tfrac{1}{\sqrt{n}})\right)^2\right]$$

By Taylor's Theorem we have now

$$f(S_n^{[i]}+\tfrac{1}{\sqrt{n}}) = f(S_n^{[i]}) + \tfrac{1}{\sqrt{n}} f'(S_n^{[i]}) + \tfrac{1}{2n} f''(\xi_+)$$

$$f(S_n^{[i]}-\tfrac{1}{\sqrt{n}}) = f(S_n^{[i]}) - \tfrac{1}{\sqrt{n}} f'(S_n^{[i]}) + \tfrac{1}{2n} f''(\xi_-)$$

We assume that f is twice differentiable and f', f'' bounded : $|f''(\xi)| \leq k \quad \forall \xi$
$|f'(\xi)| \leq k$
[the general situation can be approximated by this !]

Then we have

$$\left(f(S_n^{[i]}+\tfrac{1}{\sqrt{n}}) - f(S_n^{[i]}-\tfrac{1}{\sqrt{n}})\right)^2 =$$

$$= \Big(\tfrac{2}{\sqrt{n}} \underbrace{f'(S_n^{[i]})}_{=: R_1} + \tfrac{1}{2n}\big(\underbrace{f''(\xi_+)-f''(\xi_-)}_{=: R_2}\big)\Big)^2$$

$$= \frac{4}{n} f'(S_n^{[i]})^2 + \frac{2}{n^{3/2}} \cdot R_1 R_2 + \frac{1}{4n^2} \cdot R_2^2$$

where $|R_1| \leq k$, $|R_2| \leq 2k$

and thus

$$\mathrm{Var}(f(S_n)) \leq \frac{1}{4} \cdot n \left\{ \frac{4}{n} \overbrace{E[f'(S_n^{[i]})^2]}^{\text{independent of } i} + \frac{2}{n^{3/2}} \cdot 2k^2 + \frac{1}{4n^2} \cdot 4k^2 \right\}$$

Note that both S_n and $S_n^{[i]}$ converge to our standard Gaussian variable X; thus taking $n \to \infty$ in the above gives

$$\mathrm{Var}(f(X)) \leq E[f'(X)^2]$$

$\square$

6.10. Concentration for $\mathrm{tr}(R_A(z))$:

We apply this now to the random matrix setting: $A = (x_{ij})_{i,j=1}^{N}$ where the $\{x_{ij} \mid i \leq j\}$ are independent Gaussian random variables with $E[x_{ij}] = 0$ and

$$E[x_{ii}] = \frac{2}{N}, \quad E[x_{ij}] = \frac{1}{N} \quad (i \neq j)$$

Note that by a change of variable the constant in the Poincaré inequality is then given by $\max\{\sigma_{ij}^2 \mid i \leq j\} = \frac{2}{N}$

Thus we have for nice real-valued f:

$$\operatorname{Var}(f(A)) \le \frac{2}{N} E[\|\nabla f\|_2^2]$$

We take now

$$g(A) := \operatorname{tr}[(A - z \cdot 1)^{-1}] = \operatorname{tr}(R_A(z))$$

and want to control

$$\operatorname{Var}(g(A_N)) \quad \text{for } N \to \infty$$

Note that g is complex-valued, but we can estimate

$$|\operatorname{Var}(g(A))| = |\operatorname{Var}(\operatorname{Re} g(A) + i \operatorname{Im} g(A))|$$

$$\le 2\Big(\operatorname{Var}(\operatorname{Re} g(A)) + \operatorname{Var}(\operatorname{Im}(A$$

thus it suffices to estimate the variance of real and imaginary part.

We have, for $i < j$

$$\frac{\partial g}{\partial x_{ij}} = \frac{\partial}{\partial x_{ij}} \underbrace{\operatorname{tr}(R_A(z))}_{\frac{1}{N}\sum_{k=1}^{N} [R_A(z)]_{kk}}$$

$$= \frac{1}{N} \sum_{k=1}^{N} \frac{\partial [R_A(z)]_{kk}}{\partial x_{ij}}$$

$$= -\frac{1}{N} \sum_{k=1}^{N} \left([R_A(z)]_{ki} [R_A(z)]_{jk} + [R_A(z)]_{kj} [R_A(z)]_{ik} \right.$$

$$= -\frac{2}{N} \sum_{k=1}^{N} [R_A(z)]_{ik} [R_A(z)]_{kj}$$

(since $R_A(z)$ is symmetric, see Proof of 5.5

$$= -\frac{2}{N} [R_A(z)^2]_{ij}$$ [the same for $i=j$ with $2/N$ replaced by $1/N$]

and thus with $f = \operatorname{Re} g = \operatorname{Re} \operatorname{tr} [R_A(z)]$

$$\left| \frac{\partial f}{\partial x_{ij}} \right| = \left| \operatorname{Re} \frac{\partial g}{\partial x_{ij}} \right| = \frac{2}{N} \underbrace{\left| [R_A(z)^2]_{ij} \right|}_{\leq \|R_A(z)^2\| \leq \frac{1}{(\operatorname{Im} z)^2}}$$

$$\Rightarrow \left| \frac{\partial f}{\partial x_{ij}} \right|^2 \leq \frac{4}{N^2 \cdot (\operatorname{Im} z)^4}$$

$$\Rightarrow \operatorname{Var}(f(A)) \leq \frac{2}{N} \left\{ \sum_{i \leq j} \underbrace{\left| \frac{\partial f}{\partial x_{ij}} \right|^2}_{\leq \frac{4}{N^2 \cdot (\operatorname{Im} z)^4}} \right\}$$

$$\leq \frac{8}{N} \cdot \frac{1}{(\operatorname{Im} z)^4}$$

The same for the imaginary part, and thus

$$\operatorname{Var}[\operatorname{tr}(R_A(z))] \sim \frac{c}{N} \quad \text{for } N \to \infty$$
(*)

By Chebyshev Inequality, this implies that for any $\varepsilon > 0$

$$P_N\{A_N : |\operatorname{tr}(R_{A_N}(z)) - E(\operatorname{tr}(R_{A_N}(z))| \geq \varepsilon\}$$

$$\leq \frac{c}{N \cdot \varepsilon} \xrightarrow{N \to \infty} 0$$

Thus we get the convergence in probability of $\operatorname{tr}(R_{A_N}(z)) = S\mu_{A_N}(z)$ to

$$\lim_{N \to \infty} E[\operatorname{tr}(R_{A_N}(z))] = S\mu_W(z),$$

i.e. to the Stieltjes transform of the semicircle.

(*) The fact that $\operatorname{Var}(\operatorname{tr}(R_A(z)))$ goes to zero for $N \to \infty$ closes the gap in our proof of 5.5, in the replacement in (**). Furthermore it improves our type of convergence in Wigner's semicircle law.

6.12. Remark: 1) Note that our estimate $\mathrm{Var}(..) \sim \frac{c}{N}$ is not strong enough to get almost sure convergence; one can, however, improve our arguments to get $\mathrm{Var}(..) \sim \frac{c}{N^2}$, which implies then almost sure convergence

2) One actually has typically even exponential convergence. Such concentration estimates rely usually on so called LSI = logarithmic Sobolev inequalities. Those are of the form: a probability measure μ on $\mathbb{R}^m$ satisfies LSI with constant $c>0$, if for all nice f:

$$\underbrace{\mathrm{Ent}_\mu(f^2)}_{\text{entropy like quantity}} \leq 2c \int_{\mathbb{R}^m} \|\nabla f\|_2^2 \, d\mu$$

$$= \int_{\mathbb{R}^m} f \log f \, d\mu - \int_{\mathbb{R}^m} f \, d\mu \cdot \log \int_{\mathbb{R}^m} f \, d\mu$$

As Poincare inequality LSI are stable under tensorization and Gaussian measures satisfy LSI.

From LSI follows then concentration inequality of the form

$$P_N\{A_N : |\operatorname{tr}(R_{A_N}(z)) - E(\operatorname{tr}(R_{A_N}(z)))| \geq \varepsilon\}$$

$$\leq 2\, e^{-\frac{N\varepsilon^2}{2} \cdot \operatorname{Im} z^4}$$