Concentration of measure and the Herbst argument

Q: When is a function of independent random variables close to it's mean?

Thm (Chernoff bound): Let $X_1, \ldots, X_n$ be independent r.v.'s with $\mathbb{E}[X_i] = 0$, $\mathbb{E}[e^{\lambda X_i}] \le e^{\lambda^2 v/2}$ for a constant $v > 0$ and all $\lambda \in \mathbb{R}$.

Then for any $t > 0$, $\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \ge t\right] \le 2e^{-\frac{t^2}{2v}\cdot n}$

Proof: For any $\lambda > 0$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \ge t\right] = \mathbb{P}\left[e^{\lambda(\frac{1}{n}\sum_{i=1}^{n} X_i)} \ge e^{\lambda t}\right]$$
$$\le e^{-\lambda t}\,\mathbb{E}\left[e^{\lambda(\frac{1}{n}\sum_{i=1}^{n} X_i)}\right]$$
$$= e^{-\lambda t}\prod_{i=1}^{n}\mathbb{E}\left[e^{\frac{\lambda}{n}X_i}\right] \le e^{-\lambda t}\cdot e^{\frac{\lambda^2 v}{2n}}.$$

Set $\lambda = tn/v$. The lower bound is the same.

Want to generalize this to other functions $f(X_1, \ldots, X_n)$ of independent random variables. Two ingredients:

- Each X_i has fast enough tail decay.
- f is not too strongly influenced by any individual X_i.

Closeness of f to $\mathbb{E}[f]$ is called concentration of measure.

Good reference: Boucheron, Lugosi, Massart, "Concentration Inequalities".

Def: For a random element $X \in \mathcal{X}$ and a nonnegative function $f: \mathcal{X} \to [0, \infty)$, we define the entropy

$$\mathrm{Ent}_X[f] = \mathbb{E}[f(X)\log f(X)] - \mathbb{E}[f(X)]\cdot\log\mathbb{E}[f(X)].$$

Def: Let $\mathcal{A}$ be a class of functions $f: \mathcal{X} \to \mathbb{R}$, and $\mathcal{E}: \mathcal{A} \to [0,\infty]$. A random element $X \in \mathcal{X}$ satisfies a <u>log-Sobolev inequality</u> w.r.t. $\mathcal{E}$, with constant $c>0$, if for every $f \in \mathcal{A}$,

$$\mathrm{Ent}_X(f^2) \le 2c \cdot \mathcal{E}(f).$$

Prop: Let $X \sim \mathrm{Bernoulli}(1/2)$, i.e. $\mathbb{P}[X=x_0] = \mathbb{P}[X=x_1] = 1/2$. Let $\mathcal{A}$ be all functions $f: \{x_0,x_1\} \to \mathbb{R}$, and set $\mathcal{E}(f) = \mathbb{E}[(f(X)-f(X'))^2]$ where X' is an independent copy of X. Then X satisfies the LSI w/ $c = 1/2$.

Proof: Let $a = f(x_0)$, $b = f(x_1)$. We want to show

$$h(a,b) \equiv \frac{1}{2}\left(a^2 \log a^2 + b^2 \log b^2\right) - \frac{a^2+b^2}{2} \log \frac{a^2+b^2}{2} - \frac{1}{2}(a-b)^2 \le 0.$$

Suffices to assume $a \ge b \ge 0$. For each fixed $b \ge 0$, can check:

- $h(b,b) = 0$
- $\partial_a h(a,b)\big|_{a=b} = 0$.
- $\partial_a^2 h(a,b) \le 0$ for all $a \ge b$.

Then $h(a,b) \le h(b,b) = 0$ for all $a \ge b \ge 0$.

Prop: Let X be Bernoulli with $\mathbb{P}[X=x_0]=p$, $\mathbb{P}[X=x_1]=1-p$. Let $\mathcal{E}(f) = \mathbb{E}[(f(X)-f(X'))^2]$. Then X satisfies LSI w/ constant $\frac{1}{1-2p}\log\frac{1-p}{p}$.

Proof: See [BLG Thm 5.2].

Prop: Let $X \sim \mathcal{N}(0,1)$, $\mathcal{A}$ all continuously differentiable $f: \mathbb{R} \to \mathbb{R}$, and $\mathcal{E}(f) = \mathbb{E}[f'(X)^2]$. Then X satisfies LSI w/ constant $c=1$.

Proof: Let $P_t f(x) = \mathbb{E}\left[f\left(\sqrt{e^{-t}}x + \sqrt{1-e^{-t}}Z\right)\right]$ for $Z \sim \mathcal{N}(0,1)$.

- $P_0 f(x) = f(x)$ and $P_\infty f(x) = \mathbb{E}[f(Z)]$.
- For $X \sim \mathcal{N}(0,1)$ and all t, $\mathbb{E}[P_t f(X)] = \mathbb{E}[f(Z)]$.

  In particular, $\mathbb{E}[\partial_t P_t f(X)] = \partial_t \mathbb{E}[P_t f(X)] = 0$.
- Recall the Gaussian integration by parts formula

$$\mathbb{E}[Zf(Z)] = \int z f(z) \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int f'(z) \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \mathbb{E}[f'(Z)].$$

Then:
$$\partial_t P_t f(x) = \mathbb{E}\left[\left(-\frac{1}{2}\sqrt{e^{-t}} \cdot x + \frac{1}{2}\frac{e^{-t}}{\sqrt{1-e^{-t}}} Z\right) f'\left(\sqrt{e^{-t}}x + \sqrt{1-e^{-t}}Z\right)\right]$$
$$= -\frac{1}{2} x \cdot \partial_x P_t f(x) + \frac{1}{2} e^{-t} \mathbb{E}\left[f''\left(\sqrt{e^{-t}}x + \sqrt{1-e^{-t}}Z\right)\right]$$
$$= -\frac{1}{2} x \cdot \partial_x P_t f(x) + \frac{1}{2}\partial_x^2 P_t f(x).$$

For $f \ge 0$, implicitly evaluating all functions at $X \sim \mathcal{N}(0,1)$, write

$$\mathrm{Ent}_X(f) = \mathbb{E}[P_0 f \cdot \log P_0 f] - P_\infty f \cdot \log P_\infty f$$
$$= -\int_0^\infty \mathbb{E}[\partial_t (P_t f \log P_t f)]\, dt$$
$$= -\int_0^\infty \mathbb{E}[(1+\log P_t f) \cdot \partial_t P_t f]\, dt$$
$$= -\int_0^\infty \mathbb{E}\left[\log P_t f \cdot \left(-\frac{1}{2} X \cdot \partial_x P_t f + \frac{1}{2}\partial_x^2 P_t f\right)\right] dt$$

Apply $\mathbb{E}[X \cdot g(X) \cdot \partial_x P_t f(X)] = \mathbb{E}[\partial_x g(X) \cdot \partial_x P_t f(X) + g(X) \cdot \partial_x^2 P_t f(X)]$:

$$\mathrm{Ent}_X(f) = \frac{1}{2}\int_0^\infty \mathbb{E}\left[\frac{(\partial_x P_t f)^2}{P_t f}\right] dt$$
$$= \frac{1}{2}\int_0^\infty e^{-t}\, \mathbb{E}\left[\frac{(P_t(f'))^2}{P_t f}\right] dt.$$

Cauchy-Schwarz gives $(P_t(f'))^2 \le P_t f \cdot P_t\left(\frac{f'^2}{f}\right)$. Then

$$\mathrm{Ent}_X(f) \le \frac{1}{2}\int_0^\infty e^{-t}\, \underbrace{\mathbb{E}\left[P_t\left(\frac{f'^2}{f}\right)\right]}_{=\mathbb{E}\left[\frac{f'(X)^2}{f(X)}\right]\ \forall t} dt = \frac{1}{2}\mathbb{E}\left[\frac{f'(X)^2}{f(X)}\right].$$

Then $\mathrm{Ent}_X(f^2) \le 2\,\mathbb{E}[f'(X)^2]$ as desired.

Prop: Let $X \in \mathbb{R}$ be a continuous r.v. with density $p(x) \propto e^{-V(x)}$, where $V(x) - \frac{x^2}{2c}$ is convex for some $c>0$. Then X satisfies the LSI with constant c, for $\mathcal{E}(f) = \mathbb{E}[f'(X)^2]$.

Proof: See [AGZ Lemma 2.3.2].

Prop: Let $X_1, \dots, X_n \in \mathbb{R}$ be independent random variables, each X_i satisfying a LSI with constant $c>0$ wrt $\mathcal{E}_i$. For $f: \mathbb{R}^n \to \mathbb{R}$, define

$$\mathcal{E}(f) = \sum_{i=1}^n \mathbb{E}\left[\mathcal{E}_i\left(f(X_1, \dots, X_{i-1}, \cdot\,; X_{i+1}, \dots, X_n)\right)\right]$$

where $f(x_1, \dots, x_{i-1}, \cdot\,; x_{i+1}, \dots, x_n): \mathbb{R} \to \mathbb{R}$ is the function fixing all but X_i.

Then $X = (X_1, \dots, X_n)$ satisfies the LSI w/ constant $c>0$.

Proof: It suffices to show, for $f \ge 0$,

$$\mathrm{Ent}_X(f) \le \sum_{i=1}^n \mathbb{E}\left[\mathrm{Ent}_{X_i}\left(f(X_1, \dots, X_{i-1}, \cdot\,; X_{i+1}, \dots, X_n)\right)\right].$$

Lemma: $\mathrm{Ent}_X(f) = \sup\{\mathbb{E}[f(X) g(X)] : g \text{ satisfies } \mathbb{E}[e^{g(X)}] \le 1\}$.

Proof: Note $\mathrm{Ent}_X(cf) = c\,\mathrm{Ent}_X(f)$, for $c>0$. Thus, assume $\mathbb{E} f(X) = 1$.

Apply Young's inequality: $uv \le u\log u - u + e^v$ for $u \ge 0$. Then

$\mathbb{E}[f(X) g(X)] \le \mathrm{Ent}_X(f) - 1 + \mathbb{E}[e^{g(X)}] \le \mathrm{Ent}_X(f)$.

Taking $g(x) = \log f(X)$ yields equality.

Now, for any $g: \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathbb{E}[e^{g(X_1, \dots, X_n)}] \le 1$, set

$$g_i(x_1, \dots, x_i) = \log \mathbb{E}[e^{g(x_1, \dots, x_i, X_{i+1}, \dots, X_n)}] - \log \mathbb{E}[e^{g(x_1, \dots, x_{i-1}, X_i, \dots, X_n)}].$$

Then:

- $g(x_1, \dots, x_n) = \sum_{i=1}^n g_i(x_1, \dots, x_i) + \log \mathbb{E}[e^{g(X_1, \dots, X_n)}] \le \sum_{i=1}^n g_i(x_1, \dots, x_i)$.
- For any fixed $x_1, \dots, x_{i-1}$, $\mathbb{E}[e^{g_i(x_1, \dots, x_{i-1}, X_i)}] = 1$.

So
$$\begin{aligned}
\mathbb{E}[f(X) g(X)] &\le \sum_{i=1}^n \mathbb{E}[f(X_1, \dots, X_n)\, g_i(X_1, \dots, X_i)] \\
&= \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}_{X_i}\left[f(X_1, \dots, X_{i-1}, X_i, X_{i+1}, \dots, X_n)\, g(X_1, \dots, X_{i-1}, X_i)\right]\right] \\
&\le \sum_{i=1}^n \mathbb{E}\left[\mathrm{Ent}_{X_i}\left(f(X_1, \dots, X_{i-1}, \cdot\,; X_{i+1}, \dots, X_n)\right)\right].
\end{aligned}$$

Take a sup over $g(x)$ s.t. $\mathbb{E}[e^{g(X)}] \le 1$.

Cor: Let $X_1, \dots, X_n$ be independent continuous r.v.'s, each satisfying the LSI $\mathrm{Ent}_{X_i}(f^2) \le 2c\,\mathbb{E}[f'(X_i)^2]$. Then for differentiable $f: \mathbb{R}^n \to \mathbb{R}$, $X = (X_1, \dots, X_n)$ satisfies the LSI

$$\mathrm{Ent}_X(f^2) \le 2c\,\mathbb{E}[\|\nabla f(X)\|^2].$$

Thm: Suppose X satisfies the LSI $\mathrm{Ent}_X(f^2) \le 2c\,\mathcal{E}(f)$ for all $f \in \mathcal{A}$. Furthermore, for some $|||\cdot|||: \mathcal{A} \to [0, \infty)$ satisfying $|||\lambda f||| = |\lambda| \cdot |||f|||$, suppose $\mathcal{E}(e^{f/2}) \le |||f|||^2\, \mathbb{E}[e^{f(X)}]$ for all $f \in \mathcal{A}$. $(*)$

Then for any $f \in \mathcal{A}$ with $|||f||| < \infty$, and any $t > 0$,

$$\mathbb{P}[f(X) - \mathbb{E} f(X) \ge t] \le e^{-t^2/8c|||f|||^2}.$$

Cases of interest:

- $\mathcal{X} = \mathbb{R}^n$, $\mathcal{E}(f) = \mathbb{E}[\|\nabla f(X)\|^2]$. Then

$$\mathcal{E}(e^{f/2}) = \mathbb{E}\left[\frac{1}{4} e^{f(X)} \|\nabla f(X)\|^2\right] \le \frac{1}{4} \sup_{x\in\mathcal{X}} \|\nabla f(x)\|^2 \cdot \mathbb{E}[e^{f(X)}].$$

So $(*)$ holds for $|||f|||^2 = \frac{1}{4}\sup_{x\in\mathcal{X}}\|\nabla f(x)\|^2$.

Corollary: Suppose $X_1,\dots,X_n$ are independent, with densities $p_i(x) \propto e^{-V_i(x)}$ satisfying $V_i(x) - \frac{x_i^2}{2c}$ is convex for all i. Then for any $f:\mathbb{R}^n\to\mathbb{R}$ satisfying $\|f(x)-f(y)\| \le L|x-y|$ for all $x,y\in\mathbb{R}^n$ (f is L-Lipschitz), and $t>0$,

$$\mathbb{P}[|f(X_1,\dots,X_n) - \mathbb{E}f(X_1,\dots,X_n)| \ge t] \le 2e^{-\frac{t^2}{2cL^2}}.$$

In particular, this holds for $X_1,\dots,X_n \overset{iid}{\sim} \mathcal{N}(0,1)$ and $c=1$.

- $\mathcal{X} = \{a,b\}^n$, $X_1,\dots,X_n \overset{iid}{\sim}$ Bernoulli(p) w/ $\mathbb{P}[X_i=a]=p$, $\mathbb{P}[X_i=b]=1-p$.

$$\mathcal{E}(f) = \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}_{X_i,X_i'}\left[(f(X_1,\dots,X_i,\dots,X_n) - f(X_1,\dots,X_i',\dots,X_n))^2\right]\right].$$

Let $V(f) = \sup\{|f(x)-f(x')| : x,x' \text{ differ in one coordinate}\}$.

Apply $|e^{a/2}-e^{b/2}| \le \frac{|a-b|}{2}\cdot e^{\max(a,b)/2}$. Then

$$\mathcal{E}(e^{f/2}) \le \frac{V(f)^2}{4}\mathbb{E}\left[\exp\left(\max(f(X_1,\dots,X_i,\dots,X_n), f(X_1,\dots,X_i',\dots,X_n))\right)\right]$$
$$\le \frac{V(f)^2}{2}\mathbb{E}[e^{f(X)}].$$

So $(*)$ holds for $|||f|||^2 = \frac{V(f)^2}{2}$.

Corollary: Suppose $X_1,\dots,X_n \overset{iid}{\sim}$ Bernoulli(p). Then for any f such that $|f(x)-f(x')| \le L$ for all x,x' differing in one coordinate,

$$\mathbb{P}[|f(X_1,\dots,X_n) - \mathbb{E}f(X_1,\dots,X_n)| \ge t] \le 2e^{-t^2/4L^2c(p)}$$

where $c(p) = \frac{1}{1-2p}\log\frac{1-p}{p}$.

Proof of Thm: Let $M(\lambda) = \mathbb{E}[e^{\lambda f(X)}]$. Then $M'(\lambda) = \mathbb{E}[f(X)e^{\lambda f(X)}]$.

Applying the LSI to $e^{\lambda f/2}$,

$$\mathrm{Ent}(e^{\lambda f}) = \lambda M'(\lambda) - M(\lambda)\log M(\lambda)$$
$$\le 2c\,\mathcal{E}(e^{\lambda f/2})$$
$$\le 2c\lambda^2 \cdot |||f|||^2 M(\lambda).$$

Then $\partial_\lambda\left[\frac{\log M(\lambda)}{\lambda}\right] = \frac{M'(\lambda)}{\lambda M(\lambda)} - \frac{\log M(\lambda)}{\lambda^2} \le 2c\cdot|||f|||^2$.

We have $M(0)=1$, $M'(0) = \mathbb{E}[f(X)]$. Then $\lim_{\lambda\to 0}\frac{\log M(\lambda)}{\lambda} = \frac{M'(0)}{M(0)} = \mathbb{E}[f(X)]$. So for any $\lambda>0$,

$$\frac{\log M(\lambda)}{\lambda} \le \mathbb{E}[f(X)] + \lambda\cdot 2c|||f|||^2$$
$$\Rightarrow M(\lambda) \le e^{\lambda\mathbb{E}f(X) + \lambda^2\cdot 2c|||f|||^2}.$$

Recall $\mathbb{P}[f(X) \ge \mathbb{E}f(X)+t] \le e^{-\lambda(\mathbb{E}f(X)+t)}M(\lambda) \le e^{-\lambda t + \lambda^2\cdot 2c|||f|||^2}$.

Setting $\lambda = \frac{t}{4c|||f|||^2}$ yields the theorem.

Concentration of eigenvalues and linear spectral statistics

Lemma (Hoffman-Wielandt): Let $X,Y\in\mathbb{R}^{n\times n}$ be symmetric, and let $\lambda_1(X)\ge\lambda_2(X)\ge\dots\ge\lambda_n(X)$ and $\lambda_1(Y)\ge\lambda_2(Y)\ge\dots\ge\lambda_n(Y)$ be the ordered eigenvalues. Then

$$\sum_{i=1}^n |\lambda_i(X)-\lambda_i(Y)|^2 \le \mathrm{Tr}\,(X-Y)^2.$$

Proof: Note $\sum_{i=1}^n |\lambda_i(X)-\lambda_i(Y)|^2 = \mathrm{Tr}\, X^2 + \mathrm{Tr}\, Y^2 - 2\sum_{i=1}^n \lambda_i(X)\lambda_i(Y)$.

It suffices to show $\mathrm{Tr}\, XY \le \sum_{i=1}^n \lambda_i(X)\lambda_i(Y)$.

Assume WLOG X is diagonal, and write $Y = U\begin{pmatrix}\lambda_1(Y) & & \\ & \ddots & \\ & & \lambda_n(Y)\end{pmatrix}U^T$.

Then
$$\mathrm{Tr}\, XY = \mathrm{Tr}\begin{pmatrix}\lambda_1(X) & & \\ & \ddots & \\ & & \lambda_n(X)\end{pmatrix} U \begin{pmatrix}\lambda_1(Y) & & \\ & \ddots & \\ & & \lambda_n(Y)\end{pmatrix}U^T$$
$$= \sum_{i,j=1}^n \lambda_i(X) U_{ij} \lambda_j(Y) (U^T)_{ji}$$
$$= \sum_{i,j=1}^n \lambda_i(X)\lambda_j(Y) U_{ij}^2.$$

Let $\mathcal{A} = \{S \in \mathbb{R}^{n\times n} : \sum_i S_{ij} = 1\ \forall j,\ \sum_j S_{ij} = 1\ \forall i,\ S_{ij} \ge 0\}$.

Note that $(U_{ij}^2)_{i,j=1}^n \in \mathcal{A}$. Then
$$\mathrm{Tr}\, XY \le \sup_{S\in\mathcal{A}} \sum_{i,j=1}^n \lambda_i(X)\lambda_j(Y) S_{ij}.$$

To show RHS is maximized at $S = I$, suppose $s_{11} < 1$. Then $s_{1j} > 0$ and $s_{k1} > 0$ for some $j, k \ne 1$. Set $\delta = \min(s_{1j}, s_{k1})$ and $\tilde{S}$ s.t. $\tilde{S}_{11} = S_{11} + \delta$, $\tilde{S}_{kj} = S_{kj} + \delta$, $\tilde{S}_{1j} = S_{1j} - \delta$, $\tilde{S}_{k1} = S_{k1} - \delta$.

Then $\tilde{S} \in \mathcal{A}$, and
$$\sum_{i,j=1}^n \lambda_i(X)\lambda_j(Y)(\tilde{S}_{ij} - S_{ij}) = \delta(\lambda_1(X) - \lambda_k(X))(\lambda_1(Y) - \lambda_j(Y)) \ge 0.$$

Iteratively repeating this procedure for all diagonal entries yields the claim.

Corollary: Let $X \in \mathbb{R}^{n\times n}$ be symmetric, $g : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz. Then the map $X \mapsto g(\lambda_1(X), \dots, \lambda_n(X))$ is a $\sqrt{2}L$-Lipschitz function of the $\frac{n(n+1)}{2}$ variables $(X_{ij})_{i<j}$ and $\left(\frac{X_{ii}}{\sqrt{2}}\right)_{i=1}^n$.

Proof: $|g(\lambda_1(X), \dots, \lambda_n(X)) - g(\lambda_1(Y), \dots, \lambda_n(Y))|$
$$\le L \cdot \sqrt{\sum_{i=1}^n (\lambda_i(X) - \lambda_i(Y))^2} \le L \cdot \sqrt{\mathrm{Tr}(X-Y)^2} = L\sqrt{\sum_{i,j}(X_{ij} - Y_{ij})^2}$$
$$= \sqrt{2}\cdot L \cdot \sqrt{\sum_{i<j}(X_{ij} - Y_{ij})^2 + \sum_i \tfrac{1}{2}(X_{ii} - Y_{ii})^2}.$$

Thm: Suppose $X \in \mathbb{R}^{n\times n}$ is Wigner, where $X_{ij} = \frac{1}{\sqrt{n}} Z_{ij}$ and Z_{ij} are continuous r.v.'s satisfying the LSI w/ constant c for $i \ne j$ and $2c$ for $i = j$. (For GOE, $c = 1$.) Then for any L-Lipschitz function $f : \mathbb{R} \to \mathbb{R}$,

(a) $\mathbb{P}\left[\left|\int f(x)\,d\mu_n(x) - \mathbb{E}\int f(x)\,d\mu_n(x)\right| \ge t\right] \le 2e^{-\frac{t^2}{4cL^2}\cdot n^2}$.

(b) For any individual ordered eigenvalue $\lambda_k(X)$,
$$\mathbb{P}\left[\left|f(\lambda_k(X)) - \mathbb{E} f(\lambda_k(X))\right| \ge t\right] \le 2e^{-\frac{t^2}{4cL^2}\cdot n}.$$

Proof: (a) The function $g(\lambda_1, \dots, \lambda_n) = \frac{1}{n}\sum_{i=1}^n f(\lambda_i)$ is $\frac{L}{\sqrt{n}}$-Lipschitz:
$$|g(\lambda_1, \dots, \lambda_n) - g(\mu_1, \dots, \mu_n)| \le \frac{1}{n}\sum_{i=1}^n |f(\lambda_i) - f(\mu_i)| \le \frac{L}{n}\sum_{i=1}^n |\lambda_i - \mu_i| \le \frac{L}{\sqrt{n}}\|\lambda - \mu\|_2.$$

Then $Z \mapsto \int f(x)\,d\mu_n(x)$ is $\frac{\sqrt{2}L}{n}$-Lipschitz in $(Z_{ij})_{i<j}$ and $\left(\frac{Z_{ii}}{\sqrt{2}}\right)_{i=1}^n$ which satisfy the LSI w/ constant c. Result follows from concentration.

(b) The function $g(\lambda_1, \dots, \lambda_n) = f(\lambda_k)$ is L-Lipschitz, so $Z \mapsto f(\lambda_k(X))$ is $\frac{\sqrt{2}L}{\sqrt{n}}$-Lipschitz. Result also follows from concentration.

Thm: Suppose $X \in \mathbb{R}^{n\times n}$ is Wigner, $X_{ij} = \frac{1}{\sqrt{n}} Z_{ij}$ where
$$Z_{ij} = \begin{cases} \frac{1-p}{\sqrt{p(1-p)}} & p \\ -\frac{p}{\sqrt{p(1-p)}} & 1-p \end{cases} \qquad Z_{ii} = 0.$$

Let $f:\mathbb{R}\to\mathbb{R}$ be L-Lipschitz. Then

(a) $\mathbb{P}\left[\left|\int f(x)\,d\mu_n(x) - \mathbb{E}\int f(x)\,d\mu_n(x)\right| \geq t\right] \leq 2e^{-\frac{t^2}{L^2\tilde{c}(p)}n^2}$

(b) $\mathbb{P}\left[\left| f(\lambda_k(X)) - \mathbb{E} f(\lambda_k(X))\right| \geq t\right] \leq 2e^{-\frac{t^2}{L^2\tilde{c}(p)}\cdot n}$

where $\tilde{c}(p) = 8\cdot\frac{|1-2p|}{p(1-p)}\log\frac{1-p}{p}$.

Proof: Similar to above.

Remark: (a) indicates that fluctuations of

$$\int f(x)\,d\mu_n(x) = \frac{1}{n}\sum_{i=1}^n f(\lambda_i(X))$$

are $O_{\mathbb{P}}(\frac{1}{n})$. This is <u>different</u> from the behavior of $\frac{1}{n}\sum_{i=1}^n f(\lambda_i)$ if $\lambda_1,\ldots,\lambda_n$ were i.i.d., which would be $O_{\mathbb{P}}(\frac{1}{\sqrt{n}})$. The eigenvalues $\lambda_1(X)\geq\ldots\geq\lambda_n(X)$ of a Wigner matrix X are closer to the quantiles of the semicircle law than n i.i.d. samples from this law. This is called "eigenvalue rigidity." The $O_{\mathbb{P}}(\frac{1}{n})$ bound is sharp, and there is a CLT for $\sum_{i=1}^n f(\lambda_i(X))$.

(b) indicates that fluctuations of $\lambda_k(X)$ are $O_{\mathbb{P}}(\frac{1}{\sqrt{n}})$. For the Wigner model this bound is not sharp! For $(1-\varepsilon)n \geq k \geq \varepsilon n$, the fluctuations are in fact $O_{\mathbb{P}}(\frac{1}{n})$. For $k = O(1)$ or $n - O(1)$, the fluctuations are $O_{\mathbb{P}}(\frac{1}{n^{2/3}})$. However, this result also holds for more general models, e.g.

$X + A$ where $A\in\mathbb{R}^{n\times n}$ is symmetric, deterministic. If A is low-rank, $X+A$ can have outlier eigenvalues which do fluctuate on the scale $O_{\mathbb{P}}(\frac{1}{\sqrt{n}})$.

Thm: In the above two settings, there exist constants $C, c > 0$ such that

$\mathbb{P}[\|X\|\geq t] \leq 2e^{-c\cdot t^2 n}$ for all $t > C$.

(Here, $\|X\| = \max_i |\lambda_i| = \sup_{v:\|v\|=1} |v^T X v|$.)

Proof: For each fixed $v: \|v\|=1$, $|v^T(X-Y)v| = \left|\sum_{i,j}(X_{ij}-Y_{ij})v_i v_j\right| \leq \sqrt{\sum_{i,j}(X_{ij}-Y_{ij})^2}\cdot\sqrt{\sum_{i,j} v_i^2 v_j^2}$

$= \sqrt{2}\cdot\sqrt{\sum_{i<j}(X_{ij}-Y_{ij})^2 + \sum_i \frac{(X_{ii}-Y_{ii})^2}{2}}$. So $X\mapsto v^T X v$ is $\sqrt{\frac{2}{n}}$-Lipschitz in the entries of Z. Since $\mathbb{E}[v^T X v] = 0$, we get

$$\mathbb{P}\left[|v^T X v| \geq t\right] \leq 2e^{-\tilde{c}t^2 n} \quad \text{for a constant } \tilde{c} > 0.$$

For a small constant $\varepsilon > 0$, take an ε-net N of the sphere $\{v: \|v\|=1\}$, i.e. for all $v: \|v\|=1$, there is $w\in N$ with $\|v-w\|\leq\varepsilon$. We may take N so that $|N| \leq (\frac{2}{\varepsilon}+1)^n$. For any such v, w,

$$|v^T X v - w^T X w| \leq |v^T X(v-w)| + |(v-w)^T X w|$$
$$\leq 2\cdot\|X\|\cdot\|v-w\| \leq 2\varepsilon\cdot\|X\|.$$

Then $\|X\| = \sup_{v:\|v\|=1}|v^T X v| \leq \sup_{w\in N}|w^T X w| + 2\varepsilon\cdot\|X\|$

$\Rightarrow \|X\| \leq \frac{1}{1-2\varepsilon}\sup_{w\in N}|w^T X w|$ for $\varepsilon < \frac{1}{2}$.

Take $\varepsilon = \frac{1}{3}$. $\mathbb{P}[\|X\|\geq t] \leq \mathbb{P}[\exists w\in N: |w^T X w| \geq 3t]$

$\leq |N|\cdot 2e^{-\tilde{c}t^2 n}$

$\leq 2e^{-c\cdot t^2 n}$

for a different constant $c > 0$, and all $t > C$.

Finishing the proof of the semi-circle law

Last time, we showed $\mathbb{E}[\int x^k d\mu_n(x)] \to \int x^k d\mu(x)$ where μ is the semi-circle law. To conclude the proof, we show concentration of $\int x^k d\mu_n(x)$ around its mean.

Lemma: In the above two settings, for any fixed $k, \varepsilon > 0$, there exist constants

$$C, c > 0 \text{ s.t. } \mathbb{P}\left[\left|\int x^k d\mu_n(x) - \mathbb{E}\left[\int x^k d\mu_n(x)\right]\right| \geq \varepsilon\right] \leq C \cdot e^{-cn}.$$

Proof: Note that $x \mapsto x^k$ is not Lipschitz for $k > 1$. However, it is Lipschitz on $[-C, C]$ which contains all eigenvalues of X with high probability.

We use the idea of a Lipschitz extension: Let

$$f(x) = \begin{cases} x^k & |x| \leq C_0 \\ (\mathrm{sgn}(x) C_0)^k & |x| > C_0 \end{cases}$$

where $\mathbb{P}[\|X\| \geq t] \leq 2e^{-cnt^2}$ for all $t \geq C_0$. Then f is L-Lipschitz for a constant $L = L(k) > 0$. We get

$$\mathbb{P}\left[\left|\int f(x) d\mu_n(x) - \mathbb{E}\left[\int f(x) d\mu_n(x)\right]\right| \geq t\right] \leq 2e^{-cn^2t^2}.$$

On the event $\|X\| \leq C_0$, $\int f(x) d\mu_n(x) = \int x^k d\mu_n(x)$. So

$$\mathbb{P}\left[\left|\int x^k d\mu_n(x) - \mathbb{E}\left[\int f(x) d\mu_n(x)\right]\right| \geq t\right]$$

$$\leq 2e^{-cn^2t^2} + \mathbb{P}[\|X\| > C_0] \leq 2e^{-cn^2t^2} + 2e^{-cn}.$$

Finally, we bound

$$\left|\mathbb{E}\left[\int f(x) d\mu_n(x)\right] - \mathbb{E}\left[\int x^k d\mu_n(x)\right]\right| \leq \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\left|f(\lambda_i(X)) - \lambda_i(X)^k\right|\right]$$

$$\leq \mathbb{E}\left[\|X\|^k \cdot \mathbb{1}\{\|X\| > C_0\}\right]$$

$$= \mathbb{E}\left[\int_0^{\|X\|} kx^{k-1} dx \cdot \mathbb{1}\{\|X\| > C_0\}\right]$$

$$= \mathbb{E}\left[\int_0^\infty kx^{k-1} \mathbb{1}\{\|X\| > \max(C_0, x)\} dx\right]$$

$$= \int_0^\infty kx^{k-1} \mathbb{P}[\|X\| > \max(C_0, x)] dx$$

$$= \int_0^{C_0} kx^{k-1} \cdot 2e^{-cn} dx + \int_{C_0}^\infty kx^{k-1} \cdot 2e^{-cnx^2} dx$$

$$\leq Ce^{-cn}.$$

Hence $\mathbb{P}\left[\left|\int x^k d\mu_n(x) - \mathbb{E}\left[\int x^k d\mu_n(x)\right]\right| \geq t + Ce^{-cn}\right] \leq 2e^{-cn^2t^2} + 2e^{-cn}$

Setting $t = \varepsilon/2$ and adjusting the constants $C, c > 0$ yields the result.

Corollary: $\int x^k d\mu_n(x) \to \int x^k d\mu(x)$ almost surely for each k.

Proof: Let $m_k = \int x^k d\mu(x)$. Since $\mathbb{E}[\int x^k d\mu_n(x)] \to m_k$,

$$\mathbb{P}\left[\left|\int x^k d\mu_n(x) - m_k\right| > \varepsilon\right] \leq \mathbb{P}\left[\left|\int x^k d\mu_n(x) - \mathbb{E}\int x^k d\mu_n(x)\right| > \frac{\varepsilon}{2}\right]$$

$$\leq Ce^{-cn}$$

for all large n. Apply Borel-Cantelli.