4 Analytic Tools: Stieltjes Transform and Convergence of Measures

4.1. Remark: Recall our setting and goal: We have, for each $N \in \mathbb{N}$, selfadj. $N \times N$-random matrices, which are given by a probability measure P_N on the entries of the matrices.

For example, for the GUE we have

$$A_N = (a_{ij})_{i,j=1}^{N} \qquad a_{ij} = x_{ij} + \sqrt{-1}\, y_{ij}$$

Since $a_{ij} = \overline{a_{ji}}$ we have $y_{ii} = 0$ and as "free" variables

$$(*) \left.\begin{cases} x_{ii} & (i = 1, \dots, N) \\ x_{ij}, y_{ij} & (1 \le i < j \le N) \end{cases}\right\} N^2 \text{ many variables}$$

All those are independent and Gaussian distributed, which can be written in the compact form

$$dP_N(A) = C_N \exp\left(-N \frac{\operatorname{Tr}(A^2)}{2}\right) dA$$

normalization constant to make P_N probability ...

product of all differential

We want statements about our matrices w.r.t. this measure P_N, either in average, or in probability.

Denote by Ω_N the space of our $\underbrace{N\times N}_{\text{selfadjoint}}$ matrices, i.e.

$$\Omega_N = \{A = (x_{ij} + \sqrt{-1}\, y_{ij})_{i,j=1}^{N} \mid \begin{array}{l} x_{ii} \in \mathbb{R} \\ x_{ij}, y_{ij} \in \mathbb{R} \\ (i<j) \end{array} \}$$

$$\hat{=}\ \mathbb{R}^{N^2}$$

then P_N is a probability measure on Ω

For $A \in \Omega_N$ we consider its eigenvalues $\lambda_1, \dots, \lambda_N$, which we encode in a probability measure μ_A on $\mathbb{R}$:

$$\mu_A = \frac{1}{N}(\delta_{\lambda_1} + \dots + \delta_{\lambda_N})$$

"eigenvalue distribution of A"

Our claim is now that μ_A converges under P_N, for $N \to \infty$, to the semicircle distribution μ_W,

- in average:

$$\mu_N := \int_{\Omega_N} \mu_A \, dP_N(A) = E[\mu_A] \xrightarrow{N\to\infty} \mu_W$$

- and, stronger, in probability or almost surely.

So what we want to understand now is:

- What kind of convergence

$$\mu_N \to \mu$$

do we have here (for probability measures on $\mathbb{R}$)?

- How can we describe probability measures (on $\mathbb{R}$) and their convergence with analytic tools?

<u>4.2. Definition</u>: Let μ be a Borel measure on $\mathbb{R}$.

1) μ is <u>finite</u> if $\mu(\mathbb{R}) < \infty$
2) μ is a <u>probability measure</u> if $\mu(\mathbb{R}) = 1$
3) For a finite measure μ on $\mathbb{R}$ we define its <u>Stieltjes transform</u> S_μ on $\mathbb{C} \setminus \mathbb{R}$ by

$$S_\mu(z) = \int_{\mathbb{R}} \frac{1}{t-z} \, d\mu(t) \qquad (z \in \mathbb{C} \setminus \mathbb{R})$$

$-S_\mu = G_\mu$ is also called the <u>Cauchy transform</u>

<u>4.3. Theorem</u>: 1) Let μ be a finite measure on $\mathbb{R}$ and $S = S_\mu$ its Stieltjes transform. Then one has

i) $S : \mathbb{C}^+ \to \mathbb{C}^+$

where $\mathbb{C}^+ := \{ z \in \mathbb{C} \mid \operatorname{Im}(z) > 0 \}$

ii) S is analytic on $\mathbb{C}^+$

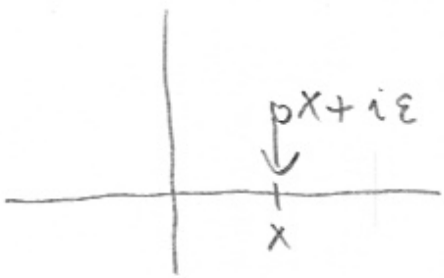
iii) $\lim\limits_{y \to \infty} iy\, S(iy) = -\mu(\mathbb{R})$

2) μ can be recovered from S_μ via the <u>Stieltjes inversion formula</u>:

for $a < b$ we have

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im} S_\mu(x + i\varepsilon)\, dx$$

$$= \mu((a,b)) + \frac{1}{2} \mu(\{a,b\})$$



$$\underbrace{\frac{1}{\pi} \operatorname{Im} S_\mu(x+i\varepsilon)}_{\text{density of measure}} dx \xrightarrow{\varepsilon \downarrow 0} d\mu(x)$$

3) In particular, we have for two finite measures μ and ν:

$$S_\mu = S_\nu \implies \mu = \nu$$

Proof: 1) Exercise

2) $\operatorname{Im} S_\mu(x+i\varepsilon) =$

$$= \int_{\mathbb{R}} \operatorname{Im}\left(\frac{1}{t-x-i\varepsilon}\right) d\mu(t)$$

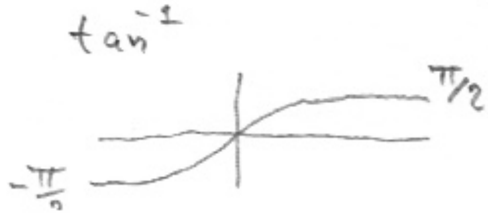
$$= \int_{\mathbb{R}} \frac{\varepsilon}{(t-x)^2+\varepsilon^2} \, d\mu(t)$$

and thus

$$\int_a^b \operatorname{Im} S_\mu(x+i\varepsilon)\, dx =$$

$$= \int_{\mathbb{R}} \underbrace{\int_a^b \frac{\varepsilon}{(t-x)^2+\varepsilon^2}\, dx}_{\displaystyle \int_{(a-t)/\varepsilon}^{(b-t)/\varepsilon} \frac{1}{x^2+1}\, dx \;=\; \tan^{-1}\left(\frac{b-t}{\varepsilon}\right) - \tan^{-1}\left(\frac{a-t}{\varepsilon}\right)} d\mu(t)$$

$$\xrightarrow{\varepsilon \searrow 0} \begin{cases} 0 & t \notin [a,b] \\ \pi/2 & t \in \{a,b\} \\ \pi & t \in (a,b) \end{cases}$$



$$\xrightarrow{\varepsilon \searrow 0} \pi \left[\mu((a,b)) + \frac{1}{2}\mu(\{a\}) + \frac{1}{2}\mu(\{b\}) \right]$$

3) Now assume $S_\mu = S_\nu$

$\Rightarrow \mu((a,b)) = \nu((a,b))$

for all open intervals s.th. a,b are neither atoms of μ nor of ν

Since there can only be countably many atoms we can write any interval as

$$(a,b) = \bigcup_{n=1}^{\infty} (a+\varepsilon_n, b-\varepsilon_n) \qquad \text{where } \varepsilon_n \searrow 0$$

and those are no atoms

By monotone convergence for measures

$$\Rightarrow \mu((a,b)) = \lim_{n\to\infty} \mu((a+\varepsilon_n, b-\varepsilon_n))$$

$$ = \lim_{n\to\infty} \nu((a+\varepsilon_n, b-\varepsilon_n))$$

$$ = \nu((a,b)) \qquad \square$$

4.4. Remark: If we put $\mu_\varepsilon = p_\varepsilon \lambda$ with (λ: Lebesgue measure)

$$p_\varepsilon(x) := \operatorname{Im} S_\mu(x+i\varepsilon)$$

$$= \int_{\mathbb{R}} \frac{\varepsilon}{(t-x)^2+\varepsilon^2}\, d\mu(t)$$

then $\mu_\varepsilon = \gamma_\varepsilon * \mu$ where γ_ε is the Cauchy distribution and we have above checked the well-known fact that $\gamma_\varepsilon * \mu \xrightarrow{\varepsilon \downarrow 0} \mu$

4.5. Proposition: Let μ be a compactly supported probability measure, say $\mu([-r, r]) = 1$ for some $r > 0$.

Then S_μ has a power series expansion (about ∞) as follows

$$S_\mu(z) = -\sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}} \qquad \text{for } |z| > r$$

where $m_n := \int_{\mathbb{R}} t^n \, d\mu(t)$ are the moments of μ.

Proof: For $z > r$ we can expand

$$\frac{1}{t-z} = -\frac{1}{z(1-\frac{t}{z})} = -\frac{1}{z} \sum_{n=0}^{\infty} \left(\frac{t}{z}\right)^n$$

$$\forall t \in [-r, r]$$

convergence is uniform, hence

$$S_\mu(z) = \int_{-r}^{r} \frac{1}{t-z} \, d\mu(t)$$

$$= -\sum_{n=0}^{\infty} \int_{-r}^{r} \frac{t^n}{z^{n+1}} \, d\mu(t)$$

$$= -\sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}} \qquad \square$$

4.6. Proposition: The Stieltjes transform S of the semicircle distribution, $d\mu_W(t) = \frac{1}{2\pi}\sqrt{t^2-4}\,dt$, is uniquely determinded by (for $z \in \mathbb{C}^+$)

- $S(z) \in \mathbb{C}^+$
- $S(z)$ is the solution of the equation

$$S(z)^2 + z\,S(z) + 1 = 0$$

Explicitly, this means

$$S(z) = \frac{-z + \sqrt{z^2-4}}{2} \qquad (z \in \mathbb{C}^+)$$

Proof: By 4.5 we know that for large $|z|$

$$S(z) = -\sum_{k=0}^{\infty} \frac{C_{2k}}{z^{2k+1}}$$

where C_n are the Catalan numbers.

By using the recursion for the Catalan numbers, this implies (compare Ass 1, Ex 2) that for large $|z|$ we have

$$S(z)^2 + z\,S(z) + 1 = 0$$

Since we know that S is analytic on $\mathbb{C}^+$, this equation is, by analytic extension, then valid for all $z \in \mathbb{C}^+$.

This equation has two solutions,

$$\frac{-z \pm \sqrt{z^2-4}}{2}, \text{ and only the one with}$$

the "+" sign is in $\mathbb{C}^+$. □

<u>4.7. Remark</u>: Proposition 4.6. gave us the Stieltjes transform of μ_W just from the knowledge of the moments.

From $S(z) = \frac{-z+\sqrt{z^2-4}}{2}$ we can then get the density of μ_W via Stieltjes inversion formula:

$$\frac{1}{\pi} \operatorname{Im} S(x+i\varepsilon) = \frac{1}{2\pi} \underbrace{\operatorname{Im} \sqrt{(x+i\varepsilon)^2 - 4}}$$

$$\xrightarrow{\varepsilon \downarrow 0} \operatorname{Im} \sqrt{x^2-4}$$

$$= \begin{cases} 0 & x > 2 \\ \sqrt{4-x^2} & x \leq 2 \end{cases}$$

$$\xrightarrow{\varepsilon \downarrow 0} \frac{1}{2\pi} \sqrt{4-x^2} \, 1_{[-2,2]}$$

Thus this analytic machinery gives an effective way to calculate a distribution from its moments (without having to know the density in advance)

4.8 Remark: Now we want to consider the convergence $\mu_N \to \mu$. We can consider (probability) measures from two (equivalent) perspectives:

i) μ gives us the measure (probability) of sets $\rightsquigarrow \mu(B)$ for $B \in \mathcal{B}(\mathbb{R})$ or just μ(intervalls)

ii) μ allows us to integrate continuous functions

$\rightsquigarrow \int f \, d\mu$ for continuous f

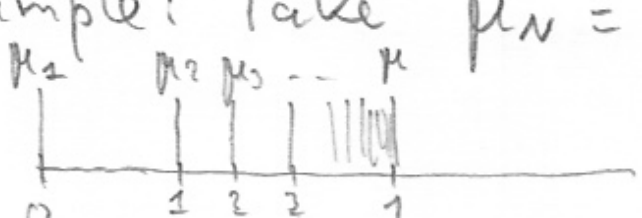
According to this there are two canonical choices for notions of convergence:

i) $\mu_N(B) \to \mu(B) \quad \forall B \in \mathcal{B}(\mathbb{R})$
or maybe for all intervalls B

ii) $\int f \, d\mu_N \to \int f \, d\mu \quad \forall$ continuous f

(i) is problematic in this generality, as it does treat atoms of the μ too restrictive

Example: Take $\mu_N = \delta_{1-\frac{1}{N}}$, $\mu = \delta_1$



Then we surely want that $\mu_N \to \mu$, but for $B = [1,2]$ we have

$$\mu_N([1,2]) = 0 \quad \forall N$$

$$\mu([1,2]) = 1$$

Thus (iii) is the better definition. But we have to be careful about which class of continuous functions we allow; we need bounded ones otherwise $\int f \, d\mu$ might not exist in general and, for compactness reasons, it is sometimes better to ignore the behavior of the measure at infinity.

4.9. Definitions: 1) We use the notations

i) $C_0(\mathbb{R}) := \{ f \in C(\mathbb{R}) \mid \lim_{|t| \to \infty} f(t) = 0 \}$

continuous fcts on $\mathbb{R}$ vanishing at $\pm\infty$

ii) $C_b(\mathbb{R}) := \{ f \in C(\mathbb{R}) \mid \exists M > 0 : |f(t)| \le M \ \forall t \in \mathbb{R} \}$

continuous bounded fcts on $\mathbb{R}$

2) Let μ and μ_N ($N \in \mathbb{N}$) be finite measures on $\mathbb{R}$. Then we say

i) μ_N <u>converges vaguely</u> to μ, denoted by $\mu_N \xrightarrow{v} \mu$, if

$$\int f(t)\, d\mu_N(t) \to \int f(t)\, d\mu(t) \qquad \forall f \in C_0(\mathbb{R})$$

ii) μ_N <u>converges weakly</u> to μ, denoted by $\mu_N \xrightarrow{w} \mu$, if

$$\int f(t)\, d\mu_N(t) \to \int f(t)\, d\mu(t) \qquad \forall f \in C_b(\mathbb{R})$$

<u>4.10 Remark:</u> 1) Note that weak convergence includes in particular that

$$\mu_N(\mathbb{R}) = \int 1\, d\mu_N(t) \to \int 1\, d\mu(t) = \mu(\mathbb{R})$$

and thus the weak limit of probability measures must again be a probability measure. For the vague convergence this is not true; there we can loose mass at infinity

Example: $\mu_N = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_N$; then

$$\int f(t)\,d\mu_N(t) = \frac{1}{2} f(1) + \frac{1}{2}\underbrace{f(N)}_{\to 0}$$

for $f \in C_0(\mathbb{R})$

$$\to \frac{1}{2} f(1)$$

$$= \int f(t)\,d\mu(t)$$

with $\mu = \frac{1}{2}\delta_1$

thus: $\underbrace{\frac{1}{2}\delta_1 + \frac{1}{2}\delta_N}_{\text{probability measure } \forall N} \xrightarrow{v} \underbrace{\frac{1}{2}\delta_1}_{\text{finite measure with } \mu(\mathbb{R}) = \frac{1}{2}}$

2) The relevance of the vague convergence, even if we are only interested in probability measures, is that the probability measures are compact in the vague topology, but not in the weak topology. E.g., in the above example, $\mu_N = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_N$ has no subsequence which converges weakly (but it has a subsequence, namely itself, which converges vaguely)

<u>4.11. Theorem</u>: The space of probability measures on $\mathbb{R}$ is pre-compact in the vague topology: Every sequence $(\mu_N)_{N\in\mathbb{N}}$ of probability measures on $\mathbb{R}$ has a subsequence which converges vaguely to a finite measure μ, with $\mu(\mathbb{R}) \le 1$.

<u>Proof</u>: 1) From a functional analytic perspective this is a special case of the Banach-Alaoglu theorem; as complex measure on $\mathbb{R}$ are the dual space of the Banach space $C_0(\mathbb{R})$, and its weak* topology is exactly the vague topology

2) From a measure theory perspective this is known as Helly's (Selection) Theorem. Here are the main ideas for the proof.

i) We describe a finite measure μ by its distribution function F_μ given by $F_\mu : \mathbb{R} \to \mathbb{R}$

$$F_\mu(t) := \mu((-\infty, t])$$

Such distribution functions are exactly functions F with the properties

- $t \mapsto F(t)$ is non-decreasing
- $0 = F(-\infty)$, $F(+\infty) < \infty$
- F is continuous on the right

ii) The vague convergence of $\mu_N \xrightarrow{v} \mu$ can also be described in terms of their distribution functions, F_N, F

$$\mu_N \xrightarrow{v} \mu \iff \begin{cases} F_N(t) \longrightarrow F(t) \\ \forall t \in \mathbb{R}, \text{ at which } F \\ \qquad \text{is continuous} \end{cases}$$

iii) Let now a sequence (μ_N) of prob. measures be given. We consider the corresponding distribution functions (F_N) and want to find a convergent subsequence (in the sense of (ii)) for those.

For this choose a countable dense subset $T = \{t_1, t_2, ..\}$ of $\mathbb{R}$. Then, by choosing subsequences of subsequences, and taking the "diagonal" subsequence, we get convergence for all $t \in T$

More precisely: Choose subsequence $(F_{N_1(m)})_m$ s.th

$$F_{N_1(m)}(t_1) \xrightarrow{m\to\infty} F_T(t_1),$$

choose then subsequence $(F_{N_2(m)})_m$ of this s.th

$$F_{N_2(m)}(t_1) \xrightarrow{m\to\infty} F_T(t_1)$$

$$F_{N_2(m)}(t_2) \xrightarrow{m\to\infty} F_T(t_2)$$

Iterating this gives subsequences $(F_{N_k(m)})_m$ s.th.

$$F_{N_k(m)}(t_i) \xrightarrow{m\to\infty} F_T(t_i) \quad \forall 1 \le i \le k$$

The diagonal subsequence $(F_{N_m(m)})_m$ converges then at all $t \in T$ to $F_T(t)$

Improve F_T to distribution function

$$F(t) := \inf\{F_T(s) \mid s \in T, s > t\}$$

and show that

- F is distribution function
- $F_{N_m(m)}(t) \xrightarrow{m\to\infty} F(t)$ at all points of continuity of F

Note that $F_N(+\infty) = 1 \quad \forall N$ gives $F(+\infty) \leq 1$, but we cannot guarantee $F(+\infty) = 1$ in general. □

<u>4.12. Remark:</u> If we want compactness in the weak topology, then we must control the mass at ∞ in a uniform way. This is given by the notion of tightness: a sequence $(\mu_N)_N$ of probability measures is <u>tight</u> if $\forall \varepsilon > 0 \;\exists\; I = [-R, R]$ s.th.

$$\mu_N(I^c) < \varepsilon \qquad \forall N$$

Then one has: Any tight sequence of probability measures has a subsequence which converges weakly (necessarily to a probability measure).

With this one can then also relate weak convergence with convergence of moments.

4.13. Def.: A probability measure μ on $\mathbb{R}$ is determined by its moments if

i) all moments $\int t^k d\mu(t) < \infty$ ($k \in \mathbb{N}$) exist

ii) μ is the only probability measure with those moments, i.e.

$$\left.\begin{array}{l} \nu \text{ prob measure} \\ \int t^k d\nu(t) = \int t^k d\mu(t) \\ \qquad \forall k \end{array}\right\} \Rightarrow \nu = \mu$$

4.14. Theorem: Let μ and μ_N ($N \in \mathbb{N}$) be probability measures for which all moments exist. Assume that μ is determined by its moments. Then the following are equivalent:

i) $\mu_N \xrightarrow{w} \mu$

ii) $\lim_{N \to \infty} \int t^k d\mu_N(t) = \int t^k d\mu(t) \quad \forall k \in \mathbb{N}$

Rough idea of proof: Note that tightness allows to control via measure theoretic arguments (like Fatou's lemma) convergence of moments

i) $\Rightarrow$ ii) weak convergence $\Rightarrow$ tightness

$\Rightarrow$ convergence of moments

ii) $\Rightarrow$ i) convergence of moments $\Rightarrow$ tightness

$\Rightarrow$ $\exists$ weakly convergent subsequence

$$\mu_{N_m} \xrightarrow{m\to\infty} \nu$$

$$\overset{i)\Rightarrow ii)}{\Longrightarrow} \quad \int t^k \, d\mu_{N_m} \xrightarrow{m\to\infty} \int t^k \, d\nu(t) \quad \forall k$$

$$\Big\downarrow (ii)\ m\to\infty$$

$$\int t^k \, d\mu(t)$$

$\Rightarrow$ μ and ν have same moments

$$\underset{\text{by moments}}{\overset{\mu \text{ determ}}{\Longrightarrow}} \quad \mu = \nu$$

In the same way all weakly convergent subsequences must converge to the same μ, thus the whole sequence converges weakly to μ.

<u>4.15. Remark</u> 1) Note that there exist also measures with all moments which are not determined by their moments. Weak convergence to them cannot be checked by just looking on convergence of moments.

Example: The log-normal distribution with density

$$d\mu(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-(\log x)^2/2} dt$$

on $[0, \infty)$

(distribution of e^X for X Gaussian) is not determined by its moments.

2) Compactly supported measures (e.g. semicircular) or also the Gaussian distribution are determined by their moments.

4.16. Theorem: Let μ and μ_N ($N \in \mathbb{N}$) be probability measures on $\mathbb{R}$. Then the following are equivalent.

i) $\mu_N \xrightarrow{w} \mu$

ii) For all $z \in \mathbb{C}^+$ we have:

$$\lim_{N \to \infty} S_{\mu_N}(z) = S_\mu(z)$$

iii) There exists a set $D \subset \mathbb{C}^+$, which has an accumulation point in $\mathbb{C}^+$, s.th.

$$\lim_{N \to \infty} S_{\mu_N}(z) = S_\mu(z) \quad \forall z \in D$$

Proof: (i) $\Rightarrow$ (ii) Assume that $\mu_N \xrightarrow{w} \mu$.

For $z \in \mathbb{C}^+$ consider $f_z : \mathbb{R} \to \mathbb{C}$ with

$$f_z(t) = \frac{1}{t-z}$$

Since $\lim_{|t|\to\infty} f_z(t) = 0$, we have

$f_z \in C_0(\mathbb{R}) \subset C_b(\mathbb{R})$ and thus, by definition of weak convergence:

$$\begin{array}{ccc} \int f_z(t)\, d\mu_N(t) & \longrightarrow & \int f_z(t)\, d\mu(t) \\ \| & & \| \\ S_{\mu_N}(z) & & S_\mu(z) \end{array}$$

(ii) $\Rightarrow$ (iii) clear

(iii) $\Rightarrow$ (i) By 4.11., we know that $(\mu_N)_N$ has a subsequence $(\mu_{N(m)})_m$ s.th.

$\mu_{N(m)} \xrightarrow[m\to\infty]{v} \nu$ for some finite measure ν with $\nu(\mathbb{R}) \leq 1$

Then, as above, we have for all $z \in D$:

$$S_\mu(z) = \lim_{m\to\infty} S_{\mu_{N(m)}}(z) = S_\nu(z)$$

Thus the analytic functions S_μ and S_ν agree on D and hence, by the identity theorem for analytic functions, also on $\mathbb{C}^+$, i.e. $S_\mu = S_\nu$, which implies, by 4.3., that $\nu = \mu$.
Thus the subsequence $(\mu_{N(m)})_m$ converges vaguely to the probability measure μ (and thus also weakly, see Ass 4, Ex 2).
In the same way, any weak cluster point of $(\mu_N)_N$ must be equal to μ, and thus the whole sequence must converge weakly to μ. $\square$

<u>4.17. Remark</u>: If we only assume that $S_{\mu_N}(z)$ converges to a limit function $S(z)$, then S must be the Stieltjes transform of a measure ν with $\nu(\mathbb{R}) \leq 1$ and we have the vague convergence $\mu_N \xrightarrow{v} \nu$