7. Stieltjes' transform of a probability measure

Definition 31. For $\mu \in \mathcal{P}(\mathbb{R})$, its *Stieltjes' transform* is defined as $G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-x}\mu(dx)$. It is well-defined on $\mathbb{C}\backslash\text{support}(\mu)$, in particular for $z \in \mathbb{H}$. If $X \sim \mu$, we can write $G_\mu(z) = \mathbf{E}\left[\frac{1}{z-X}\right]$.

Some simple observations on Stieltjes' transforms.

(a) For any $\mu \in \mathcal{P}(\mathbb{R})$, $|G_\mu(z)| \le \frac{1}{\operatorname{Im} z}$ for $z \in \mathbb{H}$.
(b) G_μ is analytic in $\mathbb{C}\backslash\text{support}(\mu)$, as can be seen by integrating over any contour (that does not enclose the support) and interchanging integrals (integrating $1/(z-x)$ gives zero by Cauchy's theorem).
(c) Suppose μ is supported on a compact interval $[-a,a]$. Then, its moments $m_k := \int x^k \mu(dx)$ satisfy $|m_k| \le a^k$ and hence $\sum m_k z^{-k-1}$ converges for $|z| > a$ and uniformly for $|z| \ge a+\delta$ for any $\delta > 0$. Hence,

$$\sum_{k=0}^{\infty} \frac{m_k}{z^{k+1}} = \mathbf{E}\left[\sum_{k=0}^{\infty} \frac{X^k}{z^k}\right] = \mathbf{E}\left[\frac{1}{z-X}\right] = G_\mu(z) \tag{9}$$

where the first equality follows by DCT. One can legitimately define $G_\mu(\infty) = 0$ and then (9) just gives the power series expansion of $w \to G_\mu(1/w)$ around 0. Since the power series coefficients are determined by the analytic function in any neighbourhood of 0, we see that if $G_\mu(z) = G_\nu(z)$ for all z in some open subset of $\mathbb{H}$, then $\mu = \nu$.
(d) For compactly supported *mu*, $G_\mu(z) \sim \frac{1}{z}$ as $z \to \infty$. If μ is not cmpactly supported, the same is true for $z = iy$ as $y \uparrow \infty$.

Equation (**??**) also shows that the Stieltjes transform is some variant of the moment generating function or the Fourier transform. Its usefulness in random matrix theory is analogous to the use of characteristic functions in proving central limit theorems. The following lemma gives analogues of Fourier inversion and Lévy's continuity theorems.

Lemma 32. *Let μ, ν be probability measures on $\mathbb{R}$.*

(1) For any $a < b$

$$\lim_{y \downarrow 0} \int_a^b -\frac{1}{\pi} \operatorname{Im}\{G_\mu(x+iy)\}dx = \mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\}.$$

(2) If $G_\mu(z) = G_\nu(z)$ for all z in an open subset of $\mathbb{H}$, then $\mu = \nu$.
(3) If $\mu_n \to \mu$, then $G_{\mu_n} \to G_\mu$ pointwise on $\mathbb{H}$.
(4) If $G_{\mu_n} \to G$ pointwise on $\mathbb{H}$ for some $G : \mathbb{H} \to \mathbb{C}$, then G is the Stieltjes' transform of a possibly defective measure. If further, $iyG(iy) \to 1$ as $y \uparrow \infty$, then, $G = G_\mu$ for a probability measure μ and $\mu_n \to \mu$.

Exercise 33. If μ has a continuous density f, then show that $f(x) = -\frac{1}{\pi}\lim_{y\downarrow 0} y \operatorname{Im}\{G_\mu(x+iy)\}$.

PROOF. (1) Observe that

$$\frac{-1}{\pi}\operatorname{Im} G_\mu(x+iy) = \frac{-1}{\pi}\int_{\mathbb{R}} \operatorname{Im}\left\{\frac{1}{x+iy-t}\right\}\mu(dt) = \int_{\mathbb{R}} \frac{1}{\pi}\frac{y}{(x-t)^2+y^2}\mu(dt).$$

The last quantity is the density of $\mu \star C_y$, where C_y is the Cauchy distribution with scale parameter y.

On some probability space, let X and Z be independent random variables such that $X \sim \mu$ and $Z \sim C_1$. Then by the above observation, we get

$$\int_a^b -\frac{1}{\pi} \operatorname{Im}\{G_\mu(x+iy)\}dx = \mathbf{P}(X+yZ \in [a,b]) = \mathbf{E}\left[\mathbf{1}_{X+yZ\in[a,b]}\right].$$

Observe that $\mathbf{1}_{X+yZ\in[a,b]} \to \mathbf{1}_{X\in(a,b)} + \mathbf{1}_{X=a,Z>0} + \mathbf{1}_{X=b,Z<0}$ as $y \downarrow 0$. Take expectations, apply DCT, and use independence of X and Z to get $\mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\}$.

(2) Follows immediately from the first part.

(3) If $\mu_n \to \mu$, then $\int f d\mu_n \to \int f d\mu$ for all bounded continuous functions f. For fixed $z \in \mathbb{H}$, the function $x \to \frac{1}{z-x}$ is bounded and continuous on $\mathbb{R}$ and hence $G_{\mu_n}(z) \to G_\mu(z)$.

(4) Conversely suppose that $G_{\mu_n} \to G$ pointwise for some function G. By Helly's selection principle, some subsequence μ_{n_k} converges vaguely to a possibly defective measure μ. As $(z-x)^{-1}$ is continuous and *vanishes at infinity*, $G_{\mu_{n_k}}(z) \to G_\mu(z)$ for all $z \in \mathbb{H}$.

Hence $G_\mu = G$ which shows that all subsequential limits have the same Stieltjes transform G. Further $iyG(iy) \to 1$ which shows that μ is a probability measure. By uniqueness of Stieltjes transforms, all subsequential limits are the same and hence $\mu_n \to \mu$. ■

Our next lemma gives a sharper version of the uniqueness theorem, by getting a bound on the Lévy distance between two probability measures in terms of the difference between their Stieltjes transforms.

8. Bounding Lévy distance in terms of Stieltjes transform

The following lemma is a quantitative statement that implies parts (2) and (4) of Lemma 32 as easy corollaries (how do you get part (4) of Lemma 32?).

Lemma 34. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$. Then, for any $y > 0$ and $\delta > 0$ we have*

$$\mathcal{D}(\mu,\nu) \le \frac{2}{\pi}\delta^{-1}y + \frac{1}{\pi}\int_{\mathbb{R}} |\operatorname{Im} G_\mu(x+iy) - \operatorname{Im} G_\nu(x+iy)|dx.$$

PROOF. Let $\mu_y = \mu \star C_y$ and $\nu_y = \nu \star C_y$. We bound the Lévy distance between μ and ν in three stages.

$$\mathcal{D}(\mu,\nu) \le \mathcal{D}(\mu_y,\mu) + \mathcal{D}(\nu_y,\nu) + \mathcal{D}(\mu_y,\nu_y).$$

By the proof of Lemma 32 we know that μ_y has density $-\pi^{-1}\operatorname{Im} G_\mu(x+iy)$ and similary for ν_y. Hence, by exercise 35

$$\mathcal{D}(\mu_y,\nu_y) \le \frac{1}{\pi}\int_{\mathbb{R}} |\operatorname{Im} G_\mu(x+iy) - \operatorname{Im} G_\nu(x+iy)|dx. \tag{10}$$

Next we control $\mathcal{D}(\mu_y,\mu)$. Let $X \sim \mu$ and $Z \sim C_1$ so that $V = X + yZ \sim \mu_y$. For $t > 0$ observe that $\mathbf{P}(Z > t) = \int_t^\infty \pi^{-1}(1+u^2)^{-1}du \le \int_t^\infty \pi^{-1}u^{-2}du = \pi^{-1}t^{-1}$. Thus, for any $\delta > 0$, we get

$$\mathbf{P}(X \le t,\, V > t+\delta) \le \mathbf{P}\left(Z > y^{-1}\delta\right) \le \pi^{-1}\delta^{-1}y$$
$$\mathbf{P}(V \le t,\, X > t+\delta) \le \mathbf{P}\left(Z < -y^{-1}\delta\right) \le \pi^{-1}\delta^{-1}y.$$

These immediately give $\mathcal{D}(\mu,\mu_y) \le \pi^{-1}\delta^{-1}y$. Similarly $\mathcal{D}(\nu,\nu_y) \le \pi^{-1}\delta^{-1}y$. Combine with (10) to get the inequality in the statement. ■

Exercise 35. Let μ and ν have densities f and g respectively. Then show that $\mathcal{D}(\mu,\nu) \le \int |f-g|$ (the latter is called the total variation distance between μ and ν).

9. Heuristic idea of the Stieltjes' transform proof of WSL for GOE

Let X_n be a GOE matrix. Let $A_n = \frac{1}{\sqrt{n}}X_n$ have eigenvalues λ_k and ESD L_n. The Stieltjes' transform of L_n is

$$G_n(z) := \int \frac{1}{z-x}L_n(dx) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{z-\lambda_k} = \frac{1}{n}\mathrm{tr}(zI-X_n)^{-1}.$$

We show that $L_n \to \mu_{s.c}$ by showing that $G_n(z) \to G_{s.c}(z)$ for all $z \in \mathbb{H}$. By Lemma **??**, this proves the claim. We are being a little vague about the mode of convergence but that will come in a moment.[6]

We introduce the following notations. We fix n for now. Y_k will denote the matrix obtained from X by deleting the k^{th} row and the k^{th} column. And $\mathbf{u}_k \in \mathbb{C}^{n-1}$ will denote the column vector got by deleting the k^{th} entry in the k^{th} column of X.

From the formulas for the entries of the inverse matrix, we know that for any M,

$$(zI-A)^{k,k} = \frac{1}{z-\frac{1}{\sqrt{n}}X_{k,k}-\frac{1}{n}\mathbf{u}_k^*(zI-\frac{1}{\sqrt{n}}Y_k)^{-1}\mathbf{u}_k}$$

and hence letting V_k denote the denominator on the right side, we can write

$$G_n(z) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{V_k}. \tag{11}$$

The key observations are

(1) Y_k is just an $(n-1)$-dimensional GOE matrix.
(2) $\mathbf{u}_k$ is a standard Gaussian vector in $(n-1)$-dimensions.
(3) Y_k and $\mathbf{u}_k$ are independent.

Therefore,

$$\begin{aligned}\mathbf{E}[V_1] &= z-\frac{1}{n}\mathbf{E}\left[\mathbf{E}\left[\mathbf{u}_1^*(zI-Y_1)^{-1}\mathbf{u}_1 \mid \frac{1}{\sqrt{n}}Y_1\right]\right] \\ &= z-\frac{1}{n}\mathbf{E}\left[\mathrm{tr}(zI-\frac{1}{\sqrt{n}}Y_1)^{-1}\right] \\ &\approx z-\mathbf{E}[G_{n-1}(z)]. \end{aligned} \tag{12}$$

provided we ignore the difference between n and $n-1$. As V_k are identically distributed, $\mathbf{E}[V_k]$ is equal to the same quantity.

Let us assume that each V_k is very close to its expectation. This will be a consequence of high dimensionality and needs justification. Then return to (11) and write

$$G_n(z) \approx \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\mathbf{E}[V_k]} = \frac{1}{z-\frac{n-1}{n}\mathbf{E}[G_{n-1}(z)]}.$$

[6]The method of Stieltjes' transform for the study of ESDs, as well as the idea for getting a recursive equation for G_n is originally due to the physicist Leonid Pastur **?**. The method was pioneered in many papers by Zhidong Bai.

There are two implications in this. Firstly, the random quantity on the left is close to the non-random quantity on the right, and hence if we assume that $\mathbf{E}[G_n(\cdot)]$ converges to some $G(\cdot)$, then so that $G_n(\cdot)$, and to the same limit. Secondly, for G we get the equation

$$G(z) = \frac{1}{z - G(z)}.$$

This reduces to the quadratic equation $G(z)^2 - zG(z) + 1 = 0$ with solutions $G(z) = (z \pm \sqrt{z^2-4})/2$. By virtue of being Stieltjes' transforms, $G_n(z) \sim z^{-1}$ as $z \to \infty$ and G must inherit this property. Thus we are forced to take $G(z) = (z - \sqrt{z^2-4})/2$ where the appropriate square root is to be chosen. By direct calculation, the Stieltjes transform of $\mu_{s.c}$ is identified to be the same. This completes the heuristic.

Exercise 36. Show that $G = G_{\mu_{s.c}}$ satisfies the equation $(G(z))^2 - zG(z) + 1 = 0$ for all $z \in \mathbb{H}$. Argue that no other Stieltjes' transform satisfies this equation. One can then write

$$G(z) = \frac{z - \sqrt{z^2-4}}{2}$$

where the branch of square root used is the one defined by $\sqrt{re^{i\theta}} = \sqrt{r}e^{i\theta/2}$ with $\theta \in (-\pi, \pi)$. Expand by Binomial theorem and verify that the even moments are given by Catalan numbers.

10. The Stieltjes' transform proof of WSL

Now for the rigorous proof. The crucial point in the heuristic that needs justification is that V_k is close to its expected value. The following two lemmas will come in handy.

Lemma 37. *Let V be a complex valued random variable and assume that almost surely, $\operatorname{Im} V \geq t$ for some constant $t > 0$. Then, for any $p > 0$*

$$\mathbf{E}\left[\,\left|\frac{1}{V} - \frac{1}{\mathbf{E}\,V}\right|^p\right] \leq t^{-2p}\mathbf{E}\left[|V - \mathbf{E}\,V|^p\right].$$

PROOF. Almost surely, $\operatorname{Im} V \geq t$ and hence $\operatorname{Im}\{\mathbf{E}\,V\} \geq t$ too. Hence, $|V| \geq t$ a.s., and $|\mathbf{E}\,V| \geq t$. Thus,

$$\left|\frac{1}{V} - \frac{1}{\mathbf{E}\,V}\right| = \frac{|V - \mathbf{E}\,V|}{|V|\,|\mathbf{E}\,V|} \leq t^{-2}|V - \mathbf{E}\,V|.$$

Raise to power p and take expectations. ■

Lemma 38. *Let $\mathbf{u}$ be an $n \times 1$ random vector where u_i are independent real or complex valued random variables with zero mean and unit variance. Let M be a non-random $n \times n$ complex matrix. Then,*

(a) $\mathbf{E}[\mathbf{u}^* M \mathbf{u}] = trM$.

(b) If in addition $m_4 := \mathbf{E}[|u_i|^4] < \infty$, then $Var(\mathbf{u}^* M \mathbf{u}) \leq (2 + m_4) tr(M^* M)$.

PROOF. Write $\mathbf{u}^* M \mathbf{u} = \sum_{i,j=1}^n M_{i,j}\bar{u}_i u_j$. When we take expectations, terms with $i \neq j$ vanish and those with $i = j$ give $M_{i,i}$. The first claim follows. To find the variance,[7] we compute the second moment $\mathbf{E}\left[|\mathbf{u}^* M \mathbf{u}|^2\right] = \sum_{i,j}\sum_{k,\ell} M_{i,j}\overline{M}_{k,\ell}\mathbf{E}[\bar{u}_i u_j \bar{u}_k u_\ell]$.

$\mathbf{E}[\bar{u}_i u_j \bar{u}_k u_\ell]$ vanishes unless each index appears at least twice. Thus, letting $m_2 = \mathbf{E}[u_1^2]$

$$\mathbf{E}[\bar{u}_i u_j \bar{u}_k u_\ell] = \delta_{i,j}\delta_{k,\ell} + \delta_{i,\ell}\delta_{j,k} + |m_2|^2\delta_{i,k}\delta_{j,\ell} + m_4\delta_{i,j,k,\ell}.$$

[7] For a complex-valued random variable Z, by $\operatorname{Var}(Z)$ we mean $\mathbf{E}[|Z - \mathbf{E}[Z]|^2] = \mathbf{E}[|Z|^2] - |\mathbf{E}[Z]|^2$. This is consistent with the usual definition if Z is real-valued.

Thus

$$\begin{aligned}
\mathbf{E}\left[|\mathbf{u}^*M\mathbf{u}|^2\right] &= \sum_{i,k} M_{i,i}\overline{M}_{k,k} + \sum_{i,j} M_{i,j}\overline{M}_{j,i} + |m_2|^2 \sum_{i,j} M_{i,j}\overline{M}_{i,j} + m_4 \sum_i M_{i,i}\overline{M}_{i,i} \\
&= (\mathrm{tr}M)^2 + \mathrm{tr}(M^*M^t) + |m_2|^2 \mathrm{tr}(M^*M) + m_4 \sum_i |M_{i,i}|^2 \\
&\leq (\mathrm{tr}M)^2 + (1+|m_2|^2+m_4)\mathrm{tr}(M^*M).
\end{aligned}$$

Observe that $|m_2|^2 \leq \mathbf{E}[|u_1|^2] \leq 1$ where equality may not hold as u_1 is allowed to be complex valued. Subtract $\mathbf{E}[\mathbf{u}^*M\mathbf{u}]^2 = (\mathrm{tr}M)^2$ to get $\mathrm{Var}(\mathbf{u}^*M\mathbf{u}) \leq (2+m_4)\mathrm{tr}(M^*M)$. ■

Now we are ready to prove Wigner's semicircle law under fourth moment assumption.

Theorem 39. *Let X_n be a Wigner matrix. Assume $m_4 = \max\{\mathbf{E}[|X_{1,2}|^4], \mathbf{E}[X_{1,1}^4]\}$ is finite. Then, $L_n \xrightarrow{P} \mu_{s.c}$ and $\overline{L}_n \to \mu_{s.c}$.*

PROOF. Let G_n and $\overline{G}_n$ denote the Stieltjes' transforms of L_n and $\overline{L}_n$ respectively. Of course, $\overline{G}_n(z) = \mathbf{E}[G_n(z)]$. Fix $z \in \mathbb{H}$. From (11) we have $G_n(z) = n^{-1}\sum_{k=1}^n 1/V_k$ where

$$V_k = (zI - X)^{k,k} = z - \frac{X_{k,k}}{\sqrt{n}} - \frac{1}{n}\mathbf{u}_k^*\left(zI - \frac{Y_k}{\sqrt{n}}\right)^{-1}\mathbf{u}_k. \tag{13}$$

Here Y_k is the $(n-1)\times(n-1)$ matrix obtained from X by deleting the k^{th} row and k^{th} column, and $\mathbf{u}_k$ is the $(n-1)\times 1$ vector obtained by deleting the k^{th} element of the k^{th} column of X. Clearly Y_k is a Wigner matrix of dimension $(n-1)$ and $\mathbf{u}_k$ is a vector of iid copies of $X_{1,2}$, and $\mathbf{u}_k$ is independent of Y_k. We rewrite (13) as

$$V_k = z - \frac{X_{k,k}}{\sqrt{n}} - \frac{1}{\sqrt{n(n-1)}}\mathbf{u}_k^*\left(z_nI - \frac{Y_k}{\sqrt{n-1}}\right)^{-1}\mathbf{u}_k, \qquad \text{where } z_n := \frac{\sqrt{n}}{\sqrt{n-1}}z. \tag{14}$$

Hence,

$$\begin{aligned}
\mathbf{E}\left[\left|\,G_n(z) - \frac{1}{\mathbf{E}[V_1]}\,\right|^2\right] &= \mathbf{E}\left[\left|\frac{1}{n}\sum_{k=1}^n\left(\frac{1}{V_k} - \frac{1}{\mathbf{E}[V_k]}\right)\right|^2\right] \\
&\leq \mathbf{E}\left[\frac{1}{n}\sum_{k=1}^n\left|\frac{1}{V_k} - \frac{1}{\mathbf{E}[V_k]}\right|^2\right] \qquad \text{(by Cauchy-Schwarz)} \\
&\leq \frac{1}{(\mathrm{Im}\,z)^4}\mathbf{E}[|V_1 - \mathbf{E}[V_1]|^2] \qquad \text{by Lemma 38.}
\end{aligned}$$

For a complex-valued random variable with finite second moment, $\mathbf{E}[|Z-c|^2]$ is minimized uniquely at $c = \mathbf{E}[Z]$. In particular we also have $|\mathbf{E}[Z]|^2 \leq \mathbf{E}[|Z|^2]$. Therefore, the above inequality implies the following two inequalities.

$$\mathrm{Var}(G_n(z)) \leq \frac{1}{(\mathrm{Im}\,z)^4}\mathbf{E}[|V_1 - \mathbf{E}[V_1]|^2] \tag{15}$$

$$\left|\,\overline{G}_n(z) - \frac{1}{\mathbf{E}[V_1]}\,\right|^2 \leq \frac{1}{(\mathrm{Im}\,z)^4}\mathbf{E}[|V_1 - \mathbf{E}[V_1]|^2]. \tag{16}$$

The next step is to compute $\mathbf{E}[V_1]$ and obtain a bound for $\mathrm{Var}(V_1)$. Firstly,

$$\begin{aligned}
\mathbf{E}[V_1] &= z - 0 - \frac{1}{\sqrt{n(n-1)}}\mathbf{E}\left[\mathbf{E}\left[\mathbf{u}_k^*\left(z_nI - \frac{Y_1}{\sqrt{n-1}}\right)^{-1}\mathbf{u}_k \,\Big|\, Y_1\right]\right] \\
&= z - \frac{\sqrt{n-1}}{\sqrt{n}}\frac{1}{n-1}\mathbf{E}\left[\mathrm{tr}\left(z_nI - \frac{Y_1}{\sqrt{n-1}}\right)^{-1}\right] \\
&= z - \sqrt{\frac{n-1}{n}}\,\overline{G}_{n-1}(z_n). \qquad (17)
\end{aligned}$$

Now, to estimate $\mathrm{Var}(V_1)$, recall that $X_{1,1}$, $\mathbf{u}_1$ and Y_1 are all independent. Write $A = (zI - \frac{Y_1}{\sqrt{n}})^{-1}$ and $B = (z_nI - \frac{Y_1}{\sqrt{n-1}})^{-1}$ and observe that if θ_j are eigenvalues of Y_1 then the eigenvalues of A and B are $(z-\theta_j/\sqrt{n})^{-1}$ and $(z-\theta_j/\sqrt{n-1})^{-1}$ both of which are bounded in absolute value by $(\mathrm{Im}\,z)^{-1}$.

Write $\mathrm{Var}(V_1)$ as $\mathbf{E}\left[\mathrm{Var}(V_1 \,|\, Y_1)\right] + \mathrm{Var}\left(\mathbf{E}[V_1 \,|\, Y_1]\right)$. We evaluate the two individually as follows. Using the expression (13) and part (b) of Lemma 38 for $\mathrm{Var}(V_1 \,|\, Y_1)$ we get

$$\mathbf{E}\left[\mathrm{Var}(V_1 \,|\, Y_1)\right] = \mathbf{E}\left[n^{-1} + n^{-2}(2+m_4)\mathrm{tr}(A^*A)\right] \le n^{-1} + m_4n^{-1}(\mathrm{Im}\,z)^{-2}.$$

Using the expression (14) we get $\mathbf{E}[V_1 \,|\, Y_1] = z - \sqrt{\frac{n-1}{n}}G_{n-1}(z_n)$ and hence

$$\mathrm{Var}\left(\mathbf{E}[V_1 \,|\, Y_1]\right) = \frac{n-1}{n}\mathrm{Var}\left(G_{n-1}(z_n)\right) \le \mathrm{Var}\left(G_{n-1}(z_n)\right).$$

Add this to the inequality for $\mathbf{E}\left[\mathrm{Var}(V_1 \,|\, Y_1)\right]$ gives a bound for $\mathrm{Var}(V_1)$ which when inserted into (15) gives

$$\mathrm{Var}(G_n(z)) \le \frac{1}{(\mathrm{Im}\,z)^4}\left(\frac{1}{n} + \frac{m_4}{n(\mathrm{Im}\,z)^2} + \mathrm{Var}\left(G_{n-1}(z_n)\right)\right).$$

Let $V_n := \sup\{\mathrm{Var}(G_n(z)) : \mathrm{Im}\,z \ge 2\}$. Observe that $V_n \le 2^{-2}$ as $|G_n(z)| \le (\mathrm{Im}\,z)^{-1}$, in particular V_n is finite. Since $\mathrm{Im}\,z_n > \mathrm{Im}\,z$, we arrive at the recursive inequality

$$V_n \le \frac{1}{2^4n} + \frac{m_4}{2^6n} + \frac{1}{2^4}V_{n-1} \le \frac{A}{n} + \frac{1}{2}V_{n-1}$$

where $A = 2^{-2} + 2^{-6}m_4$. We increased the first term from 2^4 to 2^{-2} so that $V_1 \le A$ also. Iterating this inequality gives

$$\begin{aligned}
V_n &\le \frac{A}{n} + \frac{A}{2(n-1)} + \frac{A}{2^2(n-2)} + \ldots + \frac{A}{2^{n-2}2} + \frac{A}{2^{n-1}} \\
&\le \frac{A}{n/2}\sum_{k=0}^{n/2-1}\frac{1}{2^k} + \frac{A}{2^{n/2}}\frac{n}{2} \\
&\le \frac{5A}{n} \qquad (\text{for } n \ge 10).
\end{aligned}$$

Insert this into (15) and (16) and use (17) to get

$$\sup_{\mathrm{Im}\,z\ge 2}\mathrm{Var}(G_n(z)) \le \frac{1}{n} \qquad (18)$$

$$\sup_{\mathrm{Im}\,z\ge 2}\left|\,\overline{G}_n(z) - \frac{1}{z - \sqrt{(n-1)/n}\,\overline{G}_{n-1}(z_n)}\,\right|^2 \le \frac{1}{n}. \qquad (19)$$

Convergence of $\overline{L}_n$ to semicircle: $\overline{L}_n$ is a sequence of probability measure with Stieltjes transforms $\overline{G}_n$. Let μ be any subsequential limit of $\overline{L}_n$, a priori allowed to be a defective measure. By (19) G_μ must satisfy $G_\mu(z)(z-G_\mu(z))=1$ for all z with $\operatorname{Im} z \ge 2$ (why? Justification is needed to claim that $\overline{G}_{n-1}(z_n) \to G_\mu(z)$, but one can argue this by using equicontinuity of $\overline{G}_n$ as in the next paragraph). Thus, $G_\mu(z) = (z \pm \sqrt{z^2-4})/2$. Since G_μ must be analytic in z and $G_\mu(z) \sim \mu(\mathbb{R}) z^{-1}$ as $z \to \infty$, the branch of square root is easily fixed. We get

$$G_\mu(z) = \frac{z+\sqrt{z^2-4}}{2}, \qquad \text{for } \operatorname{Im} z \ge 2$$

where the square root is the branch $\sqrt{re^{i\theta}} = \sqrt{r}e^{i\theta/2}$ with $\theta \in (-\pi,\pi)$. By exercise 36 this is precisely the Stieltjes transform of the semicircle distribution on $[-2,2]$. Thus all subsequential limits of $\overline{L}_n$ are the same and we conclude that $\overline{L}_n \to \mu_{s.c}$.

Convergence of L_n to semicircle: Without loss of generality, let X_n be defined on the same probability space for all n.[8] If $\sum 1/n_k < \infty$, then by (18) it follows that for fixed z with $\operatorname{Im} z \ge 2$ we have $G_{n_k}(z) - \overline{G}_{n_k}(z) \xrightarrow{a.s.} 0$. Take intersection over a countable dense subset S of z and invoke the convergence of $\overline{G}_n$ to conclude that $G_{n_k}(z) \to G_{s.c}(z)$ for all $z \in S$, almost surely. For a Stieltjes transform G, we have the inequality $|G'(z)| \le (\operatorname{Im} z)^{-2}$, from which we see that G_n are equicontinuous on $\{\operatorname{Im} z \ge 2\}$. Therefore we get $G_{n_k}(z) \to G(z)$ for all z with $\operatorname{Im} z \ge 2$, almost surely. Hence $L_{n_k} \xrightarrow{a.s.} \mu_{s.c}$.

Now, given any subsequence $\{n_k\}$, choose a further subsequence $\{n_{k_\ell}\}$ such that $\sum 1/n_{k_\ell} < \infty$. Then $L_{n_{k_\ell}} \xrightarrow{a.s.} \mu_{s.c}$. Thus every subsequence has an almost sure convergent sub-sub-sequence. Therefore $L_n \xrightarrow{P} \mu$. ■

Remark 40. If we had used Lemma 37 with $p=4$ instead of $p=2$ (which would force the assumption that $X_{i,j}$ have finite eighth moment), then we could get n^{-2} as a bound for $\mathbf{E}\left[\left|G_n(z) - \frac{1}{\mathbf{E}[V_1]}\right|^2\right]$. Therefore we would get almost sure convergence.

In fact, one can conclude almost sure convergence assuming only finite second moment! This requires us to use $p=1$, but then we are faced with estimating $\mathbf{E}[|V_1 - \mathbf{E}[V_1]|]$ which is more complicated than estimating the variance. Lastly, Stieltjes' transform methods are very powerful, and can be used to prove rates of convergence in Wigner's semicircle law.

Exercise 41. Prove Theorem 24 by Stieltjes transform methods. Mainly, work out the heuristic steps in the proof and arrive at an equation for the Stieltjes transform of the limiting measure and show that the equation is satisfied uniquely by the Stieltjes transform of the Marcenko-Pastur law. The full details will involve similar technicalities and may be omitted.

---

[8] The strategy used here is as follows. To show that real-valued random variables Y_n converge in probability to zero, we may first of all construct random variables Z_n on the same probability space so that $Z_n \stackrel{d}{=} Y_n$ and then show that Z_n converge in probability to zero. And for the latter, it suffices to show that any subsequence has a further subsequence that converges *almost surely* to zero.