



Centro de Investigación en Matemáticas, A.C.

---

# Ley local del semicírculo, con extensión y aplicaciones a matrices ralas

T E S I S

Que para obtener el grado de  
**Maestro en Ciencias**  
con especialidad en  
**Probabilidad y Estadística**

**P r e s e n t a:**

Lic. Diego de Jesus Caudillo Amador

**Director de tesis**

Dr. Mario Alberto Diaz Torres

A handwritten signature in black ink, appearing to read "M. A. Diaz Torres", is positioned above a horizontal line.

Autorización de la versión final

Guanajuato, Gto. 22 de noviembre del 2019



# AGRADECIMIENTOS

Agradezco a mi asesor por todo el trabajo a lo largo de estos meses, su atención y consejos.

Gracias al Centro de Investigación en Matemáticas (CIMAT), en especial a sus profesores quienes sirven como fuente de inspiración y están dispuestos a compartir su experiencia. Agradezco también al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico con una beca de maestría asignada al CVU 861333.

Gracias a mi madre porque siempre ha confiado en mi y por mantenerme con los pies en la tierra. Agradezco a mi hermana por el apoyo incondicional durante las etapas difíciles.



# RESUMEN

En teoría de la probabilidad hay un área que estudia propiedades espectrales de matrices grandes con entradas aleatorias. La propiedad más importante es la Ley del Semicírculo de Wigner, que da origen a la teoría de matrices aleatorias. Esta investigación replica resultados de esta teoría y explora sus alcances asintóticos, numéricos y prácticos.

En particular se presenta una demostración elemental de la Ley Local del Semicírculo. Esta es un refinamiento de la ley de Wigner al combinarse con la transformada inversa de Stieltjes, otro resultado límite. Esta combinación otorga cotas para el error de aproximación. Las leyes locales para matrices aleatorias son una rama reciente en la teoría y suelen vincularse con el concepto de universalidad. Se pretende que esta tesis sirva como una primera introducción al tema de leyes locales sin requerir de una notación específica ni herramientas analíticas avanzadas como es lo usual en la literatura del área. Esta intención se refleja en la simplicidad de la notación y la de la herramienta clave, una desigualdad de concentración elemental para vectores gaussianos.

Se extienden las ideas de dicha prueba para presentar resultados en la literatura sobre matrices aleatorias ralas. En esta tesis, un resultado reciente se aborda desde un punto de vista numérico para ejemplificar y complementar los resultados teóricos.

**Palabras clave:** *matrices aleatorias, ley de Wigner, leyes locales, ley local del semicírculo, GUE, desigualdades de concentración, matrices ralas*



# ÍNDICE GENERAL

<b>Resumen</b>	<b>v</b>
<b>1. Introducción</b>	<b>1</b>
1.1. La teoría de Leyes Locales . . . . .	1
1.1.1. Intuición detrás de la teoría . . . . .	2
1.1.2. Estado del arte en Leyes Locales . . . . .	2
1.2. Descripción del Contenido . . . . .	3
<b>2. Preliminares de Matrices Aleatorias</b>	<b>5</b>
2.1. La Transformada de Stieltjes . . . . .	5
2.2. Ecuación Funcional del Semicírculo . . . . .	7
2.3. Distribución Empírica Espectral . . . . .	9
2.4. Resolvente de una Matriz Aleatoria . . . . .	10
2.5. Minores Matriciales . . . . .	13
2.6. Funciones Lipschitz sobre Vectores Gaussianos . . . . .	13
2.7. Más Sobre Funciones Lipschitz . . . . .	15
<b>3. Ley Local del Semicírculo</b>	<b>17</b>
3.1. Ensamble Gaussiano Unitario . . . . .	18
3.2. La Ley del Semicírculo . . . . .	18
3.3. Ecuación Funcional . . . . .	18
3.4. Ley Global del Semicírculo . . . . .	21
3.4.1. Decaimiento del Primer Término . . . . .	21
3.4.2. Decaimiento del Segundo Término . . . . .	23
3.4.3. Decaimiento del Tercer Término . . . . .	29
3.4.4. Ley del Semicírculo de Wigner . . . . .	32
3.5. Proceso Bootstrap . . . . .	34
3.5.1. El Argumento de Continuidad . . . . .	36
3.5.2. Sucesión y Dominio de Continuidad . . . . .	37
3.5.3. Demostración del Teorema . . . . .	38
3.6. Discusión . . . . .	40
3.6.1. Comparación con la Literatura . . . . .	40
3.6.2. Interpretaciones . . . . .	41
3.6.3. Implicaciones en la Elección del Modelo . . . . .	42

<b>4. Matrices Aleatorias Ralas</b>	<b>45</b>
4.1. Modelación Estocástica de una Matriz Rala . . . . .	45
4.1.1. Parámetro de Raleza . . . . .	45
4.1.2. Decaimiento de las Entradas . . . . .	46
4.1.3. Estructura Conjunta . . . . .	48
4.1.4. Ejemplos de Modelos Matriciales Ralos . . . . .	48
4.2. Leyes Locales para Matrices Ralas Simétricas . . . . .	49
4.2.1. Recordatorio de la Notación . . . . .	50
4.2.2. Universalidad y Sesgo . . . . .	50
4.2.3. Leyes Locales Universal y Sesgada . . . . .	51
4.2.4. Leyes Locales para Matrices Ralas Multibloque . . . . .	52
4.3. Análisis de una Ley Local para Matrices Ralas Rectangulares . . . . .	53
4.3.1. La Ley de Marchenko-Pastur . . . . .	54
4.3.2. Ley Local Sesgada . . . . .	54
4.3.3. Propiedades Asintóticas de la Ley . . . . .	55
4.3.4. Análisis Desde una Perspectiva Numérica . . . . .	57
<b>A. Material Adicional</b>	<b>61</b>
A.1. Leyes Locales con el Método de Stein . . . . .	61
A.2. Código en R para Ley Sesgada . . . . .	63
<b>Referencias</b>	<b>67</b>

# CAPÍTULO 1

## INTRODUCCIÓN

Un dicho popular reza que el espectro de una matriz es como su ADN. Ciertamente es que conocer el espectro arroja mucha información sobre la matriz en sí misma. En la práctica hay matrices cuyas entradas son aleatorias: tablas de observaciones  $X$ , matrices de covarianza empírica  $X^T X$  o proveniente de gráficas con aristas aleatorias  $A_G$ . Entender las propiedades espectrales de estas matrices es un acercamiento a entender las matrices mismas.

En Teoría de la Probabilidad se estudia el espectro de matrices aleatorias desde la década de 1950, cuando Wigner demuestra la Ley del Semicírculo, un equivalente matricial al Teorema del Límite Central. En [Wig55] tomó un tipo muy especial de matrices deterministas y representó su espectro como si fuera una densidad de probabilidad. Luego supuso que la matriz era aleatoria, por lo que la densidad propuesta también era aleatoria y demostró que cuando el tamaño de la matriz crece, esa densidad converge a una densidad límite determinista. En trabajos posteriores generaliza este resultado a una clase más general de matrices simétricas. Inspirada en este resultado [Tao12] se desarrolla el área de matrices aleatorias con diferentes enfoques, metodologías e intereses.

Uno de los enfoques más recientes es el de las leyes locales, que surge en 2009 como un refinamiento a teoremas límite ya existentes con el interés de encontrar cotas de error. En la última década el enfoque de leyes locales desarrolló su propia notación y la usa para resolver problemas originales. Uno de estos problemas es el de describir matrices cuyas entradas tienen una probabilidad muy grande de ser cero. En la práctica estas matrices son conocidas como matrices ralas y son objeto común en aplicaciones de ciencia, ingeniería y aprendizaje máquina. Las leyes locales innovan en resultados para este problema al incorporar la velocidad en que la matriz se llena de valores cero dentro de la velocidad de convergencia [Erd+13b].

En esta investigación se tomó el enfoque de leyes locales como eje central y matrices ralas como interés particular. El trabajo se presenta como una introducción a la teoría y una exploración de los alcances de sus resultados a través de ejercicios numéricos o aplicaciones.

### 1.1. LA TEORÍA DE LEYES LOCALES

La teoría de leyes locales surge en el contexto de matrices aleatorias a partir del método del resolvente [ESY09], el cual es un conjunto de ideas y técnicas analítico-recursivas con base a desigualdades de largas desviaciones. Una de ellas es el entrelazado de Cauchy que relaciona el espectro de una matriz con el de una sub-matriz. A lo largo del texto se explicarán las partes de este método. El artículo [KY17] contiene una introducción muy completa a la metodología específica en leyes locales. También se trabaja en matrices aleatorias con el método combinatorio

y la teoría analítica [Tao12], de las cuáles no se tiene noticia de que tengan resultados locales.

### 1.1.1. INTUICIÓN DETRÁS DE LA TEORÍA

A continuación se presentan dos interpretaciones de la teoría de leyes locales. Una desde el punto de vista estadístico y otro desde una perspectiva probabilista.

La primera es que dada una matriz, se intenta estudiar su espectro como si fuera una densidad de probabilidad. En estadística es común estimar densidades usando kernels Gaussianos, *i.e.* convertir el histograma en una curva suave usando la densidad de la normal. En la práctica equivale a suponer que las observaciones tienen un error de medición Gaussiano. Formalmente es la convolución de la distribución empírica y una normal centrada. En el caso de matrices aleatorias las estimaciones usan kernels Cauchy. El objetivo es encontrar una resolución, o parámetro de escala Cauchy, que dé resultados estables y que al mismo tiempo aporte información significativa. Las leyes locales proveen directamente ese parámetro.

La segunda es más técnica y está basada en la transformada inversa de Stieltjes. La transformada de Stieltjes es un análogo para matrices aleatorias de la transformada de Fourier, es decir tiene una inversa y cumple un teorema de continuidad. La inversa misma es un resultado límite. La idea es combinar dos límites en uno sólo planteando una relación polinomial entre los índices. Un límite usa la continuidad de las transformadas de distribuciones. El otro límite obtiene la distribución aplicando la transformada inversa.

Por ejemplo, para la ley de Wigner se manejan densidades  $\rho_{N,\eta}(x) = (\delta_\eta * \tilde{\rho}_N)(x)$ , donde  $\delta_\eta$  es una densidad Cauchy(0,  $\eta$ ), mientras que  $\tilde{\rho}_N$  depende de los valores propios de una matriz de tamaño  $N$ . La ley de Wigner establece que existe una densidad  $\rho_{sc}$  determinista para la que se cumple

$$\lim_{\eta \rightarrow 0^+}^{\mathbf{P}} \lim_{N \rightarrow \infty}^{\mathbf{P}} \rho_{N,\eta}(x) = \rho_{sc}(x),$$

la ley Local del Semicírculo refina este resultado en

$$\lim_{\substack{N \rightarrow \infty \\ \eta \rightarrow 0^+ \\ N\eta \rightarrow \infty}}^{\mathbf{P}} \rho_{N,\eta}(x) = \rho_{sc}(x).$$

Las leyes locales se llaman así por un concepto que surge a finales de 1990, toma fuerza a principios de los 2000 y se consolida en el 2009 con [ESY09]. La idea es que los resultados asintóticos tratan de describir el comportamiento general de la distribución, de forma global o *desde afuera*. Mientras tanto hay resultados que intentan entender los resultados a pequeña escala, por ejemplo la distancia promedio entre dos valores propios consecutivos o la fluctuación del valor propio más grande. Trabajar con pequeñas resoluciones, debido a la transformada inversa de Stieltjes, implica trabajar con la distribución a pequeña escala. Esta relación pequeña escala y resultados globales es lo que le da al resultado en [ESY09] el nombre de Ley Local, que es el que tomarán después artículos similares.

### 1.1.2. ESTADO DEL ARTE EN LEYES LOCALES

El primer resultado de esta teoría es la Ley Local del Semicírculo [ESY09], que es un refinamiento a un antiguo teorema llamado Ley del Semicírculo de Wigner. En años posteriores este

resultado ha servido de inspiración a resultados más complejos como gráficas regulares [BHY16], polinomios de matrices [EKN18], matrices ralas [LS18] o ensambles no isotrópicos [KY17]. Ya ha desarrollando su propia notación [Erd+13b] y se ha adaptado a estudiar no sólo el espectro, sino también los vectores propios [Blo+14; KY17; BK16], además de que han reducido sus hipótesis hasta puntos donde ya es razonable considerarlos óptimos [GNT19].

Los resultados más recientes podrían clasificarse *grosso modo* en tres categorías. Está la idea de universalidad [EKS18], una meta-conjetura que cree que los límites no dependen de la distribución sino de la clase de simetría de la matriz, como simétricas, Hermitianas, simplécticas, rectangulares, etc. Algunos artículos tratan la falta de independencia en las entradas de la matriz, por ejemplo suponiendo distribuciones esféricas [WZ19] o Wishart no isotrópicas [KY17], es decir matrices empíricas de covarianza cuando la covarianza teórica no es múltiplo de la identidad. La otra categoría es la de gráficas aleatorias, cuyo principal interés es relacionar la distribución de las aristas con la del espectro, por ejemplo a través de la rareza [HLY19] o de la multiplicidad [DZ19].

En los últimos 5 o 6 años pueden encontrarse decenas de artículos titulados con el tema de leyes locales. La virtual totalidad de ellos utiliza una notación introducida en [Erd+13b], llamada *dominación estocástica* y representada con  $\preceq$ . Hay una introducción en forma de notas de curso [BK16] al tema de leyes locales y su notación, además los autores de [Erd+13b] anunciaron la próxima publicación de un libro sobre el proceso de Dyson y la ley local del Semicírculo. La notación  $\preceq$  permite presentar los resultados de una forma muy limpia, pero también requiere práctica y tiempo para entender las pruebas que la utilizan. Lo que se observó es que hacía falta una introducción que no necesitara de la notación y que fuera accesible a un público más amplio. En esta tesis se presenta una prueba elemental al teorema original que busca satisfacer esta necesidad.

En la actualidad la aplicación de los resultados es escasa. Un ejemplo pionero de extender la teoría hacia algo más palpable son las notas en [BK16] que grafican y discuten las consecuencias de la Ley Local del Semicírculo. En general, explorar las implicaciones de leyes locales, ya sea ilustrando los resultados teóricos o desarrollando aplicaciones a otras áreas, representa un área de oportunidad explorada en esta tesis.

## 1.2. DESCRIPCIÓN DEL CONTENIDO

Este texto es una introducción a la teoría de leyes locales para matrices aleatorias. En particular se investiga sobre la naturaleza de las leyes locales, se revisa la metodología de sus demostraciones y se buscan interpretaciones o representaciones gráficas. Luego se extiende a comprender ya realizaciones y simulaciones para revisar comportamientos límite y explorar aplicaciones

Estas preguntas se responden a lo largo del texto mediante ejemplos concretos. Para presentar las leyes locales se da una demostración elemental original de la Ley Local del Semicírculo, restringiendo las hipótesis a modelos gaussianos y las herramientas a desigualdades basadas en continuidad Lipschitz. Las observaciones numéricas, asintóticas y de simulación se estudian sobre la ley obtenida en [HLS+19] para matrices ralas, que es un objetivo clave del proyecto. Por último se extiende este análisis a datos reales para una generalización de [HLS+19] y [HLY19] a matrices ralas multi-bloque.

El documento se divide en los siguientes capítulos. El Capítulo 2, dedicado a los preliminares, define las herramientas y presenta los teoremas auxiliares necesarios para la teoría, como la transformada de Stieltjes, una desigualdad de concentración y el entrelazado de Cauchy. El Capítulo 3 demuestra de forma elemental la Ley Local del Semicírculo para matrices Gaussianas siguiendo el método del resolvente.

El Capítulo 4 analiza la ley para matrices ralas de [HLS+19]. Se extienden las ideas del Capítulo 3 al caso rectangular ralo con un enfoque más en las consecuencias que en la demostración. El Capítulo 4 se centra en entender la ley dada en [HLS+19] de forma asintótica y a través de simulaciones en función de la dimensión de la matriz, la cuadratura de la misma y la distribución de sus entradas. En su parte final explora aplicaciones en aprendizaje máquina a través del problema de clusterización.

---

# CAPÍTULO 2

## PRELIMINARES DE MATRICES ALEATORIAS

El nombre de *Matrices Aleatorias* sugiere un estudio de matrices cuyas entradas son variables aleatorias, quizás reales o complejas. Lo que se conoce como **teoría de matrices aleatorias** [Tao12; KY17] se enfoca en las propiedades espectrales de matrices grandes con entradas aleatorias, pensándolas como operadores lineales. Los operadores lineales pueden estudiarse de forma analítica (son espacios de Banach), algebraica (operan sobre espacios de Hilbert) ó como espacio algebraico (forman un álgebra no conmutativa). A su vez, estas líneas de estudio generan *sub-teorías* de Matrices Aleatorias llamadas analítica, algebraica o combinatoria, respectivamente (veáse pág. 143 en adelante de [Tao12]).

La literatura consultada para esta tesis [KY17] y la tesis misma usan el método del resolvente [Tao12] de la teoría analítica. Éste es un conjunto de ideas y procedimientos que usan argumentos analíticos, recursivos y de desigualdades de probabilidad. Estos incluyen, pero no se limitan, a funciones analíticas, puntos de atracción, matrices menores y colas sub-gaussianas. Estos conceptos se explican con detalle a lo largo del libro de T. Tao [Tao12] y para su uso particular en esta tesis se presenta este capítulo de preliminares.

El contenido del capítulo se divide a grandes rasgos en dos partes. La primera parte presenta el objeto de estudio definiendo lo que se llaman *la transformada de Stieltjes* (Sec. 2.1) y la distribución empírica espectral (Sec. 2.3), y cómo éstas se relacionan en la teoría de matrices aleatorias (Sec. 2.4). La segunda parte señala las dos herramientas clave que se usarán en el Capítulo 3. Una trata propiedades algebraicas recursivas en el espacio de matrices (Sec. 2.5) y la otra desigualdades de probabilidad de grandes desviaciones (Sec. 2.6).

### 2.1. LA TRANSFORMADA DE STIELTJES

Cuando  $X$  y  $Y$  son variables aleatorias reales con distribuciones conocidas, calcular la distribución de operaciones algebraicas de ellas no es tarea sencilla. Incluso la más sencilla  $X + Y$  requiere resolver una integral complicada. Una alternativa común en Probabilidad y Estadística es identificar las distribuciones con funciones analíticas, hacer operaciones y extraer límites en estas funciones y después identificar el resultado con alguna distribución de probabilidad. Las más comunes son la transformada ordinaria, la de Laplace y la de Fourier, conocidas respectivamente en Probabilidad y Estadística como funciones generadora de probabilidad, generadora de momentos y característica. Lo relevante de estas transformadas es que son biyecciones entre espacios dos espacios, el de las distribuciones y el de las funciones, que preservan los resultados límite, *i.e.* homeomorfismos.

En el tema de Leyes Locales [KY17] y en general en Matrices Aleatorias [Tao12] suele usarse la

transformada de Stieltjes, también conocida como transformada de Cauchy o de Cauchy-Stieltjes. Se mostrará a continuación cómo ésta a toda medida de probabilidad  $\mu$  asigna de forma unívoca una función analítica  $m_\mu(z)$  que satisface resultados límite. También se presenta una interpretación de la inversa funcional en términos de la distribución de probabilidad Cauchy, lo cuál permite entender el porqué de algunos teoremas y su comparación con procedimientos en estadística de variables aleatorias Reales.

**Definición 2.1.1.** Se denota al plano superior complejo por  $C^+$ . Es decir

$$C^+ = \{z \in \mathbb{C} \mid \text{Im } z > 0\}.$$

**Definición 2.1.2.** Sea  $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  una medida de probabilidad. La **transformada de Stieltjes de  $\mu$**  es la función  $m_\mu : C^+ \rightarrow C^+$  dada por

$$m_\mu(z) = \int_{\mathbb{R}} (x - z)^{-1} \mu(\mathbf{d}x).$$

Por extensión, si  $X$  es variable aleatoria real, entonces la **transformada de Stieltjes de  $X$**  está dada por

$$m_X(z) = \mathbb{E}[(X - z)^{-1}].$$

Es importante para una transformada que se pueda recuperar la distribución inicial.

**Teorema 2.1.1** (Fórmula de Inversión de Stieltjes). *Sea  $\mu$  medida de probabilidad real,  $m_\mu$  su transformada de Stieltjes. Si  $a$  y  $b$  son reales tales que  $a < b$  y  $\mu(\{a, b\}) = 0$ . Entonces*

$$\mu([a, b]) = \lim_{\eta \downarrow 0} \int_a^b \frac{1}{\pi} \text{Im } m_\mu(E + i\eta) \mathbf{d}E \quad (2.1)$$

Para la Demostración se puede ver las notas de Knowles [BK16] o el libro [Tao12]. La idea rectora es que de una transformada de Stieltjes se puede obtener una densidad de probabilidad en convolución con una aproximación Cauchy a una delta de Dirac.

**Lema 2.1.2.** *Sea  $m : C^+ \rightarrow C^+$  tal que  $\lim_{\text{Im } z \rightarrow \infty} zm(z) = -1$ . Si  $\eta = \text{Im } z$  y  $E = \text{Re } z$ , entonces existe una medida de probabilidad  $\rho$  que no depende de  $\eta$  tal que*

$$\frac{1}{\pi} \text{Im } m(E + i\eta) \mathbf{d}E = \rho(\mathbf{d}E) * \left( \frac{\eta}{\pi(E^2 + \eta^2)} \mathbf{d}E \right).$$

El lema anterior implica que para  $m = m_\mu$ , en el caso de ser  $\mu$  medida de probabilidad absolutamente continua, si  $1 \gg \eta > 0$ , entonces

$$\mathbf{d}\mu(E) \approx \frac{1}{\pi} \text{Im } m_\mu(\mathbf{d}E + i\eta).$$

Esta observación se puede extender a distribuciones que no tienen densidad, pero que convergen a distribuciones continuas. Para ello es necesario un resultado de continuidad, cuya demostración apela al mismo argumento de la delta aproximante.

**Teorema 2.1.3** (Continuidad de Stieltjes). *Sea  $(\mu_N)_{N \geq 0}$  una sucesión de medidas de probabilidad y  $(m_{\mu_N})_{N \geq 0}$  sus transformadas de Stieltjes. Las siguientes expresiones son equivalentes*

1. Existe  $\mathcal{K} \subset \mathbb{C}^+$  compacto infinito numerable tal que para todo  $z \in \mathcal{K}$

$$\lim_{N \rightarrow \infty} m_{\mu_N}(z) = m_{\mu}(z) \quad (2.2)$$

2. Para toda función  $f$  continua y acotada

$$\lim_{N \rightarrow \infty} \int f(E) \mu_N(\mathbf{d}E) = \int f(E) \mu(\mathbf{d}E) \quad (2.3)$$

Este resultado se cumple por una propiedad de las funciones analíticas. Si una sucesión de ellas coincide en un conjunto con un punto de acumulación, entonces coinciden en todo el dominio con cierta uniformidad. Parte de la intuición es que las funciones analíticas son límites de funciones continuas localmente uniformes. Si el lector quiere profundizar en la prueba de este teorema se sugiere consultar los teoremas de Vitali-Montel en análisis complejo, por ejemplo en [Con12].

## 2.2. ECUACIÓN FUNCIONAL DEL SEMICÍRCULO

El objetivo de la tesis y, en general de muchas áreas de probabilidad, es demostrar la convergencia de una familia de medidas de probabilidad  $\mu_N$  a una medida  $\mu$ . Al trabajar con la transformada de Stieltjes se opta por la comodidad de demostrar convergencia de funciones complejas en lugar de las más elaboradas para demostrar convergencia de medidas o de distribución.

Para la convergencia de funciones complejas hay un arsenal inmenso de herramientas [Con12], pero la que se usa en esta tesis es la de auto-estabilidad. En lugar convergencia puntual, se prueba que si una función *casí* cumple una ecuación, entonces la función está *muy cerca* de la solución.

**Definición 2.2.1.** Sea  $C = C^\infty[\mathbb{C}, \mathbb{C}]$  el conjunto de funciones analíticas complejas  $r : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  creciente y  $f_0$  en  $C$ . Si  $F : C \rightarrow \mathbb{C}$  es tal que  $F(f_0) = 0$  y para  $f \in C$

$$|F(f_0) - F(f)| < \varepsilon$$

implica que

$$\|f_0 - f\|_\infty = r(\varepsilon),$$

entonces se dice que  $F(f) = 0$  es una **ecuación funcional autoestable**.

De esta forma en lugar de probar que una familia  $\{f_N\}$  converge puntualmente a  $f_0$ , se demuestra que  $F(f_N) \rightarrow 0$  cuando  $N \rightarrow \infty$ . Un ejemplo de estas ecuaciones, es la siguiente

**Definición 2.2.2.** Sea  $m : \mathbb{C}^+ \rightarrow \mathbb{C}^+$ . La ecuación

$$1 + zm(z) + m_N^2(z) = 0 \quad (2.4)$$

es llamada **la ecuación funcional del semicírculo**.

Lo importante de esta ecuación es que tiene como solución la transformada de Stieltjes de una distribución importante.

**Proposición 2.2.1.** La ecuación (2.4) tiene una única solución analítica  $m_{SC} : \mathbb{C}^+ \rightarrow \mathbb{C}^+$  y es la transformada de Stieltjes de una medida de probabilidad con soporte compacto.

*Demostración.* La densidad de dicha medida es

$$\mu_{SC}(\mathbf{d}x) = \frac{\mathbb{1}_{[-2,2]}(x)}{2\pi} \sqrt{4-x^2} \mathbf{d}x.$$

Es un ejercicio directo de álgebra y cálculo que  $m_{SC}(z) = \int_{-2}^2 (x-z) \mu_{SC}(\mathbf{d}x)$  satisface la ecuación (2.4). La unicidad puede consultarse en el apéndice B de [BK16] o de la p. 143 en adelante de [Tao12]. ■

Nótese que (2.4) es un polinomio cuadrático y como tal tiene dos raíces. Ambas raíces son analíticas, pero sólo una de ellas satisface ser transformada de Stieltjes.

La siguiente proposición se desarrolla en esta tesis para demostrar que la ecuación (2.4) es en efecto auto-estable.

**Proposición 2.2.2** (Auto-estabilidad de la Ecuación del Semicírculo). *Sea  $z = E + i\eta \in \mathbb{C}^+$  con  $\eta \geq \eta_0 > 0$ . Si  $1 + zm_N(z) + m^2(z) = r$ , entonces*

$$|m_N(z) - m_{SC}(z)| \leq \frac{|r|}{\eta_0}. \quad (2.5)$$

*Demostración.* Por la ecuación funcional del Semicírculo

$$\begin{aligned} 1 + zm_N(z) + m^2(z) &= r \\ 1 + zm_{SC}(z) + m_{SC}^2(z) &= 0, \end{aligned}$$

restando la segunda de la primera expresión resulta

$$z(m_N(z) - m_{SC}(z)) + (m_N^2(z) - m_{SC}^2(z)) = r,$$

por diferencia de cuadrados

$$z(m_N(z) - m_{SC}(z)) + (m_N(z) + m_{SC}(z))(m_N(z) - m_{SC}(z)) = r,$$

agrupando con respecto a  $(m_N(z) - m_{SC}(z))$  y tomando valor absoluto

$$|z + m_N(z) + m_{SC}(z)| |m_N(z) - m_{SC}(z)| = |r|.$$

El punto clave de esta demostración está en que, como  $\text{Im } z > 0$ , ambas  $\text{Im } m_N(z), \text{Im } m_{SC}(z) > 0$  porque son transformadas de Stieltjes. Luego, para cualquier  $a \in \mathbb{C}$

$$|a| = \sqrt{\text{Re }^2 a + \text{Im }^2 a} \geq \sqrt{\text{Im }^2 a} = |\text{Im } a|.$$

Por tanto

$$|\text{Im } z + \text{Im } m_N(z) + \text{Im } m_{SC}(z)| |m_N(z) - m_{SC}(z)| \leq |r|,$$

y al ser transformadas de Stieltjes

$$|\text{Im } z + \text{Im } m_N(z) + \text{Im } m_{SC}(z)| = \text{Im } z + \text{Im } m_N(z) + \text{Im } m_{SC}(z) \geq \text{Im } z = \eta. \quad \blacksquare$$

Este resultado será clave en el Capítulo 3 para demostrar un teorema límite.

*Observación.* La definición 2.2.1 está en términos de  $\|\cdot\|_\infty$ . Sin embargo, las distribuciones de interés tienen soporte compacto con alta probabilidad y las funciones analíticas son siempre localmente acotadas.

## 2.3. DISTRIBUCIÓN EMPÍRICA ESPECTRAL

Si la sección anterior da un avance de a qué tipo de distribuciones se quiere converger, esta sección explicará cómo es la familia de la que se quiere encontrar convergencia.

Se eligen distribuciones de probabilidad relacionadas a operadores lineales. Los operadores lineales que han de usarse son matrices complejas con una propiedad de simetría muy importante sobre las cuales se puede usar el Teorema Espectral.

**Definición 2.3.1.** Se define el conjunto de matrices **Hermitianas**  $\mathbb{H}_N \subset M_{N \times N}(\mathbb{C})$  para  $N \in \mathbb{N}$  como

$$\mathbb{H}_N := \{X \in M_{N \times N}(\mathbb{C}) \mid X = X^*\}.$$

Como consecuencia del Teorema Espectral los valores propios o eigenvalores de una matriz Hermitiana son siempre reales y garantiza que existen  $N$ , considerando multiplicidad.

Para simplificar la notación respecto al Teorema Espectral la la siguiente definición resulta útil

**Definición 2.3.2.** Sean  $x = (x_1, \dots, x_N) \in \mathbb{C}^N$ . Se define para  $a_{ij} := x_i \mathbb{1}_{\{i=j\}}$  la matriz **diagonal**

$$\mathbf{diag}(x) \equiv \mathbf{diag}\{x_1, \dots, x_N\} := (a_{ij})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}.$$

Si una matriz Hermitiana tiene entradas aleatoria, los valores propios son aleatorios. Eugene Wigner [Wig55] construyó una distribución para el espectro de una matriz simétrica con entradas aleatorias a la que llamó *strength function*. Él trabajaba con problemas de física nuclear, pero en el artículo citado reconoció que encontrar esta distribución era un problema interesante por derecho propio.” Hoy esta distribución recibe el nombre de distribución empírica espectral [ESY09] por su semejanza a la distribución empírica de observaciones independientes de variables aleatorias reales que se usan en estadística [Bla+56; Tao12].

**Definición 2.3.3.** Sea  $X_N \in \mathbb{H}_N$ . Se denotan sus eigenvalores por  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$  y su diagonalización  $X_N = U_N \Lambda_N U_N^*$  donde  $\Lambda := \mathbf{diag}\{\lambda_1, \dots, \lambda_N\}$ .

Se define  $\mu_N \equiv \mu_{X_N}$ , la **distribución empírica espectral** como

$$\mu_{X_N} := \sum_{i=1}^N \delta_{\lambda_i},$$

en correspondencia se define la **función de distribución empírica espectral (FDEE)** como

$$F_{X_N}(x) := \mu_{X_N}((-\infty, x]) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{(-\infty, \lambda_i]}(x).$$

Nótese que la distribución  $\mu_{X_N}$  se definió para una matriz  $X_N$  fija, *i.e.*  $F_{X_N}(x)$  es determinista. Por esto mismo, si  $X_N$  es aleatoria, entonces  $F_{X_N}(x)$  también es aleatoria, pero toda su aleatoriedad está en función a la distribución de  $X_N$ . Esto hace deseable controlar la distribución de  $X_N$  mediante estocásticos matriciales (veáse [Tao12], p. 61). Para fines de esta tesis se da la siguiente definición.

**Definición 2.3.4.** Sea  $(\Omega, \mathcal{F}, \mathbb{P})$  un espacio de probabilidad, e  $I$  un conjunto de índices con topología. Se le llama un **ensamble** a un proceso estocástico  $\{X_i; i \in I\}$  tal que, para todo  $i$ ,

$$X_i : \Omega \rightarrow \text{Mat}_{M_i \times N_i}(\mathbb{C})$$

con  $M_i, N_i \in \mathbb{N}$  y  $X_i$  es medible.

Por ejemplo, un ensamble muy sencillo supone simetría, estandarización e independencia.

**Definición 2.3.5.** Sea  $\{X_N\}_{N \in \mathbb{N}}$  un ensamble tal que

- $X_N \in \mathbb{H}_N$
- $\mathbb{E}[X_{ij}] = 0$  y  $\sum_i \mathbf{Var}(X_{ij}) = 1$
- La familia  $(X_{ij}; i \geq j)$  es independiente.

Al proceso  $\{X_N\}$  se le llama **ensamble de Wigner** y  $X_N$  para  $N \in \mathbb{N}$  es una **matriz de Wigner**.

La observación obligatoria, ahora que se tienen las herramientas, es que al ensamble de Wigner se le puede asociar otro proceso estocástico.

**Definición 2.3.6.** Sea  $\{X_N\}$  un ensamble de Wigner y

$$\mathcal{P}_{\mathbb{R}} := \{\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1] \mid \mu \text{ es de probabilidad}\}.$$

Se define el **proceso de distribuciones empíricas espectrales** como el proceso estocástico  $\{\mu_{X_N}; N \in \mathbb{N}\}$  con  $\mathcal{P}_{\mathbb{R}}$  como espacio de estados.

Notar que  $\mathcal{P}_{\mathbb{R}}$  con la distancia en variación total forma un espacio métrico, convexo y completo, por lo que tiene sentido hablar de límites. De hecho la teoría de Matrices Aleatorias nace de querer estudiar el límite de  $\mu_{X_N}$  cuando  $N$  crece [KY17].

El Capítulo 3 reconstruye, con un ensamble específico, la demostración para  $\mu_{X_N} \xrightarrow{\mathbf{P}} \nu$  cuando  $N \rightarrow \infty$  donde  $\nu = \nu_0 \in \mathcal{P}_{\mathbb{R}}$ , **c.p.1**. Intuitivamente  $\nu$ , es el análogo al *límite ergódico* de una cadena de Markov, pero en este caso el proceso converge a una v.a. degenerada. Nótese que  $\nu$  es degenerada en  $\mathcal{P}_{\mathbb{R}}$ , pero la prueba mostrará que  $\nu_0$  tiene densidad en  $\mathbb{R}$  respecto a Lebesgue.

## 2.4. RESOLVENTE DE UNA MATRIZ ALEATORIA

En las secciones anteriores se presentaron la transformada de Stieltjes y la distribución empírica espectral. Ambos conceptos coinciden de forma natural en el concepto de resolvente de una matriz. En Leyes Locales este se utiliza a través de la función de Green y la transformada de la matriz.

**Definición 2.4.1.** Sea  $X_N \in \mathbb{H}_N$ . Se define  $G_{X_N} : \mathbb{C}^+ \rightarrow \mathbb{H}_N$ , la **función de Green** o **resolvente** de  $X_N$  como

$$G(z) \equiv (G_{ij})_{i,j} \equiv G_{X_N}(z) := (X_N - zI)^{-1}. \quad (2.6)$$

Su traza normalizada  $m_{X_N} : \mathbb{C}^+ \rightarrow \mathbb{C}^+$  está dada por

$$m_{X_N}(z) := \frac{1}{N} \mathbf{Tr}(G_{X_N}(z)). \quad (2.7)$$

Se suele decir [HLS+19] que  $m_{X_N}$  es la transformada de  $X_N$ . Esto es parcialmente cierto, ya que sí es la transformada de Stieltjes de la distribución empírica espectral.

**Lema 2.4.1.** *Sea  $F_{X_N}$  la FDEE de  $X_N$ . Luego*

$$m_{X_N}(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z}. \quad (2.8)$$

*En consecuencia*

$$m_{X_N}(z) = \int \frac{\mu_{X_N}(\mathbf{d}x)}{x - z} = \int_{\mathbb{R}} \frac{\mathbf{d}F_{X_N}(x)}{x - z}. \quad (2.9)$$

*Demostración.* Usando la diagonalización del teorema espectral

$$\begin{aligned} m_{X_N}(z) &= \frac{1}{N} \mathbf{Tr}((X_N - zI)^{-1}) \\ &= \frac{1}{N} \mathbf{Tr}((U_N \Lambda_N U_N^* - zI)^{-1}) \\ &= \frac{1}{N} \mathbf{Tr}(U_N (\Lambda_N - zI)^{-1} U_N^*) \\ &= \frac{1}{N} \mathbf{Tr}((\Lambda_N - zI)^{-1}) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z}. \end{aligned}$$

Usando notación para integrales de Lebesgue-Stieltjes

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} = \int \frac{\mu_{X_N}(\mathbf{d}x)}{x - z}.$$

El resto se concluye usando que, por construcción  $\mu_{X_N}(\mathbf{d}x) = \mathbf{d}F_{X_N}(x)$ . ■

Algunas propiedades que satisface el resolventes serán de utilidad a la hora de calcular continuidad Lipschitz. Por simplicidad supóngase que  $m_{X_N} : \mathbb{C} \setminus \mathbb{R} \rightarrow \mathbb{C}$ , usando la misma regla de la definición 2.4.1.

**Proposición 2.4.2.** *Sea  $X_N \in \mathbb{H}_N$  y  $m_{X_N}(z)$  definida como antes. Si  $z, w \in \mathbb{C}^+$ , entonces*

1.  $\overline{m_{X_N}(z)} = m_{X_N}(\bar{z})$
2.  $|m_{X_N}(z) - m_{X_N}(w)| \leq (\operatorname{Im}(z)\operatorname{Im}(w))^{-1}|z - w|$
3.  $|m_{X_N}(z)| \leq \frac{1}{\operatorname{Im} z}$

*Demostración.* Estas propiedades se basan en que cada eigenvalor  $\lambda_i$  es real. Por elección de  $z$  esto implica que  $|\lambda_i - z| \geq \operatorname{Im}(z) = \eta$ .

---

1. Por el lema 2.4.1 y sesquilinealidad de la conjugación

$$\begin{aligned}
 m_{X_N}(z) &= \overline{\int_{\mathbb{R}} (x - z)^{-1} \mathbf{d}F_{X_N}(x)} \\
 &= \int_{\mathbb{R}} \overline{(x - z)^{-1}} \mathbf{d}F_{X_N}(x) \\
 &= \int_{\mathbb{R}} (\overline{x - z})^{-1} \mathbf{d}F_{X_N}(x) \\
 &= \int_{\mathbb{R}} (x - \bar{z})^{-1} \mathbf{d}F_{X_N}(x).
 \end{aligned}$$

El resultado se concluye usando el lema 2.4.1 otra vez.

2. Por el lema 2.4.1

$$\begin{aligned}
 |m_{X_N}(z) - m_{X_N}(w)| &= \frac{1}{N} \left| \sum_{i=1}^N \left( \frac{1}{\lambda_i - z} - \frac{1}{\lambda_i - w} \right) \right| \\
 &= \frac{1}{N} \left| \sum_{i=1}^N \left( \frac{-w + z}{(\lambda_i - z)(\lambda_i - w)} \right) \right| \\
 &\leq \frac{1}{N} \sum_{i=1}^N \frac{|z - w|}{|\lambda_i - z| |\lambda_i - w|}.
 \end{aligned}$$

Como  $z \in \mathbb{C}^+$ , entonces usando que  $|\operatorname{Im} \alpha| \leq |\alpha|$  se tiene que

$$|\lambda_i - z| \geq |\operatorname{Im}(\lambda_i - z)| = |\operatorname{Im} z| = \operatorname{Im} z > 0.$$

Esto también se cumple para  $w$ . Luego,

$$\begin{aligned}
 \frac{1}{N} \sum_{i=1}^N \frac{|z - w|}{|\lambda_i - z| |\lambda_i - w|} &\leq \frac{1}{N} \sum_{i=1}^N \frac{|z - w|}{\operatorname{Im}(z) \operatorname{Im}(w)} \\
 &= \frac{|z - w|}{\operatorname{Im}(z) \operatorname{Im}(w)}.
 \end{aligned}$$

3. Por el lema 2.4.1

$$\begin{aligned}
 |m_{X_N}(z)| &= \left| \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} \right| \\
 &\leq \frac{1}{N} \sum_{i=1}^N \frac{1}{|\lambda_i - z|}
 \end{aligned}$$

Usando que  $|\operatorname{Im} \alpha| \leq |\alpha|$ , se tiene que  $|\lambda_i - z| \geq \operatorname{Im} z$ . Luego

$$|m_{X_N}(z)| \leq \frac{1}{N} \sum_{i=1}^N \frac{1}{\operatorname{Im} z}.$$

■

Las ideas de estas demostraciones, la diagonalización y las propiedades de traza, se repetirán en múltiples pruebas durante el texto.

## 2.5. MINORES MATRICIALES

Hasta ahora se ha mencionado en varias ocasiones que el método del resolvente tiene ideas recursivas. La principal de ellas viene dada por el Entrelazado de Cauchy, un resultado que relaciona el espectro de una matriz con el de una submatriz.

La siguiente notación es estándar en Leyes Locales [BK16] y sirve para definir los menores y controlar sus espectros

**Definición 2.5.1.** Sea  $X_N \in \mathbb{H}_N$  cualquier matriz Hermitiana.

Para  $1 \leq i \leq N$  se denota por  $X_N^{(i)}$  a la matriz Hermitiana  $(N-1) \times (N-1)$  que resulta de remover el  $i$ -ésimo renglón y la  $i$ -ésima columna. Se dice que  $X_N^{(i)}$  es una **minor principal** de  $X_N$ .

Por simplicidad en las pruebas, los índices  $X_N^{(i)} = (x_{jk})$  se preservan cuando  $j, k \neq i$ .

También se denota por  $(X_N)_{(i)}$  a la  $i$ -ésima columna de  $X_N^{(i)}$ , que es un vector en  $\mathbb{C}^{N-1}$ .

El teorema recursivo que suele utilizarse es el siguiente

**Teorema 2.5.1.** (*Entrelazado de Cauchy*) Sea  $F_{X_N}$  la FDEE de  $X_N \in \mathbb{H}_N$ . Para cualquier minor principal  $X_N^{(i)}$ , su FDEE  $F_{X_N^{(i)}}$  satisface

$$\left| NF_{X_N}(x) - (N-1)F_{X_N^{(i)}}(x) \right| \leq 1, \quad \forall x \in \mathbb{R}. \quad (2.10)$$

La prueba de este teorema se puede consultar en [Tao12].

## 2.6. FUNCIONES LIPSCHITZ SOBRE VECTORES GAUSSIANOS

La herramienta principal en el Capítulo 3 será un resultado de desigualdades de concentración para vectores Gaussianos. Ésta es atribuida a Cirelson y Sudakov en los años 1970s, aunque se sabe que Borel llegó a algo parecido [Mas00]. No se mostrará la prueba, sino las ideas generales para llegar a ella.

Si  $X \sim \mathcal{N}_{\mathbb{R}^N}(0, \sigma^2 I_N)$  y  $F(x) = Ax + b$ , entonces  $F(X) \sim \mathcal{N}_{\mathbb{R}^N}(b, \sigma^2 A^T A)$  [Mas00]. Es decir, aplicar funciones afines o lineales a un vector Gaussiano resulta en vectores Gaussianos. Como toda función diferenciable es localmente lineal, es natural preguntarse cómo afecta una función suave a un vector Gaussiano. Lo que se esperaría es que fuera *casi Gaussiano*, en algún sentido. En este proyecto se usó el sentido de la *sub-Gaussianidad* [Ban15].

**Definición 2.6.1.** Un vector aleatorio  $X$  se dice **sub-Gaussiano** si  $\mathbb{E}[|X|^2] < \infty$  y existen constantes  $C, c > 0$  tales que

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq x] \leq C \exp\{-cx^2\}.$$

La sub-Gaussianidad es equivalente a la existencia de la función generadora de momentos en una vecindad de cero [Mas00; Ban15]. Esta relación también se aprecia en la siguiente desigualdad de grandes desviaciones que a su vez inspira la herramienta que usa el Capítulo 3.

**Lema 2.6.1** (Desigualdad de Chernov). Sea  $X \sim \mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ . Para todo  $\varepsilon > 0$  se cumple que

$$\mathbb{P}[|X - \mu| \geq \varepsilon] \leq 2 \exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\}. \quad (2.11)$$

*Demostración.* La prueba usa la simetría de  $X - \mu$ , la desigualdad de Markov y la función generadora de momentos de la distribución normal. Para cualquier  $\lambda > 0$ ,

$$\begin{aligned} \mathbb{P}\left[\frac{X - \mu}{\sigma} \geq \varepsilon\right] &= \mathbb{P}\left[\exp\left\{\lambda \frac{X - \mu}{\sigma}\right\} \geq \exp\{\lambda\varepsilon\}\right] \\ &\leq \mathbb{E}\left[\exp\left\{\lambda \frac{X - \mu}{\sigma}\right\}\right] \exp\{-\lambda\varepsilon\} \\ &= \exp\left\{\frac{\lambda^2}{2}\right\} \cdot \exp\{-\lambda\varepsilon\} \\ &= \exp\left\{\frac{\lambda^2 - 2\lambda\varepsilon}{2}\right\}. \end{aligned}$$

La expresión al anterior de la exponencial se minimiza cuando  $\lambda = \varepsilon$ , en este caso se tienen las siguientes expresiones usando la simetría de la distribución normal.

$$\begin{aligned} \mathbb{P}\left[\frac{X - \mu}{\sigma} \geq \varepsilon\right] &\leq \exp\left\{-\frac{\varepsilon^2}{2}\right\}, \\ \mathbb{P}\left[\left|\frac{X - \mu}{\sigma}\right| \geq \varepsilon\right] &\leq 2 \exp\left\{-\frac{\varepsilon^2}{2}\right\}, \\ \mathbb{P}[|X - \mu| \geq \varepsilon\sigma] &\leq 2 \exp\left\{-\frac{\varepsilon^2}{2}\right\}. \end{aligned}$$

■

La construcción fue hecha resaltando que  $X = \sigma Z + \mu$ , para  $Z \sim \mathcal{N}_{\mathbb{R}}(0, 1)$  y la propiedad fuerte utilizada es que la función generadora de momentos de la normal estándar es conocida. Esto se puede entender como aplicar una función afín o lineal a una v.a. normal estándar. Esta es la intuición detrás del resultado más general que aplica para una clase amplia de funciones suaves, incluso más amplia que diferenciables.

**Definición 2.6.2.** Sean  $(S_1, d_1)$  y  $(S_2, d_2)$  espacios métricos. Se dice que una función continua  $f : S_1 \rightarrow S_2$  es *L-Lipschitz* con  $L \in \mathbb{R}^+$ , si para todos  $x, y \in S_1$  se cumple

$$d_2(f(x), f(y)) \leq L \cdot d_1(x, y).$$

Las funciones Lipschitz suelen usarse en desigualdades de grandes desviaciones [Tao12], en estos casos las variables aleatorias están valuadas en espacios normados como  $\mathbb{C}^N$ , las matrices con la norma operador o matrices con norma de Frobenius. La herramienta que se usará en el Capítulo 3 usa esta noción de suavidad. Se sugiere al lector consultar la prueba en la Sección 2.1.3 de [Tao12] o la Sección 4.2 de [Ban15].

**Teorema 2.6.2** (Concentración Gaussiana). *Sea  $X \sim \mathcal{N}_{\mathbb{R}^N}(0, I_N)$  para  $N \in \mathbb{N}$ . Si  $F : \mathbb{R}^N \rightarrow \mathbb{R}$  es una función L-Lipschitz, entonces para cualquier  $\varepsilon > 0$*

$$\mathbb{P}[|F(X) - \mathbb{E}[F(X)]| \geq \varepsilon] \leq 2 \exp\left\{-\frac{\varepsilon^2}{2L^2}\right\}. \quad (2.12)$$

Una observación importante es que la desigualdad *no* depende de  $N$ . El Capítulo 3 prueba que  $f_N(Y_N) \xrightarrow{P} 0$  cuando  $N \rightarrow \infty$  para algún vector Gaussiano  $Y_N$ . A grandes rasgos  $f_N$  sería  $L_N$ -Lipschitz por consecuencias de la proposición 2.4.2 y el objetivo será probar que  $L_N \rightarrow_N 0$ .

Otras desigualdades de concentración de medida, comunes en matrices aleatorias, son las de Talagrand, McDiarmid y Sotirov que pueden consultarse en [Tao12]. El método de Stein con expansión en cumulantes [LP+09] también puede emplearse para obtener una desigualdad de concentración [LS18]. Como una introducción más amplia y con referencias históricas al tema se sugiere el artículo [Mas00].

## 2.7. MÁS SOBRE FUNCIONES LIPSCHITZ

Las funciones Lipschitz son funciones continuas, pero no necesariamente son derivables [Con12]. Intuitivamente se puede pensar como una generalización de diferenciabilidad. Esta sección da algunas herramientas extra sobre las consecuencias algebraicas de la definición 2.6.2. Una es referente a la composición de funciones y la otra a funciones sobre el espacio de matrices.

Similar a la regla de la cadena para funciones diferenciables, las funciones Lipschitz también tienen una regla para la composición.

**Proposición 2.7.1.** *Sea  $f : S_1 \rightarrow S_2$  función  $L_f$ -Lipschitz y  $g : S_2 \rightarrow S_3$  función  $L_g$ -Lipschitz. La composición de funciones  $g \circ f : S_1 \rightarrow S_3$  es  $(L_g L_f)$ -Lipschitz.*

*Demostración.* Sean  $x, y \in S_1$  con  $a = f(x)$ ,  $b = f(y) \in S_2$ , por la definición 2.6.2 se tiene que

$$\begin{aligned} d_3(g(a), g(b)) &\leq L_g d_2(a, b) \\ &= L_g(d_2(f(x), f(y))) \\ &\leq L_g(L_f d_1(x, y)). \end{aligned}$$

■

La otra propiedad relaciona la continuidad en matrices con funciones sobre su espectro. Esta propiedad está fundamentada en el siguiente teorema cuya demostración puede encontrarse en la sección 2.1.5 de [AGZ10].

**Teorema** (Hoffman-Wielandt). *Sean  $A, B \in \mathbb{H}_N$  con espectros  $\Lambda(A)$  y  $\Lambda(B) \in \mathbb{R}^N$ , respectivamente. Luego*

$$\|\Lambda(A) - \Lambda(B)\|_2^2 \leq \mathbf{Tr}((A - B)^2).$$

Una consecuencia de este teorema requiere la siguiente proposición y se usará en el Capítulo 3. Este lema se puede consultar en la sección 2.3.1 de [AGZ10].

**Lema 2.7.2.** *Sea  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  función  $L_f$ -Lipschitz y  $\Lambda : \mathbb{H}_N \rightarrow \mathbb{R}^N$  la función que asigna a cada matriz Hermitiana su espectro ordenado. Si  $\|X\|_F^2 = \sum_{ij} X_{ij}^2$  define la distancia en  $\mathbb{H}_N$ , entonces la función  $f \circ \Lambda : \mathbb{H}_N \rightarrow \mathbb{R}$  es  $\sqrt{2}L_f$ -Lipschitz.*



# CAPÍTULO 3

## LEY LOCAL DEL SEMICÍRCULO

La Ley Local del Semicírculo sienta las bases de la teoría de leyes locales. En este capítulo se da una prueba original de este teorema en términos elementales. Concretamente se centra en el caso Gaussiano para matrices unitarias. La prueba intenta en todo momento dar el resultado de forma explícita, llevando registro de cada constante, contrario a lo usual en la literatura (veáse por ejemplo [BK16]).

Esta prueba fue construida a partir de ideas provenientes de tres fuentes con notación, hipótesis y procedimientos diferentes. Una de ellas es el primer artículo de leyes locales [ESY09], el cuál da la idea general de la demostración pero usa hipótesis muy fuertes y herramientas poco accesibles. Otro es el libro de texto de Tao [Tao12], que no demuestra la última parte de la prueba, pero da muchas herramientas y para lo que se le llama *el método del resolvente*. La tercera referencia son las notas de Knowles [BK16] que usan una notación específica del área y por sí misma intenta ser una fuente introductoria. Sin embargo, complican la demostración al intentar probar dos resultados fuertes al mismo tiempo y oscureciendo las propias ideas detrás de esta notación más sencilla.

Gracias a estas fuentes y comparándolas con el procedimiento en otras fuentes, principalmente [WZ19; KY17; Erd+13b; LS18] se observa que la prueba consiste en tres partes.

La primera es la ecuación funcional. Mediante manipulaciones algebraicas, recursiones de funcionales matriciales y tomando esperanzas se plantea una ecuación sobre la transformada de Stieltjes. Esta ecuación suele tener un punto fijo atractor.

La segunda parte es la ley global. Ésta consiste en probar que hay convergencia en probabilidad al punto fijo atractor. Para ello lo que se prueba es que el residuo de la ecuación funcional converge a cero y por ser un punto fijo atractor se tiene convergencia de transformadas.

La tercera parte es un argumento de continuidad. Mientras que la ley global se prueba lejos de la recta real, la ley local necesita resultados cerca de la recta real, para ello realiza un acercamiento que recuerda a la inducción en la que una cantidad finita de pasos optimizan la aproximación a la vez que pierde probabilidad. Esto se debe hacer en un equilibrio adecuado que por lo regular implica un polinomio en el tamaño de la matriz.

El capítulo está estructurado para abordar cada una de estas partes en una sola sección. Se añade al final una sección dedicada a comentar las consecuencias de este teorema y comparar las expresiones encontradas con aquellas en la literatura.

### 3.1. ENSAMBLE GAUSSIANO UNITARIO

El ensamble Gaussiano unitario, típicamente conocido como GUE es un modelo común en la teoría de matrices aleatorias. Consiste en matrices complejas con propiedades de simetría muy fuertes y con una distribución sencilla con entradas independientes.

**Definición 3.1.1** (GUE). Sea  $\{H_N\}_{N \in \mathbb{N}}$  una sucesión de matrices con entradas aleatorias tales que, para cada  $N \in \mathbb{N}$  se cumple lo siguiente

1.  $H_N \in \mathbb{H}_N$ , es decir  $H_N = H_N^*$
2.  $(H_N)_{i,i} \sim \mathcal{N}_{\mathbb{R}}(0, N^{-1})$
3.  $(H_N)_{i,j} \sim \mathcal{N}_{\mathbb{C}}(0, N^{-1})$  si  $i \neq j$
4. La colección  $\{(H_N)_{i,j} : 1 \leq i \leq j \leq N\}$  es una familia independiente.

*Observación.* La distribución de cada entrada de la matriz depende directamente de  $N$ .

Por simplicidad se escribirá  $H = (H_{i,j})$  en lugar de  $H_N$  cuando  $N$  se suponga fija y no genere confusión.

### 3.2. LA LEY DEL SEMICÍRCULO

El objetivo principal del capítulo es probar que el proceso espectral empírico  $\{\mu_N\}$  de  $\{H_N\}$  cumple un teorema límite (veáse sección 2.3). En la sección 2.2 se presentó la siguiente distribución, que será el límite de este proceso.

$$\mu_{SC}(\mathbf{d}x) = \frac{\mathbb{1}_{[-2,2]}(x)}{2\pi} \sqrt{4 - x^2} \mathbf{d}x.$$

A grandes rasgos lo que se quiere demostrar es que

$$\mu_N \xrightarrow[N]{\mathbb{P}} \mu_{SC}.$$

Para ello, en lugar de demostrar convergencia en distribución se probará convergencia de las transformadas de Stieltjes. Como se busca relacionar  $m_N$  con  $m_{SC}$ , se empleará el método del resolvente [Tao12; ESY09], cuyas primeras ideas comienzan a aparecer con Alice Guionnet [GZ+00].

Para la demostración se distinguen, tanto en la tesis como en la literatura [BK16; KY17; Tao12], tres partes esenciales. La primera es la ecuación funcional, que sirve para determinar el límite de convergencia. La segunda es la ley global, que es probar la convergencia usando desigualdades de concentración. La tercera es el argumento de continuidad bootstrap, que refina el resultado hasta el alcance deseado.

### 3.3. ECUACIÓN FUNCIONAL

El método se centra, como se habló en la sección 2.2, en que  $m_{SC}$  satisface una ecuación auto estable y por lo tanto basta demostrar que  $\{m_N\}$  está en el dominio de atracción de dicha ecuación. El primer paso es reescribir  $m_N$  en un formato similar a la ecuación (2.4).

Se tiene que  $m_N(z) = \frac{1}{N} \sum_{i=1}^N G_{ii}(z)$ . Conviene entonces saber qué estructura tiene  $G_{ii}(z)$ .

**Lema 3.3.1** (Complemento de Schur). *Para  $1 \leq i \leq N$  se tiene que*

$$G_{ii}(z) = \frac{1}{H_{ii} - z - (H_{\cdot,i}^{(i)})^*(H^{(i)} - zI)^{-1}(H_{\cdot,i}^{(i)})} \quad (3.1)$$

*Demostración.* Proceso de eliminación Gaussiana usando notación de matrices a bloques. ■

El término en el denominador de (3.1) es tan importante para la demostración que se le dará notación propia.

**Definición 3.3.1.** Para  $1 \leq i \leq N$  se escribe

$$X_i = X_i(z) := (H_{\cdot,i}^{(i)})^*(H^{(i)} - zI)^{-1}(H_{\cdot,i}^{(i)}).$$

La estructura de  $X_i$  es la de una forma cuadrática  $v^*Av$ . Aquí  $A$  es la matriz menor que resulta de remover el renglón y la columna  $i$ . Por su parte  $v$  es exactamente la columna  $i$ -ésima. A continuación se probará un lema que permite intuir  $X_i \approx m_N(z)$ . La idea principal para encontrar la ecuación funcional es sustituir  $X_i$  por  $m_N(z)$  en (3.1).

**Lema 3.3.2.** *Para  $1 \leq i \leq N$  se tiene que*

$$\mathbb{E}[X_i] = \sqrt{\frac{N-1}{N}} \mathbb{E} \left[ m_{N-1} \left( \sqrt{\frac{N}{N-1}} z \right) \right]. \quad (3.2)$$

*Demostración.* Se usa la distribución de las entradas de  $H$  condicionando a  $H_{\cdot,i}$

$$\begin{aligned} \mathbb{E} \left[ (H_{\cdot,i}^{(i)})^*(H^{(i)} - zI)^{-1}(H_{\cdot,i}^{(i)}) \right] &= \sum_{k,l \neq i} \mathbb{E} \left[ \overline{H_{ki}} (H^{(i)} - zI)_{kl}^{-1} H_{li} \right] \\ &= \sum_{k,l \neq i} \mathbb{E} \left[ \mathbb{E} \left[ \overline{H_{ki}} (H^{(i)} - zI)_{kl}^{-1} H_{li} \mid H_{\cdot,i} \right] \right]. \end{aligned}$$

Se usa ahora que las columnas de  $H$  son independientes.

$$\begin{aligned} \sum_{k,l \neq i} \mathbb{E} \left[ \overline{H_{ki}} H_{li} \right] \mathbb{E} \left[ (H^{(i)} - zI)_{kl}^{-1} \right] &= \sum_{k,l \neq i} \frac{\delta_{kl}}{N} \mathbb{E} \left[ (H^{(i)} - zI)_{kl}^{-1} \right] \\ &= \sum_{\substack{1 \leq k \leq N \\ k \neq i}} \frac{1}{N} \mathbb{E} \left[ (H^{(i)} - zI)_{kk}^{-1} \right] \\ &= \frac{N-1}{N} \mathbb{E} \left[ \frac{1}{N-1} \mathbf{Tr} \left( (H^{(i)} - zI)^{-1} \right) \right] \end{aligned}$$

Por construcción  $H_N^{(i)} \stackrel{d}{=} \sqrt{\frac{N-1}{N}} H_{N-1}$ . Esto implica que

$$\begin{aligned} \mathbb{E} \left[ \mathbf{Tr} \left( (H^{(i)} - zI)^{-1} \right) \right] &= \mathbb{E} \left[ \mathbf{Tr} \left( \left( \sqrt{\frac{N-1}{N}} H_{N-1} - zI \right)^{-1} \right) \right] \\ &= \mathbb{E} \left[ \mathbf{Tr} \left( \sqrt{\frac{N}{N-1}} \left( H_{N-1} - \left( \sqrt{\frac{N}{N-1}} zI \right) \right)^{-1} \right) \right]. \end{aligned}$$

Si  $\gamma \geq 1$  entonces  $\gamma\eta \geq \eta \geq \eta_0$  y por lo tanto  $\mathbb{E}[m_{N-1}(\gamma z)] = \mathbb{E} \left[ \frac{1}{N-1} \mathbf{Tr} \left( (H_{N-1} - \gamma zI)^{-1} \right) \right]$ . ■

El siguiente paso consiste en reemplazar  $X_i$  por  $m_N$  en (3.1). Para ello se define  $\dot{X}_i := X_i - \mathbb{E}[X_i]$ . Bajo esta notación se interpreta (3.1) con la idea de que  $\dot{X}_i, H_{ii}$  y  $(\mathbb{E}[X_i] - m_N(z))$  son pequeños. Es decir

$$G_{ii}(z) = \frac{1}{H_{ii} - z + \mathbb{E}[X_i] - \dot{X}_i} = \frac{1}{H_{ii} - z - m_N(z) + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i} \approx -\frac{1}{z + m_N(z)}.$$

Demostrar que son pequeños es el objetivo del teorema. Estas ideas se formalizan con un lema más que resume la sección.

**Lema 3.3.3.** *Sea  $z \in \mathbb{C}^+$ . Entonces*

$$1 + zm_N(z) + m_N^2(z) = \frac{1}{N} \sum_{i=1}^N \left( H_{ii} + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i \right) G_{ii}. \quad (3.3)$$

*Demostración.*

$$\begin{aligned} m_N(z) &= \frac{1}{N} \sum_{i=1}^N G_{ii}(z) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{-z - m_N(z) + H_{ii} + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i}. \end{aligned}$$

Por suma de fracciones, para  $a_i(a_i + b_i)b_i \neq 0$  se cumple que

$$\frac{1}{a_i + b_i} = \frac{1}{a_i} - \frac{b_i}{a_i(a_i + b_i)} = \frac{1}{a_i} - \frac{b_i}{a_i} \cdot \frac{1}{(a_i + b_i)}.$$

Para  $a_i = -(z + m_N(z))$  y  $b_i = H_{ii} + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i$ , se tiene que  $\frac{1}{a_i + b_i} = G_{ii}$  y que  $a_i$  no depende de  $N$ . Luego

$$\begin{aligned} m_N(z) &= \frac{1}{N} \sum_{i=1}^N G_{ii} \\ &= \frac{1}{N} \sum_{i=1}^N \left( -\frac{1}{z + m_N(z)} - \left( \frac{H_{ii} + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i}{-(z + m_N(z))} \right) G_{ii} \right) \\ &= -\frac{1}{z + m_N(z)} + \frac{1}{N} \sum_{i=1}^N \left( \frac{H_{ii} + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i}{z + m_N(z)} \right) G_{ii}. \end{aligned}$$

La ecuación anterior se puede multiplicar por  $z + m_N(z)$  y se tiene que

$$zm_N(z) + m_N^2(z) = -1 + \frac{1}{N} \sum_{i=1}^N \left( H_{ii} + (\mathbb{E}[X_i] - m_N(z)) - \dot{X}_i \right) G_{ii}.$$

■

Este lema será útil en el sentido que plantea la sección 2.2 para demostrar teoremas límite.

### 3.4. LEY GLOBAL DEL SEMICÍRCULO

La Ley Global del Semicírculo es el teorema demostrado por Wigner en 1955 y aquí se demuestra mediante desigualdades de concentración.

La sección anterior demostró la relación

$$1 + zm_N(z) + m_N^2(z) = r_N(z).$$

El residuo  $r_N(z)$  es aleatorio y depende de  $H_N$  así como de  $z$ , en esta sección se demostrará que

$$|r_N(z)| \xrightarrow[N \rightarrow \infty]{\mathbb{P}} 0.$$

En particular, se prueba que los términos

$$\frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii}, \quad \frac{1}{N} \sum_{i=1}^N (\mathbb{E}[X_i] - m_N(z)) G_{ii} \quad \text{y} \quad \frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii}$$

son pequeños con alta probabilidad y por lo tanto  $1 + zm_N(z) + m_N^2(z) \rightarrow^P 0$  cuando  $N \rightarrow \infty$ . La herramienta principal será la desigualdad de concentración Gaussiana, presentada en el Capítulo 2.

En pocas palabras, probar la ley de Wigner para el caso GUE se reduce a probar continuidad Lipschitz llevando registro de cada variable importante.

A lo largo de esta sección se supondrá  $z = E + \mathbf{i}\eta$  con  $\eta \geq \eta_0 > 0$ , donde  $\eta_0$  puede o no depender de  $N$ . También se usarán constantes  $0 < c < 1 < C < \infty$  que no dependan de  $N$  y que pueden cambiar de línea a línea, lo cual es notación estándar en Leyes Locales [HLS+19; BK16; ESY09]. En ocasiones se usarán sub-índices  $C_j, c_j$  para resaltar que provienen de diferentes términos. A diferencia de la literatura usual, los resultados se obtendrán manteniendo las constantes  $C$  y  $c$  independientes de  $\eta_0$ , lo que permitirá encontrar un resultado más fuerte que la ley global usando la prueba de ésta.

#### 3.4.1. DECAIMIENTO DEL PRIMER TÉRMINO

El primer término es también el más sencillo de acotar. La suma  $\frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii}$  tiene un símil con la traza, pero no es una traza. Para poder acotarla se usará la desigualdad de Hölder y que la norma  $\|\cdot\|_\infty$  es 1-Lipschitz.

**Lema 3.4.1.** *Sea  $\eta_0 > 0$  y  $G_{ii}$  definido como antes para  $1 \leq i \leq N$ . La traza está acotada de la siguiente forma.*

$$\frac{1}{N} \sum_{i=1}^N |G_{ii}| \leq \frac{1}{\eta_0}. \quad (3.4)$$

*Demostración.* Como  $H_N \in \mathbb{H}_N$  existe  $U$  matriz unitaria tal que  $H_N = U\Lambda U^*$ . Por lo tanto  $G = U^*(\Lambda - zI)^{-1}U$ . En particular  $G_{ii} = (U_{\cdot,i})^*(\Lambda - zI)^{-1}U_{\cdot,i}$ . Luego

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N |G_{ii}| &= \frac{1}{N} \sum_{i=1}^N |(U_{\cdot,i})^*(\Lambda - zI)^{-1}U_{\cdot,i}| \\ &= \frac{1}{N} \sum_{i=1}^N \left| \sum_{k=1}^N \overline{U_{k,i}} \frac{1}{\lambda_k - z} U_{k,i} \right| \\ &\leq \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N \frac{1}{|\lambda_k - z|} |\overline{U_{k,i}} U_{k,i}| \end{aligned}$$

Se continua usando que la matriz  $U$  es unitaria.

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N \frac{1}{|\lambda_k - z|} |\overline{U_{k,i}} U_{k,i}| &= \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N \frac{1}{|\lambda_k - z|} |U_{k,i}|^2 \\ &\leq \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N \frac{1}{\eta_0} |U_{k,i}|^2 \\ &= \frac{1}{N\eta_0} \sum_{i=1}^N \|U_{\cdot,i}\|^2 = \frac{N}{N\eta_0}. \end{aligned}$$

■

Esta cota se combinará con la desigualdad de Hölder y la equivalencia de normas de  $\mathbb{C}^N$ .

**Lema 3.4.2.** Sea  $h = (H_{11}, H_{22}, \dots, H_{NN})^T$ . Si  $\eta_0 > 0$ , entonces

$$\left| \frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii} \right| \leq \frac{\|h\|_\infty}{\eta_0}.$$

*Demostración.* Por la desigualdad de Hölder,

$$\begin{aligned} \left| \frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii} \right| &\leq \left( \max_i \{|H_{ii}|\} \right) \left( \frac{1}{N} \sum_{i=1}^N |G_{ii}| \right) \\ &\leq \frac{1}{\eta_0} \left( \max_i \{|H_{ii}|\} \right) \\ &= \frac{\|h\|_\infty}{\eta_0}. \end{aligned}$$

■

Estos lemas permiten acotar al tercer término en función mediante concentración Gaussiana.

**Lema 3.4.3.** Sean  $\varepsilon > 0$  y  $\eta_0 > 0$ . Para  $G$  y  $H$  definidos como antes,

$$\mathbb{P} \left[ \left| \frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii} \right| \geq \varepsilon \right] \leq 2N \exp \left\{ -\frac{\varepsilon^2}{2} N \eta_0^2 \right\}. \quad (3.5)$$

*Demostración.* Por monotonía de la probabilidad basta probar que concentración se cumple para  $\frac{1}{\eta_0} (\max_i \{|H_{ii}|\})$ . Sea  $h = (H_{11}, H_{22}, \dots, H_{NN})^T$ . Entonces  $N^{1/2}h$  es un vector con entradas normales estándar complejas y la norma uniforme es una función 1-Lipschitz. Luego, por concentración Gaussiana

$$\begin{aligned} \mathbb{P} \left[ \frac{\|h\|_\infty}{\eta_0} \geq \varepsilon \right] &= \mathbb{P} \left[ \|N^{1/2}h\|_\infty \geq N^{1/2}\eta_0\varepsilon \right] \\ &= \mathbb{P} \left[ \bigcup_i \{|N^{1/2}H_{ii}| \geq N^{1/2}\eta_0\varepsilon\} \right] \\ &\leq N \mathbb{P} \left[ |N^{1/2}H_{11}| \geq N^{1/2}\eta_0\varepsilon \right] \\ &\leq 2N \exp \left\{ -\frac{\varepsilon^2}{2} (N^{1/2}\eta_0)^2 \right\}. \end{aligned}$$

■

Esta desigualdad implica que el tercer término es pequeño con alta probabilidad cuando  $N\eta_0^2$  es grande. El resto de la sección continua con ideas similares, pero con términos más complejos.

### 3.4.2. DECAIMIENTO DEL SEGUNDO TÉRMINO

Se busca acotar con alta probabilidad a  $\frac{1}{N} \sum_{i=1}^N (\mathbb{E}[X_i] - m_N(z)) G_{ii}$ . Esto se hará en tres partes. Por una parte notar que  $\mathbb{E}[X_i]$  no depende de  $i$  ya que las entradas de  $H_N$  son i.i.d. Debido a esto el problema se reduce a acotar  $|\mathbb{E}[X_i] - m_N(z)|$ , que se hará en dos partes usando desigualdad del triángulo y  $N^{-1} \sum G_{ii}$ , consecuencia de la proposición 2.4.2.

*Observación.* Por la desigualdad del triángulo

$$|\mathbb{E}[X_i] - m_N(z)| \leq \left| \mathbb{E}[X_i] - \left( \frac{N-1}{N} \right) m_N^{(i)}(z) \right| + \left| \left( \frac{N-1}{N} \right) m_N^{(i)}(z) - m_N(z) \right|.$$

Hay dos términos a controlar. A grandes rasgos el primero refleja las fluctuaciones de  $m_{N-1}$  al rededor de su media mientras el segundo relaciona  $m_{N-1}$  con  $m_N$ . Cada uno de estos términos se acotarán usando desigualdades de concentración. Al finalizar se reunirán los resultados con la cota para la suma de los  $G_{ii}$ , obteniendo una desigualdad para el decaimiento del segundo término.

**PRIMERA PARTE: FLUCTUACIONES DE  $m_{N-1}$ .** La función  $m_N(z)$  está definida sobre  $z$ , la cual no es aleatoria. Sin embargo  $m_N(z)$  es aleatoria porque depende de  $H_N$ , la cual se puede conseguir a partir de un vector  $X$  de v.a.i.i.d. como normales estándar reales. Las fluctuaciones de  $m_N(z)$  al rededor de su media se pueden estudiar como transformaciones Lipschitz sobre el vector  $X$ .

La idea básica es que se tiene una composición de funciones Lipschitz, correspondiente al siguiente diagrama, y por lo tanto toda la función es Lipschitz.

$$\begin{array}{cccc} \mathbb{C}^{N^2} & \rightarrow & \mathbb{H}_N & \rightarrow & \mathbb{R}^N & \rightarrow & \mathbb{C}^+ \\ (x_{ij}) & \mapsto & X & \mapsto & \Lambda(X) & \mapsto & N^{-1} \sum ((\Lambda(X))_i - z)^{-1} \end{array}$$

Para la demostración de esta propiedad de continuidad se usan los resultados descritos en la sección 2.7.

**Lema 3.4.4.** *Para  $z$  fija sea  $s_z : \mathbb{H}_N \rightarrow \mathbb{C}^+$  dada por*

$$s_z(X) = \frac{1}{N} \mathbf{Tr} \left( \left( \frac{X}{\sqrt{N}} - zI \right)^{-1} \right).$$

*La función  $s_z$  es  $\left( \frac{\sqrt{2}}{N\eta_0^2} \right)$ -Lipschitz sobre  $\mathbb{R}^{N^2}$ .*

*Demostración.* Por el teorema espectral

$$\begin{aligned} s_z(X) &= \frac{1}{N} \mathbf{Tr} \left( \left( U \frac{\Lambda}{\sqrt{N}} U^* - U(zI)U^* \right)^{-1} \right) \\ &= \frac{1}{N} \mathbf{Tr} \left( \left( \frac{\Lambda}{\sqrt{N}} - zI \right)^{-1} \right) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\frac{\lambda_i}{\sqrt{N}} - z}. \end{aligned}$$

Como  $X$  es Hermitiana sólo hace falta demostrar que  $g(x) = \frac{1}{N} \sum_i \frac{1}{N^{-1/2}x_i - z}$  es Lipschitz. Por la desigualdad de Jensen se tiene que

$$\begin{aligned} |g(x) - g(y)|^2 &= \left| \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{N^{-1/2}x_i - z} - \frac{1}{N^{-1/2}y_i - z} \right) \right|^2 \\ &\leq \frac{1}{N} \sum_{i=1}^N \left| \frac{N^{-1/2}y_i - z - N^{-1/2}x_i + z}{(N^{-1/2}x_i - z)(N^{-1/2}y_i - z)} \right|^2 \\ &= \frac{1}{N} \sum_{i=1}^N \frac{N^{-1} |y_i - x_i|^2}{|N^{-1/2}x_i - z|^2 |N^{-1/2}y_i - z|^2}. \end{aligned}$$

Dado que  $x \in \mathbb{R}^N$  se tiene que  $|N^{-1/2}x_i - z|^2 = |(N^{-1/2}x_i - E) + \mathbf{i}\eta|^2 \geq \eta^2 \geq \eta_0^2$ . De forma análoga para  $y$ . Luego

$$\begin{aligned} |g(x) - g(y)|^2 &\leq \frac{1}{N^2} \sum_{i=1}^N \frac{|y_i - x_i|^2}{\eta_0^2 \eta_0^2} \\ &= \frac{1}{N^2 \eta_0^4} \|x - y\|^2. \end{aligned}$$

■

Como consecuencia directa del lema anterior se tiene la siguiente desigualdad de concentración

**Proposición 3.4.5.** *Sea  $m_N(z)$  la transformada de Stieltjes de  $H$ , para  $\text{Im } z > \eta_0 > 0$ . Si  $\varepsilon > 0$ , entonces*

$$\mathbb{P}[|m_N(z) - \mathbb{E}[m_N(z)]| \geq \varepsilon] \leq 2 \exp \left\{ -\frac{\varepsilon^2}{4} N^2 \eta_0^4 \right\}. \quad (3.6)$$

Este resultado es muy útil. Dice que  $m_N(z)$  tiene un decaimiento sub-gaussiano. En partes posteriores de este texto se encontrarán combinaciones de  $N$  y  $\eta_0$  para usar Borel-Cantelli y vincular convergencia en probabilidad con  $L_1$ , casi-segura y localmente uniforme casi-seguramente.

Para esta parte de la prueba lo que se necesita es estudiar las fluctuaciones de

$$\left| \mathbb{E}[X_i] - \left( \frac{N-1}{N} \right) m_N^{(i)}(z) \right|.$$

De manera temporal se define  $\gamma = \sqrt{\frac{N}{N-1}}$ , recordando que  $m_N^{(i)}(z) \stackrel{d}{=} \gamma m_{N-1}(\gamma z)$ .

**Lema 3.4.6.** *Para todo  $\varepsilon > 0$ ,  $N \geq 1$  y  $\eta_0 > 0$*

$$\mathbb{P} \left[ \left| \mathbb{E}[X_i] - \left( \frac{N-1}{N} \right) m_N^{(i)}(z) \right| \geq \varepsilon \right] \leq 2 \exp \left\{ -\frac{1}{4} \varepsilon^2 N^2 \eta_0^4 \right\}. \quad (3.7)$$

*Demostración.* Por igualdad en distribución

$$\begin{aligned} \mathbb{P} \left[ \left| \mathbb{E}[X_i] - \left( \frac{N-1}{N} \right) m_N^{(i)}(z) \right| \geq \varepsilon \right] &= \mathbb{P} \left[ \left| \mathbb{E}[X_i] - \left( \frac{N-1}{N} \right) \gamma m_{N-1}(\gamma z) \right| \geq \varepsilon \right] \\ &= \mathbb{P} \left[ \left| \gamma \mathbb{E}[m_{N-1}(\gamma z)] - \left( \frac{N-1}{N} \right) \gamma m_{N-1}(\gamma z) \right| \geq \varepsilon \right] \\ &= \mathbb{P} \left[ \left| \gamma \left( \mathbb{E}[m_{N-1}(\gamma z)] - m_{N-1}(\gamma z) \right) \right| \geq \varepsilon \right]. \end{aligned}$$

Por desigualdad del triángulo

$$\{\gamma|a - \gamma^{-2}b| \geq \varepsilon\} = \{\gamma|a - b| + \gamma|1 - \gamma^{-2}||b| \geq \gamma|a - \gamma^{-2}b| \geq \varepsilon\} \subset \{|a - b| + |1 - \gamma^{-2}||b| \geq \gamma^{-1}\varepsilon\}.$$

Como  $\gamma > 1$  y  $(\gamma^{-2} - 1)|\mathbb{E}[m_{N-1}(\gamma z)]| > 0$  está acotado, se puede usar monotonía de la probabilidad para obtener

$$\begin{aligned} \mathbb{P}[\gamma|\mathbb{E}[m_{N-1}(\gamma z)] - \gamma^{-2}m_{N-1}(\gamma z)| \geq \varepsilon] &\leq \mathbb{P}[|m_{N-1}(\gamma z) - \mathbb{E}[m_{N-1}(\gamma z)]| \geq \gamma^{-1}\varepsilon] \\ &\leq 2 \exp\left\{-\frac{1}{4}\gamma^{-2}\varepsilon^2 N^2(\gamma\eta_0)^4\right\} \\ &= 2 \exp\left\{-\frac{1}{4}\varepsilon^2 N^2\eta_0^4\gamma^2\right\}. \end{aligned}$$

Y al ser  $\gamma > 1$  se tiene lo que se quería demostrar. ■

Este resultado acota con alta probabilidad el término que se quiere controlar para el teorema.

**SEGUNDA PARTE: RESOLVENTE DE LOS MINORES PRINCIPALES.** El siguiente término que se quiere controlar es una relación entre las fluctuaciones de  $m_N$  y las de sus menores principales  $m_N^{(i)}$ . Dicha relación se obtiene a través del entrelazamiento de los eigen-valores. La desigualdad de concentración, por su parte, deriva de la concentración de la norma operador de  $H$ .

**Lema 3.4.7.** Sea  $\rho := \frac{\eta_0}{2} + \|H_N\|_{Op}$ . La siguiente desigualdad se cumple **c.p.1**

$$\left| m_N(z) - \left(\frac{N-1}{N}\right) m_N^{(i)}(z) \right| \leq \frac{1}{N\eta_0} + \frac{2\rho}{N\eta_0^2}. \quad (3.8)$$

*Demostración.* Sean  $F_N$  y  $F_N^{(i)}$  las distribuciones empíricas espectrales de  $H_N$  y  $H_N^{(i)}$  respectivamente. Por definición de  $\rho$  el soporte de  $F_N$  está contenido en  $(-\rho, \rho)$ .

El teorema del entrelazado de Cauchy implica que para todo  $x \in \mathbb{R}$

$$\left| NF_N(x) - (N-1)F_N^{(i)}(x) \right| \leq 1.$$

En particular el entrelazado implica que el soporte de  $F_N^{(i)}$  está contenido en el de  $F_N$ .

Se puede usar integración por partes para cambiar  $dF_N$  por la medida de Lebesgue en la transformada de Stieltjes

$$m_N(z) = \int_{-\rho}^{\rho} \frac{dF_N(x)}{x-z} = \frac{F_N(x)}{x-z} \Big|_{x=-\rho}^{\rho} - \int_{-\rho}^{\rho} \frac{F_N(x)}{(x-z)^2} dx.$$

Se mencionó antes que el entrelazado de Cauchy asegura contención de soporte, lo que implica que  $F_N(\pm\rho) = F_N^{(i)}(\pm\rho)$  y

$$\frac{F_N(x)}{x-z} \Big|_{x=-\rho}^{\rho} = \frac{F_N^{(i)}(x)}{x-z} \Big|_{x=-\rho}^{\rho}.$$

Si se quiere comparar  $F_N$  y  $\frac{N-1}{N}F_N^{(i)}$  se tiene a su vez que

$$\left| \frac{F_N(x)}{x-z} \Big|_{x=-\rho}^{\rho} - \left(\frac{N-1}{N}\right) \frac{F_N^{(i)}(x)}{x-z} \Big|_{x=-\rho}^{\rho} \right| = \frac{1}{N|\rho-z|}.$$

A continuación se usa la desigualdad del triángulo y después  $|f \cdot| \leq f| \cdot|$ .

$$\begin{aligned} \left| m_N(z) - \left( \frac{N-1}{N} \right) m_N^{(i)}(z) \right| &\leq \frac{1}{N|\rho-z|} + \left| \int_{-\rho}^{\rho} \frac{F_N(x)}{(x-z)^2} dx - \left( \frac{N-1}{N} \right) \int_{-\rho}^{\rho} \frac{F_N^{(i)}(x)}{(x-z)^2} dx \right| \\ &\leq \frac{1}{N|\rho-z|} + \int_{-\rho}^{\rho} \left| \frac{F_N(x)}{(x-z)^2} - \left( \frac{N-1}{N} \right) \frac{F_N^{(i)}(x)}{(x-z)^2} \right| dx \\ &= \frac{1}{N|\rho-z|} + \frac{1}{N} \int_{-\rho}^{\rho} \frac{|NF_N(x) - (N-1)F_N^{(i)}(x)|}{|(x-z)^2|} dx \end{aligned}$$

Por el entrelazado de Cauchy,

$$\begin{aligned} \frac{1}{N|\rho-z|} + \frac{1}{N} \int_{-\rho}^{\rho} \frac{|NF_N(x) - (N-1)F_N^{(i)}(x)|}{|(x-z)^2|} dx &\leq \frac{1}{N|\rho-z|} + \frac{1}{N} \int_{-\rho}^{\rho} \frac{dx}{|(x-z)^2|} \\ &\leq \frac{1}{N\eta_0} + \frac{1}{N} \int_{-\rho}^{\rho} \frac{dx}{\eta_0^2}. \end{aligned}$$

Como el resultado es determinista dado la norma operador, no sólo se cumple **c.p.1**, sino que es válido en todas partes.  $\blacksquare$

El lema anterior permite comparar el resolventes de matrices con sus menores en función de la norma operador. Sin embargo la norma operador de una matriz GUE es una variable aleatoria. Su distribución ha sido muy estudiada y se sabe que decae al menos sub-exponencial (veáse corolario 2.3.6 de [Tao12]). Este resultado se prueba mediante un argumento de latiz sobre la esfera unitaria en  $\mathbb{C}^N$ .

**Proposición 3.4.8.** *Sea  $H \in \mathbb{H}_N$  con entradas isotrópicas sub-gaussianas y  $\mathbb{E}[|H_{ij}|^2] = N^{-1}$ . Entonces para todo  $K > 2$*

$$\mathbb{P}\left[\|H\|_{op} \geq K\right] \leq 2 \exp\left\{-\frac{1}{2}KN\right\}. \quad (3.9)$$

En particular si  $H$  es GUE el resultado se cumple.

Esta desigualdad se usa junto con la obtenida por el entrelazado de Cauchy.

**Lema 3.4.9.** *Para  $1 \geq \varepsilon > 0$  y  $N\eta_0 > 4\varepsilon^{-1} \max\{1, \frac{1}{2} + \eta_0^{-1}\}$  se cumple*

$$\mathbb{P}\left[\frac{1}{N\eta_0} + \frac{2\rho}{N\eta_0^2} \geq \varepsilon\right] \leq 2 \exp\left\{-\frac{\varepsilon^2}{2}N\eta_0\right\}. \quad (3.10)$$

*Demostración.* Despejando la norma operado en la desigualdad se tiene que

$$\begin{aligned} \mathbb{P}\left[\frac{1}{N\eta_0} + \frac{2\rho}{N\eta_0^2} \geq \varepsilon\right] &= \mathbb{P}\left[1 + \frac{2(\eta_0/2 + \|H\|_{op})}{\eta_0} \geq N\eta_0\varepsilon\right] \\ &= \mathbb{P}\left[\left(\eta_0/2 + \|H\|_{op}\right) \geq \frac{1}{2}(N\eta_0\varepsilon - 1)\eta_0\right] \\ &= \mathbb{P}\left[\|H\|_{op} \geq \frac{1}{2}(N\eta_0\varepsilon - 2)\eta_0\right]. \end{aligned}$$

Para poder usar (3.9) se requiere que  $K > 2$ , es decir que  $(N\eta_0\varepsilon - 2)\eta_0 > 4$ . Esto equivale a

$$N\eta_0 > 2(1 + 2\eta_0^{-1})\varepsilon^{-1} = 4\left(\frac{1}{2} + \eta_0^{-1}\right)\varepsilon^{-1}.$$

Por las condiciones de hipótesis esto se cumple, así que

$$\begin{aligned} \mathbb{P}\left[\|H\|_{op} \geq \frac{1}{2}(N\eta_0\varepsilon - 2)\eta_0\right] &\leq 2 \exp\left\{-\frac{1}{2}\left(\frac{1}{2}(N\eta_0\varepsilon - 2)\eta_0\right)N\right\} \\ &= 2 \exp\left\{-\frac{1}{4}((N\eta_0)\varepsilon - 2)(N\eta_0)\right\} \end{aligned}$$

Como  $N\eta_0\varepsilon > 4$  se tiene que  $((N\eta_0)\varepsilon - 2) > 2 \geq 2\varepsilon^2 > 0$ . Puede usarse esto en el orden de la exponencial

$$\mathbb{P}\left[\|H\|_{op} \geq \frac{1}{2}(N\eta_0\varepsilon - 2)\eta_0\right] \leq 2 \exp\left\{-\frac{1}{4}(2\varepsilon^2)(N\eta_0)\right\}$$

■

En este lema se usan dos hipótesis fuertes. La primera es que  $N\eta_0^2 \rightarrow \infty$ , la cual sólo sirve para utilizar (3.9) y podría no ser sólo una cuestión de elección. Es decir, podría no ser una hipótesis importante sino sub-óptima.

La segunda hipótesis que se usa es  $N\eta_0 \rightarrow \infty$ . Esta es una condición más débil y pareciera redundante mencionarla. Sin embargo, en la demostración anterior esta hipótesis decide el signo de la variable  $K$  que ha de calcularse. Para que las desigualdades tengan sentido  $K$  debe de ser positivo.

Con respecto a este lema no es necesario prestar tanta atención. El énfasis de esta observación es que ya comienza a haber argumentos que indican que  $N\eta_0 \rightarrow \infty$  para que la ley local tenga sentido. Lo importante de estos argumentos es que no tienen que ver con grados de polinomios, sino con el signo.

Se tienen ahora ambas cotas con alta probabilidad.

TERCERA PARTE: COTA DEL SEGUNDO TÉRMINO. Combinando ahora ambos resultados se tiene una cota para el segundo término con alta probabilidad.

**Lema 3.4.10.** *Sea  $\varepsilon \in (0, 1)$ . Para  $N \geq 1$  y  $\eta_0 > 0$  tal que  $N\eta_0^3 > 1 + 4\varepsilon^{-1}$  se cumple que*

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{i=1}^N(\mathbb{E}[X_i] - m_N(z))G_{ii}\right| \geq \varepsilon\right] \leq 4 \exp\left\{-\frac{1}{16}\varepsilon^2 N\eta_0^3\right\}. \quad (3.11)$$

*Demostración.* Nótese primero que  $X_i \stackrel{d}{=} X_j$  para todo  $i, j$ , entonces

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N(\mathbb{E}[X_i] - m_N(z))G_{ii} &= \left(\frac{1}{N}\sum_{i=1}^N G_{ii}\right)(\mathbb{E}[X_1] - m_N(z)) \\ &= m_N(z)(\mathbb{E}[X_1] - m_N(z)). \end{aligned}$$

Como  $|m_N(z)| \leq \eta_0^{-1}$ , la probabilidad que se busca es  $\mathbb{P}[|\mathbb{E}[X_1] - m_N(z)| \geq \eta_0 \varepsilon]$ . Por la desigualdad del triángulo

$$|\mathbb{E}[X_1] - m_N(z)| \leq \left| \mathbb{E}[X_1] - \left( \frac{N-1}{N} \right) m_N^{(1)}(z) \right| + \left| \left( \frac{N-1}{N} \right) m_N^{(1)}(z) - m_N(z) \right| =: A + B.$$

Si  $a + b = 1$  con  $a, b > 0$  se tiene

$$\{A + B \geq \eta_0 \varepsilon\} = \{A + B < \eta_0 \varepsilon\}^c = \{A + B < (a + b)\eta_0 \varepsilon\}^c$$

El caso  $\{A < a\eta_0 \varepsilon\} \cap \{B < b\eta_0 \varepsilon\}$  es un caso particular de  $\{A + B < (a + b)\eta_0 \varepsilon\}$ . Por las leyes de DeMorgan

$$\begin{aligned} \{A + B < (a + b)\eta_0 \varepsilon\}^c &\subset (\{A < a\eta_0 \varepsilon\} \cap \{B < b\eta_0 \varepsilon\})^c \\ &= \{A < a\eta_0 \varepsilon\}^c \cup \{B < b\eta_0 \varepsilon\}^c \\ &= \{A \geq a\eta_0 \varepsilon\} \cup \{B \geq b\eta_0 \varepsilon\}. \end{aligned}$$

La parte  $A$  está controlada por (3.7) mientras que la parte  $B$  junto con la hipótesis de que  $N\eta_0^3 \rightarrow \infty$  está controlado por (3.10). Para  $a$  y  $b$  definidos como

$$a := \frac{1}{1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}} \quad \text{y} \quad b := \frac{N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}}{1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}},$$

se tiene

$$\begin{aligned} \mathbb{P}[|\mathbb{E}[X_1] - m_N(z)| \geq \eta_0 \varepsilon] &\leq \mathbb{P}[A \geq a\eta_0 \varepsilon] + \mathbb{P}[B \geq b\eta_0 \varepsilon] \\ &\leq 2 \exp \left\{ -\frac{1}{4} \varepsilon^2 \eta_0^2 \frac{N^2 \eta_0^4}{\left(1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}\right)^2} \right\} + 2 \exp \left\{ -\frac{1}{2} \varepsilon^2 \eta_0^2 \frac{N \eta_0 \left(N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}\right)^2}{\left(1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}\right)^2} \right\} \\ &\leq 4 \exp \left\{ -\frac{1}{4} \varepsilon^2 \frac{N^2 \eta_0^6}{\left(1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}\right)^2} \right\}. \end{aligned}$$

Por la hipótesis inicial de  $N\eta_0^3 > 1$  se cumple que  $\left(1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}\right) \leq 2N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}$ . Luego

$$4 \exp \left\{ -\frac{1}{4} \varepsilon^2 \frac{N^2 \eta_0^6}{\left(1 + N^{\frac{1}{2}} \eta_0^{\frac{3}{2}}\right)^2} \right\} \leq 4 \exp \left\{ -\frac{1}{16} \varepsilon^2 \frac{N^2 \eta_0^6}{N \eta_0^3} \right\}.$$

■

Esto concluye la demostración respecto al segundo término.

### 3.4.3. DECAIMIENTO DEL TERCER TÉRMINO

El tercer término es  $\frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii}$  y para tener control sobre éste han de usarse varios pasos poco obvios. La idea principal es demostrar que  $\dot{X}_i$  es localmente Lipschitz como función de  $H_N$ . Una observación importante es que la norma un vector gaussiano tiene cola sub-gaussiana. Estas ideas se combinarán mediante la ley de probabilidad total para obtener cotas con alta probabilidad para el tercer término.

La variable  $X_i$ , que aparece en el tercer término se define como una fórmula cuadrática  $v^* A v$  donde tanto  $v$  como  $A$  son arreglos lineales de v.a. normales. Por lo tanto  $|\dot{X}_i|$  son las fluctuaciones de una fórmula cuadrática al rededor de su media, lo que sugiere usar concentración Gaussiana. Sin embargo, no es obvio que una fórmula cuadrática sea Lipschitz, menos aún contemplando toda la función  $(v, A) \rightarrow v^* A v$ . A continuación se prueba que, considerando  $A$  fija, una fórmula cuadrática es localmente-Lipschitz.

**Lema 3.4.11.** *Sea  $A \in \mathbb{H}_{N-1}$  fija y  $K \in \mathbb{R}^+$ . Supóngase que  $A$  tiene una raíz  $A^{1/2}$ . Para  $v \in \mathbb{C}^{N-1}$  tal que  $\|v\| \leq K$  la función*

$$F_A(v) := \frac{v^*}{\sqrt{N}} A \frac{v}{\sqrt{N}}$$

es  $\mathbb{C}^N$  Lipschitz con constante local  $\left( \frac{2K \|A^{1/2}\|_{Op}^2}{N} \right)$ .

*Demostración.* La función  $F_A(v)$  se puede reescribir como  $F_A(v) = \frac{1}{N} \|A^{1/2}v\|^2$ .

$$\begin{aligned} |F_A(v) - F_A(w)| &= \frac{1}{N} |v^* A v - w^* A w| \\ &= \frac{1}{N} \left| \|A^{1/2}v\|^2 - \|A^{1/2}w\|^2 \right| \\ \text{(dif. de cuadrados)} \quad &= \frac{1}{N} \left| (\|A^{1/2}v\| + \|A^{1/2}w\|) (\|A^{1/2}v\| - \|A^{1/2}w\|) \right| \\ &= \frac{1}{N} (\|A^{1/2}v\| + \|A^{1/2}w\|) \cdot \left| \|A^{1/2}v\| - \|A^{1/2}w\| \right| \\ \text{(def. norma operador)} \quad &\leq \frac{1}{N} \|A^{1/2}\|_{Op} (\|v\| + \|w\|) \cdot \left| \|A^{1/2}v\| - \|A^{1/2}w\| \right| \\ &\leq \frac{2K}{N} \|A^{1/2}\|_{Op} \left| \|A^{1/2}v\| - \|A^{1/2}w\| \right| \\ \text{(des. inv. del triang.)} \quad &\leq \frac{2K}{N} \|A^{1/2}\|_{Op} \|A^{1/2}v - A^{1/2}w\| \\ &\leq \frac{2K}{N} \|A^{1/2}\|_{Op}^2 \|v - w\| \end{aligned}$$

■

En el caso de interés  $X_i = H_{(i)}^* (H^{(i)} - zI)^{-1} H_{(i)}$ , donde  $H^{(i)}$  era un menor principal de  $H_N$  y  $H_{(i)}$  era la  $i$ -ésima columna de  $H_N$  sin la entrada  $i$ . Para este caso en particular se puede acotar la constante local de Lipschitz de antes. Sin pérdida de generalidad, debido a que las columnas de  $H$  son idénticamente distribuidas, se trabajará con  $i = 1$ .

**Lema 3.4.12.** *Sea  $A = A(z) = \left( H_N^{(1)} - zI \right)^{-1} \in \mathbb{H}_{N-1}$  para  $z \in \mathbb{C}^+$ . Si  $\text{Im } z \geq \eta_0 > 0$ , entonces existe una raíz  $A^{1/2}$  tal que*

$$\|A^{1/2}\|_{Op}^2 \leq \frac{1}{\eta_0}.$$

*Demostración.* La matriz  $H_N$  es Hermitiana, luego  $H^{(1)}$  también lo es. Sea  $U$  unitaria y  $\Lambda$  diagonal real dadas por el teorema espectral. Por tanto

$$A = (U\Lambda U^* - zI)^{-1} = U^* (\Lambda - zI)^{-1} U.$$

Aquí  $(\Lambda - zI)$  es diagonal con  $\text{Im} (\Lambda - zI)_{ii} \leq -\eta_0 < 0$ . Esto permite establecer una raíz cuadrada en común y definir  $A^{1/2} = U^* (\Lambda - zI)^{-1/2} U$ . Ahora

$$\begin{aligned} \|A^{1/2}\|_{Op}^2 &= \left\| (\Lambda - zI)^{-1/2} \right\|_{Op}^2 \\ &= \left( \max_{2 \leq i \leq N} \left\{ \left| \frac{1}{\sqrt{\Lambda_{ii} - z}} \right| \right\} \right)^2 \\ &= \max_{2 \leq i \leq N} \left\{ \frac{1}{|\Lambda_{ii} - z|} \right\} \\ &\leq \frac{1}{\text{Im}(z)}. \end{aligned}$$

■

Estos resultados se comprimen en el siguiente lema como un primer avance para acotar  $X_i$ .

**Lema 3.4.13.** *Sea  $A \in \mathbb{H}_{N-1}$  con raíz cuadrada y  $K \in \mathbb{R}^+$ . Para todo  $\varepsilon > 0$  se tiene que*

$$\mathbb{P} \left[ \left| \dot{X}_1(z) \right| \geq \varepsilon \mid \|H_{(1)}\| \leq K \right] \leq 2 \exp \left\{ -\frac{\varepsilon^2}{8K^2} N^2 \eta_0^2 \right\}. \quad (3.12)$$

*Demostración.* Defínase la función

$$F_{A,K}(v) := \frac{\|A^{1/2}v\|}{N} \left( \mathbb{1}_{\{\|v\| \leq K\}} + \frac{K}{\|v\|} \mathbb{1}_{\{\|v\| > K\}} \right).$$

En el conjunto  $\{H_{(1)} \leq K\}$  la función  $F$  equivale a  $\dot{X}_1$ . Además, por los lemas anteriores  $F$  es Lipschitz. Por probabilidad total se puede considerar que  $H^{(1)}$  es fija y usar concentración Gaussiana. Es decir

$$\begin{aligned} \mathbb{P} \left[ \left| F_{H^{(1)},K}(H_1) - \mathbb{E} [F_{H^{(1)},K}(H_1)] \right| \geq \varepsilon \right] &= \mathbb{E} \left[ \mathbb{P} \left[ \left| F_{H^{(1)},K}(H_1) - \mathbb{E} [F_{H^{(1)},K}(H_1)] \right| \geq \varepsilon \mid \sigma(H^{(1)}) \right] \right] \\ &\leq \mathbb{E} \left[ 2 \exp \left\{ -\frac{\varepsilon^2}{8K^2} N^2 \eta_0^2 \right\} \right]. \end{aligned}$$

■

El siguiente lema indica que conocer la cola de  $\dot{X}_1$  es suficiente para lo que se busca. Además sirve para encontrar la forma explícita que ha de tener  $K$  en las desigualdades anteriores.

**Lema 3.4.14.** *Sean  $\cdot X_i$  la variable centrada antes definida para  $z \in \mathbb{C}^+$ , con  $\text{Im } z > \eta_0 > 0$ . Para  $\varepsilon > 0$  se cumple lo siguiente*

$$\mathbb{P} \left[ \left| \frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii} \right| \geq \varepsilon \right] \leq N \mathbb{P} \left[ \left| \dot{X}_1 \right| \geq \eta_0 \varepsilon \right] \quad (3.13)$$

*Demostración.* Por la desigualdad de Hölder

$$\begin{aligned} \left| \frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii} \right| &\leq \left( \max_i \{ |\dot{X}_i| \} \right) \left( \frac{1}{N} \sum_{i=1}^N |G_{ii}| \right) \\ &\leq \frac{1}{\eta_0} \max_i \{ |\dot{X}_i| \}. \end{aligned}$$

Por sub-aditividad

$$\begin{aligned} \mathbb{P} \left[ \left| \frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii} \right| \geq \varepsilon \right] &\leq \mathbb{P} \left[ \max_i \{ |\dot{X}_i| \} \geq \eta_0 \varepsilon \right] \\ &= \mathbb{P} \left[ \bigcup_i \{ |\dot{X}_i| \geq \eta_0 \varepsilon \} \right] \\ &\leq \sum_i \mathbb{P} \left[ |\dot{X}_i| \geq \eta_0 \varepsilon \right]. \end{aligned}$$

El resultado se deduce de que las  $X_i$  son idénticamente distribuidas. ■

Para usar que la fórmula cuadrática antes vista es localmente Lipschitz es necesario demostrar que se pueden usar regiones acotadas con alta probabilidad.

**Lema 3.4.15.** *Sea  $h \sim \mathcal{N}_{\mathbb{C}}(0, N^{-1}I_{(N-1)})$ . Si  $K > 0$  entonces*

$$\mathbb{P}[\|h\|_2 \geq K] \leq 2N \exp \left\{ -\frac{K^2}{2} \right\}. \quad (3.14)$$

*Demostración.* Todas las normas en  $\mathbb{C}^M$  son equivalentes, en particular  $\|x\|_2 \leq \sqrt{M} \|x\|_{\infty}$ . Para este caso  $M = (N-1)$  y  $\sqrt{N}h_1 \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ . En consecuencia

$$\begin{aligned} \mathbb{P}[\|h\|_2 \geq K] &\leq \mathbb{P} \left[ \sqrt{(N-1)} \|h\|_{\infty} \geq K \right] \\ &\leq \mathbb{P} \left[ \left\| \sqrt{N}h \right\|_{\infty} \geq K \right] \\ &= \mathbb{P} \left[ \bigcup_{i=1}^{N-1} \{ |\sqrt{N}h_i| \geq K \} \right] \\ &\leq (N-1) \mathbb{P} \left[ |\sqrt{N}h_1| \geq K \right] \\ &\leq 2N \exp \left\{ -\frac{K^2}{2} \right\}. \end{aligned}$$

El último resultado es sobre el decaimiento de  $X_1$  que incluye que la continuidad localmente Lipschitz y la sub-gaussianidad de la norma euclidiana de un vector gaussiano. Se usará probabilidad total sobre la norma utilizando un umbral  $K$  que sea adecuado para los cálculos. ■

**Lema 3.4.16.** *Para  $\varepsilon > 0$  y  $N \geq 1$  se cumple la siguiente desigualdad*

$$\mathbb{P} \left[ |\dot{X}_1| \geq \eta_0 \varepsilon \right] \leq 4N \exp \left\{ -\frac{\varepsilon^2}{8} N \eta_0^2 \right\} \quad (3.15)$$

*Demostración.* Para  $K > 1$ , por ley de probabilidad total

$$\begin{aligned} \mathbb{P}\left[\left|\dot{X}_1\right| \geq \eta_0 \varepsilon\right] &= \mathbb{P}\left[\left|\dot{X}_1(z)\right| \geq \varepsilon \eta_0 \mid \|H_{(1)}\| \leq K\right] \mathbb{P}[\|H_{(1)}\| \leq K] \\ &\quad + \mathbb{P}\left[\left|\dot{X}_1(z)\right| \geq \varepsilon \eta_0 \mid \|H_{(1)}\| > K\right] \mathbb{P}[\|H_{(1)}\| > K] \\ &\leq \mathbb{P}\left[\left|\dot{X}_1(z)\right| \geq \varepsilon \eta_0 \mid \|H_{(1)}\| \leq K\right] + \mathbb{P}[\|H_{(1)}\| > K] \\ &\leq 2 \exp\left\{-\frac{\varepsilon^2}{8K^2} N^2 \eta_0^4\right\} + 2N \exp\left\{-\frac{K^2}{2}\right\} \end{aligned}$$

Eligiendo específicamente  $K = N^{1/2} \eta_0$  se tiene

$$\begin{aligned} \mathbb{P}\left[\left|\dot{X}_1\right| \geq \eta_0 \varepsilon\right] &\leq 2 \exp\left\{-\frac{\varepsilon^2}{8N\eta_0^2} N^2 \eta_0^4\right\} + 2N \exp\left\{-\frac{1}{2} N \eta_0^2\right\} \\ &= 2 \exp\left\{-\frac{\varepsilon^2}{8} N \eta_0^2\right\} + 2N \exp\left\{-\frac{1}{2} N \eta_0^2\right\} \\ &\leq 2N \exp\left\{-\frac{\varepsilon^2}{8} N \eta_0^2\right\} + 2N \exp\left\{-\frac{\varepsilon^2}{8} N \eta_0^2\right\}. \end{aligned}$$

■

Nótese que en realidad estos resultados son ciertos para todo  $N$  y  $\eta_0$ . La convergencia en probabilidad se tiene bajo condiciones de  $N$  y  $\eta_0$ , pero la desigualdad sigue siendo cierta incluso si dichas condiciones no se cumplen.

**Lema 3.4.17.** *Sea  $\varepsilon > 0$ . Para todo  $N \in \mathbb{N}$  y  $\eta_0 > 0$  se cumple que*

$$\mathbb{P}\left[\frac{1}{N} \left| \sum_{i=1}^N \dot{X}_i(z) G_{ii}(z) \right| \geq \varepsilon\right] \leq 8N^2 \exp\left\{-\frac{\varepsilon^2}{8} N \eta_0^2\right\}. \quad (3.16)$$

*Demostración.* Es consecuencia directa de los lemas 3.4.14 y 3.4.16. ■

Que el tercer término tienda a cero en probabilidad requiere que  $(N\eta_0^2 - 2 \log(N)) \rightarrow \infty$ . Para ello es suficiente que  $\eta_0$  sea función polinomial de  $N$  con grado mayor a  $-\frac{1}{2}$ . Esto concluye la prueba de la convergencia del tercer término.

### 3.4.4. LEY DEL SEMICÍRCULO DE WIGNER

En 1955 a través de un trabajo seminal Wigner demostró un teorema al que hoy se le conoce como la Ley del Semicírculo de Wigner. En la teoría de leyes locales a esta ley se le llama la Ley Global del Semicírculo [BK16].

Todos los resultados hasta este momento se pueden resumir en el siguiente resultado, a partir del cual se demuestran los demás. En particular el teorema 3.4.19 es de utilidad para la última parte del capítulo ya que se usa para la parte más moderna de la Ley Local, el *argumento bootstrap*.

**Proposición 3.4.18.** Sea  $\varepsilon \in (0, 1)$ . Si  $\{z_N\}_N \subset \mathbb{C}^+$  tal que  $\frac{(\operatorname{Im} z_N)^3}{4N} < \varepsilon$ , entonces

$$\mathbb{P}[|1 + zm_N(z) + m_N^2(z)| \geq \varepsilon] \leq 16N^2 \exp \left\{ -\frac{\varepsilon^2}{128} N (\operatorname{Im} z_N)^\beta \right\}, \quad (3.17)$$

donde

$$\beta = \begin{cases} 3 & 0 < \operatorname{Im} z_N < 1, \\ 2 & 1 \leq \operatorname{Im} z_N. \end{cases}$$

*Demostración.* Por el lema 3.3.3

$$|1 + zm_N(z) + m_N^2(z)| \leq \left| \frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii} \right| + \left| \frac{1}{N} \sum_{i=1}^N (\mathbb{E}[X_i] - m_N(z)) G_{ii} \right| + \left| \frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii} \right|$$

Y por lemas 3.4.3, 3.4.10 y 3.4.17, tomando  $\eta_0 = \operatorname{Im} z_N$  se tiene que

$$\begin{aligned} \mathbb{P} \left[ \left| \frac{1}{N} \sum_{i=1}^N H_{ii} G_{ii} \right| \geq \varepsilon \right] &\leq 2N \exp \left\{ -\frac{\varepsilon^2}{2} N \eta_0^2 \right\}. \\ \mathbb{P} \left[ \left| \frac{1}{N} \sum_{i=1}^N (\mathbb{E}[X_i] - m_N(z)) G_{ii} \right| \geq \varepsilon \right] &\leq 4 \exp \left\{ -\frac{\varepsilon^2}{16} N \eta_0^3 \right\}. \\ \mathbb{P} \left[ \left| \frac{1}{N} \sum_{i=1}^N \dot{X}_i G_{ii} \right| \geq \varepsilon \right] &\leq 8N^2 \exp \left\{ -\frac{\varepsilon^2}{8} N \eta_0^2 \right\}. \end{aligned}$$

El resultado se sigue fraccionar  $\varepsilon = \left(\frac{1}{4} + \frac{1}{2} + \frac{1}{4}\right) \varepsilon$ . ■

Se usa este resultado junto con la proposición 2.2.2 para probar  $m_N \rightarrow m_{SC}$  en probabilidad.

**Teorema 3.4.19** (Ley Global del Semicírculo). Sea  $1 > \varepsilon > 0$  y  $z = E + i\eta$  tal que  $\eta \geq 1$  y  $E \in \mathbb{R}$ . Se cumple que

$$\mathbb{P}[|m_N(z) - m_{SC}(z)| \geq \varepsilon] \leq 16N^2 \exp \left\{ -\frac{\varepsilon^2}{32} N \eta^4 \right\}. \quad (3.18)$$

*Demostración.* Por la proposición 2.2.2 sobre la auto-estabilidad del semicírculo

$$\mathbb{P}[|m_N(z) - m_{SC}(z)| \geq \varepsilon] \leq \mathbb{P}[|1 + zm_N(z) + m_N^2(z)| \geq \varepsilon\eta]$$

Y la conclusión se sigue de la proposición 3.4.18 ■

La ley global es la prueba por método de resolvente de la Ley del Semicírculo de Wigner, que puede expresarse en la siguiente forma

**Teorema 3.4.20** (Ley de Wigner). Sea  $\eta_0 > 0$  fijo. Para  $N \rightarrow \infty$  se cumple

$$\lim_{N \rightarrow \infty} \sup_{\substack{-4 \leq E \leq 4 \\ \eta_0 < \eta < \eta_0^{-1}}} |m_N(E + i\eta) - m_{SC}(E + i\eta)| = 0, \quad \mathbf{c.p.1.}$$

*Demostración.* Al ser  $\eta_0$  fijo y  $\int_0^\infty x^2 e^{-cx} dx < \infty$ , por Borel-Cantelli

$$|m_N(E + i\eta) - m_{SC}(E + i\eta)| \rightarrow_N 0 \quad \mathbf{c.p.1.}$$

En particular, por ser  $\{m_N\}$  familia de funciones analíticas la convergencia es localmente uniforme casi seguramente. ■

Lo importante de estos teoremas y por lo cuales son llamados leyes *globales*, es que  $\eta > 0$  es fijo. Si quiere analizarse  $\eta$  pequeño, entonces primero se hace convergencia  $N \rightarrow \infty$  y una vez se conoce el límite, se utiliza  $\eta \rightarrow 0$ . Sin embargo, el teorema 3.4.20 no ofrece un criterio de cómo elegir  $\eta$  para un  $N$  fijo.

El resultado encontrado durante esta tesis es más fuerte que este teorema, incluso que los ofrecidos comúnmente en la literatura como [BK16; Tao12], debido al carácter cuantitativo de las pruebas. Este resultado es consecuencia directa la proposición 3.4.18.

**Proposición 3.4.21.** *Sea  $\varepsilon \in (0, 1)$ . Tómesese una sucesión  $\{E_N, \eta_N\}_{N \geq 1}$  tal que  $\eta_N \in (0, 1)$  y  $\frac{\eta_N^3}{4N} < \varepsilon$ . Existen  $N_0$  absoluto y constantes  $C > 1 > c > 0$  tal que para todo  $N > N_0$  se cumple*

$$\mathbb{P}[|m_N(E_N + i\eta_N) - m_{SC}(E_N + i\eta_N)| \geq \varepsilon] \leq C \exp\{-c\varepsilon^2 N \eta_N^5\}. \quad (3.19)$$

Lo anterior implica que si  $\lim_{N \rightarrow \infty} N \eta_N^5 = \infty$ , entonces hay convergencia en probabilidad de las transformadas. En la práctica esto indica que es posible usar la transformada de Stieltjes en escalas  $N^{-\frac{1}{5} + \tau}$  con  $\tau > 0$  pequeño. La intención de la siguiente sección es mejorar esta escala hasta  $N^{-1 + \tau}$ , que es la escala óptima en la teoría de Leyes Locales.

## 3.5. PROCESO BOOTSTRAP

La función  $m_N$  es una variable aleatoria que toma valores en el espacio de funciones analíticas. En la figura 3.5.1 se puede apreciar una realización de  $m_N$ . En esta gráfica se nota que mientras más cercano sea  $z$  a la recta real las fluctuaciones son más grandes y viceversa, mientras más alejado esté  $z$  de la recta real, más estable se intuye su imagen.

Recuérdese que los teoremas 3.4.19 y 3.4.18 demuestran para  $\text{Im}(z) \geq 1$  la desigualdad

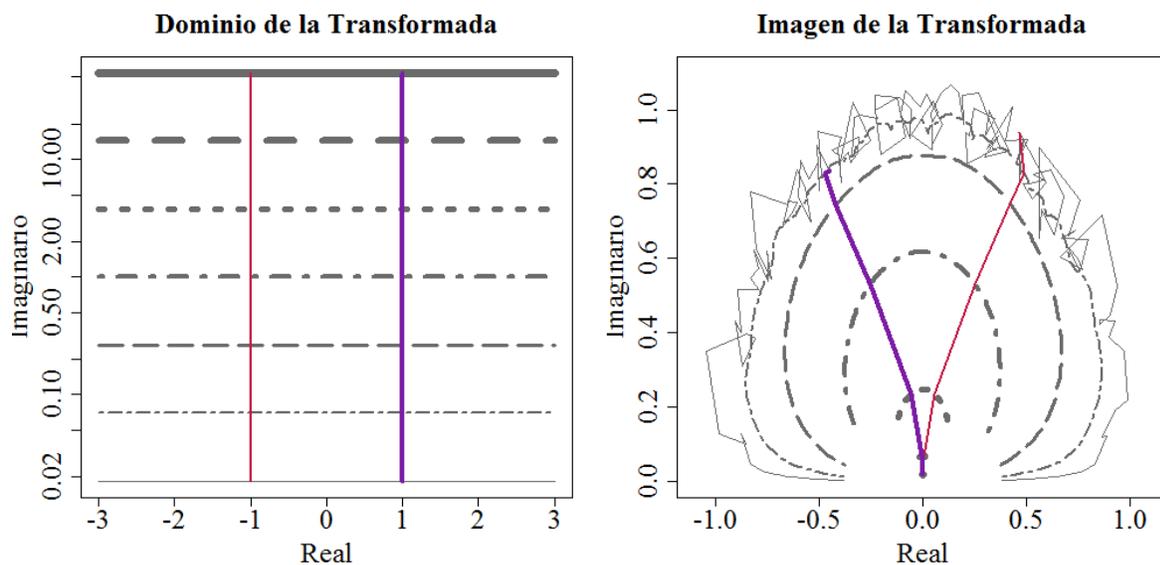
$$\mathbb{P}[|m_N(z) - m_{SC}(z)| \geq \varepsilon] \leq C \exp\{-c\varepsilon^2 N\}.$$

Este efecto se observa en la figura 3.5.2 donde se muestran varias realizaciones sobrepuestas la gráfica anterior, pero dejando fijo el valor real. Al igual que en la figura 3.5.1, para valores de  $z$  lejanos a la recta real, *i.e.*  $\eta$  grande, las realizaciones se asemejan mucho al valor predicho y las fluctuaciones son significantes para valores pequeños de  $\eta$ .

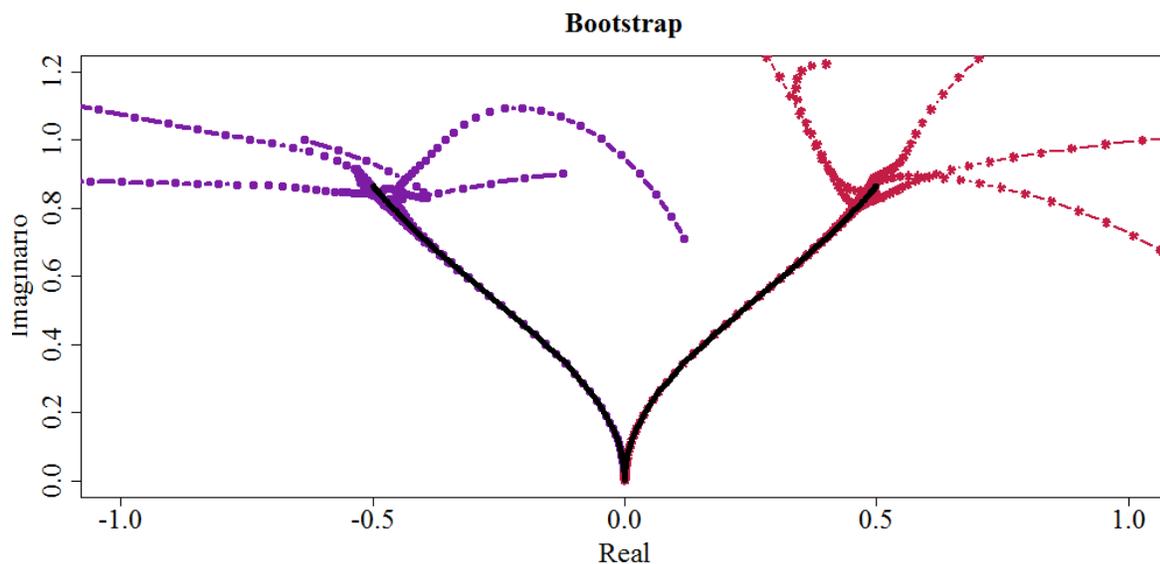
El interés de este trabajo de investigación se centra ahora en estos valores pequeños de  $\eta$ . La inquietud que se procura estudiar trata de averiguar el tamaño de estas fluctuaciones. Lo que se procura conseguir ahora es la ley local, esto es, conseguir una desigualdad similar bajo la condición de que  $\text{Im } z = \eta(N) \rightarrow 0$  cuando  $N \rightarrow \infty$ .

El procedimiento usual [ESY09; BK16; KY17] es aplicar un argumento bootstrap o de continuidad. Ésta es una técnica proveniente de Ecuaciones Diferenciales y a grandes rasgos es la idea de inducción sobre los puntos de una sucesión. Para una introducción formal y ejemplos clásicos se recomienda leer las páginas 20-26 del libro [Tao06].

A continuación se desarrolla el argumento bootstrap de leyes locales empleando un enfoque elemental. Los resultados presentados son originales de esta tesis y están lejanamente inspirados en [BK16].



**Figura 3.5.1:** Gráfica de la función  $m_N(z)$  en el plano superior complejo. La gráfica es el resultado de una realización con  $N = 512$ . A la izquierda los valores  $z \in \mathbb{C}^+$  que habrán de evaluarse y a la derecha se corresponden las líneas una vez evaluada la función.



**Figura 3.5.2:** Realizaciones de  $m_N(z)$  para  $z$  con parte real fija. A la izquierda con puntos redondos, realizaciones de  $m_N(1 + i\eta)$ . A la derecha con asteriscos, realizaciones de  $m_N(-1 + i\eta)$ . Los valores de  $\eta$  varían entre  $N^{-1}$  y  $N$ , con  $N = 512$ . En línea negra, el valor predicho por  $m_{SC}(z)$ .

### 3.5.1. EL ARGUMENTO DE CONTINUIDAD

El concepto clave en la técnica bootstrap es el paso/argumento de continuidad [Tao06]. Para el caso del semicírculo éste se obtiene de las propiedades Lipschitz de las transformadas de Stieltjes. Supóngase que ya se conoce el *error de aproximación*  $|m_N(z) - m_{SC}(z)|$  para un valor  $z$ . El paso de continuidad consiste en emplear este error para calcular el de  $\tilde{z} \approx z$ .

**Observación 3.5.1.** Sean  $z, \tilde{z} \in \mathbb{C}^+$ . Por la desigualdad del triángulo

$$|m_N(\tilde{z}) - m_{SC}(\tilde{z})| \leq |m_N(\tilde{z}) - m_N(z)| + |m_N(z) - m_{SC}(z)| + |m_{SC}(z) - m_{SC}(\tilde{z})|.$$

En esta desigualdad se puede emplear la suavidad de una transformada de Stieltjes.

**Lema 3.5.2.** Sea  $\mu$  distribución de probabilidad sobre la recta. Sean  $z, \tilde{z} \in \mathbb{C}^+$ .

$$\text{Si } \min\{\text{Im } z, \text{Im } \tilde{z}\} > N^{-1}, \text{ entonces } |m_\mu(z) - m_\mu(\tilde{z})| < N^2 |z - \tilde{z}|. \quad (3.20)$$

*Demostración.*

$$\begin{aligned} |m_N(z) - m_N(\tilde{z})| &= \left| \int \left( \frac{1}{x-z} - \frac{1}{x-\tilde{z}} \right) \mu(\mathbf{d}x) \right| \\ &\leq \int \left| \frac{z-\tilde{z}}{(x-z)(x-\tilde{z})} \right| \mu(\mathbf{d}x) \\ &= \int \frac{|z-\tilde{z}|}{|x-z||x-\tilde{z}|} \mu(\mathbf{d}x) \\ &\leq \int \frac{|z-\tilde{z}|}{\text{Im } z \cdot \text{Im } \tilde{z}} \mu(\mathbf{d}x) \\ &< \frac{|z-\tilde{z}|}{N^{-2}} \end{aligned}$$

■

En particular para  $\tilde{z} = z - \mathbf{i}N^{-3}$  se satisface

$$\begin{aligned} \left| m_N(\tilde{z}) - m_{SC}(\tilde{z}) \right| &\leq N^2 N^{-3} + |m_N(z) - m_{SC}(z)| + N^2 N^{-3} \\ &= |m_N(z) - m_{SC}(z)| + 2N^{-1}. \end{aligned}$$

Las observaciones previas tienen como consecuencia directa el siguiente resultado que contiene la idea clave del bootstrap en su forma más elemental.

**Lema 3.5.3** (Argumento de Continuidad). Sea  $\varepsilon > 0$  y  $z, \tilde{z} \in \mathbb{C}^+$  tales que  $\tilde{z} = z - \mathbf{i}N^{-3}$ . Si  $N > 2\varepsilon^{-1}$ , entonces

$$\mathbb{P}[|m_N(\tilde{z}) - m_{SC}(\tilde{z})| \geq \varepsilon] \leq \mathbb{P}[|m_N(z) - m_{SC}(z)| \leq \varepsilon - 2N^{-1}] \quad (3.21)$$

donde  $C > 1 > c > 0$  son constantes que no dependen de  $\varepsilon$  o  $N$ .

La observación importante sobre este paso es que *algo se pierde*. En particular el intervalo  $[0, \varepsilon - 2N^{-1}) \subset [0, \varepsilon)$ , por lo que se está *perdiendo* probabilidad al usar.

De hecho, si  $\text{Im } z \geq 1$  se tiene que

$$\mathbb{P}[|m_N(z) - m_{SC}(z)| \geq \varepsilon] \leq C \exp\{-c\varepsilon^2 N\},$$

mientras que, por 3.5.3 se obtiene

$$\begin{aligned} \mathbb{P}[|m_N(\tilde{z}) - m_{SC}(\tilde{z})| \geq \varepsilon] &\leq C \exp\left\{-c(\varepsilon - 2N^{-1})^2 N\right\} \\ &< C \exp\{-c\varepsilon^2 N\}. \end{aligned}$$

Mientras estos resultados son válidos e intuitivos, no son muy útiles para la prueba que se busca. Ésta necesita más estructura.

### 3.5.2. SUCESIÓN Y DOMINIO DE CONTINUIDAD

Si el argumento de continuidad es el corazón de la demostración, la elección de la sucesión y su dominio forman el esqueleto. En esta sección es técnica. Tiene como objetivo principal el presentar herramientas técnicas.

Todo proceso bootstrap necesita de la elección de dos objetos [Tao06]. El bootstrap utiliza una sucesión  $\{z_k\}$  y parte de condiciones iniciales en  $z_0$ , en este caso la ley global (teo. 3.4.19) y mediante continuidad (lem. 3.5.3), extiende el resultado de  $z_k$  a  $z_{k+1}$ . Con la sucesión se elige, a veces de forma indirecta, una región conexa de  $\mathbb{C}^+$  en la que los resultados siguen siendo válidos. Al aproximarse  $z_k$  al borde de la región es común que algunas propiedades del teorema comiencen a perderse, como la probabilidad, para el caso de leyes locales.

La siguiente definición asegura que  $\text{Im } z_0 \geq 1$ , para emplear la ley Global y que  $|z_{k+1} - z_k| = N^{-3}$ , útiles para el argumento bootstrap, el cuál también usa que  $\text{Im } z_k > N^{-1}$ .

**Definición 3.5.1.** Sea  $E \in \mathbb{R}$  fijo. Para todo  $N \in \mathbb{N}$  se define la sucesión  $\{z_k\}_{k=0}^{N^2(N-1)} \subset \mathbb{C}^+$

$$z_k = E - \mathbf{i}kN^{-3}. \quad (3.22)$$

Para definir la región de continuidad se necesita otra sucesión basada en una variable auxiliar, la cual es poco intuitiva, pero la experiencia indica que es indispensable. Como nota histórica, ésta aparece por primera vez en [Erd+13b] y es el origen de la notación propia de Leyes Locales.

**Definición 3.5.2.** Sea  $\sigma > 0$ . Se define  $\{\rho_k\}_{k=0}^{\infty}$  de tal forma que  $\sum_k 2\rho_k\sigma = \frac{1}{2}$ , por ejemplo

$$\rho_k = \frac{2^{-k}}{8\sigma}.$$

Ahora que se han definido dos sucesiones se necesita un medio para vincularlas y así definir el dominio de continuidad que se usará para la prueba.

**Proposición 3.5.4.** Sean  $\varepsilon, \sigma > 0$ . Existe  $N_\sigma \equiv N_{\sigma, \varepsilon} \in \mathbb{N}$  tal que para todo  $k \in \mathbb{N}$  y todo  $N > N_\sigma$

$$\left(N^{\rho\sigma} - \left(\frac{2}{N^{\sigma\varepsilon}}\right)N^{-1}\right) > 1. \quad (3.23)$$

*Demostración.* Dado que  $N^{1+\sigma} > 0$ , las siguientes expresiones son equivalentes

$$\begin{aligned} \left( N^{\rho_k \sigma} - \left( \frac{2}{N^\sigma \varepsilon} \right) N^{-1} \right) &> 1. \\ N^{\rho_k \sigma} N^{1+\sigma} - \left( \frac{2}{\varepsilon} \right) &> N^{1+\sigma}. \\ N^{1+\sigma} (N^{\rho_k \sigma} - 1) &> \left( \frac{2}{\varepsilon} \right). \end{aligned}$$

Además  $\rho_k \sigma > 0$ , así que  $N^{\rho_k \sigma} > 1$ . También  $N^{1+\sigma} \rightarrow \infty$ , cuando  $N \rightarrow \infty$ . Por tanto

$$\lim_{N \rightarrow \infty} N^{1+\sigma} (N^{\rho_k \sigma} - 1) = \infty.$$

■

Esta proposición permite definir el dominio conexo en el que va a ser válida la ley local.

**Definición 3.5.3.** Sean  $\varepsilon, \sigma > 0$  dados. Se define para  $X \in (N_\sigma, \infty)$

$$S_{\varepsilon, \sigma, X} := \left\{ z \in \mathbb{C}^+ : \text{Im } z > X^{-1} \right\}. \quad (3.24)$$

Un caso particular es  $S_N \equiv S_{\varepsilon, \sigma, N}$  cuando  $N \in \mathbb{N}$  y  $N > N_\sigma$ .

### 3.5.3. DEMOSTRACIÓN DEL TEOREMA

Una vez que las herramientas están completas puede demostrarse la parte principal de la Ley Local del Semicírculo.

**Lema 3.5.5.** Sean  $\varepsilon, \sigma > 0$  y  $\rho_k$  definidos como en la sección 3.5.2. Si  $N > N_\sigma$ ,  $k < N^2(N-1)$  es fijo,  $\frac{1}{2} < \alpha \leq 1$  y

$$\mathbb{P}[|m_N(z_k) - m_{SC}(z_k)| \geq N^\sigma \varepsilon] \leq C \exp \left\{ -c \varepsilon^2 N^\alpha \right\},$$

entonces

$$\mathbb{P}[|m_N(z_{k+1}) - m_{SC}(z_{k+1})| \geq N^\sigma \varepsilon] \leq C \exp \left\{ -c \varepsilon^2 N^{\alpha - 2\rho_k \sigma} \right\}. \quad (3.25)$$

*Demostración.* Por lema 3.5.3 e hipótesis se tiene que

$$\begin{aligned} \mathbb{P}[|m_N(z_{k+1}) - m_{SC}(z_{k+1})| \geq N^{\sigma + \rho_k \sigma} \varepsilon] &\leq \mathbb{P}[|m_N(z_k) - m_{SC}(z_k)| \geq N^{\sigma + \rho_k \sigma} \varepsilon - 2N^{-1}] \\ &\leq \mathbb{P}\left[|m_N(z_k) - m_{SC}(z_k)| \geq N^\sigma \left( N^{\rho_k \sigma} \varepsilon - \frac{2}{N^\sigma} N^{-1} \right)\right] \\ &\leq C \exp \left\{ -c \left( N^{\rho_k \sigma} \varepsilon - \frac{2}{N^\sigma} N^{-1} \right)^2 N^\alpha \right\}. \end{aligned}$$

Notar ahora que por elección de  $N_\sigma$  en prop. 3.5.4

$$\begin{aligned} \left( N^{\rho_k \sigma} \varepsilon - \frac{2}{N^\sigma} N^{-1} \right)^2 N^\alpha &= \left( N^{\rho_k \sigma} - \frac{2}{N^\sigma \varepsilon} N^{-1} \right)^2 \varepsilon^2 N^\alpha \\ &\geq (1)^2 \varepsilon^2 N^\alpha. \end{aligned}$$

Luego

$$\mathbb{P}[|m_N(z_{k+1}) - m_{SC}(z_{k+1})| \geq N^{\sigma+\rho_k\sigma}\varepsilon] \leq C \exp\{-c\varepsilon^2 N^\alpha\}.$$

Se requiere que esta última probabilidad esté en formato  $N^\sigma\varepsilon$  para ser compatible con la notación, aquí es donde se nota la pérdida de probabilidad. Sea  $\varepsilon' = N^{\rho_k\sigma}\varepsilon$ , por tanto  $\varepsilon = \varepsilon'N^{-\rho_k\sigma}$ . Sustituyendo se tiene

$$\mathbb{P}[|m_N(z_{k+1}) - m_{SC}(z_{k+1})| \geq N^\sigma\varepsilon'] \leq C \exp\left\{-c(\varepsilon')^2 N^{\alpha-2\rho_k\sigma}\right\}.$$

■

Aquí se atisban las ideas del bootstrap trabajando juntas. En este momento el peso de la probabilidad recae en la expresión  $N^{\alpha-2\rho_k\sigma}$ . Cada paso de continuidad impacta sobre la potencia de  $N$ . Notar que el término se añadió de forma recursiva lineal, es decir si  $\alpha_k > \frac{1}{2}$ , se puede definir

$$\alpha_{k+1} = \alpha_k - 2\rho_k\sigma = \alpha_0 - 2\sigma \sum_{l=0}^k \rho_l.$$

Si  $2\sigma \sum_{k=0}^{\infty} \rho_k < \alpha_0$ , entonces este proceso puede hacerse para cualquier  $k$ , manteniendo  $\alpha_k > 0$ . Esta es la razón por la que  $\sigma$ , y en particular  $N^\sigma$  es necesario, por que ayuda a controlar la pérdida de la velocidad. Estos comentarios se formalizan a continuación, que enuncia el resultado objetivo del capítulo.

**Teorema 3.5.6** (Ley Local del Semicírculo). *Sean  $\varepsilon > 0$  y  $\sigma > 0$ . Si  $N > N_\sigma$ , entonces*

$$\sup_{z \in S_N} \mathbb{P}[|m_N(z) - m_{SC}(z)| \geq N^\sigma\varepsilon] \leq C \exp\left\{-c\varepsilon^2 N^{\frac{1}{2}}\right\}. \quad (3.26)$$

*Demostración.* Por el lema 3.5.5 se tiene que

$$\sup_{0 \leq k \leq N^2(N-1)} \mathbb{P}[|m_N(z_k) - m_{SC}(z_k)| \geq N^\sigma\varepsilon] \leq \sup_k C \exp\left\{-c\varepsilon^2 N^{1-2\sigma \sum_{l=0}^k \rho_l}\right\}.$$

Por construcción se cumple que  $\inf_k 1 - 2\sigma \sum_{l=0}^k \rho_l = \frac{1}{2}$ , luego

$$\sup_{0 \leq k \leq N^2(N-1)} \mathbb{P}[|m_N(z_k) - m_{SC}(z_k)| \geq N^\sigma\varepsilon] \leq C \exp\left\{-c\varepsilon^2 N^{\frac{1}{2}}\right\}.$$

Para pasar al dominio conexo sólo recuérdese que si  $z = E + \mathbf{i}(kN^{-3} - tN^{-3})$  con  $t \in (0, 1)$ , entonces  $|z - z_k| < N^{-3}$  y por tanto el lema 3.5.3 es aplicable. Como consecuencia también es aplicable el lema 3.5.5 y por monotonía la cota supremo también acota al error en  $z$ . ■

Este teorema demuestra uniformidad al acercarse  $z$  a la recta real. Si se busca compatibilidad con la parte anterior del capítulo en que  $z = E + \mathbf{i}\eta$ , puede emplearse el siguiente

**Corolario.** *Sean  $\varepsilon, \sigma$  y  $\eta_N = N^{-1+\tau}$  con  $\tau \in (0, 1)$ . Si  $N > N_\sigma$ , entonces  $z = (E + \mathbf{i}\eta_N) \in S_N$  y*

$$\mathbb{P}[|m_N(z) - m_{SC}(z)| \geq \varepsilon] \leq C \exp\left\{-c\varepsilon^2 (N\eta_N)^{\frac{1}{2}}\right\}.$$

Así concluye la prueba de la Ley Local del Semicírculo.

### 3.6. DISCUSIÓN

La demostración es larga, compleja y delicada. Las pruebas fueron desarrolladas a lo largo del capítulo con un énfasis en la precisión y técnica, más que en las ideas. En contraste, esta última parte del capítulo comparte las ideas, generalidades y comentarios que surgieron durante el proyecto.

Estas ideas se agrupan en tres categorías. La primera es comparar los resultados finales con aquellos presentes en la literatura. La segunda presenta dos interpretaciones prácticas de los resultados finales, una estadística y la otra analítica. La tercera parte discute sobre cómo la prueba fue dependiente del modelo y qué partes habrían de cambiarse o se verían comprometidas en el caso de cambiarlo.

#### 3.6.1. COMPARACIÓN CON LA LITERATURA

A continuación se hace una comparación de los teoremas 3.5.6 con el estándar en la literatura.

Como referencia se usa [BK16], el cual a su vez está basado en [Erd+13b] y [Blo+14], de los que Antti Knowles es autor común. No se usará la notación propia de leyes locales, es decir la definición 2.5 de [BK16] sobre **Dominación Estocástica**. Se usará más bien una similar a la utilizada durante el capítulo.

**Teorema** (Ref: Teorema 2.6 de [BK16]). *Sea  $H$  una matriz de Wigner. Fíjese  $\tau > 0$  y defínase el dominio*

$$S_N(\tau) := \{E + \mathbf{i}\eta : |E| \leq \tau^{-1}, N^{-1+\tau} \leq \eta \leq \tau^{-1}\}.$$

*Entonces se tiene que para todo  $\sigma > 0$  existe  $N_\sigma$  tal que para todo  $N > N_\sigma$  se cumple*

$$\sup_{z \in S_N} \mathbb{P} \left[ |m_N(z) - m_{SC}(z)| \geq N^\sigma \frac{1}{N\eta} \right] \leq CN^{-D},$$

*donde  $C$  y  $D$  son constantes que no dependen de  $\sigma$ .*

Si se toma el resultado equivalente en este capítulo, tomando  $\varepsilon = (N\eta)^{-\frac{1}{5}}$  para hacer compatible la notación, se tiene

**Teorema** (Ref: Teorema 3.5.6). *Sea  $H$  una matriz GUE. Fíjese  $\tau > 0$  y defínase el dominio*

$$S_N(\tau) := \{E + \mathbf{i}\eta : |E| \in \mathbb{R}, N^{-1+\tau} < \eta\}.$$

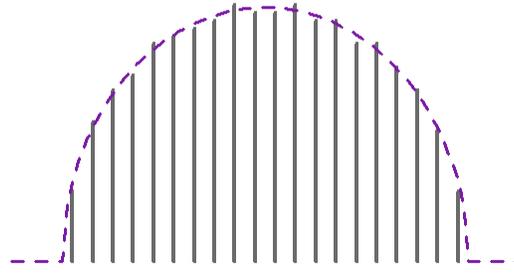
*Entonces existen  $C, c$  constantes y se tiene que para todo  $\sigma > 0$  existe  $N_\sigma$  tal que para todo  $N > N_\sigma$  se cumple*

$$\sup_{z \in S_N} \mathbb{P} \left[ |m_N(z) - m_{SC}(z)| \geq N^\sigma \left( \frac{1}{N\eta} \right)^{\frac{1}{5}} \right] \leq C \exp \left\{ -c (N\eta)^{\frac{1}{10}} \right\},$$

*donde  $C > 1 > c > 0$  son constantes que no dependen de  $\sigma$ .*

La primera observación es sobre el modelo. Dado que las matrices Wigner son más generales que el modelo específico GUE. El dominio de continuidad no es significativamente diferente, salvo que uno está acotado y otro no, quizás sea consecuencia del siguiente factor, la velocidad. Considerando que  $N\eta$  es grande [BK16] da una cota más fina, pero lo hace con menos probabilidad, pues la

## Histograma del Espectro



**Figura 3.6.1:** Histograma de una realización GUE  $N = 512$  y su comparación con la densidad del semicírculo.

exponencial es más rápida que cualquier polinomio. Nótese que el factor  $N^\sigma$  aparece en ambos teoremas y no hay indicios de que pueda retirarse. Afortunadamente estos teoremas se cumplen para cualquier  $\sigma > 0$ , de tal forma que  $N^\sigma \rightarrow 1$  cuando  $\sigma \rightarrow 0$ , sin embargo, en ambos casos  $\sup_\sigma N^\sigma = \infty$ .

Dadas estas comparaciones puede concluirse que el resultado es bastante competente, si bien no es el estándar.

### 3.6.2. INTERPRETACIONES

En este proceso de investigación se trabajó con dos perspectivas para entender la Ley Local del Semicírculo. La primera es una interpretación analítica con énfasis en la velocidad de convergencia. La otra se concentra en sus aplicaciones directas en estadística como un problema de estimación.

**INTERPRETACIÓN ANALÍTICA DE LA LEY LOCAL** El Teorema 3.5.6 puede interpretarse como un resultado que prueba no sólo convergencia, sino velocidad de convergencia. Esto permite combinar dos límites en uno.

La Ley Global (Teo. 3.4.19) y la Ley de Wigner (Teo. 3.4.20) se pueden combinar con la fórmula de inversión de Stieltjes (2.1) para afirmar que

$$\text{c.p.1} \quad \lim_{\eta \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{\mathbb{P}}{\pi} \text{Im } m_{X_N}(x + \mathbf{i}\eta) = \rho_{sc}(x),$$

Por otra parte, la Ley Local del Semicírculo (Teo. 3.5.6) combina ambos límites en

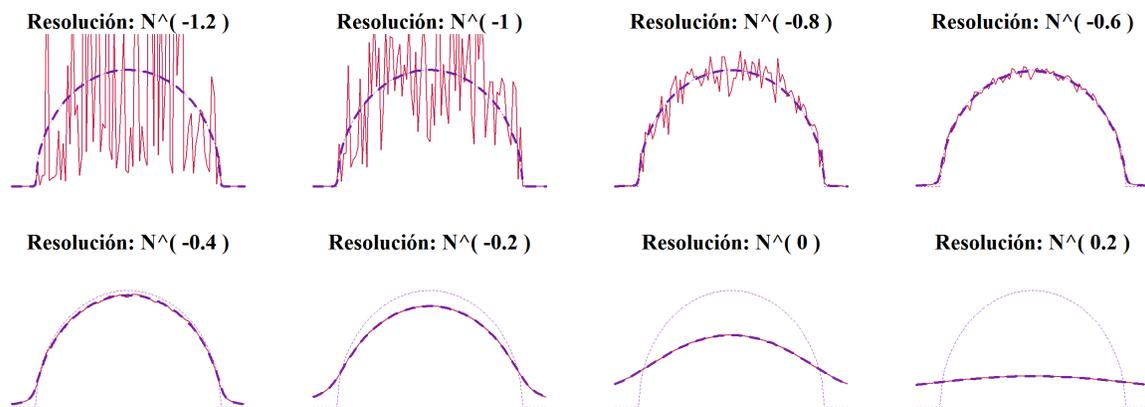
$$\lim_{\substack{N \rightarrow \infty \\ \eta \rightarrow 0 \\ N\eta^\alpha \rightarrow \infty}} \frac{\mathbb{P}}{\pi} \text{Im } m_{X_N}(x + \mathbf{i}\eta) = \rho_{sc}(x).$$

**INTERPRETACIÓN ESTADÍSTICA DE LA LEY LOCAL** Una necesidad práctica en matrices aleatorias es estimar la distribución empírica espectral. Por ejemplo, en la figura 3.6.1 se presenta una estimación para GUE usando el Teorema 3.4.20.

Al comienzo del Capítulo 2 se presenta la fórmula de inversión de Stieltjes. Recordar que la FDEE de una matriz GUE tiene un número finito de saltos (exactamente  $N$  saltos **c.p.1**), y al aplicar la fórmula de inversión lo que se obtiene es un *suavizamiento de la distribución* en base a la distribución Cauchy. Así se eliminan los átomos, se consigue diferenciabilidad y preserva suficiente estructura, aunque se pierden momentos. Este procedimiento se realiza con regularidad en estadística empleando muestras independientes y utilizando un *suavizamiento* Gaussiano. Tanto la familia de distribución Cauchy como la Gaussiana cuentan con dos parámetros: localización y escala. En estimación por kernels, como se llama a este procedimiento, se suele suponer que el parámetro de escala es cero y la dificultad radica en encontrar el parámetro de escala.

Para muestras independientes y con segundos momentos este parámetro de escala suele estar en función de  $\sqrt{N}$ , la raíz del tamaño de la muestra, lo que suele ser consecuencia del Teorema del Límite Central. Para el caso GUE, donde no hay independencia y la distribución Cauchy no tiene segundos momentos la elección se complica. La teoría de leyes locales estudia el comportamiento de la estimación cuando el parámetro de escala es pequeño, usualmente  $N^{-\alpha}$ .

En la figura 3.6.2 se presentan las transformadas de Stieltjes de una realización GUE con  $N = 512$ . Se intenta mostrar cómo diferentes parámetros de escala (o resolución) cambian la información que se obtiene de la FDEE.



**Figura 3.6.2:** Transformada Inversa de Stieltjes para  $\eta = N^{-\alpha}$ .

Línea continua (roja):  $\frac{1}{\pi} \text{Im } m_N(z)$ . Línea punteada (morada):  $\frac{1}{\pi} \text{Im } m_{SC}(z)$ . Línea punteada delgada: densidad del semicírculo.

Al aumentar la resolución, la estimación por kernels  $\frac{1}{\pi} \text{Im } m_N(z)$  gana suavidad y estabilidad, pero se aleja demasiado de cualquier uso práctico. Por el contrario, las resoluciones pequeñas son más fiables, pero muy inestables. De hecho es conocido [KY17; Erd+13b] que es inútil usar resoluciones cercanas o menores a  $N^{-1}$ , pues a esa escala dominan las fluctuaciones de las entradas de la matriz.

### 3.6.3. IMPLICACIONES EN LA ELECCIÓN DEL MODELO

Aunque la mayoría de las leyes locales en la literatura siguen los tres mismos pasos: ecuación funcional, ley global y bootstrap, la demostración dada en este capítulo depende mucho del modelo GUE.

La herramienta principal fue el Teorema 2.6.2 que ayuda a entender la cola de funciones suaves aplicadas a vectores Gaussianos. En el lema 3.3.3 se obtuvieron polinomios en entradas de un vector

Gaussiano, que son funciones suaves, por lo que el trabajo se centró en encontrar los coeficientes de Lipschitz.

Sin embargo, otras herramientas podrían usarse en esta situación, sobre todo si las entradas de la matriz no son Gaussianas. El artículo original [ESY09] emplea desigualdades de Sbolev, el libro [Tao12] usa McDiarmid y Talagrand y algunas pruebas recientes como [LS18] utilizan el método de Stein de [LP+09]. Todas estas desigualdades (de largas desviaciones) parten de un conocimiento de la cola de distribución de  $X$  y la extienden para funciones suave  $f(X)$  controlando la cola en base a la suavidad. El criterio de suavidad es el que identifica a cada desigualdad.

También hay que considerar que el modelo es Hermitiano, lo que facilitó varios cálculos, sobre todo en la demostración de 3.4.17. Para casos no simétricos se suelen emplear trucos de linearización como [HLS+19].

Elegir el modelo GUE tiene, en cambio, una consecuencia sutil. El Teorema 3.4.19 muestra que el *error* del modelo respecto al semicírculo es *sub-gaussiano*, es decir  $\mathbb{P}[X \geq x] \leq Ce^{-cx^2 f(N)}$ , donde  $f(N)$  es un polinomio de  $N$ . Aquí es donde la elección del modelo jugó un papel importante ya que cada paso de bootstrap conservaba la propiedad, pero cambiando  $f(N)$  por alguna  $g(N)$  que también era un polinomio de  $N$ , aunque  $g(N) < f(N)$ . En las notas [BK16] el bootstrap se prueba usando un decaimiento potencia, *i.e.*  $\mathbb{P}[X \geq x] \leq CN^{-cx}$ . En ambos casos, sub-gaussiano y potencia, la idea es utilizar los polinomios de  $N$  que aparecen en la distribución de las entradas del modelo  $H_n$  y explotarlas para llegar a escalas más pequeñas.

Debido a esto se recalcó al describir el modelo GUE que la distribución de las entradas sí depende de  $N$ .



# CAPÍTULO 4

## MATRICES ALEATORIAS RALAS

El Capítulo 3 presenta un resultado teórico comprendido como seminal a la Teoría de Leyes Locales [KY17], muchos de los resultados de la teoría extienden las ideas del Capítulo 3 a otros modelos de matrices, generando así varias líneas de investigación. Una de las líneas activas de esta teoría intenta explicar el comportamiento de matrices aleatorias cuyas entradas se anulan con muy alta probabilidad, así llamadas *ralas*. Dicha línea comienza con [Erd+13a] y en fechas más recientes se puede encontrar el artículo [LS18], con continuación en los *pre-prints* [HLS+19; HLY19].

El capítulo actual realiza una revisión experimental a la literatura de esta línea de investigación.

Dividido en tres secciones el capítulo presenta los modelos matriciales, algunos resultados teóricos y los complementa con experimentos. La sección 4.1 desarrolla el concepto de matriz aleatoria. La sección 4.2 compila resultados de la literatura en Leyes Locales para matrices aleatorias ralas y señala su diferencia con modelos matriciales más generales. Finalmente, en la sección 4.3, se validan de forma empírica los resultados teóricos de [HLS+19] sobre matrices aleatorias ralas rectangulares usando simulaciones.

### 4.1. MODELACIÓN ESTOCÁSTICA DE UNA MATRIZ RALA

Esta tesis propone que para definir una matriz aleatoria rala deben considerarse tres propiedades respecto a las entradas de la matriz, las cuales se desarrollan en esta sección. Dicha propuesta se basa en los supuestos teóricos y argumentación de los artículos [LS18; HL16; DZ18] y la experiencia académica del que defiende [Cau16].

#### 4.1.1. PARÁMETRO DE RALEZA

El primer concepto es sobre la proporción de entradas nulas en una matriz. El parámetro que mide esta proporción fue definido en [HL16], como continuación del trabajo en [Erd+13a; Erd+13b] sobre la gráfica de Erdős-Renyi, que se muestra más adelante como ejemplo.

**Definición 4.1.1.** Para una matriz  $X \in \text{Mat}_{M \times N}(\mathbb{C})$ , cuyas entradas son aleatorias. se define su **parámetro de raleza**  $q \equiv q(M, N)$  de tal forma que

$$q^2 := \frac{1}{M} \sum_{i=1}^M \mathbb{E} \left[ \sum_{j=1}^N \mathbb{1}_{(X_{ij} \neq 0)} \right] = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N \mathbb{P}[X_{ij} \neq 0]. \quad (4.1)$$

Se dice que  $X$  es **matriz aleatoria rala** si

$$\lim_{\min(M,N) \rightarrow \infty} \frac{q^2}{\min(M,N)} = 0, \quad (4.2)$$

y que es **matriz aleatoria densa** si  $q^2 \approx \min(M, N)$ .

Es usual pedir que  $q$  sea de orden polinomial en  $N$ . Algunos equipos de investigación [LS18; DZ18] suelen delimitar los modelos con la siguiente suposición

**Definición 4.1.2.** Sean  $0 < \phi_1 < \phi_2 < \frac{1}{2}$ , se dice que  $X$  **satisface la primera condición de rareza** cuando

$$N^{\phi_1} < q < N^{\phi_2}. \quad (4.3)$$

Para ilustrar las definiciones considérense  $M = N$  y  $\mathbb{P}[X_{ij} \neq 0] = p$  para todo  $1 \leq i, j \leq N$ . En esta situación  $q^2 = Np$  y  $0 \leq q \leq N^{1/2}$ , así en cada renglón pueden esperarse  $Np$  entradas no nulas.

Con este ejemplo, el caso  $p = N^{-\frac{1}{2}}$  satisface la primera condición, pero  $p = N^{-1}$  no.

De manera más general, si  $p < N^{-1}$ , entonces  $\lim Np < 1$ , con lo que se puede esperar no sólo que la dimensión del kernel por renglones sea alta, sino que la misma norma de la matriz (normalizada por  $\sqrt{N}$ ) sea acotada, haciendo que la distribución empírica espectral (DEE) converja a un átomo en cero, Caso diferente es el de  $p = N^{-1/2}$ , el cuál implica que  $Np = N^{1/2}$  y por lo tanto la convergencia a un átomo no aparenta ser un problema.

#### 4.1.2. DECAIMIENTO DE LAS ENTRADAS

La primera condición trata sobre el aumento en la proporción de entradas nulas:  $\mathbb{P}[X_{ij} = 0] \rightarrow 1$ . La siguiente restricción es sobre la distribución marginal de las entradas  $X_{ij}$ .

En el Capítulo 3 se probó un límite para  $X \in \mathbb{H}_N$  cuando  $X_{ij} \sim \mathcal{N}_{\mathbb{C}}(0, N^{-1})$ , usando resultados de la forma

$$\mathbb{P}[f((X_{ij})) > \varepsilon] < g(\varepsilon).$$

Los artículos del área de leyes locales que trabajan con matrices ralas [Erd+13a; HL16; LS13] imponen restricciones de este tipo a través de supuestos en los momentos usando la desigualdad de Markov.

**Teorema** (Desigualdad de Markov). *Sea  $Z > 0$  **c.p.1**, variable aleatoria. Si  $\mathbb{E}[Z] < \infty$ , entonces*

$$\mathbb{P}[Z > \varepsilon] \leq \frac{\mathbb{E}[Z]}{\varepsilon}.$$

Un supuesto en los momentos, muy útil para matrices aleatorias [ER05; Tao12; EYY12; EKY13; And+15] es el siguiente.

**Definición 4.1.3.** Si  $Z$  es v.a. tal que para algún  $\lambda > 0$  se cumple  $\mathbb{E}[e^{\lambda Z}] < \infty$ , se dice que  $Z$  es **sub-exponencial**.

Como consecuencia de esta definición se tiene el siguiente teorema (consultar [Ban15]).

**Teorema 4.1.1.** *Si  $Z$  es v.a. sub-exponencial, entonces existen constantes  $C, c > 0$  tales que para todo  $k \geq 3$*

$$\mathbb{E}[|Z|^k] \leq (Ck)^{ck}.$$

En el caso de v. a. Bernoulli se pueden calcular desigualdades de forma más o menos explícita.

**Proposición 4.1.2.** *Sea  $N \in \mathbb{N}$  y  $B_N \sim \text{Ber}(p_N)$ . Si  $q_N = \sqrt{Np_N}$ , entonces  $B_N$  tiene momentos estandarizados  $k$  de orden  $\frac{O(1)}{Nq_N^{k-2}}$ .*

*Demostración.* La versión estandarizada de  $B_N$  es

$$\dot{B}_N := \frac{B_N - p_N}{\sqrt{Np_N(1-p_N)}}.$$

Los primeros momentos cumplen  $\mathbb{E}[\dot{B}_N] = 0$  y  $\mathbb{E}[\dot{B}_N^2] = N^{-1}$ . Para momentos de orden  $k > 2$  se tiene que

$$\mathbb{E}[|\dot{B}_N|^k] = \frac{1}{(Np_N(1-p_N))^{k/2}} \left( (1-p_N)^k + p_N^{k-1}(1-p_N) \right) p_N,$$

usando que  $q_N = \sqrt{Np_N}$  se simplifica la expresión anterior

$$\begin{aligned} \mathbb{E}[|\dot{B}_N|^k] &= \frac{q^2}{Nq^k} \frac{(1-p_N)^k + p_N^{k-1}(1-p_N)}{(1-p_N)^{k/2}} \\ &= \frac{1}{Nq^{k-2}} \left( (1-p_N)^{k/2} + p_N^{k-1}(1-p_N)^{1-k/2} \right). \end{aligned}$$

La expresión al interior del paréntesis es un polinomio en  $p_N$ , si suponemos que  $|p_N| \ll 1$ , se puede aproximar por su serie de Taylor alrededor de 0. De hecho

$$\begin{aligned} \mathbb{E}[|\dot{B}_N|^k] &= \frac{1}{Nq^{k-2}} \left( 1 + \frac{k}{4}p_N + o(p_N) \right) \\ &= \frac{O(1)}{Nq_N^{k-2}}. \end{aligned}$$

■

Ahora, se necesita conectar condiciones en la cola de la distribución con la idea de rareza, por ejemplo, mediante el siguiente concepto [LS18].

**Definición 4.1.4.** Sean  $Z$  v. a. independiente a  $\{B_N\}$  donde  $B_N \sim \text{Ber}(p_N)$ . Al proceso

$$Z_N := Z \cdot B_N, \tag{4.4}$$

se le conoce como **dilución de la variable  $Z$**

La importancia de este proceso es que

$$\limsup_N \mathbb{P}[Z_N \neq 0] \leq \limsup_N \mathbb{P}[B_N \neq 0] = \limsup_N p_N = 0.$$

La dilución de una variable sub-exponencial inspira la siguiente definición.

**Definición 4.1.5** (Dada por [HL16]). Sea  $X \in \text{Mat}_{M \times N}(\mathbb{C})$  una matriz aleatoria y  $q$  su parámetro de rareza. Se dice que  $X$  **satisface la segunda condición de rareza** si existen  $C, c > 0$  tales que, para todo  $k \geq 3$

$$\sup_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \mathbb{E} \left[ \left| \frac{X_{ij} - \mathbb{E}[X_{ij}]}{\sqrt{N \text{Var}(X_{ij})}} \right|^k \right] \leq \frac{(Ck)^{ck}}{Nq^{k-2}}. \quad (4.5)$$

Esta condición controla el decaimiento de la cola de distribución y permite demostrar teoremas límite de probabilidad. En los teoremas para matrices aleatorias ralas, la rareza suele aparecer en las ecuaciones a través del término  $Nq_N^{k-2}$ .

### 4.1.3. ESTRUCTURA CONJUNTA

Dos propiedades importantes se presentaron para delimitar el objeto de trabajo. Ambas tienen en común que se definieron de forma marginal. La tercera propiedad trata la estructura con la que los ceros aparecen en la matriz.

**Definición 4.1.6.** Sea  $X \in \text{Mat}_{M \times N}(\mathbb{C})$  una matriz aleatoria y  $q$  su parámetro de rareza.

Se define el **perfil de nulidad** de  $X$  como la distribución conjunta de la familia

$$\mathcal{B}^X := \{ \mathbb{1}_{(X_{ij} \neq 0)} \mid 1 \leq i \leq M, 1 \leq j \leq N \}.$$

Dada una sucesión  $\{M_N\}_{N \in \mathbb{N}} \subset \mathbb{N}$ , el proceso estocástico matricial  $\{X_N\}_{N \in \mathbb{N}}$  con  $X_N$  v.a. valuada en  $\text{Mat}_{M_N \times N}(\mathbb{C})$  se dice que es un **ensamble raro con perfil de nulidad**  $\{\mathcal{B}^{X_N}\}_N$  si la sucesión  $\{X_N\}$  satisface las condiciones primera y segunda de rareza.

### 4.1.4. EJEMPLOS DE MODELOS MATRICIALES RALOS

El modelo más sencillo de un ensamble raro es en el que las entradas son independientes y tienen distribución de Bernoulli con parámetro común [Erd+13b].

**Definición 4.1.7** (Modelo 1: Gráfica de Erdős-Rényi ó Modelo de Gilbert). Dada una sucesión  $\{p_N\} \subset (0, 1)$  se construye el proceso estocástico matricial  $\{G(N, p_N)\}_N$  donde para cada  $N$ , la matriz  $G(N, p_N) = (a_{ij})$  es simétrica, con  $a_{ii} = 0$  en la diagonal, v.a. Bernoulli fuera de la diagonal

$$a_{ij} = \begin{cases} 0 & \text{con probabilidad } 1 - p_N \\ 1 & \text{con probabilidad } p_N \end{cases} \quad i \neq j$$

y la familia  $\{a_{ij} : 1 \leq i < j \leq N\}$  es independiente.

Por la Proposición 4.1.2, se sabe que este modelo satisface la segunda condición de rareza. También se conoce su perfil de nulidad ya que  $\mathcal{B}^{G(N, p_N)} = G(N, p_N)$  con entradas independientes. La definición de ensamble raro se completa al imponer, por ejemplo, que  $p_N = N^{-1+\varepsilon}$  y por tanto  $p_N \rightarrow 0$ .

Otro modelo es el *Stochastic Block Model* que suele usarse en detección de comunidades [HLS+19; HLY19; DZ18]. En este se considera una partición finita de los naturales en  $K$  bloques y la distribución de las entradas depende de si los índices pertenecen al mismo bloque o no.

**Definición 4.1.8** (Modelo 2: Stochastic Block Model). Dadas dos sucesiones  $\{p_N\}, \{p'_N\} \subset (0, 1)$ , un número  $K \in \mathbb{N}$  y una función  $s : \mathbb{N} \rightarrow \{1, \dots, K\}$  se construye el proceso estocástico matricial  $\{X_N\}$  donde para cada  $N$ , la matriz  $X_N = (a_{ij})$  es simétrica, con  $a_{ii} = 0$  en la diagonal, v.a. Bernoulli fuera de la diagonal

$$a_{ij} = \begin{cases} 0 & \text{si } s(i) = s(j) \text{ con probabilidad } 1 - p_N \\ 1 & \text{si } s(i) = s(j) \text{ con probabilidad } p_N \\ 0 & \text{si } s(i) \neq s(j) \text{ con probabilidad } 1 - p'_N \\ 1 & \text{si } s(i) \neq s(j) \text{ con probabilidad } p'_N \end{cases} \quad i \neq j$$

y la familia  $\{a_{ij} : 1 \leq i < j \leq N\}$  es independiente.

Se suelen imponer condiciones sobre  $p_N$  y  $q_N$  en función de  $s(\cdot)$  para cumplir las condiciones de rareza. Véase [HLY19] para más detalles.

Un ejemplo aún más complejo y que aparece a menudo en aplicaciones [Cau16] se inspira en los anteriores y en regresión logística. Suponer que el parámetro de cada v.a. Bernoulli depende directamente de los índices.

**Definición 4.1.9** (Modelo 3: Logístico, independiente, no homogéneo). Sea  $\{u_i \in \mathbb{R}^k; i \in \mathbb{N}\}$  una familia de vectores aleatorios independientes. Para cada  $i$  a  $u_i$  se le llama **las características del nodo**.

Dadas dos sucesiones aleatorias  $\{\alpha_N\}, \{\beta_N\}$  tales que  $\alpha_N \rightarrow -\infty$  y  $\beta_N = O(1)$  se construye el proceso estocástico matricial  $\{X_N\}_N$  donde para cada  $N$ , la matriz  $X_N = (a_{ij})$  es simétrica, con  $a_{ii} = 0$  en la diagonal, la familia  $\{a_{ij} : 1 \leq i < j \leq N\}$  es independiente (dados  $\alpha_N, \beta_N$  y  $\{u_i\}$ ) y fuera de la diagonal

$$a_{ij} \sim \text{Ber} \left( \frac{1}{1 + \exp \{-\alpha_N - \beta_N u_i^T u_j\}} \right).$$

Éste es un perfil de nulidad y para completar el modelo se requieren hipótesis adicionales sobre  $u_i$  y  $\alpha_N, \beta_N$ . Por ejemplo,  $u_i \sim \mathcal{N}_{\mathbb{R}^k}(0, I)$  es una hipótesis usual en aplicaciones [Cau16].

Si se piensa  $X$  como la matriz de adyacencia de una gráfica,  $u_i^T u_j$  representa una similaridad entre los nodos  $i$  y  $j$ . En esta situación el signo de  $\beta_N$  decide si se conectan nodos similares o diferentes y el valor de  $\alpha_N$  restringe la cantidad total de conexiones que habrá.

Los tres modelos presentados en este apartado están definidos en función a variables aleatorias Bernoulli independientes. En la sección 4.1.2 se dijo que la dilución de variables sub-exponenciales también son modelos válidos. Para ver más ejemplos de ensamblados raros y dilución se sugiere consultar [LS18; HLS+19; HL16].

## 4.2. LEYES LOCALES PARA MATRICES RALAS SIMÉTRICAS

En la literatura se conoce que algunas matrices aleatorias raras, si bien cumplen teoremas límite, como la Ley Local del Semicírculo del Capítulo 3, también tienen leyes locales aproximantes que funcionan aún mejor. Mientras que las leyes como la del Semicírculo reciben el nombre de *universales*, estas aproximaciones específicas a un tipo de modelo se les llama *sesgadas* (de *shift*, en inglés).

En esta sección se compilan algunos de estos resultados para matrices aleatorias raras en el contexto de universalidad y sesgo.

La sección, tras establecer la notación, consta de tres partes. Primero se presenta el fenómeno de universalidad. Después se listan algunos teoremas para modelos sencillos y se finaliza con un ejemplo más complejo de años recientes.

### 4.2.1. RECORDATORIO DE LA NOTACIÓN

Es tradición en los trabajos de leyes locales (*e.g.* [ESY09; EYY12; EKY13; Adl13; GNT15; BK16; KY17; LS18; HLS+19; WZ19]) escribir el siguiente párrafo

**Definición 4.2.1** (Convención Notacional de Leyes Locales). Se usan los símbolos  $O(\cdot)$  y  $o(\cdot)$  para la notación estándar *O-grande* y *o-pequeña*. Las notaciones  $O, o, \ll, \gg$  se refieren al límite  $N, M \rightarrow \infty$  con  $N/M = d$  fija, a menos que se especifique lo contrario. Aquí,  $a \ll b$  significa que  $a = o(b)$ . Se usan  $c$  y  $C$  para denotar constantes que no dependen de  $N$ , usualmente con la convención  $c \leq C$ . Sus valores pueden cambiar de línea a línea. Se escribe  $a \approx b$  si existe  $C \geq 1$  tal que  $C^{-1}|b| \leq a \leq C|b|$ . Se entiende por lo general que  $z = E + i\eta$ , donde  $E \in \mathbb{R}$  y  $\eta > 0$ , además  $\mathbb{C}^+ = \{z \in \mathbb{C} : \text{Im } z > 0\}$ .

Es pertinente mencionar que en la literatura reciente de leyes locales los resultados suelen presentarse en una notación conocida como *stochastic domination* (veáse definición 2.5 de [BK16]). Por razones de consistencia con el Capítulo 3, dicha notación no será usada en este texto.

Un recordatorio de la notación en esta tesis:  $\mathbb{H}_N$  y  $\mathbb{S}_N$  representan las matrices Hermitianas y simétricas de tamaño  $N \times N$ , respectivamente. Si  $X$  es una matriz,  $\mu_X$  representa su distribución empírica espectral y  $m_X(z) = \mathbf{Tr}((X - zI)^{-1})$  es su transformada de Stieltjes (veáse Capítulo 2).

De manera auxiliar se define [LS18] la siguiente región de  $\mathbb{C}^+$

$$\mathcal{E} := \{z = E + i\eta \in \mathbb{C}^+ : |E| < 3, 0 < \eta \leq 3\}.$$

Como nota cultural, la notación  $z = E + i\eta \in \mathbb{C}^+$  proviene de un contexto en física nuclear en la que  $E$  representa la energía y  $\eta$  la resolución en la medición [ESY09].

### 4.2.2. UNIVERSALIDAD Y SESGO

Hay una línea de investigación activa que tiene precedentes en la segunda mitad del siglo XX y que toma fuerza con los trabajos de Tao y Vu, entre otros, a finales de la década de los 2000s. A grandes rasgos la línea intenta generalizar teoremas límite a las clases más grandes posibles de matrices aleatorias, se le conoce como la *universalidad*. Como referencia introductoria al tema se puede citar a [HL16].

*The Wigner-Dyson-Gaudin-Mehta conjecture, or ‘bulk universality’ conjecture, states that the local statistics of the eigenvalues of random matrix ensembles should be universal in the sense that they depend only on the symmetry class of the random matrix ensemble but are otherwise independent of the law of the matrix entries.*

Cabe aclarar que por clase de simetría se refiere a si son matrices rectangulares, simétricas, complejas, reales, Hermitianas, etc.

Un ejemplo concreto de resultado universal es el de las matrices de Wigner definidas en la sección 2.3. El teorema demostrado en el Capítulo 3 se puede extender del modelo GUE al modelo Wigner imponiendo restricciones muy débiles. En el trabajo más reciente a la fecha [GNT19] se tiene lo siguiente.

**Teorema.** Si  $\{X_N\}_{N=1}^\infty$  con  $X_N \in \mathbb{H}_N$  son matrices de Wigner tales que

$$\sup_{i,j,N} \mathbb{E} \left[ \left| (X_N)_{ij} \right|^{4+\varepsilon} \right] < \infty,$$

su distribución empírica espectral satisface la ley global

$$\lim_{N \rightarrow \infty}^{\mathbb{P}} \mu_{X_N}(\mathbf{d}x) = \mu_{SC}(\mathbf{d}x), \quad \forall x \in \mathbb{R}$$

y su refinamiento, la ley local al interior del soporte,

$$\lim_{\substack{N \rightarrow \infty \\ \eta \rightarrow 0^+ \\ N\eta \rightarrow \infty}}^{\mathbb{P}} \left| \frac{1}{\pi} \operatorname{Im} m_{X_N}(E + i\eta) \mathbf{d}E - \mu_{SC}(\mathbf{d}E) \right| = 0, \quad \forall E \in (-2, 2).$$

Existen, en cambio, modelos de matrices  $\{Y_N\}_N$  que son matrices de Wigner y que por tanto satisfacen ambas leyes, pero su estructura hacen que exista una familia  $\{\nu_N\}$  determinista de medidas tales que

$$\nu_N \xrightarrow[N \rightarrow \infty]{d} \mu_{SC}.$$

Cuando lo que se busca es aproximar la distribución empírica espectral, estas medidas deterministas suelen ser mejores, es decir

$$\left| \frac{1}{\pi} \operatorname{Im} m_{Y_N}(E + i\eta) \mathbf{d}E - \nu_N(\mathbf{d}E) \right| \ll \left| \frac{1}{\pi} \operatorname{Im} m_{Y_N}(E + i\eta) \mathbf{d}E - \mu_{SC}(\mathbf{d}E) \right|.$$

A la diferencias del tipo  $\|\mu_{SC} - \nu_N\|_\infty$  ó  $|\mu_{SC}(\mathbf{d}E) - \nu_N(\mathbf{d}E)|$  se les conoce como el sesgo (*shift*) de la clase respecto al teorema universal. El sesgo se define únicamente en términos locales [LS18; HL16], pues cuando  $\eta \geq 1$ , el error producido por la continuidad de Stieltjes es más relevante (veáse la sección 2.1).

Las matrices aleatorias ralas suelen ser clases sesgadas en este sentido [HL16; LS18], como se muestra en las siguientes secciones.

### 4.2.3. LEYES LOCALES UNIVERSAL Y SESGADA

En la sección 4.1.4 se presentaron algunos modelos ralos. Para los siguientes teoremas se usan generalizaciones del modelo de Erdős-Renyi mediante dilución, como se menciona en la sección 4.1.2.

**Teorema 4.2.1** (Ley Local del Semicírculo, [Erd+13b]). *Sea  $\{H^{(N)} \in \mathbb{S}_N \mid N \in \mathbb{N}\}$  proceso matricial que satisface primera y segunda condiciones de rareza con entradas idénticamente distribuidas e independientes salvo simetría.*

*Si  $q_N$  es el parámetro de rareza de  $H^{(N)}$ , entonces para todo  $\varepsilon > 0$  pequeño y  $D > 0$  grande*

$$\sup_{z \in \mathcal{E}} \mathbb{P} \left[ |m_{H^{(N)}}(z) - m_{SC}(z)| > N^\varepsilon \left( \min \left( \frac{1}{q_N^2 \sqrt{|E-2| + \eta}}, \frac{1}{q_N} \right) + \frac{1}{N\eta} \right) \right] \leq N^{-D}. \quad (4.6)$$

En otras palabras, las matrices ralas simétricas satisfacen la Ley Local del Semicírculo en un sentido muy similar al presentado en el capítulo 3. Según la conjetura de universalidad, esto es debido a que son matrices de Wigner [HL16].

Recordar de la sección 3.3 que  $m_{SC}$  es raíz del polinomio

$$P_0(m, z) = 1 + zm + m^2.$$

Para un ensamble ralo se puede plantear el siguiente polinomio

**Definición 4.2.2.** Sea  $\{H^{(N)} \in \mathbb{S}_N \mid N \in \mathbb{N}\}$  proceso matricial que satisface primera y segunda condiciones de rareza con entradas idénticamente distribuidas e independientes salvo simetría. Supóngase que  $\mathbb{E}[H_{1N}^{(N)}] = 0$ ,  $\mathbb{E}\left[|H_{1N}^{(N)}|^2\right] = N^{-1}$  y que  $q_N$  es su parámetro de rareza.

Se define el **cuarto cumulante clásico** de las entradas como

$$\kappa_N^{(4)} := \mathbb{E}\left[|H_{1N}^{(N)}|^4\right] - 3N^{-2}.$$

Se define el polinomio  $P_1$  en función al polinomio  $P_0$  como

$$P_1(m, z, \kappa^{(4)}) := P_0(m, z) + N\kappa^{(4)}m^4. \quad (4.7)$$

La definición del polinomio  $P_1$  incorpora el sesgo que las matrices ralas tienen respecto a la ley local del semicírculo.

**Teorema 4.2.2** (Ley Local para Matrices Ralas, [LS18]). *Sea  $H^{(N)}$  como en la definición 4.2.2.*

*Para cada  $N \in \mathbb{N}$  existe  $L_N \in (2, 3)$  y  $\nu_N$  medida de probabilidad absolutamente continua con soporte  $[-L_N, L_N]$  y transformada de Stieltjes  $m_{\nu_N}$  tal que*

$$P_1(m_{\nu_N}(z), z, \kappa_N^{(4)}) = 0, \quad \forall z \in \mathbb{C}^+. \quad (4.8)$$

*Con respecto a esta medida se cumple que para todo  $\varepsilon > 0$  pequeño y  $D > 0$  grande*

$$\sup_{z \in \mathcal{E}} \mathbb{P}\left[|m_{H^{(N)}}(z) - m_{\nu_N}(z)| > N^\varepsilon \left(\frac{1}{q_N^2} + \frac{1}{N\eta}\right)\right] \leq N^{-D}. \quad (4.9)$$

La sección ?? estudia consecuencias cualitativas de imponer este tipo de sesgos.

#### 4.2.4. LEYES LOCALES PARA MATRICES RALAS MULTIBLOQUE

Un resultado reciente publicado en el *pre-print* [HLY19] es sobre el *Stochastic Block Model* presentado en la sección 4.1.4. En esta ocasión el sesgo será construido como una suma pesada de los “cuartos cumulantes”.

En esta sección  $N = K \cdot r$ , para algún  $r \in \mathbb{N}$  y los bloques son todos del mismo tamaño. Es decir, si para cada índice su bloque está etiquetado por  $s(i)$ , entonces

$$|s^{-1}(\{1, 2, \dots, N\})| = \frac{N}{K} = r.$$

**Definición 4.2.3.** Sean  $K$  y  $s$  como se describió antes y sea  $\{H^{(N)} \in \mathbb{S}_N \mid N \in \mathbb{N}\}$  proceso matricial que satisface primera y segunda condiciones de rareza con entradas independientes salvo simetría y perfil de nulidad Stochastic Block Model de tamaño  $K$  y parámetro de rareza  $q_N$ .

Supóngase que para todo  $1 \leq i, j \leq N$  las entradas están normalizadas

$$\mathbb{E}[H_{ij}^{(N)}] = 0, \quad \mathbb{E}\left[|H_{ij}^{(N)}|^2\right] = N^{-1}.$$

Se definen los **cumulantes clásicos** de las entradas como

$$\begin{aligned} \kappa^{(4)s} &:= \mathbb{E}\left[|H_{ij}^{(N)}|^4\right] - 3N^{-2}, \quad \text{si } s(i) = s(j) \\ \kappa^{(4)d} &:= \mathbb{E}\left[|H_{ij}^{(N)}|^4\right] - 3N^{-2}, \quad \text{si } s(i) \neq s(j) \end{aligned}$$

y la siguiente cantidad auxiliar

$$\xi_N := \frac{\kappa^{(4)_s} - (K - 1)\kappa^{(4)_d}}{K}.$$

Se define el polinomio  $P_2$  en función al polinomio  $P_0$  como

$$P_2(m, z, \xi) := P_0(m, z) + N\xi m^4. \tag{4.10}$$

Esta definición es usada en un artículo reciente donde se probó la siguiente ley local sesgada.

**Teorema 4.2.3** (Ley Local para Matrices a Bloques Ralas, [HLY19]). *Sea  $H^{(N)}$  como en la definición 4.2.3.*

*Para cada  $N \in \mathbb{N}$  existe  $L_N \in [2, 3)$  y  $\nu_N$  medida de probabilidad absolutamente continua con soporte  $[-L_N, L_N]$  y transformada de Stieltjes  $m_{\nu_N}$  tal que*

$$P_2(m_{\nu_N}(z), z, \xi_N) = 0, \forall z \in \mathbb{C}^+. \tag{4.11}$$

*Con respecto a esta medida se cumple que para todo  $\varepsilon > 0$  pequeño y  $D > 0$  grande*

$$\sup_{z \in \mathcal{E}} \mathbb{P} \left[ |m_{H^{(N)}}(z) - m_{\nu_N}(z)| > N^\varepsilon \left( \frac{1}{q_N^2} + \frac{1}{N\eta} \right) \right] \leq N^{-D}. \tag{4.12}$$

Para finalizar la comparación entre leyes universales y sesgadas, nótese que las leyes [sesgadas] dadas por las ecuaciones (4.7) y (4.10) presentan cierta simetría: ambas son raíces de la del polinomio  $P_0$ , la ecuación [universal] auto-estable del semicírculo, corregido el polinomio por un término en cuarto grado.

### 4.3. ANÁLISIS DE UNA LEY LOCAL PARA MATRICES RALAS RECTANGULARES

En la mayor parte de la tesis se ha trabajado con matrices cuadradas simétricas. Como recordatorio, esta elección se toma al querer estudiar el *espectro* de una matriz  $A$ , *i.e.* aquellos valores  $\lambda \in \mathbb{R}$  tales que  $(A - \lambda I)$  no sea invertible. Para ello,  $A$  tiene que pertenecer a un espacio con estructura de anillo.

Las matrices rectangulares, comunes en aplicaciones (veáse [Cau16]), no tienen una estructura de este tipo. Sin embargo, a toda matriz rectangular  $X$  se le puede asociar una matriz cuadrada  $S = X^T X$  con propiedades algebraicas relacionadas a las de  $X$ .

El interés de esta sección es estudiar matrices aleatorias ralas  $X$  a partir del espectro de  $S$ . Se parte de un resultado llamado Ley de Marchenko-Pastur [Tao12; AGZ10; Ban15], el cuál es universal [Adl13] para matrices rectangulares. Esta ley se refina, en el caso de matrices aleatorias ralas, con una ley sesgada [HLS+19] basada en una corrección del cuarto cumulante. Tras presentar la ley sesgada, la tesis estudia sus propiedades asintóticas y efectúa un estudio empírico con simulaciones.

Se espera que esta aportación, tanto cualitativa como empírica, complemente los resultados en la literatura y ayude a generar intuición sobre los alcances de leyes locales sesgadas.

### 4.3.1. LA LEY DE MARCHENKO-PASTUR

A una matriz rectangular  $X \in \text{Mat}_{M \times N}(\cdot)$ , se le asocia la matriz de **covarianza muestral**  $X^T X$ . Ésta es simétrica si  $X$  tiene entradas reales y  $X^* X$  es Hermitiana cuando las entradas son complejas.

Cuando  $X_N$  es un proceso matricial que depende de  $N$ , se suele estudiar [Tao12; ER05; KY17] el proceso empírico de su matriz de covarianza muestral. Para presenta la ley que satisface este proceso, se requiere de la siguiente definición.

**Definición 4.3.1** (Ecuación Funcional de Marchenko-Pastur). Para  $z, m \in \mathbb{C}$  y  $d \in [1, \infty)$  se define

$$P_3(m, z, d) := 1 + \left( \left( 1 - \frac{1}{d} \right) + z \right) m + zm^2. \quad (4.13)$$

La ley universal del proceso empírico es análogo a la ley del semicírculo como se probó en la primera parte del Capítulo 3.

**Teorema 4.3.1** (Teorema de Marchenko-Pastur). *Sea  $d \geq 1$  fijo y para todo  $N$  se supone  $M = dN$ . Supóngase un proceso matricial  $X_N \in \text{Mat}_{M \times N}(\mathbb{C})$  con entradas independientes e idénticamente distribuidas con decaimiento sub-exponencial.*

*Existe una medida de probabilidad  $\mu_{MP}$  absolutamente continua que depende únicamente de  $d$  de tal forma que para todo  $z \in \mathbb{C}^+$*

$$m_{X_N^* X_N}(z) \xrightarrow[N \rightarrow \infty]{\mathbb{P}} m_{\mu_{MP}}(z), \quad (4.14)$$

$$P_3(m_{\mu_{MP}}(z), z, d) = 0. \quad (4.15)$$

Este teorema también se satisface de forma local [Adl13] y se cree [HL16] que es universal. En las siguientes secciones se analizará un sesgo conocido para matrices rectangulares ralas.

### 4.3.2. LEY LOCAL SESGADA

De manera similar a cómo se presentó el sesgo en la sección 4.2, en esta sección el sesgo en la ley se presenta como una corrección al polinomio (4.13).

Para construir el término que corrige al polinomio se requiere definir lo siguiente.

**Definición 4.3.2.** Para  $d \geq 1$  sea  $\{X^{(N)} \in \text{Mat}_{dN \times N} \mid N \in \mathbb{N}\}$  proceso matricial que satisface primera y segunda condiciones de rareza con entradas idénticamente distribuidas e independientes salvo simetría, con parámetro de rareza  $q_N$ . Supóngase que sus entradas están normalizadas.

$$\mathbb{E} \left[ H_{1N}^{(N)} \right] = 0 \quad \text{y} \quad \mathbb{E} \left[ \left| H_{1N}^{(N)} \right|^2 \right] = N^{-1}.$$

Se define el **cuarto cumulante clásico** de las entradas como

$$\kappa_N^{(4)} := \mathbb{E} \left[ \left| H_{1N}^{(N)} \right|^4 \right] - 3N^{-2}.$$

Se define el polinomio  $P_4$  en función al polinomio  $P_3$  como

$$P_4(m, z, d, \kappa) := P_3(m, z, d) + N\kappa \left( \left( 1 - \frac{1}{d} \right) + zm \right)^2 m^2. \quad (4.16)$$

El teorema de la ley sesgada es el siguiente.

**Teorema 4.3.2** (Ley Local para Matrices de Covarianza Ralas, [HLS+19]). *Sea  $X^{(N)}$  como en la definición 4.3.2.*

*Para cada  $N \in \mathbb{N}$  existen  $0 < L_N^- < 1 < L_N^+ < \infty$  y  $\nu_N$  medida de probabilidad absolutamente continua, salvo por un átomo y con soporte  $\{0\} \cup [L_N^-, L_N^+]$ . La transformada de Stieltjes  $m_{\nu_N}$  de  $\nu_N$  satisface*

$$P_4(m_{\nu_N}(z), z, d, \kappa_N^{(4)}) = 0, \forall z \in \mathbb{C}^+. \quad (4.17)$$

*Con respecto a esta medida se cumple que para todo  $\varepsilon > 0$  pequeño y  $D > 0$  grande*

$$\sup_{z \in \mathcal{E}} \mathbb{P} \left[ |m_{X^{(N)}}(z) - m_{\nu_N}(z)| > N^\varepsilon \left( \frac{1}{q_N^2} + \frac{1}{N\eta} \right) \right] \leq N^{-D}. \quad (4.18)$$

Este teorema afirma que, al trabajar con matrices ralas, se puede trabajar con  $P_4$  además de con  $P_3$ . En la práctica su intención, se intuye, es preferir  $P_4$  sobre  $P_3$ . Esta tesis analiza cualitativamente el comportamiento del polinomio  $P_4$  y verifica que las simulaciones satisfacen esta ley.

### 4.3.3. PROPIEDADES ASINTÓTICAS DE LA LEY

La diferencia entre  $P_3$  y  $P_4$  es de un solo término de cuarto grado. Uno de los intereses de la tesis fue la de revisar el impacto de dicho término en las densidades de probabilidad que se derivan de ellas.

**Definición 4.3.3.** Como notación de esta tesis, cada que se hable de la **raíz** de un polinomio  $p(s, z; \Theta)$ , se refiere a la función  $s(z)$  tal que

$$p(s(z), z; \Theta) = 0, \forall z \in \mathbb{C}^+$$

y que  $\lim_{\eta \rightarrow \infty} \eta \mathbf{i} \cdot s(\eta \mathbf{i}) = -1$ , si es que esta existe.

Consistente a la transformada inversa de Stieltjes de la sección 2.1, si

$$\sigma_s(E) := \lim_{\eta \rightarrow 0^+} \frac{1}{\pi} \text{Im } s(E + \eta \mathbf{i}),$$

es densidad de probabilidad, entonces se le llama **la ley de  $p$  o la densidad subyacente de  $s$** .

Por ejemplo, en la figura 4.3.1, se muestran las leyes dadas por  $P_3$  y  $P_4$  para los mismos valores. Nótese que  $P_4$  depende de  $\kappa$ , el cuarto cumulante [clásico] de las entradas.

En esta tesis se presenta la siguiente proposición que acota de manera muy permisiva el sesgo entre estas leyes.

**Proposición 4.3.3.** *Sea  $\sigma_N(E)$  la ley de  $P_4$  y  $\sigma(E)$  la ley de Marchenko-Pastur (la raíz de  $P_3$ ). Para  $N$  grande y  $E \in \mathbb{R}$ , si  $\varepsilon > 0$ , entonces*

$$|\sigma_N(E) - \sigma(E)| = O(N^{2-\varepsilon} \kappa E^2).$$

*Demostración.* Sea  $1 \geq \eta > N^{-\frac{1-\varepsilon}{4}}$  y  $z = E + \mathbf{i}\eta$ . Por [HLS+19], se sabe que  $P_4$  tiene una raíz  $m_N(\cdot)$  cuya densidad subyacente es continua y de soporte compacto. Si  $m(\cdot)$  es la raíz de  $P_3$ , entonces se puede usar la transformada inversa de Stieltjes.

$$|\sigma_N(E) - \sigma(E)| \leq C |m_N(E + \mathbf{i}\eta) - m(E + \mathbf{i}\eta)|.$$

Como  $m$  es raíz de  $P_3$  y el polinomio  $P_3$  genera una ecuación funcional auto-estable [Tao12], se tiene que

$$|m_N(z) - m(z)| \leq C |P_3(m_N(z), z, d)|.$$

Según la ecuación (4.16)

$$P_3(m, z, d) - P_4(m, z, d, \kappa) = -N\kappa(1 - d^{-1} + zm)^2 m^2$$

y al ser  $P_4(m_N(z), z, d, \kappa) = 0$ , se tiene que

$$P_3(m_N(z), z, d) = O(N\kappa|z|^2 |m_N(z)|^4).$$

Por la sección 2.4, se tiene que

$$\begin{aligned} |m_N(z)|^4 &\leq \eta^{-4} \\ &\leq \left(N^{-\frac{1-\varepsilon}{4}}\right)^{-4} \\ &= N^{1-\varepsilon}. \end{aligned}$$

Como además

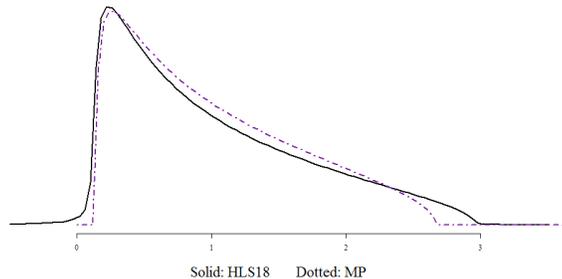
$$|z|^2 = E^2 + \eta^2 \leq E^2 + 1,$$

se pueden combinar estas cotas demostrando la proposición

$$P_3(m_N(z), z, d) = O(N\kappa E \cdot N^{1-\varepsilon}).$$

■

La primera consecuencia observable es que el sesgo es sub-lineal en  $E$ . Esto puede observarse inmediatamente en la figura 4.3.1 donde para valores chicos de  $E$  el sesgo no es evidente, pero se incrementa con valores altos de  $E$ . En pocas palabras,  $\sigma_N$  es más pesada hacia la derecha.



**Figura 4.3.1:** Comparación de la ley de  $P_3$  (sólido) contra la ley de  $P_4$  (punteado). Evaluación para  $N = 512$ ,  $\kappa$  corresponde a  $\text{Ber}(N^{-0.1})$  y  $d = M/N = 2.5$

La otra relación a considerar es  $N^2\kappa$ . Si  $\kappa = o(N^2)$ , entonces se puede asegurar que  $\sigma_N \rightarrow \sigma$ . De hecho, es en este punto donde se puede observar cómo la rareza afecta en el sesgo de la ley.

Por ejemplo, si las entradas fueran Bernoulli de parámetro  $p_N$  como los ejemplos de la sección 4.1.4, luego

$$\kappa = O\left(\frac{1}{N^2 p_N}\right),$$

por lo tanto

$$\begin{aligned} |\sigma_N(E) - \sigma(E)| &= O\left(\frac{N^{2-\varepsilon}}{N^2 p_N} E^2\right) \\ &= O\left(\frac{E^2}{N^\varepsilon p_N}\right). \end{aligned}$$

Nótese que esta es una cota superior, pero aún así hay una intuición de que el sesgo estará siempre presente y que es una función racional de  $N$ . Además se observa la importancia de que  $p_N \rightarrow 0$ , cuando  $N \rightarrow \infty$  ocurra de forma polinomial en  $N$ .

#### 4.3.4. ANÁLISIS DESDE UNA PERSPECTIVA NUMÉRICA

Otro aspecto de estudio a la ley sesgada para matrices aleatorias rectangulares ralas es el comparar la ley con realizaciones simuladas. Para ello se implementó un código en  $R$  que resolviera numéricamente el polinomio  $P_4$ .

Antes de presentar los resultados se hablará de las consideraciones técnicas al momento de implementar la solución.

##### IMPLEMENTACIÓN NUMÉRICA PARA RESOLVER LA LEY

Existen métodos numéricos para resolver polinomios complejos de grado 4. En situaciones usuales, estos métodos retornan cuatro posibles raíces y es tarea del usuario elegir una de entre ellas. Sin embargo, la ecuación (4.16) es una ecuación funcional en el que la solución es una *función analítica* y no sólo un número. Esto significa que la elección de la raíz tiene que hacerse de manera consistente para un conjunto grande, por ejemplo

$$\tilde{E}_0 = \{z = E + i\eta_0 : E \in [5, 5]\}.$$

Si bien, los métodos retornan un arreglo con cuatro valores, los primeros experimentos dejaron patente que la elección de la raíz tiene que ser un proceso más complejo. Para ilustrar esta situación, se muestran en la figura 4.3.2 las cuatro raíces para una selección de puntos en  $E_0$ .

Para valores grandes de  $\eta$  se pueden observar ciertos patrones, pero al disminuir la escala dichos patrones ya no son tan útiles. Cuéntese también que para valores de  $\eta$  en valores  $10^{-3}$ , como hay términos del tipo  $\eta^4$ , puede haber inestabilidad numérica.

Sin embargo, se encontró una solución. Por ser  $m_N(z)$  transformada de Stieltjes, para  $\eta$  suficientemente grande, exactamente una raíz va a cumplir

$$|z \cdot m_N(z) + 1| < 10^{-3}. \tag{4.19}$$

Esta consistencia se puede observar empíricamente en la tercera gráfica de la figura 4.3.2.

El método propuesto es plantear una rejilla de valores en  $C^+$ , con saltos exponenciales en sus valores imaginarios. El polinomio se resuelve para todos los valores en la rejilla, pero la elección de la raíz se hace en aquellos más alejados de la recta real.

De manera iterativa, para cada valor en la rejilla se selecciona una raíz en base a la raíz que se eligió en aquél valor justo arriba de él.

En la figura 4.3.3 puede apreciarse este método en acción para una sola columna de la rejilla (*i.e.* un valor  $E$  fijo, pero  $\eta$  variable).

En el apéndice puede consultarse el código para este algoritmo de tipo *greedy*. El respaldo teórico para este algoritmo parece residir en la relación entre la distancia combinatoria con la distancia euclidiana de las raíces [Tao12].

## RESULTADOS DE SIMULACIÓN

Ahora, dada una realización de una matriz aleatoria rala  $X_N$ , se construye el proceso empírico de su covarianza empírica.

**Definición 4.3.4.** Sea  $X_N$  una matriz aleatoria rala debidamente normalizada (veáse sección 4.1). Se define  $S_N := X_N^T X_N$  y  $s_N(z) := \text{Tr}((S_N - zI)^{-1})$ .

Sea  $\eta > 0$ . Se define la **densidad empírica de  $S_N$  con resolución  $\eta$** , como

$$\hat{\sigma}_{N,\eta}(E) := \frac{1}{\pi} \text{Im } s_N(E + i\eta).$$

De forma análoga, la **densidad subyacente de  $P_4$  con resolución  $\eta$**  es

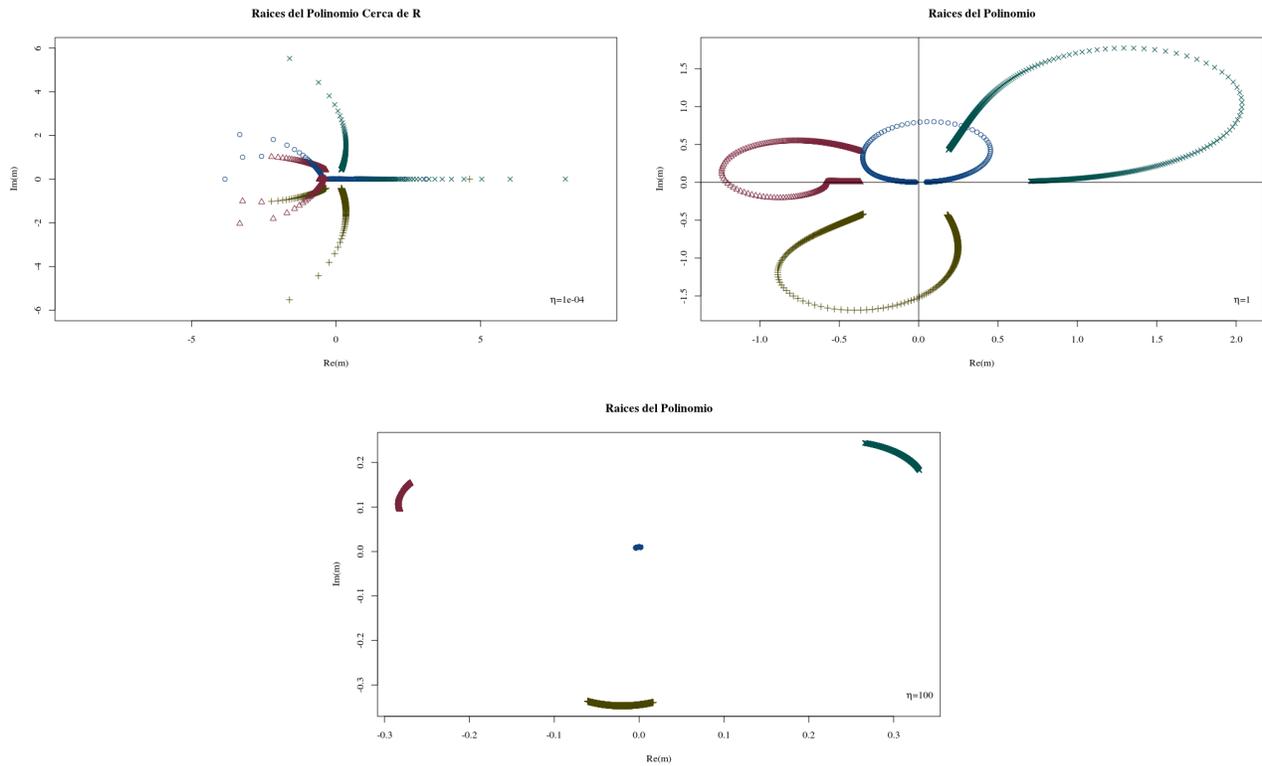
$$\sigma_{N,\eta}(E) := \frac{1}{\pi} \text{Im } \sigma(E + i\eta).$$

Durante la tesis se simularon matrices  $X_N$  ralas con entradas Bernoulli como se definen en la sección 4.1.4. El objetivo es comparar  $\hat{\sigma}_{N,\eta}(E)$  con  $\sigma_{N,\eta}(E)$  y ver si alguna de estas densidades es útil para explicar el histograma de la matriz de covarianza empírica.

En la figura 4.3.4 se observan el histograma y ambas curvas.

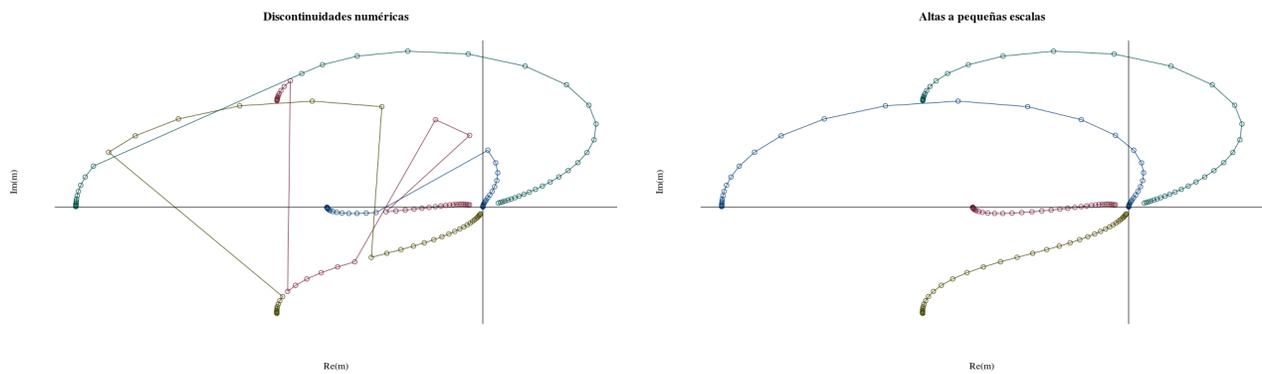
La observación principal es que con  $\eta$  pequeña, la curva  $\hat{\sigma}_{N,\eta}$  es inestable y se suaviza al aumentar el tamaño de  $\eta$ . Cuando crece  $\eta$  también se acercan las densidades empírica y subyacente. Sin embargo, este aumento hace perder información, *i.e.* ninguna de las dos se compara al histograma. Esta situación ya se había observado antes en la sección 3.6.2.

De manera empírica se puede elegir un valor de  $\eta$  entre  $N^{-.6}$  y  $N^{-.4}$  que combine cercanía entre las curvas y parecido al histograma.



**Figura 4.3.2:** Visualización de las raíces numéricas a diferentes escalas.

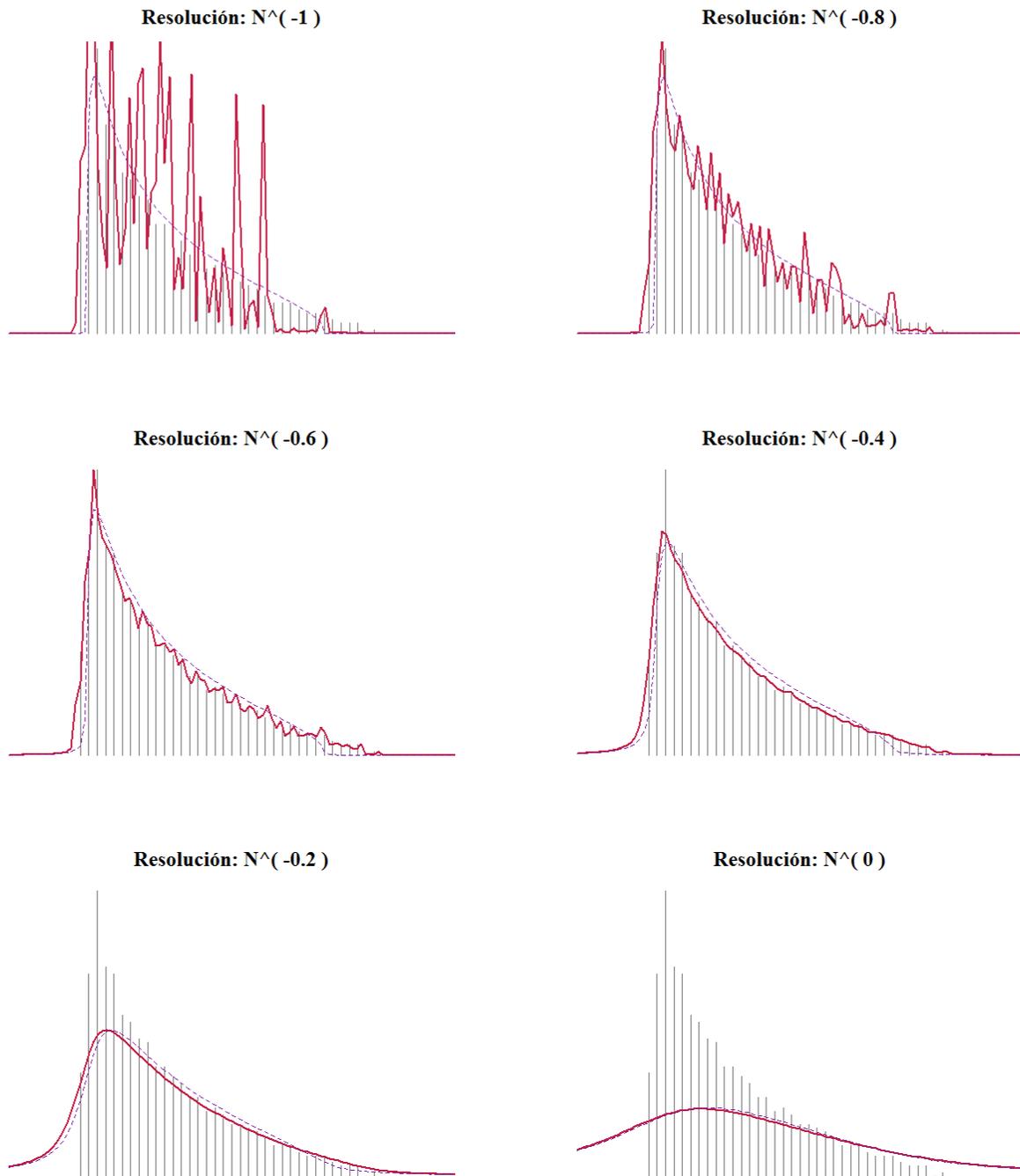
Cada color-forma corresponde al índice con el que la implementación numérica retorna esa raíz.



**Figura 4.3.3:** Curvas de consistencia numérica para un valor fijo en la parte real.

En ambas imágenes: puntos cercanos al origen corresponden a valores grandes de  $\eta$ , los puntos lejanos a los valores pequeños de  $\eta$ .  
 En la izquierda: los puntos conectados representan raíces retornadas con el mismo índice por algún método numérico.

En la derecha: los puntos conectados minimizan la longitud de cada curva.



**Figura 4.3.4:** Histograma de los valores propios de  $S_N$  comparadas con las densidades subyacente y empírica para diferentes resoluciones  $\eta$ .

Todas las imágenes:  $N = 2048$ ,  $d = 2.5$ ,  $p = N^{-0.8} \approx 0.002$ .

En línea sólida roja, densidad empírica de  $S_N$ . En línea punteada, densidad subyacente  $\sigma_{N,\eta}(\cdot)$ .

# APÉNDICE A

## MATERIAL ADICIONAL

### A.1. LEYES LOCALES CON EL MÉTODO DE STEIN

En el Capítulo 3 hay una sección dedicada a demostrar la así llamada Ley Global. Ésta prueba que la densidad del modelo matricial Gaussiano converge a una densidad límite concreta. En general, para un modelo matricial  $\{X_N\}$ , demostrar su ley global significa [KY17] encontrar una distribución  $\mu$  fija y un límite de la forma

$$m_{X_N}(\mathbf{i}\eta) \rightarrow_N^{\mathbb{P}} m_{\mu}(\mathbf{i}\eta), \quad \forall \eta \geq 1.$$

Además de la técnica de grandes desviaciones para modelos gaussianos que se usó en el Capítulo 3, existen varios modelos y diferentes técnicas conocidas para obtener este tipo de límites, de los cuáles el libro [Tao12] cubre un amplio margen.

Una de estas técnicas se presentó por primera vez en [LP+09] y ha sido usado en trabajos que estudian matrices ralas [LS18; HLS+19; HLY19]. Esta técnica se inspira en los trabajos de Charles Stein en los años 1970-80s donde desarrolla un método para probar el Teorema del Límite Central de sin usar funciones características. El método se basa en un resultado elemental que usa el Teorema de Fubini sobre la ecuación diferencial  $\frac{df(x)}{dx} = xf(x)$ , caracterización del kernel normal.

**Teorema A.1.1** (Lema de Stein). *Sean  $Z, X_1, X_2, \dots \in L^3$  donde  $Z \sim \mathcal{N}_{\mathbb{R}}(0, 1)$ . Las siguientes afirmaciones son equivalentes*

$$\forall f \in \mathcal{C}_b^1, f' \in \mathcal{C}_b \quad \mathbb{E}[f'(X_n) - X_n f(X_n)] \rightarrow 0 \quad (\text{A.1})$$

$$\forall h \in \mathcal{C}_b \quad \mathbb{E}[h(X_n) - h(Z)] \rightarrow 0. \quad (\text{A.2})$$

Pueden consultarse ejemplos del uso del teorema en [Tao12] y su demostración en el artículo original [Ste81] es sencilla de seguir pues sólo emplea teoría de la medida.

A continuación se presenta un argumento intuitivo [LS18] de por qué este método funciona para modelos matriciales y una herramienta que se usa en la literatura.

Sea  $\{X_N\}$  el modelo GUE del Capítulo 3,  $G_{X_N} \equiv G_{X_N}(z) := (X_N - zI)^{-1}$  su resolvente y  $m_{X_N} = \text{Tr}(G_{X_N})$  su transformada de Stieltjes (véase Capítulo 2).

**Observación A.1.2.** *El resolvente  $G_{X_N}$  es diferenciable respecto a  $X_N$*

$$\frac{\partial G_{X_N}}{\partial X_N} = -G_{X_N} \otimes G_{X_N},$$

con  $\otimes$  el producto tensorial de Kronecker. En particular

$$\mathbf{Tr}\left(\frac{\partial G_{X_N}}{\partial X_N}\right) = -(\mathbf{Tr}(G_{X_N}))^2 = -N^2 m_{X_N}^2(z).$$

Por otra parte el teorema A.1.1 implica que para  $Z$  normal y  $f$  suave

$$\mathbb{E}[Zf(Z)] = \mathbb{E}\left[\frac{\mathbf{d}f(Z)}{\mathbf{d}Z}\right],$$

lo que inspira en el caso matricial complejo una mímica no necesariamente correcta

$$\mathbb{E}[\mathbf{Tr}(X_N G_{X_N})] = \alpha_N \mathbb{E}\left[\mathbf{Tr}\left(\frac{\partial G_{X_N}}{\partial X_N}\right)\right], \quad (\text{A.3})$$

donde  $\alpha_N \in \mathbb{C}$ .

La diferencial del resolvente ya se tiene en términos de  $m_{X_N}$ , pero  $X_N G_{X_N}$  aún no. Como  $(X_N - zI)G_{X_N} = I$ , se tiene que  $X_N G_{X_N} = I + zG_{X_N}$ . Luego

$$\frac{1}{N} \mathbf{Tr}(X_N G_{X_N}) = 1 + z m_{X_N}(z).$$

Uniendo todos estos resultados y que la traza conmuta con la esperanza se tiene

$$\mathbb{E}[1 + z m_{X_N}(z)] = \mathbb{E}\left[\frac{1}{N} \mathbf{Tr}(X_N G_{X_N})\right] = \frac{\alpha_N}{N} \mathbb{E}\left[\mathbf{Tr}\left(\frac{\partial G_{X_N}}{\partial X_N}\right)\right] = -\frac{\alpha_N}{N} \mathbb{E}[N^2 m_{X_N}^2(z)].$$

Si resultara que  $\alpha_N = \frac{1}{N}$ , entonces

$$\mathbb{E}[1 + z m_{X_N}(z) + m_{X_N}^2(z)] = 0. \quad (\text{A.4})$$

En otras palabras, si el Lema de Stein se cumpliera en el caso matricial, entonces en promedio se estaría cumpliendo la ley del semicírculo. En realidad es más complicado que esto, lo que se hace es probar que

$$\mathbb{E}\left[|1 + z m_{X_N}(z) + m_{X_N}^2(z)|^{2D}\right] \rightarrow 0,$$

para  $D > 0$  usando la siguiente modificación al teorema A.1.1.

**Teorema A.1.3** (Lytova & Pastur,[LP+09]). Sean  $D \in \mathbb{N}$  y  $X \in L^{D+2}$  v.a. compleja con cumulantes clásicos  $\kappa_1, \dots, \kappa_{D+1}$ . Entonces, para cualquier función  $f \in \mathcal{C}_b^{D+1}$  con derivadas  $\frac{\mathbf{d}^l f}{\mathbf{d}x^l}$  acotadas, se tiene que

$$\mathbb{E}[Xf(X)] = \sum_{l=0}^D \frac{\kappa_{(l+1)}}{l!} \mathbb{E}\left[\frac{\mathbf{d}^l f}{\mathbf{d}x^l}(X)\right] + \varepsilon_D, \quad (\text{A.5})$$

donde

$$|\varepsilon_D| \leq \left(\frac{1 + (3 + 2D)^{D+2}}{(D+1)!}\right) \mathbb{E}[|X|^{D+2}] \sup_{t \in \mathbb{R}} \left|\frac{\mathbf{d}^{D+1} f}{\mathbf{d}x^{D+1}}(X)\right|.$$

Para la demostración y uso del teorema para matrices aleatorias se recomienda leer la sección 3 del artículo original [LP+09] y para su uso particular en leyes locales la sección 3.2 de [LS18].

## A.2. CÓDIGO EN R PARA LEY SESGADA

El siguiente código fue implementado y se utilizó para realizar las figuras 4.3.1 y 4.3.4.

```
# -----
# Funciones para Resolver la Ecuación Funcional en HLS18
# Autor: Diego Caudillo - Abril 2019 dcaudillo@cimat.mx
# -----

# Como usarse:
# La mayoría de las funciones son de uso interno.
# La interfaz de usuario es
#   d.HLS18(E,N,M,Kappa,eta){
# donde E es un arreglo de reales, M y N son las dimensiones de la matriz
# Kappa es el cuarto cumulante de las entradas normalizadas
# y eta>0 es la resolución deseada
#   La función regresa la densidad de HLS18 en un arreglo del mismo
# tamaño que E.

# Las 4 raíces del polinomio por lote sin orden alguno
m.4.HLS18 <- function(z,N,M,Kappa){
  d <- N/M
  D <- 1-1/d

  c.1 <-      D
  c.2 <-   N*D*D*Kappa
  c.3 <- 2*N  *D*Kappa
  c.4 <-   N   *Kappa

  r <- sapply( z , function(Z){
    t.r <- polyroot( c(
      1,
      Z + c.1,
      Z + c.2,
      Z * c.3,
      Z*Z * c.4) )
    return( t.r)
  }
)

  return(r)
}

# Algoritmo greedy para asegurar estabilidad numérica.
# Intercambio del orden de las raíces considerando longitud de línea
```

```

continuity <- function(A , B){
  if( length(A) != length(B) ) return(NULL)
  for( j in 1:(length(A)-1) ){
    for(l in 1:length(B) )
    for(k in (j+1):length(B) ){
      if(
        (Mod(B[j]-A[j]) + Mod(B[k]-A[k])) # Status Quo
        >
        (Mod(B[k]-A[j]) + Mod(B[j]-A[k])) # Change
      ){
        a <- B[k]
        B[k] <- B[j]
        B[j] <- a
      }
    }
  }
  return(B)
}

# Escoge la mejor raiz usando continuidad y cercania a (-z) lejos de R
m.HLS18 <- function(z,N,M,Kappa,high=1e2,Step=2){
  r <- sapply( z , function(Z){

    K <- max(ceiling( (log(high) - log(Im(Z)))/log(Step) ),1)
    etas <- exp( seq(log(Im(Z)),log(high),length.out=K) )[K:1]
    ZZ <- complex( real=rep(Re(Z),K) , imaginary=etas )
    m <- m.4.HLS18(ZZ,N,M,Kappa)
    for( i in 2:K ) m[,i] <- continuity( m[,i-1] , m[,i] )
    m.fin <- continuity(m[,K],m.4.HLS18(Z,N,M,Kappa))
    R.K <- m.fin[which.min( Mod(m[,1]*ZZ[1] +1) )]

    return( R.K )
  }
)

return(r)
}

# Densidad Atomo en Cero [Cauchy Escalada]
dirac.cauchy <- function(x,eta) return( (eta/pi)/( x*x + eta*eta ) )

# Curva de densidad para un nivel eta
d.HLS18 <- function(E,N,M,Kappa,eta){
  z <- complex(real = E,imaginary = eta)
  m <- m.HLS18(z,N,M,Kappa)
  dens <- pmax( (N/M)*Im(m)/pi , 0)
  atom <- dirac.cauchy(E,eta)*(1-1/d)*d
}

```

---

```
dens <- pmax(dens - atom,0)    # HLS18 si considera el atomo en cero
return( dens )
}
```

---



# BIBLIOGRAFÍA

- [Adl13] Ben Adlam. “The local Marchenko-Pastur law for sparse covariance matrices”. Bachelor’s thesis. Harvard University, 2013.
- [AGZ10] Greg W Anderson, Alice Guionnet y Ofer Zeitouni. *An introduction to random matrices*. Vol. 118. Cambridge university press, 2010.
- [And+15] Greg W. Anderson y col. “A local limit law for the empirical spectral distribution of the anticommutator of independent Wigner matrices”. En: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 51. 3. Institut Henri Poincaré. 2015, págs. 809-841.
- [Ban15] Afonso S. Bandeira. “18.S096: Concentration inequalities, scalar and matrix versions”. 2015.
- [BHY16] Roland Bauerschmidt, Jiaoyang Huang y Horng-Tzer Yau. “Local Kesten–McKay law for random regular graphs”. En: *arXiv preprint arXiv:1609.09052* (2016).
- [BK16] Florent Benaych-Georges y Antti Knowles. “Lectures on the local semicircle law for Wigner matrices”. En: *arXiv preprint arXiv:1601.04055* (2016).
- [Bla+56] Jerome Blackman y col. “An extension of the Kolmogorov distribution”. En: *The Annals of Mathematical Statistics* 27.2 (1956), págs. 513-520.
- [Blo+14] Alex Bloemendal y col. “Isotropic local laws for sample covariance and generalized Wigner matrices”. En: *Electron. J. Probab* 19.33 (2014), págs. 1-53.
- [Cau16] Diego Caudillo. “Modelos estadísticos para sistemas de recomendación”. Bachelor’s thesis. Universidad de Guanajuato, 2016.
- [Con12] John B Conway. *Functions of one complex variable II*. Vol. 159. Springer Science & Business Media, 2012.
- [DZ18] Ioana Dumitriu y Yizhe Zhu. “Sparse general Wigner-type matrices: local law and eigenvector delocalization”. En: *arXiv preprint arXiv:1808.07611* (2018).
- [DZ19] Ioana Dumitriu y Yizhe Zhu. “Spectra of random regular hypergraphs”. En: *arXiv preprint arXiv:1905.06487* (2019).
- [EKN18] Laszlo Erdos, Torben Kruger y Yuriy Nemish. “Local laws for polynomials of Wigner matrices”. En: *arXiv preprint arXiv:1804.11340* (2018).
- [EKS18] Laszlo Erdos, Torben Kruger y Dominik Schroder. “Cusp universality for random matrices I: local law and the complex Hermitian case”. En: *arXiv preprint arXiv:1809.03971* (2018).
- [EKY13] Laszlo Erdos, Antti Knowles y Horng-Tzer Yau. “Averaging fluctuations in resolvents of random band matrices”. En: *Annales Henri Poincaré*. Vol. 14. 8. Springer. 2013, págs. 1837-1926.

- 
- [ER05] Alan Edelman y N Raj Rao. “Random matrix theory”. En: *Acta Numerica* 14 (2005), págs. 233-297.
- [Erd+13a] Laszlo Erdos y col. “Spectral statistics of Erdos–Rényi graphs I: local semicircle law”. En: *The Annals of Probability* 41.3B (2013), págs. 2279-2375.
- [Erd+13b] Laszlo Erdos y col. “The local semicircle law for a general class of random matrices”. En: *Electronic Journal of Probability* 18 (2013).
- [ESY09] Laszlo Erdos, Benjamin Schlein y Horng-Tzer Yau. “Local semicircle law and complete delocalization for Wigner random matrices”. En: *Communications in Mathematical Physics* 287.2 (abr. de 2009), págs. 641-655. DOI: 10.1007/s00220-008-0636-9.
- [EYY12] Laszlo Erdos, Horng-Tzer Yau y Jun Yin. “Bulk universality for generalized Wigner matrices”. En: *Probability Theory and Related Fields* 154.1-2 (2012), págs. 341-407.
- [GNT15] Friedrich Gotze, Alexey Naumov y Alexander Tikhomirov. “Local semicircle law under moment conditions. Part II: Localization and delocalization”. En: *arXiv preprint arXiv:1511.00862* (2015).
- [GNT19] Friedrich Götze, Alexey Naumov y Alexander Tikhomirov. “Local semicircle law under fourth moment condition”. En: *Journal of Theoretical Probability* (2019), págs. 1-36.
- [GZ+00] Alice Guionnet, Ofer Zeitouni y col. “Concentration of the spectral measure for large matrices”. En: *Electronic Communications in Probability* 5 (2000), págs. 119-136.
- [HL16] Jiaoyang Huang y Benjamin Landon. “Local law and mesoscopic fluctuations of Dyson Brownian motion for general  $\beta$  and potential”. En: *arXiv preprint arXiv:1612.06306* (2016).
- [HLS+19] Jong Yun Hwang, Ji Oon Lee, Kevin Schnelli y col. “Local law and Tracy–Widom limit for sparse sample covariance matrices”. En: *The Annals of Applied Probability* 29.5 (2019), págs. 3006-3036.
- [HLY19] Jong Yun Hwang, Ji Oon Lee y Wooseok Yang. “Local law and Tracy–Widom limit for sparse stochastic block models”. En: *arXiv preprint arXiv:1903.02179* (2019).
- [KY17] Antti Knowles y Jun Yin. “Anisotropic local laws for random matrices”. En: *Probability Theory and Related Fields* 169.1-2 (2017), págs. 257-352.
- [LP+09] Anna Lytova, Leonid Pastur y col. “Central limit theorem for linear eigenvalue statistics of random matrices with independent entries”. En: *The Annals of Probability* 37.5 (2009), págs. 1778-1840.
- [LS13] Ji Oon Lee y Kevin Schnelli. “Local deformed semicircle law and complete delocalization for Wigner matrices with random potential”. En: *Journal of Mathematical Physics* 54.10 (2013), pág. 103504.
- [LS18] Ji Oon Lee y Kevin Schnelli. “Local law and Tracy–Widom limit for sparse random matrices”. En: *Probability Theory and Related Fields* 171.1-2 (2018), págs. 543-616.
- [Mas00] Pascal Massart. “Some applications of concentration inequalities to statistics”. En: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 9. 2. 2000, págs. 245-303.
- [Ste81] Charles M Stein. “Estimation of the mean of a multivariate normal distribution”. En: *The annals of Statistics* (1981), págs. 1135-1151.
-

- 
- [Tao06] Terence Tao. *Nonlinear dispersive equations: local and global analysis*. 106. American Mathematical Soc., 2006.
- [Tao12] Terence Tao. *Topics in random matrix theory*. Vol. 132. American Mathematical Soc., 2012.
- [Wig55] Eugene P. Wigner. “Characteristic vectors of bordered matrices with infinite dimensions”. En: *Annals of Mathematics* 62.3 (1955), págs. 548-564. ISSN: 0003486X. URL: <http://www.jstor.org/stable/1970079>.
- [WZ19] Jun Wen y Wang Zhou. “Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions”. En: *arXiv preprint arXiv:1901.05166* (2019).
-