



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

LA NORMA DE OPERADOR DE MATRICES ALEATORIAS
Y LA LEY DEL SEMICÍRCULO DE WIGNER

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

P R E S E N T A :

PABLO DE LA FUENTE PARRES QUINTERO

DIRECTORA DE TESIS:

DRA. LAURA CLEMENTINA ESLAVA FERNÁNDEZ

CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO, 2025



Yo no estudio para escribir, ni menos para
enseñar (que fuera en mí desmedida soberbia),
sino sólo por ver si con estudiar ignoro menos.
Así lo respondo y así lo siento.

– Sor Juana Inés de la Cruz,
Respuesta a sor Filotea de la Cruz

Agradecimientos

Esta investigación fue realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM con los proyectos IN102824 “El tiempo hacia el antepasado común más reciente” y IN108525 “Destrucción y resiliencia de árboles aleatorios”. Agradezco a la DGAPA-UNAM la beca recibida.

Aprovecho este espacio también para extender mis agradecimientos a todas las personas que me apoyaron no solo en la realización de este trabajo, sino también durante los últimos años.

A Laura Eslava por aceptar ser la tutora de esta tesis; por el apoyo, la confianza, la paciencia y el cariño; por enseñarme tanto sobre matemáticas y sobre tantas otras cosas quizás más importantes.

He tenido la fortuna de contar con un jurado excelente. Agradezco a los sinodales: Octavio Arizmendi, Ana Meda, Loiret Dosal y Sergio López, por el tiempo que dedicaron a la lectura de este trabajo, así como por sus comentarios y correcciones.

A Víctor Pérez Abreu por invitarme a cenar panuchos en Kanasín, Yucatán (la capital del panucho) y apoyarme con bibliografía, revisiones, consejos y grandes anécdotas.

A mis profesoras y profesores a lo largo de estos años. Una vez más, a Ana Meda, por adentrarme en el mundo de la probabilidad y los procesos estocásticos. De nuevo a Jana, por compartir conmigo las gráficas y por el apoyo en tiempos difíciles como la pandemia. A Gabriela Campero, Sandra Palau, Javier Fernández. A Begoña Fernán-

dez por darme la oportunidad de ser su ayudante. A Saraí Hernández Torres por las horas del café, entre otras cosas.

A José Antonio de la Peña y a Jesús Jiménez porque quizás fue en las *Matertulias* que surgió mi gusto inexplicable por las matrices; por enseñarme que un buen matemático encuentra problemas interesantes hasta en las carreteras.

A Terence Tao por siempre responder mis dudas en entradas de su blog que publicó hace 10 años.

A Darío Alatorre por aceptarme en el Laboratorio de Visualización y Sonorización del IMATE. A Jaime Lobato por invitarme al proyecto del Réquiem de Mozart en el festival El Aleph.

A mis papás y a mi hermana, y sobre todo al gato Luis, por apoyarme siempre.

A Giana, por asombrarme todos los días, por creer en mí, por compartir conmigo tantas cosas buenas. *I'm yours in insistence.*

A mis amigas y amigos de la facultad. En especial a Andrés, Diego, Emi Álvarez, porque han sido amigos y maestros. También a mis amigas y amigos de otros contextos: Charlie, Emi, Pali, Ulises, Alí. Gracias por hacer la vida más llevadera.

Índice general

1. Preliminares	9
1.1. Variables aleatorias subgaussianas	9
1.2. Árboles y números de Catalan	13
2. Suma de variables aleatorias	24
2.1. Desigualdad de Hoeffding	25
2.2. Concentración vía sus momentos	29
3. Norma de operador de matrices aleatorias	37
3.1. Técnica de la ε -red	39
3.2. Digresión: método de simetrización	48
4. Ley del semicírculo de Wigner	52
4.1. Momentos de la ley del semicírculo	53
4.2. Momentos de la distribución espectral empírica	55
4.3. Convergencia débil de la DEE	70
A. Método de truncamiento	80

Prefacio

Aunque es difícil señalar un momento puntual en el que surge la teoría de matrices aleatorias, es inevitable mencionar a Adolf Hurwitz, John Wishart y Eugene Wigner. Hurwitz introdujo y desarrolló la noción de medida invariante sobre grupos de matrices en 1897 (véase [DF17]). Por otra parte Wishart, en 1928, incorporó las matrices aleatorias a la estadística al analizar estimadores de matrices de covarianza para muestras independientes con distribución normal [Wis28]. Sin embargo, quizás el resultado más celebrado e influyente en la teoría de matrices aleatorias sea la ley de semicírculo, que obtuvo Wigner en 1955 inspirado en el estudio de sistemas cuánticos tan complicados que su comportamiento es aparentemente aleatorio; como los núcleos de átomos muy pesados, por ejemplo.

En este trabajo, primero estudiamos la norma de operador de matrices aleatorias y después demostramos la ley del semicírculo de Wigner.

Definición (Norma de operador de una matriz). *Dada una matriz M de dimensión $n \times n$ con entradas reales, no necesariamente aleatoria, definimos su norma de operador como*

$$||M||_{op} = \sup_{x \in S_n} |Mx|$$

donde S_n es la esfera unitaria en \mathbb{R}^n , es decir, $S_n = \{x \in \mathbb{R}^n : |x| = 1\}$.

Como la norma de operador de una matriz aleatoria es una variable aleatoria real, muchas técnicas y resultados de variables aleatorias reales se pueden extender para estudiarla y así, es un buen primer acercamiento a la teoría de matrices aleatorias.

La norma de operador no solo es una estadística interesante por sí misma, sino que también acota otras cantidades relevantes como los valores propios de la matriz. En la primera parte del trabajo demostramos que la norma de operador de una matriz de $n \times n$ cuyas entradas son variables aleatorias independientes se concentra en un intervalo de longitud $O(\sqrt{n})$. Este resultado se sigue cumpliendo si las matrices además son simétricas. Así, podemos concluir que los valores propios también están contenidos en ese intervalo con alta probabilidad, lo que motiva la normalización por $1/\sqrt{n}$ al estudiar el comportamiento asintótico de su espectro.

Para estudiar rigurosamente el espectro de una matriz aleatoria simétrica, introducimos el concepto de distribución espectral empírica (DEE), que nos permite estudiar a todos los valores propios simultáneamente pues le damos el mismo peso a cada uno.

Definición. (*Distribución espectral empírica*) Sea M una matriz simétrica de dimensión $n \times n$. Supongamos que sus valores propios, necesariamente reales, son $\lambda_1(M) \leq \lambda_2(M) \leq \dots \leq \lambda_n(M)$. Entonces definimos la distribución espectral empírica de M como la medida de probabilidad

$$\mu_M = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(M)}.$$

Notemos que en el caso en que las entradas de M son variables aleatorias, su DEE es una medida aleatoria.

Wigner obtuvo resultados sobre el comportamiento asintótico del espectro de matrices aleatorias pues estaba inspirado en problemas de mecánica cuántica. En esta área la energía se representa con un operador hamiltoniano y sus valores propios corresponden a los posibles valores de energía en los que se puede encontrar el sistema al realizar una medición. Al tratarse de un sistema cuántico dicho conjunto es discreto y una manera de analizarlo es observando que las propiedades estadísticas de los niveles de energía coinciden con las de los valores propios de matrices aleatorias de gran dimensión. En su trabajo pionero [Wig55], demostró que la DEE de matrices simétricas con entradas

Rademacher¹ independientes converge a la ley del semicírculo cuando la dimensión de la matriz crece a infinito. Es decir, converge a la ley definida por la densidad

$$\sigma(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbb{1}_{[-2,2]}(x),$$

un semicírculo de radio 2 centrado en el origen y adecuadamente normalizado para ser una función de densidad de probabilidad.

Por otra parte, con algunas simulaciones podemos intuir cómo se ve la DEE de matrices simétricas cuyas entradas tienen otras distribuciones. Las tres figuras siguientes muestran histogramas (normalizados) de los espectros de matrices aleatorias simétricas de dimensión $10,000 \times 10,000$. Cada matriz está normalizada por $1/\sqrt{10,000}$, justificable por la concentración de la norma de operador. Así, la única diferencia entre las tres simulaciones es la distribución de sus entradas. Notemos que las tres distribuciones tienen media 0 y varianza 1.

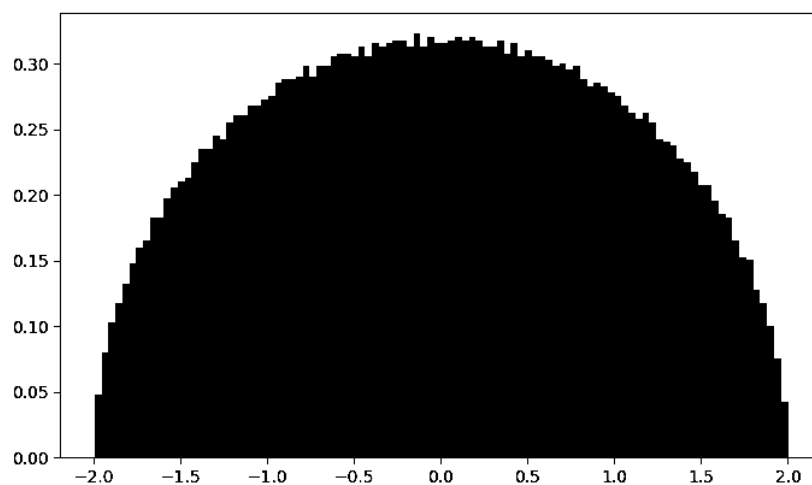


Figura 1: Histograma de la DEE de una matriz simétrica normalizada cuyas entradas tienen distribución normal estándar.

¹Una variable aleatoria X tiene distribución Rademacher si $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = 1/2$.

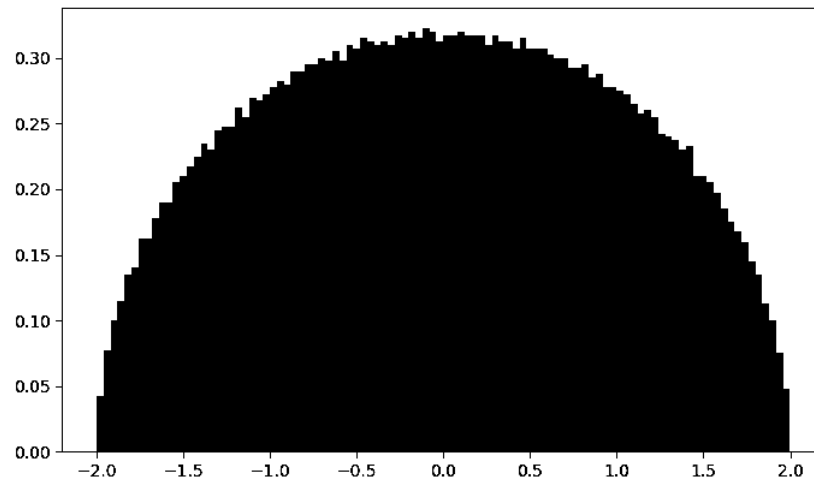


Figura 2: Histograma del espectro de una matriz simétrica normalizada cuyas entradas tienen distribución Rademacher.

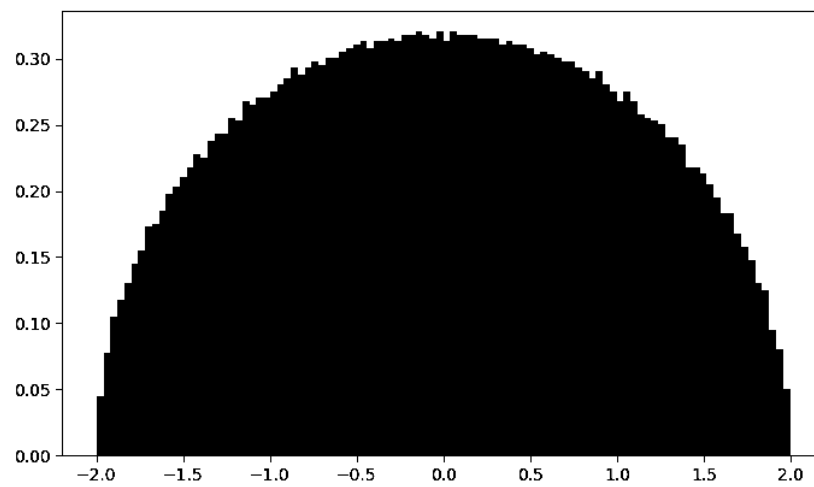


Figura 3: Histograma del espectro de una matriz simétrica normalizada cuyas entradas tienen distribución uniforme en el intervalo $(-\sqrt{3}, \sqrt{3})$.

En la segunda parte de este trabajo aprovechamos aspectos combinatorios de los momentos de la DEE para demostrar que, en efecto, la convergencia de la DEE a la ley del semicírculo ocurre independientemente de la distribución de las entradas de la matriz. Este fenómeno se conoce como universalidad y es uno de los principios fundamentales de la teoría de matrices aleatorias. En resumen, lo que demostramos es lo siguiente.

Teorema (Ley del semicírculo de Wigner). *Sea $\{M_n\}_{n \in \mathbb{N}}$ una sucesión de matrices aleatorias simétricas cuyas entradas son independientes, tienen media 0, varianza 1 y cada M_n es de dimensión $n \times n$. Entonces la sucesión de las DEE de las matrices normalizadas, es decir $\{\frac{1}{\sqrt{n}}M_n\}_{n \in \mathbb{N}}$, casi seguramente converge débilmente (en el sentido de medidas) a una medida de probabilidad determinista en \mathbb{R} : la ley del semicírculo.*

Me parece de la mayor importancia también mencionar que la teoría de matrices aleatorias ha tenido numerosos desarrollos y muy variadas aplicaciones en los últimos años. Un ejemplo curioso puede consultarse en [KS00]. En este artículo, Krbálek y Seba analizan los tiempos de llegada de camiones de transporte público en Cuernavaca. Esta ciudad, como la mayoría en México, tiene la peculiaridad de que cada conductor elige su ruta basándose en la información que tiene sobre la posición de los demás camiones. Aunque podríamos imaginar que el resultado es caótico, en realidad los camiones suelen estar espaciados de tal forma que resulta óptima. Es decir, los usuarios nunca pasan ni muy poco ni demasiado tiempo esperando a ser recogidos. Los autores observan que los camiones parecen repelerse para maximizar el número de pasajeros que llevan en cada viaje. Un efecto similar de repulsión también se puede observar en los valores propios de algunas matrices aleatorias, hecho que había demostrado ya Wigner desde sus primeros trabajos en la década de los cincuenta. A propósito de este artículo, Freeman Dyson menciona en [ABDF15]:

“Cuando un experto en mercados [financieros] me asegura que algún tipo de artilugio financiero beneficiará a la humanidad, me inclino a pensar que un conductor de autobús de Cuernavaca podría hacerlo mejor.”

Las matrices aleatorias han sido una herramienta y un objeto de estudio en áreas como análisis, estadística, física, geometría algebraica, probabilidad libre, finanzas, ciencias de la computación, biología molecular, comunicación inalámbrica y, de forma sorprendente, hasta en teoría de números (véase, por ejemplo, [\[ABDF15\]](#)). Uno debe tomar en cuenta, por supuesto, que la lista anterior no es exhaustiva y está en constante crecimiento.

Estructura de la tesis

Este trabajo está organizado en cuatro capítulos y un apéndice:

- En el primer capítulo presentamos algunos conceptos preliminares necesarios para los argumentos de secciones posteriores. Comenzamos con las variables subgaussianas, que surgen al definir qué tan concentrada está una variable aleatoria alrededor de su media. Después seguimos con los árboles y los números de Catalan, que aparecen naturalmente cuando estudiamos la distribución espectral empírica de matrices de Wigner y su comportamiento asintótico.
- En el segundo capítulo estudiamos el fenómeno de concentración que presentan las sumas de variables aleatorias bajo algunas condiciones de independencia. Para esto, aprovechamos las propiedades combinatorias de los momentos de la suma. Esta técnica es útil, por analogía, al estudiar los momentos de la DEE de matrices de Wigner.
- En el tercer capítulo desarrollamos algunos aspectos de la norma de operador de matrices aleatorias con entradas independientes. Exploramos el método de la ε -red, que es usado en general para acotar probabilidades de supremos de variables aleatorias indexadas por un conjunto infinito. Este método nos permite aproximar la cantidad que nos interesa con solo un conjunto finito de índices. Resulta esencial estudiar la concentración de la norma de operador también porque acota al espectro de la matriz. Así, podemos justificar la normalización por $1/\sqrt{n}$ en la ley del semicírculo. En este capítulo también analizamos brevemente el método de simetrización para ejemplificar las técnicas que se emplean

al estudiar la norma de operador con mayor profundidad. Este capítulo está basado principalmente en [Tao12].

- En el cuarto capítulo demostramos la convergencia de la DEE a la ley del semicírculo usando el método de los momentos. Hacemos énfasis en las propiedades combinatorias de los momentos de la DEE, donde surgen los números de Catalan. Finalmente, usando el hecho de que la densidad del semicírculo tiene soporte compacto, justificamos por qué analizar la convergencia de los momentos garantiza convergencia débil en el sentido de medidas. Aunque la idea general de la demostración es la misma que usó Wigner en [Wig55], este capítulo está basado principalmente en el artículo *The Semicircle Law and Beyond: The Shape of Spectra of Wigner Matrices* de Ioana Dumitriu que está recopilado en [BCG22]. Tratamientos similares pueden también encontrarse en [Bor19, DMRA11, GAH22].
- En el apéndice presentamos brevemente el método de truncamiento y, como ejemplo, demostramos la ley débil de los grandes números cuando las variables no tienen varianza finita. Este método permite extender resultados de variables acotadas, como los de los capítulos 2 y 3, a variables no acotadas, como los del capítulo 4. Así, podemos ver que es posible extender resultados sobre la concentración de la norma de operador a ensambles más generales de matrices aleatorias.

Capítulo 1

Preliminares

En este capítulo presentamos dos conceptos y resultados fundamentales tanto del área de probabilidad como de combinatoria.

Comenzamos estudiando variables aleatorias subgaussianas, que nos brindan un marco de referencia para describir variables significativamente concentradas alrededor de su media. Nos basamos principalmente en [Ver18]. Utilizamos estas herramientas en el capítulo 3 donde exploramos la concentración de la norma de operador de matrices aleatorias.

En la segunda sección abordamos una clase combinatoria que está asociada con los números de Catalan: los árboles planos con un número dado de vértices. Sorprendentemente los momentos impares de una variable con distribución semicircular también están dados por esta sucesión y es por esta razón que este resultado es relevante para el desarrollo del capítulo 4.

1.1. Variables aleatorias subgaussianas

Para estudiar qué tan concentrada está una variable aleatoria alrededor de su media nos interesa conocer sus colas. Es decir, dada una variable aleatoria X y un real

positivo λ , podemos analizar las colas de X , es decir, la cantidad

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda)$$

y en particular, qué tan rápido decae esa probabilidad a 0 cuando λ crece.

Como punto de partida, nos interesa entender las colas de variables gaussianas, ya que su comportamiento sirve como referencia para definir una clase más amplia de variables que exhiben concentración considerable alrededor de sus medias: las variables subgaussianas.

Definición 1.1. *Una variable aleatoria X tiene distribución gaussiana con media μ y varianza σ^2 si su función de densidad es*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x - \mu)^2/2\sigma^2\right).$$

Analicémoslas con detalle en la siguiente proposición, cuyo argumento principal es la desigualdad de Chernoff. Es decir, que para toda $t > 0$ se cumple que

$$\mathbb{P}(X > \lambda) = \mathbb{P}(e^{tX} > e^{t\lambda}) \leq e^{-t\lambda} \mathbb{E}(e^{tX}).$$

La última desigualdad se obtiene aplicando la desigualdad de Markov.

Proposición 1.2 (Colas de una variable aleatoria gaussiana). *Si X es una variable aleatoria gaussiana con media 0 y varianza σ^2 entonces para toda $\lambda > 0$*

$$\mathbb{P}(|X| > \lambda) \leq 2 e^{-\lambda^2/2\sigma^2}.$$

Demostración. Sea $\lambda > 0$. Como la función de densidad de X es simétrica alrededor de 0 entonces basta acotar la cola superior y la inferior es análoga. Es decir,

$$\mathbb{P}(|X| > \lambda) = 2 \mathbb{P}(X > \lambda).$$

Por la desigualdad de Chernoff sabemos que para toda $t > 0$ se cumple que

$$\mathbb{P}(X > \lambda) \leq e^{-t\lambda} \mathbb{E}(e^{tX}).$$

La esperanza de e^{tX} , que es la función generadora de momentos de X , se puede calcular explícitamente y obtener que $\mathbb{E}(e^{tX}) = e^{\sigma^2 t^2 / 2}$. Por lo tanto,

$$\mathbb{P}(X > \lambda) \leq e^{-t\lambda} e^{\sigma^2 t^2 / 2}.$$

Así, para encontrar la mejor cota basta encontrar la t que minimice al exponente. Derivando obtenemos que

$$\frac{d}{dt}(-t\lambda + \sigma^2 t^2 / 2) = -\lambda + \sigma^2 t$$

e igualando a 0 concluimos que la función alcanza un mínimo cuando $t = \lambda / \sigma^2$. Finalmente, evaluando en el mínimo tenemos que

$$\mathbb{P}(|X| > \lambda) = 2 \mathbb{P}(X > \lambda) \leq 2 e^{-\lambda^2 / 2\sigma^2}.$$

□

Como un primer acercamiento para estudiar resultados de concentración en variables aleatorias no necesariamente gaussianas podemos enfocarnos en aquellas que, en cierto sentido, se concentran más alrededor de su media que una gaussiana. El cálculo previo de las colas gaussianas nos motiva a definir una clase de variables cuyas colas sean controlables por una cota análoga, incluso si tienen otra distribución.

Definición 1.3 (Variable aleatoria subgaussiana). *Una variable aleatoria X es subgaussiana con factor de varianza C^2 si para toda $\lambda > 0$ se cumple que*

$$\mathbb{P}(|X| > \lambda) \leq 2 \exp(-\lambda^2 / C^2).$$

En este trabajo, no nos preocupamos por el factor de varianza óptimo, ya que nos

interesa principalmente el orden de decaimiento de las colas, el cual basta para deducir resultados de concentración. Sin embargo, hay contextos en que podría ser útil conocerlo porque el mínimo factor de varianza define una norma en el espacio de variables subgaussianas.

Uno también puede inspirarse en otras propiedades de las variables gaussianas para definir las variables subgaussianas. Es estándar utilizar los momentos de X o la función generadora de momentos de X^2 . Sin embargo, se puede concluir que estas definiciones resultan ser equivalentes salvo constantes. Un análisis detallado de estas equivalencias puede encontrarse en la Sección 2.5 de [Ver18]. Para los objetivos de este trabajo bastan ciertas implicaciones de la definición anterior, que demostramos en seguida.

Lema 1.4 (Momentos pares de una variable aleatoria subgaussiana). *Si X es una variable aleatoria subgaussiana con factor de varianza C^2 entonces para todo $k \in \mathbb{N}$ se cumple que*

$$\mathbb{E}(X^{2k}) \leq 2C^{2k}k!.$$

Demostración. Notemos que si tenemos cotas para las colas de X entonces podemos obtener cotas para las colas de X^2 de la siguiente manera:

$$\mathbb{P}(X^2 > \lambda) = \mathbb{P}(|X| > \sqrt{\lambda}) \leq 2 \exp(-\lambda/C^2).$$

Con esto ya podemos acotar sus momentos. Recordando que podemos calcular la esperanza de una variable no negativa integrando sus colas obtenemos que para toda $k \in \mathbb{N}$ se cumple que

$$\mathbb{E}(X^{2k}) = \int_0^\infty \mathbb{P}(X^{2k} > \lambda) d\lambda = \int_0^\infty \mathbb{P}(X^2 > \lambda^{1/k}) d\lambda \leq \int_0^\infty 2 \exp(-\lambda^{1/k}/C^2) d\lambda.$$

Ahora, haciendo el cambio de variable $u = \lambda^{1/k}/C^2$ tenemos que

$$2 \int_0^\infty \exp(-\lambda^{1/k}/C^2) d\lambda = 2 \int_0^\infty C^{2k} k u^{k-1} e^{-u} du = 2C^{2k} k \int_0^\infty u^{k-1} e^{-u} du.$$

Por definición, la última integral es $\Gamma(k)$ y recordando que $k\Gamma(k) = k!$ finalmente

tenemos que

$$\mathbb{E}(X^{2k}) \leq 2C^{2k}k!. \quad \square$$

Veamos ahora que haciendo uso del lema anterior podemos concluir que si X es subgaussiana entonces la función generadora de momentos de X^2 está acotada en una vecindad del 0.

Proposición 1.5 (Cota de la F.G.M. de X^2). *Si X es una variable aleatoria subgaussiana con factor de varianza C^2 entonces para todo t con $|t| \leq 1/2C^2$ se cumple que*

$$\mathbb{E}(e^{tX^2}) \leq 4.$$

Demostración. Por la expansión en serie de Taylor de la función generadora de momentos de X^2 y por el teorema de convergencia dominada sabemos que

$$\mathbb{E}(e^{tX^2}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{t^k X^{2k}}{k!}\right) = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}(X^{2k})}{k!}.$$

Ahora, aplicando la cota que obtuvimos en el lema anterior para el $2k$ -ésimo momento obtenemos que

$$\mathbb{E}(e^{tX^2}) \leq 2 \sum_{i=0}^{\infty} (tC^2)^k.$$

Como por hipótesis $|tC^2| \leq 1/2$ entonces esa serie geométrica converge a lo más a 2 y podemos concluir que

$$\mathbb{E}(e^{tX^2}) \leq 4. \quad \square$$

1.2. Árboles y números de Catalan

En esta sección analizamos cuántos árboles planos con k vértices hay. Contar árboles planos es un problema interesante en sí mismo y, además, la sucesión que los cuenta coincide con los momentos de la distribución semicircular, lo cual es clave en el capí-

tulo 4. La demostración que presentamos es esencialmente una adaptación del lema del ciclo que usan Dvoretzky y Motzkin en [DM47].

Para empezar, damos una definición recursiva de los objetos que queremos contar. Para un ejemplo de un árbol plano véase la figura 1.1.

Definición 1.6 (Árbol plano). *Decimos que una gráfica T es un árbol plano si existe un vértice $v \in T$, que llamamos raíz, y una sucesión finita, totalmente ordenada y posiblemente vacía $(T_1, \dots, T_{k(v)})$ tales que*

1. *cada T_i es un árbol plano y $v \notin T_i$,*
2. *si $i \neq j$ entonces T_i y T_j no comparten vértices ni aristas,*
3. *v es adyacente a la raíz de cada T_i y a ningún otro vértice.*

Cuando hablamos del tamaño de un árbol plano nos referimos a la cantidad de aristas que tiene. Equivalentemente, el tamaño de un árbol plano es igual a la cantidad de sus vértices menos uno.

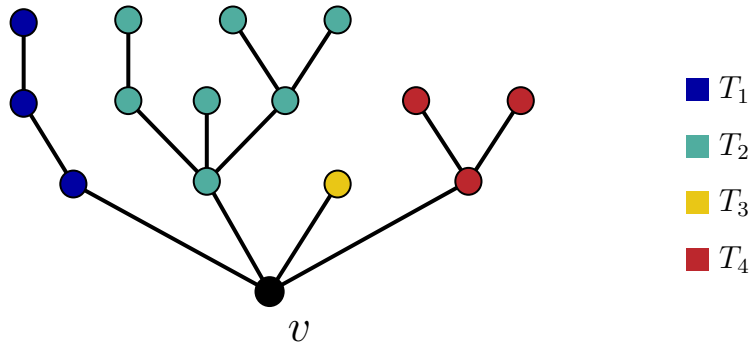


Figura 1.1: Un árbol plano con raíz v y sucesión de subárboles (T_1, T_2, T_3, T_4) .

Optamos por siempre representar a los árboles con la raíz en el vértice inferior, y los demás desplegados hacia arriba¹. Además, está implícita una ordenación de los vértices de izquierda a derecha. Ver figuras 1.1 y 1.2.

¹Conviene aclarar que no hay una orientación universalmente aceptada. Aunque en la naturaleza los árboles crecen de abajo hacia arriba, en ciencias de la computación suelen dibujarse de arriba hacia abajo. Quizás es justificable porque en occidente leemos de arriba hacia abajo; pensemos también en los árboles genealógicos.

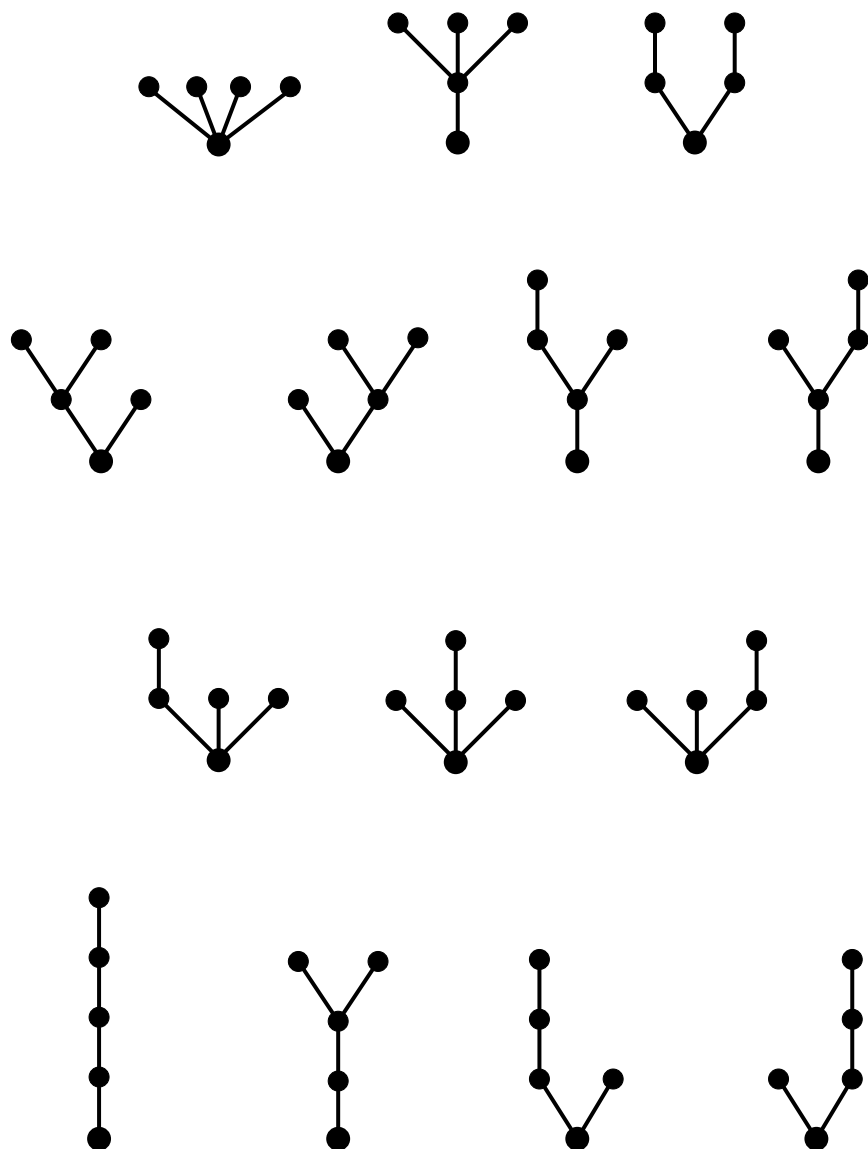


Figura 1.2: los 14 árboles planos que existen de tamaño 4. Notemos, por ejemplo, que aunque parecen similares, los últimos dos árboles son diferentes por el ordenamiento de sus vértices.

Para contar cuántos árboles planos de tamaño k hay podemos dar una biyección a un conjunto que sea más sencillo de analizar.

Si recorremos el contorno del árbol y en cada paso anotamos a qué distancia de la raíz nos encontramos, obtenemos un camino de $(0, 0)$ a $(0, 2k)$ que nunca es negativo, pues no hay vértices a distancia negativa de la raíz. El camino resultante tiene $2k$ pasos porque cada arista se recorre exactamente dos veces. Ver figura 1.3.

Los caminos que resultan de recorrer el contorno de un árbol plano se llaman caminos de Dyck, en honor al matemático alemán Walther von Dyck.

Definición 1.7 (Camino de Dyck). *Un camino de Dyck es una sucesión de longitud par, digamos $w = (w_1, w_2, \dots, w_{2k})$, que cumple que k de sus entradas son $+1$, k son -1 y además sus sumas parciales nunca son negativas, es decir, para todo $l \leq 2k$*

$$w_1 + w_2 + \dots + w_l \geq 0.$$

Aunque los definimos como sucesiones, es usual y además muy útil considerar la representación de caminos de Dyck como caminos en \mathbb{Z}^2 . Si en la sucesión hay un $+1$ entonces el camino en \mathbb{Z}^2 sube 1 unidad y si hay un -1 el camino en \mathbb{Z}^2 baja una unidad. Además, para dar una representación gráfica más intuitiva se dibujan las líneas diagonales entre cada par de pasos. De la definición se sigue que el camino va de $(0, 0)$ a $(2k, 0)$ y nunca baja del eje horizontal.

Así, con la biyección entre árboles planos y caminos de Dyck dada por recorrer su contorno, podemos concluir que contar árboles planos con k vértices es equivalente a contar caminos de Dyck con $2k$ pasos. Para resolver esto, es más sencillo volver a dar una biyección a otro conjunto. Sin embargo, esta biyección es más sencilla y es útil esencialmente por razones técnicas que se exponen en los lemas al final de la sección.

Definición 1.8 (Camino de Dyck extendido). *Un camino de Dyck extendido es una sucesión de longitud impar, digamos $w = (w_1, w_2, \dots, w_{2k+1})$, que cumple que su primera entrada $w_1 = +1$ y la sucesión $(w_2, w_3, \dots, w_{2k+1})$ es un camino de Dyck.*

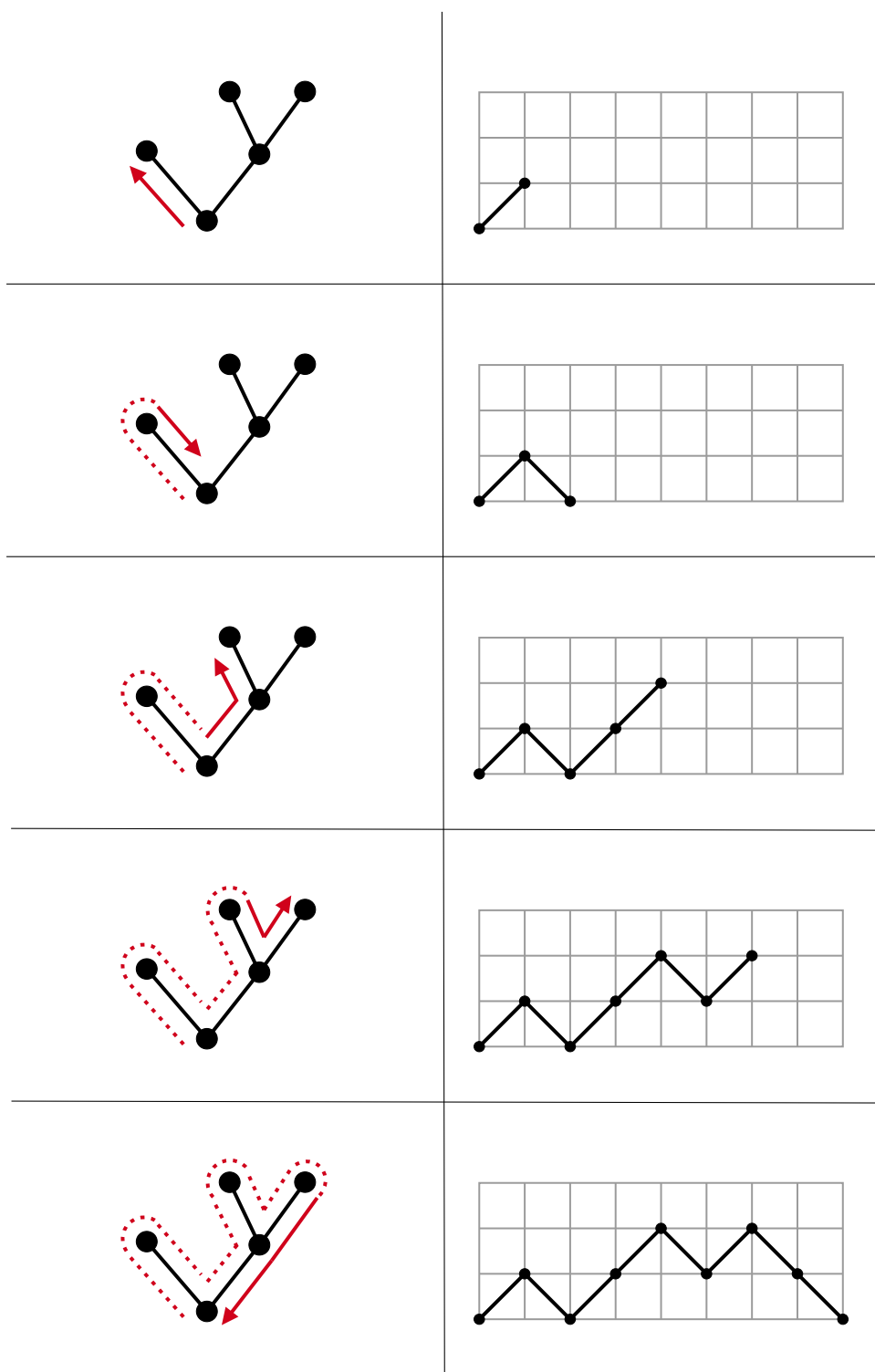


Figura 1.3: Un camino de Dyck construido por recorrer el contorno de un árbol plano. Recordemos que estamos asumiendo que las etiquetas de los vértices van de izquierda a derecha.

Notemos que hay una biyección natural entre caminos de Dyck y caminos de Dyck extendidos. Solo hay que añadir o quitar el primer paso (que siempre es un paso hacia arriba). Como un camino de Dyck nunca es negativo entonces un camino de Dyck extendido es estrictamente positivo a partir del primer paso. Ver figura 1.4.

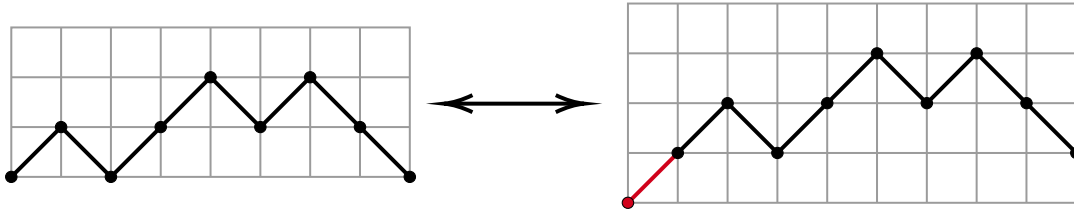


Figura 1.4: Biyección entre caminos de Dyck y caminos de Dyck extendidos

La clave ahora es dar una relación de equivalencia entre caminos que empiezan subiendo y ver que en cada clase de equivalencia solo hay uno que es estrictamente positivo después del primer paso. No es necesario considerar caminos que empiezan bajando pues estos no son caminos de Dyck extendidos.

Definición 1.9 (Permutación cíclica de un camino). *Un camino $(w_1, w_2, \dots, w_{2k+1})$ es una permutación cíclica de otro camino $(v_1, v_2, \dots, v_{2k+1})$ si existe un índice i tal que $(w_1, w_2, \dots, w_{2k+1}) = (v_i, v_{i+1}, \dots, v_{2k+1}, v_1, \dots, v_{i-1})$.*

Consideremos la relación entre caminos que empiezan subiendo en donde dos caminos están relacionados si existe una permutación cíclica que lleve uno a otro. Es esencial notar que esta es una relación de equivalencia.

Finalmente, tenemos las herramientas necesarias para demostrar cuántos árboles planos hay². En la demostración hacemos referencia a lemas cuyas demostraciones presentamos al final de la sección. Una visualización del caso $k = 3$ puede verse en la figura 1.6.

²A la sucesión $\{C_k\}_{k \in \mathbb{N}}$ que cuenta el número de árboles planos se le conoce como los números de Catalan en honor al matemático belga y francés Eugène-Charles Catalan que trabajó un problema de triangulación de polígonos en 1838. Sin embargo, esta sucesión ya había sido mencionada antes por Euler y se conocía ya en China desde aproximadamente 1730 (véase [Lar00]).

Teorema 1.10 (Número de árboles planos). *Hay $C_k = \frac{1}{k+1} \binom{2k}{k}$ árboles planos de tamaño k o, equivalentemente, con $k+1$ vértices.*

Demostración. Contar árboles planos de tamaño k es equivalente a contar caminos de Dyck extendidos con $2k+1$ pasos. Es decir, caminos de $(0,0)$ a $(2k+1,1)$ que empiezan subiendo y son estrictamente positivos a partir del primer paso. Consideramos a dos caminos que empiezan subiendo como equivalentes si existe una permutación cíclica que transforme uno en el otro. Esta relación es una relación de equivalencia y, por lo tanto, genera una partición en el espacio de caminos de $(0,0)$ a $(2k+1,1)$ que empiezan subiendo.

Ahora, cada clase de equivalencia tiene $k+1$ elementos pues por cada camino hay $k+1$ permutaciones cíclicas que empiezan subiendo y, por el lema 1.11, todas son distintas. Por el lema del ciclo (lema 1.12) sabemos que en cada clase hay un único camino que es un camino de Dyck extendido. Es decir, hay tantos caminos de Dyck extendidos como clases de equivalencia.

Finalmente, notemos que la cantidad de caminos de $(0,0)$ a $(2k+1,1)$, sin importar si son 0 o negativos en algún punto, que empiezan subiendo es $\binom{2k}{k}$. Esto ocurre porque después del primer paso el camino tiene que subir k veces y bajar k veces. Así, es suficiente escoger en cuáles de esos $2k$ pasos el camino sube. Los pasos que no se escogen son en los que el camino baja.

Podemos concluir entonces que hay $\frac{1}{k+1} \binom{2k}{k}$ caminos extendidos de Dyck de longitud $2k$ y, por lo tanto, árboles planos de tamaño k . \square

En seguida presentamos las demostraciones de los lemas que usamos para demostrar el teorema 1.10. Una de las razones de usar caminos de Dyck extendidos es que en el caso de caminos de Dyck el primer lema no se cumple.

Lema 1.11 (Toda permutación cíclica produce un camino diferente). *Sea w un camino de $(0,0)$ a $(2k+1,1)$. Si $1 < i < 2k+1$ entonces la permutación cíclica de w por i es distinta a w .*

Demostración. Sea $w = (w_1, w_2, \dots, w_{2k+1})$ un camino de $(0, 0)$ a $(2k + 1, 1)$, es decir, un camino donde $k + 1$ entradas son $+1$ y k son -1 . Supongamos, buscando una contradicción, que existe una permutación cíclica no trivial que resulta en el mismo camino, es decir, un índice i con $0 < i < 2k + 1$ tal que

$$(w_1, \dots, w_{2k+1}) = (w_i, w_{i+1}, \dots, w_{2k+1}, w_1, \dots, w_{i-1}).$$

Así, podemos asegurar que existe un mínimo entero d tal que el camino se puede expresar como d copias concatenadas de un camino de tamaño $\frac{2k+1}{d}$. Es decir,

$$(w_1, \dots, w_{2k+1}) = \underbrace{(w_1, \dots, w_{\frac{2k+1}{d}})}_{\text{copia 1}}, \underbrace{(w_{\frac{2k+1}{d}+1}, \dots, w_{2k+1})}_{\text{copia 2}}, \dots, \underbrace{(w_1, \dots, w_{\frac{2k+1}{d}})}_{\text{copia } d}.$$

Sea r al valor del camino acumulado después de $\frac{2k+1}{d}$ pasos, es decir, $r = \sum_{i=1}^{\frac{2k+1}{d}} w_i$. Como el camino está formado por d copias de un camino de longitud $\frac{2k+1}{d}$ entonces la suma total de sus pasos es dr y recordando que por hipótesis el valor final del camino es 1 tenemos

$$1 = \sum_{i=1}^{2k+1} w_i = d \sum_{i=1}^{\frac{2k+1}{d}} w_i = dr.$$

Finalmente, como d y r son enteros y $dr = 1$ podemos concluir que $d = r = 1$. Esto contradice la suposición de que la permutación cíclica no es trivial porque tener $d = 1$ implica que $i = 0$ o $i = 2k + 1$. \square

Con el segundo lema podemos afirmar que en cada clase de equivalencia hay un único camino que es un camino de Dyck extendido. Generalizaciones de este resultado son ampliamente usadas en combinatoria. Un tratamiento extensivo puede verse en [DZ90]. Quizás la primera vez que se usó fue por Dvoretzky y Motzkin en [DM47] para investigar el teorema de la votación de Bertrand.

Lema 1.12 (Lema del ciclo). *Para todo camino de $(0, 0)$ a $(2k + 1, 1)$ que empieza subiendo existe una única permutación cíclica que es un camino extendido de Dyck.*

Demostración. Sea $w = (w_1, w_2, \dots, w_{2k+1})$ un camino de $(0, 0)$ a $(2k + 1, 1)$ que empieza subiendo, es decir, $w_1 = +1$. La idea esencial es tomar la permutación cíclica que inicie en el punto mínimo para que el camino resultante sea estrictamente positivo desde el primer paso.

Sea m el valor mínimo que alcanza w y sea i el último paso en donde alcanza dicho mínimo. En el segmento $(w_i, w_{i+1}, \dots, w_{2k+1})$, el camino asciende $m + 1$ unidades, pues parte de m en el paso i y termina en 1 en el paso $2k + 1$. Además, como i es la última vez que el camino alcanza m , para todo $j > i$ la altura en el paso j es estrictamente mayor que m .

En resumen, el camino permutado empieza en 0 y vale $m + 1$ en el paso $2k + 1 - i$.

Ahora, al analizar los primeros i pasos de w podemos asegurar que el camino baja a lo más hasta m unidades porque m es el mínimo. Por lo tanto, en los últimos i pasos del camino permutado, este nunca vale 0 ni es negativo.

Así, la permutación cíclica $(w_i, w_{i+1}, \dots, w_{2k+1}, w_1, \dots, w_{i-1})$ es un camino de Dyck extendido. Ver figura 1.5.

La unicidad de la permutación cíclica se deduce del hecho de que si se empieza en cualquier otro paso entonces el valor del camino permutado es 0 o es negativo en el punto mínimo del camino original. \square

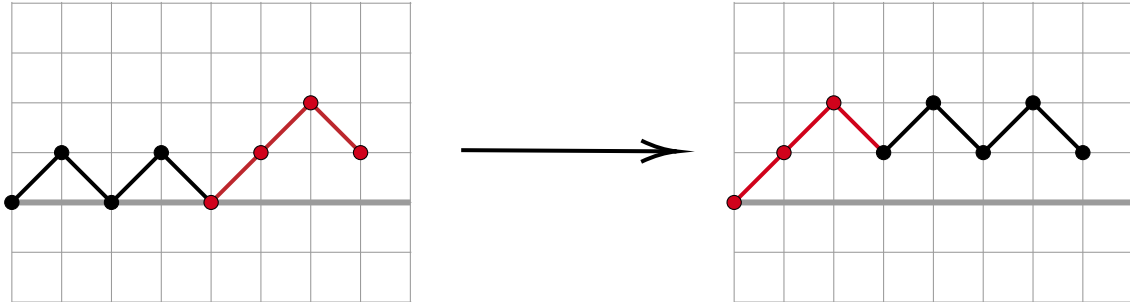
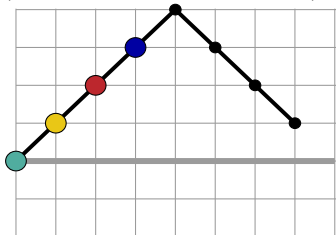
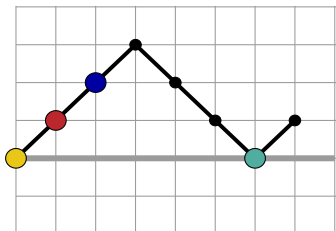


Figura 1.5: Permutación cíclica que resulta en un camino de Dyck extendido. El mínimo del camino, 0, se alcanza por última vez en el paso 4.

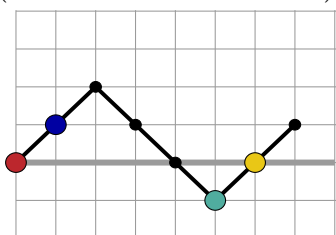
(+ + + + - - -)



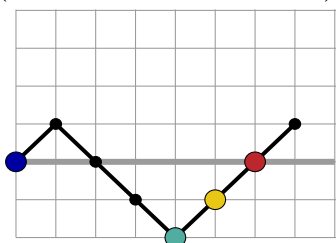
(+ + + + - - - +)



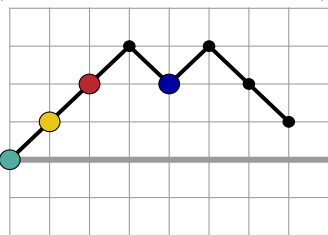
(+ + - - - + +)



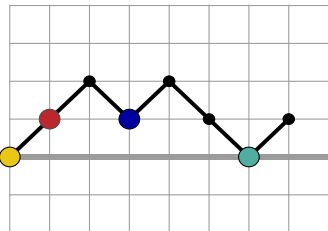
(+ - - - - + + +)



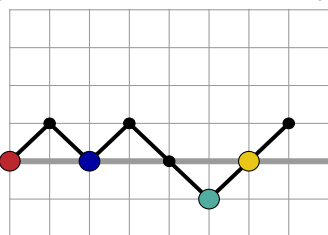
(+ + + - + - -)



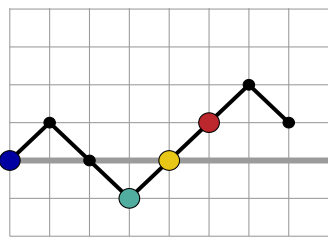
(+ + - + - - +)



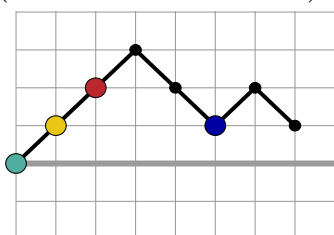
(+ - + - - + +)



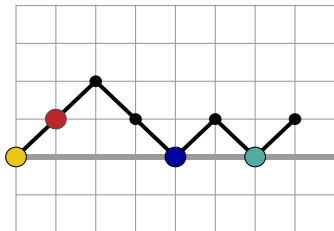
(+ - - + + + -)



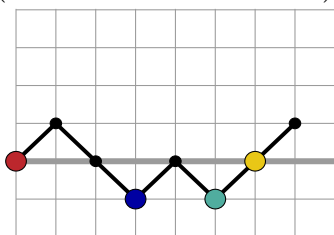
(+ + + - - + -)



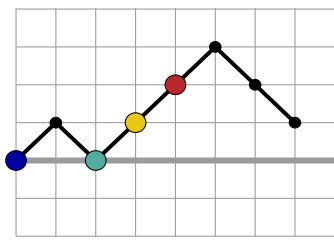
(+ + - - + - +)



(+ - - + - + +)



(+ - + + + - -)



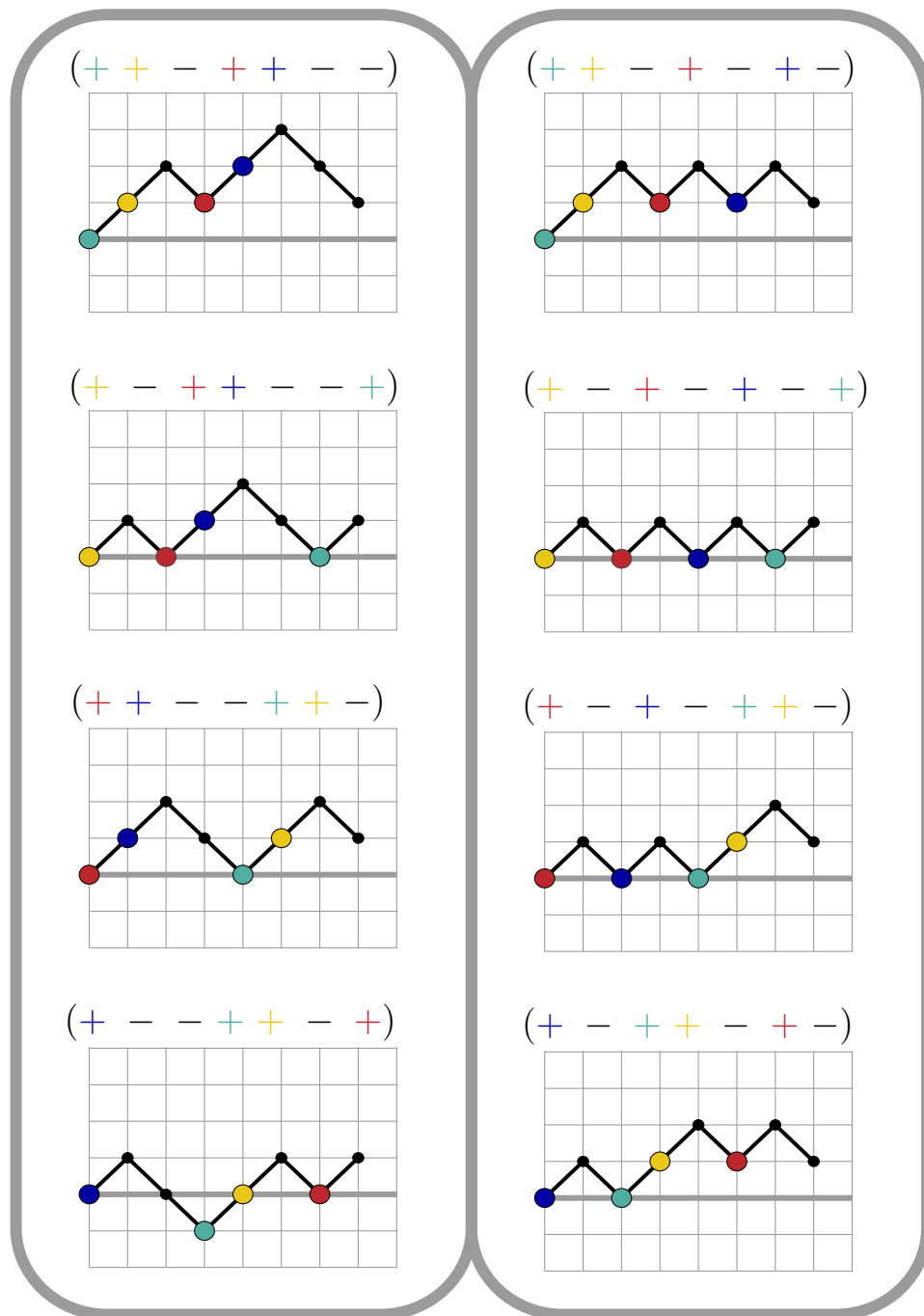


Figura 1.6: Caso $k = 3$. Es decir, caminos del $(0, 0)$ al $(7, 1)$ que empiezan subiendo. Hay 5 clases de equivalencia (rectángulos grises), cada una con 4 permutaciones cíclicas que empiezan subiendo. En cada clase el primer camino es el único que es un camino de Dyck extendido. Los vértices tienen colores para hacer más evidente la permutación cíclica en cada uno. Podemos concluir que hay $20/4 = 5$ árboles planos de tamaño 3.

Capítulo 2

Suma de variables aleatorias

En este capítulo analizamos cómo se concentra la suma de variables aleatorias que satisfacen condiciones específicas de independencia. Intuitivamente podemos pensar en el siguiente escenario: si las variables aleatorias no dependen demasiado entre sí, su suma tiende a agruparse en un rango reducido. Esto ocurre porque, al limitar la dependencia entre estas variables, se disminuye drásticamente la probabilidad de que todas alcancen valores extremos al mismo tiempo. Como consecuencia, es muy poco probable que su suma alcance valores extremos y, por lo tanto, se concentra considerablemente alrededor de su media. En un contexto más general¹ esto se conoce como el fenómeno de concentración de la medida.

En la primera sección del capítulo presentamos un resultado clásico de Hoeffding que analiza la suma de variables acotadas y conjuntamente independientes. Además, este resulta útil en el capítulo 3 al estudiar la concentración de la norma de operador de matrices aleatorias.

Como preámbulo del capítulo 4, la segunda sección contiene una extensión de este análisis a variables aleatorias independientes k a k para alguna $k \in \mathbb{N}$. Utilizando las propiedades combinatorias de los momentos de la suma podemos concluir cómo

¹Hacemos referencia, sobre todo, a desigualdades isoperimétricas y cuestiones relacionadas. El lector interesado puede consultar la monografía de Michel Talagrand [[Tal95](#)].

decaen sus colas. Esta situación resulta análoga a cuando en el capítulo 4 estudiamos la traza de las potencias de una matriz aleatoria.

2.1. Desigualdad de Hoeffding

Como un primer acercamiento al fenómeno de concentración de la medida, presentamos un resultado que obtuvo Wassily Hoeffding en 1963: la suma de variables aleatorias acotadas y conjuntamente independientes es subgaussiana. La demostración que en seguida presentamos es una adaptación de la que publicó en [Hoe63]. Su esencia reside en usar la desigualdad de Chernoff y después usar una cota de la función generadora de momentos de cada variable aleatoria. Así, primero presentamos dicha cota como un lema.

Lema 2.1 (Lema de Hoeffding). *Si X es una variable aleatoria acotada casi seguramente en el intervalo $[a, b]$ entonces para toda $t > 0$ se cumple que*

$$\mathbb{E} \left(e^{t(X - \mathbb{E}(X))} \right) \leq \exp \left(\frac{1}{8} t^2 (b - a)^2 \right).$$

Demostración. Sea $t > 0$ y X una variable aleatoria tal que $a \leq X \leq b$ casi seguramente. Sin pérdida de generalidad, suponemos que $\mathbb{E}(X) = 0$.

Como la función $f(x) = e^{tx}$ es convexa se cumple que

$$e^{tx} \leq \frac{b - x}{b - a} e^{ta} + \frac{x - a}{b - a} e^{tb}.$$

Esto se puede interpretar como que su gráfica está acotada superiormente en (a, b) por el segmento de recta que une a (a, e^{at}) con (b, e^{bt}) (ver figura 2.1).

Como X está en $[a, b]$ casi seguramente y $\mathbb{E}(X) = 0$ entonces tras un cálculo explícito obtenemos que

$$\mathbb{E} \left(e^{tX} \right) \leq \frac{b}{b - a} e^{ta} + \frac{-a}{b - a} e^{tb}.$$

Para el siguiente argumento es útil tener esa cota en forma exponencial. Es decir,

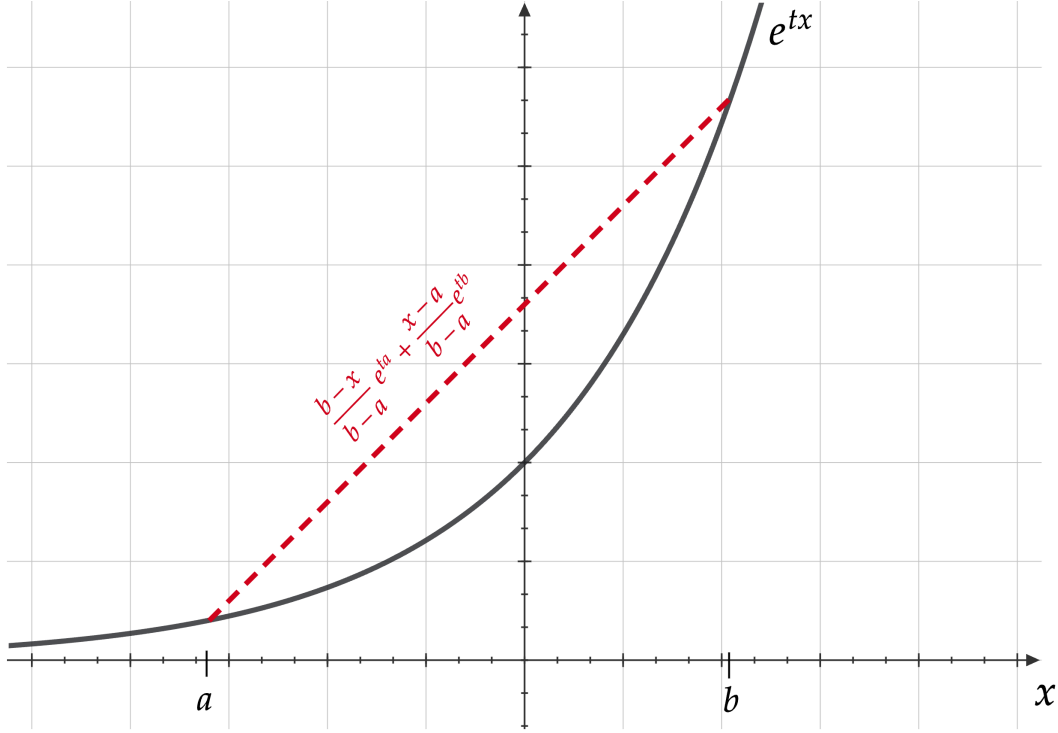


Figura 2.1: interpretación gráfica de la convexidad de e^{tx} .

buscamos una función L evaluada en $t(b-a)$ tal que la cota sea $e^{L(t(b-a))}$. Veamos entonces cuál es la función L que buscamos:

$$\begin{aligned}
 \frac{b}{b-a}e^{ta} + \frac{-a}{b-a}e^{tb} &= \left(1 + \frac{a}{b-a}\right)e^{ta} + \frac{-a}{b-a}e^{tb} \\
 &= e^{ta} \left(1 + \frac{a}{b-a} + \frac{-a}{b-a}e^{t(b-a)}\right) \\
 &= \exp \left(ta + \log \left(1 + \frac{a}{b-a} + \frac{-a}{b-a}e^{t(b-a)} \right) \right) \\
 &= \exp \left(\frac{t(b-a)a}{b-a} + \log \left(1 + \frac{a}{b-a} + \frac{-a}{b-a}e^{t(b-a)} \right) \right).
 \end{aligned}$$

Concluimos así que $\mathbb{E}(e^{tX}) \leq e^{L(t(b-a))}$ en donde

$$L(x) = \frac{xa}{b-a} + \log \left(1 + \frac{a - ae^x}{b-a} \right).$$

Ahora, usando la serie de Taylor de $L(x)$ alrededor de 0 hasta el término lineal y expresando el residuo en la forma de Lagrange obtenemos que

$$L(x) = L(0) + L'(0)x + \frac{1}{2}L''(\theta)x^2 \quad (2.1)$$

para alguna $\theta \in (0, x)$.

Calculando explícitamente las derivadas de $L(x)$ tenemos que

$$L'(x) = \frac{a}{b-a} + \frac{-ae^x}{b-ae^x} \quad y \quad L''(x) = \frac{abe^x}{(b-ae^x)^2}$$

y, por lo tanto, $L(0) = L'(0) = 0$. Así, considerando la expresión (2.1), la cota ahora queda como

$$\mathbb{E}(e^{tX}) \leq \exp\left(\frac{1}{2}L''(\theta)(t(b-a))^2\right) \quad (2.2)$$

para alguna $\theta \in (0, x)$.

Notemos ahora que de la desigualdad $0 \leq (b + ae^\theta)^2$ se sigue que $\frac{abe^\theta}{(b-ae^\theta)^2} \leq 1/4$ para toda θ . Así, podemos concluir que L'' está uniformemente acotada por $1/4$ y finalmente, sustituyendo en la ecuación (2.2) obtenemos que

$$\mathbb{E}(e^{tX}) \leq \exp\left(\frac{1}{8}(t(b-a))^2\right). \quad \square$$

Con este lema podemos demostrar el resultado principal de la sección.

Teorema 2.2 (Desigualdad de Hoeffding). *Sean X_1, \dots, X_n variables aleatorias conjuntamente independientes tales que para cada $i \in \{1, \dots, n\}$ se cumple que $X_i \in [a_i, b_i]$ casi seguramente y $S_n = \sum_{i=1}^n X_i$. Entonces*

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq \lambda) \leq 2 \exp(-2\lambda^2/\sigma_n^2)$$

donde $\sigma_n^2 = \sum_{i=1}^n |b_i - a_i|^2$.

Demostración. Sin pérdida de generalidad asumimos que para cada $i \in \{1, \dots, n\}$

$\mathbb{E}(X_i) = 0$ y, por lo tanto, $\mathbb{E}(S_n) = 0$. Analicemos primero la cola superior, es decir, la cantidad $\mathbb{P}(S_n \geq \lambda)$.

Por la desigualdad de Chernoff, sabemos que para cada $t > 0$ se cumple que

$$\mathbb{P}(S_n \geq \lambda) \leq \mathbb{E}(e^{tS_n}) e^{-t\lambda} = \mathbb{E}(e^{t \sum_{i=1}^n X_i}) e^{-t\lambda}$$

y por la independencia conjunta de las X_i podemos concluir que

$$\mathbb{P}(S_n \geq \lambda) \leq e^{-t\lambda} \prod_{i=1}^n \mathbb{E}(e^{tX_i}).$$

Ahora, para acotar la función generadora de momentos de cada X_i usamos el lema anterior para obtener que

$$\mathbb{P}(S_n \geq \lambda) \leq e^{-t\lambda} \prod_{i=1}^n \exp\left(\frac{1}{8}t^2(b_i - a_i)^2\right) = \exp\left(-t\lambda + \frac{1}{8}t^2 \sum_{i=1}^n |b_i - a_i|^2\right).$$

Recordemos que para simplificar la notación denotamos como σ_n^2 a la constante $\sum_{i=1}^n |b_i - a_i|^2$ y así podemos escribir la cota como

$$\mathbb{P}(S_n \geq \lambda) \leq \exp\left(-t\lambda + \frac{1}{8}t^2\sigma_n^2\right).$$

Para encontrar la mejor cota derivamos la expresión respecto a t e igualamos a 0 para obtener que la función alcanza un mínimo en $t = 4\lambda/\sigma_n^2$. Evaluando en esa t la expresión queda como

$$\mathbb{P}(S_n \geq \lambda) \leq \exp(-2\lambda^2/\sigma_n^2).$$

Para acotar la cola inferior podemos hacer el mismo análisis pero considerando ahora a la variable $-S_n$ y obtenemos la misma cota. Es decir,

$$\mathbb{P}(S_n \leq -\lambda) = \mathbb{P}(-S_n \geq \lambda) \leq \exp(-2\lambda^2/\sigma_n^2).$$

Finalmente, tenemos que

$$\mathbb{P}(|S_n| \geq \lambda) = \mathbb{P}(S_n \geq \lambda) + \mathbb{P}(S_n \leq -\lambda) \leq 2 \exp(-2\lambda^2/\sigma_n^2). \quad \square$$

Notemos que para que esa cota no sea trivial, es decir menor que uno, debe ocurrir que λ sea suficientemente grande. En particular, se debe cumplir que $\lambda^2 \geq \sigma_n^2 \log(2)/2$.

Aunque en la práctica puede ser útil calcular explícitamente la constante σ_n^2 , para el estudio de resultados asintóticos es más útil normalizar S_n de tal manera que la cota no dependa de n . Para esto, consideremos una constante C suficientemente grande que cumpla que toda X_i esté acotada casi seguramente en el intervalo $(-C, C)$. Con un cálculo explícito podemos ver que $\sigma_n^2 = 4nC^2$ y por la desigualdad de Hoeffding aplicada a $\lambda\sqrt{n}$ obtenemos

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \exp(-\lambda^2/2C^2).$$

Así, podemos concluir que si consideramos la normalización $\frac{1}{\sqrt{n}}S_n$ obtenemos una variable subgaussiana.

2.2. Concentración vía sus momentos

Veamos ahora que podemos obtener desigualdades de concentración de la suma de variables aleatorias acotadas analizando sus momentos y después usando una forma de la desigualdad de Markov.

Si bien esta técnica es interesante por sí misma porque aprovecha los aspectos combinatorios de los momentos de la suma, también resulta más útil para generalizar estos resultados a matrices aleatorias. Aunque la generalización no es directa, resulta más sencilla hacerlo por esta vía que aludir a la función generadora de momentos. Un tratamiento a profundidad de esta idea puede verse, por ejemplo, en [Tao12].

Además, con el análisis de los momentos de la suma podemos obtener resultados

de concentración bajo condiciones más débiles de independencia entre las variables aleatorias que sumamos. En particular analizamos el caso en que las variables son independientes k a k .

Definición 2.3 (Independencia k a k). *Sea $k \in \mathbb{N}$. Decimos que un conjunto de $n \geq k$ variables aleatorias es independiente k a k si para cualquier subconjunto de k variables distintas su distribución conjunta es el producto de sus distribuciones marginales.*

Al final de la sección vemos que si asumimos independencia conjunta podemos recuperar el resultado de concentración que nos da la desigualdad de Hoeffding.

Como un primer acercamiento analicemos el k -ésimo momento (donde k es par) de la suma de variables aleatorias que cumplan ser independientes k a k .

Proposición 2.4 (Momentos pares de la suma de v.a. independientes k a k). *Sea k un entero par. Sean X_1, \dots, X_n variables aleatorias independientes k a k , acotadas casi seguramente y con k -ésimo momento finito. Si denotamos $S_n = \sum_{i=1}^n X_i$ entonces*

$$\mathbb{E}(S_n^k) \leq 2(enk/2)^{k/2}$$

cuando n es suficientemente grande.

Demostración. Sin pérdida de generalidad podemos centrar cada variable para que $\mathbb{E}(X_i) = 0$, normalizarla para que $\text{Var}(X_i) = 1$ y considerar una C suficientemente grande para que todas resulten acotadas casi seguramente en $(-C, C)$.

Por la linealidad de la esperanza, podemos expresar al k -ésimo momento de S_n como

$$\mathbb{E}(S_n^k) = \sum_{1 \leq i_1, \dots, i_k \leq n} \mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_k}).$$

Observemos que los índices i_j se pueden repetir. Si, por el contrario, en algún sumando algún índice aparece exactamente una vez entonces, por la hipótesis de independencia y porque $\mathbb{E}(X_i) = 0$, todo el sumando es 0. Así, es suficiente estudiar solo los sumandos que cumplen que todas las X_i que aparecen estén al menos 2 veces.

Veamos dos casos particulares de aquellos sumandos que sí aportan a la suma. Primero supongamos que en un sumando todos los índices que aparecen lo hacen exactamente 2 veces, o dicho de otra manera que el término consta de exactamente $k/2$ variables donde todas tienen exponente 2. Entonces recordando nuestras hipótesis, en especial la independencia k a k , tenemos que

$$\mathbb{E} \left(X_{i_1}^2 X_{i_2}^2 \cdots X_{i_{k/2}}^2 \right) = \mathbb{E} (X_{i_1}^2) \cdots \mathbb{E} (X_{i_{k/2}}^2) = 1.$$

Ahora, otro caso posible es que todos los índices sean el mismo, o dicho de otra forma, que solo haya una X_i con exponente k . Entonces podemos acotar ese sumando de la siguiente manera:

$$\mathbb{E}(X_i^k) = \mathbb{E} (X_i^2 X_i^{k-2}) \leq \mathbb{E} (X_i^2 C^{k-2}) = C^{k-2} \mathbb{E} (X_i^2) = C^{k-2}. \quad (2.3)$$

Para estudiar el caso general supongamos que en un término aparecen exactamente $k/2 - r$ variables distintas y cada una de ellas aparece al menos dos veces. Digamos, para simplificar, que cada X_i aparece exactamente $2 + \alpha_i$ veces. Así, nuestro término se ve como

$$\mathbb{E} \left(X_{i_1}^{2+\alpha_{i_1}} X_{i_2}^{2+\alpha_{i_2}} \cdots X_{i_{k/2-r}}^{2+\alpha_{i_{k/2-r}}} \right).$$

Preparándonos para el siguiente argumento observemos que, como estamos analizando el k -ésimo momento, la suma de los exponentes de cualquier término siempre es igual a k . Es decir, $k = \sum_{j=1}^{k/2-r} (2 + \alpha_{i_j})$, o bien, $k = k - 2r + \sum_{j=1}^{k/2-r} \alpha_{i_j}$. Por lo tanto, podemos concluir que la suma de las α_i siempre es igual a $2r$.

Análogo a lo que hicimos en (2.3) tenemos que

$$\mathbb{E} (X_i^{2+\alpha_i}) \leq \mathbb{E} (X_i^2 C^{\alpha_i}) = C^{\alpha_i} \mathbb{E} (X_i^2) = C^{\alpha_i}.$$

Así, por la hipótesis de independencia y las observaciones que hicimos, podemos acotar

la esperanza del producto como

$$\begin{aligned}
 \mathbb{E} \left(X_{i_1}^{2+\alpha_{i_1}} \cdots X_{i_{k/2-r}}^{2+\alpha_{i_{k/2-r}}} \right) &= \prod_{j=1}^{k/2-r} \mathbb{E} \left(X_{i_j}^{2+\alpha_{i_j}} \right) \\
 &\leq \prod_{j=1}^{k/2-r} C^{\alpha_{i_j}} \\
 &= C^{\sum_{j=1}^{k/2-r} \alpha_{i_j}} \\
 &= C^{2r}.
 \end{aligned}$$

Notemos que obtuvimos una cota que depende únicamente de cuántos índices distintos aparecen en ese término. Por lo tanto, podemos agrupar los sumandos de S_k según la cantidad de variables distintas presentes en cada uno. Esto nos permite concluir que

$$\mathbb{E} (S_n^k) \leq \sum_{r=0}^{k/2-1} C^{2r} N_r \quad (2.4)$$

donde N_r denota el número de sumandos distintos que contienen exactamente $k/2 - r$ variables diferentes. Calcular N_r explícitamente, y por lo tanto acotar el k -ésimo momento de la suma, se reduce a un problema puramente combinatorio: ¿de cuántas maneras pueden seleccionarse exactamente $k/2 - r$ índices del conjunto $\{1, \dots, n\}$ tal que cada índice seleccionado aparezca al menos dos veces?

Como una primera cota para N_r , notemos que hay $\binom{n}{k/2-r}$ maneras de elegir $k/2 - r$ índices del conjunto $\{1, \dots, n\}$ y luego cada uno de los k términos del producto de las variables elegidas puede tener, en principio, cualquiera de esos $k/2 - r$ índices. Aunque de esta forma estamos contando muchos casos que no nos interesan. En particular, estamos contando casos en donde hay índices que solo aparecen una vez y casos en donde aparecen menos de $k/2 - r$ índices. Así, tenemos que

$$N_r \leq \binom{n}{k/2-r} (k/2-r)^k.$$

Esta cota definitivamente no es la mejor, pero es suficientemente útil para obtener resultados de concentración relevantes. Si ahora acotamos el coeficiente binomial como

$$\binom{n}{k/2-r} \leq \frac{n^{k/2-r}}{(k/2-r)!}$$

obtenemos que

$$N_r \leq \frac{n^{k/2-r}}{(k/2-r)!} (k/2-r)^k.$$

Para simplificar esa última expresión podemos acotar el factorial utilizando una forma burda de la fórmula de Stirling. De la expansión de Taylor de e^x tenemos que para toda $x \geq 0$ y para toda $n \geq 0$ se cumple que $\frac{x^n}{n!} \leq e^x$. Si ahora hacemos $x = n$ obtenemos que $\frac{1}{n!} \leq e^n n^{-n}$. En particular, cuando $n = k/2 - r$ tenemos la siguiente desigualdad: $\frac{1}{(k/2-r)!} \leq e^{k/2-r} (k/2-r)^{r-k/2}$. Por lo tanto,

$$\begin{aligned} N_r &\leq e^{k/2-r} (k/2-r)^{r-k/2} n^{k/2-r} (k/2-r)^k \\ &= (en)^{k/2-r} (k/2-r)^{r+k/2} \\ &\leq (en)^{k/2-r} (k/2)^{r+k/2}. \end{aligned} \tag{2.5}$$

Así, considerando (2.4) y (2.5) podemos concluir que

$$\begin{aligned} \mathbb{E}(S_n^k) &\leq \sum_{r=0}^{k/2-1} (en)^{k/2-r} (k/2)^{r+k/2} C^{2r} \\ &= (enk/2)^{k/2} \sum_{r=0}^{k/2-1} \left(\frac{k C^2}{2en} \right)^r. \end{aligned}$$

Notemos que bajo la suposición $C^2 \leq n/k$ el término dentro de la suma se simplifica considerablemente. Asumir esto no debería preocuparnos demasiado pues en el fondo siempre estamos pensando que n es suficientemente grande. De ahora en adelante

asumimos esa condición. Así, finalmente obtenemos que

$$\begin{aligned}
 \mathbb{E}(S_n^k) &\leq (enk/2)^{k/2} \sum_{r=0}^{k/2-1} \left(\frac{1}{2e}\right)^r \\
 &\leq (enk/2)^{k/2} \sum_{r=0}^{\infty} \left(\frac{1}{2e}\right)^r \\
 &= (enk/2)^{k/2} \left(\frac{1}{1 - (\frac{1}{2e})}\right) \\
 &\leq 2 (enk/2)^{k/2}. \quad \square
 \end{aligned}$$

Tomando en cuenta esta cota del k -ésimo momento de S_n podemos usar una forma de la desigualdad de Markov para obtener una desigualdad de concentración de S_n . Es aquí donde es útil que k sea par pues tenemos que $|S_n|^k = S_n^k$.

Teorema 2.5 (Concentración de la suma de v.a. independientes k a k). *Con las mismas hipótesis de la proposición anterior se cumple que*

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \left(\frac{\sqrt{ek/2}}{\lambda}\right)^k.$$

Demostración. Por la desigualdad de Markov tenemos que para toda $\lambda > 0$

$$\mathbb{P}(|S_n| \geq \lambda) = \mathbb{P}(|S_n|^k \geq \lambda^k) \leq \frac{\mathbb{E}(|S_n|^k)}{\lambda^k}.$$

Como k es par podemos quitar el valor absoluto de S_n y usando la cota que obtuvimos para el k -ésimo momento en la proposición 2.4 podemos concluir que

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq \frac{2 \left(\frac{enk}{2}\right)^{k/2}}{(\lambda\sqrt{n})^k} = 2 \left(\frac{\sqrt{ek/2}}{\lambda}\right)^k. \quad \square$$

Es decir, con independencia k a k para una k par fija, obtenemos que las colas de $\frac{1}{\sqrt{n}}S_n$ tienen decaimiento polinomial en λ (del orden de λ^{-k}).

Veamos ahora, aunque la desigualdad de Hoeffding ya establece este hecho, que bajo independencia conjunta es posible lograr un decaimiento de orden exponencial cuadrático (subgaussiano). Pese a que en la proposición anterior hemos ya expuesto la esencia del método, incluimos este desarrollo para insistir en que el resultado de Hoeffding (salvo constantes) también puede obtenerse mediante este enfoque.

Corolario 2.6 (Concentración de la suma de v.a. conjuntamente independientes).

Sean X_1, \dots, X_n variables aleatorias conjuntamente independientes, acotadas casi seguramente y $S_n = \sum_{i=1}^n X_i$. Entonces

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \exp(-\lambda^2/e^2).$$

Demostración. Sin pérdida de generalidad asumimos que $\mathbb{E}(X_i) = 0$, $\text{Var}(X_i) = 1$ y consideramos una C suficientemente grande para que todas las X_i resulten acotadas casi seguramente en $(-C, C)$.

Notemos que como las X_i son conjuntamente independientes entonces son independientes k a k para toda $k \in \mathbb{N}$ y, por lo tanto, la cota

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \left(\frac{\sqrt{ek/2}}{\lambda} \right)^k$$

se cumple para toda k par que cumpla que $C^2 \leq n/k$, la condición que asumimos en la demostración de la proposición 2.4. Una vez más, esta condición es equivalente a pedir que n sea suficientemente grande. Tomando en cuenta esto, podemos optimizar k para obtener una buena cota. Consideremos la cota como una función de variable real, es decir la función

$$f(x) = 2 \left(\frac{\sqrt{ex/2}}{\lambda} \right)^x$$

con $x \in \mathbb{R}$.

Podemos juntar todas las constantes definiendo $c = \frac{e/2}{\lambda^2}$, y escribir f de la siguiente

forma:

$$f(x) = 2 \exp \left(\frac{1}{2} x \log(cx) \right).$$

Notemos que f es una función convexa y, por lo tanto, tiene a lo más un mínimo global. Ahora, es suficiente minimizar el exponente para encontrar el mínimo de f . Derivando e igualando a 0 obtenemos

$$0 = \frac{d}{dx} x \log(cx) = \log(cx) + 1,$$

lo que implica que

$$x = \frac{1}{ce} = \frac{1}{\left(\frac{e/2}{\lambda^2}\right) e} = \frac{2\lambda^2}{e^2}.$$

Por lo tanto, podemos considerar un entero par cercano a $\frac{2\lambda^2}{e^2}$. Sea $k = \lfloor \frac{2\lambda^2}{e^2} \rfloor$ o $k = \lfloor \frac{2\lambda^2}{e^2} \rfloor + 1$ dependiendo de su paridad.

Sustituyendo esa k en la condición que estamos asumiendo sobre n nos permite concluir que específicamente es necesario que $n \geq \frac{2\lambda^2 C^2}{e^2}$.

Habiendo encontrado un entero par cerca del mínimo, podemos ahora evaluar la cota. Como $k \leq \frac{2\lambda^2}{e^2}$ entonces tenemos que

$$\begin{aligned} \mathbb{P}(|S_n| \geq \lambda\sqrt{n}) &\leq 2 \left(\frac{\sqrt{e \left(\frac{2\lambda^2}{e^2}\right) / 2}}{\lambda} \right)^{\left(\frac{2\lambda^2}{e^2}\right)} \\ &= 2 \left(\frac{1}{e} \right)^{\left(\frac{\lambda^2}{e^2}\right)} \\ &= 2 \exp \left(- \left(\frac{1}{e^2} \right) \lambda^2 \right). \end{aligned} \quad \square$$

En resumen, nuevamente concluimos que si tenemos independencia conjunta y n es suficientemente grande entonces podemos afirmar que $\frac{1}{\sqrt{n}} S_n$ es subgaussiana.

Capítulo 3

Norma de operador de matrices aleatorias

En este capítulo estudiamos la norma de operador de matrices aleatorias. Además de ser una estadística interesante por sí misma, en este trabajo es particularmente relevante por su relación con el espectro de la matriz, que estudiamos en el capítulo [4](#).

En analogía al capítulo [2](#), primero abordamos el caso de matrices cuyas entradas son variables aleatorias independientes y después obtenemos resultados sobre su concentración. Para esto introducimos la técnica de la ε -red, un concepto propio de espacios métricos, y la usamos para analizar cómo decaen las colas de la norma de operador de estas matrices. Esta técnica consiste en aproximar la norma de operador con un conjunto finito pero significativo de vectores, en lugar de considerar a todos los vectores de \mathbb{R}^n .

En la segunda sección de este capítulo hacemos una breve digresión para estudiar el método de simetrización. Exponemos algunos resultados de [\[Lat05\]](#) para ejemplificar la diversidad y riqueza en las técnicas empleadas en el estudio de la norma de operador de matrices aleatorias. En particular, establecemos cotas relacionadas con la esperanza de la norma de operador de matrices aleatorias con entradas i.i.d. bajo el supuesto

adicional que sus entradas son variables aleatorias con distribuciones simétricas.

Comencemos por definir la norma de operador de una matriz. Si pensamos a una matriz como un operador en \mathbb{R}^n , nos interesa conocer cuánto es lo máximo que un vector cambia de longitud después de multiplicarlo por esa matriz. Como una matriz define a un operador lineal entonces es suficiente considerar a los vectores unitarios. Así, la siguiente definición resulta natural.

Definición 3.1 (Norma de operador de una matriz). *Dada una matriz M de dimensión $n \times n$ con entradas reales, no necesariamente aleatoria, definimos su norma de operador como*

$$||M||_{op} = \sup_{x \in S_n} |Mx|$$

donde S_n es la esfera unitaria en \mathbb{R}^n , es decir, $S_n = \{x \in \mathbb{R}^n : |x| = 1\}$.

Una propiedad interesante, y que es útil al estudiar el espectro de matrices aleatorias, es que la norma de operador acota a la norma de cualquier valor propio. Veámoslo con la siguiente proposición.

Proposición 3.2 (La norma de operador acota al espectro). *Sea M una matriz de dimensión $n \times n$ con entradas reales. Entonces para todo valor propio λ de M se cumple que*

$$|\lambda| \leq ||M||_{op}.$$

Demostración. Sea x un vector propio correspondiente al valor propio λ . Sin pérdida de generalidad podemos asumir que $|x| = 1$. Así, tenemos que

$$|\lambda| = |\lambda x| = |Mx| \leq ||M||_{op}. \quad \square$$

Notemos que cuando las entradas de M son variables aleatorias, entonces su norma de operador es una variable aleatoria real. Por lo tanto, al obtener resultados de concentración de $||M||_{op}$ estamos obteniendo también resultados de concentración del espectro de M .

3.1. Técnica de la ε -red

Como una primera observación, notemos que como la esfera unitaria en \mathbb{R}^n es compacta, multiplicar por M y tomar la norma en \mathbb{R}^n son funciones continuas, entonces el supremo en la definición es en realidad un máximo. Sin embargo, es habitual definir la norma de operador como un supremo en analogía a cómo se define en contextos más generales, a saber, para operadores en espacios de Hilbert.

Ahora, para estudiar qué tan concentrada es la norma de operador de una matriz aleatoria M notemos que para cualquier $\lambda > 0$ ocurre que $\max_{x \in S_n} |Mx| > \lambda$ si y solo si existe un vector $x \in \mathbb{R}^n$ con $|x| = 1$ que cumple que $|Mx| > \lambda$. Expresado como una igualdad de eventos tenemos que

$$\{\|M\|_{op} > \lambda\} = \bigcup_{x \in S_n} \{|Mx| > \lambda\}.$$

Aunque podríamos pensar directamente en usar la subaditividad de la probabilidad hay que notar que esa unión no es numerable. Sin embargo, el método de la ε -red nos da una alternativa para esta situación. Esta técnica consiste en tomar una cantidad finita de vectores de S_n que nos dé una buena aproximación de la norma de operador de M . Así, para cada vector x que escojamos podemos acotar $\mathbb{P}(|Mx| > \lambda)$ y, ahora sí, usar subaditividad para obtener resultados sobre la concentración de $\|M\|_{op}$.

Para esto, primero necesitamos tener buenas cotas para la probabilidad del evento $|Mx| > \lambda$. En la siguiente proposición demostramos que podemos lograrlo aplicando los resultados de concentración de la medida que desarrollamos en el capítulo 2.

Es relevante notar que de ahora en adelante, así como lo hicimos en el capítulo 2, asumimos que las variables están acotadas casi seguramente. Sin embargo, en el capítulo 4, al estudiar la ley del semicírculo consideramos variables aleatorias no necesariamente acotadas. Con el método de truncamiento, que exponemos en el apéndice A, es posible extender estos resultados de concentración de la norma de operador a clases más generales de matrices aleatorias.

Proposición 3.3 (Concentración de $|Mx|$). *Sea M una matriz de dimensión $n \times n$ tal que sus entradas, ξ_{ij} , son variables aleatorias reales, independientes, con $\mathbb{E}(\xi_{i,j}) = 0$ y $|\xi_{i,j}| \leq 1$ casi seguramente. Sea x un vector unitario en \mathbb{R}^n . Entonces para λ suficientemente grande (mayor a una constante que no depende de n) se cumple que*

$$\mathbb{P}(|Mx| \geq \lambda\sqrt{n}) \leq \exp(-\lambda^2 n / C^2)$$

con C una constante positiva.

Demostración. Notemos que $(Mx)_i$, la i -ésima entrada del vector Mx , está dado por el producto punto de la i -ésima fila de M con x , es decir, $(Mx)_i = \sum_{j=1}^n \xi_{ij}x_j$. Veamos ahora que se cumplen las hipótesis de la desigualdad de Hoeffding (teorema 2.2) para esa suma de variables aleatorias. Cada sumando es una variable aleatoria acotada casi seguramente pues $|\xi_{ij}| \leq 1$ y entonces $|\xi_{ij}x_j| \leq |x_j|$. Además los sumandos son independientes pues las variables aleatorias ξ_{ij} lo son. Entonces podemos concluir que

$$\mathbb{P}(|(Mx)_i| \geq \lambda) \leq 2 \exp(-2\lambda^2 / \sigma_n^2)$$

donde σ_n^2 está definido como en la proposición 2.2. En este caso, como $|\xi_{ij}x_j| \leq |x_j|$ y x es un vector unitario entonces

$$\sigma_n^2 = \sum_{j=1}^n (2|x_j|)^2 = 4 \sum_{j=1}^n |x_j|^2 = 4.$$

Así, tenemos que

$$\mathbb{P}(|(Mx)_i| \geq \lambda) \leq 2 \exp(-\lambda^2 / 2),$$

es decir, que para cada $i \in \{1, \dots, n\}$ la variable aleatoria $(Mx)_i$ es subgaussiana con factor de varianza 2. Ahora, recordemos que la proposición 1.5 ofrece una cota para la función generadora de momentos de $|(Mx)_i|^2$ en una vecindad del 0. En particular, en $[-\frac{1}{4}, \frac{1}{4}]$ está acotada por 4 y así tenemos que

$$\mathbb{E}(e^{\frac{1}{4}|(Mx)_i|^2}) \leq 4. \tag{3.1}$$

Como las variables aleatorias $|(Mx)_i|^2$ son independientes, pues las ξ_{ij} lo son, y además $|Mx|^2 = \sum_{i=1}^n |(Mx)_i|^2$ entonces multiplicando las desigualdades que obtenemos en (3.1) para cada $i \in \{1, \dots, n\}$ tenemos que

$$\mathbb{E} \left(e^{\frac{1}{4}|Mx|^2} \right) = \prod_{i=1}^n \mathbb{E} \left(e^{\frac{1}{4} |(Mx)_i|^2} \right) \leq 4^n. \quad (3.2)$$

Por lo tanto, usando la desigualdad de Chernoff y (3.2) podemos concluir que

$$\mathbb{P}(|Mx| \geq \lambda\sqrt{n}) \leq \mathbb{E} \left(e^{\frac{1}{4}|Mx|^2} \right) e^{-\frac{1}{4}\lambda^2 n} \leq \exp \left(n \log(4) - \frac{1}{4}\lambda^2 n \right).$$

Para simplificar el exponente en la cota notemos que si $\lambda > 6$ entonces se cumple que $\frac{\lambda^2}{4} - \log(4) \geq \lambda^2/5$ y, por lo tanto,

$$\exp \left(- \left(\frac{\lambda^2}{4} - \log(4) \right) n \right) \leq \exp \left(-\lambda^2 n/5 \right).$$

Finalmente, podemos concluir que si $\lambda > 6$ entonces

$$\mathbb{P}(|Mx| \geq \lambda\sqrt{n}) \leq \exp \left(-\lambda^2 n/5 \right). \quad \square$$

Ahora que contamos con cotas precisas para el evento $|Mx| > \lambda$, podemos dar un argumento propio de espacios métricos para finalmente usar la subaditividad de la probabilidad. La idea esencial es elegir un subconjunto representativo de S_n que sea finito. Para esto, introducimos el concepto de ε -red.

Definición 3.4 (ε -red). *Un conjunto $\Sigma \subseteq \mathbb{R}^n$ es una ε -red si para cada par de puntos distintos en Σ , digamos x y y , se cumple que $|x - y| \geq \varepsilon$.*

En nuestro caso usamos una ε -red específicamente sobre la esfera unitaria en \mathbb{R}^n . Es decir, consideramos una ε -red donde además todos los vectores en ella tienen norma 1. Veámoslo con detalle en el siguiente lema.

Lema 3.5. *Sea $0 < \varepsilon < 1$ y $\Sigma \subseteq S_n$ una ε -red en la esfera unitaria de \mathbb{R}^n que cumpla ser maximal respecto a la contención de conjuntos. Entonces para toda matriz*

aleatoria M con las hipótesis de la proposición anterior y para toda $\lambda > 0$ se da la siguiente contención de eventos

$$\{\|M\|_{op} > \lambda\} \subseteq \bigcup_{z \in \Sigma} \{|Mz| > (1 - \varepsilon)\lambda\}.$$

Demostración. Sea $\lambda > 0$ y supongamos que $\|M\|_{op} > \lambda$. Entonces es suficiente demostrar que existe un vector $y \in \Sigma$ que cumple que $|My| > (1 - \varepsilon)\lambda$.

Como hemos ya mencionado, el supremo en la definición de la norma de operador es en realidad un máximo, y por lo tanto podemos encontrar un vector, digamos $x \in \mathbb{R}^n$, tal que $|Mx| = \|M\|_{op}$. Aunque x no está necesariamente en Σ , sí podemos considerar un vector $y \in \Sigma$ suficientemente cercano y este es el que cumple lo que buscamos. Para esto, notemos que existe un $y \in \Sigma$ tal que $|x - y| \leq \varepsilon$. Si no existe tal, entonces $\Sigma \cup \{x\}$ contradice a la maximalidad de Σ . Por lo tanto, tenemos que

$$|M(x - y)| = |x - y| M \left(\frac{1}{|x - y|} (x - y) \right) \leq |x - y| \|M\|_{op} \leq \varepsilon \|M\|_{op}.$$

Usando la desigualdad del triángulo obtenemos $|Mx| - |My| \leq |M(x - y)|$, y recordando que $|Mx| = \|M\|_{op}$ finalmente tenemos que

$$|My| \geq (1 - \varepsilon)\|M\|_{op} > (1 - \varepsilon)\lambda$$

donde $y \in \Sigma$, estableciendo la contención deseada. \square

Lo anterior nos permite reducir el evento que nos interesa a una unión finita de eventos. Así usar la subaditividad de la probabilidad ya es una idea viable. Sin embargo, esa unión podría seguir siendo muy grande. Notemos que entre más pequeña sea ε tendremos mejores cotas para $\mathbb{P}(|Mx| > \lambda)$ pero Σ tendrá más elementos. Por eso es necesario proceder con cuidado. A continuación presentamos una cota para la cardinalidad de Σ .

Proposición 3.6 (Cardinalidad de una ε -red en la esfera unitaria). *Si $0 < \varepsilon < 1$ y*

$\Sigma \subseteq S_n$ es una ε -red entonces

$$|\Sigma| \leq \left(1 + \frac{2}{\varepsilon}\right)^n.$$

Demostración. Consideremos el conjunto de bolas de radio $\varepsilon/2$ con centro en cada punto de Σ . Por hipótesis, como Σ es una ε -red, estas bolas son disjuntas entre sí.

Ahora, sea $x \in \Sigma$ y $y \in \mathbb{R}^n$ tal que $|x - y| \leq \varepsilon/2$. Como x está en la esfera unitaria entonces $|x| = 1$ y, por la desigualdad del triángulo, podemos concluir que $|y| \leq 1 + \varepsilon/2$. Por lo tanto, todas las bolas con centro en un punto de Σ y radio $\varepsilon/2$ están contenidas en una bola con centro en el origen y radio $1 + \varepsilon/2$ (ver figura 3.1).

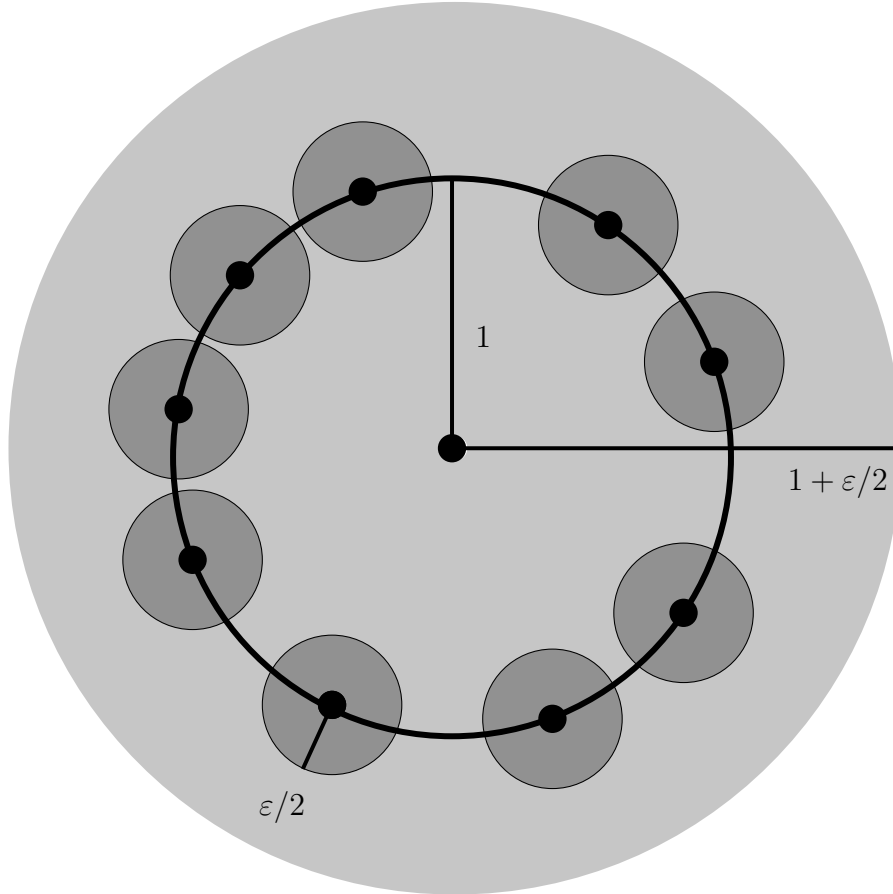


Figura 3.1: bolas de radio $\varepsilon/2$ (gris oscuro), con centros en una ε -red en la esfera unitaria en \mathbb{R}^2 (puntos negros), cubiertas por una bola de radio $1 + \varepsilon/2$ (gris claro).

Así, sabemos que el volumen de la unión de las bolas (que es la suma de sus volúmenes porque son disjuntas) es menor o igual al volumen de la región que las contiene. Es decir, si $V(\cdot)$ denota el volumen y $B(x, r)$ denota a la bola con centro en x y radio r entonces

$$\sum_{y \in \Sigma} V(B(y, \varepsilon/2)) \leq V(B(0, 1 + \varepsilon/2)). \quad (3.3)$$

Recordemos que en \mathbb{R}^n el volumen de una bola de radio r es Kr^n donde K es el volumen de la bola unitaria. Por lo tanto, la desigualdad (3.3) se traduce a que $|\Sigma|K(\varepsilon/2)^n \leq K(1 + \varepsilon/2)^n$ y así concluimos que

$$|\Sigma| \leq \frac{(2 + \varepsilon)^n}{\varepsilon^n} = \left(1 + \frac{2}{\varepsilon}\right)^n. \quad \square$$

Tomando en cuenta los resultados que hemos obtenido, ya podemos enunciar la cota que buscamos.

Teorema 3.7 (Concentración de la norma de ensambles independientes.). *Sea M una matriz aleatoria de dimensión $n \times n$ con entradas independientes ξ_{ij} que cumplen que $\mathbb{E}(\xi_{ij}) = 0$ y $|\xi_{ij}| \leq 1$ casi seguramente. Entonces para λ suficientemente grande (mayor a una constante que no depende de n) se cumple que*

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq \exp(-\lambda^2 n / C^2)$$

con C una constante positiva.

Demostración. Sea Σ una $\frac{1}{2}$ -red maximal en S_n . Por el lema 3.5 y subaditividad tenemos que

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq \sum_{x \in \Sigma} \mathbb{P}(|Mx| \geq \lambda\sqrt{n}/2).$$

Ahora, por la proposición 3.3, si $\lambda > 6$ podemos acotar cada uno de los sumandos para obtener que

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq |\Sigma| \exp\left(\frac{-\lambda^2 n}{4 \cdot 5}\right).$$

Por la proposición 3.6 sabemos que $|\Sigma| \leq 5^n$ y así concluimos que

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq 5^n \exp(-\lambda^2 n/20) = \exp\left(-\left(\lambda^2/20 - \log(5)\right)n\right).$$

El término que viene de la cardinalidad de Σ se puede absorber por la exponencial. Es decir, notando que para toda $\lambda > 14$ se cumple que $\lambda^2/20 - \log(5) \geq \lambda^2/24$ entonces se sigue que $\exp\left(-\left(\lambda^2/20 - \log(5)\right)n\right) \leq \exp(-\lambda^2 n/24)$.

Finalmente, podemos concluir que si $\lambda > 14$ entonces

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq \exp(-\lambda^2 n/24). \quad \square$$

El método de la ε -red se puede utilizar también en otros contextos en los que se quiera acotar el supremo de un conjunto infinito de variables aleatorias. Por ejemplo, podría usarse cuando se estudian supremos de procesos estocásticos. Hay varias sutilezas en el método y que en nuestro caso se cumplen principalmente por la independencia de las entradas de la matriz. Por ejemplo, en este caso no es muy complicado obtener cotas buenas para los elementos del conjunto del que nos interesa el supremo, es decir, los eventos $\{|Mx| > \lambda\}$. También, notemos que $\varepsilon = 1/2$ fue suficientemente bueno, pero en otros contextos elegir un ε adecuado podría requerir argumentos más sofisticados.

Concentración del ensamble de Wigner

En el estudio de matrices aleatorias son de particular relevancia las matrices simétricas pues todos sus valores propios son reales. Son el objeto principal de estudio del capítulo 4 donde demostramos la ley del semicírculo. Se les nombra en honor a Eugene Wigner que las estudió, por ejemplo, en [Wig55].

Definición 3.8 (Ensamble de Wigner). *Sea M una matriz cuyas entradas ξ_{ij} son variables aleatorias reales, idénticamente distribuidas e independientes si $i \neq j$. Decimos que M está en el ensamble de Wigner si cumple que $\mathbb{E}(\xi_{ij}) = 0$, $\text{Var}(\xi_{ij}) = 1$*

y es simétrica, es decir, $\xi_{ij} = \xi_{ji}$.

Aunque no es relevante para nuestros propósitos, el ensamble de Wigner se suele definir de manera más general. A decir, las entradas diagonales pueden tener distinta distribución que las entradas no diagonales. Más aún, las entradas diagonales pueden tener varianza (finita) distinta de 1. También, podríamos distinguir el caso en que la matriz, en lugar de ser real simétrica, fuera compleja hermitiana (igual a su transpuesta conjugada).

Veamos que conociendo la concentración de ensambles con entradas independientes. (teorema 3.7) podemos dar también resultados de concentración de la norma de operador de matrices en el ensamble de Wigner.

Corolario 3.9 (Concentración de la norma de operador de matrices de Wigner). *Sea M es una matriz de $n \times n$ en el ensamble de Wigner con entradas ξ_{ij} que cumplen que $|\xi_{ij}| \leq 1$ casi seguramente. Entonces para λ suficientemente grande (mayor a una constante que no depende de n) se cumple que*

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq 2\exp(-\lambda^2 n/C^2)$$

con C una constante positiva.

Demostración. Sea M^+ la matriz cuyas entradas son iguales a las de M en su parte triangular superior (incluyendo la diagonal) y 0 debajo de la diagonal. De forma similar, sea M^- la matriz que coincide con M debajo de la diagonal y es 0 en su parte triangular superior. Por lo tanto, M^+ y M^- son matrices cuyas entradas son variables aleatorias independientes y $M = M^+ + M^-$.

Por la desigualdad del triángulo de la norma de operador tenemos que

$$\|M\|_{op} \leq \|M^+\|_{op} + \|M^-\|_{op}.$$

Por lo tanto, si $\|M\|_{op} > \lambda\sqrt{n}$ entonces al menos una de las normas de M^+ o M^- es mayor que $\lambda\sqrt{n}/2$. Para ver esto, supongamos que $\|M\|_{op} > \lambda\sqrt{n}$ y, sin pérdida de

generalidad, supongamos que $\|M^+\|_{op} \leq \lambda\sqrt{n}/2$. Entonces

$$\lambda\sqrt{n} < \|M\|_{op} \leq \|M^+\|_{op} + \|M^-\|_{op} \leq \lambda\sqrt{n}/2 + \|M^-\|_{op}$$

y, por lo tanto, $\|M^-\|_{op} > \lambda\sqrt{n}/2$.

Es decir, probamos la siguiente contención de eventos

$$\{\|M\|_{op} > \lambda\sqrt{n}\} \subseteq \{\|M^+\|_{op} > \lambda\sqrt{n}/2\} \cup \{\|M^-\|_{op} > \lambda\sqrt{n}/2\},$$

y por la subaditividad de la probabilidad tenemos que

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq \mathbb{P}(\|M^+\|_{op} > \lambda\sqrt{n}/2) + \mathbb{P}(\|M^-\|_{op} > \lambda\sqrt{n}/2).$$

Ahora, por el teorema 3.7 podemos acotar las colas de la norma de M^+ y M^- , y afirmar que si $\lambda > 14$ entonces se cumple que

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq \exp(-\lambda^2 n/4 \cdot 24) + \exp(-\lambda^2 n/4 \cdot 24).$$

Finalmente, podemos concluir que

$$\mathbb{P}(\|M\|_{op} > \lambda\sqrt{n}) \leq 2 \exp(-\lambda^2 n/96). \quad \square$$

Notemos que estos resultados de concentración son similares a los que obtuvimos cuando estudiamos la suma de variables independientes en el capítulo 2. Sin embargo, al tratarse de la norma de operador de matrices aleatorias, estamos considerando n sumas de variables aleatorias (una por fila) y, por lo tanto, tenemos una dependencia de n en el exponente. Es decir, al normalizar por \sqrt{n} obtenemos de nuevo una variable subgaussiana, pero entre más grande es n más rápido decaen las colas en función de λ .

3.2. Digresión: método de simetrización

Veamos en esta sección que en algunos casos podemos reducir el estudio de matrices aleatorias al caso particular en que sus entradas son variables aleatorias simétricas (no confundir con que la matriz sea simétrica). Esta técnica facilita, por ejemplo, estimar la esperanza de la norma de operador.

Para simplificar la notación, si M es una matriz aleatoria, denotamos como $\mathbb{E}(M)$ a la matriz que resulta de tomar la esperanza de cada una de sus entradas.

Lema 3.10. *Sea M una matriz aleatoria tal que $\mathbb{E}(M) = 0$ y \tilde{M} una copia independiente de M . Entonces se cumple que*

$$\mathbb{E}(\|M\|_{op}) \leq \mathbb{E}(\|M - \tilde{M}\|_{op}).$$

Demostración. Condicionando con M tenemos que

$$\mathbb{E}(M - \tilde{M} \mid M) = M - \mathbb{E}(\tilde{M}) = M$$

y como la norma de operador es convexa (pues todas las normas lo son) entonces por la desigualdad de Jensen para la esperanza condicional podemos afirmar que

$$\|M\|_{op} = \|\mathbb{E}(M - \tilde{M} \mid M)\|_{op} \leq \mathbb{E}(\|M - \tilde{M}\|_{op} \mid M).$$

Finalmente, podemos calcular la esperanza de ambos lados de la desigualdad para obtener que

$$\mathbb{E}(\|M\|_{op}) \leq \mathbb{E}(\|M - \tilde{M}\|_{op}). \quad \square$$

El lema anterior es de utilidad porque $M - \tilde{M}$ es una matriz aleatoria simétrica en el sentido que $M - \tilde{M} \stackrel{d}{=} \tilde{M} - M$. Así, para acotar la esperanza de la norma de operador de M es suficiente acotar la esperanza de la norma de operador de $M - \tilde{M}$, lo cual puede ser más sencillo considerando que sus entradas son variables aleatorias simétricas.

Extendiendo esta idea, podemos usar técnicas más sofisticadas para agregar hipótesis sobre nuestra matriz que podrían ser útiles en algunas situaciones. Con el siguiente ejemplo veamos que podemos elegir aleatoriamente el signo de cada entrada y el precio a pagar es solo una constante.

Proposición 3.11 (Desigualdad de simetrización). *Sea M una matriz aleatoria con entradas independientes ξ_{ij} , y E una matriz aleatoria con entradas ε_{ij} tales que $\mathbb{P}(\varepsilon_{ij} = 1) = \mathbb{P}(\varepsilon_{ij} = -1) = 1/2$. Si E y M son independientes y $\mathbb{E}(M) = 0$ entonces se cumple que*

$$\mathbb{E}(\|M\|_{op}) \leq 2 \mathbb{E}(\|M \cdot E\|_{op})$$

en donde \cdot denota el producto de Hadamard (es decir, entrada a entrada).

Demostración. Sea \tilde{M} , con entradas $\tilde{\xi}_{ij}$, una copia independiente de M . Como entrada a entrada se cumple que $(\xi_{ij} - \tilde{\xi}_{ij}) \stackrel{d}{=} (\xi_{ij} - \tilde{\xi}_{ij})\varepsilon_{ij}$ entonces en general

$$(M - \tilde{M}) \stackrel{d}{=} (M - \tilde{M}) \cdot E$$

y de la proposición 3.10 podemos concluir que

$$\mathbb{E}(\|M\|_{op}) \leq \mathbb{E}(\|M - \tilde{M}\|_{op}) = \mathbb{E}(\|(M - \tilde{M}) \cdot E\|_{op}).$$

Ahora, como el producto entrada a entrada es distributivo y por la desigualdad del triángulo de la norma de operador tenemos que

$$\|(M - \tilde{M}) \cdot E\|_{op} \leq \|M \cdot E\|_{op} + \|\tilde{M} \cdot E\|_{op}$$

y como $M \cdot E \stackrel{d}{=} \tilde{M} \cdot E$ entonces finalmente concluimos que

$$\mathbb{E}(\|M\|_{op}) \leq 2 \mathbb{E}(\|M \cdot E\|_{op}). \quad \square$$

Para terminar esta sección veamos que, de manera similar, podemos multiplicar cada entrada por una variable aleatoria normal y el precio a pagar es también una

constante.

Proposición 3.12 (Desigualdad de simetrización gaussiana). *Sea M una matriz aleatoria con entradas ξ_{ij} con $\mathbb{E}(M) = 0$ y G una matriz aleatoria con entradas g_{ij} tales que $g_{ij} \sim \mathcal{N}(0, 1)$. Entonces se cumple que*

$$\mathbb{E}(\|M\|_{op}) \leq \sqrt{2\pi} \mathbb{E}(\|M \cdot G\|_{op}).$$

Demostración. Notemos como una variable aleatoria normal es simétrica entonces la podemos ver como su signo por su magnitud. Es decir, si consideramos matrices E y $|G|$ donde E tiene entradas ε_{ij} con $\mathbb{P}(\varepsilon_{ij} = 1) = \mathbb{P}(\varepsilon_{ij} = -1) = 1/2$ y $|G|$ tiene entradas $|g_{ij}|$ entonces podemos concluir que $G \stackrel{d}{=} E \cdot |G|$. Además E y $|G|$ son independientes y, por lo tanto,

$$M \cdot G \stackrel{d}{=} M \cdot E \cdot |G|.$$

Ahora, condicionando con M y E obtenemos que

$$\mathbb{E}(M \cdot E \cdot |G| \mid M, E) = \mathbb{E}(|G|) \cdot M \cdot E = \sqrt{2/\pi} (M \cdot E)$$

pues $|G|$ tiene entradas que se distribuyen como el valor absoluto de una variable normal estándar y podemos calcular explícitamente que su esperanza es $\sqrt{2/\pi}$. Por la desigualdad de Jensen para la esperanza condicional podemos afirmar que

$$\sqrt{2/\pi} \|M \cdot E\|_{op} = \|\mathbb{E}(M \cdot E \cdot |G| \mid M, E)\|_{op} \leq \mathbb{E}(\|M \cdot E \cdot |G|\|_{op} \mid M, E).$$

Ahora, calculando la esperanza de ambos lados de la desigualdad obtenemos que

$$\sqrt{2/\pi} \mathbb{E}(\|M \cdot E\|_{op}) \leq \mathbb{E}(\|M \cdot E \cdot |G|\|_{op}) = \mathbb{E}(\|M \cdot G\|_{op}).$$

Finalmente, por la proposición 3.11 podemos concluir que

$$\mathbb{E}(\|M\|_{op}) \leq 2 \mathbb{E}(\|M \cdot E\|_{op}) \leq 2\sqrt{\pi/2} \mathbb{E}(\|M \cdot G\|_{op}). \quad \square$$

Con un poco más de trabajo, estas tres proposiciones se pueden usar para acotar la esperanza de la norma de operador de matrices aleatorias. La clave es aprovechar estas proposiciones para usar resultados de concentración específicamente de variables aleatorias gaussianas y después usar un argumento de ε -redes, como lo hicimos en la sección anterior. Los detalles pueden consultarse también en [Lat05].

Capítulo 4

Ley del semicírculo de Wigner

El objetivo principal de este capítulo es ver cuál es el comportamiento asintótico del espectro de matrices en el ensamble de Wigner. Es decir, matrices cuyas entradas son variables aleatorias ξ_{ij} que cumplen que $\xi_{ij} = \xi_{ji}$. Véase la definición 3.8. Podemos analizar qué pasa con sus valores propios cuando la dimensión de la matriz crece a infinito.

Para poder estudiar rigurosamente el espectro de una matriz de dimensión $n \times n$, no necesariamente aleatoria, definimos una medida de probabilidad con n átomos que representan sus valores propios. Podemos pensar, naturalmente, en darle el mismo peso, $1/n$, a cada uno de sus valores propios.

Definición 4.1 (Distribución espectral empírica). *Sea M una matriz simétrica de dimensión $n \times n$ y supongamos que sus valores propios, necesariamente reales, son $\lambda_1(M) \leq \lambda_2(M) \leq \dots \leq \lambda_n(M)$. Entonces definimos su distribución espectral empírica (DEE) como la medida de probabilidad*

$$\mu_M = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(M)}.$$

En nuestro análisis las entradas de la matriz simétrica M son variables aleatorias y, por lo tanto, su DEE es una medida aleatoria (una variable aleatoria en el espacio de

medidas de probabilidad de \mathbb{R}). Así, los momentos de la DEE son variables aleatorias reales.

En este capítulo, el objetivo principal es demostrar que cuando la dimensión de la matriz tiende a infinito su DEE converge a una medida de probabilidad determinista. Wigner demostró en su artículo de 1955 [Wig55] que la medida de probabilidad determinista a la que converge la DEE es la ley del semicírculo, cuya densidad está definida por un semicírculo de -2 a 2 normalizado para integrar 1.

Definición 4.2 (Ley del semicírculo). *Una variable aleatoria X tiene distribución semicircular si su función de densidad está dada por*

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

Denotamos como σ a la ley de una variable semicircular.

Notemos que, por el factor de normalización, la gráfica de la densidad de una variable semicircular es en realidad media elipse.

Para demostrar la convergencia casi segura de la DEE, que es en un sentido débil de medidas, usamos el método de los momentos. Es decir, demostramos que los momentos de la DEE convergen casi seguramente a los momentos de la ley del semicírculo. Al final del capítulo damos una justificación rigurosa de por qué esto y el hecho de que la densidad de la ley del semicírculo tiene soporte compacto implican la convergencia casi segura (en el sentido débil de medidas) de la DEE a la ley del semicírculo.

4.1. Momentos de la ley del semicírculo

Veamos primero cuáles son los momentos de una variable aleatoria con distribución semicircular. Como la función de densidad de la distribución semicircular es simétrica entonces los momentos impares son 0 y los pares, como vemos en seguida, resultan ser los números de Catalan. La presencia de esta sucesión, que vimos por primera vez en la sección 1.2 al estudiar árboles planos, es lo que permite usar técnicas combinatorias

al analizar el comportamiento asintótico de la DEE.

Proposición 4.3 (Momentos de la ley del semicírculo). *Si X es una variable aleatoria con distribución semicircular entonces para toda $k \in \mathbb{N}$ se cumple que*

$$\mathbb{E}(X^{2k}) = C_k \quad \text{y} \quad \mathbb{E}(X^{2k+1}) = 0,$$

donde C_k es el k -ésimo número de Catalan, es decir, $C_k = \frac{1}{k+1} \binom{2k}{k}$.

Demostración. Sea $\sigma(x)$ la función de densidad de X , una variable con distribución semicircular. Notemos que como $\sigma(x)$ es una función par entonces $x^{2k+1}\sigma(x)$ es una función impar y, por lo tanto, los momentos impares de X son 0.

Para los momentos pares notemos que como $x^{2k}\sigma(x)$ es par entonces

$$\int_{-\infty}^{\infty} x^{2k} \sigma(x) dx = \frac{1}{2\pi} \int_{-2}^2 x^{2k} \sqrt{4-x^2} dx = \frac{1}{\pi} \int_0^2 x^{2k} \sqrt{4-x^2} dx.$$

Haciendo el cambio de variable $u = (x/2)^2$ obtenemos que

$$\frac{1}{\pi} \int_0^2 x^{2k} \sqrt{4-x^2} dx = \frac{2^{2k+1}}{\pi} \int_0^1 u^{k-1/2} (1-u)^{1/2} du = \frac{2^{2k+1}}{\pi} B\left(k + \frac{1}{2}, \frac{3}{2}\right)$$

donde B es la función beta. Ahora, aludiendo a la relación entre la función beta y la función gamma tenemos que

$$B\left(k + \frac{1}{2}, \frac{3}{2}\right) = \frac{\Gamma(k + \frac{1}{2}) \Gamma(\frac{3}{2})}{\Gamma(k+2)}. \quad (4.1)$$

Para simplificar el término $\Gamma(k + \frac{1}{2})$ usamos la propiedad de la función gamma

$\Gamma(z+1) = z\Gamma(z)$. Aplicándola iteradamente obtenemos que

$$\begin{aligned} \Gamma\left(k + \frac{1}{2}\right) &= \left(\frac{2k-1}{2}\right) \left(\frac{2k-3}{2}\right) \cdots \left(\frac{3}{2}\right) \left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right) \\ &= \frac{1}{2^k} (2k-1)(2k-3) \cdots 3 \Gamma\left(\frac{1}{2}\right). \end{aligned} \quad (4.2)$$

Podemos multiplicar ambos lados de la igualdad 4.2 por $(2k)(2k-2)(2k-4)\cdots 2$. Del lado derecho tenemos un $(2k)!$ y del lado izquierdo podemos factorizar un 2 de cada término y obtener $2^k k!$. Es decir,

$$2^k k! \Gamma(k + \tfrac{1}{2}) = \frac{1}{2^k} (2k)! \Gamma(\tfrac{1}{2}).$$

Reordenando los términos concluimos que

$$\Gamma(k + \tfrac{1}{2}) = \frac{1}{2^{2k}} \frac{(2k)!}{k!} \Gamma(\tfrac{1}{2}).$$

Ahora, para simplificar la ecuación 4.1 solo falta sustituir lo que calculamos y usar que $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{1}{2}\sqrt{\pi}$ y que $\Gamma(k+2) = (k+1)!$. Así obtenemos que

$$B(k + \tfrac{1}{2}, \tfrac{3}{2}) = \frac{(2k)! \pi}{2^{2k+1} k! (k+1)!}$$

y, finalmente,

$$\int_{-\infty}^{\infty} x^{2k} \sigma(x) dx = \frac{(2k)!}{k! (k+1)!} = \frac{1}{k+1} \binom{2k}{k} = C_k. \quad \square$$

4.2. Momentos de la distribución espectral empírica

Ahora nos enfocamos en los momentos de las distribuciones espectrales empíricas de una sucesión de matrices de Wigner cuya dimensión tiende a infinito. En esta sección demostramos que los momentos de la DEE convergen a los momentos de la ley del semicírculo que calculamos en la sección anterior. Esto es la esencia del método de los momentos. El análisis que presentamos a continuación es similar al que presenta Wigner en [Wig55]. Sin embargo, hacemos uso explícito del lenguaje de la teoría de gráficas, lo que hace más evidentes y claros a los aspectos combinatorios que se aprovechan en la demostración.

Como vimos en el capítulo 3, la norma de operador de una matriz de Wigner de dimensión $n \times n$, y por lo tanto su espectro, se concentra considerablemente en un intervalo de tamaño $O(\sqrt{n})$. Por esta razón al normalizar a las matrices por $1/\sqrt{n}$ podemos comparar sus DEE independientemente de su dimensión.

Consideremos entonces una sucesión $\overline{M}_n = \frac{1}{\sqrt{n}} M_n$ que cumpla que M_n es una matriz de Wigner de dimensión $n \times n$. De ahora en adelante suponemos que las entradas de las matrices, ξ_{ij} , tienen todos sus momentos finitos, es decir, cumplen que $\mathbb{E}(\xi_{ij}^k) < \infty$ para cada $k \in \mathbb{N}$. Recordemos además que de la definición del ensamble de Wigner en 3.8 tenemos que $\mathbb{E}(\xi_{ij}) = 0$ y $\text{Var}(\xi_{ij}) = 1$.

En la versión más general de este teorema no se requiere la finitud de todos los momentos y la varianza en las entradas diagonales puede ser distinta de 1. Véase [BCG22].

Para hacer más sencilla la notación, denotamos como $m_{n,k}$ al k -ésimo momento de la DEE de la n -ésima matriz de la sucesión.

Veamos ahora qué forma tienen los momentos de la DEE (que, recordemos, son variables aleatorias reales). Por su definición como suma de medidas de Dirac tenemos

$$m_{n,k} := \mathbb{E}_{\mu_{\overline{M}_n}}(x^k) = \sum_{i=1}^n \frac{1}{n} (\lambda_i(\overline{M}_n))^k.$$

Uno puede probar por inducción que si $\lambda_i(\overline{M}_n)$ es un valor propio de la matriz \overline{M}_n entonces $(\lambda_i(\overline{M}_n))^k$ es un valor propio de la matriz \overline{M}_n^k . Por lo tanto, podemos expresar los momentos de la DEE como trazas de potencias de las matrices \overline{M}_n . Es decir,

$$m_{n,k} = \frac{1}{n} \text{tr}(\overline{M}_n^k).$$

Ahora, recordando que $\overline{M}_n = \frac{1}{\sqrt{n}} M_n$ y calculando explícitamente la traza de M_n^k , podemos expresar $m_{n,k}$ en términos de las entradas de la matriz (las variables ξ_{ij}) de

la siguiente manera:

$$m_{n,k} = \frac{1}{n^{k/2+1}} \operatorname{tr} (M_n^k) = \frac{1}{n^{k/2+1}} \sum_{1 \leq i_1, \dots, i_k \leq n} \xi_{i_1 i_2} \xi_{i_2 i_3} \cdots \xi_{i_k i_1}.$$

Convergencia en esperanza de los momentos de la DEE

Para ver a dónde convergen los momentos $m_{n,k}$ cuando $n \rightarrow \infty$ primero estudiamos su esperanza. Es importante notar que la esperanza es respecto a la medida de la matriz \overline{M}_n . Así, todas las veces que en este capítulo mencionemos una esperanza, varianza o covarianza es respecto a la medida de la matriz \overline{M}_n . Para no cargar demasiado la notación, no lo denotamos explícitamente.

Por la linealidad de la esperanza, podemos expresar la esperanza de los momentos como

$$\mathbb{E}(m_{n,k}) = \frac{1}{n^{k/2+1}} \sum_{1 \leq i_1, \dots, i_k \leq n} \mathbb{E}(\xi_{i_1 i_2} \xi_{i_2 i_3} \cdots \xi_{i_k i_1}). \quad (4.3)$$

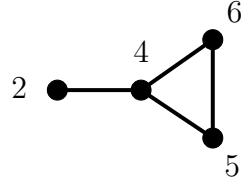
La clave de la demostración es traducir este análisis al lenguaje de gráficas. Para esto, a cada palabra $I = (i_1, i_2, \dots, i_k)$ que representa al término $\xi_{i_1 i_2} \xi_{i_2 i_3} \cdots \xi_{i_k i_1}$ le podemos asociar una gráfica G_I de la siguiente manera: por cada índice distinto que aparece en la palabra I hay un vértice y hay una arista entre dos de ellos si aparecen consecutivamente en I . Las gráficas podrían tener lazos (si aparecen variables de la forma ξ_{ii}) pero no aristas múltiples. Para un ejemplo véase la figura 4.1.

Hay que hacer énfasis en que cada palabra I define una caminata cerrada en G_I siguiendo el orden en que aparece cada índice en la palabra. Así en (4.3) al calcular la esperanza de los momentos una suma sobre todas las posibles palabras de longitud k se convierte en una suma sobre todas las posibles caminatas cerradas de longitud k sobre gráficas con a lo más k vértices y cuyos vértices tienen etiquetas en el conjunto $\{1, \dots, n\}$.

$$I = (2, 4, 5, 6, 4)$$

$$\xi_I = \xi_{24}\xi_{45}\xi_{56}\xi_{64}\xi_{42}$$

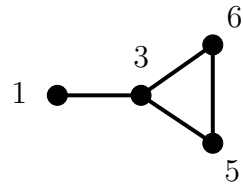
$G_I :$



$$J = (1, 3, 5, 6, 3)$$

$$\xi_J = \xi_{13}\xi_{35}\xi_{56}\xi_{63}\xi_{31}$$

$G_J :$



$$K = (3, 1, 3, 6, 5)$$

$$\xi_K = \xi_{31}\xi_{13}\xi_{36}\xi_{65}\xi_{53}$$

$G_K :$

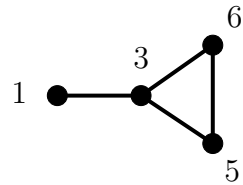


Figura 4.1: Palabras de longitud $k = 5$ y sus gráficas asociadas. Notemos que I y J definen a la misma gráfica pero con diferentes etiquetas. Notemos también que J y K definen exactamente a la misma gráfica aunque son palabras distintas.

De ahora en adelante, si $I = (i_1, i_2, \dots, i_k)$ entonces denotamos como ξ_I al producto $\xi_{i_1 i_2} \xi_{i_2 i_3} \dots \xi_{i_k i_1}$. De esta manera, para cada palabra I tenemos una expresión de la esperanza de ξ_I en términos de cuántas veces se recorre cada arista en la caminata definida por I . Es decir, si α_{ij} es el número de veces que la caminata pasa por la arista ij entonces

$$\mathbb{E}(\xi_I) = \prod_{ij} \mathbb{E}(\xi_{ij}^{\alpha_{ij}})$$

donde el producto corre sobre todas las aristas de G_I . Notemos que estamos usando la hipótesis de que la matriz es simétrica al no distinguir la dirección de las aristas y la hipótesis de independencia para ver la esperanza del producto como el producto de las esperanzas.

Si la caminata que define I en G_I recorre alguna arista ij una sola vez, es decir $\alpha_{ij} = 1$, entonces $\mathbb{E}(\xi_I) = 0$ pues, por hipótesis, las entradas de la matriz \overline{M}_n tienen esperanza 0. Por esta razón, de ahora en adelante solo consideramos caminatas en las que cada arista se recorra al menos dos veces. Recordemos que cuando estudiamos sumas de variables aleatorias en la sección 2.2 teníamos una situación casi análoga.

Veamos en seguida que suponer esta condición va a restringir el número de vértices en las gráficas que consideramos.

Lema 4.4 (Cota del número de vértices en una gráfica). *Sea $|V(G_I)|$ el número de vértices en la gráfica G_I . Si la caminata definida por I recorre cada arista al menos dos veces entonces $|V(G_I)| \leq k/2 + 1$.*

Demostración. Notemos que la caminata que define I en G_I contiene a todos sus vértices y por lo tanto G_I es una gráfica conexa. Así, si denotamos como $|E(G_I)|$ al número de aristas de G_I podemos afirmar que $|V(G_I)| - 1 \leq |E(G_I)|$.

Ahora, como cada arista se cruza al menos dos veces en la caminata de longitud k entonces $2|E(G_I)| \leq k$.

Juntando las dos desigualdades obtenemos que $2(|V(G_I)| - 1) \leq 2|E(G_I)| \leq k$ y, por lo tanto, $|V(G_I)| \leq k/2 + 1$. \square

La clave de lo que sigue es notar que la única dependencia en n al calcular (4.3) es el número de maneras de etiquetar los vértices. Así, podemos demostrar que las únicas gráficas que aportan a la esperanza cuando $n \rightarrow \infty$ son aquellas que alcanzan la igualdad en el lema anterior. Lo presentamos en seguida con todo detalle.

Lema 4.5 (Aportación de las gráficas con pocos vértices). *Sea \mathcal{H}_r el conjunto de palabras cuyas gráficas asociadas tienen exactamente r vértices. Si $r < k/2 + 1$ entonces la aportación de las palabras en \mathcal{H}_r a la esperanza de $m_{n,k}$ es despreciable en el límite, es decir,*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{k/2+1}} \sum_{I \in \mathcal{H}_r} \mathbb{E}(\xi_I) = 0.$$

Demostración. Sea $r < k/2 + 1$. Consideremos $I \in \mathcal{H}_r$ y consideremos a los vértices de G_I con algunas etiquetas fijas. Entonces el número de caminatas de longitud k sobre G_I es una función de k y por lo tanto, una constante. Ahora, como estamos asumiendo que todos los momentos de las variables aleatorias ξ_{ij} son finitos entonces la contribución a la esperanza de cada caminata también es una constante.

Es decir, la única dependencia en n viene de considerar cuántas maneras hay de etiquetar a los vértices de G_I con etiquetas en el conjunto $\{1, \dots, n\}$. Para un primer vértice hay n opciones, para un segundo hay $n - 1$ y así hasta llegar al último. Por lo tanto, hay

$$n(n-1) \cdots (n-(r-1)) = O(n^r)$$

maneras de etiquetar los vértices de G_I . Como por hipótesis $r < k/2 + 1$ entonces podemos asegurar que las maneras de etiquetar los vértices de G_I son $o(n^{k/2+1})$. Por lo tanto, concluimos que

$$\lim_{n \rightarrow \infty} \frac{1}{n^{k/2+1}} \sum_{I \in \mathcal{H}_r} \mathbb{E}(\xi_I) = \lim_{n \rightarrow \infty} f(k) \frac{o(n^{k/2+1})}{n^{k/2+1}} = 0$$

donde $f(k)$ es una función que solo depende de k . □

Con este resultado ya podemos concluir cuál es el comportamiento asintótico de la

esperanza de los momentos impares de la DEE.

Corolario 4.6 (Esperanza de los momentos impares de la DEE). *Si k es un entero impar entonces $\lim_{n \rightarrow \infty} \mathbb{E}(m_{n,k}) = 0$.*

Demostración. Si k es impar entonces $k/2 + 1$ no es un entero y por lo tanto, no existe una palabra I tal que G_I tenga exactamente $k/2 + 1$ vértices. Se sigue entonces del lema 4.5 que $\lim_{n \rightarrow \infty} \mathbb{E}(m_{n,k}) = 0$. \square

Lo que resta ahora es, cuando k es par, analizar las gráficas que tengan exactamente $k/2 + 1$ vértices. Si esto ocurre entonces se alcanza la igualdad en las dos desigualdades de la demostración del lema 4.4. Es decir, en las caminatas definidas por las palabras, cada arista se recorre exactamente dos veces y la cantidad de aristas que tiene es exactamente una más que la cantidad de vértices. Como además la gráfica es conexa entonces podemos asegurar que es un árbol.

Notemos además que la caminata que define I en G_I está en realidad explorando el contorno del árbol y, por la biyección dada en la sección 1.2, lo define únicamente como árbol plano. Es decir, sin considerar etiquetas, hay tantas caminatas de este tipo como árboles planos con $k/2 + 1$ vértices.

Veamos ahora específicamente cuál es la aportación de estos términos a la esperanza del k -ésimo momento de la DEE.

Lema 4.7 (Aportación de las gráficas con el máximo número de vértices). *Sea k un entero par y $\mathcal{H}_{k/2+1}$ el conjunto de palabras cuyas gráficas asociadas tienen exactamente $k/2 + 1$ vértices. Entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{k/2+1}} \sum_{I \in \mathcal{H}_{k/2+1}} \mathbb{E}(\xi_I) = C_{k/2}$$

donde $C_{k/2}$ es el $k/2$ -ésimo número de Catalan.

Demostración. Sea I una palabra tal que su gráfica asociada G_I tiene exactamente $k/2 + 1$ vértices. Por lo tanto, en la caminata que define I en G_I cada arista se cruza

exactamente 2 veces. Así tenemos que

$$\mathbb{E}(\xi_I) = \prod_{ij} \mathbb{E}(\xi_{ij}^2) = 1$$

donde el producto es sobre las aristas de G_I , y es igual a 1 por la hipótesis de que las variables aleatorias ξ_{ij} tiene varianza 1.

Por lo tanto, la suma de las esperanzas de las variables ξ_I es igual a la cardinalidad del conjunto $\mathcal{H}_{k/2+1}$. Recordemos que hay tantas palabras cuyas gráficas asociadas tienen $k/2 + 1$ vértices como árboles planos de $k/2 + 1$ vértices etiquetados con el conjunto $\{1, \dots, n\}$. En la sección 1.2 demostramos que el número de árboles planos con $k/2 + 1$ vértices es $C_{k/2}$. Por lo tanto, solo falta contar de cuántas maneras podemos etiquetar sus vértices. Para un primer vértice tenemos n opciones de etiqueta, para un segundo tenemos $n - 1$ y así sucesivamente. Por lo que podemos concluir que hay

$$n(n - 1) \cdots (n - k/2) = n^{k/2+1}(1 + o(1))$$

maneras de etiquetar a un árbol plano con $k/2 + 1$ vértices. Por lo tanto,

$$\sum_{I \in \mathcal{H}_{k/2+1}} \mathbb{E}(\xi_I) = |\mathcal{H}_{k/2+1}| = C_{k/2} n^{k/2+1} (1 + o(1)).$$

Concluimos entonces que

$$\lim_{n \rightarrow \infty} \frac{1}{n^{k/2+1}} \sum_{I \in \mathcal{H}_{k/2+1}} \mathbb{E}(\xi_I) = \lim_{n \rightarrow \infty} C_{k/2} (1 + o(1)) = C_{k/2}. \quad \square$$

Finalmente tenemos que, en el límite, la esperanza de los momentos pares es igual a los números de Catalan.

A modo de resumen, en el siguiente teorema enunciamos lo que demostramos.

Teorema 4.8 (Convergencia de la esperanza de los momentos de la DEE). *Sea $\{M_n\}_{n \in \mathbb{N}}$ una sucesión de matrices de Wigner donde M_n es de dimensión $n \times n$*

y $m_{n,k}$ el k -ésimo momento de la DEE de la matriz $\frac{1}{\sqrt{n}}M_n$. Entonces

$$\lim_{n \rightarrow \infty} \mathbb{E}(m_{n,k}) = \begin{cases} C_{k/2} & \text{si } k \text{ es par,} \\ 0 & \text{si } k \text{ es impar.} \end{cases}$$

Varianza de los momentos de la DEE

Para lograr demostrar la convergencia de los momentos de la DEE, y no solo de su esperanza, veamos que son variables aleatorias muy concentradas alrededor de sus medias. Para esto, primero notemos que podemos expresar su varianza en términos de las covarianzas de las variables ξ_I . Es decir,

$$\text{Var}(m_{n,k}) = \text{Cov}\left(\frac{1}{n^{k/2+1}} \sum_I \xi_I, \frac{1}{n^{k/2+1}} \sum_J \xi_J\right) = \frac{1}{n^{k+2}} \sum_I \sum_J \text{Cov}(\xi_I, \xi_J)$$

donde las sumas recorren todas las palabras del conjunto $\{1, \dots, n\}$ con longitud k .

Podemos proceder de manera similar a cuando analizamos su esperanza. Para cada sumando de la varianza, digamos $\text{Cov}(\xi_I, \xi_J)$, consideramos la gráfica que resulta de unir a las gráficas G_I y G_J . Es decir, la gráfica $G_{I \cup J}$ tiene como vértices a la unión de los vértices de ambas gráficas y hay una arista entre dos de ellos si esa arista ya estaba en G_I o si ya estaba en G_J . La gráfica $G_{I \cup J}$ puede tener lazos pero no aristas múltiples.

El análisis siguiente es casi análogo a cuando analizamos la esperanza. Como una primera observación notemos que si G_I y G_J no comparten aristas entonces las variables ξ_I y ξ_J son independientes y, por lo tanto, $\text{Cov}(\xi_I, \xi_J) = 0$.

Además notemos que las caminatas que definen I y J sobre G_I y G_J , respectivamente, definen una caminata cerrada en $G_{I \cup J}$ de la siguiente manera: se recorre la caminata definida por I en G_I hasta encontrar un vértice de G_J por primera vez. Después se recorre la caminata cerrada definida por J hasta regresar a ese vértice. En seguida,

se reanuda la caminata I hasta regresar al vértice donde se inició. Nos referimos a esta caminata como la concatenación de las caminatas definidas por I y J . Para un ejemplo la figura 4.2.

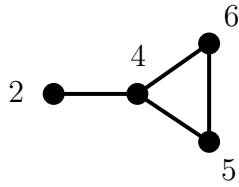
$$I = (2, 4, 5, 6, 4)$$

$$J = (3, 6, 5, 3, 1)$$

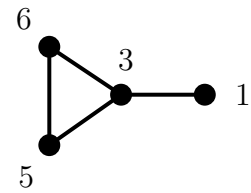
$$\xi_I = \xi_{24}\xi_{45}\xi_{56}\xi_{64}\xi_{42}$$

$$\xi_J = \xi_{36}\xi_{65}\xi_{53}\xi_{31}\xi_{13}$$

$G_I :$



$G_J :$



$G_{I \cup J} :$

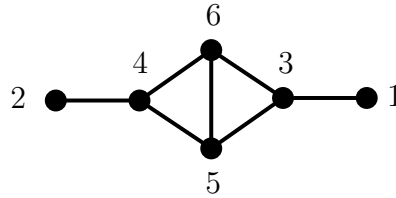


Figura 4.2: Ejemplo de una gráfica como la unión de dos. Como las gráficas G_I y G_J tienen a la arista 56 en común entonces existe una caminata cerrada en $G_{I \cup J}$. En este caso, la concatenación de I y J define a la caminata 24531365642 en $G_{I \cup J}$.

La covarianza también es 0 si la caminata en $G_{I \cup J}$ definida por la concatenación de I y J recorre alguna arista una única vez. Estas dos condiciones nos dan una restricción sobre el número de vértices en $G_{I \cup J}$.

Lema 4.9 (Cota del número de vértices en una unión de gráficas). *Sea $|V(G_{I \cup J})|$ el número de vértices de la gráfica $G_{I \cup J}$. Si G_I y G_J comparten al menos una arista y la caminata definida por la concatenación de I y J recorre cada arista al menos dos veces entonces*

$$|V(G_{I \cup J})| \leq k + 1.$$

Demostración. Notemos que las caminatas que definen I y J en sus respectivas gráficas G_I y G_J son de longitud k (contando multiplicidades). Por lo tanto, la caminata definida por la concatenación será de longitud $2k$ (contando multiplicidades). Así, como cada arista debe recorrerse al menos dos veces entonces, denotando $|E(G_{I \cup J})|$ al número de aristas en $G_{I \cup J}$, se cumple que $|E(G_{I \cup J})| \leq k$.

Ahora, como G_I y G_J son conexas y además comparten al menos una arista entonces existe una caminata cerrada en $G_{I \cup J}$ (la definida por la concatenación) y, por lo tanto, es conexa. Así, tenemos que $|V(G_{I \cup J})| - 1 \leq |E(G_{I \cup J})|$.

Juntando las dos desigualdades obtenemos que $|V(G_{I \cup J})| \leq k + 1$. \square

Proposición 4.10 (Varianza asintótica de los momentos de la DEE). *Para toda $k \in \mathbb{N}$ se cumple que*

$$\lim_{n \rightarrow \infty} \text{Var}(m_{n,k}) = 0.$$

Demostración. Consideremos palabras I y J de longitud k . Supongamos que G_I y G_J tienen al menos una arista en común y que las caminatas que definen I y J en G_I y G_J , respectivamente, recorren cada arista al menos dos veces. El número de palabras que cumplen esto es una función de k y el número de caminatas de longitud k en G_I y G_J también es una función de k . Además, la aportación a la varianza de cada una de estas caminatas concatenadas es una constante, pues por hipótesis las variables ξ_{ij} tienen momentos finitos. Por lo tanto, la única dependencia en n es la

cantidad de maneras en las que se pueden etiquetar los vértices de $G_{I \cup J}$ con etiquetas en $\{1, \dots, n\}$. Por el lema 4.9 sabemos que el número de vértices en $G_{I \cup J}$ es a lo más $k + 1$ y por lo tanto, hay $O(n^{k+1})$ maneras de etiquetar sus vértices. Así, recordando que $\text{Var}(m_{n,k}) = \frac{1}{n^{k+2}} \sum_I \sum_J \text{Cov}(\xi_I, \xi_J)$, podemos concluir que

$$\lim_{n \rightarrow \infty} \text{Var}(m_{n,k}) = \lim_{n \rightarrow \infty} f(k) \frac{O(n^{k+1})}{n^{k+2}} = 0$$

donde $f(k)$ es una función que solo depende de k . □

Notemos que en realidad demostramos algo un poco más fuerte pues podemos asegurar que la varianza de $m_{n,k}$ se va a 0 al menos tan rápido como $1/n$. Sin embargo, en seguida argumentamos que el orden de magnitud correcto es en realidad $1/n^2$.

Proposición 4.11 (Orden de magnitud de la varianza de los momentos de la DEE).

Para toda $k \in \mathbb{N}$ se cumple que

$$\text{Var}(m_{n,k}) = O(1/n^2).$$

Demostración. Veamos que si la gráfica definida por un par de palabras I y J tiene exactamente $k + 1$ vértices entonces $\text{Cov}(\xi_I, \xi_J) = 0$. Esto implica que las gráficas de las palabras que sí aportan a la covarianza tienen a lo más k vértices y, recordando la demostración de la proposición 4.10, podemos concluir que

$$\text{Var}(m_{n,k}) = f(k) \frac{O(n^k)}{n^{k+2}} = O(1/n^2)$$

donde $f(k)$ es una función que solo depende de k .

Supongamos que existen palabras I y J tales que la gráfica, $G_{I \cup J}$, tiene exactamente $k + 1$ vértices. Recordando la demostración de la proposición 4.9, esto implica que $G_{I \cup J}$ tiene exactamente k aristas y, por lo tanto, es un árbol. Como G_I y G_J son gráficas conexas podemos concluir que también son árboles.

Recordemos que las palabras I y J definen caminatas cerradas en G_I y G_J , respectiva-

mente, en donde cada arista se recorre exactamente dos veces. Ahora, la concatenación $I \cup J$ define una caminata cerrada en $G_{I \cup J}$ que recorre cada una de sus arista exactamente dos veces. Por lo tanto, no puede ocurrir que G_I y G_J compartan alguna arista, pues si lo hicieran, habría una arista en $G_{I \cup J}$ que se recorre al menos cuatro veces. Así, concluimos que las variables aleatorias ξ_I y ξ_J son independientes y, por lo tanto, $\text{Cov}(\xi_I, \xi_J) = 0$. \square

Sabiendo ya que los momentos de la DEE son variables aleatorias muy concentradas alrededor de sus medias, podemos usar la desigualdad de Chebyshev para concluir que convergen en probabilidad a los momentos del semicírculo.

En las siguientes proposiciones analizamos el caso de los momentos pares. Para los momentos impares el análisis es análogo cambiando el respectivo número de Catalan por 0.

Proposición 4.12 (Convergencia en probabilidad de los momentos de la DEE). *Para todo $k \in \mathbb{N}$ par y para toda $\varepsilon > 0$ se cumple que*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) = 0.$$

Demostración. Sea $\varepsilon > 0$. Por 4.8 sabemos que $\mathbb{E}(m_{n,k}) \rightarrow C_{k/2}$ y entonces podemos afirmar que existe una N tal que si $n > N$ entonces $|\mathbb{E}(m_{n,k}) - C_{k/2}| < \varepsilon/2$.

Además, notemos que si $|m_{n,k} - C_{k/2}| > \varepsilon$ y $|\mathbb{E}(m_{n,k}) - C_{k/2}| < \varepsilon/2$ entonces podemos asegurar que $|m_{n,k} - \mathbb{E}(m_{n,k})| > \varepsilon/2$ pues de la desigualdad del triángulo se sigue que

$$\begin{aligned} \varepsilon &< |m_{n,k} - C_{k/2}| \\ &\leq |m_{n,k} - \mathbb{E}(m_{n,k})| + |\mathbb{E}(m_{n,k}) - C_{k/2}| \\ &< |m_{n,k} - \mathbb{E}(m_{n,k})| + \varepsilon/2. \end{aligned}$$

Por lo tanto, si $n > N$ entonces

$$\mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) \leq \mathbb{P}(|m_{n,k} - \mathbb{E}(m_{n,k})| > \varepsilon/2)$$

y así, por la desigualdad de Chebyshev y la proposición 4.10, podemos concluir que cuando $n \rightarrow \infty$

$$\mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) \leq \frac{4}{\varepsilon^2} \text{Var}(m_{n,k}) \rightarrow 0. \quad \square$$

Si conocemos el orden de magnitud al que la varianza de los momentos tiende a 0 entonces podemos decir incluso más. Usando el hecho de que $\text{Var}(m_{n,k}) = O(1/n^2)$ y el lema de Borel-Cantelli, que en seguida enunciamos sin demostrar, podemos garantizar que la convergencia de los momentos de la DEE a los momentos del semicírculo es casi segura.

Lema 4.13 (Borel-Cantelli). *Sea $\{E_i\}_{i \in \mathbb{N}}$ una sucesión de eventos que cumple que $\sum_{i=0}^{\infty} \mathbb{P}(E_i) < \infty$. Entonces con probabilidad 1 solo ocurre una cantidad finita de ellos.*

Antes de proceder, hay que notar que cada $m_{n,k}$ es una función de la matriz M_n y todas esas matrices viven, en principio, en espacios de probabilidad distintos. Por esta razón no tiene sentido hablar de la convergencia casi segura de los momentos. Sin embargo, podemos darle sentido si acoplamos a las matrices M_n de la siguiente manera: consideremos un arreglo de variables aleatorias ξ_{ij} con $i, j \in \mathbb{N}$ y tal que $\xi_{ij} = \xi_{ji}$. Definimos a la matriz M_n como la matriz cuyas entradas son a las primeras $n \times n$ variables aleatorias de este arreglo infinito. Podemos pensar que tenemos una matriz aleatoria simétrica infinita y cada matriz M_n es la matriz que se obtiene de solo considerar la primera porción de ella. Podemos visualizar este acoplamiento como si la sucesión de matrices estuviera, de cierta manera, anidada. Véase la figura 4.3.

Una vez más, analizamos el caso de los momentos pares. Para los momentos impares es análogo cambiando el respectivo número de Catalan por 0.

Proposición 4.14 (Convergencia casi segura de los momentos de la DEE). *Para todo $k \in \mathbb{N}$ par se cumple que*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} m_{n,k} = C_{k/2}\right) = 1.$$

Demostración. Notemos que para demostrar que $m_{n,k}$ converge casi seguramente a

El diagrama muestra una matriz infinita de elementos ξ_{ij} organizada en filas y columnas. Las primeras cuatro filas y columnas están agrupadas en submatrices destacadas con bordes de colores:

- M_1 (rojo): el elemento ξ_{11} .
- M_2 (azul): los elementos $\xi_{12}, \xi_{21}, \xi_{22}$.
- M_3 (verde): los elementos $\xi_{13}, \xi_{23}, \xi_{31}, \xi_{32}, \xi_{33}$.
- M_4 (amarillo): los elementos $\xi_{14}, \xi_{24}, \xi_{34}, \xi_{41}, \xi_{42}, \xi_{43}, \xi_{44}$.

 Las filas y columnas continúan con elementos $\xi_{15}, \xi_{25}, \xi_{35}, \xi_{45}, \xi_{55}, \dots$ y los elementos de la diagonal inferior son $\vdots, \vdots, \vdots, \vdots, \vdots, \ddots$. A la derecha de la matriz, se listan las submatrices con sus respectivos colores y dimensiones:

- M_1 (1×1)
- M_2 (2×2)
- M_3 (3×3)
- M_4 (4×4)

Figura 4.3: Las cuatro primeras matrices en el acoplamiento de las matrices M_n .

$C_{k/2}$ es suficiente verificar que con probabilidad 1 para toda $\varepsilon > 0$ existe solo un número finito de índices n tales que $|m_{n,k} - C_{k/2}| > \varepsilon$.

Por el lema de Borel-Cantelli entonces es suficiente demostrar que para toda $\varepsilon > 0$ se cumple que

$$\sum_{n=1}^{\infty} \mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) < \infty.$$

Similar a la demostración de la convergencia en probabilidad de los momentos de la DEE, por la desigualdad del triángulo podemos asegurar que existe $N > 0$ tal que si $n > N$ entonces $|\mathbb{E}(m_{n,k}) - C_{k/2}| < \varepsilon/2$ y, por lo tanto,

$$\mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) \leq \mathbb{P}(|m_{n,k} - \mathbb{E}(m_{n,k})| > \varepsilon/2).$$

Ahora por la desigualdad de Chebyshev podemos concluir que si $n > N$ entonces se cumple que

$$\mathbb{P}(|m_{n,k} - \mathbb{E}(m_{n,k})| > \varepsilon/2) \leq \frac{4}{\varepsilon^2} \text{Var}(m_{n,k}). \quad (4.4)$$

Así, sumando sobre toda $n \in \mathbb{N}$ y considerando (4.4) para toda $n > N$ obtenemos

$$\sum_{n=1}^{\infty} \mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) \leq \sum_{n=1}^N \mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) + \frac{4}{\varepsilon^2} \sum_{n=N+1}^{\infty} \text{Var}(m_{n,k}).$$

El primer término en la cota es finito porque es una suma finita y el segundo término es finito porque en 4.11 demostramos que $\text{Var}(m_{n,k}) = O(1/n^2)$. Por lo tanto, concluimos que

$$\sum_{n=1}^{\infty} \mathbb{P}(|m_{n,k} - C_{k/2}| > \varepsilon) < \infty$$

y así, aseguramos que los momentos pares convergen casi seguramente a los números de Catalan. \square

4.3. Convergencia débil de la DEE

En esta sección justificamos por qué la convergencia de los momentos de la DEE es la parte clave para asegurar que la DEE converge débilmente, en el sentido de medidas, a la ley del semicírculo. Esta técnica es ampliamente usada en la teoría de matrices aleatorias porque, como hemos visto, traduce problemas de convergencia de medidas a problemas combinatorios. En relevante notar que para esto utilizamos constantemente que la densidad de la ley del semicírculo tiene soporte compacto. Sin embargo, es posible usar el método de los momentos siempre que la medida límite esté únicamente determinada por sus momentos. Para este análisis más general referimos al lector a la sección 3.3 de [FK23].

Definición 4.15 (Convergencia débil de medidas). *Sea $\{\mu_n\}_{n \in \mathbb{N}}$ una sucesión de medidas de probabilidad de \mathbb{R} . La sucesión converge débilmente a la medida μ si para cada función $f : \mathbb{R} \rightarrow \mathbb{R}$ continua y acotada se cumple que*

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu.$$

Así, intuitivamente, queremos que las medidas conserven la media de las funciones en

un clase suficientemente interesante.

En nuestro caso, las medidas que queremos analizar son las definidas por la DEE de la sucesión de matrices de Wigner. Es decir, tomamos $\mu_n = \mu_{\overline{M}_n}$. Ahora, como estas son medidas aleatorias entonces las cantidades $\int f d\mu_n$ son variables aleatorias reales y, por lo tanto, tenemos que especificar el modo en que convergen.

Definición 4.16. (*Convergencia débil de medidas aleatorias*) Sea $\{\mu_n\}_{n \in \mathbb{N}}$ una sucesión de medidas aleatorias de probabilidad de \mathbb{R} . La sucesión converge débilmente en probabilidad (respectivamente, casi seguramente) a μ , una medida aleatoria de probabilidad en \mathbb{R} , si para cada función $f : \mathbb{R} \rightarrow \mathbb{R}$ continua y acotada se cumple que la sucesión de variables aleatorias reales $\int f d\mu_n$ converge en probabilidad (respectivamente, casi seguramente) a la variable aleatoria real $\int f d\mu$.

En nuestro caso, la medida límite es la ley del semicírculo, que es una medida de probabilidad determinista en \mathbb{R} .

Convergencia débil en probabilidad de la DEE

Para hacer más sencilla la notación, dada una medida de probabilidad μ y una función f definimos

$$\langle \mu, f \rangle = \int f d\mu.$$

Veamos algunos ejemplos para familiarizarnos con esta notación. Recordemos que denotamos como σ a la ley de una variable semicircular. Entonces, por ejemplo, tenemos que dada una función $f : \mathbb{R} \rightarrow \mathbb{R}$

$$\langle \sigma, f \rangle = \frac{1}{2\pi} \int_{-2}^2 f(x) \sqrt{4 - x^2} dx$$

y, recordando la definición de la DEE en [4.1](#),

$$\langle \mu_{\overline{M}_n}, f \rangle = \frac{1}{n} \sum_{i=1}^n f(\lambda_i(\overline{M}_n)).$$

Con esta notación, la convergencia en probabilidad de los momentos de DEE a los momentos del semicírculo (proposición 4.12) se traduce a que para toda $\varepsilon > 0$ y para toda $k \in \mathbb{N}$ se cumple que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\langle \mu_{\overline{M}_n}, x^k \rangle - \langle \sigma, x^k \rangle| > \varepsilon) = 0.$$

Veamos ahora, en el siguiente lema, que podemos acotar la aportación de los momentos de la DEE fuera de un intervalo. Este lema es útil porque al demostrar la convergencia débil de la DEE podemos reducir el análisis a funciones con soporte en este intervalo. Esto nos permite aproximar a las funciones con polinomios y usar las convergencias de los momentos que ya demostramos.

Lema 4.17 (Aportación de los momentos de la DEE fuera de un intervalo). *Sea $R > 4$, entonces para toda $k \in \mathbb{N}$ se cumple que*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle > \varepsilon) = 0.$$

Demostración. Por la desigualdad de Markov tenemos que

$$\mathbb{P}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle) \quad (4.5)$$

donde la esperanza es respecto a la medida de la matriz \overline{M}_n .

Ahora, queremos acotar esa esperanza con algo que involucre a $\mathbb{E}(\langle \mu_{\overline{M}_n}, x^{2k} \rangle)$, pues, como demostramos en la proposición 4.8 sabemos que la esperanza de los momentos pares converge a los números de Catalan y su forma cerrada nos permite controlar esos términos.

Notemos que si $|x| > R$ entonces $x^{2k} > R^k |x|^k$ y, por lo tanto,

$$R^k \mathbb{E}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle) \leq \mathbb{E}(\langle \mu_{\overline{M}_n}, x^{2k} \mathbb{1}_{\{|x| > R\}} \rangle). \quad (4.6)$$

Así, usando (4.5), (4.6) y que $\mathbb{E}(\langle \mu_{\overline{M}_n}, x^{2k} \mathbb{1}_{\{|x|>R\}} \rangle) \leq \mathbb{E}(\langle \mu_{\overline{M}_n}, x^{2k} \rangle)$, obtenemos

$$\mathbb{P}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x|>R\}} \rangle > \varepsilon) \leq \frac{1}{\varepsilon R^k} \mathbb{E}(\langle \mu_{\overline{M}_n}, x^{2k} \rangle).$$

Como por la proposición 4.8 sabemos que los momentos pares de la DEE convergen a los números de Catalan, tomando límite superior de ambos lados podemos concluir que

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x|>R\}} \rangle > \varepsilon) \leq \frac{1}{\varepsilon} \frac{C_k}{R^k}.$$

Notemos ahora que el k -ésimo número de Catalan es menor que 4^k pues

$$C_k = \frac{1}{k+1} \binom{2k}{k} \leq \binom{2k}{k} \leq \sum_{i=0}^{2k} \binom{2k}{i} = 2^{2k}.$$

Finalmente, tenemos que para toda $k \in \mathbb{N}$ se cumple que

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x|>R\}} \rangle > \varepsilon) \leq \frac{1}{\varepsilon} \frac{4^k}{R^k}. \quad (4.7)$$

Veamos ahora que ese límite superior es igual a 0. La sucesión de probabilidades en el lado izquierdo de la desigualdad (4.7) es no decreciente en k , pues cuando $|x| > R > 1$ entonces $|x|^k$ es creciente en k . Sin embargo, por la elección de R tenemos que $4/R < 1$ y, por lo tanto, el lado derecho de la desigualdad tiende a 0 cuando $k \rightarrow \infty$. Es decir, tenemos una sucesión no decreciente acotada término a término por una sucesión que decrece a 0. Entonces la única posibilidad es que para toda $k \in \mathbb{N}$ el lado izquierdo sea 0. Así, concluimos que para toda $k \in \mathbb{N}$ se cumple que

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x|>R\}} \rangle > \varepsilon) = 0. \quad \square$$

En el siguiente corolario del lema anterior podemos ver que basta considerar a las funciones con soporte en el intervalo $[-R, R]$ para demostrar la convergencia débil en probabilidad que buscamos.

Corolario 4.18 (Es suficiente considerar funciones con soporte en un intervalo). *Sea $f : \mathbb{R} \rightarrow \mathbb{R}$ una función continua y acotada. Entonces su aportación fuera del intervalo $[-R, R]$ es despreciable, es decir,*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\langle \mu_{\overline{M}_n}, f \mathbb{1}_{\{|x| > R\}} \rangle > \varepsilon) = 0$$

Demostración. Como f es una función acotada consideremos una $M \in \mathbb{R}$ tal que $f(x) \leq M$ para toda x . Por lo tanto, $\langle \mu_{\overline{M}_n}, f \mathbb{1}_{\{|x| > R\}} \rangle \leq M \langle \mu_{\overline{M}_n}, \mathbb{1}_{\{|x| > R\}} \rangle$.

Así, tenemos que $\mathbb{P}(\langle \mu_{\overline{M}_n}, f \mathbb{1}_{\{|x| > R\}} \rangle > \varepsilon) \leq \mathbb{P}(\langle \mu_{\overline{M}_n}, \mathbb{1}_{\{|x| > R\}} \rangle > \frac{\varepsilon}{M})$ que, por el lema 4.17 (caso $k = 0$) tiende a 0 cuando $n \rightarrow \infty$. \square

Por esta razón, de ahora en adelante consideramos funciones continuas con soporte en $[-R, R]$.

Para lo que sigue es esencial que el conjunto de polinomios es denso (respecto a la métrica uniforme) en las funciones con soporte en un intervalo cerrado. Es decir, el teorema de aproximación de Weierstrass, que en seguida enunciamos sin demostrar.

Teorema 4.19 (Aproximación de Weierstrass). *Sea $\varepsilon > 0$ y f una función con soporte en el intervalo $[-R, R]$. Entonces existe un polinomio p_ε tal que*

$$\sup_{x \in [-R, R]} |f(x) - p_\varepsilon(x)| < \varepsilon.$$

Aplicar este teorema al análisis de la convergencia de la DEE es el punto clave del método de los momentos. Como la integral es un operador lineal, entonces la integral $\int f d\mu_{\overline{M}_n}$ se puede aproximar como una combinación lineal de los momentos, que ya estudiamos. Veámoslo con detalle en el siguiente teorema.

Teorema 4.20 (Convergencia débil en probabilidad a la ley del semicírculo). *Para toda $\varepsilon > 0$ y para toda función $f : \mathbb{R} \rightarrow \mathbb{R}$ continua y con soporte en $[-R, R]$ se cumple que*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle| > \varepsilon) = 0.$$

Demostración. Sea $\varepsilon > 0$ y f una función continua con soporte en $[-R, R]$. Por el teorema de aproximación de Weierstrass podemos garantizar la existencia de un polinomio

$$p(x) = \sum_{i=0}^m c_i x^i,$$

cuyo grado m y coeficientes c_i dependen de ε , que cumple que

$$\sup_{x \in [-R, R]} |f(x) - p(x)| < \varepsilon/4.$$

La clave ahora es incluir a los términos $\langle \mu_{\overline{M}_n}, p \rangle$ y $\langle \sigma, p \rangle$ en nuestro desarrollo, pues esto nos permite aprovechar el análisis que ya hicimos sobre los momentos de $\mu_{\overline{M}_n}$ y de σ . Para esto los sumamos y restamos, y por la linealidad de la integral obtenemos que

$$\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle = \langle \mu_{\overline{M}_n}, f - p \rangle - \langle \sigma, f - p \rangle + \langle \mu_{\overline{M}_n}, p \rangle - \langle \sigma, p \rangle. \quad (4.8)$$

Recordemos que el polinomio p es una buena aproximación de f en el intervalo $[-R, R]$. Por lo tanto, conviene analizar los primeros dos términos de (4.8) considerando por separado lo que pasa dentro y fuera de $[-R, R]$. Usando el hecho de que la función f y la densidad de σ valen 0 fuera de ese intervalo, tenemos que

$$\begin{aligned} \langle \mu_{\overline{M}_n}, f - p \rangle - \langle \sigma, f - p \rangle &= \langle \mu_{\overline{M}_n}, (f - p) \mathbb{1}_{\{x \leq R\}} \rangle + \langle \mu_{\overline{M}_n}, (f - p) \mathbb{1}_{\{x > R\}} \rangle \\ &\quad - \langle \sigma, (f - p) \mathbb{1}_{\{x \leq R\}} \rangle - \langle \sigma, (f - p) \mathbb{1}_{\{x > R\}} \rangle \\ &= \langle \mu_{\overline{M}_n}, (f - p) \mathbb{1}_{\{x \leq R\}} \rangle - \langle \mu_{\overline{M}_n}, p \mathbb{1}_{\{x > R\}} \rangle \\ &\quad - \langle \sigma, (f - p) \mathbb{1}_{\{x \leq R\}} \rangle. \end{aligned} \quad (4.9)$$

Considerando el valor absoluto y que, por la elección de p_ε , el primero y el último

término de (4.9) están acotados por $\varepsilon/4$ finalmente tenemos que

$$\begin{aligned} |\langle \mu_{\overline{M}_n}, f - p \rangle - \langle \sigma, f - p \rangle| &\leq |\langle \mu_{\overline{M}_n}, (f - p) \mathbb{1}_{\{x \leq R\}} \rangle| + |\langle \mu_{\overline{M}_n}, p \mathbb{1}_{\{x > R\}} \rangle| \\ &\quad + |\langle \sigma, (f - p) \mathbb{1}_{\{x \leq R\}} \rangle| \\ &\leq \varepsilon/2 + |\langle \mu_{\overline{M}_n}, p \mathbb{1}_{\{x > R\}} \rangle|. \end{aligned}$$

Finalmente, tomando en cuenta el valor absoluto de la desigualdad (4.8) podemos concluir que

$$|\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle| \leq \varepsilon/2 + |\langle \mu_{\overline{M}_n}, p \mathbb{1}_{\{x > R\}} \rangle| + |\langle \mu_{\overline{M}_n}, p \rangle - \langle \sigma, p \rangle|. \quad (4.10)$$

Por la desigualdad del triángulo tenemos que si $|\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle|$ es mayor que ε entonces alguno de los últimos dos términos de la cota en (4.10) debe ser mayor que $\varepsilon/4$. Por lo tanto,

$$\begin{aligned} \mathbb{P}(|\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle| > \varepsilon) &\leq \mathbb{P}(|\langle \mu_{\overline{M}_n}, p \mathbb{1}_{\{x > R\}} \rangle| > \varepsilon/4) \\ &\quad + \mathbb{P}(|\langle \mu_{\overline{M}_n}, p \rangle - \langle \sigma, p \rangle| > \varepsilon/4). \end{aligned} \quad (4.11)$$

Lo que sigue ahora es aprovechar al polinomio p_ε y la linealidad de la integral para expresar los dos sumandos en la cota de (4.11) en términos de momentos. La primera parte la podemos acotar gracias a que acotamos a los momentos de la DEE fuera de $[-R, R]$ (lema 4.17) y la segunda gracias a la convergencia en probabilidad de los momentos de la DEE a los momentos de la ley del semicírculo (proposición 4.12). En seguida lo analizamos con todo detalle.

Para el primer sumando en (4.11) notemos que

$$\begin{aligned} |\langle \mu_{\overline{M}_n}, p_\varepsilon \mathbb{1}_{\{x > R\}} \rangle| &= \left| \sum_{i=0}^m c_{i,\varepsilon} \langle \mu_{\overline{M}_n}, x^i \mathbb{1}_{\{x > R\}} \rangle \right| \\ &\leq \sum_{i=0}^m |c_{i,\varepsilon}| |\langle \mu_{\overline{M}_n}, x^i \mathbb{1}_{\{x > R\}} \rangle|. \end{aligned} \quad (4.12)$$

Además, si la suma en (4.12) es mayor a $\varepsilon/4$ entonces alguno de sus $m+1$ sumandos debe ser mayor que $\frac{\varepsilon}{4(m+1)}$ y, por lo tanto, tenemos que

$$\mathbb{P} \left(\sum_{i=0}^m |c_{i,\varepsilon}| \langle \mu_{\overline{M}_n}, x^i \mathbb{1}_{\{x>R\}} \rangle > \varepsilon/4 \right) \leq \sum_{i=0}^m \mathbb{P} \left(\langle \mu_{\overline{M}_n}, x^i \mathbb{1}_{\{x>R\}} \rangle > \frac{\varepsilon}{4(m+1)|c_{i,\varepsilon}|} \right).$$

Por el lema 4.17 sabemos que cada sumando tiende a 0 cuando $n \rightarrow \infty$ y así podemos concluir que

$$\lim_{n \rightarrow \infty} \mathbb{P} (|\langle \mu_{\overline{M}_n}, p_\varepsilon \mathbb{1}_{\{x>R\}} \rangle| > \varepsilon/4) = 0.$$

Procedemos de manera análoga con el segundo sumando de (4.11). Expandiendo el polinomio p_ε para dejar todo en término de los momentos tenemos que

$$|\langle \mu_{\overline{M}_n}, p_\varepsilon \rangle - \langle \sigma, p_\varepsilon \rangle| \leq \sum_{i=0}^m |c_{i,\varepsilon}| |\langle \mu_{\overline{M}_n}, x^i \rangle - \langle \sigma, x^i \rangle|.$$

Por un argumento análogo,

$$\mathbb{P} \left(\sum_{i=0}^m |c_{i,\varepsilon}| |\langle \mu_{\overline{M}_n}, p_\varepsilon \rangle - \langle \sigma, p_\varepsilon \rangle| > \varepsilon/4 \right) \leq \sum_{i=0}^m \mathbb{P} \left(|\langle \mu_{\overline{M}_n}, x^i \rangle - \langle \sigma, x^i \rangle| > \frac{\varepsilon}{4(m+1)|c_{i,\varepsilon}|} \right)$$

y como los momentos de la DEE convergen en probabilidad a los momentos de la ley del semicírculo (proposición 4.12) entonces cada sumando tiende a 0 cuando $n \rightarrow \infty$. Podemos concluir entonces que

$$\lim_{n \rightarrow \infty} \mathbb{P} (|\langle \mu_{\overline{M}_n}, p_\varepsilon \rangle - \langle \sigma, p_\varepsilon \rangle| > \varepsilon/4) = 0.$$

Como hemos demostrado que los dos términos en (4.11) tienden a 0 cuando $n \rightarrow \infty$ finalmente tenemos que

$$\lim_{n \rightarrow \infty} \mathbb{P} (|\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle| > \varepsilon) = 0.$$

Es decir, la sucesión de las DEE converge débilmente en probabilidad a la ley del semicírculo. \square

Convergencia débil casi segura de la DEE

Finalmente demostramos que la convergencia débil de la sucesión de las DEE a la ley del semicírculo se puede mejorar a convergencia casi segura. El análisis es análogo a la convergencia débil en probabilidad. Primero presentamos un lema que nos permite asegurar que casi seguramente, en el límite las DEE son medidas concentradas en un intervalo. Este lema es un análogo casi seguro del lema 4.17.

Lema 4.21 (Aportación de los momentos de la DEE fuera de un intervalo). *Sea $R > 4$, entonces para toda $k \in \mathbb{N}$ se cumple casi seguramente que*

$$\limsup_{n \rightarrow \infty} \langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle = 0.$$

Demostración. Notemos que si $|x| > R$ entonces $|x|^{-k} < R^{-k}$ y por lo tanto,

$$|x|^k = |x|^{k-2k} |x|^{2k} < R^{-k} x^{2k}.$$

Así, tenemos que $\limsup_{n \rightarrow \infty} \langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle \leq R^{-k} \langle \mu_{\overline{M}_n}, x^{2k} \rangle$. Por la proposición 4.14 sabemos que los momentos pares de la DEE convergen casi seguramente a los números de Catalan y así podemos concluir que casi seguramente

$$\limsup_{n \rightarrow \infty} \langle \mu_{\overline{M}_n}, |x|^k \mathbb{1}_{\{|x| > R\}} \rangle \leq \frac{C_k}{R^k}.$$

Al igual que en el lema 4.17, esto nos permite concluir que para cada $k \in \mathbb{N}$ ese límite superior debe ser 0. □

De igual manera, el lema anterior nos permite considerar solo a las funciones con soporte en $[-R, R]$, pues en el límite su aportación fuera de este es 0 casi seguramente.

Teorema 4.22 (Convergencia débil casi segura a la ley del semicírculo). *Para toda función $f : \mathbb{R} \rightarrow \mathbb{R}$ continua y con soporte en $[-R, R]$ se cumple que*

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \langle \mu_{\overline{M}_n}, f \rangle = \langle \sigma, f \rangle \right) = 1.$$

Demostración. Veamos que $|\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle| \rightarrow 0$ casi seguramente.

Como f tiene soporte en $[-R, R]$ consideremos un polinomio p que la aproxime uniformemente ahí. Podemos garantizar la existencia de p por el teorema de aproximación de Weierstrass.

Análogo a la demostración de la proposición 4.20 podemos aludir a la desigualdad del triángulo y obtener que

$$\begin{aligned} |\langle \mu_{\overline{M}_n}, f \rangle - \langle \sigma, f \rangle| &\leq |\langle \mu_{\overline{M}_n}, (f - p)\mathbb{1}_{\{x \leq R\}} \rangle| + |\langle \sigma, (f - p)\mathbb{1}_{\{x \leq R\}} \rangle| \\ &\quad + |\langle \mu_{\overline{M}_n}, p\mathbb{1}_{\{x > R\}} \rangle| + |\langle \mu_{\overline{M}_n}, p \rangle - \langle \sigma, p \rangle|. \end{aligned} \quad (4.13)$$

Los primeros dos términos de (4.13) se pueden hacer arbitrariamente pequeños pues p es una aproximación uniforme de f en el intervalo $[-R, R]$.

Para los últimos dos términos notemos que como la integral es lineal entonces podemos expresarlos en término de momentos. Así, el tercer término de (4.13) tiende a 0 casi seguramente por el lema 4.21. El último término tiende a 0 casi seguramente pues los momentos de la DEE tienden casi seguramente a los momentos de la ley del semicírculo (proposición 4.14).

Concluimos así que la sucesión de las DEE converge débilmente casi seguramente a la ley del semicírculo. \square

Apéndice A

Método de truncamiento

Los resultados analizados en los capítulos 2 y 3 se restringieron a variables aleatorias acotadas y, sin embargo, al demostrar la convergencia a la ley del semicírculo solo asumimos que las entradas de las matrices estuvieran centradas y tuvieran varianza 1. Mediante el método de truncamiento es posible extender resultados a variables no acotadas que satisfacen ciertas condiciones de decaimiento en sus colas. En particular, esta técnica suele simplificar la generalización de propiedades de variables acotadas a variables subgaussianas. Así, se puede ver, por ejemplo, que el fenómeno de concentración de la norma de operador se presenta también en contextos más generales y es posible justificar la normalización por \sqrt{n} en la convergencia a la ley del semicírculo para una clase más amplia de matrices aleatorias.

Cuando hablamos de truncar a una variable aleatoria X por un parámetro $N > 0$ nos referimos a descomponerla como

$$X = X_{\leq N} + X_{> N}$$

donde $X_{\leq N} = X \mathbb{1}_{\{|X| \leq N\}}$ y $X_{> N} = X \mathbb{1}_{\{|X| > N\}}$.

Esta descomposición es útil porque los resultados de concentración establecidos para variables acotadas son directamente aplicables al primer término, $X_{\leq N}$. La sutileza

del método, por lo tanto, radica en controlar adecuadamente al término $X_{>N}$, o sea, a las colas de X .

Es útil notar que del teorema de convergencia monótona se sigue que

$$\lim_{N \rightarrow \infty} \mathbb{E}(|X_{>N}|) = \mathbb{E}\left(\lim_{N \rightarrow \infty} |X_{>N}|\right) = 0. \quad (\text{A.1})$$

Por lo tanto, para controlar el decaimiento de este término a veces es suficiente con escoger adecuadamente al parámetro N y usar cotas del primer momento como la desigualdad de Markov.

Otra observación relevante es que si X tiene esperanza finita entonces $X_{\leq N}$ tiene segundo momento finito, pues

$$\mathbb{E}(|X_{\leq N}|^2) \leq \mathbb{E}(N|X_{\leq N}|) = N\mathbb{E}(|X_{\leq N}|) \leq N\mathbb{E}(X) < \infty \quad (\text{A.2})$$

y, por lo tanto, también tiene varianza finita.

Con tan solo estas propiedades podemos ya extender muchos resultados de variables acotadas a variables no acotadas. Veamos un ejemplo de un teorema clásico para ilustrar la utilidad del método de truncamiento.

Teorema A.1 (Ley débil de los grandes números). *Sean $\{X_n\}_{n \in \mathbb{N}}$ variables aleatorias independientes e idénticamente distribuidas con $X_i \stackrel{d}{=} X$ para cada i . Si $S_n = \sum_{i=1}^n X_i$ entonces $\frac{1}{n}S_n$ converge en probabilidad a $\mathbb{E}(X)$. Es decir, para cada $\varepsilon > 0$ se cumple que*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}(X)\right| > \varepsilon\right) = 0.$$

Demostración. Sea $\varepsilon > 0$. Vamos a asumir sin perder generalidad que $\mathbb{E}(X) = 0$. Así, ahora tenemos que demostrar que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n| > n\varepsilon) = 0.$$

Notemos primero que si X tiene segundo momento finito entonces el resultado se

sigue directamente de la desigualdad de Chebyshev pues como $\mathbb{E}(S_n) = n\mathbb{E}(X) = 0$ y $\text{Var}(S_n) = n\mathbb{E}(X^2)$ entonces

$$\mathbb{P}(|S_n| > \varepsilon n) \leq \frac{\mathbb{E}(X^2)}{n\varepsilon^2} \rightarrow 0.$$

Sin embargo, si X no tiene segundo momento finito entonces no podemos usar ese argumento directamente. En cambio usamos el método de truncamiento. Trunquemos a cada X_i por un parámetro $N > 0$, que elegiremos más adelante, y consideremos las sumas de sus partes acotadas y no acotadas. Es decir, sea $S_{n,\leq N} = \sum_{i=1}^n X_{i,\leq N}$ y $S_{n,>N} = \sum_{i=1}^n X_{i,>N}$.

Veamos ahora que si $|S_n| > n\varepsilon$ entonces al menos una de las partes del truncamiento es mayor que $\frac{n\varepsilon}{2}$. Para demostrar esto, supongamos que $|S_n| > n\varepsilon$ y, sin pérdida de generalidad, supongamos que $|S_{\leq N}| \leq \frac{n\varepsilon}{2}$. Entonces, por la desigualdad del triángulo,

$$n\varepsilon < |S_n| \leq |S_{\leq N}| + |S_{>N}| \leq \frac{n\varepsilon}{2} + |S_{>N}|$$

y, por lo tanto, $|S_{>N}| > \frac{n\varepsilon}{2}$. Visto de otra manera, probamos la siguiente contención de eventos:

$$\{|S_n| > n\varepsilon\} \subseteq \left\{|S_{\leq N}| > \frac{n\varepsilon}{2}\right\} \cup \left\{|S_{>N}| > \frac{n\varepsilon}{2}\right\}.$$

Así, por la monotonía y la subaditividad de la probabilidad tenemos que

$$\mathbb{P}(|S_n| > n\varepsilon) \leq \mathbb{P}\left(|S_{\leq N}| > \frac{n\varepsilon}{2}\right) + \mathbb{P}\left(|S_{>N}| > \frac{n\varepsilon}{2}\right). \quad (\text{A.3})$$

Ahora basta ver que los dos sumandos tienden a 0 cuando $n \rightarrow \infty$. Para esto notemos que de la hipótesis de independencia de las X_i y como asumimos que están centradas podemos concluir que

$$\mathbb{E}(|S_{n,\leq N}|^2) = \mathbb{E}\left(\left|\sum_{i=1}^n \sum_{j=1}^n X_{i,\leq N} X_{j,\leq N}\right|\right) = n \mathbb{E}(|X_{\leq N}|^2).$$

Como en (A.2) demostramos que $X_{i,\leq N}$ tiene segundo momento finito, entonces podemos asegurar que $S_{n,\leq N}$ también tiene segundo momento finito. Por lo tanto, podemos usar la desigualdad de Chebyshev para $S_{n,\leq N}$ y obtener que

$$\mathbb{P}\left(|S_{n,\leq N}| \geq \frac{\varepsilon}{2}n\right) \leq \frac{4\text{Var}(X_{\leq N})}{\varepsilon^2 n} \leq \frac{4N\mathbb{E}(|X_{\leq N}|)}{\varepsilon^2 n}, \quad (\text{A.4})$$

que tiende a 0 cuando n tiende a infinito.

Para controlar $S_{n,>N}$ podemos usar la desigualdad de Markov y obtener

$$\mathbb{P}\left(|S_{n,>N}| \geq \frac{\varepsilon}{2}n\right) \leq \frac{2}{\varepsilon} \mathbb{E}(|X_{>N}|). \quad (\text{A.5})$$

Por lo tanto, usando (A.3), (A.4) y (A.5) obtenemos que

$$\mathbb{P}(|S_n| \geq \varepsilon n) \leq \frac{4N\mathbb{E}(|X_{\leq N}|)}{\varepsilon^2 n} + \frac{2}{\varepsilon} \mathbb{E}(|X_{>N}|).$$

Por (A.1), si N es suficientemente grande podemos hacer $\mathbb{E}(|X_{\leq N}|) - \mathbb{E}(X)$ tan pequeño como queramos. En particular, recordando que asumimos $\mathbb{E}(X) = 0$, podemos hacer $\mathbb{E}(|X_{\leq N}|) \leq \frac{\varepsilon}{4}$.

Sea $\delta > 0$. Como sabemos que $\mathbb{E}(|X_{>N}|) \rightarrow 0$ cuando $N \rightarrow \infty$ entonces podemos considerar N suficientemente grande para que $\mathbb{E}(|X_{>N}|) \leq \delta$. Así, si consideramos una N apropiada para que se cumplan ambas condiciones obtenemos que

$$\mathbb{P}(|S_n| \geq \varepsilon n) \leq \frac{N}{\varepsilon n} + \frac{2}{\varepsilon} \delta.$$

Tomando el límite cuando n tiende a infinito tenemos que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n| \geq \varepsilon n) \leq \frac{2}{\varepsilon} \delta.$$

Como δ es un real positivo arbitrario entonces finalmente podemos concluir que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n| \geq \varepsilon n) = 0. \quad \square$$

Extender resultados de variables acotadas a variables no acotadas usualmente requiere métodos más sofisticados. Sin embargo, la filosofía subyacente queda bien ilustrada con el ejemplo anterior: utilizar los resultados para variables acotadas en la parte acotada del truncamiento y controlar adecuadamente las colas.

Bibliografía

Los números en color azul muestran en qué página se cita cada trabajo.

- [ABDF15] Gernot Akemann, Jinho Baik, and Philippe Di Francesco. *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, 09 2015. [5](#), [6](#)
- [BCG22] A. Borodin, I. Corwin, and A. Guionnet, editors. *Random Matrices*. American Mathematical Society and IAS/Park City Mathematics Institute, 2022. [8](#), [56](#)
- [Bor19] C. Bordenave. IMPA Lecture Notes on Random Matrix Theory, 2019. Lecture notes. [8](#)
- [DF17] Persi Diaconis and Peter J. Forrester. Hurwitz and the origins of random matrix theory in mathematics. *Random Matrices: Theory and Applications*, 06(01):1730001, 2017. [1](#)
- [DM47] A. Dvoretzky and T. Motzkin. A Problem of Arrangements. *Duke Mathematical Journal*, 14(2):305–313, 1947. [14](#), [20](#)
- [DMRA11] J. A. Domínguez Molina and H. Rocha Arteaga. El Teorema de Wigner para Matrices Aleatorias. *Miscelánea Matemática*, 52:31–51, 2011. [8](#)
- [DZ90] N. Dershowitz and S. Zaks. The Cycle Lemma and Some Applications. *European Journal of Combinatorics*, 11(1):35–40, 1990. [20](#)

BIBLIOGRAFÍA

- [FK23] Michael Fleermann and Werner Kirsch. Proof Methods in Random Matrix Theory. *Probability Surveys*, 20:291 – 381, 2023. [70](#)
- [GAH22] A. Guionnet, J. Augeri, and J. Husson. *An Introduction to Random Matrices*. Cambridge University Press, 2022. [8](#)
- [Hoe63] W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. [25](#)
- [KS00] M. Krbálek and P. Seba. The Statistical Properties of the City Transport in Cuernavaca (Mexico) and Random Matrix Ensembles. *Journal of Physics A: Mathematics and General*, 33(26):L229–L234, 2000. [5](#)
- [Lar00] P. J. Larcombe. The 18th Century Chinese Discovery of the Catalan Numbers. *Mathematical Spectrum*, 32(1):5–7, 1999/2000. [18](#)
- [Lat05] R. Latała. Some Estimates of Norms of Random Matrices. *Proceedings of the American Mathematical Society*, 133(5):1273–1282, 2005. [37](#), [51](#)
- [Tal95] Michel Talagrand. Concentration of Measure and Isoperimetric Inequalities in Product Spaces. *Publications Mathématiques de l’Institut des Hautes Études Scientifiques*, 81:73 – 205, 1995. [24](#)
- [Tao12] T. Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. [8](#), [29](#)
- [Ver18] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018. [9](#), [12](#)
- [Wig55] E. P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. [2](#), [8](#), [45](#), [53](#), [55](#)

BIBLIOGRAFÍA

- [Wis28] J. Wishart. The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, 20A(1/2):32–52, 1928. [1](#)

Guía de notación

S_n	Esfera unitaria de \mathbb{R}^n .
$\lambda_i(M)$	El i -ésimo valor propio de la matriz M , si esta tiene valores propios reales y los ordenamos de menor a mayor.
$\langle \mu, f \rangle$	$\int f d\mu$.
\mathbb{E}	Esperanza.
\mathbb{P}	Probabilidad.
μ_M	Distribución espectral empírica (DEE) de la matriz M .
σ	Ley del semicírculo.
$\ M\ _{op}$	Norma de operador de la matriz M .