Concentración de medidas de probabilidad

Marco Tulio Gaxiola Leyva

Agosto 2012

Contenido

Prefacio

La concentración de medida se refiere al hecho de que una función de un número grande de variables aleatorias tienda a concentrar su valor en un rango relativamente pequeño alrededor de cierto número, esto bajo ciertas condiciones de suavidad de dicha función. En esta tesis las condiciones de suavidad que se les pide a las funciones serán los supuestos de *convexidad y 1-Lipschitz.*

El primero y más común de los resultados de concentración es para promedios. Éste establece que "El promedio de variables aleatorias independientes acotadas está concentrado alrededor de su esperanza". Escrito de otra forma: "Dada una colección de variables aleatorias independientes y acotadas $X_1,\ X_2,\ \ldots,\ X_n$, si $\bar{\mathbf{X}} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ es una nueva variable aleatoria con valor esperado $\mathbf{E}[\bar{\mathbf{X}}]$, ¿Cuál es la probabilidad de que $\bar{\mathbf{X}}$ se desvíe considerablemente de su valor esperado?".

El fenómeno de la concentración de medida fue presentado en los años setenta por V. Milman [8]. Es de un gran interés en aplicaciones, en diversas áreas, como la geometría, análisis funcional e integración en dimensiones infinitas, matemáticas discretas y especialmente en teoría de probabilidad. El fenómeno de la concentración de la medida se desarrolló gracias a los trabajos de Milman y Gromov [5], Maurey [7], Pisier [9], Talagrand [11], [12], Ledoux [6], y otros.

En forma más compleja, el fenómeno de la concentración de la medida subyace ampliamente nuestro mundo físico. Como sabemos, el mundo está hecho de partículas microscópicas que son gobernadas por las leyes de la probabilidad. La razón por la cual las propiedades macroscópicas de las grandes uniones de estas partículas parecieran ser determinísticas cuando son observadas a nuestras escalas, es precisamente la concentración de la medida: pues aunque las propiedades de estas partículas microscópicas son modeladas con distintas probabilidades, al unir un número considerable partículas (al grado de ser visibles) las probabilidades de estas propiedades tienden a concentrarse.

El principal objetivo de esta tesis es ilustrar el uso de la concentración de medidas de probabilidad en la teoría de las matrices aleatorias. Las matrices aleatorias son un campo de estudio de las matemáticas, que actualmente está siendo ampliamente investigado, debido a sus aplicaciones tanto teóricas como prácticas.

Nuestro segundo objetivo es exponer algunos resultados de concentración de medidas de probabilidad en ciertos espacios de probabilidad.

La organización de esta tesis es la siguiente: en el Capítulo 1 se exponen algunos resultados básicos en probabilidad, así como algunas desigualdades que serán útiles para el desarrollo del mismo, tales como la desigualdad de Markov y la desigualdad de Jensen. En el Capítulo 2 se presentan resultados de concentración de medida de probabilidad: primero mostrando la concentración de medida en el espacio de probabilidad conocido como el cubo booleano. En este capítulo también se expone un resultado de concentración para funciones 1-Lipschitz evaluadas en variables aleatorias independientes, las cuales poseen una distribución normal estándar. Y para cerrar el Capítulo 2 se presenta el teorema nombrado como desigualdad de concentración de Talagrand, el cual es uno de los resultados más importantes del fenómeno de concentración de medida. El Capítulo 3 contiene aplicaciones de la desigualdad de concentración de Talagrand en teoría de matrices aleatorias, dentro de las cuales se encuentra una demostración de la parte de convergencia casi segura en el Teorema de Wigner. En el Apéndice A aparecen algunos resultados importantes de cálculo, que son utilizados en algunas de las demostraciones de esta tesis. Por último en el Apéndice B se encuentran las demostraciones de los lemas, teoremas, corolarios y proposiciones que se enuncian en el desarrollo de este trabajo.

Capítulo 1

Preliminares

1.1 Motivación

En probabilidad, la idea básica de concentración de medida se puede observar en el experimento aleatorio más simple: el lanzamiento de una moneda. Si lanzamos una moneda (legal) no podemos asegurar qué resultado obtendremos: "águila" o "sol". Sin embargo, supongamos que lanzamos la misma moneda "muchas" veces, digamos mil veces. Ahora en el experimento se torna *fuertemente predecible* que el número de águilas que obtendremos estará *concentrado* alrededor de los 500. Veamos esto de manera formal:

Sean $X_1, ..., X_n \in \{0, 1\}$ donde $X_i,\ 1 \leq i \leq n$ representa el i-ésimo resultado de lanzar una moneda (0 si obtenemos sol y 1 si obtenemos águila) y supongamos que $\mathbf{P}(X_i = 0) = \mathbf{P}(X_i = 1) = 1/2$. Entonces el número de águilas que obtendremos, al lanzar n veces una moneda, estará representado por $S_n = X_1 + \cdots + X_n$, entonces tenemos lo siguiente:

$$\mathbf{P}(|S_n - n/2| \geq \varepsilon) \leq 2e^{-2\varepsilon^2/n}, \tag{1.1}$$

véase Teorema 21.

Tomemos $\varepsilon = 0.1n$, de acuerdo con la desigualdad (1.1) se tienen los siguientes tres casos particulares

$$\mathbf{P}(|S_{100} - 50| \geq 10) \leq 2\,(0.98)^{100} = 0.270,$$

$$\mathbf{P}\left(\left|S_{500}-250\right| \geq 50\right) \leq 2\left(0.98\right)^{500}=0.0001,$$

$$\mathbf{P}\left(\left|S_{1000}-500\right| \geq 100\right) \leq 2\left(0.98\right)^{1000}=4.122\times 10^{-9}.$$

La siguiente figura muestra la gráfica de la función $p\left(n\right)=2e^{-(0.1)^2 n}$

Observemos que conforme crece el número de lanzamientos de la moneda, la probabilidad de que el número de águilas se desvíe un 10% del valor esperado va disminuyendo considerablemente.

Con el fin de generalizar (1.1) en el tema de concentración de medida, se requiere un cambio de perspectiva: simplemente considerar la variable aleatoria $S_n=\sum_{i=1}^{n} X_i$ como una *función* de las variables aleatorias individuales X_i, *i.e.* $S_n=F\left(X_1,...,X_n\right)$ y buscar qué condiciones pedirle a F para que se cumplan desigualdades del tipo (1.1).

1.2 Conceptos y resultados básicos de probabilidad

En esta sección se presentan conceptos y resultados básicos de probabilidad que son necesarios para la lectura de este trabajo. El lector interesado en ahondar en estos temas puede consultar [2].

Recordemos que un espacio de probabilidad es una terna $(\Omega,\mathcal{F},\mathbf{P})$, donde a Ω le llamamos el espacio muestral, $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω y $\mathbf{P}$ es una medida sobre $\mathcal{F}$, tal que, $\mathbf{P}\left(\Omega\right)=1$. A $\mathbf{P}$ se le llama medida de probabilidad. Como ejemplo de espacio de probabilidad podemos dar el espacio que corresponde al experimento de sacar una bola al azar de una caja que contiene n bolas. Empecemos por tomar a Ω como un conjunto finito de n puntos, es decir n representa el número de bolas que se encuentran dentro de la caja, $\mathcal{F}$ será la colección de todos los subconjuntos de Ω y $\mathbf{P}$ es la medida de probabilidad que asigna a $A\in\mathcal{F}$ la probabilidad $\mathbf{P}(A)=k/n$, si A tiene exactamente k elementos.

Definición 1 *A la σ-álgebra más pequeña de subconjuntos de $\mathbb{R}$ que contiene a todos los conjuntos abiertos de $\mathbb{R}$, le llamamos la σ-álgebra de Borel y la denotamos por $\mathfrak{B}(\mathbb{R})$. A los elementos de $\mathfrak{B}(\mathbb{R})$ le llamamos los borelianos.*

Definición 2 *Una variable aleatoria X en $(\Omega, \mathcal{F}, \mathbf{P})$ es una función*

$$X : \Omega \longrightarrow \mathbb{R},$$

medible relativa a las σ-álgebras $\mathcal{F}$ y $\mathfrak{B}(\mathbb{R})$, es decir, $X^{-1}(A) \in \mathcal{F}$, para todo $A \in \mathfrak{B}(\mathbb{R})$.

Si tenemos X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ y $A \in \mathcal{F}$ entonces definimos

$$\mathbf{P}_X(A) := \mathbf{P}\left(X^{-1}(A)\right) = \mathbf{P}(X \in A).$$

Definición 3 *Sean $X_1, \ldots, X_n$ variables aleatorias en $(\Omega, \mathcal{F}, \mathbf{P})$. Entonces diremos que X_1, $\ldots, X_n$ son independientes si y solo si para todo conjunto $B_1, \ldots, B_n \in \mathfrak{B}(\mathbb{R})$, tenemos que*

$$\mathbf{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1) \cdots \mathbf{P}(X_n \in B_n).$$

Definición 4 *Si X es variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$, la esperanza de X se define como*

$$\mathbf{E}(X) = \int_{\Omega} x d\mathbf{P}_X(x),$$

siempre que la integral exista.

El siguiente resultado nos será de gran ayuda en las pruebas de los teoremas importantes de este trabajo.

Proposición 5 *Sea X una variable aleatoria no negativa en $(\Omega, \mathcal{F}, \mathbf{P})$, entonces*

$$\mathbf{E}(X) = \int_0^\infty \mathbf{P}(X \geq \lambda)\, d\lambda.$$

Definición 6 *Sea X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$. Si $k > 0$ y si $\mathbf{E}|X|^k < \infty$, entonces a el número $\mathbf{E}(X^k)$ se le llama el k-ésimo momento de X; a $\mathbf{E}|X|^k$ se le llama el k-ésimo momento absoluto de X.*

Definición 7 *Si X es una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$, a $\mathbf{E}(X - \mathbf{E}(X))^k$ se le llama el k-ésimo momento central de X; a $\mathbf{E}|X - \mathbf{E}(X)|^k$ se le llama el k-ésimo momento central absoluto de X.*

Definición 8 *Al primer momento $(k = 1)$ $\mathbf{E}(X)$ se le llama media de X. El segundo momento central $\sigma^2 = \mathbf{E}\left[(X - \mathbf{E}(X))^2\right]$ se le llama varianza de X y se denota $Var(X)$, y a su raíz cuadrada positiva se le llama desviación estándar.*

Definición 9 *Sea X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$, la mediana $\mathbf{M}(X)$ es un número tal que,*

$$\mathbf{P}(X \leq \mathbf{M}(X)) \geq 1/2 \quad y \quad \mathbf{P}(X \geq \mathbf{M}(X)) \geq 1/2.$$

La mediana siempre existe.

Una propiedad de la mediana está dada por el siguiente resultado.

Lema 10 *Sea X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$ tal que $\mathbf{P}(X \leq \mathbf{M}(X)) = \mathbf{P}(X \geq \mathbf{M}(X)) = \frac{1}{2}$. Entonces la mediana $\mathbf{M}(X)$ minimiza $\mathbf{E}|X - C|$ como función de C.*

Proposición 11 *Sea X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$ con segundo momento y tal que $\mathbf{P}(X \leq \mathbf{M}(X)) = \mathbf{P}(X \geq \mathbf{M}(X)) = \frac{1}{2}$. Entonces*

$$|\mathbf{M}(X) - \mathbf{E}(X)| \leq \sqrt{\mathrm{Var}(X)}.$$

Teorema 12 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes en $(\Omega, \mathcal{F}, \mathbf{P})$. Si X_i es no negativa o $\mathbf{E}(X_i)$ existe para toda $i = 1, 2, \ldots, n$, entonces $\mathbf{E}(X_1 \cdots X_n)$ existe y es igual a $\mathbf{E}(X_1) \cdots \mathbf{E}(X_n)$.*

Lema 13 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes en $(\Omega, \mathcal{F}, \mathbf{P})$ con $X_i \sim N(\mu_i, \sigma_i^2)$, i.e. X_i tiene distribución normal con media μ_i y varianza σ_i^2, $1 \leq i \leq n$ y sean $c_1, \ldots, c_n$ constantes reales. Entonces*

$$c_1 X_1 + \cdots + c_n X_n \stackrel{D}{=} N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

1.3 Desigualdades

1.3.1 Desigualdades básicas en probabilidad

Teorema 14 ***(Desigualdad de Markov).*** *Sea X variable aleatoria no negativa en $(\Omega, \mathcal{F}, \mathbf{P})$. Entonces*

$$\mathbf{P}(X \geq \lambda) \leq \frac{\mathbf{E}(X)}{\lambda},$$

para todo $\lambda > 0$.

Teorema 15 ***(Desigualdad de Chebyshev).*** *Si X es una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$ con media μ y varianza σ^2. Entonces, para todo número real $\lambda > 0$*

$$\mathbf{P}(|X - \mu| \geq \lambda) \leq \frac{\sigma^2}{\lambda^2}.$$

1.3.2 Desigualdad de Jensen

Definición 16 *Una función $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ es convexa si para todo $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ y $t \in [0, 1]$, se tiene que*

$$F((1-t)\mathbf{x} + t\mathbf{y}) \leq (1-t) F(\mathbf{x}) + tF(\mathbf{y}).$$

Teorema 17 ***(Desigualdad de Jensen).*** *Sea* $(\Omega, \mathcal{F}, \mathbf{P})$ *un espacio de probabilidad, sea* F *una función real integrable con valores en un intervalo* $I \subseteq \mathbb{R}$ *y sea* $\varphi : I \longrightarrow \mathbb{R}$ *una función convexa. Entonces*

$$\varphi\left(\int_\Omega F d\mathbf{P}\right) \leq \int_\Omega \varphi \circ F d\mathbf{P}.$$

Corolario 18 *Sean* a, b *números reales con* $a < b$, $I = [a,b]$, $F : I \longrightarrow I$ *una función integrable, y sea* $\varphi : I \longrightarrow \mathbb{R}$ *una función convexa, entonces*

$$\varphi\left(\int_a^b F(x)\frac{dx}{b-a}\right) \leq \frac{1}{b-a}\int_a^b \varphi\left(F(x)\right) dx.$$

Corolario 19 *Sea* $I \subseteq \mathbb{R}$ *un intervalo,* $\varphi : I \longrightarrow \mathbb{R}$ *una función convexa y sea* X *una variable aleatoria acotada. Entonces*

$$\varphi(\mathbf{E}(X)) \leq \mathbf{E}(\varphi(X)).$$

1.3.3 Desigualdad de Hoeffding

La desigualdad de Hoeffding proporciona una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado.

Lema 20 *Sea* X *una variable aleatoria en* $\left([a,b], \mathfrak{B}([a,b]), \mathbf{P}\right)$. *Si* X *tiene media* 0, *entonces* $\mathbf{E}e^{tX} \leq e^{t^2(b-a)^2/8}$.

Teorema 21 ***(Desigualdad de Hoeffding).*** *Sean* X_1, ..., X_n *variables aleatorias independientes y acotadas con* $a_i \leq X_i \leq b_i$. *Sea* $S_n = X_1 + \cdots + X_n$. *Entonces para todo* $\varepsilon > 0$

$$\mathbf{P}\left(S_n - \mathbf{E}S_n \geq \varepsilon\right) \leq e^{-2\varepsilon^2/D} \quad y \quad \mathbf{P}\left(S_n - \mathbf{E}S_n \leq -\varepsilon\right) \leq e^{-2\varepsilon^2/D},$$

donde $D = \sum_{i=1}^n (b_i - a_i)^2$.

Capítulo 2

Concentración de medidas de probabilidad

El objetivo de este capítulo es mostrar algunos resultados de la concentración de medidas de probabilidad. En la primer sección se describe al cubo booleano como espacio de probabilidad, además se prueban dos teoremas que hablan de cómo están concentrados los puntos y las funciones *1-Lipschitz* en el cubo booleano. En la segunda sección se presenta la desigualdad de concentración gaussiana, la cual establece que una función *1-Lipschitz* cuyas entradas sean variables aleatorias independientes con distribución normal estándar se concentra alrededor de su valor esperado. Y por último en la tercera sección de este capítulo se muestra la desigualdad de concentración de Talagrand, el cual es uno de los resultados más importantes del fenómeno de concentración de medida. A diferencia de la desigualdad de concentración gaussiana, la desigualdad de concentración de Talagrand no condiciona a que las variables aleatorias tengan distribución normal estándar, en lugar de dicha condición se pide que las variables aleatorias sean acotadas, además de pedir que la función *1-Lipschitz* sea *convexa.*

2.1 Concentración en el cubo booleano

El objetivo de esta sección es mostrar cómo se presenta el fenómeno de la concentración de medida en un espacio de probabilidad en específico. El cubo booleano, $I_n = \{0,1\}^n$, $n \geq 1$, es el conjunto de las 2^n sucesiones de longitud n cuyas entradas son 0's y 1's.

Para $\mathbf{x},\ \mathbf{y} \in\ I_n$ donde $\mathbf{x} = (x_1,\ x_2,\ \ldots,\ x_n)$ y $\mathbf{y} = (y_1,\ y_2,\ \ldots,\ y_n)$, se define la distancia de Hamming de $\mathbf{x}$ a $\mathbf{y}$ como

$$d_H(\mathbf{x}, \mathbf{y}) = |\{i : x_i \neq y_i, \quad i = 1, ..., n\}|,$$

donde, para un conjunto finito A, $|A|$ denota la cardinalidad de A ($i.e$ el número de elementos que tiene A).

En la Proposición 42 se demuestra que d es métrica.

Observemos que $(I_n, \mathcal{P}(I_n), \mathbf{P}_n)$ es un espacio de probabilidad, donde $\mathcal{P}(I_n)$ es el conjunto potencia de I_n, $\mathbf{P}_n(\{\mathbf{x}\}) = 2^{-n},\ \forall x \in\ I_n$. Nótese que si $A \subseteq I_n$ entonces $\mathbf{P}_n(A) = |A|2^{-n}$, además $\mathbf{P}_n(I_n) = |I_n|\, 2^{-n} = 2^n 2^{-n} = 1$. En virtud de que I_n es finito, todas las funciones $F : I_n \longrightarrow \mathbb{R}$ son variables aleatorias ya que todas son medibles, y todas las integrales son sumas, $i.e.$

$$\int_{I_n} F d\mathbf{P}_n = \frac{1}{2^n} \sum_{\mathbf{x} \in I_n} F(\mathbf{x})$$

2.1.1 Concentración puntual

Ya que tenemos una métrica definida en el espacio de probabilidad del cubo booleano, podemos entonces describir cómo es que se concentran los puntos de este conforme la dimensión del espacio va creciendo. Los siguientes resultados nos aclaran este hecho.

Definición 22 *La distancia de un punto* $\mathbf{x} \in \mathbb{R}^n$ *a un conjunto* $A \subseteq \mathbb{R}^n$ *se define como*

$$d(\mathbf{x}, A) = \inf_{\mathbf{y} \in A} \{d(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in A\},$$

donde $d(\cdot, \cdot)$ *es una distancia en* $\mathbb{R}^n$.

Teorema 23 *Sea A un subconjunto no vacío de I_n y sea $F : I_n \longrightarrow \mathbb{R}$, $F(x) = d_H(x, A)$. Entonces, para todo $t > 0$, se tiene que*

$$\int_{I_n} e^{tF} d\mathbf{P}_n \leq \frac{1}{\mathbf{P}_n(A)} \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4} \right)^n . \tag{2.1}$$

Demostración. Sea

$$c(t) = \frac{1}{2} + \frac{e^t + e^{-t}}{4}.$$

Para hacer la demostración usaremos inducción sobre la dimensión n del cubo. En primer lugar, si $n = 1$, tenemos que A puede consistir de uno o dos puntos. Si A tiene 1 punto, entonces $\mathbf{P}_1(A) = 1/2$ y

$$\int_{I_1} e^{tF} d\mathbf{P}_1 = \frac{1}{2}(e^{tF(0)} + e^{tF(1)}) = \frac{1}{2} + \frac{1}{2}e^t \leq 2c(t),$$

con lo que (2.1) se cumple, ahora si A tiene 2 puntos, entonces $\mathbf{P}_1(A) = 1$ y

$$\int_{I_1} e^{tF} d\mathbf{P}_1 = \frac{1}{2}(e^{tF(0)} + e^{tF(1)}) = \frac{1}{2}(e^{t(0)} + e^{t(0)}) = 1 \leq c(t)$$

y se cumple (2.1).

Ahora supongamos que $n > 1$. Podemos separar el cubo I_n de la siguiente manera $I_n = I_n^1 \cup I_n^0$ donde

$$I_n^1 = \{\mathbf{x} \in I_n : x_n = 1\} \text{ e } I_n^0 = \{\mathbf{x} \in I_n : x_n = 0\}.$$

Notemos que I_n^1 e I_n^0 se pueden identificar con el cubo I_{n-1}. Definamos los subconjuntos $A_1, A_0 \subset I_{n-1}$ con

$$A_0 = \{\mathbf{x} \in I_{n-1} : (\mathbf{x}, 0) \in A\} \text{ y } A_1 = \{\mathbf{x} \in I_{n-1} : (\mathbf{x}, 1) \in A\}.$$

Tenemos que $|A| = |A_0| + |A_1|$, o en términos de medida

$$\mathbf{P}_n(A) = \frac{\mathbf{P}_{n-1}(A_0) + \mathbf{P}_{n-1}(A_1)}{2}.$$

Para un punto $\mathbf{x} \in I_n$, sea $\bar{\mathbf{x}} \in I_{n-1}$ el punto $\mathbf{x}$ pero omitiendo su última coordenada. Con lo anterior se tiene que

$$d_H(\mathbf{x}, A) = \min\{d_H(\bar{\mathbf{x}}, A_1), d_H(\bar{\mathbf{x}}, A_0) + 1\}$$

para todo $\mathbf{x} \in I_n^1$ y

$$d_H(\mathbf{x}, A) = \min\{d_H(\bar{\mathbf{x}},A_0), d_H(\bar{\mathbf{x}},A_1) + 1\}$$

para todo $\mathbf{x} \in I_n^0$.

Definamos F_0 y F_1 por $F_0(\mathbf{x}) = \ d_H(\mathbf{x},A_0)$ y $F_1(\mathbf{x}) = \ d_H(\mathbf{x},A_1)$ para $\mathbf{x} \in I_{n-1}$. Entonces

$$\begin{aligned}
\int_{I_n} e^{tF} d\mathbf{P}_n &= \int_{I_n^1} e^{tF} d\mathbf{P}_n + \int_{I_n^0} e^{tF} d\mathbf{P}_n = \frac{1}{2^n} \sum_{\mathbf{x} \in I_n^1} \exp\{t d_H(\mathbf{x}, A)\} + \frac{1}{2^n} \sum_{\mathbf{x} \in I_n^0} \exp\{t d_H(\mathbf{x}, A)\} \\
&= \frac{1}{2^n} \sum_{\mathbf{x} \in I_n^1} \exp\{\min\{t d_H(\bar{\mathbf{x}}, A_1), t + t d_H(\bar{\mathbf{x}}, A_0)\}\} \\
&\quad + \frac{1}{2^n} \sum_{\mathbf{x} \in I_n^0} \exp\{\min\{t d_H(\bar{\mathbf{x}}, A_0), t + t d_H(\bar{\mathbf{x}}, A_1)\}\} \\
&= \frac{1}{2^n} \sum_{\mathbf{x} \in I_n^1} \min\{\exp\{t d_H(\bar{\mathbf{x}}, A_1), e^t \exp\{t d_H(\bar{\mathbf{x}}, A_0)\}\} \\
&\quad + \frac{1}{2^n} \sum_{\mathbf{x} \in I_n^0} \min\{\exp\{t d_H(\bar{\mathbf{x}}, A_0)\}, e^t \exp\{t d_H(\bar{\mathbf{x}}, A_1)\}\} \\
&= \frac{1}{2} (\frac{1}{2^{n-1}} \sum_{\mathbf{x} \in I_n^1} \min\{\exp\{t d_H(\bar{\mathbf{x}}, A_1), e^t \exp\{t d_H(\bar{\mathbf{x}}, A_0)\}\}) \\
&\quad + \frac{1}{2} (\frac{1}{2^{n-1}} \sum_{\mathbf{x} \in I_n^0} \min\{\exp\{t d_H(\bar{\mathbf{x}}, A_0)\}, e^t \exp\{t d_H(\bar{\mathbf{x}}, A_1)\}\}) \\
&= \frac{1}{2} \int_{I_{n-1}} \min\{e^{tF_1}, e^t e^{tF_0}\} d\mathbf{P}_{n-1} + \frac{1}{2} \int_{I_{n-1}} \min\{e^{tF_0}, e^t e^{tF_1}\} d\mathbf{P}_{n-1}.
\end{aligned}$$

La integral del mínimo es siempre menor que el mínimo de las integrales. Además usando la hipótesis de inducción tenemos que

$$\begin{aligned}
\int_{I_n} e^{tF} d\mathbf{P}_n &\leq \frac{1}{2} \min\left\{ \int_{I_{n-1}} e^{tF_1} d\mathbf{P}_{n-1}, e^t \int_{I_{n-1}} e^{tF_0} d\mathbf{P}_{n-1} \right\} \\
&\quad + \frac{1}{2} \min\left\{ \int_{I_{n-1}} e^{tF_0} d\mathbf{P}_{n-1}, e^t \int_{I_{n-1}} e^{tF_1} d\mathbf{P}_{n-1} \right\} \\
&\leq \frac{1}{2} \min\left\{ \frac{c^{n-1}(t)}{\mathbf{P}_{n-1}(A_1)}, e^t \frac{c^{n-1}(t)}{\mathbf{P}_{n-1}(A_0)} \right\} + \frac{1}{2} \min\left\{ \frac{c^{n-1}(t)}{\mathbf{P}_{n-1}(A_0)}, e^t \frac{c^{n-1}(t)}{\mathbf{P}_{n-1}(A_1)} \right\} \\
&= \frac{c^{n-1}(t)}{\mathbf{P}_n(A)} \left(\frac{1}{2} \min\left\{ \frac{\mathbf{P}_n(A)}{\mathbf{P}_{n-1}(A_1)}, e^t \frac{\mathbf{P}_n(A)}{\mathbf{P}_{n-1}(A_0)} \right\} \right. \\
&\quad \left. + \frac{1}{2} \min\left\{ \frac{\mathbf{P}_n(A)}{\mathbf{P}_{n-1}(A_0)}, e^t \frac{\mathbf{P}_n(A)}{\mathbf{P}_{n-1}(A_1)} \right\} \right).
\end{aligned}$$

Escribamos

$$a_0 = \frac{\mathbf{P}_{n-1}(A_0)}{\mathbf{P}_n(A)} \text{ y } a_1 = \frac{\mathbf{P}_{n-1}(A_1)}{\mathbf{P}_n(A)}$$

con lo que $a_0, a_1 \geq 0$ y $a_0 + a_1 = 2$. Entonces todo se reduce a hacernos la pregunta de cuál es el máximo valor posible de

$$\frac{1}{2}\min\left\{a_1^{-1}, e^t a_0^{-1}\right\} + \frac{1}{2}\min\{a_0^{-1}, e^t a_1^{-1}\} \tag{2.2}$$

Si $a_0 = 0$ y $a_1 = 2$, el valor de (2.2) es $\frac{1}{4} + \frac{e^t}{4} \leq c(t)$ si $a_0 = 2$ y $a_1 = 0$ se obtiene el mismo valor. Ahora supongamos que $a_0, a_1 > 0$. Si tuviéramos que $a_0 = a_1 = 1$ el valor de (2.2) sería $\frac{1}{2} + \frac{1}{2} \leq c(t)$, entonces, sin pérdida de generalidad, supongamos que $a_0 = 1 + b$ y $a_1 = 1 - b$ para algun $0 < b < 1$. Entonces $a_0^{-1} < e^t a_1^{-1}$, si $a_1^{-1} < e^t a_0^{-1}$ el valor de (2.2) sería

$$\frac{1}{2}(1-b)^{-1} + \frac{1}{2}(1+b)^{-1} = \frac{1}{1-b^2}$$

si incrementamos el valor de b, estaríamos incrementando el valor de (2.2), lo cual es una contradicción. Si $a_1^{-1} > e^t a_0^{-1}$, entonces el valor de (2.2) sería

$$\frac{1}{2}e^t a_0^{-1} + \frac{1}{2}a_0^{-1} = \frac{e^t - 1}{2\,(1+b)}$$

si disminuimos b, el valor de (2.2) aumentaría, lo cual es una contradicción. Por lo tanto en el punto máximo tenemos que

$$a_1^{-1} = e^t a_0^{-1}, \ i.e. \ a_0 = \frac{2e^t}{1+e^t} \text{ y } a_1 = \frac{2}{1+e^t}$$

con lo que el valor de (2.2) sería

$$\frac{1}{2}a_1^{-1} + \frac{1}{2}a_0^{-1} = \frac{1}{2}\left(\frac{1+e^t}{2}\right) + \frac{1}{2}\left(\frac{1+e^t}{2e^t}\right) = \frac{1}{2} + \frac{e^t + e^{-t}}{4} = c(t)$$

con lo que tenemos que

$$\int_{I_n} e^{tF} d\mathbf{P}_n \leq \frac{c^{n-1}(t)}{\mathbf{P}_n(A)} c(t) = \frac{c^n(t)}{\mathbf{P}_n(A)}$$

■

Los siguientes dos corolarios son resultados de suma importancia, pues ilustran de una manera formal el fenómeno de concentración de medida en el espacio de probabilidad del cubo booleano.

Corolario 24 *Sea* $A \subset I_n$ *un conjunto no vacío y sea* $F : I_n \longrightarrow \mathbb{R}$, $F(\mathbf{x}) = d_H(\mathbf{x}, A)$. *Entonces para todo* $t > 0$, *se tiene que*

$$\int_{I_n} e^{tF} d\mathbf{P}_n \leq \frac{1}{\mathbf{P}_n(A)} e^{t^2 n/4}.$$

Corolario 25 *Sea* $A \subset I_n$ *un conjunto no vacío. Entonces para cualquier* $\varepsilon > 0$, *tenemos que* $\mathbf{P}_n\left(\{\mathbf{x} \in I_n : d_H(\mathbf{x}, A) \geq \varepsilon\sqrt{n}\}\right) \leq \frac{e^{-\varepsilon^2}}{\mathbf{P}_n(A)}$.

Una ilustración de lo que dice el corolario anterior es la siguiente: si $\mathbf{P}_n(A) = 0.01$, *i.e.* A contiene el 1% de los puntos del cubo I_n, y escogemos $\varepsilon = 2\sqrt{\ln 10} \approx 3.035$, según el corolario $\mathbf{P}_n\left(\{\mathbf{x} \in I_n : d_H(\mathbf{x}, A) \geq 3.035\sqrt{n}\}\right) \leq \frac{e^{-4\ln 10}}{0.01} = 0.01$, con lo que concluimos que el 99% de los puntos del cubo I_n están a una distancia menor que $3.035\sqrt{n}$ de A. Tomando en cuenta que $\max\{d_H(\mathbf{x}, \mathbf{y}) | \ \mathbf{x}, \mathbf{y} \in I_n\} = n$, resulta sorprendente entonces cómo los resultados arriba mostrados nos dicen que conforme crece la dimensión del cubo booleano, todos los puntos de éste están muy cerca los unos de los otros.

2.1.2 Concentración de funciones

En nuestro siguiente resultado mostraremos cómo es que al crecer la dimensión del espacio (en este caso del cubo booleano) ciertas funciones se "concentran" alrededor de su mediana.

Definición 26 *Una función* $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ *es* 1-Lipschitz si

$$|F(\mathbf{x}) - F(\mathbf{y})| \leq |\mathbf{x} - \mathbf{y}|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

donde usamos la métrica euclideana sobre $\mathbb{R}^n$.

Teorema 27 *Sea* $F : I_n \longrightarrow \mathbb{R}$, *una función* 1-Lipschitz. *Sea* $\mathbf{M}(F)$ *la mediana de* F. *Entonces para todo* $\varepsilon > 0$,

$$\mathbf{P}_n\left(\left\{\mathbf{x} \in I_n : |F(\mathbf{x}) - \mathbf{M}(F)| \ \geq \varepsilon\sqrt{n}\right\}\right) \leq 4e^{-\varepsilon^2}.$$

Demostración. Consideremos los siguientes dos conjuntos:

$$A_+ = \{\mathbf{x} \in I_n : F(\mathbf{x}) \geq \mathbf{M}(F)\} \text{ y } A_- = \{\mathbf{x} \in I_n : F(\mathbf{x}) \leq \mathbf{M}(F)\},$$

entonces $\mathbf{P}_n(A_+)$, $\mathbf{P}_n(A_-) \geq 1/2$. Sea $A_+(\varepsilon)$ la $\varepsilon\sqrt{n}$-vecindad de A_+ y $A_-(\varepsilon)$ la $\varepsilon\sqrt{n}$-vecindad de A_-, *i.e.*

$$A_+(\varepsilon) = \{\mathbf{x} \in I_n : d_H(\mathbf{x}, A_+) < \varepsilon\sqrt{n}\} \text{ y } A_-(\varepsilon) = \{\mathbf{x} \in I_n : d_H(\mathbf{x}, A_-) < \varepsilon\sqrt{n}\}.$$

Por el Corolario 25

$$\mathbf{P}_n(A_+(\varepsilon)) > 1 - 2e^{-\varepsilon^2} \text{ y } \mathbf{P}_n(A_-(\varepsilon)) > 1 - 2e^{-\varepsilon^2},$$

por lo tanto, para $A(\varepsilon) = A_+(\varepsilon) \cap A_-(\varepsilon)$ tenemos que

$$\mathbf{P}_n(A(\varepsilon)) > 1 - 4e^{-\varepsilon^2}. \tag{2.3}$$

Ahora si $\mathbf{x} \in A(\varepsilon)$, entonces $\mathbf{x} \in A_+(\varepsilon) \cap A_-(\varepsilon)$, entonces como $\mathbf{x} \in A_+(\varepsilon)$, entonces existe $\mathbf{y}_1 \in A_+$ tal que $F(\mathbf{y}_1) \geq \mathbf{M}(F)$ y además

$$\mathbf{M}(F) - F(\mathbf{x}) \leq F(\mathbf{y}_1) - F(\mathbf{x}) \leq d_H(\mathbf{x}, \mathbf{y}_1) < \varepsilon\sqrt{n}, \tag{2.4}$$

y como $\mathbf{x} \in A_-(\varepsilon)$ entonces existe $\mathbf{y}_2 \in A_-$ tal que $F(\mathbf{y}_2) \leq \mathbf{M}(F)$ y

$$F(\mathbf{x}) - \mathbf{M}(F) \leq F(\mathbf{x}) - F(\mathbf{y}_2) \leq d_H(\mathbf{x}, \mathbf{y}_2) < \varepsilon\sqrt{n}, \tag{2.5}$$

por lo que por (2.4) y (2.5) concluimos que si $\mathbf{x} \in A(\varepsilon)$ entonces $|F(\mathbf{x}) - \mathbf{M}(F)| \ < \varepsilon\sqrt{n}$, de donde obtenemos que

$$A(\varepsilon) \subseteq \left\{\mathbf{x} \in I_n : |F(\mathbf{x}) - \mathbf{M}(F)| \ < \varepsilon\sqrt{n}\right\},$$

de donde por (2.3) y esta última contención se sigue que

$$\mathbf{P}_n \left(\{ \mathbf{x} \in I_n : |F(\mathbf{x}) - \mathbf{M}(F)| \ < \varepsilon\sqrt{n} \} \right) \geq 1 - 4e^{-\varepsilon^2}$$

con lo que se concluye que

$$\mathbf{P}_n \left(\{ \mathbf{x} \in I_n : |F(\mathbf{x}) - \mathbf{M}(F)| \ \geq \varepsilon\sqrt{n} \} \right) \leq 4e^{-\varepsilon^2}.$$

■

Para ejemplificar el resultado anterior, sea $\varepsilon = \sqrt{\ln 400} \approx 2.45$, entonces

$$\mathbf{P}_n \left(\{ \mathbf{x} \in I_n : |F(\mathbf{x}) - \mathbf{M}(F)| \ \geq 2.45\sqrt{n} \} \right) \leq 4e^{-\ln 400} = 0.01,$$

con lo que concluimos que para el 99% de los puntos de I_n, la función F no se desvía de su mediana por más de $2.45\sqrt{n}$.

2.2 Desigualdad de concentración gaussiana

En esta sección se presenta un resultado de concentración de medida de probabilidad, nombrado como desigualdad de concentración gaussiana. Este resultado establece que si tenemos una función *1-Lipschitz* cuyas entradas sean variables aleatorias independientes con distribución normal estándar, entonces ésta se concentrará alrededor de su esperanza.

Teorema 28 ***(Desigualdad de concentración gaussiana)****.* *Sea $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas, $X_1 \sim N(0,1)$, sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vector aleatorio y sea $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ una función* 1-Lipschitz *diferenciable. Entonces existe una constante positiva c tal que para todo $\lambda > 0$ tenemos que*

$$\mathbf{P}\left(|F(\mathbf{X}) - \mathbf{E}F(\mathbf{X})| \geq \lambda \right) \leq 2\exp\left(-c\lambda^2\right).$$

Demostración. Supongamos que $\mathbf{E}F(\mathbf{X}) = 0$, entonces es suficiente probar que

$$\mathbf{P}(F(\mathbf{X}) \geq \lambda) \leq \exp\left(-c\lambda^2\right), \tag{2.6}$$

ya que al ser $-F$ 1-Lipschitz, por (2.6) se tendría que

$$\mathbf{P}(F(\mathbf{X}) \leq -\lambda) = \mathbf{P}(-F(\mathbf{X}) \geq \lambda) \leq \exp\left(-c\lambda^2\right),$$

en consecuencia se sigue que

$$\mathbf{P}(|F(\mathbf{X})| \geq \lambda) = \mathbf{P}(F(\mathbf{X}) \geq \lambda) + \mathbf{P}(F(\mathbf{X}) \leq -\lambda) \leq 2\exp\left(-c\lambda^2\right).$$

La diferenciabilidad de F y la cota de Lipschitz sobre F implica que

$$|\nabla F(\mathbf{x})| \leq 1, \tag{2.7}$$

para todo $\mathbf{x} \in \mathbb{R}^n$.

Tenemos que, para todo $t > 0$,

$$\mathbf{P}(F(\mathbf{X}) \geq \lambda) = \mathbf{P}\left(e^{tF(\mathbf{X})} \geq e^{t\lambda}\right) \leq \frac{\mathbf{E}e^{tF(\mathbf{X})}}{e^{t\lambda}},$$

por la desigualdad de Markov, por lo que es suficiente probar que

$$\mathbf{E}e^{tF(\mathbf{X})} \leq e^{c_0 t^2}, \tag{2.8}$$

para alguna constante $c_0 > 0$ y para todo $t > 0$. Ya que se tendría que

$$\mathbf{P}(F(\mathbf{X}) \geq \lambda) \leq e^{c_0 t^2 - t\lambda},$$

cuyo valor mínimo se obtiene en $t = \frac{\lambda}{2c_0}$ lo que da que

$$\begin{aligned}
\mathbf{P}(F(\mathbf{X}) \geq \lambda) &\leq \exp\left(c_0 \frac{\lambda^2}{4c_0^2} - \frac{\lambda^2}{2c_0}\right) \\
&= \exp\left(-\frac{\lambda^2}{4c_0}\right) \\
&= \exp\left(-c\lambda^2\right).
\end{aligned}$$

Sea $\mathbf{Y}$ una copia independiente de $\mathbf{X}$, como $\mathbf{E}F(\mathbf{Y}) = 0$, tenemos que por la desigualdad de Jensen (Corolario 19)

$$\mathbf{E}\exp(-tF(\mathbf{Y})) \geq \exp(-t\mathbf{E}F(\mathbf{Y})) = 1,$$

por lo que

$$\mathbf{E}\exp(tF(\mathbf{X})) \leq \mathbf{E}\exp(t(F(\mathbf{X}) - F(\mathbf{Y}))).$$

Luego, gracias al Teorema fundamental del cálculo, tenemos que

$$F(\mathbf{X}) - F(\mathbf{Y}) = \int_0^{\pi/2} \frac{d}{d\theta} F(\mathbf{Y}\cos\theta + \mathbf{X}\sin\theta)\, d\theta.$$

Sea $\mathbf{X}_\theta = \mathbf{Y}\cos\theta + \mathbf{X}\sin\theta$, observemos que $\mathbf{X}_\theta$ es un vector aleatorio gaussiano (Lema 13), lo mismo que su derivada $\mathbf{X}'_\theta = -\mathbf{Y}\sin\theta + \mathbf{X}\cos\theta$.

Usando el Corolario 18 de la Desigualdad de Jensen, se sigue que

$$\begin{aligned} \exp(t[F(\mathbf{X}) - F(\mathbf{Y})]) &= \exp\left(\left[\int_0^{\pi/2} \left(\frac{\pi t}{2}\frac{d}{d\theta}F(\mathbf{X}_\theta)\right)\frac{d\theta}{\pi/2}\right]\right) \\ &\leq \frac{2}{\pi}\int_0^{\pi/2} \exp\left(\frac{\pi t}{2}\frac{d}{d\theta}F(\mathbf{X}_\theta)\right) d\theta, \end{aligned}$$

aplicando la regla de la cadena y tomando las esperanzas a ambos lados de la desigualdad:

$$\mathbf{E}\exp(t(F(\mathbf{X}) - F(\mathbf{Y}))) \leq \frac{2}{\pi}\mathbf{E}\int_0^{\pi/2} \exp\left(\frac{\pi t}{2}\nabla F(\mathbf{X}_\theta)\cdot \mathbf{X}'_\theta\right) d\theta,$$

de donde, aplicando el Teorema Fubini-Tonnelli se tiene que

$$\mathbf{E}\int_0^{\pi/2} \exp\left(\frac{\pi t}{2}\nabla F(\mathbf{X}_\theta)\cdot \mathbf{X}'_\theta\right) d\theta = \int_0^{\pi/2} \mathbf{E}\exp\left(\frac{\pi t}{2}\nabla F(\mathbf{X}_\theta)\cdot \mathbf{X}'_\theta\right) d\theta,$$

de donde obtenemos que

$$\mathbf{E}\exp(t(F(\mathbf{X}) - F(\mathbf{Y}))) \leq \frac{2}{\pi}\int_0^{\pi/2} \mathbf{E}\exp\left(\frac{\pi t}{2}\nabla F(\mathbf{X}_\theta)\cdot \mathbf{X}'_\theta\right) d\theta.$$

Como ya se había dicho anteriormente, $\mathbf{X}'_\theta$ es un vector aleatorio gaussiano, por lo que, según el Lema 13 $\frac{\pi}{2}\nabla F(\mathbf{X}_\theta)\cdot \mathbf{X}'_\theta$ también es normal con desviación estándar σ menor o igual a $\frac{\pi}{2}\left|\nabla F(\mathbf{X}_\theta)\right|$, y por (2.7), tenemos que $\sigma \leq \frac{\pi}{2}$, por lo que

$$\begin{aligned}
\mathbf{E}\exp\left(t\left(F(\mathbf{X})-F(\mathbf{Y})\right)\right) &\leq \frac{2}{\pi}\int_0^{\pi/2}\mathbf{E}\exp\left(t\frac{\pi}{2}\nabla F(\mathbf{X}_\theta)\cdot\mathbf{X}_\theta'\right)d\theta \\
&= \frac{2}{\pi}\int_0^{\pi/2}e^{\frac{1}{2}\sigma^2t^2}d\theta \\
&= e^{\frac{1}{2}\sigma^2t^2}\leq\exp\left(\frac{\pi^2}{8}t^2\right)=\exp\left(c_0t^2\right),
\end{aligned}$$

donde $c_0 = \pi^2/8$ de donde obtenemos (2.8).

Por último, si $\mathbf{E}F(X) \neq 0$ definamos la variable aleatoria $G(X) = F(X) - \mathbf{E}F(X)$. Observemos que G es una función 1-Lipschitz diferenciable con $\mathbf{E}G(X) = 0$, de donde obtenemos el resultado deseado. ∎

2.3 Desigualdad de concentración de Talagrand

El resultado de concentración de medida que presentamos en esta sección es uno de los resultados más importantes del fenómeno de la concentración de medida. A diferencia de la desigualdad de concentración gaussiana, en la desigualdad de concentración de Talagrand no se pide que las variables aleatorias tengan alguna distribución en especial, se pide la condición adicional de que la función 1-Lipschitz sea convexa además de que las variables aleatorias sean acotadas. Este resultado fue probado por primera vez por Talagrand [12], y aquí lo probaremos usando la misma idea de la prueba hecha por él. El resultado se basa en la hipótesis de que las funciones son convexas, por lo que usaremos varios resultados sobre envolventes convexas para probar nuestro teorema principal.

2.3.1 Envolventes convexas

En la demostración de la Desigualdad de concentración de Talagrand utilizaremos un lema muy importante: el Lema 37. El cual se demuestra en esta subsección. Para la demostración de este lema utilizaremos algunos resultados previos acerca de envolventes convexas. Por lo que se presentan primero algunas definiciones referentes a convexidad, como lo es la distancia combinatoria, además de presentar un resultado que nos dice que el doble de la distancia combinatoria de un punto a un conjunto convexo es mayor que la distancia euclideana de ese punto a dicho conjunto, hecho que servirá para simplificar la demostración del Lema 37.

Definición 29 *Un conjunto $A \subseteq \mathbb{R}^n$ es **convexo** si y solo si, para todo $\mathbf{x},\ \mathbf{y} \in A$ se cumple que*

$$t\mathbf{x} + (1-t)\mathbf{y} \in A,$$

para todo $t \in [0,1]$.

Definición 30 *La **envolvente convexa** $W(A)$ de un conjunto $A \subseteq \mathbb{R}^n$, es el conjunto convexo más pequeño que contiene a A (i.e. si W_1 es un conjunto convexo tal que $A \subseteq W_1$, entonces $W(A) \subseteq W_1$).*

Si $A \subseteq \mathbb{R}^n$ es un conjunto finito, y $A = \{\mathbf{a}_1,\ \mathbf{a}_2,\ \ldots, \mathbf{a}_k\}$ entonces su envolvente convexa se puede escribir como

$$W(A) = \left\{ \sum_{i=1}^{k} \alpha_i w_i : w_i \in A \text{ y } \sum_{i=1}^{k} \alpha_i = 1 \right\},$$

para ver la demostración de este hecho, el lector interesado puede consultar el Corolario 2.3.1 de [10].

Definición 31 *Sea $\mathbf{x} \in \mathbb{R}^n$ con $\mathbf{x} = (x_1,\ x_2,\ \ldots,\ x_n)$, definimos el **soporte combinatorio** de un conjunto $A \subseteq \mathbb{R}^n$ relativo a $\mathbf{x}$ como*

$$U_A(\mathbf{x}) = \{\mathbf{w} \in \{0,1\}^n \mid \exists \mathbf{z} \in A \text{ tal que } z_i = x_i \text{ si } w_i = 0,\ 1 \leq i \leq n\},$$

donde $\mathbf{z} = (z_1,\ z_2,\ \ldots, z_n)$ y $\mathbf{w} = (w_1,\ w_2,\ \ldots, w_n)$. Diremos que $\mathbf{w}$ soporta a $\mathbf{z}$ *si $z_i = x_i$ cuando $w_i = 0$, para toda i, $1 \leq i \leq n$.*

Definición 32 *Sea $A \subseteq \mathbb{R}^n$ y sea $\mathbf{x} \in \mathbb{R}^n$ entonces se define la **envolvente combinatoria** $V_A(\mathbf{x})$ de A relativa a $\mathbf{x}$ a la envolvente convexa de $U_A(\mathbf{x})$.*

Proposición 33 *Sea* $U_A(\mathbf{x})$ *como en la Definición 31, entonces* $\mathbf{0} \in U_A(\mathbf{x})$ *si y solo si* $\mathbf{x} \in A$.

Definición 34 *Definimos la* ***distancia combinatoria*** $d_c(\mathbf{x}, A)$ *como la distancia entre* $V_A(\mathbf{x})$ *y el origen, es decir,*

$$d_c(\mathbf{x}, A) = d(\mathbf{0}, V_A(\mathbf{x})) = \inf_{\mathbf{y} \in V_A(\mathbf{x})} |\mathbf{y}| .$$

Lema 35 *Sea* $\mathbf{I} = [-1, 1] \subset \mathbb{R}$ *y* A *un subconjunto convexo de* $\mathbf{I}^n \subset \mathbb{R}^n$, *y sea* $\mathbf{x} \in \mathbf{I}^n$, *entonces*

$$d(\mathbf{x}, A) \leq 2d_c(\mathbf{x}, A).$$

Lema 36 *Sea* $\mathbf{x} = (\bar{\mathbf{x}}, x_n) \in \mathbf{I}^n$ *con* $\bar{\mathbf{x}} \in \mathbf{I}^{n-1}$ *y,* $x_n \in \mathbf{I}$. *Sea* $A \subseteq \mathbf{I}^n$, *definamos los siguientes subconjuntos de* $\mathbf{I}^{n-1}$, $A_{x_n} := \{\bar{\mathbf{z}} \in \mathbf{I}^{n-1} : (\bar{\mathbf{z}}, x_n) \in A\}$, $B := \{\bar{\mathbf{z}} \in \mathbf{I}^{n-1} : (\bar{\mathbf{z}}, t) \in A$ *para algún* $t \in \mathbf{I}\}$. *Entonces tenemos que para cualquier* $0 \leq \lambda \leq 1$, *se cumple que*

$$d_c^2\left(\mathbf{x}, A\right) \leq (1-\lambda)^2 + \lambda d_c^2(\bar{\mathbf{x}}, A_{x_n}) + (1-\lambda) d_c^2(\bar{\mathbf{x}}, B). \tag{2.9}$$

Lema 37 *Sea* $\mathbf{X} = (X_1,\ X_2,\ \ldots,\ X_n)$ *un vector aleatorio en* $(\mathbf{I}^n, \mathfrak{B}(\mathbf{I}^n), \mathbf{P}_n)$ *y sea* A *un subconjunto convexo de* $\mathbf{I}^n$. *Entonces existe* $c > 0$ *tal que*

$$\mathbf{E} \exp(c d^2\left(\mathbf{X}, A\right)) \leq \frac{1}{\mathbf{P}_n(\mathbf{X} \in A)}. \tag{2.10}$$

Demostración. Para probar (2.10) es suficiente probar, por el Lema 35, que

$$\mathbf{E} \exp(\bar{c} d_c^2\left(\mathbf{X}, A\right)) \leq \frac{1}{\mathbf{P}_n(\mathbf{X} \in A)}, \tag{2.11}$$

haciendo $\bar{c} = 4c$.

Verifiquemos primero el caso de dimensión 1. En este caso, $d_c(\mathbf{X}, A)$ es igual a 1 cuando $\mathbf{X} \notin A$ y 0 si $\mathbf{X} \in A$, por lo que $d_c(\mathbf{X}, A) \leq 1$ por lo que (2.11) se sigue de tomar $\bar{c} =$

$\ln\left(\frac{1}{\mathbf{P}_n(\mathbf{X}\in A)}\right)$. Ahora supongamos que $n>1$ y que el resultado funciona para $n-1$. Fijemos X_n y tomemos la esperanza condicional

$$\mathbf{E}\left(\exp\left(\bar{c}d_c^2\left(\mathbf{X},A\right)\right)\mid X_n\right),$$

usando el Lema 36 (para algún $\lambda\in(0,1)$ dependiendo de X_n) se tiene que

$$\mathbf{E}\exp\left(\bar{c}d_c^2\left(\mathbf{X},A\right)\right)\leq e^{\bar{c}(1-\lambda)^2}\mathbf{E}\left(\left(e^{\bar{c}d_c(\bar{\mathbf{X}},A_{X_n})^2}\right)^{\lambda}\left(e^{\bar{c}d_c(\bar{\mathbf{X}},B)^2}\right)^{1-\lambda}\mid X_n\right),$$

aplicando la desigualdad de Hölder, la parte derecha de la desigualdad anterior es acotada por

$$e^{\bar{c}(1-\lambda)^2}\mathbf{E}\left(e^{\bar{c}d_c(\bar{\mathbf{X}},A_{X_n})^2}\mid X_n\right)^{\lambda}\mathbf{E}\left(e^{\bar{c}d_c(\bar{\mathbf{X}},B)^2}\mid X_n\right)^{1-\lambda},$$

y por la hipótesis de inducción podemos acotar lo anterior con

$$e^{\bar{c}(1-\lambda)^2}\frac{1}{\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in A_{X_n}\mid X_n)^{\lambda}\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in B\mid X_n)^{1-\lambda}},$$

lo cual podemos reescribir como (nótese que el evento $\bar{\mathbf{X}}\in B$ es independiente de X_n)

$$\frac{1}{\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in B)}e^{\bar{c}(1-\lambda)^2}r^{-\lambda},$$

donde $r:=\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in A_{X_n}\mid X_n)/\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in B)$. Observemos que $0\leq r\leq 1$, por lo que podemos aplicar la desigualdad elemental (ver Proposición 43)

$$\inf_{0\leq\lambda\leq 1}e^{\bar{c}(1-\lambda)^2}r^{-\lambda}\leq 2-r,$$

con lo que concluimos que

$$\mathbf{E}\left(\exp\left(\bar{c}d_c^2\left(\mathbf{X},A\right)\right)\mid X_n\right)\leq\frac{1}{\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in B)}\left(2-\frac{\mathbf{P}_{n-1}(\bar{\mathbf{X}}\in A_{X_n}\mid X_n)}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}}\in B\right)}\right),$$

calculando esperanzas con respecto a X_n tenemos que

$$\mathbf{E}\left(\exp\left(\bar{c}d_c^2\left(\mathbf{X},A\right)\right)\right)\leq\frac{1}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}}\in B\right)}\left(2-\frac{\mathbf{P}_n(\mathbf{X}\in A)}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}}\in B\right)}\right),$$

usando la desigualdad $x(2-x)\leq 1, 0\leq x\leq 2$ con $x:=\frac{\mathbf{P}_n(X\in A)}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}}\in B\right)}$ tenemos

$$\frac{\mathbf{P}_n\left(\mathbf{X}\in A\right)}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}}\in B\right)}\left(2-\frac{\mathbf{P}_n\left(\mathbf{X}\in A\right)}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}}\in B\right)}\right)\leq 1,$$

de donde

$$\frac{1}{\mathbf{P}_{n-1}(\bar{\mathbf{X}} \in B)} \left(2 - \frac{\mathbf{P}_n\left(\mathbf{X} \in A\right)}{\mathbf{P}_{n-1}\left(\bar{\mathbf{X}} \in B\right)} \right) \leq \frac{1}{\mathbf{P}_n\left(\mathbf{X} \in A\right)},$$

lo que concluye la inducción. ■

2.3.2 Desigualdad de Talagrand

En esta sección se demuestra uno de los resultados más importantes del fenómeno de concentración de medida el cual está dado por la desigualdad de Talagrand. Con esta desigualdad probaremos la concentración de la norma operador y la convergencia casi segura del Teorema de Wigner para matrices aleatorias.

Teorema 38 ***(Desigualdad de concentración de Talagrand).*** *Sea $K > 0$, y sean $X_1,\ X_2,\ \ldots,\ X_n$ variables aleatorias independientes con $|X_i| \leq K$ para todo $1 \leq i \leq n$. Sea $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ una función* 1-Lipschitz *convexa. Entonces, existen constantes positivas c y C tales que para todo $\lambda > 0$ tenemos que*

$$\mathbf{P}\left(|F(\mathbf{X}) - \mathbf{M}\left(F\left(\mathbf{X}\right)\right)| \geq \lambda K\right) \leq C \exp\left(-c\lambda^2\right), \tag{2.12}$$

y

$$\mathbf{P}\left(|F(\mathbf{X}) - \mathbf{E}\left(F\left(\mathbf{X}\right)\right)| \geq \lambda K\right) \leq C \exp\left(-c\lambda^2\right), \tag{2.13}$$

donde $\mathbf{X} = (X_1,\ X_2,\ \ldots,\ X_n)$ y $\mathbf{M}\left(F\left(\mathbf{X}\right)\right)$ es la mediana de $F(\mathbf{X})$ (ver Definición 9).

Demostración. Podemos suponer, sin pérdida de generalidad, que $K = 1$. En este caso $\mathbf{X}$ toma valores en el conjunto convexo $\mathbf{I}^n \subset \mathbb{R}^n$. Para $x \in \mathbb{R}$, sea el conjunto convexo

$$A := \left\{\mathbf{z} \in \mathbf{I}^n \mid F(\mathbf{z}) \leq x\right\},$$

sea $y = x + \lambda$, $\lambda > 0$, entonces

$$A_\lambda = \left\{\mathbf{z} \in \mathbf{I}^n \mid F(\mathbf{z}) < x + \lambda\right\},$$

por lo que

$$A_\lambda^c = \left\{\mathbf{z} \in \mathbf{I}^n \mid F(\mathbf{z}) \geq y\right\} \subseteq \left\{\mathbf{z} \in \mathbf{I}^n \mid d(\mathbf{z}, A) \geq \lambda\right\},$$

Ahora tenemos que

$$\mathbf{P}\left(\mathbf{X} \notin A_\lambda\right) \leq \mathbf{P}\left(d(\mathbf{X}, A) \geq \lambda\right) = \mathbf{P}\left(cd^2\left(\mathbf{X}, A\right) \geq c\lambda^2\right) = \mathbf{P}\left(e^{cd^2(\mathbf{X}, A)} \geq e^{c\lambda^2}\right),$$

por la desigualdad de Markov se tiene que

$$\mathbf{P}\left(e^{cd^2(\mathbf{X}, A)} \geq e^{c\lambda^2}\right) \leq e^{-c\lambda^2}\mathbf{E}\exp\left(cd^2\left(\mathbf{X}, A\right)\right),$$

con lo que, aplicando (2.10)

$$e^{-c\lambda^2}\mathbf{E}\exp\left(cd^2\left(\mathbf{X}, A\right)\right) \leq e^{-c\lambda^2}\mathbf{P}\left(\mathbf{X} \in A\right)^{-1},$$

por lo tanto

$$\mathbf{P}\left(\mathbf{X} \in A\right)\mathbf{P}\left(\mathbf{X} \notin A_\lambda\right) \leq e^{-c\lambda^2},$$

si tenemos que $x < y$, entonces

$$\mathbf{P}\left(F(\mathbf{X}) \leq x\right)\mathbf{P}\left(F(\mathbf{X}) \geq y\right) \leq e^{-c\lambda^2}, \tag{2.14}$$

hagamos $x = \mathbf{M}\left(F\left(\mathbf{X}\right)\right)$ y $y = \mathbf{M}\left(F\left(\mathbf{X}\right)\right) + \lambda$, entonces

$$\mathbf{P}\left(F(\mathbf{X}) \leq \mathbf{M}\left(F\left(\mathbf{X}\right)\right)\right)\mathbf{P}\left(F(\mathbf{X}) \geq \mathbf{M}\left(F\left(\mathbf{X}\right)\right) + \lambda\right) \leq e^{-c\lambda^2},$$

Dado que $\mathbf{P}\left(F(\mathbf{X}) \leq \mathbf{M}\left(F\left(\mathbf{X}\right)\right)\right) \geq \frac{1}{2}$, se sigue entonces que

$$\mathbf{P}\left(F(\mathbf{X}) - \mathbf{M}\left(F\left(\mathbf{X}\right)\right) \geq \lambda\right) \leq 2e^{-c\lambda^2},$$

Consideremos $-F\left(X\right)$, se cumple que $M\left(-F\left(X\right)\right) = -M\left(F\left(X\right)\right)$ por lo que $\mathbf{P}(-F(\mathbf{X}) \geq -\mathbf{M}\left(F\left(\mathbf{X}\right)\right) + \lambda) = \mathbf{P}\left(F(\mathbf{X}) \leq \mathbf{M}\left(F\left(\mathbf{X}\right)\right) - \lambda\right)$, si hacemos $x = \mathbf{M}\left(F\left(\mathbf{X}\right)\right) - \lambda$ y $y = \mathbf{M}\left(F\left(\mathbf{X}\right)\right)$
sustituyendo en (2.14) concluimos que

$$\mathbf{P}\left(F(\mathbf{X}) - \mathbf{M}\left(F\left(\mathbf{X}\right)\right) \geq -\lambda\right) \leq 2e^{-c\lambda^2},$$

con lo que (2.12) se cumple.

Antes de demostrar (2.13) observemos que se cumple la siguiente desigualdad entre la media y la mediana:

$$|\mathbf{E}(X) - \mathbf{M}(X)| \leq \delta, \tag{2.15}$$

donde $\delta = \frac{1}{2}\frac{\sqrt{\pi}KC}{\sqrt{c}}$. Para ver esto notemos que

$$\begin{aligned}
|\mathbf{E}(X) - \mathbf{M}(X)| &= |\mathbf{E}(X - \mathbf{M}(X))| \\
&\leq \mathbf{E}|X - \mathbf{M}(X)| \\
&= \int_0^\infty \mathbf{P}(|X - \mathbf{M}(X)| \geq \lambda)\, d\lambda \\
&\leq C\int_0^\infty \exp\left(-\frac{c}{K^2}\lambda^2\right) d\lambda \\
&= \frac{1}{2}\frac{\sqrt{\pi}KC}{\sqrt{c}}.
\end{aligned}$$

Ahora, para demostrar (2.13) observemos que por la desigualdad $|a| - |b| \leq |a+b|$, para todo $a, b \in \mathbb{R}$, se sigue que

$$|F(\mathbf{X}) - \mathbf{E}(F(\mathbf{X}))| - |\mathbf{M}(F(\mathbf{X})) - \mathbf{E}(F(\mathbf{X}))| \leq |F(\mathbf{X}) - \mathbf{M}(F(\mathbf{X}))|,$$

por lo que tenemos que

$$\mathbf{P}(|F(\mathbf{X}) - \mathbf{E}(F(\mathbf{X}))| \geq \lambda + |\mathbf{M}(F(\mathbf{X})) - \mathbf{E}(F(\mathbf{X}))|) \leq \mathbf{P}(|F(\mathbf{X}) - \mathbf{M}(F(\mathbf{X}))| \geq \lambda),$$

y por (2.15)

$$\mathbf{P}(|F(\mathbf{X}) - \mathbf{E}(F(\mathbf{X}))|) \geq \delta + \lambda \leq \mathbf{P}(|F(\mathbf{X}) - \mathbf{M}(F(\mathbf{X}))| \geq \lambda),$$

además, gracias a (2.12)

$$\mathbf{P}(|F(\mathbf{X}) - \mathbf{E}(F(\mathbf{X}))| \geq \delta + \lambda) \leq C_1 e^{-c_1\lambda^2}, \tag{2.16}$$

Observemos que si $\lambda < \delta$, hagamos $C_1 = e^{c_1\delta^2}$, entonces

$$\mathbf{P}(|F(\mathbf{X}) - \mathbf{E}(F(\mathbf{X}))| \geq \lambda) \leq 1 = C_1 e^{-c_1\lambda^2} \leq C_1 e^{-c_1 k_1 \lambda^2},$$

por lo que solo nos falta demostrar el caso $\lambda \geq \delta$. Tenemos que si $\lambda - \delta \geq 0$, por (2.16)

$$\mathbf{P}\left(\left|F(\mathbf{X}) - \mathbf{E}\left(F\left(\mathbf{X}\right)\right)\right| \geq \lambda\right) = \mathbf{P}\left(\left|F(\mathbf{X}) - \mathbf{E}\left(F\left(\mathbf{X}\right)\right)\right| \geq \delta + \lambda - \delta\right) \leq C_1 \exp\left[-c_1\left(\lambda - \delta\right)^2\right],$$

y por la Proposición 44, existen $k_1 > 0$ y $b \in \mathbb{R}$ tales que

$$\begin{aligned}
\mathbf{P}\left(\left|F(\mathbf{X}) - \mathbf{E}\left(F\left(\mathbf{X}\right)\right)\right| \geq \lambda\right) &\leq C_1 \exp\left(-c_1 k_1 \lambda^2 - c_1 b\right) \\
&= C e^{-c\lambda^2},
\end{aligned}$$

de donde (2.13) se cumple. ∎

Capítulo 3

Aplicaciones en matrices aleatorias

3.1 Concentración de la norma operador

Una *matriz aleatoria* es una matriz cuyas componentes son variables aleatorias. Dada una matriz aleatoria A, existen muchas propiedades de ésta que se pueden estudiar, como por ejemplo, los eigenvalores o valores propios, la traza, el determinante, etc. En este capítulo nos centraremos en estudiar el comportamiento de una propiedad básica, la *norma operador* de la matriz A, la cual se define como

$$\|A\|_{op} := \sup_{\mathbf{x}\in\mathbb{R}^n:|\mathbf{x}|=1} |A\mathbf{x}| .$$

Esta cantidad por sí sola es interesante, pero además sirve como una cota básica de otras cantidades de interés de la matriz. Por ejemplo los eigenvalores de la matriz tienen tamaño a lo más $\|A\|_{op}$. Con base en lo anterior podemos entonces decir que es importante encontrar cómo es que se concentran los valores de las normas operador de las matrices aleatorias.

En esta sección se muestra un resultado que nos dice cómo es que la norma operador, vista como función, se concentra alrededor de su media. Para dicho fin utilizaremos el resultado del Teorema 38, quedando a la vista la utilidad de las desigualdades de concentración.

Proposición 39 ***(Concentración de la norma operador).*** *Sea $A = (X_{ij})$ una matriz aleatoria de $n \times n$ donde las entradas X_{ij} son independientes con media cero y acotadas uniformemente por 1. Entonces para cualquier $\lambda > 0$, tenemos que*

$$\mathbf{P}\left(\left|\|A\|_{op} - \mathbf{M}\left(\|A\|_{op}\right)\right| \geq \lambda\right) \leq C\exp\left(-c\lambda^2\right),$$

para algunas constantes $C, c > 0$. El resultado se mantiene si reemplazamos $\mathbf{M}\left(\|A\|_{op}\right)$ por $\mathbf{E}\,\|A\|_{op}$.

Demostración. Podemos ver a $\|A\|_{op}$ como una función $F\left((X_{ij})_{1\leq i,j\leq n}\right)$ de las variables aleatorias independientes X_{ij}, con lo que F es una función de $\mathbb{R}^{n^2}$ a $\mathbb{R}$. La convexidad de la norma operador nos muestra que F es convexa.

Ahora observemos que, $|A\mathbf{x}|$ esta acotado, por lo que, por la propiedad del supremo, para algún $\mathbf{x} \in \mathbb{R}^n$

$$\begin{aligned}
\|A\|_{op} &= \sqrt{\left(\sum_{j=1}^{n} X_{1j}x_j\right)^2 + \left(\sum_{j=1}^{n} X_{2j}x_j\right)^2 + \cdots + \left(\sum_{j=1}^{n} X_{nj}x_j\right)^2} \\
&= \sqrt{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} X_{ij}x_j\right)^2},
\end{aligned}$$

por la desigualdad de Cauchy-Schwarz

$$\|A\|_{op} \leq \sqrt{\sum_{i=1}^{n}\left[\left(\sum_{j=1}^{n} X_{ij}^2\right)\left(\sum_{j=1}^{n} x_j^2\right)\right]} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij}^2}. \tag{3.1}$$

Sea $B = (Y_{ij})$ una matriz aleatoria de $n \times n$ con entradas Y_{ij} variables aleatorias independientes con media cero y acotadas uniformemente por 1. Entonces, por la desigualdad del triángulo, tenemos que

$$\begin{aligned}
|F(A) - F(B)| &= \left|F\left((X_{ij})_{1\leq i,j\leq n}\right) - F\left((Y_{ij})_{1\leq i,j\leq n}\right)\right| \\
&= \left|\|A\|_{op} - \|B\|_{op}\right| \\
&\leq \|A - B\|_{op}` \\
&\leq \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{ij} - Y_{ij})} \\
&= |A - B|,
\end{aligned}$$

por lo que F es *1-Lipschitz*. Por lo tanto el resultado se sigue de la desigualdad de Talagrand (Teorema 38). ■

3.2 El Teorema de Wigner

En esta sección presentaremos el Teorema de Wigner, el cual es uno de los resultados más importantes en la teoría de las matrices aleatorias. Este teorema nos dice que la distribución espectral empírica de una matriz aleatoria hermitiana, cuyas entradas cumplen ciertas características, converge a la distribución del semicírculo. Además de presentar el teorema, se expondrá de que manera se utiliza la Desigualdad de concentración de Talagrand en la demostración de la convergencia casi segura en el teorema.

Una *matriz de Wigner* es una matriz aleatoria A^n de $n \times n$ con elementos reales o complejos tales que A_{ij}^n, $1 \leq i \leq j \leq n$ son independientes y A^n es hermitiana, i.e. $A_{ij}^n = \overline{A_{ji}^n}$.

Sea $F(x)$ la función de distribución del semicírculo, *i.e.*,

$$F(x) = \int_{-\infty}^{x} \frac{1}{2\pi} \sqrt{4-t^2}\,\mathbf{1}_{[-2,2]}(t)dt. \tag{3.2}$$

Teorema 40 ***(Teorema de Wigner).*** *Sea A^n una matriz de Wigner tal que A_{ii}^n son variables aleatorias independientes e idénticamente distribuidas, $1 \leq i \leq n$, $\mathbf{E}\left|A_{ij}^n\right|^2 = 1/n$, $1 \leq i < j \leq n$. Entonces la distribución espectral empírica de $\frac{1}{\sqrt{n}}A^n$ converge casi seguramente a la distribución del semicírculo, es decir, para cada $x \in \mathbb{R}$, cuando $n \longrightarrow \infty$*

$$F_{A^n}(x) := \frac{1}{n}\left|\{i : \lambda_{n,i} \leq x,\ 1 \leq i \leq n\}\right| \xrightarrow{c.s.} F(x),$$

donde $\lambda_{n,i}$, $1 \leq i \leq n$, son los valores propios de la matriz A^n.

En lo que sigue supondremos, sin pérdida de generalidad, que las variables A_{ij}^n son acotadas y con media igual a cero, el lector interesado en ver a detalle la justificación de lo que se acaba de mencionar puede ver [3] páginas de la 15 a la 26. Con estas condiciones se puede seguir la demostración dada en [4] del Teorema de Wigner, utilizando herramientas de teoría de gráficas y combinatoria.

En esta sección demostraremos que

$$\sum_{k=1}^{\infty} \operatorname{Var}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right) < \infty, \tag{3.3}$$

utilizando la concentración de medida, en lugar de teoría de gráficas. Con lo anterior, siguiendo los pasos en [4], quedará demostrada la convergencia casi segura del Teorema 40.

Sea $f(x) = x^k$, observemos que f restringida a un conjunto compacto es L-Lipschitz en $\mathbb{R}$, $L > 0$ (F es L-Lipschitz si y solo si $\frac{1}{L}F$ es 1-Lipschitz) entonces por Lema 2.3.1 de [1] $\operatorname{tr}(A^n)^k$ es L_1-Lipschitz para alguna constante $L_1 > 0$. Ahora, para demostrar (3.3), primero por definición tenemos que

$$\operatorname{Var}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right) = \mathbf{E}\left|\frac{1}{n}\operatorname{tr}\left(A^n\right)^k - \mathbf{E}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right)\right|^2$$

de donde por la Proposición 5 se tiene que

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right) &= \int_0^{\infty} \mathbf{P}\left(\left|\frac{1}{n}\operatorname{tr}\left(A^n\right)^k - \mathbf{E}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right)\right| \geq \sqrt{\lambda}\right) d\lambda \\ &= \int_0^{\infty} \mathbf{P}\left(\frac{1}{L_1}\left|\operatorname{tr}\left(A^n\right)^k - \mathbf{E}\left(\operatorname{tr}\left(A^n\right)^k\right)\right| \geq \frac{\sqrt{\lambda}}{L_1}n\right) d\lambda.\end{aligned}$$

Entonces dado que f también es convexa, entonces por Lema 4.4.12 de [1], $\operatorname{tr}(A^n)^k$ es convexa, entonces por el Teorema 38 tenemos que existen $c > 0$ y $C > 0$ tales que

$$\operatorname{Var}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right) \leq \int_0^{\infty} C\exp\left(-c\frac{\lambda}{L_1^2}n^2\right) d\lambda = \frac{C_0}{n^2},$$

donde C_0 es una constante positiva. Por lo anterior obtenemos que

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{1}{n}\operatorname{tr}\left(A^n\right)^k\right) < \infty.$$

Para el caso de k impar, siguiendo la sugerencia de Tao [13], utilizando el hecho de que $\|A^n\|_{op} \xrightarrow{c.s.} 2$ (Teorema 2.3.4. [13]), escogemos un $c > 2$, entonces la matriz $A^n + cI_n$ es una matriz positiva definida, casi seguramente; por lo que utilizando los mismos argumentos de arriba tenemos que $\mathrm{tr}(A^n + cI_n)^k$ es L_2-Lipschitz y convexa. Además de manera similar a la demostración de la convergencia en esperanza dada en [4], de la distribución espectral empírica a la distribución del semicírculo se puede demostrar que

$$\mathbf{E}\left(\frac{1}{n}\mathrm{tr}\left(A^n + cI_n\right)^k\right) \underset{n\to\infty}{\longrightarrow} \mathbf{E}\left(X+c\right)^k,$$

donde $X \sim F$ dada en (3.2).

Ahora aplicando la desigualdad de Talagrand (Teorema 38), obtenemos que

$$\begin{aligned} &\int_0^\infty \mathbf{P}\left(\left|\mathrm{tr}\left(A^n + cI_n\right)^k - \mathbf{E}\left(\mathrm{tr}\left(A^n + cI_n\right)^k\right)\right| > \frac{\sqrt{\lambda}}{L_2}n\right) d\lambda \\ \leq &\int_0^\infty C\exp\left(-c\frac{\lambda}{L_2^2}n^2\right) d\lambda = \frac{C_1}{n^2}, \end{aligned}$$

donde C_1 es una constante positiva. Lo que nos da la convergencia casi segura

$$\mathbf{E}\left(\frac{1}{n}\mathrm{tr}\left(A^n + cI_n\right)^k\right) \xrightarrow[n\to\infty]{c.s.} \mathbf{E}\left(X+c\right)^k,$$

Por lo tanto por el método de momentos (ver [4]) se sigue que

$$F_{A^n+cI_n}\left(x\right) \xrightarrow[n\to\infty]{c.s.} F_{X+c}\left(x\right).$$

En virtud de que

$$F_{A^n+cI_n}\left(x\right) = F_{A^n}\left(x-c\right) \quad \text{y} \quad F_{X+c}\left(x\right) = F_X\left(x-c\right),$$

se sigue que

$$F_{A^n}\left(x\right) \xrightarrow[n\to\infty]{c.s.} F_X\left(x\right).$$

Apéndice A

Resultados auxiliares

En este apéndice se muestran unos resultados de cálculo que son útiles en algunas de las demostraciones de esta tesis.

A.1 Recta básica de la gráfica de una función convexa

Lema 41 *Sea* $\mathbf{I} \subseteq \mathbb{R}$ *un intervalo, sea* $\varphi : \mathbf{I} \longrightarrow \mathbb{R}$ *una función convexa y sea* c *un punto interior de* $\mathbf{I}$. *Entonces existe un* $\alpha \in \mathbb{R}$ *tal que*

$$\varphi(x) \geq \alpha\,(x-c) + \boldsymbol{\varphi}(c) \qquad \forall x \in I.$$

Demostración. Como $\boldsymbol{\varphi}$ es convexa, sus derivadas unilaterales $\boldsymbol{\varphi}'_-(c)$ y $\boldsymbol{\varphi}'_+(c)$ derivada por derecha y derivada por izquierda respectivamente existen en el punto $c \in \mathbf{I}$ y satisfacen la siguiente desigualdad:

$$\boldsymbol{\varphi}'_-(c) \leq \boldsymbol{\varphi}'_+(c).$$

Tomemos α como cualquier número del intervalo$\left[\boldsymbol{\varphi}'_-(c), \boldsymbol{\varphi}'_+(c)\right]$. Tenemos que si $x < c$, entonces

$$\alpha \geq \boldsymbol{\varphi}'_-(c) \geq \frac{\boldsymbol{\varphi}(x) - \boldsymbol{\varphi}(c)}{x-c},$$

lo cual, multiplicando por $x - c < 0$ nos da que

$$\boldsymbol{\varphi}(x) \geq \alpha(x-c) + \boldsymbol{\varphi}(c).$$

Ahora, si $x > c$, entonces

$$\alpha \leq \boldsymbol{\varphi}'_{+}(c) \leq \frac{\boldsymbol{\varphi}(x) - \boldsymbol{\varphi}(c)}{x - c},$$

lo que, multiplicando por $x - c > 0$ nos da que

$$\boldsymbol{\varphi}(x) \geq \alpha(x - c) + \boldsymbol{\varphi}(c).$$

■

A.2 El cubo booleano como espacio métrico.

Obsérvese que $\mathbf{x},\ \mathbf{y} \in I_n$

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \{i : x_i \neq y_i, \quad i = 1, ..., n\} \\ &= |x_1 - y_1| + \cdots + |x_n - y_n| \,. \end{aligned}$$

Proposición 42 *Sea $I_n = \{0,1\}^n$ el cubo booleano. La dupla (I_n, d) es un espacio métrico, donde d es la "distancia de Hamming", la cual se define como:*

$$d(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + \cdots + |x_n - y_n| \,,$$

para $\mathbf{x}$ y $\mathbf{y}$ en I_n.

Demostración. Sean $\mathbf{x}, \mathbf{y}$ como arriba y $\mathbf{z} = (z_1, z_2, ..., z_n)$, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in I_n$, entonces si $x_i = y_i$, entonces $|x_i - y_i| = 0$ y si $x_i \neq y_i$ tenemos que $|x_i - y_i| = 1$, de donde obtenemos que $d(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + \cdots + |x_n - y_n| \geq 0$ y $d(\mathbf{x}, \mathbf{y}) = 0$ si y solo si $|x_1 - y_1| + \cdots + |x_n - y_n| = 0$ lo cual solo pasa cuando $x_i = y_i$ para toda $i = 1, \ldots, n$,i.e. solo cuando $\mathbf{x} = \mathbf{y}$. Ahora observemos que

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= |x_1 - y_1| + \cdots + |x_n - y_n| \\ &= |y_1 - x_1| + \cdots + |y_n - x_n| \\ &= d(\mathbf{y}, \mathbf{x}), \end{aligned}$$

por lo que solo nos falta demostrar la desigualdad del triángulo para ésta métrica. Tenemos que

$$\begin{aligned} d(\mathbf{x},\mathbf{z}) &= |x_1 - z_1| + \cdots + |x_n - z_n| \\ &= |x_1 - y_1 + y_1 - z_1| + \cdots + |x_n - y_n + y_n - z_n| \\ &\leq |x_1 - y_1| + |y_1 - z_1| \cdots + |x_n - y_n| + |y_n - z_n| \\ &= d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}). \end{aligned}$$

■

A.3 Desigualdades elementales

Proposición 43 *Para* $0 \leq r \leq 1$, $c > 0$

$$\inf_{0\leq\lambda\leq 1} e^{c(1-\lambda)^2} r^{-\lambda} \leq 2 - r. \tag{A.1}$$

Demostración. Si derivamos $e^{c(1-\lambda)^2} r^{-\lambda}$ con respecto a λ e igualamos a cero, se tiene que $\lambda = 1 + 2\log r$ es un punto mínimo pues la segunda derivada en λ es mayor que cero. Si $\lambda \in (0,1]$ y si $r \geq e^{-1/2}$ entonces tenemos que $1 + 2\log r \geq 0$, por lo que si $0 \leq r < e^{-1/2}$ entonces $\lambda = 0$, de donde, haciendo $c = 1/4$, se sigue que

$$1.28 \approx e^{1/4} \leq 2 - e^{-1/2} \approx 1.4,$$

con lo que se cumpliría A.1.

Entonces si $\lambda = 1 + 2\log r,\ r \in \left[e^{-1/2}, 1\right]$ y $c = 1/4$, tendremos que

$$e^{1/4(1-(1+2\log r))^2} r^{-1-2\log r} \leq 2 - r,$$

lo cual ocurre si y solo si

$$(\log r)^2 + (-1 - 2\log r)\log r = -(\log r)^2 - \log r \leq \log(2 - r). \tag{A.2}$$

Para verificar que A.2 se cumple, sea

$$F(r) = \log(2-r) + \log r + (\log r)^2 .$$

Observemos que $F''(r) \geq 0$ para $r \in \left[e^{-1/2}, 1\right]$, por lo que $F'(r)$ es creciente, además como $F'(e^{-1/2}) = -1.7$ y $F(1) = 0$, entonces $F'(r) \leq 0$ para todo $r \in \left[e^{-1/2}, 1\right]$, lo que quiere decir que F es creciente, y como $F(e^{-1/2}) = 0.08$ y $F(1) = 0$, entonces $F(r) \geq 0$ para todo $r \in \left[e^{-1/2}, 1\right]$, con lo que se cumple A.2, y esto completa la demostración. ■

Proposición 44 *Para todo* $0 < x \in \mathbb{R}$ *y toda constante* $a \in \mathbb{R}$, *existen constantes* $k > 0$ *y* $b \in \mathbb{R}$ *las cuales cumplen que*

$$(x-a)^2 \geq kx^2 - b. \tag{A.3}$$

Demostración. Para que (A.3) se cumpla se tiene que cumplir que

$$(1-k)\,x^2 - (2a)x + \left(a^2 + b\right) \geq 0,$$

lo que se cumple si y solo si se cumple que

$$1-k > 0 \quad \text{y} \quad 4a^2 - 4\,(1-k)\left(a^2+b\right) < 0,$$

de donde se puede ver que existen k y b que cumplan con (A.3). Por ejemplo, si se elige un k, tal que $0 < k < 1$, se debe de buscar un b tal que

$$\frac{a^2}{a^2+b} < 1-k < 1$$

lo que cumple cualquier cualquier $b > \frac{a^2}{1-k} - a^2$. ■

Apéndice B

Demostraciones

En este apéndice se encuentran las demostraciones de los teoremas, proposiciones, lemas y corolarios que se enuncian en el desarrollo de esta tesis.

Demostración de la Proposición 5. Por la definición de probabilidad de un evento tenemos entonces la siguiente igualdad

$$\begin{aligned}\int_0^\infty \mathbf{P}(X \geq \lambda)\, d\lambda &= \int_0^\infty \mathbf{E}\mathbf{1}_{\{X\geq\lambda\}}(X)\, d\lambda \\ &= \int_0^\infty \int_\Omega \mathbf{1}_{\{X\geq\lambda\}}(x)\, d\mathbf{P}_X(x)\, d\lambda \\ &= \int_0^\infty \int_\lambda^\infty d\mathbf{P}_X(x)\, d\lambda\end{aligned}$$

y por el Teorema de Fubini-Tonelli se sigue que

$$\begin{aligned}\int_0^\infty \mathbf{P}(X \geq \lambda)\, d\lambda &= \int_0^\infty \left(\int_0^x d\lambda\right) d\mathbf{P}_X(x) \\ &= \int_0^\infty x d\mathbf{P}_X(x) \\ &= \int_\Omega X d\mathbf{P} \\ &= \mathbf{E}(X).\end{aligned}$$

■

Demostración de el Lema 10. Observemos que

$$\begin{aligned}
\mathbf{E}\left|X-C\right|-\mathbf{E}\left|X-\mathbf{M}(X)\right| &= \int_{-\infty}^{\infty}\left|x-C\right|d\mathbf{P}_X\left(x\right)-\int_{-\infty}^{\infty}\left|x-\mathbf{M}(X)\right|d\mathbf{P}_X\left(x\right)\\
&= \int_{-\infty}^{c}-\left(x-C\right)d\mathbf{P}_X\left(x\right)+\int_{c}^{\infty}\left(x-C\right)d\mathbf{P}_X\left(x\right)\\
&\quad+\int_{-\infty}^{\mathbf{M}(X)}\left(x-\mathbf{M}(X)\right)d\mathbf{P}_X\left(x\right)\\
&\quad-\int_{\mathbf{M}(X)}^{\infty}\left(x-\mathbf{M}(X)\right)d\mathbf{P}_X\left(x\right)\\
&= -\int_{-\infty}^{c}xd\mathbf{P}_X\left(x\right)+C\int_{-\infty}^{c}d\mathbf{P}_X\left(x\right)\\
&\quad+\int_{c}^{\infty}xd\mathbf{P}_X\left(x\right)-\int_{c}^{\infty}Cd\mathbf{P}_X\left(x\right)\\
&\quad+\int_{-\infty}^{\mathbf{M}(X)}xd\mathbf{P}_X\left(x\right)-\mathbf{M}(X)\int_{-\infty}^{\mathbf{M}(X)}d\mathbf{P}_X\left(x\right)\\
&\quad-\int_{\mathbf{M}(X)}^{\infty}xd\mathbf{P}_X\left(x\right)+\mathbf{M}(X)\int_{\mathbf{M}(X)}^{\infty}d\mathbf{P}_X\left(x\right)\\
&= -\int_{-\infty}^{c}xd\mathbf{P}_X\left(x\right)+C\int_{-\infty}^{c}d\mathbf{P}_X\left(x\right)\\
&\quad+\int_{c}^{\infty}xd\mathbf{P}_X\left(x\right)-\int_{c}^{\infty}Cd\mathbf{P}_X\left(x\right)\\
&\quad+\int_{-\infty}^{\mathbf{M}(X)}xd\mathbf{P}_X\left(x\right)-\int_{\mathbf{M}(X)}^{\infty}xd\mathbf{P}_X\left(x\right),
\end{aligned}$$

a partir de aquí demostraremos el caso $\mathbf{M}(X)<C$, para el caso $C<\mathbf{M}(X)$ es similar.

Entonces sea $\mathbf{M}(X) < C$, de la última igualdad tenemos que

$$\begin{aligned}
\mathbf{E}\left|X-C\right|-\mathbf{E}\left|X-\mathbf{M}(X)\right| &= -\int_{-\infty}^{c} xd\mathbf{P}_X\left(x\right)+C\int_{-\infty}^{\mathbf{M}(X)} d\mathbf{P}_X\left(x\right) \\
&\quad +C\int_{\mathbf{M}(X)}^{\infty} d\mathbf{P}_X\left(x\right)+\int_{c}^{\infty} xd\mathbf{P}_X\left(x\right) \\
&\quad -C\int_{\mathbf{M}(X)}^{\infty} d\mathbf{P}_X\left(x\right)+C\int_{\mathbf{M}(X)}^{C} d\mathbf{P}_X\left(x\right) \\
&\quad +\int_{-\infty}^{\mathbf{M}(X)} xd\mathbf{P}_X\left(x\right)-\int_{\mathbf{M}(X)}^{\infty} xd\mathbf{P}_X\left(x\right) \\
&= -\int_{-\infty}^{c} xd\mathbf{P}_X\left(x\right)+2C\int_{\mathbf{M}(X)}^{C} d\mathbf{P}_X\left(x\right) \\
&\quad +\int_{c}^{\infty} xd\mathbf{P}_X\left(x\right)+\int_{-\infty}^{\mathbf{M}(X)} xd\mathbf{P}_X\left(x\right) \\
&\quad -\int_{\mathbf{M}(X)}^{\infty} xd\mathbf{P}_X\left(x\right) \\
&= -2\int_{\mathbf{M}(X)}^{C} xd\mathbf{P}_X\left(x\right)+2C\int_{\mathbf{M}(X)}^{C} d\mathbf{P}_X\left(x\right) \\
&= 2\int_{\mathbf{M}(X)}^{C} \left(C-x\right)d\mathbf{P}_X\left(x\right),
\end{aligned}$$

por lo que tenemos la siguiente igualdad

$$\mathbf{E}\left|X-C\right|=\mathbf{E}\left|X-\mathbf{M}(X)\right|+2\int_{\mathbf{M}(X)}^{C}\left(C-x\right)d\mathbf{P}_X\left(x\right),$$

de donde se sigue que $C=\mathbf{M}(X)$ minimiza $\mathbf{E}\left|X-C\right|$. ■

Demostración de la Proposición 11. Tenemos que

$$\left|\mathbf{E}(X)-\mathbf{M}(X)\right|=\left|\mathbf{E}(X-\mathbf{M}(X))\right|\leq\mathbf{E}\left|X-\mathbf{M}(X)\right|,$$

y por el Lema 10, tenemos que

$$\begin{aligned}
\left|\mathbf{E}(X)-\mathbf{M}(X)\right| &\leq \mathbf{E}\left|X-\mathbf{M}(X)\right|\leq\mathbf{E}\left|X-\mathbf{E}(X)\right| \\
&= \mathbf{E}\sqrt{\left(X-\mathbf{E}\left(X\right)\right)^2}
\end{aligned}$$

y por el Teorema 17 (la función $-\sqrt{x}$ es convexa) tenemos que

$$\left|\mathbf{E}(X)-\mathbf{M}(X)\right|\leq\sqrt{\mathbf{E}\left(X-\mathbf{E}\left(X\right)\right)^2}=\sqrt{\mathrm{Var}\left(X\right)}.$$

■

Demostración del Lema 13 Por la independencia de $X_1, \ldots, X_n$ tenemos que

$$\begin{aligned}\mathbf{E}\exp\left[t\left(c_1X_1+\ldots+c_nX_n\right)\right] &= \mathbf{E}\left[\exp\left(tc_1X_1\right)\cdots\exp\left(tc_nX_n\right)\right] \\ &= \mathbf{E}\exp\left(tc_1X_n\right)\cdots\mathbf{E}\exp\left(tc_nX_n\right).\end{aligned}$$

Por lo tanto

$$\begin{aligned}\mathbf{E}\exp\left(tc_1X_n\right)\cdots\mathbf{E}\exp\left(tc_nX_n\right) &= \exp\left(\mu_1c_1+\frac{\sigma_1^2c_1^2}{2}t^2\right)\cdots\exp\left(\mu_nc_n+\frac{\sigma_n^2c_n^2}{2}t^2\right) \\ &= \exp\left(\sum_{i=1}^n\mu_ic_i+\sum_{i=1}^n\frac{\sigma_i^2c_i^2}{2}t^2\right),\end{aligned}$$

de donde tenemos que $c_1X_1+\ldots+c_nX_n$ es Gaussiana con media $\sum_{i=1}^n\mu_ic_i$ y varianza $\sum_{i=1}^n\sigma_i^2c_i^2$. ■

Demostración del Teorema 14. Tenemos la desigualdad

$$\lambda\mathbf{1}(X\geq\lambda)\leq X,$$

tomando las esperanzas a ambos lados de la desigualdad tenemos que

$$\begin{aligned}\mathbf{E}\left[\lambda\mathbf{1}(X\geq\lambda)\right] &\leq \mathbf{E}(X) \\ \lambda\mathbf{E}\left[\mathbf{1}(X\geq\lambda)\right] &\leq \mathbf{E}(X) \\ \lambda\mathbf{P}\left(X\geq\lambda\right) &\leq \mathbf{E}(X) \\ \mathbf{P}\left(X\geq\lambda\right) &\leq \frac{\mathbf{E}(X)}{\lambda}.\end{aligned}$$

■

Demostración del Teorema 17. Sean

$$a=\inf I, \qquad b=\sup I, \qquad c=\int_\Omega Fd\mathbf{P}.$$

Demostremos que $c\in I$. Observemos que si $(a,b)=(-\infty,\infty)$, como F es integrable, se tiene que $c\in I$. Ahora si $a\in I$, entonces $F\geq a$ en todo punto de Ω, por lo que

$$\int_\Omega Fd\mathbf{P}\geq\int_\Omega ad\mathbf{P}=a.$$

Si $b \in I$, entonces $F \leq b$ en todo punto de Ω, por lo que

$$\int_\Omega F d\mathbf{P} \leq \int_\Omega b d\mathbf{P} = b.$$

Ahora supongamos que $-\infty \neq a \notin I$. Entonces $F > a$ en todo punto de Ω. Si $\int_\Omega F d\mathbf{P} = a$, entonces de $\int_\Omega (F-a) d\mathbf{P} = 0$ se tuviera que $F = a$ casi en todas partes en Ω, lo cual es una contradicción. Por lo tanto $\int_\Omega F d\mathbf{P} > a$. De manera similar tenemos que si $\infty \neq b \notin I$, entonces $F < b$ en todo punto de Ω, por lo que $\int_\Omega F d\mathbf{P} < b$, con lo que queda demostrado que $c \in I$.

Siguiendo con la demostración, supongamos que c es un punto interior de I. Sea $\alpha \in \left[\varphi'_-(c), \varphi'_+(c)\right]$. Entonces, por el lema 41

$$\varphi(t) \geq \varphi(c) + \alpha(t-c) \qquad \forall t \in I,$$

al sustituir t por $F(x)$ se obtiene

$$(\varphi \circ F)(x) \geq \varphi(c) + \alpha\left(F(x) - c\right) \qquad \forall x \in \Omega,$$

lo cual, integrando sobre Ω con respecto de $\mathbf{P}$ nos da que

$$\int_\Omega (\varphi \circ F)\, d\mathbf{P} \geq \varphi(c) + \alpha\left(\int_\Omega F d\mathbf{P} - c\right) = \varphi(c),$$

con lo que únicamente nos queda por examinar el caso donde c es uno de los extremos de I, que entonces pertenece a I. Supongamos, sin pérdida de generalidad, que $c = a$, lo cual implica que $F = a$ casi en todas partes. Se tiene entonces que

$$\int_\Omega (\varphi \circ F)\, d\mathbf{P} = \varphi(a) \int_\Omega d\mathbf{P} = \varphi(a) = \varphi(c),$$

con lo que se concluye la prueba. ∎

Demostración del Corolario 18. Aplicando el Teorema 17 al espacio de probabilidad $([a,b]$, $\mathfrak{B}([a,b]), \mathbf{P})$, donde $d\mathbf{P} = \frac{dx}{b-a}$. ∎

Demostración del Lema 20. Si escribimos $x = \lambda a + (1-\lambda) b$, $0 \leq \lambda \leq 1$, la convexidad de e^{tx} nos dice que

$$e^{tx} \leq \lambda e^{ta} + (1-\lambda) e^{tb}.$$

Sea $\lambda = (b-x) / (b-a)$ entonces se sigue que

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

Tomando la esperanza en ambos lados de la desigualdad, llegamos a

$$\begin{aligned} \mathbf{E} e^{tX} &\leq \frac{b - \mathbf{E}X}{b-a} e^{ta} + \frac{\mathbf{E}X - a}{b-a} e^{tb} \\ &= \frac{be^{ta} - ae^{tb}}{b-a} \\ &\leq e^{t^2 (b-a)^2 / 8}. \end{aligned}$$

Esta última desigualdad se demostrará a continuación. Sea $p = -a / (b-a)$, entonces

$$\begin{aligned} \mathbf{E} e^{tX} &\leq \frac{be^{ta} - ae^{tb}}{b-a} \\ &= \left(1 - p + pe^{t(b-a)}\right) e^{-pt(b-a)} \\ &= e^{\phi(u)}, \end{aligned}$$

donde $u = t(b-a)$ y $\phi(u) = -pu + \log(1 - p + pe^{u})$. Dado que

$$\phi'(u) = -p + \frac{p}{p + (1-p) e^{-u}},$$

por lo tanto $\phi(0) = \phi'(0) = 0$. Más aún

$$\phi''(u) = \frac{p(1-p) e^{-u}}{(p + (1-p) e^{-u})^2} \leq \frac{1}{4},$$

observe que el máximo valor de la función $z / \left(p + \frac{1}{p} z\right)^2$ con $z = p(1-p) e^{-u}$, es $\frac{1}{4}$. Ahora, por el Teorema de Taylor se tiene que para algún $\theta \in [0, u]$

$$\begin{aligned} \phi(u) &= \phi(0) + u\phi'(0) + \frac{u^2}{2} \phi''(\theta) \\ &\leq \frac{u^2}{8} \\ &= \frac{t^2 (b-a)^2}{8}. \end{aligned}$$

■

Demostración del Teorema 21. Solo probaremos la cota inferior (la prueba para la cota superior es similar). Sea $Y_i = X_i - \mathbf{E}X_i$, entonces las Y_i son variables aleatorias independientes. con media cero y rango $[a_i - \mathbf{E}X_i, b_i - \mathbf{E}X_i]$. Para cualquier $t > 0$, se sigue

$$\mathbf{P}\left(S_n - \mathbf{E}S_n \geq \varepsilon\right) = \mathbf{P}\left(Y_1 + \cdots + Y_n \geq \varepsilon\right) = \mathbf{P}\left(e^{t(Y_1+\cdots+Y_n)} \geq e^{t\varepsilon}\right) \leq \frac{\mathbf{E}e^{t(Y_1+\cdots+Y_n)}}{e^{t\varepsilon}},$$

esto último es por la Desigualdad de Markov (Teorema 14). Por la independencia de los Y_i y usando el Lema 20 para cada Y_i, obtenemos que

$$\begin{aligned}
\mathbf{P}\left(S_n - \mathbf{E}S_n \geq \varepsilon\right) &\leq \frac{\mathbf{E}e^{tY_1} \cdots \mathbf{E}e^{tY_n}}{e^{t\varepsilon}} \\
&\leq \frac{e^{t^2(b_1-a_1)^2/8} \cdots e^{t^2(b_n-a_n)^2/8}}{e^{t\varepsilon}}.
\end{aligned}$$

sea $t = 4\varepsilon/D$, se sigue que

$$\mathbf{P}\left(S_n - \mathbf{E}S_n \geq \varepsilon\right) \leq e^{-2\varepsilon^2/D}.$$

■

Demostración del Corolario 24. Tenemos que

$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots \quad \text{y} \quad e^{-t} = 1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \cdots,$$

entonces

$$\begin{aligned}
\frac{1}{2} + \frac{e^t + e^{-t}}{4} &= \frac{1}{2} + \left(\frac{1+1}{4}\right) + \left(\frac{t-t}{4}\right) + \left(\frac{t^2+t^2}{4(2!)}\right) + \cdots \\
&= 1 + \frac{t^2}{4} + \frac{t^4}{2(4!)} + \cdots + \frac{t^{2k}}{2(2k)!} + \cdots \\
&\leq 1 + \frac{t^2}{4} + \cdots + \frac{t^{2k}}{4^k k!} + \cdots = e^{t^2/4},
\end{aligned}$$

por lo que el resultado se sigue del Teorema 23. ■

Demostración del Corolario 25. Tenemos que $d(\mathbf{x}, A) \geq \varepsilon\sqrt{n} \Longrightarrow td(\mathbf{x}, A) \geq t\varepsilon\sqrt{n}$, de donde obtenemos lo siguiente

$$\int_{I_n} e^{td(\mathbf{x},A)} d\mathbf{P}_n \geq \int_{\{\mathbf{x}\in I_n : d(\mathbf{x}.A)\geq\varepsilon\sqrt{n}\}} e^{td(\mathbf{x},A)} d\mathbf{P}_n \geq \mathbf{P}_n\{\mathbf{x} \in I_n : d(\mathbf{x}, A) \geq \varepsilon\sqrt{n}\} e^{t\varepsilon\sqrt{n}},$$

por lo que

$$\mathbf{P}_n\{\mathbf{x} \in I_n : d(\mathbf{x}, A) \geq \varepsilon\sqrt{n}\} \leq e^{-t\varepsilon\sqrt{n}} \int_{I_n} e^{td(\mathbf{x},A)} d\mathbf{P}_n.$$

Como

$$\int_{I_n} e^{td(\mathbf{x},A)} d\mathbf{P}_n \leq \frac{1}{\mathbf{P}_n(A)} e^{t^2 n/4},$$

por el Corolario 24, entonces

$$\mathbf{P}_n\{\mathbf{x} \in I_n : d(\mathbf{x}, A) \geq \varepsilon\sqrt{n}\} \leq \frac{e^{-t\varepsilon\sqrt{n}}}{\mathbf{P}_n(A)} e^{t^2 n/4}.$$

Haciendo $t = 2\varepsilon/\sqrt{n}$,

$$\frac{e^{-t\varepsilon\sqrt{n}}}{\mathbf{P}_n(A)} e^{t^2 n/4} = \frac{e^{-(\frac{2\varepsilon}{\sqrt{n}})\varepsilon\sqrt{n}}}{\mathbf{P}_n(A)} e^{(\frac{2\varepsilon}{\sqrt{n}})n/4} = \frac{e^{-\varepsilon^2}}{\mathbf{P}_n(A)},$$

con lo que queda completa la demostración. ■

Demostración de la Proposición 33. Primero supongamos que $\mathbf{w} = \mathbf{0} \in U_A(\mathbf{x})$, lo que significa que $w_i = 0$ para todo $1 \leq i \leq n$, entonces por la definición de $U_A(\mathbf{x})$ existe un $\mathbf{z} \in A$, tal que $z_i = x_i$ para todo $1 \leq i \leq n$, por lo que $\mathbf{x} \in A$. Ahora supongamos que $\mathbf{x} \in A$, entonces $w_i = 0$ para todo $1 \leq i \leq n$, por lo que $\mathbf{w} = \mathbf{0} \in U_A(\mathbf{x})$. ■

Demostración del Lema 35. Supongamos que $d_c(\mathbf{x}, A) = r$. Entonces $d(V_A(\mathbf{x}), 0) = r$, y como $V_A(\mathbf{x})$ es cerrado, entonces existe $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in V_A(\mathbf{x})$ tal que $|\mathbf{t}| = r$. Observemos que $U_A(\mathbf{x}) \subseteq \{0,1\}^n$ entonces $U_A(\mathbf{x})$ es finito, supongamos que

$$U_A(\mathbf{x}) = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\},$$

ahora bien $\mathbf{t}$ está en la envolvente convexa de $U_A(\mathbf{x})$ por lo que es una combinación convexa de elementos en $U_A(\mathbf{x})$ (*i.e.* $\mathbf{t} = \sum_{i=1}^{k} \alpha_i \mathbf{w}_i$, donde $\alpha_i \geq 0$, $\sum_{i=1}^{k} \alpha_i = 1$ y $w_i \in U_A(\mathbf{x})$). En virtud de que, por definición de $U_A(\mathbf{x})$, para todo $w_i \in U_A(\mathbf{x})$ existe $z_{w_i} \in A - \mathbf{x}$ (donde $A - \mathbf{x} = \{(a_1 - x_1, a_2 - x_2, \ldots, a_n - x_n) : (a_1, a_2, \ldots, a_n) \in A\}$), tal que $(\mathbf{z}_{w_i})_j$ (la j-ésima coordenada de $\mathbf{z}_{w_i}$) es distinta de cero, lo que implica que $(\mathbf{w}_i)_j$ es distinto de cero también. Como A y $\mathbf{x}$ viven ambos en $[-1,1]^n$, entonces cada coordenada de cada elemento de $A-\mathbf{x}$ tiene valor absoluto menor o igual a 2. Por lo tanto $\left|(\mathbf{z}_{w_i})_j\right| \leq 2\,(\mathbf{w}_i)_j$ (si $(w_i)_j$ es 0 entonces $(\mathbf{z}_{w_i})_j$ también es 0; si $(\mathbf{w}_i)_j$ es 1 entonces $2\,(\mathbf{w}_i)_j = 2 \geq \left|(\mathbf{z}_{w_i})_j\right|$). Ahora, sea

$$\mathbf{z}_t = \sum_{i=1}^{k} \alpha_i \mathbf{z}_{w_i}.$$

Como A es convexo, $A-\mathbf{x}$ es convexo y por lo tanto $\mathbf{z}_t \in A-\mathbf{x}$. Observemos que, si $\mathbf{z}_t = (y_1,\ y_2,\ \ldots,\ y_n)$ entonces $|y_i| \leq 2t_i$, por lo tanto $|\mathbf{z}_t| \leq 2\,|\mathbf{t}| \leq 2r$, y como $\mathbf{z}_t = \mathbf{a}-\mathbf{x}$ para alguna $\mathbf{a} \in A$, por lo que $|\mathbf{a}-\mathbf{x}| \leq 2r$, de donde se sigue que

$$d\,(\mathbf{x},A) = \inf_{\mathbf{a}\in A} \{|\mathbf{a}-\mathbf{x}|\} \leq 2r.$$

■

Demostración del Lema 36. Primero observemos que si $\bar{\mathbf{t}} \in U_{A_{x_n}}(\bar{\mathbf{x}})$, entonces por la definición de A_{x_n} y $U_{A_{x_n}}(\bar{\mathbf{x}})$, tenemos que $(\bar{\mathbf{t}},0) \in U_A\,(\mathbf{x})$ y si $\bar{\mathbf{u}} \in U_B(\bar{\mathbf{x}})$ entonces $(\bar{\mathbf{u}},1) \in U_A(\bar{\mathbf{x}})$. Con lo anterior tenemos que si $\bar{\mathbf{t}} \in V_{A_{x_n}}(\bar{\mathbf{x}})$ y $\bar{\mathbf{u}} \in V_B(\bar{\mathbf{x}})$, entonces para todo $0 \leq \lambda \leq 1$,

$$(\lambda\,(\bar{\mathbf{t}},0) + (1-\lambda)\,(\bar{\mathbf{u}},1)) = (\lambda\bar{\mathbf{t}} + (1-\lambda)\,\bar{\mathbf{u}}\ ,1-\lambda) \in V_A(\mathbf{x}).$$

Por lo tanto, por la definición de distancia combinatoria tenemos que

$$d_c^2\,(\mathbf{x},A) \leq (1-\lambda)^2 + |\lambda t + (1-\lambda)u|^2\,,$$

por la convexidad de la función cuadrada y la desigualdad del triángulo se sigue que

$$d_c^2\,(\mathbf{x},A) \leq (1-\lambda)^2 + \lambda\,|t|^2 + (1-\lambda)\,|u|^2\,,$$

por lo tanto

$$d_c^2\,(\mathbf{x},A) \leq (1-\lambda)^2 + \lambda d_c^2(\bar{\mathbf{x}},A_{x_n}) + (1-\lambda) d_c^2(\bar{\mathbf{x}},B).$$

■

Bibliografía

[1] Anderson, G. W., Guionnet, A., Zeitounni, O. (2009). *An introduction to random matrices.* Cambridge Studies in Advanced Mathematics.

[2] Ash R. B., Doléans-Dade, C. A. (2000). *Probability and measure theory.* Academic Press.

[3] Bai, Z., Silverstein J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices 2nd edition.* Springer.

[4] Domínguez Molina, J. A., Rocha Arteaga, A. (2011). *El teorema de Wigner para matrices aleatorias.* Miscelánea Matemática de la Sociedad Mexicana de Matemáticas no. 52.

[5] Gromov, M., Milman, V.D. (1983). *A topological application of the isoperimetrical inequality.* American J. of Math., Vol. 105, N4, 843-854.

[6] Ledoux, M. (2001). *The concentration of measure phenomenon.* Surveys and Monographs, 89.

[7] Maurey, B. Véase *http://65.54.113.26/Author/1011386/bernard-maurey.*

[8] Milman, V. D. (1971). *A new proof of the theorem of A. Dvoretzky on sections of convex bodies.* Functional Analysis and its Applications, vol. 5, no. 4, 28-37.

[9] Pisier, J. G. G. Véase *http://www.gap-system.org/~history/Biographies/Pisier.html.*

[10] Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press.

[11] Talagrand, M. (1996). *A New Look at Independence.* The Annals of Probability, Vol. 24.

[12] Talagrand, M. (1995). *Concentration of measure and isoperimetric inequalities in product spaces.* Publications mathématiques de l'I.H.É.S., tome 81.

[13] Tao, T. (2012). *Topics in Random Matrix Theory.* Graduate Studies in Mathematics. Volume 132, American Mathematical Society.