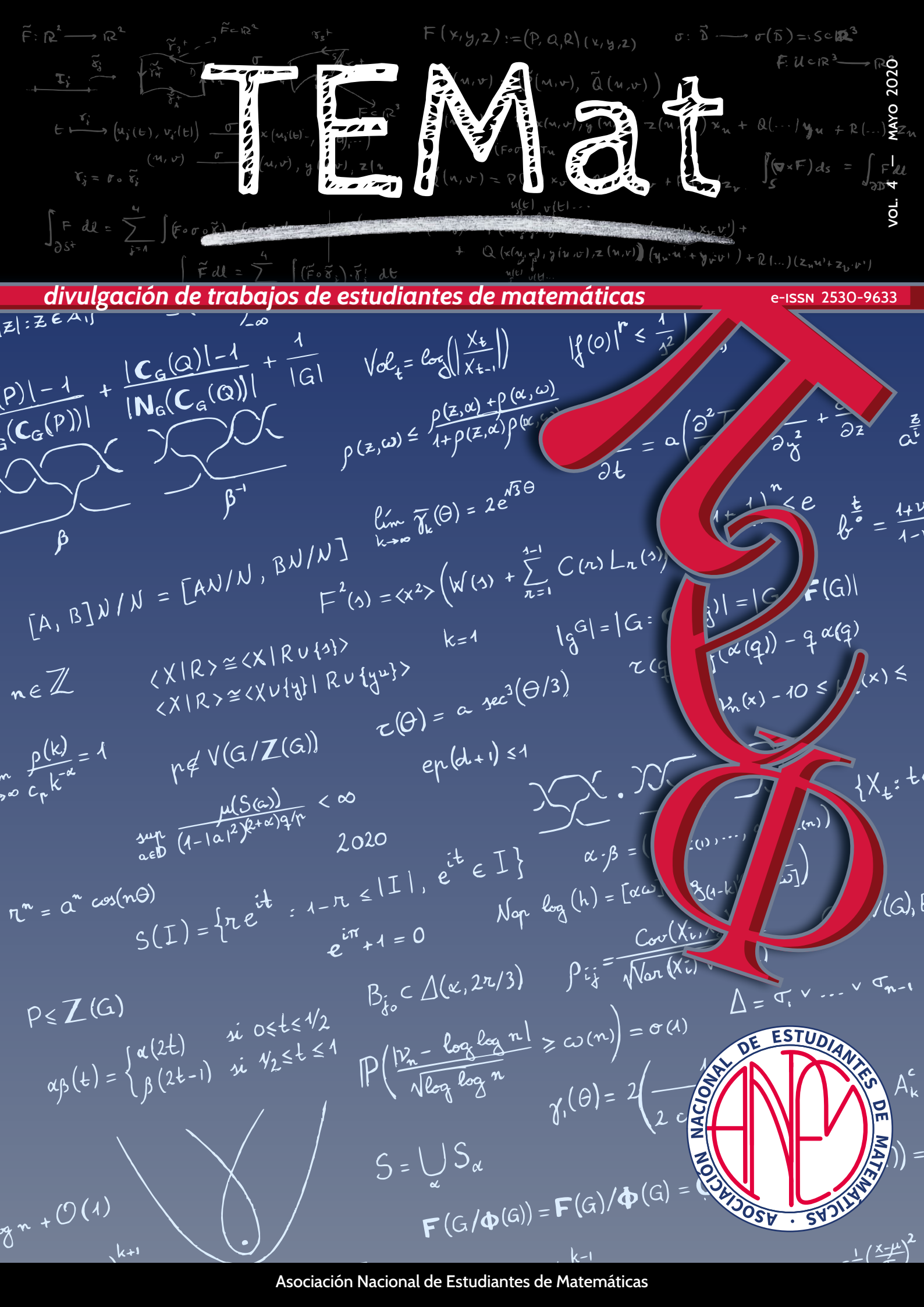
TEMat

VOL. 4 — MAYO 2020

divulgación de trabajos de estudiantes de matemáticas

e-ISSN 2530-9633



Asociación Nacional de Estudiantes de Matemáticas

TEMat

divulgación de trabajos de estudiantes de matemáticas

volumen 4
mayo de 2020

https://temat.es/volumen/2020/
http://www.anem.es/

Una iniciativa de la
Asociación Nacional de Estudiantes de Matemáticas

Publica

Asociación Nacional de Estudiantes de Matemáticas
Plaza de las Ciencias, 3
Despacho 525, Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 – Madrid

temat@temat.es
publicaciones@anem.es
contacto@anem.es

Colabora

Real Sociedad Matemática Española
Plaza de las Ciencias, 3
Despacho 525, Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 – Madrid

Diseño de portada: Roberto Berná Larrosa, rbernalarrosa@gmail.com

TEMat, divulgación de trabajos de estudiantes de matemáticas – volumen 4 – mayo de 2020
e-ISSN: 2530-9633
https://temat.es/

Equipo

Revisiones externas

En este volumen han colaborado realizando revisiones externas:

Yago Antolín, Universidad Autónoma de Madrid
Jorge Betancor, Universidad de La Laguna
Gonzalo Cao Labora, Universitat Politècnica de Catalunya
Daniel Eceizabarrena, BCAM - Basque Center for Applied Mathematics
Ramon Esteban Romero, Universitat de València
Raquel Fernández Peralta, Universitat de les Illes Balears
Antonio Galbis, Universitat de València
Adrián Llinares, Universidad Autónoma de Madrid
Mateu Morro
Marialaura Noce, Universidad del País Vasco / Euskal Herriko Unibertsitatea y Università degli Studi di Palermo
Guillem Perarnau, Universitat Politècnica de Catalunya
Ángeles Ruiz González, Highlands School Sevilla
Juan Francisco Torres Sancho, Universidad de Salamanca

Sobre TEMat

TEMat es una revista de divulgación de trabajos de estudiantes de matemáticas publicada sin ánimo de lucro por la Asociación Nacional de Estudiantes de Matemáticas. Se busca publicar trabajos divulgativos de matemáticas, escritos principalmente (pero no exclusivamente) por estudiantes, de todo tipo: breves reseñas, introducciones a temas de investigación complejos, o artículos explicando las bases e incluso algún pequeño resultado de trabajos desarrollados por estudiantes.

TEMat persigue el doble objetivo de dar visibilidad a la calidad y diversidad de los trabajos realizados por estudiantes de matemáticas en los centros españoles a la vez que permite a los estudiantes publicar sus primeros artículos, familiarizándose así con el proceso de redacción, revisión y corrección que va asociado a la actividad investigadora.

Se contemplan para su publicación artículos escritos en castellano de todas las áreas de las matemáticas, incluyendo álgebra, análisis, ciencias de la computación, combinatoria, educación matemática, estadística, geometría, teoría de números y cualquier otra área de las matemáticas (puras y aplicadas), así como aplicaciones científicas o tecnológicas en las que las matemáticas jueguen un papel central.

Índice general

Carta del presidente de la ANEM

Pese a las duras circunstancias que actualmente vivimos, *TEMat* sigue adelante con su cuarto volumen de este ambicioso proyecto único a nivel internacional y consolidado como una parte fundamental de la ANEM.

En cada edición son mayores el interés por *TEMat* y el número de artículos recibidos por parte del comité editorial, muestra innegable de la importancia que ha obtenido *TEMat* en estos cuatro años de vida y la que está por venir. Ejemplo de ello son los volúmenes monográficos, de los cuales el primero, de la *Escuela-Taller de Análisis Funcional*, está previsto que se publique este año.

Ahora, más que nunca, es necesaria la defensa de la investigación científica abierta y por eso seguimos apostando por la involucración del estudiantado en esta para su futuro profesional. Creemos firmemente que la educación y la investigación son bases indispensables de nuestra sociedad y las matemáticas son un pilar fundamental para el desarrollo de un ecosistema científico de primer nivel.

No podría terminar sin dar las gracias a todos los miembros del comité editorial por su incansable esfuerzo para que este ilusionante proyecto siga creciendo, así como a los autores y revisores de los artículos, sin los cuales nada de esto sería posible. Más que nunca, gracias por todo el esfuerzo para que esta edición vea la luz. Esperamos que, quienes no lo hayáis hecho ya, os apuntéis a enviar vuestros artículos a *TEMat* o a revisar los de vuestros compañeros para que esta increíble iniciativa siga adelante.

Alfonso Márquez Martínez,
presidente de la ANEM.

Sevilla, mayo de 2020.

TEMat

Este trabajo fue galardonado con el primer premio en la edición de 2019 del Premi Poincaré, entregado por la Facultat de Matemàtiques i Estadística de la Universitat Politècnica de Catalunya.

Estudio de una familia de curvas formadas inductivamente a partir de una construcción geométrica

Pau Redon Muñoz[a]
Universitat Politècnica de Catalunya (UPC)
pauredonm@gmail.com

Resumen: Este artículo nace de preguntarse cuál es la curva equidistante a una parábola y a su foco. Veremos que la respuesta a esta pregunta es una parte de la cúbica de Tschirnhausen, una curva estudiada en el siglo XVII por varios matemáticos. Para obtener esta curva, aplicaremos a la parábola y a su foco una construcción geométrica que también puede ser aplicada en la cúbica de Tschirnhausen y al foco de la parábola inicial para obtener otra curva, y a las curvas subsecuentes. Parte del artículo se centrará en estudiar las propiedades de esta familia de curvas.

Abstract: This article is built over asking what is the curve equidistant from a parabola and its focus. We will see that the answer to this question is a part of Tschirnhausen's cubic, a curve studied by various mathematicians during the 17th century. To obtain this curve, we will apply to a parabola and its focus a geometric contruction that can also be applied to Tschirnhausen's cubic and the initial parabola's focus to obtain another curve, and to the subsequent curves. Part of the paper will be dedicated to studying the properties of this family of curves.

Palabras clave: geometría euclídea, geometría algebraica, curvas, espirales sinusoidales.

MSC2010: 14H50.

Recibido: 18 de septiembre de 2019.
Aceptado: 30 de abril de 2020.

Agradecimientos: Quiero agradecer a todos los que me han ayudado a realizar este artículo, en especial a mi tutora del Trabajo de Investigación de Bachillerato, Yolanda Segarra, y al jurado de los Premios Poincaré por escoger mi trabajo como ganador y darme la oportunidad de publicarlo en esta revista.

Referencia: Redon Muñoz, Pau. «Estudio de una familia de curvas formadas inductivamente a partir de una construcción geométrica». En: *TEMat*, 4 (2020), págs. 1-14. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p1.

[a]Este trabajo fue realizado como Trabajo de investigación de Bachillerato en el Institut d'Alella.

1. Introducción

Se cree que las secciones cónicas fueron definidas por primera vez por el geómetra griego Menaechmus (380-320 a. C.) como parte de su solución al problema de duplicar el cubo. La definición dada por Menaechmus difiere de la definición actual de sección cónica y es la siguiente. La sección cónica es determinada por la intersección de un cono con un plano perpendicular a una generatriz cualquiera de este cono. El tipo de cónica definida dependerá del ángulo formado en el vértice del cono: si el ángulo es obtuso, recto o agudo la intersección será una hipérbola, una parábola o una elipse, respectivamente.

Más tarde, Apolonio de Perga (262-190 a. C.) haría un importante progreso en la teoría de las secciones cónicas en su más famosa obra *Sobre las secciones cónicas*. Este estudio de Apolonio hizo posible extender la definición anterior de Menaechmus a la usada actualmente. Demostró que, independientemente del ángulo formado en el vértice de un doble cono, la intersección de este con cualquier plano resulta en una sección cónica.

A principios del siglo XVII, aparece la geometría analítica y se introducen los sistemas de coordenadas en la geometría, por obra de personajes como Desargues, Kepler, Descartes y Fermat. Kepler empezó a usar el termino «foco» de una parábola, y fue el primero en enunciar rigurosamente el concepto de continuidad (el cambio continuo de una entidad matemática de un estado a otro) para tratar la parábola como el caso límite entre una elipse y una hipérbola cuando uno de los dos focos se va al infinito. Con esta introducción del término «foco» se da paso a una de las definiciones de la parábola más conocidas y usadas actualmente: la parábola es el lugar geométrico formado por los puntos del plano que están a la misma distancia de un punto llamado foco y una recta llamada directriz.

La definición de parábola que acabamos de ver es la base de este artículo, ya que es la definición que ha inspirado la pregunta siguiente: «¿Cuál es la curva formada por los puntos equidistantes a una parábola y a su foco?». Veremos más adelante en el artículo que la respuesta a esta pregunta coincide con una parte de una curva estudiada en finales del siglo XVII por Tschirnhausen, L'Hôpital y Catalan.

Para estudiar esta curva, proponemos una definición de un lugar geométrico, y, más tarde en el artículo, veremos que esta definición responde a la pregunta formulada en el párrafo anterior.

Definición 1. Sea C una curva derivable en todos sus puntos[1], y sea F un punto coplanario con C pero exterior a esta. Llamamos $f(C,F)$ al lugar geométrico de los puntos que se obtienen haciendo para cada $P \in C$ la intersección entre la mediatriz del segmento FP y la recta normal (o las rectas normales, si hay más de una) a la curva en el punto F. ◂

Este artículo tendrá dos partes. En la primera parte analizaremos el lugar geométrico $\gamma = f(\Gamma, F)$ para cualquier parábola Γ con foco F. En esta parte veremos la propiedad de equidistancia de la curva $f(\Gamma, F)$ a Γ y F, y veremos la similitud de la cúbica de Tschirnhausen con γ. En la segunda parte del artículo estudiaremos la familia de curvas a las que llamaremos γ_k, definidas de la siguiente manera: γ_0 es una línea recta arbitraria en el plano, y F un punto exterior a esta línea; entonces $\gamma_k = f(\gamma_{k-1}, F)$ para $k \in \mathbb{N}$. Observemos que, con esta definición, $\gamma_1 = \Gamma$.

Durante todo el artículo supondremos que estamos trabajando en el plano euclídeo.

2. La curva γ

Todas las parábolas son semejantes, lo que significa que, dadas dos parábolas arbitrarias, siempre podemos transformar una en la otra mediante transformaciones rígidas (traslaciones y rotaciones) u homotecias. Entonces, sin perdida de generalidad, durante esta sección trabajaremos con la forma general de la parábola con el eje de simetría vertical, $y = a(x-h)^2 + k$, con k, h y a reales y $a \neq 0$.

[1] Una curva plana C es derivable en todo punto si admite unas ecuaciones paramétricas $x = f(t)$, $y = g(t)$ de modo que $f(t)$ y $g(t)$ sean funciones derivables para todo t.

2.1. Ecuación paramétrica de γ

Para encontrar la ecuación de γ, usaremos la definición 1. Consideremos una parábola Γ de ecuación $\Gamma:\ y = a(x-h)^2 + k$. Es conocido que el foco de esta parábola tiene coordenadas $F = (h,\ k + 1/4a)$. Para empezar, encontraremos la ecuación de la mediatriz entre el foco y un punto $P = (t,\ a(t-h)^2 + k)$ de la parábola. Entonces, encontraremos la ecuación de la línea normal a la parábola en el punto P. Finalmente, encontraremos la intersección Q entre estas dos líneas rectas y, como sabemos que $Q \in \gamma$, basta con tratar la variable t del punto P como el parámetro de la ecuación paramétrica.

La ecuación de la mediatriz entre P y F es

$$(x-h)^2 + \left(y - \left(k + \frac{1}{4a}\right)\right)^2 = (x-t)^2 + \left(y - \left(a(t-h)^2 + k\right)\right)^2.$$

La ecuación de la recta normal a la parábola en el punto P es la siguiente:

$$y = a(t-h)^2 + k + \frac{1}{2a} + \frac{x-h}{2a(h-t)}.$$

Ahora, para encontrar el punto de intersección de las dos rectas tenemos que resolver el sistema dado por las ecuaciones de las dos rectas:

$$\begin{cases} y = \dfrac{x-h}{2a(h-t)} + k + a(h-t)^2 + \dfrac{1}{2a}, \\ (x-h)^2 + \left(y - \left(k + \dfrac{1}{4a}\right)\right)^2 = (x-t)^2 + (y - (a(t-h)^2 + k))^2. \end{cases}$$

Resolviendo este sistema resulta que γ es una curva de ecuaciones paramétricas derivables, con parámetro $t \in \mathbb{R}$:

$$\begin{aligned} x(t) &= \frac{3(t-h)}{4} + h - a^2(t-h)^3, \\ y(t) &= \frac{12a^2(h-t)^2 + 8ak + 1}{8a}. \end{aligned}$$

2.2. Representación gráfica de γ

Para conocer las propiedades de la curva, es importante conocer su representación gráfica. Para hacerlo, usaremos la parábola de ecuación $y = x^2$. Entonces, las ecuaciones paramétricas de la curva γ respectiva a esta parábola son

$$\begin{aligned} x(t) &= \frac{3t}{4} - t^3, \\ y(t) &= \frac{12t^2 + 1}{8}, \end{aligned}$$

y la representación de estas ecuaciones se muestra en la figura 1.

2.3. Propiedad de equidistancia de γ

En esta sección responderemos a la pregunta formulada en la introducción: «¿Cuál es la curva formada por los puntos equidistantes a una parábola y a su foco?» Demostraremos que la respuesta es una parte de la curva γ y veremos cuál es.

Observación 2. Veamos en qué casos la construcción aplicada en dos puntos $P_1 = (t_1,\ a(t_1-h)^2 + k)$ y $P_2 = (t_2,\ a(t_2-h)^2 + k)$ de la parábola diferentes resulta en el mismo punto Q de la curva γ. Cuando esto pasa, tenemos que

$$x(t_1) = x(t_2), \quad y(t_1) = y(t_2),$$

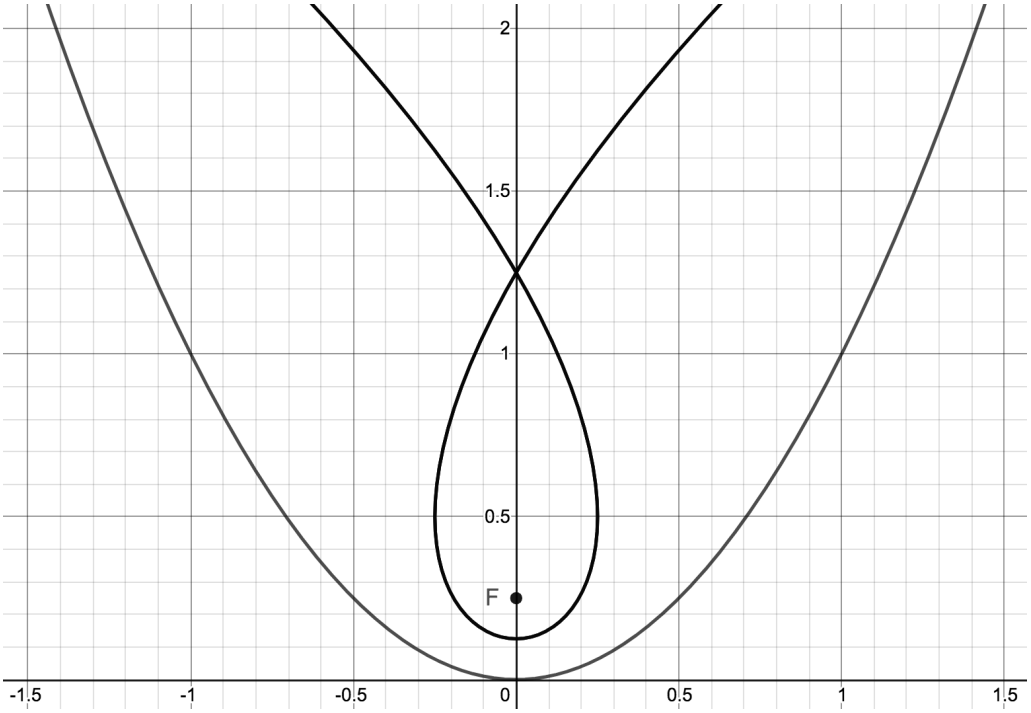


Figura 1: Representación gráfica de γ con su respectiva parábola y foco.

es decir,

$$\begin{cases} \dfrac{3(t_1-h)}{4} + h - a^2(t_1-h)^3 = \dfrac{3(t_2-h)}{4} + h - a^2(t_2-h)^3, \\ \dfrac{12a^2(h-t_1)^2 + 8ak + 1}{8a} = \dfrac{12a^2(h-t_2)^2 + 8ak + 1}{8a}. \end{cases}$$

Resolviendo el sistema en función de t_1 y t_2, sin perdida de generalidad, llegamos a que $t_1 = h - \sqrt{3}/2a$ y $t_2 = h + \sqrt{3}/2a$. El punto resultante de aplicar la construcción en P_1 o en P_2 es $Q = (h,\ 5a/4 + k)$. El punto Q es el único punto con la propiedad que buscábamos.

Si $\vec{x}(t) = (x(t), y(t))$, derivando esta función en los puntos t_1 y t_2 vemos que la curva admite dos tangentes distintas en el punto Q. ◂

Definición 3. Sean $x(t)$ e $y(t)$ las ecuaciones que definen la curva γ para una parábola genérica con ecuación $y = a(x-h)^2 + k$. Definimos el bucle de la curva γ como el conjunto de puntos $(x(t), y(t))$ tales que $h - \sqrt{3}/2a \leq t \leq h + \sqrt{3}/2a$. ◂

Nota **4.** Debido a la observación 2, tenemos que el bucle de γ forma una curva cerrada, ya que es continua y los dos extremos de la curva son el mismo punto. ◂

Ahora enunciaremos y demostraremos dos lemas que usaremos en la demostración de la propiedad de equidistancia de γ.

Lema 5. *Sea $\Gamma:\ y = a(x-h)^2 + k$, y sea γ la curva formada a partir de esta parábola usando la definición 1. Separamos el plano real en dos semiplanos $S_1 = \{(x,y) \in \mathbb{R}^2 : x \leq h\}$ y $S_2 = \{(x,y) \in \mathbb{R}^2 : x \geq h\}$. Consideremos el punto $P = (t, a(t-h)^2 + k)$ y el punto Q de γ obtenido al aplicar la construcción a P, $Q = (x(t), y(t))$. Entonces, P y Q están en el mismo semiplano si y solo si Q forma parte del bucle de γ.*

Demostración. Que Q pertenezca al bucle de γ es equivalente a decir que $a^2(t-h)^2 \leq 3/4$. La función $x(t)$ se puede reescribir como

$$x(t) = (t-h)\left(\frac{3}{4} - a^2(t-h)^2\right) + h.$$

Así, es fácil ver que, si $t \leq h$, entonces $x(t) \leq h$ si y solo si $a^2(t-h)^2 \leq 3/4$, y análogamente para $t \geq h$. ∎

Lema 6. *Consideremos la parábola $\Gamma: y = a(x-h)^2 + k$ y los dos semiplanos S_1 y S_2 del lema anterior. Sean $P = (x_1, y_1)$ y $Q = (x_2, y_2)$ dos puntos diferentes de la parábola Γ en el mismo semiplano S_i, con x_1 y x_2 diferentes a h. Las rectas r_1 y r_2 normales a la parábola en los puntos P y Q, respectivamente, se cruzan en el semiplano S_j, con $j \neq i$.*

Demostración. Supongamos sin perdida de generalidad que tanto P como Q están en el semiplano S_1, y que $a > 0$. Entonces, $x_1 < h$ y $x_2 < h$. También supongamos que $x_1 > x_2$, así que $y_1 < y_2$. Decimos que $y_1 + d = y_2$ para algún $d > 0$.

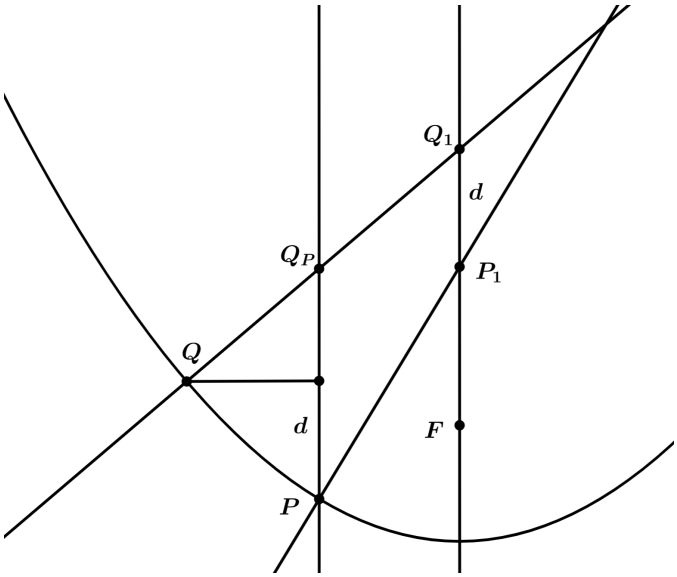


Figura 2: Diagrama de la construcción del lema 6.

Sean P_1 y Q_1 los puntos de intersección de las rectas normales r_1 y r_2 con la recta $x = h$. Se puede calcular[2] que $P_1 = (h, y_1 + 1/2a)$ y $Q_1 = (h, y_1 + d + 1/2a)$. De aquí deducimos que $|P_1Q_1| = d$, y Q_1 está por encima de P_1.

Entonces, sea $Q_P = (x_1, t)$ el punto de intersección de la recta r_2 con la recta $x = x_1$. Como Q_P está en el segmento QQ_1, tenemos que $y_1 + d < t < y_1 + d + 1/2a$ y, en consecuencia, $|PQ_P| > d = |P_1Q_1|$, con Q_P por encima de P.

Finalmente, consideramos las dos rectas paralelas $x = x_1$ y $x = h$, y las intersecciones de estas dos rectas con r_1 y r_2, que son los puntos P y P_1, y Q_P y Q_1, respectivamente. Como $|PQ_P| > |P_1Q_1|$, podemos aplicar el teorema de Tales para concluir que r_1 y r_2 se cruzan en el semiplano S_2. ∎

Teorema 7. *El bucle de la curva γ construida a partir de la parábola $\Gamma: y = a(x-h)^2 + k$ con foco F es el lugar geométrico de todos los puntos que están a igual distancia de Γ que de F.*

Demostración. Para demostrar este teorema, primero veremos que, si un punto cumple la propiedad de equidistancia, entonces pertenece a γ, y después hallaremos los puntos de γ que satisfacen la propiedad.

Sea R un punto que no pertenece a la parábola. Consideramos el círculo más pequeño con centro R tal que este círculo tenga algún punto en común con la parábola. Llamamos P a uno de estos puntos en común (si hay más de uno, escogemos un punto arbitrario entre ellos). Como la parábola es una función continua y derivable, la circunferencia será tangente a la parábola en el punto P, lo cual implica que R pertenece a la recta normal a Γ en P. Como hemos considerado la circunferencia más pequeña posible, la distancia de R a la parábola es la misma que la distancia de R a P. Entonces, el punto R cumple la propiedad de equidistancia solo si está a la misma distancia de P que de F, o, lo que es equivalente, que R se halla en la mediatriz entre P y F. Esto implica que $R \in \gamma$.

[2] Este cálculo está hecho en el trabajo original en el que está basado este artículo [5].

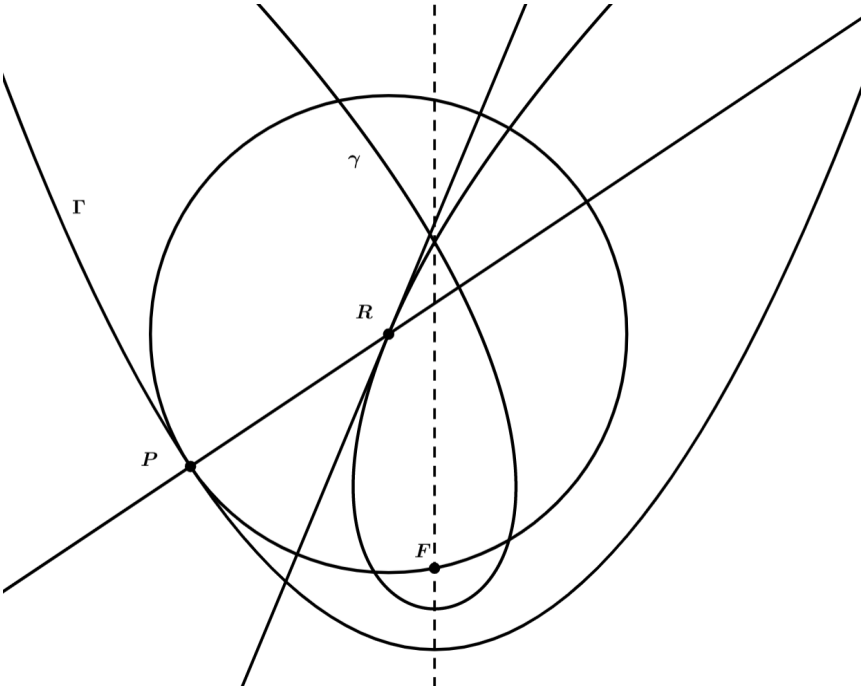


Figura 3: Diagrama de la primera parte del teorema 7.

Ahora hallaremos los puntos de γ que cumplen la propiedad deseada. Sea A un punto de γ. Sabemos por la construcción de γ que existe un punto B de la parábola Γ tal que la recta normal a Γ en B y la mediatriz del segmento BF se cortan en A. Entonces, como $|BA| = |AF|$, el punto A cumple la propiedad de equidistancia si y solo si el punto de Γ más cercano a A es B.

Supongamos que A no pertenece al bucle de γ. Por el lema 5, A y B no están en el mismo semiplano. Entonces, B no puede ser el punto de Γ más cercano a A, ya que podemos comprobar que el punto B', simétrico a B respeto a la recta $x = h$, pertenece a Γ y $|B'A| < |BA|$.

Ahora, supongamos que A pertenece al bucle de γ. Si A es el punto medio entre F y el vértice de la parábola, A cumple la propiedad de equidistancia ya que es el único punto de γ tal que su punto más cercano de la parábola es su vértice[3]. Entonces, supongamos que A es un punto del bucle diferente del punto medio entre F y el vértice. Por el lema 5, el punto B está en el mismo semiplano que A; por el lema 6, B es el único punto de este semiplano, excluyendo el vértice de la parábola, tal que la recta normal a Γ en B pasa por A. Entonces, $|BA|$ es la distancia entre A y la parábola, y hemos demostrado el teorema. ∎

2.4. Semejanza de γ con la cúbica de Tschirnhausen

La cúbica de Tschirnhausen[4], también llamada trisectriz de Catalan y cúbica de l'Hôpital, es una curva estudiada por Tschirnhausen en 1690 y por l'Hôpital en 1696. Esta curva se caracteriza por las ecuaciones paramétricas siguientes:

$$x = a(1 - 3t^2),$$
$$y = at(3 - t^2),$$

y su ecuación polar[4] es

$$\tau(\theta) = a \sec^3(\theta/3).$$

[3] Es fácil de comprobar, se deja como ejercicio al lector.

[4] Para representar un punto P en coordenadas polares usaremos $P = (r, \alpha)$, donde r es la distancia al origen y α el ángulo que forma la parte positiva del eje X con PO, en sentido antihorario. Definimos la ecuación polar de una curva como la distancia r en función del ángulo α, $r = f(\alpha)$.

Queremos demostrar que esta curva es semejante a la curva γ; para ello, demostraremos primero los siguientes dos lemas.

Lema 8. *Todas las curvas γ son semejantes. Esto significa que se puede transformar cualquier curva γ en otra usando solo transformaciones rígidas y homotecias.*

Demostración. Sabemos que todas las parábolas son semejantes. Entonces, consideremos dos curvas γ_1 y γ_2 con sus respectivas parábolas Γ_1 y Γ_2. Si aplicamos a todo el sistema formado por Γ_1 y Γ_2 las transformaciones necesarias para transformar la primera parábola en la segunda, también transformaremos γ_1 en γ_2, ya que la curva gamma está formada por una construcción geométrica basada solo en la parábola. Entonces, todas las curvas γ son semejantes. ∎

Lema 9. *Todas las cúbicas de Tschirnhausen son semejantes.*

Demostración. Sean τ_1 y τ_2 dos cúbicas de Tschirnhausen, con ecuaciones polares $\tau_1(\theta) = a\sec^3(\theta/3)$ y $\tau_2(\theta) = b\sec^3(\theta/3)$. Si aplicamos una homotecia a τ_2 de razón a/b y centro de homotecia en el origen de coordenadas, obtendremos la curva τ_1, así que todas las cúbicas de Tschirnhausen son semejantes. ∎

Teorema 10. *La cúbica de Tschirnhausen es semejante a γ.*

Demostración. Usaremos los lemas anteriores para demostrar que las dos curvas son similares. Demostraremos que la cúbica de Tschirnhausen para un valor concreto de a es semejante a una curva γ concreta. Escogeremos la cúbica de Tschirnhausen con $a = 1/8$, y la curva γ con parámetros $a = 1$, $k = h = 0$. Transformaremos la cúbica de Tschirnhausen en γ a partir de transformaciones rígidas. Primero usaremos una rotación de 90°, y más tarde una translación.

Para hacer la rotación, multiplicaremos las coordenadas de la ecuación paramétrica por la matriz de rotación:

$$\left(\frac{1}{8}(1-3t^2),\ \frac{1}{8}t(3-t^2)\right)\begin{pmatrix} \cos 90^\circ & -\operatorname{sen} 90^\circ \\ \operatorname{sen} 90^\circ & \cos 90^\circ \end{pmatrix} = \left(\frac{1}{8}t(3-t^2),\ \frac{1}{8}(3t^2-1)\right).$$

Podemos hacer el cambio de variable $t = 2s$ para obtener que

$$x = \frac{3s}{4} - s^3, \quad y = \frac{1}{8}(12s^2 - 1)).$$

Ahora, aplicamos la translación dada por el vector $\vec{v} = (0,\ 1/4)$ a la curva. Nos quedan las siguientes ecuaciones:

$$x = \frac{3s}{4} - s^3, \quad y = \frac{12s^2+1}{8}.$$

Como podemos ver, estas ecuaciones paramétricas son las correspondientes a la curva γ con parámetros $a = 1$, $k = h = 0$, así que, si tenemos en cuenta los dos lemas anteriores, el teorema está demostrado. ∎

3. Generalización de la curva γ

En la sección anterior hemos hablado de la cúbica de Tschirnhausen, intentando responder la pregunta de la equidistancia que hemos planteado en la introducción. Para responder a esta pregunta, hemos dado una construcción geométrica en la definición 1. En esta tercera sección generalizaremos el concepto de curva γ tal y como hemos mencionado en la introducción.

Recordemos que $F = (0,\ 0)$, $\gamma_0 :\ y = -1$ y $\gamma_k = f(\gamma_{k-1},\ F)$. A continuación demostraremos las siguientes tres propiedades de la curva γ_k (para $k \geq 1$):

1. Sea P_0 un punto cualquiera de la directriz γ_0. Para cada punto P_{k-1} ($k \geq 1$) que pertenece a la curva γ_{k-1}, llamamos P_k al punto resultante de aplicar la construcción explicada en la definición 1 en el punto P_{k-1}. Entonces, si $A = (0,\ -1)$, la siguiente relación de ángulos se cumple:

$$\angle AFP_0 = \angle P_{k-1}FP_k.$$

2. La ecuación polar que define a la curva γ_k es la siguiente:

$$\gamma_k(\theta) = 2\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{k+1}\right)}\right)^{k+1}.$$

3. La mediatriz entre los puntos P_{k-1} y F es tangente a la curva γ_k en el punto P_k.

Haremos juntamente la demostración de estas tres propiedades. Será una demostración por inducción fuerte, donde la hipótesis inductiva será suponer que las tres propiedades mencionadas se cumplen para los números naturales menores o iguales que k.

Demostración.
Caso base.
Demostraremos el caso base para $k = 1$. Recordemos que $A = (0, -1)$. El punto P_0 pertenece a la recta $\gamma_0 : y = -1$. Definimos a P_0 con el ángulo α que forma el segmento FP_0 con el segmento FA.

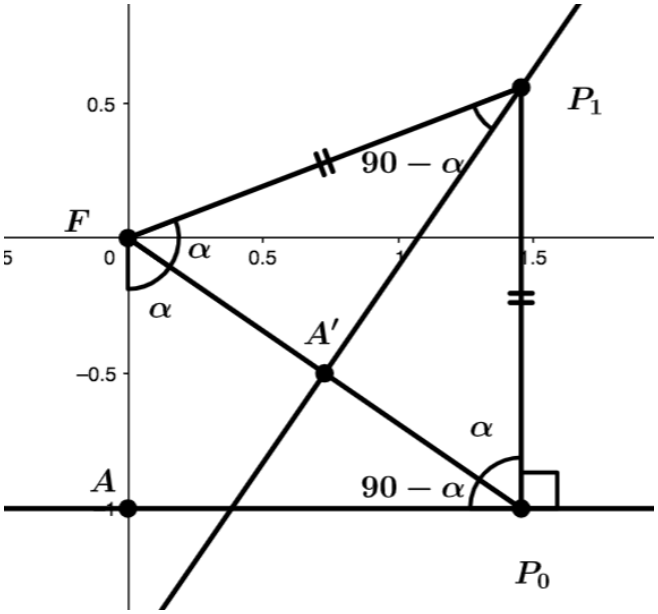


Figura 4: Diagrama de la construcción utilizada para caso base.

Como podemos ver en la figura 4, el punto P_1 es la intersección entre la mediatriz de FP_0 y la perpendicular a γ_0 en el punto P_0. Ya que el ángulo $\angle FAP_0$ es recto, $\angle AP_0F = 90° - \alpha$ y $\angle FP_0P_1 = \alpha$. Entonces, como P_1 está en la mediatriz de PF_0, sabemos que $\angle P_0FP_1 = \alpha$, lo que demuestra la primera propiedad para el caso base.

Definimos A' como el punto medio de FP_0. Sabemos que $|FA| = 1$, así que $|FP_0| = 1/\cos\alpha$. Entonces, $|FA'| = |FP_0|/2 = 1/(2\cos\alpha)$ y, si nos fijamos en el triángulo rectángulo $\triangle FA'P_1$, vemos también que $|FP_1| = |FA'|/\cos\alpha = 2(1/(2\cos\alpha))^2$. Llamaremos θ al ángulo que define el punto P_1 en forma polar. Entonces, $2\alpha = \angle AFP_1 = \pi/2 + \theta$, de lo que deducimos que $\alpha = (\theta + \pi/2)/2$. Finalmente, escribimos la fórmula para la distancia $|FP_1|$ en función del ángulo θ, que coincide con la ecuación polar de γ_1:

$$\gamma_1(\theta) = 2\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{2}\right)}\right)^2.$$

La segunda propiedad está demostrada para el caso base.

Sabemos que γ_1 es una parábola, y es conocido que la mediatriz entre el foco de la parábola y cualquier punto de su directriz es tangente a la parábola, así que la tercera parte de la hipótesis inductiva también se cumple.

Paso inductivo.
Supongamos que las tres propiedades están demostradas para todo entero entre 1 y k, los dos incluidos. Ahora las demostraremos para $k+1$.

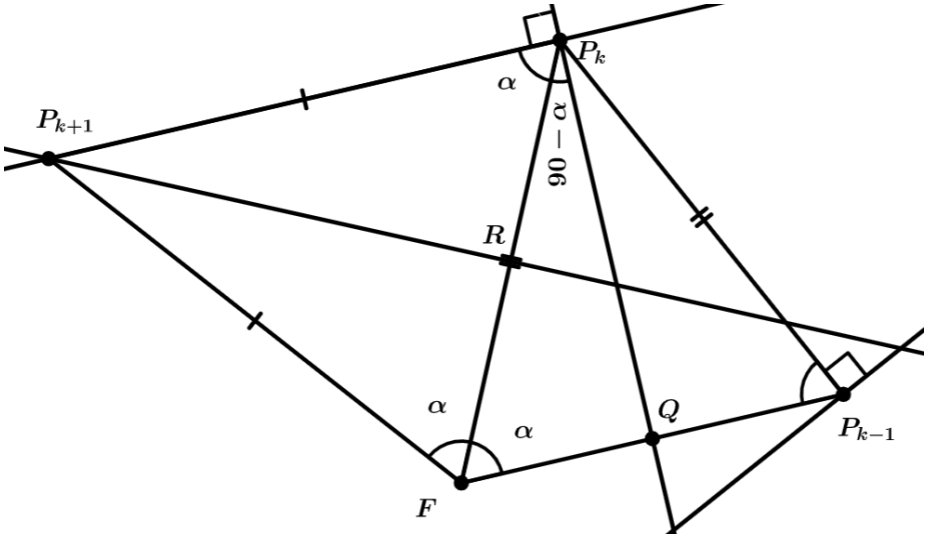


Figura 5: Diagrama de la construcción usada en el paso inductivo.

Sabemos por la tercera propiedad de la hipótesis inductiva que la mediatriz entre P_{k-1} y F es tangente a la curva γ_k en el punto P_k. Entonces, la recta perpendicular a esta mediatriz en el punto P_k será la recta normal a la curva γ_k que necesitamos para encontrar el punto P_{k+1} a partir de la construcción ya definida. El punto P_{k+1} es la intersección entre esta recta normal y la mediatriz entre P_k y F.

Llamamos Q y R a los puntos medios de los segmentos $P_{k-1}F$ y P_kF, respectivamente, como podemos observar en la figura 5. Por la primera propiedad de la hipótesis de inducción, $\angle P_{k-1}FP_k = \alpha$. Como $\angle P_kQF$ es recto, tenemos que $\angle FP_kQ = 90° - \alpha$, $\angle P_{k+1}P_kF = \alpha$ y, finalmente, $P_kFP_{k+1} = \alpha$, lo cual demuestra la primera propiedad para $k+1$.

Recordemos que $\alpha = (\theta_1 + \pi/2)/2$, donde θ_1 es el ángulo que FP_1 forma con el eje de abscisas. En consecuencia, $\theta_1 = 2\alpha - \pi/2$. Si llamamos θ al ángulo que FP_k forma con el eje de abscisas, podemos afirmar que $\theta = \theta_1 + (k-1)\alpha$ ya que, por la primera propiedad de la hipótesis, $\angle P_{n-1}FP_n = \alpha$ para todo n natural menor que $k+1$. Entonces, $\theta = (k+1)\alpha - \pi/2$. En consecuencia, también sabemos que FP_{k+1} forma un ángulo de valor $\alpha(k+2) - \pi/2$ con la parte positiva del eje de abscisas. Por la segunda parte de la hipótesis de inducción, sabemos que

$$|FP_k| = \gamma_k(\theta) = 2\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{k+1}\right)}\right)^{k+1}.$$

Si aplicamos el cambio de variable $\theta = \alpha(k+1) - \pi/2$, tenemos que

$$|FP_k| = \gamma_k\left(\alpha(k+1) - \frac{\pi}{2}\right) = 2\left(\frac{1}{2\cos\alpha}\right)^{k+1}.$$

Tenemos que $|FR| = |FP_k|/2$ y, considerando el triángulo rectángulo $\triangle FRP_{k+1}$, sabemos que $|FP_{k+1}| = |FP_k|/(2\cos\alpha)$. Entonces,

$$\gamma_{k+1}\left(\alpha(k+2) - \frac{\pi}{2}\right) = |FP_{k+1}| = \frac{\gamma_k\left(\alpha(k+1) - \frac{\pi}{2}\right)}{2\cos\alpha} = 2\left(\frac{1}{2\cos\alpha}\right)^{k+2}$$

Si ahora deshacemos el cambio de variable $\alpha = (\theta + \pi/2)/(k+2)$, tenemos que

$$\gamma_{k+1}(\theta) = 2\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{k+2}\right)}\right)^{k+2}.$$

Por lo tanto, la segunda propiedad ha sido demostrada para $k+1$.

Para finalizar esta demostración nos falta demostrar la tercera propiedad para $k+1$, que dice que la mediatriz entre los puntos P_k y F es tangente a la curva γ_{k+1} en el punto P_{k+1}. Lo demostraremos de la siguiente manera. Sabemos que la mediatriz de P_kF y γ_{k+1} se intersecan en el punto P_{k+1} por la construcción que hemos definido. Entonces, tenemos que ver que las derivadas de las dos curvas tienen el mismo valor en el punto P_{k+1}.

El punto P_k en función del ángulo α en coordenadas polares es $(2(1/(2\cos\alpha))^{k+1},\ \alpha(k+1)-\pi/2)$.

Sabemos que la ecuación polar de una recta r es $r(\beta)=a\sec(\beta-\beta_0)$, donde a es la distancia mínima de r al origen y β_0 es el ángulo que forma la perpendicular a r que pasa por el origen con el eje de las X positivo. Ya que el ángulo que forma FP_k con el eje de abscisas positivo es $\alpha(k+1)-\pi/2$ y la distancia mínima a la recta es $2(1/(2\cos\alpha))^{k+1}$, la ecuación polar de la mediatriz es

$$m(\theta)=\frac{\left(\frac{1}{2\cos\alpha}\right)^{k+1}}{\cos(\theta-(\alpha(k+1)-\pi/2))}.$$

La derivada de la mediatriz respecto a θ es, entonces,

$$m'(\theta)=\left(\frac{1}{2\cos\alpha}\right)^{k+1}\frac{\operatorname{sen}(\theta-(\alpha(k+1)-\pi/2)}{\cos^2(\theta-(\alpha(k+1)-\pi/2)},$$

y el valor de la derivada en el punto P_{k+1}, que está definido por el ángulo $\theta=\alpha(k+2)-\pi/2$, es

$$m'(\alpha(k+2)-\pi/2)=\left(\frac{1}{2\cos\alpha}\right)^{k+1}\frac{\operatorname{sen}(\alpha)}{\cos^2(\alpha)}.$$

Ahora derivamos la curva $\gamma_{k+1}(\theta)$ respecto a θ:

$$\gamma'_{k+1}(\theta)=\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{k+2}\right)}\right)^{k+1}\frac{\operatorname{sen}\left(\frac{\theta+\pi/2}{k+2}\right)}{\cos^2\left(\frac{\theta+\pi/2}{k+2}\right)}.$$

Finalmente, el valor de esta derivada en el punto P_{k+1} es

$$\gamma'_{k+1}(\alpha(k+2)-\pi/2)=\left(\frac{1}{2\cos\alpha}\right)^{k+1}\frac{\operatorname{sen}(\alpha)}{\cos^2(\alpha)}.$$

Como las dos derivadas en el punto P_{k+1} tienen el mismo valor, sabemos que la mediatriz es tangente a γ_{k+1}, lo que demuestra la tercera propiedad y concluye la demostración. ∎

Entonces, la curva γ_k tiene la ecuación polar siguiente:

$$\gamma_k(\theta)=2\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{k+1}\right)}\right)^{k+1}.$$

Observación 11. Es fácil ver que la curva polar γ_k tiene período $2\pi(k+1)$, ya que $\cos x$ tiene período 2π. El valor mínimo de $|\gamma_k(\theta)|$ es adoptado cuando[5]

$$\cos\left(\frac{\theta+\pi/2}{k+1}\right)=\pm1 \implies \theta=-\frac{\pi}{2}+n\pi(k+1),\quad n\in\mathbb{Z},$$

y la curva no está definida para

$$\cos\left(\frac{\theta+\pi/2}{k+1}\right)=0 \implies \theta=-\frac{\pi}{2}+\left(n+\frac{1}{2}\right)\pi(k+1),\quad n\in\mathbb{Z}.$$ ◂

[5]El punto en coordenadas polares (θ,r) para $r<0$ se define como el punto simétrico a $(\theta,-r)$ respeto el origen de coordenadas.

Definición 12. Llamaremos *ala negativa* de γ_k a la curva

$$\gamma_k(\theta), \quad \theta \in \left(-\frac{\pi}{2} + \left(n - \frac{1}{2}\right)\pi(k+1),\ -\frac{\pi}{2} + n\pi(k+1)\right]$$

para todo $n \in \mathbb{Z}$.

Llamaremos *ala positiva* de γ_k a la curva

$$\gamma_k(\theta), \quad \theta \in \left[-\frac{\pi}{2} + n\pi(k+1),\ -\frac{\pi}{2} + \left(n + \frac{1}{2}\right)\pi(k+1)\right)$$

para todo $n \in \mathbb{Z}$. ◂

Se deja como ejercicio para el lector ver que cualquier punto de γ_k pertenece a una de las dos alas.

3.1. La curva γ_k como espiral sinusoidal

Una espiral sinusoidal es una curva de la forma

$$r^n = a^n \cos(n\theta),$$

donde $na \neq 0$ con n racional y a real. Si escogemos $n = -1/(m+1)$, $a = 1/2^m$ y $\theta = \alpha + \pi/2$, con m natural, tenemos la siguiente curva:

$$r = 2\left(\frac{1}{2\cos\left(\frac{\alpha+\pi/2}{m+1}\right)}\right)^{m+1}.$$

Entonces, podemos concluir que las curvas γ_k forman parte de la familia de las espirales sinusoidales, en concreto las espirales en que $n = -1/m$ con m natural.

3.2. Representación gráfica de γ_k

Ahora que hemos demostrado la ecuación polar de la curva γ_k podemos representarla en el plano cartesiano, como veremos en la figura 6.

Esta figura está hecha con la calculadora gráfica Desmos [2]. Esta calculadora ha sido muy útil para enunciar las propiedades de la curva γ_k, sobre todo para las siguientes secciones del artículo. En el enlace de la bibliografía se puede ver la representación de la curva γ_k para cualquier k y una ayuda visual para la sección 3.4.

3.3. Intersecciones de γ_k con el círculo $r(\theta) = 2$

Observando la figura 6, vemos que las curvas γ_k cruzan el círculo $r(\theta) = 2$ en unos ángulos concretos. El teorema 13, además de ser interesante como propiedad, será útil en la sección 3.4.

Teorema 13. *Hay seis posibles puntos de intersección de γ_k con la circunferencia centrada en el origen de radio* 2.

Demostración. Para demostrar este teorema determinaremos los ángulos θ tales que $\gamma_k(\theta) = \pm 2$ para cada valor de k. Aislaremos θ de la ecuación siguiente:

$$2\left(\frac{1}{2\cos\left(\frac{\theta+\pi/2}{k+1}\right)}\right)^{k+1} = \pm 2.$$

La expresión anterior es equivalente[6] a

$$\cos\left(\frac{\theta + \pi/2}{k+1}\right) = \pm\frac{1}{2}.$$

[6] Esto se puede comprobar fácilmente haciendo un análisis por casos.

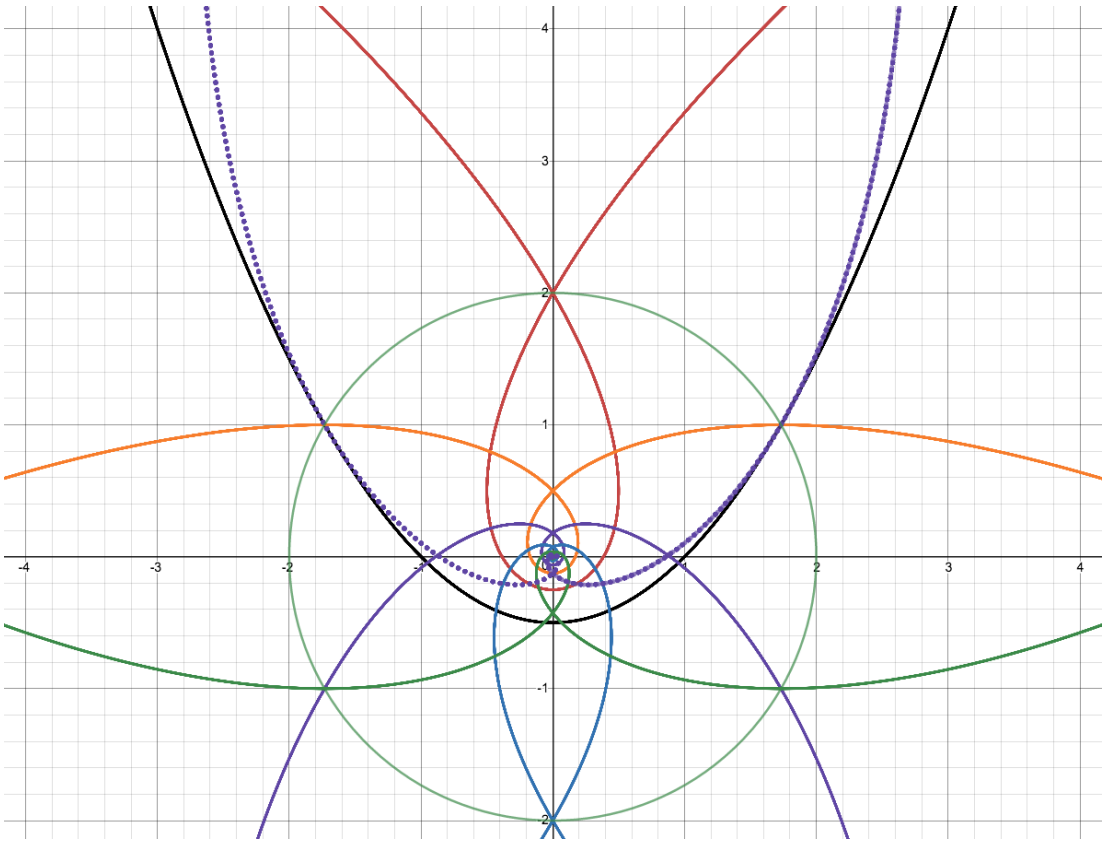


Figura 6: Podemos ver γ_1 en negro, γ_2 en rojo, γ_3 en naranja, γ_4 en violeta, γ_5 en azul, γ_6 en verde oscuro y γ_7 en violeta punteado.

Ahora consideramos dos casos:

- $\frac{\theta+\pi/2}{k+1} = \pi n + \frac{\pi}{3}$ para cualquier $n \in \mathbb{Z}$.
  Si aislamos θ tenemos que
  $$\theta = -\frac{\pi}{2} + \pi(k+1)\left(n + \frac{1}{3}\right).$$
  Como k y n son enteros, concluimos que los ángulos resultantes son de la forma $\pi\ell/3 - \pi/2$ para valores de ℓ enteros.
  Los puntos definidos por estos ángulos pertenecen a el ala positiva de γ_k, ya que
  $$-\frac{\pi}{2} + \pi(k+1)\left(n + \frac{1}{3}\right) \in \left[-\frac{\pi}{2} + n\pi(k+1),\ -\frac{\pi}{2} + \left(n + \frac{1}{2}\right)\pi(k+1)\right).$$
- $\frac{\theta+\pi/2}{k+1} = \pi n - \frac{\pi}{3}$ para cualquier $n \in \mathbb{Z}$.
  Hacemos lo mismo en este caso:
  $$\theta = -\frac{\pi}{2} + \pi(k+1)\left(n - \frac{1}{3}\right).$$
  Igual que en el caso anterior, los valores de los ángulos que buscamos son de la forma $\theta = \pi\ell/3 - \pi/2$ para ℓ entero.
  Y, análogamente, los puntos definidos por estos ángulos pertenecen a el ala negativa de γ_k.

Concluimos que solo hay seis ángulos para los que la curva γ_k puede intersecar la circunferencia de radio 2, y estos son
$$\frac{\pi}{2},\quad \frac{5\pi}{6},\quad \frac{7\pi}{6},\quad \frac{3\pi}{2},\quad \frac{11\pi}{6} \quad \text{y} \quad \frac{\pi}{6}.$$
■

3.4. γ_k cuando k tiende a infinito

Explorando la curva γ_k para diferentes valores de k en la web Desmos [2], es fácil darse cuenta que, para valores de k suficientemente grandes, las alas de γ_k forman espirales. En esta sección veremos cuál es la espiral a la que cada ala de γ_k se aproxima cuando k tiende a infinito.

Como dice la intuición y puede ser demostrado fácilmente, las dos alas de la curva γ_k son simétricas respeto el eje vertical. Entonces, sin pérdida de generalidad, estudiaremos el ala positiva de γ_k.

Para k muy grande, vemos que los valores que adopta $\gamma_k(\theta)$ son arbitrariamente pequeños alrededor de $\theta = 0$. Para resolver este problema y poder trabajar mejor con el ala positiva de γ_k para cualquier k, definiremos una rotación[7] $\bar{\gamma}_k$ de γ_k, definida con la relación $\bar{\gamma}_k(\theta) = \gamma_k(\theta + \pi(k+1)/3 - \pi/2)$, tal que la intersección de el ala positiva de γ_k con la circunferencia $r(\theta) = 2$ se encuentre en el eje de abscisas:

$$\bar{\gamma}_k(\theta) = 2\left(\frac{1}{2\cos\left(\frac{\pi}{3} + \frac{\theta}{k+1}\right)}\right)^{k+1}.$$

La curva $\bar{\gamma}_k$ es similar a la curva γ_k ya que es una rotación de ella, y podemos ver que $\bar{\gamma}_k(0) = 2$ para todo k.

Finalmente, podemos calcular el límite cuando k tiende a infinito,

$$\lim_{k\to\infty} \bar{\gamma}_k(\theta) = \lim_{k\to\infty} 2\left(\frac{1}{2\cos\left(\frac{\pi}{3} + \frac{\theta}{k}\right)}\right)^{k}.$$

Usando que $\cos(a+b) = \cos a \cos b - \operatorname{sen} a \operatorname{sen} b$, tenemos que

$$\lim_{k\to\infty} \bar{\gamma}_k(\theta) = 2\lim_{k\to\infty}\left(\frac{1}{\cos\frac{\theta}{k} - \sqrt{3}\operatorname{sen}\frac{\theta}{k}}\right)^{k}.$$

Es conocido que, si $\lim_{x\to a} f(x) = 1$ y $\lim_{x\to a} g(x) = \infty$, entonces $\lim_{x\to a} f(x)^{g(x)} = \lim_{x\to a} \mathrm{e}^{(f(x)-1)g(x)}$. Usando esta propiedad,

$$\lim_{k\to\infty} \bar{\gamma}_k(\theta) = 2\lim_{k\to\infty} \mathrm{e}^{\left(\frac{1}{\cos(\theta/k)-\sqrt{3}\operatorname{sen}(\theta/k)}-1\right)k} = 2\mathrm{e}^{\lim_{k\to\infty}\left(\frac{1}{\cos(\theta/k)-\sqrt{3}\operatorname{sen}(\theta/k)}-1\right)k}.$$

Trabajando con el límite del exponente y simplificando, tenemos que

$$\lim_{k\to\infty}\left(\frac{1}{\cos\frac{\theta}{k} - \sqrt{3}\operatorname{sen}\frac{\theta}{k}} - 1\right)k = \lim_{k\to\infty} k\left(1 - \cos\frac{\theta}{k}\right) + \lim_{k\to\infty} k\left(\sqrt{3}\operatorname{sen}\frac{\theta}{k}\right) = 0 + \sqrt{3}\theta$$

Entonces, podemos concluir que

$$\lim_{k\to\infty} \bar{\gamma}_k(\theta) = 2\mathrm{e}^{\sqrt{3}\theta}.$$

La curva en forma polar $r(\theta) = a\mathrm{e}^{b\theta}$ con a, b reales diferentes de 0 es una espiral logarítmica. Por lo tanto, cuando k tiende a infinito, las dos alas de la curva γ_k se aproximan a dos espirales logarítmicas de la forma $r(\theta) = \pm 2\mathrm{e}^{\sqrt{3}(\theta-\alpha)}$, para algún α real.

[7] En coordenadas polares, el punto (r, θ) representa una rotación respeto el origen de coordenadas del punto $(r, \theta + \alpha)$.

Referencias

[1] Burton, David M. *The History of Mathematics: An Introduction*. 7.ª ed. Nueva York: McGraw-Hill, 2011. ISBN: 978-0-07-338315-6.

[2] *Desmos, Calculadora gráfica*. URL: https://www.desmos.com/calculator/aorz8y1ldc.

[3] Ferréol, Robert. *Tschirnhausen's cubic*. En: *Encyclopédie des formes mathématiques remarquables*. 2017. URL: https://www.mathcurve.com/courbes2d.gb/tschirnhausen/tschirnhausen.shtml.

[4] Lawrence, J. Dennis. *A Catalog of Special Plane Curves*. Mineola, Nueva York: Dover Publications, 1972. ISBN: 978-0-486-60288-2.

[5] Redon, Pau. *La corba gamma*. Trabajo de Investigación de Bachillerato. 2019. URL: https://fme.upc.edu/ca/premi-poincare/edicions-anteriors/Premi-poincare-2019/arxius/poincare-2019-memoria-poincare_077-la_corba_gamma.pdf.

[6] Weisstein, Eric W. *Sinusoidal spirals*. En: *MathWorld*. A Wolfram Web Resource. URL: http://mathworld.wolfram.com/SinusoidalSpiral.html.

TEMat

Este trabajo fue galardonado con el segundo premio en la edición de 2017 del Premi Poincaré, entregado por la Facultat de Matemàtiques i Estadística de la Universitat Politècnica de Catalunya.

Estudio del origen del número e y de sus aplicaciones en diversos campos de las matemáticas

Pablo Nicolás Martínez
Universitat Politècnica de Catalunya (UPC)
pnico.martinez@gmail.com

Resumen: En este artículo se analiza el origen del número e a lo largo de la historia mediante el estudio de los escritos originales de autores que, de una forma u otra, intervienen en el origen y contribuyen a la definición de e. Se analizan también algunas de sus aplicaciones en fórmulas relevantes mediante el estudio de los textos originales correspondientes a su primera aparición.

Abstract: In this paper, we analize the origin of number e throughout history. We do so through the study of the original texts of authors who, one way or another, contributed to the origin and definition of e. Furthermore, we also present some of its applications in relevant formulae through the study of the original texts corresponding to its first appearance.

Palabras clave: origen de e, historia de las matemáticas, definición de e.

MSC2010: 01A99, 97A30.

Recibido: 11 de marzo de 2018.
Aceptado: 27 de agosto de 2019.

Agradecimientos: Me gustaría agradecer y destacar la participación e implicación de mi tutora de bachillerato, M.ª Trinidad Cámara Meseguer, quien me ha proporcionado todas las ayudas posibles para poder realizar este proyecto.

Quiero agradecer también la colaboración de D. Víctor Jiménez López, D. José Ginés Espín Buendía, D.ª Carmen Noemí Zoroa Alonso, D. Alfredo Marín Pérez y D. Jaime Colchero Paetz por contestar nuestras preguntas sobre las aplicaciones del número e en sus respectivos campos de trabajo.

Por último, quiero agradecer a Juan Carlos Ferre Ruiz su ayuda corrigiendo nuestras traducciones de los textos del latín.

Referencia: Nicolás Martínez, Pablo. «Estudio del origen del número e y de sus aplicaciones en diversos campos de las matemáticas». En: *TEMat*, 4 (2020), págs. 15-26. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p15.

1. Introducción

Existen determinadas constantes en matemáticas que, por su relevancia o utilidad, reciben un nombre particular que las distingue. Ejemplos son el caso de π, que representa el cociente entre la longitud de una circunferencia y su diámetro, o el número áureo φ. Mientras que las definiciones de constantes elementales como π o φ son relativamente intuitivas, el caso de e es diferente, siendo necesarios conceptos de cálculo infinitesimal. Por tanto, es natural preguntarse cómo surge de manera histórica el número e hasta llegar a la definición conocida actualmente, así como el por qué de su utilidad. Tomando esto como motivación, nos proponemos los siguientes objetivos [16]:

- Encontrar el proceso que se sigue a lo largo de la historia hasta definir e.
- Investigar sus principales aplicaciones en distintas ramas de las matemáticas, así como en otras áreas relacionadas.

Para comenzar a trabajar en la historia, partimos de las siguientes hipótesis en base a conocimientos anteriores. Uno de los primeros autores que podría haber utilizado e puede ser Napier, ya que los logaritmos conocidos como neperianos tienen base e. También pensamos que Euler está posiblemente involucrado en su definición, ya que corrientemente e es conocido como el «número de Euler». Finalmente, como e aparece de forma natural en problemas de derivadas e integrales, consideramos a Leibniz como otro de los autores a investigar, por ser uno de los padres del cálculo infinitesimal. Mientras trabajábamos en estos tres autores, fuimos encontrando otros matemáticos que también habían participado de una forma más o menos directa en la definición de e. En esta búsqueda resultaron de gran ayuda algunos textos sobre historia de la matemática, especialmente autores como Cajori [3] o Ivory [9].

Para elegir las aplicaciones de e a estudiar, hemos preguntado a siete profesores de la Universidad de Murcia por la aplicación más importante de e en sus respectivas áreas de trabajo. Hemos recibido respuesta de dos profesores del área de análisis matemático, dos del Departamento de Estadística e Investigación Operativa y uno del Departamento de Física. Los dos profesores de las áreas de álgebra y geometría a los que preguntamos no contestaron.

En ambas partes, la metodología utilizada ha sido la misma. Una vez seleccionado el autor o aplicación, nos hemos remitido a la publicación original en la que aparece de alguna forma e, hemos seleccionado las partes de estos textos necesarias para comprender la demostración que hace el autor, hemos realizado las traducciones del idioma original (generalmente, latín, inglés y francés) al castellano y, en base a estas traducciones, hemos convertido las demostraciones originales a notación moderna.

2. Historia del número e

2.1. John Napier y la primera definición del logaritmo

El primer autor que hemos estudiado ha sido John Napier, conocido por ser la primera persona en definir el logaritmo.

Para realizar este estudio, nos hemos basado en las definiciones que da en su obra *Mirifici Logarithmorum Canonis Descriptio* [15]. Para definir el logaritmo, Napier introduce en esta obra dos tipos de movimientos, que podemos ver en la figura 1.

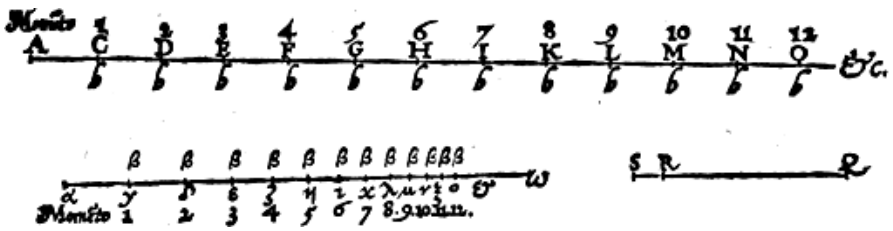


Figura 1: Imagen de los movimientos que introduce Napier [15, pág. 4].

El primero de ellos es un movimiento con velocidad constante, mientras que, en el segundo movimiento, la velocidad es proporcional a la distancia que falta por recorrer. Para el primer movimiento, Napier muestra cómo construir la progresión que define la distancia recorrida en un instante a partir de la distancia recorrida en el instante anterior. Esta progresión se puede expresar por la relación de recurrencia

$$x_{n+1} = d + x_n,$$

donde d es la distancia entre dos términos contiguos y $n \in \mathbb{N}$ (a lo largo de todo el artículo se asumirá que $0 \in \mathbb{N}$). Como $x_0 = 0$, por inducción se demuestra que

$$x_n = n \cdot d. \tag{1}$$

En el segundo movimiento, la progresión indica el espacio que falta por recorrer en un instante en función del que faltaba por recorrer en el instante anterior. La relación de recurrencia en este caso es

$$y_{n+1} = (1-k)\, y_n, \tag{2}$$

donde k es la razón de decrecimiento y $n \in \mathbb{N}$. Napier impone que la distancia original del segmento es la longitud total del segmento, que tiene por extremos los puntos α y ω (ver figura 1). Se tiene entonces la condición $y_0 = [\alpha\omega]$, donde $[\alpha\omega]$ denota la longitud del segmento que une α y ω. Por inducción, es fácil demostrar que

$$y_n = [\alpha\omega]\,(1-k)^n. \tag{3}$$

Para llegar a la definición de logaritmo, además de definir estos dos movimientos, Napier obliga a que ambos tengan la misma velocidad inicial. Como estamos trabajando con progresiones y no con funciones continuas, hemos obtenido esta velocidad inicial como la variación del espacio recorrido entre $n = 1$ y $n = 0$. Obtenemos que, para el primer movimiento, esta velocidad es d y, para el segundo, es $[\alpha\omega] \cdot k$. Sustituyendo en (1) y en (3), obtenemos

$$x_n = [\alpha\omega] \cdot k \cdot n, \qquad\qquad y_n = [\alpha\omega]\,(1-k)^n,$$

donde $n \in \mathbb{N}$. Napier define el logaritmo de una cantidad h, que denotaremos de ahora en adelante por $\operatorname{Nap\,log}(h)$, como la distancia recorrida en el primer movimiento (es decir, x_n) cuando en el segundo movimiento (respectivamente y_n) resta por recorrer una distancia h. Esta definición puede ser expresada en notación moderna mediante el sistema de ecuaciones

$$\begin{cases} h = y_n, \\ \operatorname{Nap\,log}(h) = x_n; \end{cases} \iff \begin{cases} h = [\alpha\omega]\,(1-k)^n, \\ \operatorname{Nap\,log}(h) = [\alpha\omega] \cdot k \cdot n. \end{cases}$$

Despejando n en la primera ecuación y sustituyendo en la segunda, llegamos a la expresión de $\operatorname{Nap\,log}(h)$ en función de la definición actual de logaritmo,

$$\operatorname{Nap\,log}(h) = [\alpha\omega]\, \log_{(1-k)^{1/k}}\left(\frac{h}{[\alpha\omega]}\right). \tag{4}$$

Un dato interesante es que en la base del logaritmo solamente influye la k escogida. En esta obra, Napier no hace referencia al valor de k que usa para la construcción de sus tablas. Para encontrar este dato hay que avanzar hasta su obra *Mirifici Logarithmorun Canonis Constructio*, donde fija $k = 10^{-7}$ [14, pág. 12]. Esta elección aparentemente extraña hay que entenderla sabiendo que, en sus orígenes, el logaritmo no era una operación, sino una herramienta. El objetivo principal es la construcción de tablas, y el método para hacerlo viene dado por la relación de recurrencia (2), donde para obtener el siguiente término de la sucesión es necesario multiplicar el término actual por la razón de decrecimiento y restarlo. La multiplicación por 10^{-7} (y, en general, por 10^n con $n \in \mathbb{Z}$) es muy sencilla de realizar y reduce los cálculos para la construcción de tablas[1].

[1] La elección en concreto del valor 10^7 se debe a los trabajos de Regiomontano. El motivo por el que utiliza esta cifra es para tener una exactitud en los cálculos que hiciera hasta la séptima cifra, ya que no se conocían todavía las fracciones decimales [11]. Precisamente por los trabajos de Regiomontano, Napier escoge como razón 10^7 [17], aunque en otras tablas calculadas posteriormente en el *Mirifici Logarithmorun Canonis Constructio* usa como razones 10^5 y $5 \cdot 10^4$.

Aplicando este valor a la base del logaritmo que aparece en (4) encontramos un valor dado por

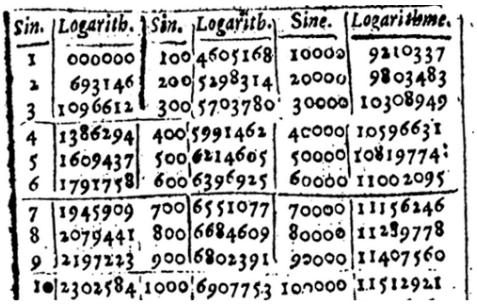
$$\left(1-10^{-7}\right)^{10^7}=\left(1-\frac{1}{10^7}\right)^{10^7}\simeq\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n=\mathrm{e}^{-1},$$

que está directamente relacionado con e. Es por esto que podemos considerar a Napier como el primero que trabaja con un valor aproximado de esta constante.

2.2. Elaboradores de tablas

Tras la primera definición de los logaritmos que da Napier, son muchos los matemáticos que añaden modificaciones al método de Napier para mejorar algunos inconvenientes que presentan sus tablas. El principal inconveniente que presentan las tablas de Napier es que sus logaritmos son negativos para valores de x superiores a 10^7, cosa que podemos observar de la expresión analítica del logaritmo en la ecuación (4) con $[\alpha\omega]=10^7$.

La primera tabla que hemos estudiado y que aporta una mejora significativa fue introducida en el anexo de la segunda edición de la traducción al inglés del *Mirifici Logarithmorun Canonis Descriptio* [13], que se atribuye al matemático William Oughtred. Antes de continuar, diremos que en ninguna de las tablas que vamos a citar en este apartado viene expuesto un procedimiento que indique cómo se obtuvieron, así que, para deducir las funciones que dan los valores de dichas tablas, nos hemos basado en comprobaciones numéricas y afirmaciones de otros autores como Cajori [3] o Ivory [9]. Así, en las tablas de Oughtred (figura 2, izquierda) hemos comprobado que los valores se aproximan a $10^6\ln(x)$ (figura 2, derecha).



x	$10^6\ln(x)$	x	$10^6\ln(x)$
1	0	6	1 791 759
2	693 147	7	1 945 910
3	1 098 612	8	2 079 441
4	1 386 294	9	2 197 224
5	1 609 437	10	2 302 585

Figura 2: Extracto de la tablas atribuidas a W. Oughtred [13, pág. 12] (izquierda) y tabla con el valor de $10^6\ln(x)$ hasta $x=10$ (derecha). Los valores se han obtenido truncando la primera cifra decimal de cada número.

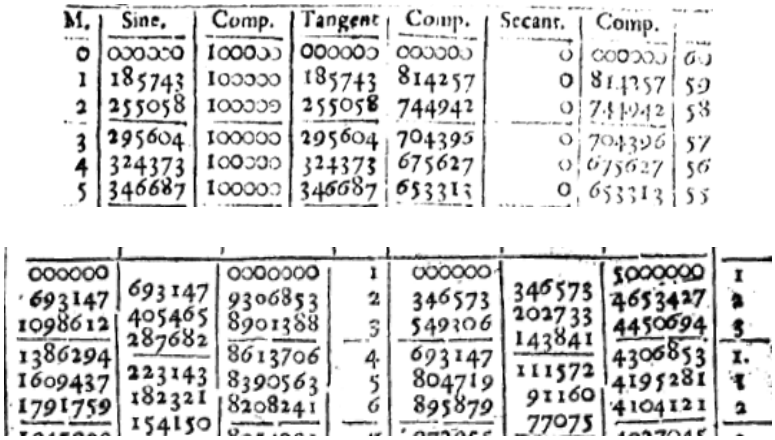
El segundo autor que hemos estudiado ha sido John Speidell, quien elaboró distintas versiones de tablas de logaritmos. Las primeras tablas que elabora y que encontramos en *New Logarithmes* [19] son una modificación de las de Napier para hacer que los logaritmos tomen siempre valores positivos. Tomando como cierta la afirmación de Ivory [9, págs. 714-715], hemos obtenido que los logaritmos de Speidell (que denominaremos $\mathrm{Sp}\log(x)$) cumplen que

$$10^8=\mathrm{Nap}\log(x)+100\ \mathrm{Sp}\log(x).$$

De la expresión del logaritmo de Napier (4) podemos obtener el valor del logaritmo de Speidell como

$$\mathrm{Sp}\log(x)=10^5\ln\left(\frac{x}{454}\right).$$

A partir de 1622, Speidell incluyó en las nuevas ediciones del *New Logarithmes* [20] otra tabla en la que calcula los logaritmos de los números enteros entre 1 y 1000. Mediante otra comprobación, hemos visto que los logaritmos de esta segunda tabla corresponden con los de $10^6\ln(x)$.



M.	Sine.	Comp.	Tangent	Comp.	Secant.	Comp.	
0	000000	100000	000000	000000	0	000000	60
1	185743	100000	185743	814257	0	814357	59
2	255058	100000	255058	744942	0	744942	58
3	295604	100000	295604	704395	0	704396	57
4	324373	100000	324373	675627	0	675627	56
5	346687	100000	346687	653313	0	653313	55

000000		0000000	1	000000		5000000	1
693147	693147	9306853	2	346573	346573	4653427	2
1098612	405465	8901388	3	549306	202733	4450694	3
1386294	287682	8613706	4	693147	143841	4306853	4
1609437	223143	8390563	5	804719	111572	4195281	5
1791759	182321	8208241	6	895879	91160	4104121	6
	154150				77075		

Figura 3: La primera figura (arriba) muestra recortes de las primeras tablas que presenta Speidell, mientras que la segunda (abajo) corresponde a las segundas tablas.

2.3. Leibniz y e en la integral de la función hiperbólica

Una de las apariciones más naturales de e es en los problemas relacionados con el cálculo integral y diferencial. Por ello, una figura que podría resultar interesante estudiar en nuestra búsqueda del origen de e es Gottfried Leibniz, padre del cálculo integral junto a Isaac Newton.

En el análisis del siglo XVII destaca el afán de calcular el área bajo las curvas. Este problema se conoce tradicionalmente como el de «cuadrar» la curva. En concreto, el área bajo la hipérbola $f(x) = k/x$ fue estudiada por matemáticos como De Sarasa, Huygens, Newton, Mercator, James Gregory y Leibniz. Desde el trabajo *Solutio problematis a R. P. Marino Mersenno Minimo propositi*, de Alphonse Antoine de Sarasa [18], ya se sabía que había que escribir el área en términos de un logaritmo. Pero, aunque todos estos matemáticos encontraron métodos correctos (principalmente mediante series), la primera vez que se referencia la base de dicho logaritmo es en una serie de cartas [8] entre Huygens y Leibniz sobre un problema que planteó el último en las *Acta Eroditorum* [10]. El problema es, a grandes rasgos, hallar las ecuaciones que rigen el movimiento de un cuerpo en caída libre que sufre una resistencia cuadrática causada por rozamiento con aire. A lo largo de la sección denotaremos por t el tiempo de caída, v la velocidad instantánea del cuerpo y a la velocidad límite de caída. Asumiendo que $t_0 = 0$ y que el cuerpo comienza a caer en reposo, por lo que $v_0 = 0$, Leibniz llega a plantear la integral

$$t = \int_0^v \frac{a}{a^2 - v'^2}\,\mathrm{d}v'. \tag{5}$$

Huygens es el primero que da un resultado, utilizando series, pero se equivoca[2], y es Leibniz el que da los resultados correctos en la contestación, también en forma de series. Estos resultados dados por Leibniz son equivalentes a las formas habituales de hoy en día, que son

$$t = \int_0^v \frac{a}{a^2 - v'^2}\,\mathrm{d}v' = \frac{a}{2}\ln\left(\frac{a+v}{a-v}\right). \tag{6}$$

En otra carta sobre la naturaleza de la función logarítmica y su relación con la función exponencial, Leibniz

[2] En la correspondencia se plantea, además de la expresión (5) para hallar el tiempo t, una expresión semejante para hallar el espacio recorrido s. El error que comete Huygens es cambiar los resultados de las dos integrales planteadas.

hace referencia a la primera integral en (6) para el caso $a = 1$ y escribe la expresión[3]

$$b^{\frac{t}{\cdot}} = \frac{1+v}{1-v},$$

que involucra una constante b. Respecto a este valor Leibniz escribe: «b siendo una gran constante cuyo logaritmo es 1, siendo 0 el logaritmo de 1» [8, pág. 76]. Si hacemos los mismos cálculos pero usando métodos actuales, obtenemos que

$$\mathrm{e}^{\frac{t}{1/2}} = \frac{1+v}{1-v}.$$

Comparando ambos resultados y la afirmación dada por Leibniz, tenemos la creencia de que Leibniz conoce e pero lo denomina b. Otros autores, como Cajori [2, pág. 493], también toman b como la base de los logaritmos naturales o neperianos.

2.4. Jacob Bernoulli y el problema del interés compuesto

Hasta este momento, hemos visto la aparición de logaritmos en cuya base aparece una constante relacionada con e, e incluso la presencia de una constante llamada b de la que tenemos indicios de que sea precisamente e. Aún así, en ningún momento se ha hecho referencia al valor de esta constante. La primera vez que se obtiene algo de información sobre el valor de e es de forma casual mediante acotaciones de su valor. Sorprendentemente, esta acotación no se presenta en ningún tratado de logaritmos, área bajo la hipérbola o similares, sino en el artículo «Quæstiones Nonnullæ de usuris, cum solutione Problematis de Sorte Alearum, propositi in Ephem. Gall. A. 1685» de Jacob Bernoulli [1], en el estudio de un problema de interés compuesto. Dada una cantidad inicial invertida a rédito compuesto anual, el problema consiste en calcular el dinero que se obtendría si se cobrasen los intereses en periodos más cortos y, en última instancia, momentáneamente. Sea supuesto que se invierte un euro con un interés anual del 100 %. Si el interés se cobra anualmente, al final del año se tendrán 2 euros. Si, en cambio, se cobran los intereses cada 6 meses con un interés proporcional del 50 % se obtienen 2,25 euros. El proceso de cobrar los intereses momentáneamente se obtiene mediante un paso al límite, siguiendo la definición de derivada o tasa de valor instantáneo.

Bernoulli obtiene la solución dada por una serie en la que a es el capital inicial y b el rédito anual. Además, calcula una cota inferior y otra superior de esta solución:

$$a + b + \frac{b^2}{2a} < a\sum_{n=0}^{\infty} \frac{b^n}{a^n\, n!} < a + b + \frac{b^2}{2a-b}.$$

Mostramos la forma de obtener ambas cotas.

1. Respecto a la primera, la suma $\sum_{i=0}^{n} \frac{b^i}{a^i\, i!}$ es de términos estrictamente positivos, por lo que la sucesión de sumas parciales $s_n = a\sum_{i=0}^{n} \frac{b^i}{a^i\, i!}$ es estrictamente creciente. Por tanto, $s_2 < \lim_{n\to\infty} s_n$. Se tiene entonces que
$$a + b + \frac{b^2}{2a} = s_2 < \lim_{n\to\infty} s_n = a\sum_{i=0}^{\infty} \frac{b^i}{a^i\, i!}.$$

2. Para la segunda cota, sea notado que $2^{n-1} \leqslant n!$ siempre que $n \geqslant 2$, y la desigualdad es estricta si $n \geqslant 3$. Aplicando esta relación a las sumas parciales y llevándolas al límite[4],
$$a\sum_{n=0}^{\infty} \frac{b^n}{a^n\, n!} < a + b + 2a\sum_{n=2}^{\infty} \frac{b^n}{2^n\, a^n} = a + b + \frac{\frac{b^2}{2a}}{1-\frac{b}{2a}} = a + b + \frac{b^2}{2a-b}.$$

[3] Hemos reproducido la expresión exactamente como la hemos encontrado en el documento consultado [8, pág. 283]. Definitivamente, se puede observar el punto en la expresión original. La exponencial en una base b aparece anteriormente en otra carta [8, pág. 276] y presenta el mismo punto. Sin embargo, el punto no está presente en todas las funciones exponenciales que escribe; por ejemplo, también escribe la ecuación $x^x + x = 30$, en la que no se observa [8, pág. 276]. Esto puede llevar a pensar que la base b es especial y, por tanto, siempre que se usa de base para una función exponencial lleva un punto debajo del argumento, o que dicho punto tiene otro tipo de significado independiente de b. El significado se deja a interpretación del lector.

[4] Para asegurar la convergencia de la serie para la primera igualdad, es necesario tener que $\left|\frac{b}{2a}\right| < 1$. Sin embargo, Bernoulli no remarca esta condición en su artículo [1] y asume directamente la convergencia.

En base a esta expresión, Bernoulli afirma que, si $a = b$, entonces el valor del capital estaría entre $2{,}5a$ y $3a$, lo que se puede comprobar sustituyendo en las cotas. Como a es un valor positivo, entonces la equivalencia

$$2{,}5a < a\sum_{n=0}^{\infty}\frac{1}{n!} < 3a \iff 2{,}5 < \sum_{n=0}^{\infty}\frac{1}{n!} < 3$$

es cierta. Sabiendo con nuestros conocimientos actuales que $\mathrm{e} = \sum_{n=0}^{\infty}\frac{1}{n!}$, este resultado prueba que $2{,}5 < \mathrm{e} < 3$. Sin embargo, Bernoulli no conoce la existencia de e ni de una constante parecida, por lo que, como decíamos, parece que acotar su valor ha sido más un golpe de casualidad por el problema en el que trabajaba que algo intencionado.

2.5. Leonhard Euler y el desarrollo en serie de la función exponencial

A pesar de las apariciones ya comentadas, todavía nadie ha definido la constante e. Según Cajori [2, pág. 493], el primer autor que lo hace es Leonhard Euler. Si nos referimos estrictamente al nombre «e», Euler ya lo utiliza en sus primeros artículos, como en «Meditatio in experimenta explosione tormentorum nuper instituta» [4], llegando incluso a dar un valor aproximado. En este caso, la definición que utiliza es la misma que la que dio Leibniz. Pero la primera vez que define e tal y como aparece en la expresión dada por Bernoulli es en su *Introductio in Analysin Infinitorum* [5]. En el séptimo capítulo de esta obra, Euler indaga en el desarrollo de funciones exponenciales como series infinitas. A lo largo de todo el texto, emplea de forma velada la aproximación lineal de funciones derivables en un punto. En la época de Euler, el cálculo infinitesimal moderno (es decir, el desarrollado a partir de la definición de límite) no existía. Por tanto, conceptos como la derivabilidad no estaban claros del todo, y lo expuesto a continuación debe ser entendido como un intento de formalizar la demostración original.

Euler escribe que, si ω es arbitrariamente pequeño, lo que quiere decir que $\omega \to 0$, entonces se tiene que $a^{\omega} = 1 + k\,\omega$, siendo k una constante dependiente de la base a. Con nuestro conocimiento moderno, esta afirmación es equivalente a formular la aproximación lineal de la función a^x en un entorno del punto $x = 0$, por lo que se tiene que

$$a^{\omega} = a^0 + f'(0)(\omega - 0) + R_1(f) = 1 + k\,\omega + R_1(a^x),$$

siendo $f(x) = a^x$ y $R_1(a^x) = o(\omega)$ el resto de orden 1 de la aproximación lineal. Se puede observar que $k = \left(a^x\right)'(0) = \ln a$, resultado que Euler obtiene posteriormente en el libro.

Euler introduce ahora las nuevas variables z, $i \in \mathbb{R}$,[5] relacionadas con ω mediante $z = \omega \cdot i$. A partir de ahora, Euler considera, sin hacer mención explícita de ello, la cantidad z como un número real fijo. Como $\omega \to 0$, concluye entonces que $i \to \infty$. Sustituyendo esto en la primera igualdad llega a que

$$a^{\frac{z}{i}} = 1 + k\,\frac{z}{i} \iff a^z = \left(1 + k\,\frac{z}{i}\right)^i. \tag{7}$$

Mediante el binomio de Newton y la aplicación de los actuales límites del tipo $\frac{\infty}{\infty}$, llega a que toda función exponencial es de la forma

$$a^z = \sum_{n=0}^{\infty}\frac{k^n z^n}{n!}.$$

Es capaz de demostrar también que, sabiendo la k asociada a una base a, se puede hallar el desarrollo en serie de cualquier función exponencial con otra base b. En la demostración original iguala las dos funciones exponenciales $a^z = b^y$, despeja z y la sustituye en la serie de a^z:

$$b^y = a^{y\cdot\log_a b} = \sum_{n=0}^{\infty}\frac{k^n y^n (\log_a b)^n}{n!}.$$

[5] Aprovechamos para recordar que, en la época de Euler, el concepto de número complejo tampoco estaba desarrollado completamente. Por tanto, la constante i no representaba ningún número particular y podía ser utilizada como una variable más sin lugar a confusión (la elección de i para representar un número tal que $i \to +\infty$ parece estar motivada por la palabra « *infinitorum*», que es la traducción al latín de «infinito»). De hecho, posteriormente Euler escribe $\sqrt{-1}$ en lugar de i. En este artículo se evitará la confusión entre la constante $\sqrt{-1}$ y la variable i tipográficamente, escribiendo i para denotar $\sqrt{-1}$.

Una vez concluido esto y tras otras demostraciones correspondientes a la función logarítmica, Euler introduce la definición de e. No menciona explícitamente el argumento que le lleva a hacerlo, por lo que presentamos aquí uno posible. Ya que sabiendo la k asociada a una base a podemos hallar el desarrollo en serie de cualquier función exponencial, entonces buscaremos la base con la k más simple, es decir, la base con $k = 1$. Esta elección provoca que dicha base sea

$$a^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \implies a = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

A esta base privilegiada la llama e, y da como valor aproximado e = 2,7182818284590452353602 8... Aplicando esta definición a la ecuación (7) obtenemos que

$$\mathrm{e}^z = \lim_{i\to\infty} \left(1 + \frac{z}{i}\right)^i. \tag{8}$$

Euler no lo hace, pero de aquí es fácil deducir otra definición de e al sustituir $z = 1$. Esta definición, que es la presentada usualmente en bachillerato, es

$$\mathrm{e} = \lim_{i\to\infty} \left(1 + \frac{1}{i}\right)^i.$$

3. Aplicaciones de e en ramas de la matemática

3.1. Identidad de Euler

La primera aplicación que hemos estudiado es la identidad de Euler, que relaciona las constantes conocidas actualmente como i, π y e mediante

$$\mathrm{e}^{\mathrm{i}\pi} + 1 = 0.$$

Como esta identidad es una consecuencia directa de la fórmula de Euler, hemos decidido analizar el origen de esta última. La encontramos en la obra de Leonhard Euler *Introductio in Analysin Infinitorum* [5] de 1748, donde se obtiene en el capítulo VIII tras manipular las funciones seno y coseno. Comienza utilizando la fórmula de De Moivre,

$$(\cos x + \mathrm{i}\,\mathrm{sen}\, x)^n = \cos(nx) + \mathrm{i}\,\mathrm{sen}(nx), \qquad n \in \mathbb{N},$$

que se puede demostrar por inducción. A partir de esta relación es evidente obtener que

$$\cos(nx) = \frac{(\cos x + \mathrm{i}\,\mathrm{sen}\, x)^n + (\cos x - \mathrm{i}\,\mathrm{sen}\, x)^n}{2}, \tag{9}$$

$$\mathrm{sen}(nx) = \frac{(\cos x + \mathrm{i}\,\mathrm{sen}\, x)^n - (\cos x - \mathrm{i}\,\mathrm{sen}\, x)^n}{2\mathrm{i}}. \tag{10}$$

Euler toma de nuevo $nx = v$ como una cantidad real fija y hace tender x a 0. Esto fuerza que $n \to \infty$. A continuación, aproxima las funciones seno y coseno por sus respectivas aproximaciones lineales, obteniendo $\cos x \simeq 1$ y $\mathrm{sen}\, x \simeq x = \frac{v}{n}$. Estas expresiones dan lugar por sustitución en las ecuaciones (9) y (10) a

$$\cos v = \frac{1}{2}\left[\left(1 + \frac{v\,\mathrm{i}}{n}\right)^n + \left(1 - \frac{v\,\mathrm{i}}{n}\right)^n\right], \qquad \mathrm{sen}\, v = \frac{1}{2\mathrm{i}}\left[\left(1 + \frac{v\,\mathrm{i}}{n}\right)^n - \left(1 - \frac{v\mathrm{i}}{n}\right)^n\right].$$

Como ha empleado que $n \to \infty$, por la ecuación (8) concluye las conocidas como identidades de Euler,

$$\cos x = \frac{\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x}}{2} \quad \text{y} \quad \mathrm{sen}\, x = \frac{\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}}{2\mathrm{i}}.$$

Euler emplea estas identidades para demostrar la fórmula de Euler. Se obtiene que

$$\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\,\mathrm{sen}\, x.$$

La identidad de Euler se obtiene para el caso $x = \pi$:

$$\mathrm{e}^{\mathrm{i}\pi} + 1 = 0.$$

3.2. Serie de Fourier

La serie de Fourier permite representar muchas funciones periódicas como combinación lineal de funciones exponenciales complejas.

Hemos encontrado la primera versión de la serie de Fourier en el artículo «Mémoire sur la propagation de la chaleur dans les corps solides» [7] presentado a la Societé Philomatique de Paris, en el que indaga sobre la propagación del calor. Tomando la temperatura T como una función escalar dependiente de los parámetros espaciales x, y, z y del tiempo t, es decir, $T(x, y, z, t)$, la ecuación de Fourier afirma que

$$\frac{\partial T}{\partial t} = a\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right), \tag{11}$$

donde a es una constante que, en principio, depende de las variables x, y, z, t.

Las condiciones del problema se restringen considerando un caso particular al que aplica esta ecuación. Supone que el medio de transmisión es homogéneo, es decir, la constante a es independiente de las variables x, y, z, t. Considera el problema en una placa situada en el plano xy de longitud 2 en el eje y e indefinidamente larga en el eje x, por lo que $x \in \mathbb{R}$, $y \in [-1, 1]$ y z es constante. Por último, se considera que el sistema está en estado estacionario, lo que quiere decir que T es independiente de t. Todo esto hace que la ecuación (11) pase a ser escrita como

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0,$$

que es conocida como ecuación de Laplace. Mediante el método de separación de variables, expresando la temperatura como $T(x, y) = M(x)N(y)$, hemos obtenido varias soluciones al problema, dadas por las ecuaciones diferenciales ordinarias

$$\frac{\partial^2 M}{\partial x^2} - \lambda M = 0, \qquad \frac{\partial^2 N}{\partial y^2} + \lambda N = 0.$$

Dichas soluciones con $\lambda \geqslant 0$ vienen dadas por

$$M(x) = c\,\mathrm{e}^{\sqrt{\lambda}x} + d\,\mathrm{e}^{-\sqrt{\lambda}x}, \qquad N(y) = C\cos\left(\sqrt{\lambda}x\right) + D\,\mathrm{sen}\left(\sqrt{\lambda}x\right).$$

Como la ecuación se satisface para cualquier elección de λ, se puede escoger una solución general como superposición de soluciones para λ_i, obteniendo la solución más general

$$T(x, y) = \sum_{i=1}^{\infty} \left(b_i \mathrm{e}^{\sqrt{\lambda_i}x} + a_i \mathrm{e}^{-\sqrt{\lambda_i}x}\right)\cos\left(\sqrt{\lambda_i}x\right) + \left(B_i \mathrm{e}^{\sqrt{\lambda_i}x} + A_i \mathrm{e}^{-\sqrt{\lambda_i}x}\right)\mathrm{sen}\left(\sqrt{\lambda_i}x\right).$$

Ahora impone simetría de la función respecto a la variable y, por lo que los términos senoidales deben desaparecer. Así, $A_i = B_i = 0$. Además, la temperatura debe ser una función acotada para toda la placa, por lo que $b_i = 0$. La solución general tras eliminar estas condiciones es

$$T(x, y) = \sum_{i=1}^{\infty} a_i\,\mathrm{e}^{-\sqrt{\lambda_i}\,x}\cos\left(\sqrt{\lambda_i}\,y\right),$$

siendo los valores λ_i con $i \in \mathbb{N}$ constantes de valores indeterminados.

Aplicando la condición de frontera $T(x, \pm 1) = 0$, que es equivalente a afirmar que la temperatura en el borde de la placa es la misma que en el exterior, obtiene los valores de las constantes

$$\sqrt{\lambda_i} = \frac{2i+1}{2}\,\pi.$$

Con esto ya queda totalmente determinado el perfil de temperaturas estacionario de la placa. El objetivo es estudiar ahora el perfil para diferentes valores de x. El que estudia, en particular, es para $x = 0$, obteniendo la función de temperatura

$$\varphi(y) = T(0, y) = \sum_{i=1}^{\infty} a_i\,\cos\left(\frac{2i+1}{2}\,\pi y\right),$$

que es lo que se conoce actualmente como serie de Fourier. Aplicando ahora la ortonormalidad[6] de las funciones $\left\{\cos\left(\frac{2i+1}{2}\pi y\right)\right\}_{i\in\mathbb{N}}$ en el intervalo $[-1,1]$ en el que está definida la función φ, se obtiene la expresión de los coeficientes

$$a_i = \int_{-1}^{1} \varphi(y)\,\cos\left(\frac{2i+1}{2}\,\pi y\right)\mathrm{d}y.$$

La serie de Fourier completa aparece en su *Théorie analytique de la chaleur* [6, pág. 257], de 1822. No hemos encontrado un primer autor que reformule la serie de Fourier en función de la función exponencial, pero usando las identidades de Euler es sencillo hallar que

$$\varphi(x) = \sum_{n=-\infty}^{\infty} \mathrm{e}^{\frac{2\pi n}{T}\mathrm{i}x} \cdot \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \varphi(x)\,\mathrm{e}^{-\frac{2\pi n}{T}\mathrm{i}x}\,\mathrm{d}x.$$

3.3. Distribución normal

La última aplicación estudiada ha sido la función de densidad de la distribución normal, dada su importancia debida al teorema central del límite.

La primera vez que la hemos encontrado ha sido en *The Doctrine of Chances*, de Abraham de Moivre [12, págs. 243-254]. Se presenta como una forma de aproximar la distribución binomial para un gran número de experimentos n y sin que la expresión resultante dependa de dicho número de experimentos. Divide el estudio en dos partes, primero para una distribución binomial[7] $X \sim \mathrm{Bin}\left(n, \frac{1}{2}\right)$ y después una distribución con probabilidad arbitraria $X \sim \mathrm{Bin}\left(n, \frac{a}{a+b}\right)$[8]. En ambos casos, el proceso que sigue es el mismo.

De Moivre trabaja con probabilidades mediante casos favorables y casos desfavorables y aplica lo que conocemos hoy como regla de Laplace. Por tanto, trabajaremos con una función de casos favorables de nuestra variable aleatoria, que denotaremos como C. Fav. $[X = x]$, para indicar el número de casos en los que la variable aleatoria X toma el valor x. Con esta notación, y sabiendo que el número de casos totales es 2^n, la función de probabilidad viene dada en términos de la regla de Laplace por

$$\Pr[X = x] = \frac{\text{C. Fav. } [X = x]}{2^n}.$$

Primero, De Moivre aproxima el valor de la probabilidad más alta de la distribución, que corresponde al término central $x = n/2$, obteniendo que

$$\frac{\text{C. Fav. } [X = n/2]}{2^n} = \frac{1}{2^n}\binom{n}{n/2} \simeq \frac{2A(n-1)^n}{n^n\sqrt{n-1}}.$$

Para valores de n muy grandes obtiene que $\sqrt{n-1} \simeq \sqrt{n}$ y también emplea la aproximación de e como

$$\frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n \simeq \mathrm{e}^{-1}.$$

[6] Aprovechamos para recordar que, sobre el espacio vectorial de funciones $\mathcal{C}^0[a,b]$ siendo $a < b$, la forma lineal

$$\langle f, g\rangle = \int_a^b f(x)\,g(x)\,\mathrm{d}x$$

es un producto escalar entre funciones. Lo que estamos afirmando, entonces, respecto a la ortonormalidad de las funciones coseno es que

$$\left\langle \cos\left(\frac{2i+1}{2}\pi y\right), \cos\left(\frac{2j+1}{2}\pi y\right)\right\rangle = \int_{-1}^{1} \cos\left(\frac{2i+1}{2}\pi y\right)\cos\left(\frac{2j+1}{2}\pi y\right)\mathrm{d}y = \delta_{i,j}.$$

[7] A lo largo del texto se denotará por $X \sim \mathrm{Bin}(n,p)$ a una distribución binomial de n ensayos con probabilidad de éxito p.

[8] De Moivre no trabaja con la probabilidad p, sino con el número de casos favorables a y desfavorables b. Probar que $p = \frac{a}{a+b}$ es obvio a partir de la regla de Laplace.

Con todo esto, la aproximación final es

$$\frac{2A(n-1)^n}{n^n\sqrt{n-1}} \simeq \frac{2B}{\sqrt{n}}.$$

De Moivre indica que Stirling halla posteriormente el valor de B como $B = \frac{1}{\sqrt{2\pi}}$.

A continuación, calcula el cociente entre un término cualquiera del binomio y el término central. Lo aproxima indirectamente mediante logaritmos, obteniendo para n grande que

$$\ln\left(\frac{\text{C. Fav. } [X = x]}{\text{C. Fav. } [X = n/2]}\right) \simeq -\frac{2\ell^2}{n},$$

donde ℓ es la distancia que nos alejamos de la media de $X \sim \text{Bin}\left(n, \frac{1}{2}\right)$, es decir, $\ell = |x - \frac{n}{2}|$. Así, la probabilidad de un evento cualquiera viene dada por

$$\begin{aligned}\Pr[X = x] &= \frac{\text{C. Fav. } [X = x]}{2^n} \\ &= \frac{\text{C. Fav. } [X = n/2]}{2^n} \cdot \frac{\text{C. Fav. } [X = x]}{\text{C. Fav. } [X = n/2]} \\ &\simeq \frac{2}{\sqrt{2\pi n}}\, \mathrm{e}^{-\frac{2\ell^2}{n}}.\end{aligned}$$

Repitiendo el cálculo para una distribución binomial de probabilidad arbitraria $X \sim \text{Bin}\left(n, \frac{a}{a+b}\right)$, obtiene aproximaciones de los cocientes anteriores como

$$\frac{\text{C. Fav. } [X = np]}{(a+b)^n} \simeq \frac{a+b}{\sqrt{2\pi abn}}, \qquad \ln\left(\frac{\text{C. Fav. } [X = x]}{\text{C. Fav. } [X = np]}\right) \simeq -\frac{(a+b)^2}{2abn}\ell^2.$$

Siguiendo la misma deducción que en el caso con probabilidad $p = 1/2$, la función de probabilidad para la distribución binomial arbitraria viene dada por

$$\Pr[X = x] = \frac{a+b}{\sqrt{2\pi abn}}\, \mathrm{e}^{-\frac{(a+b)^2}{2abn}\ell^2},$$

donde ahora tenemos que $\ell = x - n\frac{a}{a+b}$. En este caso, recordando que la media en una distribución binomial viene dada por $\mu = np = n\,\frac{a}{a+b}$ y la varianza por $\sigma = \sqrt{npq} = \frac{\sqrt{abn}}{a+b}$, la función de probabilidad viene dada por

$$\Pr[X = x] = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

que es precisamente la función de probabilidad de la distribución normal.

4. Conclusiones

Tanto los motivos que llevan a la definición de e como sus aplicaciones son muy variadas. En el origen hemos encontrado aproximaciones por diversas áreas: construcción de tablas de logaritmos, problemas de cálculo de áreas bajo la hipérbola, desarrollos en serie de funciones exponenciales e incluso problemas de economía. También hemos visto su rango de aplicaciones en análisis y estadística, estando presente en teoremas clave de ambas ramas.

Las aplicaciones que hemos expuesto no son más que un pequeño ejemplo de las muchas que se pueden estudiar. Este campo queda abierto a más autores que deseen profundizar más en las aplicaciones de e. Por otro lado, también se puede profundizar en el origen buscando y analizando más autores que hayan hecho aportes a la definición de e. Actualmente estamos trabajando en esta tarea, revisando autores presentes en el proyecto y estudiando otros nuevos.

Referencias

[1] Bernoulli, Jakob. «Quæstiones Nonnullæ de usuris, cum solutione Problematis de Sorte Alearum, propositi in Ephem. Gall. A. 1685». En: *Acta Eroditorum* 9 (1690), págs. 219-223. URL: https://hdl.handle.net/2027/ucm.5324324906?urlappend=%3Bseq=249.

[2] Cajori, Florian. *A History of mathematical notations*. Reimpresión de la edición de 1929. Vol. II. Chicago: The Open Court Company, 1952. URL: https://archive.org/details/in.ernet.dli.2015.88254/.

[3] Cajori, Florian. *A History of Mathematics*. 5.ª ed. Rhode Island: American Mathematical Society, 1991. ISBN: 978-0-8218-2102-2.

[4] Euler, Leonhard. «Meditatio in experimenta explosione tormentorum nuper instituta». En: *Euler Archive - All Works*. 853. Obra póstuma publicada en 1862. 1728. URL: https://scholarlycommons.pacific.edu/euler-works/853/.

[5] Euler, Leonhard. *Introductio in Analysin Infinitorum*. Reimpresión del año 2000. Sevilla: Real Sociedad Matemática Española, 1748.

[6] Fourier, Jean-Baptiste Joseph. *Théorie analytique de la chaleur*. París: Firmin Didot, Père et Fils, 1822. URL: https://archive.org/details/thorieanalytiq00four/.

[7] Fourier, Jean-Baptiste Joseph. «Mémoire sur la propagation de la chaleur dans les corps solides». En: *Oeuvres de Fourier*. Ed. por Darboux, Jean Gaston. Vol. 2. París: Gauthier-Villars et Fils, 1890, págs. 213-221. URL: http://gallica.bnf.fr/ark:/12148/bpt6k33707/.

[8] Gerhardt, Carl Immanuel. *Leibnizens Matematische Schriften*. Vol. 2. Berlín: Verlag von A. Asher & Comp., 1849. URL: https://archive.org/details/leibnizensmathe01leibgoog/.

[9] Ivory, James. «An account of a table of logarithms. Published by John Speidell, an eminent English Mathematician, in the year 1619, and afterwards in the year 1628, under the title of New Logarithms, extracted from and out of those of Lord Napier». En: *Scriptores Logarithmici. A collection of several curious tracts on the nature and construction of logarithms*. Ed. por Hutton, Charles. Vol. VI. Londres: R. Wilks, 1807. URL: https://books.google.com/books?id=65YiAQAAMAAJ&printsec=frontcover.

[10] Leibniz, Gottfried Wilhelm. «Schediasma de resistentia Medii, & Motu projectorum gravium in medio resistente». En: *Acta Eroditorum* (1689), págs. 38-47. URL: https://archive.org/details/bub_gb_YPRaAAAAQAAJ/.

[11] Méndez, Hubert. «Capítulo VII: Funciones trigonométricas». En: *Tópicos de matemática elemental*. San José: Editorial EUNED, 2000. ISBN: 978-9977-64-641-1.

[12] Moivre, Abraham de. *The Doctrine of Chances. A method of calculating the probabilities of events in play*. 3.ª ed. Londres: Printed for A. Millar, in the Strand, 1756. URL: https://archive.org/details/doctrineofchance00moiv/.

[13] Napier, John. *A Description of the Admirable Table of Logarithms*. Londres, 1618.

[14] Napier, John. *The Construction of the Wonderful Canon of Logarithms*. Trad. por RaeMacDonald, William. Edimburgo y Londres: William Blackwood and sons, 1889. URL: https://archive.org/details/cu31924085321093.

[15] Napier, John. *Mirifici Logarithmorum Canonis Descriptio*. Trad. por Bruce, Ian. 2012. URL: http://www.17centurymaths.com/contents/napiercontents.html.

[16] Nicolás Martínez, Pablo. *Estudio del origen del número e y de sus aplicaciones en diversos campos de las matemáticas*. Trabajo de Investigación de Bachillerato. 2017. URL: https://fme.upc.edu/ca/premi-poincare/edicions-anteriors/premi-poincare-2017/treballs-guanyadors-2017/kronecker.pdf.

[17] Requena Fraile, Ángel. *Cuarto centenario del Mirifici Logarithmorun Canonis Descriptio*. 2014. URL: http://www.divulgamat.net/index.php?option=com_content&view=article&id=16133.

[18] Sarasa, Alphonse Antoine de. *Solutio problematis a R. P. Marino Mersenno Minimo propositi*. Amberes: Apud Ioannem et Iacobum Meursios, 1649. URL: https://archive.org/details/bub_gb_TG-i3Dzr7QoC/.

[19] Speidell, John. *New Logarithmes*. Early English Books Online. Reproducción del original por ProQuest. Londres, 1619. ISBN: 978-1-171-27615-9.

[20] Speidell, John. *New Logarithmes*. 1622.

TEMat

El problema de la palabra en los grupos de trenzas

Javier Aguilar Martín
Universidad de Sevilla (US)
javiecija96@gmail.com

Resumen: El problema de la palabra es uno de los problemas más importantes en teoría combinatoria de grupos. En este artículo presentamos una familia de grupos, los grupos de trenzas, donde es posible resolverlo, junto con uno de los algoritmos más eficientes que existen para ello.

Abstract: The word problem is one of the most important problems in combinatorial group theory. In this paper we present a family of groups, the braid groups, in which it is possible to solve it, together with one of the most efficient algorithms for that purpose.

Palabras clave: trenzas, grupo, problema de la palabra, algoritmo.

MSC2010: 20F36.

Recibido: 24 de julio de 2019.
Aceptado: 20 de noviembre de 2019.

Agradecimientos: Quiero agradecer a mis directores de TFG, Juan González-Meneses y Ramón Flores Díaz, por el apoyo y conocimiento aportado, que me permitieron desarrollar el trabajo y extraer de él este artículo.

Referencia: Aguilar Martín, Javier. «El problema de la palabra en los grupos de trenzas». En: *TEMat*, 4 (2020), págs. 27-42. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p27.

1. Introducción

El problema de la palabra es uno de los problemas fundamentales de la teoría combinatoria de grupos propuestos por Max Dehn [7]. Este problema consiste en, dado un grupo G con una presentación finita $\langle S \mid R \rangle$ y dados dos elementos a y b de G expresados como productos de los elementos de S y sus inversos, decidir si $a = b$ como elementos del grupo o, equivalentemente, si $ab^{-1} = 1$.

El nombre de este problema proviene de que podemos considerar el alfabeto $\Sigma = S \cup S^{-1}$, donde S^{-1} representa el conjunto formado por los inversos de los elementos de S, y ver G como un lenguaje sobre Σ, en el que dos palabras a y b representarán el mismo elemento si y solo si se puede transformar a en b mediante un número finito de pasos usando las reglas de reescritura proporcionadas por las relaciones de R junto con la cancelación de inversos.

El propio Dehn describió algoritmos para resolver el problema de la palabra en grupos fundamentales de 2-variedades orientables cerradas con género mayor o igual que 2 [8]. Sin embargo, en 1955 Pyotr Novikov encontró ejemplos de grupos finitamente presentados donde el problema de la palabra era indecidible [18], es decir, que no se puede diseñar un algoritmo que lo resuelva. A pesar de esto, hay gran cantidad de grupos donde el problema de la palabra sí es resoluble. Ejemplos claros de ello son los grupos finitos y los grupos libres. Aquí estudiaremos los grupos de trenzas, que aparecen en numerosas ramas de las matemáticas, como el álgebra, la topología, la criptografía y el análisis, y en los cuales el problema de la palabra es resoluble.

En este artículo, basado en el TFG de Javier Aguilar Martín [1], empezamos con algunos preliminares sobre teoría de grupos y topología. A continuación introducimos los grupos de trenzas, destacando los subgrupos de trenzas puras y las presentaciones de dichos grupos. Por último, describimos una estructura en los grupos de trenzas basada en unos monoides incluidos en los grupos. Esta estructura permite resolver el problema de la palabra.

2. Preliminares

Dedicaremos esta sección a hablar sobre monoides, explicar qué son las presentaciones de grupos y dar una pequeña introducción a la teoría de homotopía.

2.1. Monoides

Empezaremos hablando de monoides, puesto que dentro del grupo de trenzas hay un monoide importante que nos permitirá desarrollar el algoritmo para resolver el problema de la palabra, además de que tendrán relevancia en las presentaciones de grupos.

Definición 1. Un **monoide** es un par $(S, *)$, donde S es un conjunto y $* \colon S \times S \to S$ es una operación binaria que satisface las siguientes propiedades:

- Asociatividad, es decir, para cualesquiera $a, b, c \in S$, $(a * b) * c = a * (b * c)$.
- Existencia de elemento neutro, es decir, existe $e \in S$ tal que, para todo $a \in S$, $e * a = a * e = a$.

Habitualmente el símbolo de la operación será omitido y nos referiremos a S como monoide, entendiéndose que en realidad es el par anterior. Un **homomorfismo de monoides** $f \colon S \to T$ es una aplicación que verifica que $f(ab) = f(a)f(b)$ para todo $a, b \in S$. ◂

Observación 2. Un monoide es análogo a un grupo, pero sin requerir la existencia de elementos inversos, esto es, no exigimos que para todo $s \in S$ exista $s^{-1} \in S$ tal que $ss^{-1} = s^{-1}s = e$. ◂

Ejemplo 3. En el contexto de los lenguajes formales es habitual llamar **alfabeto** a cualquier conjunto finito X. Esto es porque podemos considerar el monoide de palabras en este conjunto. Por ejemplo, sea $X = \{a, b\}$. El **lenguaje** generado por X se denota X^* y se corresponde con todas las **palabras** (cadenas de texto) formadas con las letras a y b, así como la palabra vacía ε. La operación que dota de estructura de monoide al lenguaje es la concatenación. ◂

2.2. Presentaciones de grupos

Definición 4. Sea X un conjunto, L un grupo e $i\colon X \to L$ una función. Diremos que (L,i) es **libre** en X si para todo grupo H y toda función $f\colon X \to H$ existe un único homomorfismo $\varphi\colon L \to H$ de modo que el siguiente diagrama conmuta.

$$\begin{array}{ccc} X & \xrightarrow{f} & H \\ {\scriptstyle i}\downarrow & \nearrow_{\exists!\varphi} & \\ L & & \end{array}$$

◂

Observación 5. Sea $H = \langle Y\rangle$ un grupo. Sea X un conjunto con $|X| > |Y|$ y (L,i) un grupo libre en X. Como $|X| > |Y|$, podemos tomar $f\colon X \to H$ tal que $Y \subseteq f(X)$. Entonces, existe $\varphi\colon L \to H$ homomorfismo sobreyectivo y tenemos que $H \cong L/\ker\varphi$. ◂

Definición 6. Una **presentación** de H es un grupo libre (L,i) y un subgrupo $N \trianglelefteq L$ tal que $L/N \cong H$. ◂

Teorema 7 ([2, §II.5]). *Dado un conjunto X, existe (L,i) libre en X.*

Demostración. Sea X^{-1} un conjunto en biyección con X mediante $x \mapsto x^{-1}$. Por abuso de notación, a la inversa también la denotamos $x \mapsto x^{-1}$. Sea $(X \cup X^{-1})^*$ el monoide de palabras en $X \cup X^{-1}$ con la concatenación. Decimos que $w = y_1 \cdots y_n \in (X \cup X^{-1})^*$ es **reducida** si para todo $1 \leq i \leq n-1$ tenemos que $y_i \neq y_{i+1}^{-1}$. Si w no es reducida, entonces $w' = y_1 \cdots y_{i-1} y_{i+2} \cdots y_n$ se ha obtenido a través de una **reducción elemental** y escribimos $w \to w'$.

Dos palabras son equivalentes si se puede obtener una a partir de la otra mediante reducciones elementales o mediante el proceso inverso de introducir un par de la forma $y_i y_i^{-1}$. Es fácil comprobar que esta relación entre palabras es una relación de equivalencia.

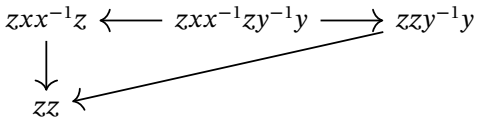


Figura 1: Palabras equivalentes.

Entonces se tiene que $F(X) = (X \cup X^{-1})^*/\sim$, siendo $\sim$ la relación de equivalencia anterior, es un grupo libre en X y existe una única palabra reducida en cada clase de equivalencia. Para la segunda afirmación, ver el libro de Lyndon y Schupp [15, Capítulo 1, §1]. Para la primera tenemos que probar dos cosas: que $F(X)$ es un grupo y que es libre. La propiedad asociativa y la existencia de elemento neutro se heredan del monoide $(X \cup X^{-1})^*$. Dado un elemento representado por una palabra reducida $w = y_1^{\varepsilon_1} \cdots y_n^{\varepsilon_n}$, es inmediato comprobar que el inverso está representado por la palabra $w^{-1} = y_n^{-\varepsilon_n} \cdots y_1^{-\varepsilon_1}$.

Probamos a continuación que $F(X)$ es libre. Consideramos $i\colon X \to F(X)$ la aplicación que envía cada elemento a su clase de equivalencia (los elementos de X, de hecho, son palabras reducidas). Sean G un grupo y $f\colon X \to G$ una función. Definimos $\varphi\colon F(X) \to G$ a partir de las palabras reducidas como $x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n} \mapsto f(x_{i_1})^{\varepsilon_1} \cdots f(x_{i_n})^{\varepsilon_n}$. Cualquier homomorfismo de $F(X) \to G$ que haga que el diagrama conmute debe cumplir esa definición, luego es único. ∎

Observación 8. El grupo libre en X es único salvo isomorfismo, esto es, si (L_1,i_1) y (L_2,i_2) son libres en X, entonces $L_1 \cong L_2$. ◂

Volvemos a las presentaciones. Dado X, el grupo libre en X existe y es único salvo isomorfismo. Lo denotamos $\langle X \mid \rangle$.

Sea G un grupo. Dado $R \subseteq G$, $\langle R^G\rangle$ denota el subgrupo generado por todos los G-conjugados de R, que coincide con el menor subgrupo normal que contiene a R:

$$\bigcap_{N \trianglelefteq G, R \subseteq N} N.$$

Dado un conjunto X y $R \subseteq \langle X \mid \rangle$, denotamos por $\langle X \mid R \rangle$ al grupo $\langle X \mid \rangle / \langle R^{\langle X \mid \rangle} \rangle$. Por lo visto anteriormente, para todo grupo G existen un conjunto X y $R \subseteq \langle X \mid \rangle$ tales que $\langle X \mid R \rangle \cong G$.

Ejemplo 9.

1. Para $n \geq 1$, consideremos la presentación $\langle x \mid x^n \rangle$. Este es el grupo generado por un elemento de orden n, por lo que este grupo es isomorfo al grupo cíclico aditivo $\mathbb{Z}/n\mathbb{Z}$. A menudo, las relaciones se escriben como ecuaciones, de modo que esta presentación podría haberse escrito como $\langle x \mid x^n = 1 \rangle$.
2. La presentación del grupo libre abeliano de rango n, $\mathbb{Z}^n$, está dada por $\langle x_1, \dots, x_n \mid [x_i, x_j] = 1 \rangle$, donde $[x_i, x_j] = x_i x_j x_i^{-1} x_j^{-1}$ es el conmutador de los elementos x_i y x_j. Es decir, la presentación expresa que el grupo está generado por n elementos que conmutan entre sí y ninguna relación más, por lo que efectivamente es el grupo libre abeliano de rango n. Escribiendo las relaciones como ecuaciones podemos también despejar y escribir la presentación anterior como $\langle x_1, \dots, x_n \mid x_i x_j = x_j x_i \rangle$. ◂

2.3. Transformaciones de Tietze

Vamos a concentrarnos a partir de ahora en presentaciones **finitas**, es decir, tanto X como R serán conjuntos finitos. Se conocen como **transformaciones de Tietze** los siguientes isomorfismos:

1. Si $s \in \langle R^{\langle X \mid \rangle} \rangle$, entonces $\langle X \mid R \rangle \cong \langle X \mid R \cup \{s\} \rangle$.
2. Si $y \notin X$ y $u \in \langle X \mid \rangle$, entonces $\langle X \mid R \rangle \cong \langle X \cup \{y\} \mid R \cup \{yu\} \rangle$.

Teorema 10 ([15, Capítulo 2, Proposición 2.1]). *Si $\langle X \mid R \rangle \cong \langle X \mid R' \rangle$ son presentaciones finitas, entonces se puede obtener una presentación de la otra mediante una sucesión finita de transformaciones de Tietze.*

Volviendo al problema de la palabra, sean $\langle X \mid R \rangle$ una presentación finita de un grupo G y $w \in (X \cup X^{-1})^*$. Entonces, $w =_G 1$ si y solo si $w \in \langle R^{\langle X \mid \rangle} \rangle$ si y solo si $w = \prod_{i=1}^{M} p_i r_i^{\varepsilon_i} p_i^{-1}$ en $\langle X \mid \rangle$ para algún $M \in \mathbb{N}$, algunos $r_i \in R$ y algunos $p_i \in \langle X \mid \rangle$, con $i \in \{1, \dots, M\}$ y $\varepsilon_i = \pm 1$.

2.4. Relación entre monoides y grupos

De forma análoga a como se hace para grupos, podemos considerar los generadores de un monoide y una presentación de un monoide mediante generadores y relaciones. También se definen de forma análoga los morfismos entre monoides. De hecho, la definición de monoide libre es análoga a la de grupos, siendo el monoide libre en un conjunto X el lenguaje generado por este conjunto.

Definición 11. Dado un monoide S con presentación $\langle M \mid R \rangle$, su **grupo de fracciones** $G(S)$ es el grupo con presentación $\langle M \mid R \rangle$. ◂

Existe una aplicación natural de un monoide en su grupo de fracciones. Sin embargo, esta aplicación no siempre es inyectiva, pues la existencia de inverso en el grupo puede hacer que dos elementos distintos del monoide representen el mismo elemento del grupo de fracciones. Por ejemplo, si consideramos la presentación $\langle a, b, c \mid ab = cb \rangle$, los elementos a y c son distintos en el monoide; sin embargo, en el grupo son el mismo, pues multiplicando a la derecha por b^{-1} en la relación obtenemos $a = c$. De aquí que consideremos la siguiente definición.

Definición 12. Decimos que un monoide S se **inyecta** en su grupo de fracciones $G(S)$ si el morfismo de monoides $\iota : S \to G(S)$ dado por $\iota(a) = a$ es inyectivo. ◂

Definición 13. Decimos que un monoide S satisface las **condiciones de Ore** [19] si se cumple lo siguiente:

- S es cancelativo, es decir, $xay = xby$ implica $a = b$ para todo $x, y, a, b \in S$.
- Para todo $a, b \in S$ existen $a', b' \in S$ tales que $aa' = bb'$ (existe un múltiplo común). ◂

Proposición 14 ([6, Teorema 1.23]). *Si un monoide satisface las condiciones de Ore, entonces se inyecta en su grupo de fracciones.*

2.5. Teoría de homotopía

Para definir algunos conceptos necesitaremos ciertas nociones básicas de teoría de homotopía que procedemos a enunciar.

Definición 15. Sean $f, g : X \to Y$ funciones continuas y sea $I = [0,1]$. Decimos que f y g son **homotópicas**, denotado $f \simeq g$, si existe $H : X \times I \to Y$ continua tal que $H(x,0) = f(x)$ y $H(x,1) = g(x)$. A la aplicación H la llamamos **homotopía** entre f y g. Si $A \subseteq X$, decimos que f y g son **homotópicas relativamente** a A si la homotopía H cumple además que $H(a,t) = f(a) = g(a)$ para todo $a \in A$. Si A es un conjunto unitario, hablamos de **homotopía basada**. ◂

Lema 16. *Ser homotópicas es una relación de equivalencia (también serlo relativamente).*

Demostración. Vamos a demostrar que la relación de homotopía es reflexiva, simétrica y transitiva.

- Reflexiva: $f \simeq f$ mediante $H(x,t) = f(x)$.
- Simétrica: si $f \simeq g$ mediante H, entonces $g \simeq f$ mediante $\widetilde{H}(x,t) = H(x,1-t)$. Se cumple que

$$\widetilde{H}(x,0) = H(x,1) = g(x),$$
$$\widetilde{H}(x,1) = H(x,0) = f(x).$$

  Es inmediato probar que $\widetilde{H}$ es continua.
- Transitiva: sea $f \simeq g$ mediante F y $g \simeq h$ mediante G. Entonces, $f \simeq h$ mediante

$$H(x,t) = \begin{cases} F(x,2t) & \text{si } 0 \le t \le 1/2, \\ G(x,2t-1) & \text{si } 1/2 \le t \le 1, \end{cases} \Longrightarrow \begin{cases} H(x,0) = F(x,0) = f(x), \\ H(x,1) = G(x,1) = h(x). \end{cases}$$

  Está bien definida en $t = 1/2$ pues $F(x,1) = g(x) = G(x,0)$. Es inmediato probar la continuidad de H y adaptar la demostración al caso relativo. ∎

2.6. Caminos

Definición 17. Dado un espacio topológico X, un **camino** entre x e y pertenecientes a X es una aplicación continua $\alpha : I \to X$ tal que $\alpha(0) = x$ y $\alpha(1) = y$, siendo $I = [0,1]$. ◂

Definición 18. Dados $\alpha, \beta : I \to X$ dos caminos con $\alpha(0) = x, \alpha(1) = \beta(0) = y, \beta(1) = z$, se llama **concatenación** de α y β al camino definido como

$$\alpha\beta(t) = \begin{cases} \alpha(2t) & \text{si } 0 \le t \le 1/2, \\ \beta(2t-1) & \text{si } 1/2 \le t \le 1. \end{cases}$$ ◂

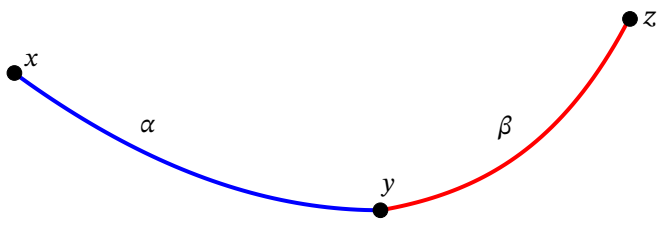


Figura 2: La concatenación $\alpha\beta$ es el camino resultante desde x hasta z.

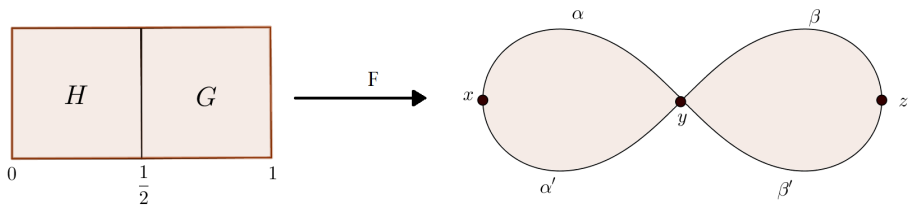
Nota 19. Si α y β son caminos homotópicos relativamente a $\{0,1\}$, decimos que son **equivalentes** y escribimos $\alpha \sim \beta$ en lugar de $\alpha \simeq \beta$ relativamente a $\{0,1\}$. ◂

Lema 20. *La concatenación es compatible con la equivalencia de caminos, esto es, si $\alpha \sim \alpha'$ y $\beta \sim \beta'$, entonces $\alpha\beta \sim \alpha'\beta'$.*

Demostración. Sean $\alpha \sim \alpha'$ y $\beta \sim \beta'$ caminos con $\alpha(0) = \alpha'(0) = x$, $\alpha(1) = \alpha'(1) = \beta(0) = \beta'(0) = y$ y $\beta(1) = \beta'(1) = z$. Como $\alpha \sim \alpha'$, existe una homotopía H entre α y α' relativa a $\{0,1\}$. Como $\beta \sim \beta'$, existe una homotopía G entre β y β' relativa a $\{0,1\}$.

Sea $F\colon I \times I \to X$ la homotopía resultante de «unir» las dos anteriores,

$$F(t,s) = \begin{cases} H(2t,s) & \text{si } 0 \le t \le 1/2, \\ G(2t-1,s) & \text{si } 1/2 \le t \le 1. \end{cases}$$



$F(1/2,s)$ está bien definida pues, al ser H y G relativas a $\{0,1\}$, se tiene que $H(1,s) = y = G(0,s)$. Tenemos que

$$F(t,0) = \begin{cases} H(2t,0) = \alpha(2t) & \text{si } 0 \le t \le 1/2, \\ G(2t-1,0) = \beta(2t-1) & \text{si } 1/2 \le t \le 1, \end{cases}$$

que es justamente la concatenación de α y β. Análogamente, $F(t,1) = (\alpha'\beta')(t)$. Finalmente,

$$\left.\begin{array}{l} F(0,s) = H(0,s) = x \\ F(1,s) = G(1,s) = z \end{array}\right\} \implies \alpha\beta \sim \alpha'\beta'. \qquad \blacksquare$$

Proposición 21. *Se cumplen las siguientes propiedades de la concatenación con respecto a la equivalencia de caminos:*

1. *Propiedad asociativa:* $(\alpha\beta)\gamma \sim \alpha(\beta\gamma)$.
2. *Elemento neutro: si c_x es el camino constante x y α es un camino entre x e y, entonces $c_x\alpha \sim \alpha \sim \alpha c_y$.*
3. *Elemento inverso: si α es un camino entre x e y y $\overline{\alpha}\colon I \to X$ es el camino $\overline{\alpha}(t) = \alpha(1-t)$ (camino opuesto), entonces $\alpha\overline{\alpha} \sim c_x$ y $\overline{\alpha}\alpha \sim c_y$.*

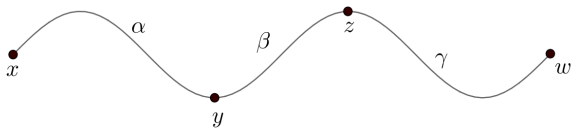


Figura 3: La triple concatenación está bien definida salvo homotopía.

Demostración.

1. Por definición, tenemos por un lado que

$$\alpha(\beta\gamma)(t) = \begin{cases} \alpha(2t) & \text{si } 0 \le t \le 1/2, \\ \beta\gamma(2t-1) & \text{si } 1/2 \le t \le 1, \end{cases} = \begin{cases} \alpha(2t) & \text{si } 0 \le t \le 1/2, \\ \beta(4t-2) & \text{si } 1/2 \le t \le 3/4, \\ \gamma(4t-3) & \text{si } 3/4 \le t \le 1. \end{cases}$$

   Por otro lado,

$$(\alpha\beta)\gamma(t) = \begin{cases} (\alpha\beta)(2t) & \text{si } 0 \le t \le 1/2, \\ \gamma(2t-1) & \text{si } 1/2 \le t \le 1, \end{cases} = \begin{cases} \alpha(4t) & \text{si } 0 \le t \le 1/4, \\ \beta(4t-1) & \text{si } 1/4 \le t \le 1/2, \\ \gamma(2t-1) & \text{si } 1/2 \le t \le 1. \end{cases}$$

Sea, por lo tanto,

$$F(t,s)=\begin{cases}\alpha\left(\dfrac{4t}{s+1}\right) & \text{si } 0\le t\le \dfrac{s+1}{4},\\ \beta\left(4t-(s+1)\right) & \text{si } \dfrac{s+1}{4}\le t\le \dfrac{s+2}{4},\\ \gamma\left(\dfrac{4t-s-2}{2-s}\right) & \text{si } \dfrac{s+2}{4}\le t\le 1.\end{cases}$$

Se tiene que F es una homotopía relativa a $\{0,1\}$ entre $F(t,0)=(\alpha\beta)\gamma(t)$ y $F(t,1)=\alpha(\beta\gamma)(t)$.

2. Para probar que $c_x\alpha\sim\alpha$ definimos la homotopía

$$F(t,s)=\begin{cases}x & \text{si } 0\le t\le \dfrac{1-s}{2},\\ \alpha\left(\dfrac{2t+s-1}{s+1}\right) & \text{si } \dfrac{1-s}{2}\le t\le 1.\end{cases}$$

La relación $\alpha\sim\alpha c_y$ se demuestra de manera análoga.

3. Se tiene que

$$\alpha\overline{\alpha}(t)=\begin{cases}\alpha(2t) & \text{si } 0\le t\le 1/2,\\ \overline{\alpha}(2t-1)=\alpha(2-2t) & \text{si } 1/2\le t\le 1\end{cases}$$

define una homotopía relativa a $\{0,1\}$ entre $\alpha\overline{\alpha}$ y c_x. Análogamente, se puede encontrar una homotopía relativa a $\{0,1\}$ entre $\overline{\alpha}\alpha$ y c_y. ∎

3. Grupos de trenzas

Aunque el término *grupo de trenzas* fue acuñado por Artin [3] en 1925, estos grupos ya fueron considerados por Hurwitz [14] en 1891 como lo que en terminología moderna se llamaría «grupo fundamental de espacios de configuración de n puntos en el plano complejo». En 1935, Magnus [16] consideró el mismo grupo desde el punto de vista de los *mapping class groups*. Markoff [17] dio una aproximación totalmente algebraica.

Esta variedad de definiciones permite estudiar los grupos de trenzas desde perspectivas muy distintas, lo cual aporta una gran riqueza a la teoría. Aquí veremos la definición más geométrica y mundana, además de su presentación, que también puede tomarse como definición algebraica del grupo.

3.1. Trenzas como colección de cuerdas

Empezamos dando la definición más gráfica e intuitiva, consistente en visualizar las trenzas como cuerdas que se entrelazan.

Definición 22. Sea $n\ge 1$ un entero. Denotemos Σ_n al grupo simétrico sobre n elementos. Sean n puntos $P_1,\dots,P_n$ en $\mathbb{C}$ (se puede suponer que $P_k=k$ para todo $1\le k\le n$). Se define una **trenza geométrica de n cuerdas** como una n-upla $\beta=(\beta_1,\dots,\beta_n)$ de caminos $\beta_k\colon [0,1]\to\mathbb{C}\times[0,1]$ tal que

- $\beta_k(t)=(\alpha_k(t),t)$, donde $\alpha_k(0)=P_k$ para todo $1\le k\le n$,
- existe una permutación $\tau=\tau(\beta)\in\Sigma_n$ tal que $\alpha_k(1)=P_{\tau(k)}$ para todo $1\le k\le n$, llamada **permutación inducida por** β, y
- $\alpha_k(t)\ne\alpha_\ell(t)$ para todo $k\ne\ell$ y para todo $t\in[0,1]$.

Si la permutación inducida por β es el elemento neutro de Σ_n, es decir, si $\beta_k(1)=(P_k,1)$ para todo $1\le k\le n$, entonces decimos que la trenza geométrica es **pura**.

Dos trenzas geométricas α y β se dicen **homotópicas** si existe una familia continua de trenzas $\{\gamma_s\}_{s\in[0,1]}$ de modo que $\gamma_0=\alpha$ y $\gamma_1=\beta$. Es decir, dos trenzas geométricas son homotópicas si son homotópicas como colección de caminos relativamente a los puntos extremos. Consideraremos que dos trenzas geométricas son la misma si son homotópicas, y a la clase de homotopía de una trenza geométrica de n cuerdas la

llamaremos **trenza de n cuerdas**. Nótese que, si α y β son homotópicas, entonces $\tau(\alpha) = \tau(\beta)$, así que diremos que una trenza es **pura** si los elementos de su clase de homotopía son trenzas geométricas puras. ◂

La forma de un dibujo tridimensional de una trenza geométrica se puede observar en la figura 4.

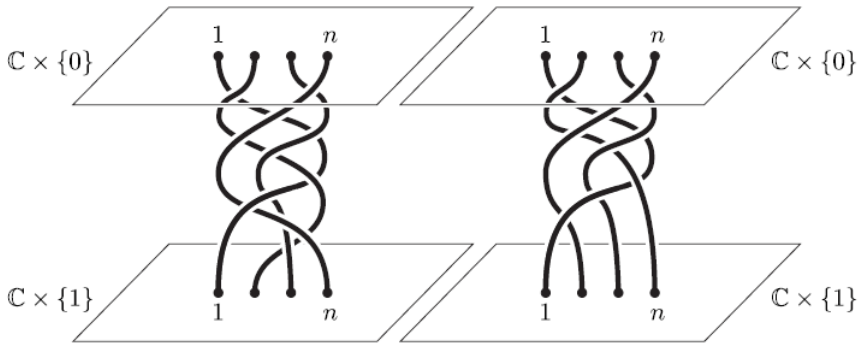


Figura 4: Una trenza geométrica pura y una trenza geométrica no pura.

Observación 23. Para cada $t \in [0,1]$, el plano $\mathbb{C} \times \{t\}$ es atravesado una sola vez por cada cuerda de la trenza. ◂

Normalmente, se representan las trenzas como su proyección en $\mathbb{R} \times [0,1]$ (posiblemente seguida de una rotación de 90°, ver figura 7). Los puntos en los que la proyección de dos cuerdas coincida los representaremos como en la figura 5 para conservar la información de cuál cruzaba originalmente por encima. Salvo homotopía, podemos suponer que la proyección tiene un número finito de puntos de cruce, en los cuales solo intervienen dos cuerdas. Además, podemos suponer también que los cruces ocurren a distintas alturas, es decir, para distintos valores de $t \in [0,1]$. En la figura 7 se ilustra la proyección de la trenza no pura de la figura 4.

Figura 5: Cruce positivo y cruce negativo, respectivamente.

Definición 24. Se definen los **generadores estándar** o **generadores de Artin** como las trenzas σ_i con $1 \leq i \leq n-1$ indicadas en la figura 6. ◂

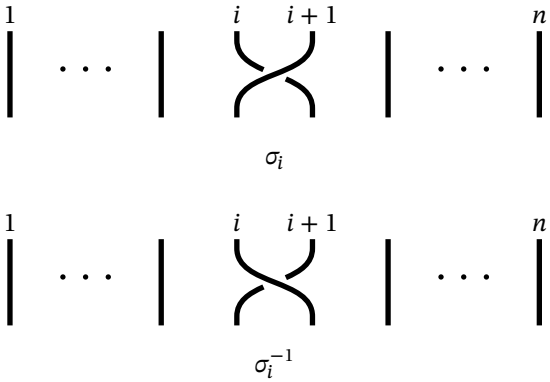


Figura 6: Generador de Artin y su inverso.

A partir de las observaciones anteriores, está claro que cualquier trenza se puede construir como concatenación de los generadores de Artin.

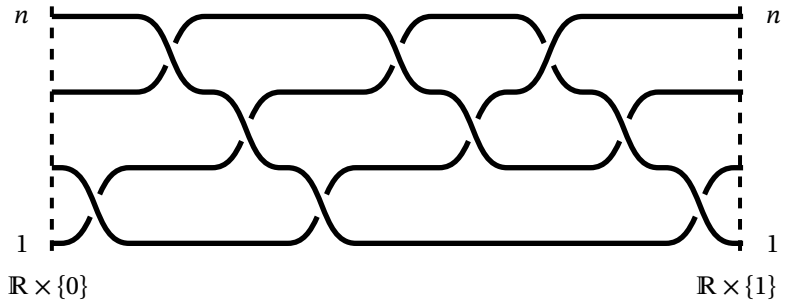


Figura 7: Ejemplo de representación plana.

3.2. Estructura de grupo

Una de las características más importantes del conjunto de clases de homotopía de trenzas es que puede dotarse de estructura de grupo para cada n. Para ello, definiremos el producto de trenzas.

Definición 25. El producto de dos trenzas $\alpha = (\alpha_1, \dots, \alpha_n)$ y $\beta = (\beta_1, \dots, \beta_n)$ se define como la trenza

$$\alpha \cdot \beta = (\alpha_1 \beta_{\tau(1)}, \dots, \alpha_n \beta_{\tau(n)}),$$

donde $\tau = \tau(\alpha)$. Es decir, el producto de dos trenzas en el mismo número de cuerdas es su concatenación, en la cual se recorre en primer lugar α y después β. En la figura 8 se ilustra un ejemplo. En ocasiones omitiremos el punto y escribiremos simplemente $\alpha\beta$. Asimismo, denotaremos $\alpha^n = \underbrace{\alpha \cdots \alpha}_{n \text{ veces}}$. ◂

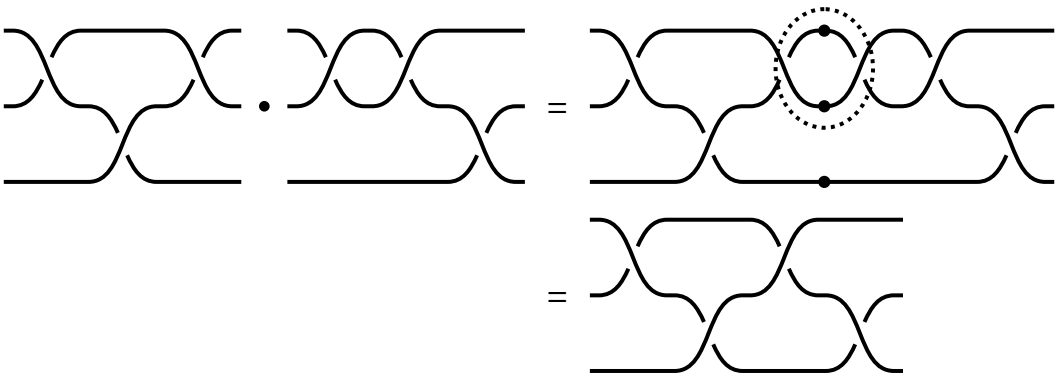


Figura 8: Producto de dos trenzas.

Denotemos B_n al conjunto de clases de homotopía de trenzas de n cuerdas y PB_n al conjunto de clases de homotopía de trenzas puras de n cuerdas. Es evidente que la multiplicación anterior induce una operación en B_n (y por tanto en PB_n); es más, se tiene el siguiente resultado.

Proposición 26. *El conjunto B_n dotado de esta operación tiene estructura de grupo. El resultado también es cierto para PB_n.*

Al grupo B_n se le llama **grupo de trenzas de *n* cuerdas** y a PB_n se le llama **grupo de trenzas puras de *n* cuerdas**.

Demostración de la proposición 26. Sean α y β dos trenzas con representantes $a = (a_1, \dots, a_n)$ y $b = (b_1, \dots, b_n)$, respectivamente. En primer lugar, veamos que la operación está bien definida, es decir, que $a \cdot b$ es una trenza geométrica y, por tanto, podemos definir $\alpha\beta = [a \cdot b]$. Sea $\tau = \tau(a)$ la permutación inducida por a. Como $a_k b_{\tau(k)}(0) = a_k(0) = (P_k, 0)$ para todo $1 \leq k \leq n$, se cumple la primera propiedad de la definición 22. Para la segunda, basta observar que la nueva permutación es $\tau(a \cdot b) = \tau(b) \circ \tau(a)$. En particular, si a y b son puras, entonces la permutación inducida por el producto también es la identidad, por lo que el producto es una trenza pura. Por último, si $t \in [0, 1/2]$, entonces $a_k \beta_{\tau(k)} = a_k(2t)$, y si $t \in [1/2, 1]$, $\alpha_k \beta_{\tau(k)} = \beta_{\tau(k)}(2t-1)$ para todo $1 \leq k \leq n$, por lo que se tiene claramente la tercera propiedad.

Por otra parte, si a' y b' son otros representantes de α y β, respectivamente, se tiene que $[a' \cdot b'] = [a \cdot b]$ por las propiedades de la homotopía de caminos con respecto a la concatenación.

Veamos ahora la estructura de grupo. Tenemos que probar que la operación es asociativa, pero esto se deduce de que la concatenación de caminos es asociativa salvo homotopía. Tenemos claramente que la identidad es la trenza constante representada por $\mathrm{Id} = (\mathrm{Id}_1, \dots, \mathrm{Id}_n)$, donde Id_k denota el camino (P_k, t) para $t \in [0, 1]$ y para $1 \leq k \leq n$. Finalmente, dada $\alpha = [(a_1, \dots, a_n)]$ con permutación inducida τ, se tiene que $\alpha^{-1} = [(\overline{a}_{\tau^{-1}(1)}, \dots, \overline{a}_{\tau^{-1}(n)})]$, donde $\overline{a}_k$ denota el camino que es opuesto a a_k en la primera coordenada y que es idéntico a a_k en la segunda coordenada.

En efecto, usando las propiedades de homotopía de caminos con respecto al camino opuesto tenemos que

$$\alpha\alpha^{-1} = [(a_1, \dots, a_n) \cdot (\overline{a}_{\tau^{-1}(1)}, \dots, \overline{a}_{\tau^{-1}(n)})] = [(a_1 \overline{a}_{\tau(\tau^{-1}(1))}, \dots, a_n \overline{a}_{\tau(\tau^{-1}(n))})] = [\mathrm{Id}].$$

Análogamente se prueba que $\alpha^{-1}\alpha = [\mathrm{Id}]$. ∎

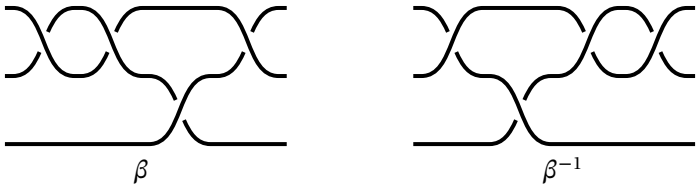


Figura 9: Una trenza y su inversa.

3.3. Presentación del grupo

Una de las características mejor conocidas de los grupos de trenzas es su presentación finita descubierta por Artin [4]. Ya hemos mencionado los generadores $\sigma_1, \dots, \sigma_{n-1} \in B_n$ en la definición 24. La presentación completa sería la siguiente:

(1)
$$B_n = \left\langle \sigma_1, \dots, \sigma_{n-1} \;\middle|\; \begin{array}{ll} \sigma_i\sigma_j = \sigma_j\sigma_i, & |i-j| > 1 \\ \sigma_i\sigma_j\sigma_i = \sigma_j\sigma_i\sigma_j, & |i-j| = 1 \end{array} \right\rangle.$$

La prueba de la completitud de esta presentación puede encontrarse en el artículo de Magnus [16].

Vamos a dar también la presentación del grupo de trenzas puras, en concreto la dada por Birman [5] (ver también el artículo de González-Meneses y Silvero [13]), pues nos será más útil para probar ciertos resultados. La presentación original fue dada por Artin [4]. Así pues, definimos los **generadores de Birman**

(2)
$$A_{ij} = \sigma_{j-1} \dots \sigma_{i+1} \sigma_i^2 \sigma_{i+1}^{-1} \dots \sigma_{j-1}^{-1} \; (1 \leq i < j \leq n)$$

y las relaciones

$$\begin{aligned}
A_{ij}^{-1} A_{rs} A_{ij} &= A_{rs} && (i < j < r < s) \text{ o bien } (r+1 < i < j < s), \\
A_{ij}^{-1} A_{js} A_{ij} &= A_{is} A_{js} A_{is}^{-1} && (i < j < s), \\
A_{ij}^{-1} A_{is} A_{ij} &= A_{is} A_{js} A_{is} A_{js}^{-1} A_{is}^{-1} && (i < j < s), \\
A_{ij}^{-1} A_{rs} A_{ij} &= A_{is} A_{js} A_{is}^{-1} A_{js}^{-1} A_{rs} A_{js} A_{is} A_{js}^{-1} A_{is}^{-1} && (i+1 < r < j < s).
\end{aligned}$$

En la figura 10 se puede observar qué trenza representa geométricamente el generador A_{ij}.

Nota 27. Cuando $j = i+1$, $A_{ij} = \sigma_i^2$. ◂

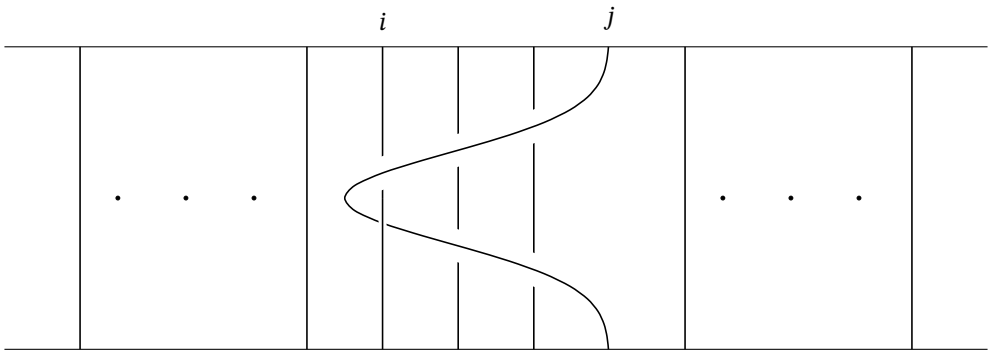


Figura 10: Interpretación geométrica de la trenza A_{ij}.

4. Algoritmo de Garside

El objetivo de esta sección es proporcionar una *forma normal* para las trenzas, es decir, una forma «estándar» de escribirlas, de modo que para ver si dos palabras representan la misma trenza sea suficiente calcular sus formas normales y comprobar si son iguales. Para llegar hasta esa forma normal estudiaremos la *estructura de Garside* del grupo de trenzas, para lo cual seguiremos la referencia «The braid group and other groups» [12]. Los resultados referentes a *Word Processing in Groups* [11] se pueden encontrar en el capítulo 9 de dicho libro.

4.1. Formas normales

Obsérvese que la presentación (1) solo involucra potencias positivas de los generadores. Por tanto, se puede considerar el monoide B_n^+ determinado por esa misma presentación. Los elementos de B_n^+ son palabras en $\sigma_1, \dots, \sigma_{n-1}$ (pero no sus inversos), y dos palabras son equivalentes si y solo si una puede obtenerse de la otra reemplazando reiteradamente subpalabras de la forma $\sigma_i\sigma_j$ con $|i-j|>1$ (respectivamente, $\sigma_i\sigma_j\sigma_i$ con $|i-j|=1$) por $\sigma_j\sigma_i$ (respectivamente, $\sigma_j\sigma_i\sigma_j$).

Definición 28. El monoide B_n^+ se denomina **monoide de las trenzas positivas** y sus palabras son llamadas **trenzas positivas**. ◂

En el monoide B_n^+ hay un orden parcial natural.

Definición 29. Definimos en B_n^+ el orden parcial $\preccurlyeq$ tal que, dadas $a, b \in B_n^+$, $a \preccurlyeq b$ si $ac = b$ para alguna $c \in B_n^+$. Decimos en ese caso que a es un **prefijo** de b. Escribimos $a \prec b$ si c no es trivial. Si además $a \neq 1$, decimos que a es un **prefijo propio** de b. ◂

Antes de continuar debemos probar que la relación que hemos definido es realmente un orden parcial.

Lema 30. *La relación $\preccurlyeq$ es una relación de orden.*

Demostración. Dada $x \in B_n^+$, se tiene que $x \preccurlyeq x \cdot 1 = x$, por lo que se cumple la propiedad reflexiva. Si $x, y \in B_n^+$ con $x \preccurlyeq y$ e $y \preccurlyeq x$, entonces tenemos que $y = xa$ y $x = yb$ para algunos $a, b \in B_n^+$, así que $y = yba$. Como en el monoide de trenzas positivas las relaciones son homogéneas, la longitud de las palabras dentro de una clase de equivalencia es constante , pero $\text{long}(yab) = \text{long}(y) + \text{long}(a) + \text{long}(b)$, por lo que tenemos necesariamente que $a = b = 1$, de donde $x = y$, cumpliéndose la propiedad antisimétrica. Por último, supongamos que $x \preccurlyeq y \preccurlyeq z$ para ciertas $x, y, z \in B_n^+$. Entonces $y = xa$ y $z = yb$ para $a, b \in B_n^+$. Sustituyendo, $z = xab$, por lo que $x \preccurlyeq z$, lo que prueba la propiedad transitiva. ∎

Nótese que $\preccurlyeq$ es invariante por multiplicación a izquierda, esto es, $a \preccurlyeq b$ implica $xa \preccurlyeq xb$ para todo $a, b, x \in B_n^+$.

Dado tal orden parcial, uno podría preguntarse si existe un único máximo común divisor o mínimo común múltiplo con respecto a $\preccurlyeq$. Esto es, dadas $a, b \in B_n^+$, ¿existe un único $d \in B_n^+$ tal que $d \preccurlyeq a$, $d \preccurlyeq b$ y $d' \preccurlyeq d$ para todo d' prefijo común de a y b? ¿Y existe un único $m \in B_n^+$ tal que $a \preccurlyeq m$, $b \preccurlyeq m$ y $m \preccurlyeq m'$ para todo m' que tenga a a y a b como prefijos? En tales casos, escribimos $d = a \wedge b$ y $m = a \vee b$. Nótese que también tendríamos $xd = xa \wedge xb$ y $xm = xa \vee xb$ para todo $x \in B_n^+$.

Nota 31. Análogamente podríamos definir el orden parcial de **sufijos**, $\succcurlyeq$, invariante por multiplicación a derecha. Nótese que este orden no es equivalente al de prefijos, puesto que $b \succcurlyeq a$ no implica en general $a \preccurlyeq b$ ni recíprocamente. Por ejemplo, $\sigma_1 \preccurlyeq \sigma_1\sigma_2$, pero claramente $\sigma_1\sigma_2 \not\succcurlyeq \sigma_1$. ◂

Proposición 32 ([12, Teorema 1.2]). *El mínimo común múltiplo de los generadores σ_i y σ_j viene dado por*

$$\sigma_i \vee \sigma_j = \begin{cases} \sigma_i\sigma_j & si\ |i-j| > 1, \\ \sigma_i\sigma_j\sigma_i & si\ |i-j| = 1. \end{cases}$$

En la prueba original se prueba también que B_n^+ es cancelativo, es decir, $xay = xby$ implica $a = b$ para todo $a, b, x, y \in B_n^+$. Este resultado permite probar que todo par de elementos de B_n^+ tiene un único mínimo común múltiplo y un único máximo común divisor, tal como hace Dehornoy [9].

Como las relaciones de (1) son homogéneas, palabras equivalentes en B_n^+ tienen la misma longitud, por lo que la longitud de una trenza positiva se define como la longitud de cualquier palabra que la represente.

Garside después estudia el siguiente elemento especial.

Definición 33. La **trenza fundamental** de n cuerdas es la trenza

$$\Delta_n = \sigma_1(\sigma_2\sigma_1)\cdots(\sigma_{n-1}\cdots\sigma_1).$$

Cuando n se sobreentiende, escribimos simplemente Δ. ◂

Proposición 34. *Se verifican las siguientes propiedades:*

1. $\Delta = \sigma_1 \vee \ldots \vee \sigma_{n-1}$ [12, Lema 1].
2. $\sigma_i\Delta = \Delta\sigma_{n-i}$ *para todo* $i = 1, \ldots, n-1$ [12, Lema 4].

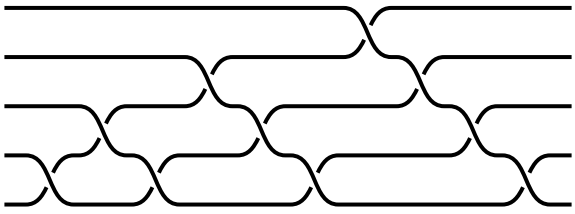


Figura 11: La trenza fundamental Δ_5.

Lema 35. *Se tiene para $k > i \geq 1$ que* $\sigma_i(\sigma_k \cdots \sigma_1) = (\sigma_k \cdots \sigma_1)\sigma_{i+1}$.

Demostración. Usando las relaciones del monoide de trenzas positivas,

$$\begin{aligned} \sigma_i(\sigma_k \cdots \sigma_1) &= \sigma_i(\sigma_k \cdots \sigma_{i+2})(\sigma_{i+1}\sigma_i)(\sigma_{i-1} \cdots \sigma_1) \\ &= (\sigma_k \cdots \sigma_{i+2})(\sigma_i\sigma_{i+1}\sigma_i)(\sigma_{i-1} \cdots \sigma_1) \\ &= (\sigma_k \cdots \sigma_{i+2})(\sigma_{i+1}\sigma_i\sigma_{i+1})(\sigma_{i-1} \cdots \sigma_1) \\ &= (\sigma_k \cdots \sigma_{i+2})(\sigma_{i+1}\sigma_i)(\sigma_{i-1} \cdots \sigma_1)\sigma_{i+i} \\ &= (\sigma_k \cdots \sigma_1)\sigma_{i+1}. \end{aligned}$$

■

A partir de este resultado podemos deducir las siguientes propiedades sobre Δ.

Proposición 36. *Se cumplen las siguientes propiedades:*

1. $\sigma_1, \dots, \sigma_{n-1}$ *son también sufijos de* Δ.
2. Δ^2 *conmuta con todo elemento de* B_n^+.
3. *Para todo* $a \in B_n^+$ *se tiene que* $a \preccurlyeq \Delta^m$ *y* $\Delta^m \succcurlyeq a$, *donde* $m \geq 0$ *es la longitud de* a.

Demostración.

1. En primer lugar, por el lema 35 se tiene para $k > i \geq 1$ que $\sigma_i(\sigma_k \cdots \sigma_1) = (\sigma_k \cdots \sigma_1)\sigma_{i+1}$. Así pues, para expresar σ_i como sufijo de Δ hacemos lo siguiente. Si $i = 1$, entonces por la definición de Δ ya tenemos que es un sufijo. Si $1 < i \leq n-1$, partimos de
$$\Delta = \sigma_1(\sigma_2\sigma_1)\cdots(\sigma_{n-i}\dots\underline{\sigma_1})\cdots(\sigma_{n-1}\dots\sigma_1).$$
Procedemos a desplazar a la derecha el σ_1 subrayado tal como hemos hecho anteriormente. En cada paso irá aumentando el índice en una unidad. Por tanto, como hay $n-(n-i-1) = i+1$ bloques que se dejan atrás, obtenemos $\sigma_{1+i-1} = \sigma_i$, es decir,
$$\begin{aligned}\Delta &= \sigma_1(\sigma_2\sigma_1)\cdots(\sigma_{n-i}\dots\underline{\sigma_1})(\sigma_{n-i+1}\cdots\sigma_1)\cdots(\sigma_{n-1}\dots\sigma_1)\\ &= \sigma_1(\sigma_2\sigma_1)\cdots(\sigma_{n-i}\cdots\sigma_2)(\sigma_{n-i+1}\cdots\sigma_1\underline{\sigma_2})\cdots(\sigma_{n-1}\dots\sigma_1)\\ &= \sigma_1(\sigma_2\sigma_1)\cdots(\sigma_{n-i}\cdots\sigma_2)(\sigma_{n-i+1}\cdots\sigma_1)\cdots(\sigma_{n-1}\dots\sigma_1)\underline{\sigma_i}.\end{aligned}$$
2. Basta probar que Δ^2 conmuta con σ_i para todo $1 \leq i \leq n-1$. Como $\sigma_i\Delta = \Delta\sigma_{n-i}$ y $\sigma_{n-i}\Delta = \Delta\sigma_i$, se tiene que
$$\sigma_i\Delta^2 = \Delta\sigma_{n-i}\Delta = \Delta^2\sigma_i.$$
3. Probamos que $a \preccurlyeq \Delta^m$, donde m es la longitud de a, por inducción en m. Evidentemente, $1 \preccurlyeq \Delta^0 = 1$. Para una palabra de longitud 1 también está claro, porque $\sigma_i \preccurlyeq \Delta$ para todo $1 \leq i \leq n-1$ por definición de mínimo común múltiplo. Supongamos ahora que, para una palabra $a \in B_n^+$ de longitud $m-1$, se tiene el resultado. Entonces, cualquier palabra de longitud m será de la forma $\sigma_j a$ para algún $1 \leq j \leq n-1$. Así que, usando la invarianza por multiplicación a izquierda y el caso $m = 1$,
$$a \preccurlyeq \Delta^{m-1} \implies \sigma_j a \preccurlyeq \sigma_j\Delta^{m-1} = \Delta^{m-1}\sigma_t \preccurlyeq \Delta^{m-1}\Delta = \Delta^m,$$
donde o bien $t = j$, o bien $t = n-j$, dependiendo de la paridad de m. De forma análoga, usando la invarianza por multiplicación a derecha se prueba que $\Delta^m \succcurlyeq a$. ∎

Esto tiene importantes implicaciones. Como todo par de elementos de B_n^+ tiene un múltiplo común y B_n^+ es cancelativo, las condiciones de Ore (definición 13) implican que B_n^+ se inyecta en su grupo de fracciones, que es precisamente B_n. Por lo tanto, B_n^+ no es solamente un monoide definido algebraicamente, sino que puede ser considerado como un submonoide de B_n formado por las trenzas que pueden ser escritas solo con potencias positivas de los generadores.

Las propiedades anteriores implican que el orden parcial $\preccurlyeq$ (respectivamente, $\succcurlyeq$) puede ser extendido a B_n de la siguiente manera: dadas $a, b \in B_n$, $a \preccurlyeq b$ (respectivamente, $a \succcurlyeq b$) si $ac = b$ (respectivamente, $b = ca$) para algún $c \in B_n^+$. Esto da un orden parcial que es invariante por multiplicación a izquierda (respectivamente, a derecha), y el cual admite un único mínimo común múltiplo y un único máximo común divisor. Este hecho podrá ser probado una vez definida la *forma normal de Garside* en la sección a continuación.

4.2. Solución al problema de la palabra

Garside dio una nueva solución al problema de la palabra en los grupos de trenzas de la siguiente manera. Recordemos que para todo $i = 1, \dots, n-1$ se tiene que $\Delta \succcurlyeq \sigma_i$ por la proposición 36, apartado 1, esto es, $\Delta = X_i\sigma_i$ para algún $X_i \in B_n^+$. Dada una trenza escrita como una palabra en $\sigma_1, \dots, \sigma_{n-1}$ y sus inversos, se puede reemplazar cada aparición de σ_i^{-1} por $\Delta^{-1}X_i$. Conjugar una trenza positiva por Δ sigue dando una

trenza positiva por la proposición 34, apartado 2, así que podemos mover todas las apariciones de Δ^{-1} a la izquierda, de la siguiente forma: si encontramos $\sigma_j\Delta^{-1}$ $(1 \leq j \leq n-1)$, entonces por la proposición 34 sabemos que $\Delta\sigma_j = \sigma_{n-j}\Delta$, si y solo si $\sigma_j\Delta^{-1} = \Delta^{-1}\sigma_{n-j}$, por lo que podemos sustituir $\sigma_j\Delta^{-1}$ por $\Delta^{-1}\sigma_{n-j}$. Esto muestra que toda trenza puede ser escrita como $\Delta^p A$ para algún $p \in \mathbb{Z}$ y algún $A \in B_n^+$. Además, si $\Delta \preccurlyeq A$, podemos reemplazar Δ^p por Δ^{p+1} y A por $\Delta^{-1}A$. Esto reduce la longitud de A, así que solo puede hacerse una cantidad finita de veces. Por tanto, toda trenza puede descomponerse *de manera única*, como $\Delta^p A$, donde $p \in \mathbb{Z}$, $A \in B_n^+$ y $\Delta \not\preccurlyeq A$. Efectivamente, si tuviéramos dos expresiones $\Delta^p A = \Delta^q B$ con $p < q$ en las condiciones anteriores, dividiendo por Δ^p tendríamos que $A = \Delta^{q-p}B$, lo cual contradice el hecho de que A no tenga a Δ como prefijo. Análogamente para $p > q$, luego $p = q$ y $A = B$.

Definición 37. En base a lo comentado en el párrafo anterior, definimos la **forma normal de Garside** de una palabra $w \in B_n$ como $w = \Delta^p A$, donde $p \in \mathbb{Z}$, $A \in B_n^+$ y $\Delta \not\preccurlyeq A$. ◂

Esta forma normal permite resolver el problema de la palabra, ya que se pueden enumerar todas las palabras positivas que representan la trenza positiva A reiterando las relaciones del monoide de trenzas positivas de todas las formas posibles. Esta fue la solución dada por Garside [12]. Sin embargo, no es muy satisfactoria, ya que da lugar a un algoritmo altamente ineficiente.

Elrifai y Morton [10] lo mejoraron definiendo la **forma normal a la izquierda** de una trenza. Basta tomar la descomposición $\Delta^p A$ y después definir

$$\begin{aligned} a_1 &= A \wedge \Delta, \\ a_i &= (a_{i-1}^{-1} \cdots a_1^{-1} A) \wedge \Delta, \ \forall\, i > 1. \end{aligned}$$

Nótese que existe un $r \geq 0$ tal que $a_i = 1$ para todo $i > r$, ya que la longitud de $a_{i-1}^{-1} \cdots a_1^{-1}A$ es estrictamente decreciente. De esta forma, toda trenza puede ser escrita de manera única como

$$\Delta^p a_1 \cdots a_r,$$

donde los a_i son los definidos anteriormente, los cuales por definición son unos prefijos propios de Δ, es decir, $1 \prec a_i \prec \Delta$, y además se puede demostrar que $(a_i a_{i+1}) \wedge \Delta = a_i$ para todo $i = 1, \ldots, r-1$ [11]. Esta es la anteriormente mencionada forma normal a la izquierda de la trenza. Los prefijos positivos de Δ son llamados **trenzas simples** o **trenzas de permutación**. El nombre no es casual, ya que, como prueba Thurston [11], estas trenzas representan en cierto modo a las permutaciones que inducen. Estas trenzas se caracterizan por ser aquellas en las que los generadores σ_i aparecen con exponente no superior a 1 en todas sus expresiones en términos de los generadores de Artin. Por tanto, la forma normal a la izquierda de una trenza es una descomposición única como producto de una potencia de Δ y una sucesión de elementos simples propios. Thurston [11] mostró que esta forma normal puede ser calculada en tiempo $O(\ell^2 n \log n)$ para una palabra de ℓ letras en B_n.

En [11] se puede encontrar además una forma más práctica de llevar a cabo el algoritmo de encontrar la forma normal a la izquierda, la cual utilizareos en el ejemplo 38. Antes de explicarla, vamos a introducir algo de nomenclatura. Dadas dos trenzas simples positivas A y B, decimos que un prefijo no trivial $b \preccurlyeq B$ **se puede pasar** de B a A si Ab es simple y, en tal caso, **pasar** b de B a A consiste en realizar las transformaciones $A \to Ab$ y $B \to b^{-1}B$. Con esto presente, el algoritmo consiste en lo siguiente:

1. Una vez tenemos una palabra $w \in B_n$ en forma normal de Garside $w = \Delta^p A$, si $A = 1$, entonces no hay nada que hacer. En caso contrario, dividimos A en bloques formados por elementos simples, digamos,
$$A = a_{1,0}a_{2,0} \ldots a_{m,0}.$$
2. En el paso $t \geq 0$ tenemos A expresada en bloques de elementos simples como
$$A = a_{1,t}a_{2,t} \ldots a_{m,t}.$$
En este paso buscamos el primer par $a_{i,t}a_{i+1,t}$ de modo que se pueda pasar algún prefijo de $a_{i+1,t}$ a $a_{i,t}$ y lo pasamos. Esto nos dará la descomposición
$$A = a_{1,t+1}a_{2,t+1} \ldots a_{m,t+1}.$$
3. Volvemos al paso 2 y reiteramos hasta que no quede ningún par que verifique la condición.

Este proceso, naturalmente, termina, porque el vector formado por las longitudes de los bloques aumenta en cada paso su orden lexicográfico, el cual está acotado por $(m, 0, \dots, 0)$, donde m es la longitud de A. La forma normal a la izquierda se obtendrá eliminando los bloques triviales (que necesariamente estarán al final).

Alternativamente, podríamos empezar con una descomposición $w = \Delta^q A$ con $A \in B_n^+$, pero sin asegurarnos de que $\Delta \not\preccurlyeq A$, pues Δ aparecería al acumular elementos simples en caso de ser prefijo de A, y podríamos enviarlo al bloque de Δ^q. En cualquier caso, este proceso acabará con la forma normal a la izquierda, pues no poder pasar ninguna letra del bloque a_{i+1} al bloque a_i es equivalente a que $a_i = (a_i a_{i+1}) \wedge \Delta$.

Ejemplo 38. En B_4, sean $\alpha_1 = \sigma_1\sigma_2^{-1}\sigma_3$ y $\alpha_2 = \sigma_3\sigma_1\sigma_1\sigma_2\sigma_1$. Queremos comprobar si α_1 y α_2 representan el mismo elemento. Lo primero que debemos hacer es eliminar el exponente negativo de α_1. Para ello, tenemos que expresar $\Delta = \Delta_4 = \sigma_1(\sigma_2\sigma_1)(\sigma_3\sigma_2\sigma_1)$ de forma que tenga a σ_2 como sufijo. Esto es sencillo, pues basta usar la técnica de la demostración del primer apartado de la proposición 36 para escribir

$$\Delta = \sigma_1(\sigma_2)(\sigma_3\sigma_2\sigma_1)\sigma_2.$$

Así pues, $\sigma_2^{-1} = \Delta^{-1}\sigma_1\sigma_2\sigma_3\sigma_2\sigma_1$, de modo que $\alpha_1 = \sigma_1\Delta^{-1}\sigma_1\sigma_2\sigma_3\sigma_2\sigma_1\sigma_3$. Usando la proposición 34, pasamos Δ^{-1} a la izquierda:

$$\alpha_1 = \Delta^{-1}\sigma_3\sigma_1\sigma_2\sigma_3\sigma_2\sigma_1\sigma_3.$$

Ahora vamos a hacer la separación de bloques en las palabras positivas. Empezamos con α_2. Vamos a dividirla en los bloques $b_{1,0} = \sigma_3\sigma_1$ y $b_{2,0} = \sigma_1\sigma_2\sigma_1$, que son claramente trenzas simples. En general, se puede comenzar por bloques de una sola letra para evitar esta verificación. Así, obtenemos

$$\alpha_2 = b_{1,0}b_{2,0} = (\sigma_3\sigma_1)(\sigma_1\sigma_2\sigma_1).$$

Aparentemente no podemos pasar ninguna letra de $b_{2,0}$ a $b_{1,0}$, pues aparecería σ_1 dos veces seguidas. Sin embargo, recordemos que las relaciones de (1) nos dan $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$. Por lo tanto, reescribimos α_2 y continuamos con

$$\alpha_2 = b_{1,1}b_{2,1} = (\sigma_3\sigma_1\sigma_2\sigma_1)(\sigma_2).$$

Ahora tenemos la situación inversa: aparentemente, podríamos añadir σ_2 al primer bloque, pero utilizando la misma relación de la presentación del grupo de trenzas que antes, nos aparecería σ_2 dos veces consecutivas, por lo que hemos finalizado el proceso y $\alpha_2 = \Delta^0 b_1 b_2$ con $b_1 = b_{1,1}$ y $b_2 = b_{2,1}$. Obsérvese que el bloque que hemos pasado a la izquierda $(\sigma_2\sigma_1)$ se corresponde con $b_{2,0} \wedge (b_{1,0}^{-1}\Delta)$ y el bloque resultante $(\sigma_3\sigma_1\sigma_2\sigma_1)$ se corresponde con $\alpha_2 \wedge \Delta$ en el algoritmo original de Elrifai y Morton. Además, esta claro que ninguno de los factores es una potencia de Δ. ◂

Referencias

[1] Aguilar Martín, Javier. *El problema de la palabra en los grupos de trenzas*. Trabajo de Fin de Grado. Universidad de Sevilla, 2018. URL: https://hdl.handle.net/11441/77489.

[2] Aluffi, Paolo. *Algebra: Chapter 0*. Vol. 104. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2009. https://doi.org/10.1090/gsm/104.

[3] Artin, Emil. «Theorie der Zöpfe». En: *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* 4.1 (1925), págs. 47-72. ISSN: 0025-5858. https://doi.org/10.1007/BF02950718.

[4] Artin, Emil. «Theory of braids». En: *Annals of Mathematics. Second Series* 48 (1947), págs. 101-126. ISSN: 0003-486X. https://doi.org/10.2307/1969218.

[5] Birman, Joan S. *Braids, Links, and Mapping Class Groups*. Vol. 82. Annals of Mathematics Studies. Princeton: Princeton University Press, 2016. https://doi.org/10.1515/9781400881420.

[6] Clifford, Alfred H. y Preston, Gordon B. *The algebraic theory of semigroups. Vol. I*. Mathematical Surveys, No. 7. Providence, Rhode Island: American Mathematical Society, 1961. ISBN: 978-0-8218-0272-4.

[7] Dehn, Max. «Über unendliche diskontinuierliche Gruppen». En: *Mathematische Annalen* 71.1 (1911), págs. 116-144. issn: 0025-5831. https://doi.org/10.1007/BF01456932.

[8] Dehn, Max. «Transformation der Kurven auf zweiseitigen Flächen». En: *Mathematische Annalen* 72.3 (1912), págs. 413-421. issn: 0025-5831. https://doi.org/10.1007/BF01456725.

[9] Dehornoy, Patrick. «Groupes de Garside». En: *Annales Scientifiques de l'École Normale Supérieure. Quatrième Série* 35.2 (2002), págs. 267-306. issn: 0012-9593. https://doi.org/10.1016/S0012-9593(02)01090-X.

[10] Elrifai, Elsayed A. y Morton, Hugh R. «Algorithms for positive braids». En: *The Quarterly Journal of Mathematics. Oxford. Second Series* 45.180 (1994), págs. 479-497. issn: 0033-5606. https://doi.org/10.1093/qmath/45.4.479.

[11] Epstein, David B. A.; Cannon, James W.; Holt, Derek F.; Levy, Silvio V. F.; Paterson, Michael S., y Thurston, William P. *Word Processing in Groups*. Boca Raton, Florida: CRC Press, 1992. isbn: 978-0-86720-244-1.

[12] Garside, Frank A. «The braid group and other groups». En: *The Quarterly Journal of Mathematics. Oxford. Second Series* 20 (1969), págs. 235-254. issn: 0033-5606. https://doi.org/10.1093/qmath/20.1.235.

[13] González-Meneses, Juan y Silvero, Marithania. «Polynomial braid combing». En: *Mathematics of Computation* 88.318 (2019), págs. 2027-2045. issn: 0025-5718. https://doi.org/10.1090/mcom/3392.

[14] Hurwitz, Adolf. «Ueber Riemann'sche Flächen mit gegebenen Verzweigungspunkten». En: *Mathematische Annalen* 39.1 (1891), págs. 1-60. issn: 0025-5831. https://doi.org/10.1007/BF01199469.

[15] Lyndon, Roger C. y Schupp, Paul E. *Combinatorial Group Theory*. Vol. 89. Classics in Mathematics. Berlin, Heidelberg: Springer, 2001. https://doi.org/10.1007/978-3-642-61896-3.

[16] Magnus, Wilhelm. «Über Automorphismen von Fundamentalgruppen berandeter Flächen». En: *Mathematische Annalen* 109.1 (1934), págs. 617-646. issn: 0025-5831. https://doi.org/10.1007/BF01449158.

[17] Markoff Jr., Andrey. «Foundations of the algebraic theory of tresses». En: *Russian mathematicians in the 20th century* (2003), págs. 614-621.

[18] Novikov, Petr Sergeevič. *On the algorithmic unsolvability of the word problem in group theory*. Vol. 44. Trudy Matematicheskogo Instituta imeni V. A. Steklova. Moscú: Izdatelstvo Akademii Nauk SSSR, 1955. url: http://mi.mathnet.ru/eng/tm1180.

[19] Ore, Oystein. «Linear equations in non-commutative fields». En: *Annals of Mathematics. Second Series* 32.3 (1931), págs. 463-477. issn: 0003-486X. https://doi.org/10.2307/1968245.

TEMat

Cuestiones existenciales en combinatoria y teoría de números: el método probabilístico

Ismael Morales López
Universidad Autónoma de Madrid
ismael.moralesl@estudiante.uam.es

Resumen: La probabilidad es una rama de las matemáticas indispensable en la formulación de muchos fenómenos físicos y, en general, de procesos que contengan algún tipo de arbitrariedad. Un aspecto menos conocido de esta es el poder que puede llegar a tener en cuestiones de naturaleza discreta.

El método probabilístico es una herramienta que parte de una idea muy limpia y prometedora. Con el fin de demostrar la existencia de un objeto C caracterizado por una determinada propiedad P, se embebe C en un espacio de probabilidad y se demuestra que el suceso correspondiente a tener la propiedad P ocurre con probabilidad positiva.

El objetivo de este artículo es sentar la base teórica tras la cual subyace esta técnica y presentar varias aplicaciones en problemas relacionados con la combinatoria y la teoría de números.

Abstract: Probability is an indispensable branch of mathematics when it comes to formulating many physical phenomena and, in general, processes which contain some arbitrariness. A less-known aspect is the power it can achieve when tackling questions of discrete nature.

The probabilistic method is a tool based on a neat and promising idea. With the aim of proving the existence of an object C characterised by some property P, C is embedded in a probability space and it is proved that the event corresponding to having property P holds with positive probability.

The objective of this paper is to set the theoretical background that rests behind this technique and to present some applications in problems with a combinatorial or number-theoretic flavour.

Palabras clave: espacio y función de probabilidad, independencia entre sucesos, elección aleatoria y uniforme, distribución, variable aleatoria, esperanza, varianza, método probabilístico, propiedad local, comportamiento asintótico.

MSC2010: 05C15, 05D40, 11E99.

Recibido: 6 de mayo de 2019.
Aceptado: 27 de abril de 2020.

Referencia: Morales López, Ismael. «Cuestiones existenciales en combinatoria y teoría de números: el método probabilístico». En: *TEMat*, 4 (2020), págs. 43-65. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p43.

1. Introducción histórica y motivación

La probabilidad es una de las áreas de las matemáticas que crece a mayor velocidad. Por un lado, ha resultado ser imprescindible en la formulación matemática de nociones como arbitrariedad e incertidumbre en ramas científicas como la informática teórica y la física estadística. También ha sido crucial en el desarrollo de la combinatoria y la teoría de grafos, aunque ciertamente no se puede decir que estos avances fuesen esperados. Al menos no eran parte de las inquietudes reflejadas en el sexto de los 23 problemas que planteó Hilbert para la conferencia de París del *Congreso Internacional de Matemáticos* de 1900, en el que se hablaba de la necesidad de axiomatizar la teoría de la probabilidad.

La influencia de la probabilidad en combinatoria, teoría de grafos y otras ramas como teoría de números, geometría convexa y finita o combinatoria aditiva se debe al método que da nombre al artículo. Comienza siendo un argumento muy concreto pero admite diferentes niveles de sofisticación y generalidad. Vagamente, consiste en demostrar que entre una colección de objetos existe al menos uno, digamos C, con una propiedad P determinada. Para ello, estos objetos pasarán a ser los elementos de un cierto espacio de probabilidad en el cual el suceso correspondiente a tener la propiedad P ocurre con probabilidad positiva, garantizando la existencia de C.

Sin embargo, cuando se consideran versiones más generales y sofisticadas de este argumento, es mucho más difícil entender el fondo, como ocurre con la demostración del teorema de Green-Tao [24], el cual afirma que existen progresiones aritméticas de primos arbitrariamente largas. Habiendo enunciado el resultado anterior, conviene recalcar que la probabilidad parece ser una pieza fundamental en el estudio de los números enteros. Este es un hecho tan sorprendente como el papel que juega el análisis complejo, inevitable en cualquier demostración del teorema de los números primos. Pero no nos perdamos en nuestro afán de conocimiento y remontémonos a mediados del siglo XX para exponer un problema de combinatoria que servirá para explicar el origen del método probabilístico en su versión más simple.

Definición 1. Dados dos enteros positivos n, m, definimos el **número de Ramsey** $R(n, m)$ como el mínimo entero positivo R tal que, para toda coloración de las aristas de un grafo R-completo (véase la definición 11) con los colores azul y rojo, existe un subgrafo n-completo de aristas rojas o un subgrafo m-completo de aristas azules. ◂

Para que lo anterior sea una definición, hay que demostrar que tales números existen. El lector puede encontrar una demostración de este hecho en el libro de Bollobás [6, capítulo 6], donde se prueba que

$$R(n, m) \leq R(n, m-1) + R(n-1, m), \tag{1}$$

que, por las relaciones de recurrencia que satisfacen los coeficientes binomiales, nos lleva a que

$$R(n, m) \leq \binom{n+m-2}{m-1} \quad \text{si } n, m \geq 2. \tag{2}$$

Cuando $n = m = k$, se tienen los números de Ramsey diagonales $R(k, k)$. Se comprueba fácilmente que $R(1, 1) = 1$, $R(2, 2) = 2$ y $R(3, 3) = 6$. Sin embargo, en general no es sencillo calcular explícitamente estos valores. El valor de $R(4, 4) = 18$ también puede hallarse a mano, pero el de $R(5, 5)$ se desconoce por el momento: solo se conoce la estimación $43 \leq R(5, 5) \leq 48$ [44]. Como consecuencia de la ecuación (2) en el caso $n = m = k$, se tiene que

$$R(k, k) < 4^{k-1} \quad \text{si } k \geq 2. \tag{3}$$

La demostración de la ecuación (1) es inductiva y solo hace uso del principio del palomar. De hecho, la misma prueba proporciona incluso un algoritmo para encontrar un subgrafo completo monocromático de manera recursiva. Este argumento aparece en el artículo original de Ramsey [31], donde se introducen versiones más generales del problema que estamos discutiendo con la motivación de encontrar un algoritmo que determine la validez de ciertas fórmulas de una lógica de primer orden. Por otro lado, en este artículo se plantea probar una cota inferior explícita para los números de Ramsey, concretamente que

$$2^{k/2} \leq R(k, k) \quad \text{si } k \geq 2, \tag{4}$$

en el ejercicio 4. Lo interesante es que la demostración es de naturaleza existencial y no produce explícitamente la coloración de un grafo de al menos $2^{k/2}$ vértices sin subgrafos k-completos monocromáticos. En efecto, se colorean las aristas uniformemente al azar y se prueba que el grafo verificará esta propiedad con probabilidad positiva, que corresponde al esquema de demostración que seguiremos en las secciones 3 y 6.

Esta diferencia entre las demostraciones de las ecuaciones (3) y (4) también explica el nombre del artículo, porque las cuestiones que tratamos son únicamente existenciales, en contraposición con las constructivas.

La cota de la ecuación (4) fue descubierta por Erdős en 1947 [16]. Comenzamos hablando de Erdős porque es considerado el pionero de lo llamado posteriormente *método probabilístico*. Bien es cierto que Szele [37] aplicó un argumento probabilístico a un problema de combinatoria en un artículo de 1943 (considerada la primera aplicación del método probabilístico en combinatoria) y que, como analizaremos en la sección 5, Turán [40] también habría empleado técnicas de probabilidad para probar un resultado de teoría analítica de números en 1934, el cual trataremos con detalle en el teorema 21. Sin embargo, es Erdős quien entendió de verdad su potencia y lo aplicó recurrentemente en múltiples resultados. Alon [1] explicó por qué considera que esta es una de las mayores contribuciones de Erdős y añade que él siempre estaba más interesado en discutir nuevos problemas que en evaluar el mérito que tendrían a largo plazo sus resultados. Por ello, resalta que durante la celebración de su 80.º cumpleaños en Keszthely, Hungría, Erdős dijera que creía que esta técnica viviría mucho después de él.

Una objeción de carácter técnico sobre este método puede ser que en algunos casos parece superfluo enfocar el problema desde un punto de vista probabilístico. De hecho, al menos en las demostraciones de las secciones 3 y 4, para probar la existencia de una configuración con una característica determinada, se prueba que la cantidad de configuraciones sin dicha propiedad es menor que el número total de configuraciones posibles, forzando la existencia de al menos una con el requerimiento deseado. Además, en esto se basa la proposición 8, el único ingrediente de dichas secciones. Sin embargo, el lenguaje probabilístico permite simplificar los cálculos y, más importante, resulta inevitable para diseñar un método más potente con condiciones más técnicas de independencia como las del lema local de Lovász, discutido en la sección 6. La justificación para esto último es que la noción de independencia se reconoce fácilmente desde el punto de vista de la combinatoria pero se explota propiamente con las técnicas probabilísticas. Además, aunque no es el caso de este artículo, a veces conviene considerar distribuciones no uniformes, como puede verse en el libro de Alon y Spencer [3, teorema 3.2.1, pág 29, y teorema 1, pág 41].

Por otro lado, aunque por simplicidad restrinjamos nuestra exposición a cuestiones que involucren o permitan reducciones a espacios de probabilidad finitos, no todos los objetos sobre los que podemos aplicar esta técnica son de naturaleza discreta. En efecto, hay espacios mucho más complejos de naturaleza analítica o geométrica que vienen equipados de una medida gracias a la cual pueden recrearse estos argumentos de tipo probabilístico. Hay una rica colección en Wikipedia [41] donde, por ejemplo, se mencionan resultados tan icónicos como el teorema fundamental del álgebra, el teorema de Picard y el teorema de aproximación de Weierstrass. Estos ejemplos de teoremas demostrables con técnicas de probabilidad no fueron probados por primera vez así, sino que estas demostraciones vinieron *a posteriori*. Por el contrario, hay muchos ejemplos de objetos en matemáticas cuya existencia se prueba primero por medio de un argumento de este estilo antes de poder encontrar ejemplos concretos (más discusión en el libro de Alon y Spencer [3, capítulo 9]). Este es el caso de los «*expanders*», un tipo de grafos de gran interés en áreas ligadas a la informática como la teoría de códigos [35], de los cuales no se encontraron construcciones explícitas hasta 1973 [26]. De todas formas, aún en caso de solo haber probado la existencia de un cierto objeto por medio de estos argumentos, en general estos pueden dar lugar a algoritmos probabilísticos efectivos que puedan además desaleatorizarse para construir un tal objeto de manera determinista. Este enfoque práctico, discutido también por Alon y Spencer [3, capítulo 6], solo volverá a mencionarse en contextos más concretos al final de la sección 6.

Habiendo quedado clara la importancia del enfoque probabilístico, hacemos un pequeño resumen de la exposición. En la siguiente sección recordaremos algunas definiciones básicas. Aunque el artículo sea prácticamente autocontenido, es importante tener cierta familiaridad con el cálculo de probabilidades en espacios finitos y, en particular, con el significado de los coeficientes binomiales. Para adquirir o recordar estas técnicas de combinatoria se recomiendan los apuntes de Fernández y Fernández [22], por la selección de ejemplos tratados y su cautivadora redacción. Las distribuciones consideradas son siempre uniformes y, por tanto, los cálculos de probabilidades se harán empleando la regla de Laplace y, a lo largo del artículo, las manipulaciones involucrarán identidades que enunciaremos con antelación, como la proposición 7 o el lema 18. En esta sección 2 también se enuncia el resultado fundamental, recogido en la proposición 8, sobre el cual se apoya la primera versión que desarrollaremos sobre el método probabilístico y que se empleará en las secciones 3 y 4 para resolver problemas relacionados con la teoría de grafos y la combinatoria aditiva.

En la sección 5 se estudiará una aplicación a la teoría de números. Se considera la función ν, donde $\nu(m)$ es la cantidad de divisores primos de m. A través de sus propiedades aritméticas, se estudian la esperanza y la varianza de esta variable aleatoria en segmentos $\{1, \dots, n\}$ de $\mathbb{N}$ para obtener información sobre su distribución cuando $n \to \infty$ por medio de un argumento de Turán. Se cierra la sección con el enunciado del teorema de Erdős-Kac, que describe completamente cómo se distribuye ν, y después se discute en qué consiste el método de momentos en el contexto de la demostración de dicho resultado.

Como adelantamos, en la sección 6 se estudiará el lema local de Lovász. Primero se discutirá la motivación, el significado detrás del lema y se ofrecerá una demostración de la versión general de este lema. Después daremos una versión que es más transparente y fácil de usar, desde el punto de vista de las aplicaciones de este artículo, para aplicarlo en problemas relacionados con la coloración de hipergrafos. Uno de estos corolarios es bastante sorprendente ya que la coloración se considera en $\mathbb{R}$ y se conjetura que el uso de argumentos topológicos es irremplazable. En esta sección también se discute brevemente el aspecto algorítmico tanto del lema local de Lovász como de las estimaciones de los números de Ramsey. Por último, en el apéndice A se resuelven todos los problemas planteados a lo largo de la exposición.

La redacción de este artículo y los temas tratados están profundamente influidos por el estilo y las motivaciones del artículo de Chen [11], los apuntes de Loh [25] y Riordan [32] y, especialmente, por los libros de Alon y Spencer [3] y Tao y Vu [39].

2. Fundamentos del método probabilístico

Comenzamos fijando una notación que se empleará en todo el artículo.

Notación. Se denota por $\mathbb{Z}^+$ el conjunto de enteros positivos. Dado $n \in \mathbb{Z}^+$, denotamos por $[n]$ el conjunto de enteros k tales que $1 \leq k \leq n$. ◂

En esta sección fijaremos nociones fundamentales de probabilidad y daremos la primera herramienta con la que trabajar problemas de combinatoria desde un punto de vista probabilístico en la proposición 8. Las definiciones que se presentan suelen aparecer en libros de probabilidad en un contexto más general, lo cual no es necesario para nuestro propósito. Por lo tanto, haremos un desarrollo *ad hoc* de algunos objetos y nociones asociados a espacios de probabilidad finitos. Se recuerda que, dado un conjunto Ω, se denota por $\mathcal{P}(\Omega)$ a la colección de subconjuntos de Ω, que incluye a $\varnothing$ y Ω.

Definición 2. Un **espacio de probabilidad** es un par $(\Omega, \mathbb{P})$, que consta de un conjunto Ω finito (no vacío) denominado **espacio muestral**, y de una función $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$, denominada **función de probabilidad**, que verifica que

- $\mathbb{P}(\Omega) = 1$ y
- dada una colección $\{A_i\}_{i=1}^m$ de subconjuntos de Ω disjuntos entre sí dos a dos, se tiene que

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbb{P}(A_k).$$ ◂

Podemos entender cada uno de los elementos del espacio muestral Ω como un posible resultado de un experimento. Por ejemplo, adelantando el primer ejemplo de la siguiente sección, cada $\omega \in \Omega$ puede ser la relación final entre los jugadores de un torneo en el cual se ha determinado al azar el resultado de cada uno de los enfrentamientos. Un suceso S es la colección de resultados ω con una determinada característica, como puede ser la propiedad de tener un jugador que haya ganado a todos. Con el propósito de esta exposición, haremos la siguiente identificación en el lenguaje: diremos que un resultado $\omega \in \Omega$ de un experimento aleatorio tiene las características descritas por el suceso $S \subseteq \Omega$ si $\omega \in S$.

Otro posible ejemplo es el siguiente. Tomamos un número uniformemente al azar en $\{1, \dots, 100\}$ y consideramos el suceso S que viene dado por la propiedad de ser par. Es decir, $S = \{x \in [100] : x \text{ es par}\} = \{2, 4, \dots, 98, 100\}$ y x verifica la propiedad S si y solo si $x \in S$. En casos tan sencillos parece algo innecesariamente confuso, pero, cuando se traten más ejemplos concretos, se podrá comprobar que este abuso de lenguaje resulta ser en realidad una simplificación. Otra justificación para no referirnos a S como un subconjunto de Ω sino como una propiedad es el hecho de que todos los sucesos que nos interesan aparecerán como propiedades y no como una colección explícita de elementos.

En lo que sigue, todas las definiciones y proposiciones parten de un espacio de probabilidad $(\Omega, \mathbb{P})$.

Definición 3. Una **variable aleatoria** es una función $X\colon \Omega \to \mathbb{R}$. ◂

Obedeciendo a la intuición que hemos presentado sobre cada suceso como una colección de resultados ω con una determinada propiedad, vamos a fijar la siguiente notación estándar.

Notación. Sea P una propiedad que pueda adquirir un número real y sea X una variable aleatoria. El suceso $\{\omega \in \Omega : X(\omega) \text{ verifica } P\}$ se escribirá abreviadamente como $\{X \text{ verifica } P\}$. En particular, si $T \subseteq \mathbb{R}$, entonces el suceso $X^{-1}(T) = \{\omega \in \Omega : X(\omega) \in T\}$ se abreviará por $\{X \in T\}$. ◂

Por tanto, lo anterior nos servirá para introducir sucesos en Ω por medio de una variable aleatoria X, ya que una tal propiedad P induce una propiedad en Ω considerando la preimagen por X.

Definición 4. Decimos que n sucesos $S_1, \dots, S_n$ de Ω son **independientes** si, para todo subconjunto $J \subseteq [n]$,

$$\mathbb{P}\left(\bigcap_{j\in J} S_j\right) = \prod_{j\in J} \mathbb{P}(S_j).$$ ◂

Definición 5. Dado $A \subseteq \Omega$, la variable aleatoria **indicatriz** (o indicadora) de A, denotada por X_A, es

$$X_A(\omega) = \begin{cases} 0 & \text{si } \omega \notin A, \\ 1 & \text{si } \omega \in A. \end{cases}$$ ◂

Es oportuno adelantar que las variables aleatorias que consideraremos serán sumas de indicatrices.

Definición 6. Sea X una variable aleatoria que toma los valores $a_1, \dots, a_n$. Entonces, la **esperanza** de X es

$$\mathbb{E}[X] = \sum_{k=1}^{n} a_k \mathbb{P}(X = a_k).$$ ◂

Proposición 7. *En todo espacio de probabilidad $(\Omega, \mathbb{P})$ se verifican estas propiedades:*

1. (Subaditividad de la función de probabilidad) *Dados los sucesos $A_1, \dots, A_n$, se tiene que*

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mathbb{P}(A_k),$$

   con igualdad si los sucesos son disjuntos dos a dos.
2. (Linealidad de la esperanza) *Dadas $X_1, \dots, X_k$ variables aleatorias, entonces*

$$\mathbb{E}\left[\sum_{k=1}^{n} X_k\right] = \sum_{k=1}^{n} \mathbb{E}[X_k].$$

Proposición 8 (Resultados de existencia). *En todo espacio de probabilidad $(\Omega, \mathbb{P})$ se verifica lo siguiente:*

1. *Dados los sucesos $A_1, \dots, A_n$, si tenemos que $\sum_{k=1}^{n} \mathbb{P}(A_k) < 1$, entonces se tiene que $\bigcup_{k=1}^{n} A_k \neq \Omega$, o, equivalentemente por las leyes de De Morgan, $\bigcap_{k=1}^{n} A_k^c \neq \emptyset$. De hecho, la probabilidad del suceso anterior es positiva. Por ello, en tal caso se dice que los sucesos $A_1, \dots, A_n$ no cubren todo el espacio de probabilidad.*
2. *Sea X una variable aleatoria. Entonces, existe $\omega \in \Omega$ tal que $X(\omega) \geq \mathbb{E}[X]$.*

Observación 9. Con la notación del primer apartado de la proposición 8, los sucesos $A_1, \dots, A_n$ corresponderían en la práctica a sucesos «malos» que queremos evitar. Así, la proposición 8 nos asegura bajo esa hipótesis que, con probabilidad positiva, ninguno de los sucesos «malos» ocurre. ◂

Observación 10. En el segundo punto de la proposición 8 también puede garantizarse, por razones análogas, la existencia de un elemento $\omega' \in \Omega$ tal que $X(\omega') \leq \mathbb{E}[X]$. ◂

Es conveniente remarcar sobre la proposición 8 que el segundo resultado implica el primero. Aún así, es preferible separarlos en las aplicaciones, ya que el primero refleja un objetivo que volverá a aparecer cuando tratemos el lema local de Lovász en la sección 6 y que consiste en buscar condiciones suficientes bajo las cuales varios sucesos $A_1, A_2, \dots, A_n$ no cubren todo el espacio de probabilidad ambiente. El lema anterior mejora este punto de la proposición 8 complementándolo con ciertas condiciones de independencia entre los sucesos $A_1, A_2, \dots, A_n$ involucrados.

Por otro lado, el segundo resultado permite deducir información sobre la distribución de una variable aleatoria por medio de su esperanza, un método que emplearemos en la sección 4 y que se complementará en la sección 5 con el estudio de la varianza.

Para comprobar que el segundo enunciado de la proposición 8 implica el primero, basta considerar una variable aleatoria contador X que indique cuántos de los sucesos A_i se cumplen para cada uno de los $\omega \in \Omega$ y aplicar la observación 10. Tomando $X = \sum_{k=1}^{n} X_{A_k}$ se observa, por la proposición 7, que la media de X es $\mathbb{E}[X] = \sum_{k=1}^{n} \mathbb{P}(A_k)$ y que un $\omega \in \Omega$ pertenece a $\bigcap_{k=1}^{n} A_k^c$ si y solo $X(\omega) < 1$.

La distinción que hacemos entre estos dos objetos en la proposición 8, una función de probabilidad y una variable aleatoria, también tiene interés desde el punto de vista de la construcción de un espacio de probabilidad. Siempre que hablamos de una elección estamos introduciendo un tal espacio, y cuando añadimos más información sobre nuestra elección (como, por ejemplo, exigir que las elecciones se realicen de manera uniforme o que sean independientes), lo que estamos haciendo es detallar más la información de la función de probabilidad o de las diferentes variables aleatorias de dicho espacio. Por ahora, no es necesario entender este párrafo, pero conviene tenerlo en cuenta cuando veamos ejemplos más concretos, ya que los espacios de probabilidad involucrados no se harán explícitos en el sentido de la definición 2.

3. Demostraciones de existencia directa

Ya tenemos todos los ingredientes para analizar la primera aplicación del método probabilístico, que consistirá en la construcción de un torneo con unas propiedades determinadas. Hay una infinidad de técnicas que se engloban dentro del método probabilístico y aquí pasaremos a analizar la más sencilla de aplicar, que consiste en explotar de manera directa la proposición 8. Aunque es sorprendente la cantidad de problemas que permite resolver, sigue siendo muy primitiva y tiene muchas limitaciones porque se aplica bajo condiciones muy restrictivas. En ese sentido, mejoraremos la técnica en la sección 6 con el lema local de Lovász introduciendo condiciones de independencia.

Recordamos la noción de grafo (simple) porque será útil a lo largo del artículo.

Definición 11. Un **grafo** es un par de conjuntos $G = (V(G), E(G))$, donde $V(G)$ es el conjunto de vértices y $E(G)$ es el conjunto de aristas, tal que $E(G) \subseteq \{\{v_1, v_2\} : v_1 \neq v_2 \in V(G)\}$. Decimos que G es n-completo si $|V(G)| = n$ y el número de aristas es máximo, es decir, $E(G) = \{\{v_1, v_2\} : v_1 \neq v_2 \in V(G)\}$ y $|E(G)| = \binom{n}{2}$. ◂

Un **torneo** de n jugadores consta, primeramente, de un grafo n-completo G. Naturalmente, identificamos los vértices de G con los jugadores del torneo y cada arista $\{i, j\}$ simboliza el enfrentamiento entre los correspondientes jugadores i y j. Para determinar completamente el torneo, se debe apuntar en cada arista $\{i, j\}$ el ganador de ese enfrentamiento (no se admiten empates). Por tanto, para cada $n \geq 1$, es obvio que hay $2^{\binom{n}{2}}$ torneos posibles con n jugadores.

Definición 12. Un torneo de n jugadores cumple la propiedad S_k si $n > k$ y si, para cada colección L de k jugadores, existe otro jugador fuera de L que ha ganado a todos los jugadores de L. ◂

Según cuenta Erdős [17], Schütte le plantea la pregunta de si, para todo k positivo, existen torneos con la propiedad S_k. Finalmente, Erdős consiguió dar una demostración con respuesta afirmativa en 1963 [18], y dicha prueba involucra simples argumentos de probabilidad que ilustran muy bien la filosofía del método probabilístico. Esencialmente, vamos a comprobar que, para valores lo suficientemente grandes de n, será muy probable que un torneo de n jugadores elegido al azar cumpla la propiedad S_k.

Proposición 13. *Si* $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, *entonces existe un torneo de n jugadores con la propiedad* S_k.

Demostración. Para diseñar un torneo de n jugadores solamente hay que asignar un resultado a cada uno de los $\binom{n}{2}$ enfrentamientos. Vamos a construirlo de forma que cada resultado es elegido uniformemente (probabilidad 1/2 para cada sentido) y de manera independiente. Pasaremos después a demostrar que la probabilidad de que un torneo cumpla la propiedad S_k es positiva. Para que el torneo cumpla la propiedad S_k, tiene que cumplir $\binom{n}{k}$ condiciones, a saber, que para cada grupo de k jugadores exista otro jugador en el torneo que gane a todo este grupo. Vamos ahora a introducir los sucesos malos, en el sentido de la observación 9. Ordenamos de 1 a $\binom{n}{k}$ los grupos de k jugadores y al i-ésimo, denotado por L_i, le asignamos el suceso $A_i = \{\text{No existe un jugador en el torneo que gane a los } k \text{ jugadores de } L_i\}$. Vamos a comprobar que existe un torneo de n jugadores con la propiedad S_k o, equivalentemente, que los sucesos A_i no cubren todo el espacio de probabilidad descrito. Esto se hará por medio del primer punto de la proposición 8.

Es sencillo realizar el cálculo de $\mathbb{P}(A_i)$ para cada $1 \leq i \leq \binom{n}{k}$. La probabilidad de que un jugador fijo (y fuera de L_i) no gane a todos los jugadores de L_i es igual a $1-2^{-k}$, ya que a cada jugador le gana de manera independiente y con probabilidad 1/2 (véase la definición 4). Para que el suceso A_i se dé, debe darse la condición anterior para cada uno de los $n-k$ jugadores fuera de L_i. Como tales $n-k$ sucesos son independientes y de probabilidad $1-2^{-k}$, se deduce que $\mathbb{P}(A_i) = (1-2^{-k})^{n-k}$. Así que, por la igualdad anterior y la hipótesis del enunciado,

$$\sum_{i=1}^{\binom{n}{k}} \mathbb{P}(A_i) = \binom{n}{k}\left(1-2^{-k}\right)^{n-k} < 1,$$

de modo que $\mathbb{P}(\bigcap_{i=1}^{\binom{n}{k}} A_i^c) > 0$ por el primer punto de la proposición 8. ∎

Esta proposición resuelve la pregunta de Schütte ya que, para k fijo, basta tomar $n > k$ lo suficientemente grande para aplicar este resultado. La existencia de tal n está garantizada porque $\binom{n}{k}(1-2^{-k})^{n-k} \to 0$ cuando $n \to \infty$. De hecho, teniendo en cuenta este límite y la prueba de esta proposición, notamos que la probabilidad de que un torneo aleatorio de n jugadores tenga la propiedad S_k tiende a 1 según n tiende a infinito.

3.1. Ejercicios

Proponemos algunos ejercicios, extraídos del libro de Alon y Spencer [3], el artículo de Chen [11] y las notas de Loh [25], para invitar al lector a familiarizarse con la técnica basada en la proposición 8 antes de analizar otras más avanzadas. Todas las soluciones están en el apéndice A del final del artículo. Se recuerda primero la definición de grafo bipartito para los ejercicios 1 y 3.

Definición 14. Decimos que un grafo G es **bipartito** si $V(G)$ es la unión de dos conjuntos A y B disjuntos y no vacíos tales que $E(G) \subseteq \{\{a,b\} : a \in A, b \in B\}$. Dado $n \in \mathbb{Z}^+$, denotamos por $\mathcal{K}_{n,n}$ el grafo bipartito con $A = \{a_i : i \in [n]\}$, $B = \{b_j : j \in [n]\}$ y aristas $\{\{a_i, b_j\} : i, j \in [n]\}$. ◂

Ejercicio 1. Todo grafo de m aristas contiene un subgrafo bipartito con al menos $m/2$ aristas. ◂

Ejercicio 2. En la Duma hay 1600 delegados, que han formado 16 000 comités de 80 personas cada uno. Prueba que hay dos comités con, al menos, cuatro delegados en común. ◂

Ejercicio 3. Cualquier subgrafo de $2n$ vértices y $n^2 - n + 1$ aristas del grafo bipartito $\mathcal{K}_{n,n}$ admite una partición de sus vértices en n parejas de manera que haya una arista entre cada pareja de vértices. ◂

Ejercicio 4. Se pide probar la ecuación (4), es decir, que el número de Ramsey $R(k,k)$ (definición 1) cumple que $R(k,k) \geq 2^{k/2}$ si $k \geq 2$. ◂

Ejercicio 5. Si en una cuadrícula rectangular de dimensiones $n \times n$ se colocan los números del 1 al n de manera que cada uno de ellos aparezca n veces en la cuadrícula, entonces existe una fila o una columna con al menos $\sqrt{n}$ números distintos. ◂

4. Combinatoria aditiva

Vamos a pasar a analizar un resultado de teoría combinatoria de números. La combinatoria aditiva es un inmenso campo de investigación y motivar una introducción sobre las cuestiones que abarca es una tarea que se escapa de las intenciones de este artículo; se recomienda el libro de Tao y Vu [39]. Expondremos primero un ejemplo que se debe a Erdős [19] y, después de la demostración, discutiremos brevemente la evolución del problema hasta ahora planteado con diferentes parámetros o en otros contextos.

Definición 15. Un conjunto de enteros se dice **libre de sumas** si no existen tres números en dicho conjunto, digamos x, y, z tales que $x + y = z$. ◂

Teorema 16. *Todo conjunto $A = \{a_1, \dots, a_n\} \subseteq \mathbb{Z}$ de enteros no nulos admite un subconjunto libre de sumas con al menos $n/3$ elementos.*

Antes de exponer la demostración, es conveniente adelantar cómo se va a desarrollar. Se va trabajar en la aritmética modular del anillo $\mathbb{Z}/p\mathbb{Z}$, para cierto primo p. Este anillo también es un cuerpo, de modo que los elementos no nulos tienen inverso multiplicativo.

Notación. Para p primo, $z \in \mathbb{Z}/p\mathbb{Z}$ y $W \subseteq \mathbb{Z}/p\mathbb{Z}$, denotamos por zW el subconjunto $\{zw : w \in W\}$. ◂

Para probar el teorema 16, primero se tomará un primo grande para que todos los enteros sean distintos módulo ese primo. Además, consideraremos un subconjunto B de $[p-1]$ libre de sumas de tamaño mayor que un tercio del total. Esto nos da muchos subconjuntos libres de sumas, los que son de la forma cB con c no nulo, y alguno de ellos deberá tener una intersección grande con A, que será libre de sumas.

Demostración del teorema 16. Conservamos la notación de la discusión inmediatamente anterior. Sea p un primo de la forma $3k+2$, para cierto entero k, tal que todos los elementos de A sean distintos módulo p. Por ejemplo, es suficiente que $p > 2\,\text{máx}_{a\in A}|a|$. Consideremos el anillo $\mathbb{Z}/p\mathbb{Z}$ y el siguiente subconjunto libre de sumas con $k+1 > p/3$ elementos:

$$B = \{k+1, \dots, 2k+1\}.$$

Dado que estamos en un cuerpo, para todo elemento a de A, existen exactamente $k+1$ elementos x del conjunto $[p-1]$ tales que $ax \in B$. Ahora, tomando un elemento c de $[p-1]$ uniformemente al azar, se considera la variable aleatoria $X(c) = |A \cap c^{-1}B|$, que es expresable como la suma de indicadores $\sum_{i=1}^{n} X_{a_i^{-1}B}$ (definición 5). Aplicando la linealidad de la esperanza,

$$\mathbb{E}[X] = \sum_{k=1}^{n} \mathbb{E}[X_{a_k^{-1}B}] = \sum_{k=1}^{n} \mathbb{P}(X_{a_k^{-1}B} = 1) = \sum_{k=1}^{n} \mathbb{P}\left(c \in a_k^{-1}B\right) = \sum_{k=1}^{n} \frac{|a_k^{-1}B|}{p-1} = n\frac{k+1}{p-1} > \frac{n}{3},$$

y queda finalizada la demostración por el segundo punto de la proposición 8, ya que para cierto $c \in [p-1]$ se deberá tener que $A \cap c^{-1}B$ es un subconjunto de A libre de sumas con más de $n/3$ elementos. ∎

No se sabe si el resultado anterior es cierto o no reemplazando $n/3$ por $n/3 + 10$ en el enunciado. Sin embargo, sí que podría reemplazarse por $(n+2)/3$, según prueba Bourgain [9] con técnicas de análisis de Fourier. Por otro lado, sí se sabe que $1/3$ es la mayor fracción que se puede poner en el teorema 16 (véase el artículo de Eberhard, Green y Manners [15]).

Nótese que tanto la definición 15 como el teorema 16 pueden reformularse para un grupo G cualquiera, en lugar de $\mathbb{Z}$. Es natural preguntarse qué familias de grupos y qué fracciones pueden considerarse en el teorema 16 de modo que siga siendo cierto. Es decir, queremos entender las familias de grupos $\mathcal{F}$ para las cuales existe una fracción $c > 0$ de forma que para todo grupo G en $\mathcal{F}$ y todo $A \subseteq G$ de n elementos exista un subconjunto de A con al menos nc elementos que sea libre de productos.

Para esta cuestión, tras mirar cuidadosamente la demostración del teorema 16, puede tomarse $\mathcal{F}_1$ compuesta por $\mathbb{Z}$ y los grupos cíclicos $\mathbb{Z}/q\mathbb{Z}$, para q primo, junto con la constante $c = 1/3$. Además, Alon y Kleitman [2] probaron que puede tomarse la familia $\mathcal{F}_2$ de grupos abelianos y la fracción $c = 2/7$ (menor que $1/3$, pero óptima para esta familia más grande).

Sin embargo, Gowers [23] encuentra una familia $\mathcal{F}_3$ para la cual no puede asegurarse tal fracción $c > 0$ absoluta. En efecto, se comprueba para todo primo q que el grupo lineal proyectivo de $\mathbb{F}_q^2$, denotado por $\mathrm{PSL}_2(\mathbb{F}_q)$ y con orden n_q, no tiene subconjuntos libres de productos con más de $2n_q^{8/9}$ elementos. Nótese que, para cada fracción $c > 0$ fija, la cantidad $2n_q^{8/9}$ es menor que cn_q si q es lo suficientemente grande.

El ingrediente principal en la prueba se debe a un resultado de Frobenius de 1897 por el cual, esencialmente, estos grupos $\mathrm{PSL}_2(\mathbb{F}_q)$ no tienen representaciones no triviales de dimensión baja (véase el libro de Davidoff, Sarnak y Valette [13, teorema 3.5.1]). En este sentido, estos grupos están muy lejos de ser conmutativos porque todas las representaciones irreducibles de un grupo abeliano tienen dimensión 1. La moraleja es que las técnicas probabilísticas y el análisis de Fourier en grupos no conmutativos tienen muy distinta naturaleza al caso abeliano.

Por último, planteamos al lector el siguiente ejercicio extraído del libro de Djukić *et al.* [14, sección 3.40.2].

Ejercicio 6. Sea $N \in \mathbb{Z}^+$. Consideramos el grupo $G = \mathbb{Z}/N^2\mathbb{Z}$ junto con un subconjunto $A \subseteq G$ de N elementos. Comprueba que existe una colección B de N elementos de G de manera que $|A+B| \geq |G|/2$. ◂

5. El método del segundo momento en teoría de números

Hemos hablado de la media o esperanza $\mathbb{E}[X]$ de una variable aleatoria X y de cómo su cálculo nos daba una idea primitiva de cómo puede ser X, al menos en un sentido débil reflejado en el segundo punto de la proposición 8. Dando un paso más en el estudio de una variable aleatoria X, estudiaremos tanto $\mathbb{E}[X]$ como su varianza $\mathrm{Var}(X)$, definida por $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Una posible inquietud es intentar entender cómo estas dos cantidades delimitan las posibles distribuciones de X. Por ejemplo, si $\mathrm{Var}(X) = 0$, entonces X es constante e igual a $\mathbb{E}[X]$. De manera general, $\mathrm{Var}(X)$ restringe las desviaciones en distribución que pueda tener X con respecto a la constante $\mathbb{E}[X]$, como recoge más cuantitativamente la desigualdad de Chebyshov.

Lema 17 (Desigualdad de Chebyshov). *Sea X una variable aleatoria. Se tiene que, para todo c positivo,*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq c\sqrt{\mathrm{Var}(X)}\right) \leq \frac{1}{c^2}.$$

Insistimos en que la naturaleza de las conclusiones obtenidas con el método empleado en esta sección son diferentes a las del resto del artículo. En el caso de la sección 3, solo podía concluirse que $X \geq \mathbb{E}[X]$ con probabilidad positiva apelando al segundo punto de la proposición 8. Aquí se considerará también la varianza y se podrá inferir que X estará en un intervalo centrado en $\mathbb{E}[X]$ con alta probabilidad, obteniendo no solamente resultados de existencia sino también de concentración.

Un hecho conveniente sobre la esperanza es que es lineal en la suma de variables aleatorias (proposición 7), pero la varianza de una suma equivaldrá a la suma de varianzas salvo un término que mide la dependencia entre dichas variables (lema 18). Dicha medida cuantitativa de la dependencia que puede haber entre dos variables aleatorias X, Y viene dada por su covarianza $\mathrm{Cov}(X,Y)$, definida por $\mathbb{E}\left[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])\right]$. Si X e Y son independientes, entonces $\mathrm{Cov}(X,Y) = 0$, pero el recíproco no es cierto.

Lema 18. *Dadas las variables aleatorias $X_1, \ldots, X_n$, se tiene la siguiente identidad:*

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

El momento de orden k de una variable aleatoria se define por $\mathbb{E}[X^k]$, y el método de momentos consiste, brevemente, en obtener información sobre la distribución de una variable aleatoria a partir del cálculo o estimación de sus momentos. La desigualdad de Chebyshov es un resultado que va en esa dirección y decimos que en esta sección emplearemos el método del segundo momento para adelantar que estudiaremos una variable aleatoria a través de su media $\mathbb{E}[X]$ y su varianza, que puede calcularse como $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Una de las grandes familias de problemas que aparecen en teoría de números es aquella en la que se pretende estudiar cuestiones relacionadas con los números primos. Hay múltiples enfoques al estudio de estos números; en este caso, entenderemos un poco mejor cuáles son su densidad y distribución. Para ello, vamos a necesitar una estimación de Mertens [27], el teorema 20, sobre series asociadas a los números primos, para la cual hay pruebas más directas combinando la sumación de Abel y la fórmula de Stirling. Antes de ello, introducimos una notación asintótica que se usará especialmente en esta sección pero también en alguna estimación de la sección 6. Podrá verse que es muy cómoda y conveniente.

Definición 19. Sean $f\colon \mathbb{Z}^+ \to \mathbb{R}$ y $g\colon \mathbb{Z}^+ \to (0,\infty)$.

- Escribimos $f = \mathcal{O}(g)$ si existe $C > 0$ tal que $|f(n)| \leq Cg(n)$ para todo $n \in \mathbb{Z}^+$.
- Escribimos $f = o(g)$ si $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. ◂

Por ejemplo, $10 - n^2 = \mathcal{O}(n^2)$ y $45/n^2 = o(1/n)$. Se recuerda también que el uso del símbolo de igualdad en la definición 19 es un abuso de notación. Por otro lado, se observa que, si $f_1, f_2\colon \mathbb{Z}^+ \to \mathbb{R}$ y $g\colon \mathbb{Z}^+ \to (0,\infty)$ verifican que $f_1 = \mathcal{O}(g)$ y $f_2 = \mathcal{O}(g)$, entonces para todos los reales c_1, c_2 se tiene que $c_1 f_1 + c_2 f_2 = \mathcal{O}(g)$.

Teorema 20 (Mertens). *La suma de los inversos de los primos menores que un cierto entero n cumple lo siguiente:*

$$\sum_{p\leq n} \frac{1}{p} = \log\log n + \mathcal{O}(1).$$

Sea x un entero positivo. Denotamos por $\nu(x)$ el número de primos que dividen a x. El siguiente resultado fue demostrado por primera vez por Hardy y Ramanujan en 1920, pero la prueba que trataremos, ofrecida por Turán en 1934 [40], es menos compleja y tiene un papel crucial en el posterior desarrollo de estas técnicas probabilísticas en teoría de números. Vagamente, viene a decir que la cantidad de factores primos que se espera de un entero positivo en $[n]$ es $\log\log n$.

Teorema 21. *Sea $\omega\colon \mathbb{Z}^+ \to (0,\infty]$ con $\omega(n) \to \infty$. Entonces, la cantidad de valores de x en $[n]$ tales que $|\nu(x) - \log\log n| \geq \omega(n)\sqrt{\log\log n}$ es $o(n)$.*

Demostración. Queremos estudiar los x de un conjunto $[n]$ que verifican una cierta propiedad. Para estudiar la densidad de tales x en $[n]$, se considera una distribución uniforme en este espacio. Es decir, dado $A \subseteq [n]$, se tiene que $\mathbb{P}(A) = |A|/n$. Para cada n, la función $\nu_n\colon [n] \to \mathbb{R}$ definida por $x \mapsto \nu(x)$ es una variable aleatoria en $[n]$. En términos probabilísticos, lo que se pretende probar es lo siguiente:

$$\mathbb{P}\left(\frac{|\nu_n - \log\log n|}{\sqrt{\log\log n}} \geq \omega(n)\right) = o(1). \tag{5}$$

Vamos a denotar la indicatriz de $\{\text{múltiplos de } m\} \cap [n]$ por $X_{m,n}$. Es decir, dado $x \in [n]$, $X_{m,n}(x) = 1$ si m divide x, y $X_{m,n}(x) = 0$ en otro caso.

Es directo ver que $\nu_n = \sum_{p\leq n} X_{p,n}$, donde la suma se toma sobre primos p. Sin embargo, por razones técnicas, no trabajaremos con esta variable aleatoria. El problema que surge al estudiar la suma de indicatrices $X_{p,n}$ es que su varianza no es lo suficientemente estable en n porque la suma de covarianzas que aparece en el lema 18 para realizar el cálculo de $\mathrm{Var}(\nu_n)$ no es negligible. En su lugar, consideraremos la variable aleatoria $\mu_n = \sum_{p\leq n^{1/10}} X_{p,n}$, la cual tiene el mismo comportamiento asintótico que ν_n por un sencillo argumento de conteo que nos lleva al siguiente control entre ambas.

Lema 22. *Para todo $x \in [n]$, $\nu_n(x) - 10 \leq \mu_n(x) \leq \nu_n(x)$.*

Demostración. La segunda desigualdad es inmediata porque $\nu_n(x)$ cuenta los divisores primos de x pero $\mu_n(x)$ solo aquellos divisores primos de x no mayores que $n^{1/10}$. La primera desigualdad se deduce de que ningún $x \leq n$ pueda tener más de diez divisores primos mayores que $n^{1/10}$. ∎

El siguiente objetivo es estimar $\mathbb{E}[\mu_n]$ y $\mathrm{Var}(\mu_n)$ cuando $n \to \infty$. La función $\log\log n$ emergerá en ambos cálculos haciendo uso del teorema 20.

Lema 23. $\mathbb{E}[\mu_n] = \log\log n + \mathcal{O}(1)$.

Demostración. Recuérdese que $r - 1 < \lfloor r \rfloor \leq r$ para todo $r \in \mathbb{R}$. Se observa que $\mathbb{E}[X_{p,n}] = \lfloor n/p \rfloor / n$, ya que la cantidad de múltiplos de p en $[n]$ es $\lfloor n/p \rfloor$. Luego $1/p - 1/n < \mathbb{E}[X_{p,n}] \leq 1/p$ y, por tanto, $\mathbb{E}[X_{p,n}] = 1/p + \mathcal{O}(1/n)$. Aplicando la linealidad de la esperanza y, en la última igualdad, el teorema 20, concluimos que

$$\mathbb{E}[\mu_n] = \sum_{p \leq n^{1/10}} \left(\frac{1}{p} + \mathcal{O}\left(\frac{1}{n}\right)\right) = \sum_{p \leq n^{1/10}} \frac{1}{p} + n^{1/10} \cdot \mathcal{O}\left(\frac{1}{n}\right) = \log\log n + \mathcal{O}(1). \qquad \blacksquare$$

Lema 24. $\mathrm{Var}(\mu_n) = \log\log n + \mathcal{O}(1)$.

Demostración. Tenemos que $\mathrm{Var}(X_{p,n}) = \mathbb{E}[X_{p,n}^2] - \mathbb{E}[X_{p,n}]^2 = (1 - 1/p)/p + \mathcal{O}(1/n)$. Por tanto,

$$\sum_{p \leq n^{1/10}} \mathrm{Var}(X_{p,n}) = \sum_{p \leq n^{1/10}} \frac{1}{p} - \sum_{p \leq n^{1/10}} \frac{1}{p^2} + n^{1/10} \mathcal{O}\left(\frac{1}{n}\right) = \log\log n + \mathcal{O}(1), \tag{6}$$

donde hemos usado, análogamente al lema anterior, el teorema 20 y el hecho de que tenemos que $\sum_{p \leq n} 1/p^2 \leq \sum_{k=1}^{n} 1/k^2 = \mathcal{O}(1)$, ya que esta es una serie convergente. Sobre el cálculo de las covarianzas, tenemos que

$$\begin{aligned}\mathrm{Cov}(X_{p,n}, X_{q,n}) &= \mathbb{E}[X_{p,n}X_{q,n}] - \mathbb{E}[X_p]\,\mathbb{E}[X_q] = \mathbb{E}[X_{pq,n}] - \mathbb{E}[X_p]\,\mathbb{E}[X_q] \\ &= \frac{1}{pq} + \mathcal{O}\left(\frac{1}{n}\right) - \left(\frac{1}{p} + \mathcal{O}\left(\frac{1}{n}\right)\right)\left(\frac{1}{q} + \mathcal{O}\left(\frac{1}{n}\right)\right) = \left(1 + \frac{1}{p} + \frac{1}{q}\right) \cdot \mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}\left(\frac{1}{n}\right).\end{aligned}$$

Lo siguiente será probar que $\sum_{p \neq q \leq n^{1/10}} \mathrm{Cov}(X_{p,n}, X_{q,n}) = o(1)$. Efectivamente,

$$\sum_{p \neq q \leq n^{1/10}} \mathrm{Cov}(X_{p,n}, X_{q,n}) = \sum_{p \neq q \leq n^{1/10}} \mathcal{O}\left(\frac{1}{n}\right) = n^{2/10} \cdot \mathcal{O}\left(\frac{1}{n}\right) = o(1). \tag{7}$$

Nótese que es en el cálculo anterior donde precisamente se explota el hecho de que el rango de primos considerado está acotado por $n^{1/10}$ y no solo por n. Así, la dependencia entre las variables aleatorias $X_{p,n}$ es muy pequeña y no altera (asintóticamente) la varianza de μ_n. Por el lema 18 y las estimaciones (6) y (7),

$$\mathrm{Var}(\mu_n) = \sum_{p \leq n^{1/10}} \mathrm{Var}(X_{p,n}) + \sum_{p \neq q \leq n^{1/10}} \mathrm{Cov}(X_{p,n}, X_{q,n}) = \log\log n + \mathcal{O}(1). \qquad \blacksquare$$

Ya tenemos todos los ingredientes para demostrar la ecuación (5). De hecho, vamos a comprobar que existe N_0, que solo depende de la función ω, tal que para todo $n \geq N_0$ se cumple que

$$\mathbb{P}\left(\frac{|\nu_n - \log\log n|}{\sqrt{\log\log n}} \geq \omega(n) \right) \leq \frac{32}{\omega(n)^2}, \tag{8}$$

lo cual finaliza la prueba porque $\omega(n) \to \infty$ cuando $n \to \infty$. Teniendo en cuenta los dos lemas anteriores y que $\omega(n) \to \infty$ cuando $n \to \infty$, se sigue que existe N_0 tal que, para todo $n \geq N_0$, las tres siguientes condiciones se dan:

$$\frac{10}{\sqrt{\log\log n}} \leq \frac{\omega(n)}{2}, \quad \frac{|\mathbb{E}[\mu_n] - \log\log n|}{\sqrt{\log\log n}} \leq \frac{\omega(n)}{4}, \quad \frac{\mathrm{Var}(\mu_n)}{\log\log n} \leq 2.$$

Empleando el lema por el cual $|\nu_n - \mu_n| \leq 10$, se deduce que para todo $n \geq N_0$ se cumple lo siguiente:

$$\begin{aligned}\mathbb{P}\left(\frac{|\nu_n - \log\log n|}{\sqrt{\log\log n}} \geq \omega(n) \right) &\leq \mathbb{P}\left(\frac{10 + |\mu_n - \log\log n|}{\sqrt{\log\log n}} \geq \omega(n) \right) \leq \mathbb{P}\left(\frac{|\mu_n - \log\log n|}{\sqrt{\log\log n}} \geq \frac{\omega(n)}{2} \right) \\ &\leq \mathbb{P}\left(\frac{|\mu_n - \mathbb{E}[\mu_n]| + |\mathbb{E}[\mu_n] - \log\log n|}{\sqrt{\log\log n}} \geq \frac{\omega(n)}{2} \right) \\ &\leq \mathbb{P}\left(\frac{|\mu_n - \mathbb{E}[\mu_n]|}{\sqrt{\log\log n}} \geq \frac{\omega(n)}{4} \right) \leq \mathbb{P}\left(\frac{|\mu_n - \mathbb{E}[\mu_n]|}{\sqrt{\mathrm{Var}(\mu_n)}} \geq \frac{\omega(n)}{4\sqrt{2}} \right).\end{aligned}$$

Finalmente, por el lema 17, la última expresión está acotada por $32/\omega(n)^2$, demostrando así (8). $\blacksquare$

Este teorema admite una extensión que da una descripción más precisa de la distribución de ν, la cual se comporta, en el límite, como una normal de media y varianza iguales a $\log\log n$. Este es un teorema de Erdős y Kac, de 1940 [20].

Teorema 25. *Sea $\lambda \in \mathbb{R}$ fijo. Entonces, se tiene que*

$$\lim_{n\to\infty} \frac{1}{n} \left| \left\{ \nu \geq \log\log n + \lambda\sqrt{\log\log n} \right\} \cap [n] \right| = \int_\lambda^\infty \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}\, \mathrm{d}x.$$

Recordemos que una distribución N es normal (o también conocida como gaussiana) de media μ y varianza σ^2 si y solo si es de la forma $N = \mu + \sigma N_0$, donde N_0 sigue una distribución normal estándar, esto es, de media 0 y varianza 1. Equivalentemente, el primer caso se da si $\frac{N-\mu}{\sigma}$ sigue un distribución normal estándar. La función de densidad de N_0 es $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, así que $\mathbb{P}(N_0 \geq c) = \int_c^\infty f(x)\,\mathrm{d}x$ para todo $c \in [-\infty, \infty]$. Podemos reescribir el límite del teorema de Erdős y Kac de manera más esclarecedora:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\nu_n - \log\log n}{\sqrt{\log\log n}} \geq \lambda \right) = \mathbb{P}(N_0 \geq \lambda), \quad \text{para todo } \lambda \in \mathbb{R},$$

en forma de una convergencia en distribución. El hecho más importante en el tipo de argumentos involucrados en la prueba del teorema 25 es que hay variables aleatorias X, como la gaussiana, que están totalmente determinadas por sus momentos en el siguiente sentido: cualquier variable aleatoria Y con momentos $\mathbb{E}[Y^k] = \mathbb{E}[X^k]$ para todo $k \in \mathbb{N}$ cumple que $\mathbb{P}(X = Y) = 1$. El lector puede dirigirse al libro de Billingsley [5, capítulo 30] para ver condiciones suficientes en las que la situación anterior se da, esencialmente cuando la sucesión de momentos crece lo suficientemente despacio para que la función $M_X(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}XT}]$ pueda estudiarse como la función generatriz que contiene como coeficientes de su desarrollo en serie de potencias a la sucesión $\{\mathbb{E}[X^k]/k!\}_{k\geq 0}$. La importancia de la función $M_X(t)$ radica en que, dadas las variables aleatorias X_n e Y, la convergencia de X_n a Y en distribución puede reducirse a comprobar la convergencia puntual de las funciones $M_{X_n}(t)$ a $M_Y(t)$, lo cual también puede reducirse a probar, para cada $k \geq 1$, la convergencia de $\mathbb{E}[X_n^k]$ a $\mathbb{E}[Y^k]$, una tarea mucho más sencilla. Al igual que en esta sección hemos aplicado el *método del segundo momento*, la sección 3 podría llamarse *método del primer momento* y en la demostración del teorema 25 se aplica el *método de momentos* descrito con más detalle en el libro de Billingsley [5, teorema 30.3]. Otra idea importante en esta demostración, que ya aparece en el teorema 21 y que es muy común en probabilidad, es el método de truncamiento. En ocasiones, una variable aleatoria en su totalidad oscila demasiado, como ocurre con la varianza de ν_n, y por ello es mejor aproximarla, considerando μ_n. Para estimar los momentos de orden $n \geq 3$ de ν_n y probar el teorema 25 se emplean las mismas técnicas que en la prueba del teorema 21, pero el truncamiento sí que resulta ser más sutil y se basa en el teorema 20.

Aprovechando que a esta alturas hemos descrito pruebas probabilísticas de distinta naturaleza, queremos indicar que, entendiendo el método probabilístico en este sentido amplio que se considera en la introducción, la primera aplicación de este método se le atribuye a Emil Borel [8], uno de los padres fundadores de la teoría de la medida y la probabilidad. En el artículo anterior se prueba que casi todos los números reales son normales.

6. Lema local de Lovász

Una de las debilidades del argumento empleado en la proposición 13 es que requiere condiciones muy fuertes para ser aplicado y, en el sentido de la observación 9, las probabilidades de los sucesos malos deben ser muy pequeñas para que se pueda proceder con esta técnica. En efecto, la conclusión del resultado no es solo que exista algún torneo con la propiedad S_k, sino que casi todos la verifican cuando el número de vértices es muy grande. Esto sugiere investigar nuevas formas de asegurar que ciertos sucesos malos no cubran un espacio de probabilidad.

Por otro lado, hay otro fenómeno muy usual en problemas de combinatoria de interés como las coloraciones: la independencia de sucesos. Obsérvese, por la definición 4, que si d sucesos malos tienen probabilidad a lo sumo $p < 1$ y son independientes, entonces, con probabilidad al menos $(1-p)^d > 0$, ninguno de ellos ocurre. Así que ya conocemos dos condiciones extremas bajo las cuales una colección finita de sucesos no cubre un espacio de probabilidad: o bien no lo cubren individualmente y son independientes o bien sus probabilidades son muy pequeñas en el sentido del primer punto de la proposición 8. Ambos casos son demasiado favorables. El lema local de Lovász permite cuantificar y explotar casos intermedios.

Antes de enunciar este lema necesitamos introducir una nueva noción de independencia en la definición 26 diferente de la que aparece en la definición 4, para la cual usaremos la siguiente notación.

Notación. La probabilidad $\mathbb{P}(A \mid B)$ de un suceso A condicionado al suceso B viene dada por $\mathbb{P}(A \cap B)/\mathbb{P}(B)$ y está definida solo cuando $\mathbb{P}(B) > 0$. Nótese que A y B son independientes si y solo $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Es por esto que la notación de probabilidad condicionada es más cómoda y contiene más significado. Así que escribiremos identidades del estilo de $\mathbb{P}(A \mid B) \geq c_1$ o $\mathbb{P}(A \mid B) = c_2$, que son simplemente una abreviación para $\mathbb{P}(A \cap B) \geq c_1 \mathbb{P}(B)$ y $\mathbb{P}(A \cap B) = c_2 \mathbb{P}(B)$, respectivamente. En el caso degenerado de que $\mathbb{P}(B) = 0$, se tiene que $\mathbb{P}(A \cap B) = 0$ también y ambas desigualdades son trivialmente ciertas. ◂

Definición 26. Decimos que un suceso S es independiente de una colección $\{S_k\}_{k=1}^n$ de sucesos si para todo $I \subseteq [n]$ se tiene que $\mathbb{P}(S \mid \bigcap_{k \in I} S_k) = \mathbb{P}(S)$. ◂

Definición 27. Dados $A_1, \dots, A_n$ sucesos en un espacio de probabilidad, su **grafo de dependencia** es un grafo G (definición 11) cuyo conjunto de vértices es $\{1, \dots, n\}$ (que debemos pensar como $\{A_1, \dots, A_n\}$) y cuyas aristas indican la posible dependencia entre los correspondientes sucesos. De manera precisa, A_k es independiente de la colección $\{A_i\}_{\{k,i\} \notin E(G)}$ para todo $k \in [n]$. ◂

Observación 28. Se remarca que la condición anterior, que A_k sea independiente de la colección $\{A_i\}_{\{k,i\} \notin E(G)}$, es mucho más débil que la de imponer que los sucesos de la colección $\bigcup_{\{k,i\} \notin E(G)} \{A_i\} \cup \{A_k\}$ sean independientes, recuérdese la definición 4. ◂

Observación 29. También se verifica que S es independiente de $\{S_k, S_k^c\}_{k=1}^n$ si lo es de $\{S_k\}_{k=1}^n$. ◂

El siguiente resultado apareció en un artículo de Erdős y Lovász en 1975 [21] y a modo de motivación adelantamos que nos llevará directamente al lema 31, que cuantifica esta intuición: si una colección de sucesos malos tiene probabilidad positiva de no ocurrir y cada uno de ellos tiene una probabilidad pequeña en comparación con el número de sucesos de los que depende, entonces, con probabilidad positiva, ninguno de ellos sucede.

Lema 30 (Versión general del lema local de Lovász). *Sean $A_1, \dots, A_n$ sucesos en un espacio de probabilidad $(\Omega, \mathbb{P})$ con grafo de dependencia G. Supongamos que existen números reales $\{x_i\}_{i=1}^n \subseteq [0,1)$ tales que*

$$\mathbb{P}(A_i) \leq x_i \prod_{\{i,j\} \in E(G)} (1 - x_j), \quad \text{para todo } i \in [n].$$

Entonces, se tiene que

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) \geq \prod_{i=1}^n (1 - x_i) > 0.$$

Demostración. Vamos a asumir ahora el siguiente enunciado que probaremos más tarde:

$$\mathbb{P}\left(A_i^c \,\middle|\, \bigcap_{j \in J} A_j^c\right) \geq 1 - x_i, \ \forall J \subseteq [n], \forall i \notin J. \tag{9}$$

De esta forma, es fácil terminar la prueba del lema:

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k^c\right) = \mathbb{P}(A_1^c)\,\mathbb{P}(A_2^c \mid A_1^c)\,\mathbb{P}(A_3^c \mid A_1^c \cap A_2^c) \cdots \mathbb{P}\left(A_n^c \,\middle|\, \bigcap_{j=1}^{n-1} A_j^c\right) \geq (1-x_1)(1-x_2)(1-x_3)\cdots(1-x_n).$$

Procederemos por inducción en $|J|$ para demostrar el siguiente enunciado equivalente a (9):

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j\in J} A_j^c\right) \leq x_i,\ \forall J \subseteq [n], \forall\, i \notin J. \tag{10}$$

El caso base es $|J| = 0$. Se recuerda que la intersección vacía de subconjuntos es el conjunto total. Así,

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j\in J} A_j^c\right) = \mathbb{P}(A_i) \leq x_i \prod_{\{i,j\}\in E(G)} (1 - x_j) \leq x_i \quad \forall\, i \in [n].$$

Para el paso inductivo, vamos a asumir que la propiedad se verifica para todo subconjunto J con $|J| < m$ (hipótesis de inducción fuerte) y vamos a comprobarlo para cualquier subconjunto $J \subseteq [n]$ de tamaño $|J| = m$. Dado $i \notin J$, definimos $J_{i,1} = \{j \in J : \{i, j\} \in E(G)\} = \{u_1, \ldots, u_b\}$, con b elementos, y $J_{i,2} = J - J_{i,1}$, con $m - b$ elementos. También definimos $B_i = \bigcap_{j\in J_{i,1}} A_j^c$ y $C_i = \bigcap_{j\in J_{i,2}} A_j^c$. Reordenando,

$$\mathbb{P}(A_i \mid B_i \cap C_i) = \frac{\mathbb{P}(A_i \cap B_i \cap C_i)}{\mathbb{P}(B_i \cap C_i)} = \frac{\mathbb{P}(A_i \cap B_i \cap C_i)}{\mathbb{P}(C_i)} \frac{\mathbb{P}(C_i)}{\mathbb{P}(B_i \cap C_i)} = \frac{\mathbb{P}(A_i \cap B_i \mid C_i)}{\mathbb{P}(B_i \mid C_i)}. \tag{11}$$

Teniendo en cuenta que A_i es independiente de C_i, por la observación 29 después de la definición 26,

$$\mathbb{P}(A_i \cap B_i \mid C_i) = \frac{\mathbb{P}(A_i \cap B_i \cap C_i)\mathbb{P}(A_i)}{\mathbb{P}(A_i \cap C_i)} \leq \mathbb{P}(A_i) \leq x_i \prod_{\{i,j\}\in E(G)} (1 - x_j), \tag{12}$$

lo cual nos da esta cota superior para el numerador de (11). Ahora pasamos a analizar el denominador,

$$\mathbb{P}(B_i \mid C_i) = \mathbb{P}(A_{u_1}^c \cap A_{u_2}^c \cap \cdots \cap A_{u_b}^c \mid C_i) = \prod_{k=1}^{b} \mathbb{P}\left(A_{u_k}^c \mid C_i \cap A_{u_1}^c \cap \cdots \cap A_{u_{k-1}}^c\right). \tag{13}$$

Teniendo en cuenta que cada uno de los conjuntos $C_i \cap A_{u_1}^c \cap \cdots \cap A_{u_{k-1}}^c$, con $k \in [b]$, es intersección de $(m - b) + k - 1 < m$ conjuntos de la colección inicial $\{A_i\}_{i=1}^n$, podemos aplicar la hipótesis de inducción a cada factor del lado derecho de (13) y la observación de que $\{u_1, \ldots, u_b\} \subseteq \{j : \{i, j\} \in E(G)\}$ para deducir que

$$\mathbb{P}(B_i \mid C_i) = \prod_{k=1}^{b} \mathbb{P}\left(A_{u_k}^c \mid C \cap A_{u_1}^c \cap \cdots \cap A_{u_{k-1}}^c\right) \geq \prod_{k=1}^{b} (1 - x_{u_k}) \geq \prod_{\{i,j\}\in E(G)} (1 - x_j). \tag{14}$$

Finalmente, se observa que (11), (12) y (14) implican que $\mathbb{P}(A_i \mid B_i \cap C_i) \leq x_i$, es decir, (10). ∎

Es necesario plantearse por qué llamamos «lema local» a este resultado. Siguiendo la discusión anterior, es valioso entender el problema general de garantizar que una colección de sucesos en un espacio de probabilidad no cubra todo el espacio. Con la ayuda de este lema no tendríamos que comprobar una condición que involucre a todos los sucesos A_i para concluir que $\mathbb{P}(\bigcap_{k=1}^n A_k^c) > 0$, sino que la condición en cada uno de los A_i es local porque solamente aparecen involucrados aquellos sucesos que puedan depender de A_i. El lema 30 tiene su propio interés pero, a primera vista, parece difícil de usar porque involucra muchos pesos x_i y muchas condiciones a comprobar. Una versión más homogénea y manejable del lema local de Lovász, que será suficiente para las aplicaciones que veremos, es la siguiente.

Lema 31 (Versión simétrica del lema local de Lovász). *Sean $A_1, \ldots, A_n$ sucesos en un espacio de probabilidad $(\Omega, \mathbb{P})$ con grafo de dependencia G. Supongamos que cada vértice de G es extremo de a lo sumo d aristas y que, además, para algún $p \in [0, 1]$ se cumple que $\mathbb{P}(A_k) \leq p$ para todo $k \in [n]$. Si, además, se tiene que $\mathrm{e}p(d + 1) \leq 1$, entonces*

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k^c\right) > 0.$$

Por simplicidad, dejamos enunciado el siguiente lema sencillo.

Lema 32. *Sea n un entero distinto de 0 y -1. Entonces, $\left(1 + \frac{1}{n}\right)^n \leq \mathrm{e}$ si y solo si $n > 0$.*

Demostración del lema 31. Es consecuencia de la versión general. Consideramos los pesos $x_i = \frac{1}{d+1}$. De esta forma se comprueba que, para todo $i \in [n]$:

$$x_i \prod_{\{i,j\}\in E(G)} (1 - x_j) = \frac{1}{d+1}\left(\frac{d}{d+1}\right)^d \geq \frac{1}{d+1}\frac{1}{\mathrm{e}} \geq p \geq \mathbb{P}(A_i),$$

donde hemos aplicado el lema 32 con $n = d + 1 > 0$. Por tanto, el resultado se sigue directamente del lema 30. ∎

Como curiosidad, la constante e que aparece en el lema 31 no es arbitraria, y Shearer [34] demostró que es la mejor que puede considerarse, es decir, la más pequeña para que siga siendo cierto el teorema. Conviene analizar ahora una aplicación estándar de este lema antes de proponer algunos ejercicios.

Definición 33. Un hipergrafo H es un par (V, E), donde V y E son conjuntos tales que $E \subseteq \mathcal{P}(V)$. Llamamos a los elementos de V vértices, y a los elementos de E hiperaristas (o simplemente aristas). Decimos que H es finito si V es finito. Decimos que dos aristas $e, e' \in E$ del hipergrafo se cortan si comparten algún vértice, es decir, si $e \cap e' \neq \emptyset$. Una coloración del hipergrafo (V, E) con k colores es una función $c : V \to \{1, \dots, k\}$. Decimos que un hipergrafo es 2-coloreable si admite una coloración con dos colores que no contenga ninguna arista monocromática. ◂

Teorema 34. *Sea (V, E) un hipergrafo finito que verifica las dos siguientes propiedades:*

1. *Toda arista $e \in E$ tiene al menos k vértices.*
2. *Cada arista $e \in E$ interseca a lo sumo a otras d aristas.*

Si se cumple que $\mathrm{e}(d + 1) \leq 2^{k-1}$, entonces el hipergrafo (V, E) es 2-coloreable.

Demostración. Coloreamos al azar cada vértice con probabilidad 1/2 y de manera independiente. Se considera, para cada arista a, el suceso $S_a = \{\text{La arista } a \text{ es monocromática}\}$. Claramente, si la arista a tiene c_a vértices, entonces $\mathbb{P}(S_a) = 2/2^{c_a} \leq 1/2^{k-1}$ porque $c_a \geq k$. Además, S_a es independiente de la colección de sucesos $\{S_b : b \in E, a \cap b = \emptyset\}$, así que S_a es independiente de una colección de sucesos que contiene a todos los de S quitando aquellos a los que interseca, a lo sumo d. Como $\mathrm{e}(d + 1)/2^{k-1} \leq 1$, se sigue por el lema 31 que $\bigcap_{a\in E} S_a^c \neq \emptyset$ y que, por tanto, existe tal coloración. ∎

Una observación importante es que la desigualdad $\mathrm{e}(d + 1) \leq 2^{k-1}$ del teorema 34 no depende del tamaño de V sino del parámetro local d. Si quisiéramos demostrar el teorema 34 con un argumento de existencia directa al estilo de la sección 3, entonces no habría una manera obvia de aprovechar la segunda condición del enunciado de este teorema. Si empleáramos solo la primera condición y buscáramos aplicar la primera parte de la proposición 8 a lo bruto, necesitaríamos que se cumpliera que $|E(V)| < 2^{k-1}$. Esta desigualdad será, generalmente, mucho más fuerte que la anterior.

Para exhibir esta diferencia en un ejemplo concreto, tomamos dos enteros positivos $k \leq n$ y consideramos el hipergrafo de vértices $V = \mathbb{Z}/n\mathbb{Z}$ y aristas $\{i, i + 1, \dots, i + k - 1\}_{i\in[n]}$ (se recuerda que estos enteros se suman módulo n). Para probar que este hipergrafo es 2-coloreable con un argumento de existencia directa necesitamos $n < 2^{k-1}$ y, en particular, $k > \log_2(n)$, lo cual deja muchos casos sin cubrir. Para aplicar el lema local de Lovász, notando que podemos tomar $d + 1 = 2k$, basta con tener $2\mathrm{e}k \leq 2^{k-1}$, lo cual se cumplirá siempre que $k \geq 7$. Esta desigualdad no involucra a n y, además, prueba la 2-coloración de prácticamente toda la familia de hipergrafos descrita, con parámetros $k \leq n$.

Queremos resaltar que el hipergrafo del párrafo anterior es 2-coloreable si $k \geq 3$ o si n es par y $k = 2$ (coloreando de manera alternada los vértices). Pero esta observación depende de la propia estructura global del grafo y lo que se pretende con estos argumentos probabilísticos es probar que sea inevitable la presencia de cierta estructura (ser 2-coloreable) recurriendo únicamente a ciertas condiciones débiles del grafo que aún den mucha libertad para su posible aspecto (como ocurre con las dos condiciones impuestas en el teorema 34).

En el ejemplo anterior es evidente la potencia del lema local de Lovász. La razón por la que funciona mucho mejor que la proposición 8 es que las dependencias entre los correspondientes sucesos malos eran poco comunes (por ejemplo, $1 + d = 2k$ no dependía del tamaño n del grafo). Sin embargo, en la solución

del ejercicio 4, donde se prueba que $R(k,k) \geq 2^{k/2}$ si $k \geq 2$, sí se trabaja con sucesos malos que presentan mucha dependencia mutua. Al final de la solución se comenta que, con el mismo método, el dado por la proposición 8, y solo usando mejores estimaciones de los coeficientes binomiales (por ejemplo, por medio de la fórmula de Stirling), se obtiene que

$$R(k,k) \geq \frac{k \cdot 2^{k/2}}{\sqrt{2}\,\mathrm{e}}(1+o(1)).$$

Lo interesante es que, con el lema local de Lovász (véase el libro de Alon y Spencer [3, teorema 5.3.1], junto con los comentarios posteriores), se llega a que

$$R(k,k) \geq \frac{\sqrt{2}\cdot k \cdot 2^{k/2}}{\mathrm{e}}(1+o(1)),$$

que solo mejora en un factor 2 porque el alto número de dependencias entre los sucesos malos no puede producir una mejora significativa con respecto al método directo. Esta última cota inferior, descubierta por Spencer [36], es la mejor que se conoce hasta el momento.

Planteamos dos ejercicios, extraídos de [33] y Wikipedia [42], para invitar al lector a emplear el lema 31 antes de exponer una última aplicación más elaborada. Las soluciones pueden encontrarse en el apéndice A.

Ejercicio 7. Sean k y n enteros positivos tales que $k \geq 3\sqrt{n}$. Entonces, es posible pintar las aristas del grafo completo $\mathcal{K}_n$ con k colores de manera que no haya ningún triángulo monocromático. ◂

Ejercicio 8. Se consideran $11n$ puntos sobre una circunferencia y se pintan con n colores de forma que haya 11 puntos de cada color. Se pide probar que pueden elegirse n puntos de distinto color y mutuamente no consecutivos en la circunferencia. ◂

Vamos a pasar ahora a analizar un resultado del cual no se conoce ninguna demostración que no involucre consideraciones probabilísticas. Un aspecto interesante añadido a la demostración del siguiente resultado es que emplea el lema local de Lovász para resolver la versión finita del problema y dota al espacio de configuraciones de una topología, según la cual es compacto, para concluir con la versión global.

Teorema 35. *Sean m, k dos enteros positivos tales que*

$$\mathrm{e}\,(m(m-1)+1)\,k\left(1-\frac{1}{k}\right)^m \leq 1. \tag{15}$$

Entonces, para cualquier conjunto de m reales $W = \{x_1, \ldots, x_m\}$, existe una coloración $c : \mathbb{R} \to [k]$ tal que toda traslación $x + W$, con $x \in \mathbb{R}$, contiene los k colores.

Miremos detenidamente el enunciado anterior. Lo sorprendente es que, fijado el número de colores k, siempre podemos tomar m lo suficientemente grande para que se cumpla la conclusión del teorema. Este resultado es el teorema 5.2.2 del libro de Alon y Spencer [3] y, tal y como se observa después en este libro, puede tomarse $m \geq (3+o(1))k\log k$ para asegurar que se cumple la ecuación (15).

Demostración del teorema 35. Observamos que aparecen términos similares a los del teorema 34. Fijemos el conjunto W. Hay que encontrar una coloración del hipergrafo $H = (\mathbb{R}, E)$, donde $E = \{x + W : x \in \mathbb{R}\}$, de manera que cada arista verifique una propiedad. Dicha propiedad no será, como en el teorema 34, no ser monocromática, sino contener todos los k colores. Lo común con la prueba del teorema 34 es que las coloraciones aleatorias las haremos uniformemente al azar y que en este caso también es fácil tener un control de la correspondiente condición local de dependencia.

Fijada una arista $u \in E$, hay a lo sumo otras $m(m-1)$ aristas que intersecan a u. Vamos a probar esto último. Fijamos $y_1 \in \mathbb{R}$. Para que las dos aristas distintas $y_1 + W, y_2 + W$ ($y_2 \in \mathbb{R}$) se corten, deben existir dos elementos $x_i, x_j \in W$ tales que $y_1 + x_i = y_2 + x_j$. Es decir, $y_2 = x_i + y_1 - x_j$ y hay a lo sumo $m(m-1)$ tales y_2, porque hay como mucho uno por cada par ordenado (x_i, x_j) de $W \times W$ con $x_i \neq x_j$. Con el objetivo de aplicar el lema 31, tomamos $d = (m-1)m$.

Ahora introducimos los sucesos malos. Para cada arista $f \in E$ se define $S_f = \{f \text{ no contiene los } k \text{ colores}\}$. Por la proposición 7, la probabilidad de cada suceso S_f admite la siguiente cota sencilla:

$$S_f = \bigcup_{i=1}^{k} \{f \text{ no contiene el color } i\} \implies \mathbb{P}(S_f) \leq \sum_{i=1}^{k} \mathbb{P}\left(f \text{ no contiene el color } i\right) = k\left(1 - \frac{1}{k}\right)^m.$$

Así que, nuevamente con la mirada en el lema 31, definimos $p = k(1-1/k)^m$. Como se cumple que $ep(d+1) \leq 1$, todo apunta a que el teorema 35 sea una aplicación del lema local de Lovász análoga al teorema 34, pero hay una gran diferencia que quizá el lector atento haya notado, y es que el hipergrafo que estamos considerando no es finito, ni tampoco la cantidad de sucesos malos. Recordamos que el resultado 31 es cierto cuando se considera una cantidad finita de sucesos, así que no podemos aplicarlo directamente como antes.

Sin embargo, no está todo perdido porque podemos aproximar el teorema 35 por sus versiones finitas. En lugar de considerar todas las traslaciones de W, tomaremos un subconjunto $A \subseteq \mathbb{R}$ finito y probaremos que existe una coloración de $\mathbb{R}$ de tal manera que todas las traslaciones de W por elementos de A contengan los k colores. Ahora, el hipergrafo H_A tiene aristas $E_A = \{x + W : x \in A\}$, vértices $V_A = \bigcup_{x \in A}(x + W)$ y sucesos malos $\{S_f : f \in E_A\}$. Con la notación del lema 31 y por lo comentado anteriormente, podemos tomar $d = (m-1)m$ y $p = k(1-1/k)^m$, ya que en este nuevo subgrafo finito de H nos encontramos en unas condiciones locales más débiles y cada S_f sigue teniendo la misma probabilidad. Por tanto, ahora sí podemos concluir con la existencia de tal coloración porque la ecuación (15) equivale a la desigualdad dada en el teorema 34.

El espacio de posibles coloraciones de H es $[k]^{\mathbb{R}}$, que corresponde a un producto cartesiano $\prod_{i \in I} Q_i$ de conjuntos idénticos $Q_i = [k]$ con $I = \mathbb{R}$ como conjunto de índices. Para cada $A \subseteq \mathbb{R}$, llamaremos $C_A \subseteq [k]^{\mathbb{R}}$ al conjunto de coloraciones de $\mathbb{R}$ para las cuales todas las traslaciones $a + W$, con $a \in A$, contienen a todos los colores. Hemos demostrado que C_A es no vacío si A es finito, porque basta considerar la coloración de los vértices de H_A dada por el argumento existencial anterior y, después, se puede pintar el resto de $\mathbb{R}$ de cualquier forma. Esto se debe a que los vértices de H_A son exactamente los reales que aparecen en las traslaciones de W por elementos de A.

En definitiva, hemos comprobado que C_A es no vacío para todo $A \subseteq \mathbb{R}$ finito y queremos probar que $C_{\mathbb{R}}$ es no vacío. Para pasar del caso finito al de todo $\mathbb{R}$, recurriremos a un argumento topológico que consiste en entender mejor el espacio $[k]^{\mathbb{R}}$ y los subconjuntos C_A. Con ese objetivo vamos a introducir, sin demostración, un lema básico y un teorema de topología.

Lema 36. *Sea K un espacio topológico compacto. Sea $\{U_i\}_{i \in I}$ una colección de cerrados de K con todas las intersecciones finitas no vacías, es decir, para todo $n \in \mathbb{Z}^+$ y cualesquiera $i_1, \ldots, i_n \in I$, tenemos que $\bigcap_{k=1}^{n} U_{i_k} \neq \emptyset$. Entonces, $\bigcap_{i \in I} U_i \neq \emptyset$.*

Este lema procede reformular la definición usual de espacio compacto tomando el paso al complementario en todos los conjuntos involucrados. De esta forma, reemplazamos abiertos y uniones por cerrados e intersecciones, respectivamente. El siguiente resultado es conocido como el teorema de Tychonoff y es equivalente al axioma de la elección.

Teorema 37. *El producto de una colección arbitraria de espacios topológicos compactos, dotado de la topología producto, es compacto.*

Como consecuencia de este teorema, tenemos que el espacio de coloraciones $[k]^{\mathbb{R}}$ con la topología producto es compacto si consideramos cada espacio $[k]$ equipado de la topología discreta. Además, con esta topología, los conjuntos $C_{\{a\}}$, con $a \in \mathbb{R}$, son cerrados y cumplen las condiciones del lema anterior, ya que cada intersección finita es igual a $C_A \neq \emptyset$ para cierto $A \subseteq \mathbb{R}$ finito. Por tanto, por la compacidad del espacio $[k]^{\mathbb{R}}$, se deduce que $C_{\mathbb{R}} = \bigcap_{a \in \mathbb{R}} C_{\{a\}} \neq \emptyset$, como queríamos. ∎

El argumento topológico empleado al final de la demostración es bastante común en matemáticas, es un argumento de compacidad por el cual podemos tratar a los compactos como conjuntos finitos. La siguiente cita de René Thom, de su libro *Stabilité Structurelle et Morphogenèse*, hace referencia a esta idea: «La topologie est précisément la discipline mathématique qui permet le passage du local au global».

Además de las fuentes de problemas ya mencionadas, muchas más aplicaciones del método probabilístico pueden encontrarse en el libro de Alon y Spencer [3]. Más en particular, se recomienda la colección de problemas de existencia directa de Boppana [7] y el libro de Molloy y Reed [28], con aplicaciones del lema local de Lovász en cuestiones de coloraciones de grafos.

Nótese que las aplicaciones del lema 31 que hemos tratado contienen como conclusión la existencia de un cierto objeto pero, en la práctica, no permiten saber cómo poder encontrarlo de manera eficiente. Esto se debe a que la prueba del lema 30 involucra argumentos puramente probabilísticos. Por ejemplo, en el teorema 35 se prueba la existencia de una coloración pero no aparece ninguna pista sobre cómo podría ser un algoritmo para hallarla explícitamente en un caso concreto. Es cierto que, aún así, estas demostraciones probabilísticas pueden dar, cuando las probabilidades involucradas son lo suficientemente buenas, un algoritmo aleatorio con el objetivo anterior y que puede simplemente consistir en sortear configuraciones con alguna distribución adecuada varias veces y comprobar si alguna cumple lo deseado.

La línea de investigación que plantea este punto de vista algorítmico comenzó en 1991, cuando Beck [4] probó una versión constructiva del lema local de Lovász con condiciones más restrictivas. Posteriormente, Moser [29] dio una demostración constructiva del lema 30 en 2009, que sería mejorada en 2010 por Moser y Tardos [30] para cubrir casi todos los casos que la versión no constructiva del lema resuelve. Puede consultarse en el blog de Tao [38] una breve discusión del argumento de compresión de entropía utilizado en estos dos últimos trabajos aplicado al *Problema de satisfacibilidad booleana* [43], la versión algorítmica de un problema de lógica sintáctica que consiste en encontrar, dadas varias fórmulas de una cierta forma, una asignación de valores de verdad a sus variables proposicionales para que todas las fórmulas se satisfagan.

Para finalizar, resumimos brevemente lo que se conoce sobre la versión algorítmica de otro de los resutados de naturaleza existencial que hemos tratado sobre estimaciones de los números diagonales de Ramsey. En el ejercicio 4 se planteó probar que $R(k,k) \geq 2^{k/2}$ si $k \geq 2$. Antes de explicar el problema algorítmico que plantea esta desigualdad, vamos a cambiar un poco su formulación. Decimos que un grafo es m-Ramsey si no tiene ningún subgrafo completo de m vértices ni tampoco un subgrafo de m vértices sin aristas. Así que $R(k,k)-1$ sería el máximo entero n para el cual existe un grafo k-Ramsey de n vértices. Con esta formulación, la desigualdad anterior refleja el hecho de que, para cada $n \geq 2$, existe un grafo de n vértices que es $2\log n$-Ramsey. Como se remarca en la introducción, la prueba original de Erdős en 1947 [16] (descrita en la solución del ejercicio 4 en el apéndice A) no produce explícitamente tal grafo, aunque es necesario aclarar dicha noción de explicitud porque no es evidente y, de hecho, ha ido evolucionando.

El propio Erdős ofreció cien dólares (desde su muerte, Graham mantiene esta oferta) a quien pudiera dar, para cierta constante C, ejemplos de grafos $C\log n$-Ramsey de n vértices para un n general. Por entonces, lo noción de explicitud era diferente y, tras la era computacional, se relajó considerablemente hasta llegar a la siguiente interpretación. De manera imprecisa, decimos que un grafo G de n vértices es dado de forma explícita si para dos vértices u y v cualesquiera de G se puede determinar eficientemente si existe una arista entre u y v. Dicha eficiencia significa, cuantitativamente, que existe un algoritmo que pueda realizar la correspondiente tarea en una cantidad de tiempo que sea polinomial en $\log n$, ya que $\log n$ es aproximadamente el número de bits necesarios para codificar cada vértice de G.

Con este planteamiento, ha habido bastante progreso en los últimos años y los mejores resultados actualmente se deben a Chattopadhyay y Zuckerman [10] y Cohen [12]. En ambos se dan explícitamente grafos $2^{(\log\log n)^c}$-Ramsey de n vértices, para una constante c absoluta. Con la intención de comparar la cota de este resultado con la propuesta por Erdős, observamos que, para $c=1$, un grafo $C\cdot 2^{(\log\log n)^c}$-Ramsey es también $C\log n$-Ramsey, pero el c que se asegura en los artículos anteriores no es tan pequeño, así que aún nadie ha podido llevarse los cien dólares.

El desarrollo de esta cuestión a lo largo de los años también exhibe el poder de la aleatoriedad. Con un breve argumento probabilístico quedó resuelta esta cuestión sobre los números de Ramsey a nivel existencial. Sin embargo, en su versión constructiva, incluso con una noción de explicitud relajada a nivel computacional, constituye aún un problema abierto.

Referencias

[1] Alon, Noga. «Paul Erdős and the Probabilistic Method». En: *Notices of the American Mathematical Society* 62.3 (2015), págs. 226-230. issn: 0002-9920. https://doi.org/10.1090/noti1223.

[2] Alon, Noga y Kleitman, Daniel J. «Sum-free subsets». En: *A tribute to Paul Erdős*. Ed. por Baker, Alan; Bollobás, Béla, y Hajnal, András. Cambridge: Cambridge University Press, 1990, págs. 13-26. https://doi.org/10.1017/CBO9780511983917.003.

[3] Alon, Noga y Spencer, Joel H. *The probabilistic method*. With an appendix on the life and work of Paul Erdős. 2.ª ed. Wiley-Interscience Series in Discrete Mathematics and Optimization. Nueva York: Wiley-Interscience, 2000. https://doi.org/10.1002/0471722154.

[4] Beck, József. «An algorithmic approach to the Lovász local lemma. I». En: *Random Structures & Algorithms* 2.4 (1991), págs. 343-365. issn: 1042-9832. https://doi.org/10.1002/rsa.3240020402.

[5] Billingsley, Patrick. *Probability and measure*. 3.ª ed. Wiley Series in Probability and Mathematical Statistics. Nueva York: Wiley-Interscience, 1995. isbn: 978-0-471-00710-4.

[6] Bollobás, Béla. *Graph theory. An introductory course*. Vol. 63. Graduate Texts in Mathematics. Nueva York-Berlín: Springer-Verlag, 1979. https://doi.org/10.1007/978-1-4612-9967-7.

[7] Boppana, Ravi. *Unexpected Uses of Probability*. 2005. url: http://www.aops.com/Forum/viewtopic.php?p=1943887#p1943887/.

[8] Borel, Émile. «Les probabilités dénombrables et leurs applications arithmétiques». En: *Rendiconti del Circolo Matematico di Palermo* 27 (1909), págs. 247-271. https://doi.org/10.1007/BF03019651.

[9] Bourgain, Jean. «Estimates related to sumfree subsets of sets of integers». En: *Israel Journal of Mathematics* 97 (1997), págs. 71-92. issn: 0021-2172. https://doi.org/10.1007/BF02774027.

[10] Chattopadhyay, Eshan y Zuckerman, David. «Explicit two-source extractors and resilient functions». En: *Annals of Mathematics. Second Series* 189.3 (2019), págs. 653-705. issn: 0003-486X. https://doi.org/10.4007/annals.2019.189.3.1.

[11] Chen, Evan. «Expected Uses of Probability». En: *Mathematical reflections* 6 (2014). url: https://web.evanchen.cc/handouts/ProbabilisticMethod/ProbabilisticMethod.pdf.

[12] Cohen, Gil. «Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs». En: *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*. Nueva York: ACM, 2016, págs. 278-284. https://doi.org/10.1145/2897518.2897530.

[13] Davidoff, Giuliana; Sarnak, Peter, y Valette, Alain. *Elementary number theory, group theory, and Ramanujan graphs*. Vol. 55. London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 2003. https://doi.org/10.1017/CBO9780511615825.

[14] Djukić, Dušan; Janković, Vladimir; Matić, Ivan, y Petrović, Nikola. *The IMO Compendium*. 2.ª ed. Problem Books in Mathematics. Nueva York: Springer-Verlag, 2011. https://doi.org/10.1007/978-1-4419-9854-5.

[15] Eberhard, Sean; Green, Ben, y Manners, Freddie. «Sets of integers with no large sum-free subset». En: *Annals of Mathematics. Second Series* 180.2 (2014), págs. 621-652. issn: 0003-486X. https://doi.org/10.4007/annals.2014.180.2.5.

[16] Erdös, Paul. «Some remarks on the theory of graphs». En: *Bulletin of the American Mathematical Society* 53 (1947), págs. 292-294. issn: 0002-9904. https://doi.org/10.1090/S0002-9904-1947-08785-1.

[17] Erdős, Paul. «Applications of probability to combinatorial problems». En: *Colloquium on Combinatorial Methods in Probability Theory*. 1962, págs. 90-92. url: https://users.renyi.hu/~p_erdos/Erdos.html.

[18] Erdős, Paul. «On a problem in graph theory». En: *The Mathematical Gazette* 47 (1963), págs. 220-223. issn: 0025-5572. https://doi.org/10.2307/3613396.

[19] Erdős, Paul. «Extremal problems in number theory». En: *Proceedings of Symposia in Pure Mathematics*. Vol. VIII. Providence, Rhode Island: American Mathematical Society, 1965, págs. 181-189. https://doi.org/10.1090/pspum/008.

[20] Erdös, Paul y Kac, Mark. «The Gaussian law of errors in the theory of additive number theoretic functions». En: *American Journal of Mathematics* 62 (1940), págs. 738-742. ISSN: 0002-9327. https://doi.org/10.2307/2371483.

[21] Erdős, Paul y Lovász, László. «Problems and results on 3-chromatic hypergraphs and some related questions». En: *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday)*. Vol. II. Colloquia Mathematica Societatis Janós Bolyai 10. Amsterdam, 1975. URL: https://www.renyi.hu/~p_erdos/1975-34.pdf.

[22] Fernández, Pablo y Fernández, José Luis. «El discreto encanto de la matemática». 2018. URL: http://verso.mat.uam.es/~pablo.fernandez/.

[23] Gowers, William T. «Quasirandom groups». En: *Combinatorics, Probability and Computing* 17.3 (2008), págs. 363-387. ISSN: 0963-5483. https://doi.org/10.1017/S0963548307008826.

[24] Green, Ben y Tao, Terence. «The primes contain arbitrarily long arithmetic progressions». En: *Annals of Mathematics. Second Series* 167.2 (2008), págs. 481-547. ISSN: 0003-486X. https://doi.org/10.4007/annals.2008.167.481.

[25] Loh, Po-Shen. *Probabilistic method in combinatorics*. 2009. URL: http://www.math.cmu.edu/~ploh/public_html/olympiad.shtml.

[26] Margulis, Grigori A. «Explicit constructions of expanders». En: *Problemy Peredači Informacii* 9.4 (1973), págs. 71-80. ISSN: 0555-2923.

[27] Mertens, Franz. «Ein Beitrag zur analytischen Zahlentheorie». En: *Journal für die reine und angewandte Mathematik* 78 (1874), págs. 46-62. ISSN: 0075-4102. https://doi.org/10.1515/crll.1874.78.46.

[28] Molloy, Michael y Reed, Bruce. *Graph colouring and the probabilistic method*. Vol. 23. Algorithms and Combinatorics. Berlín: Springer-Verlag, 2002. https://doi.org/10.1007/978-3-642-04016-0.

[29] Moser, Robin A. «A constructive proof of the Lovász local lemma». En: *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*. Nueva York: ACM, 2009, págs. 343-350. https://doi.org/10.1145/1536414.1536462.

[30] Moser, Robin A. y Tardos, Gábor. «A constructive proof of the general Lovász local lemma». En: *Journal of the ACM* 57.2 (2010). ISSN: 0004-5411. https://doi.org/10.1145/1667053.1667060.

[31] Ramsey, Frank P. «On a Problem of Formal Logic». En: *Proceedings of the London Mathematical Society. Second Series* 30.4 (1929), págs. 264-286. ISSN: 0024-6115. https://doi.org/10.1112/plms/s2-30.1.264.

[32] Riordan, Oliver. *Lecture notes on Probabilistic Combinatorics*. 2019. URL: https://courses.maths.ox.ac.uk/node/view_material/41048.

[33] Riordan, Oliver. *Oxford materials about the course on Probabilistic Combinatorics*. 2019. URL: https://courses.maths.ox.ac.uk/node/view_material/41304.

[34] Shearer, Jean B. «On a problem of Spencer». En: *Combinatorica* 5.3 (1985), págs. 241-245. ISSN: 0209-9683. https://doi.org/10.1007/BF02579368.

[35] Sipser, Michael y Spielman, Daniel A. «Expander codes». En: *IEEE Transactions on Information Theory* 42.6 (1996). Codes and complexity, págs. 1710-1722. ISSN: 0018-9448. https://doi.org/10.1109/18.556667.

[36] Spencer, Joel. «Ramsey's Theorem — A New Lower Bound». En: *Journal of Combinatorial Theory. Series A* 18.1 (1975), págs. 108-115. ISSN: 0097-3165. https://doi.org/10.1016/0097-3165(75)90071-0.

[37] Szele, Tibor. «Kombinatorikai vizsgálatok az irányitott teljes gráffal». En: *Matematikai és Fizikai Lapok* 50 (1943), págs. 223-256.

[38] Tao, Terence. *Moser's entropy compression argument*. 2009. URL: https://terrytao.wordpress.com/2009/08/05/mosers-entropy-compression-argument/.

[39] Tao, Terence y Vu, Van. *Additive combinatorics*. Vol. 105. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2006. https://doi.org/10.1017/CBO9780511755149.

[40] Turán, Paul. «On a Theorem of Hardy and Ramanujan». En: *The Journal of the London Mathematical Society* 9.4 (1934), págs. 274-276. issn: 0024-6107. https://doi.org/10.1112/jlms/s1-9.4.274.

[41] Wikipedia. *List of probabilistic proofs of non-probabilistic theorems*. En: *Wikipedia, The Free Encyclopedia*. 2019. url: https://en.wikipedia.org/w/index.php?title=List_of_probabilistic_proofs_of_non-probabilistic_theorems&oldid=910281376.

[42] Wikipedia. *Lovász local lemma*. En: *Wikipedia, The Free Encyclopedia*. 2019. url: https://en.wikipedia.org/w/index.php?title=Lov%5C%C3%5C%A1sz_local_lemma&oldid=895002723.

[43] Wikipedia. *Problema de satisfacibilidad booleana*. En: *Wikipedia, La enciclopedia libre*. 2019. url: https://es.wikipedia.org/w/index.php?title=Problema_de_satisfacibilidad_booleana&oldid=117829273.

[44] Wikipedia. *Ramsey's theorem*. En: *Wikipedia, The Free Encyclopedia*. 2020. url: https://en.wikipedia.org/w/index.php?title=Ramsey%5C%27s_theorem&oldid=937680845.

A. Anexo: soluciones a los ejercicios planteados

Solución del ejercicio 1. Con el objetivo de construir un subgrafo bipartito de G, consideramos una partición aleatoria de los vértices de G en dos conjuntos A y B. Esta partición se dará a partir de elegir de manera independiente para cada vértice de G el conjunto al que pertenecen, A o B, cada uno con probabilidad $1/2$. Para cada tal partición de $V(G)$ aleatoria, quitamos las aristas de G que unan dos vértices de A o dos vértices de B, dando así lugar a un subgrafo bipartito G_{bip} de G con partición de vértices $V(G_{\text{bip}}) = A \cup B$ (con la notación de la definición 14). Sea X la variable aleatoria correspondiente al número de aristas de este grafo bipartito G_{bip} aleatorio. Probaremos que $\mathbb{E}[X] = m/2$ y la conclusión del ejercicio 1 se seguirá directamente del segundo punto de la proposición 8. Vemos que $X = \sum_{e \in E(G)} X_{e \in E(G_{\text{bip}})}$ y, además, dado $e \in E(G)$, $\mathbb{P}(e \in E(G_{\text{bip}})) = \mathbb{P}(\text{los extremos de } e \text{ están en el mismo conjunto de la partición}) = 1/2$. Por la proposición 7, $\mathbb{E}[X] = m/2$. ∎

Solución del ejercicio 2. Tomaremos al azar un par de comités. Sea X el número de delegados en común que tienen ambos comités, de modo que $X = X_1 + ... + X_{1600}$, donde cada X_i es una indicatriz que toma el valor 1 si y solo si el i-ésimo delegado está en ambos comités seleccionados. Supongamos que el delegado i-ésimo aparece en a_i comités. Entonces, por un argumento de doble conteo, $a_1 + ... + a_{1600} = 16000 \cdot 80$. Para estimar $\mathbb{E}[X]$ también es necesario hacer uso de un caso particular de la desigualdad de Jensen.

Lema 38 (Desigualdad de Jensen). *Dada una función $f : \mathbb{R} \to \mathbb{R}$ convexa, para cualesquiera $x_1, ..., x_n \in \mathbb{R}$ se cumple lo siguiente:*

$$\sum_{k=1}^{n} \frac{f(x_k)}{n} \geq f\left(\sum_{k=1}^{n} \frac{x_k}{n}\right).$$

Vamos a aplicar este lema a la función $f(x) = x(x-1)/2$ que, por tener segunda derivada positiva, es convexa. Notemos que, si $n \geq 0$ es entero, entonces $f(n) = \binom{n}{2}$. Tenemos que

$$\mathbb{E}[X] = \sum_{i=1}^{1600} \mathbb{P}(X_i = 1) = \frac{\sum_{k=1}^{1600} f(a_i)}{\binom{16000}{2}} \geq \frac{1600 \cdot f\left(\frac{16000 \cdot 80}{1600}\right)}{\binom{16000}{2}} \approx 3{,}995 > 3$$

y, por la proposición 8, se garantiza la existencia de un par P de comités tal que $X(P) \geq \mathbb{E}[X] > 3$ y, como X toma valores enteros, $X(P) \geq 4$ como queríamos. ∎

Solución del ejercicio 3. Denotamos por G cualquier subgrafo de $\mathcal{K}_{n,n}$ (véase la definición 14). Una partición de las descritas en el enunciado consiste en elegir una permutación $\sigma \in S_n = \text{Sim}([n])$ tal que $\{a_i, b_{\sigma(i)}\}$ es una arista de G para todo $i \in [n]$. Para probar que existe tal permutación, elegiremos $\sigma \in \text{Sim}([n])$ uniformemente al azar, de entre las $n!$ permutaciones posibles, y probaremos que la variable aleatoria X definida por $X(\sigma) = |\{i \in [n] : \{a_i, b_{\sigma(i)}\} \in E(G)\}|$ tiene esperanza mayor que $n-1$. Notemos que $X = \sum_{e \in E(G)} X_e$, donde cada variable X_e es la indicatriz del suceso $\{\exists\, i \in [n] : e = \{a_i, b_{\sigma(i)}\}\}$. Se calcula fácilmente que $\mathbb{P}(X_e = 1) = (n-1)!/n! = 1/n$ y, por la proposición 7, $\mathbb{E}[X] = (n^2 - n + 1)/n > n - 1$. Por

tanto, por el segundo punto de la proposición 8, para algún emparejamiento σ tendremos que $X(\sigma) > n-1$ y, por tanto, $X(\sigma) = n$, lo cual termina la solución del ejercicio.

Queremos remarcar que el valor de n^2-n+1 es el mejor posible. En efecto, puede considerarse un subgrafo G en el que $\{a_i, b_j\}$ sea una arista si y solo si $j \neq 1$. En este caso, tenemos $n(n-1) = n^2 - n$ aristas y no es posible hacer un emparejamiento de los descritos. Como curiosidad, en este caso la variable aleatoria X sería constante e igual a $n-1$. ∎

Solución del ejercicio 4. Teniendo en cuenta que $R(2,2) = 2$, la desigualdad se cumple para $k = 2$. Vamos a comprobarla para $k \geq 3$ y, para ello, realizaremos la coloración de un grafo n-completo uniformemente al azar y probaremos que, con probabilidad positiva, no tendrá ningún subgrafo k-completo monocromático si $n \leq 2^{k/2}$. Tenemos $\binom{n}{k}$ subgrafos completos de k elementos en el grafo completo $\mathcal{K}_n$. Los enumeramos y, para cada uno de ellos, definimos el suceso malo $A_i = \{$el subgrafo k-completo i-ésimo es monocromático$\}$. Se tiene que $\mathbb{P}(A_i) = 2^{1-\binom{k}{2}}$. Observamos la siguiente cota para los coeficientes binomiales: $\binom{n}{k} \leq n^k/k!$, directo porque $\binom{n}{k} = \frac{1}{k!}\prod_{i=0}^{k-1}(n-i)$. Entonces,

$$\sum_{1\leq i\leq \binom{n}{k}} \mathbb{P}(A_i) = \binom{n}{k}\frac{2}{2^{\binom{k}{2}}} \leq \frac{n^k 2^{k/2+1}}{k! 2^{k^2/2}} \leq \frac{2^{k/2+1}}{k!} < 1,$$

usando que $k \geq 3$ en el último paso. La conclusión se sigue del primer punto de la proposición 8.

Tras el teorema 3.1.1 del libro de Alon y Spencer [3] se observa que puede tomarse $n = (1+o(1))k2^{k/2}/(\mathrm{e}\sqrt{2})$ para que se siga cumpliendo la desigualdad $\binom{n}{k}2^{1-\binom{k}{2}} < 1$. Esto puede hacerse estimando mejor los coeficientes binomiales con la fórmula de Stirling. De este modo, la misma prueba que hemos dado, acompañada de un mejor análisis de los posibles n como acabamos de comentar, nos lleva a concluir que $R(k,k) \geq (1+o(1))k2^{k/2}/(\mathrm{e}\sqrt{2})$, mejorando claramente la anterior desigualdad. ∎

Solución del ejercicio 5. Por la simetría de filas y columnas en el problema anterior, nos referiremos a ellas indistintamente como líneas. Vamos escoger una línea al azar y uniformente para después estudiar la variable aleatoria X, que devuelve, para cada línea L, la cantidad $X(L)$ de números distintos en dicha línea.

Tenemos que X es la suma de n variables indicatrices, $X = X_1 + \ldots + X_n$, donde cada X_k vale 1 en cada línea que contenga a k y 0 en el resto. Para cada k,

$$\mathbb{E}[X_k] = \frac{1}{2n}\left(\sum_{L \text{ línea}} X_k(L)\right) = \frac{|\{L \text{ línea} : L \text{ contiene a } k\}|}{2n}.$$

Digamos que k aparece en a_k filas y en b_k columnas exactamente. Entonces, k aparece en exactamente $a_k + b_k$ líneas. Por otro lado, tenemos que k aparece n veces en la cuadrícula y dichas casillas deben estar en una de las a_k filas y en una de las b_k columnas anteriormente descritas. Así que k puede aparecer en a lo sumo $a_k b_k$ casillas. Por la versión más simple de la desigualdad aritmético-geométrica tenemos que $a_k + b_k \geq 2\sqrt{a_k b_k} \geq 2\sqrt{n}$. Por tanto, para cada $k \in [n]$, $\mathbb{E}[X_k] = (a_k + b_k)/2n \geq 1/\sqrt{n}$. Finalmente, $\mathbb{E}[X] = \sum_{k=1}^n \mathbb{E}[X_k] \geq \sum_{k=1}^n 1/\sqrt{n} = \sqrt{n}$, así que, por el segundo punto de la proposición 8, para alguna línea L tenemos que $X(L) \geq \sqrt{n}$, como queríamos. ∎

Solución del ejercicio 6. El enunciado es trivial para $N = 1$. Estudiemos ahora el caso $N > 1$. Vamos a hacer N elecciones de elementos de G, con posibles repeticiones, y a formar así nuestro conjunto B. Aplicaremos el segundo punto de la proposición 8 a la variable aleatoria $X = |A + B|$, que verifica que $\mathbb{E}[X] \geq N^2/2$. Podemos escribir X como suma de indicadores X_k, con $k \in G$, que toman el valor 1 si y solo si $k \in A + B$. Esto se da precisamente cuando $k - a \in B$ para algún $a \in A$, es decir, si alguno de tales N elementos está en B. Así que $\mathbb{P}(X_k = 1) = 1 - (1 - N/N^2)^N$. Finalmente, $\mathbb{E}[X] = N^2(1 - (1 - 1/N)^N) \geq N^2(1 - 1/\mathrm{e}) \geq N^2/2$, haciendo uso del lema 32 para $n = -N < -1$. ∎

Solución del ejercicio 7. Vamos a proceder de manera estándar. Consideramos una coloración que asocie a cada arista del grafo un color de manera uniforme e independiente al resto. Los malos sucesos que vamos a estudiar, de los cuales queremos probar que su unión no cubre todo el espacio de probabilidad, son precisamente aquellos en los que alguno de los triángulos es monocromático. No necesitamos saber

cuántos sucesos malos hay, aunque sea inmediato de contar. Solamente necesitamos comprobar la condición local. La probabilidad de que un triángulo cualquiera sea monocromático es $p = k/k^3 = 1/k^2$. Además, el suceso que consiste en que un triángulo T sea monocromático es dependiente del suceso que consiste en que el triángulo $T' \neq T$ sea monocromático si y solo si T y T' tienen una arista en común. Fijado T, tenemos $3(n-3)$ tales T', así que $d = 3n-9$ con la notación del lema 31 y $\mathrm{e}p(d+1) \leq \mathrm{e}\frac{1}{9n}(3n) < 1$. ∎

Solución del ejercicio 8. Vamos a optar por elegir, para cada uno de los n colores, uno de los once puntos de dicho color uniformemente al azar. Es fácil darse cuenta de cuáles son los sucesos malos: para cada pareja de puntos consecutivos con distinto color, un suceso malo sería el que corresponda a seleccionar ambos puntos tras realizar la anterior elección aleatoria. Llamaremos pareja especial a una tal pareja de puntos.

Consideramos un conjunto de índices I que enumera las parejas especiales. Para cada $i \in I$, se denota por P_i la pareja especial asociada al índice i, y por S_i el suceso malo correspondiente que describimos antes. Para todo i, se tiene que $p = \mathbb{P}(A_i) = 1/11^2$, ya que cada punto de la pareja P_i tiene probabilidad $1/11$ de ser elegido y cada elección es independiente (porque tienen distintos colores, por asunción).

Además, podemos tener una cota en el número de dependencias. Consideremos una pareja P_i especial, cuyos puntos tienen los colores C_1 y C_2. Ahora afirmamos que A_j va a ser independiente de la familia de sucesos a los que A_i pertenece si y solo si los colores de su pareja P_i son ambos distintos de C_1 y C_2. Esto se debe a que la elección de cada uno de los n puntos se hace para cada color de manera independiente. Ahora vamos a contar cuántas parejas especiales P_k van a tener alguno de sus colores igual a C_1 o C_2. Cada una de esas parejas P_k debe contener un punto D_k de color C_1 o C_2 (hay a lo sumo 22 puntos con estas condiciones) y el otro punto E_k que forme la pareja $P_k = \{D_k, E_k\}$ debe ser consecutivo a D_k. Así que, fijado D_k, hay como máximo dos opciones para E_k, y por ello hay como máximo $22 \times 2 = 44$ tales parejas especiales, de entre las cuales debemos eliminar la propia pareja del suceso A_j fijada inicialmente. Con la notación del lema 31, hemos comprobado que podemos tomar $p = 1/11^2$ y $d+1 = 44$. Se comprueba simplemente que $\mathrm{e}p(d+1) < 1$ y, así, por el lema 31, habrá una posible elección de n puntos tal que ninguno de los sucesos malos se da, que es precisamente la conclusión del enunciado. ∎

TEMat

q-medidas de Carleson en espacios de Hardy H^p y Bergman A^p_α

Tanausú Aguilar
Universidad de Sevilla
taguilar@us.es

Sergi Baena
Universitat de Barcelona
sergibaena@ub.edu

Carlos Cruz
Universitat de Barcelona
ccruz@ub.edu

Jordi Lendínez
Universitat Autònoma de Barcelona
jordi.lendinez@kantarmedia.com

Alejandro Molero
Universitat Autònoma de Barcelona
amolero@mat.uab.cat

Marco Praderio
Universitat Autònoma de Barcelona
PraderioM@Gmail.com

Resumen: La caracterización de las q-medidas de Carleson para espacios de funciones holomorfas es un problema clásico en el análisis matemático. Presentamos un análisis de la caracterización clásica de este tipo de medidas para el espacio de Hardy H^p y el espacio de Bergman A^p_α.

Abstract: The characterization of q-Carleson measures for holomorphic function spaces is a classical problem in mathematical analysis. We give a survey of the classical characterization of this type of measures for the Hardy H^p and Bergman A^p_α spaces.

Palabras clave: q-medidas de Carleson, espacios de Hardy, espacios de Bergman, operador maximal, espacios de Hilbert con núcleo reproductor.

MSC2010: 30J99, 31A10.

Recibido: 26 de marzo de 2019.
Aceptado: 27 de marzo de 2020.

Agradecimientos: Queremos agradecer al Prof. J. A. Peláez por su inestimable ayuda en el análisis de este problema y en la depuración de la presentación de nuestro trabajo. También agradecemos a los organizadores de la última edición del *Workshop on Complex Analysis and Operator Theory*, así como a la Spanish Network on Complex Analysis and Operator Theory.

Referencia: Aguilar, Tanausú; Baena, Sergi; Cruz, Carlos; Lendínez, Jordi; Molero, Alejandro y Praderio, Marco. «q-medidas de Carleson en espacios de Hardy H^p y Bergman A^p_α». En: *TEMat*, 4 (2020), págs. 67-82. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p67.

1. Introducción

La caracterización de las q-medidas de Carleson para espacios de funciones holomorfas es un problema clásico en el análisis matemático. Estas fueron introducidas por L. Carleson en la década de 1960 [3, 4] para caracterizar las sucesiones de interpolación en el álgebra de Banach H^∞, formada por las funciones holomorfas acotadas en el disco unidad abierto $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, y dar una solución al problema de la corona.

Definición 1. Sea μ una medida de Borel positiva y $\mathcal{F}$ un espacio cuasinormado de funciones analíticas en $\mathbb{D}$. Diremos que μ es una q-medida de Carleson para $\mathcal{F}$ si el operador

$$\mathrm{Id} : \mathcal{F} \to L^q(\mu)$$

es acotado. ◂

La importancia de este tipo de medidas es ampliamente conocida. Existen múltiples aplicaciones en análisis armónico o en ecuaciones diferenciales, por ejemplo en la solución de la ecuación $\bar\partial$ dada por Jones [12]. Como observa Jones [11], existe un destacable uso de este tipo de medidas en la caracterización de funciones de oscilación media acotada (BMO por sus siglas en inglés). Fefferman y Stein [8] demuestran que $\phi \in \mathrm{BMO}(\mathbb{R})$ si y solo si su extensión armónica a $\mathbb{R}^2_+$ satisface que $y\nabla\phi\,\mathrm{d}x\,\mathrm{d}y$ es una medida de Carleson. Esto implica que el dual de H^1 se identifica con BMO. Sin embargo, hay que notar que este resultado se puede obtener sin el uso de este tipo de medidas (ver el libro de Pavlović [14, cap. 6]).

Presentamos un análisis de la caracterización clásica de este tipo de medidas, dada por Carleson, para el espacio de Hardy H^p y el espacio de Bergman con peso A^p_α en el disco unidad con $p \le q$. Para el caso $q < p$ son necesarias otras técnicas que exceden la pretensión de este texto. Además, analizamos algunas propiedades del operador maximal de Hörmander [10] para obtener el resultado. La importancia del esquema de demostración que desarrollamos en este artículo radica en que no requiere conocimientos avanzados. El resultado que mostramos se puede obtener a través del teorema de interpolación de Marcinkiewicz (ver el artículo de Bernard [1, Teorema 3.8]), pero ello implica un conocimiento más avanzado de teoría de la interpolación. Nótese que el esquema para el espacio de Hardy se puede adaptar al espacio de Bergman (ver también el artículo de Peláez y Rättyä [15, Theorem 4.19]). Sin embargo, optamos por otra vía también sencilla.

El artículo se divide en cuatro secciones. En la primera repasamos algunas definiciones básicas de los espacios de Hilbert con núcleo reproductor y definimos los espacios de Hardy y Bergman clásicos. En la segunda sección, definimos y analizamos las cajas de Carleson y el operador maximal de Hörmander. Finalizamos con una sección para la caracterización en cada espacio.

A lo largo del artículo utilizaremos la siguiente notación. Denotaremos por $(H, \langle\cdot,\cdot\rangle)$ a un espacio de Hilbert de funciones holomorfas con producto escalar $\langle\cdot,\cdot\rangle$, y por $\mathbb{T}$ a la frontera del disco unidad $\mathbb{D}$. El conjunto de funciones holomorfas en $\mathbb{D}$ estará denotado por $\mathcal{O}(\mathbb{D})$. Además, si $T : X \to Y$ es un operador lineal y continuo entre los espacios de Banach X e Y, denotaremos su norma por $\|T\|_{(X,Y)}$. Por otro lado, dados $A, B > 0$, si existe $c > 0$ tal que $A \le cB$, escribiremos $A \lesssim B$. Si $A \lesssim B$ y $B \lesssim A$, escribiremos $A \asymp B$.

Finalmente, sugerimos al lector interesado en profundizar en el teorema de interpolación ver el artículo de Bernard [1]; al interesado en los espacios de Bergman, leer los libros de Duren y Schuster [7] y Hedenmalm, Korenblum y Zhu [9], y al interesado en los espacios de Hardy, los libros de Duren [6], Koosis [13] y Pavlović [14]. Para profundizar en medidas de Carleson y aplicaciones de estas, se recomienda ver los artículos de Carleson [3, 4], Fefferman y Stein [8], Jones [11, 12] y Peláez y Rättyä [15]. Para el operador maximal de Hörmander se recomienda leer su artículo original [10], y para conceptos básicos de análisis complejo aconsejamos los libros de Bruna y Cufí [2] y Rudin [17]. Cabe destacar que los artículos de Fefferman y Stein [8] y Hörmander [10] están enfocados a varias variables complejas, por lo que recomendamos al lector que consulte los libros de Rudin [16] y Scheidemann [18] previamente a su lectura.

2. Espacios de Hilbert con núcleo reproductor

Consideremos H un espacio de Hilbert, $H \subset \mathcal{O}(\mathbb{D})$, tal que para todo punto $z \in \mathbb{D}$ el operador evaluación $L_z : H \to \mathbb{C}$ dado por $L_z(f) = f(z)$ está acotado. El teorema de representación de Riesz [17] asegura la existencia de una única función K_z tal que $\langle f, K_z \rangle = L_z(f) = f(z)$ para todo $z \in \mathbb{D}$. Dicha función se denomina *núcleo reproductor de* H.

Dada $(e_n)_{n\in\mathbb{N}}$ una base ortonormal de H, su núcleo reproductor será

$$K_z(\xi) = \sum_{n=1}^{\infty} e_n(\xi)\overline{e_n(z)}, \quad \xi, z \in \mathbb{D}.$$

Nos centraremos en la caracterización de las q-medidas de Carleson para los espacios de Hardy H^p y los espacios de Bergman A^p_α.

Dadas la función $f \in \mathcal{O}(\mathbb{D})$ y constantes $r \in (0,1)$ y $p \in (0,\infty)$, definimos la función

$$M_p(r,f) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(re^{i\theta})|^p \, d\theta\right)^{1/p}.$$

Definición 2. Sea $0 < p < \infty$. Definimos el **espacio de Hardy** H^p como el espacio de funciones $f \in \mathcal{O}(\mathbb{D})$ tales que

$$\|f\|_{H^p} = \sup_{0<r<1} M_p(r,f) = \lim_{r\to 1^-} M_p(r,f) < \infty. \qquad \blacktriangleleft$$

Nótese que la última igualdad es debida al teorema de convexidad de Hardy (véase el libro de Conway [5, cap. 6]). Por otro lado, sea $dA(re^{it}) = \frac{r}{\pi}\, dr\, dt$ y consideremos un peso w (es decir, $w \in L^1(\mathbb{D}, dA)$ con $w \geq 0$).

Definición 3. Sea $0 < p < \infty$. Definimos el **espacio de Bergman con peso** w, A^p_w, como el espacio de funciones $f \in \mathcal{O}(\mathbb{D})$ tales que

$$\|f\|_{A^p_w} = \left(\int_{\mathbb{D}} |f(z)|^p \, w(z)\, dA(z)\right)^{1/p} < \infty. \qquad \blacktriangleleft$$

En particular, nos centraremos en los pesos

$$w(z) = w_\alpha(z) = (\alpha+1)(1-|z|^2)^\alpha, \quad \alpha > -1,$$

por lo que de ahora en adelante denotaremos $A^p_w = A^p_\alpha$ para estos pesos.

Notemos ahora que tanto el espacio de Hardy H^2 como el espacio de Bergman A^2_α tienen estructura de espacio de Hilbert con los productos escalares respectivos,

$$\langle f, g\rangle_{H^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\left(e^{i\theta}\right)\overline{g}\left(e^{i\theta}\right) d\theta \qquad \text{y} \qquad \langle f, g\rangle_{A^2_\alpha} = \int_{\mathbb{D}} f(z)\,\overline{g}(z) w_\alpha(z)\, dA(z).$$

Estos espacios admiten una base ortonormal dada por los monomios normalizados. En el caso del espacio de Hardy H^2 tenemos que $e_n(z) = z^n$ y la expresión de su núcleo reproductor queda como

$$K_z(\xi) = \sum_{n=1}^{\infty} \xi^n \overline{z}^n = \frac{1}{1-\xi\overline{z}},$$

el cual normalizando queda

$$k_z(\xi) = \frac{K_z(\xi)}{\|K_z\|} = \frac{\left(1-|z|^2\right)^{1/2}}{1-\xi\overline{z}},$$

donde $\|K_z\|^2 = \left(1-|z|^2\right)^{-1}$. En el caso del espacio de Bergman A^2_α, los elementos de la base de monomios normalizados son

$$e_n(z) = \frac{z^n}{\sqrt{2\displaystyle\int_0^1 \rho^{2n+1} w_\alpha(\rho)\,\mathrm{d}\rho}},$$

los cuales nos permiten calcular el núcleo reproductor de A^2_α como

$$K_z(\xi) = \sum_{n=1}^{\infty} \frac{\xi^n \overline{z^n}}{2(\alpha+1)\displaystyle\int_0^1 \rho^{2n+1}(1-\rho^2)^\alpha\,\mathrm{d}\rho} = \frac{1}{(1-\xi\overline{z})^{2+\alpha}}.$$

Finalmente, normalizando el núcleo reproductor obtenemos que

$$k_z(\xi) = \frac{\left(1-|z|^2\right)^{(2+\alpha)/2}}{(1-\xi\overline{z})^{2+\alpha}}.$$

3. Cajas de Carleson y operador maximal de Hörmander

En esta sección introduciremos la noción de caja de Carleson asociada a un arco de $\mathbb{T}$ o a un punto del disco unidad $\mathbb{D}$. Posteriormente, hablaremos sobre el operador maximal de Hörmander, el cual nos será útil en la sección 4.

Definición 4. Sea $I \subset \mathbb{T}$ un intervalo. La **caja de Carleson** asociada al intervalo I es el conjunto

$$S(I) = \left\{re^{it} : 1-r \le |I|,\ e^{it} \in I\right\},$$

donde $|I|$ denota la medida de Lebesgue normalizada del intervalo I. De forma similar, se puede definir la caja de Carleson para cualquier punto $a = |a|e^{i\theta} \in \mathbb{D}\setminus\{0\}$ como

$$S(a) = \left\{re^{it} : 1-r \le 1-|a|,\ |\arg(\overline{a}z)| = |t-\theta| \le \frac{1-|a|}{2}\right\}.$$

Además, si definimos el arco $I_a \subset \mathbb{T}$ como

$$I_a = \left\{e^{it} : |t-\theta| \le \frac{1-|a|}{2}\right\},$$

tenemos que $|I_a| = 1-|a|$ y el arco asociado a la caja de Carleson $S(a)$ es I_a (ver figura 1). Vemos entonces que existe una biyección entre los conjuntos I_a y $S(a)$, $a \in \mathbb{D}\setminus\{0\}$. En particular, $S(a) = S(I_a)$. ◂

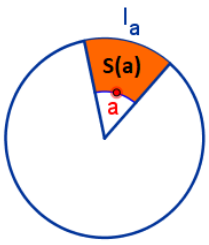


Figura 1: Caja de Carleson.

Una propiedad interesante de las cajas de Carleson es el siguiente resultado técnico, el cual nos permite obtener una condición necesaria para describir las q-medidas de Carleson en los espacios mencionados.

Lema 5. *Sean $a \in \mathbb{D} \setminus \{0\}$ y $z \in S(a)$. Entonces,*

$$\frac{1-|a|^2}{2} \leq |1-\overline{a}z| \leq (1-|a|^2)\frac{\sqrt{5}}{2}.$$

Demostración. Nótese que $(1+|a|)/2 \leq 1$, por lo que la primera desigualdad se obtiene de

$$|1-\overline{a}z| \geq 1-|az| \geq 1-|a| \geq (1-|a|)\left(\frac{1+|a|}{2}\right) = \frac{1-|a|^2}{2}.$$

Veamos la segunda desigualdad. Sea $\theta = \arg(\overline{a}z)$. Tenemos que

$$|1-\overline{a}z|^2 = (1-|\overline{a}z|)^2 + 4|a||z|(\operatorname{sen}(\theta/2))^2 \leq (1-|az|)^2 + |\theta|^2,$$

donde estamos usando que $|\operatorname{sen}(t)| \leq |t|$ para todo $t \in \mathbb{R}$. Finalmente, como $z \in S(a)$,

$$(1-|az|)^2 + |\theta|^2 \leq (1-|az|)^2 + \frac{(1-|a|)^2}{4} \leq (1-|a|^2)^2 + \frac{(1-|a|)^2}{4} \leq \frac{5}{4}(1-|a|^2)^2. \qquad \blacksquare$$

La siguiente clase de medidas quedarán caracterizadas cuando $\alpha = q/p$ como las q-medidas de Carleson para H^p.

Definición 6. Sea $\alpha > 0$. Si μ es una medida de Borel positiva, definimos

$$\|\mu\|_\alpha = \sup_{a\in\mathbb{D}} \frac{\mu(S(a))}{(1-|a|^2)^\alpha}.$$

Proposición 7. *Sean $0 < p, q < \infty$. Si μ es una q-medida de Carleson para H^p, existe una constante $c = c(p,q) > 0$ tal que*

$$\|\mu\|_{q/p} \leq c\,\|\mathrm{Id}\|^q_{(H^p, L^q(\mu))} < \infty.$$

Demostración. Por hipótesis, tenemos que $\mathrm{Id}: H^p \to L^q(\mu)$ es acotado. Tomamos $a \in \mathbb{D}$ y consideramos la familia de funciones $f_a(z) = k_a(z)^{2/p}$, donde k_a consiste en el núcleo reproductor normalizado

$$k_a(z) = \frac{(1-|a|^2)^{1/2}}{1-\overline{a}z}, \quad z \in \mathbb{D}.$$

Entonces, como $\|k_a\|_{H^2} = 1$,

$$\|\mathrm{Id}\|^q_{(H^p,L^q(\mu))} = \|k_a\|^{2q/p}_{H^2}\,\|\mathrm{Id}\|^q_{(H^p,L^q(\mu))} = \|f_a\|^q_{H^p}\,\|\mathrm{Id}\|^q_{(H^p,L^q(\mu))} \geq \|\mathrm{Id}(f_a)\|^q_{L^q(\mu)} = \|f_a\|^q_{L^q(\mu)}. \tag{1}$$

Ahora bien, $f_a \in H^p$ y

$$\begin{aligned}\|f_a\|^q_{L^q(\mu)} &= \int_{\mathbb{D}} |f_a(z)|^q\,\mathrm{d}\mu(z) \geq \int_{S(a)} |f_a(z)|^q\,\mathrm{d}\mu(z) = \int_{S(a)} |k_a(z)|^{2q/p}\,\mathrm{d}\mu(z)\\ &= \int_{S(a)} \left|\frac{(1-|a|^2)^{1/2}}{1-\overline{a}z}\right|^{2q/p} \mathrm{d}\mu(z) \geq \left(\frac{2}{\sqrt{5}(1-|a|^2)^{1/2}}\right)^{2q/p} \mu(S(a)) = (4/5)^{q/p}\frac{\mu(S(a))}{(1-|a|^2)^{q/p}},\end{aligned} \tag{2}$$

donde en la última desigualdad hemos usado que $|1-\overline{a}z| \leq \frac{\sqrt{5}}{2}(1-|a|^2)$ para $z \in S(a)$ (ver lema 5). Como esto es cierto para toda $a \in \mathbb{D}$, de (1) y (2), y tomando $c = (4/5)^{-q/p}$, deducimos que

$$c\,\|\mathrm{Id}\|^q_{(H^p,L^q(\mu))} \geq \sup_{a\in\mathbb{D}}\left(\frac{\mu(S(a))}{(1-|a|^2)^{q/p}}\right) = \|\mu\|_{q/p}. \qquad \blacksquare$$

Más adelante veremos que, si $q \geq p$, entonces la inclusión $H^p \subset L^q(\mu)$ es acotada y existe $C > 0$ tal que $\|\mu\|_{q/p} \geq C\,\|\mathrm{Id}\|^q_{(H^p,L^q(\mu))}$. Por lo tanto, en este caso, $\|\mu\|_{q/p} \asymp \|\mathrm{Id}\|^q_{(H^p,L^q(\mu))}$ (véase el teorema 13). Para eso, necesitaremos el siguiente lema de recubrimiento.

Lema 8 (Lema de recubrimiento 1). *Sea $A \subset \mathbb{D}$, $A \neq \varnothing$. Supongamos que no existe una sucesión infinita de puntos $(z_n)_{n\in\mathbb{N}} \subset A$ tales que los intervalos asociados $(I_{z_n})_{n\in\mathbb{N}}$ sean disjuntos dos a dos. Entonces, existe una sucesión finita de puntos $z_1, \dots, z_m \in A$ tales que los arcos $I_1, \dots, I_m$ son disjuntos y*

$$A \subset \bigcup_{n=1}^{m} \{z \in \mathbb{D} : I_z \subset 5I_{z_n}\},$$

donde $5I_{z_n} = \left\{e^{it} : |t-\theta_n| \leq \frac{5(1-|z_n|)}{2}\right\}$.

Demostración. Para la prueba vamos a construir una sucesión de puntos $z_1, \dots, z_m \in A$ que va a satisfacer la hipótesis que queremos.

Sean $A_1 = A$ y $\rho_1 = \inf\{|z| : z \in A_1\}$. Si $0 \in A_1$, elegimos $z_1 = 0$; en caso contrario, elegimos $z_1 \in A_1$ tal que $|z_1| \leq (\rho_1+1)/2$, lo cual es posible ya que, si $|z| > (\rho_1+1)/2$ para todo $z \in A_1$, entonces tendríamos que

$$\frac{\rho_1+1}{2} \leq \inf\{|z| : z \in A_1\} = \rho_1 < 1,$$

y llegaríamos a una contradicción. Sea ahora $A_2 = \{z \in A_1 : I_z \cap I_{z_1} = \varnothing\}$. Si $A_2 = \varnothing$, escogemos $m = 1$ y ya tenemos la sucesión buscada. Si $A_2 \neq \varnothing$, definimos $\rho_2 = \inf\{|z| : z \in A_2\}$ y elegimos $z_2 \in A_2$ tal que $|z_2| \leq (\rho_2+1)/2$.

De esta manera, se puede crear una sucesión de puntos $z_1, z_2, z_3, \dots$ tales que $z_n \in A_n$ y $|z_n| \leq (\rho_n+1)/2$, donde $A_n = \{z \in A_{n-1} : I_z \cap I_{z_{n-1}} = \varnothing\} \neq \varnothing$ y $\rho_n = \inf\{|z| : z \in A_n\}$ para $n \geq 2$. Por hipótesis, el proceso de escoger las z_n debe parar (es decir, existe un $n \in \mathbb{N}$ tal que $A_n = \varnothing$, por lo que ya tendríamos la sucesión buscada), ya que no existe ninguna colección infinita de puntos $(z_n)_{n\in\mathbb{N}} \subset A$ tales que los intervalos asociados $(I_{z_n})_{n\in\mathbb{N}}$ sean disjuntos dos a dos.

Por lo tanto, existe un entero $m \geq 1$ tal que $A_n \neq \varnothing$ para todo $1 \leq n \leq m$ pero $A_{m+1} = \varnothing$. Entonces, dado $z \in A$, existe algún $n_0 \in \{1, \dots, m\}$ tal que $I_z \cap I_{z_{n_0}} \neq \varnothing$. Fijemos el subíndice n_0 (si hubiera más de uno, escogemos n_0 como el mínimo entre estos subíndices). Obsérvese que, si $n_0 > 1$, entonces $I_z \cap I_{z_j} = \varnothing$ para todo $j \in \{1, \dots, n_0-1\}$. Por tanto,

$$|z| \geq \rho_{n_0} = \inf\{|z| : z \in A, I_z \cap I_{z_j} = \varnothing,\ 1 \leq j \leq n_0-1\}.$$

Así pues, como $2|z_{n_0}| \leq \rho_{n_0} + 1$,

$$|I_z| = 1-|z| \leq 1-\rho_{n_0} \leq 1+1-2|z_{n_0}| = 2(1-|z_{n_0}|) = 2|I_{z_{n_0}}|.$$

Concluimos que $I_z \cap I_{z_{n_0}} \neq \varnothing$ y $|I_z| \leq 2|I_{z_{n_0}}|$. Por lo tanto, $I_z \subset 5I_{z_{n_0}}$.

Si $n_0 = 1$, hacemos el mismo argumento pero ahora con ρ_1 y z_1 en lugar de ρ_{n_0} y z_{n_0}. ∎

Para acabar esta sección, vamos a introducir el operador maximal de Hörmander y daremos algunos resultados interesantes sobre este operador.

Definición 9. Sean $\varphi \in L^1(\mathbb{T})$ y $z \in \mathbb{D}$. El **operador maximal de Hörmander** se define como

$$\tilde{\mathcal{M}}(\varphi)(z) = \sup_{I \supset I_z} \frac{1}{|I|} \int_I |\varphi(e^{it})|\, dt. \qquad \blacktriangleleft$$

El siguiente resultado nos da una manera equivalente de escribir el operador de Hörmander.

Lema 10. *Si $z = re^{i\theta} \in \mathbb{D}$, definimos*

$$\mathcal{M}_r(\varphi)(e^{i\theta}) = \sup_{I \ni e^{i\theta},\ |I| \geq |I_z|} \frac{1}{|I|} \int_I |\varphi(e^{it})|\, dt.$$

Entonces, $\mathcal{M}_r(\varphi)(e^{i\theta}) \asymp \tilde{\mathcal{M}}(\varphi)(z)$ para todo $z = re^{i\theta} \in \mathbb{D}$.

Demostración. Observamos que para todo arco $I \supset I_z$ tenemos que $|I| \geq |I_z|$. Además, $e^{i\theta} \in I_z \subset I$, de lo que se obtiene la primera desigualdad

$$\tilde{\mathcal{M}}(\varphi)(z) \leq \mathcal{M}_r(\varphi)(e^{i\theta}).$$

Para ver la segunda desigualdad, sea I un intervalo tal que $e^{i\theta} \in I$ y $|I| \geq |I_z|$. Entonces, $e^{i\theta} \in I_z \cap I \neq \emptyset$. Así pues, existe un arco $\tilde{I} \supset I$ tal que $|\tilde{I}| = 2|I|$ y $I_z \subset \tilde{I}$. Por lo tanto,

$$\frac{1}{|I|}\int_I |\varphi(e^{it})|\,dt \leq 2\frac{1}{2|I|}\int_{\tilde{I}} |\varphi(e^{it})|\,dt = 2\frac{1}{|\tilde{I}|}\int_{\tilde{I}} |\varphi(e^{it})|\,dt \leq 2\tilde{\mathcal{M}}(\varphi)(z),$$

y la desigualdad que queremos ver se sigue tomando supremo sobre los arcos I tales que $e^{i\theta} \in I$ y $|I| \geq |I_z|$. ∎

La ventaja de esta equivalencia es que podemos probar una desigualdad puntual para este operador cuando actúa sobre espacios de Hardy.

Lema 11. *Sea $p > 0$. Existe una constante $C = C(p)$ tal que*

$$|f(z)|^p \leq C\tilde{\mathcal{M}}(|f|^p)(z),$$

para todo $z \in \mathbb{D}$ y para toda $f \in H^p$.

Demostración. En virtud del lema 10, basta probar la desigualdad

$$|f(z)|^p \leq C'\mathcal{M}_r(|f|^p)(e^{i\theta}), \quad z \in \mathbb{D},\ f \in H^p,$$

donde $C' = C'(p)$. Dada $f \in H^p$, observemos que, al ser f holomorfa, puede probarse que $|f|^p$ es subarmónica en el disco unidad. Por otro lado, recordemos que, si denotamos por

$$P_r(t) = \frac{1-r^2}{1-2r\cos t + r^2}, \quad t \in [-\pi, \pi],$$

el *núcleo de Poisson* con $r \in (0,1)$ y tomamos $g \in L^1(\mathbb{T})$, la función

$$u_g(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta - t)g(e^{it})\,dt, \quad \theta \in [-\pi, \pi],$$

es una función armónica en $\mathbb{D}$ cuyo límite radial en $\mathbb{T}$ existe y coincide con $g(z)$ para casi todo $z \in \mathbb{T}$ [17, Teorema 11.16]. Como $f \in H^p$, puede probarse que $|f|^p$ restringida a $\mathbb{T}$ es una función de $L^1(\mathbb{T})$. Así,

$$u_{|f|^p}(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta - t)|f(e^{it})|^p\,dt$$

es una función armónica que coincide con $|f|^p$ para casi todo punto en $\mathbb{T}$. Pero, como $|f|^p$ es subarmónica, concluimos por definición que

$$|f(re^{i\theta})|^p \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta - t)|f(e^{it})|^p\,dt.$$

Supongamos que $r \leq 1/2$. Se sigue que $P_r(t) \leq 3$ para todo $t \in [-\pi, \pi]$. De esta manera, sea $z = re^{i\theta} \in \mathbb{D}$. Entonces,

$$|f(z)|^p \leq 3\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(e^{it})|^p\,dt \leq 3\sup_{e^{i\theta}\in I, |I|\geq|I_z|}\frac{1}{|I|}\int_I |f(e^{it})|^p\,dt = 3\mathcal{M}_r(|f|^p)(e^{i\theta}).$$

Por otro lado, si tomamos $1 > r > 1/2$, observamos que no se puede acotar el núcleo de Poisson uniformemente en t y en r, debido a que, para $t = 0$, tenemos que $P_r(0) \to \infty$ cuando $r \to 1^-$. En este caso procedemos de la siguiente manera. Para todo $n = 1, 2, 3, \ldots$, definimos $t_n = 2^{n-1}\pi(1-r)$.

Notemos que $(t_n)_{n\in\mathbb{N}}$ es una sucesión creciente en la que $t_{n+1} = 2t_n$ y $t_1 = \pi(1-r)$. Así, al tener que $1-r < 1/2$, existe un único natural N tal que $t_N < \pi/2$ y $t_{N+1} \geq \pi/2$. De esta forma (ver figura 2), definimos

$$J_n = [-t_n, t_n], \quad n = 1, 2, \dots, N+1,$$

y

$$G_n = \begin{cases} J_1 & \text{si } n = 1, \\ J_n \setminus J_{n-1} & \text{si } n = 2, 3, \dots, N, \\ [-\pi, \pi] \setminus J_{N+1} & \text{si } n = N+1, \end{cases}$$

y vamos a encontrar una cota conveniente para cada G_n. Nótese que, para cada $n = 1, 2, \dots, N+1$,

$$P_r(t) \leq \frac{1}{4^{n-2}(1-r)}, \quad t \in G_n.$$

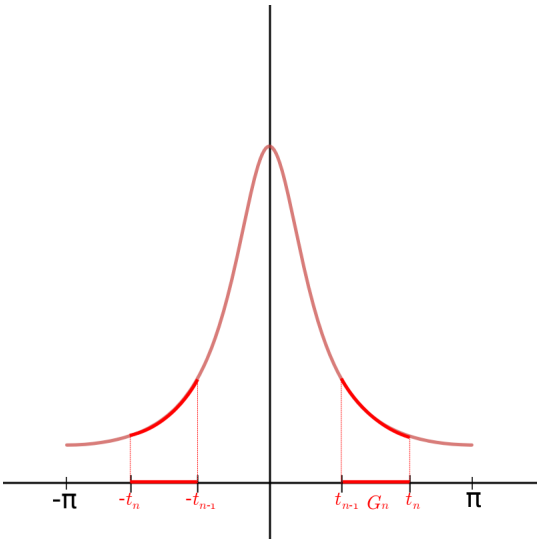


Figura 2: Conjunto G_n y núcleo de Poisson.

En efecto, fijado r, el núcleo de Poisson es una función par en t y decreciente para $t \in [0, \pi]$. Así,

$$\begin{aligned} P_r(t) &\leq P_r(0), && t \in G_1, \text{ y} \\ P_r(t) &\leq P_r(t_{n-1}), && t \in G_n \ \forall\, n \geq 2. \end{aligned}$$

Además, como $\cos t \leq 1 - 2t^2/\pi^2$, se sigue que

$$P_r(t_{n-1}) = \frac{1-r^2}{(1-r)^2 + 2r(1-\cos t_{n-1})} \leq \frac{1-r^2}{\frac{4rt_{n-1}^2}{\pi^2}} \leq \frac{1-r^2}{\frac{2t_{n-1}^2}{\pi^2}} = \frac{8(1+r)}{4^n(1-r)} \leq \frac{1}{4^{n-2}(1-r)}.$$

Obtenemos que

$$\begin{aligned} |f(re^{i\theta})|^p &\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta - t)|f(e^{it})|^p \,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(t-\theta)|f(e^{it})|^p \,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(t)|f(e^{i(t+\theta)})|^p \,dt = \frac{1}{2\pi}\sum_{n=1}^{N+1}\int_{G_n} P_r(t)|f(e^{i(t+\theta)})|^p \,dt \\ &\leq \frac{1}{2\pi}\sum_{n=1}^{N+1}\frac{1}{4^{n-2}(1-r)}\int_{G_n}|f(e^{i(t+\theta)})|^p \,dt \leq 8\sum_{n=1}^{N+1}\frac{1}{2^n}\frac{1}{|J_n|}\int_{J_n}|f(e^{i(t+\theta)})|^p \,dt \\ &\leq 8\left(\sup_{e^{i\theta}\in I, |I|\geq|I_z|}\frac{1}{|I|}\int_I |f(e^{it})|^p \,dt\right)\left(\sum_{n=1}^{\infty}\frac{1}{2^n}\right) = 16\mathcal{M}_r(|f|^p)(e^{i\theta}). \end{aligned}$$

■

Para el operador de Hörmander se tiene la siguiente desigualdad débil.

Lema 12 (Desigualdad débil). *Sean $q \geq p > 0$ y sea μ una medida de Borel positiva que satisface que $\|\mu\|_{p/q} < \infty$. Entonces, existe una constante $c = c(p,q) > 0$ tal que, para cada función $\varphi \in L^1(\mathbb{T})$,*

$$\mu(\{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi)(z) > \lambda\}) \leq c \frac{\|\mu\|_{q/p}}{\lambda^{q/p}} \|\varphi\|_{L^1(\mathbb{T})}^{q/p}, \quad \lambda > 0.$$

Demostración. Tomamos $\lambda > 0$ y definimos $E_\lambda = \{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi)(z) > \lambda\}$. Si $E_\lambda = \emptyset$, el resultado es trivial. Supongamos, pues, que $E_\lambda \neq \emptyset$. Para cualquier $\varepsilon > 0$, definimos los conjuntos

$$A_\lambda^\varepsilon = \left\{ z \in \mathbb{D} \ : \ \int_{I_z} |\varphi(e^{it})| \, dt > \lambda(|I_z| + \varepsilon) \right\} \quad \text{y} \quad B_\lambda^\varepsilon = \left\{ z \in \mathbb{D} : I_z \subset I_w, \text{ para algún } w \in A_\lambda^\varepsilon \right\}.$$

Nótese que los conjuntos A_λ^ε y B_λ^ε crecen si ε decrece. Además, tenemos que

$$E_\lambda \subseteq \bigcup_{\varepsilon > 0} B_\lambda^\varepsilon,$$

ya que para $z \in E_\lambda$, por la definición de supremo, existen $\varepsilon > 0$ y $w \in \mathbb{D}$ tales que $w \in A_\lambda^\varepsilon$ y $I_z \subset I_w$; en consecuencia, $z \in B_\lambda^\varepsilon$. Por lo tanto,

$$\mu(E_\lambda) \leq \lim_{\varepsilon \to 0^+} \mu(B_\lambda^\varepsilon).$$

Ahora observemos que, para cada $\varepsilon > 0$, no hay infinitos puntos $(z_n)_{n \in \mathbb{N}}$ en A_λ^ε tales que los intervalos $(I_{z_n})_{n \in \mathbb{N}}$ sean disjuntos. En caso contrario, si $(z_n)_{n \in \mathbb{N}} \subset A_\lambda^\varepsilon$, entonces

$$\lambda(|I_{z_n}| + \varepsilon) < \int_{I_{z_n}} |\varphi(e^{it})| \, dt, \quad \forall \, n \in \mathbb{N},$$

y obtenemos que

$$\lambda \sum_{n=1}^{\infty} (|I_{z_n}| + \varepsilon) < \sum_{n=1}^{\infty} \int_{I_{z_n}} |\varphi(e^{it})| \, dt = \int_{\bigcup_{n=1}^{\infty} I_{z_n}} |\varphi(e^{it})| \, dt \leq \int_{\mathbb{T}} |\varphi(e^{it})| \, dt = \|\varphi\|_{L^1(\mathbb{T})} < +\infty, \tag{3}$$

lo cual no es posible. Como $E_\lambda \neq \emptyset$, tenemos que $A_\lambda^\varepsilon \neq \emptyset$ para ε suficientemente pequeño. Así pues, el lema de recubrimiento 1 (lema 8) nos asegura que existen puntos $z_1, \ldots, z_m \in A_\lambda^\varepsilon$ tales que

$$A_\lambda^\varepsilon \subset \bigcup_{n=1}^{m} \{z \in \mathbb{D} : I_z \subset 5I_{z_n}\}.$$

De aquí se sigue que

$$B_\lambda^\varepsilon \subset \bigcup_{n=1}^{m} \{z \in \mathbb{D} : I_z \subset 5I_{z_n}\},$$

ya que si $z \in B_\lambda^\varepsilon$, entonces $I_z \subset I_w$ para algún $w \in A_\lambda^\varepsilon$, y $I_w \subset 5I_{z_n}$ para algún $n \in \{1, \ldots, m\}$. Además, dado que $I_{z_1} \subset I_{z_2}$ implica que $S(I_{z_1}) \subset S(I_{z_2})$ para $z_1, z_2 \in \mathbb{D}$, y tenemos que $q \geq p$ y la medida de Lebesgue es doblante, deducimos que

$$\begin{aligned} \mu(B_\lambda^\varepsilon) &\leq \sum_{n=1}^{m} \mu(\{z \in \mathbb{D} : I_z \subset 5I_{z_n}\}) \leq \sum_{n=1}^{m} \mu(S(5I_{z_n})) \leq c \|\mu\|_{q/p} \sum_{n=1}^{m} |5I_{z_n}|^{q/p} \\ &\leq c \|\mu\|_{q/p} \left(\sum_{n=1}^{m} |I_{z_n}| \right)^{q/p} \leq c \|\mu\|_{q/p} \|\varphi\|_{L^1(\mathbb{T})}^{q/p} \lambda^{-q/p}, \end{aligned}$$

donde en la última desigualdad hemos usado (3). Finalmente, haciendo tender ε a 0, obtenemos el resultado. ∎

4. El espacio de Hardy H^p

En esta sección probaremos la caracterización de las q-medidas de Carleson de H^p con $p \le q$. El caso $p > q$ no se estudia en este artículo, pero su análisis se puede encontrar en distintos libros de las referencias.

Teorema 13. *Sean $0 < p \le q < \infty$ y μ una medida de Borel en $\mathbb{D}$. Entonces, las siguientes afirmaciones son equivalentes:*

(I) *Para cada $\alpha > \frac{1}{p}$, el operador*

$$\left[\tilde{\mathcal{M}}((\cdot)^{1/\alpha})\right]^\alpha : L^p(\mathbb{T}) \to L^q(\mu)$$

está acotado;

(II) *μ es una q-medida de Carleson para H^p;*

(III) $\|\mu\|_{q/p} = \sup_{I\subset\mathbb{T}} \frac{\mu(S(I))}{|I|^{q/p}} < \infty$.

Además,

$$\|\mathrm{Id}\|^q_{(H^p,L^q(\mu))} \asymp \left\|\left[\tilde{\mathcal{M}}((\cdot)^{1/\alpha})\right]^\alpha\right\|^q_{(L^p(\mathbb{T}),L^q(\mu))} \asymp \|\mu\|_{q/p}.$$

Demostración. Veamos que (III) $\Rightarrow$ (I). Para cada $\lambda > 0$ y $\varphi \in L^p(\mathbb{T})$ definimos la función

$$\varphi_{1/\alpha,\lambda}(e^{i\theta}) = \begin{cases} |\varphi(e^{i\theta})|^{1/\alpha} & \text{si } |\varphi(e^{i\theta})|^{1/\alpha} > \lambda/2, \\ 0 & \text{si } |\varphi(e^{i\theta})|^{1/\alpha} \le \lambda/2, \end{cases}$$

y se satisface la inclusión

$$\left\{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi^{1/\alpha})(z) > \lambda\right\} \subset \left\{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi_{1/\alpha,\lambda})(z) > \frac{\lambda}{2}\right\}$$

ya que $\varphi^{1/\alpha} = \varphi^{1/\alpha}\chi_{\{|\varphi(e^{i\theta})|^{1/\alpha}>\lambda/2\}} + \varphi^{1/\alpha}\chi_{\{|\varphi(e^{i\theta})|^{1/\alpha}\le\lambda/2\}}$. Ahora bien, por el teorema de Fubini y la anterior inclusión de conjuntos,

$$\begin{aligned} \int_{\mathbb{D}} (\tilde{\mathcal{M}}(\varphi^{1/\alpha})(z))^{\alpha q}\,d\mu(z) &= \int_{\mathbb{D}} \left(\alpha q \int_0^{(\tilde{\mathcal{M}}(\varphi^{1/\alpha})(z))} \lambda^{\alpha q-1}\,d\lambda\right) d\mu(z) \\ &= \alpha q \int_0^\infty \lambda^{\alpha q-1}\mu\left(\left\{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi^{1/\alpha})(z) > \lambda\right\}\right) d\lambda \\ &\le \alpha q \int_0^\infty \lambda^{\alpha q-1}\mu\left(\left\{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi_{1/\alpha,\lambda})(z) > \frac{\lambda}{2}\right\}\right) d\lambda. \end{aligned} \tag{4}$$

Además, de la desigualdad débil disponible en el lema 12 se sigue que

$$\begin{aligned} &\int_0^\infty \lambda^{\alpha q-1}\mu\left(\left\{z \in \mathbb{D} : \tilde{\mathcal{M}}(\varphi_{1/\alpha,\lambda})(z) > \frac{\lambda}{2}\right\}\right) d\lambda \\ &\quad\le C\|\mu\|_{q/p} \int_0^\infty \lambda^{q(\alpha-1/p)-1}\|\varphi_{1/\alpha,\lambda}\|^{q/p}_{L^1(\mathbb{T})}\,d\lambda \\ &\quad= C\|\mu\|_{q/p} \int_0^\infty \lambda^{q(\alpha-1/p)-1}\left(\int_0^{2\pi} |\varphi(e^{i\theta})|^{1/\alpha}\chi_{\{|\varphi(e^{i\theta})|^{1/\alpha}>\lambda/2\}}\,d\theta\right)^{q/p} d\lambda. \end{aligned} \tag{5}$$

Finalmente, empleando la desigualdad de Minkowski (cuando $q > p$) y el teorema de Fubini (cuando $p = q$) se tiene que

$$\begin{aligned}
&\int_0^\infty \lambda^{q(\alpha-1/p)-1}\left(\int_0^{2\pi} |\varphi(e^{i\theta})|^{1/\alpha}\chi_{\{|\varphi(e^{i\theta})|^{1/\alpha}>\lambda/2\}}\,d\theta\right)^{q/p} d\lambda \\
&\qquad \le \left(\int_0^{2\pi} |\varphi(e^{i\theta})|^{1/\alpha}\left(\int_0^\infty \lambda^{q(\alpha-1/p)-1}\chi_{\{|\varphi(e^{i\theta})|^{1/\alpha}>\lambda/2\}}\,d\lambda\right)^{p/q} d\theta\right)^{q/p} \\
&\qquad = \left(\int_0^{2\pi} |\varphi(e^{i\theta})|^{1/\alpha}\left(\int_0^{2|\varphi(e^{i\theta})|^{1/\alpha}} \lambda^{q(\alpha-1/p)-1}\,d\lambda\right)^{p/q} d\theta\right)^{q/p} \\
&\qquad \asymp \left(\int_0^{2\pi} |\varphi(e^{i\theta})|^{p}\,d\theta\right)^{q/p} = \|\varphi\|^q_{L^p(\mathbb{T})}.
\end{aligned} \tag{6}$$

Así pues, de (4), (5) y (6) se sigue que

$$\int_{\mathbb{D}} (\tilde{\mathcal{M}}(\varphi^{1/\alpha})(z))^{\alpha q}\,d\mu(z) \lesssim \|\mu\|_{q/p}\|\varphi\|^q_{L^p(\mathbb{T})},$$

lo que prueba que el operador

$$\left[\tilde{\mathcal{M}}((\cdot)^{1/\alpha})\right]^\alpha : L^p(\mathbb{T}) \to L^q(\mu)$$

está acotado y, además,

$$\left\|\left[\tilde{\mathcal{M}}((\cdot)^{1/\alpha})\right]^\alpha\right\|^q_{(L^p(\mathbb{T}),L^q(\mu))} \lesssim \|\mu\|_{q/p}.$$

Ahora, veamos que (i) ⇒ (ii). Sea $f \in H^p$. Como $\alpha > 1/p$, tenemos que $H^p \subset H^{1/\alpha}$. Además, tenemos por el lema 11 que

$$|f(z)| \lesssim [\tilde{\mathcal{M}}(|f|^{1/\alpha})(z)]^\alpha.$$

De esta forma,

$$\|\mathrm{Id}(f)\|_{L^q(\mu)} = \|f\|_{L^q(\mu)} \lesssim \left\|[\tilde{\mathcal{M}}(|f|^{1/\alpha})]^\alpha\right\|_{L^q(\mu)} \le \left\|[\tilde{\mathcal{M}}((\cdot)^{1/\alpha})]^\alpha\right\|_{(L^p(\mathbb{T}),L^q(\mu))}\|f\|_{L^p(\mathbb{T})}.$$

Así pues, como $\|f\|_{L^p(\mathbb{T})} \lesssim \|f\|_{H^p}$, el operador identidad es acotado de H^p a $L^q(\mu)$. Además,

$$\|\mathrm{Id}\|_{H^p\to L^q(\mu)} \lesssim \left\|[\tilde{\mathcal{M}}((\cdot)^{1/\alpha})]^\alpha\right\|_{(L^p(\mathbb{T}),L^q(\mu))}.$$

Por último, el caso (ii) ⇒ (iii) corresponde a la proposición 7. ∎

5. El espacio de Bergman A^p_α

En esta sección vamos a ver que podemos caracterizar las q-medidas de Carleson para los espacios de Bergman A^p_α de forma geométrica vía las cajas de Carleson sobre intervalos o vía discos pseudohiperbólicos. Para ello necesitaremos un lema de recubrimiento y otro de armonicidad.

Definición 14 (Distancia pseudohiperbólica). Definimos la distancia pseudohiperbólica entre los puntos $z, \omega \in \mathbb{D}$ como

$$\rho(z,\omega) = \left|\frac{z-\omega}{1-\overline{z}\omega}\right|.$$

En particular, definimos el disco pseudohiperbólico de centro z y radio r como

$$\Delta(z,r) = \{\omega \in \mathbb{D} : \rho(z,\omega) < r\}.$$ ◂

Una propiedad importante de esta distancia es la desigualdad triangular fuerte.

Proposición 15 (Desigualdad triangular fuerte). *Sean $z, \omega \in \mathbb{D}$. Entonces, tenemos que*

$$\rho(z,\omega) \leq \frac{\rho(z,\alpha)+\rho(\alpha,\omega)}{1+\rho(z,\alpha)\rho(\alpha,\omega)}$$

para todo $\alpha \in \mathbb{D}$.

Demostración. Sea $\alpha \in \mathbb{D}$ y sea

$$\varphi_\alpha(z) = \frac{\alpha - z}{1-\overline{\alpha}z}, \quad z \in \mathbb{D},$$

el automorfismo del disco que verifica que

$$|\varphi_\alpha(z)| = \rho(\alpha, z) \quad \text{y} \quad \varphi_\alpha(z) = \varphi_\alpha^{-1}(z).$$

Observemos que, entonces, para cualesquiera $z, \omega \in \mathbb{D}$,

$$\varphi_\alpha(\alpha) = 0 \quad \text{y} \quad \rho(z,\omega) = \rho(\varphi_\alpha(z), \varphi_\alpha(\omega)). \tag{7}$$

Así pues, nos es suficiente ver que la desigualdad triangular anterior se satisface para $z, w \in \mathbb{D}$ y $\alpha = 0$, ya que entonces, si $\alpha \neq 0$,

$$\rho(z,\omega) = \rho(\varphi_\alpha(z), \varphi_\alpha(\omega)) \leq \frac{\rho(\varphi_\alpha(z),0)+\rho(0,\varphi_\alpha(\omega))}{1+\rho(\varphi_\alpha(z),0)\rho(0,\varphi_\alpha(\omega))} = \frac{\rho(z,\alpha)+\rho(\alpha,\omega)}{1+\rho(z,\alpha)\rho(\alpha,\omega)}.$$

Así pues, consideremos $|z| = a$, $|w| = b$ y sea $\theta = \arg(\overline{zw})$. Entonces, la desiguldad triangular fuerte para estos valores es equivalente a

$$\frac{a^2+b^2-2ab\cos(\theta)}{1+a^2b^2-2ab\cos(\theta)} \leq \frac{(a+b)^2}{(1+ab)^2}.$$

Reescribiendo la parte izquierda de la desigualdad anterior, tenemos que

$$\frac{a^2+b^2-2ab\cos(\theta)}{1+a^2b^2-2ab\cos(\theta)} = 1 - \frac{(1-a^2)(1-b^2)}{1+a^2b^2-2ab\cos(\theta)} \leq 1 - \frac{(1-a^2)(1-b^2)}{1+a^2b^2+2ab} = \frac{(a+b)^2}{(1+ab)^2},$$

de lo que se sigue la desigualdad esperada. ∎

Observemos que de la desigualdad triangular fuerte se sigue que, para todo $z, \omega, \alpha \in \mathbb{D}$,

$$\rho(z,\omega) \leq \rho(z,\alpha) + \rho(\alpha,\omega),$$

por lo que tenemos que ρ es una distancia. El siguiente lema será necesario para la prueba de la caracterización y se trata de un lema de recubrimiento del disco unidad por discos pseudohiperbólicos.

Lema 16 (Lema de recubrimiento 2). *Dado $r \in (0,1)$, existe una secuencia de puntos $(\alpha_n)_{n\in\mathbb{N}} \subset \mathbb{D}$ tales que $\mathbb{D} = \bigcup_{n=1}^\infty \Delta(\alpha_n, r)$. Además, para todo $s \in (0,1)$ tenemos que cada punto $z \in \mathbb{D}$ pertenece a lo sumo a $N = N(s,r)$ discos $\Delta(\alpha_n, s)$.*

Demostración. Sea $(B_j)_{j\in\mathbb{N}}$ una sucesión de discos pseudohiperbólicos de radio $r/3$ tales que $\mathbb{D} = \bigcup_{j=1}^\infty B_j$. Cogemos $D_1 = B_1$, $D_2 = B_{j_2}$, donde j_2 es el primer número natural tal que $D_1 \cap B_{j_2} = \emptyset$, y vamos eligiendo los discos $(D_n)_{n\in\mathbb{N}}$ de forma recursiva. Sea $(\alpha_n)_{n\in\mathbb{N}}$ el centro pseudohiperbólico de $(D_n)_{n\in\mathbb{N}}$. Queremos ver que $\mathbb{D} = \bigcup_{n=1}^\infty \Delta(\alpha_n, r)$. Para probar esa igualdad, vamos a considerar que existe un punto $\alpha \in \mathbb{D} \setminus \bigcup_{n=1}^\infty \Delta(\alpha_n, r)$ y llegaremos a una contradicción. Si existe este punto, entonces tenemos que

(a) $\Delta(\alpha, 2r/3) \cap \Delta(\alpha_n, r/3) = \emptyset$ para todo $n \in \mathbb{N}$, y

(b) existe $j_0 \in \mathbb{N}$ tal que $\alpha \in B_{j_0}$; por lo tanto, $B_{j_0} \subset \Delta(\alpha, 2r/3)$.

En particular, tenemos que $B_{j_0} \notin (D_n)_{n\in\mathbb{N}}$, pero de (a) y (b) se sigue que $B_{j_0} \cap D_n = \varnothing$ para todo n, lo cual nos da una contradicción y, por lo tanto, $\mathbb{D} = \bigcup_{n=1}^{\infty} \Delta(\alpha_n, r)$.

Ahora vamos a ver la segunda parte. Sean $s \in (0,1)$ y $z \in \mathbb{D}$. Definimos el conjunto

$$E(z) = \{\alpha_k \in \mathbb{D} : \varphi_z(\alpha_k) \in \Delta(0,s) = D(0,s)\},$$

donde

$$\varphi_z(\omega) = \frac{z-\omega}{1-\overline{z}\omega} \tag{8}$$

es el automorfismo del disco que verifica que

$$|\varphi_z(\alpha_k)| < s \Longleftrightarrow \rho(\alpha_k, z) < s.$$

Sean $\alpha_k, \alpha_j \in E(z)$ y denotemos $w_k = \varphi_z(\alpha_k)$ y $w_j = \varphi_z(\alpha_j)$. Observemos que, entonces,

$$\rho(\omega_k, \omega_j) = \rho(\alpha_k, \alpha_j) \geq \frac{2r}{3}$$

y, como $|\omega_k| \leq s$, tenemos que

$$\rho(\omega_k, \omega_j) = \frac{|\omega_k - \omega_j|}{|1-\overline{\omega_k}\omega_j|} \leq \frac{|\omega_k - \omega_j|}{1-s^2},$$

de donde concluimos que $|\omega_k - \omega_j| \geq 2r(1-s^2)/3$ y que los discos

$$D_k = D\left(\omega_k, \frac{r(1-s^2)}{3}\right) = D\left(\varphi_z(\alpha_k), \frac{r(1-s^2)}{3}\right)$$

son disjuntos. Así pues, para cada $\alpha_k \in E(z)$ existe un disco D_k tal que si cogemos otro centro pseudohiperbólico $\alpha_j \in E(z)$, tenemos que $D_k \cap D_j = \varnothing$. Ahora bien, sea $\zeta \in D_k$. Tenemos que

$$|\zeta| \leq |\omega_k| + |\zeta - \omega_k| < s + \frac{r(1-s^2)}{3} \leq 1.$$

Usando esta cota obtenemos que

$$\#(E(z))\left[\pi\frac{r^2(1-s^2)^2}{9}\right] = \sum_{\alpha_k \in E(z)} |D_k| \leq |D(0,1)| \leq \pi,$$

por lo que

$$\#(E(z)) \leq \frac{9}{r^2(1-s^2)^2},$$

donde vemos que, efectivamente, la cota de $\#(E(z))$ depende únicamente de r y s y, por consiguiente, $\#(E(z)) \leq N(s,r)$. ∎

Para la caracterización necesitaremos también el siguiente lema. Recordemos que, si f es una función holomorfa en $\mathbb{D}$, entonces, para $0 < p < \infty$, tenemos que $|f|^p$ es subarmónica en $\mathbb{D}$, luego para todo $0 < s < 1$ tenemos que

$$|f(0)|^p \leq \frac{1}{s^2}\int_{D(0,s)} |f(\zeta)|^p \, \mathrm{d}A(\zeta).$$

Esta desigualdad puede reformularse en términos de los discos hiperbólicos como sigue.

Lema 17. *Sean $f \in \mathcal{O}(\mathbb{D})$, $0 < p < \infty$, $\alpha > -1$, $0 < s < 1$. Entonces, existe una constante $c(s,\alpha)$ de manera que*

$$|f(a)|^p \leq \frac{c(s,\alpha)}{(1-|a|^2)^{(2+\alpha)}}\int_{\Delta(a,s)} |f(\zeta)|^p (1-|\zeta|^2)^\alpha \, \mathrm{d}A(\zeta), \quad a \in \mathbb{D}.$$

Demostración. Sea $a \in \mathbb{D}$. Definimos $\varphi_a \in \mathrm{Aut}(\mathbb{D})$ como en (8). Por la subarmonicidad de $|f|^p$, tenemos que

$$|f(a)|^p = |f \circ \varphi_a(0)|^p \leq \frac{1}{s^2} \int_{\Delta(0,s)} |f \circ \varphi_a(\zeta)|^p \, \mathrm{d}A(\zeta) = \frac{c(s)}{s^2} \int_{\Delta(a,s)} |f(\zeta)|^p |\varphi_a'(\zeta)|^2 \, \mathrm{d}A(\zeta)$$
$$= \frac{c(s)}{s^2} \int_{\Delta(a,s)} |f(\zeta)|^p \frac{(1-|a|^2)^2}{|1-\overline{a}\zeta|^4} \, \mathrm{d}A(\zeta),$$

donde hemos hecho el cambio de variable $\zeta = \varphi_a(\zeta) = \varphi_a^{-1}(\zeta)$. Como $\xi \in \Delta(a,s)$, se tiene que $1-|a|^2 \asymp |1-\overline{a}\zeta|$, por lo que se sigue la desigualdad

$$|f(a)|^p \leq \frac{c(s)}{(1-|a|^2)^2} \int_{\Delta(a,s)} |f(\zeta)|^p \, \mathrm{d}A(\zeta).$$

Finalmente, nótese que también tenemos que $1-|a|^2 \asymp 1-|\zeta|^2$, lo que implica que

$$|f(a)|^p \leq \frac{c(s,\alpha)}{(1-|a|^2)^2} \int_{\Delta(a,s)} |f(\zeta)|^p \frac{(1-|\zeta|^2)^\alpha}{(1-|a|^2)^\alpha} \, \mathrm{d}A(\zeta). \qquad \blacksquare$$

Pasemos a demostrar la caracterización de las q-medidas de Carleson en los espacios A^p_α. El resto de casos se pueden encontrar en los libros que aparecen en las referencias.

Teorema 18. *Sean $0 < p \leq q < \infty$, $\alpha > -1$ y μ una medida de Borel positiva en $\mathbb{D}$. Entonces, los siguientes enunciados son equivalentes:*

(I) *μ es una q-medida de Carleson para A^p_α.*

(II) *Para todo $r \in (0,1)$ tenemos que*

$$\sup_{a \in \mathbb{D}} \frac{\mu(\Delta(a,r))}{(1-|a|^2)^{(2+\alpha)q/p}} < \infty.$$

(III) *Existe $r \in (0,1)$ tal que*

$$\sup_{a \in \mathbb{D}} \frac{\mu(\Delta(a,r))}{(1-|a|^2)^{(2+\alpha)q/p}} < \infty.$$

(IV) *Si $S(a)$ es la caja de Carleson asociada a $a \in \mathbb{D}$, entonces*

$$\sup_{a \in \mathbb{D}} \frac{\mu(S(a))}{(1-|a|^2)^{(2+\alpha)q/p}} < \infty.$$

Observación 19. La condición (IV) se puede reformular como sigue:

$$\sup_{I \subset \mathbb{T}} \frac{\mu(S(I))}{|I|^{(2+\alpha)q/p}} < \infty. \qquad \blacktriangleleft$$

Demostración. Empezamos por demostrar que (I) $\Rightarrow$ (II). Nótese que el esquema es análogo al realizado en la prueba de la implicación (II) $\Rightarrow$ (III) del teorema 13, pero ahora tomando el núcleo reproductor normalizado de A^2_α

$$k_a(z) = \frac{(1-|a|^2)^{(2+\alpha)/2}}{(1-\overline{a}z)^{2+\alpha}}, \quad a, z \in \mathbb{D},$$

el cual tiene norma 1 en A^α_2 y no tiene ceros. Podemos ver que la función $f_a(z) = (k_a(z))^{2/p}$ es holomorfa en el disco unidad y que también tiene norma A^p_α igual a 1, así que podemos aplicar nuestra hipótesis de que μ es una q-medida de Carleson y que, dado $r \in (0,1)$, entonces $\Delta(a,r) \subset \mathbb{D}$, para así obtener que

$$1 = \|f_a\|_{A^p_\alpha} \gtrsim \|f_a\|_{L^q} = \left(\int_{\mathbb{D}} |f_a(z)|^q \, \mathrm{d}\mu(z) \right)^{1/q} = \left(\int_{\mathbb{D}} |k_a(z)|^{2q/p} \, \mathrm{d}\mu(z) \right)^{1/q}$$
$$\gtrsim \left(\int_{\Delta(a,r)} |k_a(z)|^{2q/p} \, \mathrm{d}\mu(z) \right)^{1/q} \gtrsim \frac{1}{(1-|a|^2)^{(2+\alpha)/p}} \left(\int_{\Delta(a,r)} \mathrm{d}\mu(z) \right)^{1/q} \asymp \frac{\mu(\Delta(a,r))^{1/q}}{(1-|a|^2)^{(2+\alpha)/p}}.$$

Esto nos da la cota

$$\frac{\mu(\Delta(a,r))}{(1-|a|^2)^{(2+\alpha)q/p}} < C,$$

donde C no depende de a. Tomando el supremo sobre $a \in \mathbb{D}$, obtenemos para $r \in (0,1)$ que

$$\sup_{a\in\mathbb{D}} \frac{\mu(\Delta(a,r))}{(1-|a|^2)^{(2+\alpha)q/p}} < \infty.$$

Para ver que (I) ⇒ (IV) podemos usar exactamente el mismo proceso que para ver que (I) ⇒ (II) cambiando $\Delta(a,r)$ por la caja de Carleson $S(a)$, y llegamos a la desigualdad

$$\frac{\mu(S(a))}{(1-|a|^2)^{(2+\alpha)q/p}} < C;$$

como la constante no depende de a, podemos poner supremos sobre $a \in \mathbb{D}$, obteniendo así el resultado deseado.

La condición (II) ⇒ (III) es trivial. Veamos que (III) ⇒ (I), es decir, queremos ver que la condición

$$\sup_{a\in\mathbb{D}} \frac{\mu\left(\Delta(a,r)\right)}{(1-|a|^2)^{(2+\alpha)q/p}} < \infty$$

para algún $0 < r < 1$ implica que μ es una q-medida de Carleson para A^p_α. Sea $(\alpha_n)_{n\in\mathbb{N}}$ la sucesión generada en el lema 16 para r. Entonces,

$$\int_{\mathbb{D}} |f(z)|^q \,\mathrm{d}\mu(z) = \int_{\bigcup_{n=1}^\infty \Delta(\alpha_n,r)} |f(z)|^q \,\mathrm{d}\mu(z) \le \sum_{n=1}^\infty \int_{\Delta(\alpha_n,r)} |f(z)|^q \,\mathrm{d}\mu(z) \le \sum_{n=1}^\infty |f(\tilde{z}_n)|^q \mu\left(\Delta(\alpha_n,r)\right),$$

donde $\tilde{z}_k \in \overline{\Delta(\alpha_k,r)}$ y

$$|f(\tilde{z}_n)| = \max_{z\in\overline{\Delta(\alpha_n,r)}} |f(z)|.$$

Nótese que este máximo está bien definido ya que ningún disco cerrado toca la frontera del disco unidad. Esto se debe a que, si algún punto se encuentra en la frontera, entonces $\rho(\alpha_n,z)=1$, lo cual es imposible porque suponemos que $r<1$. Tomemos $s \in (0,1)$ tal que $s < (1-r)/2$ y, usando el lema 17, obtenemos que, para $\alpha > -1$,

$$\begin{aligned}
\sum_{n=1}^\infty \left(|f(\tilde{z}_n)|^p \mu\left(\Delta(\alpha_n,r)\right)^{p/q}\right)^{q/p} &\lesssim \sum_{n=1}^\infty \left(\frac{\mu\left(\Delta(\alpha_n,r)\right)^{p/q}}{(1-|\tilde{z}_n|^2)^{\alpha+2}} \int_{\Delta(\tilde{z}_n,s)} |f(\zeta)|^p (1-|\zeta|^2)^\alpha \,\mathrm{d}A(\zeta)\right)^{q/p} \\
&\lesssim \sum_{n=1}^\infty \left(\frac{(1-|\alpha_n|^2)^{\alpha+2}}{(1-|\tilde{z}_n|^2)^{\alpha+2}} \int_{\Delta\left(\tilde{z}_n,\frac{1+r}{2}\right)} |f(\zeta)|^p (1-|\zeta|^2)^\alpha \,\mathrm{d}A(\zeta)\right)^{q/p} \\
&\lesssim \left(\sum_{n=1}^\infty \int_{\Delta\left(\alpha_n,\frac{1+r}{2}\right)} |f(\zeta)|^p (1-|\zeta|^2)^\alpha \,\mathrm{d}A(\zeta)\right)^{q/p} \\
&\simeq \left(\int_{\mathbb{D}} \left(\sum_{n=1}^\infty \chi_{\Delta\left(\alpha_n,\frac{1+r}{2}\right)}(\zeta)\right) |f(\zeta)|^p (1-|\zeta|^2)^\alpha \,\mathrm{d}A(\zeta)\right)^{q/p} \lesssim N(r,s)^{q/p} \|f\|^q_{A^p_\alpha}.
\end{aligned}$$

Para acabar, veamos que (IV) ⇒ (III). Sea $a \in \mathbb{D}$ tal que $|a| > 1/3$. Escogemos $r < (1-|a|)/4$. Entonces, tenemos que

$$\Delta(a,r) \subset D\left(a, \frac{1}{2}(1-|a|)\right) \subset \mathrm{S}(a^*), \quad a^* = \frac{3|a|-1}{2} \mathrm{e}^{\mathrm{i}\arg(a)}.$$

Finalmente, por hipótesis,

$$\mu\left(\Delta(a,r)\right) \le \mu(\mathrm{S}(a^*)) \le C(1-|a^*|^2)^{(2+\alpha)q/p} \le C\left(\frac{9}{2}\right)^{(2+\alpha)q/p} (1-|a|^2)^{(2+\alpha)q/p}. \qquad \blacksquare$$

Referencias

[1] Bernard, Calista. «Interpolation Theorems and Applications». En: *Chicago, IL: University of Chicago Mathematics REU* (2013).

[2] Bruna, Joaquim y Cufí, Julià. *Complex analysis*. EMS Textbooks in Mathematics. Translated from the Catalan by Ignacio Monreal. European Mathematical Society (EMS), Zürich, 2013. https://doi.org/10.4171/111.

[3] Carleson, Lennart. «An interpolation problem for bounded analytic functions». En: *American Journal of Mathematics* 80 (1958), págs. 921-930. issn: 0002-9327. https://doi.org/10.2307/2372840.

[4] Carleson, Lennart. «Interpolations by bounded analytic functions and the corona problem». En: *Annals of Mathematics. Second Series* 76 (1962), págs. 547-559. issn: 0003-486X. https://doi.org/10.2307/1970375.

[5] Conway, John B. *Functions of one complex variable*. Second. Graduate Texts in Mathematics 11. Springer-Verlag, New York-Berlin, 1978. isbn: 978-0-387-90328-6.

[6] Duren, Peter. *Theory of H^p spaces*. Pure and Applied Mathematics 38. Academic Press, New York-London, 1970.

[7] Duren, Peter y Schuster, Alexander. *Bergman spaces*. Mathematical Surveys and Monographs 100. American Mathematical Society, Providence, RI, 2004. https://doi.org/10.1090/surv/100.

[8] Fefferman, Charles y Stein, Elias M. «H^p spaces of several variables». En: *Acta Mathematica* 129.3-4 (1972), págs. 137-193. issn: 0001-5962. https://doi.org/10.1007/BF02392215.

[9] Hedenmalm, Håkan; Korenblum, Boris, y Zhu, Kehe. *Theory of Bergman spaces*. Graduate Texts in Mathematics 199. Springer-Verlag, New York, 2000. https://doi.org/10.1007/978-1-4612-0497-8.

[10] Hörmander, Lars. «L^p estimates for (pluri-) subharmonic functions». En: *Mathematica Scandinavica* 20 (1967), págs. 65-78. issn: 0025-5521. https://doi.org/10.7146/math.scand.a-10821.

[11] Jones, Peter W. «Carleson measures and the Fefferman-Stein decomposition of BMO(R)». En: *Annals of Mathematics. Second Series* 111.1 (1980), págs. 197-208. issn: 0003-486X. https://doi.org/10.2307/1971197.

[12] Jones, Peter W. «L^∞ estimates for the $\bar{\partial}$ problem in a half-plane». En: *Acta Mathematica* 150.1-2 (1983), págs. 137-152. issn: 0001-5962. https://doi.org/10.1007/BF02392970.

[13] Koosis, Paul. *Introduction to H_p spaces*. Second. Cambridge Tracts in Mathematics 115. With two appendices by V. P. Havin [Viktor Petrovich Khavin]. Cambridge University Press, Cambridge, 1998. isbn: 978-0-521-45521-3.

[14] Pavlović, Miroslav. *Function classes on the unit disc*. De Gruyter Studies in Mathematics 52. An introduction. De Gruyter, Berlin, 2014. isbn: 978-3-11-028123-1; 978-3-11-028190-3.

[15] Peláez, José Ángel y Rättyä, Jouni. «Weighted Bergman spaces induced by rapidly increasing weights». En: *Memoirs of the American Mathematical Society* 227.1066 (2014), págs. vi+124. issn: 0065-9266.

[16] Rudin, Walter. *Function theory in the unit ball of* $\mathbb{C}^n$. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science] 241. Springer-Verlag, New York-Berlin, 1980. isbn: 978-0-387-90514-3.

[17] Rudin, Walter. *Real and complex analysis*. Third. McGraw-Hill Book Co., New York, 1987. isbn: 978-0-07-054234-1.

[18] Scheidemann, Volker. *Introduction to complex analysis in several variables*. Birkhäuser Verlag, Basel, 2005. https://doi.org/10.1007/3-7643-7491-8.

TEMat

Análisis de señales complejas: correlaciones de largo alcance y propiedades multifractales

Alberto Martín Aguilar
Universidad de Málaga
alberto_martin93@hotmail.com

Resumen: Muchos sistemas dinámicos no lineales, naturales y artificiales, exhiben como salida observable series temporales complejas que reflejan las propiedades de la dinámica subyacente, que normalmente no se conoce explícitamente. Por lo tanto, el análisis adecuado de estas señales es fundamental para la caracterización de los sistemas que las originan.

En general, las series suelen ser no estacionarias, exhiben correlaciones de largo alcance, y también presentan propiedades fractales o multifractales. En este trabajo, presentamos herramientas recientes para el estudio de este tipo de señales y las aplicamos para caracterizar señales complejas naturales (señales de variaciones de temperatura en el océano) y artificiales (índices bursátiles de distintas compañías españolas e internacionales).

Abstract: Many non-linear dynamical systems, both natural and artificial, exhibit complex time series as observable output. These series reflect the properties of the underlying dynamics, which are not usually explicitly known. Therefore, an adequate analysis of these signals is fundamental for the characterization of the systems that originate them.

In general, the series are usually non-stationary, present long-range correlations, and also satisfy fractal or multifractal properties. In this work, we present recent tools for the study of this type of signals and we apply them to characterize both natural complex signals (time series of pressure variations in the ocean) and artificial signals (stock indices of different Spanish and international companies).

Palabras clave: señal, fractales, DFA, multifractal.

MSC2010: 62M10, 28A80.

Recibido: 21 de febrero de 2019.
Aceptado: 28 de noviembre de 2019.

Agradecimientos: Quisiera agradecer a mi familia y amigos su apoyo durante todos estos años de estudio y su dedicación a la hora de aconsejarme y motivarme.

Igualmente, quisiera agradecer a los profesores que me han guiado durante todos estos años, y a Pedro J. Carpena Sánchez y Manuel Gómez Extremera, de la Universidad de Málaga, en particular, por su ayuda con este trabajo.

Referencia: Martín Aguilar, Alberto. «Análisis de señales complejas: correlaciones de largo alcance y propiedades multifractales». En: *TEMat*, 4 (2020), págs. 83-99. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p83.

1. Introducción

En este trabajo se pretende estudiar la dinámica subyacente de algunas de las series temporales presentes en los mercados financieros, así como otras series temporales en la naturaleza. Esto es debido a la habitual presencia en la naturaleza de series temporales que pueden parecer aleatorias, pero que, sin embargo, esconden una dinámica compleja (por ejemplo, los sistemas fisiológicos, como el cardíaco, o el cerebral; el clima; los mercados financieros; etc.), que se refleja en la presencia de correlaciones de largo alcance o propiedades fractales [1, 19-22].

Muchos sistemas dinámicos producen como salida series no estacionarias y, por lo tanto, son difíciles de analizar con métodos tradicionales. Como conocemos, la función de autocorrelación es únicamente aplicable a series estacionarias, dando como resultado un correlograma, el cual refleja las correlaciones de largo alcance en la serie, en el caso de haberlas. Sin embargo, la función de autocorrelación no puede ser aplicada a series no estacionarias, ya que el resultado nos produce un correlograma plano, que no aporta ningún tipo de información. Además, la función de autocorrelación es muy sensible al tamaño de las series estacionarias. Por lo tanto, es necesario recurrir a otros métodos que puedan ser utilizados en todo tipo de series.

En este trabajo, se presenta el método de análisis de fluctuaciones sin tendencia (DFA, del inglés), que estudia cómo se comportan las fluctuaciones de la señal (el momento de orden 2) en función de la escala de observación de esta, originariamente introducido por Peng *et al.* [21]. Si las fluctuaciones escalan en ley de potencias, el DFA proporciona como resultado un exponente, el cual determinará la existencia o no de correlaciones de largo alcance en la serie temporal. Igualmente, este método es generalizado mediante el DFA multifractal (MF-DFA), el cual utiliza otros momentos de las fluctuaciones para estudiar más en profundidad la serie a analizar. De esta forma, cuantificamos las correlaciones de una serie y obtenemos propiedades fractales en esta, en el caso de estar presentes. Se puede encontrar más información en el trabajo de Martín Aguilar [18], donde se han estudiado otros aspectos como la influencia de la longitud de la serie o el estudio de la función de autocorrelación en las series de módulo y signo

2. Fractales

El concepto de conjunto fractal, dado a conocer por B. Mandelbrot, fue creado para dar cabida en la estructura formal de la geometría a diversos objetos, denominados entonces *monstruos geométricos*, cuyo comportamiento se apartaba radicalmente del mostrado por los conjuntos que estaban en la base conceptual de esta disciplina. Pese a sus inicios en el ámbito estricto de las matemáticas, los fractales han traspasado ampliamente las fronteras de lo abstracto para servir de modelos en múltiples campos de la ciencia [4, 15, 22].

Las propiedades que presentan frente a cambios de escala fueron utilizadas, por ejemplo, por Gutemberg y Richter para establecer su ley de distribución de intensidades de terremotos, y por Richardson para establecer que los perfiles costeros poseían propiedades similares a algunas de las curvas de von Koch. La hidrología, la geografía, la geofísica, la ecología y la economía han sido campos pioneros en el desarrollo de modelos basados en los fractales. Los éxitos en estos campos y, a veces quizás, la imposibilidad de encontrar otro ámbito más adecuado han conducido, por un lado, a una considerable ampliación de los campos de aplicación y, por otro, a incrementar enormemente la base teórica en la que se apoyan.

Un primer ejemplo de un conjunto fractal (véase el libro de Feder [11]) se encuentra en el intento de medir la longitud de la costa de Noruega. Para responder a esta pregunta, tenemos que tener en cuenta que en la costa nos encontramos con salientes de ríos, islas, valles, etc. Esto escapa de la geometría usual y se tienen que emplear técnicas diferentes para establecer la longitud de esta.

Un primer intento para medir la costa de Noruega podría ser medir el número de pasos de longitud δ que se dan, $N(\delta)$, y ver la distancia total desde un extremo de la costa al otro, $L = N(\delta) \times \delta$. Este proceso podría ser repetido numerosas veces con diferente longitud δ.

Sin embargo, este proceso tiene numerosos problemas teniendo en cuenta a las islas o a los salientes y entrantes de los ríos. Por ello, otro segundo método sería dividir la costa en cuadrados, con lado de longitud δ. La longitud de la costa de Noruega es bien conocida, es por ello que el resultado no debería

depender del método empleado. Mediante este segundo proceso, parecido al anterior, si disminuimos δ, se podría esperar un incremento proporcional al número de cuadrados necesarios para cubrir toda la costa, para así tender el resultado a la longitud real. Sin embargo, este no es el caso, ya que al disminuir la longitud δ, la distancia aumenta según una línea recta, escribiendo en escala logarítmica ambos ejes.

Si se realiza una regresión lineal en este ejemplo, vemos que la recta se puede aproximar mediante

$$L(\delta) = a\delta^{1-D},$$

con $D > 1$, al cual llamaremos **dimensión fractal**. Como este, hay numerosos ejemplos estudiados, como la costa de Inglaterra por Mandelbrot, la costa de Australia, etc. Fórmulas del tipo

$$y = ax^b$$

son las denominadas **leyes de potencias**. Para una introducción más formal del concepto de fractal, sería necesario introducir también conceptos como la dimensión de Hausdorff. Sin embargo, este aspecto se aleja del contenido de este trabajo, y se puede consultar en el libro de Feder [11].

Definición 1. Se dice que un objeto geométrico es **fractal** si su forma no varía independientemente de la escala mediante la cual se observe. ◂

La anterior definición es general y se aplica a todos los ámbitos donde se utilicen los fractales. Sin embargo, en la estadística en concreto, no se comprueba de dicha forma, ya que las series temporales no pueden llegar a ser iguales a diferente escala. Por ello, una serie temporal fractal se considera autosimilar cuando el «aumento» de una parte es equivalente al total, comparando propiedades estadísticas y no geométricas.

El concepto de proceso autosimilar fue introducido por Kolmogorov en 1941. Sin embargo, la definición era tan teórica que no fue tomada en cuenta hasta 1969, cuando Mandelbrot la introdujo en la estadística. Así, decimos que un cuerpo geométrico es autosimilar si observamos la misma estructura geométrica independientemente de la distancia a la cual miramos el cuerpo.

Definición 2. Decimos que un cuerpo geométrico es **autosimilar** si se puede escribir como unión de copias redimensionadas de sí mismo, siendo el redimensionamiento uniforme en todas las direcciones del cuerpo. ◂

Sin embargo, los sistemas presentes en la naturaleza son más complejos y requieren de una teoría más extensa de fractales. Esto se debe a que posiblemente no haya un único exponente en el conjunto, y este presente diversos fractales entrelazados, con diferentes dimensiones fractales. Esto nos deriva en la necesidad de introducir el concepto de *conjunto multifractal*, a diferencia de los que tienen un único exponente, que se denominarán *conjuntos monofractales*.

2.1. Multifractales

El objetivo de introducir el concepto de multifractal es el incluir en el esquema a conjuntos aún más complejos, que presentan leyes de escalado múltiples. Pueden encontrarse aplicaciones físicas de estos objetos en campos como la física de altas energías, la meteorología, las ciencias medioambientales y otros muchos en los que actualmente se trabaja activamente con ellos.

Este concepto fue inicialmente introducido por Mandelbrot dentro del contexto del estudio de la turbulencia y fue desarrollado y extendido por él mismo a muchos otros campos. No será desarrollada toda la teoría presentada en su momento, ya que difiere del tema de estudio del modelo que presentaremos en la siguiente sección. Se puede consultar el desarrollo completo en la tesis de Faleiro [10] y el libro de Feder [11].

La idea principal del concepto multifractal es la posibilidad de encontrar diferentes escalados en un mismo conjunto. Es decir, supongamos que tenemos un conjunto con dimensión fractal D, y lo podemos dividir en diferentes subconjuntos fractales

$$S = \bigcup_{\alpha} S_\alpha,$$

donde cada uno de los S_α tendrá dimensión fractal $f(\alpha) \leq D$. Con esto, en el capítulo 6 del libro de Feder [11] se presenta que la medida μ_α en una celda de tamaño λ sigue una ley de potencias anteriormente descrita, es decir, $\mu_\alpha = \delta^\alpha$, y, por tanto, la medida M del conjunto S satisface la siguiente igualdad:

$$M_d(q,\delta) = \int \rho(\alpha)\,\mathrm{d}\alpha \delta^{-f(\alpha)} \delta^{\alpha q} \delta^d = \int \rho(\alpha)\,\mathrm{d}\alpha \delta^{q\alpha - f(\alpha)+d},$$

donde $\rho(\alpha)\,\mathrm{d}\alpha$ es el número de conjuntos entre S_α y $S_{\alpha+\mathrm{d}\alpha}$. De aquí se obtiene que, asintóticamente, la integral es finita si se cumple que d es igual a $\tau(q)$, donde

$$\tau(q) = f(\alpha(q)) - q\alpha(q),$$

siendo $\alpha(q)$ la solución de

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\{q\alpha - f(\alpha)\}\Big|_{\alpha=\alpha(q)} = 0.$$

Finalmente, haciendo uso de la transformada de Legendre, se define como espectro multifractal a la representación de las siguientes unidades en los correspondientes ejes:

$$\beta(q) = \frac{\mathrm{d}}{\mathrm{d}q}\tau(q),$$
$$f(\alpha(q)) = q\beta(q) - \tau(q),$$

donde $f(\beta(q))$ representa la dimensión fractal para cada uno de los momentos q. Esta gráfica nos permitirá identificar la presencia de multifractalidad. Así, si el análisis resulta en un espectro multifractal de una gran amplitud, esto nos revela que existe multifractalidad en la serie, al haber más exponentes de escalado distinto. En el caso de obtener un espectro de poca amplitud, la serie sería considerada monofractal.

Concretamente, estas fórmulas anteriores serán utilizadas en el modelo considerado en este trabajo para calcular el espectro multifractal de las series temporales analizadas. De esta forma, se estudiará la dinámica de estas y la complejidad presente. La interpretación de los parámetros anteriores se puede consultar en el libro de Feder [11, capítulo 6].

3. Series temporales

Para describir el mecanismo que genera una determinada serie, tenemos que suponer que existe una variable aleatoria subyacente para cada instante de tiempo. Es decir, la existencia de un proceso estocástico.

Definición 3. Un **proceso estocástico** es una colección de variables aleatorias $\{X_t : t \in T\}$, ordenadas según el subíndice t que, en general, se suele identificar con el tiempo. ◂

Con esto, ya se puede establecer el concepto de serie temporal.

Definición 4. Una **serie temporal** es una realización de un proceso estocástico $\{X_t : t \in T\}$. ◂

Para cada una de las variables aleatorias del proceso estocástico, podemos definir una función de probabilidad con una función de densidad asociada. Por tanto, cada variable aleatoria tiene una media, una varianza y covarianza para estudiar la dependencia (o independencia) de las observaciones entre sí. Sin embargo, se suele utilizar en mayor medida la denominada función de correlación, definida como

$$\rho_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\,\mathrm{Var}(X_j)}}.$$

Una de las propiedades deseadas es la regularidad en el tiempo.

Definición 5. Se dice que un proceso estocástico $\{X_t : t \in T\}$ es **fuertemente estacionario** si la función de distribución multivariante de $\{X_i, X_{i+1}, X_{i+2}, \ldots, X_{i+k-1}\}$ es idéntica a la función de distribución de $\{X_j, X_{j+1}, X_{j+2}, \ldots, X_{j+k-1}\}$, para todo i, j y para todo $k > 0$. ◂

Sin embargo, este concepto es muy restrictivo si se trabaja con series reales, ya que esto es difícil de obtener. En lugar de pedir la estacionariedad fuerte, se establece el concepto de estacionariedad débil.

Definición 6. Se dice que un proceso estocástico $\{X_t : t \in T\}$ es **débilmente estacionario** si satisface las siguientes condiciones:

- $\mathbb{E}(X_t) = \mu$, para todo $t \in T$;
- $\mathrm{Var}(X_t) = \sigma^2$, para todo $t \in T$;
- $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$, para todo $t \in T$. ◂

A partir de ahora, se dirá que el proceso es estacionario refiriéndose a un proceso con estacionariedad débil.

Un caso particular de la función de correlación es la función de autocorrelación, la cual estudia la dependencia o independencia de una señal consigo misma tras haberla desplazado k posiciones. Nótese que esta únicamente tiene sentido para procesos estocásticos estacionarios, siendo este el principal motivo del desarrollo de este trabajo. Se define la función de autocorrelación como

$$\rho(k) = \frac{\mathrm{Cov}(X_i, X_{i+k})}{\sqrt{\mathrm{Var}(X_i)\,\mathrm{Var}(X_{i+k})}}.$$

Una vez introducido el concepto de estacionariedad, podemos definir el concepto de ruido blanco.

Definición 7. Decimos que un proceso estocástico es un **ruido blanco** si cumple la siguientes condiciones:

- Es estacionario.
- Es incorrelado (tiene correlación 0 entre sus observaciones).
- Tiene media cero. ◂

Definición 8. Decimos que un proceso estocástico es un ruido blanco **estricto** si cumplen las siguientes condiciones:

- Las observaciones tienen la misma distribución.
- Las observaciones son independientes.
- Tiene media cero. ◂

El concepto de estacionariedad definido anteriormente puede ser aplicado igualmente a los incrementos de un proceso.

Definición 9. Decimos que $\{X_t : t \in T\}$ tiene incrementos estacionarios si, para cualquier $k \geq 1$ y para cualesquiera k puntos $t_1, \ldots, t_k$, la distribución de

$$(X_{t_1+c} - X_{t_1+c-1}, \ldots, X_{t_k+c} - X_{t_k+c-1})$$

no depende de $c \in \mathbb{R}$. ◂

4. Series temporales con propiedades fractales

Introduciremos los conceptos de ruido fraccionario gaussiano y movimiento fraccionario browniano. Los procesos estocásticos de este tipo poseen las propiedades fractales presentadas en la primera sección, y se supondrá que este tipo de procesos son capaces de modelar series temporales que presenten esta estructura fractal.

Definición 10. Sea $\{X_t : t \in T\}$ un proceso estocástico estacionario. Se dice que X_t es un proceso estacionario con larga memoria, o dependencia de largo alcance, o un proceso estacionario de correlaciones de largo alcance, si existen un número $\alpha \in (0, 1)$ y una constante $c_p > 0$ tales que

$$\lim_{k\to\infty} \frac{\rho(k)}{c_p k^{-\alpha}} = 1. \quad ◂$$

En el caso de procesos estocásticos, la autosimilitud no se define como vimos en la primera sección, ya que este era un concepto determinista. Las series temporales llevan implícita una aleatoriedad, lo que hace que nos fijemos en la distribución del proceso.

Definición 11. Sea $\{X_t : t \in T\}$ un proceso estocástico. Decimos que $\{X_t : t \in T\}$ es autosimilar con parámetro de autosimilitud $H > 0$ si, para cada número $c > 0$, el proceso reescalado con escala temporal ct, $c^{-H}X_{ct}$, tiene la misma distribución que el proceso X_t. ◂

Esto quiere decir que, para cualesquiera momentos en el tiempo $t_1, \dots, t_k$ y cualquier constante $c > 0$, tenemos que la distribución de $c^{-H}(X_{ct_1}, X_{ct_2}, \dots, X_{ct_k})$ es la misma que la distribución de $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$.

Con estos conceptos [2, 16], la correlación de proceso estocástico autosimilar es de la forma siguiente, dependiente del exponente H:

$$\rho(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right]. \tag{1}$$

Para estudiar el comportamiento asintótico de $\rho(k)$, utilizamos el desarrollo de Taylor, obteniendo que

$$\rho(k) \simeq H(2H-1)k^{2H-2} = \frac{H(2H-1)}{k^{2-2H}}. \tag{2}$$

Por lo tanto, al ser una ley de potencias, no posee una escala característica, poseyendo así las propiedades fractales. Definimos el exponente de la autocorrelación γ a partir del exponente de caída como

$$\gamma = 2 - 2H,$$

con $\gamma \in (0, 2)$. Para el caso en que $1/2 < H < 1$, tenemos que la correlación se va a cero y

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty,$$

y, por tanto, el proceso tiene correlaciones de largo alcance. Para el caso en que $H = 1/2$, tenemos que las correlaciones son cero y, por tanto, las observaciones X_i son incorreladas. Para el caso en que $0 < H < 1/2$, la suma de las correlaciones es convergente y, además,

$$\sum_{k=-\infty}^{\infty} \rho(k) = 0.$$

Como podemos imaginar, la aleatoriedad está presente en todos los fenómenos que ocurren en la naturaleza. Por lo tanto, sería deseable disponer de las herramientas necesarias para replicar esta aleatoriedad. Una de las técnicas más usadas para reproducir el movimiento (aleatorio) que produce una determinada partícula es el movimiento browniano.

Definición 12. Diremos que un proceso estocástico $\{X_t : t \in T\}$ es un proceso de Wiener o un **movimiento browniano** si cumple las siguientes propiedades:

- $X_0 = 0$.
- X tiene trayectorias continuas (el conjunto de puntos donde X es discontinuo tiene probabilidad cero).
- X tiene incrementos independientes y estacionarios con media 0.
- Para cada $t > 0$, la variable aleatoria X_t sigue una distribución $N(0, t)$. ◂

Supongamos ahora que tenemos un proceso Y_t que es autosimilar con incrementos $X_i = Y_i - Y_{i-1}$ estacionarios. Supongamos que los incrementos tienen media 0 y que cada X_t es una variable aleatoria normal. Por lo tanto, su distribución está totalmente definida por la media y las covarianzas. Se puede demostrar que, para cada $H \in (0, 1)$, existirá un único proceso gaussiano X_t, siendo estos los incrementos estacionarios de un proceso autosimilar Y_t.

Definición 13. Decimos que un proceso estocástico $\{X_t : t \in T\}$ es un ruido fraccionario gaussiano si es un proceso gaussiano y está definido como los incrementos de un proceso autosimilar Y_t. Se denomina movimiento browniano fraccionario al proceso Y_t, y se denota por $B_H(t)$. ◂

Si consideramos el caso $H = 1/2$, tenemos que, por la fórmula (1), las variables son normales e independientes. Es decir, para $H = 1/2$, el proceso es un movimiento browniano. Igualmente, se puede demostrar que, para el caso $H = 1/2$, el proceso es autosimilar. Para los casos $H > 1/2$, podemos observar en la ecuación (2) que las correlaciones son de largo alcance, ya que el exponente del denominador es menor que 1. Para los casos donde $H < 1/2$, tenemos que las correlaciones son de corto alcance, ya que la correlación decae mediante una ley de potencias con exponente mayor que 1.

5. Técnicas de análisis de fluctuaciones

A continuación, se introducirá el método de *análisis de fluctuaciones sin tendencia* (DFA, por sus siglas en inglés), el cual será el modelo central de estudio de las series temporales planteadas. Este estudiará las variaciones existentes en una serie temporal a distintas escalas, logrando un resultado mucho más suavizado y parecido a una ley de potencias que el logrado mediante la función de autocorrelación.

Igualmente, hablaremos del concepto de multifractal, que será utilizado para el estudio de las diferentes dinámicas que pueden estar presentes en una señal. Para ello, se utilizará la teoría presentada anteriormente y se aplicará al caso de las series temporales. De la misma forma, se describirá el algoritmo utilizado. Por último se hablará de un método de generación de series temporales a partir de su espectro de potencias.

5.1. Análisis de fluctuaciones sin tendencia (DFA)

El siguiente método que vamos a presentar es el denominado *detrended fluctuation analysis* (DFA), introducido por Peng *et al.* [21] para el estudio de las correlaciones de largo alcance en las secuencias de las cadenas de ADN. El método DFA es una modificación del FA (*fluctuation analysis*), ya que, en ese caso, se elimina la tendencia de la serie en cada ventana en el tiempo. Por lo tanto, es capaz de eliminar los efectos de la no estacionariedad y es menos sensible al tamaño finito de la serie [3, 6, 16].

Al igual que el FA, en el caso de que la serie escale (es decir, sea invariante frente a cambios de escala) y tenga las propiedades fractales, el método DFA también da como resultado un exponente que refleja las propiedades de escalado (es decir, autosimilitud) y las correlaciones de largo alcance de la serie analizada, en el caso de que la serie posea el escalado correspondiente. Por lo tanto, el DFA es una técnica de análisis de fluctuaciones (variaciones) que permite calcular de forma directa el exponente de la señal temporal a estudiar.

El algoritmo de aplicación de este método es el siguiente:

Paso 1: Supongamos que queremos analizar la serie X_t, con $t = 1, \dots, N$. En primer lugar, calculamos la media de las observaciones,

$$\overline{X} = \frac{1}{N} \sum_{t=1}^{N} X_t.$$

Paso 2: Una vez tenemos la media de la serie completa, calculamos la serie acumulada (o integrada), pero substrayendo la media en cada una de las observaciones:

$$Y_i = \sum_{j=1}^{i} (X_j - \overline{X}), \qquad i = 1, \dots, N.$$

Paso 3: A continuación, se divide la serie integrada en ventanas (subconjuntos ordenados de puntos consecutivos de la serie) de tamaño k. Para cada una de las ventanas, se calcula una regresión lineal con los puntos que están en dicha ventana, lo cual representa la tendencia lineal en la ventana correspondiente:

$$\overline{Y}_\ell^j(k) = Y_j - y_j(k),$$

donde $y_j(k)$ es el ajuste lineal de los datos para la ventana ℓ. De esta forma, estamos quitándole la tendencia lineal en cada una de las ventanas. Con ese ajuste, eliminamos la tendencia y el ruido que pueda contener la serie a analizar. Tras el ajuste lineal, las variaciones de los residuos serán estudiados posteriormente.

Paso 4: Para cada una de las ventanas de tamaño k, se calcula la raíz cuadrática media de la señal integrada, una vez eliminada la tendencia:

$$F_\ell^2(k) = \frac{1}{k}\sum_{j=1}^{k} \overline{Y_\ell^j(k)}^2 .$$

Paso 5: Una vez que se ha realizado el paso 4 para toda la serie, se calcula la media de todos los $F_\ell(k)$ para un determinado tamaño de ventana k:

$$F(k) = \sqrt{\frac{1}{N/k}\sum_{\ell=1}^{N/k} F_\ell^2(k)}.$$

Paso 6: $F(k)$ describe como se comportan las fluctuaciones (momento de orden 2) alrededor de la tendencia local en función de la escala. Si la señal tiene autosimilitud, se obtendrá la relación

$$F(k) \approx ak^\alpha, \tag{3}$$

cuyo exponente α podrá ser calculado llevando a cabo un ajuste lineal de los datos en escala logarítmica,

$$\log(F(k)) \approx c + \alpha \log(k).$$

Para una serie con correlaciones (tanto de largo como de corto alcance), se debe cumplir que la fluctuación escale como una ley de potencias en función del tamaño de ventana (3). En el caso del DFA, la serie temporal presenta anticorrelaciones si $\alpha < 0{,}5$; es ruidosa o no correlacionada si $\alpha = 0{,}5$, y presenta correlaciones de largo alcance si $\alpha \in (0{,}5, 2)$. En particular, para el caso $\alpha = 1{,}5$, la serie corresponde al movimiento browniano presentado anteriormente.

Este nuevo método de cálculo tiene una relación con la conocida función de autocorrelación. Tal y como señalan Höll y Kantz [16], la relación entre la función de autocorrelación y el exponente DFA viene dada por la siguiente identidad:

$$F^2(s) = \langle x^2 \rangle \left(W(s) + \sum_{r=1}^{s-1} C(r) L_r(s) \right).$$

5.2. DFA multifractal

Como ya comentamos anteriormente, los conjuntos fractales son aquellos conjuntos formados por otros subconjuntos, a su vez igualmente fractales entrelazados, pero con diferente dimensión fractal. En ese caso, se necesita realizar un análisis multifractal de la serie temporal para tener una descripción completa de esta, obteniendo un espectro multifractal que nos revelará la existencia de monofractalidad o multifractalidad en la serie. El análisis que se realizará estará basado en el formalismo presentado en la sección 2, que fue utilizado inicialmente por Kantelhardt *et al.* [17] en el estudio de series complejas. El método es una simple generalización del DFA, cambiando los momentos de las fluctuaciones de la serie. El algoritmo es el siguiente:

Paso 1: Supongamos que queremos analizar la serie X_t, con $t = 1, \dots, N$. En primer lugar, calculamos la media de las observaciones,

$$\overline{X} = \frac{1}{N}\sum_{t=1}^{N} X_t.$$

Paso 2: Una vez tenemos la media de la serie completa, calculamos la serie acumulada (o integrada), pero substrayendo la media en cada una de las observaciones

$$Y_i = \sum_{j=1}^{i} (X_j - \overline{X}), \qquad i = 1, \dots, N.$$

Paso 3: A continuación, se divide la serie integrada en ventanas de tamaño k. Para cada una de las ventanas, se calcula una regresión lineal con los puntos que están en dicha ventana, lo cual representa la tendencia lineal en la ventana correspondiente:

$$\overline{Y}_\ell^j(k) = Y_j - y_j(k),$$

donde $y_j(k)$ es el ajuste lineal de los datos que están en la ventana ℓ. De esta forma, se quita la tendencia lineal en cada una de las ventanas. Con ese ajuste, nos eliminamos la tendencia y el ruido que pueda contener la serie a analizar.

Paso 4: Para cada una de las ventanas de tamaño k, se calcula la raíz cuadrática media de la señal integrada, una vez eliminada la tendencia:

$$F_\ell(k) = \sqrt{\frac{1}{k}\sum_{j=1}^{k} \overline{Y_\ell^j(k)}^2}.$$

Paso 5: Una vez que se ha realizado el paso 4 anterior para toda la serie, se realiza la media de todos los $F_\ell(k)$ para un determinado tamaño de ventana k y para cada orden q:

$$F_q(k) = \left\{ \frac{1}{N/k} \sum_{\ell=1}^{N/k} \left[F_\ell^2(k)\right]^{q/2} \right\}^{1/q},$$

donde, para $q = 2$, tenemos como resultado el exponente resultante de aplicar el método DFA. Una vez realizado este paso, estamos interesados en conocer el comportamiento de $F_q(k)$ con respecto a q. Es por ello que este proceso tiene que ser repetido para cada orden q.

Paso 6: Existirá un escalado y la serie temporal tendrá autosimilitud si encontramos la siguiente relación:

$$F_q(k) \approx ak^{h(q)},$$

donde $h(q)$ se obtiene, para cada q, mediante un ajuste lineal de los datos en escala logarítmica,

$$\log(F_q(k)) \approx c + h(q)\log(k).$$

En general, el exponente $h(q)$ dependerá de q. Para series monofractales, $h(q)$ será independiente de q, ya que el escalado de las fluctuaciones será igual para todos los tamaños de ventana y, por tanto, dará el mismo exponente en todos los órdenes.

Una vez realizado el algoritmo completo, los resultados se suelen presentar de la misma manera que la utilizada en la sección 2. Para ello, para cada $h(q)$ obtenido, calculamos la función

$$\tau(q) = qh(q) - 1. \tag{4}$$

Utilizando, por tanto, el procedimiento presentado en la sección 2, tenemos que el espectro multifractal vendrá dado por

$$\alpha = h(q) + qh'(q), \qquad f(\alpha) = q[\alpha - h(q)] + 1,$$

donde $f(\alpha)$ denota la dimensión del subconjunto de la serie caracterizada por α, la cual está relacionada con τ de la ecuación (4) mediante $\alpha = \tau'(q)$.

6. Generación de series temporales con propiedades fractales

Con los métodos anteriores se pueden analizar las diferentes series temporales presentes en la naturaleza. Sin embargo, es necesario comprobar los resultados que se obtienen con series que tienen una dinámica conocida [**2**]. Dicha comprobación se realiza generando una serie temporal con el mismo exponente DFA, replicando así la dinámica de la serie temporal a analizar y, de esta forma, comparando los espectros multifractales de ambas series aislándolas de todos los efectos exteriores a estas.

Definición 14. Sea $x(t)$ una señal. La transformada de Fourier y la transformada inversa se definen, respectivamente, por las expresiones

$$y(\xi) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{i}\xi t}\,\mathrm{d}t, \qquad x(t) = \int_{-\infty}^{\infty} y(\xi)\mathrm{e}^{\mathrm{i}\xi t}\,\mathrm{d}\xi. \quad \blacktriangleleft$$

Definición 15. Se define la **densidad espectral** (o espectro de potencias) de una señal como la función que representa la distribución de la potencia de la señal en el dominio de frecuencias de esta, calculada mediante la transformada de Fourier de la serie. Se suele denotar por $S(\xi) = |\mathcal{F}(x)|^2$, donde $\mathcal{F}(x)$ es la transformada de Fourier de la señal $x(t)$. $\blacktriangleleft$

Con esas herramientas, a partir de una señal podemos calcular su espectro de potencias, y viceversa. La idea del siguiente algoritmo es la creación de una señal cuyo espectro de potencias, calculado mediante la transformada de Fourier, sea plano. Con eso, se modificará el espectro de potencias para que adopte la forma que necesitamos para comparar la señal. Una vez realizado este último paso, se calculará la transformada inversa del espectro de potencias modificado para obtener la señal resultante con el espectro multifractal buscado. Para ello, utilizamos el teorema de Wiener-Khinchin:

Teorema 16 (teorema de Wiener-Khinchin). *Si un proceso estocástico es estacionario, su densidad espectral de potencias se puede expresar como la transformada de Fourier de la función de autocorrelación*

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau.$$

Demostración. La demostración se obtiene mediante una serie de igualdades. Supongamos en primer lugar que el proceso tiene media 0 y que su covarianza es de la forma $\mathrm{Cov}(x(n), x(m)) = \gamma(n-m)$. Por lo tanto,

$$\begin{aligned}
S(\omega) &= \lim_{T\to\infty} \frac{1}{2T}\mathbb{E}\left[|\mathcal{F}_T(\omega)|^2\right] = \lim_{T\to\infty} \frac{1}{2T}\mathbb{E}\left[\left(\int_{-T}^{T} x(\zeta)\mathrm{e}^{-\mathrm{i}\omega\zeta}\,\mathrm{d}\zeta\right)\left(\int_{-T}^{T} x(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau\right)^*\right] \\
&= \lim_{T\to\infty} \frac{1}{2T}\mathbb{E}\left[\int_{-T}^{T}\int_{-T}^{T} x(\zeta)x(\tau)\mathrm{e}^{-\mathrm{i}\omega(\zeta-\tau)}\,\mathrm{d}\tau\,\mathrm{d}\zeta\right] = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} \mathbb{E}[x(\zeta)x(\tau)]\mathrm{e}^{-\mathrm{i}\omega(\zeta-\tau)}\,\mathrm{d}\tau\,\mathrm{d}\zeta \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R(\zeta-\tau)\mathrm{e}^{-\mathrm{i}\omega(\zeta-\tau)}\,\mathrm{d}\tau\,\mathrm{d}\zeta = \lim_{T\to\infty}\lim_{M\to\infty} \frac{1}{2T}\int_{-T}^{T}\int_{-M}^{M} R(\zeta-\tau)\mathrm{e}^{-\mathrm{i}\omega(\zeta-\tau)}\,\mathrm{d}\tau\,\mathrm{d}\zeta \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\lim_{M\to\infty}\int_{-M}^{M} R(\zeta-\tau)\mathrm{e}^{-\mathrm{i}\omega(\zeta-\tau)}\,\mathrm{d}\tau\,\mathrm{d}\zeta = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\left(\int_{-\infty}^{\infty} R(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau\right)\mathrm{d}\zeta \\
&= \left(\int_{-\infty}^{\infty} R(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau\right)\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\mathrm{d}\zeta = \int_{-\infty}^{\infty} R(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau. \quad \blacksquare
\end{aligned}$$

Existen numerosos algoritmos para la generación de series. Uno de ellos es el siguiente:

Paso 1: Creamos un vector con la misma longitud que la serie y con valor 1 en todos sus puntos.

Paso 2: Multiplicamos el vector por

$$\mathrm{Re}(\vartheta(f)) = f^{-(2\alpha-1)/2}\cos(2\pi u), \qquad \mathrm{Im}(\vartheta(f)) = f^{-(2\alpha-1)/2}\,\mathrm{sen}(2\pi u),$$

donde α es el exponente obtenido del análisis DFA de la serie, u son números aleatorios que siguen una distribución $U(0,1)$ y $\mathrm{Re}(\vartheta(i))$ y $\mathrm{Im}(\vartheta(i))$ son la parte real e imaginaria de la transformada de Fourier, respectivamente.

Paso 3: Calculamos la transformada inversa para obtener la serie resultante,

$$x(i) = \mathcal{F}^{-1}[\vartheta(i)].$$

6.1. Generación de series temporales con *crossover*

Mediante los métodos anteriores, generamos una señal que escala de una única forma. Sin embargo, los sistemas dinámicos reales producen series que escalan de diferente manera, generando *crossovers* (tamaño de ventana en el cual se produce un cambio de pendiente en el escalado de la serie) en el espectro de exponentes en algunas ocasiones. Por lo tanto, es necesario un nuevo algoritmo para modelar dichas series [5].

En el caso de querer generar una serie con dos exponentes de escalado, uno de ellos para escalas menores que el *crossover* y otro para las escalas de mayores que este, hacemos una diferenciación en la definición del espectro de potencias. Así, si el *crossover* se presenta en t_c, entonces multiplicamos el espectro por

$$Q(f) = \begin{cases} f^{-(2\alpha_\ell - 1)/2} & \text{si } f \leq f_c, \\ f_c^{(\alpha_s - \alpha_\ell)} f^{-(2\alpha_s - 1)/2} & \text{si } f > f_c, \end{cases}$$

donde $f_c = t_c^{-1}$ es el inverso del *crossover* y representa el cambio de tendencia en el espectro de potencias, y α_s y α_ℓ denotan los exponentes a corta y larga escala de la serie, respectivamente. De esta forma, conseguimos crear un espectro de potencias que es continuo.

7. Resultados

En esta sección, se describirán los resultados de las diferentes series reales que hemos seleccionado para su análisis. En primer lugar, se han elegido series financieras debido a la disponibilidad de estas de forma libre. Sin embargo, como veremos más adelante, estas series no son muy largas, debido a la digitalización relativamente reciente, que hace que no haya tantos datos disponibles como nos gustaría.

Por ello, se ha elegido también la serie de las diferencias de presiones entre las Azores e Islandia, ya que dicha señal tiene un número mayor de datos, aparte de suponer una aplicación directa de estos métodos en el ámbito de la física [12, 13].

7.1. Series financieras

En esta sección se explicarán los diferentes resultados obtenidos para las compañías analizadas. Se hacen tres distinciones en las series analizadas: empresas en el mercado español, empresas de EE. UU. e índices bursátiles. Los datos han sido obtenidos en línea [8, 9].

Se espera una dinámica más compleja en este último mercado, ya que es el tomado como referencia por todos los inversores al ser la Bolsa de Nueva York (Wall Street) el mercado con mayor número de transacciones, seguido de Londres y Tokyo. Al seleccionar las empresas para analizar, se han tenido en cuenta los siguientes criterios:

- Disponibilidad de los datos: al obtener las series temporales, numerosas de ellas presentaban numerosos huecos de cotización en algunos días, lo cual podía inducir a errores en los análisis.
- Diferentes sectores: se han estudiado empresas de energía, financieras, consumo y tecnológicas.

Igualmente, también se han examinado los retornos absolutos de la serie y la volatilidad de estos.

Definición 17. Se define el **retorno** (o incremento) de una serie temporal mediante la diferencia de las observaciones. Igualmente, se define la **volatilidad** de una serie temporal como el logaritmo del valor absoluto de las variaciones. Así,

$$r_t = X_t - X_{t-1} \quad \text{y} \quad \text{Vol}_t = \log\left(\left|\frac{X_t}{X_{t-1}}\right|\right). \qquad \blacktriangleleft$$

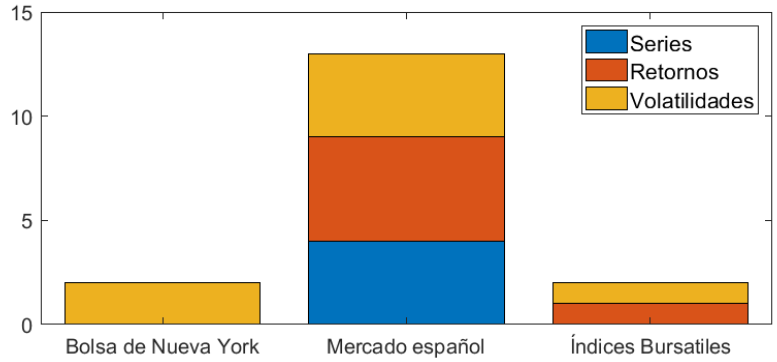


Figura 1: Número de series que no escalan correctamente en cada uno de los activos financieros analizados.

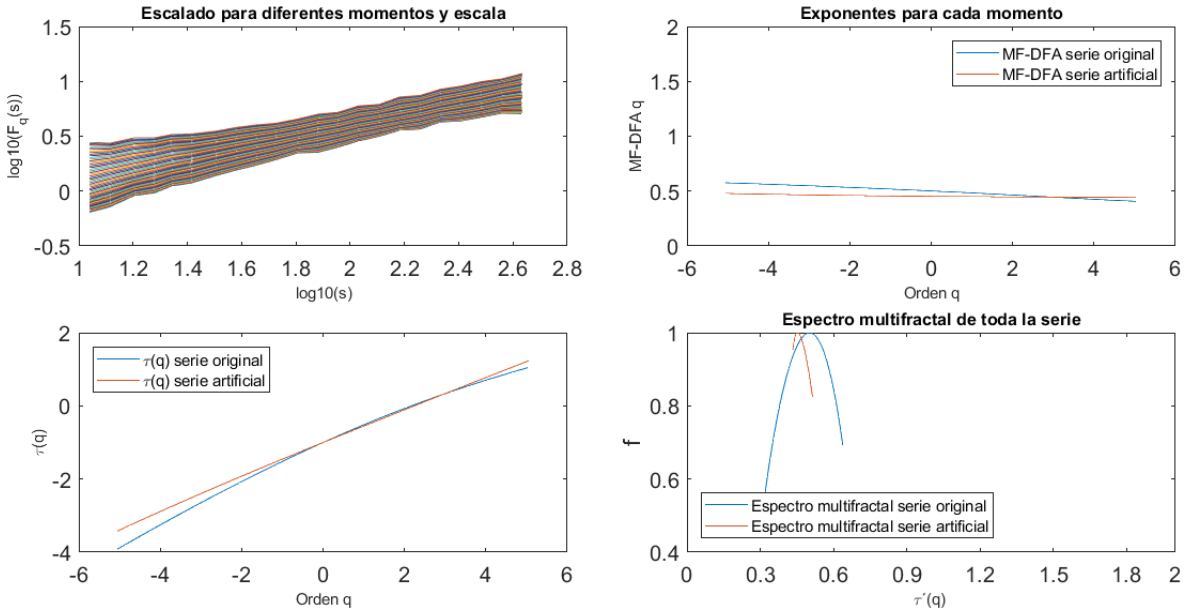


Figura 2: Análisis multifractal de la serie de volatilidades de Repsol.

El estudio de los retornos y la volatilidad es muy importante en el ámbito financiero, al determinar estos la viabilidad de una inversión en numerosas ocasiones. Las series seleccionadas son las siguientes:

- Empresas cotizadas en el mercado español: Acciona, BBVA, Banco Santander, Endesa, Iberdrola, Melia Hotels International, Red Eléctrica Española, Telefónica, Repsol, Viscofan.
- Empresas cotizadas en EE. UU.: Apple, The Boing Company, Citigroup, Ebay, Starbucks Corporation.
- Índices bursátiles: IBEX 35, S&P 500, Dow Jones Industrial Average, Nasdaq Composite.

No se presentarán todos los resultados que se obtengan de estas, ya que muchas de ellas suelen tener un comportamiento parecido a otras. Como resumen general, concluimos que se produce un escalado generalizado en las series analizadas. Es notable el caso de las cotizaciones bursátiles, ya que se produce un escalado correcto (se obtiene un buen ajuste en la regresión lineal anteriormente expuesta) en el 80 % de los casos. Igualmente, vemos cómo las volatilidades tienen un comportamiento peor, ya que el 40 % de estas no escalan correctamente. La figura 1 nos muestra los resultados por mercado.

Como podemos apreciar en la figura 1, el mercado español presenta numerosas series que no escalan correctamente, siendo el número muy parecido en los tres tipos de series. Esto puede ser debido a la técnica de *análisis técnico* que se produce en los EE. UU., el mayor volumen de negocio y la cualificación de los inversores.

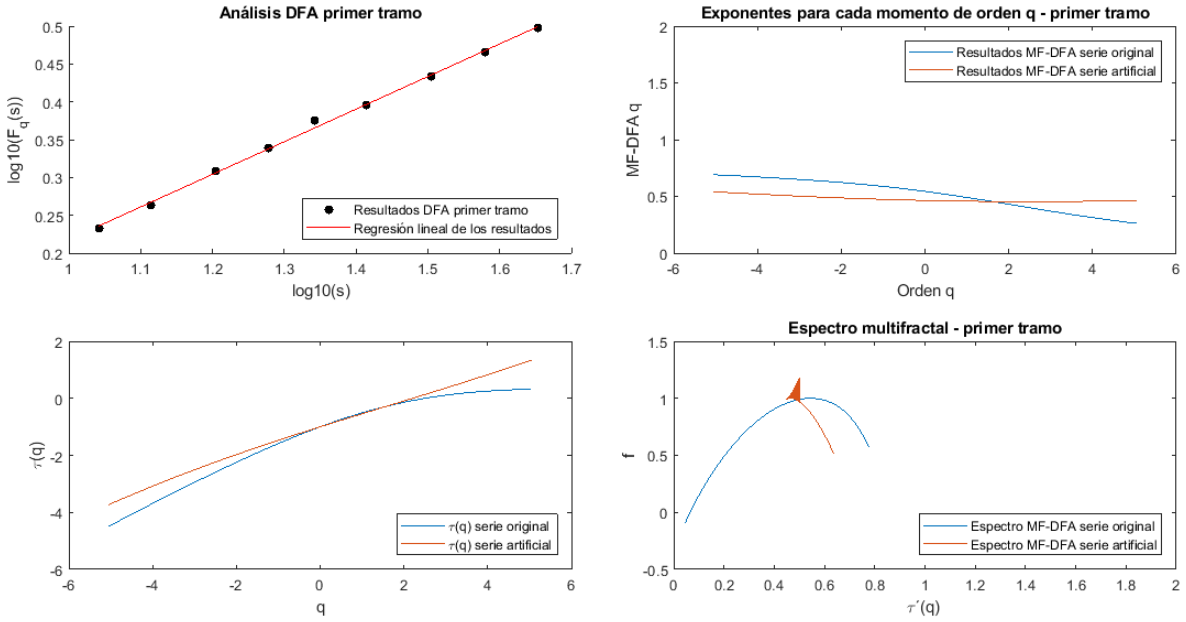


Figura 3: Análisis multifractal del primer tramo de la serie de volatilidades de Repsol.

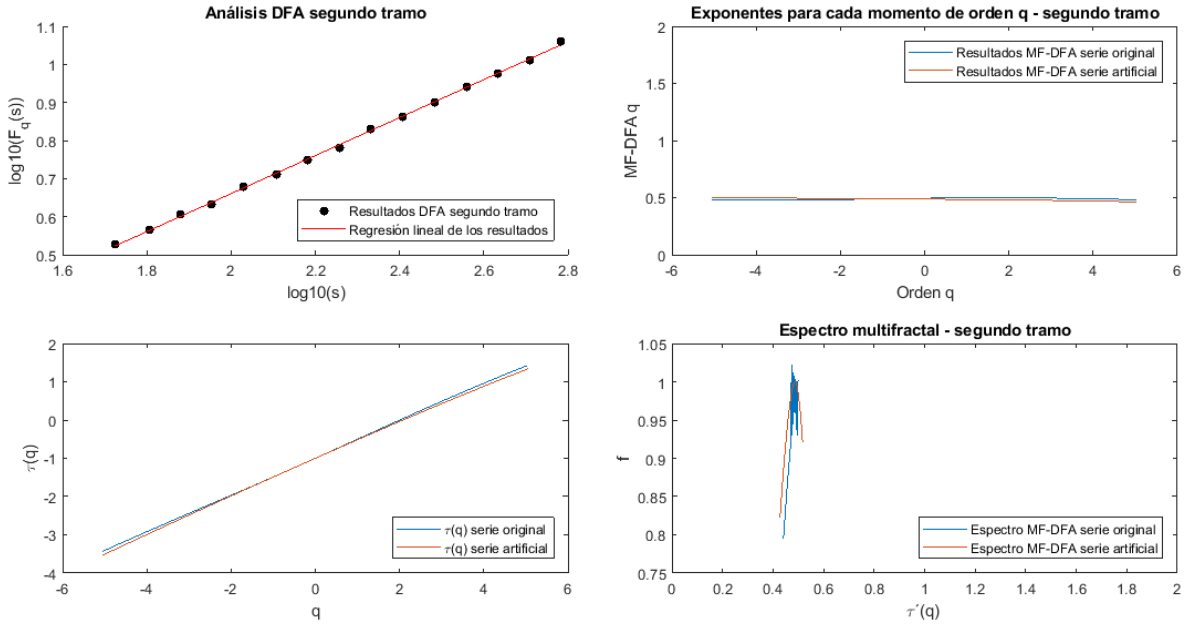


Figura 4: Análisis multifractal del segundo tramo de la serie de volatilidades de Repsol.

Concluimos que alrededor del 40 % de las series que tienen propiedades de escalado presentan propiedades monofractales. Asimismo, un 55 % de estas tienen características multifractales, debido a su amplio espectro multifractal. En el análisis de cada una de las series temporales, se ha realizado un análisis multifractal, que nos revela la posible existencia de algún *crossover* y la multifractalidad de la serie en cada uno de los tramos determinados por los posibles *crossover*. Así, nos podemos encontrar con la existencia de un *crossover* en la serie temporal de las volatilidades de Repsol, como observamos en la figura 2, donde nos encontramos con un cambio de pendiente alrededor del punto 2 en el gráfico izquierdo superior.

Esto implica una dinámica muy compleja, ya que, en el corto plazo (en el primer tramo de la serie dado por la figura 3), la serie temporal es multifractal, mientras que, en el largo plazo (segundo tramo dado por la figura 4), la serie es monofractal.

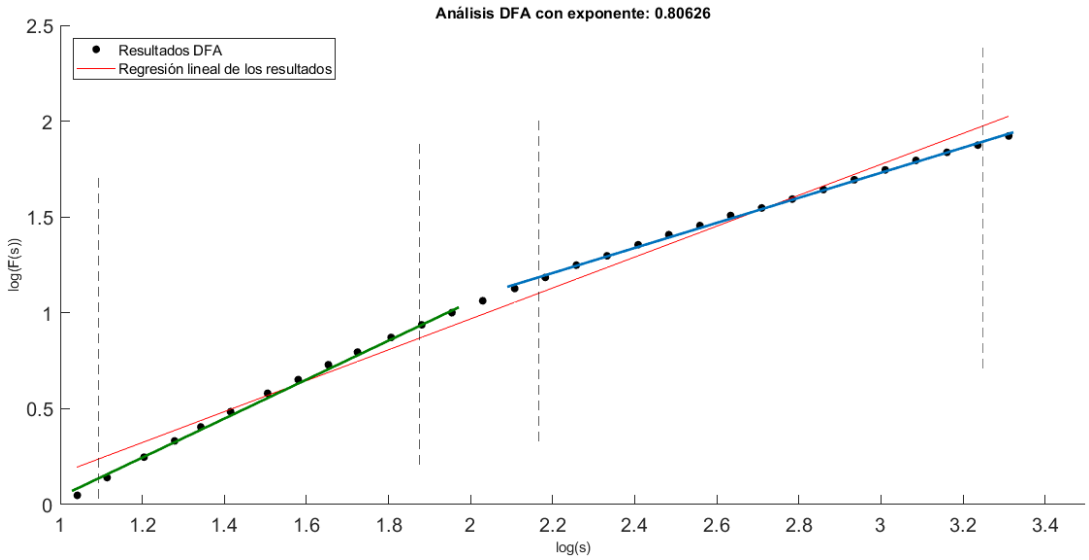


Figura 5: Método DFA aplicado a la serie.

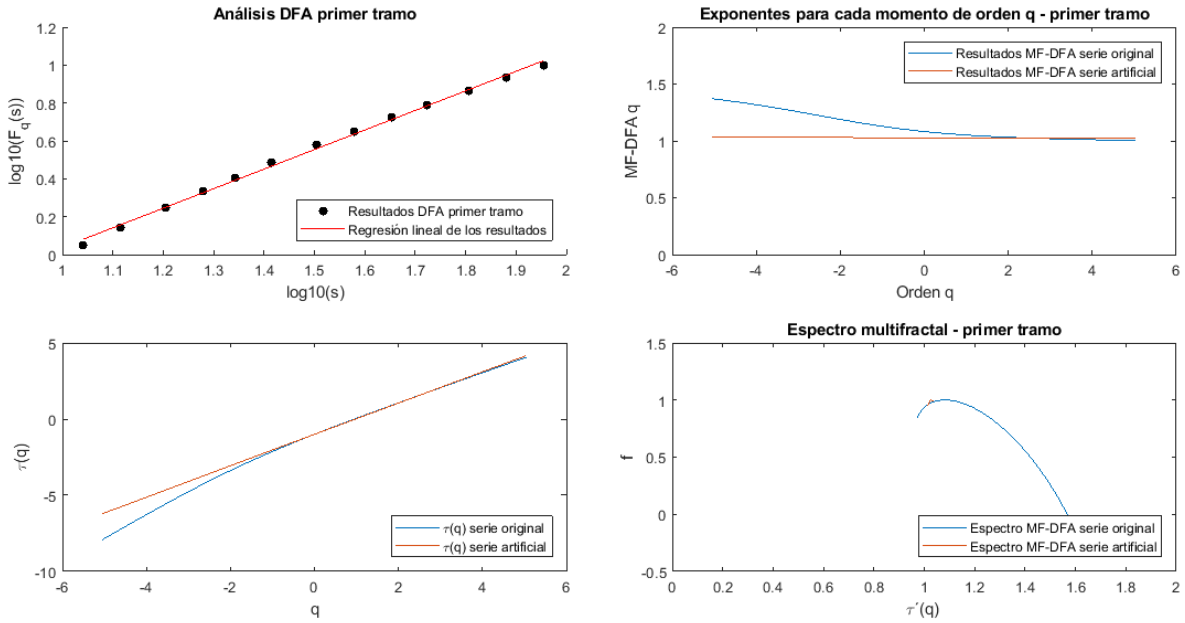


Figura 6: Análisis multifractal del primer tramo de la serie.

7.2. Variaciones de presión en el Océano Atlántico Norte

Procedemos ahora a describir los resultados de la serie temporal de las oscilaciones del atlántico norte, (NAO, North Atlantic Oscilations, en inglés), siendo estos datos obtenidos de [7]. Este es un fenómeno climático que recoge las fluctuaciones de la diferencia de presión atmosférica entre Islandia y las Azores. La NAO se descubre en los 1920s por Gilbert Walker. Siendo similar al fenómeno de El Niño en el océano Pacífico, la NAO es una de las más importantes conductoras de las fluctuaciones climáticas en el Noratlántico y climas húmedos vecinos [12, 13].

Calculamos en primer lugar el DFA de la serie y vemos su posible escalado en la figura 5. Podemos observar un pequeño cambio de tendencia alrededor de la escala $2 = \log(100)$ (en escala logarítmica). En este caso, obtenemos un exponente DFA de 1,0408 por debajo de las escalas de tamaño 100, y un exponente DFA de 0,6627 para escalas con un tamaño mayor que 100.

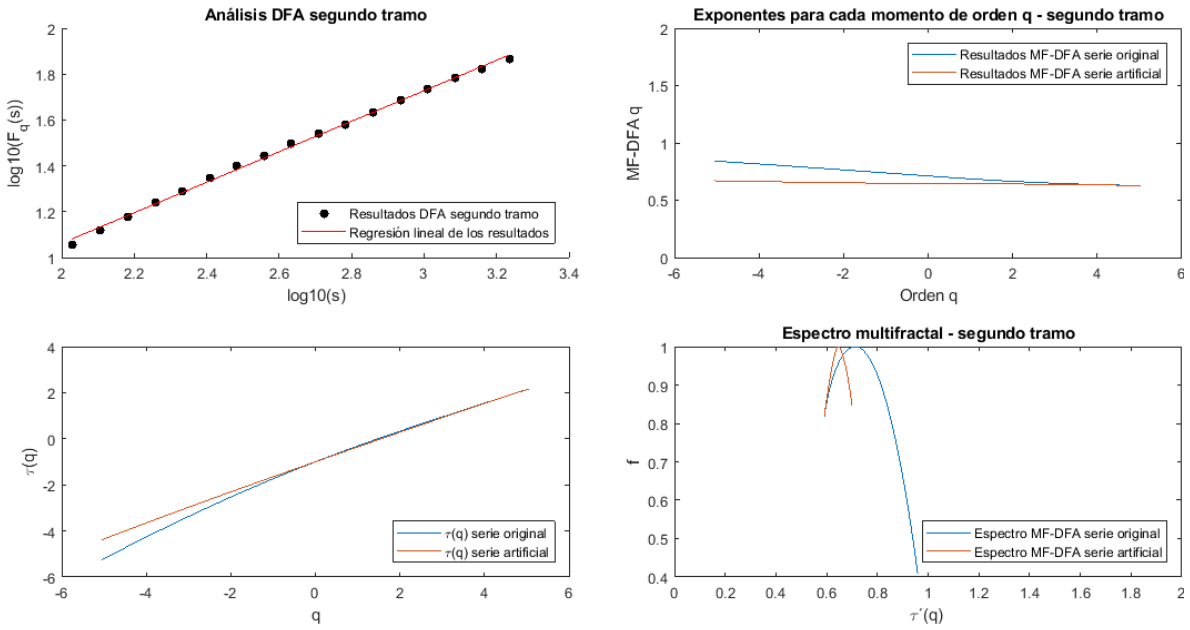


Figura 7: Análisis multifractal del segundo tramo de la serie.

De esta forma, la serie escala correctamente con un *crossover*. Procedemos ahora a realizar el cálculo del multifractal. Para ello, se emplearía el método multifractal ya presentado, y se pintaría $F_q(s)$ vs q en escala logarítmica, obteniendo un cambio de tendencia en la señal alrededor del punto 2.

Por este motivo, realizaremos un análisis multifractal de cada parte de la serie. Concretamente, tomaremos los tamaños de ventana de 1 a 2 y, por otro lado, de 2 a 3,3, y calcularemos su espectro multifractal. Esto será realizado para el estudio por separado de ambas dinámicas presentes en la serie, y estudiar la mono- o multifractalidad de cada una de ellas. En el primer tramo, obtenemos los resultados dados en la figura 6.

Observamos cómo la serie sigue teniendo un espectro multifractal más amplio que la serie de control (serie artificial generada) en ese primer tramo, teniendo esta un rango de 0,59 en comparación con la serie de control, la cual posee un rango de 0,05. Por lo tanto, concluimos que la serie es multifractal en el corto plazo (para ventanas de tamaño menor que 100).

Se puede observar en la figura 7 el mismo resultado en el segundo tramo de la serie. Vemos cómo la serie de control tiene un espectro más estrecho que la segunda parte de la serie. Cada una tiene un rango de 0,16 y 0,36, respectivamente, indicándonos de nuevo la multifractalidad para ventanas con un tamaño mayor que 100.

8. Conclusiones

Las principales aportaciones y conclusiones de este trabajo son las siguientes:

1. En este trabajo se ha introducido el método DFA, que mejora otros métodos previos empleados en el análisis de diferentes series temporales complejas, con el fin de estudiar su dinámica y las propiedades fractales presentes en ellas. El DFA es aplicable tanto a series estacionarias como no estacionarias y se ha relacionado el resultado que nos produce el DFA con la función de autocorrelación para series estacionarias. Igualmente, se ha introducido el DFA multifractal, útil para analizar series más complejas, el cual analiza las propiedades de escala de los diferentes momentos de las fluctuactiones de las series.
2. Se han presentado los algoritmos necesarios para generar series de control utilizando la transformada de Fourier, así como el espectro de potencias. Con estos algoritmos, somos capaces de generar series con una dinámica monofractal concreta. Estas series sintéticas se usan para modelar las series reales, y son utilizadas posteriormente como referencia para comparar con los resultados de las señales reales analizadas.

3. Los métodos DFA y DFA multifractal han sido empleados para analizar numerosas cotizaciones bursátiles de los mercados financieros, así como los retornos y las volatilidades de estas. Estas señales son conocidas por su complejidad y sus características fractales. Se concluye que alrededor de un 70 % poseen propiedades fractales, y aproximadamente un 40 % son multifractales. Eso nos revela una dinámica compleja presente en la mayoría de las series temporales. Igualmente, observamos series con propiedades fractales distintas a pequeñas y grandes escalas temporales, con una transición entre ambas (*crossover*) en una escala intermedia, lo que sugiere la existencia de dos mecanismos distintos que controlan la dinámica a corto y largo plazo.
4. Finalmente, se ha analizado el índice NAO, el cual mide las variaciones de presión en diferentes puntos del Atlántico Norte. Los resultados revelan una dinámica muy compleja, presentando un cambio de tendencia en cada estación del año y propiedades multifractales en cada una de ellas.

Referencias

[1] Allegrini, Paolo; Barbi, Maria; Grigolini, Paolo, y West, Bruce J. «Dynamical model for DNA sequences». En: *Physical Review E* 52.5 (1995), págs. 5281-5296. https://doi.org/10.1103/PhysRevE.52.5281.

[2] Beran, Jan. *Statistics for long-memory processes*. Vol. 61. Monographs on Statistics and Applied Probability. Chapman y Hall, New York, 1994. ISBN: 978-0-412-04901-9.

[3] Bryce, Robert M. y Sprague, Kevin B. «Revisiting detrended fluctuation analysis». En: *Scientific Reports* 2 (2012). https://doi.org/10.1038/srep00315.

[4] Bunde, Armin y Havlin, Shlomo, eds. *Fractals in science*. Berlín: Springer-Verlag, 1994. ISBN: 978-3-540-56221-4.

[5] Carpena, Pedro; Coronado, Ana; Carretero-Campos, Concepción; Bernaola-Galván, Pedro, y Ivanov, Plamen. «First Passage time properties of correlated time series with scale-invariant behaviour with crossovers in the scaling». En: *Time Series Analysis and Forecasting*. Ed. por Rojas, Ignacio y Pomares, Héctor. Springer, 2016, págs. 89-102. https://doi.org/10.1007/978-3-319-28725-6.

[6] Coronado, Ana V. y Carpena, Pedro. «Size Effects on Correlation Measures». En: *Journal of Biological Physics* 31.1 (2005), págs. 121-133. ISSN: 1573-0689. https://doi.org/10.1007/s10867-005-3126-8.

[7] *Datos de la serie NAO*. URL: http://www.cpc.ncep.noaa.gov.

[8] *Datos de las series financieras de España*. URL: https://www.infobolsa.es.

[9] *Datos de las series financieras internacionales*. URL: http://www.investing.com.

[10] Faleiro, Eduardo. *Estructura Multifractal y Aplicaciones de las fluctuaciones en cascadas atmosféricas producidas por rayos cósmicos*. Tesis doctoral. Universidad Complutense de Madrid, 1998. ISBN: 978-84-669-1571-7. URL: https://eprints.ucm.es/3123/.

[11] Feder, Jens. *Fractals*. Physics of Solids and Liquids. With a foreword by Benoit B. Mandelbrot. Plenum Press, New York, 1988. https://doi.org/10.1007/978-1-4899-2124-6.

[12] Fernández, Isabel; Hernández, Carmen N., y Pacheco, José M. «Is the North Atlantic Oscillation just a pink noise?» En: *Physica A: Statistical Mechanics and its Applications* 323 (2003), págs. 705-714. ISSN: 0378-4371. https://doi.org/10.1016/S0378-4371(03)00056-6.

[13] Fernández, Isabel; Pacheco, José M., y Quintana, María P. «Pinkness of the North Atlantic Oscillation signal revisited». En: *Physica A: Statistical Mechanics and its Applications* 389.24 (2010), págs. 5801-5807. ISSN: 0378-4371. https://doi.org/10.1016/j.physa.2010.08.003.

[14] Halsey, Thomas C.; Jensen, Mogens H.; Kadanoff, Leo P.; Procaccia, Itamar, y Shraiman, Boris I. «Fractal measures and their singularities: The characterization of strange sets». En: *Physical Review A* 33.2 (1986), págs. 1141-1151. https://doi.org/10.1103/PhysRevA.33.1141.

[15] Hastings, Harold M. y Sugihara, George. *Fractals: A User's Guide for the Natural Sciences*. Oxford: Oxford University Press, 1993. ISBN: 978-0-19-854597-2.

[16] Höll, Marc y Kantz, Holger. «The relationship between the detrendend fluctuation analysis and the autocorrelation function of a signal». En: *The European Physical Journal B* 88.12 (2015), pág. 327. issn: 1434-6036. https://doi.org/10.1140/epjb/e2015-60721-1.

[17] Kantelhardt, Jan W.; Zschiegner, Stephan A.; Koscielny-Bunde, Eva; Havlin, Shlomo; Bunde, Armin, y Stanley, H. Eugene. «Multifractal detrended fluctuation analysis of nonstationary time series». En: *Physica A: Statistical Mechanics and its Applications* 316.1 (2002), págs. 87-114. issn: 0378-4371. https://doi.org/https://doi.org/10.1016/S0378-4371(02)01383-3.

[18] Martín Aguilar, Alberto. *Análisis de señales complejas: correlaciones de largo alcance y propiedades multifractales*. Trabajo Fin de Máster. Universidad de Málaga, 2018.

[19] Moura, Francisco A. B. F. de y Lyra, Marcelo L. «Delocalization in the 1D Anderson Model with Long-Range Correlated Disorder». En: *Physical Review Letters* 81.17 (1998), págs. 3735-3738. https://doi.org/10.1103/PhysRevLett.81.3735.

[20] Peng, Chung-Kang; Buldyrev, Sergey V.; Goldberger, Aary; Halvin, Shlomo; Sciortino, F.; Simons, M., y Stanley, H. Eugene. «Long-range correlations in nucleotide sequences». En: *Nature* 356 (1992), págs. 168-170. https://doi.org/10.1038/356168a0.

[21] Peng, Chung-Kang; Buldyrev, Sergey V.; Havlin, Shlomo; Simons, Michael; Stanley, H. Eugene, y Goldberger, Ary L. «Mosaic organization of DNA nucleotides». En: *Physical Review E* 49.2 (1994), págs. 1685-1689. https://doi.org/10.1103/PhysRevE.49.1685.

[22] Seuront, Laurent. *Fractals and Multifractals in Ecology and Aquatic Science*. Boca Raton, Florida: CRC Press, 2010. https://doi.org/10.1201/9781420004243.

TEMat

Grafo asociado a los tamaños de las clases de conjugación de un grupo finito

A la memoria de Carlo Casolo

Víctor Sotomayor
Centro Universitario EDEM - Valencia
vsotomayor@edem.es

Resumen: En el presente trabajo asociamos a los tamaños de las clases de conjugación de un grupo finito G el denominado *grafo primo* $\Delta(G)$: los vértices son los números primos que dividen a algún tamaño de clase de G, y dos primos p y q forman una arista si pq divide a algún tamaño de clase. Nuestro objetivo en este artículo es doble: por un lado, mostrar algunos resultados básicos en esta área de investigación, y por otro lado, demostrar un bonito e importante teorema de S. Dolfi sobre la inexistencia de *conjuntos independientes* de tres vértices en este grafo cuando G es resoluble; es decir, dados tres vértices en $\Delta(G)$, siempre existen al menos dos de ellos conectados.

Abstract: In this paper we attach to the set of conjugacy class sizes of a finite group G the so-called *prime graph* $\Delta(G)$: the vertices are the prime numbers that divide some conjugacy class size of G, and two primes p and q form and edge whenever pq divides some class size. Our objective in this note is twofold: on the one hand, to show some basic results within this research area; on the other hand, to prove a nice and important theorem due to S. Dolfi about the non-existence of *independent sets* of three vertices in this graph when G is soluble; that is, given three distinct vertices of $\Delta(G)$, there always exist at least two of them which are connected.

Palabras clave: grupos finitos, clases de conjugación, grafos.

MSC2010: 20E45.

Recibido: 9 de agosto de 2019.
Aceptado: 15 de enero de 2020.

Agradecimientos: Este trabajo ha sido financiado por el contrato predoctoral ACIF/2016/170 de la Generalitat Valenciana, España. Me gustaría agradecer a los revisores sus cuidadosas revisiones y sus sugerencias de mejora.

Referencia: Sotomayor, Víctor. «Grafo asociado a los tamaños de las clases de conjugación de un grupo finito». En: *TEMat*, 4 (2020), págs. 101-113. ISSN: 2530-9633. URL: https://temat.es/articulo/2020-p101.

1. Introducción

En el presente trabajo, todos los grupos considerados son finitos. El estudio de la relación existente entre ciertos números naturales asociados a un grupo y su estructura es un amplio tema de investigación dentro de la teoría de grupos finitos. Concretamente, los tamaños de las clases de conjugación han tenido una especial relevancia en las últimas décadas. En el artículo de Ortiz Sotomayor [9] se muestra una introducción preliminar a esta área, estudiándose cómo algunas propiedades aritméticas concretas de los tamaños de clase controlan la estructura del grupo.

En esta línea, diversos autores están considerando recientemente un novedoso enfoque: grafos asociados a los tamaños de clase de un grupo G. Uno de los grafos que más interés está teniendo es el denominado *grafo primo* $\Delta(G)$, donde el conjunto de vértices $V(G)$ son los números primos que dividen a algún tamaño de clase de G, y el conjunto de aristas $E(G)$ está formado por pares $\{p, q\}$ de vértices p y q tales que su producto pq divide a algún tamaño de clase.

Dentro de este contexto, dos preguntas naturales que surgen son las siguientes: dado un grafo primo asociado a los tamaños de clase un grupo G, ¿qué podemos decir sobre la estructura de G? Y ¿qué grafos ocurren como grafos primos asociados a los tamaños de clase de G? Respecto a la primera pregunta veremos que, por ejemplo, un grafo primo $\Delta(G)$ no contiene como vértice a un primo divisor del orden de G si y solo si G tiene un p-subgrupo de Sylow contenido en su centro. También veremos que si dos vértices p y q no son adyacentes en $\Delta(G)$, entonces G tiene p-subgrupos de Sylow abelianos o q-subgrupos de Sylow abelianos, entre otras propiedades. No obstante, el principal resultado que probaremos en este trabajo responde a la segunda de las preguntas planteadas anteriormente. Antes de enunciarlo, recordamos que un *conjunto independiente* de un grafo $\mathcal{G}$ es un subconjunto π del conjunto de vértices tal que ningún par de elementos de π son adyacentes en $\mathcal{G}$.

Teorema A. *Si G es un grupo resoluble y p, q y r son tres números primos distintos tales que $\{p, q, r\} \subseteq V(G)$, entonces al menos dos de ellos forman una arista en $\Delta(G)$. En particular, $\Delta(G)$ no contiene conjuntos independientes de tres vértices.*

Este célebre teorema fue probado inicialmente por S. Dolfi [3, Theorem 16]. Sin embargo, algunos años más tarde el mismo autor solucionó algunos errores que había en su prueba [4]. Además, Dolfi [4] también proporcionó una demostración del mismo resultado eliminando la hipótesis de la resolubilidad del grupo, aunque haciendo uso de la clasificación de los grupos finitos simples. Señalamos también que la prueba del teorema A que presentamos en este artículo utiliza algunos razonamientos alternativos respecto de la original dada por Dolfi [3] y corregida posteriormente [4].

El teorema A tiene importantes consecuencias. La primera de ellas es que, en cierto sentido, el grafo $\Delta(G)$ tiende a poseer muchas aristas. Otras dos consecuencias inmediatas que se desprenden de dicho resultado vienen dadas en el corolario que aparece a continuación. Recordamos primeramente algunas definiciones que aparecen de la teoría de grafos. Una *componente conexa* de un grafo $\mathcal{G}$ es un subgrafo $\mathcal{H}$ donde todo par de vértices están conectados por algún camino y no hay aristas entre $\mathcal{H}$ y el resto del grafo $\mathcal{G}$. El *diámetro* de un grafo es la máxima distancia entre dos de sus vértices. Un grafo es *completo* cuando todo par de vértices están unidos por una arista.

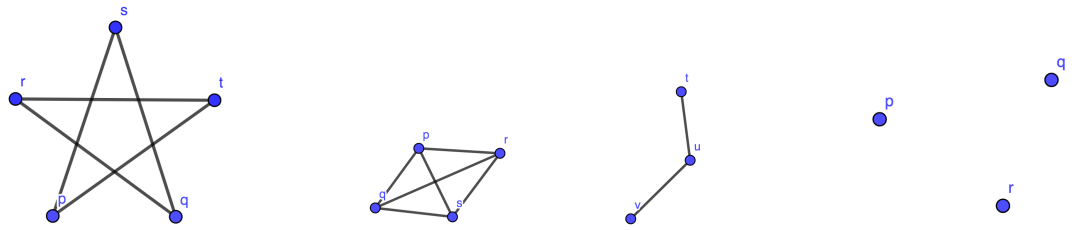
Corolario B. *Sea G un grupo resoluble. Entonces, el número de componentes conexas de $\Delta(G)$ es a lo sumo 2. Además,*

- *si $\Delta(G)$ es conexo, entonces su diámetro es a lo sumo 3;*
- *si $\Delta(G)$ es disconexo, entonces tiene dos componentes conexas que son grafos completos.*

Hemos mencionado anteriormente que el teorema A es cierto sin la hipótesis de la resolubilidad de G debido al trabajo de Dolfi [4], luego el corolario B se cumple realmente también para cualquier grupo G. Sin embargo, este hecho ya era conocido [3, Remark 8, Theorem 17] incluso antes de publicarse el trabajo de Dolfi [4].

Recordemos que, dado un grafo $\mathcal{G}$, el *grafo complementario* $\overline{\mathcal{G}}$ es el grafo que tiene el mismo conjunto de vértices que $\mathcal{G}$ y dos vértices forman una arista en $\overline{\mathcal{G}}$ si y solo si no forman una arista en $\mathcal{G}$. En otras palabras, es el mismo grafo pero con las aristas intercambiadas.

Así pues, el principal teorema de Dolfi [4] puede reescribirse como sigue: para cualquier grupo G, el grafo $\overline{\Delta}(G)$ no contiene ciclos de longitud 3. Recientemente, en el trabajo conjunto entre Dolfi, Pacifici, Sanus y Sotomayor [5], se ha probado que dicho teorema admite una versión mucho más general: el grafo $\overline{\Delta}(G)$ no contiene ciclos de longitud impar. Este tipo de grafos se llaman también grafos *bipartitos*, es decir, grafos cuyos conjuntos de vértices pueden dividirse en dos conjuntos independientes disjuntos de tal forma que toda arista une a un elemento de cada conjunto. Observemos que estos subconjuntos de vértices (llamados *cliques*) inducen subgrafos completos en $\Delta(G)$, luego dicho teorema nos dice que el grafo $\Delta(G)$ contiene dos cliques disjuntos cuya unión es $V(G)$. Por tanto, el segundo caso del corolario B se corresponde con la situación de dos cliques no conexos entre sí, mientras que en el primer caso existe alguna arista entre ambos cliques. Como consecuencia directa de los razonamientos anteriores, podemos afirmar que los grafos de la figura 1 no son posibles grafos $\Delta(G)$ para ningún grupo G.



(a) El grafo complementario es un ciclo de longitud 5.

(b) Consta de dos componentes conexas, pero una no es completa.

(c) Es un conjunto independiente de tamaño 3.

Figura 1: Grafos que no ocurren como $\Delta(G)$ para ningún grupo G.

Finalmente, destacamos que en el artículo recopilatorio de Lewis [8] puede encontrarse una excelente colección de resultados, aparte de los aquí presentes, sobre diversos grafos asociados a diferentes conjuntos de números naturales relativos a un grupo. Notemos que dicho trabajo fue publicado en 2008, por lo que muchos resultados que han sido probados en esta última década no aparecen en él.

La estructura del artículo es la siguiente: en las secciones 2 y 3 exponemos algunos resultados previos genéricos de la teoría de grupos y específicos de tamaños de clase, respectivamente. Dichos resultados serán necesarios para demostrar el teorema A en la sección 4, mientras que la última sección la dedicamos a la prueba del corolario B.

Notación. Utilizaremos notación y terminología estándar del contexto de la teoría de grupos finitos, las cuales siguen principalmente los libros de Doerk y Hawkes [2] y Isaacs [6]. No obstante, recordamos a continuación los principales conceptos que aparecerán frecuentemente. Para un grupo G, sea x^G la *clase de conjugación* de un elemento $x \in G$, es decir, $x^G = \{x^g \mid g \in G\} = \{g^{-1}xg \mid g \in G\}$. El tamaño de este conjunto lo denotaremos por $|x^G|$. El conjunto de primos divisores de un número natural n lo denotaremos por $\pi(n)$. En particular, $\pi(G)$ es el conjunto de primos divisores del orden de G. Para un número primo p, escribiremos $\mathrm{Syl}_p(G)$ para denotar al conjunto de p-subgrupos de Sylow de G. Usaremos $\mathbf{Z}(G)$ para referirnos al centro de un grupo G. Si H y K son subconjuntos de G, entonces el centralizador y normalizador en K de H los denotaremos por $\mathbf{C}_K(H)$ y $\mathbf{N}_K(H)$, respectivamente. Escribiremos $\langle H_1, H_2, \ldots, H_n\rangle$ para referirnos al menor grupo que contenga a los subgrupos $H_1, H_2, \ldots, H_n$. Si N es un subgrupo normal de G, escribiremos $N \trianglelefteq G$. Finalmente, usaremos la notación $\leqslant$ y $<$ para la inclusión e inclusión propia de subgrupos, mientras que $\subseteq$ y $\subset$ serán inclusión e inclusión propia de subconjuntos, respectivamente. ◂

2. Preliminares de grupos

Comenzamos esta sección recordando de manera breve algunos conceptos y propiedades básicas de la teoría de grupos.

Dado un grupo G, la célebre *identidad de Dedekind* afirma que, si U, V y W son subconjuntos de G tales que $V \subseteq U$, entonces $U \cap VW = V(U \cap W)$.

Diremos que un subgrupo H es *característico* en G, y escribiremos H car G, si $\phi(H) = H$ para cualquier automorfismo ϕ de G. Es fácil ver que todo subgrupo característico es normal ya que la conjugación es un caso particular de automorfismo de un grupo. Además, si H_1 car $H_2 \trianglelefteq G$, entonces $H_1 \trianglelefteq G$.

Dados dos elementos $a, b \in G$, definimos el *conmutador* de a y b como $[a,b] := a^{-1}b^{-1}ab$. Si A y B son dos subgrupos de G, entonces $[A,B] := \langle [a,b] \mid a \in A, b \in B\rangle \leqslant G$. En particular, $G' := [G,G]$ es el *subgrupo derivado* de G, el cual es característico en G. Notemos que $a \in \mathbf{C}_G(b)$ es equivalente a $[a,b] = 1$, y $G' = 1$ si y solo si G es abeliano. Análogamente, se define $G^{(2)} := [G', G']$ y, en general, $G^{(i)} := [G^{(i-1)}, G^{(i-1)}]$. Es fácil ver que un grupo es resoluble si y solo si existe un número natural n tal que $G^{(n)} = 1$. Un par de sencillas propiedades de los conmutadores que serán útiles son las siguientes.

Lema 1 ([2, A - Lemma 7.4 (a-c)]). *Sean A y B dos subgrupos de un grupo G y N un subgrupo normal de G. Entonces,*

1. $[A,B] = [B,A] \trianglelefteq \langle A,B\rangle$;
2. $[A,B]N/N = [AN/N, BN/N]$.

Sea P un p-subgrupo de Sylow de G para un primo p. Definimos el *subgrupo p-radical* $\mathbf{O}_p(G)$ de G como la intersección de todos los conjugados de P en G. Claramente, $\mathbf{O}_p(G)$ es un p-subgrupo normal de G. De hecho, es el mayor p-subgrupo normal de G, ya que cualquier otro p-subgrupo normal de G estará contenido en todos los conjugados de P por el segundo y el tercer teorema de Sylow. De manera análoga, definimos el subgrupo $\mathbf{O}_{p'}(G)$ como el mayor subgrupo normal de G cuyo orden es no divisible por p.

Un grupo es *nilpotente* si es el producto directo de sus subgrupos de Sylow. El *subgrupo de Fitting* $\mathbf{F}(G)$ de G es el producto de los subgrupos p-radicales de G para cada primo p. Como todos ellos son normales en G y tienen intersección trivial, en particular forman un producto directo, luego $\mathbf{F}(G)$ es un subgrupo nilpotente y normal en G. Sea N un subgrupo nilpotente y normal en G. Observemos que $\mathbf{O}_p(N)$ es el único p-subgrupo de Sylow de N por el segundo teorema de Sylow, luego $\mathbf{O}_p(N)$ car $N \trianglelefteq G$. Se sigue que $\mathbf{O}_p(N) \leqslant \mathbf{O}_p(G)$ para cada primo p, luego $N \leqslant \mathbf{F}(G)$ y por tanto $\mathbf{F}(G)$ es el mayor subgrupo normal y nilpotente de G.

La intersección de todos los subgrupos maximales de G es el *subgrupo de Frattini* $\mathbf{\Phi}(G)$ de G, el cual es un subgrupo característico de G. Las principales propiedades del subgrupo $\mathbf{\Phi}(G)$ son las siguientes.

Lema 2. *Si G es un grupo, entonces se cumplen las siguientes propiedades:*

1. *Si $H \leqslant G$ tal que $G = H\mathbf{\Phi}(G)$, entonces $G = H$.*
2. (Argumento de Frattini) *Si $N \trianglelefteq G$ y $P \in \mathrm{Syl}_p(N)$, entonces $G = N\mathbf{N}_G(P)$.*
3. *$\mathbf{\Phi}(G)$ es un subgrupo nilpotente. En particular, $\mathbf{\Phi}(G) \leqslant \mathbf{F}(G)$.*
4. *Si $N \trianglelefteq G$ y $U \leqslant G$ tal que $N \leqslant \mathbf{\Phi}(U)$, entonces $N \leqslant \mathbf{\Phi}(G)$.*
5. *Si $N \trianglelefteq G$, entonces $\mathbf{\Phi}(N) \leqslant \mathbf{\Phi}(G)$.*
6. *Si G es p-grupo y $\mathbf{\Phi}(G) = 1$, entonces G es abeliano.*

Demostración. 1. Supongamos por reducción al absurdo que $H < G$, luego existe un subgrupo maximal M de G tal que $H \leqslant M$. Pero $\mathbf{\Phi}(G) \leqslant M$ por definición, luego $H\mathbf{\Phi}(G) \leqslant M < G$, lo cual contradice la hipótesis $G = H\mathbf{\Phi}(G)$.

2. Sea $x \in G$. Entonces $P^x \leqslant N^x = N$ porque N es normal en G. Como $|P^x| = |P|$, entonces vuelve a ser un p-subgrupo de Sylow de N y, por el segundo teorema de Sylow, existe $y \in N$ tal que $P^x = P^y$. Así, $xy^{-1} \in \mathbf{N}_G(P)$. Como $x = (xy^{-1})y$ y x era arbitrario, concluimos que $G = \mathbf{N}_G(P)N$.

3. Si $P \in \mathrm{Syl}_p(\mathbf{\Phi}(G))$, como $\mathbf{\Phi}(G)$ es normal en G, entonces por el argumento de Frattini tenemos que $G = \mathbf{\Phi}(G)\mathbf{N}_G(P)$. Por el punto 1 deducimos que $G = \mathbf{N}_G(P)$, luego P es normal en G y, en particular, en $\mathbf{\Phi}(G)$, para cada primo p. La última afirmación se debe a que $\mathbf{F}(G)$ es el mayor subgrupo normal y nilpotente de G.

4. Supongamos que $N \not\leqslant \mathbf{\Phi}(G)$, por lo que existe un subgrupo maximal M de G tal que $N \not\leqslant M$. Entonces, $M < MN \leqslant G$ y, por maximalidad de M, deducimos que $G = MN$. Por tanto, usando la hipótesis $N \leqslant \mathbf{\Phi}(U)$ y la identidad de Dedekind, tenemos que $U = U \cap G = U \cap MN = N(U \cap M)$. Consecuentemente,

$U = N(U \cap M) \leqslant \mathbf{\Phi}(U)(U \cap M)$, luego $U = U \cap M$ por la propiedad del punto 1. Así, $N \leqslant U \leqslant M$, lo cual es una contradicción.

5. Como $\mathbf{\Phi}(N)$ car $N \trianglelefteq G$, entonces $\mathbf{\Phi}(N) \trianglelefteq G$. Aplicando el punto 4 con N y $\mathbf{\Phi}(N)$, obtenemos la tesis.

6. Sea M un subgrupo maximal arbitrario de G. Entonces, M es normal en G y $|G : M| = p$. Se sigue que G/M es abeliano, luego usando el lema 1 (2) deducimos $[G/M, G/M] = [G,G]M/M = 1$. Por tanto, $G' \leqslant M$, para todo M maximal de G. Se sigue que $G' \leqslant \mathbf{\Phi}(G) = 1$, luego G es abeliano. ∎

La estructura de los normales minimales resolubles de un grupo G será utilizada frecuentemente en la prueba del principal teorema del trabajo. En este sentido, el sencillo resultado que viene a continuación es fundamental. Es una de las razones por las cuales imponemos la hipótesis de resolubilidad del grupo en el teorema A.

Lema 3. *Sea N un subgrupo normal minimal resoluble de un grupo G. Entonces, N es abeliano y $|N|$ es una potencia de primo.*

Demostración. Como N es resoluble, entonces existe un número natural n tal que $N^{(n)} = 1$. En particular, el subgrupo derivado $N' < N$. Pero N' car $N \trianglelefteq G$, luego $N' \trianglelefteq G$. Como N es normal minimal de G, deducimos que $N' = 1$ y N es abeliano. En particular, N es nilpotente. Si p y q son dos primos divisores de $|N|$, entonces $\mathbf{O}_p(N)$ y $\mathbf{O}_q(N)$ son subgrupos de Sylow de N únicos. Luego son característicos en $N \trianglelefteq G$. Se sigue que ambos son normales en G y, por la minimalidad de N, deducimos que solamente uno de ellos puede ser igual a N. Por tanto, $|N|$ es una potencia de primo. ∎

El subgrupo de G generado por sus normales minimales es el *subgrupo socle* $\mathrm{Soc}(G)$. Como dos normales minimales distintos tienen intersección trivial, es fácil ver que $\mathrm{Soc}(G)$ es el producto directo de un subconjunto de normales minimales de G. En particular, si G es resoluble, el subgrupo $\mathrm{Soc}(G)$ es abeliano por el lema 3. De hecho, se cumplen las siguientes importantes igualdades, que serán de mucha utilidad en la demostración del teorema A. La prueba utiliza técnicas elementales pero no es inmediata, luego será referida.

Lema 4 ([2, A - Theorem 10.6 (c) (ii)]). *Sea G un grupo resoluble. Entonces,*

$$\mathbf{F}(G/\mathbf{\Phi}(G)) = \mathbf{F}(G)/\mathbf{\Phi}(G) = \mathbf{C}_{G/\mathbf{\Phi}(G)}(\mathbf{F}(G)/\mathbf{\Phi}(G)) = \mathrm{Soc}(\mathbf{F}(G)/\mathbf{\Phi}(G)).$$

Una situación significativa que aparecerá en numerosas ocasiones en el resto del trabajo es la siguiente. Diremos que un subgrupo A de un grupo G es *complementado en G* si existe un subgrupo H tal que $G = AH$ con $A \cap H = 1$. En particular, si un p-subgrupo de Sylow de G es complementado por un subgrupo normal de G, diremos que G es *p-nilpotente*.

A continuación, presentamos algunos resultados preliminares que serán necesarios para el resto del trabajo. Para agilizar la lectura, algunas de las pruebas serán referidas. El siguiente resultado se le atribuye a Gaschütz y nos da una condición suficiente para que un subgrupo normal abeliano sea complementado. Aunque es un resultado clásico en la teoría de grupos, incluimos su corta demostración.

Lema 5. *Si A es un subgrupo abeliano normal de un grupo G tal que $A \cap \mathbf{\Phi}(G) = 1$, entonces A es complementado en G.*

Demostración. Sea H un subgrupo de orden mínimo dentro del conjunto no vacío $\{T \leqslant G : G = TA\}$. Como A es abeliano y normal en G, entonces $H \cap A$ es centralizado por A y normal en H. Por tanto, $H \cap A \trianglelefteq AH = G$. Supongamos que $H \cap A$ no está contenido en $\mathbf{\Phi}(H)$. Entonces, existe un subgrupo maximal M de H tal que $H \cap A \nleqslant M$, luego $M < M(H \cap A) \leqslant H$. Por maximalidad de M en H, tenemos que $H = M(H \cap A)$. Como $H \cap A \leqslant A$, entonces $G = HA = M(H \cap A)A = MA$ con $|M| < |H|$, lo cual contradice la minimalidad de H. Deducimos que $H \cap A \leqslant \mathbf{\Phi}(H)$ y usando el lema 2 (4) obtenemos que $H \cap A \leqslant A \cap \mathbf{\Phi}(G) = 1$. ∎

El próximo resultado también proporciona una condición suficiente para que un subgrupo normal sea complementado. Se trata del célebre teorema de Schur-Zassenhaus.

Teorema 6 ([2, A - Theorem 11.3]). *Sea A un subgrupo normal de un grupo G tal que $|A|$ y $|G/A|$ son coprimos. Entonces, A es complementado en G.*

En particular, el teorema de Schur-Zassenhaus determina una condición suficiente para que un grupo G sea p-nilpotente. La siguiente proposición también va en esta línea y es un resultado clásico de Burnside.

Proposición 7 ([6, Theorem 9.13]). *Sean G un grupo y P un p-subgrupo de Sylow de G. Si $P \leqslant \mathbf{Z}(\mathbf{N}_G(P))$, entonces G es p-nilpotente.*

Como aplicación del teorema 6 también mostramos la siguiente interesante propiedad del subgrupo de Frattini.

Lema 8. *Sea G un grupo. Entonces, $\pi(\mathbf{\Phi}(G)) \subseteq \pi(G/\mathbf{\Phi}(G))$.*

Demostración. Por reducción al absurdo, supongamos que existe $p \in \pi(\mathbf{\Phi}(G)) \setminus \pi(G/\mathbf{\Phi}(G))$. Por el primer teorema de Sylow, existe $P \in \mathrm{Syl}_p(\mathbf{\Phi}(G))$. Como $|G| = |G/\mathbf{\Phi}(G)| \cdot |\mathbf{\Phi}(G)|$, entonces por órdenes también $P \in \mathrm{Syl}_p(G)$. Sabemos que $\mathbf{\Phi}(G)$ es un subgrupo nilpotente por el lema 2 (3), luego P es un subgrupo característico de $\mathbf{\Phi}(G)$, el cual es normal en G. Deducimos que $P = \mathbf{O}_p(G)$. Aplicando el teorema 6 obtenemos que $G = PH$ para cierto subgrupo H de G tal que $P \cap H = 1$. Pero entonces $G = PH \leqslant \mathbf{\Phi}(G)H$. Utilizando el lema 2 (1), deducimos que $G = H$, lo cual es imposible. ∎

Lema 9. *Sea G un grupo.*

1. *Si H es un subgrupo de G tal que $G = \bigcup_{g \in G} H^g$, entonces $G = H$.*
2. *Si $G = A \cup B$ con A y B subgrupos de G, entonces $G = A$ o $G = B$.*

Demostración. 1. Se sigue de unos sencillos argumentos de conteo [6, Corollary 4.16].

2. Supongamos que $A < G$ y $B < G$. Entonces, $|G : A| \geq 2$ y $|G : B| \geq 2$, de donde deducimos que $|A| \leq |G|/2$ y $|B| \leq |G|/2$. Como posiblemente coincidirán en algunos elementos, además del elemento identidad, obtenemos que $|G| \leq |A| + |B| \leq (|G|/2 - 1) + (|G|/2 - 1) + 1 = |G| - 1$, lo cual es una contradicción. ∎

Cerramos esta sección con un resultado clásico de acción coprima.

Lema 10 ([2, A - Proposition 12.5]). *Supongamos que P y Q son subgrupos de un grupo G de tal forma que Q actúa por conjugación sobre P. Si $|P|$ y $|Q|$ son coprimos, entonces $P = [P, Q]\,\mathbf{C}_P(Q)$.*

3. Preliminares de tamaños de clases de conjugación

Comenzamos esta sección presentando unas propiedades elementales. Recordemos que $|x^G| = |G : \mathbf{C}_G(x)|$ para cualquier elemento $x \in G$ por el teorema de la órbita-estabilizador.

Lema 11. *Sea G un grupo. Entonces, se cumplen las siguientes propiedades:*

1. *Si $N \trianglelefteq G$, entonces $|x^N|$ divide a $|x^G|$ para todo $x \in N$.*
2. *Si $N \trianglelefteq G$, entonces $|(xN)^{G/N}|$ divide a $|x^G|$ para todo $x \in G$.*
3. *Si N y M son subgrupos normales de G tales que $M \cap N = 1$, entonces $\pi(|x^G|) \cup \pi(|y^G|) \subseteq \pi(|(xy)^G|)$ para todo $x \in N$ e $y \in M$.*
4. *Si A y B son dos subgrupos de G y $g \in G$, entonces $\mathbf{C}_A(B)^g = \mathbf{C}_{A^g}(B^g)$.*
5. *Si un primo p no divide a $|x^G|$ y $P \in \mathrm{Syl}_p(G)$, entonces $x \in \mathbf{C}_G(P^g)$ para algún $g \in G$.*

Demostración. Los dos primeros apartados se corresponden con el lema 5 del artículo de Ortiz Sotomayor [9]. Por tanto, nos centramos en la prueba de los apartados 3, 4 y 5.

3. Sea $g \in \mathbf{C}_G(xy)$, con $x \in N$ e $y \in M$. Entonces, $xy = (xy)^g = x^g y^g$, luego $x^{-1}x^g = y(y^g)^{-1}$. Como M y N son normales, entonces $x^{-1}x^g = y(y^g)^{-1} \in M \cap N = 1$, luego $x = x^g$ e $y = y^g$. Deducimos que $g \in \mathbf{C}_G(x) \cap \mathbf{C}_G(y)$ y, por tanto, $\mathbf{C}_G(xy) = \mathbf{C}_G(x) \cap \mathbf{C}_G(y)$, ya que la otra inclusión es evidente. Como consecuencia, obtenemos que $\mathbf{C}_G(xy) \leqslant \mathbf{C}_G(x) \leqslant G$, luego por transitividad $|x^G| = |G : \mathbf{C}_G(x)|$ divide a $|G : \mathbf{C}_G(xy)| = |(xy)^G|$ y, análogamente, $|y^G|$ divide a $|(xy)^G|$. La inclusión de conjuntos del enunciado queda probada.

4. Sea $x \in \mathbf{C}_A(B)^g$, luego $x = c^g$ para cierto $c \in \mathbf{C}_A(B)$. Como $c \in A$, entonces $x = c^g \in A^g$. Además, para cualquier $b \in B$, se cumple que $[x, b^g] = [c^g, b^g] = [c, b]^g = 1$ ya que $c \in \mathbf{C}_A(B)$. Por tanto, $x \in \mathbf{C}_{A^g}(B^g)$. Recíprocamente, si $x \in \mathbf{C}_{A^g}(B^g)$, entonces $x \in A^g$ y $[x, b^g] = 1$ para todo $b \in B$. Así, $x = a^g$ para cierto $a \in A$ y $1 = [a^g, b^g] = [a, b]^g$. Por tanto, $[a, b] = 1$ y $a \in \mathbf{C}_A(B)$, luego $x = a^g \in \mathbf{C}_A(B)^g$.

5. Observemos que $|G| = |x^G| \cdot |\mathbf{C}_G(x)|$. Por tanto, como p no divide a $|x^G|$, necesariamente la mayor potencia de p que divide a $|G|$ es exactamente la misma que la que divide a $|\mathbf{C}_G(x)|$. Deducimos por el primer teorema de Sylow que $\mathbf{C}_G(x)$ tiene un p-subgrupo de Sylow P_0 y, por órdenes, obtenemos que $P_0 \in \mathrm{Syl}_p(G)$. Además, por el segundo teorema de Sylow, $P_0 = P^g$ para algún $g \in G$, luego $P^g \leqslant \mathbf{C}_G(x)$ y $x \in \mathbf{C}_G(P^g)$. ∎

Como hemos comentado en la introducción, una de las primeras situaciones a estudiar es cuando $\Delta(G)$ no posee cierta arista. El siguiente resultado proporciona información estructural de G en este contexto. Cabe destacar que fue demostrado por Itô en 1953 [7], mucho antes de conocerse la noción de grafo primo asociado a los tamaños de clase.

Teorema 12. *Sean G un grupo y $p, q \in V(G)$. Si $\{p, q\} \notin E(G)$, entonces (salvo intercambio de p y q) G es p-nilpotente con p-subgrupos de Sylow abelianos.*

Demostración. Tomamos $P \in \mathrm{Syl}_p(G)$, $Q \in \mathrm{Syl}_q(G)$ y $g \in G$. Si $|g^G|$ es no divisible por p, entonces por el lema 11 (4) y (5) tenemos que $g \in \mathbf{C}_G(P^h) = \mathbf{C}_G(P)^h$ para cierto $h \in G$. Si p divide a $|g^G|$, entonces q no divide a dicho tamaño de clase, ya que $\{p, q\} \notin E(G)$, luego análogamente $g \in \mathbf{C}_G(Q^k) = \mathbf{C}_G(Q)^k$ para cierto $k \in G$. Por tanto,

$$G = \bigcup_{h \in G} \mathbf{C}_G(P)^h \cup \bigcup_{k \in G} \mathbf{C}_G(Q)^k.$$

Observemos que el número de conjugados en G de un subgrupo $H \leqslant G$ coincide con $|G : \mathbf{N}_G(H)|$. Además, el elemento trivial está en todos los términos de la unión anterior, luego por un argumento de conteo obtenemos que

$$|G| \leq (|\mathbf{C}_G(P)| - 1)\,|G : \mathbf{N}_G(\mathbf{C}_G(P))| + (|\mathbf{C}_G(Q)| - 1)\,|G : \mathbf{N}_G(\mathbf{C}_G(Q))| + 1.$$

Dividiendo por $|G|$ a ambos lados, tenemos que

$$1 \leq \frac{|\mathbf{C}_G(P)| - 1}{|\mathbf{N}_G(\mathbf{C}_G(P))|} + \frac{|\mathbf{C}_G(Q)| - 1}{|\mathbf{N}_G(\mathbf{C}_G(Q))|} + \frac{1}{|G|}.$$

Denotemos $n_1 = |\mathbf{N}_G(\mathbf{C}_G(P))|$ y $n_2 = |\mathbf{N}_G(\mathbf{C}_G(Q))|$. Si $\mathbf{C}_G(P) < \mathbf{N}_G(\mathbf{C}_G(P))$ y $\mathbf{C}_G(Q) < \mathbf{N}_G(\mathbf{C}_G(Q))$, entonces $1 \leq \frac{1}{2} - \frac{1}{n_1} + \frac{1}{2} - \frac{1}{n_2} + \frac{1}{|G|}$, luego $\frac{|G|}{n_1} + \frac{|G|}{n_2} \leq 1$, lo cual es claramente una contradicción pues ambas fracciones son índices de subgrupos. Consecuentemente, podemos suponer, por ejemplo, que $\mathbf{C}_G(P) = \mathbf{N}_G(\mathbf{C}_G(P))$. Así, $P \leqslant \mathbf{N}_G(P) \leqslant \mathbf{N}_G(\mathbf{C}_G(P)) = \mathbf{C}_G(P)$, donde la segunda inclusión se da debido al lema 11 (4). Por tanto, P es abeliano y G es p-nilpotente por la proposición 7. ∎

Ejemplo 13. En la figura 1 (c) vimos que $\Delta(G)$ no puede constar solamente de tres vértices aislados, para ningún grupo G. Sin embargo, sí que puede constar solamente de dos vértices aislados. Sea G un grupo simétrico de grado 3. Es fácil comprobar que los tamaños de clase de G son $\{1, 2, 3\}$, luego $\Delta(G)$ está formado por los vértices 2 y 3, los cuales no son adyacentes. Además, esto es un ejemplo de un grupo que satisface las hipótesis del teorema 12. Más concretamente, G es 2-nilpotente. ◂

Otra situación clave a analizar es cuando un primo p divisor del orden de G no es vértice de $\Delta(G)$, en otras palabras, cuando todos los tamaños de clase de G son no divisibles por p. El siguiente lema caracteriza esta propiedad.

Lema 14. *Sean G un grupo, $p \in \pi(G)$ y $P \in \mathrm{Syl}_p(G)$. Entonces, $p \notin V(G)$ si y solo si $P \leqslant \mathbf{Z}(G)$.*

Demostración. Si $P \leqslant \mathbf{Z}(G)$, entonces claramente $P \leqslant \mathbf{C}_G(x) \leqslant G$ para todo $x \in G$, luego $|x^G| = |G : \mathbf{C}_G(x)|$ divide a $|G : P|$, de donde deducimos que p no divide a ningún tamaño de clase de G. Recíprocamente, si $p \notin V(G)$, entonces usando el lema 11 (4) y (5) tenemos que cada $x \in \mathbf{C}_G(P^g) = \mathbf{C}_G(P)^g$ con $g \in G$. Así, $G = \bigcup_{g \in G} \mathbf{C}_G(P)^g$ y usando el lema 9 (1) obtenemos que $G = \mathbf{C}_G(P)$, luego $P \leqslant \mathbf{Z}(G)$. ∎

El resultado posterior nos da una condición suficiente para que $\Delta(G)$ posea un clique.

Lema 15. *Sea G un grupo tal que $G/\mathbf{F}(G)$ es abeliano. Entonces, existe $g \in G$ tal que $\pi(|g^G|)$ contiene a todos los primos que dividen a $|G/\mathbf{F}(G)|$.*

Demostración. Observemos que G es claramente resoluble. Razonamos por inducción sobre $|G|$. Supongamos en primer lugar que $\mathbf{\Phi}(G) \neq 1$. Observemos que $\mathbf{F}(G/\mathbf{\Phi}(G)) = \mathbf{F}(G)/\mathbf{\Phi}(G)$ por el lema 4. Por tanto, el grupo cociente $\overline{G} \coloneqq G/\mathbf{\Phi}(G)$ cumple que $\overline{G}/\mathbf{F}(\overline{G}) = (G/\mathbf{\Phi}(G))/(\mathbf{F}(G)/\mathbf{\Phi}(G))$ es isomorfo a $G/\mathbf{F}(G)$ por el tercer teorema de isomorfía, siendo este último grupo abeliano por hipótesis. Aplicando la hipótesis de inducción, tenemos que para cierto elemento $\overline{g} \in \overline{G}$ se cumple que $\pi(\overline{G}/\mathbf{F}(\overline{G})) = \pi(G/\mathbf{F}(G))$ está contenido en $\pi(|\overline{g}^{\overline{G}}|) \subseteq \pi(|g^G|)$, donde la última inclusión es en virtud del lema 11 (2).

Por tanto, podemos suponer que $\mathbf{\Phi}(G) = 1$. Usando el lema 4, podemos afirmar que, para ciertos normales minimales $N_1, \dots, N_r$ de G, se da que $\mathbf{F}(G) = \mathrm{Soc}(G) = N_1 \times \cdots \times N_r$. Además, al ser G resoluble, entonces también son resolubles todos los N_i anteriores, luego son todos abelianos por el lema 3. Luego $\mathbf{F}(G)$ es abeliano y, por el lema 5, tenemos que $G = \mathbf{F}(G)H$ para cierto subgrupo H con $\mathbf{F}(G) \cap H = 1$. Veamos que para cada $i \in \{1, \dots, r\}$, existe $1 \neq x_i \in N_i$ tal que $\mathbf{C}_H(x_i) = \mathbf{C}_H(N_i)$. Sea $1 \neq x_i \in N_i$. Claramente, $\mathbf{C}_H(N_i) \leqslant \mathbf{C}_H(x_i)$. Por otra parte, sea $h \in \mathbf{C}_H(x_i)$. Entonces, $1 \neq x_i \in \mathbf{C}_{N_i}(h)$ y este subgrupo es centralizado por $\mathbf{F}(G)$, ya que $\mathbf{F}(G)$ es abeliano. Además, $\mathbf{C}_{N_i}(h)$ es normalizado por H, debido al lema 11 (4) y a que H es abeliano, ya que $G/\mathbf{F}(G) = H\mathbf{F}(G)/\mathbf{F}(G)$ es isomorfo por el segundo teorema de isomorfía a $H/(H \cap \mathbf{F}(G)) = H$. Así, $\mathbf{C}_{N_i}(h)$ es normal en $\mathbf{F}(G)H = G$ y $1 \neq \mathbf{C}_{N_i}(h) \leqslant N_i$. Por minimalidad de N_i, deducimos que $N_i = \mathbf{C}_{N_i}(h)$, por lo que $h \in \mathbf{C}_H(N_i)$ y, por tanto, $\mathbf{C}_H(x_i) \leqslant \mathbf{C}_H(N_i)$, luego queda probada la igualdad de ambos subgrupos.

Finalmente, considerando los elementos $1 \neq x_i$ anteriores, tomamos $g \coloneqq (x_1, \dots, x_r) \in N_1 \times \cdots \times N_r = \mathbf{F}(G)$. Por ser $\mathbf{F}(G)$ abeliano tenemos, usando la identidad de Dedekind, que

$$\mathbf{C}_G(g) = \mathbf{C}_G(g) \cap G = \mathbf{C}_G(g) \cap \mathbf{F}(G)H = \mathbf{F}(G)(\mathbf{C}_G(g) \cap H) = \mathbf{F}(G)\,\mathbf{C}_H(g).$$

Además,

$$\mathbf{C}_H(g) \leqslant \mathbf{C}_H(x_1) \cap \cdots \cap \mathbf{C}_H(x_r) = \mathbf{C}_H(N_1) \cap \cdots \cap \mathbf{C}_H(N_r) \leqslant \mathbf{C}_H(\mathbf{F}(G)) \leqslant H \cap \mathbf{C}_G(\mathbf{F}(G)) = H \cap \mathbf{F}(G) = 1,$$

donde la última igualdad es por el lema 4. Deducimos que $|g^G| = |G : \mathbf{C}_G(g)| = |G : \mathbf{F}(G)|$. ∎

Cuando razonemos en la demostración del teorema A con cocientes del grupo sobre normales minimales nos será de mucha utilidad el siguiente resultado.

Lema 16. *Sea M un subgrupo normal minimal de un grupo G y sea p un primo tal que $p \in V(G) \setminus V(G/M)$. Si M es abeliano y $\mathbf{Z}(G) = \mathbf{\Phi}(G) = 1$, entonces o bien $M \in \mathrm{Syl}_p(G)$, o bien $|M|$ es no divisible por p y $\mathbf{C}_M(P) = 1$ con $P \in \mathrm{Syl}_p(G)$. En este último caso, $p \in \pi(|x^G|)$ para todo $1 \neq x \in M$.*

Demostración. Supongamos que $M \notin \mathrm{Syl}_p(G)$. Como $\mathbf{\Phi}(G) = 1$, por el lema 5 tenemos que $G = MH$ para cierto subgrupo H de G tal que $M \cap H = 1$. Observemos que $G/M = HM/M \cong H/(H \cap M) = H$. Si $p \notin \pi(G/M)$, como $|G| = |M| \cdot |H|$ y $|M|$ es una potencia de primo por el lema 3, entonces necesariamente $M \in \mathrm{Syl}_p(G)$, lo cual no puede suceder. Luego $p \in \pi(G/M)$. Así, por hipótesis, $p \in \pi(G/M) \setminus V(G/M)$ y por el lema 14 deducimos que $H \cong G/M$ tiene un p-subgrupo de Sylow central, digamos P_1.

Supongamos que M es un p-grupo. Sea $P \coloneqq P_1M$. Tenemos que $M \cap P_1 \leqslant M \cap H = 1$, y por órdenes deducimos que $P \in \mathrm{Syl}_p(G)$. Además, $M \leqslant \mathbf{N}_G(P)$ ya que $M \leqslant P$ y, claramente, $H \leqslant \mathbf{N}_G(P)$ pues $P_1 \leqslant \mathbf{Z}(H)$ y M es normal en G, luego $P \trianglelefteq MH = G$. En virtud del lema 2 (5) obtenemos que $\mathbf{\Phi}(P) \leqslant \mathbf{\Phi}(G) = 1$, luego P es abeliano por el lema 2 (6). Se sigue que P_1 conmuta con M, luego $P_1 \leqslant \mathbf{C}_G(H) \cap \mathbf{C}_G(M) \leqslant \mathbf{Z}(G) = 1$, por lo que $P = M$, lo cual es una contradicción.

Por tanto, M tiene orden no divisible por p. En este caso tenemos que $P_1 \in \mathrm{Syl}_p(G)$. Observemos que $\mathbf{C}_M(P_1)$ es un subgrupo normal de $G = MH$, ya que M es abeliano y P_1 es central en H (lema 11 (4)). Como M es normal minimal de G, entonces o bien $M = \mathbf{C}_M(P_1)$, o $\mathbf{C}_M(P_1) = 1$. El primer caso implica que $P_1 \leqslant \mathbf{Z}(G) = 1$ y, como $P_1 \in \mathrm{Syl}_p(G)$, esto contradice nuestra hipótesis $p \in V(G)$. Por tanto, $\mathbf{C}_M(P_1) = 1$.

Sea $1 \neq x \in M$. Si $p \notin \pi(|x^G|)$, entonces por el lema 11 (5) tenemos que $x \in \mathbf{C}_G(P_1^g) \cap M = \mathbf{C}_M(P_1^g)$ para cierto $g \in G$. Aplicando el lema 11 (4) obtenemos que $x^{g^{-1}} \in \mathbf{C}_M(P_1^g)^{g^{-1}} = \mathbf{C}_M(P_1) = 1$, luego $x = 1$, la contradicción final. ∎

Finalizamos la sección con la proposición 18, la cual es fundamental para demostrar el teorema A. Se trata de una unificación de los dos apartados del siguiente resultado. La demostración de este requiere de técnicas más avanzadas, por lo que será referida.

Proposición 17. *Sean G un grupo y $p, q \in V(G)$ tales que $\{p, q\} \notin E(G)$. Sean $P \in \mathrm{Syl}_p(G)$ y $Q \in \mathrm{Syl}_q(G)$. Sea M un subgrupo normal de G.*

1. *Si $1 \neq |M| = t^n$ con t primo, $\mathbf{C}_M(P) = 1$ y M es abeliano y complementado en G, entonces $\mathbf{O}_q(G) = Q \cap \mathbf{C}_G(M) \leqslant \mathbf{Z}(\mathbf{C}_G(M))$.*
2. *Si $M = \mathbf{O}_{p'}([G, P])$ es normal minimal de G y Q no es normal en G, entonces*
   a) *$\overline{G} := G/\mathbf{C}_G(M)$ cumple que $\overline{G}/\mathbf{F}(\overline{G})$ y $\mathbf{F}(\overline{G})$ son cíclicos, $\overline{P} \leqslant \mathbf{F}(\overline{G})$ y $\overline{Q} \cap \mathbf{F}(\overline{G}) = 1$, y*
   b) *$\overline{G} = \overline{Q} \cdot \overline{T}$, donde $\overline{T} := \langle \overline{x} \in \overline{G} \mid q \in \pi(|\overline{x}^{\overline{G}}|) \rangle$.*

Demostración. El apartado 1 se sigue del trabajo de Casolo *et al.* [1, Proposition 3.1], mientras que el segundo apartado es exactamente lo que muestran Dolfi *et al.* [5, Proposition 2.5]. ∎

Proposición 18. *Sean G un grupo y $p, q \in V(G)$ tales que $\{p, q\} \notin E(G)$. Sean $P \in \mathrm{Syl}_p(G)$ y $Q \in \mathrm{Syl}_q(G)$. Supongamos que M es un subgrupo normal minimal abeliano de G tal que $p \notin V(G/M)$, $|M|$ es no divisible por p y M es complementado en G. Entonces, $\mathbf{O}_q(G) = Q \cap \mathbf{C}_G(M) \leqslant \mathbf{Z}(\mathbf{C}_G(M))$.*

Además, si Q no es normal en G, entonces

1. *$\overline{G} := G/\mathbf{C}_G(M)$ cumple que $\overline{G}/\mathbf{F}(\overline{G})$ y $\mathbf{F}(\overline{G})$ son cíclicos, $\overline{P} \leqslant \mathbf{F}(\overline{G})$ y $\overline{Q} \cap \mathbf{F}(\overline{G}) = 1$, y*
2. *$\overline{G} = \overline{Q} \cdot \overline{T}$, donde $\overline{T} := \langle \overline{x} \in \overline{G} \mid q \in \pi(|\overline{x}^{\overline{G}}|) \rangle$.*

Demostración. En primer lugar, observemos que M es complementado por hipótesis, luego existe $H \leqslant G$ tal que $G = HM$ con $H \cap M = 1$. Así, H es isomorfo a G/M por el segundo teorema de isomorfía. Notemos que $p \in \pi(G/M)$, ya que $|G| = |H| \cdot |M|$ y M tiene orden no divisible por p. Como estamos suponiendo que $p \notin V(G/M)$, entonces H tiene un p-subgrupo de Sylow P_0 central por el lema 14. Como M tiene orden coprimo con p, entonces, en particular, $P_0 \in \mathrm{Syl}_p(G)$. Pero P es conjugado en G a P_0 por el segundo teorema de Sylow, luego podríamos trabajar desde el principio con P_0 en lugar de P, ya que el enunciado no depende de la elección del p-subgrupo de Sylow. Por tanto, salvo conjugación, podemos suponer que el p-subgrupo de Sylow central de H es P. Deducimos que $\mathbf{C}_M(P)$ es centralizado por M (por ser M abeliano) y normalizado por H, ya que $P \leqslant \mathbf{Z}(H)$ y podemos aplicar el lema 11 (4). Como M es normal minimal de G, entonces o bien $\mathbf{C}_M(P) = M$, o $\mathbf{C}_M(P) = 1$. En el primer caso obtenemos que $P \leqslant \mathbf{Z}(G)$ ya que $G = HM$, pero esto y el lema 14 nos llevan a la contradicción $p \notin V(G)$, ya que $P \in \mathrm{Syl}_p(G)$. Por tanto, $\mathbf{C}_M(P) = 1$. En virtud del lema 3, $|M|$ es una potencia de primo, luego podemos aplicar la proposición 17 (1) y obtenemos que $\mathbf{O}_q(G) = Q \cap \mathbf{C}_G(M) \leqslant \mathbf{Z}(\mathbf{C}_G(M))$.

Por otra parte, como $p \notin V(G/M)$, de nuevo por el lema 14 tenemos que $PM/M \leqslant \mathbf{Z}(G/M)$, luego $1 = [G/M, PM/M] = [G, P]M/M$ por el lema 1 (2) y $[G, P] \leqslant M$. Pero $[G, P] \trianglelefteq \langle G, P \rangle = G$ por el lema 1 (1) y M es normal minimal de G. Como p es vértice de $\Delta(G)$, entonces, en virtud del lema 14, obtenemos que $1 \neq [G, P]$ y deducimos que $M = [G, P]$, luego trivialmente $M = \mathbf{O}_{p'}([G, P])$ ya que p no divide a $|M|$. Luego si Q no es normal en G, entonces estamos en disposición de aplicar la proposición 17 (2), lo que nos proporciona las afirmaciones 1 y 2. ∎

4. Prueba del teorema A

Demostración del teorema A. Sean p, q, r tres vértices cualesquiera del grafo $\Delta(G)$, con p, q y r números primos distintos. Nuestro objetivo es demostrar que existe alguna arista entre ellos. Observemos que cualquier subgrupo o grupo cociente de G es claramente resoluble. Trabajaremos por contraejemplo minimal, esto es, supongamos que $\{p, q, r\}$ es un conjunto independiente de $\Delta(G)$ pero todo grupo G_0 tal que $|G_0| < |G|$ y $\{p, q, r\} \subseteq V(G_0)$ cumple que $\{p, q, r\}$ no es un conjunto independiente de $\Delta(G_0)$. Sean $P \in \mathrm{Syl}_p(G)$, $Q \in \mathrm{Syl}_q(G)$ y $R \in \mathrm{Syl}_r(G)$. Dividimos la demostración en una serie de pasos.

Paso 1. Si $1 \neq N$ es un normal de G tal que $\{p,q,r\} \subseteq \pi(G/N)$, entonces $\{p,q,r\} \cap V(G/N) \subsetneq \{p,q,r\}$.

Por reducción al absurdo, supongamos que se da la igualdad entre ambos conjuntos, esto es, que $\{p,q,r\}$ son vértices de $\Delta(G/N)$. Por minimalidad de G, tenemos que dos de ellos son adyacentes en G/N, es decir, existiría $xN \in G/N$ tal que, por ejemplo, $pq \in \pi(|(xN)^{G/N}|) \subseteq \pi(|x^G|)$, donde la última inclusión se debe al lema 11 (2), lo cual contradice nuestras hipótesis. Por tanto, como $\{p,q,r\} \cap V(G/N)$ es siempre un subconjunto de $\{p,q,r\}$ y no son iguales, la inclusión es estricta.

Paso 2. Podemos suponer $\mathbf{\Phi}(G) = 1$. En particular, todo subgrupo normal en G y abeliano es complementado en G.

Supongamos que $\mathbf{\Phi}(G) \neq 1$. Observemos que $\pi(\mathbf{\Phi}(G)) \subseteq \pi(G/\mathbf{\Phi}(G))$ debido al lema 8 y, como además sabemos que $|G| = |G/\mathbf{\Phi}(G)| \cdot |\mathbf{\Phi}(G)|$, entonces $\pi(G) = \pi(G/\mathbf{\Phi}(G)) \cup \pi(\mathbf{\Phi}(G)) = \pi(G/\mathbf{\Phi}(G))$. Por tanto, $\{p,q,r\} \subseteq \pi(G/\mathbf{\Phi}(G))$. Por el paso 1, podemos suponer sin pérdida de generalidad que $p \notin V(G/\mathbf{\Phi}(G))$. Así, $P\mathbf{\Phi}(G)/\mathbf{\Phi}(G) \leqslant \mathbf{Z}(G/\mathbf{\Phi}(G))$ por el lema 14, luego $P\mathbf{\Phi}(G)$ es normal en G y claramente $P \in \mathrm{Syl}_p(P\mathbf{\Phi}(G))$. Aplicando el lema 2 (1) y (2) deducimos que $G = P\mathbf{\Phi}(G)\mathbf{N}_G(P) = P\mathbf{N}_G(P) = \mathbf{N}_G(P)$, luego P es normal en G. Por el teorema 6 tenemos que $G = PH$ para cierto subgrupo H de G tal que $P \cap H = 1$. Consecuentemente, H actúa por conjugación sobre P y $|H| = |G : P|$ es coprimo con $|P|$. Por el lema 10 se sigue que $P = [P,H]\mathbf{C}_P(H)$, luego $G = [P,H]\mathbf{C}_P(H)H$. Como $P\mathbf{\Phi}(G)/\mathbf{\Phi}(G) \leqslant \mathbf{Z}(G/\mathbf{\Phi}(G))$, obtenemos en particular que $1 = [P\mathbf{\Phi}(G)/\mathbf{\Phi}(G), H\mathbf{\Phi}(G)/\mathbf{\Phi}(G)] = [P,H]\mathbf{\Phi}(G)/\mathbf{\Phi}(G)$, donde la última igualdad se debe al lema 1 (2). Así, $G = [P,H]\mathbf{C}_P(H)H \leqslant \mathbf{\Phi}(G)\mathbf{C}_P(H)H = H\mathbf{C}_P(H)$ por el lema 2 (1) y, por tanto, $|G| = |H| \cdot |\mathbf{C}_P(H)|$. Por órdenes, necesariamente $P = \mathbf{C}_P(H)$, por lo que $H \trianglelefteq G$. Además, $G = P \times H$ pues $P \cap H = 1$.

Como $p \in V(G)$ por hipótesis, necesariamente $\mathbf{Z}(P) < P$ por el lema 14. Escogemos $x \in P \setminus \mathbf{Z}(P)$, el cual claramente verifica que $1 \neq |x^G| = |x^P| = |P : \mathbf{C}_P(x)|$, por lo que $p \in \pi(|x^G|)$. Por otro lado, por órdenes, cualquier q-subgrupo de Sylow Q_0 de H es un q-subgrupo de Sylow de G. Deducimos por el lema 14 que $q \in \pi(|y^H|)$ para algún $y \in H$, pues si $Q_0 \leqslant \mathbf{Z}(H)$ entonces $Q \leqslant \mathbf{Z}(G)$ y obtendríamos la contradicción $q \notin V(G)$ de nuevo por el lema 14. Aplicando el lema 11 (1) obtenemos que $q \in \pi(|y^G|)$, luego $pq \in \pi(|(xy)^G|)$ por el lema 11 (3). Esta contradicción implica que necesariamente $\mathbf{\Phi}(G) = 1$.

Finalmente, como $\mathbf{\Phi}(G) = 1$, aplicando el lema 5 con cualquier subgrupo normal abeliano de G obtenemos la última afirmación.

Paso 3. Podemos suponer que $\mathbf{Z}(G) = 1$.

Supongamos que $\mathbf{Z}(G) \neq 1$ y consideremos el grupo cociente $G/\mathbf{Z}(G)$. Claramente, $\{p,q,r\} \subseteq \pi(G/\mathbf{Z}(G))$, ya que de otra forma por órdenes tendríamos que $\mathbf{Z}(G)$ contendría un t-subgrupo de Sylow de G con $t \in \{p,q,r\}$, lo cual es una contradicción por el lema 14. Nuevamente por el paso 1, podemos suponer que $V(G/\mathbf{Z}(G))$ no contiene a alguno de los primos $\{p,q,r\}$. Sin pérdida de generalidad, supongamos que $p \notin V(G/\mathbf{Z}(G))$.

Observemos que $\mathbf{Z}(G)$ es complementado en G por el paso 2, esto es, $G = \mathbf{Z}(G)L$ para algún $L \leqslant G$ con $\mathbf{Z}(G) \cap L = 1$. Pero entonces L es claramente normal en $\mathbf{Z}(G)L = G$, luego $G = \mathbf{Z}(G) \times L$. Como L es isomorfo a $G/\mathbf{Z}(G)$ por el segundo teorema de isomorfía y $p \notin V(G/\mathbf{Z}(G))$, entonces por el lema 14 tenemos que L tiene un p-subgrupo de Sylow central, digamos L_p. Sea Z_p el p-subgrupo de Sylow de $\mathbf{Z}(G)$. Entonces, claramente $Z_p \times L_p \in \mathrm{Syl}_p(G)$ es central en G, contradiciendo nuestras hipótesis.

Paso 4. A lo sumo uno de los subgrupos contenidos en $\{P,Q,R\}$ es normal en G.

Por reducción al absurdo, supongamos que P y Q son ambos normales en G. Como $\{p,q\} \notin E(G)$, entonces pq no divide a ningún tamaño de clase de G. Así, si $g \in G$ cumple que $p \in \pi(|g^G|)$, entonces q no divide a dicho tamaño de clase, luego por el lema 11 (5) obtenemos que $g \in \mathbf{C}_G(Q)$ pues Q es normal en G. Si p no divide a $|g^G|$, entonces, análogamente, $g \in \mathbf{C}_G(P)$. Deducimos que $G = \mathbf{C}_G(P) \cup \mathbf{C}_G(Q)$. Notemos que ambos subgrupos son propios en G; de otra forma, tendríamos que p o q no son vértices de $\Delta(G)$ por el lema 14. Pero G no puede ser unión de dos subgrupos propios por el lema 9 (2), lo cual es una contradicción.

Paso 5. G tiene a lo sumo dos normales minimales distintos.

Supongamos que N_1, N_2, N_3 son tres normales minimales distintos de G. Al ser G resoluble, entonces por el lema 3 cada uno de ellos es un s_i-grupo abeliano para ciertos primos s_i, con $1 \leq i \leq 3$.

Si $t \in \{p,q,r\} \setminus \pi(G/N_i)$ para algún i, entonces por órdenes necesariamente N_i es un t-subgrupo de Sylow de G. Usando el paso 4, deducimos que existe como máximo un $i \in \{1,2,3\}$ tal que $\pi(G/N_i)$ no contiene a algún primo de $\{p,q,r\}$. Por tanto, sin pérdida de generalidad, podemos suponer que tanto $\pi(G/N_1)$ como $\pi(G/N_2)$ contienen a $\{p,q,r\}$ y, por el paso 1, podemos asumir que $V(G/N_1)$ y $V(G/N_2)$ no contienen a alguno de dichos primos.

Supongamos en primer lugar que $V(G/N_1)$ y $V(G/N_2)$ no contienen al mismo primo, digamos q. Entonces, por el lema 14 tenemos que $G/N_1 \times G/N_2$ tiene un q-subgrupo de Sylow central. Consideremos ahora el homomorfismo $\phi\colon G \to G/N_1 \times G/N_2$ con $\phi(g) = (gN_1, gN_2)$ para todo $g \in G$, el cual cumple que $\ker(\phi) = N_1 \cap N_2$ y, por el primer teorema de isomorfía, $G/\ker(\phi) \cong \mathrm{Im}(\phi) \leqslant G/N_1 \times G/N_2$. Por tanto, $G/(N_1 \cap N_2)$ es isomorfo a un subgrupo de dicho producto directo. Como claramente $N_1 \cap N_2 = 1$, entonces $G = G/(N_1 \cap N_2)$ también tiene un q-subgrupo de Sylow central, es decir, $q \notin V(G)$ por el lema 14, lo cual es una contradicción.

Podemos por tanto suponer que $p \notin V(G/N_1)$ y $q \notin V(G/N_2)$. Distinguimos ahora varios casos posibles. Si N_1 es un p-grupo y N_2 es un q-grupo, entonces por el lema 16 deducimos que $P = N_1$ y $Q = N_2$, lo cual contradice el paso 4. Si $|N_1|$ es coprimo con p y $|N_2|$ es coprimo con q, entonces por el lema 16 obtenemos que $p \in \pi(|x^G|)$ y $q \in \pi(|y^G|)$ para todo $1 \neq x \in N_1$ y $1 \neq y \in N_2$. Utilizando el lema 11 (3) obtenemos la contradicción $pq \in \pi(|(xy)^G|)$. Finalmente, supongamos que $N_1 = P$ y N_2 tiene orden no divisible por q. Entonces, por órdenes y por el paso 4 tenemos que necesariamente $\{p,q,r\} \subseteq \pi(G/N_3)$. Asimismo, $V(G/N_3)$ pierde a alguno de dichos primos por el paso 1, digamos $t \in \{p,q,r\}$. Además, t no divide a $|N_3|$ por el lema 16 y el paso 4. Si $t \neq q$, entonces por el lema 16 nuevamente podemos encontrar elementos en N_2 y N_3 tales que el tamaño de clase en G del producto de ambos es divisible por tq, contradicción. Luego podemos afirmar que $t = q$ y, por tanto, $G/N_2 \times G/N_3$ tiene un q-subgrupo de Sylow central. Repitiendo el argumento del párrafo anterior, deducimos que G también tiene un q-subgrupo de Sylow central, lo cual no puede ocurrir por el lema 14 y las hipótesis del teorema.

Paso 6. $\mathbf{F}(G)$ no es normal minimal de G.

Supongamos que $\mathbf{F}(G)$ es un normal minimal de G, luego es un t-grupo abeliano para algún primo t por el lema 3. Observemos que $\mathbf{F}(G)$ es complementado en G por el paso 2, es decir, existe $H \leqslant G$ tal que $G = H\mathbf{F}(G)$ con $\mathbf{F}(G) \cap H = 1$. Por tanto, $\mathbf{C}_H(\mathbf{F}(G)) = \mathbf{C}_G(\mathbf{F}(G)) \cap H = \mathbf{F}(G) \cap H = 1$, donde la penúltima igualdad se debe al lema 4.

Supongamos que alguno de los primos $\{p,q,r\}$ no está contenido en $\pi(G/\mathbf{F}(G))$. Entonces, por órdenes deducimos que $\mathbf{F}(G)$ es un t-subgrupo de Sylow de G, con $t \in \{p,q,r\}$. Podemos asumir sin pérdida de generalidad que $t = p$, luego $\mathbf{F}(G) = P$. Si $1 \neq x \in H$ cumple que $p \notin \pi(|x^G|)$, como $P \trianglelefteq G$, entonces por el lema 11 (5) obtenemos que $x \in H \cap \mathbf{C}_G(P) = \mathbf{C}_H(P) = 1$, lo cual no puede ocurrir. Luego $p \in \pi(|x^G|)$ para todo $1 \neq x \in H$. Por hipótesis, deducimos que q y r no dividen a $|x^G|$ para todo $x \in H$. Además, H es isomorfo a $G/\mathbf{F}(G)$ por el segundo teorema de isomorfía, luego existe una biyección entre $|x^H|$ y $|(x\,\mathbf{F}(G))^{G/\mathbf{F}(G)}|$ para todo $x \in H$. Como este último tamaño de clase divide a $|x^G|$ por el lema 11 (2), deducimos que $\{q,r\} \cap \pi(|x^H|) = \varnothing$ para todo $x \in H$. Así, $H \cong G/\mathbf{F}(G)$ tiene un q-subgrupo de Sylow central y un r-subgrupo de Sylow central por el lema 14. Deducimos que $\{q,r\} \cap V(G/\mathbf{F}(G)) = \varnothing$. Aplicando el lema 16 llegamos a que $\{q,r\} \subseteq \pi(|y^G|)$ para todo $1 \neq y \in \mathbf{F}(G)$, lo cual es una contradicción.

Por tanto, $\{p,q,r\} \subseteq \pi(G/\mathbf{F}(G))$ y, por el paso 1, podemos suponer que, por ejemplo, p no pertenece a $V(G/\mathbf{F}(G))$. En virtud del lema 16 tenemos además que $t \neq p$. Aplicando la proposición 18 con las aristas $\{p,q\}$ y $\{p,r\}$ obtenemos que $\mathbf{O}_q(G) = Q \cap \mathbf{C}_G(\mathbf{F}(G))$ y $\mathbf{O}_r(G) = R \cap \mathbf{C}_G(\mathbf{F}(G))$. Pero $\mathbf{C}_G(\mathbf{F}(G)) = \mathbf{F}(G)$ debido al lema 4 y es un t-grupo normal minimal, luego podemos asumir, por ejemplo, que $\mathbf{O}_q(G) = 1$ y, por tanto, Q no es normal en G. Si denotamos $\overline{G} \coloneqq G/\mathbf{C}_G(\mathbf{F}(G))$, entonces la proposición 18 (1) nos dice que $\overline{G}/\mathbf{F}(\overline{G})$ es cíclico con $\overline{Q} \cap \mathbf{F}(\overline{G}) = \mathbf{O}_q(\overline{G}) = 1$. Pero $|\overline{G}| = |\overline{G}/\mathbf{F}(\overline{G})| \cdot |\mathbf{F}(\overline{G})|$ y, por órdenes, deducimos que $q \in \pi(\overline{G}/\mathbf{F}(\overline{G}))$. Además, $\overline{G} = G/\mathbf{C}_G(\mathbf{F}(G)) = H\,\mathbf{C}_G(\mathbf{F}(G))/\mathbf{C}_G(\mathbf{F}(G))$ es isomorfo a $H/\mathbf{C}_H(\mathbf{F}(G)) = H$ por el segundo teorema de isomorfía, luego $H/\mathbf{F}(H)$ es cíclico con $q \in \pi(H/\mathbf{F}(H))$.

Supongamos que $\mathbf{O}_r(G) = 1$ también. Entonces, R tampoco sería normal en G y, nuevamente, la proposición 18 (1) nos proporcionaría que $r \in \pi(H/\mathbf{F}(H))$. Aplicando el lema 15 tenemos que qr divide a $|g^H|$ para algún $g \in H$ y, como H es isomorfo a $G/\mathbf{F}(G)$, por el lema 11 (2) obtenemos que $qr \in \pi(|(g\,\mathbf{F}(G))^{G/\mathbf{F}(G)}|) \subseteq \pi(|g^G|)$. Por tanto, $\{q,r\} \in E(G)$, una contradicción con nuestras hipótesis.

Consecuentemente, $1 \neq \mathbf{O}_r(G)$ y, como $\mathbf{F}(G)$ es normal minimal de G, necesariamente tenemos que $\mathbf{F}(G) = \mathbf{O}_r(G)$. Aplicando el segundo apartado de la proposición 18 con la arista $\{p,q\}$ y recordando que $\mathbf{C}_G(\mathbf{F}(G)) = \mathbf{F}(G)$, obtenemos que $G/\mathbf{F}(G) = \overline{G} = \overline{Q} \cdot \overline{T}$, donde $\overline{T} = \langle \overline{x} \in \overline{G} \mid q \in \pi(|\overline{x}^{\overline{G}}|)\rangle$. Observemos que $\overline{T} \neq 1$ claramente, pues de otra forma $\overline{G} = \overline{Q} = \mathbf{F}(\overline{G})$ y, por otro lado, teníamos por la proposición 18 (1) que $\overline{Q} \cap \mathbf{F}(\overline{G}) = 1$, lo cual es una contradicción. Escogemos un generador $1 \neq \overline{x} = x\,\mathbf{F}(G)$ de $\overline{T}$, el cual verifica que $q \in \pi(|(x\,\mathbf{F}(G))^{G/\mathbf{F}(G)}|)$, y por el lema 11 (2) deducimos que $q \in \pi(|x^G|)$. Entonces, por hipótesis $r \notin \pi(|x^G|)$. Por el lema 11 (5) y la definición de $\mathbf{O}_r(G)$, obtenemos que $x \in \mathbf{C}_G(R^g) \leqslant \mathbf{C}_G(\mathbf{O}_r(G)) \leqslant \mathbf{C}_G(\mathbf{F}(G)) = \mathbf{F}(G)$, lo cual es claramente una contradicción.

Paso 7. $\mathbf{F}(G) = P \times N$ con P y N normales minimales abelianos de G, $q \notin V(G/N)$ y q coprimo con $|N|$.

Por los pasos 5 y 6 junto con el lema 4 obtenemos que $\mathbf{F}(G) = \mathrm{Soc}(G) = N_1 \times N_2$ con N_1 y N_2 normales minimales distintos de G. Además, por el lema 3, ambos son abelianos y, por tanto, también lo es $\mathbf{F}(G)$. En virtud del paso 4, necesariamente $\{p,q,r\}$ está contenido en $\pi(G/N_1)$ o $\pi(G/N_2)$.

Supongamos primeramente que $\{p,q,r\} \subseteq \pi(G/N_2)$ pero, por ejemplo, $p \notin \pi(G/N_1)$. Entonces, por órdenes, $N_1 = P$, y por el paso 1 tenemos que $\{p,q,r\}$ no está contenido en $V(G/N_2)$. Recordemos que, por el primer teorema de isomorfía, podemos ver $G = G/(N_1 \cap N_2)$ como un subgrupo de $G/N_1 \times G/N_2$. Si p es el primo no contenido en $V(G/N_2)$, entonces por el lema 14 tenemos que $G/N_1 \times G/N_2$ tiene un p-subgrupo de Sylow central, y por el mismo lema llegaríamos a la contradicción $p \notin V(G)$. Por tanto, podemos suponer, por ejemplo, que $q \notin V(G/N_2)$. Además, si N_2 fuese un q-grupo, entonces por el lema 16 tendríamos que $N_2 = Q$, lo cual contradice el paso 4. Por tanto, $|N_2|$ es coprimo con q y tenemos la estructura de $\mathbf{F}(G)$ deseada.

Supongamos ahora que $\{p,q,r\}$ está contenido en $\pi(G/N_1)$ y $\pi(G/N_2)$. Nuevamente, el paso 1 nos lleva a que $\{p,q,r\}$ no está contenido en $V(G/N_1)$ ni en $V(G/N_2)$. Distinguimos ahora una serie de casos.

Si $V(G/N_1)$ y $V(G/N_2)$ no contienen al mismo primo, digamos p, entonces razonando como en el segundo párrafo deducimos que G (el cual es isomorfo a un subgrupo de $G/N_1 \times G/N_2$) tiene un p-subgrupo de Sylow central, lo cual es imposible.

Por tanto, podemos suponer que $p \notin V(G/N_1)$ y $q \notin V(G/N_2)$. Claramente no puede darse que N_1 sea un p-grupo y N_2 sea un q-grupo, por el lema 16 y el paso 4. Si $|N_1|$ es coprimo con p y $|N_2|$ es coprimo con q, el lema 16 nos dice que todo elemento de N_1 tiene tamaño de clase divisible por p y todo elemento de N_2 tiene tamaño de clase divisible por q. Usando el lema 11 (3) podemos encontrar un elemento cuyo tamaño de clase en G es divisible por pq, lo cual contradice nuestras hipótesis. Por tanto, por el lema 16, podemos asumir que $N_1 = P \in \mathrm{Syl}_p(G)$ y $q \notin V(G/N_2)$ con q coprimo con $|N_2|$, lo cual es la estructura de $\mathbf{F}(G)$ que buscábamos.

Paso 8. Contradicción final.

Por el paso 7 tenemos que $\mathbf{F}(G) = P \times N$ con P y N normales minimales abelianos de G, $q \notin V(G/N)$ y q coprimo con $|N|$. Notemos que R no es normal en G por el paso 4. Aplicando la proposición 18 (2) con G y N y la arista $\{q,r\}$, obtenemos que $G/\mathbf{C}_G(N) = (R\,\mathbf{C}_G(N)/\mathbf{C}_G(N)) \cdot (T/\mathbf{C}_G(N))$, donde $T/\mathbf{C}_G(N)$ es el grupo generado por los elementos $x\,\mathbf{C}_G(N)$ de $G/\mathbf{C}_G(N)$ (con $x \in G$) tales que $r \in \pi(|(x\,\mathbf{C}_G(N))^{G/\mathbf{C}_G(N)}|)$. Por el lema 11 (2) tenemos entonces que $r \in \pi(|x^G|)$. Aplicando las hipótesis del teorema llegamos a que $p \notin \pi(|x^G|)$ y, por el lema 11 (5), obtenemos que $x \in \mathbf{C}_G(P)$ ya que P es normal en G. Por tanto, $T/\mathbf{C}_G(N) \leqslant \mathbf{C}_G(P)\,\mathbf{C}_G(N)/\mathbf{C}_G(N)$ y, entonces, $G/\mathbf{C}_G(N) = (R\,\mathbf{C}_G(N)/\mathbf{C}_G(N)) \cdot (\mathbf{C}_G(P)\,\mathbf{C}_G(N)/\mathbf{C}_G(N))$. Así, tenemos que $G = R\,\mathbf{C}_G(P)\,\mathbf{C}_G(N)$. Usando la proposición 18 con G y N y la arista $\{p,q\}$ obtenemos que $P = P \cap \mathbf{C}_G(N) \leqslant \mathbf{Z}(\mathbf{C}_G(N))$, luego todo elemento de $\mathbf{C}_G(N)$ conmuta con P, esto es, $\mathbf{C}_G(N) \leqslant \mathbf{C}_G(P)$ y, como $G = R\,\mathbf{C}_G(P)\,\mathbf{C}_G(N)$, deducimos que $G = R\,\mathbf{C}_G(P)$. En consecuencia, $\mathbf{C}_P(R)$ centraliza a R y es centralizado por $\mathbf{C}_G(P)$, luego $\mathbf{C}_P(R) \leqslant \mathbf{Z}(G) = 1$.

Podemos aplicar la proposición 17 (1) con G y P y la arista $\{q,r\}$, ya que P es complementado en G por el paso 2, con lo que obtenemos que $\mathbf{O}_q(G) = Q \cap \mathbf{C}_G(P)$. Como $G = R\,\mathbf{C}_G(P)$, por órdenes, cualquier q-subgrupo de Sylow Q_0 de $\mathbf{C}_G(P)$ lo será también de G. Por el segundo teorema de Sylow, $Q_0 = Q^u$ para algún $u \in G$, luego $Q^u \leqslant \mathbf{C}_G(P)$ y, por el lema 11 (4), obtenemos que $Q \leqslant \mathbf{C}_G(P^{u^{-1}}) = \mathbf{C}_G(P)$ pues $P \trianglelefteq G$. Por tanto, $\mathbf{O}_q(G) = Q \cap \mathbf{C}_G(P) = Q$, y esta última contradicción con el paso 4 finaliza la prueba. ∎

Observemos que la hipótesis de la resolubilidad del grupo no ha sido necesaria hasta el paso 5, luego los argumentos de los pasos 1 a 4 son válidos para grupos no resolubles también.

5. Prueba del corolario B

Demostración del corolario B. El número de componentes conexas de $\Delta(G)$ es claramente menor o igual que dos, pues en otro caso podríamos encontrar tres vértices, cada uno en una componente conexa distinta, que formarían un conjunto independiente, lo cual contradice el teorema A.

Supongamos ahora, por contradicción, que el diámetro es de longitud al menos 4. Esto significa que existiría un camino de la forma p-q-r-s-t para ciertos primos, de tal forma que no existe otro camino más corto que una p con t. Entonces, $\{p, r, t\}$ sería un conjunto independiente de tamaño 3 y tendríamos una contradicción de nuevo con el teorema A.

Finalmente, supongamos que $\Delta(G)$ es disconexo. Por la primera afirmación del corolario, tenemos que $\Delta(G)$ está formado por dos componentes conexas π_1 y π_2. Supongamos que alguna de ellas no es completa, esto es, que existen p y q en (por ejemplo) π_1 tales que $\{p, q\} \notin E(G)$. Sea $r \in \pi_2$ arbitrario. Entonces, $\{p, q, r\}$ es un conjunto independiente de tamaño 3, lo cual no puede ocurrir. Esto finaliza la demostración. ∎

Referencias

[1] Casolo, Carlo; Dolfi, Silvio; Pacifici, Emanuele, y Sanus, Lucia. «Incomplete vertices in the prime graph on conjugacy class sizes of finite groups». En: *Journal of Algebra* 376 (2013), págs. 46-57. issn: 0021-8693. https://doi.org/10.1016/j.jalgebra.2012.10.012.

[2] Doerk, Klaus y Hawkes, Trevor. *Finite soluble groups*. Vol. 4. De Gruyter Expositions in Mathematics. Berlin: Walter de Gruyter & Co., 1992. https://doi.org/10.1515/9783110870138.

[3] Dolfi, Silvio. «Arithmetical conditions on the length of the conjugacy classes of a finite group». En: *Journal of Algebra* 174.3 (1995), págs. 753-771. issn: 0021-8693. https://doi.org/10.1006/jabr.1995.1151.

[4] Dolfi, Silvio. «On independent sets in the class graph of a finite group». En: *Journal of Algebra* 303.1 (2006), págs. 216-224. issn: 0021-8693. https://doi.org/10.1016/j.jalgebra.2005.09.006.

[5] Dolfi, Silvio; Pacifici, Emanuele; Sanus, Lucia, y Sotomayor, Víctor. «The prime graph on class sizes of a finite group has a bipartite complement». En: *Journal of Algebra* 542 (2020), págs. 35-42. issn: 0021-8693. https://doi.org/10.1016/j.jalgebra.2019.09.022.

[6] Isaacs, Irving Martin. *Algebra: A Graduate Course*. Pacific Grove, California: Brooks/Cole Thompson Learning, 1994. isbn: 978-0-534-19002-6.

[7] Itô, Noboru. «On finite groups with given conjugate types. I». En: *Nagoya Mathematical Journal* 6 (1953), págs. 17-28. issn: 0027-7630. url: http://projecteuclid.org/euclid.nmj/1118799473.

[8] Lewis, Mark L. «An overview of graphs associated with character degrees and conjugacy class sizes in finite groups». En: *The Rocky Mountain Journal of Mathematics* 38.1 (2008), págs. 175-211. issn: 0035-7596. https://doi.org/10.1216/RMJ-2008-38-1-175.

[9] Ortiz Sotomayor, Víctor Manuel. «Influencia de los tamaños de clase en grupos finitos». En: *TEMat* 1 (2017), págs. 45-51. issn: 2530-9633. url: https://temat.es/articulo/2017-p45/.

TEMat, volumen 4. Mayo de 2020.
e-ISSN: 2530-9633

Publicado con la colaboración de la
Real Sociedad Matemática Española