
ÍNDICE

Editorial	2
-----------------	---

EL MATEMÁTICO DEL NÚMERO

Los revolucionarios franceses del análisis matemático	3
---	---

AXIOMAS, TEOREMAS Y ALGO MÁS

Anatomía de una Ecuación	6
El Teorema de Representación de Bruno de Finetti	17

ATERRIZANDO IDEAS

El algoritmo de Metropolis-Hastings	21
Gráficas químicas	31
Árboles y máquinas inteligentes	37

ACTIVA TUS NEURONAS

La contraseña	41
Frutería	41
Camino del monje	41
La cuerda y el tubo	42
Beer pong	42

ZONA OLÍMPICA

Lista de problemas	43
Pregunta de Erdős	44

EN EL HORIZONTE

Teoría de grupos aplicada al modelamiento del cubo de Rubik	45
---	----

TENDIENDO AL INFINITO

Del ITAM a Viena: entrevista con Juan Pablo Aguilera	59
--	----



Consejo Académico

Claudia Gómez Wulschner
Gustavo Preciado Rosas

Consejo Editorial

Director

José Carlos Zamorano Pérez

Tesorero

José Luis Meza Orozco

Secretario

Eduardo Antonio Peña Sandoval

Edición

Gerardo Antonio López Ruiz
Gian Carlo Diluvi
Ilan Jinich Fainsod
José Murguía Fuentes
Mario Enrique Carranza Barragán
Víctor Toledo Cortés

Redes sociales

Roberto Galán

Diseño web

Adrián Tame Jacobo



Se terminó de imprimir en Otoño del 2016, en la imprenta:
MULTIGRÁFICA PUBLICITARIA S.A. de C.V.
Avenida 15, Col. Granjas Esmeralda, México D.F., C.P. 09810

El tiraje fue de 1500 ejemplares.
Todos los derechos reservados. Prohibida la reproducción total o parcial de cualquier artículo o imagen sin la autorización del Consejo Editorial.
Los artículos son responsabilidad del autor y no reflejan necesariamente el punto de vista del Consejo Editorial.
Esta revista es gratuita.

Editorial

Las matemáticas distan mucho de ser una creación humana o una mera invención de un grupo privilegiado de personas que lograron sobrepasar la barrera de lo imposible para brindarle al mundo leyes capaces de desentrañar lo más profundo de nuestra naturaleza. Lo que sucede realmente es que las leyes matemáticas simplemente “son”; independientemente del hombre y su agudo intelecto. A lo más que puede aspirar el ser humano es descubrir la existencia de esas leyes y llegar a dominarlas.

Nuestra labor como jóvenes matemáticos y actuarios se torna desafiante y aparentemente imposible ¿cómo dominar una ciencia tan vasta y compleja? El trabajo duro, nuestra dedicación y pasión por abrir caminos, que aparentemente eran intransitables, buscando lograr un mejor desarrollo de nuestra sociedad; es lo que nos permite dominar ciertos aspectos del extenso universo que son las matemáticas.

Agradecimientos

A la División Académica de Actuaría, Estadística y Matemáticas del ITAM, en especial a Beatriz Rumbos, Claudia Gómez y Gustavo Preciado. A la Dirección Escolar del ITAM, específicamente a Patricia Medina. Gracias a Poisson y Fourier, representaciones de los alumnos de Actuaría y Matemáticas Aplicadas, respectivamente, por el apoyo brindado. Agradecemos también al Fondo de Organizaciones Estudiantiles y al Consejo Universitario de Honor y Excelencia.

$$\sum_{n=1}^{\infty} \sum_{k=10^{n-1}}^{10^n-1} \frac{k}{10^{n(k-10^{n-1}+1)+9} \sum_{\ell=1}^{n-1} 10^{\ell-1} \ell}$$

<http://laberintos.itam.mx>
laberintos@itam.mx

Imagen de portada:

Jana Fainsod Perelman

Los revolucionarios franceses del análisis matemático

Mariana Menchero

Estudiantes de matemáticas aplicadas del ITAM

Gerardo Antonio López

Estudiantes de matemáticas aplicadas del ITAM

“No me imagino una vida más plena que una vida dedicada a la matemática”

Augustin Louis Cauchy

Introducción

Así como los revolucionarios franceses usaron libertad, igualdad y fraternidad como máximas durante la Revolución Francesa de 1789; convergencia, continuidad e integración es el lema de la revolución que los matemáticos franceses Cauchy, Fourier, Lebesgue y Borel, tuvieron en el Análisis Matemático.

Las preguntas que tuvieron sobre los conceptos de convergencia, continuidad e integración ¿cuándo una sucesión es convergente?, ¿qué significa que una función sea continua?, ¿cómo asignarle una medida a un conjunto?, fueron fundamentales para el desarrollo del Análisis. Si bien no fueron los únicos o los primeros en hacernos estas preguntas, nuestras respuestas fueron revolucionarias e influenciaron otras áreas de las matemáticas, por lo que, al considerar todos nuestros trabajos en conjunto, se puede demostrar fácilmente que ellos, los matemáticos franceses, hicieron una revolución en el Análisis Matemático.



Figura 1: Cauchy, Borel, Fourier, Lebesgue

Continuidad: ¿Qué significa que una función sea continua?

Cauchy, en su obra *Course d'Analyse* definió una función continua como una función en la que un incremento muy pequeño en la variable siempre produce un incremento muy pequeño en la función. En este libro, publicado en 1821, Cauchy fue el primero en utilizar los argumentos con ϵ y también en demostrar una versión más general de Teorema del Valor Intermedio que había demostrado Bolzano en 1817.

La definición formal de continuidad es:

Una función $f : R^n \rightarrow R$ es continua en el dominio de la función si:
 $\forall \epsilon > 0 \quad \exists \delta > 0 \text{ tal que } \forall x \in R : \text{Si } |x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \epsilon$

Convergencia: ¿Cuándo una sucesión es convergente?

Respecto al problema de convergencia. Cauchy y Fourier hicieron aportaciones importantes. Cauchy demostró que:

Una sucesión $(X_n), n \in \mathbf{N}$ converge $\Leftrightarrow \exists$ una $n \in \mathbf{N}$ tal que $\forall n, m \geq N$.

Por su parte, Fourier demostró que una función periódica y continua se puede representar como una serie de senos y cosenos. Este descubrimiento lo hizo en 1807 mientras intentaba resolver la ecuación del calor, la cual describe la variación de la temperatura en una región dada durante un periodo de tiempo. Su trabajo en esta rama de estudio lo llevó a contribuciones como la transformada de Fourier, que sería fundamental en el desarrollo de la teoría de probabilidad, procesamiento de imágenes y mecánica cuántica. Las series de Fourier dieron origen al Análisis de Fourier, que es el estudio de la manera en la que ciertas funciones pueden representarse o aproximarse por funciones trigonométricas. Estas aportaciones le granjearían la admiración de matemáticos menores como Riemann y Borel (a comparación de Fourier, claro está).

Integración: ¿Cómo asignarle una medida a un conjunto?

En la era moderna, Cauchy fue el primero en dar una definición rigurosa de integral en 1823. Su definición es muy similar a la integral de Riemann, que fue publicada en 1854, sólo que considera particiones uniformes.

El problema con la integral de Riemann es que existen funciones que no son Riemann-integrables, como la función de Dirichlet, la cual está definida en el intervalo $[0,1]$ y vale 0 en los irracionales y 1 en los racionales. Sin embargo, a principios del siglo XX, Lebesgue desarrolló una integral que extiende este concepto a una clase más grande de funciones, que incluye a la función de Dirichlet.

Lebesgue demostró que si una función acotada y definida en un intervalo cerrado es Riemann-integrable, entonces es Lebesgue integrable. Así, la integral de Lebesgue es una extensión de la integral de Riemann (y otra mejora) que además introduce el concepto de medida de un conjunto en su construcción.

El concepto de medida que introdujo Lebesgue en 1901 permite asignarle una medida a clase suficientemente grande de subconjuntos R^n . Lebesgue fue alumno de Borel, quien introdujo el concepto de σ - álgebra de Borel. La σ - álgebra de Borel es la σ - álgebra generada por conjuntos abiertos. A partir de esta σ - álgebra se puede construir la medida de Borel. Asimismo, Borel, que también ayudó a clarificar la definición de Cantor de conjuntos medibles, fue un parteaguas en la formalización de la teoría de la probabilidad con la publicación de *Eléments de la théorie des probabilités* (1909).

Anatomía de una Ecuación

Hugo Harleston Aguirre

Estudiante de Física, Facultad de Ciencias de la UNAM

Introducción.

“*Es un formulazo*”: una frase muy común que puede alegrar a muchos y frustrar a otros. Muchas veces, cuando se tiene que resolver un problema de una tarea, podemos seguir una receta facilísima: primero, hay que sustituir algunos datos conocidos en una expresión sencilla; luego hay que despejar la variable que no se conozca en la misma; y por último calcular su valor con la operación que obtengamos en el despeje. Siguiendo siempre estos pasos, podremos terminar una tarea en muy poco tiempo, ser felices y pasar el resto del día sin preocupaciones.

Pero, ¿de dónde vienen esos “*formulazos*”? ¿Por qué podemos usar tal o cual expresión para poder resolver un problema que, sin dicha fórmula, sería muchísimo más complicado? ¿Qué hay detrás de una ecuación?

Es sencillo. Cada “*formulazo*” que utilizamos en nuestra “vida diaria” (no todos tenemos que utilizar la ley de Ohm, $V=R \cdot i$, como nuestro pan de cada día), tiene detrás suyo a un mundo entero de desarrollo matemático y, en el caso particular de este artículo, un profundo entendimiento de distintas ramas de la Física.

En general, una ecuación, o fórmula, es la representación más sencilla que se ha podido encontrar de algún aspecto de una ciencia en particular. De hecho, muchas personas han dedicado sus vidas a encontrar expresiones de esta naturaleza; basta con ver la película “*The Theory of Everything*” para darnos cuenta de que Stephen Hawking ha querido llevar esto al límite encontrando una ecuación “que describa al universo”, uniendo a las teorías de la Relatividad y la Mecánica Cuántica en una sola.

Un ejemplo con el que todos hemos trabajado en algún momento es la famosa Segunda Ley de Newton. Desde que entramos a la secundaria, los profesores nos enseñan la famosa cantaleta: “*La Fuerza es igual a la masa de un objeto multiplicada por su aceleración*”, o, mejor dicho, “*Efe es igual a eme por a*”.

$$\vec{F} = m * \vec{a}. \quad (1)$$

Entonces, si un día vamos a comer tacos y el amable taquero nos dice: “Si me dices cuál es la Segunda Ley de Newton te regalo dos extra de Pastor”, nosotros contestamos alegremente “Efe es igual a eme por a”, y nos llenamos de alegría creyendo que tendremos tacos gratis en nuestras vidas; pero, para nuestra enorme sorpresa, él nos avienta una tortilla y nos pide que nos retiremos de su fino establecimiento. “Van a ahuyentar a mis clientes”, argumenta.

El señor taquero, obviamente, sabe que esa cantaleta en definitiva no es la Segunda Ley de Newton, sino una expresión, o “*formulazo*”, que nos enseñaron desde que éramos pequeños para poder resolver problemas relacionados con el concepto de Fuerza.

La Segunda Ley de Newton es muchísimo más que eso. Una expresión que se encuentra en los libros, y que es considerada más completa, suele ser: “El cambio con respecto al tiempo del momento lineal de una partícula (o un sistema de partículas) es proporcional al cambio de su velocidad con respecto al tiempo”. [1]

$$\frac{d\vec{p}}{dt} \propto \frac{d\vec{v}}{dt}. \quad (2)$$

Aquí encontramos conceptos como “Fuerza” (capacidad de realizar un trabajo, o movimiento, que también se representa como el cambio del momento lineal con respecto al tiempo¹), y aceleración (cambio de la velocidad con respecto al tiempo). Además, la constante de proporcionalidad en la expresión (2) es nada más y nada menos que la masa de la partícula o el sistema de partículas con el que se trabaje.

Sin embargo, podemos ir todavía más lejos al decir que la velocidad es el cambio de la posición con respecto al tiempo. Por lo tanto, la aceleración es la segunda derivada de la posición con respecto al tiempo:

$$\frac{d\vec{p}}{dt} = m * \frac{d^2\vec{x}}{dt^2}. \quad (3)$$

Hemos transformado la ecuación (1) en una ecuación diferencial de segundo orden, y la expresión tan sencilla que conocimos en secundaria ya no tiene tanto atractivo. Aunque uno creería que ya no hay mucho que hacer, la ecuación (3) es apenas el inicio de un largo camino que, si uno lo sigue hasta el final, llegará a descubrir que comienza nada más y nada menos con la creación de una bestia tremenda: el *Cálculo Diferencial e Integral*, inventado por Isaac Newton. [1][2]

¹Aunque no es como tal la Segunda Ley de Newton, la definición de “Fuerza” sí está dada por la expresión $F=m*a$. [1][2]

La Segunda Ley de Newton es tan solo un ejemplo que demuestra lo profundo que puede llegar a ser el significado de un “*formulazo*”. Pero, para ilustrarlo de una mejor manera, veremos más a fondo otro caso que es menos conocido y que trata sobre un tema que rodea nuestras vidas cotidianas: la electricidad.

Un Ejemplo Particular.

Uno de los objetos de estudio más importantes de la Física es el Electromagnetismo –al grado de que Albert Einstein, estudiándolo, llega a formular su teoría de la Relatividad- y, dentro de éste, lo primero que se estudia es la Electroestática. Ésta se encarga de entender a fondo todas las interacciones que se dan entre cargas eléctricas en reposo. Como en toda rama de la Física, existen expresiones muy particulares que nos permiten trabajar de manera muy sencilla con los problemas que pueden presentarse dentro de la misma. El “*formulazo*” que estudiaremos aquí es uno conocido como la Ecuación de Poisson:

$$\nabla^2 V(\vec{r}) = \frac{-\rho(\vec{r})}{\varepsilon_0}. \quad (4)$$

Aquí, $\nabla^2 V(\vec{r})$ es el Laplaciano del potencial eléctrico; $\rho(\vec{r})$ es la densidad volumétrica de carga eléctrica (la cual es una función de la posición, \vec{r}); y ε_0 es la constante de permitividad en el vacío (esta constante describe cómo un campo eléctrico se ve afectado en el vacío). [3][4]

Dentro de la ecuación (4) podemos encontrar resumidos, en general, todos los conceptos importantes de la Electroestática. Pero para poder entender cuáles son esos, y qué significa cada elemento de la expresión anterior, tenemos que mencionar los conceptos más básicos de la Electroestática. El primer concepto que vale la pena mencionar es uno que también nos enseñan a cantar en secundaria, o tal vez ya en la prepa: la Ley de Coulomb.

$$\vec{F} = \frac{kQ_1Q_2}{r^2} \hat{u}_r. \quad (5)$$

Esta ley nos permite calcular a la fuerza eléctrica, \vec{F} , que ejerce una carga, Q_1 , sobre otra, Q_2 . Como se puede ver en (5), la magnitud de la fuerza dependerá de una constante, k , y será inversamente proporcional al cuadrado de la distancia, r entre las cargas. La dirección de la fuerza será, justamente, en la dirección radial entre las cargas. La constante k está definida como:

$$k = \frac{1}{4\pi\varepsilon_0}. \quad (6)$$

Supongamos que ahora llamamos a Q_2 nuestra carga de prueba. Ésta será una carga que nosotros podemos llevar en nuestra bolsa en todo momento para utilizar cuando queramos, y para quitarle peso y poder manejarla más fácilmente, en vez de Q_2 , la llamaremos q .

Utilizaremos a q para definir un nuevo concepto a partir de la Ley de Coulomb: el campo eléctrico. Éste será la fuerza eléctrica ejercida por unidad de carga. Su dirección será la misma que aquella en la que la fuerza eléctrica actuaría sobre nuestra carga de prueba, q .²

Entonces, una expresión correcta para el campo eléctrico es:

$$\vec{E} = \lim_{q \rightarrow 0} \frac{\vec{F}}{q} = \frac{kQ_1}{r^2} \hat{u}_r. \quad (7)$$

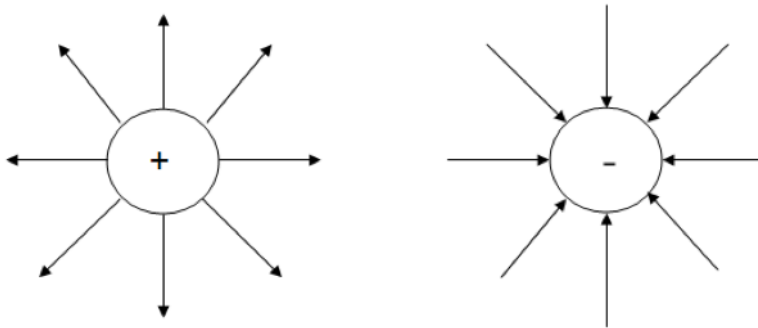


Figura 1: Se muestra lo que se conoce como “líneas de campo”. Éstas nos ayudan a dibujar el campo eléctrico generado por una carga y ver en qué dirección va el mismo. En la figura se pueden ver las líneas de campo que generan una carga puntual positiva (izquierda) y una negativa (derecha).

Ya tenemos definidos dos conceptos básicos de la Electrostática: la fuerza y el campo. Entonces surge una nueva pregunta: si yo tengo una superficie cerrada sobre la que pasa un campo eléctrico, ¿cómo puedo saber “cuánto” campo está pasando sobre la misma? [3][4]

Ese “cuánto”, la cantidad de campo eléctrico que pasa sobre una unidad de área de una superficie cerrada arbitraria, se conoce como el flujo del campo eléctrico.

²Es importante tener en cuenta que las cargas eléctricas tienen signo. La dirección del campo generado por una carga puntual negativa será justamente hacia dicha carga, mientras que la dirección del campo generado por una carga puntual positiva será al revés, hacia afuera de la carga (Figura 1).

Supongamos que ponemos a nuestra carga de prueba, q , en un punto cualquiera del espacio. Cerca de ella colocamos una esfera hueca, de radio R , y queremos conocer cuál es el flujo del campo sobre la misma.

Si dibujamos las líneas del campo que genera q , tendremos algo igual a lo que vemos en la Figura 1, del lado izquierdo. Por la manera en la que tenemos colocada nuestra esfera, las líneas de campo entrarán por un lado de la misma y saldrán por el otro (Figura 2). Por lo tanto, el flujo del campo sobre la superficie será cero, pues todo lo que entra, vuelve a salir.

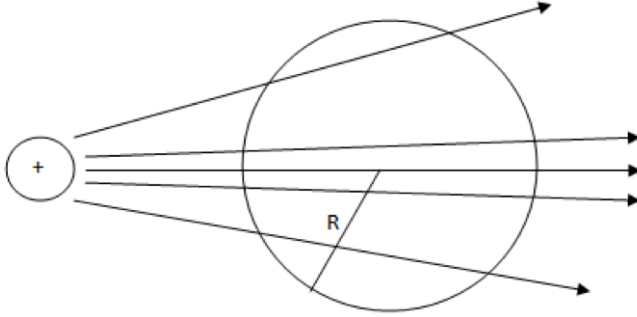


Figura 2: Algunas líneas de campo generadas por una carga puntual positiva atraviesan a una esfera hueca de radio R . Como el campo entra y sale de la esfera, el flujo total sobre la misma es cero. El tamaño de la carga está exagerado, para ilustrar mejor el concepto.

Por lo tanto, para poder calcular el flujo del campo eléctrico sobre una superficie, necesitamos encerrar a la fuente del campo con ella. De esta manera, el campo solamente saldrá de la superficie (si la fuente es positiva), o solamente entrará (si tenemos una fuente negativa). Así, el flujo no se anulará al volver a salir de la superficie, y tendremos un valor neto distinto de cero.

Para poder calcular dicho valor, existe otra de esas expresiones que llamamos “*formulazos*”, pero que en realidad tienen un contexto inmenso: la Ley de Gauss.

$$\phi_E = \varepsilon_0 \oint_S \vec{E} \cdot d\vec{S} = Q_t. \quad (8)$$

La Ley de Gauss dice que el flujo eléctrico, ϕ_E , es igual a la integral sobre la superficie cerrada, S , del campo por el diferencial de área de la superficie, y que todo esto es igual a la carga total encerrada, Q_t , por la superficie que se utilice. [3][4]

La superficie con la que trabajemos depende enteramente de nosotros. Obviamente, dependiendo de la geometría del problema con el que trabajemos, vale la pena escoger alguna

superficie sobre otra. Sin embargo, uno puede escoger la superficie cerrada que desee. Veamos esto con un ejemplo:

Tenemos, en vez de una carga puntual como hasta ahora, una lámina infinita que tiene una densidad superficial de carga positiva y uniforme. Esta lámina genera un campo uniforme, el cual viajará en dirección horizontal, de izquierda a derecha, si colocamos de ese lado, a una distancia D , una lámina igual que la primera, pero con una densidad de carga negativa. Para poder calcular el flujo de dicho campo, ¿qué superficie nos conviene utilizar?

Como el campo viaja solamente en una dirección, y no en todas, una esfera no será lo más conveniente esta vez. Pero, ¿qué tal un cilindro? Este se puede dividir en tres partes: las dos tapas y el cuerpo. Podemos ver la Figura 3 para tener una mejor idea de cómo funciona el problema.

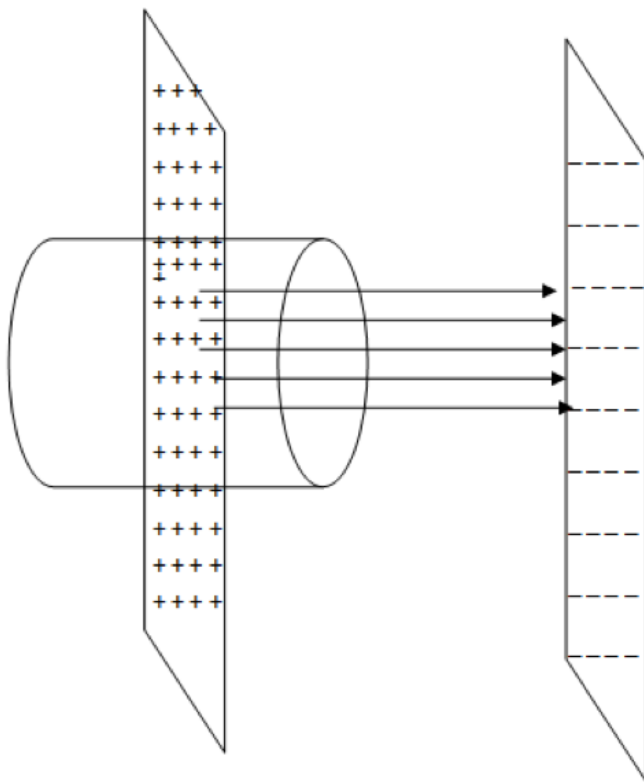


Figura 3: Se muestra una placa infinita con una densidad de carga uniforme y positiva. Se utiliza un cilindro para encerrar un pedazo de la misma, y las líneas de campo se ven solamente del lado derecho porque el campo “viaja” de la placa con carga positiva a la placa con carga negativa, así que antes de ella no habrá nada.

Es importante recordar que la superficie que se decida utilizar siempre deberá encerrar a la fuente del campo. Como en este caso se trabaja con una densidad superficial y uniforme, no se tiene que encerrar toda la placa. [3]

Se puede ver que en la tapa izquierda del cilindro el flujo será cero, pues no hay campo pasando por la misma. Luego, sobre el cuerpo del cilindro el flujo también valdrá cero, pues las líneas de campo son paralelas al mismo, y por lo tanto, no pasaran a través de esa superficie. Así que la única componente que nos queda de nuestra superficie es la tapa derecha del cilindro. Utilizando la ecuación (8), y tomando en cuenta que el campo es constante, llegamos a la conclusión de que el flujo sobre dicha tapa será: $\varepsilon_0 * E * A$, donde A es el área de la tapa del cilindro. Pero, además, esto será igual a la carga total encerrada, la cuál será simplemente el producto del área y la densidad superficial de carga, que denotaremos como σ .

Por lo tanto:

$$E = \frac{\sigma}{\varepsilon_0}. \quad (9)$$

La ecuación (9), arroja otro resultado importante en la Electrostática: una expresión para el campo eléctrico que depende solamente de una densidad de carga y de una constante.

Estamos a punto de dar el paso más importante, pues vamos a utilizar un teorema, que es “importantísimo” en el Cálculo Diferencial e Integral, y que nos permitirá llegar, por fin, a la Ecuación de Poisson. Sin embargo, para poder utilizarlo, tenemos que construir y definir algunos conceptos todavía. Podemos imaginarnos al campo eléctrico como un campo vectorial (las líneas de campo serán justamente los vectores del mismo). Se dice que un campo es conservativo si su rotacional es cero. En otras palabras, un campo será conservativo si la fuerza que éste induce sobre una partícula no depende de la trayectoria que ésta siga dentro del campo; en vez de eso, dependerá solamente del punto A en el que se encuentre, y del punto B al que se mueva. [5]

Esto nos permite pensar que si una partícula que se encuentra en un campo conservativo se mueve sobre una trayectoria cerrada, la integral de la Fuerza sobre dicha curva será nula.

$$\oint F \cdot dl = 0. \quad (10)$$

¿De qué nos sirve esto? Es sencillo, podemos partir de este concepto para definir uno nuevo: el potencial. Una consecuencia de la ecuación (10) es que si nosotros trabajamos ahora con una trayectoria que no sea cerrada, el resultado de la integral sobre la curva será constante. Es decir, si nosotros integramos del punto A al punto B , obtendremos un resultado: V . Pero si integramos ahora del punto B al punto A (utilizando una trayectoria distinta), para que la ecuación (10) se cumpla, entonces el resultado deberá ser $-V$. Entonces $V - V = 0$.

Eso es algo importantísimo, pues podemos decir entonces que (derivando), la fuerza en un campo conservativo será igual al gradiente negativo de un campo escalar V .

$$\vec{F} = -\vec{\nabla}V. \quad (11)$$

Cuál es nuestra sorpresa al decir que un campo eléctrico generado por cargas en reposo (un campo electrostático), es conservativo, y por lo tanto, podemos reescribir la ecuación (11) como:

$$\vec{E} = -\vec{\nabla}V. \quad (12)$$

En electrostática, a V se le conoce como el Potencial Eléctrico, y es el trabajo que un campo eléctrico puede realizar para mover a una carga desde el punto A , hasta el punto B . [3][4][5]

Ahora sí podemos utilizar el “teorema importantísimo” que mencionamos previamente para dar el paso más importante del proceso: subir una dimensión. Hasta el momento, hemos trabajado con superficies. Pero, en el mundo real, hay más de dos dimensiones. ¿Podemos relacionar de alguna manera una superficie con un volumen? La respuesta es: ¡claro! Como trabajamos con superficies cerradas, de algún modo podemos relacionarlas con el volumen que éstas encierran. Aquí entra el súper teorema importantísimo que queremos introducir desde hace rato: el Teorema de la Divergencia de Gauss (Gauss, por lo visto, fue alguien muy famoso). Este teorema nos permite relacionar áreas con volúmenes de la siguiente manera:

$$\iint_S \vec{E} \cdot d\vec{S} = \iiint_V \nabla \cdot \vec{E} dV. \quad (13)$$

La integral de un campo sobre una superficie cerrada es igual a la integral del gradiente de dicho campo sobre el volumen que encierra dicha superficie (S es la “frontera” de V).³ [5]

¡Esto es algo maravilloso! Si nos fijamos del lado izquierdo de la igualdad, tenemos nada más y nada menos que a la Ley de Gauss (a excepción de la constante, ϵ_0), y por lo tanto, esa expresión es igual también a la carga total encerrada por la superficie con la que trabajemos: Q_t .

Por otro lado, si nosotros nos fijamos en el volumen que encierra la superficie, podemos suponer que, al existir una carga encerrada en dicho volumen, tenemos una densidad volumétrica de carga (a la cual habíamos llamado “ ρ ” al inicio del artículo). Entonces, si nosotros integramos la densidad volumétrica que tenemos por unidad de volumen dentro de nuestra área de

³La demostración del Teorema de la Divergencia de Gauss puede encontrarse prácticamente en cualquier libro de Cálculo de más de una variable, por ejemplo, en [5].

interés (el volumen encerrado por nuestra superficie), tendremos como resultado, de nuevo, ¡a nuestra carga total encerrada en ese volumen! Así que podemos reescribir, gracias a esto y al Teorema de la divergencia de Gauss, a la expresión (8) como:

$$\epsilon_0 \iiint_V \nabla \cdot \vec{E} dV = \int_V \rho dV. \quad (14)$$

Nos estamos acercando, al fin, a nuestra meta. Podemos hacer uso del Teorema Fundamental del Cálculo para derivar ambos lados y “eliminar” a nuestra integral, y después, pasamos dividiendo a ϵ_0 , para obtener:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}. \quad (15)$$

¡Ya casi se acaba el tormento! Recordemos que el campo eléctrico en este caso es conservativo, entonces podemos sustituir la expresión (12) en la (15) para obtener:

$$\nabla \cdot (-\nabla \cdot \vec{V}) = \frac{\rho}{\epsilon_0}. \quad (16)$$

¡Ah! ¡Ya está! Simplificamos y multiplicamos todo por un -1 para obtener, al fin, la siguiente ecuación:

$$\nabla^2 V(\vec{r}) = \frac{-\rho(\vec{r})}{\epsilon_0}. \quad (17)$$

Agregamos nada más la notación vectorial al potencial y a la densidad porque, efectivamente, ambos dependen de la posición en la que nos encontremos en el volumen con el que estamos trabajando.

¡Maravilloso! Después de definiciones y construcciones eternas, llegamos a lo que queríamos: la Ecuación de Poisson. Un “*formulazo*” que, como se puede apreciar aquí, tiene detrás de él a toda una rama de la ciencia: la Electrostática. Dentro de esa expresión tan pequeña caben conceptos enormes como la Fuerza Eléctrica, el Campo Eléctrico, la Ley de Coulomb, la Ley de Gauss, el concepto de Carga y el de Densidad (superficial y volumétrica), el Teorema de la Divergencia de Gauss, y, como vimos un poco a la hora de trabajar con superficies y volúmenes, ¡muchísima geometría! La Ecuación de Poisson se puede ver, en términos generales, como un gran resumen de lo que es la Electrostática, una de las ramas más importantes de una de las más grandes áreas de estudio de la Física: el Electromagnetismo.

Conclusión.

Los “*formulazos*” son la prueba perfecta de que cualquier rama de la Ciencia se puede simplificar y expresar de una manera elegante y en ocasiones extremadamente sencilla. Sin embargo, el hecho de poder utilizar uno implica una cantidad de trabajo previo enorme que se ha realizado a lo largo de mucho tiempo, y trae consigo una cantidad de información tal que pocas veces la entendemos por completo. Muchas veces, la sencillez de estas expresiones nos hace olvidar lo importantes y útiles que son y, en particular, lo profundo de su significado.

Lo que hay detrás de cada ecuación y fórmula que conocemos es muchas veces algo infinitamente más interesante que los problemas que estamos acostumbrados a resolver con ellas. Incluso hay libros dedicados explícitamente a hablar sobre la importancia que muchas de estas expresiones han tenido a lo largo del tiempo y su impacto en la humanidad (basta con hablar un poco sobre el Teorema de Pitágoras para entender esto). Por ejemplo, existe una ecuación que resume en una sencilla línea la cantidad de información que puede tener un mensaje codificado (el coco de los hackers). También existen otras que nos permiten, literalmente, viajar al espacio y a otros planetas (conservación de la energía, para el mundo), y otras tantas que permiten analizar hasta cierto punto el comportamiento que tendrá el mercado de valores bajo ciertas condiciones específicas. ¡Incluso tenemos expresiones que nos permiten modelar desde situaciones climáticas, hasta células humanas! Las posibilidades son infinitas. [6]

En fin. Así como la de los seres humanos, la anatomía de una ecuación puede ser extremadamente compleja, y estar llena de secretos e información útil que nos puede permitir entender al mundo de una mejor manera y también nuestro papel en el mismo.

Referencias

- [1] Goldstein, H. *Classical Mechanics*. Addison Wesley, Estados Unidos de América, 1980, Capítulo 1.
- [2] Alonso, M. Finn, E. *Fundamental University Physics, Volume I: Mechanics*. Addison Wesley Publishing Company, primera edición, Estados Unidos de América, 1967.
- [3] Purcell, Edward M. *Electricity and Magnetism, Berkeley Physics Course - Volume 2*. Education Development Center, primera edición, Estados Unidos de América, 1965.
- [4] Jackson, J. D. *Classical Electrodynamics*. Wiley, segunda edición, Estados Unidos de América, 1975.
- [5] Apostol, T. *Calculus, Volume II*. Wiley, segunda edición, Estados Unidos de América, 1967.
- [6] Stewart, I. *17 Ecuaciones Que Cambiaron El Mundo*. Ediciones Culturales Paidós, primera edición, México, 2015.

El Teorema de Representación de Bruno de Finetti

Eduardo Gutiérrez Peña
*Investigador del IIMAS, UNAM*¹

Introducción

Bruno de Finetti hizo importantes aportaciones a la Probabilidad y a la Estadística. Nació en Innsbruck, Austria, el 13 de junio de 1906 y murió en Roma, Italia, el 20 de julio de 1985. Hijo de padres italianos, recibió su educación y desarrolló prácticamente todo su trabajo en instituciones italianas. Se graduó de la carrera de Matemáticas Aplicadas en 1927. Para entonces ya tenía un marcado interés en la Teoría de la Probabilidad, y poco después se interesó en la Estadística. Cabe mencionar que también trabajó como actuario (durante 15 años, de 1931 a 1946).

Sus contribuciones a la Probabilidad incluyen resultados sobre procesos estocásticos con incrementos independientes y sobre sucesiones de variables aleatorias intercambiables. Por lo que se refiere a la Estadística, entre sus aportaciones podemos mencionar su trabajo relativo a la interpretación subjetiva de la probabilidad, el enfoque predictivo y la Teoría de la Decisión. Su labor fue fundamental para el desarrollo y consolidación de la Teoría Bayesiana durante la primera mitad del Siglo XX. Cifarelli & Regazzini (1996) describen con detalle las contribuciones de de Finetti y el impacto que ha tenido su trabajo en las áreas mencionadas. El lector interesado puede consultar también el obituario escrito por Lindley (1986), así como el sitio <http://www.brunodefinetti.it/>.

De Finetti publicó inicialmente en italiano y en francés, por lo que su trabajo no fue conocido en el mundo angloparlante sino hasta principios de la década de 1950, cuando Leonard Jimmie Savage (otro pionero de la Estadística Bayesiana moderna) lo “descubrió” (Good, 1977). Por su parte, de Finetti comenzó a desarrollar su visión de la Estadística Bayesiana sin estar al tanto del trabajo que sobre la probabilidad subjetiva y la toma de decisiones racional Frank Ramsey había realizado en Inglaterra unos años antes.

De acuerdo con Bernardo (1997), las aportaciones de de Finetti más trascendentes para el desarrollo de la Estadística contemporánea son: (a) la formalización del concepto de probabilidad como grado de creencia, que permite un tratamiento riguroso de la Probabilidad que se deduce a partir de la Teoría de la Decisión; y (b) el concepto de intercambiabilidad que, a través de los teoremas de representación, permite integrar en un paradigma unificado los conceptos estadísticos frecuentistas asociados a modelos paramétricos con el concepto de probabilidad como grado de creencia. Cada uno de estos dos puntos proporciona una justificación del paradigma Bayesiano.

¹ Actualmente realizando una estancia sabática en el Departamento Académico de Estadística del ITAM.

El propósito de esta nota es discutir el concepto de intercambiabilidad y Teorema de Representación (en su versión más simple), así como comentar sobre su importancia para clarificar el proceso de Inferencia Estadística.

El Concepto de Intercambiabilidad

Definición 1. Las variables aleatorias X_1, \dots, X_n son (*finitamente*) intercambiables bajo una medida de probabilidad P si la distribución inducida por P satisface

$$p(x_1, \dots, x_n) = p(x_{\pi(1)}, \dots, x_{\pi(n)})$$

para toda permutación π definida sobre el conjunto $\{1, 2, \dots, n\}$. En otras palabras, las “etiquetas” que identifican a cada una de las variables no proporcionan información alguna. Es claro que si las variables aleatorias X_1, \dots, X_n son independientes e idénticamente distribuidas entonces son intercambiables. El ejemplo siguiente muestra que X_1, \dots, X_n pueden ser intercambiables a pesar de no ser independientes.

Ejemplo. Sea $\mathbf{X} = (X_1, \dots, X_n)$ un vector aleatorio con distribución Normal Multivariada $N_n(\mathbf{0}, \Sigma)$ y supongamos que los elementos de la matriz de varianzas-covarianzas están dados por $\Sigma_{ii} = \sigma^2$ ($i = 1, \dots, n$) y $\Sigma_{ij} = \rho\sigma^2$ ($i \neq j$), con $\sigma^2 > 0$ y $0 \leq \rho < 1$. Consideremos los siguientes dos casos:

(i) La matriz Σ es diagonal ($\rho = 0$). Entonces las variables aleatorias X_1, \dots, X_n son independientes e idénticamente distribuidas, y por lo tanto intercambiables.

(ii) La matriz Σ no es diagonal ($0 < \rho < 1$). En este caso las variables aleatorias X_1, \dots, X_n no son independientes, pero siguen siendo intercambiables.

Definición 2. La sucesión infinita de variables aleatorias X_1, X_2, \dots es (*infinitamente*) intercambiable si toda subsucesión finita es intercambiable en el sentido de la Definición 1.

El Teorema

Por simplicidad, en esta nota discutiremos el teorema solamente para el caso de variables dicotómicas. Sea X_1, X_2, \dots una sucesión infinita de variables aleatorias intercambiables que toman valores en el conjunto $\{0, 1\}$. El Teorema de de Finetti proporciona una representación de la distribución conjunta de cualquier colección finita de dichas variables, digamos X_1, \dots, X_n .

Teorema (de Finetti). Si X_1, X_2, \dots es una sucesión infinita de variables aleatorias definidas sobre $\{0, 1\}$ e intercambiables con respecto a la medida de probabilidad P , entonces existe una función de distribución Q tal que la distribución conjunta $p(x_1, \dots, x_n)$ tiene la forma

$$p(x_1, \dots, x_n) = \int_0^1 \left\{ \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \right\} dQ(\theta),$$

donde $Q(\theta) = \lim_{n \rightarrow \infty} \Pr(Y_n/n \leq \theta)$, con $Y_n = X_1 + \dots + X_n$, y $\theta = \lim_{n \rightarrow \infty} Y_n/n$ (c. s.).

La demostración de este resultado y la del corolario discutido más adelante pueden encontrarse, por ejemplo, en Bernardo & Smith (2000).

El teorema nos dice que el modelo predictivo para una sucesión intercambiable de variables aleatorias binarias puede ser descrito en términos de una situación en la que:

- (i) condicional en el valor de una variable aleatoria, θ , las variables aleatorias X_i se consideran independientes con distribución Bernoulli;
- (ii) a θ se le asigna una distribución de probabilidad Q .

Por la Ley de los Grandes Números, $\theta = \lim_{n \rightarrow \infty} Y_n/n$ (c. s.), de manera que Q puede interpretarse como una descripción de los juicios acerca del límite de la frecuencia relativa de los “éxitos” en una sucesión de ensayos Bernoulli.

Corolario. Si X_1, X_2, \dots es una sucesión infinita de variables aleatorias definidas sobre $\{0, 1\}$ e intercambiables con respecto a la medida de probabilidad P , entonces la distribución condicional $p(x_{m+1}, \dots, x_n | x_1, \dots, x_m)$ es de la forma

$$p(x_{m+1}, \dots, x_n | x_1, \dots, x_m) = \int_0^1 \left\{ \prod_{i=m+1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \right\} dQ(\theta | x_1, \dots, x_m)$$

donde $1 \leq m < n$,

$$dQ(\theta | x_1, \dots, x_m) = \frac{\left\{ \prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i} \right\} dQ(\theta)}{\int_0^1 \left\{ \prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i} \right\} dQ(\theta)} \quad (1)$$

y $Q(\theta) = \lim_{n \rightarrow \infty} \Pr(Y_n/n \leq \theta)$.

La expresión (1) no es más que una versión del Teorema de Bayes. Notemos que la forma de la representación no cambia. En la terminología usual, la *distribución inicial* $Q(\theta)$ ha sido actualizada a través del Teorema de Bayes, obteniéndose la *distribución final* $Q(\theta | x_1, \dots, x_m)$. La distribución predictiva final, $p(x_{m+1}, \dots, x_n | x_1, \dots, x_m)$, nos permite derivar la correspondiente distribución predictiva de cualquier otra variable definida en términos de las observaciones futuras. Por ejemplo, dado $X_1 = x_1, \dots, X_m = x_m$, la distribución predictiva de $Y_{n-m} = X_{m+1} + \dots + X_n$ es de la forma

$$p(y_{n-m} | x_1, \dots, x_m) = \int_0^1 \binom{n-m}{y_{n-m}} \theta^{y_{n-m}} (1 - \theta)^{(n-m)-y_{n-m}} dQ(\theta | x_1, \dots, x_m).$$

Un caso particular importante es la frecuencia relativa de los “éxitos” en una muestra grande. Los resultados anteriores implican que

$$\lim_{n \rightarrow \infty} \Pr(Y_{n-m}/(n-m) \leq \theta | x_1, \dots, x_m) = Q(\theta | x_1, \dots, x_m).$$

Así, la distribución final del parámetro θ puede verse como un caso límite de la distribución predictiva final de una variable aleatoria observable: la frecuencia relativa de “éxitos”.

El Teorema de Representación puede extenderse al caso de variables aleatorias sobre espacios más generales (Bernardo & Smith, 2000).

Conclusiones

El Teorema de Representación tiene un significado muy profundo desde el punto de vista de la modelación subjetiva. Partiendo de una noción más elemental, como lo es el concepto de intercambiabilidad, el teorema no sólo permite justificar el supuesto usual de que las observaciones sean independientes e idénticamente distribuidas *condicional en un parámetro* θ , sino que además permite considerar a dicho parámetro como una variable aleatoria cuya distribución refleja nuestro estado de información sobre su valor desconocido. Lo anterior es la esencia de la Inferencia Estadística Bayesiana.

Referencias

- Bernardo, J.M. (1997). Bruno de Finetti en la Estadística Compemporánea. *Reporte Técnico*. Universidad de Valencia, España. <http://www.uv.es/~bernardo/DeFinetti.pdf>
- Bernardo, J.M. & Smith, A.F.M. (2000). *Bayesian Theory*. Chichester: Wiley.
- Cifarelli, D.M. & Regazzini, E. (1996). De Finetti's Contributions to Probability and Statistics. *Statistical Science*, Vol. 11, No. 4, 253–282.
- de Finetti, B. (1974/1975). *Theory of Probability I/II* (translation from Italian by A. Machi and A.F.M. Smith of a 1970 book), 2 volumes, New York: Wiley.
- Good. I.J. (1977). Review of *Theory of Probability I/II* by Bruno de Finetti. *Bulletin of the American Mathematical Society*, Vol. 83, No. 1, 94–97.
- Lindley, D.V. (1986). “Bruno de Finetti, 1906-1985 (Obituary)”. *Journal of the Royal Statistical Society*, Series A, Vol. 149, p. 252 .

El algoritmo de Metropolis-Hastings

Christian Alvarado

Estudiantes de Matemáticas Aplicadas y Actuaría del ITAM

Gian Carlo Diluvi

Estudiantes de Matemáticas Aplicadas del ITAM

Demian Espinosa

Estudiantes de Matemáticas Aplicadas e Ingeniería Industrial del ITAM

Introducción

El Algoritmo de Metropolis-Hastings (M-H) es una de las técnicas más populares de muestreo estadístico. Forma parte de una familia de métodos llamados de Monte Carlo, y se basa en generar una cadena de Markov que tenga, como distribución límite, la distribución objetivo (de la que se quiere generar una muestra).

Los métodos de Monte Carlo son una basta clase de algoritmos computacionales que se basan en la generación repetida de números aleatorios para obtener estimaciones de interés. Pueden ser utilizados para aproximar, por ejemplo, π , volúmenes en dimensiones altas, o para simular juegos aleatorios. Es muy común utilizar alguno de estos métodos, en conjunto con una cadena de Markov, para generar muestras de una distribución objetivo. La idea es forzar a esta cadena a tener la distribución objetivo como distribución límite, y generar números que provengan de esta cadena. Los algoritmos que usan este procedimiento se clasifican como Markov chain Monte Carlo (MCMC).

La historia de M-H se remonta a 1953, cuando Nicholas Metropolis (1915-1999), junto con los equipos de esposo y esposa de Marshall y Ariana Rosenbluth y de Edward y Augusta Teller, publicaron un artículo en el que mostraban las primeras simulaciones de un líquido. Poco sabían ellos que el método que desarrollaron, originalmente llamado Algoritmo de Metropolis, tendría un impacto profundo en la Estadística contemporánea (ver [4]). Sin embargo, pasaron casi veinte años antes de que se hiciera mayor investigación estadística en el tema. Eventualmente, en 1970, W. Keith Hastings generalizó este algoritmo¹ (más adelante se explicará precisamente cómo). Empero, el auge del método (ahora conocido como el algoritmo de Metropolis-Hastings) no llegó sino hasta varios años después. En 1990, Gelfand y Smith, basados en el artículo de Geman y Geman (1984), publicaron el método conocido como Gibbs sampler. Éste sí tuvo mucha mayor popularidad, y en 1992 Andrew Gelman probó que el método de Gibbs sampler era, de hecho, un caso particular de M-H. Fue así como, finalmente, el algoritmo de M-H logró alcanzar su merecido reconocimiento².

¹Ver [1] para más información.

²Para una historia más detallada, ver [2].

En el presente artículo exploraremos el algoritmo de M-H tanto teórica como prácticamente. Comenzaremos dando una explicación teórica del mismo, primero intuitiva y después con mayor formalidad. Después, mencionaremos una aplicación, a nivel conceptual, a la inferencia Bayesiana. Finalmente, explicaremos cómo se lleva a cabo la implementación computacional, y mostraremos un ejemplo práctico de muestreo univariado.

Cabe mencionar que este artículo no pretende ser de naturaleza formal, sino informativa. Si bien se presentará un bosquejo de la prueba de por qué el método funciona, ésta es más bien de carácter intuitiva y cubrirá solo un caso particular. Si el lector desea ahondar más en el tema, puede consultar la bibliografía sugerida (especialmente [3] y [5]).

Marco Teórico

Una idea intuitiva

Antes de entrar a detalle en la parte técnica, motivaremos la intuición detrás del algoritmo con el siguiente ejemplo, propuesto por John Kruschke en su libro *Doing Bayesian Data Analysis*³.

Consideremos a un político que está haciendo campaña en una cadena de islas. Éste pasa cada día en una de las islas, y cada una de éstas tiene una cierta población. El político quiere que el número de días que esté en cada isla sea proporcional a la cantidad relativa de población que representa esta isla del total. El problema surge de que el político no conoce la población total. Lo que sí conoce es la población de la isla en la que esté, así como la de las islas aledañas, pues puede preguntarle la población a los alcaldes de dichas islas. Sus asesores políticos le sugieren que haga lo siguiente: empieza la campaña en una isla escogida al azar. Cada día va a lanzar una moneda justa; si sale cara, propone que el siguiente día vayan a la isla a la derecha de la actual. En otro caso, propone la isla a la izquierda. Si la isla propuesta tiene más población que la actual, entonces al día siguiente visitan esa isla. En caso contrario, definimos P_{actual} y $P_{\text{propuesta}}$ como la población de la isla actual y la población de la isla propuesta, respectivamente. Entonces se visita la isla propuesta con probabilidad $\frac{P_{\text{propuesta}}}{P_{\text{actual}}}$ (esto último se puede simular con una ruleta, por ejemplo). Lo impresionante del algoritmo propuesto por sus asesores es que, en el largo plazo, la probabilidad de que esté en una isla concuerda con la población relativa de esa isla respecto al total. Y note que en ningún momento fue necesario calcular o estimar la población total. Este es, a grandes rasgos, el algoritmo de M-H⁴.

El lector debe notar que, claramente, la población de cada isla **no** es un número entre 0 y 1, como sí lo es la población relativa, es decir, la que nos interesa conocer. Lo increíble de este algoritmo es que, en realidad, esto no nos importa. Lo que nos interesa es la población de la isla propuesta con respecto a la isla en la que nos encontramos. Este detalle será de gran importancia. Finalmente, note que el hecho de moverse a la isla propuesta con certeza si tiene mayor población que la actual y con cierta probabilidad si no es así se puede escribir

³Para más detalles, ver [3], pp 119-120.

⁴En realidad este sería el algoritmo de Metropolis et al., que también es un caso particular del algoritmo de M-H.

de la siguiente manera:

$$r = P(\text{Moverse a la isla propuesta}) = \min\left(\frac{P_{\text{propuesta}}}{P_{\text{actual}}}, 1\right),$$

pues, si $P_{\text{propuesta}} > P_{\text{actual}}$, entonces el cociente vale 1 y visitamos la isla propuesta con certeza. Este cociente, que podemos pensar como $r = r(P_{\text{actual}}, P_{\text{propuesta}})$, se conoce como *cociente de Hastings*. Más adelante veremos la versión más general de éste, pero su propósito es el mismo.

Formalizando la idea intuitiva

Para explicar con mayor detalle por qué este algoritmo funciona, desarrollaremos las transiciones del ejemplo anterior. Supongamos que estamos en un estado (en una isla) θ . La probabilidad de moverse al estado $\theta + 1$ (la isla de la derecha), denotado por $P(\theta \rightarrow \theta + 1)$, es la probabilidad de proponer ese estado (que la moneda salga cara, por ejemplo) por la probabilidad de aceptarla (que depende de la población de dicha isla). Es decir,

$$P(\theta \rightarrow \theta + 1) = 0.5 \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right). \text{ Haciendo cuentas similares, podemos obtener}$$

$$P(\theta \rightarrow \theta - 1) = 0.5 \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right). \text{ Y así, la probabilidad de quedarse en el mismo estado } \theta \text{ es el complemento de irse a uno de los otros dos estados, es decir}$$

$$P(\theta \rightarrow \theta) = 0.5 \left[1 - \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right)\right] + 0.5 \left[1 - \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right)\right].$$

Podemos representar estas transiciones en una matriz. La posición $T_{i,j}$ es la probabilidad de ir del estado i al estado j . En este caso particular, una submatriz de esta cadena que muestra de la fila $\theta - 2$ a $\theta + 2$ y de las columnas $\theta - 2$ a $\theta + 2$ se ve así:

$$\begin{bmatrix} \ddots & P(\theta - 2 \rightarrow \theta - 1) & 0 & 0 & 0 \\ \ddots & P(\theta - 1 \rightarrow \theta - 1) & P(\theta - 1 \rightarrow \theta) & 0 & 0 \\ 0 & P(\theta \rightarrow \theta - 1) & P(\theta \rightarrow \theta) & P(\theta \rightarrow \theta + 1) & 0 \\ 0 & 0 & P(\theta + 1 \rightarrow \theta) & P(\theta + 1 \rightarrow \theta + 1) & \ddots \\ 0 & 0 & 0 & P(\theta + 2 \rightarrow \theta + 1) & \ddots \end{bmatrix}.$$

Sustituyendo las ecuaciones obtenidas en el párrafo anterior, la submatriz se ve así:

$$\begin{bmatrix} \ddots & 0.5 \min\left(\frac{P(\theta-1)}{P(\theta-2)}, 1\right) & 0 & 0 & 0 \\ \ddots & \alpha_1 & 0.5 \min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) & 0 & 0 \\ 0 & 0.5 \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right) & \alpha_2 & 0.5 \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right) & 0 \\ 0 & 0 & 0.5 \min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) & \alpha_3 & \ddots \\ 0 & 0 & 0 & 0.5 \min\left(\frac{P(\theta+1)}{P(\theta+2)}, 1\right) & \ddots \end{bmatrix},$$

donde

$$\begin{aligned}\alpha_1 &= 0.5 \left[1 - \min \left(\frac{P(\theta - 2)}{P(\theta - 1)}, 1 \right) \right] + 0.5 \left[1 - \min \left(\frac{P(\theta)}{P(\theta - 1)}, 1 \right) \right] \\ \alpha_2 &= 0.5 \left[1 - \min \left(\frac{P(\theta - 1)}{P(\theta)}, 1 \right) \right] + 0.5 \left[1 - \min \left(\frac{P(\theta + 1)}{P(\theta)}, 1 \right) \right] \\ \alpha_3 &= 0.5 \left[1 - \min \left(\frac{P(\theta)}{P(\theta + 1)}, 1 \right) \right] + 0.5 \left[1 - \min \left(\frac{P(\theta + 2)}{P(\theta + 1)}, 1 \right) \right].\end{aligned}$$

La ventaja de tener la matriz de transiciones es que podemos calcular las probabilidades de ir a cada isla en cualquier número de pasos: sólo hace falta multiplicar el vector que indique en dónde estamos por dicha matriz elevada al número de pasos. En virtud de esto, probaremos que la distribución objetivo, es decir, la población relativa de cada isla, es una distribución estacionaria. Para ello, considere el vector w de dicha distribución. Éste se ve así:

$$w = \frac{1}{Z} (\dots, P(\theta - 1), P(\theta), P(\theta + 1), \dots),$$

donde $Z = \sum_{\theta} P(\theta)$ es un factor de normalización (de hecho es la población total, que desconocemos) y cada componente θ de w es una isla distinta. Demostraremos que el componente θ de $w \cdot T$ es el mismo que el de w , para cualquier componente θ . Recordemos que el componente θ de $w \cdot T$ es $\sum_r w_r T_{r\theta}$ (simplemente aplicando la definición de multiplicación de matrices). Haciendo cuentas, tenemos que

$$\begin{aligned}\sum_r w_r T_{r\theta} &= \frac{P(\theta - 1)}{Z} \times 0.5 \min \left(\frac{P(\theta)}{P(\theta - 1)}, 1 \right) \\ &\quad + \frac{P(\theta)}{Z} \times \left(0.5 \left[1 - \min \left(\frac{P(\theta - 1)}{P(\theta)}, 1 \right) \right] \right) \\ &\quad + 0.5 \left[1 - \min \left(\frac{P(\theta + 1)}{P(\theta)}, 1 \right) \right] \\ &\quad + \frac{P(\theta)}{Z} \times 0.5 \min \left(\frac{P(\theta)}{P(\theta + 1)}, 1 \right).\end{aligned}$$

Hay 4 casos distintos: (i) $P(\theta) > P(\theta - 1)$ y $P(\theta) > P(\theta + 1)$; (ii) $P(\theta) < P(\theta - 1)$ y $P(\theta) < P(\theta + 1)$; los casos (iii) y (iv) son sus respectivos complementos. En cualquiera de los cuatro casos se llega a la misma respuesta, es decir,

$$\sum_r w_r T_{r\theta} = \frac{P(\theta)}{Z}.$$

Esto nos dice que la distribución objetivo es estacionaria, como queríamos. Sin embargo, esto no nos garantiza que cuando hagamos correr la cadena ésta eventualmente converja a la distribución estacionaria; tampoco nos garantiza que esta distribución sea la única distribución estacionaria a la que podría llegar la cadena. Dichas demostraciones son por demás técnicas y rebasan los propósitos de este artículo. Si el lector desea consultarlas de cualquier manera, éstas se encuentran en [5]. Note, sin embargo, que la intuición nos dice que esto debe ser cierto: estamos recorriendo todos los posibles estados. Cuando llegamos a un estado de alta probabilidad, es probable que nos quedemos ahí y que, por lo tanto, tomemos más muestras de él. Así, la cadena tenderá a quedarse donde haya una mayor concentración de probabilidad. Eso es precisamente lo que queremos.

Generalización

El algoritmo descrito antes es un caso muy particular del algoritmo de M-H. En el ejemplo anterior consideramos *(i)* un conjunto discreto de estados (las islas), *(ii)* en una sola dimensión y *(iii)* con solo dos propuestas de movimiento por estado (isla izquierda o derecha, y además ambas con la misma probabilidad). El algoritmo general sirve también para *(i)* distribuciones continuas, *(ii)* en cualquier número de dimensiones, y *(iii)* con una distribución de propuesta de movimientos más general. Sin embargo, el principio es el mismo.

Recordemos que el método requiere una distribución $P(x)$ que sea proporcional a la distribución objetivo $\pi(x)$ (conocer la población por islas, por ejemplo) de la cual se quiere muestrear. Ésta puede estar dada sobre un espacio multidimensional que puede ser discreto o continuo, y no necesita estar normalizada, es decir, que sume o integre uno, pero sí se necesita poder calcular el valor de $P(x)$ para cualquier candidato que vayamos obteniendo. Resumiendo, tendremos

$$\pi(x) \propto P(x) \quad \forall x \in \text{Sop}(\pi)$$

donde, además, $P(x)$ es conocido (o calculable) para cualquier x . También se requiere una distribución Q de la cual, conociendo $x \in \text{Sop}(\pi)$, podamos muestrear. Dicha Q viene a representar el volado que realizaba el político para decidir qué isla proponer, y es por eso llamada la distribución de propuestas. Luego, el algoritmo genérico consistirá en escoger un $x_0 \in \text{Sop}(\pi)$ arbitrario, generar $y \sim Q|x_0$, fijar $x_1 = y$ con probabilidad $\frac{P(Y) \cdot Q(X_{i-1}|Y)}{P(X_{i-1}) \cdot Q(Y|X_{i-1})}$ y repetir el proceso tantas veces como se quiera. Podemos resumir esto en pseudocódigo como sigue:

Algoritmo de Metropolis–Hastings

1. Escoge X_0 un valor arbitrario en el soporte de π
2. Para $i=1, 2, \dots, n$
 - Genera una observación Y de $Q|X_{i-1}$
 - Determina $r = \frac{P(Y) \cdot Q(X_{i-1}|Y)}{P(X_{i-1}) \cdot Q(Y|X_{i-1})}$
 - Determina $\alpha = \min(1, r)$
 - Escoge $X_i = Y$ con probabilidad α
 - $X_i = X_{i-1}$ con probabilidad $1 - \alpha$
4. Fin

La idea de incluir $\frac{Q(x|y)}{Q(y|x)}$ en el cociente de Hastings fue precisamente la aportación de Hastings, y la razón por la que dicho cociente lleva su nombre. Básicamente lo que hace es considerar a la distribución de propuestas también, de manera que la elección de ésta se pueda realizar de manera más general, pero aún garantizando que la distribución objetivo sea estacionaria.

Aplicaciones

Inferencia bayesiana

El algoritmo de Metropolis-Hastings, entre otros, permitió el reciente despegue de la inferencia Bayesiana. Al hacer inferencia paramétrica, uno supone que el fenómeno bajo estudio se manifiesta a través de datos que se pueden pensar como realizaciones de una cierta variable aleatoria, misma que pertenece a alguna familia paramétrica de modelos de probabilidad, donde lo único desconocido es el parámetro. De esta manera, describir al fenómeno equivale a estudiar dicho parámetro. Desde la perspectiva Bayesiana, se tiene un grado de creencia (de antemano) del comportamiento del parámetro. Esta creencia se puede pensar como una probabilidad, es decir, nuestra creencia de que el parámetro sea de acuerdo a cierta θ fija la denotamos por $P(\theta)$. Esta creencia sobre cada posible valor del parámetro se llama la distribución *a priori*. Uno de los métodos principales en inferencia Bayesiana es actualizar esta distribución a partir de nuevos datos disponibles. Esto se logra a través de la regla de Bayes, ya que si se nos presentan ciertos datos D , también llamados evidencia, entonces nuestras nuevas creencias sobre el modelo, también llamada distribución *a posteriori*, cambian y están dadas por:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int_{\Theta} P(D|\theta)P(\theta)d\theta}.$$

Con un modelo ya dado, muchas veces es factible poder calcular la verosimilitud, $P(D|\theta)$.

Sin embargo, la integral que está en el denominador tiende a ser muy complicada. Uno podría pensar en alguna aproximación numérica, pero si el parámetro vive en un espacio de dimensión alta, esto puede ser computacionalmente costoso e ineficiente. Empero, uno puede darse cuenta de que este factor, que es constante, es igual para cada θ . En realidad está ahí como un factor de normalización. Podemos entonces escribir

$$P(\theta|D) \propto P(D|\theta)P(\theta).$$

Esto es precisamente lo que necesitamos para aplicar M-H: una distribución que sea, si no igual, por lo menos proporcional a la distribución objetivo. He ahí la importancia de no tener que conocer precisamente la distribución objetivo.

Implementación

El lector probablemente haya notado la complejidad y dificultad inherentes a la teoría detrás del algoritmo de M-H. Sin embargo, esto se compensa de una manera satisfactoria: la implementación computacional del algoritmo es muy sencilla. Basta proponer una distribución de propuestas $Q(y|x)$ cuyo soporte sea al menos el soporte de la distribución objetivo, $\pi(x)$, una condición inicial arbitraria, e iterar sucesivamente. Algo que el lector debe notar es que la elección de la distribución de propuestas es algo que decide, según su juicio, el que realiza el estudio en cuestión. Esto se debe de realizar con cuidado, y dependerá del caso. Otro detalle que hay que mencionar es que, si bien el algoritmo, en el límite, nos genera muestras que provengan de la distribución objetivo, las primeras de ellas dependen altamente del valor inicial que se escoja. Por ello es costumbre, para compensar este hecho, desechar las primeras muestras hasta cierto valor. Este procedimiento se conoce como *burn in*.

Con esto, ya estamos listos para mostrar un ejemplo del algoritmo M-H en acción.

Muestreo de una densidad Beta

Si el lector ya llevó Cálculo de Probabilidades I (o Probabilidad), de igual manera probablemente no recuerde la distribución Beta. Por ello, la definiremos rápidamente y mostraremos su función de densidad:

Definición. Decimos que la variable aleatoria X sigue una distribución Beta con parámetros θ_1 y θ_2 si su función de densidad está dada por

$$f_X(x) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1} \mathbb{1}_{(0,1)}(x).$$

A continuación generaremos, utilizando el lenguaje estadístico R, una muestra de una variable aleatoria $X \sim \text{Beta}(2.7, 6.3)$ ⁵. En este caso, nuestra distribución objetivo π es precisamente la Beta. Requerimos además una distribución candidata $Q(y|x)$ que “cubra” el dominio de la Beta, i.e., el $(0,1)$. Uno claramente se ve tentado a escoger la más sencilla: la uniforme en $(0,1)$. Así, $Q \sim \mathcal{U}(0,1)$. Note que, en este caso, $Q(y|x)$ no depende de x . Esto implica que la razón de Hastings sea simplemente

$$r(x, y) = \frac{\pi(y)}{\pi(x)}.$$

A continuación mostramos el código de la implementación del algoritmo:

```
set.seed(230395)

n <- 10000      # Fijamos el número de iteraciones
X <- runif(1)   # Inicializamos la cadena de Markov con una condición inicial
```

⁵Este ejemplo fue tomado de [5], y el código fue adaptado de uno visto en el curso de Simulación, en el semestre de Primavera 2016.

```

# arbitraria
for(i in 2:n){
  y <- runif(1) # Generamos con la distribución Q(y/x)
  r <- dbeta(y,2.7,6.3)/dbeta(X[i-1],2.7,6.3)
  r <- min(1,r)
  # Generamos la probabilidad de aceptación de M-H
  xnew <- ifelse(runif(1) < r, y , X[i-1]) # Decidimos si cambiar o no
  X = c(X, xnew) # Actualizamos nuestro vector
}

# Ahora realizamos el burn in:
X <- X[5001:length(X)]

```

Para ver gráficamente qué tan bien se aproxima nuestra muestra a la verdadera distribución, graficamos la función de densidad de los datos (creada con la función `density` de R) y sobreponemos la densidad teórica de dicha Beta con una línea roja:

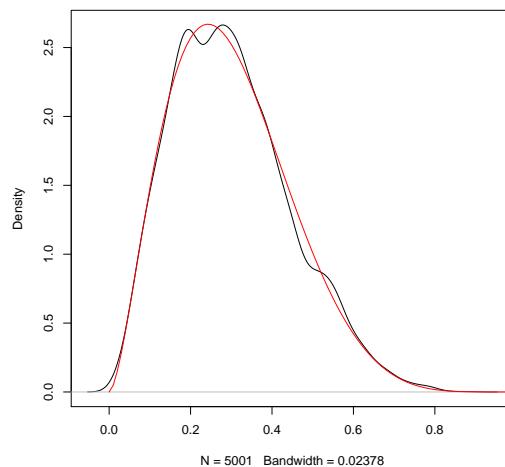


Figura 1: Gráfica de la densidad de los datos, X , vs la densidad teórica en rojo.

Como podemos ver, con un tamaño de muestra de $n = 5,000$ (después del burn in), la densidad de nuestros datos es prácticamente igual a la densidad teórica. Para corroborar lo que vemos gráficamente, utilizaremos la prueba de bondad de ajuste de Kolmogorov-Smirnov⁶, incluida en la librería `randtoolbox` de R. Rechazaremos la hipótesis nula (los datos sí provienen de una Beta(2.7,6.3)) si el p -value es menor a 0.05. Mostramos los resultados:

⁶Para más información, ver <http://mathworld.wolfram.com/Kolmogorov-SmirnovTest.html>.

```
ks.test(X,"pbeta", a,b)

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  X
## D = 0.014141, p-value = 0.27
## alternative hypothesis: two-sided
```

Note que el p -value sí fue mayor a 0.05, por lo que no podemos rechazar la hipótesis nula. Concluimos entonces que los datos sí provienen de una distribución Beta(2.7,6.3). Este ejemplo nos sirve para darnos una idea lo verdaderamente sencillo de implementar y, a su vez, poderoso, que es el algoritmo de M-H.

Conclusiones

Hay varias conclusiones y comentarios que realizar. En primer lugar, el lector debe haber notado la complejidad teórica del algoritmo. Justificar, con todo detalle y formalidad, por qué funciona en el más general de los casos es una tarea colosal. Como se mencionó anteriormente, un libro que incluye todos los detalles técnicos es el de Robert y Casella, [5]. En segundo lugar, y como también se mencionó, la ventaja del algoritmo de M-H es su sencilla implementación computacional y sus brillantes resultados. Algo que también es importante mencionar es que la verdadera utilidad de este algoritmo no se encuentra en ejemplos como el presentado en este artículo. Si se desea obtener muestras de una cierta distribución en una variable, hay una variedad de métodos conceptualmente más sencillos e igual de fáciles de implementar⁷ que generan tan buenos resultados como M-H. Sin embargo, es sencillo visualizar y entender el algoritmo con ejemplos de esta índole. Empero, el lector no debe perder de vista que M-H, desde su concepción y posterior desarrollo, fue creado para resolver problemas de naturaleza mucho más complicada. Finalmente, si bien el algoritmo de M-H se dio a conocer en general hasta hace poco tiempo, éste verdaderamente ha ganado una inmensa popularidad. Según SIAM News, el algoritmo de M-H es uno de los diez algoritmos más importantes del último siglo⁸. Es por ello relevante conocer su planteamiento teórico e implementación, incluso aunque hoy en día haya multiplicidad de paquetes (ya sea en R o cualquier otro lenguaje) que lo implementan automáticamente.

⁷En el curso de Simulación del ITAM se estudian, con todo detalle, dichos métodos.

⁸<https://www.siam.org/pdf/news/637.pdf>, consultado el 30 de mayo de 2016.

Referencias

- [1] Hastings, W. K., 1970, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, publicado en *Biometrika*, Vol. 57, No. 1. (Apr., 1970), pp. 97-109. URL: <http://www2.stat.duke.edu/~scs/Courses/Stat376/Papers/Basic/Hastings1970.pdf>, consultado el 29 de mayo de 2016.
- [2] Hitchcock, David. 2003. *A History of the Metropolis-Hastings Algorithm*, publicado en *The American Statistician*, Vol. 57, No. 4, pp. 254-257. Taylor & Francis. Ltd. URL: <http://www.jstor.org/stable/30037292>, consultado el 3 de mayo de 2016.
- [3] Kruschke, John K., 2011. *Doing Bayesian Data Analysis*, Academic Press, Indiana, USA.
- [4] Metropolis, N. and Rosenbluth, A. and Rosenbluth, M. and Teller, A. and Teller, E. 1953. *Equation of State Calculations by Fast Computing Machines*. URL: <https://ui.adsabs.harvard.edu/?abs/1953JChPh..21.1087M/abstract>, consultado el 3 de mayo de 2016.
- [5] Robert, Christian y Casella, George, 2004. *Monte Carlo Statistical Methods*, Springer, USA.

Gráficas químicas

Ramón Espinosa Armenta

Profesor del Departamento Académico de Matemáticas del ITAM

Ilan Jinich Fainsod

Estudiante de Actuaría y Matemáticas Aplicadas del ITAM

Adrián Tame Jacobo

Estudiante Matemáticas Aplicadas del ITAM

Introducción

El estudio de gráficas comienza cuando Leonhard Euler publica "*Solutio problematis ad geometriam situs pertinentis*" en 1736 en el cual se discute si es posible cruzar todos los puentes sobre el río Pregel de la ciudad de Königsberg (hoy Kaliningrado). Eventualmente, los estudios de Euler sobre los vértices y las aristas de un gráfica fueron generalizados por Augustin-Louis Cauchy y Simon Antoine Jean L'Huilier y son conocidos hoy como la base de la topología.

Otro matemático y científico importante en el desarrollo de la teoría de gráficas y especialmente sus aplicaciones a la química fue Arthur Cayley, quien estaba interesado más que nada en una clase de gráficas que hoy llamamos árboles. Las implicaciones que tuvieron sus estudios fueron fundamentales para el avance de la química teórica, y la relevancia de la teoría de gráficas en varias áreas.

Hoy en día, el estudio de la teoría de gráficas químicas es de gran interés para la comunidad científica.

El propósito de este artículo es utilizar conceptos de teoría de gráficas para definir estructuras químicas y sus propiedades. También se demostrarán resultados relacionados con propiedades importantes de las estructuras.

Gráficas

Definición. Una gráfica es una pareja $G = (V, E)$ donde V es un conjunto finito no vacío cuyos elementos se llaman vértices y E es un conjunto cuyos elementos son subconjuntos de cardinalidad 2 de vértices cuyos elementos se llaman aristas.

Definición. La matriz de incidencia de G es la matriz $M = (M_{ik}), i = 1, \dots, |V|; k = 1, \dots, |E|$ donde

$$M_{ik} = \begin{cases} 1 & \text{si } e_k \text{ incide en } i \\ 0 & \text{en otro caso} \end{cases}$$

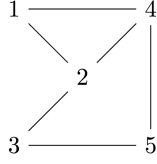


Figura 1: Ejemplo de una gráfica.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Figura 2: Matriz de la gráfica en la Figura 1.

Definición. El grado de un vértice $v \in V$ es el número de aristas que inciden en v , se denota $d(v)$.

Una conclusión inmediata de esto es:

$$d(i) = \sum_{k=1}^{|E|} m_{ik}$$

Teorema. Primer teorema de teoría de gráficas

Si G es una gráfica Entonces:

$$\sum_{i \in V} d(i) = 2|E|$$

Demostración. Dado que cada arista e_k coincide en exactamente 2 vértices, se tiene que:

$$\sum_{i=1}^{|V|} m_{ik} = 2$$

Se sigue que:

$$\sum_{i \in V} d(i) = \sum_{i=1}^{|V|} \sum_{k=1}^{|E|} m_{ik} = \sum_{k=1}^{|E|} \sum_{i=1}^{|V|} m_{ik} = \sum_{k=1}^{|E|} 2m_{ik} = 2|E|$$

//

Definición. Un camino es una sucesión finita de vértices, aristas, de forma alternada:

$$w = (v_1, e_1, v_2, e_2, \dots, e_{k-1}, v_k)$$

Donde:

$$e_i = v_i v_{i+1}$$

Se acostumbra describir un camino por la sucesión de vértices.

Definición. Una trayectoria es un camino donde no se repiten aristas y vértices.

Definición. Un ciclo es un camino donde no se repiten aristas, su origen y destino coinciden y además su origen y vértices internos son distintos.

Definición. Una gráfica es conexa si para toda pareja de vértices en la gráfica existe una trayectoria.

Definición. Un árbol es una gráfica acíclica (no contiene ningún ciclo) y conexa.

Teorema. Si T es un árbol con más de dos vértices entonces debe tener al menos dos vértices de grado 1.

Demostración. Sea P una trayectoria maximal (No está contenida en ninguna otra trayectoria), donde $v_i, i = 1, \dots, k$ son los vértices de la trayectoria.

Supongamos $d(v_i) > 1$ para toda i .

Sabemos que v_2 es vecino v_1 y además sabemos que v_1 no es vecino de ningún vértice v_i con $i \geq 3$ en P .

Por lo tanto existe $w \in V$ tal que $wv_1 \in E$

$$\text{Entonces } P \subset P + w$$

Por lo tanto P no es maximal, lo cual contradice la hipótesis de que P es maximal.

Es análogo para v_k .

//

Teorema. Si T es un árbol entonces:

$$|E| = |V| - 1$$

Demostración. Por inducción sobre $n = |V|$

Si $n=1$, T es un árbol trivial (solo tiene un vértice) entonces $|E(T)| = 0 = 1 - 1$

Supongamos el teorema cierto para cualquier árbol con n vértices.

Sea T un árbol con $n+1$ vértices.

Sea v un vértice de grado 1 en T .

Por lo tanto $\hat{T} = T - v$ es un árbol con n vértices

Por hipótesis $|E(\hat{T})| = n - 1$

$$\text{Por lo tanto } |E(T)| = |E(\hat{T})| + 1 = n = |V(T)| - 1$$

//



Figura 3: Diagrama Estructural de H_2O (Alexander Brown, 1864).

Un Poco de Química

Una molécula química consta de un conjunto finito de átomos unidos por medio de enlaces químicos. Por ejemplo H_2O .

Una fórmula o diagrama estructural es una representación gráfica de la composición de una molécula, que muestra la distribución espacial de los átomos. Una importante diferencia entre un diagrama estructural y una fórmula química estándar es que existen varias moléculas (como la de butano e isobutano) que aunque tienen la misma fórmula, sus vínculos son diferentes.

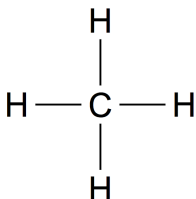


Figura 4: Diagrama Estructural del Metano

La fórmula estructural tiene diferentes representaciones, entre ellas las fórmulas semidesarrolladas, diagramas de Lewis, pero para los propósitos de este artículo utilizaremos el formato línea-ángulo estándar.

Los isómeros químicos son moléculas que tienen la misma fórmula química pero distintas propiedades.

Un isómero químico pertenece a la familia de los alcanos si satisface que:

1. Todo átomo es un átomo de carbono o de hidrógeno.
2. Todo átomo de carbono tienen valencia (grado) 4 y todo átomo de hidrógeno tiene valencia 1.
3. El diagrama estructural es acíclico y conexo (i.e. es un árbol).

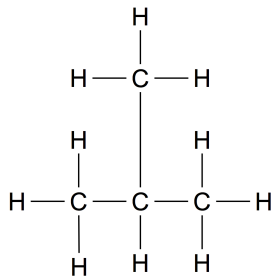


Figura 5: Diagrama Estructural del isobutano

Gráficas Químicas

Teorema. (Cayley) Si una molécula de la forma C_rH_s es un alcano, entonces $s = 2r + 2$

Demostración. Sea $G=(V,E)$ la gráfica que representa el diagrama estructural de C_rH_s . Separamos el conjunto de vértices en carbonos e hidrógenos, de la siguiente manera:

$$V = C \cup H$$

Donde C es el conjunto de vértices que son carbonos y H es el de hidrógenos.

Como un vértice no puede ser carbono e hidrógeno al mismo tiempo, se tiene que:

$$C \cap H = \emptyset$$

Por hipótesis:

$$|C| = r$$

$$|H| = s$$

Si n es la cardinalidad de V se tiene que:

$$n = |V| = r + s$$

Como la gráfica es un árbol, entonces:

$$|E| = n - 1 = r + s - 1$$

Por otra parte,

$$\sum_{v \in V} d(v) = \sum_{v \in C} d(v) + \sum_{v \in H} d(v) = 4r + s$$

También,

$$\sum_{v \in V} d(v) = 2|E| = 2r + 2s - 2$$

Por lo tanto $4r + s = 2r + 2s - 2$ entonces $2r + 2 = s$.

//

Comentarios Finales

Teoría de gráficas es un tema extremadamente amplio que puede abarcar varios aspectos de la vida cotidiana y cuenta con una gran variedad de aplicaciones, como en este caso la química.

Si al lector le interesa el tema, recomendamos consultar el libro de “*Introduction to Graph Theory*” de Douglas D. West[3] y el artículo de “*Contando Árboles Filogenético*” de Ramón Espinosa Armenta[4].

Referencias

- [1] ”Graph Theory”. 2016. Math.Fau.Edu. Accessed September 9 2016. <http://math.fau.edu/locke/GRAPHTHE.HTM>.
- [2] Espinosa, R., “Contando árboles filogenéticos.” *Laberintos e Infinitos* 38 (2015): 6-11.
- [3] West, Douglas B. *Introduction to Graph Theory*, Pearson, 2001.
- [4] Espinosa, R., *Matemáticas Discretas*, Editorial Alfaomega, 2010.
- [5] N.L. Biggs, E.K. Lloyd and R.J. Wilson. *Graph Theory 1736-1936*. Clarendon Press, Oxford, 1976.
- [6] Brown, A. C. “On the Theory of Isometric Compounds.” *Transactions of the Royal Society of Edinburgh* 23 (1864): 707-719
- [7] Cayley, A. “On the Mathematical Theory of Isomers.” *Philosophical Magazine* (4) 47 (1874): 444-446
- [8] Sylvester, J. J. “Chemistry and Algebra” *Nature* 17 (1877): 284

Árboles y máquinas inteligentes

Héctor Garrido

Estudiante de Economía y Matemáticas Aplicadas del ITAM

Introducción

Mucho se ha dicho de la Inteligencia Artificial, pero ¿en qué medida la realidad se ha puesto a la par con la ficción?

Más de la que algunos creen. Con sistemas de reconocimiento de lenguaje, clasificación de clientes bancarios a partir de enormes bases de datos, máquinas capaces de vencer a campeones de ajedrez a partir de retroalimentación propia, recomendaciones de compras en base a nuestros patrones de consumo y asistencia en la identificación de nuevos tipos de células cancerígenas, la inteligencia artificial ha ido permeando en nuestro quehacer diario de manera sutil pero sustancial. Esto ha despertado discusiones en campos de estudio tan diversos como computación, estadística, filosofía, epistemología y ética [4].

Una de las ramas de la inteligencia artificial, conocida como aprendizaje de máquina (machine learning) hace uso de la estadística, el álgebra lineal y la ciencia computacional para hacer que una máquina mejore su desempeño en determinada tarea a medida que pasa el tiempo. Un ejemplo claro sería intentar predecir el índice de precios dados los datos trimestrales de la última década que se actualiza con los datos reales regularmente.

En este artículo nos enfocaremos en uno de los métodos de aprendizaje de machine learning: los árboles de decisión con parte de sus fundamentos teóricos y un ejemplo.

Definiciones y conceptos

Para formalizar el concepto de aprendizaje al punto de poder trasladarlo a la computadora y hacer que esta lo comprenda se necesita de una estructura que sea consistente y se apegue al rigor matemático [1]. Para ello, comenzaremos con la definición clave de aprendizaje:

Un programa computacional **aprende** de la experiencia E respecto a cierta tarea T con medida de desempeño P si su desempeño en tareas en T , medido por P mejora con la experiencia E .

Como buscamos predecir valores futuros a partir de los datos disponibles en el presente hay un rango de error para nuestros pronósticos, algo que es propio de la estadística que por definición busca definir fenómenos que aparecen a través de datos que presentan variabilidad. Aunado a esto está el **sesgo inductivo**, el cual básicamente nos dice qué cosas da por supuestas determinado método para llegar a un resultado concreto. Esto, a su vez, se encuentra conectado con la **hipótesis de aprendizaje inductivo** que establece que toda hipótesis que

se vea adecuada aproximando sobre un conjunto suficientemente grande de ejemplos observados también se aproximará adecuadamente a la evaluación de ejemplos no observados futuros.

Esto, sin embargo, plantea una serie de preguntas a la hora de llevarlo a la práctica, ¿cuál es el tamaño adecuado del conjunto de ejemplos observados para dar ese paso inductivo? ¿se puede cuantificar el grado de exactitud con el que mi función final predecirá los ejemplos futuros? ¿cuáles serían las consecuencias de un pronóstico incorrecto?, estas y otras cuestiones son competencia tanto de la estadística como de la teoría de la información y el campo del caso en el que se esté trabajando, y una forma de cuantificar la incertidumbre que imponen nuestros datos es la medida de entropía. A su vez, la ganancia de forma simétrica mide que tan bien determinada variable de nuestros datos observados nos separa mejor los datos para su clasificación.

$$Entropía(S) = \sum_{i=1}^c -p_i \log_2(p_i)$$

$$Ganancia(S, A) = Entropía(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropía(S_v)$$

Dadas estas definiciones generales podemos proceder al método de **árboles de decisión**: un método capaz de sobreponerse al ruido presente en los datos. Normalmente se orienta a una función objetivo discreta.

¿Cómo funciona? De las variables toma aquella con la mayor ganancia. Una vez clasificados, procede con la nueva variable de mayor ganancia sujeta al subconjunto de información en el que se encuentra clasificado y así sucesivamente, dejando valores predictivos explícitos cuando los datos disponibles ya no presentan variabilidad en determinado nivel de clasificación. A continuación se muestra una representación del algoritmo ID3, que es su versión más simple (excluyendo casos triviales sin variabilidad) y un ejemplo para verlo con mayor claridad:

```
ID3(S, R, V)
S: ejemplos, datos observados.
R: variable por predecir
V: conjunto de variables entrenadas por el árbol
v ← variable ak tal que max({Ganancia(S1), ..., Ganancia(Sn)}) = Sk
variable de decisión para R ← v
para cada valor v(i) en v
agregar nueva rama bajo R correspondiente a v = v(i)
Si = {ejemplos en S tal que v = v(i)}
si Si = ∅ agregar nodo final bajo esta rama con etiqueta = valor más
común R en S
en caso contrario agregar subárbol bajo esta rama ID3(Si, R, V - {v})
```

Caso aplicado: encuestas electorales

En este ejemplo, por medio del método de árboles de decisión buscaremos una regla de comportamiento de los individuos en torno a las elecciones, suponiendo que los factores que inciden en sus decisiones (A, B, C) son tanto su nivel socioeconómico (bajo, medio, alto) como el partido de origen del dirigente en turno de la localidad (A,B,C).

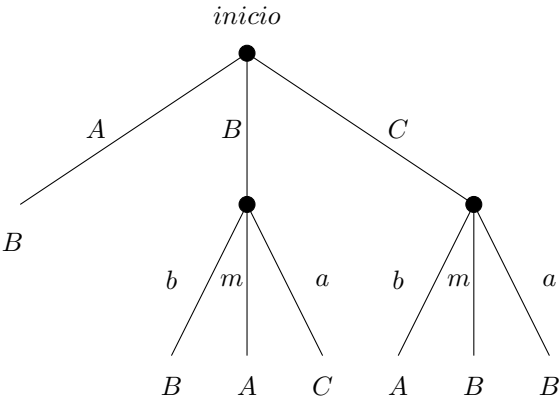
Nivel	Anterior	Respuesta
Bajo	A	B
Bajo	A	B
Bajo	C	A
Medio	B	A
Medio	C	B
Alto	B	C

Cuadro 1: Datos de encuestas electorales

$Ganancia(S, Nivel) = 1,46 - (0,5)0,91 - (0,33)1 - (0,16)0 = 0,68$

$Ganancia(S, Anterior) = 1,46 - (0,33)0 - (0,33)1 - (0,33)1 = 0,79$

Comparando los niveles de ganancia por variable es mejor tomar como primera variable a Anterior y, dado que solo queda una variable más esta sería la segunda en turno para clasificar. Nótese que ya no fue necesario continuar clasificando en la rama A, pues los datos disponibles ya no presentaban variabilidad, con único valor Respuesta igual a B. Para las otras ramas se prosiguió con la clasificación y eventualmente les fueron asignadas un valor predictivo de acuerdo al algoritmo.



A pesar de la simplicidad de este ejemplo, es evidente que se puede llevar a n variables con m posibles estados por variable, siendo incluso factible con datos en los números reales y otras extensiones más. Con suficiente imaginación podrá verse que puede fortalecerse de campos como Estadística Bayesiana, teoría de juegos y optimización [2].

Consideraciones finales

A medida que ha avanzado el tiempo, el crecimiento exponencial de las capacidades de cómputo que tenemos disponibles, junto con un respaldo teórico que se ha ido acumulando a lo largo de los años, nos han permitido expandir el alcance práctico de este campo, pero debemos que tener cuidado de no abusar de él. Cuando se hace un descubrimiento o desarrollo de alguna nueva tecnología tendemos a olvidarnos que estas son un medio y no un fin para hacer las cosas, y las viejas prácticas no deben de quedar relegadas si no que, al contrario, deben de complementarse mutuamente.

Sin embargo, no pretendo desincentivar a nadie con esto. Este campo, con lo íntimamente relacionado que está a la programación y su carácter autodidacta, nos ofrece un bagaje de conocimiento inmenso que siempre está en constante aumento dadas las contribuciones diarias de otros entusiastas en el tema [3] (más ahora con el ascenso de software *open source*).

Referencias

- [1] Tom M. Mitchell, *Machine Learning*, 1997
- [2] Andrew Gelman et al, *Bayesian Data Analysis*, 2004
- [3] <https://www.kaggle.com/>
- [4] <https://www.quora.com/>

Activa tus neuronas

La contraseña

Un grupo de policías investiga el cuartel general de un grupo de delincuentes. Quieren infiltrarse, pero necesitan la contraseña, así que vigilan para intentar averiguarla. Un tipo se acerca a la puerta. Desde el interior le dicen: “18”. Contesta: “9”. La puerta se abre y le dejan pasar. Llega otro. Le dicen: “8” y contesta: “4”. También le dejan entrar. Llega un tercero. Al número “14” contesta “7” y le abren la puerta.

Los policías creen haber dado con la clave: sólo hay que dividir entre dos el número que digan. Así pues, deciden enviar a un agente de incógnito. Al llegar a la puerta le dicen “0”. Contesta: “0”. La puerta no sólo no se abre, sino que le disparan y lo matan. Lo vuelven a probar con otro agente. Desde dentro se oye: “6”. Contesta: “3”. Matan a este agente también. ¿Cuál es el error que cometieron los policías?

Frutería

Tienes una frutería y te han repartido tres cajas: una tiene sólo manzanas; otra, sólo naranjas; la tercera, manzanas y naranjas. Cada caja tiene una etiqueta: “manzanas”, “naranjas” y “manzanas y naranjas”. Ninguna de las cajas tiene la etiqueta que le corresponde.



¿Cómo puedes saber la fruta que contiene cada una de las cajas sacando una sola pieza de una sola de ellas?

Camino del monje

Un monje parte al amanecer de su monasterio hasta la cima de una montaña tras un camino de varias horas. Se queda a descansar y a dormir, y sale por la mañana de la montaña a la misma hora para regresar a su monasterio.

Es posible que no tardara lo mismo en ir que en volver y da igual que su velocidad no fuera constante o cuándo y cuántas veces se parara a descansar: el monje pasó por algún punto del camino exactamente a la misma hora, pero con un día de diferencia. ¿Por qué?

La cuerda y el tubo

Una cuerda está enrollada de forma simétrica alrededor de una barra circular. La cuerda da la vuelta exactamente cuatro veces alrededor de la barra, que tiene una circunferencia de 4 centímetros y una longitud de 12 centímetros. Averigua cuánto mide la cuerda.



HINT: La perspectiva es importante.

Beer Pong

Un triángulo de 10 vasos está hecho de forma que el vértice apunta hacia arriba. ¿Cuántos vasos rojos deberán moverse como mínimo para que el triángulo apunte hacia abajo?



HINT: La respuesta no es 6.

Zona Olímpica

1. Encuentra las soluciones de enteros positivos de la ecuación

$$x^{x+y} = y^{y-x}.$$

2. Sea $f : [0, 1] \rightarrow [0, 1]$ una función continua tal que $f(f(x)) = 1$ para todo $x \in [0, 1]$. Determina todos los posibles valores de

$$\int_0^1 f(x) dx.$$

3. Encuentra todas las funciones $f : \mathbb{R} \rightarrow \mathbb{R}$ que satisfagan

$$f(x^2 - y^2) = (x - y)(f(x) + f(y)).$$

4. Prueba que las raíces del polinomio con coeficientes complejos dado por

$$p(z) = z^7 + 7z^4 + 4z + 1$$

están adentro del disco de radio 2 centrado en el origen.

5. Prueba que para cada número natural n existen n enteros consecutivos tal que cada uno de ellos es divisible por dos primos distintos.
6. Para una sucesión $a_1, a_2, \dots, a_n, \dots$, define la n -ésima suma parcial, $S_n = a_1 + a_2 + \dots + a_n$. Prueba que, para cualquier entero positivo n ,

$$\sum_{k=0}^n \binom{n}{k} a_{k+1} = \frac{2^n}{n+1} S_{n+1}.$$

7. Un polinomio real no-cero $p(x)$ satisface $p(x) = p'(x)p''(x)$ para cualquier x . ¿Cuál es el coeficiente principal de $p(x)$?
8. Para $x > 0$, define $f(x) = x^x$. Encuentra todos los valores de x para los cuales $f(x) = f'(x)$.

Pregunta de Erdős

Sea a_1, a_2, a_3, \dots una sucesión de números reales positivos. Se tiene que para algún entero positivo s ,

$$a_n = \max \{a_k + a_{n-k} \text{ tal que } 1 \leq k \leq n-1\}$$

para todo $n > s$. Demuestre que existen enteros positivos ℓ y N , con $\ell \leq s$, tales que $a_n = a_\ell + a_{n-\ell}$ para todo $n \geq N$.

Teoría de grupos aplicada al modelamiento del cubo de Rubik

Alonso González Valdés

Estudiantes de Matemáticas Aplicadas, ITAM

Mario Enrique Carranza Barragán

Estudiantes de Matemáticas Aplicadas, ITAM

“El álgebra es generosa; a menudo da más de lo que se le pide.”

Jean Le Rond d’Álambert

Introducción y objetivo general

El cubo Rubik fue inventado por el arquitecto Ernő Rubik en 1974, no como un medio de entretenimiento, sino como una herramienta que mejorase sus clases universitarias. Posteriormente habría una infinidad de ellos en los estantes de jugueterías a nivel mundial. Pero luego, por la dificultad que muchos hallaban al intentar solucionarlo, recobraría el interés académico, en particular el de matemáticos. Y durante ya un tiempo han habido varias implementaciones de algoritmos de solución y modelado. Más aún, por la relación entre las matemáticas y la computación, muchos de estos algoritmos están diseñados para computadoras y no para seres humanos.

Nuestro objetivo es explorar el cubo Rubik usando las matemáticas proporcionadas por la Teoría de Grupos. Primero daremos una representación del cubo (un modelo del cubo físico), y luego daremos pie a una representación de los *movimientos* que se pueden realizar con el cubo, para finalmente llegar a un algoritmo para solucionar el rompecabezas.

Modelo del cubo físico

Las 26 piezas que conforman el cubo se dividen en tres tipos: los **centros** (que sólo tienen un color asignado), las **aristas** (que tienen dos colores asignados), y los **vértices** (las esquinas, a quienes les corresponden tres colores). Al espacio que una pieza ocupa, sin importar el tipo, se le llama **cubículo**. De esta forma, como las piezas centrales nunca cambian de cubículo, las usamos como punto de referencia para los cubículos que rodean cada una. Cuando uno realiza movimientos con el cubo, las piezas se mueven de cubículo en cubículo. Por lo anterior, podemos decir que el cubo está resuelto si las piezas están en su cubículo original, y si los colores de cada pieza están en su orientación correcta (de nuevo respecto al centro).

Nuestro modelo consiste en una **4-tupla** (ρ, v, σ, w) , donde ρ (respectivamente σ), describe cómo los vértices (respectivamente las aristas) están permutados, y v (y w), describe cómo los vértices (y las aristas) están orientadas. Más aún, imaginemos que las piezas y los cubículos están enumerados. Cuando una pieza está en su posición natural, es porque su número asociado es igual al de su cubículo.

Cualquier configuración del cubo arrojará dos permutaciones (ignorando las orientaciones de las piezas): $\rho \in S_8$, que corresponde a cómo los vértices han sido permutados, y $\sigma \in S_{12}$, que representa cómo las aristas fueron permutadas.

A su vez, para las orientaciones de cada pieza (la distribución de los colores) enumeramos cada una de las caras o colores del cubo asociados a las piezas movibles, a saber, aristas y vértices (véase Figura 1). 0, 1, 2 para los vértices (pues tienen tres colores), y 0, 1 para las aristas (pues tienen dos posibles colores). Por esta razón, muy naturalmente trabajamos con \mathbb{Z}_3 y \mathbb{Z}_2 , respectivamente. O más bien:

- $w = (w_1, w_2, \dots, w_{12}) \in \mathbb{Z}_2^{12}$, es decir, $w = \{0, 1\}^{12}$ para la configuración de las 12 aristas, donde w_i corresponde al i -ésima arista y el color que ésta contiene.
- $v = (v_1, v_2, \dots, v_8) \in \mathbb{Z}_3^8$, es decir, $v = \{0, 1, 2\}^8$ para la configuración de los 8 vértices, donde v_j corresponde a la j -ésimo vértice y el color que éste contiene.

Para simplificar la descripción de un cubo y poder ver rápidamente la posición relativa y sus orientaciones, convenimos en nombrar las aristas y vértices de una forma única establecida por una convención arbitraria. Teniendo el cubo de frente comenzamos a nombrar los vértices. Este nombramiento se puede ver claramente en la Figura 1. Se comienza por las aristas superiores y se recorre en el sentido de las manecillas del reloj. Después hacemos lo mismo con las inferiores. Para las aristas procedemos de igual manera: primero las 4 de arriba, las 4 de en medio y finalmente las 4 de abajo. Además, toda estampa tiene marcado lo que llamaremos número de orientación de la estampa que, para simplificar, se abreviará como NDO. Observamos en la imagen que las caras superior e inferior tienen puros ceros en sus NDO (tanto de los vértices como las aristas), y dos de las aristas tanto de la cara izquierda como la derecha. Esto nos será útil para identificar rápidamente en qué orientación se encuentra determinado vértice o arista.

Podríamos simplificar todo lo anterior con la siguiente definición:

Definición. Sea X una configuración del cubo. El **vector de descripción del estado** X es $(\rho, v, \sigma, w) \in S_8 \times \mathbb{Z}_3^8 \times S_{12} \times \mathbb{Z}_2^{12}$, donde:

- $\rho \in S_8$ codifica la permutación de los vértices respecto a los centro inamovibles; $\rho(i) = j$ si la arista j está en el cubículo i .
- $v \in \mathbb{Z}_3^8$, donde el número v_i corresponde a la orientación del vértice j en el cubículo i .
- $\sigma \in S_{12}$ codifica la permutación de las aristas respecto a los centro inamovibles; $\sigma(i) = j$ si el vértice j está en el cubículo i .

0.1234567891011121314151617181920212223242526272829303132333435363738394041

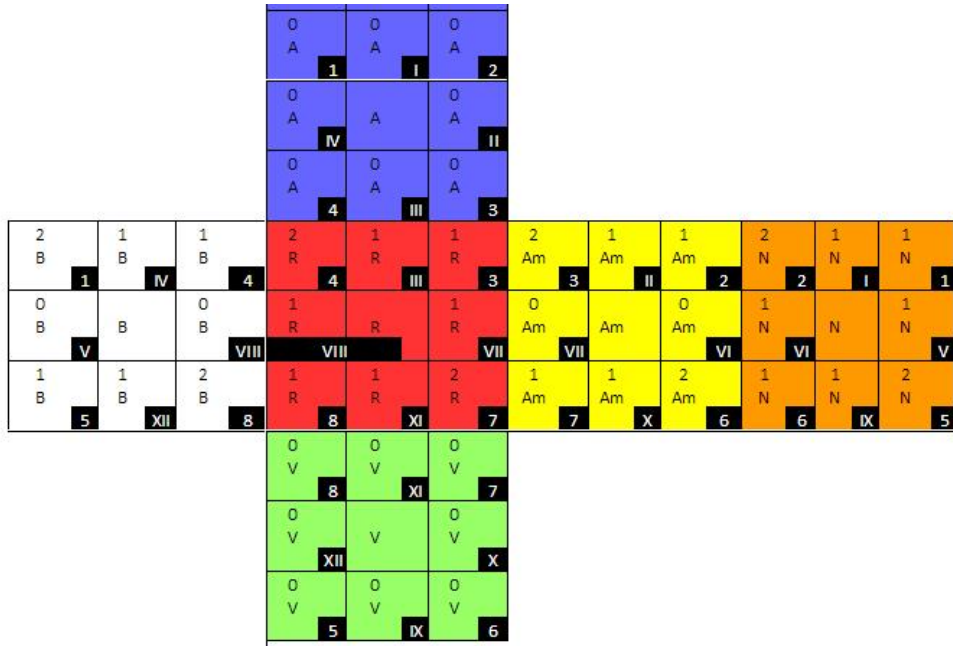


Figura 1: Esquema del cubo con etiquetas de esquinas y aristas

- $w \in \mathbb{Z}_2^{12}$, donde el número v_j corresponde a la orientación de la arista j en el cubículo i .

Ejemplo. *Notación:* Para el siguiente ejemplo, para representar una permutación $\rho \in S_j$, se usará un vector donde, si en la entrada i -ésima hay una k , significa que $\rho(i) = k$.

A) De tupla a representación gráfica

Primero veamos, dada una 4-tupla, cómo llegar a su representación gráfica.

Consideremos:

- $\rho = (6, 1, 4, 3, 5, 2, 8, 7) \in S_8$.
- $\vec{x} = (1, 0, 2, 2, 2, 0, 2, 0) \in \mathbb{Z}_3^8$.
- $\tau = (10, 5, 1, 6, 8, 7, 9, 12, 11, 4, 2, 3) \in S_{12}$.
- $\vec{y} = (0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1) \in \mathbb{Z}_2^{12}$,

que conforman el vector de estado $(\rho, \vec{x}, \tau, \vec{y})$.

Tomemos el vector de permutaciones para los vértices. Observemos que tiene un 4 en la tercera entrada, por lo que el vértice 4 (Azul, Blanco, Rojo) está en el lugar correspondiente al vértice 3 (Azul, Rojo, Amarillo). Además veamos que, en el lugar 1 de las orientaciones de vértices, hay un 1 de orientación. Luego, el color 0 (en este caso Azul) va en la posición 1 del vértice 3 (el Rojo); así, todos los colores quedan desplazados a la derecha, y de esta manera tenemos que el Azul está en donde debería ir el Rojo, el Blanco donde debería ir el Amarillo, y el Rojo donde debería ir el Azul (por “debería” nos referimos al color que tendría si el cubo estuviese resuelto. Esto se comprueba fácilmente viendo si el color de cada estampa de la esquina coincide con sus respectivos centros).

Este mismo análisis aplica para las aristas, donde es un poco más fácil entender y aplicar las orientaciones en tanto que son solo dos. Ejemplifiquemos con la entrada 2, en la que aparece un 5, por lo que en donde debería estar la (Azul, Amarillo) encontramos la (Blanco, Negro). Además, en el vector de orientaciones vemos en la entrada 2 un 0, por lo que el color 0 de (Blanco, Negro) cae en la posición 0 de (Azul, Amarillo). Tenemos Blanco donde debería ir Azul, y Negro donde debería ir Amarillo. De esta manera el vector de estado $(\rho, \vec{x}, \tau, \vec{y})$ representa el siguiente cubo:

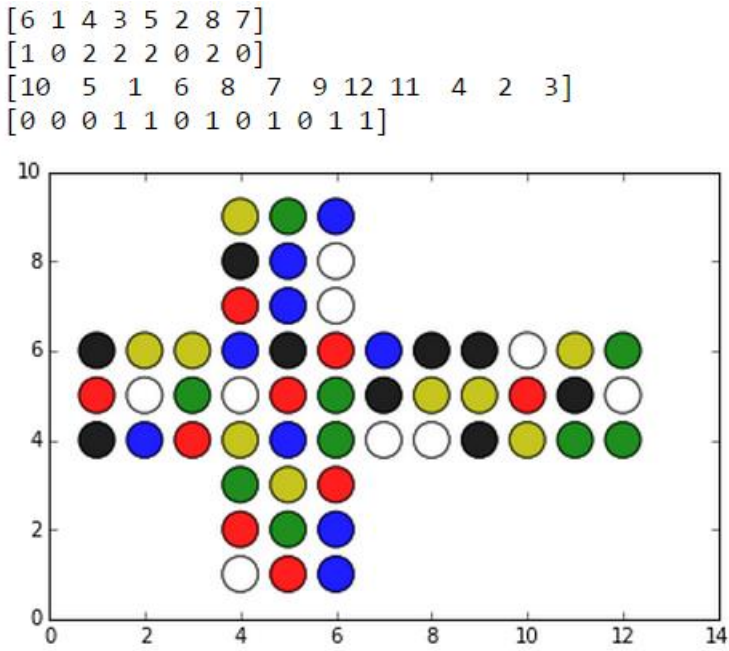


Figura 2: Ejemplo de cubo

B) De representación gráfica a tupla

Hagamos el mismo análisis pero inverso, es decir, representar el vector de estado partiendo de la imagen:

Comencemos, por conveniencia visual en la imagen, con el vértice 7. En ella se encuentra el vértice con estampas (Verde, Rojo, Blanco). Estos colores corresponden al vértice 8 (observamos esto al buscar qué esquina es la que comparten estos centros y luego identificamos su número con ayuda de la tabla). Entonces, $\rho(7) = 8$, i.e., en la séptima posición aparece el 8. Veamos que el color 0 (Verde) cae en el que debería ser Rojo, quien tiene NDO 2. Por lo tanto $v(7) = 2$.

Para terminar, veamos que en la posición de la arista 12 (Verde, Rojo) tenemos la arista (Azul, Amarillo), esto es, la arista 2. Finalmente, el color 0 de la arista 2 (Azul) no está en la posición 0 de la 12 (Verde), por lo que la orientación es 1.

Claramente nuestro modelo tiene una mayor cantidad de configuraciones posibles que la que tiene en realidad un cubo físico. Estamos tomando en cuenta también las posiciones ilegales del cubo: aquellas que son imposibles de alcanzar con movimientos válidos (más adelante veremos un modelo para esto). Por ejemplo, es fácil ver que el vector de estado $(\epsilon_8, (0, \dots, 0), \epsilon_{12}, (1, 0, \dots, 0))$ (ϵ_i es la función identidad en S_i) es imposible de alcanzar, aunque sí cumple con la definición. Primero daremos una definición informal del conjunto que caracteriza al cubo Rubik, RC_3 . Luego pasaremos a formalizarla.

Definición. *Conjuntos válido e inválido del cubo.*

Al conjunto $RC_3^* = S_8 \times \mathbb{Z}_3^8 \times S_{12} \times \mathbb{Z}_2^{12}$ se le llama **conjunto inválido** del cubo.

Al conjunto $RC_3 = \{X \in RC_3^* : X \text{ es una configuración legal de cubo}\}$ se le llama **conjunto válido** del cubo.

Es fácil ver que, aún siendo esto una presentación informal de estos conjuntos, $RC_3 \subset RC_3^*$; de hecho, $|RC_3| = 8!12!3^8 2^{12} \approx 4,3 \times 10^{19} < 5,1 \times 10^{20} \approx 2^{27} 3^{14} 5^3 7^2 11 = |RC_3^*|$ (más adelante probaremos esto). Es entonces lógico preguntarse ¿cómo caracterizar únicamente las configuraciones legales del cubo?

No solo eso, nosotros no intentamos engañar a nadie: es claro que RC_3^* es un grupo, puesto que $S_8 \times \mathbb{Z}_3^8 \times S_{12} \times \mathbb{Z}_2^{12}$ forma una estructura algebraica con las operaciones respectivas de cada elemento en el producto cartesiano (operando componente a componente). Empero, hasta ahora no hemos dado como tal la definición de un grupo, junto con la operación respectiva. Primero un poco más de información:

Los movimientos del Cubo: La Notación de Singmaster

La siguiente tabla muestra la notación más utilizada para representar algunos de los movimientos del cubo, así como el efecto que estos tienen en el cubo mismo. Más aún, a estos seis movimientos les llamaremos de ahora en más **movimientos fundamentales**.

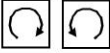




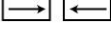
notation (Singmaster)	pictorial (view from front)	description of basic move (clockwise/counterclockwise refers to viewing the face straight-on)
F, F^{-1}		F = quarter turn of front face in the clockwise direction. F^{-1} = quarter turn of front face in the counterclockwise direction.
B, B^{-1}		B = quarter turn of back face in the clockwise direction. B^{-1} = quarter turn of back face in the counterclockwise direction.
R, R^{-1}		R = quarter turn of right face in the clockwise direction. R^{-1} = quarter turn of right face in the counterclockwise direction.
L, L^{-1}		L = quarter turn of left face in the clockwise direction. L^{-1} = quarter turn of left face in the counterclockwise direction.
U, U^{-1}		U = quarter turn of up face in the clockwise direction. U^{-1} = quarter turn of up face in the counterclockwise direction.
D, D^{-1}		D = quarter turn of down face in the clockwise direction. D^{-1} = quarter turn of down face in the counterclockwise direction.

Figura 3: Movimientos fundamentales o rotaciones

Siguiendo la misma notación de Singmaster, si consideramos a $X \in \{R, L, U, D, F, B\}$, X^n , con $n \in \mathbb{N}$, es equivalente a realizar n veces el movimiento fundamental X . A su vez X^{-1} , es el movimiento fundamental X en sentido contrario.

Omitiremos la demostración de los dos siguientes resultados, pero son muy sencillos de probar y es posible hallar las pruebas en la bibliografía.

Olvidándonos un poco del modelo que hemos realizado de las configuraciones, es fácil ver que los movimientos del cubo Rubik son un grupo, al que denotaremos, a falta de imaginación, como (\mathbb{G}, \cdot) . Aquí, la operación " \cdot " para $R, L \in \{X : X \text{ es un movimiento del cubo Rubik}\}$ $R \cdot L$ representa realizar la sucesión de movimientos R y luego L sobre el cubo.

Teorema. $\mathbb{G} = \langle D, U, R, L, F, B \rangle$

Para dar el paso importante y llegar a los temas centrales que nos conciernen, primero revisemos una definición más:

Definición. Una **acción derecha de un grupo** $(G, *)$ en un conjunto A (no vacío), es una función $\circ : A \times G \rightarrow A$, tal que:

1. $(a \circ g_1) \circ g_2 = a \circ (g_1 * g_2)$, para toda $g_1, g_2 \in G$, y para toda $a \in A$
2. $a \circ e = a$, para toda $a \in A$, con e , neutro en G .

Cuando tenemos una acción de un grupo G en un conjunto A , decimos simplemente que G actúa en A .

Y por fin, el teorema o definición formal de RC_3 .

Teorema. El Grupo del Cubo Rubik RC_3 El grupo (\mathbb{G}, \cdot) actúa sobre el conjunto $RC_3^* = S_8 \times \mathbb{Z}_3^8 \times S_{12} \times \mathbb{Z}_2^{12}$ a través de los siguientes mapeos:

$$D, U, R, L, F, B : RC_3^* \times \mathbb{G} \rightarrow RC_3^*$$

, donde, para $x_0 = (\rho, (x_1, \dots, x_8), \tau, (y_1, \dots, y_{12})) \in RC_3^*$, arbitrario.

$$x_0 \circ D = ((6 \ 7 \ 8 \ 5)\rho, (x_1, x_2, x_3, x_4, x_8, x_5, x_6, x_7), \\ (10 \ 11 \ 12 \ 9)\tau, (y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_{12}, y_9, y_{10}, y_{11}))$$

$$x_0 \circ U = ((4 \ 3 \ 2 \ 1)\rho, (x_2, x_3, x_4, x_1, x_5, x_6, x_7, x_8), \\ (4 \ 3 \ 2 \ 1)\tau, (y_2, y_3, y_4, y_1, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}))$$

$$x_0 \circ R = ((3 \ 7 \ 6 \ 2)\rho, (x_1, x_6 + 1, x_2 + 2, x_4, x_5, x_7 + 2, x_3 + 1, x_8), \\ (7 \ 10 \ 6 \ 2)\tau, (y_1, y_6 + 1, y_3, y_4, y_5, y_{10} + 1, y_2 + 1, y_8, y_9, y_7 + 1, y_{11}, y_{12}))$$

$$x_0 \circ L = ((5 \ 8 \ 4 \ 1)\rho, (x_4 + 2, x_2, x_3, x_8 + 1, x_1 + 1, x_6, x_7, x_5 + 2), \\ (5 \ 12 \ 8 \ 4)\tau, (y_1, y_2, y_3, y_8 + 1, y_4 + 1, y_6, y_7, y_{12} + 1, y_9, y_{10}, y_{11}, y_5 + 1))$$

$$x_0 \circ F = ((4 \ 8 \ 7 \ 3)\rho, (x_1, x_2, x_7 + 1, x_3 + 2, x_5, x_6, x_8 + 2, x_4 + 1), \\ (8 \ 11 \ 7 \ 3)\tau, (y_1, y_2, y_7, y_4, y_5, y_6, y_{11}, y_3, y_9, y_{10}, y_8, y_{12}))$$

$$x_0 \circ B = ((2 \ 6 \ 5 \ 1)\rho, (x_5 + 1, x_1 + 2, x_3, x_4, x_6 + 2, x_2 + 1, x_7, x_8), \\ (6 \ 9 \ 5 \ 1)\tau, (y_5, y_2, y_3, y_4, y_9, y_1, y_7, y_8, y_6, y_{10}, y_{11}, y_{12}))$$

Luego, si tomamos a la identidad $\vec{\epsilon} = (\epsilon_8, (0, \dots, 0), \epsilon_{12}, (0, \dots, 0)) \in RC_3^*$, el conjunto válido del cubo se define como:

$$RC_3 = \{\vec{\epsilon} \circ M \mid M \in \mathbb{G}\}$$

Definimos la operación binaria “ $*$ ” sobre RC_3 dada por:

Para $y_0 = \vec{\epsilon} \circ M$, para alguna $M \in \mathbb{G}$

Para $z_0 = \vec{\epsilon} \circ N$, para alguna $N \in \mathbb{G}$,

$y_0 * z_0 = \vec{\epsilon} \circ (M \cdot N) \in RC_3$

Unas cuantas observaciones antes de la prueba: en el teorema anterior se escribe la composición de las permutaciones como la simple concatenación de éstas, omitiendo el símbolo; además, éstas están escritas como ciclos. Por último, recordemos que la sumas se realizan en módulo 3 o 2, donde sea pertinente.

Intuitivamente lo que el teorema anterior nos dice es que cualquier vector de estado puede ser alcanzado a través de una sucesión finita de rotaciones válidas aplicadas al vector de estado inicial. Más aún, “multiplicar” dos vectores de estado válidos es equivalente a realizar en ese mismo orden ambas sucesiones de movimientos, partiendo del estado inicial.

Demostración. Primero observemos que $\mathbb{G} = \langle D, U, R, L, F, B \rangle$, por lo que solo es necesario establecer el mapeo con estos seis movimientos, y con ello decir que \mathbb{G} actúa sobre todo el conjunto RC_3^* .

Consideremos una configuración arbitraria $x_0 = (\rho, \vec{x}, \tau, \vec{y}) \in RC_3^*$, y el movimiento $M_1 \in \mathbb{G}$. Al aplicárselo obtenemos una nueva configuración o vector de estado, digamos $x_0 \circ M_1 \in RC_3^*$. Si tenemos un nuevo movimiento de cubo, $M_2 \in \mathbb{G}$, y se vuelve a aplicar a la última configuración obtenida, obtendríamos una nueva: $(x_0 \circ M_1) \circ M_2 \in RC_3^*$, pero esto equivale a realizarle a la configuración x_0 el movimiento $(M_1 \cdot M_2) \in RC_3^*$, es decir, $x_0 \circ (M_1 \cdot M_2) \in RC_3^*$. Ahora, consideremos el **movimiento nulo** (la identidad en \mathbb{G}), tal que al operarlo con cualquier vector de estado obtendríamos ese mismo vector, es decir: $x_0 \circ I = x_0$. Concluimos que \circ es una acción de grupo. \square

Para probar que RC_3 es grupo, usaremos el siguiente teorema que da una caracterización de todos los elementos que están en él.

Teorema. Primer Teorema Fundamental del Cubo

Un vector de estado $(\rho, v, \sigma, w) \in S_8 \times \mathbb{Z}_3^8 \times S_{12} \times \mathbb{Z}_2^{12} = RC_3^*$ representa una configuración legal del cubo si, y solo si, satisface las siguientes propiedades:

- $sign(\rho) = sign(\sigma)$,
- $v_1 + v_2 + \dots + v_8 \equiv 0 \pmod{8}$,
- $w_1 + w_2 + \dots + w_{12} \equiv 0 \pmod{2}$.

Notación: $sign(\cdot)$ se refiere a la paridad de la permutación

Demostración. (i) Primero tenemos que las tres condiciones son necesarias. Para esto vamos a probar que se satisfacen en el estado original (resuelto) del cubo, y que se mantienen bajo los 6 movimientos básicos R, L, U, D, F y B . Cuando el cubo está resuelto le corresponde el vector $(\varepsilon, \vec{0}, \varepsilon, \vec{0})$ y las tres condiciones se satisfacen. Cada permutación en los seis movimientos básicos es un ciclo de orden 4 (par) y tiene signo -1 . La suma de los componentes de cada vector de orientación de las esquinas es 0 o 6, que son divisibles entre 3. La suma de los componentes de cada vector de orientación de aristas es 0 o 4 que es divisible entre 2.

Si X es una configuración legal con vector de posición (ρ, v, σ, w) que satisface las tres condiciones, entonces después de aplicarle uno de los seis movimientos básicos, estas condiciones todavía se cumplen:

- La primera se cumple ya que cada movimiento causa al mismo tiempo ciclos de orden 4 en los cubículos de las esquinas y las aristas, por lo tanto de misma paridad.

- La segunda se cumple porque U y D no cambian nada, mientras que R, L, F y B causan simultáneamente que dos componentes aumenten en $1(\text{mod } 3)$ y dos componentes se reduzcan en $1(\text{mod } 3)$.
- La tercera se cumple por la misma lógica. Los movimientos U, D, F y B no cambian nada, mientras que R y L incrementan y reducen dos componentes cada uno en $1(\text{mod } 2)$.

Dado que toda configuración legal del cubo se obtiene aplicando estos movimientos al cubo resuelto, entonces las tres propiedades se cumplen para cada configuración legal.

(ii) Para probar que las tres condiciones son suficientes, tenemos que probar que cualquier vector de estado (ρ, v, σ, w) que las satisface puede resolverse usando los seis movimientos básicos. Usando el método por capas, se tiene este resultado:

Sea X una configuración correspondiente a (ρ, v, σ, w) . Primero, sin pérdida de generalidad asumimos que ρ y σ tienen paridad igual a 1 (en caso contrario solo aplicamos un movimiento de la cara y quedarían como queremos). Esto significa que la permutación de esquinas es par, por lo que podríamos regresarlas a sus ubicaciones iniciales usando un ciclo de movimientos de orden 3. Análogamente para las aristas. Ya que podemos aplicar cualquier ciclo de orden 3 en esquinas o aristas, podemos regresar cualquier configuración a su posición original. Llamemos esta nueva configuración X' . Como los movimientos básicos conservan las condiciones del primer teorema fundamental, el vector de estado (ρ', v', σ', w') de X' también las satisface y en este caso tenemos que $\rho' = \varepsilon, \sigma' = \varepsilon$.

Ahora falta mostrar que podemos voltear los cubículos a sus orientaciones correctas.

La tercera condición nos indica que tenemos que voltear un número par de aristas. Como tenemos series de movimientos para voltear cualquier par de aristas, entonces esta parte la podemos resolver. Para las esquinas, la segunda condición nos dice que el número de vueltas en el sentido de las manecillas que tenemos que darles es igual al número de vueltas en sentido contrario módulo 3. Así que primero volteamos cualquier par en los dos sentidos a sus orientaciones originales. El resultado va a ser que todas las vueltas que nos faltan van a ocurrir en tripletes del mismo sentido. Estos se pueden resolver usando una secuencia de movimientos de esquinas que ya conocemos.

Podemos concluir que X es una configuración que podemos resolver. \square

A merced del teorema anterior (la parte (i) de la demostración), podemos caracterizar más fácilmente al conjunto de estados válidos del cubo: $RC_3 = \{x \in RC_3^* : x \text{ cumple con el Teorema Fundamental del Cubo I}\}$, y que éste sea grupo con la operación proporcionada más arriba: las propiedades de grupo simplemente se “heredan” del grupo de movimientos del cubo, ya que \mathbb{G} es asociativo, tiene inversos, y neutros.

Dada la explicación anterior, podemos fácilmente llegar al siguiente corolario:

Teorema. *El tamaño del grupo del cubo*

El número de posiciones legales e ilegales es:

- $|RC_3| = \frac{|RC_3^*|}{12} = 2^{27} \cdot 3^{14} \cdot 5^3 \cdot 7^2 \cdot 11 \simeq 4,3 \cdot 10^{19}$.
- $|RC_3^*| = 8! \cdot 12! \cdot 3^8 \cdot 2^{12}$.

Demostración. Ya que $RC^*_3 = S_8 \times S_{12} \times \mathbb{Z}_3^8 \times \mathbb{Z}_2^{12} \Rightarrow |RC^*_3| = |S_8||S_{12}||\mathbb{Z}_3^8||\mathbb{Z}_2^{12}| = 8! \cdot 12! \cdot 3^8 \cdot 2^{12}$ Para las posiciones legales, este número se reduce en:

- La mitad por la primera condición del primer teorema fundamental, ya que habrá el mismo número de permutaciones pares que impares.
- Un tercio por la segunda condición, ya que la orientación de 7 esquinas se puede escoger arbitrariamente, pero esto determina la orientación de la esquina restante.
- La mitad por la tercera condición con el mismo razonamiento aplicado a las 12 aristas.

Concluimos que $|RC_3| = \frac{|RC^*_3|}{12}$. □

Replantando las proposiciones anteriores viéndolas como sucesiones de movimientos, tenemos inmediatamente lo siguiente:

Teorema. Segundo Teorema Fundamental del Cubo

Una secuencia de movimientos es posible si, y solo si, se satisfacen las siguientes tres proposiciones:

- *La permutación de las esquinas y las aristas (como producto) es par.*
- *El número de esquinas que giramos en el sentido de las manecillas es el mismo que el número de esquinas que giramos en sentido contrario mod 3.*
- *El número de aristas invertidas es par.*

Corolario. *Cada una de las siguientes configuraciones son imposibles de obtener a partir del estado resuelto del cubo mediante movimientos legales:*

1. *Exactamente dos esquinas intercambian lugares.*
2. *Exactamente dos aristas intercambian lugares.*
3. *Exactamente una arista se invierte.*
4. *Exactamente una esquina se voltea.*
5. *Exactamente dos esquinas se voltean en la misma dirección.*

Gracias a todo el trabajo anterior, podemos simplemente referirnos a un vector de estado $g_0 \in RC_3$, siendo generado por los movimientos correspondientes (no únicos) que llevaron el vector de estado resuelto en RC_3 a g_0 . Por esta razón, para la presentación del siguiente algoritmo, simplemente nos referiremos a $g_0 \in \mathbb{G}$, haciendo referencia a que para este vector existe $M \in \mathbb{G}$ tal que $g_0 = \vec{\epsilon} \circ M$.

El Algoritmo de Thistlethwaite

El siguiente método ha tenido muchas repercusiones importantes para la investigación y diseño de algoritmos de solución del cubo Rubik. Cabe destacar su importancia y belleza teórica que, por su propia naturaleza, lo invalidan para el uso humano, no así para las computadoras.

0.1234567891011121314151617181920212223242526272829303132333435363738394041

Cuando el algoritmo ha sido mejorado e implementado con herramientas más sofisticadas de programación se puede lograr una solución en un promedio de 20 movimientos.

La idea detrás del algoritmo es dividir el problema de resolución del cubo en problemas mucho más sencillos de completar (de nuevo, por una máquina). Para esto se va haciendo un “escalamiento” entre cada subgrupo, restringiendo cada vez más el número de movimientos permitidos.

Los subgrupos son:

- $G_0 = \mathbb{G} = \langle L, R, F, B, U, D \rangle$, todo el grupo del cubo está aquí.
- $G_1 = \langle L, R, F^2, B^2, U, D \rangle$, subgrupo de G_0 ,
- $G_2 = \langle L^2, R^2, F^2, B^2, U, D \rangle$, subgrupo de G_1 ,
- $G_3 = \langle L^2, R^2, F^2, B^2, U^2, D^2 \rangle$, subgrupo de G_2 ,
- $G_4 = \{e\}$, el cubo resuelto.

$G_i, 1 \leq i \leq 4$ corresponde a todos los estados del cubo que pueden ser alcanzados usando los movimientos en su conjunto generador.

Intuitivamente lo que queremos hacer es ir maniobrando a lo largo de la cadena de subgrupos $G_0 > G_1 > G_2 > G_3 > G_4$. Supongamos un caso ideal en el que el rompecabezas fue **bien encriptado** (revuelto), y que solo podemos llegar a este estado usando todos los seis movimientos que generan a G_0 . Para pasar a G_1 , usamos los movimientos en el conjunto generador de G_0 , a saber, L, R, F, B, U, D . Una vez que llegamos a G_1 , usando únicamente movimientos en el conjunto generador de G_1 (L, R, F^2, B^2, U, D), pasamos a G_2 ; y así, sucesivamente hasta llegar a G_4 : el cubo resuelto.

Ahora bien, para realizar la transición entre cada subgrupo, nos fijamos en las clases laterales. En álgebra moderna se demuestra que, teniendo cualquier subgrupo de un grupo, las clases laterales forman una partición del grupo. Thistlethwaite primeramente diseñó unas tablas que representaran las clases laterales; y para cada elemento (cada clase lateral) en cada grupo cociente (o en cada tabla) halló la serie de movimientos que te llevarían dicho elemento al subgrupo más pequeño.

Lo anterior lo podríamos formalizar de la siguiente manera:

Sea $g_0 \in G_0$, un estado bien encriptado del rompecabezas.

Transición de G_0 a G_1

- Ya que $G_1 < G_0$, podemos formar una partición de G_0 , a partir de la unión de las clases laterales, es decir, $G_0 = \bigcup_{i=1}^{r_0} g_{i,0}G_1$, con $g_{i,0} \in G_0$, para toda $1 \leq i \leq r_0$, para alguna $r_0 > 1$.
- Luego, existe $x_1 \in G_0$, tal que $g_0 \in x_1G_1$, o bien, $g_0 = x_1g_1$, para alguna $g_1 \in G_1$.

Se sigue que $g_1 = x_1^{-1}g_0 \in G_1$. Nótese que para llegar a G_1 , solo se usan movimientos en G_0 , pues $x_1^{-1} \in G_0$.

Transición de G_1 a G_2

0.1234567891011121314151617181920212223242526272829303132333435363738394041

- Ya que $G_2 < G_1$, podemos formar una partición de G_1 , a partir de la unión de las clases laterales, es decir, $G_1 = \bigcup_{i=1}^{r_1} g_{i,1}G_2$, con $g_{i,1} \in G_1$, para toda $1 \leq i \leq r_1$, para alguna $r_1 > 1$.
- Luego, existe $x_2 \in G_1$, tal que $g_1 \in x_2G_2$, o bien, $g_1 = x_2g_2$, para alguna $g_2 \in G_2$.

Se sigue que $g_2 = x_2^{-1}g_1 = x_2^{-1}x_1^{-1}g_0 \in G_2$. De nueva cuenta, solo usamos movimientos en G_1 , para llegar a G_2 , puesto que $x_2^{-1}x_1^{-1} \in G_1$.

Transición de G_2 a G_3

- Ya que $G_3 < G_2$, podemos formar una partición de G_2 , a partir de la unión de las clases laterales, es decir, $G_2 = \bigcup_{i=1}^{r_2} g_{i,2}G_3$, con $g_{i,2} \in G_2$, para toda $1 \leq i \leq r_2$, para alguna $r_2 > 1$.
- Luego, existe $x_3 \in G_2$, tal que $g_2 \in x_3G_3$, o bien, $g_2 = x_3g_3$, para alguna $g_3 \in G_3$.

Se sigue que $g_3 = x_3^{-1}g_2 = x_3^{-1}x_2^{-1}x_1^{-1}g_0 \in G_3$.

Transición de G_3 a $G_4 = \{e\}$

- Consideremos a g_3^{-1} , inverso de g_3 . Simplemente, $e = g_3^{-1}g_3 \in G_4$. Por último, redefinimos: $x_4 = g_3$,

y se sigue que $e = x_4^{-1}x_3^{-1}x_2^{-1}x_1^{-1}g_0$

Conclusión

El empleo de la estructura algebraica tiene dos finalidades muy claras y significativas. En primer lugar, la posibilidad de contar los estados posibles del cubo y darnos una idea de lo ingenuo que puede ser emplear métodos exhaustivos para la solución de este rompecabezas. La estructura por esquinas, aristas y centros puede extenderse y formar otros tipos de piezas, las mismas que son necesarias para pensar en cubos de 5×5 o 7×7 . Este tema es abordado con más detalle en el iPython Notebook. En segundo, la resolución del cubo Rubik. Al emplear el algoritmo Thistlethwaite es indispensable utilizar el hecho de que un grupo puede ser particionado en sus clases laterales. Es asombroso lo idónea que resulta la estructura para resolverlo. Da la impresión de ser *ad hoc*, si bien no es así. Ahí el sentido de nuestro epígrafe: el álgebra no solo nos da una forma bella de representar usando plenamente sus resultados teóricos, sino que por su generosidad nos revela multitud de propiedades interesantes, la posibilidad de extender nuestro modelo, contar y finalmente resolver de forma eficiente. El algoritmo fue implementado en iPython; el lector programador curioso siéntase libre de mandarnos un correo electrónico a mcarranzabar@gmail.com para revisar el iPython Notebook.

Agradecemos a la profesora Marcela González Peláez, del Departamento Académico de Matemáticas del ITAM, por sus valiosos comentarios para la realización de este artículo.

Referencias

- [1] Mulholland, Jamie. *Permutation Puzzles: A Mathematical Perspective*. Lecture Notes, 2013
- [2] Fraleigh, John B. *A First Course in Abstract Algebra*. Pearson, 2003
- [3] Joyner, W.D. *Mathematics of the Rubik's cube*. Lecture Notes, 2010



Del ITAM a Viena

Entrevista con Juan Pablo Aguilera

Laberintos & Infinitos (L&I): Cuéntanos un poco de ti

Juan Pablo Aguilera (JPA): Estudié Matemáticas Aplicadas en el ITAM, formé parte de la revista Laberintos & Infinitos, fui miembro de la representación de mi carrera “Aleph” y realicé mi tesis sobre cómo interpretar lógicas de demostrabilidad en espacios topológicos con el profesor David Fernández Duque. Actualmente llevo un año haciendo mi doctorado en Viena.

L&I: ¿Tienes algún tipo de beca mientras estudias el doctorado?

JPA: Estoy contratado como investigador por la universidad, lugar donde me desempeño en diversos temas, tales como lógica, y otros en los que me interesa adentrarme como el “Análisis Ordinal”, que consiste en intentar asignar a diferentes teorías matemáticas una norma que muestre qué tan fuerte son con respecto a otras teorías. Por ejemplo, una teoría debería ser más fuerte que otra si puede probar más cosas.

L&I: ¿El análisis ordinal emplea algún tipo de métrica?

JPA: No es una métrica lo que se emplea, sino algo llamado números ordinales, estos son una extensión de los números naturales dónde vas contando uno, dos, tres para luego llegar al primer número infinito que se llama omega, luego sigues omega más uno, omega más dos para luego llegar a dos copias de los naturales, de ahí se siguen omega dos veces, omega omega veces, omega a la omega, el límite de eso se llama épsilon cero, y ese es la norma de la teoría de números que va ligado al principio de inducción, si limitas inducción a cierto tipo de fórmulas obtienes una norma más pequeña.

L&I: ¿El tema lo aprendiste por optativas impartidas en el ITAM o cómo lograste desarrollar un interés en él?

JPA: Creo que juntándome con profesores, particularmente de la mano del profesor David Fernández Duque fue que desarrollé el gusto por este tema ya que no había optativas que abordaran el análisis ordinal. Comenzó a disminuir el trabajo sobre el tema hace algunos años pero está resurgiendo.

L&I: ¿A qué te refieres con que está “resurgiendo”?

JPA: Me refiero a que el proyecto de análisis ordinal empezó a morir hace unos 20 años ya que se estaba volviendo muy complicado, y la gente no podía avanzar dado que toma mucho tiempo aprender lo que ya se había hecho. Sin embargo, desde hace poco se han ido reconstruyendo los resultados que ya se tenían con técnicas más sencillas.

L&I: ¿Qué se busca lograr con la investigación que estás realizando?

JPA: La idea es tener una clara noción de que principios estás usando para demostrar qué cosa, me refiero a que si puedes demostrar algo; saber qué es lo mínimo que necesitas para hacerlo. Otro problema es que sabemos, desde hace casi cien años, que en cualquier teoría matemática decente se tienen enunciados que no se pueden demostrar o decidir, al principio la existencia de estos enunciados se entendía como un accidente teórico, pero con el tiempo se encuentran más y más de ellos, por lo que se ha vuelto un problema cuando alguien quiere trabajar en algo y de pronto se encuentra con que todo es indecidible, no puede demostrar nada porque nada se puede refutar o demostrar. Una idea razonable sería tomar axiomas que parezcan naturales para completar las teorías con las que se trabajan, pero no es muy claro cómo debería abordarse eso, porque si algo es independiente podrías tomar eso o su negación como un axioma. Entonces hay que ver cómo es el panorama de los axiomas, cuál implica cuál e ir viendo los resultados de implicación de ellos como evidencia de lo que resulta más natural aceptar. Por ejemplo, hay axiomas que dicen que existen ciertos conjuntos muy grandes, llamados de “grandes cardinales”, que parecen estar muy retirados de las matemáticas mundanas, pero si supones que existen esos conjuntos muy grandes entonces obtienes propiedades de los números reales; se puede demostrar que sin la existencia de esos conjuntos, no se pueden demostrar las propiedades.

L&I: ¿Hay algún otro proyecto que tengas entre manos o que te interese desarrollar?

JPA: Hay un proyecto que estoy empezando a estudiar, va de la mano con el tema anterior pero a una escala mucho más grande, con teorías más fuertes. La idea es formar modelos para teorías matemáticas muy fuertes que sean muy canónicos. Por ejemplo, un modelo para teoría de números son los números naturales, todo lo que la teoría de números prueba es verdadero acerca de los números naturales. Un modelo para la teoría de análisis son los números reales porque lo que pruebas en análisis es cierto acerca de los números reales. Sin embargo, si se estudia la teoría de conjuntos no queda muy claro con qué estás trabajando, La idea no es construir modelos donde todo lo que pruebes sea cierto y viceversa (de hecho no es posible, por el teorema de Gödel). La idea es construir modelos que satisfagan todos los enunciados demostrables en la teoría y que además sean lo más canónicos posibles. Por ejemplo, existen muchos modelos, no estándares, de la teoría de números (“modelos” se refiere a estructuras que satisfacen los axiomas de la teoría, en el mismo sentido que \mathbb{Q} satisface los axiomas de la teoría de grupos), pero los números naturales son el modelo canónico. En particular, cualquier modelo debe contenerlos. La idea es algo similar. Otro ejemplo es el modelo de la teoría de conjuntos de Gödel (se llama L): todo otro modelo debe contenerlo. Más precisamente, una teoría T prueba que una teoría U es consistente si T prueba que existe un modelo para U . Por ejemplo, usando conjuntos puedes construir los números naturales, que forman un modelo para la teoría de números, o los complejos, que forman un modelo para la teoría de campos algebraicamente cerrados. Por lo tanto, la teoría de conjuntos prueba la consistencia de la teoría de números y la teoría de campos algebraicamente cerrados.

L&I: ¿Cuál es tu siguiente paso o qué te llama la atención después de estos temas?

JPA: Creo que el siguiente paso es adentrarme, en mayor medida, en cada uno de los temas antes mencionados. La evolución de los proyectos es como subir una escalera, no porque vaya de poco en poco, sino porque va subiendo.

L&I: ¿Tu vocación es la investigación?

JPA: Sin duda. En particular la investigación sobre los dos temas anteriores, los cuales son muy cercanos en espíritu pero con técnicas muy distintas que me van a tomar bastante tiempo.

L&I: Aproximadamente ¿cuánto tiempo toma estar a la par en cuánto a información de la investigación que está realizando?

JPA: No estoy seguro, pero el último resultado que tienen en este segundo tema denominado “teoría de modelos internos” de las teorías muy fuertes está en un artículo ya escrito que todavía no se publica, el cual consiste de 700 páginas.

L&I: ¿El ITAM te preparó bien para lo que estás enfrentando o sientes que le falta algo?

JPA: A la carrera le faltan muchas materias abstractas, por ejemplo Lógica, no vemos casi nada de eso y tuve que aprenderlo por mi cuenta, esta es un área bastante grande y no es algo rápido de estudiar. En lo general, alguien que vaya a otras áreas de Matemáticas puede que tenga problemas porque no vemos nada de Topología, no vemos cosas de Geometría Diferencial o Geometría Algebraica. En lo referente a Topología sé que es muy importante, ya que se usa en casi todas las áreas y lo vemos muy poco, yo lo tuve que aprender para mi tesis. Ayuda mucho el hecho de que llevemos Análisis Matemático y recomendaría que si hay materias al respecto los alumnos intenten llevarlas para mejor su comprensión sobre todos estos temas. A veces es fácil ignorar algunas habilidades que te deja el ITAM con los cursos de Estudios Generales, por los menos Ideas y Problemas, que te ayudan a aprender a expresarte con fluidez y exponer tus pensamientos de manera clara.

L&I: ¿Cuáles son las mejores herramientas que dejó el ITAM para tu vida como investigador?

JPA: Quizá ser ordenado a la hora de administrar mi tiempo y decidir la forma en que debo acomodar mis actividades diarias. Son cosas que tal vez no nos hubiera dejado una universidad que no te presione todo el tiempo.

L&I: ¿Por qué decidiste estudiar Matemáticas Aplicadas?

JPA: Originalmente me interesaba mucho la Economía por lo que inicié en el plan conjunto de Matemáticas Aplicadas y Economía, pero después me empezó a gustar menos la Economía, y después menos Matemáticas Aplicadas y más Matemáticas Puras. Creo que desde primer semestre, más específicamente desde que empezamos a llevar Álgebra, donde vemos conjuntos, me llamaron la atención los temas matemáticos más abstractos. Sentía que había mucho por explorar y que no estábamos viendo en clase. De cualquier forma en la biblioteca hay algunos libros al respecto.

L&I: ¿No pensaste estudiar Matemáticas Aplicadas en el ITAM y Matemáticas Puras en la UNAM?

JPA: Sí lo pensé de hecho alguna vez hablé con el doctor Carlos Bosch y me dijo que a esas alturas no valía la pena, que lo mejor que podía hacer era terminar rápido e irme a estudiar un posgrado en el extranjero. Yo creo que fue un buen consejo porque ya en el posgrado puedes estudiar justo lo que a ti te interesa, además de que un posgrado es un nivel más alto por lo que avanzas más rápido.

L&I: ¿Cuáles fueron los cursos que más te gustaron y los que menos durante la carrera?

JPA: Lo que más me gustó fueron las clases de Cálculo Numérico con José Luis Morales, el Seminario de Teoría de Decisión con Levent Ülkü e Ideas con Javier García Salcedo. Lo que no me gustó en absoluto fueron las materias de Historia y Problemas de la Realidad Mexicana ya que, aún cuando Ideas y Problemas me agradaron bastante, a estas últimas dos no les vi mucho el caso. Tampoco me gustó que en varias materias avanzadas de Computación y de Estadística no se emplearan muchas de las herramientas matemáticas que ya nos habían dado otras clases.

L&I: ¿Recomendarías a un joven que estudiara Matemáticas Aplicadas en el ITAM?

JPA: Sí lo recomendaría porque las clases de matemáticas son buenas y te preparan bien. Además de eso tiene un ritmo muy rápido y la gente de aquí siempre es muy activa, eso te ayuda para que seas más activo tú, creo que avanzas muy rápido si sientes que tienes que ir rápido todo el tiempo. Llevamos un plan de estudios de seis materias por semestre y todo el tiempo debes estar haciendo cosas.

L&I: ¿Cuál fue tu mejor semestre o cuál es el mejor recuerdo que guardes del ITAM?

JPA: Mi primer semestre fue muy bueno porque estaba muy determinado, muy dispuesto a aprender y trabajé mucho. Pero los últimos semestres también fueron muy buenos porque llevaba menos materias y estaba trabajando en la tesis, la cual me gustaba mucho.

L&I: ¿Cómo fue tu participación en la revista de Laberintos Infinitos?

JPA: Primero empecé respondiendo una pregunta de Erdős; por cierto, invito a la gente a que las respondan porque la mayoría de las veces nadie las intenta y muchas veces son menos difíciles de lo que parecen. Después escribí unos artículos, que viéndolo en retrospectiva eran para mí porque eran sobre temas que yo estaba aprendiendo y los escribía para organizarme pues todo lo hacía por mi cuenta y en ese momento no tenía otras fuentes, tenía que ir escribiendo para procesar todo bien y después me decidía enviarlo a la revista. Tiempo después me uní a la revista, estuve trabajando allí, motivando a la gente a escribir y revisando artículos. Creo que pude haber hecho mi servicio social en otro lado, pero disfruté bastante la experiencia en Laberintos & Infinitos.

L&I: ¿Cómo decidiste irte a Viena?

JPA: Originalmente estaba muy decidido a investigar en teoría de la demostración, pero de eso no hay mucho en Estados Unidos, además de que no tenía muchas ganas de irme ahí porque me parecía más interesante Europa, y le pregunté al profesor David Fernández Duque quién me ayudó a buscar diferentes candidatos de los cuáles elegí Viena porque me pareció una ciudad más agradable y era mejor el trato. En Europa, a los estudiantes de doctorado los tratan como investigadores así que perciben un salario.

L&I: ¿Es reconocido el nombre del ITAM en lo referente a Matemáticas Aplicadas a nivel internacional?

JPA: No realmente, si conocen algún lugar es la UNAM, pero quizá en algunos años; sobre todo en lógica que está empezando a haber más itamitas.

L&I: ¿Cuál es tu matemático favorito?

JPA: Gödel es mi matemático favorito, él empezó muchas áreas. En parte, la razón por la que pudo hacerlo es que, una vez que había iniciado una, no la agotaba, que sería la opción sencilla, porque él era el experto en ella. Por el contrario, buscaba resolver un problema distinto.

L&I: ¿Cómo te ves de aquí en 10 años?

JPA: Me veo investigando, espero como profesor en alguna universidad y supongo que tendré algunos intereses distintos, no en el sentido académico sino en el personal, por lo que no me preocuparía tanto por una ciudad tan cultural como Viena sino por una ciudad más tranquila.

L&I: ¿Dónde te gustaría impartir clases en un futuro?

JPA: Quizás en Berkeley, siempre ha sido el centro de la lógica, pero muy probablemente no sería la única opción.

L&I: ¿Qué te dirías a ti mismo el primer día de clases?

JPA: Serían dos consejos, el primero es: no tengas miedo de estudiar más allá de lo que se ve en clase. El segundo consejo sería: estudia matemáticas abstractas porque va a ser muy útil como topología y análisis funcional.

L&I: ¿Algo que quieras agregar?

JPA: Si alguien quisiera estudiar algo fuera del plan de estudios yo recomendaría topología y teoría de conjuntos. Además invito a cualquier persona que esté interesada en lógica y quiera hablar o preguntar algo al respecto a enviarme un correo a jpabloaguilera@live.com.

¿Te gustaría escribir para LABERINTOS & INFINITOS ?

Te invitamos a colaborar con nosotros, sin importar de qué universidad seas, mandándonos tu artículo sobre lo que más te apasione de las matemáticas.

¿Qué es lo que te gusta de ellas?

¿Cuáles son las aplicaciones más interesantes que pueden tener?

¿Quién es tu matemático favorito y por qué?

¿Tienes algún teorema predilecto y quieres compartirlo?

¿Qué representan las matemáticas más allá de los números?

Además, si eres alumno del ITAM, participas para obtener el Premio Cauchy, el cual se otorgará al mejor artículo por número y será determinado por la mesa directiva de la publicación. El ganador será galardonado con una cantidad de \$1000 pesos.

No dejes pasar la oportunidad.

¡Esperamos recibir tus artículos!

Envía tu artículo a
laberintos@itam.mx

