ÍNDICE

1.6180339887498948482045868343656381177203091798057628621354486227052604628189 0

LABERINTOS INFINITOS, Año 2018, No. 46, enero – abril 2018, es una publicación cuatrimestral editada por el Instituto Tecnológico Autónomo de México, a través de las Direcciones de Actuaría y Matemáticas del ITAM. Calle Río Hondo No. 1, Col. Progreso Tizapán, Delegación Álvaro Obregón, C.P. 01080, Tel. 56284000 ext 1732, www.itam.mx, **raulranirezriba@gmail.com**. Editor responsable: Ilan Jinich Fainsod. Reservas de Derechos al Uso Exclusivo No. 04-2016-112313125200-102, ISSN: en trámite, ambos otorgados por el Instituto Nacional del Derecho de Autor, Licitud de Título y contenido en trámite, otorgado por la Comisión Calificadora de Publicaciones y Revistas Ilustradas de la Secretaría de Gobernación. Permiso SEPOMEX en trámite.

Editorial

Si la aritmética es poder contar hasta veinte sin quitarte los zapatos, el álgebra es sumar peras con manzanas, la geometría es darle sentido a los garabatos, la combinatoria es jugar a las sillas musicales sin música, la topología es comer una dona que en realidad es una taza y el cálculo es ver cosas con lupas. Entonces; ¿qué son las matemáticas?

Agradecimientos

A la División Académica de Actuaría, Estadística y Matemáticas del ITAM. En especial a Beatriz Rumbos, Claudia Gómez, César Luis García. A la Dirección Escolar del ITAM, específicamente a Patricia Medina. Gracias a Pearson y Galois, 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.

Dedicatoria

Este número es dedicado a la memoria de Patricia Medina. De parte de todo el equipo editorial y del consejo académico agradecemos la ayuda y el apoyo todos estos años. La maestra Medina fue un pilar fundamental para el desarrollo de nuestra institución y una mujer ejemplar.

ϕ

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

Imagen de portada:
Andrea Harris

1.6180339887498948482045868343656381177203091798057628621354486227052604628189O

Nicolas Bourbaki

Nicolas Bourbaki
Profesor Emérito de École Normale Superior en Paris[1]

"En el centro de nuestro universo se encuentran los grandes tipos de estructuras ... y serán llamados las estructuras madre. ... Más allá de este primer núcleo, aparecen las estructuras que podrían ser llamadas múltiples estructuras. Implican dos o más de las grandes estructuras-madres no en simple yuxtaposición (que no produciría nada nuevo) sino combinado orgánicamente por uno o más axiomas que establecen una conexión entre ellos. ... Más adelante llegamos finalmente a las teorías propiamente llamada particulares. En estas los elementos de los conjuntos considerados, los cuales en las estructuras generales han permanecido totalmente indeterminados, obteniendo una individualidad más definitivamente caracterizada." [2]
Nicolas Bourbaki

Uno de los matemáticos más prolíficos e interesantes del siglo pasado es Nicolas Bourbaki. O bueno, lo sería, si de verdad esa persona existiera como tal. El seudónimo Nicolas Bourbaki fue adoptado por varios matemáticos franceses en 1934 para representar la esencia de un matemático contemporáneo, y eventualmente se volvió un grupo de gran importancia con su principal obra *Elementos de matemática*, un tratado que intentaba redefinir y capturar el conocimiento matemático hasta ese punto en la historia. Hoy en día el grupo Bourbaki, oficialmente conocidos como *Association des collaborateurs de Nicolas Bourbaki*, todavía tiene una oficina en la École Normale Superior en París, en donde el grupo originó.

[1]Laberintos e Infinitos agradece inmensamente a Adrian Tame por su ayuda en realizar el artículo.
[2]Tomado de [7]

Al comienzo del siglo 20, la Academia de ciencias de París ya no era uno de los pilares más importantes del avance matemático como lo había sido por mucha de su historia. Los matemáticos Alemanes habían presentado los más importantes avances en el área durante un tiempo, entre ellos Carl Friedrich Gauss y Bernhard Riemann. La primera guerra mundial además tomo la vida de muchos de los matemáticos de la época, dejando un vació considerable en el campo. Cabe recalcar que los matemáticos franceses de los años treinta no tenían a un gran matemático o área por seguir. Esto cambió cuando André Weil escribió a la academia de ciencias de Paris, sobre Nicolas Bourbaki. Hablando a detalle sobre este importante matemático, como la primera guerra mundial interrumpió su trabajo y tuvo que escapar Paris, y eventualmente regresó, pero sus avances no eran considerados por ningún partido. Como se encontró ocupado por otras cosas en la revolución de 1917, impreso en su trabajo en una institución en Poldavia pero tuvo que emigrar a Iran por la guerra civil, y para sobrevivir daba clases en un café sobre un juego de cartas en el cual era bastante bueno.

Todo esto, una ficción. El nombre y la historia detrás fue imaginado por un pequeño grupo de matemáticos Franceses con el liderazgo de André Weil con el propósito de publicar bajo el, creando un personaje cuyo trabajo entero sería la base rigurosa y esencial de la cual las matemáticas posteriores se podrían basar. Todo esto comenzando con tres matemáticos inconformes con la realidad y el estado de las matemáticas al comienzo de los 30s.

Al salir de una conferencia impartida por Raoul Housson en 1934, los entonces estudiantes Jean Delsarte, Henri Cartan y André Weil quedaron atónitos por la falta de rigurosidad y errores lógicos presentados. El área análisis de esa época en Francia que se enseñaba en las universidades estaba basado principalmente en el trabajo de Jacques Hadamard, Emile Picard y Edouard Goursat, matemáticos de tiempos pasados que habían quedado atrás comparados con muchos de los matemáticos Alemanes y de otros lugares en Europa. Su trabajo fue seminal, pero habían importantes avances recientes que no estaban tomados en consideración para el temario general de análisis. Por lo tanto, en la primera junta del grupo Bourbaki que tomó lugar en un pequeño café parisino, quedó arreglado que se consideraría intentar crear un nuevo trato del análisis, con un punto de vista más moderno y puntual, tomando como base la rigurosidad que ellos consideraban que hacía falta en el área.

En las siguientes juntas del grupo, se definieron importantes cosas como los miembros oficiales del mismo, incluyendo a Henri Cartan, Claude Chevalley, Jean Coulomb, Jean Delsarte, Jean Dieudonné, Charles Ehresmann, Szolem Mandelbrojt, René de Possel, y claro, André Weil. Todos ellos conocidos por haber estudiado en la École Normale Supérieure. Se discutiría también donde y cuando tomarían lugar las juntas y cuales serían los intereses principales de ellas. Mientras que el tratado de análisis era el punto focal y final, no era una hazaña pequeña, y tenía que ser desglosada en varias juntas posteriores. Poco a poco el grupo empezó a generar interés externo, e invitaciones fueron enviadas poco a poco a conocidos o matemáticos importantes que los miembos originales consideraban podrían expandir el grupo de manera positiva. Eventualmente, matemáticos como Samuel Eilenberg , Alexander Grothendieck, Pierre Samuel, Laurent Schwartz, Jean-Pierre Serre, Serge Lang, y Armand Borel se unieron

al grupo, todos trabajando por su parte en sus propios proyectos además de ayudando al grupo Bourbaki con sus metas. Una curiosa parte de la pertenencia del grupo es que a los 50 años de edad, se les pedía a los miembros de este que se retiraran, para mantener un aire de juventud en el grupo y que no hubiera una sola persona con demasiado poder dentro de el.

Por los problemas de la guerra, empezando en 1935 el grupo solamente se juntaba dos o tres veces al año, y se adoptó un método diferente para la creación del libro de análisis. Se les pidió a miembros individuales que construyeran capítulos por su cuenta, para poder luego editarlos y revisarlos en grupo, y generar cohesión entre ellos. Este método fue increíblemente efectivo pero tuvo un imprevisto enorme, siendo este que el libro poco a poco se iba expandiendo más y más. Los capítulos asignados a los miembros empezaron siendo sencillos, pero eventualmente empezaron a tener que dividirse, ya que por ejemplo un capítulo entero de topología era necesario para explicar los conceptos de análisis funcional presentes posteriormente. es así como nació el tratado *Eléments de mathématique*, el trabajo culminar de Nicolas Bourbaki. Originalmente el tratado de análisis, se volvió en un libro de increíble profundidad tocando básicamente todos los temas de las matemáticas entonces actuales de importancia. Cada capítulo había pasado por tantas revisiones y ediciones que eran completamente diferentes de los originalmente publicados por cada autor, esto generando un tratado que verdaderamente no le pertenecía a nadie menos al grupo en general. Algo interesante es que el título usa el término *mathématique* en vez de *mathématiques*, representando unidad entre todos los temas presentados, hablando de la matemática, y no de las matemáticas.

Eléments de mathématique está compuesto de varios volúmenes, empezando con lo "básico", la teoría de conjuntos, siguiéndose de volúmenes enteros dedicados al estudio de álgebra, topología, funciones reales de una variable, y terminando con volúmenes dedicados al álgebra de los grupos de Lie, la teoría de integración, y la historia de las matemáticas. Empezando con publicaciones en 1939 y continuando con creaciones, expansiones y ediciones a los volúmenes y capítulos presentes, el tratado es uno de los más completos y rigurosos presentados en los últimos siglos. Han habido ediciones del mismo tan recientemente como el año 2016, con la inclusión de un nuevo volumen, topología algebraica. Es un libro en constante revisión y re-revisión, incorporando nuevos avances en las áreas relevantes y editando errores pasados, y por lo tanto, *Eléments de mathématique* sigue incompleto hasta hoy en día.

El impacto del grupo Nicolas Bourbaki ha sido enorme en las matemáticas contemporáneas. Posteriormente a los años 50, se había adoptado la rigurosidad presentada por el libro y trabajos de Bourbaki como el estándar de calidad sobre el cual se deberían de hacer las matemáticas contemporáneas, y desde los años 60, *todos los textos publicados seriamente en cualquier área de las matemáticas empezaron a seguir sus estándares de calidad.* Esto solamente 20 años después de la publicación del inicio de su trabajo seminal. En París, se tiene todavía el seminario Bourbaki, donde se presentan los avances mas importantes en las matemáticas. Es claro también que el grupo sigue vivo, y publicando, bajo el mismo nombre, expandiendo el siempre creciendo texto por el cual el grupo es famoso.

Referencias

[1] Marcus du Sautoy. *"A Brief History of Mathematics."* Podcast. Nicolas Bourbaki. BBC Radio 4, 25 Junio 2010. Web. 26 Marzo 2018.

[2] Beaulieu, Liliane. *"A Parisian Café and Ten Proto-Bourbaki Meetings (1934–35)."* Mathematical Intelligencer 15 (1993), 27–35.

[3] Atiyah, Michael (2000, November 17). Book Review [Reseña del libro *Bourbaki, A Secret Society of Mathematicians and The Artist and the Mathematician*]. American Mathematical Society, Volumen 54, 1150.

[4] Chouchan, Michéle. *Nicolas Bourbaki: Faits et legendes.* AgrenteuilCedex: Editions du Choix, 1995.

[5] Weil, André. *The Apprenticeship of a Mathematician.* Traducido por Jennifer Gage. Basel; Boston: Birkhauser Verlag, 1992.

[6] Gamez, J. (2014, July 20). *Nicolas Bourbaki, ¿un matemático?*. Ultima revisión en Abril 01, 2018. `http://www.matematicasdigitales.com/nicolas-bourbaki-un-matematico/`

[7] Melvecs. *La Sociedad Secreta Bourbaki*. `https://melvecsblog.wordpress.com/2016/10/27/la-sociedad-secreta-bourbaki/`

Aproximando a la razón áurea y a otros números.

Gerardo González Robert
Licenciado en Matemáticas Aplicadas por el ITAM

Resumen

Partiendo de un problema simple sobre sucesiones de Fibonacci generalizadas y la razón áurea, presentamos algunas ideas sobre la rama de la teoría de los números llamada aproximación diofantina (o diofántica).

1 Introducción y planteamiento del problema

La sucesión de Fibonacci ($(F_n)_{n=0}^{\infty}$) y la razón áurea (ϕ) son dos elementos que, al estar presentes tanto en las matemáticas puras como en la naturaleza e inclusive en el arte, han ganado reconocimiento alrededor del mundo. Su sencillez los hace aptos para la matemática recreativa, pero, al mismo tiempo, tienen una estructura suficientemente rica para ser base de investigación matemática seria.

La **sucesión de Fibonacci** $(F_n)_{n=0}^{\infty}$ es la sucesión de enteros no negativos definida por

$$F_0 = 0,\ F_1 = 1,\ F_n = F_{n-1} + F_{n-2}.$$

Mientras que la **razón áurea**, también conocida como razón dorada o número de oro, es el real $\phi = \frac{1+\sqrt{5}}{2}$. Una de las tantas relaciones entre ambos conceptos es que

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \phi. \tag{1}$$

Una manera de generalizar la sucesión de Fibonacci es cambiando las condiciones iniciales, de tal manera que satisfaga:

I. $|f_0| + |f_1| > 0$.

II. Para[1] $n \in \mathbb{N}$ se cumple $f_{n+1} = f_{n-1} + f_{n-1}$.

donde a la pareja (f_0, f_1) se le conoce como **semilla**.

A partir de (1) nos planteamos la siguiente pregunta.

Problema 1. *Sea $(f_n)_{n=0}^{\infty}$ una sucesión generalizada de Fibonacci. ¿Podemos determinar el límite de $\frac{f_{n+1}}{f_n}$ cuando n tiende a infinito?*

Si bien llegar a la respuesta puede tomar tan solo unos minutos, utilizaremos el Problema 1 para mostrar algunos aspectos de la teoría de los números. A lo largo de la siguiente sección, motivaremos el uso de fracciones continuadas y presentaremos algunas de sus propiedades, resolveremos el Problema 1 y reflexionaremos un poco sobre preguntas relacionadas. Luego, en la tercera sección, resolveremos el Problema 1 de otra manera.

[1] Llamaremos números naturales, $\mathbb{N}$, a los enteros no negativos.

2 Primera solución

Sea $(f_n)_{n=0}^{\infty}$ una sucesión de Fibonacci generalizada. Dejando a un lado las posibles indeterminaciones, la recurrencia que determina a $(f_n)_{n=0}^{\infty}$ nos dice que

$$\frac{f_{n+1}}{f_n} = 1 + \frac{f_{n-1}}{f_n} = 1 + \frac{1}{1 + \frac{f_{n-2}}{f_{n-1}}} = 1 + \frac{1}{1 + \frac{1}{1 + \frac{f_{n-3}}{f_{n-2}}}} = \ldots = 1 + \cfrac{1}{1 + \cfrac{1}{\ddots\, 1 + \frac{1}{\frac{f_1}{f_0}}}}.$$

Las expresiones anteriores se llaman fracciones continuadas[2]. La primera solución al Problema 1, es una aplicación directa de la teoría de de fracciones continuas. El objetivo es ver qué tan buena es la aproximación a ϕ usando este método.
Supondremos que, sin pérdida de generalidad, $f := (f_n)_{n=0}^{\infty}$ satisface ciertas condiciones. Primero, f es un múltiplo de $(F_n)_{n=0}^{\infty}$ cuando $f_0 = 0$ y $f_{n+1}/f_n = F_{n+1}/F_n$ para toda n. Obtenemos una igualdad similar cuando $f_1 = 0$. Además, si $f_0 f_1 \neq 0$ y $\frac{f_1}{f_0} = -\frac{F_{j+1}}{F_j}$ para alguna $j \in \mathbb{N}$, entonces f vuelve a ser un múltiplo $(F_n)_{n=0}^{\infty}$ (la recurrencia implica $f_j = 0$). Así, pensaremos que $(f_n)_{n=0}^{\infty}$ satisface

$$f_0 \neq 0, \quad f_1 \neq 0, \quad \forall j \in \mathbb{N} \quad \frac{f_1}{f_0} \neq -\frac{F_{j+1}}{F_j}. \tag{2}$$

2.1. Fracciones continuadas

Recordemos algunas propiedades básicas de las fracciones continuadas. Para más detalles, consultar las siguientes referencias: ([Hardy Wright], 2009) ([Niven et al.], 1991) y ([Khinchin], 1961).
Una **fracción continuada** es una expresión de la forma

$$[a_0; a_1, a_2, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}, \tag{3}$$

con una interpretación intuitiva cuando tenemos una cantidad finita de términos $a_0, \ldots, a_n$. Diremos que la fracción continuada es **regular** si a_0 es un entero y a_n es un entero no negativo para cada $n \in \mathbb{N}$. Bajo estas condiciones, podemos entender a (3) como un límite de números racionales

$$[a_0, a_1, a_2, \ldots] = \lim_{n\to\infty} [a_0; a_1, \ldots, a_n].$$

Llamaremos **convergentes** al lado derecho de la expresión anterior.

[2]También se les conoce como *fracciones continuas*.

Las fracciones continuadas establecen una biyección entre los números irracionales y cierta familia de sucesiones de enteros.

Teorema 2.1. *Para cada irracional α existe una y sólo una sucesión de enteros $(a_n)_{n=0}^{\infty}$ con $a_n \geq 1$ para cada $n \in \mathbb{N}$ tal que*

$$\alpha = [a_0; a_1, a_2, \ldots]. \tag{4}$$

Además, para cada sucesión en $\mathbb{Z}$, $(a_n)_{n=0}^{\infty}$, con $a_n \geq 1$ para $n \in \mathbb{N}$ existe un irracional α tal que (4) se satisface.

El Teorema 2.1 se sigue de estudiar a la sucesión $[a_0; a_1, \ldots]$. Aunque dejamos la prueba al lector, los puntos importantes están contenidos en el Lema 2.2.

Lema 2.2. *Sea $[a_0; a_1, a_2, \ldots]$ una fracción continuada regular y $(p_n)_{n=0}^{\infty}$ y $(q_n)_{n=0}^{\infty}$ sucesiones de enteros tales que*

$$\forall n \in \mathbb{N} \quad \frac{p_n}{q_n} = [a_0; a_1, \ldots, a_n].$$

Se tienen las siguientes propiedades.

I. *Definiendo p_{-1} y q_{-1}, se tiene que*

$$\begin{pmatrix} p_{-1} & p_0 \\ q_{-1} & q_0 \end{pmatrix} = \begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix}, \quad \forall n \in \mathbb{N} \quad \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} a_{n+1} \\ 1 \end{pmatrix}. \tag{5}$$

II. *Los términos p_n y q_n son coprimos; de hecho,*

$$\forall n \in \mathbb{N} \quad q_n p_{n-1} - q_{n-1} p_n = (-1)^n. \tag{6}$$

III. *La sucesión $(q_n)_{n=0}^{\infty}$ es positiva y crece exponencialmente.*

IV. *Para cada $n \in \mathbb{N}$ y $x \in \mathbb{R} \setminus \left\{ \frac{q_{j-1}}{j_n} : j \in \{1, \ldots, n\} \right\}$ tenemos que*

$$[a_0; a_1, \ldots, a_n, x] = \frac{p_n x + p_{n-1}}{q_n x + q_{n-1}} = \frac{p_n}{q_n} + \frac{(-1)^n}{q_n^2 \left(x + \frac{q_{n-1}}{q_n} \right)}. \tag{7}$$

V. *Los convergentes de orden par son crecientes y los de orden impar, decrecientes; luego,*

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \ldots < \alpha < \ldots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}. \tag{8}$$

Un ejemplo sencillo es la fracción $[1; 1, 1, \ldots]$. Llamando $x_0 > 1$ al límite, se tiene que

$$x_0 = 1 + x_0^{-1}$$

y, como $x_0 > 1$, $x_0 = \phi$. Además, las recurrencias de los convergentes dan

$$\forall n \in \mathbb{N} \quad p_n = F_{n+2}, \quad q_n = F_{n+1}. \tag{9}$$

El Lema 2.2 nos dice que los cocientes de términos consecutivos oscilan alrededor de ϕ.

2.2. Solución del Problema 1

De la solución del Problema 1, obtenemos el siguiente resultado.

Teorema 2.3. *Sea $(f_n)_{n=0}^{\infty}$ una sucesión de Fibonacci con semilla $(a,b) \neq (0,0)$. Entonces,*

$$\lim_{n\to\infty} \frac{f_{n+1}}{f_n} = \begin{cases} \phi & si \;\; b \neq -\phi^{-1}a, \\ -\phi^{-1} & si \;\; b = -\phi^{-1}a. \end{cases}$$

Demostración. Sea $(f_n)_{n=0}^{\infty}$ una sucesión generalizada de Fibonacci que, sin pérdida de generalidad, satisface (2). Uilizando (7) con $x = \frac{f_1}{f_0}$, tenemos que

$$\forall n \in \mathbb{N} \quad \frac{f_{n+1}}{f_n} = [\underbrace{1;1,\ldots,1}_{n \text{ veces}}, x] = \frac{F_{n+2}}{F_{n+1}} + \frac{(-1)^n}{F_{n+2}^2\left(x + \frac{F_{n+1}}{F_{n+2}}\right)} \tag{10}$$

Caso I. $x \neq -\phi^{-1}$. Por (1), existe $\eta > 0$ tal que

$$\forall n \in \mathbb{N} \quad \left| x + \frac{F_n}{F_{n+1}} \right| > \eta.$$

Tomando el límite cuando n tiende a infinito en (10), concluimos que

$$\lim_{n\to\infty} \frac{f_{n+1}}{f_n} = \lim_{n\to\infty} \frac{F_{n+2}}{F_{n+1}} + \frac{(-1)^n}{F_{n+2}^2\left(x + \frac{F_{n+1}}{F_{n+2}}\right)} = \phi.$$

Caso II. $x = -\phi^{-1}$. Por $-\phi^{-1} = 1 - \phi$, para cada $n \in \mathbb{N}$ tenemos

$$\frac{f_{n+1}}{f_n} = \frac{F_{n+2}}{F_{n+1}} + \frac{(-1)^n}{F_{n+2}^2\left(-\phi^{-1} + \frac{F_{n+1}}{F_{n+2}}\right)} = \frac{F_{n+2}}{F_{n+1}} + \frac{(-1)^n}{F_{n+2}^2\left(\frac{F_{n+3}}{F_{n+2}} - \phi\right)}. \tag{11}$$

Observemos el segundo sumando del lado derecho. Aplicando (7), (9) y $\phi = [1;1,1,\ldots]$, llegamos a

$$\phi = [\underbrace{1;1,\ldots,1}_{n+1 \text{ veces}}, \phi] = \frac{F_{n+3}}{F_{n+2}} + \frac{(-1)^{n+1}}{F_{n+2}^2\left(\phi + \frac{F_{n+1}}{F_{n+2}}\right)}.$$

Así, reordenando los términos, tenemos que

$$(-1)^n F_{n+2}^2 \left(\phi - \frac{F_{n+3}}{F_{n+2}}\right) = -\frac{1}{\phi + \frac{F_{n+2}}{F_{n+3}}} \underset{n\to\infty}{\to} -\frac{1}{\phi + \phi^{-1}} = -\frac{1}{\sqrt{5}}. \tag{12}$$

En la primera igualdad hemos usado (8). El resultado se obtiene de sustituir a (12) en (11) y tomar el límite cuando $n \to \infty$. □

2.3. Problemas relacionados

El estudio minucioso de ciertas aproximaciones racionales a ϕ conduce a la solución del Problema 1. Para ver lo qué ocurre en general, empecemos con un irracional $\alpha = [a_0; a_1, a_2, \ldots]$ y llamemos $(q_n)_{n=0}^{\infty}$ y $(p_n)_{n=0}^{\infty}$ a las sucesiones de numeradores y denominadores de sus convergentes. Utilizaremos $\|x\|$ para referirnos a la distancia de $x \in \mathbb{R}$ al entero más cercano. Por ejemplo, $\|0.3\| = \|1.7\| = 0.3$.
A partir de (7) se llega a

$$\forall n \in \mathbb{N} \;\; q_n \|q_n \alpha\| < 1,$$

lo cual indica la cercanía a un real fijo con racionales de denominador acotado y muestra que $\mathbb{Q}$ es denso en $\mathbb{R}$.
lo cual indica la cercanía a un real fijo con racionales de denominador acotado. Así, una manera de determinar la aproximación a un número real mediante racionales es encontrando la mínima $c > 0$ tal que $q\|q\alpha\| < c$. En símbolos, nos interesa calcular

$$\liminf_{q \to \infty} q\|q\alpha\|. \tag{13}$$

Las fracciones continuadas simplifican esta tarea. Dado un real α, decimos que $\frac{p'}{q'} \in \mathbb{Q}$ es una **mejor aproximación** de α si

$$\forall q \in \mathbb{N} \quad \forall p \in \mathbb{Z} \quad 0 < q < q' \implies |q'\alpha - p'| < |q\alpha - p|.$$

Teorema 2.4. *Sea $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Un racional es una mejor aproximación de α si y sólo si es un convergente de α.*

Por el Teorema 2.4, cuando $q \in \mathbb{N} \setminus \{q_j : j \geq 0\}$ existe n tal que

$$q_n < q < q_{n+1}, \quad \|q_n\alpha\| < \|q\alpha\|.$$

Multiplicando las desigualdades llegamos a $q_n\|q_n\alpha\| \leq q\|q\alpha\|$. Entonces, podemos calcular el límite (13) fijándonos sólo en $(q_n)_{n=0}^{\infty}$:

$$\liminf_{q \to \infty} q\|q\alpha\| = \liminf_{n \to \infty} q_n\|q_n\alpha\| \leq 1.$$

Al resolver el Problema 1, calculamos el límite correspondiente a la razón áurea obteniendo como resultado

$$\liminf_{q \to \infty} q\|q\phi\| = \frac{1}{\sqrt{5}}. \tag{14}$$

Encontramos que ϕ es el número real más difícil de aproximar mediante racionales.

Teorema 2.5 (A. Hurwitz, 1891)**.** *Para todo $x \in \mathbb{R}$ $\liminf_{q\to\infty} q\|qx\| \leq \frac{1}{\sqrt{5}}$. De hecho, $(\sqrt{5})^{-1}$ es la mejor cota posible, esto es*

$$\text{ínf}\left\{c \in \mathbb{R} : \forall x \in \mathbb{R} \quad \liminf_{q \to \infty} q\|qx\| \leq c\right\} = \frac{1}{\sqrt{5}}.$$

Demostración. Probaremos un refinamiento de É. Borel (1903): para cada $j \in \mathbb{N}$ se cumple

$$\text{mín}\,\{q_{j-1}\|q_{j-1}x\|,\ q_j\|q_jx\|,\ q_{j+1}\|q_{j+1}x\|\} < \frac{1}{\sqrt{5}}. \tag{15}$$

Podemos suponer que $x > 0$. Si (15) fuera falsa, definiendo $\lambda_j = \frac{q_{j+1}}{q_j}$, tendríamos que

$$\frac{1}{\sqrt{5}q_{j-1}q_j} \leq \left|x - \frac{p_{j-1}}{q_{j-1}}\right| + \left|x - \frac{p_j}{q_j}\right| \implies \lambda_j + \frac{1}{\lambda_j} < \sqrt{5} \implies (\lambda_j - \phi)\left(\lambda_j + \phi^{-1}\right) < 0.$$

La desigualdad estricta se debe a la irracionalidad de $\sqrt{5}$. La segunda implicación se sigue de multiplicar por $\lambda_j > 0$ y factorizar. El mismo argumento muestra que $\frac{q_j}{q_{j-1}} = \lambda_{j-1} < \phi$. Sin embargo, la recurrencia de los denominadores nos lleva a la contradicción

$$\phi > \lambda_j = a_j + \frac{1}{\lambda_{j-1}} > 1 + \frac{1}{\phi} = \phi.$$

Por lo tanto, (15) debe ser cierta. La segunda afirmación sigue de (14). □

Volviendo al irracional $\alpha = [a_0; a_1, a_2, \ldots]$ y utilizando (7), con $x = [a_n; a_{n+1}, \ldots] > 1$ demostramos que $\liminf_n q_n\|q_n\alpha\| > 0$ ocurre si y sólo si $(a_n)_{n=0}^{\infty}$ es acotada. Un real con esta propiedad se llama **mal aproximable** y el conjunto de números mal aproximables se denota por **Bad**, el cual es de primera categoría en el sentido de Baire. **Bad**, a pesar de ser no numerable, es relativamente chico.

Teorema 2.6 (Borel,1909)**.** *Casi todo real tiene una fracción continuada con términos no acotados.*

El límite inferior da pie a un conjunto conocido como el **espectro de Lagrange** definido por:

$$\Lambda := \{\liminf_{q\to\infty} q\|q\alpha\| : \alpha \in \mathbb{R}\}.$$

Ahora, ¿qué pasa si consideramos a dos reales en lugar de uno? Esto es, para $\alpha, \beta \in \mathbb{R}$ buscamos

$$\liminf_{q\to\infty} q\|q\alpha\|\|q\beta\|.$$

Se intentará analizar este nuevo problema más complejo.

Conjetura 2.1 (Littlewood, 1942)**.** *Para todos $\alpha, \beta \in \mathbb{R}$ tenemos que*

$$\liminf_{q\to\infty} q\|q\alpha\|\|q\beta\| = 0.$$

Por el Teorema 2.6, casi todas las parejas de reales cumplen con la Conjetura de Littlewood. Hasta ahora, el resultado más fuerte al respecto se debe a Manfred Einsiedler, Anatole Katok y Elon Lindenstrauss. Ellos establecieron que el conjunto de parejas que violan la conjetura es (en un sentido técnico) muy chico. Esta investigación le valió a Elon Lindenstrauss la Medalla Fields en 2010.

3 Segunda solución

La siguiente identidad, llamada Fórmula de Binet, es un ejericio de inducción

$$\forall n \in \mathbb{N} \quad F_n = \frac{\phi^n - (-\phi^{-1})^n}{\phi - \phi^{-1}}. \tag{16}$$

Haciendo manipulaciones algebráicas llegamos a

$$\forall n \in \mathbb{N} \quad \frac{F_{n+1}}{F_n} - \phi = \frac{1}{\phi^{2n}} \frac{(-1)^n}{\phi} \delta_n \quad \text{con} \quad \lim_{n\to\infty} \delta_n = 1 \tag{17}$$

y concluimos (1). Ahora, tomemos una sucesión generalizada de Fibonacci, $(f_n)_{n=0}^{\infty}$, con semilla (a, b), $b \neq 0$. Inductivamente, se tiene que $\forall n \in \mathbb{N} \quad f_n = F_{n-1}a + F_n b$. Sea $x = \frac{a}{b}$. Llegamos a una solución completa al Problema 1 tras aplicar (17) al cociente

$$\frac{f_{n+1}}{f_n} = \frac{aF_n + F_{n+1}b}{F_{n-1}a + F_n b} = \frac{F_n}{F_{n-1}} \left(\frac{x + \frac{F_{n+1}}{F_n}}{x + \frac{F_n}{F_{n-1}}} \right).$$

Como los racionales F_{n+1}/F_n son distintos, el cociente anterior tiene sentido para n grande. Cuando $x \neq -\phi^{-1}$ el segundo factor en la igualdad anterior converge a 1 y el producto, a ϕ. En cambio, cuando $x = -\phi^{-1}$ usamos la identidad $\phi = 1 + \phi^{-1}$ para concluir que el segundo factor converge a $-\phi^{-2}$. Así, concluimos el Teorema 2.3.
Esta estrategia nos permite resolver preguntas similares, por ejemplo:

Problema 2. *Sean $(g_n)_{n=0}^{\infty}$ y $(h_n)_{n=0}^{\infty}$ sucesiones generalizadas de Fibonacci. ¿Cuál es el límite, si existe, de g_n/h_n cuando n tiende a infinito?*

Agradecimientos. Gracias a Santiago Cabello Tueme y a los editores por sus comentarios.

Referencias

[Ba] Alan Baker. 2012. *A Comprehensive Course in Number Theory.* Cambridge: Cambridge University Press.

[Bu] Yann Bugeaud. *Around the Littlewood conjecture in Diophantine approximation.* Publ. Math. Besançon (2014): 5-18.

[Ca] John W.S. Cassels, (1957), *Diophantine Approximation.* Cambridge Tracts in Mathematics and Mathematical Physics, No. 45., Cambridge University Press, New York.

[Hardy Wright] Godfrey H. Hardy y Edward M. Wright. (2009) *An Introduction to the Theory of Numbers.* Sexta ed. Inglaterra: Oxford.

[Khinchin] Alexander Ya. Khinchin. 1961. *Continued Fractions.* Reimpresión 2006. Nueva York: Dover Publications.

[Niven et al.] Ivan Niven, Hugh Montgomery y Herbert Zuckerman. 1991. *An Introduction to the Theory of Numbers.* Quinta ed. Nueva York: John Wiley & Sons.

El espacio vectorial de los números reales sobre los números racionales

Ramón Espinosa Armenta
Departamento de Matemáticas, ITAM

Introducción

El propósito de esta nota es mostrar que el espacio vectorial de los números reales sobre el campo de los números racionales tiene dimensión infinita. Aunque este resultado es conocido, la demostración no aparece en los libros usuales de Álgebra Lineal. Una de las razones podría ser que la demostración requiere conceptos y resultados acerca de conjuntos infinitos, los cuales no se ven en cursos básicos.

En este trabajo supondremos que el lector está familiarizado con la teoría de espacios vectoriales (ver por ejemplo [2], capítulo 1). En la sección 2 veremos los conceptos, ejemplos y resultados acerca de conjuntos infinitos necesarios para entender la demostración del teorema principal, la cual veremos en la sección 3. En la última sección veremos un subespacio interesante del espacio vectorial de los números reales sobre los racionales.

Conjuntos infinitos

Se dice que dos conjuntos no vacíos A y B tienen la misma cardinalidad si existe una función biyectiva de A en B. En este caso escribimos $A \sim B$.

Se dice que un conjunto A es **finito** si es vacío o si existe $n \in \mathbb{N}$ tal que $A \sim \{1, \ldots, n\}$.Se dice que un conjunto A es **infinito** si no es finito.

Algunos ejemplos de conjuntos infinitos son el conjunto de los números naturales $\mathbb{N} = \{1, 2, 3, \ldots\}$, el de los números enteros $\mathbb{Z}$, el de los números racionales $\mathbb{Q}$ y el de los números reales $\mathbb{R}$.

Se dice que un conjunto infinito A es **numerable** si $A \sim \mathbb{N}$. Se dice que A es **a lo más numerable** si es finito o numerable.

Ejemplo 1. *El conjunto de los números naturales pares* $2\mathbb{N} = \{2, 4, 6, \ldots\}$ *es numerable, pues la función* $f(n) = 2n$ *es una función biyectiva de* $\mathbb{N}$ *en* $2\mathbb{N}$. △

Ejemplo 2. *La función* $f : \mathbb{N} \to \mathbb{Z}$ *definida por*

$$f(n) = \begin{cases} n/2 & \text{si } n \text{ es par;} \\ (1-n)/2 & \text{si } n \text{ es impar.} \end{cases}$$

es biyectiva (ver [1], ejemplo 4.34), lo cual muestra que el conjunto de los números enteros $\mathbb{Z}$ es numerable. △

Enunciamos a continuación un resultado importante para determinar si un conjunto es a lo más numerable. El lector puede consultar la demostración en [1], Corolario 4.3.

Teorema 1. *Si A es un conjunto numerable y B es otro conjunto para el cual existe una función inyectiva $f : B \to A$, entonces B es a lo más numerable.*

El siguiente teorema muestra que el producto cartesiano de conjuntos numerables es numerable.

Teorema 2. *Si A y B son conjuntos numerables, entonces $A \times B$ es numerable.*

Demostración. Como A y B son numerables, podemos escribir

$$A = \{a_1, a_2, a_3, \ldots\} \quad \text{y} \quad B = \{b_1, b_2, b_3, \ldots\}.$$

Sea $f : A \times B \to \mathbb{N}$ definida por

$$f(a_n, b_m) = 2^n \cdot 3^m.$$

Si $f(n, m) = f(p, q)$, entonces $2^n \cdot 3^m = 2^p \cdot 3^q$, de ahí que, por el teorema fundamental de la aritmética, $n = p$ y $m = q$, por lo tanto $(n, m) = (p, q)$, es decir, f es inyectiva, por lo que por el teorema anterior $A \times B$ es a lo más numerable. Como además $A \times B$ es infinito, se sigue que $A \times B$ es numerable. □

La demostración del siguiente corolario se puede hacer por inducción matemática.

Corolario 1. *Si $A_1, A_2, \ldots, A_n$ son conjuntos numerables, entonces*

$$A_1 \times A_2 \times \cdots \times A_n$$

es numerable.

Ejemplo 3. *Cada número racional se puede escribir de manera única como m/n, donde $m \in \mathbb{Z}$, $n \in \mathbb{N}$ y $mcd(m, n) = 1$. Sea $f : \mathbb{Q} \to \mathbb{Z} \times \mathbb{N}$ definida por $f(m/n) = (m, n)$. Por el teorema 2, $\mathbb{Z} \times \mathbb{N}$ es numerable, por otra parte es claro que f es inyectiva, por lo que por el teorema 1, $\mathbb{Q}$ es a lo más numerable. Además, como $\mathbb{Q}$ es infinito, se sigue que $\mathbb{Q}$ es numerable.* △

Ejemplo 4. *Por el ejemplo anterior, $\mathbb{Q}$ es numerable; por lo que por el corolario 1, $\mathbb{Q}^n$ también es numerable para todo número natural n.* △

No todos los conjuntos infinitos son numerables. En 1891 el matemático alemán Georg Cantor demostró que el intervalo $I = (0, 1)$ es no numerable, utilizando una ingeniosa demostración que el lector puede consultar en [1] (Teorema 4.17).

El siguiente ejemplo muestra que $I \sim \mathbb{R}$ y por lo tanto $\mathbb{R}$ es no numerable.

Ejemplo 5. *Sea* $f : (0, 1) \to \mathbb{R}$ *definida por*

$$f(x) = \frac{2x - 1}{2x(1 - x)}$$

Se deja al lector verificar que f *es biyectiva, por lo tanto* $I \sim \mathbb{R}$. △

El resultado principal

Si F es un campo y K es un subcampo de F, es fácil ver que F es un espacio vectorial sobre el campo K. Por ejemplo, $\mathbb{C}$ es un espacio vectorial sobre sí mismo (de dimensión 1), pero también es un espacio vectorial sobre el campo de los números reales (de dimensión 2).

El siguiente teorema muestra que el espacio vectorial $\mathbb{R}$ sobre el campo $\mathbb{Q}$ no es de dimensión finita.

Teorema 3. *El espacio vectorial de los números reales sobre los números racionales tiene dimensión infinita.*

Demostración. Supongamos que tiene dimensión finita y sea $\mathcal{B} = \{x_1, \ldots, x_n\}$ una base. Sea $f : \mathbb{Q}^n \to \mathbb{R}$ definida por:

$$f(r_1, \ldots, r_n) = \sum_{i=1}^{n} r_i x_i$$

Como $\mathcal{B}$ es base de $\mathbb{R}$ se tiene que para todo $x \in \mathbb{R}$, existen $r_1, \ldots, r_n \in \mathbb{Q}$ tales que

$$x = \sum_{i=1}^{n} r_i x_i$$

De ahí que $f(r_1, \ldots, r_n) = x$ y por lo tanto f es suprayectiva.
Por otra parte, si $f(r_1, \ldots, r_n) = f(s_1, \ldots, s_n)$ entonces

$$\sum_{i=1}^{n} r_i x_i = \sum_{i=1}^{n} s_i x_i$$

y de ahí que $r_i = s_i$ para todo $i = 1, \ldots, n$, porque la representación de un elemento en términos de la base es única. Por lo tanto $(r_1, \ldots, r_n) = (s_1, \ldots, s_n)$, lo cual muestra que f es inyectiva.
Por lo tanto f es una biyección de $\mathbb{Q}^n$ en $\mathbb{R}$, lo cual no es posible porque $\mathbb{Q}^n$ es numerable y $\mathbb{R}$ no lo es. Lo cual prueba que $\mathbb{R}$ como espacio vectorial sobre $\mathbb{Q}$ tiene dimensión infinita. □

Un subespacio interesante

Veremos a continuación un subespacio de dimensión finita del espacio vectorial $\mathbb{R}$ sobre el campo $\mathbb{Q}$.

Teorema 4. *El conjunto*

$$W = \{a + b\sqrt{2} \ \mid \ a, b \in \mathbb{Q}\}$$

es un subespacio del espacio vectorial de los números reales sobre los números racionales. Además $\dim W = 2$.

Demostración. Observemos primero que $0 = 0 + 0\sqrt{2}$, por lo que $0 \in W$. Supongamos ahora que $a + b\sqrt{2} \in W$ y $c + d\sqrt{2} \in W$. Por lo tanto

$$(a + b\sqrt{2}) + (c + d\sqrt{2}) = (a + c) + (b + d)\sqrt{2} \in W,$$

pues $(a+c) \in \mathbb{Q}$ y $(b+d) \in \mathbb{Q}$, por lo que la suma es cerrada en W. Por último, si $a+b\sqrt{2} \in W$ y $c \in \mathbb{Q}$, entonces

$$c(a + b\sqrt{2}) = (ca) + (cb)\sqrt{2} \in W,$$

pues $ca \in \mathbb{Q}$ y $cb \in \mathbb{Q}$, lo cual muestra que W es cerrado bajo el producto por escalares, con lo cual concluimos que W es subespacio de $\mathbb{R}$ sobre el campo $\mathbb{Q}$.
Por último, es claro que el conjunto $\mathcal{B} = \{1, \sqrt{2}\}$ genera a W. Para ver que es linealmente independiente sean $a, b \in \mathbb{Q}$ tales que $a + b\sqrt{2} = 0$. Si $b \neq 0$, entonces $\sqrt{2} = -a/b \in \mathbb{Q}$, lo cual no es posible, porque $\sqrt{2}$ es irracional. Por lo tanto $b = 0$, y de ahí que también $a = 0$, lo cual muestra que $\mathcal{B}$ es base de W, y por lo tanto $\dim W = 2$. □

Agradecimientos

Agradezco el apoyo de la Asociación Mexicana de Cultura, A. C. y del Instituto Tecnológico Autónomo de México (ITAM), para la realización de este trabajo.

Referencias

[1] Espinosa, R., *Matemáticas Discretas*, 2a Edición, Editorial Alfaomega, 2017.

[2] Friedberg, S. H., Insel, A. J. y Spence, L. E., *Linear Algebra*, 4th ed. Prentice Hall, 2003.

Experimentos con sumas exponenciales incompletas

Mario Cortina Borja
University College London

En crespa tempestad del oro undoso...
(Quevedo)

Introducción

En este artículo estudiamos sumas exponenciales incompletas[1] de la forma general

$$\mathcal{S} := S_f(N_0, N_1) = \sum_{j=N_0}^{N_1} e(f(j)) \tag{1}$$

donde $0 \leq N_0 < N_1$ son enteros, $e(x)$ denota $e^{2\pi ix}$ y f es una función real con dominio en los enteros positivos. Esta clase de sumas parciales surge en teoría de números y tiene un pedigrí histórico impresionante al haber sido estudiada originalmente por Jakob Bernoulli, Euler y Gauss [Gray (1997)].

En este artículo sólo revisaremos versiones determinísticas de $\mathcal{S} := S_f(N_0, N_1)$, aunque hay un campo amplio de investigación que incluye componentes estocásticas en f. Por ejemplo [Loxton (1985)] desarrolla una versión del teorema del límite central para $\mathcal{S}$ con formas de f que incluyen una sucesión de variables aleatorias i.i.d. uniformemente y [Alonso–Sanz (2010)] estudia espirales de Euler con memoria basadas en $\mathcal{S}$ resultantes de un proceso estocástico. Al margen de esta clase de trabajos, [Loxton (1981)] y [Angell (sin fecha)] muestran que del análisis de sumas parciales exponenciales se deriva fácilmente un aspecto informal, prácticamente lúdico[2]. Este es el núcleo del presente artículo y consiste en examinar cinco clases de $\mathcal{S}$:

1. Sumas exponenciales incompletas gaussianas (espirales de Euler)
2. Sumas exponenciales basadas en múltiplos de raíz cuadrada
3. Sumas exponenciales basadas en funciones logarítmicas
4. Sumas exponenciales resultantes en espirales asintóticas
5. Sumas exponenciales basadas en una función cúbica con coeficientes basados en fechas

La siguiente sección describe cómo obtener generalmente gráficas de $\mathcal{S} := S_f(N_0, N_1)$. Las siguientes secciones analizan los cinco casos determinísticos mencionados.

[1] La distinción entre sumas exponenciales completas e incompletas se centra en que las segundas ocurren sobre un conjunto de números naturales acotado por una desigualdad y las primeras lo hacen sobre todos los residuos de una función, módulo un entero positivo.

[2] Angell acota lúcidamente: "Real–life applications of these ideas? Forget it. This is mathematics for fun. Nothing wrong with that."

Gráficas de sumas exponenciales incompletas

Con base en [Loxton (1983)][3] definimos la gráfica $\mathcal{G}$ de una suma exponencial $\mathcal{S}$ como la sucesión de sumas parciales dibujada sobre el plano complejo al unir puntos sucesivos de $(\Re(\mathcal{S}), \Im(\mathcal{S}))$ con segmentos rectos.Estas gráficas pueden tener una complejidad enorme y es difícil predecir su comportamiento en función de f aún para formas simples de f con cambios relativamente pequeños en sus parámetros. En algunos casos, los valores de N_0 y N_1 igualmente inducen cambios difíciles de caracterizar en $\mathcal{G}$.

Como se mencionó en la Introducción, es posible incluir un elemento estocástico en $\mathcal{S}$ pero en el presente artículo consideramos solamente ejemplos deteminísticos. En las siguientes secciones analizamos ejemplos de cinco clases interesantes de $\mathcal{G}$.

Sumas exponenciales incompletas gaussianas (espirales de Euler)

La conexión entre sumas exponenciales y espirales fue estudiada definitivamente por Gauss, y primariamente, por Jakob Bernoulli en 1694 y por Euler 50 años después; [Levien (2008)] cuenta sucintamente esta historia. Siguiendo a [Lehmer (1976)], definimos la suma gaussiana incompleta en función de potencias de enteros escaladas como

$$G_N(m;p) \equiv S_f(0, m-1) = \sum_{j=0}^{m-1} \exp(2\pi\, i\, j^p/N) \tag{2}$$

con N un entero positivo, $p \geq 0$ y $1 \leq m < N$. Si $p > 1$, $G_N(m;p)$ es una suma incompleta gaussiana de orden p y $G_N(N,2)$ denota la suma incompleta ordinaria gaussiana. Para valores grandes de N, [Lehmer (1976)] muestra que casi todos los valores de $G_N(m,2)$, con $m < N/2$ se encuentran en una vecindad de $\sqrt{N}(1+i)/4$.

Sea $\mathcal{G} = \{G_N(0,2) \ldots, G_N(N,2)\}$ la gráfica con vértices dirigidos en el conjunto de valores $G_N(\cdot,2)$ y aristas en el conjunto de N vectores en el plano complejo uniendo $G_N(j,2)$ a $G_N(j+1,2)$. $\mathcal{G}$ empieza en el origen del plano complejo y es localmente lineal, no una curva suave. Para $m = \mathcal{O}\left(\sqrt{N}\right)$ tenemos que las primeras m aristas aproximan una clotoide (o doble espiral) de longitud ∞ cuya curvatura es proporcional a la longitud de sus aristas. Desde luego, como nota [Lehmer (1976)], la espiral tiene longitud ∞ de manera que estas comparaciones con $\mathcal{G}$ son útiles sólo en el contexto de las gráficas que mostramos a continuación.

La Figura 1 muestra cuatro ejemplos de esta clase de sumas exponenciales definida en la ecuación (2) con $m = N = 1024$ para hacer notar la variabilidad causada por valores de $p \geq 1$.

[3] Aunque modificando ligeramente su notación

Figura 1: $G_{1024}\,(1024, p)$ con cuatro valores de p

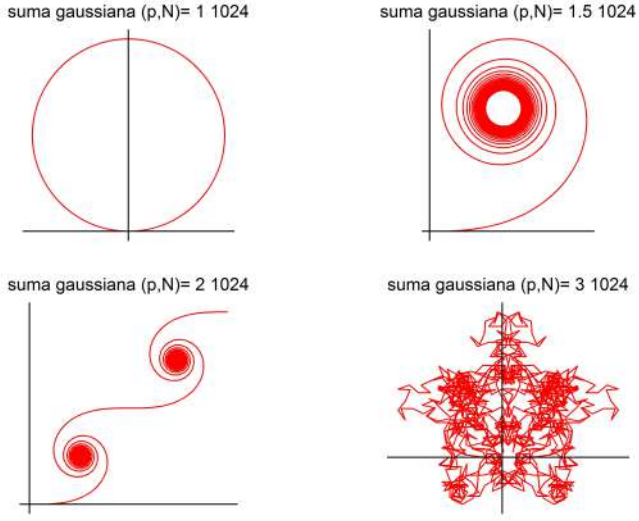
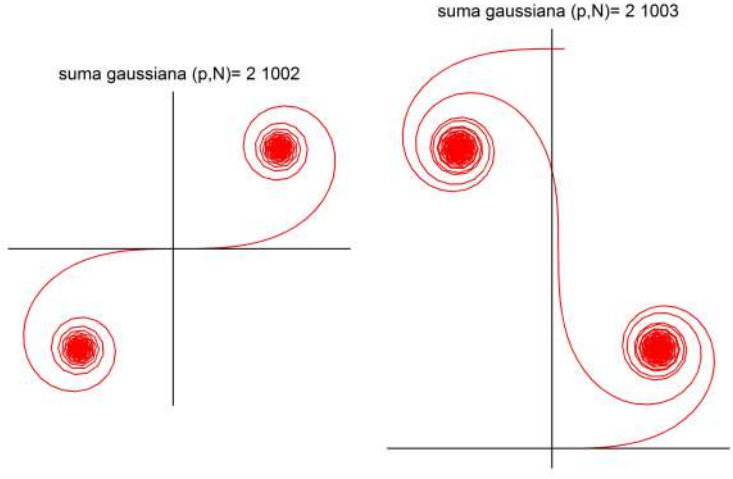


Figura 2: $G_N\,(2, N)$ con dos valores consecutivos de N



Como muestra [Paris (2005)] el caso gaussiano ordinario, i.e. $p = 2$, resulta generalmente en dos espirales simétricas formando la curva paramétrica llamada clotoide, espiral de Euler,

o espiral de Cornú[4]. [Paris (2005)] investiga la construcción de aproximaciones asintóticas para este caso. En la Figura 2 mostramos dos ejemplos de esta convergencia y notamos (i) el cambio cualitativo correspondiente a dos valores consecutivos de N y (ii) que $G_{1002}\left(2;1002\right)$ genera una clotoide doblemente transversal. Con razón, [Gray (1997)] no duda en describir la clotoide como 'una de las curvas más elegantes' en la historia de las Matemáticas.

Sumas exponenciales basadas en múltiplos de raíz cuadrada

Un caso interesante de $\mathcal{S}$ es la suma exponencial incompleta

$$R_t\left(N\right) \equiv S_f\left(0,N\right) = \sum_{j=0}^{N} \exp\left(2\pi i t\sqrt{j}\right) \tag{3}$$

con t un número real positivo. La $\mathcal{G}$ correspondiente ha sido estudiada en en [Loxton (1983)] especialmente respecto a la construcción de aproximaciones para los puntos de condensación de las espirales resultantes. La teoría expuesta en [Loxton (1983)] va más allá del alcance del presente artículo de manera que nos limitaremos a ilustrar cambios en $R_t\left(N\right)$ en función de t y N en la ecuación (3). Las Figuras 3 y 4 varían t en potencias de 10 para $N = 500$ y $N = 5000$. Para $t = 1$ se ve una convergencia inmediata a una espiral mientras que para $t = 10$ esta convergencia se alcanza luego de una caminata que se antoja aleatoria. En ambos casos, si $N \to \infty$, estas espirales cubrirán totalmente el plano complejo. Las trazas para $t = 100$ y $t = 1000$ son mucho más complicadas y no es fácil predecir si convergerán en una espiral total o cuándo lo harían.

La Figura 5 muestra la evolución de las trazas de $R_t\left(N\right)$ al incrementar N fijando $t = 1024$: las $\mathcal{G}$ resultantes son subconjuntos de las gráficas subsecuentes con valores crecientes de N: esto es especialmente notable al comparar, por ejemplo la traza de $R_{1024}\left(2048\right)$ con la de $R_{1024}\left(4096\right)$, y ésta con la de $R_{1024}\left(40096\right)$.

Sumas exponenciales basadas en funciones logarítmicas

La espiral logarítmica, una de las curvas más conocidas en Matemáticas, se puede aproximar a partir de la suma parcial en la ecuación (1) con $f\left(j\right) = \log\left(j\right)$. Como es bien sabido, esta espiral se manifiesta frecuentemente en la naturaleza y el arte. La Figura 6 muestra un ejemplo de esta curva, cuyas propiedades principales se comentan en [Wells (1991)].

[4]En honor del físico francés Marie Alfred Cornu

Figura 3: $R_t\,(500)$ con t variando en potencias de 10

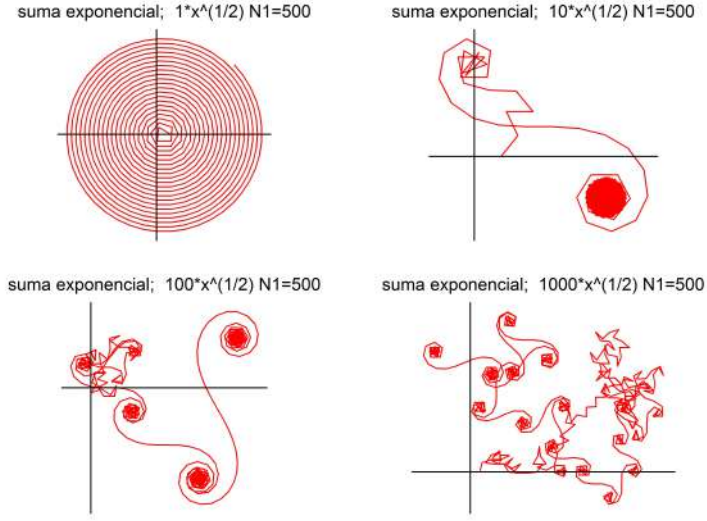


Figura 4: $R_t\,(5000)$ con t variando en potencias de 10

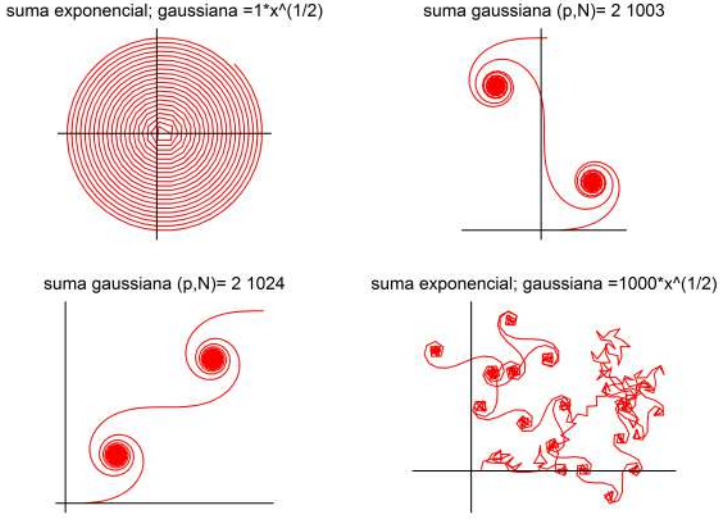
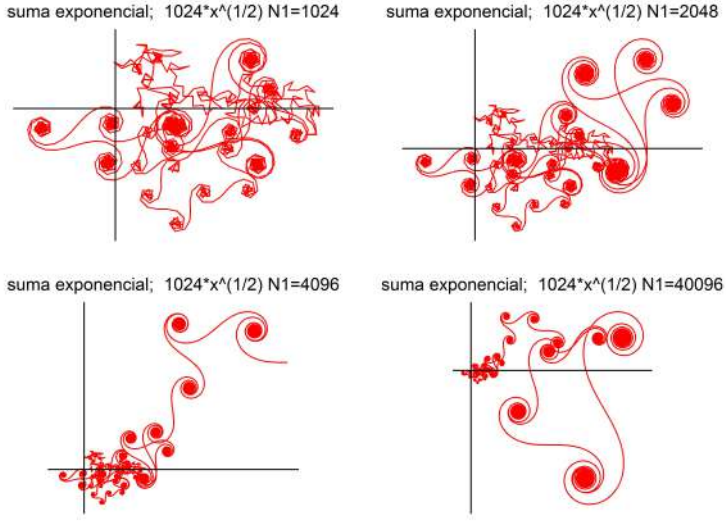


Figura 5: $R_{1024}(N)$ con N variando en potencias de 2



Es natural preguntarse cómo cambiaría la forma de esta espiral si usamos, por ejemplo, funciones potencia, i.e.

$$L_u(N) \equiv S_f(1,N) = \sum_{j=1}^{N} \exp\left(2\pi i (\log j)^u\right) \tag{4}$$

donde $u \geq 1$. Probablemente el ejemplo más celebrado de $L_u(N)$ es la curva llamada *Monstruo de Loch Ness* definida primeramente por [Loxton (1981)] para $u = 4$ en la ecuación (4) y que aparece, para $N = 5000$ en la Figura 7.

Aparte de su extraña, intrínseca belleza esta curva impresiona por su aparente impredecibilidad: ¿Por qué, en común con las curvas mencionadas en la sección anterior, parece empezar caóticamente? ¿Qué determina el inicio y la posición de cada sub–espiral? ¿Cuál es su conducta si $N \to \infty$? ¿Cuál es la relación entre u y el número de sub–espirales? Usando los argumentos desarrollados por [Loxton (1981)] es posible contestar aproximadamente algunas de estas preguntas.

[Loxton (1981)] nota que sucesión de ángulos entre segmentos consecutivos en $\mathcal{G}$ para el monstruo de Loch Ness es:

$$\Phi := \left\{\phi_j = 2\,\pi\,\left[(\log(j+1))^4 - (\log(j))^4\right]\right\}$$

y si $j \to \infty$ converge a 0. Esto implica que esta suma parcial exponencial resulta en una espiral que eventualmente cubre todo el plano complejo a partir del centro de una última sub–espiral.

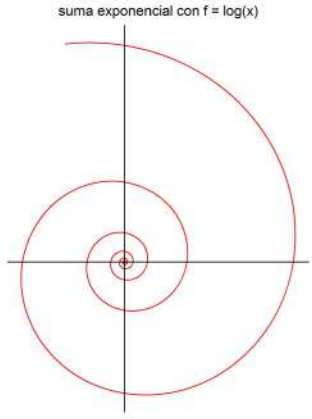


Figura 6: Espiral logarítmica

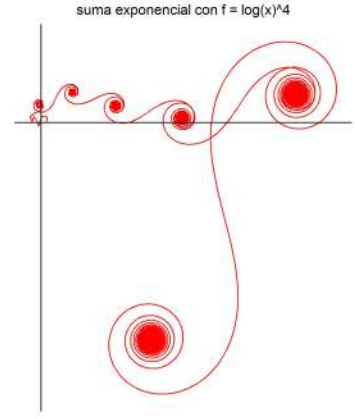


Figura 7: Monstruo de Loch Ness: $f(j) = (\log(j))^4$

Es posible mostrar que $\mathcal{G}$ correspondiente a $L_4(N)$ tiene sólo las seis sub–espirales que aparecen en la Figura 7. Si el ángulo $\phi_j/2\,\pi$ está en la vecindad de un número entero tenemos que $\mathcal{G}$ cambia de dirección y aproxima una curva que se mueve lentamente ligando sub–espirales hasta alcanzar la siguiente, que puede ser la última. Al contrario, si la distancia entre $\phi_j/2\,\pi$ y un entero positivo es relativamente grande tenemos que la gráfica $\mathcal{G}$ cambia de dirección rápidamente alrededor de un punto de atracción. Esto resulta en una especie de hoyo negro expansivo del que sólo es posible salir cuando la sucesión $\Phi/2\,\pi$ alcanza a estar relativamente cerca de un entero positivo, lo cual es imposible $\forall u \geq 1$ cuando $N \to \infty$. La primera gráfica de la Figura 1 muestra algo localmente parecido a tal curva.

[Loxton (1981)] define a $\mathcal{G}$ para $\mathcal{S} := L_7(N)$ como *La vía láctea* la cual aparece con $N = 5000$ en la Figura 8 y es obviamente más complicada que el Monstruo de Loch Ness, $(u = 4)$.
A continuación estudiamos empíricamente la complejidad de $\{L_u(N)\}$. Primeramente, definimos el grado de complejidad de $\mathcal{G}$ como el número de enteros en la sucesión $\Phi/2\,\pi$ con valores relativamente grandes de N. La Tabla 1 muestra el número de enteros diferentes en $\Phi/2\,\pi$, i.e. el número de sub–espirales en $\mathcal{G}$, para $N = 10^6$ y $1 \leq u \leq 7$. Note que u no es necesariamente un entero positivo, de manera que se tiene un continuo en los resultados reportados en la Tabla 1.

Tabla 1: Número de sub–espirales en $L_u(10^6)$, para $1 \leq u \leq 7$

u	1	2	3	4	5	6	7
sub–espirales	1	1	2	6	24	256	768

Al incrementar N, el número de sub–espirales crece rápidamente siempre y cuando no se haya alcanzado su número máximo. Por ejemplo, para observar la (sexta y última) sub–espiral con

$u = 4$ se necesitan menos de $N = 5000$ aristas en $\mathcal{G}$. En general esta cota superior para $L_u(N)$ es el máximo número entero z tal que $\phi_j / 2\pi \leq z$. La Figura 8 contiene menos de 768 sub–espirales por estar basada en $N_1 = 5000$, no en $N_1 = 10^6$ como se construyó la Tabla 1, y desde luego hay más de 768 sub–espirales en versiones de $\mathcal{G}$ con $u = 7$ si $N \gg 10^6$: por ejemplo con $N = 10^7$ observamos empíricamente 810 sub–espirales.

Si u crece, tenemos una mayor complejidad y una convergencia mucho más lenta pero igualmente inexorable hacia una sub–espiral dominante. Con $u = 8$ y $N = 10^8$ observamos 6008 sub–espirales; para $u = 9$ con $N = 10^9$, 50654. Si N crece, eventualmente la última sub–espiral cubrirá totalmente al plano complejo: el monstruo se transforma en un Leviatán que se devora a sí mismo, la vía láctea desaparece en un hoyo negro[5].

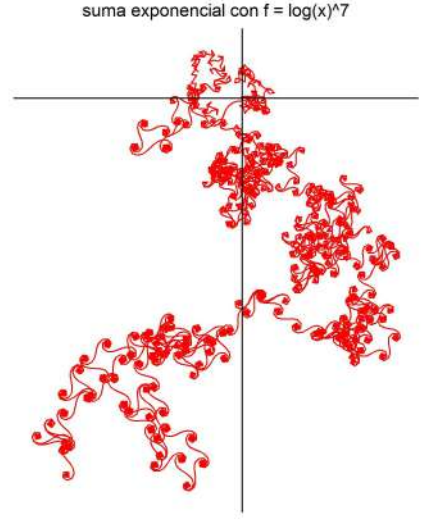
Sumas exponenciales resultantes en espirales asintóticas

[Paris (2009)] define una nueva clase de sumas parciales exponenciales que siempre alcanzan una asíntota. La forma general depende de tres parámetros; usando la notación de Paris se define como:

$$A_p(\theta; m) \equiv S_f(p, \theta, m;\, 0, \infty) = {\sum_{j=0}^{\infty}}' \exp\left[-\frac{j}{m} - i\,\theta \exp\left(\frac{-p\,j}{m}\right)\right] \tag{5}$$

donde p, θ y m son cantidades positivas y m no es necesariamente un número entero; la suma primada denota que su primer término se divide por 2. En esta clase, $\mathcal{G}$ inicia en el origen y alcanza necesariamente su asíntota para j suficientemente grande al salir de la última sub–espiral. Su grado de complejidad depende de las combinaciones de los tres parámetros de f.

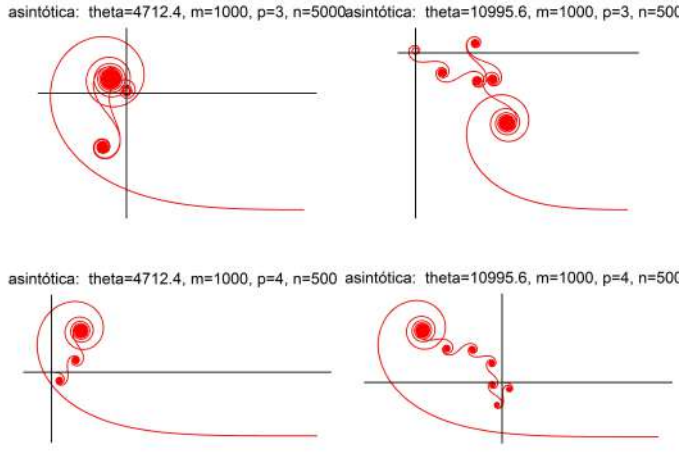
Figura 8: La Vía Láctea: $L_7(5000)$



La Figura 9 ilustra cuatro ejemplos. Los valores de θ son múltiplos de π: 1500 y 3500, para $p = 3$ y $p = 4$ y $m = 1000$; las gráficas se dibujaron con 5000 puntos. [Paris (2009)] desarrolla aproximaciones interesantes para predecir el comportamiento de $\mathcal{G}$ en esta clase, pero sus argumentos exceden al enfoque del presente artículo.

[5] *O quam cito transit gloria mundi* (Thomas à Kempis, De Imitatione Christi)

Figura 9: Cuatro sumas parciales exponenciales asintóticas

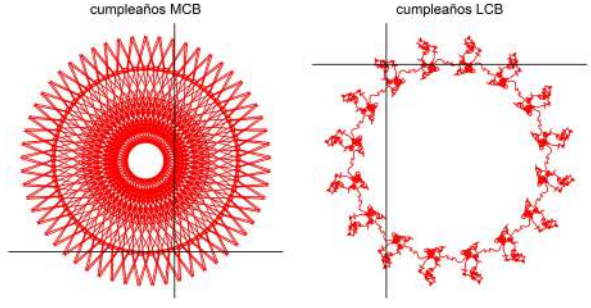


Sumas exponenciales incompletas relativas a fechas

Nuestro último ejemplo de f en la ecuación (1) se refiere a un modelo descrito por [Angell (sin fecha)] y que busca expresar $\mathcal{G}$ en función de fechas, por ejemplo cumpleaños, a partir de la ecuación

$$F\left(d,m,a;N_0,N_1\right) \equiv S_f\left(N_0,N_1\right) = \sum_{j=N_0}^{N_1} e\left(\frac{j}{d}+\frac{j^2}{m}+\frac{j^3}{a}\right) \tag{6}$$

Figura 10: Sumas parciales exponenciales definidas por un polinomio cúbico con fechas



donde d, m, a corresponden a el día, mes y año definidos por a lo más dos dígitos para una fecha particular. Por ejemplo, el 3 de mayo de 1989 se puede representar con $d = 3$, $m = 5$, y $a = 89$. Las trazas $\mathcal{G}$ resultantes exhiben una variedad enorme: a veces son algo parecido a lo que un espirógrafo genera, otras son más difíciles de caracterizar[6] y dependen tanto de N_0 y N_1 como de d, m y a. La Figura 10 muestra dos ejemplos de estos casos, ambos con $N_0 = 1$ y $N_1 = 5000$.

Le recomendamos al lector a realizar sus propios experimentos tal vez sobreinterpretando[7] las $\mathcal{G}$s correspondientes a fechas especiales...

Conclusión

Las trazas de sumas parciales exponenciales en el plano complejo son objetos interesantes y complicados, con una historia fascinante y fáciles de programar. Por ejemplo, una función en **R**, escrita por el autor, que permite reproducir todas las figuras de este artículo está disponible en `https://github.com/MarioCortinaBorja/Sumas-Parciales-Exponenciales`.

Finalmente, queremos notar la presencia de espirales tal vez no muy diferentes de las aquí descritas en *El árbol de la vida* de Gustav Klimt[8] que aparece en la Figura 11. Desde luego, muchas otras obras de arte contienen representaciones de espirales e invitamos al lector a nombrar su favorita en este contexto.

Figura 11: Gustav Klimt (1862–1918) El árbol de la vida (1905): Museo de Artes Aplicadas, Viena

El campo del presente artículo es muy estrecho: simplemente nos limitamos a mostrar clases de gráficas interesantes que pueden programarse fácilmente en, por ejemplo, **R**. Sin embargo, es un mínimo ejemplo de la idea central en la obra de G. H. Hardy: las Matemáticas más bellas son [tal vez] aquellas sin aplicaciones prácticas [Hardy (1940)].

[6] Hasta donde sabemos, no se ha intentado un estudio formal de esta clase de $\mathcal{G}$

[7] Con licencia poética, ¡nada más!

[8] By Gustav Klimt - The Yorck Project: 10.000 Meisterwerke der Malerei. DVD-ROM, 2002. ISBN 3936122202. Distributed by DIRECTMEDIA Publishing GmbH., Public Domain, `https://commons.wikimedia.org/w/index.php?curid=153464`

Referencias

[1] Alonso–Sanz, R. Curlicues with memory. *International Journal of Bifurcation and Chaos*, **20**, 2225–2240, 2010.

[2] Angell D. Exponential sums. `https://www.maths.unsw.edu.au/about/exponential-sums` (descargado el 2 de noviembre de 2017)

[3] Gray A. *Modern Differential Geometry of Curves and Surfaces with* Mathematica. Boca Ratón, Forida: CRC Press, 2a edición, 1997.

[4] Hardy G.H. *A Mathematician's Apology.* Cambridge: Cambridge University Press, 1940.

[5] Lehmer D.H. Incomplete Gauss sums. *Mathematika*, **23**, 125–135, 1976.

[6] Levien R. (2008) *The Euler spiral: a mathematical history.* Technical report UCB/EECS-2008-111, University of California at Berkeley. `https://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-111.pdf`, descargado el 2 de noviembre de 2017.

[7] Loxton J.H. Captain Cook and the Loch Ness monster. *James Cook Mathematical Notes*, **27**, 3060–3064, 1981.

[8] Loxton J.H. The graphs of exponential sums. *Mathematika*, **30**, 153–163, 1983.

[9] Loxton J.H. The distribution of exponential sums. *Mathematika*, **32**, 16–25, 1985.

[10] Paris R.B. An asymptotic approximation for incomplete Gauss sums. *Journal of Computational and Applied Mathematics*, **180**, 461–477, 2005.

[11] Paris R.B. The asymptotics of a new exponential sum. *Journal of Computational and Applied Mathematics*, **223**, 314–325, 2009.

[12] Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* Londres: Penguin, 1991.

El algoritmo Gibbs sampler

Jorge Francisco de la Vega Góngora
Profesor del Departamento de Estadística del ITAM

"All probabilities are conditional on something, and to be useful they must condition on the right thing."
Frank Harrell

El *Gibbs Sampler* (GS), o muestreador de Gibbs como le llamaríamos en español, es uno de los avances en el ámbito estadístico más notables del siglo XX que ha facilitado un crecimiento exponencial en las aplicaciones prácticas, especialmente aquellas que tienen un enfoque Bayesiano. Las ideas asociadas al GS permiten descomponer problemas muy complejos en conjuntos de problemas más pequeños y fáciles de resolver. Este algoritmo es una de las piezas claves que permitió conectar una serie de ideas que resultaron en la consolidación de la Estadística Bayesiana y su aplicación intensiva en la ciencia actual.

Introducción

Los métodos de Monte Carlo (MC) estudian la generación de una muestra de observaciones $\boldsymbol{\theta}_1, \ldots \boldsymbol{\theta}_m$ provenientes de una función de distribución $F(\boldsymbol{\theta})$ específica, que puede ser univariada o multivariada. Usualmente tales muestras se utilizan para simular el comportamiento aleatorio de fenómenos o sistemas complejos y tratar de comprender su funcionamiento bajo un ambiente controlado y poder realizar inferencias.

En ese contexto se han propuesto una gran variedad de algoritmos y métodos. La primera generación de métodos MC, fueron desarrollados por Nicholas Metropolis, Stanislaw Ulam y John Von Neumann, entre otros, durante los años 40's y 50's. Se basan en generar muestras *independientes* de la distribución de interés, a partir de números uniformes independientes, o a través de métodos indirectos, como el método de aceptación y rechazo, que se asemeja al lanzamiento de dardos en una región y 'aceptando' los puntos que caen en la región de interés. Aún cuando estos métodos pueden ser eficientes y fáciles de implementar en el caso univariado, resultan complejos en dimensiones mayores.

Otro conjunto de métodos MC fueron desarrollados por el mismo Metropolis, en conjunto con el par de matrimonios de científicos Rosenblueth y Teller, para obtener muestras de la distribución de interés, pero en donde las observaciones ya no son independientes. Estos métodos se basan en generar observaciones de una cadena de Markov ergódica cuya distribución límite es la distribución de interés. El algoritmo de Metropolis fue perfeccionado en 1970 por W. Keith Hastings, quien lo generalizó y le dio la forma final al algoritmo de Metropolis-Hastings que se reconoce hasta hoy y que se conoce como *Markov Chain Monte Carlo* (MCMC). Más detalle sobre esta parte de la historia se puede consultar, por ejemplo en [6].

Aun cuando se ha probado que el GS es un caso particular del algoritmo de Metropolis-Hastings, aquel tuvo un origen diferente y no fue hasta varios años después que se ha mostrado la conexión entre los dos.

El procesamiento de imágenes y su análisis es un tema relevante para los militares, la industria de la automatización y el diagnóstico médico. Las imágenes borrosas requieren el procesamiento de señales, eliminación de ruido y afinación para hacerlas reconocibles. A principios de la década de los 80's, Ulf Grenader en Brown diseñaba modelos matemáticos para imágenes médicas explorando el efecto que un pixel puede tener en unos cuantos de sus vecinos. Los cálculos involucraban fácilmente más de un millón de variables desconocidas. Stuart Geman atendía el seminario de Grenader y junto con su hermano Donald Geman trataban de restaurar una fotografía borrosa de una señal en el camino.

Stuart y Donald inventaron una variante MC que fue particularmente adecuada para problemas de restauración de imágenes. Decidieron nombrar el método por analogía del problema que estaban resolviendo al de la configuración de una caja de chocolates. Un regalo popular para el día de las madres en esa época era la "*Whitman's sampler*" que contenía dentro de la caja un diagrama que identificaba el relleno oculto de cada chocolate. Como el diagrama era muy parecido a una matriz de variables ocultas, decidieron llamar al método como el **Gibbs Sampler**, ya que en su problema utilizaban la distribución de Gibbs, en honor al físico americano del siglo XIX, Josiah Williard Gibbs.

El propósito de esta nota es explicar, de manera sencilla e intuitiva, y sin entrar en formalidad matemática estricta, las ideas fundamentales del algoritmo y mostrar con ejemplos algunas de sus aplicaciones prácticas.

Marco Teórico

Conceptos y algoritmo

En simples palabras, la idea del GS es que las distribuciones condicionales determinan las distribuciones marginales. Si $\pi(\boldsymbol{\theta})$ es una densidad conjunta donde $\boldsymbol{\theta}$ es un vector particionado en $p > 1$ subvectores $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ el problema a resolver es obtener la densidad marginal para un subconjunto de las variables en el vector $\boldsymbol{\theta}$, digamos $\boldsymbol{\theta}_i$:

$$\pi_i(\boldsymbol{\theta}_i) = \int \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{-i},$$

donde $\boldsymbol{\theta}_{-i} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_p)$. La integral anterior puede ser muy difícil de obtener (analítica o numéricamente).

El Gibbs Sampler permite generar una muestra $\boldsymbol{\theta}_i^{(1)}, \ldots . \boldsymbol{\theta}_i^{(m)}$ de la distribución $\pi_1(\boldsymbol{\theta}_i)$ sin requerir explícitamente la distribución π_1, a partir de las distribuciones condicionales:

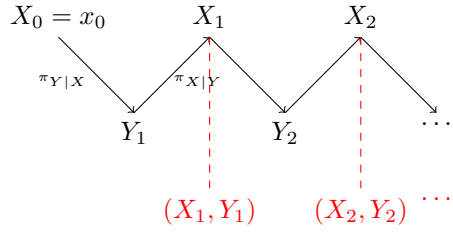
$$\pi(\boldsymbol{\theta}_j | \boldsymbol{\theta}_{-j}), \quad j = 1, \ldots, p$$

iterando sobre los p subvectores de la partición del vector $\boldsymbol{\theta}$. En particular, permite obtener una muestra de la distribución completa $\pi(\boldsymbol{\theta})$ a partir de las distribuciones marginales.

Para aterrizar ideas, consideremos el caso bidimensional, $(X, Y) \sim \pi(x, y)$. El algoritmo propone generar una sucesión de parejas (X_j, Y_j) del siguiente modo. Para $j = 1, 2, \ldots, n$:

1. Se inicializa $X_0 = x_0$.
2. Genera $Y_j \sim \pi_{Y|X}(\cdot|X_{j-1})$.
3. Genera $X_j \sim \pi_{X|Y}(\cdot|Y_j)$
4. Hacer $j := j + 1$.

Una representación gráfica del flujo del algoritmo se muestra a continuación:



La sucesión de valores $X_0, Y_0, X_1, Y_1, \ldots, X_k, Y_k$ se conoce como **sucesión de Gibbs**. Hay dos maneras de tomar muestras para π_X: ya sea tomando los valores de la sucesión de Gibbs a partir de una k grande, lo que daría una muestra de observaciones de π_X no independientes, o bien, si se generan m sucesiones de Gibbs independientes de longitud k, y tomando las "últimas" observaciones de esas m sucesiones, suponiendo nuevamente que k es suficientemente grande.

Para el caso con más de dos variables, se particiona el vector $\boldsymbol{\theta}$ en k componentes $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_k)$, y si podemos simular observaciones de $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i})$, entonces podemos aplicar el siguiente algoritmo. Para $j = 1, \ldots, n$:

1. Inicializar $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \ldots, \boldsymbol{\theta}_k^{(0)})$ en un punto relativamente arbitrario.

2. Obtener $\boldsymbol{\theta}^{(j)}$ a partir de $\boldsymbol{\theta}^{(j-1)}$ generando

$$\begin{aligned}
\boldsymbol{\theta}_1^{(j)} &\sim \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(j-1)},\ldots,\boldsymbol{\theta}_k^{(j-1)})\\
\boldsymbol{\theta}_2^{(j)} &\sim \pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(j)},\boldsymbol{\theta}_3^{(j-1)},\ldots,\boldsymbol{\theta}_k^{(j-1)})\\
\boldsymbol{\theta}_3^{(j)} &\sim \pi(\boldsymbol{\theta}_3|\boldsymbol{\theta}_1^{(j)},\boldsymbol{\theta}_2^{(j)},\ldots,\boldsymbol{\theta}_k^{(j-1)})\\
&\vdots\\
\boldsymbol{\theta}_{k-1}^{(j)} &\sim \pi(\boldsymbol{\theta}_{k-1}|\boldsymbol{\theta}_1^{(j)},\boldsymbol{\theta}_2^{(j)},\ldots,\boldsymbol{\theta}_k^{(j-1)})\\
\boldsymbol{\theta}_k^{(j)} &\sim \pi(\boldsymbol{\theta}_k|\boldsymbol{\theta}_1^{(j)},\boldsymbol{\theta}_2^{(j)},\ldots,\boldsymbol{\theta}_{k-1}^{(j)})
\end{aligned}$$

3. Hacer $j := j+1$ y repetir a partir del paso 2 hasta alcanzar n.

Funcionamiento del algoritmo

En su artículo, Stuart y Donald Geman [8] demuestran la convergencia del GS. Sin embargo, su demostración es oscura y compleja, utilizando lenguaje no estadístico y haciendo uso de conceptos de construcción de imágenes digitales, redes neuronales y sistemas expertos.

La naturaleza Markoviana del GS es mostrada en el siguiente ejemplo de Casella y George [4], que seguiremos aquí. No pretende ser la prueba del algoritmo sino ilustrar su fundamento a través de un caso particular.

Consideramos $\boldsymbol{\theta} = (X, Y)$ dos variables Bernoulli con distribución conjunta dada por

$$P(X = i|Y = j) = p_{ij} \text{ con } p_{ij} \geq 0, \quad p_{11} + p_{12} + p_{21} + p_{22} = 1.$$

La distribución marginal de X, por ejemplo, es Bernoulli con parámetro $p_{21} + p_{22}$.

Por otra parte, las distribuciones condicionales se pueden expresar en dos matrices, $A_{x|y}$ y $A_{y|x}$ con los siguientes elementos:

$$A_{x|y} = \begin{bmatrix} \frac{p_{11}}{p_{11}+p_{21}} & \frac{p_{21}}{p_{11}+p_{21}} \\ \frac{p_{12}}{p_{12}+p_{22}} & \frac{p_{22}}{p_{12}+p_{22}} \end{bmatrix} \qquad \text{y} \qquad A_{y|x} = \begin{bmatrix} \frac{p_{11}}{p_{11}+p_{12}} & \frac{p_{12}}{p_{11}+p_{12}} \\ \frac{p_{21}}{p_{21}+p_{22}} & \frac{p_{22}}{p_{21}+p_{22}} \end{bmatrix}.$$

Por ejemplo, $P(Y = 1|X = 0)$, el elemento (1,2) de $A_{y|x}$:

$$P(Y = 1|X = 0) = \frac{P(Y = 1, X = 0)}{P(X = 0)} = \frac{p_{12}}{p_{11} + p_{12}}.$$

Para generar una sucesión de Gibbs, $Y_0, X_0, Y_1, X_1, \ldots, Y_k, X_k, \ldots$ las matrices $A_{y|x}$ y $A_{x|y}$ se pueden considerar como las matrices de transición dando las probabilidades condicionales $P(Y = i|X = j)$ y viceversa.

Si queremos obtener la distribución marginal de X, nos concentramos sólo en las observaciones X's de la sucesión de Gibbs. Podemos pensar entonces en la cadena de Markov que lleva de los estados iniciales de X a estados posteriores de X, pasando por los estados de Y, con matriz de transición:

$$A_{x|x} = A_{y|x}A_{x|y}.$$

Entonces podemos ver que la matriz de transición en k pasos, que nos da las probabilidades $P(X_k = i|X_0 = j)$ es $A^k_{x|x}$. La distribución marginal de X_k se puede escribir como $\pi_k = (\pi_k(0), \pi_k(1))$ y entonces podemos ver que se cumple la siguiente relación de recurrencia:

$$\pi_k = \pi_0 A^k_{x|x} = (\pi_0 A^{k-1}_{x|x})A_{x|x} = \pi_{k-1}A_{x|x}.$$

Si las entradas de $A_{x|x}$ son positivas y dada cualquier distribución inicial π_0, π_k converge a una única distribución π que satisface la condición de punto fijo [1]: $\pi = \pi A_{x|x}$. Entonces, si la sucesión de Gibbs converge, la π que satisface esta recurrencia es la distribución marginal de X, $\pi = \pi_x$.

Ejemplos y Aplicaciones

GS en modelos jerárquicos

Comenzaremos con una de las aplicaciones más comunes del GS. Dennis Lindley junto con uno de sus estudiantes, Adrian F. M. Smith, descompusieron varios problemas complejos en modelos jerárquicos, en donde resulta muy fácil aplicar el GS ya que se cuenta con las condicionales de manera explícita.

Consideraremos el modelo jerárquico Binomial-Beta:

$$\begin{aligned} X|\theta &\sim \mathbf{Bin}\,(n, \theta) \\ \theta &\sim \mathcal{B}e\,(a, b)\,. \end{aligned}$$

En este modelo densidad conjunta de X y θ se puede escribir como

$$\begin{aligned} f(X,\theta) = f(X|\theta)f(\theta) = f(\theta|X)f(X) &= \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\ f(\theta|X) = \frac{f(X,\theta)}{f(X)} &\propto \binom{n}{x}\theta^{x+a-1}(1-\theta)^{n-x+b-1}. \end{aligned}$$

Entonces $\theta|X \sim \mathcal{B}e\,(x + a, n - x + b)$ (beta y binomial son familias conjugada). Así que para generar de la distribución conjunta (X, θ) se puede generar de $f(X|\theta)$ y de $f(\theta|X)$. También podemos tener la distribución marginal de X. Queremos obtener muestras de la distribución conjunta de los datos y el parámetro. En el siguiente cálculo numérico se considera $X|\theta$ una binomial con $n = 15$, y para obtener θ utilizamos $a = 3$ y $b = 7$. El siguiente código de `R` obtiene una muestra de tamaño 5,000 para los pares (X, θ).

[1] Esto se puede revisar en cualquier libro básico de procesos estocásticos, por ejemplo, [2].

```r
set.seed(12345)
nsim <- 5000
n <- 15         #Parámetro dado
a <- 3; b <- 7 #Parámetros de Beta
Theta <- rbeta(1,a,b) #Valores iniciales
X <- rbinom(1,n,Theta[1])
for(i in 2:nsim){
        X[i] <- rbinom(1,n,Theta[1])
        Theta[i] <- rbeta(1,a+X[i],n-X[i]+b)
}
```

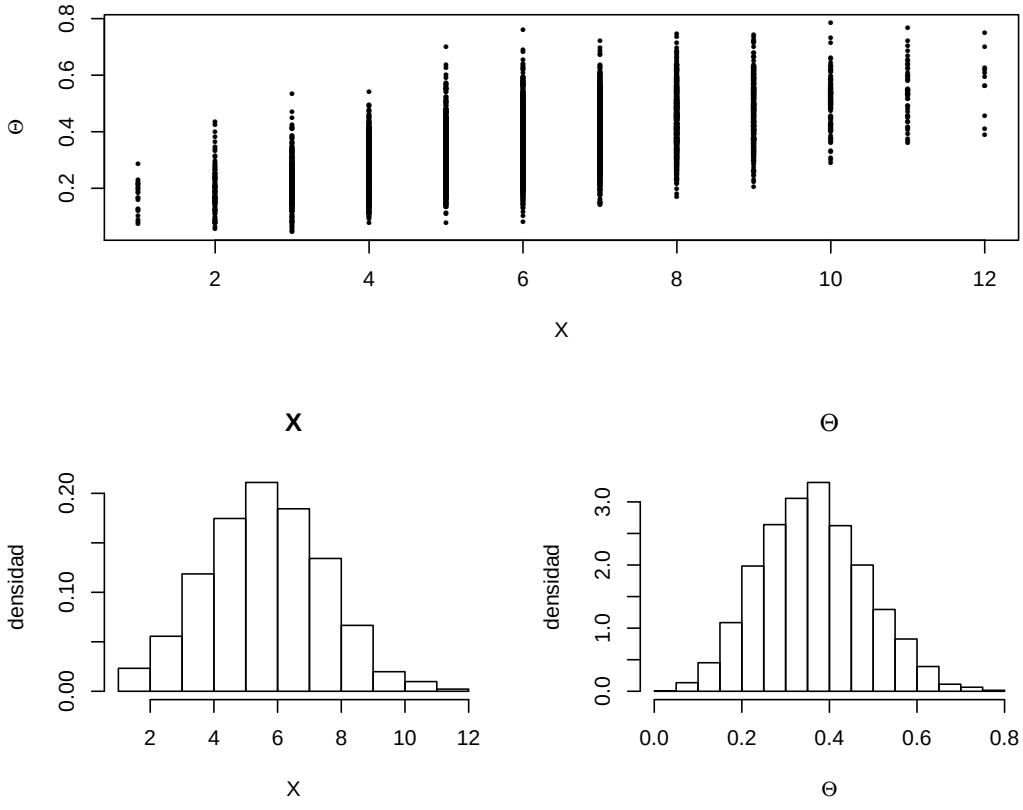


Figura 1: Simulación de la distribución conjunta de (X, θ) y sus distribuciones marginales

Modelo de Ising

El siguiente ejemplo muestra una aplicación del GS que utiliza una distribución de Gibbs como en el artículo de Geman y Geman [8]. El modelo de Ising fue propuesto como problema de tesis por Whilhem Lenz a Ernst Ising, su alumno, que lo resolvió en 1925. Es un modelo matemático de ferromagnetismo en mecánica estadística, pero que también ha sido usado para el procesamiento de imágenes digitales. El modelo de Ising se estudia, entre otras razones, por sus propiedades de transición de fase.

En una gráfica retícula de $n \times n$ como la siguiente de 4×4:

σ_{11}	σ_{12}	σ_{13}	σ_{14}
σ_{21}	σ_{22}	σ_{23}	σ_{24}
σ_{31}	σ_{32}	σ_{33}	σ_{34}
σ_{41}	σ_{42}	σ_{43}	σ_{44}

A cada vértice v se le asigna un color, por ejemplo blanco con valor +1 o rojo con valor -1. Una configuración $\boldsymbol{\sigma} = (\sigma_v)$ es una asignación de colores en los vértices de la retícula. Hay n^2 vértices y 2^{n^2} posibles configuraciones. Cada configuración tiene una energía asociada, dada por la función $E(\boldsymbol{\sigma}) = -\sum_{v \sim w} \sigma_v \sigma_w$, donde la relación '$\sim$' es de vecindad (cada vértice tiene cuatro vecinos: arriba, abajo, izquierda y derecha, excepto en la frontera de la retícula). La configuración dada en el ejemplo de arriba tiene energía $E(\boldsymbol{\sigma}) = 12$.

La distribución de Gibss es una distribución de probabilidad sobre el conjunto de configuraciones con un parámetro T:

$$\pi(\boldsymbol{\sigma}) = \frac{e^{-E(\boldsymbol{\sigma})/T}}{\sum_{\boldsymbol{\tau}} e^{-E(\boldsymbol{\tau})/T}}.$$

El parámetro T tiene interpretación física de temperatura. En este modelo, si la temperatura es infinita, la distribución de Gibbs es uniforme sobre el conjunto de configuraciones. En dos dimensiones, el sistema entra en un cambio radical de conducta a una temperatura crítica de $T = 2.269$. Los detalles se pueden ver en [10].

El GS se puede usar fácilmente para simular este modelo. Dada una configuración $\boldsymbol{\sigma}$, un vértice v se escoge uniformemente al azar. El color en ese sitio se actualiza de la distribución condicional de ese vértice dados los otros vértices de la configuración $\boldsymbol{\sigma}$.
Si denotamos σ_{-k} a la configuración sin el vértice k, y $\boldsymbol{\sigma}^+$ a la configuración con un +1 en el vértice k y $\boldsymbol{\sigma}^-$ a la configuración con un -1 en ese vértice, entonces podemos calcular la probabilidad condicional:

$$
\begin{aligned}
P(\sigma_k = +1 | \sigma_{-k}) &= \frac{P(\boldsymbol{\sigma}^+)}{P(\boldsymbol{\sigma}_{-k})} = \frac{P(\boldsymbol{\sigma}^+)}{P(\boldsymbol{\sigma}^+) + P(\boldsymbol{\sigma}^-)} && (1)\\
&= \frac{\exp\{-E(\boldsymbol{\sigma}^+)/T\}}{\exp\{-E(\boldsymbol{\sigma}^+)/T\} + \exp\{-E(\boldsymbol{\sigma}^-)/T\}} && (2)\\
&= \frac{1}{1 + \exp\{(E(\boldsymbol{\sigma}^+) - E(\boldsymbol{\sigma}^-))/T\}} && (3)\\
&= \frac{1}{1 + \exp\{-2\sum_{i\sim k} \sigma_i\}}. && (4)
\end{aligned}
$$

La última igualdad se obtiene notando que en $E(\boldsymbol{\sigma}^+)$ y $E(\boldsymbol{\sigma}^-)$ se pueden separar los sumandos que no son vecinos del vértice k y se cancelan entre sí, sobreviviendo sólo los vértices vecinos de k.

Implementamos el GS tomando un vértice al azar y actualizando cada vértice de acuerdo a las distribuciones condicionales obtenidas.

```
g <- 150 #Elementos de la retícula
valT <- c(Inf,2.267574,1,-1) #valores de temperatura
final <- list(NULL)

for(v in valT){
nsim <- 100000 #número de iteraciones
M <- matrix(sample(c(-1,1),(g+2)^2,rep=T),nrow=g+2) #retícula
M[c(1,g+2), ] <- 0 #asigna 0 a los extremos de la retícula valor 0
M[ ,c(1,g+2)] <- 0

for(m in 1:nsim){
i <- sample(2:(g+1),1) #muestrea renglón y columma
j <- sample(2:(g+1),1)
E <- M[i,j+1] + M[i,j-1] + M[i-1,j] + M[i+1,j] #calcula la energía del vértice
p <- 1/(1+exp(-2*E/v)) #probabilidad del vértice
if(runif(1)<p) M[i,j] <- 1 else M[i,j] <- -1 #actualiza el color del vértice
}
final[[match(v,valT)]] <- M[2:(g+1),2:(g+1)]
}
```

La simulación del modelo de Ising se realizó generando 100,000 iteraciones en un grid de 150×150 para cuatro valores de la temperatura. En este contexto, el espacio de estados de las configuraciones tiene 2^{22500} elementos. La simulación prácticamente no toma tiempo de ejecución en una laptop y se puede observar cómo impacta la temperatura en la concentración de energía.

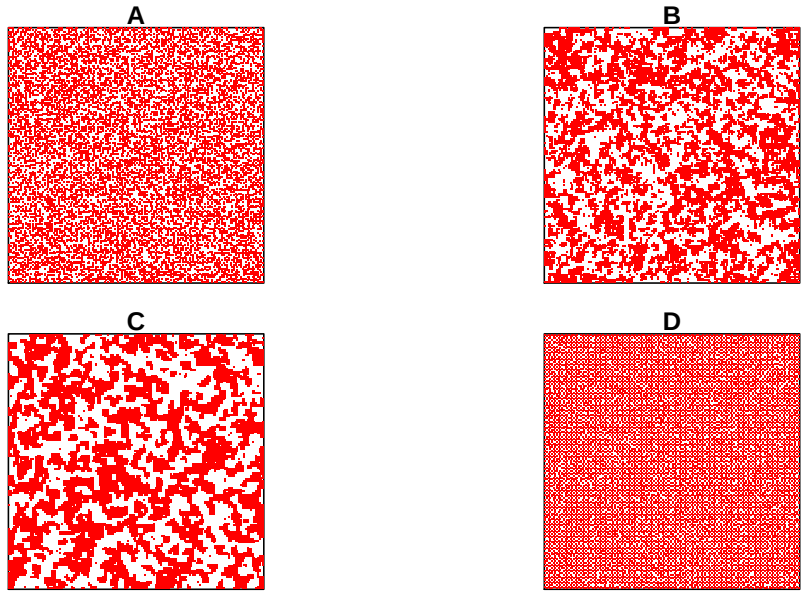


Figura 2: Simulación del modelo de Ising para un modelo con grid 150^2. A: $T = \infty$, B: 2.267574, C: 1, D: -1

Análisis de Punto de Cambio

En este último ejemplo consideramos un proceso Poisson con punto de cambio, esto es, un punto en el tiempo donde la tasa de eventos cambia:

$$X_t \sim \begin{cases} \mathcal{P}(\mu t) & 0 < t \leq k \\ \mathcal{P}(\lambda t) & t > k \end{cases}.$$

Dada una muestra de n observaciones del proceso anterior, el problema consiste en estimar μ, λ y k. Este es un proceso muy común y que ha sido ampliamente estudiado. Hay varias opciones de especificación de modelos Bayesianos para resolverlo. Una aplicación concreta del problema anterior se relaciona con los datos publicados en el artículo de Jarret [9]. Los datos corresponden a las fechas de 191 explosiones en minas de carbón que resultaron en más de 10 fatalidades desde marzo 15, 1851 hasta marzo 22, 1962. Los datos se encuentran en `coal`, en el paquete `boot` de `R`.

Los datos muestran un cambio en el número promedio de desastres por año alrededor de 1900. Los números anuales de accidentes se muestran a continuación:

1.6180339887498948482045868343656381177203091798057628621354486227052604628189037

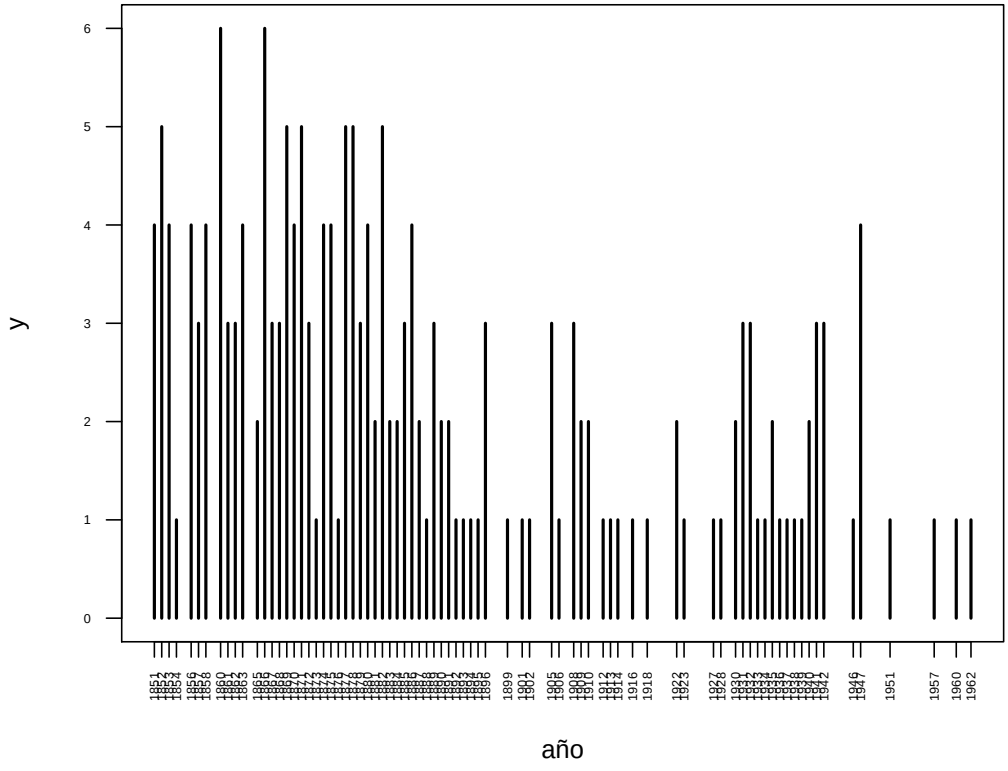


Figura 3: Frecuencia de accidentes fatales en minas de carbón entre 15/03/1851 y 22/03/1962.

```
y

  [1] 4 5 4 1 0 4 3 4 0 6 3 3 4 0 2 6 3 3 5 4 5 3 1 4 4 1 5 5 3 4 2 5 2 2 3
 [36] 4 2 1 3 2 2 1 1 1 1 3 0 0 1 0 1 1 0 0 3 1 0 3 2 2 0 1 1 1 0 1 0 1 0 0
 [71] 0 2 1 0 0 0 1 1 0 2 3 3 1 1 2 1 1 1 1 2 3 3 0 0 0 1 4 0 0 0 1 0 0 0 0
[106] 0 1 0 0 1 0 1
```

Modelamos estos datos de la siguiente manera. Definimos Y_i como el número de desastres en el año i (1851=1). Entonces, si k es el año punto de cambio,

$$Y_i \sim \mathcal{P}(\mu), \qquad i = 1, \ldots k,$$
$$Y_i \sim \mathcal{P}(\lambda), \qquad i = k+1, \ldots, n$$

Hay $n = 112$ observaciones terminando en el año 1962. El parámetro es $\boldsymbol{\theta} = (k, \mu, \lambda)$. Se requiere estimar como distribución objetivo la posterior $\pi(k, \mu, \lambda|y)$, y en particular, la distribución posterior de k, $\pi(k|y, \mu, \lambda)$.

Se puede construir un modelo Bayesiano con las siguientes distribuciones iniciales *independientes*:

$$\begin{aligned} k &\sim U\{1, 2, \ldots, n\}, \\ \mu &\sim \mathcal{G}(a_1, b_1), \\ \lambda &\sim \mathcal{G}(a_2, b_2) \end{aligned}$$

donde a_1, a_2, b_1 y b_2 son hiperparámetros que se pueden fijar o bien, se pueden considerar a su vez aleatorios. Sean $S_k = \sum_{i=1}^{k} Y_i$ el total de eventos hasta el año k y $S_k^{'} = S_n - S_k$ el total de eventos entre el año $k+1$ y el año n. Para aplicar GS, se necesita especificar las distribuciones condicionales posteriores. Las densidades condicionales para k, μ y λ están dadas por:

$$\begin{aligned} \mu|y, k &\sim \mathcal{G}(a_1 + S_k, k + b_1) \\ \lambda|y, k &\sim \mathcal{G}\left(a_2 + S_k^{'}, n - k + b_2\right) \\ k|y, \mu, \lambda &\sim \frac{L(Y|k, \mu, \lambda)}{\sum_{j=1}^{n} L(Y|j, \mu, \lambda)} \end{aligned}$$

donde L es la función de verosimilitud

$$L(Y|k, \mu, \lambda) = e^{k(\lambda - \mu)} \left(\frac{\mu}{\lambda}\right)^{S_k}.$$

A continuación se muestra el código en `R` para llevar a cabo la simulación.

```
#Simulación de los datos para la distribución del punto de cambio.
n <- length(y) #longitud de los datos
m <- 2000      #longitud de la cadena
mu <- lambda <- k <- numeric(m)
L <- numeric(n) #inicializo vector de valores de verosimilitud
k[1] <- sample(1:n,1) #valor inicial del punto de cambio (seleccionado al azar)
mu[1] <- 1
lambda[1] <- 1
b1 <- b2 <- 1 #valores de los hiperparámetros
a1 <- a2 <- 2

for (i in 2:m){
        #genera mu
        mu[i] <- rgamma(1, shape = a1 + sum(y[1:k[i-1]]), rate = k[i-1] + b1)
         #genera lambda
        lambda[i] <- rgamma(1, shape = a2 + sum(y)-sum(y[1:k[i-1]]),
                            rate= n - k[i-1] + b2)
        for(j in 1:n){
        L[j] <- exp((lambda[i]-mu[i])*j) * (mu[i]/lambda[i])^sum(y[1:j])
}
L <- L/sum(L)
```

```
#genera k de la distribución discreta L en 1:n
k[i] <- sample(1:n,prob=L,size=1)
}
```

La Figura 4 muestra las gráficas de las trazas de las cadenas para los tres parámetros muestreados, y la Figura 5 muestra histogramas de las densidades marginales obtenidas.

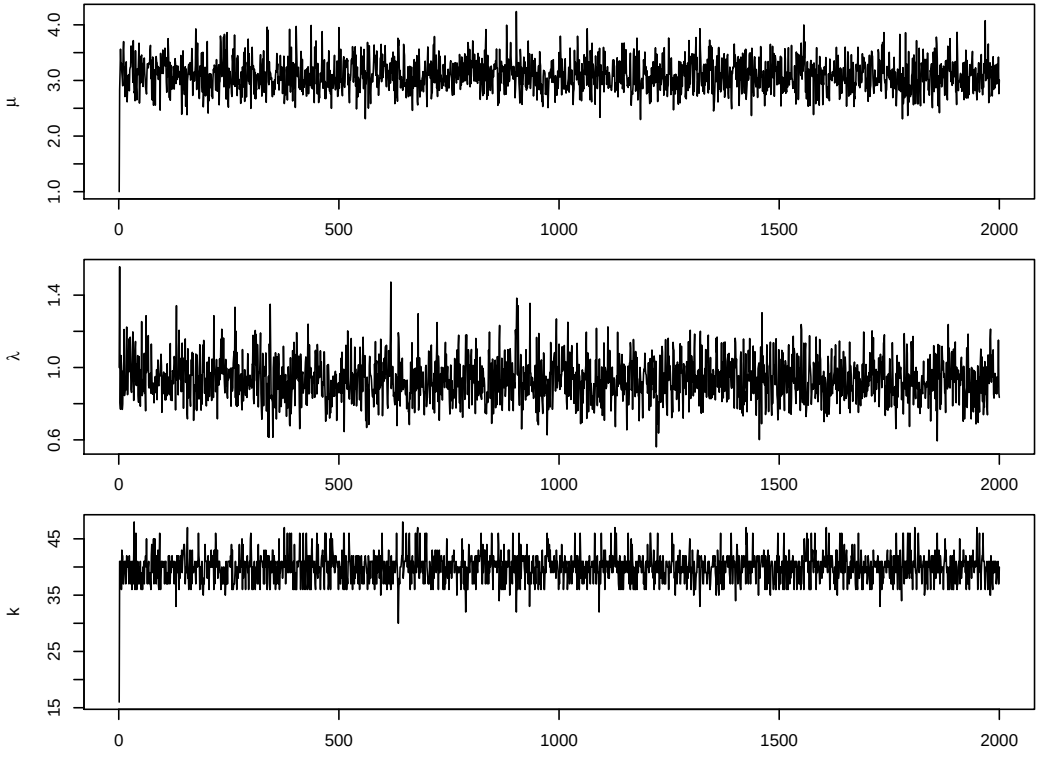


Figura 4: Trazas de las cadenas generadas por GS para los tres parámetros considerados: μ, λ y k

Entonces el punto de cambio está alrededor de $\hat{k} = 40$, que corresponde al año $1851 + 40 = 1891$. De 1851 a 1890 la media Poisson es alrededor de $\hat{\mu} \approx 3.1$ y del año 1891 hacia adelante la media es $\hat{\lambda} \approx 0.93$.

```r
# histogramas de la salida del Gibbs sampler:
b <- 500 #burn-in
par(mfrow=c(1,3))
label1 <- paste("mu=",round(mean(mu[b:m]),1))
label2 <- paste("lambda=", round(mean(lambda[b:m]),1))
hist(mu[b:m], main="", xlab=label1, prob=TRUE) #mu posterior
hist(lambda[b:m], main="", xlab=label2, prob=TRUE) #lambda posterior
hist(k[b:m], breaks = min(k[b:m]):max(k[b:m]), prob=TRUE, main="", xlab = "Punto de cambio k")
```

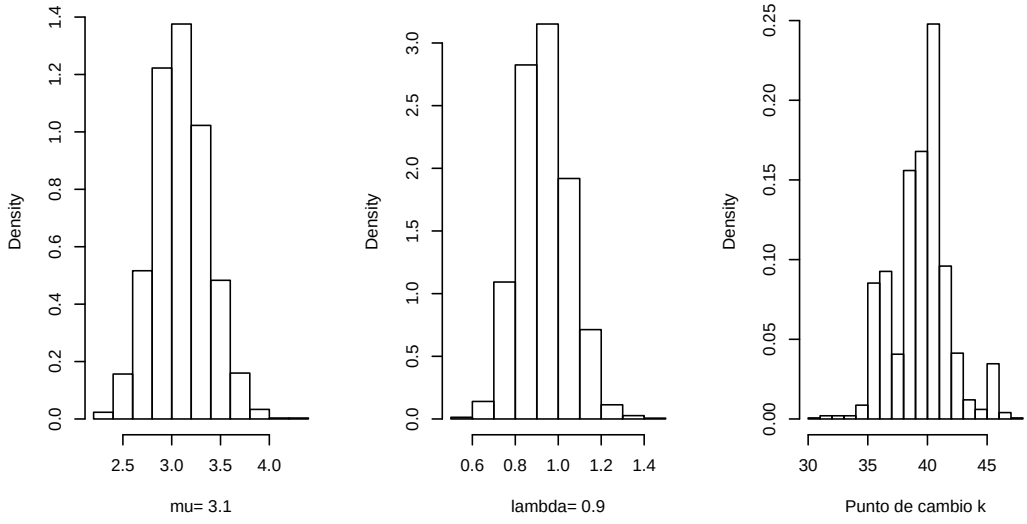


Figura 5: Histogramas de las distribuciones marginales para los tres parámetros estimados

Conclusiones

El Gibbs sampler es una herramienta poderosa de simulación que ha tenido un impacto significativo en la investigación científica en general, permitiendo que una gran variedad de planteamientos bayesianos puedan resolverse computacionalmente y junto con el algoritmo de Metropolis-Hastings, cambiaron por completo el paradigma estadístico para resolver problemas.

Hay varios aspectos relevantes que deben ser estudiados con más detalle. Por ejemplo, ¿por cuánto tiempo es necesario simular la cadena para considerar que las observaciones ya están lo suficientemente cerca a la distribución objetivo? De acuerdo a Charles Geyer [11], fuerte contribuidor a la teoría de MCMC, no se requiere esperar, o en otras palabras, lo que se conoce como el período de *burn-in* es innecesario.

En esta nota se habló poco de las herramientas computacionales y todos los ejemplos se hicieron en `R`. Sin embargo, existen muchas otros programas que se han realizado ex-profeso para el cálculo del GS, entre ellas se encuentran `BUGS` (*Bayesian Analysis Using the Gibbs Sampler*), desarrollado por David Spiegelhalter en Cambridge desde 1989 y liberado en 1991. De aquí se deriva `OpenBUGS` que tiene una interface a `R`. También se cuenta con `JAGS` (*Just Another Gibbs Sampler*) y con `Stan`, que tiene soporte para varios lenguajes de programación.

Esta y otras herramientas computacionales importantes son cubiertas en cursos académicos como Estadística Computacional, Simulación Estocástica, Métodos computacionales Bayesianos, entre otras posibilidades. Estas herramientas deberían formar parte del herramental cotidiano de todos los estadístic@s y los ahora llamados *científic@s de datos*.

Referencias

[1] Robert, Christian y Casella, George. *Monte Carlo statistical Methods*. Springer, USA, 2004.

[2] Dobrow, Robert. *Introduction to Stochastic Processes with R*. Wiley, USA, 2016.

[3] Bertsch Mcgrayne, Sharon. *The Theory That Would Not Die*. Yale University Press, USA, 2011.

[4] Casella, George, George, Edward I. "Explaining the Gibbs sampler", *The American Statistician* **46** (1992):167-174.

[5] Eckhardt, Roger. "Stan Ulam, John von Neumann and the Monte Carlo Method", *Los Alamos Science* **15** Special Issue (1987):131-137.

[6] Alvarado, Christian, Diluvi, Gian Carlo y Espinosa, Demian "El algoritmo de Metropolis-Hastings", *Laberintos e Infinitos* **41** (2016):21-29.

[7] Metropolis, N., Rosenblueth, A., Rosenblueth, M., Teller, A. and Teller, E. "Equation of State Calculations by Fast Computing Machines", *J. Chem. Phys.* **21** (1953):1087-1092.

[8] Geman, S. and Geman, D. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **6** (1984):721-741.

[9] Jarret, R. G. "A note on the intervals between coal-mining disasters.", *Biometrika* **66** (1979):191-193.

[10] "Square-lattice Ising model." *Wikipedia*. Consultado el 16 de marzo de 2018. `https://en.wikipedia.org/wiki/Square-lattice_Ising_model`

[11] "Burn-in is unnecessary." *Charles Geyer*. Consultado el 10 de enero de 2018. `http://users.stat.umn.edu/~geyer/mcmc/burn.html`

Diseño Bayesiano de experimentos

Gian Carlo Diluvi
Exalumno de Matemáticas Aplicadas del ITAM

To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.
-Ronald Fisher

Introducción

El diseño de experimentos se refiere a la selección de todos los aspectos relevantes de un experimento, y se realiza antes de la recolección de los datos (la cual generalmente está restringida por los recursos disponibles). Estos aspectos son, principalmente, cuántas observaciones del experimento se pueden realizar, qué variables independientes se medirán, a qué niveles (si son factores, por ejemplo), cuántas observaciones hacer por cada variable y nivel, así como qué tipo de modelo se planteará, entre muchos otros. Además, se debe especificar un criterio de optimalidad del experimento. La importancia del diseño experimental recae en la eficiencia ante restricciones de recursos, ya que permite extraer la mayor cantidad de información de calidad utilizando los recursos disponibles.

El diseño de experimentos ha pasado por varias etapas, comenzando con el trabajo de Ronald Fisher [8] en las primeras décadas del Siglo XX. Hoy en día las técnicas del diseño experimental han sido estudiadas y desarrolladas por un sinfín de personajes. Más aún, las aplicaciones del diseño de experimentos han evolucionado, superando por mucho sus orígenes agriculturales (las aplicaciones discutidas por Fisher). Prácticamente todas las áreas científicas y de ingeniería han realizado experimentos diseñados estadísticamente.

En cuanto al enfoque Bayesiano para el diseño de experimentos, es importante destacar que éste en general ha tenido un desarrollo mucho más lento que el de la contraparte clásica. Esto se debe principalmente a la complejidad matemática que ocurre en el paradigma Bayesiano; hasta hace un par de años, las soluciones que se buscaban eran analíticas y, por lo tanto, escasas. Si bien con el descubrimiento de los diversos métodos computacionales esto ha mejorado, también es cierto que el avance ha sido menor que el de otras áreas de la Estadística Bayesiana.

El propósito de este artículo es plantear el problema de diseño de experimentos como uno de decisión, para así resolverlo (a nivel teórico) desde un enfoque Bayesiano (ver [6]). En particular, se pensará que el modelo a utilizar para modelar los datos es un modelo lineal generalizado [7, 12]. También se revisará el método ACE de Overstall y Woods [15, 16], uno de los algoritmos más recientes para encontrar diseños óptimos Bayesianos. Finalmente, se implementará este algoritmo para encontrar un diseño óptimo de un ejemplo tomado de [21].

Si se desea ahondar en alguno de los temas presentados, se recomienda ampliamente al lector revisar Montgomery [13] para un panorama general del tema y Chaloner y Verdinelli [5] para una versión Bayesiana, así como la bibliografía ahí citada.

Diseño Bayesiano de experimentos

Ya que el diseño de experimentos se realiza antes de recopilar los datos, éste se puede considerar por lo menos implícitamente Bayesiano. [5], siguiendo la idea de [11, pp. 20-21], presentan un enfoque del diseño experimental basado en la Teoría de la Decisión.

Supongamos que nuestro experimento involucra p variables y que el tamaño de muestra permitido es de n observaciones. La i-ésima observación será y_i, la cual será generada por los valores $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T \in \mathcal{X} \subset \mathbb{R}^p$ de las covariables. Un diseño $\eta = \{x_1, ..., x_n\} \in \mathcal{H}$ se refiere a la colección de todos los valores que toman todas las covariables. La matriz de diseño de un experimento η se define como $X = (x_{ij})$, donde i indexa las observaciones y j las variables. Dicho diseño ocasionará que se observen los datos $y = (y_1, ..., y_n)^T \in \mathcal{Y}$. Además se supondrá que y tiene función de probabilidad generalizada $p(y \mid \theta)$, donde el parámetro[1] $\theta \in \Theta$ tiene distribución inicial $p(\theta)$. Se supondrá que y es una observación de una variable aleatoria Y perteneciente a la familia exponencial de distribuciones.

Por otro lado, se considerará una función de pérdida $l(\eta, y, \theta)$. Es de suma importancia que ésta refleje adecuadamente los objetivos del experimento (e.g. obtener la mejor predicción o minimizar la varianza de los estimadores). Así pues, juntando todo lo anterior tenemos que la pérdida esperada de un diseño cualquiera $\eta \in \mathcal{H}$ es

$$\mathbb{E}\left[l(\eta, y, \theta)\right] = \int_{\mathcal{Y}} \int_{\Theta} l(\eta, y, \theta)\, p(\theta \mid y, \eta)\, p(y \mid \eta)\, d\theta\, dy.$$

Como $p(\theta \mid y, \eta)\, p(y \mid \eta) = p(\theta, y \mid \eta)$, el diseño óptimo η^* es entonces

$$\eta^* = \arg\min_{\eta \in \mathcal{H}} \int_{\mathcal{Y}} \int_{\Theta} l(\eta, y, \theta)\, p(\theta, y \mid \eta)\, d\theta\, dy. \tag{1}$$

En principio esa es la solución del problema de diseño de experimentos. Sin embargo, hay algunas observaciones que realizar. En primer lugar, la elección de la función de pérdida o utilidad es un tema sumamente delicado. No hay un consenso sobre una elección idónea, aunque en la literatura existen diversos artículos dedicados a discutir las opciones más populares; se ahondará en este tema más adelante.

Por otro lado, hay que resaltar de nuevo la complejidad computacional en el cálculo de la pérdida esperada (1). Como Woods et al. mencionan en [21], existen algunos problemas que se presentan en la práctica:

[1] Puede ser un parámetro de varias dimensiones.

a) La evaluación de l puede ser sumamente complicada, ya que puede depender de la distribución final de θ; en muchas ocasiones solo es posible obtener valores de ésta numéricamente, e incluso en ese caso hacerlo no es trivial.

b) Las integrales en (1) tienden a ser de dimensión alta (la dimensión del parámetro más el número de observaciones), por lo que su cálculo se debe realizar con métodos numéricos ingeniosos y que logren sobrepasar la maldición de la dimensión.[2]

Funciones de pérdida

Hay una gran variedad de funciones de pérdida que se utilizan comúnmente para problemas de diseño experimental (ver [5, Capítulos 2.2–2.5]). Para propósitos de este artículo se utilizará

$$l_{\mathrm{D}}(\eta, y, \theta) = \log\left(\frac{1}{p(\theta \mid y, \eta)}\right). \tag{2}$$

Generalmente se define

$$\Phi_{\mathrm{D}}(\eta) = \int_{\mathcal{Y}} \int_{\Theta} \log\left(\frac{1}{p(\theta \mid y, \eta)}\right) p(\theta, y \mid \eta)\, d\theta\, dy. \tag{3}$$

Dicha función corresponde a la información esperada de Shannon de la distribución final de θ, y minimizarla es equivalente a maximizar la distancia esperada de Kullback-Leibler entre las distribuciones inicial y final [5, p. 5]. Un diseño óptimo bajo dicha función de pérdida se conoce como D-óptimo; en general, dicha función da lugar a la D-optimalidad Bayesiana. El nombre se debe a la similitud con la contraparte clásica, la cual también cuenta con su propia D-optimalidad. En ese caso el criterio de optimalidad es minimizar el determinante de la matriz de información de Fisher del modelo. Chaloner y Verdinelli [5] argumentan que la D-optimalidad Bayesiana es equivalente a minimizar el determinante de la matriz de información de Fisher del modelo, más la matriz de precisión inicial de θ.

Método ACE

Si bien la teoría detrás del diseño Bayesiano de experimentos ha sido estudiada desde hace varias décadas, el desarrollo de métodos computacionales eficientes que permitan encontrar diseños que minimicen (1) ha avanzado poco, por lo menos hasta años recientes. Por lo general se conocían soluciones a casos con características particulares. Sin embargo, en los últimos años se ha logrado abordar, por lo menos en parte, este problema.

Overstall y Woods proponen en [15] su método de *intercambio aproximado de coordenadas* (ACE por sus siglas en inglés), que logra encontrar diseños óptimos para una gran variedad

[2] En el contexto de estimar integrales, la maldición de la dimensión se refiere a que, conforme aumenta la dimensión, el trabajo computacional incrementa exponencialmente.

de problemas de diseño experimental. Además, tiene la ventaja de no remontarse a estimaciones asintóticas de la distribución final o de la función de pérdida. La idea general es que el usuario ingrese un diseño inicial, cuyas entradas serán recorridas una por una, decidiendo en cada iteración si la entrada en cuestión debe o no ser cambiada (y por qué otra cantidad). El Código 1 muestra los pasos básicos del método ACE, y fue tomado de [21]. Para mayor detalle acerca de los pasos intermedios, se recomienda revisar [15, 21]

La convergencia del método se determina en el mismo espíritu que el utilizado para los métodos MCMC. En particular, se utilizan análisis gráficos que comparan el número de iteración con alguna aproximación a la utilidad esperada del diseño en cuestión. Dicha aproximación generalmente se logra mediante métodos de Monte Carlo; por ejemplo, siguiendo la notación de la sección anterior, se puede estimar

$$\Phi(\eta) = \int_{\mathcal{Y}} \int_{\Theta} l(\eta, y, \theta) p(\theta, y \,|\, \eta) d\theta \, dy \qquad \text{con} \qquad \frac{1}{N} \sum_{k=1}^{N} l(\eta, y, \theta), \tag{4}$$

donde $(y, \theta) \sim p(\theta, y \,|\, \eta)$ y N es grande (usualmente alrededor de 20,000). Para generar el vector (y, θ) se aprovecha el hecho de que $p(\theta, y \,|\, \eta) = p(y \,|\, \theta, \eta) p(\theta \,|\, \eta)$, y ambas distribuciones son totalmente conocidas.

```
Input : Diseño inicial η = (η_ij) de tamaño n × p.
        Enteros positivos Q y B
Output: Diseño Φ-óptimo

repeat{ # Hasta convergencia
 for(i in 1:n)
  for(j in 1:p)
   Genera un diseño de relleno unidimensional
    ζ_ij = {x_ij^1, ..., x_ij^Q} en X_j ⊂ ℝ
     for(k in 1:Q)
      Evalúa Φ̃_ij(x_ij^k | η) mediante aproximación de Monte Carlo con una muestra de tamaño
            B
     end
    Construye un emulador unidimensional Φ̂(x)
    Encuentra x̂ = arg min_{x∈X_j} Φ̂(x)
    Encuentra p̂ = p̂(η, x̂) utilizando el Código 2
    Define η_ij = x̂ con probabilidad p̂
  end
 end
}
```

Código 1: Método de intercambio aproximado de coordenadas (ACE), propuesto por [15].

```
Input : Diseño actual η = (η_ij)
        Coordenada propuesta x̂
        Entero positivo B̃
Output: Probabilidad posterior p̂ de que Φ̃_ij(x̂ | η) < Φ̃(η)

Define η_p como el diseño obtenido al reemplazar la ij-ésima entrada de η con x̂
for(k in 1:B̃)
```

```
8  |  Genera θ̃ ∼ p(θ)
9  |  Genera y_1 ∼ p(y | θ, η_p) y y_2 ∼ p(y | θ, η)
10 |  Define L_1k = l(η_p, y_1, θ̃) y L_2k = l(η, y_2, θ̃)
11 | end
12 | Supón que L_1k ∼ N(b_1 + b_2, a) y L_2k ∼ N(b_1, a)
13 | Considerando L_1k y L_2k como si fueran datos, calcula la probabilidad posterior p̂ de
       que b_2 < 0
```

Código 2: Algoritmo de aceptación/rechazo utilizado para encontrar la probabilidad de aceptación de la línea 15 del Código 1, tomado de [21].

Algo que se debe de notar es la enorme cantidad de argumentos iniciales que recibe dicho método ($\eta, B, Q, \tilde{B}$, la función de pérdida, las distribuciones iniciales de los parámetros desconocidos). Ello habla de que no solo el método es sofisticado en sí, sino que implementarlo también lo es. Esto habla del difícil acceso a algoritmos de punta que permitan encontrar diseños óptimos con mayor flexibilidad.

Diseño para modelo de regresión logística

La distribución binomial se utiliza para describir el comportamiento de respuestas binarias, es decir, que solo tomen dos valores: 0 (fracaso) y 1 (éxito). Debido a la gran variedad de fenómenos que satisfacen esta característica, dicha distribución ha sido altamente estudiada. La familia de modelos lineales generalizados que utilizan una respuesta binaria son de especial interés. En particular, el modelo de regresión logística, derivado de utilizar la función logit como liga, es muy utilizado en la práctica por las propiedades estadísticas inherentes a utilizar la liga canónica.

Woods et al. [20] encuentran el diseño óptimo para un ejemplo industrial. En particular, una empresa de tecnología alimentaria quería empaquetar papas en un ambiente de atmósfera protectora, con el objetivo de incrementar la vida útil de éstas. El experimento estudiaba el efecto de tres covariables–concentración de vitaminas en el sumergimiento de pre-empacado y los niveles de dos gases en la atmósfera protectora–en diversas variables binarias, entre las que se encuentra la presencia o no de líquido en el paquete después de 7 días. El experimento consta de 16 corridas.

En [21], los autores retoman el ejemplo del empaquetado de papas. Para ello, suponen que la relación que existe entre la respuesta y las covariables se puede describir mediante un modelo de regresión logística. Sea $Y_i \sim \text{Bernoulli}(\pi(x_i))$ la respuesta de la i-ésima corrida del experimento con valores $x_i = (x_{i1}, x_{i2}, x_{i3})^T$ de las covariables, $i = 1, 2, ..., 16$. El predictor lineal η será una función que incluye términos cruzados de primer orden, así como términos cuadráticos, en las variables, además de los términos lineales, es decir,

$$\eta_i = \beta_0 + \sum_{j=1}^{3} \beta_j x_{ij} + \sum_{j=1}^{3} \sum_{k=j}^{3} \beta_{ik} x_{ij} x_{ik}. \tag{5}$$

Aquí, $\beta_0, ..., \beta_3, \beta_{11}, \beta_{12}, ..., \beta_{33}$ son los coeficientes (desconocidos) a ser estimados después de

la recolección de los datos. De esta forma, el modelo es

$$\pi(x_i) = \frac{1}{1 + e^{-\eta_i}} \tag{6}$$

o, alternativamente,

$$\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \sum_{j=1}^{3} \beta_j x_{ij} + \sum_{j=1}^{3}\sum_{k=j}^{3} \beta_{ik} x_{ij} x_{ik}. \tag{7}$$

Con el propósito de ilustrar el funcionamiento del método, los autores asumen distribuciones iniciales uniformes para los parámetros del modelo; en particular,

$$\beta_1, \beta_2 \sim \mathcal{U}(2,6), \qquad \beta_0, \beta_3, \beta_{jk} \sim \mathcal{U}(-2,2) \quad \text{para } j,k = 1,2,3. \tag{8}$$

Además, el espacio de valores que pueden tomar las covariables considerado por los autores es $\mathcal{X} = [-1.2872, 1.2872] \times [-1.2872, 1.2872] \times [-1.2872, 1.2872]$, abreviado $[-1.2872, 1.2872]^3$.

Se replicaron los resultados obtenidos por los autores utilizando el método previamente mencionado, así como los parámetros que proponen Woods et al.[3] Estos son $B = 1,000$ y $Q = 10$ en el método ACE (Código 1) y $\tilde{B} = 20,000$ en el algoritmo de aceptación y rechazo (Código 2). Se encontró el diseño D-óptimo Bayesiano para 16 corridas, es decir, aquel que minimiza la ecuación (3).

La Figura 1 muestra las proyecciones en dos dimensiones de las variables, así como la proyección en una dimensión en forma de densidad. Para corroborar la convergencia del algoritmo, la Figura 2 muestra una aproximación a la pérdida esperada del diseño en cuestión para cada una de las iteraciones de éste. Es inmediato observar que la pérdida esperada es prácticamente igual para las últimas iteraciones, lo que indica que, en efecto, el método convergió.

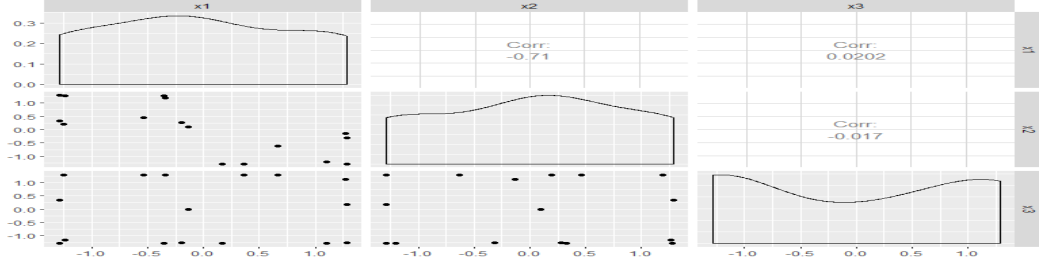


Figura 1: Se muestran proyecciones en una y dos dimensiones de las tres variables del modelo de regresión logística (7), así como su correlación.

[3]Esto se llevó a cabo en el lenguaje de programación R y utilizando el paquete `acebayes` [16] escrito por los mismos autores.

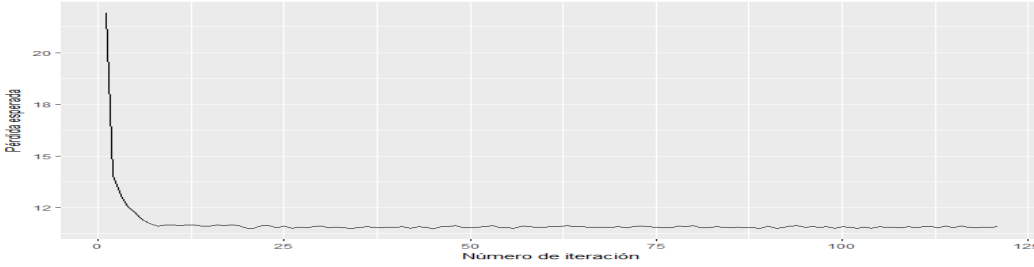


Figura 2: Para la regresión logística, se muestran aproximaciones a la utilidad esperada de los diseños en cada iteración del algoritmo para corroborar convergencia.

Comentarios finales

La importancia del diseño de experimentos radica en que éste permite que la recolección de datos se realice acorde a algún criterio de optimalidad, el cual debe reflejar el objetivo final del estudio en cuestión. Esto eleva la calidad de cualquier análisis inferencial hecho con base en los datos, pues garantiza la validez de la multiplicidad de supuestos necesarios.

El problema de diseño experimental se puede plantear, de manera natural, como un problema de decisión. Sin embargo, inmediatamente aparecen complicaciones en la formulación; en particular, en la mayoría de los casos no existen soluciones analíticas. Más aún, dichas soluciones son difíciles de encontrar incluso mediante métodos computacionales.

El método ACE es uno de los algoritmos más recientes para encontrar diseños óptimos [15], y es particularmente útil en el marco de modelos lineales generalizados y criterios de optimalidad Bayesiana populares, como se mostró en el ejemplo de regresión logística en este artículo.

Futuras investigaciones pueden estar enfocadas a encontrar nuevos métodos que faciliten la resolución de este tipo de problemas, es decir, cuya implementación sea más accesible y, a la vez, se puedan utilizar para diseños con un número alto de corridas y funciones de pérdida más generales.

Referencias

[1] Albert, James et al. 2000. *Generalized Linear Models: A Bayesian Perspective*. 1era edición. New York: Marcel Dekker Inc.

[2] Atkinson, A. C. y Woods, D. C. 2013. *Designs for Generalized Linear Models*. 1era edición. Chapman and Hall. Handbook of Design and Analysis of Experiments.

[3] Box, G. E. P., Hunter, J. S. y Hunter, W. G. 2005. *Statistics for Experimenters: Design, Innovation, and Discovery*. 2da edición. John Wiley & Sons. Wiley Series in Probability and Statistics.

[4] Box, G. E. P. y Liu, P. Y. T. (1999) *Statistics as a Catalyst to Learning by Scientific Method Part I–An Example*. Journal of Quality Technology. 31(1), 1-15.

[5] Chaloner, K. y Verdinelli, I. (1995) *Bayesian Experimental Design: A Review*. Statistical Science. 10(3), 273-304.

[6] Diluvi, G. C. (2017) *Teoría de la Decisión y Estadística Bayesiana*. Lab & Inf. 44, 6-13.

[7] —. (2017) *Modelos lineales generalizados: un enfoque Bayesiano*. Lab & Inf. 45, 36-45.

[8] Fisher, R. A. 1935. *The Design of Experiments*. 8va ed. NY: Hafner Publishing Co.

[9] Johnson, R. T. y Montgomery, D. C. (2009) *Designing Experiments for Nonlinear Models — An Introduction*. Quality and Reliability Engineering International. 26(8), 431-441.

[10] Khuri, A. I., Mukherjee, B., Sinha, B. K. y Ghosh, M. (2006) *Design Issues for Generalized Linear Models: A Review*. Statistical Science. 21(3), 376-399.

[11] Lindley, D. V. 1987. *Bayesian Statistics. A Review*. Society for Industrial Mathematics. CBMS-NSF Regional Conference Series in Applied Mathematics.

[12] McCullagh, P. y Nelder, J. A. 1983. *Generalized Linear Models*. 2da ed. CH.

[13] Montgomery, D. C. 2012. *Design and Analysis of Experiments*. 8va ed. Wiley & Sons.

[14] Nelder, J. A. y Wedderburn, R. W. A. (1972). *Generalized Linear Models*. Journal of the Royal Statistical Society. 135(3), 370-384.

[15] Overstall, A. M. y Woods, D. C. (2016). *Bayesian Design of Experiments using Approximate Coordinate Exchange*. Technometrics In Press. 1-13.

[16] Overstall, A. M., Woods, D. C. y Adamou, M. (2017). *acebayes: Optimal Bayesian Experimental Design using the ACE Algorithm* [Manual de software informático]. Descargado de `https://CRAN.R-project.org/package=acebayes`. (R package version 1.4.1).

[17] R Core Team (2017). *R: A Language and Environment for Statistical Computing* [Manual de software informático]. Vienna, Austria. Descargado de `https://www.R-project.org/`

[18] Rasmussen, C. E. y Williams, C. K. I. 2006. *Gaussian Processes for Machine Learning*. 1era edición. The MIT Press.

[19] Welch, W. J., Mitchell, T. J. y Wynn, H. P. (1989). *Design and Analysis of Computer Experiments*. Statistical Science. 4(4), 409-435.

[20] Woods, D. C., Lewis, S. M., Eccleston J. A. y Russell K. G. (2006). *Designs for Generalized Linear Models With Several Variables and Model Uncertainty*. Technometrics. 48(2), 284-292.

[21] Woods, D. C., Overstall, A., Adamou, M. y Waite, T. W. (2017). *Bayesian design of experiments for generalized linear models and dimensional analysis with industrial and scientific application*. Quality Engineering. 29(1), 91-103.

Métodos de Galerkin discontinuos para leyes de conservación hiperbólicas[1]

Mariana Harris
Estudiante de Matemáticas Aplicadas del ITAM

Introducción

Las leyes de conservación hiperbólicas son un tipo de ecuaciones diferenciales parciales que son usadas para modelar ondas que se propagan a una velocidad finita. Por ejemplo, las leyes de conservación son importantes en la mecánica de fluidos y en la dinámica de gases; sin embargo, es muy difícil encontrar soluciones analíticas para estas ecuaciones; por esto, el estudio y desarrollo de métodos numéricos para resolverlas es de sumo interés.

Una clase de métodos que son usados para resolver las leyes de conservación hiperbólicas son los métodos de Galerkin discontinuos (DG), propuestos por primera vez por Reed y Hill [5] en 1973 como solución numérica de la ecuación del transporte del neutrón. Desde entonces, se ha dado una gran variedad de investigación que ha permitido el desarrollo de nuevos métodos DG.

El presente trabajo busca introducir al lector a los métodos DG, explicando algunas formulaciones usadas hasta ahora; como el método DG con discretización Runge-Kutta en el tiempo[6] (RKDG), los métodos implícitos DG en el tiempo y espacio [4] y el método DG Lax Wendroff local (LW) [2].

Leyes de Conservación

Una ley de conservación, en una dimensión espacial, es una ecuación de la forma

$$q_t + f(q)_x = 0, \tag{1}$$

dónde $t \in \mathbb{R}$ es el tiempo, $x \in \mathbb{R}$ es la coordenada espacial uni-dimensional, $q(t,x) : \mathbb{R}^+ \times \mathbb{R} \mapsto \mathbb{R}^n$ es el vector de n variables conservadas (por ejemplo masa y energía) y $f(q(t,x)) : \mathbb{R}^n \mapsto \mathbb{R}^n$ es la función del flujo.

[1]Este artículo es producto del programa de investigación de verano (REU Iowa State University 2017) NSF Grant DMS-1457443. Asesores: Dr. James Rossmanith, Caleb Logemann. Colaboradores: Ian Pelakh, Camille Felton, Stephan Nelson y Mariana Harris

Una característica de las ecuaciones del tipo (1) es que la cantidad total de $q(t,x)$ en un intervalo $[a,b]$ sólo puede ser modificada mediante el flujo por la frontera $x=a$ y $x=b$. Matemáticamente, por el Teorema Fundamental del Cálculo (TFC) se tiene

$$\frac{d}{dt}\int_a^b q(t,x)dx = f(q(t,a)) - f(q(t,b)). \tag{2}$$

Figura 1: Ley de Conservación en intervalo $[a,b]$

Notemos que en $x=a$ se tiene un flujo positivo, lo que corresponde a un aumento en $q(t,x)$, mientras que en $x=b$ el flujo apunta hacia afuera del intervalo; por esto, el flujo es negativo, lo que corresponde a un decremento en la cantidad de $q(t,x)$.

Leyes de Conservación Hiperbólicas

Definición 1. *Una ley de conservación lineal ($q_t + Aq_x = 0$) es hiperbólica si la matriz A es diagonalizable con eigenvalores reales.*

Algunos ejemplos son la ecuación del transporte $u_t + vu_x = 0$ y la ecuación de la onda $u_{tt} - c^2 u_{xx} = 0$.

Definición 2. *Una ley de conservación no lineal ($q_t + f(q)_x = 0$) es hiperbólica si la matriz jacobiana del flujo,*

$$A(q) = \frac{\partial}{\partial q} f(q),$$

es diagonalizable con eigenvalores reales.

Ejemplos de leyes de conservación hiperbólicas no lineales son las ecuaciones de aguas someras y las ecuaciones de Euler compresibles.

Métodos de Galerkin discontinuos

Como ya se mencionó en la introducción, los métodos DG son métodos numéricos que se utilizan para resolver leyes de conservación hiperbólicas. Dado un problema de valores iniciales para un sistema de esta clase:

$$q_t + f(q)_x = 0, \quad x \in [a,b], \quad q(x,0) = q_0(x), \tag{3}$$

dividimos el dominio $[a,b]$ en un número finito de celdas, dónde la celda i está dada por

$$T_i = \left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}\right]$$

con centro en $x_i = a + \left(i - \frac{1}{2}\right)$. Definimos un paso en el espacio como $\Delta x = \frac{b-a}{N}$, dónde N es el número de celdas, y un paso en el tiempo como Δt, con $t^{n+1} = t^n + \Delta t$.

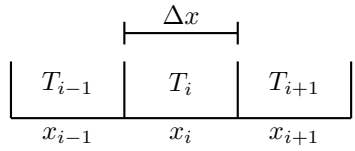


Figura 2: Discretización en el espacio

Procedemos integrando en el tiempo y espacio sobre una celda

$$\frac{1}{\Delta x}\int_{t^n}^{t^n+\Delta t}\int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} \left[q_t + f(q)_x\right] dx\, dt = 0.$$

Luego, definiendo $Q_i^n := \frac{1}{\Delta x}\int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} q(t^n, x)dx$ y $F_{i-\frac{1}{2}}^{n+\frac{1}{2}} := \frac{1}{\Delta t}\int_{t^n}^{t^n+\Delta t} f(q(t, x_{i-\frac{1}{2}}))dt$, obtenemos una expresión para actualizar en el tiempo,

$$Q_i^{n+1} := Q_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+\frac{1}{2}}^{n+\frac{1}{2}} - F_{i-\frac{1}{2}}^{n-\frac{1}{2}}\right). \tag{4}$$

La expresión (4) nos indica cómo actualizar Q^n; sin embargo, no es posible obtener $F_{i\pm\frac{1}{2}}^{n\pm\frac{1}{2}}$, ya que no conocemos la solución exacta para $q(x,t)$. Por esto, procedemos a aproximar la solución en cada celda con un polinomio de grado M que no tiene que ser continuo de una celda a otra.

La solución aproximada en una celda $\mathcal{T}_i$ está dada por la expresión

$$q^{\Delta x}(t,x)\Big|_{\mathcal{T}_i} = \sum_{k=1}^{M} Q_i^{(k)}(t)\phi^{(k)}(\xi),$$

dónde ϕ es una base ortogonal, por ejemplo los polinomios de Legendre.

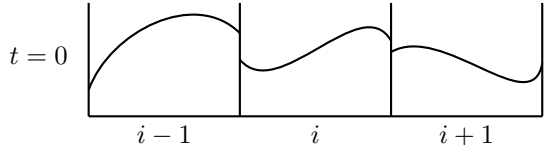


Figura 3: Aproximación usando polinomios discontinuos entre celdas.

Discretización DG en el espacio

Una primera estrategia podría ser utilizar la discretizacion en el espacio propuesta por el método DG, y posteriormente proceder con un método para resolver ecuaciones diferenciales ordinarias como, por ejemplo, un método Runge Kutta (RKDG).
En este caso empezamos con un cambio de variable al escribir x en términos de ξ tal que $x = x_i + \frac{\Delta x}{2}\xi$, así se escribe $q_t + f(q)_x = 0$ como $q_t + \frac{2}{\Delta x} f(q)_\xi = 0$. Después multiplicamos por la función prueba ϕ

$$\frac{1}{2}\int_{-1}^{1} \phi^{(k)} q_t + \frac{1}{\Delta x}\int_{-1}^{1} \phi^{(k)} f(q)_\xi = 0,$$

e integramos por partes en ξ,

$$\frac{1}{2}\int_{-1}^{1} \phi^{(k)} q_t d\xi - \frac{1}{\Delta x}\int_{-1}^{1} \phi_\xi^{(k)} f(q) d\xi + \frac{1}{\Delta x}\left[\phi^{(k)}(1)\mathcal{F}_{i+\frac{1}{2}} - \phi^{(k)}(-1)\mathcal{F}_{i-\frac{1}{2}}\right] = 0.$$

Sustituyendo q con la aproximación $q^{\Delta x}(t,x)\Big|_{\mathcal{T}_i} = \sum_{k=1}^{M} Q_i^{(k)}(t)\phi^{(k)}(\xi)$ obtenemos,

$$\frac{\partial}{\partial t} Q^{(k)} = \frac{1}{\Delta x}\left(\int_{-1}^{1} \phi_\xi^{(k)} f(q^{\Delta x}) d\xi - \left(\phi^k(1)\mathcal{F}_{i+\frac{1}{2}} + (\phi^k(-1)\mathcal{F}_{i-\frac{1}{2}}\right)\right), \tag{5}$$

dónde $\mathcal{F}_{i\pm\frac{1}{2}}$ representa el flujo numérico y $\int_{-1}^{1} \phi_\xi^{(k)} f(q^{\Delta x}) d\xi$ puede ser resuelto numéricamente.
De esta manera se obtiene un sistema de ecuaciones diferenciales ordinarias (5) de la forma $\frac{\partial}{\partial t} Q = \mathcal{L}(Q)$ en el que se puede usar un método numérico explícito como, por ejemplo, Runge-Kutta.

Discretización DG en el tiempo y el espacio

Otra forma de proceder podría ser usando una discretización DG en el espacio y en el tiempo. En este caso se tiene un método implícito en dónde la solución aproximada ahora está dada por

$$q^{\Delta x}\Big|_{T_i^{n+\frac{1}{2}}} = \sum_{\ell=1}^{M(M+1)/2} Q_i^{(\ell)}(t)\psi^{(\ell)}(\tau,\xi).$$

Tal que $\psi(\tau,\xi)$ es una base ortonormal.
Al tomar $\xi = \frac{2}{\Delta x}(x - x_i)$, y $\tau = \frac{2}{\Delta t}(t - t_i)$. Escribimos (3) en términos de ξ y τ para obtener,

$$q_\tau + \frac{\Delta x}{\Delta x} f(q)_\xi = 0. \tag{6}$$

Luego multiplicamos (6) por la función prueba ψ, integramos en el tiempo, el espacio y normalizamos

$$\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \psi q_\tau d\xi d\tau + \frac{\triangle t}{4 \triangle x}\int_{-1}^{1}\int_{-1}^{1} \psi f(q)_\xi d\xi d\tau = 0.$$

Luego, integrando por partes obtenemos

$$-\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}\psi_\tau q_\xi d\tau - \frac{\triangle t}{4\triangle x}\int_{-1}^{1}\int_{-1}^{1}\psi_\xi f(q)d\xi d\tau$$
$$+\frac{1}{4}\int_{-1}^{1}\left[\psi(\tau=1,\xi)q(\tau=1,\xi)-\psi(\tau=-1,\xi)q(\tau=-1,\xi)\right]d\xi$$
$$+\frac{\triangle t}{4\triangle x}\int_{-1}^{1}\left[\psi(\tau,\xi=1)\mathcal{F}_{i+\frac{1}{2}}(\tau)-\psi(\tau,\xi=-1)\mathcal{F}_{i-\frac{1}{2}}(\tau)\right]=0,$$

donde

$$q(\tau=1,\xi)=\lim_{\tau\to 1}q^{\triangle x}(\tau,x)_{T_i^{n+\frac{1}{2}}} \quad \text{y} \quad q(\tau=-1,\xi)=\lim_{\tau\to 1}q^{\triangle x}(\tau,x)_{T_i^{n-\frac{1}{2}}}.$$

El problema con ésta formulación es que, a pesar de que se tiene la estabilidad incondicional de los métodos implícitos, para obtener la solución aproximada se necesita resolver un sistema de ecuaciones de tamaño $N*\frac{M(M+1)}{2}$ dónde N es el número de celdas y M el orden de los polinomios base. Este sistema es muy grande; por esto, resolverlo implica un costo computacional muy alto.

Lax-Wendroff DG Local

Para no enfrentarnos al sistema de tamaño $N*\frac{M(M+1)}{2}$ mencionado anteriormente, podemos usar el método de Lax-Wendroff DG local, este consiste en dos pasos: un paso predictor y otro corrector.

Paso predictor

En este paso se trata a cada celda como independiente de las celdas vecinas, de tal forma que no hay interacción entre ellas. Se procede entonces discretizando en el tiempo y espacio usando el método implícito descrito en el apartado anterior, nada más que ahora se tienen que resolver N sistemas independientes de tamaño $\frac{M(M+1)}{2}$.
Multiplicamos (6) por la función ψ e integramos en el tiempo y espacio:

$$\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}\psi\left(q_\tau+\frac{\triangle t}{\triangle x}f(q)_\xi\right)d\xi d\tau=0.$$

Para desacoplar las celdas integramos por partes en τ y luego integramos de regreso. Durante este último paso se tiene que aproximar a través del interior de la celda, esto genera una discontinuidad de salto.

$$\int_{-1}^{1}\int_{-1}^{1}\underline{\psi}\,q\left(\sum_{\ell=1}^{\frac{M(M+1)}{2}}W_i^{(\ell)}\psi^{(\ell)}\right)_\tau+\frac{\Delta t}{\Delta x}\underline{\psi}\,f\left(\sum_{\ell=1}^{\frac{M(M+1)}{2}}W_i^{(\ell)}\psi^{(\ell)}\right)_\xi d\xi d\tau$$

$$+\underbrace{\int_{-1}^{1}\underline{\psi}(-1,\xi)\left(q(-1,\xi)-\underline{\phi}^{T}(\xi)\underline{Q}_{i}^{n}\right)d\xi}_{\text{Discontinuidad de salto}}=0.$$

De lo anterior se obtiene un sistema de ecuaciones de la forma:

$$A\,W=B(Q)$$

en el caso lineal y

$$A(W)W=B(Q)$$

en el caso no lineal. Resolvemos para W por cada celda y así obtenemos una primera solución.

Paso corrector

La solución W que se obtiene en el paso predictor no es una solución correcta, ya que cada celda es independiente del resto del dominio. Para corregir esto agregamos el flujo de las celdas adyacentes, al utilizar un esquema del tipo Lax-Wendroff [4].

$$\begin{aligned}Q_{i}^{(k)\,n+1}=&Q_{i}^{(k)\,n}+\frac{\Delta t}{2\Delta x}\int_{-1}^{1}\int_{-1}^{1}\phi_{\xi}^{(k)}f(q)d\xi d\tau\\&-\frac{\Delta t}{2\Delta x}\left[\int_{-1}^{1}\phi^{(k)}(\xi=1)\,\mathcal{F}_{i+\frac{1}{2}}(\tau)d\tau-\int_{-1}^{1}\phi^{(k)}(\xi=-1)\,\mathcal{F}_{i-\frac{1}{2}}d\tau\right].\end{aligned}$$

Dónde $\mathcal{F}_{i\pm\frac{1}{2}}$ es el flujo numérico que tiene que cumplir con las condiciones de consistencia y conservación. Finalmente, obtenemos una solución numérica Q_{i}^{n+1} para nuestro sistema de conservación (3).

Comentarios finales

Los métodos DG han resultado ser muy útiles en la práctica para resolver leyes de conservación hiperbólicas. Una de las ventajas de éstos métodos es que para soluciones suficientemente suaves el orden del método aumenta al subir el grado de los polinomios de la base; por esto se pueden obtener muy buenas aproximaciones.

Un problema frecuente en la práctica es que, muchas veces, para condiciones iniciales continuas, las leyes de conservación hiperbólicas generan choques (soluciones con discontinuidades). Estas discontinuidades, como consecuencia, pueden generar oscilaciones numéricas en el método; por esto, hay que tener cuidado con este tipo de problemas. Es común usar funciones limitadoras para suavizar las oscilaciones generadas por las discontinuidades y de esta forma obtener buenas aproximaciones que además sean físicamente plausibles.

Referencias

[1] G. Gassner, F.Lörcher, y C.-D. Munz, *A discontinuous Galerkin Scheme based on a space-time expansion II. Viscous flow equations in multi dimensions,* J.Sci. Comp **34**(2008), 260-286.

[2] G. Gassner, M. Dumbser, F. Hindenlang, and C.-D. Munz , Explicit one-step time discretizations for discontinuous Galerkin and finite volume schemes based on local predictors , J. Comput. Physics, 230 (2011), pp. 4232–4247.

[3] F.Lörcher,G. Gassner, y C.-D. Munz, *A discontinuous Galerkin Scheme based on a space-time expansion I. Inviscid commplressible flow equations in one space dimensions,* J.Sci. Comp **32**(2007), 175-199.

[4] J. Qiu, M. Dumbser, and C.-W. Shu *The discontinuous Galerkin method with Lax-Wendroff type time discretizations,* Comput. Methods Appl. Mech. Engr **194** (2005) 4528-4543

[5] W.H. Reed y T. Hill, *Triangular mesh methods for the neutron transport equation,* Los Alamos Report LA-UR-73-479, 1973.

[6] C.-W. Shu and S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes , Journal of Computational Physics, 77 (1988), 439-471

Métodos para validación de modelos de Valor en Riesgo

Samuel Ramos Pérez
Estudiante de Actuaría del ITAM

Introducción

En finanzas existen distintos modelos para la medición del riesgo, uno de ellos es el modelo del Valor en Riesgo (VaR) que, en breves términos es una prueba para estimar la pérdida máxima esperada, para el k-ésimo cuantil, de la distribución de las pérdidas y ganancias de la empresa. Es decir, para un nivel de confianza se calcula la región de rechazo que nos dará la marca a mercado (Mark to Market) que debería estar por encima o por debajo de la pérdida que se espera tener en un intervalo de tiempo, normalmente un día. Antes de continuar mencionaré que, aunque cuantitavamente el modelo estima las pérdidas, los modelos para la validación utilizan ambas colas de la distribución de ganancias y pérdidas, es decir, se hace un cálculo para pérdidas y "pérdidas negativas" (ganancias), lo que nos llevará a considerar pruebas de dos colas, así, en este artículo las pérdidas se tomarán tanto positivas como negativas, al final se justificará esta decisión.

La metodología da como resultado un monto para el cual al final del día se espera tener una pérdida (o ganancia) no mayor (o menor) a ésta, dado un nivel de confianza fijado por la empresa o quien esté realizando el modelo. Ahora, el problema de inferencia estadística de estimación por regiones nos dice que, para cierto nivel de confianza, por ejemplo .99, en cien observaciones, tendríamos a lo más una que no estima el parámetro que deseamos conocer, en este caso, en uno de cien días la pérdida sobrepasaría el límite estimado.

Por otro lado, en el sentido del artículo y en uso de la teoría de toma de desiciones, existen métodos para la validación de estos modelos de medición del riesgo que llevan por nombre Backtesting. Éstos sirven para validar los modelos que una empresa tenga del valor en riesgo; el Backtesting se ocupa de realizar un análisis retrospectivo de los datos, básicamente, se analizan las observaciones del VaR (pérdida esperada) y la pérdida real; entonces una observación se añade si sobrepasó la estimación proporcionada por el resultado del VaR.

A continuación se presenta un método de Backtesting que posteriormente sera modificado para refinación.

Backtesting por el método de Kupiec

Este método toma como principal variable el tiempo en el que sucede una observación (excepción del VaR) proponiendo el siguiente estadístico:

$$LR_{pf} = -2ln[p(1-p)^{\nu-1}] + 2ln\left[\left(\frac{1}{\nu}\right)\left(1-\frac{1}{\nu}\right)^{\nu-1}\right]$$

1.6180339887498948482045868343656381177203091798057628621354486227052604628189

donde ν es el tiempo en días entre que se empieza a contar y la excedencia del VaR, p es la probabilidad de que suceda una observación en nuestros datos y LR_{pf} representa el negativo del doble logaritmo del cociente de verosimilitudes generalizado bajo la H_o (hipótesis nula) del VaR, este estadístico se distribuye como una variable aleatoria Ji-Cuadrada con un grado de libertad.

Podemos modificar este estadístico para definir uno que considere más tiempos de ocurrencia y que al mismo tiempo pruebe la independencia de cada evento, pues Kupiec solo se concentra en la primera excedencia. Así definimos:

$$LR_{pf_i} = -2\,ln[p(1-p)^{\nu_i - 1}] + 2\,ln\Big[\Big(\frac{1}{\nu_i}\Big)\Big(1-\frac{1}{\nu_i}\Big)^{\nu_i - 1}\Big]$$

donde ν_i es el tiempo entre las excepciones i e $i-1$ y p es la probabilidad de que suceda una observación en nuestros datos.

Por lo que si definimos ahora:

$$LR_{ind^*} = \sum_{i=2}^{n}\left[-2\,ln\left(\frac{p(1-p)^{\nu_i-1}}{(\frac{1}{\nu_i})(1-\frac{1}{\nu_i})^{\nu_i-1}}\right)\right] - 2\,ln\left(\frac{p(1-p)^{\nu-1}}{(\frac{1}{\nu_i})(1-\frac{1}{\nu})^{\nu-1}}\right).$$

Como la suma de todos los estadísticos para los que sucede una excepción en un horizonte de tiempo $i \in \{1, ..., n\}$.

Por una parte, el estadístico de independencia se distribuye como una variable aleatoria Ji-cuadrada con n grados de libertad que al sumarlo con estadístico de Kupiec original tenemos una prueba más potente del modelo, obteniendo $LR_{comb} = LR_{Kupiec} + LR_{ind^*}$.
LR_{comb} un estadístico que se distribuye como una variable aleatoria Ji-Cuadrada con $n+1$ grados de libertad.

Como observación, para las pruebas de Backtesting es necesario que al momento de realizar el modelo de VaR el nivel de confianza no sea tan elevado, pues si quisiéramos validarlo tendríamos muy pocas observaciones (excedencias) algo que nos imposibilitaría decir si lo estamos sobrevalorando, infravalorando o requeríamos de más datos (de los 252 anuales que se utilizan comúnmente) para poder rechazar o no nuestro modelo de VaR.

Backtesting por el método generalizado de Christoffersen

Adicional al método propuesto por Kupiec, Christoffersen analiza la independencia entre cada observación (exceder el VaR diario calculado con nivel de confianza) y la probabilidad de que suceda una dependiendo del resultado el día anterior, es decir, el modelo se condiciona a la variación de los datos, añadiendo un componente estocástico.

Entonces, se define el estadístico:

$$LR_{ind} = -2\,ln\big[(1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})}\big] + 2\,ln\big[(1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}\big]$$

donde:

T_{ij} es el número de días en los que hay una excedencia (o no) al VaR (j) y el evento anterior (i) e.g. T_{01} número de días entre los que se presenta una excedencia del VaR dado que el día anterior no la hubo, T_{11} número de días entre los que se presentó una excedencia del VaR dado que el día anterior también hubo una y π_i la probabilidad de observar una excedencia del VaR del dia i.

Además

$$\pi_0 = \frac{T_{01}}{T_{00}+T_{01}} \qquad \pi_1 = \frac{T_{11}}{T_{10}+T_{11}} \qquad \pi = \frac{T_{01}+T_{11}}{T_{00}+T_{10}+T_{01}+T_{11}}$$

π es al considerar que las excedencias son independientes entre días y LR_{ind} representa el negativo del doble logaritmo del cociente de verosimilitudes generalizado bajo la misma H_0 de Kupiec y el supuesto de que cada evento de excedencia del VaR ocurre de forma independiente, este estadístico se distribuye como una Ji-Cuadrada con un grado de libertad, al igual que el estadístico de Kupiec.

Entonces, al combinar ambos métodos obtenemos $LR_{cond} = LR_{Kupiec} + LR_{ind}$.
Un estadístico que se distribuye como una variable aleatoria Ji-cuadrada con dos grados de libertad y sirve para la misma prueba que el modelo de Kupiec.

Es importante mencionar que la prueba de independencia de Christoffersen solo obtiene resultados si existen observaciones en días consecutivos, $T_{11} \neq 0$, ya que solo analiza la independencia del proceso en forma estocástica de primer orden, pero paradógicamente, si hubiera más observaciones que fueran consecutivas, el modelo del VaR no estaría respondiendo a cambios en el mercado.

Otro problema de este modelo es la generalización de la independencia de las excedencias del VaR diario, ya que si no se estuviera considerano este supuesto, T_{11} sería mayor que α, invalidando las distintas π y por consiguiente el estadístico asociado.

Adicional

Existen métodos de backtesting basados en regresiones y son utilizados en casos en los que el VaR no está reaccionando a la volatilidad del mercado, se están teniendo más excedencias y de forma más consecutiva que antes o cuando la volatilidad cambia de manera significativa en periodos cortos de tiempo, e.g. alguna crisis. Estos métodos dan idea de cómo modificar el modelo del VaR que se esté aplicando. La idea principal de estos modelos de regresión es añadir información de las variables para crear el vector de variables independientes.

Conclusión

Recordemos que el nivel de confianza utilizado es nuestro modelo de VaR no debe ser tan alto si es que se desea poder realizar un análisis de éste y por supuesto mucho menos que sea bajo, pues esto no solo nos da más información para validar el modelo sino también nos ayuda a esperar menos pérdidas, respectivamente. Además éstos modelos no brinda las herramientas para decidir se se está colocando suficiente capital de riesgo o para no desestimar el mismo.

Al analizar ambas colas de la distribución, que es considerar las pérdidas y ganancias (pérdidas negativas) a las que pueda estar sujeta una compañía, se puede saber si se está teniendo demasiado flujo de efectivo o la deuda no representa las obligaciones esperadas, así se evita que en días corrientes no se pueda cumplir con las responsabilidades o exista efectivo ocioso.

Referencias

[1] Carol, Alezander. (2008). *Market Risk Analysis, volume IV, value at risk models. John Wiley & Sons, Chichester. 385*

[2] Jorion, Philippe. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. McGraw Hill, United States.

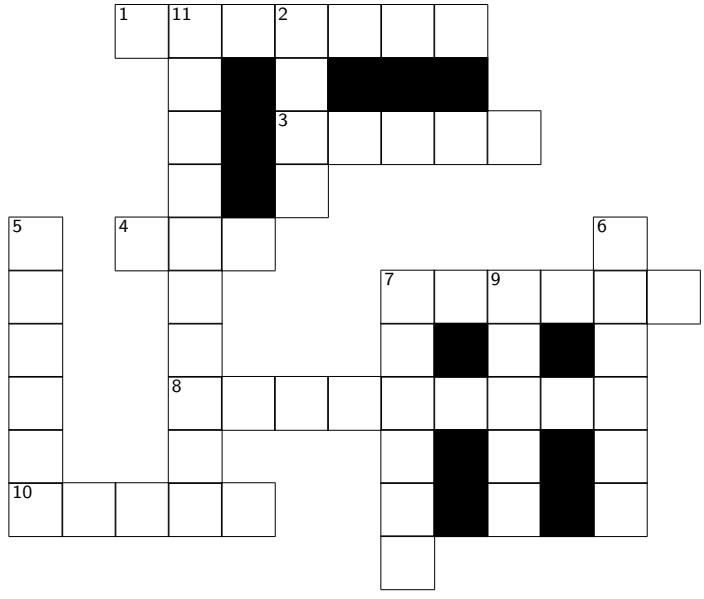
[3] Christoffersen, Peter & Pelletier, Denis. (2003). *Backtesting Value-at-Risk: A Duration-Based Approach*

[4] Basle Committee on Banking Supervision. (1996). *Supervisory Framework for the use of "Backtesting" in Conjunction with the Internal Models Aprroach to Market Risk Capital Requirements*

[5] Nieppola, Olli. (2009). *Backtesting Value-at-Risk Models*

Activa tus neuronas

Personajes matemáticos



Horizontal 1 Creador del metodo Simplex 3 Mentor de Ramanujan 4 Lo arrestaron porque pensaron que era un espía alemán 7 Teorema de la curva simple cerrada 8 $f(x) = 1$ si $x \in \mathbb{Q}$ y $f(x) = 0$ si $x \in \mathbb{I}$ 10 Matemático famoso por su sombrero y el de una bruja.

Vertical 2 Libro de investigación de operaciones 5 Maquina Enigma 6 $\aleph_0$ 9 Funciones elípticas 11 ¡Eureka!

Retos matemáticos

1. ¿Qué resalta sobre el número 6210001000?

2. ¿Cuál es el menor número además del 1 que aparece al menos 6 veces en el triangulo de Pascal?

3. ¿Falso o verdadero? Existe una función real valuada que es continua en exactamente un punto.

4. ¿Qué tienen en común Ronald Graham, Péter Frankl y Claude E. Shannon?

1.6180339887498948482045868343656381177203091798057628621354486227052604628189062

5. ¿Quién dijo "Die ganzen Zahlen hay der liebe Gott gemacht, alles andere ist Menschenwerk" y qué significa?

6. ¿Cuál es el teorema mencionado en la pelicula del Mago de Oz?

Enigmas matemáticos

1. Tienes a 7 generales y una caja fuerte con muchos candados. Asignas las llaves de la caja de forma que cualquier conjunto de 4 generales tiene las suficientes llaves para abrir la caja, al mismo tiempo ningún conjunto de tres generales tiene suficientes llaves. ¿Cuántos candados necesitas? ¿Cuántas llaves recibe cada general?

2. Considere un tablero de ajedrez de 5×5, suponga que tiene 5 reinas y 3 peones. ¿Es posible acomodar las 8 piezas de forma que ninguna reina ataque a ningún peón?

3. Dado que $x^{x^{x^{x^{x\cdots}}}} = 2$, encuentre x.

4. Claudia, mientras daba clase de álgebra lineal, encontró una botella de Vignere con el siguiente mensaje:

   ¿VB XM CCXRM QJAPZBPQE QI ZCLWKN HWZW EOA ZFBRUTHQPFA QME

   GMAYQQW, EOA ZFBRUTHQPFA PWFC TN RCFQVO LR QI EISCV?

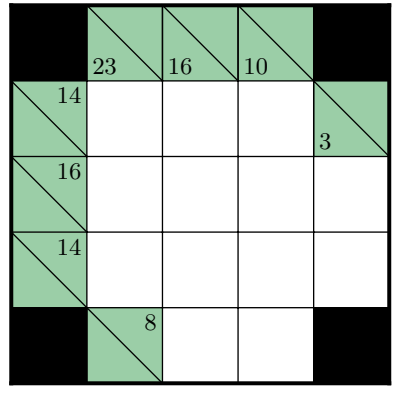
   RQ UHABQW FNMABX ZIF RIGMFOBVHIF, ME AIGJUNBBQW CNMAAT SV YF

   UHABQI: YF UHABQI RQ AHMGC, TN RIGMFOBVHI YI OWLN QIOWKOT.

   Hasta arriba de la botella dice ∞, como tu eres su alumno favorito te pide ayuda para descifrarlo, ¿qué dice el mensaje?

Juegos matemáticos

Resuelve el siguiente Kakuro



Zona Olímpica

1. Se suman los dígitos del producto $1 \times 2 \times \cdots \times 2018$. Se vuelven a sumar los dígitos del número resultante y se continúa de esta forma hasta obtener una sola cifra ¿cuál es?

2. Sea A una matriz cuadrada de tamaño n con entradas complejas. Se dice que A es hermitiana si y solo si $a_{i_j} = \bar{a}_{j_i}$[1]. Demuestra que para cualquier matriz cuadrada compleja M existen matrices A, B hermitianas tales que

$$M = A + iB.$$

3. Prueba que en cualquier reunión de gente (número finito de personas), hay al menos dos personas que conocen al mismo número de gente. Suponga que si la persona A conoce a la persona B eso implica que B conoce a A (la relación es reflexiva).

4. Prueba que la sucesión $\{a_n\}_{n\in\mathbb{N}}$ dada por

$$a_n = \sqrt{1 + \sqrt{2 + \sqrt{3 + \cdots + \sqrt{n}}}}$$

converge.

5. Sea $f : \mathbb{R} \to \mathbb{R}$ una función continua. Para $x \in \mathbb{R}$ se define

$$g(x) = 2f(x) \int_0^x f(t)dt.$$

Demostrar que si $g(x)$ es no creciente en todo $\mathbb{R}$, entonces f es idénticamente cero para toda x.

6. Sean a y b números reales tal que

$$9a^2 + 8ab + 7b^2 \leq 6.$$

Prueba que $7a + 5b + 12ab \leq 9$.

Pregunta de Erdös

Un par ordenado (x, y) de enteros es *punto primitivo* si el máximo común divisor de x y y es 1. Dado un conjunto finito S de puntos primitivos, demostrar que existe un entero positivo n y números enteros $a_0, a_1, \ldots, a_n$ tales que para cada (x, y) de S se cumple que

$$a_0x^n + a_1x^{n-1}y + a_2x^{n-2}y^2 + \cdots + a_{n-1}xy^{n-1} + a_ny^n = 1.$$

[1] Dado un número complejo $z = a + bi$, se define el conjugado como $\bar{z} = a - bi$