LABERINTOS & INFINITOS

Revista de matemáticas y actuaría del ITAM

PRIMAVERA 2021

ÍNDICE

LABERINTOS INFINITOS, Año 2021, No. 54, febrero 2021, 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: Dan 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

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 Magdalena Barba. Gracias a Basilea y Jaime Borel, 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.

$$\ln\left(x^2\right) = x$$

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

Diseño de portada:
Pilar Ramos Francia Paz

Imagen de portada:
Mariana de Jesús Alcocer Piña

-0.703467422498391652049818601859902130342928431034

Alexander Grothendieck

José Ángel Rodríguez
Alumno de Matemáticas Aplicadas del ITAM

"Matemáticas: El honor del espíritu humano"
Carl Jacobi

El siglo XX estuvo repleto de grandes matemáticos, para convencerse de esto uno puede ver los miembros de las distintas generaciones del grupo Bourbaki. De estas personas destaca Alexander Grothendieck por su personalidad, su trabajo y sus fuertes convicciones. Dos términos o características que con frecuncia se usan para describir a un gran matemático son la intuición y la dedicación. El segundo es muy claro y se refiere a poner las horas de trabajo y ensuciarse las manos cuando hay que hacerlo. El primero, sin embargo, es un poco más abstracto y a menudo se confunde simplemente con talento, con la habilidad innata de entender con mayor facilidad que otras personas los objetos de las matemáticas. Para Grothendieck la interpretación que se presentó anteriormente estaba muy lejos de la verdad. Si por intuición nos referimos a la "visión" que uno tiene, podemos caracterizar a la de Grothendieck como un niño curioso que no tiene miedo a hacer preguntas ni a estar mal, pero que solamente quiere entender. Grothendieck incluso mencionó alguna vez que lo que realmente era su mayor interés no eran las consecuencias o implicaciones que tuviera algún resultado, sino poder entender mejor la forma de las estructuras con las que estuviera trabajando.

Grothendieck a los 22 años

Puede haber discusiones sobre cuáles son las cualidades que hacen que el trabajo de un matemático o una matemática destaque sobre el de los demás, pero algo en lo que podemos estar de acuerdo es que sean las que sean Alexander Grothendieck las tenía dominadas. Fue considerado como uno de los mejores matemáticos del siglo XX y tuvo contribuciones de suma importancia en distintas áreas de las matemáticas. Su vida como matemático se puede dividir en dos periodos: el primero, el que constituyó su trabajo hasta su tesis doctoral, y el segundo, se refiere a sus numerosos e importantes trabajos realizados durante la llamada "Época Dorada".

Grothendieck fue alumno de Jean Dieudonné, con quien trabajaría por mucho tiempo; y Laurent Schwartz, a quien fue referido por Henri Cartan, quien le pidió a Schwartz que le ayudara a Grothendieck tras ver que el joven no estaba listo para el material que se impartió en el seminario de Cartan. Podría parecer extraño que dos de los mejores matemáticos de la época mostraran tanto interés en un joven que parecía tener tantas carencias en sus conocimientos. Este interés se debió en gran parte a que Grothendieck desarrolló por su cuenta

lo que parecía ser una variante de la teoría de integración de Lebesgue y por este motivo lo enviaron a Francia para mostrarle su trabajo a Cartan, esto antes de empezar su doctorado. Dieudonné y Schwartz, ambos maestros de Grothendieck en la Universidad de Nancy, publicaron un artículo sobre espacios vectoriales topológicos con catorce problemas abiertos que los autores no habían logrado probar. Unos meses después Grothendieck los había resuelto todos.

Después de este tiempo, el trabajo de Grothendieck se concentró principalmente en Geometría Algebraica, que a grandes rasgos se enfoca en las propiedades geométricas de las variedades: conjunto de raíces de una familia de polinomios.Esta área cambió fundamentalmente con el trabajo de Grothendieck. A inicios de la segunda mitad del siglo XX, Grothendieck y sus colaboradores empezaron a trabajar en las conjeturas de Weil, que son una serie de problemas que se enfocan en Geometría Algebraica y Teoría de Números. El trabajo que hicieron en este programa llevó al desarrollo del concepto de un **esquema** (scheme), que lleva a un cambio de lenguaje para el trabajo en Geometría Algebraica que permite hacer conexiones con Topología, Álgebra Homológica y Teoría de Categorías que lleva a la llamada Geometría Algebraica moderna o Geometría Algebraica Functorial. Este fue un cambio verdaderamente revolucionario que llevó a un gran progreso en las conjeturas de Weil y el desarrollo de nuevas teorías. En esta época se escribió la colosal obra *Élements de géometrie algébriqué* con ayuda de su maestro Dieudonné. Dos de estos problemas serían probados después: uno por Grothendieck y el otro por su alumno más destacado Pierre Deligne.

Como es de esperarse, a Grothendieck le ofrecieron muchos trabajos y le otorgaron muchos premios. Sin embargo, muchos de estos los terminaría rechazando por cuestión de principios. Entre los casos más notables de esto está la medalla Fields. Grothendieck rechazó el premió y se negó a asistir a la ceremonia como protesta de algunas de las acciones tomadas por la Unión Soviética. Otro ejemplo es la renuncia de su puesto en el IHES (*Institut des Hautes Études Scientifiques*) por no estar de acuerdo en que hubiera fondos militares. Esto se puede explicar en parte por la vida que llevaron sus padres. Independientemente del motivo, lo importante es la perspectiva que tenía Grothendieck de las matemáticas que no solamente se pronunciaba en contra de su uso con fines militares, sino que no tenía interés alguno en cualquier otra forma de acercarse a las matemáticas que apreciar la belleza que en ellas se esconde y tratar de entenderlas mejor.

Referencias

[1] Ribenboim, P. *Excerpt from the Grothendieck I Knew: Telling, Not Hiding, Not Judging.* American Mathematical Society, 2019.

[2] Bosch, C. García, C. "De la luz a la sombra", *Miscelánea Matemática* 62 (2016): 45-61.

[3] "Alexander Grothendieck." *Wikipedia*. Consultado el 13 de abril de 2021. `https://en.wikipedia.org/wiki/Alexander_Grothendieck`

Sobre la curiosidad, esferas y sus sombras

Sergio Iván Arroyo Giles
Egresado de Matemáticas Aplicadas, ITAM
Francisco Castañeda Ruan
Egresado de Matemáticas Aplicadas, ITAM

"Mathematical thinking is the basis of all of the sciences.
You cannot be a scientist without learning mathematical thinking."
Sam Nelson

Pero, ¿por qué?

Las matemáticas son especiales entre todas las áreas del conocimiento, pues en ellas se pueden hacer aseveraciones indudablemente verdaderas o completamente falsas. Para las otras ciencias, ya sean naturales o sociales, es inclusive peligroso afirmar que algo es cierto sin lugar a dudas. Dado lo anterior, quizá la única aseveración que se puede hacer con plena confianza es que toda persona que estudia, o estudió, matemáticas ha escuchado la frase "Yo odiaba las matemáticas" como respuesta casi inmediata al escuchar sobre su elección de carrera.

Las causas de esta repulsión visceral son múltiples, y las hojas de esta revista no alcanzarían para cubrir con detalle cada una de ellas. Sin embargo, aprovechemos este espacio para resaltar una de ellas: la ilusión de arbitrariedad. La justificación que gran parte de las personas que "odiaban las matemáticas" nos han dado se reduce a la idea de que las matemáticas son un conjunto de fórmulas y definiciones arbitrarias cuyo único propósito es ser memorizadas y regurgitadas en un examen. Concepción que, para los que hemos cursado alguna asignatura de matemáticas por los últimos cuatro años, suena diametralmente opuesta a la realidad.

¿Por qué es este el caso? Desde nuestra perspectiva, una de las características comunes a la mayoría de compañeros en la carrera es la curiosidad, esa molestia en la parte de atrás de la cabeza que no expira hasta que se pronuncian las palabras mágicas: "Pero, ¿por qué?". Y, solo hasta el momento en que se haya echado a andar esta curiosidad incisiva es que la ilusión de arbitrariedad comienza a resquebrajarse.

La maravilla de esta curiosidad es que se puede encender con cualquier pregunta. Para uno de nosotros, la primera instancia de curiosidad ocurrió cuando se le presentó la fórmula del volumen de una esfera: $V = \frac{4}{3}\pi r^3$. La aparición de π y r eran comunes y esperadas, pero "¿Por qué $\frac{4}{3}$?", preguntaba un pequeño estudiante de secundaria que esperaba ver un

entero en vez de un racional en su fórmula. La confusión se volvió aún más grande cuando nadie supo decirle exactamente por qué $\frac{4}{3}$ era el coeficiente indicado para el volumen.

Imagínate la sorpresa del estudiante primerizo cuando finalmente llegó al curso de Cálculo Diferencial e Integral II y la gran revelación fue que $\frac{4}{3}$ era tan solo la constante de integración que resultaba de girar un círculo alrededor de un eje. Fue un momento menos climático de lo que aquel estudiante esperaba.

Pero, afortunadamente, las clases de universidad estuvieron repletas de momentos curiosos que resultaron ser más interesantes que éste. La lección es que la curiosidad a veces nos provee de mayor perspectiva sobre cómo operan los objetos con los que trabajamos, y otras veces a conclusiones poco emocionantes. Sin embargo, aún a pesar de la segunda posibilidad, creemos que fomentar la curiosidad es fundamental para romper el estigma social sobre las matemáticas.

¿$4\pi r^2$?

Regresemos a la esfera, la fórmula del volumen es parte del dominio público, pero la fórmula del área superficial es mucho menos conocida. Para una esfera de radio r, el área superficial de ésta está dada por la expresión

$$4\pi r^2. \tag{1}$$

Vista de manera superficial esta expresión puede parecer tan arbitraria como todas las demás. Necesitamos una chispa para encender la curiosidad. Para un círculo de radio r, el área de dicha esfera se calcula mediante

$$\pi r^2 \tag{2}$$

Es aquí donde comienza la curiosidad.

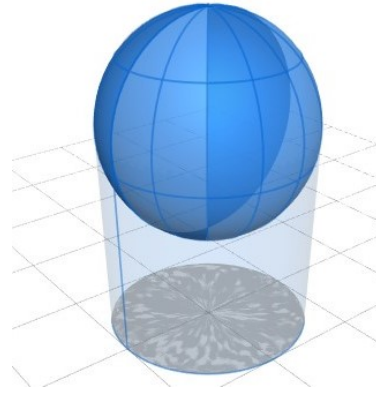


Fig. 1 La esfera y su sombra

¿Significa esto que una esfera tiene la misma área que cuatro círculos del mismo radio? Sí. ¿Entonces se puede construir una esfera utilizando solo cuatro círculos? Esta pregunta y la anterior no son tan distintas, pero la segunda requiere mayor atención. Antes de continuar leyendo toma un segundo y piensa cómo puede ser esto posible. Si tienes la oportunidad, forma círculos con papel o plastilina e intenta recrear una esfera de algún modo. Si tu acercamiento es pegar los círculos uno con otro, el hecho de que los círculos sean planos y la esfera causará problemas. Otro acercamiento puede ser cortar el círculo en partes más pequeñas e intentar reconstruir la esfera, pero la pregunta ahora se torna en cómo ordenar las piezas y ese es un problema aparte.

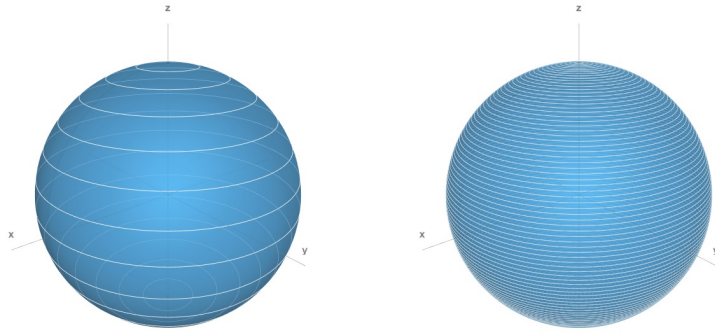


Fig. 2 Esferas con 15 cortes (izquierda) y 200 cortes (derecha)

Los argumentos más comunes para afirmar que el área superficial de una esfera es $4\pi r^2$, o sea, las *demostraciones*, usualmente utilizan herramientas de cálculo, por ejemplo, integración o proyección sobre un cilindro. Pero, aún cuando estas demostraciones explican por qué el área superficial es $4\pi r^2$, ninguna de ellas hace evidente cómo cuatro círculos forman una esfera.

En otras palabras, ¿por qué el área superficial de una esfera igual a la de cuatro círculos? O, reinterpretando la pregunta, ¿por qué la superficie de una esfera es igual a 4 veces su sombra? [1] (Fig. 1). La demostración que se desarrolla en el presente texto intenta responder a esta pregunta. Ciertamente no es la más "elegante", o la que usa las herramientas más sofisticadas, pero sí la que responde a la curiosidad.

La prueba

Considere una esfera de radio r, fijo, y realice cortes paralelos al plano xy, tal como se muestra en la figura 2. Entre dos cortes contiguos hay anillos circulares que identificaremos de acuerdo al ángulo θ que forman la línea que une al centro de la esfera con un punto del anillo y el eje z, es decir, el anillo R_θ corresponde al ángulo θ (véase 3). Además, denote a la apertura angular por $\Delta\theta$. Puede ser de utilidad pensar este problema como si fuera uno de integración: mientras $\Delta\theta$ se hace más pequeño, la suma del área de la superficie de los anillos se acerca cada vez más a la superficie de la esfera. Es decir, si observa lateralmente nuestra esfera, la superficie que calcularemos como una aproximación será aquella que resulta de unir "linealmente" dos cortes contiguos. Una mejor explicación gráfica se encuentra en la figura 4.

De acuerdo, con las definiciones anteriores, se tiene que θ toma valores entre 0 y π. Mientras que la apertura angular $\Delta\theta$ se hará más pequeña cada vez que realicemos más cortes regulares a lo largo de la superficie de la esfera y así las aproximaciones de la misma serán más precisas. Tal vez te resulte tentador interpretar la frase "$\Delta\theta$ cada vez más peque-

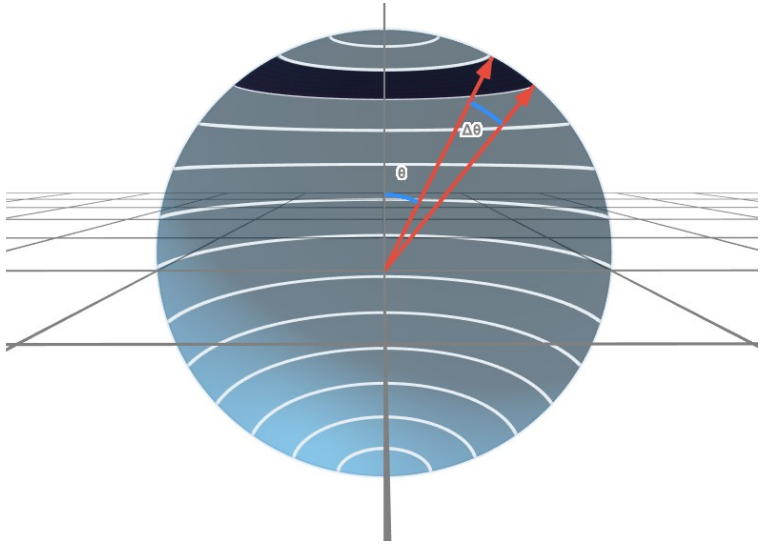


Fig. 3 El anillo R_θ se identifica de acuerdo al ángulo θ y el ángulo de cada anillo es $\Delta\theta$

ño" como la tendencia $\Delta\theta \to 0$ de un límite adecuado; sin embargo, toda tentación por integrar o tomar límites debe ser puesta a un lado, pues el camino que tomaremos muestra de forma más evidente la relación entre la superficie y su sombra.

Pregunta 1. *¿Cuál es la circunferencia de cada anillo (en términos de r y θ)?*

Para cada anillo R_θ calcularemos la circunferencia del anillo interno. Para ello, necesitamos encontrar el radio del anillo interno o, en otras palabras, hallar la longitud del cateto opuesto del triángulo formado por el eje z y el vector director del anillo R_θ (fig. 5). Por construcción, el ángulo es θ de modo que el radio interno que buscamos es

$$r\sin(\theta). \tag{3}$$

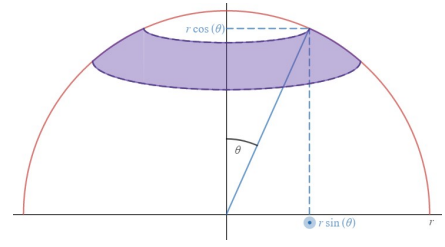


Fig. 5 Radio interno

De esta manera, la circunferencia del anillo R_θ es

$$2\pi r\sin(\theta). \tag{4}$$

Note que al multiplicar esta cantidad por $r\Delta\theta$, que es la longitud del arco generado por la apertura del anillo, obtiene una aproximación de la superficie del anillo, es decir

$$\mathcal{S}\left(R_\theta\right) = 2\pi r^2\sin(\theta)\Delta\theta. \tag{5}$$

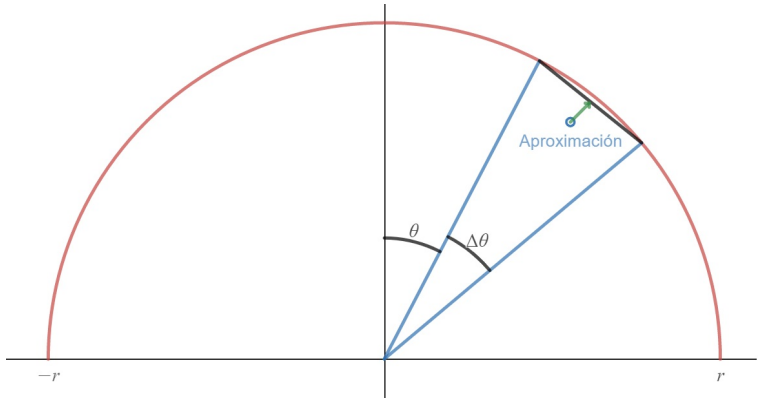


Fig. 4 Aproximación *lineal* de la superficie

Hasta este punto, si considera a $\Delta\theta$ como un diferencial e integramos sobre el dominio de θ, obtendríamos el mágico resultado. Sin embargo, tomaremos un camino alterno para mostrar la relación que perseguimos.

En orden de conseguir esa relación, se plantea lo siguiente:

Pregunta 2. *¿Cuál es el área de la sombra sobre el plano xy de uno de estos anillos R_θ (en términos de r, θ y $\Delta\theta$)?*

Recuerde que para obtener el área de un anillo bidimensional hay que encontrar el radio interno y externo que lo forman. Para esto, en la pregunta 1 ya hemos calculado el radio interno del anillo superficial, que coincide con el radio interno de la sombra del anillo. Por otro lado, note que el radio externo resulta ser la longitud del cateto opuesto del triángulo formado por el centro de la esfera y el corte externo de en la superficie de la esfera (véase la fig. 6). De este modo, el área de la sombra del anillo, denotada por $A(R_\theta)$, resulta ser

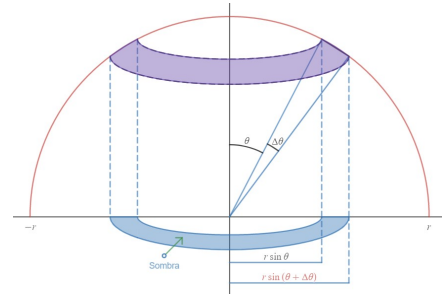


Fig. 6 Sombra de un anillo

$$A\left(R_\theta\right) = \pi r^2 \sin^2(\theta + \Delta\theta) - \pi r^2 \sin^2(\theta) \tag{6}$$

$$= \pi r^2 \left(\sin^2(\theta + \Delta\theta) - \sin^2(\theta)\right). \tag{7}$$

Para la siguiente pregunta tome en cuenta que el área de cada la sombre de un anillo es exactamente la mitad de la superficie *aproximada* (pregunta 1) de algún anillo distinto al que genera dicha sombra. Es decir:

Pregunta 3. *Para la sombra de un anillo cualquiera, ¿cuál es el anillo que tiene el doble de área?*

Ayuda replantear esta pregunta en otros términos: Sea θ_1 fijo. Encontrar θ_2 que satisfaga

$$A\left(R_{\theta_1}\right) = \frac{1}{2}\mathcal{S}\left(R_{\theta_2}\right) \tag{8}$$

donde $A\left(R_{\theta_1}\right)$ es el área de la sombra generada por el anillo de ángulo θ_1; y $\mathcal{S}\left(R_{\theta_2}\right)$ es el área de la superficie del anillo generado por el ángulo θ_2.

De la relación en (8), se obtiene que:

$$\pi r^2\left(\sin^2\left(\theta_1+\Delta\theta\right)-\sin^2\left(\theta_1\right)\right) = \frac{1}{2}(2\pi r^2\sin\left(\theta_2\right)\Delta\theta), \tag{9}$$

$$\iff \sin^2\left(\theta_1+\Delta\theta\right)-\sin^2\left(\theta_1\right) = \sin\left(\theta_2\right)\Delta\theta. \tag{10}$$

Si expande la suma de ángulo dentro del seno, consigue que:

$$\sin\left(\theta_1+\Delta\theta\right) = \sin\left(\theta_1\right)\cos(\Delta\theta)+\sin(\Delta\theta)\cos\left(\theta_1\right), \tag{11}$$

$$\approx \sin\left(\theta_1\right)+\cos\left(\theta_1\right)\Delta\theta. \tag{12}$$

Tome en cuenta que en (12) hemos considerado las aproximaciones $\cos(\Delta\theta)\approx 1$ y $\sin(\Delta\theta)\approx \Delta\theta$, pues se ha de considerar $\Delta\theta$ pequeña. Por lo tanto, al sustituir la aproximación (12) en el lado izquierdo de (10), se tiene que

$$\left(\sin\left(\theta_1\right)+\cos\left(\theta_1\right)\Delta\theta\right)^2-\sin^2\left(\theta_1\right) = 2\sin\left(\theta_1\right)\cos\left(\theta_1\right)\Delta\theta+\cos^2\left(\theta_1\right)(\Delta\theta)^2. \tag{13}$$

En la última expresión se puede remover el término $\cos^2\left(\theta_1\right)(\Delta\theta)^2$ pues su aportación se vuelve completamente marginal si toma una apertura de ángulo cercana a 0. De esta manera, se obtiene la siguiente relación:

$$\sin\left(2\theta_1\right) = \sin\left(\theta_2\right) \tag{14}$$

Es decir, $\theta_2 = 2\theta_1$ con la restricción[1] de tomar $\theta_1\in[0,\pi/2]$ para que tengamos que $\theta_2\in[0,\pi]$. En resumen, hemos demostrado que

$$A\left(R_\theta\right) = \frac{1}{2}\mathcal{S}\left(R_{2\theta}\right), \qquad \forall\theta\in\left[0,\frac{\pi}{2}\right]. \tag{15}$$

Pregunta 4. *Utiliza el resultado anterior para encontrar una correspondencia entre las sombras de los anillos en el hemisferio norte de la esfera y los anillos pares, es decir, cada 2 anillos en la superficie de la misma*

Notemos que la sombra de los anillos en el hemisferio norte generan un círculo de radio r. Además, cada sombra viene de un anillo cuyo ángulo es un múltiplo (entero) de la apertura angular, y gracias al resultado anterior, tenemos la siguiente relación:

[1] La función seno es inyectiva de 0 a $\pi/2$.

$$A\left(R_{k(\Delta\theta)}\right) = \frac{1}{2}\mathcal{S}\left(R_{2k(\Delta\theta)}\right). \tag{16}$$

La relación que se obtiene se interpreta como que la sombra producida por los anillos del hemisferio norte (un círculo de radio r) de la esfera es igual a la mitad de la suma de la superficie de los anillos pares (uno sí y uno no).

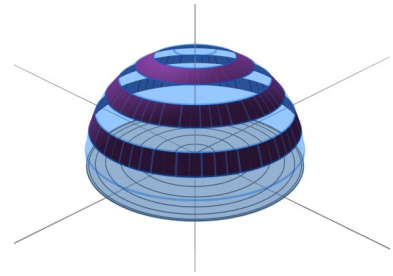


Fig. 7 La sombra de los anillos del hemisferio norte y algunos anillos *pares*

Pregunta 5. *¿Por qué lo anterior implica que el área de la sombra es la cuarta parte de la superficie?*

Note que el anillo superficial de ángulo θ en el hemisferio norte, tiene su simétrico en el ángulo $\pi - \theta$. Esto quiere decir que la sombra de dos anillos simétricos es cubierta por el anillo de ángulo 2θ. Es decir, gracias a la identidad (15) se tiene que

$$A\left(R_\theta\right) = S\left(R_{\pi-\theta}\right) \tag{17}$$

$$\implies S\left(R_\theta\right) + A\left(R_{\pi-\theta}\right) = \mathcal{S}\left(R_{2\theta}\right) \tag{18}$$

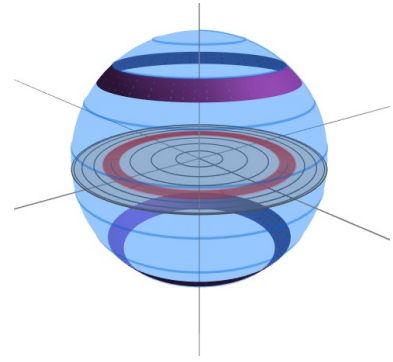


Fig. 8 Anillos simétricos

Luego, la mitad de la superficie está cubierta por lo anillos pares por lo que:

$$\mathcal{S}\left(R_{\text{pares}}\right) = \frac{1}{2}\mathbf{SE}, \tag{19}$$

donde $\mathbf{SE}$ representa la superficie total de la esfera. Por lo tanto, debido a la relación encontrada en la pregunta 4, se tiene que:

$$A_{\text{sombra}} = \frac{1}{2}\mathcal{S}\left(R_{\text{pares}}\right) = \frac{1}{4}\mathbf{SE}, \tag{20}$$

donde A_{sombra} representa el área de la sombra de la esfera. Finalmente, como la sombra es un círculo de radio r, el área de la sombra es πr^2, por lo que:

$$\mathbf{SE} = 4\pi r^2. \qquad \blacksquare \tag{21}$$

3blue1brown

La idea de los pasos seguidos en la prueba anterior se basan en un vídeo interactivo [1] del canal *3blue1brown* en el que se guía la demostración en forma de preguntas, pero sin dar la respuesta a ellas. A este tipo de resolución de problemas se les conoce como *ejercicios guiados*.

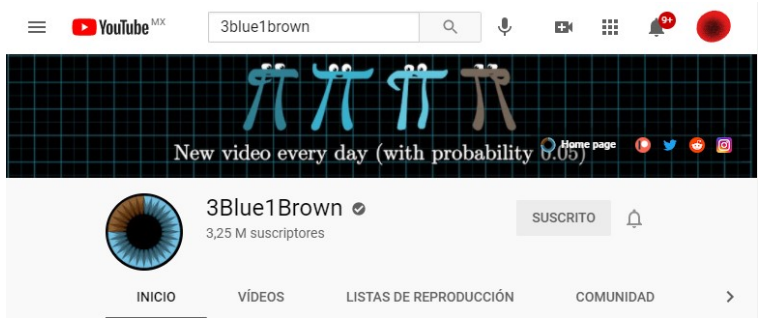


Fig. 9 Canal de **3blue1brown**

Dentro de la inmensidad de canales en Youtube con contenido matemático se encuentran *Julioprofe*, *Numberphile*, *Derivando*, entro otros. Uno de ellos, creado por Grant Sanderson, *3blue1brown* (fig. 9) es un canal que, en palabras de su creador, es una combinación de matemáticas y entretenimiento. El propósito de dicho canal es explicar problemas difíciles dentro de la matemática mediante animaciones y bajo simple cambios de perspectiva. En opinión de los autores que redactan estas líneas, los vídeos mantienen esa creatividad y curiosidad de la que se habló en párrafos anteriores.

En este canal podrán encontrar de todo, desde matemáticas recreativas, como la solución a las torres de Hanoi, hasta visualizaciones de conceptos más prácticos dentro de la matemática aplicada, tales como la simulación de una epidemia, aproximaciones mediante series de Fourier; o también soluciones a problemas de olimpiadas matemáticas. En fin, la única limitante de los vídeos de este canal es su propia naturaleza de ser finitos. Invitamos a los lectores a que se maravillen del basto contenido ofrecido por este canal y que, de paso, interactúen con algunas otras plataformas dedicadas a la divulgación de contenido matemático de forma accesible pero con el suficiente rigor.

Conclusiones

En una clase de matemáticas es común que los ejercicios vistos en clase sirvan como base para las tareas o los exámenes, ya sean en el resultado mismo o en el procedimiento. Esto puede impulsar en el joven estudiante la idea de que las demostraciones solo son útiles en tanto pueden contribuir a un resultado más grande. Evidencia suficiente para esta aseveración se puede encontrar en la famosa frase "¿Y esto va a venir en el examen?" y la inmediata pérdida de interés en más de una persona cuando la respuesta es negativa.

En cierto sentido, este fenómeno es tan solo un eco distante de la arbitrariedad que discutíamos al principio de este artículo. Si una demostración no tiene valor en si misma, entonces tan solo es otra herramienta arbitraria más destinada a usarse para otro propósito desconocido aún más grande. Y bajo este razonamiento podríamos haber navegado cien

leguas en el mar infinito del conocimiento y aún así sentir que no llegamos a ningún lado.

Combatir y aplacar la arbitrariedad es entender la belleza de los resultados tan solo por lo que significan, y no solo por lo que no pueden traer más adelante. Es redescubrir la fórmula del área superficial de una esfera tan solo por querer hacer evidente la relación con su sombra. En el gran mar del conocimiento este no es un resultado que no conociéramos antes, mas el valor de la prueba no está en la fórmula, sino en su capacidad de satisfacer la curiosidad. Porque la curiosidad, tan personal como es, es uno de los mejores motivadores que existen; es nuestra propia linterna que nos guía en nuestra travesía por el mar infinito que son las matemáticas.

Y para aquellos que encuentran su pasión en la parte "aplicada", de Matemáticas Aplicadas, esa linterna es aún más importante pues surcarán más de un solo mar en su travesía. Las matemáticas son tan nobles que encajan en cualquier campo: computación, economía, estadística, química, lógica, inteligencia artificial, medicina, hasta dentro de la filosofía se encuentran los primeros fundamentos lógicos. Puro o aplicado, es fácil reconocer a quien disfruta y ama esta ciencia exacta con tan mala reputación. Basta con encender una chispa de curiosidad frente a sus ojos. Esa cuestión no lo abandonará hasta haberla saciado y, con suerte, en el camino habrá encendido otras tantas llamaradas por si mismo.

Así como el profesor Guillermo Grabinsky dijo una vez [2]:
"Amen las matemáticas, hagan matemáticas lo mejor que puedan, disfruten las matemáticas, déjense deslumbrar. [...] Yo sigo deslumbrado."

Referencias

[1] 3blue1brown. But why is a sphere's surface area four times its shadow?, Dic 2018. Disponible en este link https://www.youtube.com/watch?v=GNcFjFmqEc8.

[2] Mesa redonda: Matemáticas puras vs aplicadas, Sep 2019. Disponible en este link https://web.facebook.com/watch/live/?v=498676644027547ref=watch_permalink.

Software

Las imágenes del presente texto fueron creadas en desmos.com y math3d.org.

Efemérides y Polinomios de Chebyshev
Un Vistazo a la Teoría de Aproximación en la Mecánica Celeste

Julieta Rivero González
Estudiante de Matemáticas Aplicadas

Las efemérides planetarias, en el contexto de la astronomía, son tablas que permiten determinar la posición de un objeto celeste o un satélite artificial con respecto a la Tierra en un momento dado, así como su trayectoria a lo largo del tiempo. Dichas tablas fueron herramientas indispensables para la navegación en siglos pasados y actualmente, siguen siendo utilizadas con frecuencia en proyectos de navegación e ingeniería espacial, así como en la predicción y el análisis de fenómenos astronómicos.

El primer registro que se tiene de la existencia de una efemérides es en la astronomía babilónica, alrededor de un milenio A.C.; desde entonces, estas herramientas han sido perfeccionadas por notables astrónomos de diversas civilizaciones. Entre las efemérides más destacables, por su uso y precisión a lo largo de la historia, están las siguientes: las *Handy Tables* de Ptolomeo, las cuales marcaron un precedente en el cálculo de las posiciones del Sol, la Luna y los planetas; las Tablas de Toledo, elaboradas por un grupo de astrónomos árabes en España en el año 1080 D.C. con base en la investigación astronómica islámica; las Tablas Alfonsinas del siglo XIII, que fueron ampliamente usadas por casi 300 años; las Tablas Pruténicas, las primeras efemérides que consideraban la teoría copernicana; y las Tablas Rudolfinas de Kepler, las cuales ya tomaban en cuenta el movimiento elíptico de los planetas.

Actualmente, las efemérides son generadas a partir de algoritmos que calculan las posiciones de los planetas, asteroides, cometas y satélites con una gran precisión. Los datos en estas tablas, además de ser capaces de estimar la trayectoria de un objeto, también permiten el cálculo de otras propiedades de los cuerpos celestes en un tiempo pasado o futuro, como su velocidad y aceleración orbital, el ángulo de elongación[1], el diámetro aparente, el brillo y su tránsito planetario[2].

El problema matemático detrás de generar una efeméride se resume en determinar una función que aproxime la posición del objeto celeste con respecto a la Tierra en un tiempo determinado. Para algunos lectores, no será sorpresa que la respuesta a este problema hace uso de la Teoría de Aproximación, en particular, de los polinomios de Chebyshev; por otro lado, quienes no están tan familiarizados con este campo del análisis matemático, descubrirán en los próximos párrafos por qué el uso de este método de aproximación es idóneo en este contexto.

Los polinomios de Chebyshev son una familia de polinomios ortogonales[3] que, como se muestra a continuación, se construyen a partir de la fórmula de De Moivre, para valores de x en el intervalo cerrado $[-1, 1]$.

$$T_n(x) = \cos(n\theta), \quad \theta = \arc\cos(x)$$

[1] En astronomía, se conoce como elongación al ángulo formado entre un planeta y el sol con respecto a la Tierra.

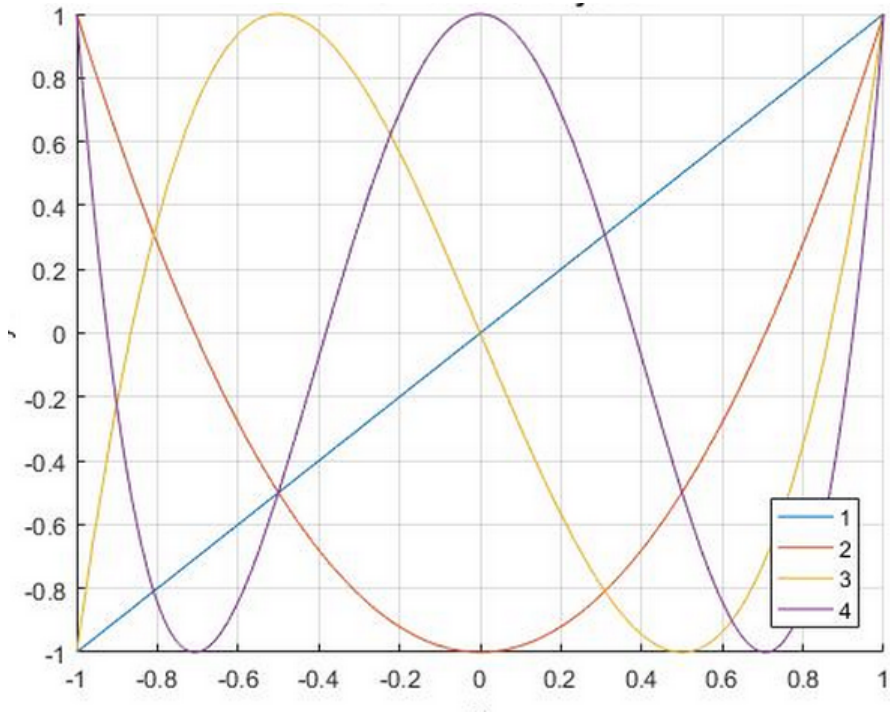
[2] Este término se refiere a cuando un astro se posiciona entre el Sol y la Tierra.

[3] Dos funciones f, g son ortogonales si su producto escalar es nulo.

A partir de esta relación, es posible definir de manera recursiva a estos polinomios como sigue

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{n+1}(x) = 2x * T_n(x) - T_{n-1}(x)$$

En particular, la representación gráfica de los primeros cuatro polinomios de Chebyshev es útil para entender la idea detrás de la fórmula recursiva dada anteriormente, ver Gráfica 1.



Gráfica 1. Los cuatro primeros polinomios de Chebyshev[4]

Además, analizando un poco más a fondo la imagen anterior y considerando la definición dada de los polinomios de Chebyshev, podemos llegar a las siguientes conclusiones: el grado de cada polinomio depende de n; el n-ésimo polinomio de Chebyshev tiene n raíces en $[-1, 1]$, las cuales corresponden a los puntos $x_n = \cos\left(\frac{(2n-1)\pi}{2n}\right)$; los valores extremos de este tipo de polinomios son 1 o -1.

Por otra parte, la razón por la cual los polinomios de Chebyshev son excelentes candidatos para la aproximación de una función, se debe a la siguiente proposición:

$$P(x) = \frac{1}{2^{k-1}} * T_k(x)$$

es el polinomio mónico de grado k con mínima norma en el intervalo $[-1, 1]$[5].

Es decir, de todos los polinomios de grado k, el que alcanza un mínimo valor con respecto

[4]Para aquellos que estén interesados en explorar más a fondo la representación gráfica de los polinomios de Chebyshev, en la referencia correspondiente a Gráfica 1 podrán encontrar el código para generar esta imagen.

[5]Un polinomio mónico de grado k es aquel cuyo coeficiente principal es 1.

a la norma del supremo es proporcional al k-ésimo polinomio de Chebyshev. A su vez, esta propiedad implica que cualquier polinomio P de orden k, definido en el intervalo $[-1, 1]$, con coeficiente a_k en el término de mayor orden, cumple con que

$$\|P\|_\infty = \max_{x\in[-1,1]}(|P(x)|) \geq \frac{|a_k|}{2^{k-1}}$$

Las características y propiedades descritas anteriormente nos permiten hacer algunas observaciones importantes: en primer lugar, las raíces de estos polinomios, también conocidos como nodos de Chebyshev, pueden ser usados como nodos de interpolación para funciones a aproximar; por otra parte, determinando coeficientes apropiados, una combinación lineal de polinomios de Chebyshev es útil para la aproximación precisa[6] de una función objetivo; finalmente, como los polinomios son funciones continuas y diferenciables, la primera y segunda derivada de los polinomios de Chebyshev nos dan información relevante sobre la función a la que aproximan.

Así pues, los polinomios de Chebyshev cumplen con una variedad de propiedades que los hacen únicos, versátiles y fáciles de aplicar en algoritmos numéricos. Por estas razones, esta familia de polinomios no solo aparece recurrentemente en la teoría e investigación, sino que ha sido usada en diversos problemas prácticos, tanto en el contexto de la computación, la física y la mecánica, así como en el contexto particular de la mecánica celeste.

A continuación, se describirá brevemente el uso que se le da a los polinomios de Chebyshev en el caso particular de las efemérides para conseguir una función que aproxima la posición de un cuerpo celeste a través del tiempo. En primer lugar, se considera la serie definida como

$$p(x) = \sum_{n=0}^{N} p_n T_n(x),$$

con $x \in [-1, 1]$ y T_n el n-ésimo polinomio de Chebyshev, la cual describe la posición del objeto considerado en el tiempo. Esta función tiene por derivada a

$$v(x) = \sum_{n=0}^{N} v_n T_n(x),$$

con $v_n = 0$, función que aproxima a la velocidad a la que viaja el objeto. Finalmente, la segunda derivada de esta función está dada por

$$a(x) = \sum_{n=0}^{N} a_n T_n(x),$$

con $a_n = 0 = a_{n-1}$, que describe su aceleración.

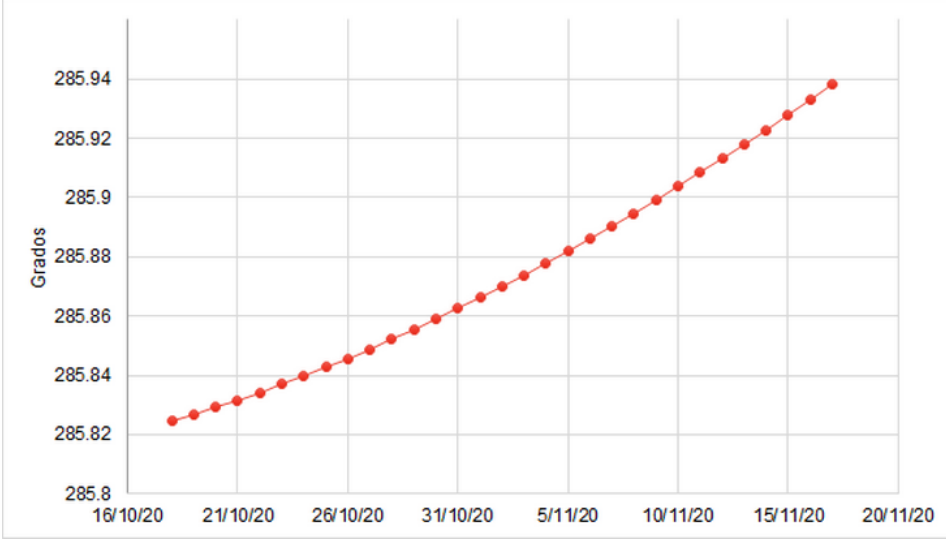
De esta manera, determinando los coeficientes p_n, es posible describir la posición, la velocidad y la aceleración de un determinado cuerpo celeste en un punto del tiempo t, a lo largo de un periodo $[t_0, t_1]$, tomando $x = -a + 2\frac{t-t_0}{t_1-t_0}$.

[6] Precisa en el sentido de que el error de aproximación con respecto a la función original es pequeño.

El método usual para determinar los coeficientes de una determinada función aproximante de la trayectoria de un cuerpo celeste particular no se discutirá en este artículo, dado que este se basa en la resolución de un sistema de ecuaciones que no se relaciona directamente con los polinomios de Chebyshev. Sin embargo, para aquellos lectores que estén interesados en dicho método, en el artículo *"Numerical Representation of Planetary Ephemerides"*, Newhall lo describe con detalle[7].

Finalmente, a manera de conclusión de este artículo y teniendo en mente la información que el lector ha adquirido a lo largo del mismo, es pertinente mencionar que el centro de investigación *NASA Jet Propulsion Laboratory* ha desarrollado un sitio web, llamado *JPL HORIZON System*, que permite a cualquier cibernauta consultar efemérides de alta precisión, de diversos objetos celestes en el sistema solar[8] a través de largos periodos de tiempo pasados y futuros. Estos resultados son generados al instante a partir de los métodos y herramientas matemáticos descritos anteriormente.

Para ilustrar la poderosa herramienta que esta página[9] ofrece y reafirmar al lector que sí existen aplicaciones concretas a partir de las abstractas ideas con las que uno puede toparse en un curso de matemáticas, me tomé la libertad de elaborar las siguientes gráficas[10] que describen la posición de Saturno en un periodo de 30 días[11].



Gráfica 2. Ascensión Recta de Saturno con respecto a la Tierra del 18 de octubre al 17 de noviembre del 2020

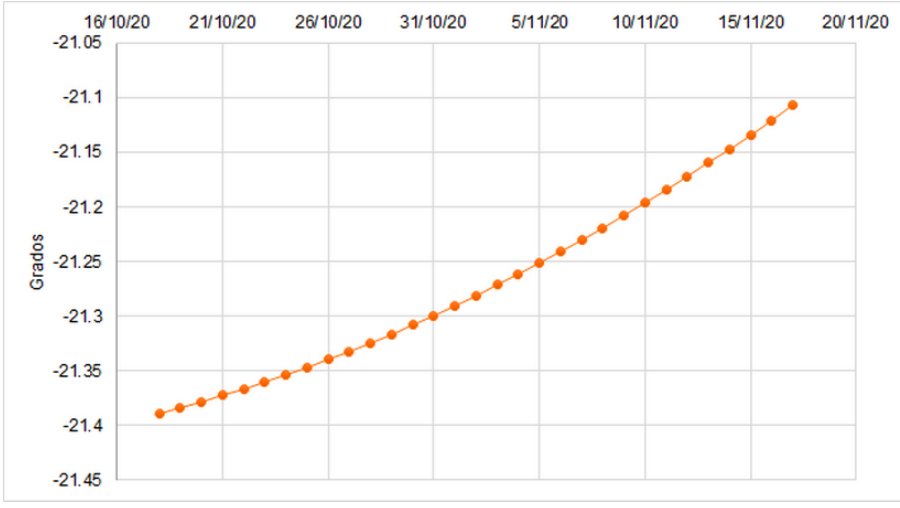
[7] El enlace puede encontrarse en la lista de referencias.

[8] Incluyendo a 10,166,065 asteroides, 3679 cometas, 209 satélites naturales, 8 planetas y el Sol.

[9] `https://ssd.jpl.nasa.gov/?horizons`

[10] Ambas gráficas fueron elaboradas a partir de los datos encontrados en `https://ssd.jpl.nasa.gov/horizons.cgi#results`

[11] La ascensión recta y la declinación son coordenadas astrológicas que se miden en ángulos y se usan para determinar la posición de un objeto en el espacio con respecto a la Tierra.



Gráfica 3. Declinación de Saturno con respecto a la Tierra del 18 de octubre al 17 de noviembre del 2020

Referencias

[1] Wikipedia (2020). *Ephemeris*. https://en.wikipedia.org/wiki/Ephemeris

[2] W. Thompson (1994). *Chebyshev Polynomials: After the Spelling the Rest is Easy*. Computers in Physics. Vol(8), 161-165. https://aip.scitation.org/doi/pdf/10.1063/1.4823278

[3] A. Franco (2016). *Polinomios de Chebyshev*. http://www.sc.ehu.es/sbweb/fisica3/especial/chebyshev/chebyshev.html

[4] F. Bach (2019). *Polynomial magic I: Chebyshev polynomials*. https://francisbach.com/chebyshev-polynomials/

[5] E. Byhova, A. Tamarov (1978). *Representation on planetary satellite ephemerides by Chebyshev polynomials*. http://articles.adsabs.harvard.edu//full/1978SvAL....4..203B/0000203.000.html

[6] X. Newhall (1989). *Numerical Representation of Planetary Ephemerides*. http://articles.adsabs.harvard.edu//full/1989CeMec..45..305N/0000305.000.html

[7] NASA. *Jet Propulsion Laboratory (s.f). HORIZON System*. https://ssd.jpl.nasa.gov/?horizons

Dimensiones no enteras y curvas con área

Francisco Aramburu
Estudiante de Matemáticas Aplicadas e Ingeniería en Computación

"La esencia de las matemáticas reside en su libertad"
George Cantor

Introducción

Más allá de definir formalmente y demostrar nociones de dimensiones fractales, que es bastante tedioso, éste artículo va a motivar la definición de dimensión de fractal o dimensión de Hausdorff de manera intuitiva y se mostrán ejemplos de dos dimensiones fractales particularmente interesantes.

Un poco de historia

George Cantor, uno de los matemáticos más renombrados de finales del s.XIX, pasó mucho tiempo intentando demostrar la hipótesis del continuo y durante este tiempo llegó al famoso resultado que afirma que la cardinalidad del intervalo unitario en $\mathbb{R}$ es la misma que la cardinalidad de $\mathbb{R}$. Sabiendo esto, la comunidad metemática de la época se preguntó de la existencia de una función con dominio en el intervalo unitario y contradominio en una curva que cubra toda el área del cuadrado unitario. En 1890 el italiano Guiseppe Peano encontró una curva con tales caracteristicas y después de horas y horas de meditación decidió llamarla "Curva de Peano", al año siguiente el matemático David Hilbert descubre otra curva de llenado de espacio hoy conocida como "Curva de Hilbert". Pero, ¿qué tiene que ver esto con dimensiones no enteras? Para poder contestar la pregunta anterior primero debemos motivar la noción de dimensión de Hausdroff.

Dimensión de Hausdorff

Intuitivaemte, sabemos que una recta tiene una longitud mayor a cero y un área igual a cero, un cuadrado tiene un área mayor a cero y un volumen igual a cero. Ahora, como buenos matemáticos ¿tiene sentido por la longitud de un cuadrado? La respuesta es que sí, y que ésta longitud es infinita. La afirmación anterior nace del siguiente razonamiento: si en un cuadrado hay infinitas rectas que lo cubren, al sacar cada una de esas rectas del cuadrado y unirlas unas con otras llegaremos a una nueva recta de longitud infinita (ver Figura 1).

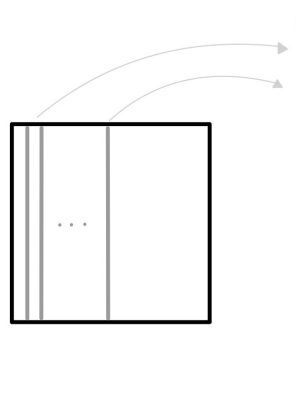


Figura 1: Construcción de recta infinita a partir de un cuadrado

Lo mismo podemos decir sobre un cubo y su área, si laminamos el cubo en cuadrados terminamos con un cantidad infinita no numerable de cuadrados con área 1 por lo que podemos unir los cuadrados arista con arista que generará un área infinita, por lo que el cubo tiene también un área infinita. La siguiente tabla resume lo explicado anteriormente.

	Longitud (med_1)	Área (med_2)	Volumen (med_3)
Segmento (Dim_1)	> 0	0	0
Cuadrado (Dim_2)	∞	> 0	0
Cubo (Dim_3)	∞	∞	> 0

Félix Hausdroff observó que para un objeto de dimensión n, con n en los naturales, el valor de med_{n-1} es infinito y el valor de med_{n+1} es igual a cero. De cierta forma la dimensión del objeto es un parteaguas en los valores de las medidas. De lo anterior nace la noción de dimesión de Hausdroff, si cambiamos la n por una k en los reales, cuando k es la dimensión de un objeto se cumple que $med_{\lfloor k \rfloor}$ es infinita y $med_{\lceil k \rceil}$ es cero.

Ahora bien, supongamos que queremos reescalar un objeto T por un factor de 2. Si el objeto es de una dimensión entonces la longitud será dos veces la longitud original. Si el objeto es un cuadrado entonces el área final será 4 veces el área original y si fuera un cubo el volumen final será 8 veces el volumen original. Generalizando el objeto a una dimensión k y un factor de reescalamiento x llegamos a la siguente igualdad:

$$med_k(r_x T) = x^k med_k(T) \tag{1}$$

y es fácil ver que

$$med_k(r_x(T_1 + T_2)) = x^k med_k(T_1) + x^k med_k(T_2). \quad (2)$$

De ahora en adelante, por simplicidad, llamaremos dimensión a la dimensión de Hausdorff.

Conjunto de Cantor

La construcción del conjunto de Cantor es como sigue:

1. Tomar una recta de longitud 1;
2. Dividir la recta en tres y eliminar segundo tercio;
3. Repetir el segundo paso con los segmentos restantes.

La Figura 2 muestra 3 iteracciones de la construcción del conjunto.

Figura 2: Construcción del conjunto de Cantor

Dado que el ternario de Cantor está contenido en la recta, tiene sentido preguntarnos por la longitud del ternario en cada iteración.

Paso	Cantidad de segmentos	Longitud individual	Longitud total
1	2	$1/3$	$2/3$
2	4	$1/9$	$4/9$
3	8	$1/27$	$8/27$
...	...	...	...
n	2^n	$1/3^n$	$2^n/3^n$

Ahora, un punto tiene dimensión cero, al igual que una cantidad infinita numerable de puntos. Se puede mostrar que el conjunto de cantor tiene una cantidad de puntos infinita no numerable por lo que concluimos que la dimensión del conjunto de Cantor es mayor a cero. Ahora en cuanto a la longitud, tomando el límite cuando n tiende a infitinto vemos

que la longitud total se va a cero entonces el conjunto tiene dimensión menor a uno y mayor a cero.

Sea D la dimensión que buscamos y T el conjunto de Cantor. En iteración cada uno de los segmentos se reescala a un factor de $1/3$ dos veces. Usando (1) y (2) después de la primer iteración tendremos:

$$med_D(T) = \left(\tfrac{1}{3}\right)^D med_D(T) + \left(\tfrac{1}{3}\right)^D med_D(T) = 2\left(\tfrac{1}{3}\right)^D med_D(T)$$

Como $med_D(T)$ no es cero ni infinito, pues D es la dimensión, despejamos

$$1 = 2\left(\tfrac{1}{3}\right)^D$$

entonces

$$D = ln(2)/ln(3) = 0.6309...$$

es la dimensión del conjunto de Cantor que, aparte de ser no entera, es irracional.

Curva de Hilbert

Ahora un ejemplo que a mi parecer es de los más interesantes, la curva de Hilbert. Una curva de llenado de espacio. La construcción de la curva de Hilbert es como sigue:

1. Dividir al cuadrado unitario en cuatro cuadrados iguales y unir sus centros con una recta;
2. Reescalar la recta y poner una copia en cada uno de los 4 sub-cuadrados;
3. Rotar las figuras inferior izquierda y derecha noventa grados a la derecha e izquierda respectivamente;
4. Unir con una recta las figuras;
5. Repetir los pasos 2, 3 y 4 con la nueva curva.

En la Figura 3 se muestra el inicio de la construcción.

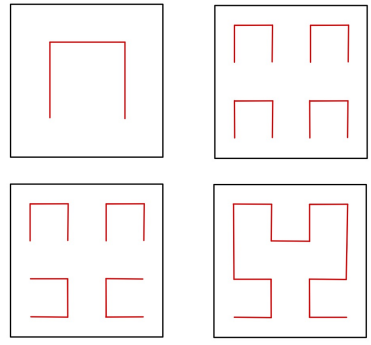


Figura 3: Primer iteración de la curva de Hilbert

Y a continuación se muestran la segunda, tercera y cuarta iteración.

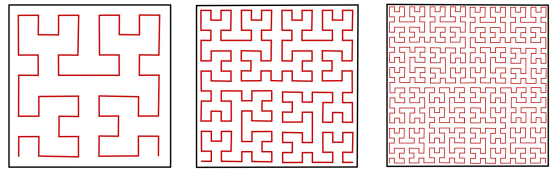


Figura 4: Iteraciones 2, 3 y 4 de construcción de la curva de Hilbert

Ahora exploramos cómo va cambiando la longitud de la curva en cada iterarción.

Paso	Cantidad de segmentos	Longitud individual	Longitud total
1	$2^2 - 1$	$1/2$	$3/2$
2	$2^4 - 1$	$1/4$	$15/4$
3	$2^6 - 1$	$1/2^3$	$63/8$
4	$2^8 - 1$	$1/2^4$	$255/16$
...	...	...	...
n	$2^{2n} - 1$	$1/2^n$	$(2^{2n} - 1)/2^n$

Tomamos el límite de la longitud total.

$$\lim_{x\to\infty} \frac{2^{2n}-1}{2^n} = \infty$$

por lo que concluimos que la dimensión de la curva de Hilbert es mayor a uno. Ahora, como la curva nunca sale del cuadrado unitario y tiene longitud infinita, basándonos en lo

explicado sobre la Figura 1 intuimos que la dimensión de la curva es 2 y su área es 1. En otras palabras, el área de los puntos del cuadrado que no cubre la curva de Hilbert es cero. Es decir, la curva de Hilbert es tan compleja que, de cierta forma, es una superficie. Lo anterior se puede demostar pero sale de los objetivos de éste artículo.

Conclusión

El pensar que un objeto tiene una dimensión no entera o que existan curvas con área puede tomarse como algo bonito pero que se queda en plano teórico, increiblemente pasa todo lo contrario. Por ejemplo, la extencón tridimensional del conjunto de Cantor, mejor conocido como esponja de Menger, ha motivado a usar dimensiones fractales para describir o caracerizar densidad de piedras y pordria usarse para describir estructuras biológicas complejas como los alveolos pulmonares. La curva de Hilbert, aparte de servir para tener un muy buen juego de *snake*, usualmente se usa para transformar sonido en imágenes y viceversa, esto porque al subir o bajar la sensibilidad del muestreo de un sonido, los puntos que representan cada sección de la onda sonora quedan cerca en la imagen, cosa que no pasaría si mapeáramos una onda a una imagen usando lineas rectas. La principal motivación de éste artículo fue el primer libro de la bibliografía, que apliamente recomiendo si el lector disfrutó los temas tratados a lo largo de este texto.

Referencias

[1] Benoit Mandelbrot. *La geometría Fractal de la Naturaleza*. W. H. Freeman and Company, 1982.

[2] "The Uncountability of the Unit Interval" 8/04/2021
`https://arxiv.org/abs/1209.51`

[3] "Introduction to dimension theory and fractal geometry: Fractal dimensions and measures" 5/04/2021
`http://pi.math.cornell.edu/~erin/docs/dimension.pdf`

[4] "Topological dimensions, Hausdorff dimensions and fractals" 2/04/2021
`https://u.math.biu.ac.il/~megereli/final_topology.pdf`

[5] Maria Isabel Binimelis Brasa. *Una nueva manera de ver el mundo-Geometría fractal*. RBA, 2011.

[6] Gustavo Ernesto Piñeiro. *La esfera que quiso ser infinita-Las paradojas de la medida*. RBA, 2014.

Paradoja de Bertrand Russell en teoría de conjuntos

David Isaac López Romero
Estudiante de Matemáticas Aplicadas y Actuaría

Introducción

Para hablar de paradojas, es necesario saber qué es una paradoja. Una paradoja es aquel razonamiento que, aunque parezca correcto, genera una contradicción. En matemáticas, las paradojas surgen de algún planteamiento erróneo, una deducción incorrecta o una operación con error lógico.
La paradoja de Russell es una contradicción que se manifiesta en la teoría de conjuntos, la cual constituye la base de las matemáticas. En el siglo XIX, Georg Cantor, junto al matemático Gottlob Frege, formula la primera teoría de conjuntos. (Huertas Sánchez & Manzano Arjona, 2002) La teoría de conjuntos de Cantor se puede sintetizar como sigue: "un conjunto es cualquier colección [llamado] C de objetos determinados y bien distintos [llamados] x de nuestra percepción o nuestro pensamiento (que se denominan elementos de C), reunidos en un todo." (Hernández H., 2003) El conjunto no es la única clase que existe, pues pueden existir clases más amplias que otras. Por ejemplo, se dice que una colección de los elementos de y es subconjunto de C si sus elementos y cumplen con las características del conjunto C, más algunas otras propiedades.

Paradoja de Russell

Bertrand Russell propone que no puede existir un conjunto de todos los conjuntos que no se contienen a sí mismos como elementos. Es complicado entender esta paradoja por su proposición formal. En la literatura se utiliza un ejemplo muy conocido, la paradoja del barbero. Sin embargo, en esta ocasión se muestra uno distinto.

Imaginemos un pueblo chico donde solo hay un médico -cirujano- que opera a todas las personas de ese pueblo, puesto que no se pueden operar a sí mismas. Cuando el cirujano requiere de alguna cirugía, él no puede llevar a cabo la operación porque no se puede operar a sí mismo; de este modo, el cirujano necesita que otro especialista lo opere. En conclusión, no existe una persona que opere a todas las personas.

Una vez que se tiene la intuición del problema, surge la siguiente pregunta: ¿cómo se observa lo anterior en términos matemáticos? Russell plantea que que no puede existir un conjunto de todos los conjuntos que no se contienen a sí mismos como elementos, como anteriormente se comenta.

Primero, se plantea la siguiente situación: "Para toda propiedad $\Phi(X)$ definible en la teoría, existe un conjunto Y cuyos elementos son exactamente los conjuntos X que cumplen la propiedad $\Phi(X)$." Es decir, formalmente, la teoría postula la existencia del conjunto:

$$Y = \{X | \, \phi(X)\}$$

Russell, a partir de lo anterior, descubre que al no haber límites en el conjunto, se puede dar la siguiente equivalencia:

$$\phi(X) \equiv X \notin X$$

En otros términos, la propiedad $\phi(x)$ expone que algún conjunto X no se debe tener a sí mismo como elemento; entonces, si esto es posible, debe existir un conjunto cualquiera que contenga a todos los conjuntos que no se contengan a sí mismos, sin ninguna excepción. De este modo, se puede plantear la existencia del conjunto con la especificación anterior, es decir, el conjunto Y es aquel que contiene a todos los conjuntos que no se contienen a sí mismos.

$$Y = \{x | x \notin x\}$$

Entonces se define a Y de la siguiente manera.

$$\forall x \in Y \iff x \notin x$$

A manera de extensión, el conjunto Y se compone de los elementos x que no sean elementos de sí mismos. Como Y se compone de conjuntos, la definición se modifica, y todas las posiciones de x se sustituyen por un conjunto.

$$\forall X \in Y \iff X \notin X$$

Entonces, se entiende que todo conjunto X pertenece al conjunto Y si, y solo si, el conjunto no es elemento de sí mismo.

Sin embargo, cuando X se sustituye por el conjunto Y, la ecuación no puede ser válida.

$$Y \in Y \iff Y \notin Y$$

Lo anterior expone que el conjunto Y es elemento de Y si, y solo si, Y no es elemento de Y. Es evidente que se genera una contradicción; por consiguiente, Bertrand Russell demostró que no es posible el conjunto de todos los conjuntos que no se contienen a sí mismos.

Resoluciones a la paradoja

Bertrand Russell no detuvo su trabajo en descubrir la paradoja, sino que, con ayuda de Alfred N. Whitehead se expone la resolución a la contradicción; no obstante, fue poco aceptada en el mundo matemático pues era un postulado complejo. (Russell & Whitehead, 2010) Más tarde, Ernst Zermelo y Adolf Fraenkel reformularon toda la matemática y crearon una

nueva teoría de conjuntos, la cual se compuso de diez axiomas (Hernández H., 2003) (Universidad de Buenos Aires, 2012). Así pues, surge una nueva cuestión. ¿Qué axiomas resuelven la paradoja de Bertrand Russell?

Si la siguiente sentencia permitía la paradoja de Russell:

> "Para toda propiedad $\Phi(X)$ definible en la teoría, existe un conjunto Y cuyos elementos son exactamente los conjuntos X que cumplen $\Phi(X)$."

Era necesario que la propiedad de $\Phi(X)$ se restringiera; de esta forma, se postuló un nuevo axioma a la nueva teoría de conjuntos, también conocido como "esquema axiomático de especificación". Dicho axioma postula que los conjuntos se obtienen de otros conjuntos ya dados. En otras palabras, todo conjunto tiene un conjunto más grande que lo contiene; por consiguiente, la fórmula $x \in x$ no es posible, ya que un conjunto no se puede extraer de sí mismo. (Hernández H., 2003)

Este nuevo axioma, permitió que el postulado se escriba de una nueva manera:

> "Para toda propiedad $\Phi(X)$ definible en la teoría y todo conjunto U, existe un conjunto Y cuyos elementos son exactamente los elementos $X \in U$ que cumplen $\Phi(X)$."

El nuevo objeto U es el conjunto que tiene a todos los objetos, según el contexto dado. A dicho elemento se le llamó universo o conjunto universal. Entonces el conjunto Y que se definió con anterioridad se reescribe del siguiente modo:

$$Y = \{X \in U | \phi(X)\}$$

Entonces, el conjunto Y no es conjunto de todos los conjuntos que no se contienen a sí mismos como elementos, sino que es un conjunto de todos los conjuntos de un sector bien definido. En resumen, Y es un conjunto que contiene todos los elementos a evaluar o estudiar.

$$\nexists X : (\forall u | (u \in X))$$

Conclusiones

La teoría de conjuntos ha cambiado a lo largo de los años con diversos axiomas para evitar posibles contradicciones o paradojas. Como la teoría conjuntista es la base de la matemática actual, se necesita de una base sólida para su continua elaboración e investigación.

En este caso, la paradoja de Russell no se puede resolver, ya que es imposible que el problema no genere alguna contradicción; sin embargo, se puede evitar el conflicto. Zermelo y Fraenkel limitaron la teoría inicial de conjuntos de Cantor con el fin de que la paradoja de Russell no se pueda plantear.

Referencias

[1] Hernández H., F., *Teoría de conjuntos*, Universidad Nacional Autónoma de México, México, DF, 2003.

[2] Huertas Sánchez, A. y Manzano Arjona, *Teoría de conjuntos*, Universidad Complutense de Madrid, https://webs.ucm.es/info/pslogica/teoriaconjuntos.pdf, Madrid, España, Febrero 2002.

[3] Russell, B. y Whitehead, A. N., *Principia Mathematica*, Routledge Classics, Reino Unido, 2010

[4] Universidad de Buenos Aires *Teoría axiomática*, Departamento de Matemática, http://www.dm.uba.ar/materias/optativas/topicos_de_logica/2012/1/Capitulo_1-Bibliografia.pdf, 2012

Introducción a las Pruebas de Primalidad

Carlos Galeana Hernández
Alumno de Matemáticas Aplicadas

Introducción

Ya en 1798, Gauss, en sus *Disquisitiones Arithmeticae*, auguraba: ."El problema de distinguir entre números primos y números compuestos es uno de las más importantes y útiles de la aritmética"[1].
En efecto, en 1977 se desarrolló el sistema criptográfico RSA [3], que ha pasado la prueba del tiempo y aún es usado para garantizar la confidencialidad de las transacciones electrónicas. RSA es un algoritmo fuértemente ligado a los números primos: el proceso de cifrado requiere de números primos grandes (del orden los trescientos dígitos), las llamadas *llaves públicas*.
Generar números primos de tal magnitud es un problema muy estudiado, y casi siempre involucra resolver un problema en apariencia más sencillo, pero igualmente escencial; el mismo que interesó a Gauss en uno de los textos más importantes sobre Teoría de Números: ¿Cómo saber si un número dado es primo o compuesto?
En este artículo describiremos una solución a este problema, estudiando primero los conceptos y propiedades de números primos que hacen posible su funcionamiento.

Números primos

Recordemos una definición esencial para nuestros próximos esfuerzos, la de divisibilidad. Decimos que un número entero a divide a otro número b (y escribimos $a \mid b$) si existe un tercer número c tal que $ac = b$. Además, a a y c los llamamos divisores de b. Por ejemplo, $2 \mid 8$ (pues $2 \times 4 = 8$), $3 \mid 15$, pero 6 no divide a 21.
Por otro lado, decimos que un número p es primo si sus únicos divisores son 1 y p. Consideremos por ejemplo al número 18 que tiene 6 divisores: 1, 2, 3, 6, 9 y 18. Efectivamente, 1 y 18 son divisores de 18, pero también lo son 2, 3, 6, y 9, de tal forma que 18 no es un número primo (a este tipo de números solemos llamarle compuestos). En contraste 19 tiene solo dos divisores: 1 y 19, lo que lo convierte en un número primo.

Una Primera Prueba de Primalidad

Usemos la definición de número primo para diseñar un primer algoritmo que identifique si un número es primo o compuesto (prueba de primalidad). Un número n es primo si sus únicos divisores son 1 y él mismo. Bastará entonces que probemos, uno por uno, con todos los números entre 2 y $n-1$, si alguno de ellos es divisor de n, sabemos que n no es primo; si, por otro lado, no encontramos ningún divisor de n, concluimos que n es primo.

Algoritmo 1: Primera Prueba de Primalidad

Entrada: Un numero natural n
Salida: Devuelve PRIMO si n es primo, COMPUESTO en caso contrario
Procedimiento `EsPrimo(`n`)`

```
para i entre 2 y −1 hacer
    si i divide a n entonces
        devolver COMPUESTO ; // Encontramos un divisor de n
    fin
fin
devolver PRIMO // No encontramos un divisor de n, concluimos que n
    es primo
```

Esta es nuestra primera prueba de primalidad. Notemos que este algoritmo determina si el número n es compuesto encontrado un número i, entre 2 y $n-1$, que divida a n. De existir, dicho número i funge como, llamémosle así, testigo de la no-primalidad de n.

Es importante resaltar que en el peor de los casos, cuando n es primo, se hará la prueba de divisibilidad con cada uno de los números entre 2 y $n-1$ sin encontrar divisores. Esta primera prueba de primalidad tiene entonces una complejidad computacional de $O(n)$.

¿Es posible diseñar una prueba de primalidad más eficiente?

Es fácil ver que si $ab = c$, se debe cumplir que $a \leq \sqrt{c}$ o $b \leq \sqrt{c}$ (de lo contrario, si ambos divisores fueran mayores a $\sqrt{c}$, el producto ab sería mayor a c). Esta afirmación nos permite concluir que si n es un número compuesto, entonces n tiene un divisor menor o igual a $\sqrt{n}$ que es distinto de 1 y de n. Con esto podemos reducir la complejidad de nuestra prueba de primalidad a $O(\sqrt{n})$.

Algoritmo 2: Segunda Prueba de Primalidad

Entrada: Un numero natural n
Salida: Devuelve PRIMO si n es primo, COMPUESTO en caso contrario
Procedimiento `EsPrimo(`n`)`

```
para i entre 2 y √n hacer
    si i divide a n entonces
        devolver COMPUESTO ; // Encontramos un divisor de n
    fin
fin
devolver PRIMO // No encontramos un divisor de n, concluimos que n
    es primo
```

El procedimiento a seguir es el mismo, simplemente hemos notado que basta con buscar divisores menores o iguales a $\sqrt{n}$. Aún con este algoritmo mejorado, nuestra prueba de primalidad es poco eficiente y no es viable para números tan grandes como los usados por RSA. Para diseñar una mejor prueba de primalidad requerimos de más conceptos de Teoría de Números.

Conceptos de Teoría de Números

Nuestro principal objetivo para diseñar una mejor prueba de primalidad será determinar una manerá más eficiente de econtrar testigos de la no primalidad de un número. Para ello permítamonos estudiar brevemente algunos conceptos de Teoría de Números y Teoría de Anillos. Comencemos por el concepto de congruencia.

Definición 1. *Sean $a, b, m \in \mathbb{Z}$, con $m \neq 0$. Decimos que a y b son congruentes módulo m si m divide a $b - a$. Esta relación se expresa $a \equiv b \pmod{m}$.*

La relación de congruencia módulo m define una relación de equivalencia: en efecto es simétrica, reflexiva y transitiva. De este modo podemos definir $\overline{a}$, la clase de equivalencia de a módulo m, como $\overline{a} = \{x \in \mathbb{Z} \mid a \equiv x \pmod{m}\}$. A este conjunto también se le conoce como clase de congruencia. Por ejemplo, si $m = 2$, tenemos dos clases de congruencia: $\overline{0}$, el conjunto de todos los números pares, y $\overline{1}$, el conjunto de todos los números impares. En general hay m clases de congruencia modulo m, a saber: $\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{m-2}, \overline{m-1}$.

Definición 2. *El conjunto de todas las clases de coungrencia módulo m se denota por $\mathbb{Z}/m\mathbb{Z}$.*

$\mathbb{Z}/m\mathbb{Z}$ define un anillo con la suma y producto definidos como sigue: para $\overline{a}, \overline{b} \in \mathbb{Z}/m\mathbb{Z}$, $\overline{a} + \overline{b} = \overline{a+b}$ y $\overline{a}\overline{b} = \overline{ab}$. Esta definición parece indicar que el valor de la suma y producto depende de los valores númericos de a y b y no de sus clases de coungruencia. La siguiente proposición es suficiente para ver que el valor de la suma y producto así definidos depende solo de las clases de congruencia de a y b.

Proposición. Sean $a, b, c, d \in \mathbb{Z}$. Si $a \equiv c \pmod{m}$ y $b \equiv d \pmod{m}$, entonces $a + b \equiv c + d \pmod{m}$ y $ab \equiv cd \pmod{m}$.

Demostración. Una importante propiedad de divisibilidad es que si $n \mid x$ y $n \mid y$, entonces $n \mid x + y$.

Dado que $m \mid a - c$ y $m \mid b - d$, concluimos que $m \mid (a - c) + (b - d) = (a + b) - (c + d)$; por lo tanto $a + b \equiv c + d \pmod{m}$. Por otro lado $ab - cd = ab - cb + cb - cd = b(a - c) + c(b - d)$. m divide a ambos terminos de esta suma, por lo que $m \mid ab - cd$ y por lo tanto $ab \equiv cd \pmod{m}$.

Además, $\overline{0}$ funge como neutro aditivo, mientras que $\overline{1}$, a su vez, juega el papel de neutro multiplicativo. Todo elemento $\overline{a} \in \mathbb{Z}/m\mathbb{Z}$ tiene inverso aditivo: $\overline{m - a}$. Así pues, a $\mathbb{Z}/m\mathbb{Z}$ también se le denomina como el anillo de los enteros módulo m.

Recordemos que las unidades de un anillo son aquellos elementos del mismo que tienen inverso multiplicativo. Para el caso particular de $\mathbb{Z}/m\mathbb{Z}$, las unidades son $\overline{a} \in \mathbb{Z}/m\mathbb{Z}$ tal que existe x que cumple $\overline{ax} = \overline{1}$. Como vimos, esto es equivalente a $ax \equiv 1 \pmod{m}$. Es posible demostrar que esta ecuación tiene solución si y solo si a es coprimo con m, es decir, si a y m no tienen divisores en común además de 1. De tal forma que $\overline{a}$ es una unidad de $\mathbb{Z}/m\mathbb{Z}$ si y solo si a es coprimo con m. Estamos en condiciones para enunciar un teorema escencial para nuestro propósito de distinguir números primos de números compuestos.

Definición 3. *Sea $m \in \mathbb{Z}^+$, definimos la función $\phi(m)$ como el número de enteros entre 1 y $m - 1$ que son coprimos con m.*

Por ejemplo, $\phi(8) = 4$, pues $1, 3, 5$ y 7 son coprimos con 8. Es fácil ver que si p es primo, $\phi(p) = p-1$. Por lo que vimos arriba, $\mathbb{Z}/m\mathbb{Z}$ tiene entonces $\phi(m)$ unidades. Si p es un número primo, $\mathbb{Z}/p\mathbb{Z}$ tiene $\phi(p) = p - 1$ unidades.

Teorema 1 (Teorema de Euler)**.** *Si a y m son coprimos, entonces $a^{\phi(m)} \equiv 1 \pmod{m}$. Teorema de Euler. Si a y m son coprimos, entonces $a^{\phi(m)} \equiv 1 \pmod{m}$.*

Demostración. Por propiedades de anillos, las unidades de $\mathbb{Z}/m\mathbb{Z}$ forman un grupo de orden $\phi(m)$. Como a es coprimo con m, $\overline{a}$ es una unidad del anillo y se cumple que $\overline{a}^{\phi(m)} = \overline{1}$ o, equivalentemente, $a^{\phi(m)} \equiv 1 \pmod{m}$. □

Corolario (Pequeño Teorema de Fermat)**.** *Sean $a, p \in \mathbb{Z}$, con p primo. Si p no divide a a, entonces $a^{p-1} \equiv 1 \pmod{p}$.*

Demostración. Es consecuencia directa del Teorema anterior: como p no divide a a y p es primo, a y p son coprimos. Además, como vimos antes, $\phi(p) = p - 1$. Aplicando el Teorema de Euler obtenemos la congruencia que buscamos: $a^{p-1} \equiv 1 \pmod{m}$. □

Por último, demostremos la siguiente proposición que nos será de gran utilidad.

Proposición. *Si p es primo, entonces $x^2 \equiv 1 \pmod{p}$ si y solo si $x \equiv 1 \pmod{p}$ o $x \equiv -1 \pmod{p}$.*

Demostración. Para esta demostración requerimos de una importante propiedad de los números primos: sean $a, b \in \mathbb{Z}$, si $p \mid ab$, entonces $p \mid a$ o $p \mid b$.
Supongamos que $x^2 \equiv 1 \pmod{p}$, entonces $p \mid x^2 - 1 = (x+1)(x-1)$. Esto implica que $p \mid x+1$ o $p \mid x-1$; escribiendo esto en términos de congruencia módulo p: $x \equiv -1 \pmod{p}$ o $x \equiv 1 \pmod{p}$.
Por otro lado, si $x \equiv 1 \pmod{p}$ o $x \equiv -1 \pmod{p}$, naturalmente tenemos que $x^2 \equiv 1 \pmod{p}$; con lo que concluimos la demostración.
Resulta entonces que en $\mathbb{Z}/p\mathbb{Z}$, con p primo, lás unicas ráices cudradas de $\overline{1}$ son las triviales: $\overline{1}$ y $\overline{-1}$. □

Prueba de Primalidad Probabilística de Miller-Rabin

Permítamonos resaltar las dos propiedades demostradas en la sección anterior que nos permitirán describir una segunda y mucho más eficiente prueba de primalidad:

Teorema 2 (Pequeño Teorema de Fermat)**.** *Sean $a, p \in \mathbb{Z}$, con p primo. Si p no divide a a, entonces $a^{p-1} \equiv 1 \pmod{p}$.*
***No existen raíces cuadradas no triviales de la unidad en** $\mathbb{Z}/p\mathbb{Z}$**.** Si p es primo, entonces $x^2 \equiv 1 \pmod{p}$ si y solo sí $x \equiv 1 \pmod{p}$ o $x \equiv -1 \pmod{p}$.*

Es valioso resaltar que si tomamos un número compuesto, estas propiedades no necesariamente se cumplen. Por ejemplo, tomemos 8, que a todas luces es un número compuesto. Resulta que 5 es una raíz cuadrada no trivial de la únidad: $5^2 = 25 \equiv 1 \pmod{p}$. Tampoco se cumple la ecuación del Pequeño Teorema de Fermat: $5^7 \equiv 5 \pmod{8}$.
Tenemos entonces dos nuevas maneras de buscar testigos de la no-primalidad de un número:

1. Si $n \in \mathbb{Z}^+$, $a \in \{2, 3, \ldots, n-1\}$, y $a^{n-1} \not\equiv 1 \pmod{n}$, entonces n es un número compuesto. a es testigo de la no primalidad de n pues muestra que no se cumple el Pequeño Teorema de Fermat.

2. Si $n \in \mathbb{Z}^+$, $b \in \{2, 3, \ldots, n-1\}$, y $b^2 \equiv 1 \pmod{n}$, entonces n es un número compuesto. b es testigo de la no primalidad de n pues muestra que no se cumple la propiedad de raíces no triviales en $\mathbb{Z}/n\mathbb{Z}$.

En nuestra primera prueba de primalidad buscábamos testigos de manera secuencial, de 2 a $n-1$. Para esta prueba de primalidad elijamos los potenciales testigos de manera aleatoria; más adelante veremos que esta forma de proceder es bastante eficiente.
De tal forma que, para saber si $n \in \mathbb{Z}^+$ es primo, procederemos de la siguiente forma:

1. Elegimos $a \in \{2, 3, \ldots, n-1\}$ aleatoriamente.

2. Escribimos a $n-1$ como $n-1 = 2^t u$, con t, u enteros no negativos y u impar.

3. Calcularemos la sucesión $x_0, x_1, \ldots, x_t$. $x_0 = a^u$ y para $i \in \{1, 2, \ldots, t\}$ $x_i = x_{i-1}^2 \pmod{m}$. Notemos que $x_t \equiv a^{2^t u} \equiv a^{n-1} \pmod{n}$.

4. Si $i \in \{1, 2, \ldots, t\}$, $x_i \equiv 1 \pmod{n}$ y además $x_{i-1} \neq 1, -1$, hemos encontrado una raíz cuadrada no trivial de la unidad módulo n. Concluimos entonces que n no es primo.

5. Si $x_t = a^{n-1} \not\equiv 1 \pmod{n}$, entonces n no cumple la ecuación del Pequeño Teorema de Fermat, por lo que n es un número compuesto.

A continuación mostramos el pseudocódigo de este algoritmo.

Algoritmo 3: Testigo de No-primalidad

Entrada: Un numero natural n y $a \in \{2, 3, \ldots, n-1\}$
Salida: Devuelve SÍ si a es testigo de la no-primalidad de n, NO en caso contrario.
Procedimiento `Testigo(`n, a`)`

```
x_0 = a^u (mód n)
para i de 1 a t hacer
    x_i = x_{i-1}^2 (mód n)
    si x_i ≡ 1 (mód n) entonces
        devolver SÍ // Encontramos una raíz no trivial.
    fin
fin
si x_t ≠ 1 (mód n) entonces
    devolver SÍ // No se cumple el Pequeño Teorema de Fermat.
si no
    devolver NO // a no atestigua la no-primalidad de n.
fin
```

Está claro que el costo computacional de este algoritmo es el de obtener los elementos de la secuencia $x_0, x_1, \ldots, x_t$. Si tenemos cuidado de obtener el valor de a^u de manera eficiente (utilizando un método conocido como exponenciación binaria), la complejidad computacional de este procedimiento es $O(\log n)$.

Ahora bien, este procedimiento solo prueba con un potencial testigo. La prueba de primalidad de Miller-Rabin (propuesta por Michael O. Rabin en 1980, basada en el trabajo previo de Gary L. Miller) [4] consiste en elegir s potenciales testigos de manera aleatoria, si alguno de ellos en efecto atestigua la no-primalidad de n, el algoritmo concluye que n es compuesto; en caso contrario se concluye que n es primo.

Algoritmo 4: Prueba de Primalidad de Miller-Rabin

Entrada: Un numero natural n y $s \in \{1, 2, \ldots, n-1\}$
Salida: COMPUESTO si se encuentra un testigo de la no-primalidad de n, PRIMO en caso contrario
Procedimiento `EsPrimo(`n, s`)`
  para i de 1 a s **hacer**
    Elegir a aleatoriamente en $\{2, 3, \ldots, n-1\}$
    si `Testigo(`n, a`)` **entonces**
      devolver COMPUESTO
      `// Encontramos un testigo de la no primalidad de` n.
    fin
  fin
  devolver PRIMO`// No encontramos testigos.`

Notemos lo siguiente: cuando la prueba de primalidad devuelve COMPUESTO, tenemos total certeza de que n no es primo, pues hemos encontrado que n no cumple la ecuación del Pequeño Teorema de Fermat o tiene raíces no triviales de la unidad. Por otro lado, cuando el procedimiento devuelve PRIMO, no tenemos total certeza de que el número sea en efecto primo, sólo sabemos que entre los s valores elegidos aleatoriamente no hay testigos de la no-primalidad de n. Este algoritmo es entonces propenso a errores, pero el único tipo de error que comete es asumir erroneamente que n es primo.

Lo cierto es que es posible demostrar que incluso con valores pequeños de s, tan pequeños como 3, la Prueba de Primalidad de Miller-Rabin es muy poco propensa a errores [2]. Así pues, aunque está prueba de primalidad tienen una pequeña probabilidad de cometer errores, su complejidad computacional es, como vimos, $O(\log n)$, mucho mejor que la de nuestras primeras dos pruebas de primalidad. No entraremos en detalles sobre la probabilidad de fallo de este algoritmo, pero vale la pena reconocer que este algoritmo probabilístico, como RSA, ha pasado la prueba del tiempo y sigue siendo la prueba de probabilidad más utilizada.

Bibliografía

[1] Derbyshire, John. *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics.* Joseph Henry Press, 2003.

[2] Gauss, Carl Friedrich. *Disquisitiones Arithmeticae*. 1966.

[3] T.H. Cormen. *Introduction to Algorithms*. MIT Press, 2009.

[4] Ireland, Kenneth F., and Michael I. Rosen. *A Classical Introduction to Modern Number Theory*. Springer, 2011.

[5] Rabin, Michael O. "Probabilistic Algorithm for Testing Primality." *Journal of Number Theory* 12, no. 1 (1980): 128-38.

[6] M. D. Kelly. *The RSA Algorithm: A Mathematical History of the Ubiquitous Cryptological Algorithm*. 2009.

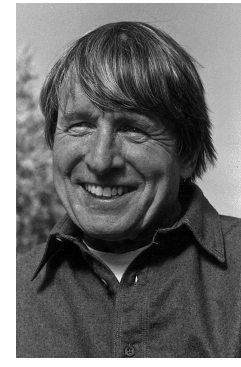
Algoritmos genéticos

Rebeca Angulo Rojas
Estudiante de Matemáticas Aplicadas

En términos muy generales se puede definir la **computación evolutiva** como una familia de modelos computacionales inspirados en la evolución. Formalmente, el término de computación evolutiva se refiere al estudio de los fundamentos y aplicaciones de ciertas técnicas heurísticas inspiradas en los principios de la evolución natural. Estas técnicas heurísticas pueden clasificarse en tres grandes categorías:

- Algoritmos genéticos
- Estrategias de evolución
- Programación evolutiva

Algoritmos genéticos.



A mediados de la década de los años sesenta, el Dr. John Henry Holland, mejor conocido como el padre del algoritmo genético, desarrolló una técnica de programación que se adaptaba muy bien a la evolución. El propósito original de Holland era estudiar de un modo formal, el fenómeno de la adaptación tal y como ocurre en la naturaleza, y desarrollar vías para extrapolar esos mecanismos de adaptación natural a los sistemas computacionales (H. Holland, *Genetic Algorithms*).

En 1975 con apoyo de trabajos anteriores y en colaboración con otros investigadores y alumnos, logró presentar en el libro *Adaptación en Sistemas Naturales y Artificiales* el **algoritmo genético** como una abstracción de la evolución biológica, con esto Holland fue el primero en intentar colocar la computación evolutiva sobre una base teórica firme, utilizando la mutación, la selección y el cruzamiento, simulando el proceso de la evolución biológica como estrategia para resolver problemas. John Holland trabajo por más de 40 años en la investigación de la la teoría y práctica de los algoritmos genéticos. En la década de los años ochenta los algoritmos genéticos comenzaban a aplicarse en una amplia variedad de áreas, desde problemas matemáticos abstractos como el problema de la mochila *bin-packing*, la coloración de grafos, hasta problemas de control de flujo, reconocimiento y clasificación de patrones y optimización estructural (Sancho Caparrini, *Algoritmos Genéticos*).

Los algoritmos genéticos se convirtieron solucionadores de problemas en distintas áreas, tal como la optimización, tanto en espacios discretos como en espacios continuos. En un breve

resumen, los algoritmos genéticos son algoritmos de optimización, búsqueda y aprendizaje inspirados en los procesos de evolución natural y evolución genética.

La evolución es un proceso que opera sobre los cromosomas. Estos cromosomas pueden ser considerados como herramientas orgánicas que codifican la vida. Toda la información contenida en los cromosomas se conoce como genotipo, sin embargo dicha información puede o no manifestarse en el individuo. El fenotipo es el resultado de la expresión del genotipo conjuntamente con la influencia del medio y factores epigenéticos (caracteres observables).(cita)

La **selección natural**, expuesta en la teoría de la evolución biológica por Charles Darwin (1859), es un mecanismo que relaciona los cromosomas (genotipo) con el fenotipo y otorga a los individuos más adaptados un mayor número de oportunidades de reproducirse. Los procesos evolutivos tienen lugar durante la etapa de reproducción, aunque existe una larga serie de mecanismos que afectan a la reproducción, el más común es la mutación, causante de que los cromosomas en la descendencia sean diferentes a los de los padres y el cruce o recombinación, que combina los cromosomas de los padres para producir la descendencia («Selección natural»). Rigurosamente podemos definir la selección natural como el proceso que se da en una población de entidades biológicas cuando se cumplen las tres condiciones siguientes:

1. **Variación fenotípica entre los individuos de una población:** los distintos individuos de una población difieren en sus caracteres observables.
2. **Eficacia biológica diferencial asociada a la variación:** ciertos fenotipos o variantes están asociados a una mayor descendencia y/o una mayor supervivencia.
3. **La herencia de la variación:** permite la transmisión de los fenotipos seleccionados a la siguiente generación.

Si en una población de organismos se dan estas tres condiciones, entonces se sigue necesariamente un cambio en la composición genética de la población por selección natural. La selección es, por lo tanto, el proceso que resulta de las tres premisas citadas. Y esto es lógicamente cierto tanto en éste como en cualquier otro mundo imaginable.

En un algoritmo genético para alcanzar la solución a un problema se parte de un conjunto inicial de individuos, llamado población, el cual es generado de manera aleatoria. Cada uno de estos individuos representa una posible solución al problema. En el campo de los algoritmos genéticos, se construye una función objetivo mejor conocida como *fitness function* y se definen los paisajes del potencial (*fitness landscapes* o *adaptive landscapes*), los cuales son evaluaciones de la función objetivo para todos las soluciones candidatas. Para evaluar las soluciones en un algoritmo genético se utiliza la **función de evaluación**, la cual establece una medida numérica, mejor conocida como bondad de ajuste de una solución.

Esta medición permite controlar el número de selecciones, cruces y copias. En la naturaleza este ajuste puede entenderse como la probabilidad de que un individuo sobreviva hasta la edad de reproducción y se reproduzca.

Representación

La codificación utilizada en los algoritmos genéticos es una abstracción del genotipo de los individuos. Existen varias formas de representar a los cromosomas, la más sencilla es la binaria, aunque no siempre resulta ser la más natural, en estos casos se utilizan valores reales. (Gestal et al., Introducción a los Algoritmos genéticos y la Programación Genética).

El genotipo de un individuo en un algoritmo genético que emplea cadenas de bits es sencillamente, la configuración de bits del cromosoma de ese individuo. Aquí la noción de fenotipo no aparece en el contexto de los algoritmos genéticos, aunque en avances recientes, hay algoritmos que poseen un nivel "genotípico" y uno "fenotípico", por ejemplo, la cadena de bits que codifica una red neuronal. En la Figura 1 se muestra un ejemplo de un individuo binario que codifica 3 parámetros.

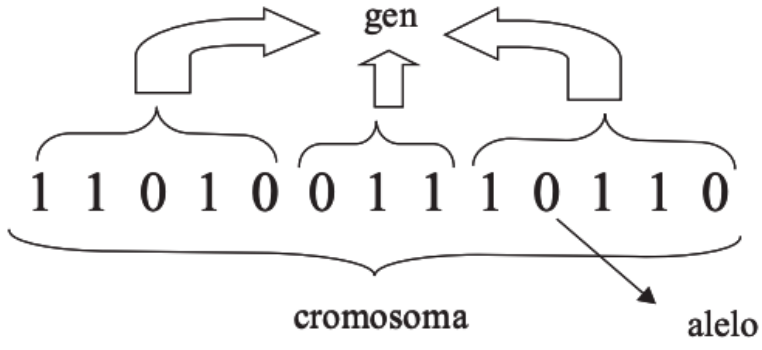


Figura 1: Cromosoma representado como cadena de bits. Tomada de *Introducción a los Algoritmos genéticos y la Programación Genética*. Cada uno de los bits pertenecientes a un gen recibe el nombre de alelo.

Operadores Genéticos

Una generación se obtiene a partir de la anterior por medio de operadores, que son mejor conocidos como **operadores genéticos**. Los más empleados son los operadores de selección, cruce, copia y mutación (Gestal et al., Introducción a los Algoritmos genéticos y la Programación Genética).
El pseudocódigo de un algoritmo genético se puede ver de la siguiente forma (Sancho Caparrini, *Algoritmos Genéticos*).

Crea población inicial

Evalúa los cromosomas de la población inicial
while no se cumpla el criterio de terminación **do**
    Selección de los cromosomas más aptos en la nueva población
    Cruzamiento de los cromosomas
    Mutación de los cromosomas
    Evaluación de los cromosomas
end while
Devuelve la mejor solución (la más apta) en la población

En cuanto a los criterios de terminación o parada, los más utilizados son (Sancho Caparrini, *Algoritmos Genéticos*):

- Los mejores individuos de la población representan soluciones suficientemente buenas para el problema que se desea resolver.
- La población ha convergido. Un gen ha convergido cuando el noventa y cinco porciento de la población tiene el mismo valor (caso de codificaciones binarias) o valores dentro de un rango especificado (otro tipo de codificaciones).
- Se ha alcanzado el número de generaciones máximo especificado.

Selección

La selección es el mecanismo por el cual son seleccionados los individuos que serán los padres de la siguiente generación. Similarmente a lo que ocurre en la selección natural, se otorga un mayor número de oportunidades de reproducción a los individuos más aptos. Es importante destacar que no se deben eliminar por completo las opciones de reproducción de los individuos menos aptos, pues en pocas generaciones la población podría volverse homogénea. (Gestal et al., *Introducción a los Algoritmos genéticos y la Programación Genética*).

Existen diversas formas de realizar una selección:

- **Selección por truncamiento:** elegimos a los padres entre los mejores k cromosomas de la población.
- **Selección de torneos:** se eligen subgrupos de individuos de la población y los integrantes de cada subgrupo compiten entre ellos. Sólo se elige a un individuo de cada subgrupo para la reproducción.
- **Selección por ruleta:** cada padre se elige con una probabilidad proporcional a su desempeño en relación con la población.
- **Selección por jerárquica:** los individuos atraviesan diversas rondas de selección en cada generación. Las evaluaciones de los primeros niveles son más sencillas, mientras que aquellos que sobreviven hasta niveles más altos son evaluados más rigurosamente.

Los algoritmos de selección pueden ser divididos en dos grupos: probabilísticos, en este grupo se encuentran los algoritmos de selección por ruleta, y determinísticos, en este grupo se encuentran los algoritmos de selección por jerarquías, muestreo determinístico, estado uniforme, sobrante estocástico, entre otros.

Una vez que se han elegido a los individuos más aptos por medio de la selección, estos deben ser cruzados y mutados para formar una nueva generación de individuos.

Cruce

El cruce consiste, como en el caso biológico, en un intercambio de material genético entre dos cromosomas de dos padres y a partir de esto se genera una descendencia. Al igual que en la selección existen diversas formas de hacer un cruce, entre ellos se encuentran:

- Cruce de un solo punto
- Cruce de dos puntos
- Cruce uniforme

Figura 2: cruce de un solo punto. Tomada de *Algorithms for Optimization*

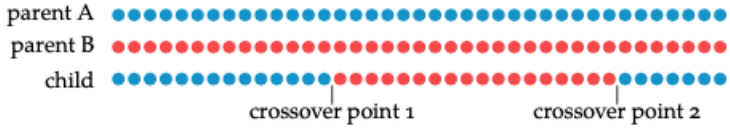


Figura 3:cruce de dos puntos. Tomada de *Algorithms for Optimization*

Figura 4: cruce uniforme. Tomada de *Algorithms for Optimization*

La idea principal del cruce se basa en que si se toman dos individuos correctamente adaptados al medio y se obtiene una descendencia que comparta genes de ambos, al compartir las características buenas de dos individuos, la descendencia, o al menos parte de ella, debería tener una mayor bondad que cada uno de los padres por separado. Sin embargo, puede

suceder que el cruce no agrupe las mejores características en uno de los hijos y por tanto la descendencia tendrá un peor ajuste que los padres. Por lo que optando por una estrategia de cruce no destructiva garantizamos que pasen a la siguiente generación los mejores individuos. Si aún en esta circunstancia se opta por insertar a la descendencia dado que los genes de los padres continuarán en la población, dispersos y posiblemente modificados por la mutación, en posteriores cruces se puede volver a obtener a los padres, recuperando así la bondad previamente perdida. (Salazar Larico, *Algoritmos Genéticos*).

Mutación

Análogo a la mutación biológica, una mutación en un algoritmo genético también causa pequeñas alteraciones en puntos determinados de la codificación del individuo. Este operador produce variaciones de modo aleatorio en un cromosoma. Cada bit tiene una pequeña probabilidad de ser invertido. Para un cromosoma con m bits, esta tasa de mutación generalmente se establece como $1/m$. En el caso de tratar con codificaciones binarias, la mutación consiste simplemente en negar un bit.

La mutación suele realizarse después del cruce. Por lo general primero se seleccionan dos individuos de la población para realizar el cruce y si el cruce tiene éxito entonces uno de los descendientes, o ambos, se muta con cierta probabilidad. En la siguiente figura podemos observar un ejemplo de mutación y cruce.

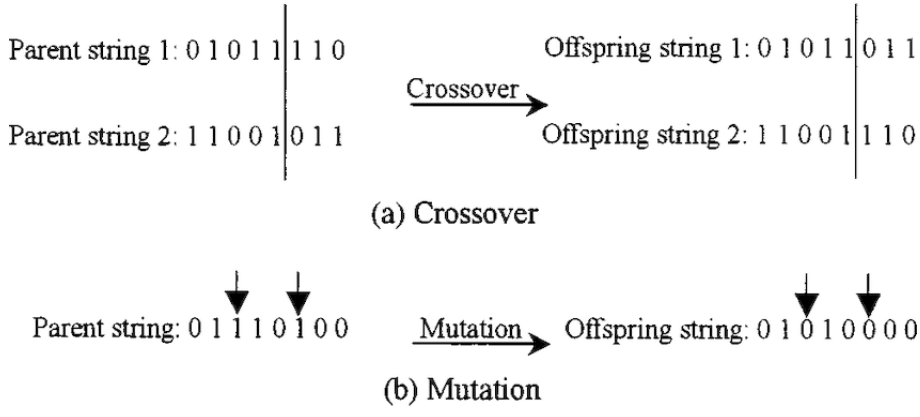


Figura 5: Crossover and mutation operations in genetic algorithm Nota. Tomado de Two-Stepped Evolutionary Algorithm and Its Application to Stability Analysis of Slopes

Reemplazo

Cuando trabajamos con una población que no es temporal, y sobre la que se realizan selecciones e inserciones, deberá tenerse en cuenta que para insertar un nuevo individuo deberá de eliminarse previamente otro de la población. A esta operación se le conoce como reemplazo.

(Gestal et al., *Introducción a los Algoritmos genéticos y la Programación Genética*).

Existen diferentes métodos de reemplazo:

- **Aleatorio:** el nuevo individuo se inserta en un lugar escogido de manera aleatoria en la población.
- **Reemplazo de padres:** se obtiene espacio para la nueva descendencia liberando el espacio ocupado por los padres.
- **Reemplazo de similares:** una vez obtenido el ajuste de la descendencia se selecciona un grupo de individuos de la población con un ajuste similar. Y se reemplazan aleatoriamente los que sean necesarios.
- **Reemplazo de los peores:** entre un porcentaje de los peores individuos de la población se seleccionan aleatoriamente los necesarios para dejar sitio a la descendencia.

Copia

Se trata de un operador de tipo asexual, pues consiste simplemente en la copia de un individuo en la nueva generación. La copia, es otra estrategia reproductiva para la obtención de una nueva generación a partir de la anterior. Un determinado número de individuos pasa, sin sufrir ninguna variación, directamente a la siguiente generación (Gestal et al., *Introducción a los Algoritmos genéticos y la Programación Genética*).

Elitismo

El elitismo es un caso particular del operador de copia, consiste en copiar siempre al mejor o mejores individuos. (Gestal et al., *Introducción a los Algoritmos genéticos y la Programación Genética*).

Ventajas y restricciones

Los algoritmos presentan algunas ventajas, al igual que ciertas restricciones en relación a otros algoritmos tradicionales de optimización. La principal es que a diferencia de otros algoritmos de optimización, los algoritmos genéticos buscan una población de puntos, no un único punto y además los algoritmos genéticos emplean la función objetivo, es decir, no necesitan derivadas ni otros objetos complementarios, lo cual muchas veces resulta costoso y difícil de calcular. Por otro lado, uno de los mayores obstáculos en los algoritmos genéticos suele ser la elección de la función objetivo, de ella depende que se alcancen soluciones más aptas en el problema. Otros de los parámetros que deben elegirse con mucho cuidado son el tamaño de la población, la tasa de mutación y cruce, dado que una mala elección de estas podría traer problemas en la convergencia del algoritmo.

Uno los problemas que suelen suceder en problemas con poblaciones pequeñas se conoce como convergencia prematura, esta ocurre cuando un individuo es más apto que el resto, entonces este ha de reproducir muchos más individuos que otros, esto reduce la diversidad en la población y provoca que el algoritmo converja a un óptimo local que representa este individuo en lugar de buscar el paisaje del potencial (*fitness landscape*) para poder encontrar la solución global óptima (*Genetic Algorithms: The Reality*).

Aplicaciones

Algunas de las aplicaciones de los algoritmos genéticos se señalan a continuación:

- Optimización
- Aprendizaje de máquina
- Economía
- Ecología: modelización de fenómenos ecológicos.
- Medicina
- Criptoanálisis
- Robótica: actualmente los algoritmos genéticos son utilizados para crear robots cuyo comportamiento imita a un humano en diversas tareas. Los algoritmos genéticos son técnicas de búsqueda adaptativas que se utilizan para aprender estructuras de conocimiento de alto rendimiento.
- Procesamiento de imagen

Durante muchos años, a través de la observación de la evolución y adaptación de distintas especies naturales, se logró adaptar esos sistemas naturales en sistemas artificiales, a esto se le conoce como biónica. Actualmente hay un gran desarrollo en torno a la computación evolutiva, algoritmos genéticos y programación evolutiva y que, como mencionamos en un principio, han logrado aprovechar el proceso de evolución para la resolución de distintos problemas.

Referencias

[1] Ghaheri, Ali, Saeed Shoar, Mohammad Naderan, y Sayed Shahabuddin Hoseini. "The Applications of Genetic Algorithms in Medicine", *Oman Medical Journal 30* nº 6 (2015).

[2] "Introduction to Genetic Algorithms — Including Example Code. Medium" 15 de marzo de 2021. `https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3`

[3] "Algorithms for Optimization — The MIT Press." *The MIT Press*. 17 de marzo de 2021. https://mitpress.mit.edu/books/algorithms-optimization

[4] "Biónica." *Wikipedia*. 17 de marzo de 2021. https://es.wikipedia.org/w/index.php?title=Binica&oldid=131648658

[5] "Selección natural" *Wikipedia*. 15 de marzo de 2021. https://es.wikipedia.org/w/index.php?title=Seleccin_natural&oldid=133998532

[6] "Genetic Algorithms: The Reality." *Wikipedia/Jstor (opcional)*. 3 de abril de 2021 https://www.doc.ic.ac.uk/project/examples/2005/163/g0516312/Algorithms/Reality.html

[7] "Genome.gov.Haploide — NHGRI" 3 de abril de 2021 https://www.genome.gov/es/genetics-glossary/Haploide

[8] "Why Genetic Algorithms Is Popular than Other Heuristic Algorithms?" *ResearchGate*. 3 de abril de 2021. https://www.researchgate.net/post/Why-genetic-algorithms-is-popular-than-other-heuristic-algorithms

[9] Cheng-Xiang, Yang, y Feng Xia-Ting. "Two-Stepped Evolutionary Algorithm and Its Application to Stability Analysis of Slopes", *Journal of Computing in Civil Engineering* 18 (2016)

[10] "John Henry Holland: Rogue Scientist" *Michigan Engineering*. 3 de abril de 2021 https://news.engin.umich.edu/2016/11/john-holland/

[11] "Introduction to Optimization with Genetic Algorithm." 3 de abril de 2021. https://towardsdatascience.com/introduction-to-optimization-with-genetic-algorithm-2f5001d9964b

[12] "Algoritmos Genéticos." 3 de abril de 2021 http://www.revistasbolivianas.org.bo/pdf/rits/n1/n1a07.pdf

[13] "Algoritmos Genéticos." 2 de abril de 2021. http://www.cs.us.es/~fsancho/?e=65

[14] H. Holland, John. "Genetic Algorithms", *Scientific American* número (1992)

[15] "A comparison of genetic programming and genetic algorithms for auto-tuning mobile robot motion control." 2 de abril de 2021. https://doi.org/10.1109/DELTA.2002.994686

[16] Gestal, Marcos, Daniel Rivero, Juan Ramon Rabuñal, Julian Dorado, y Alejandro Pazos. *Introducción a los Algoritmos genéticos y la Programación Genética*.

[17] J. Kochenderfeer, Mykel, y Tim A. Wheeler. *Algorithms for Optimization*. Cambridge, Massachusetts, 2019.

[18] Khalid Jebari. *Selection Methods for Genetic Algorithms*. Abdelmalek Essaâdi University, 2013.

Simulación de algunos teoremas de probabilidad y estadística

Samantha Arzate González
Ex-alumna de Matemáticas Aplicadas y Actuaría
Mariana Martínez Aguilar
Ex-alumna de Matemáticas Aplicadas
Jorge Méndez García
Alumno de Actuaría
Tonantzin Real Rojas
Alumna de Matemáticas Aplicadas

Introducción

Debido a que nuestro cerebro no fue diseñado para manejar grandes cantidades de números o información al mismo tiempo, como sí puede hacerlo una computadora hoy en día, resulta relevante el papel que juega la simulación a la hora de ilustrar o ejemplificar conceptos más complejos. Este trabajo pretende facilitar la comprensión de algunos resultados estadísticos haciendo uso de técnicas de simulación.

Datos

Todas las simulaciones del **Teorema del Límite Central** y de **intervalos de confianza** fueron realizadas con las funciones de R que generan valores de las distribuciones presentadas en este trabajo. Todos estos valores fueron generados con la semilla 123. La semilla usada para el cálculo de las integrales de las $f_{i,n}$ fue 18. Usamos como n la secuencia del 10 al 1000 con brincos de tamaño 10. Por último, en el problema captura y recaptura, que se resuelve con un enfoque bayesiano, se utiliza la semilla 165696 para generar las muestras aleatorias de tamaño 100,000 de la distribución a priori.

Métodos

Por un lado, para las simulaciones del **Teorema del Límite Central** se programó una función que generó N muestras de tamaño n de las distribuciones uniforme, exponencial, Poisson y binomial y otra función que plasmó en una sola imagen las densidades y la media de todas las muestras para que se facilitara la presentación de resultados.
Además, para analizar los supuestos del TLC, se construyeron, a través de R, tres escenarios diferentes en los que se violara el supuesto a revisar y se respetaran los otros dos. Posteriormente, se intentó forzar una aproximación normal sobre las medias muéstrales para cada caso y se comparó con los datos reales obtenidos.
Por otro lado, para las simulaciones de los **intervalos de confianza**, se programó una función que contó los intervalos que contenían al parámetro poblacional y otra función que graficó

esta situación. Para la cobertura de los intervalos de confianza, se programó otra función en la que, para cada valor de $\theta \in [0,1]$, se mostrara el porcentaje de cobertura del intervalo de confianza correspondiente.
En el caso del **Teorema de la convergencia dominada de Lebesgue (TCDL)** primero se dan unas definiciones previas necesarias para entender el teorema. Luego se dan 3 ejemplos de dicho teorema. Para poder calcular las integrales de f_i y de $f_{i,n}$ (para las n mencionadas en la sección anterior), se usó el Método de Monte Carlo Crudo. Para cada n se calculó el intervalo de confianza de la forma $\left(2\hat{\theta} - \hat{\theta}_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}_{\alpha/2}\right)$ con $\alpha = 0.1$. El $\hat{\theta}_{1-\alpha/2}$ para cada n se calculó con los cuantiles de la muestra de las integrales de $f_{i,n}$ de 1 a n para todas las n consideradas. Esto permite visualizar que no solo la integral converge puntualmente, sino que los intervalos de confianza sí incluyen la integral límite. En un ejemplo se muestra que el TCDL no se cumple si no se cumplen todos los supuestos.
Por último, en la sección dedicada al **teorema de Bayes**, se comienza con el enunciamiento del teorema. Procedemos a explicar la relevancia de este resultado para el desarrollo de la inferencia bayesiana y presentamos los pasos a seguir para aplicar esta metodología de análisis de datos. Después, presentamos un ejemplo teórico de la aplicación del teorema y su importancia para actualizar la probabilidad de que algo es verdad dadas las evidencias. Al final, como ejemplo práctico, presentamos el método de Computación Bayesiana Aproximada que es un enfoque que estima la función de verosimilitud basado en simulaciones.

Resultados

Teorema del Límite Central

Uno de los resultados más relevantes y más usados de la teoría de Probabilidad y Estadística es el **Teorema del Límite Central**, este resultado nos ha facilitado conocer la distribución aproximada de la media de una variable aleatoria cuando $n \to \infty$; aplicable, por ejemplo, en procesos estocásticos con los Procesos de Poisson Compuestos o como motivación para las simulaciones de Monte Carlo.

Teorema 1 (Teorema del Límite Central (TLC))**.** *Sea X_1,X_2, ...,X_n un conjunto de variables aleatorias, independientes e idénticamente distribuidas con media μ y varianza $0 < \sigma^2 < \infty$. Sea*

$$S_n = X_1 + ... + X_n$$

Entonces,

$$\lim_{n\to\infty} \Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z).$$

Es decir, si n es suficientemente grande, la variable aleatoria

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

tiene aproximadamente una distribución normal con $\mu_{\bar{X}} = \mu$ y $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$.

Para comprender mejor el teorema, a continuación presentamos la simulación de cuatro distribuciones (dos continuas y dos discretas), cada una con diferentes tamaños de muestra n y diferente número de repeticiones por simulación.

Variables aleatorias continuas

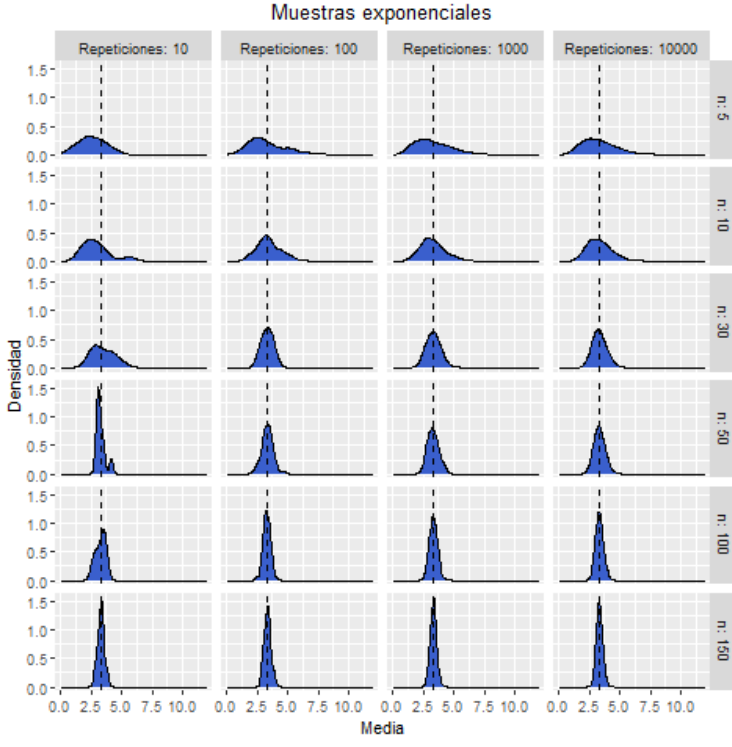


Figura 1: Variables aleatorias exponenciales

Notemos que a pesar de que las densidades provienen de dos diferentes variables aleatorias continuas, conforme la n aumenta, la densidad se asemeja a la de una campana gaussiana. Tanto en la Figura 1 como en la Figura 2 podemos notar que la mejor aproximación a una distribución normal sucede cuando las repeticiones son 10000 y a partir de un tamaño de muestra $n = 30$ pues las densidades siguen más claramente una normal. De igual forma, podemos notar que entre menos repeticiones y menor tamaño de muestra, las densidades no se parecen a una normal. Por último podemos ver que la media en la mayor parte de las gráficas corta a la mitad a las densidades; es decir, hay simetría.

Variables aleatorias discretas

Lo que se puede decir con estas distribuciones es muy similar a lo que sucede con las dis-

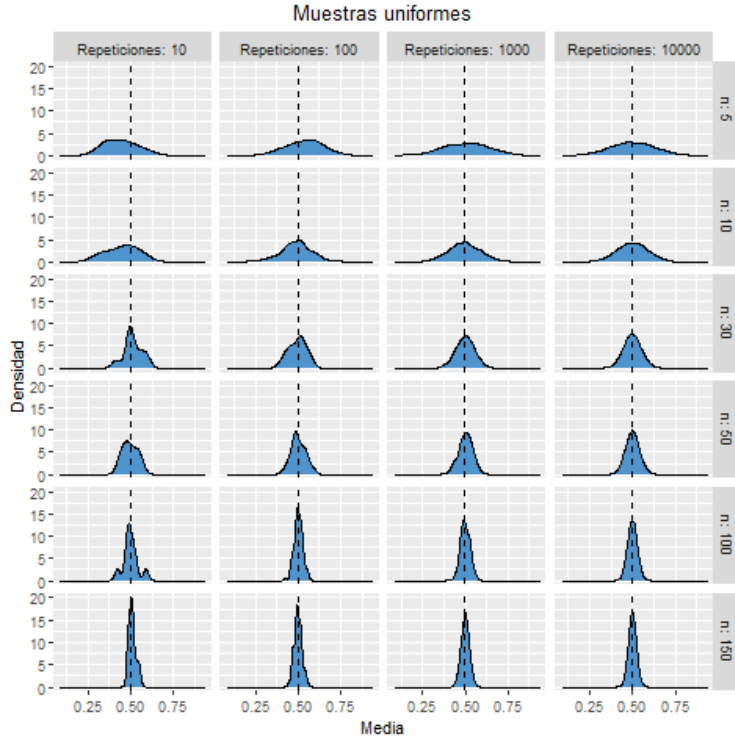


Figura 2: Variables aleatorias Uniformes (0,1)

tribuciones continuas: a mayores repeticiones y tamaño de muestra, mejor aproximación a una densidad normal. En la Figura 4 llama la atención la densidad que se forma con las repeticiones 1000 y 10000 y $n = 5$ y $n = 10$; sin embargo, en ambos casos casos, conforme el tamaño de la muestra crece, la densidad deja de tener varias ondas y, como bien menciona el TLC, se asemeja a una distribución normal. Situación que desde un inicio es más clara para la Figura 3.

¿Qué pasaría si la varianza no fuera finita?

Como vimos anteriormente, el Teorema del Límite Central (en su versión original) requiere que los datos a aproximar con una distribución normal provengan de una distribución con varianza finita. La distribución Cauchy, la cual es un caso especial de la distribución t, es un ejemplo de distribución con varianza no finita, de hecho, la distribución Cauchy ni siquiera tiene media finita. A continuación podemos ver la gráfica de la densidad Cauchy (naranja) junto a la gráfica de la densidad de una distribución Normal estándar (verde) para ayudarnos a compararlas. Notemos que la distribución Cauchy tiene un pico más corto y estrecho que la distribución normal, pero tiene colas más anchas.

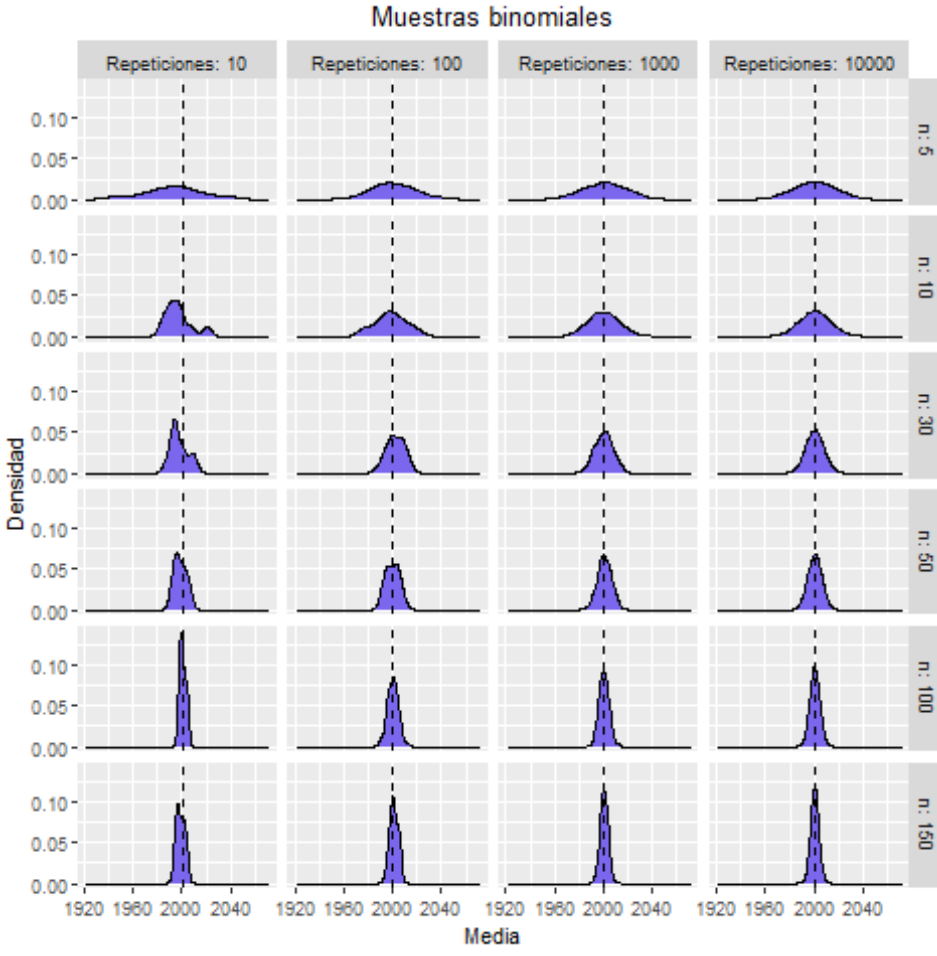


Figura 3: Variables aleatorias Binomial

Para ilustrar el problema que existiría al tratar de aproximar variables Cauchy con el Teorema del Límite Central, generamos 1000 muestras de tamaño 100 a partir de una distribución Cauchy(0,1) y calculamos la media de cada muestra obtenida. Observar estas 1000 medias muestrales debería darnos una idea de la distribución de las mismas. A continuación, las estadísticas descriptivas obtenidas a partir de nuestras 1000 muestras:

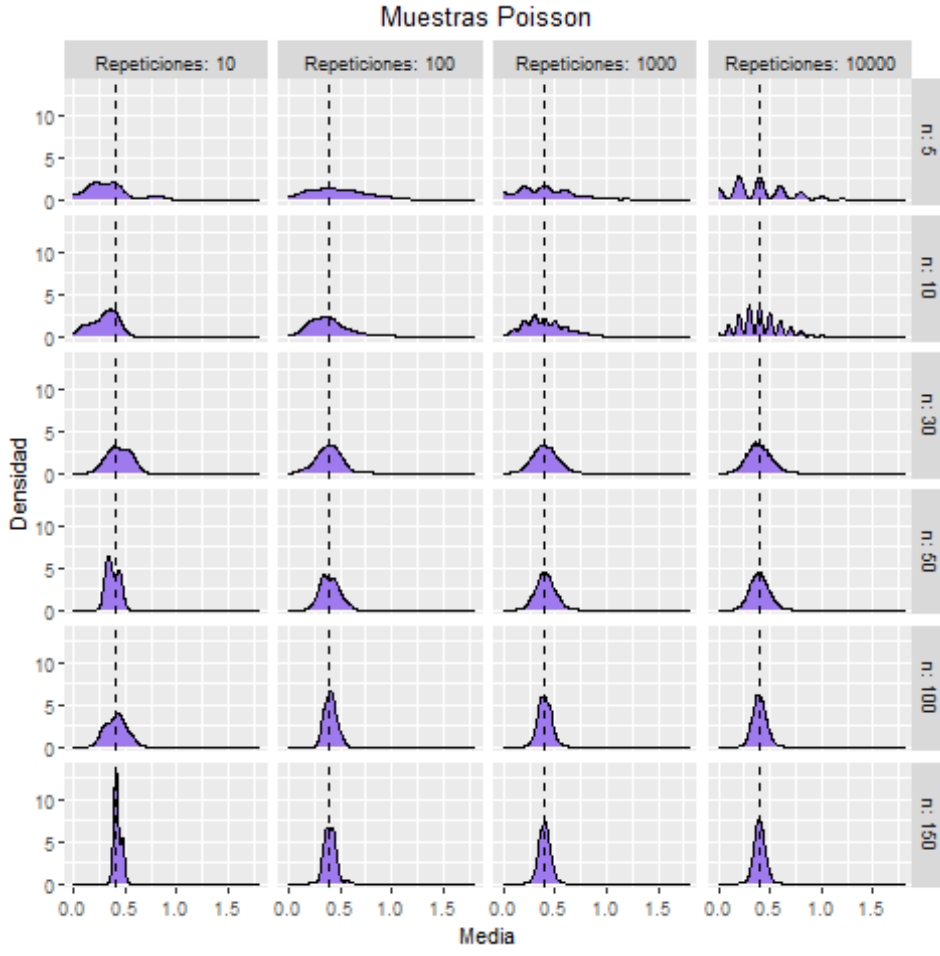


Figura 4: Variables aleatorias Poisson

Media	Desv.Est.	Mínimo	Q1	Mediana	Q3	Máximo
3.857e-03	1.502e+01	-2.948e+02	-9.416e-01	3.549e-02	9.561e-01	2.069e+02

Cuadro 1: Estadísticas descriptivas sobre las medias muestrales obtenidas, fijando la semilla en 123

Notemos que la desviación estándar es mucho mayor que el rango intercuartílico, y los valores mínimos y máximos parecen ser valores atípicos muy extremos.
A continuación, podemos ver la gráfica de caja de las medias de estas 1000 muestras. Cabe

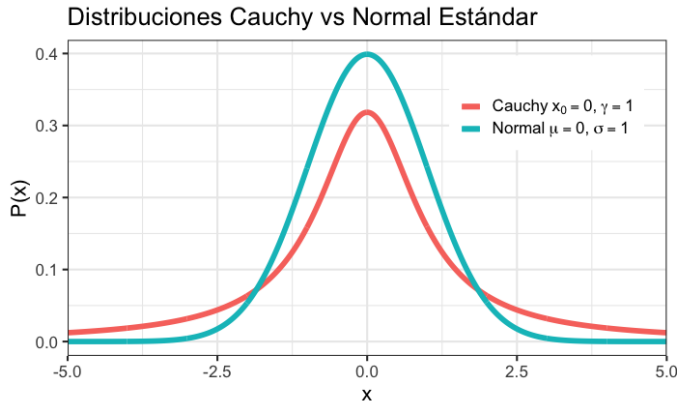


Figura 5: Aunque la distribución Cauchy es más baja que la distribución Normal, sus colas son más anchas

resaltar que es bastante inusual, con una caja muy delgada y varios valores atípicos extremos.

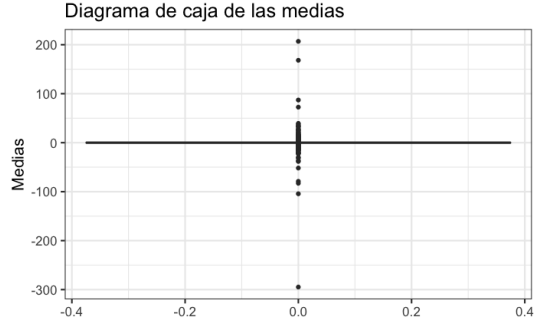


Figura 6: Diagrama de caja para las medias de las muestras obtenidas

Aquí hay un histograma de las 1000 medias muestrales que ilustra la distribución muestral de las medias para muestras de tamaño 100 de la distribución de Cauchy. Además, para facilitar la comparación, sobrepusimos una distribución normal con la misma media y desviación estándar de las 1000 medias simuladas. Así, podemos concluir que, efectivamente, ¡esta distribución muestral está lejos de ser normal!

¿Y si las variables a aproximar fueran dependientes entre sí?

Para ilustrar la falla en la aproximación de la distribución muestral a través del TLC, generaremos otras 1000 muestras tamaño 100 de una distribución Uniforme(0,1), pero simuladas de tal manera que dichas 100 variables aleatorias estén correlacionadas.

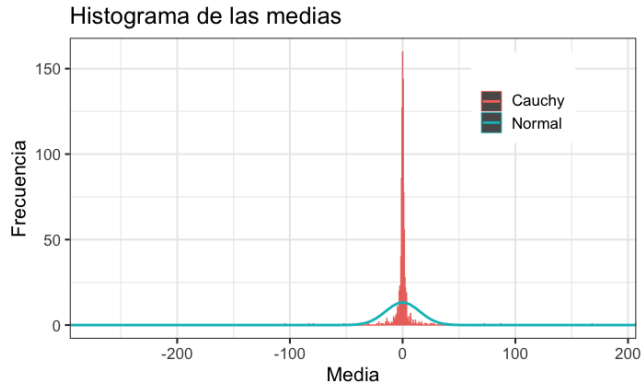


Figura 7: Hist. de las medias muestrales obtenidas vs frec. normales esperadas

Así, con la intención de introducir dependencia a las variables generadas, decidimos utilizar generadores lineales congruenciales (GLC) a la hora de simularlas. Una de las deficiencias en la calidad de los GLC es que las muestras generadas pueden mostrar cierto nivel de dependencia serial, en particular autocorrelación.

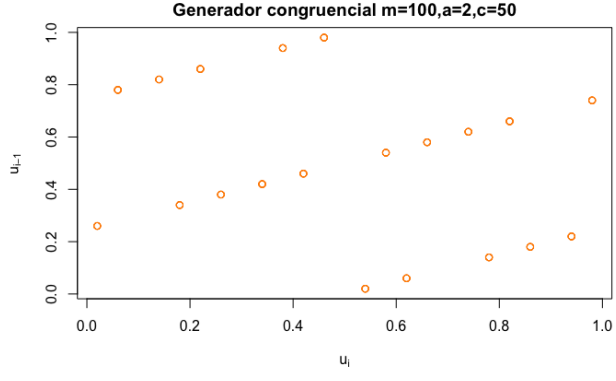


Figura 8: Ejemplo de simulación de variables aleatorias U(0,1) a través de GLM

Una vez generadas nuestras muestras, corroboramos que efectivamente hayan indicios de dependencia lineal entre las variables mediante el siguiente diagrama de correlaciones. (Notar que para facilitar su visualización, nos limitamos a presentar en este archivo el diagrama de correlación para las primeras 20 variables simuladas).
Finalmente, tras calcular las medias de las muestras obtenidas, intentamos ajustar una distribución normal a los datos obtenidos, como lo haríamos según el TLC. Sin embargo, como era de esperarse, el ajuste no fue nada bueno y podemos verlo gráficamente en la siguiente figura.

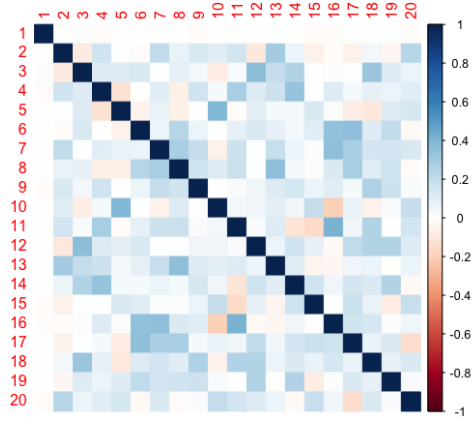


Figura 9: Las variables simuladas tienen correlaciones distintas a cero

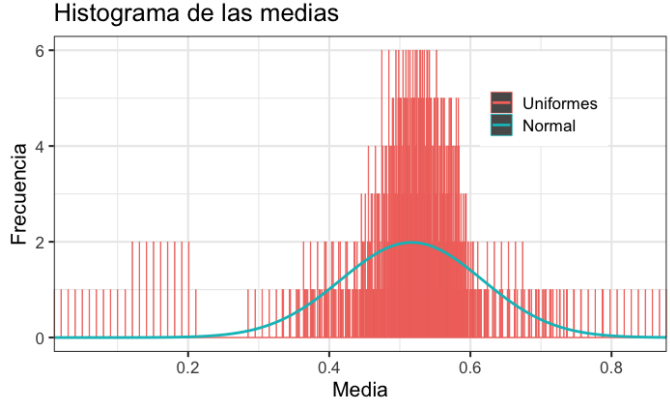


Figura 10: Comparación entre el histograma de las medias muestrales obtenidas y las frecuencias normales esperadas.

¿Y si no se distribuyeran idénticamente?

El último de los supuestos del TLC que nos queda por analizar es el supuesto de que todas las variables de la sucesión, deben ser idénticamente distribuidas. Así, con la intención de ver qué podría pasar si este supuesto no se cumple, generamos, una vez más, 1000 muestras de tamaño 100, pero esta vez a partir de distribuciones Binomiales con $n = j$ y $p = 1/j$ para cada $j = 1, ..., 100$.
Luego, tras calcular las medias de las muestras obtenidas y tratar de ajustar una distribución normal a su histograma, obtuvimos el siguiente resultado:
A pesar de que el histograma obtenido parece tener una forma acampanada, el ajuste normal

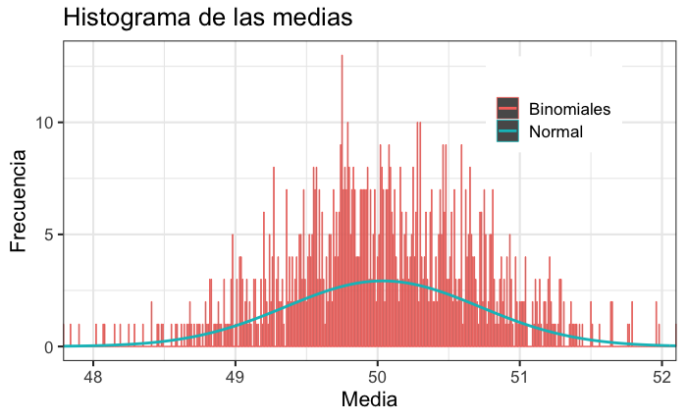


Figura 11: Comparación entre el histograma de las medias de la muestra de tamaño 1000, contra las frecuencias normales esperadas. (Semilla 123)

con la media y la varianza recomendadas por el TLC claramente no son suficientes para otorgar un buen ajuste.

Intervalos de Confianza

La idea que se tiene acerca de los **intervalos de confianza** (IC) para la media poblacional es que, como solo conocemos la media muestral, un IC para μ del $100(1-\alpha)\,\%$ significa que la media real μ estará contenida en el $100(1-\alpha)\,\%$ de los intervalos de la media de las muestras que se tomen de la población completa.
Recordemos que para una $X_1, X_2, ..., X_n$ muestra aleatoria $N(\mu, \sigma^2)$ con σ^2 conocida, el intervalo de confianza del $100(1-\alpha)\,\%$ para μ está dado por:

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

donde $z_{1-\frac{\alpha}{2}}$ es el percentil $1-\frac{\alpha}{2}$ de una $N(0,1)$

Para entender mejor a qué se refiere, visualicemos primero las siguientes figuras de variables aleatorias normales estándar:

En ambas gráficas, se muestran 200 IC (líneas negras y rojas) para las medias muestrales correspondientes (puntos negros y rojos) donde la línea horizontal que pasa por 0 se refiere a la media teórica pues las variables fueron normales estándar. Notemos que en la Figura 12, prácticamente la mitad de los intervalos presentados son rojos; es decir, dichos intervalos no contienen a $\mu = 0$. De las 10000 muestras que se generaron, el $50.18\,\%$ de los IC contuvieron a μ, lo cual es aproximadamente el mismo porcentaje de los 200 IC mostrados en la esta

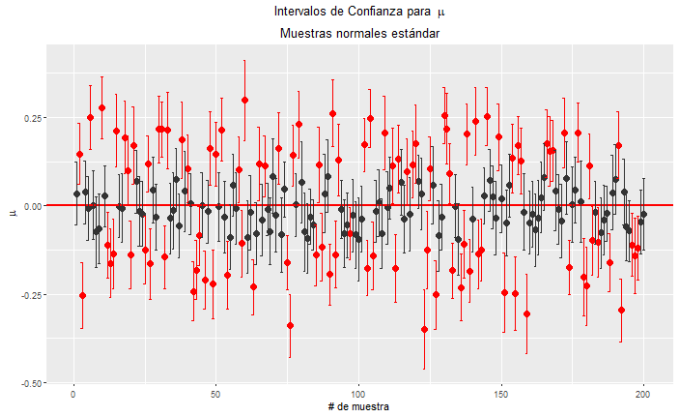


Figura 12: Intervalo de confianza del 50 %

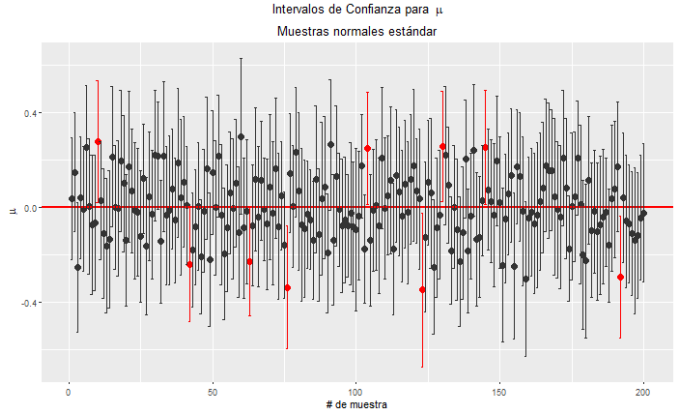


Figura 13: Intervalo de confianza del 95 %

figura. En cuanto a la Figura 13 podemos ver que únicamente 9 de los 200 IC presentados no contienen a $\mu = 0$, lo que significa que el 95.5 % de los IC sí contienen al parámetro. En este caso, se obtuvo que de las 10000 muestras que se generaron, el 94.86 % de los IC contuvieron a μ.

Las figuras 12 y 13 nos permitieron ver que los IC del $1 - \alpha$ definido para μ de variables aleatorias normales estándar, fueron acorde a lo que dice la teoría. A continuación mostraremos qué pasa en el caso de variables aleatorias binomiales.

Para este caso tenemos que si $X_1, X_2, ..., X_n$ es una muestra aleatoria $Bin(n, \theta)$ cuando n es

grande, el intervalo de confianza del $100(1-\alpha)\,\%$ para θ está dado por:

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

donde $z_{1-\frac{\alpha}{2}}$ es el percentil $1-\frac{\alpha}{2}$ de una $N(0,1)$.

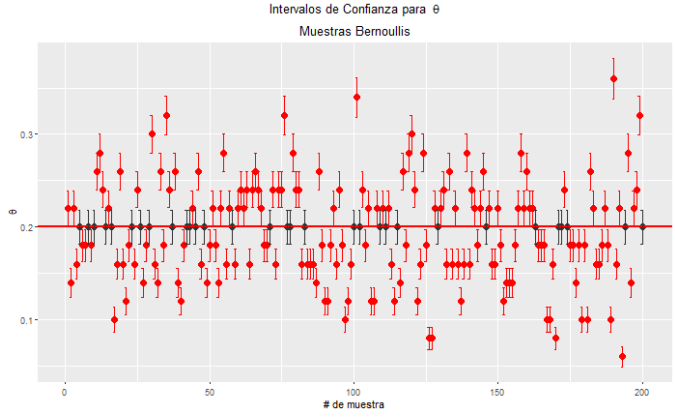


Figura 14: Intervalo de confianza del 25 %

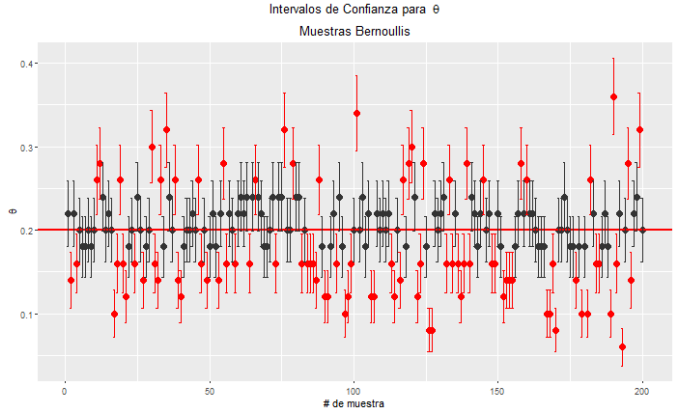


Figura 15: Intervalo de confianza del 50 %

Variables aleatorias Binomial(50,0.2)

Observemos que la Figura 14 tiene muchos más IC que no contienen a $\theta = 0.2$ que la Figura 15, lo cual tiene sentido puesto que la primera figura es un IC del 25 % mientras que el segundo es del 50 %. De las 10000 muestras generadas, para el IC del 25 %, únicamente el 13.35 % de los intervalos contuvieron a la media teórica $\theta = 0.2$, mientras que para el IC del 50 %, el

50.38 % lo contuvieron. Pero, ¡cómo es posible que un IC supuestamente del 25 % únicamente contenga al parámetro θ en el 13.35 % de los casos! ¿Será que sucede algo similar si elegimos otro intervalo? Veamos el caso del IC del 95 %:

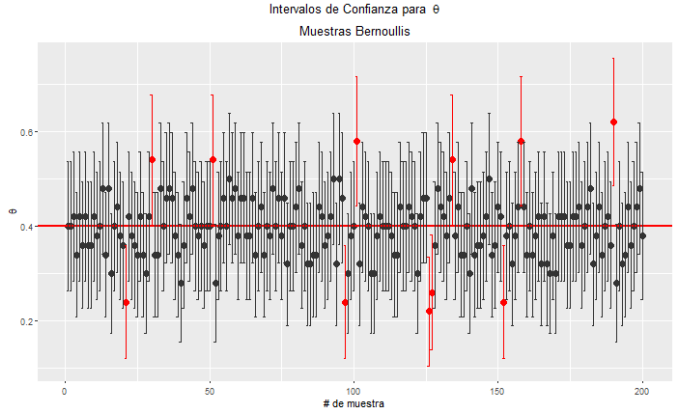


Figura 16: Intervalo de confianza del 95 % para Bin(50,0.4)

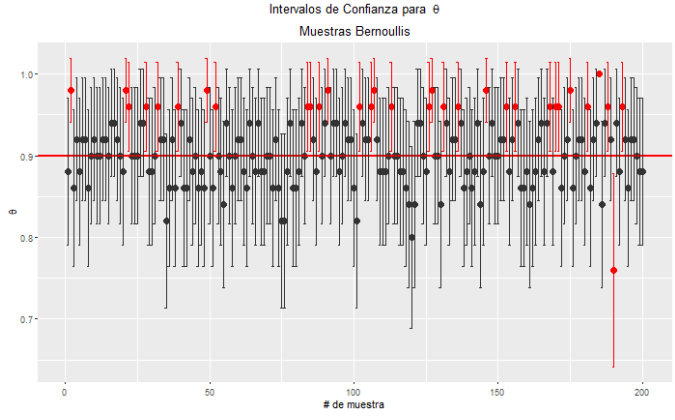


Figura 17: Intervalo de confianza del 95 % para Bin(50,0.9)

Es evidente que la Figura 17 tiene muchos más IC que no contienen al parámetro real $\theta = 0.9$ que la Figura 16 pues en ese caso solo 11 de 200 IC no contienen a $\theta = 0.4$; es decir, el 94.5 % de los intervalos sí contienen a θ. De las 10000 muestras que se generaron, el 94.5 % de los IC referentes a las variables Bin(50,0.4) sí incluyeron a $\theta = 0.4$, lo cual era de esperarse puesto que el IC buscado era del 95 %. Sin embargo, para las variables Bin(50,0.9) únicamente el 88 % de los intervalos contuvieron a $\theta = 0.9$. Pero, ¿por qué pasa esto si nada más cambiamos a θ?

Para entender por qué sucede lo que hemos mencionado, mostraremos la cobertura de los IC para el caso binomial para diferentes tamaños de muestra n y valores de θ.

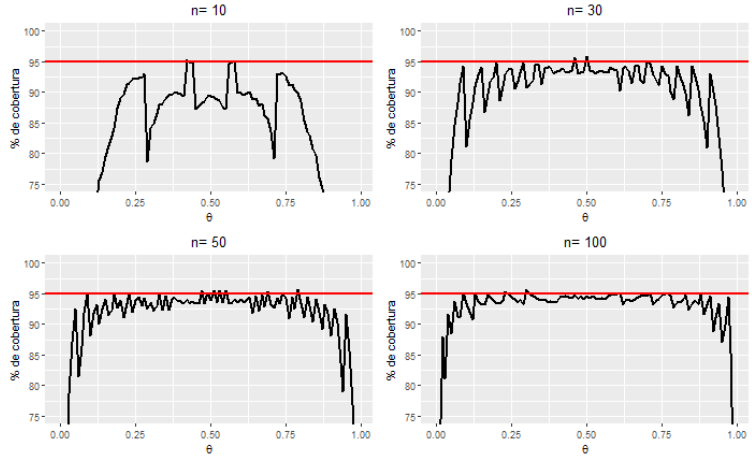


Figura 18: Cobertura de intervalos de confianza para θ

En la Figura 18 hay cuatro gráficos que muestran la cobertura de intervalos de confianza del 95 % para todos los valores de θ de la distribución Binomial y con diferentes tamaños de muestra n. Notemos que conforme n aumenta, la cobertura de los IC aumenta para más valores de θ. Cuando $n = 10$, únicamente dos valores (pareciera $\theta = 0.45$ y $\theta = 0.55$) alcanzan el porcentaje de cobertura del 95 % mientras que para todos los demás valores de θ la cobertura es mucho menor que la deseada. En el caso de $n = 30$, hay más valores de θ para los cuales se alcanza e inclusive se pasa del 95 %. Cuando $n = 50$, hay varios picos que alcanzan la cobertura del 95 %; sin embargo, podemos ver que para valores de θ cercanos a 0 y a 1, el porcentaje de cobertura alcanza 88 % e incluso valores menores al 75 %. Finalmente, cuando $n = 100$ el rango de valores de θ para los cuales se alcanza una cobertura del 95 % es mucho mayor que para el resto de los tamaños de muestra n presentados. Sobre todo podemos ver que cuando $\theta \approx 1$, el porcentaje de cobertura aumentó significativamente con respecto a todas las demás n.

Teorema de la convergencia dominada de Lebesgue

El teorema que vamos a discutir en esta sección es uno de los teoremas estrella de la teoría de la medida. Primero vamos a hablar un poco sobre qué es teoría de la medida, qué hace, luego vamos a dar unas nociones básicas de esta rama tan abstracta de las matemáticas. Estas nociones son necesarias para poder entender lo que el teorema de la convergencia dominada de Lebesgue (TCDL) dice. No vamos a incluir demostraciones pues no es el punto de este trabajo, pero si el curioso lector quiere saber más sobre esto le podemos recomendar que lea a Grabinsky (2016) en [**Grabs**] o a Tao (2011) en [**Tao**]. Si el lector quiere saber más sobre el impacto que tiene la teoría de la medida le podemos recomendar el libro de Ash y Doleans-Dade (1999) en [1]. Todas estas referencias las puede encontrar en este trabajo en

la sección de bibliografía. Para poder entender mejor nuestros dos teoremas estrella y como introducción a la teoría de la medida en general, vamos a empezar dando unas definiciones básicas de esta bella rama.

Definición 1. *Una clase no vacía del conjunto potencia de* $\mathcal{X}$, $\mathcal{S} \subseteq \mathcal{P}(\mathcal{X})$ *se llama* σ*-álgebra si cumple con los siguientes puntos:*

- $E, F \in \mathcal{S} \Rightarrow E \setminus F \in \mathcal{S}$
- *Si* $E_1, E_2, \ldots$ *es una sucesión de elementos en* $\mathcal{S}$, *entonces* $\bigcup_{k\in\mathbb{N}} E_k \in \mathcal{S}$
- $\mathcal{X} \in \mathcal{S}$

Tener esta definición ayuda mucho pues, tal vez nos topemos con una clase de conjuntos de $\mathcal{X}$, pero que no necesariamente sea una σ-álgebra pero que sí querríamos definir una medida sobre ella. En tal caso trabajaríamos con la σ-álgebra generada.

Definición 2. *Supongamos que* $\mathcal{E}$ *es una clase de conjuntos de* $\mathcal{X}$ *(es decir, un conjunto cuyos elementos son subconjuntos de elementos de* $\mathcal{X}$*). En este caso* $\mathcal{E}$ *puede o no formar una* σ*-álgebra. Entonces siempre va a existir una única* σ*-álgebra denotada como* $S(\mathcal{E})$ *con las siguientes propiedades:*

- $\mathcal{E} \subseteq S(\mathcal{E})$
- $S(\mathcal{E})$ *es la menor (con menor elementos)* σ*-álgebra que contiene a* $\mathcal{E}$

A esta σ*-álgebra le vamos a llamar la* σ*-álgebra generada por* $\mathcal{E}$.

A continuación definimos a los Borelianos, una de las σ-álgebras más importantes.

Definición 3. *La* σ*-álgebra de Borel en* $\mathbb{R}$ *se define como:*

$$\mathcal{B}_{\mathbb{R}} = S(\mathcal{A}),$$

donde $\mathcal{A}$ *son los conjuntos abiertos de* $\mathbb{R}$.

Definición 4. *Una vez que tengamos a* $\mathcal{X}$ *y a una* σ*-álgebra definida con esto, entonces la pareja* $(\mathcal{X}, \mathcal{S})$ *se llama espacio medible.*

Es decir, ya tenemos una pareja para poder "medir".

Definición 5. *Sean* $(\mathcal{X}, \mathcal{S})$ *y* $\left(\hat{\mathcal{X}}, \hat{\mathcal{S}}\right)$ *dos espacios medibles. Decimos que una función* $f : \mathcal{X} \to \hat{\mathcal{X}}$ *es* $\mathcal{S} - \hat{\mathcal{S}}$ *medible si* $f^{-1}\left(\hat{E}\right) \in \mathcal{S}$ *para cualquier* $\hat{E} \in \hat{\mathcal{S}}$.

Para simplificar notación, diremos que una función f es $\mathcal{S}$ medible si $\left(\hat{\mathcal{X}}, \hat{\mathcal{S}}\right)$ es $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Diremos que f es Borel-medible si también $(\mathcal{X}, \mathcal{S})$ es $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
Lo que vamos a definir a continuación muy probablemente impresione a nuestro querido lector pues a raíz de esta definición vamos a poder escribir $[-\infty, \infty]$ sin ningún problema.

Definición 6. *Definimos a la recta real extendida al agregarle dos símbolos,* $-\infty$ *y* ∞*, es decir* $\overline{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty\} \bigcup \{\infty\}$.

Con estos dos símbolos las operaciones válidas y propiedades en $\overline{\mathbb{R}}$ son: no se puede dividir entre $-\infty$, ∞ o 0, $-\infty < x < \infty$ para cualquier $x \in \mathbb{R}$ y respetamos el orden usual de $\mathbb{R}$, $(\pm\infty) + x = x + (\pm\infty) = \pm\infty$, $\infty + \infty = \infty$, $(\pm\infty)(\pm\infty) = \infty$ y $(\pm\infty)(\mp\infty) = -\infty$. Con la recta real extendida podemos hacer definiciones de los Borelianos, funciones medibles, etc. como las hicimos para la recta real. Dadas todas estas definiciones básicas ya podemos definir lo que es una medida.

Definición 7. *Sea* $(\mathcal{X}, \mathcal{S})$ *un espacio medible fijo. Una medida,* μ*, es una función conjuntista, es decir* $\mu : \mathcal{S} \to \overline{\mathbb{R}}$ *tal que:*

- $\mu(E) \geq 0$ *para todo* $E \in \mathcal{S}$
- $\mu(\emptyset) = 0$
- *La medida de la unión de elementos disjuntos es la suma de la medida de estos elementos, es decir:*

$$\mu\left(\bigsqcup_{k\in\mathbb{N}} E_k\right) = \sum_{k\in\mathbb{N}} \mu(E_k)$$

Todos estos puntos son lo que nos esperaríamos de una medida (hablando coloquialmente). ¿Qué tiene de distinto esto? La respuesta es que nunca especificamos quién es el espacio medible $(\mathcal{X}, \mathcal{S})$, puede ser un conjunto abstracto, un conjunto de funciones, un conjunto de cosas concretas, etc. En este trabajo nos vamos a centrar en la pregunta ¿cuándo podemos "meter" un límite a una integral? Para esto vamos a por fin enunciar el teorema de la convergencia dominada de Lebesgue (TCDL).

Teorema 2 (Teorema de la convergencia dominada de Lebesgue (TCDL))**.** *Sea* $(\mathcal{X}, \mathcal{S}, \mu)$ *un espacio de medida y* $(f_n)_{n=1}^{\infty}$ *una sucesión en la cerradura de* $(\mathcal{X}, \mathcal{S})$ *supongamos que existe una función* g *positiva y que* $\int |g| d\mu < \infty$*. Supongamos que* $|f_n| \leq g$ *casi donde sea con respecto a* μ *y para todo* $n \in \mathbb{N}$*, entonces si* $f(x) = \lim_{n\to\infty} f_n(x)$*:*

1. $f \in \mathcal{L}_1(\mu)$*, es decir,* $\int |f| d\mu < \infty$
2. *Para todo* $E \in \mathcal{S}$ *tenemos que* $\int_E f\, d\mu = \lim_{n\to\infty} \int_E f_n\, d\mu$.

Vamos a ejemplificar este teorema. El querido lector se podrá dar cuenta que las funciones de probabilidad acumulada son un caso particular de medidas, se le invita a revisar que cumplen los 3 puntos de la definición de medida. Por ende vamos a usarlas para ejemplificar el TCDL. Vamos a usar integración de Monte Carlo crudo y luego con variadas antitéticas. Dentro de estas funciones hay una que no cumple con los supuestos del TCDL y por ende no podemos "meter" el límite a la integral. Usando simulación vamos a graficar esto para que el querido lector pueda ver que sí son necesarios todos los supuestos del teorema. Notamos que la función que no cumple con los supuestos es $f_{3,n}$ pues no hay función alguna que la domine por el

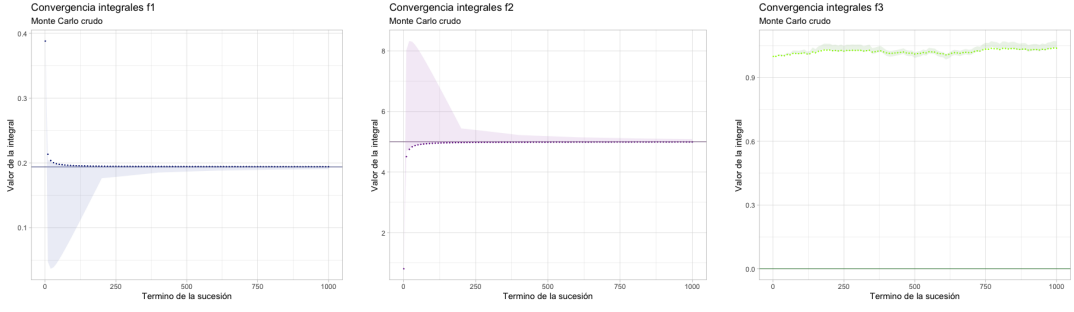
factor n si $0 < x \leq 1/n$. Además vamos a hacer uso del teorema de la modificación de la integral pues todas las funciones son Riemann-integrables. Es decir, podemos "modificar"la integral de la siguiente manera usando la derivada de la medida.

$$\int_E f(x)\, d\mu = \int_E f(x)\, \mu'(x)\, dx.$$

Sucesión de funciones	Función límite	Medida usada
$f_{1,n} = \frac{nx+x}{n}$	$f_1 = x$	Distribución exponencial con $\lambda = 2$
$f_{2,n} = \frac{nx^3-10nx^2+25nx-10x+25}{1+nx}$	$f_2 = (x-5)^2$	Distribución gamma con $\alpha = 5$ y $\beta = 1$
$f_{3,n} = \begin{cases} n & si \quad 0 < x \leq 1/n \\ 0 & otro \quad caso \end{cases}$	$f_3 = \begin{cases} \infty & si \quad x = 0 \\ 0 & otro \quad caso \end{cases}$	Distribución uniforme en $[0,1]$

Figura 19: Ejemplos simulados para el TCDL

Usando la técnica de Monte Carlo Crudo obtenemos las siguientes gráficas. Podemos ver la convergencia en cada uno de los casos mencionados arriba. Pusimos intervalos de confianza del estilo $\left(2\hat{\theta} - \hat{\theta}_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}_{\alpha/2}\right)$ con $\alpha = 0.1$ para mostrar que, efectivamente el límite teórico se encuentra dentro de estas bandas además que la estimación puntual parece que sí converge.



f_1 f_2 f_3

Figura 20: Ejemplos simulados TCDL

Con puntitos representamos los valores que las sucesiones de integrales (las integrales de cada f_n) van tomando para cada n y con líneas el valor aproximado de la integral de la función límite. Como podemos ver la convergencia de la sucesión verde (relacionada con $f_{3,n}$) es distinta al valor de la integral de su función límite f_3. Esto se debe a que esta sucesión de

funciones no cumple con los supuestos necesarios para el TCDL, es decir para "meter" el límite a la integral.
Con la gráfica **??** ya podemos tener más claro que, efectivamente, no podemos "meter" límites a integrales, pues ambos son procesos límites y no es tan trivial intercambiarlos. De hecho, ni las bandas de los intervalos de confianza incluyen a la integral del límite.
Las funciones que "dominan" a $f_{1,n}$ y $f_{2,n}$ son:

$$\begin{aligned} g_1 &= x+1 \\ g_2 &= (x-5)^2 \end{aligned}$$

El curioso lector puede confirmar que, efectivamente, $\int |g_i|\, d\mu < \infty$ para las distintas medidas usadas para $i = 1, 2$. También podemos apreciar en esta segunda gráfica que, independientemente de cuál técnica de integración por simulación usemos, obtenemos resultados similares.

Teorema de Bayes

Bayes (1763) analiza el siguiente experimento mental: se lanza una bola sobre una mesa y se marca la posición de la bola. Después, se lanzan más bolas y en cada lanzamiento registramos si la bola cayó a la izquierda o a la derecha, adelante o atrás, de la primera bola lanzada. Thomas Bayes calculó la probabilidad condicional de la posición de la primera bola dada la secuencia de lanzamientos de las demás bolas y notó que entre más pelotas eran lanzadas, más actualizaba su idea de dónde estaba la primera bola. Aunque no estaba completamente seguro de la localización, con cada nueva pieza de evidencia, podía estar cada vez estaba más seguro de la posición real. Esta es la idea detrás de todo que dio origen a lo que conocemos como Teorema de Bayes.

Teorema 3 (Teorema de Bayes)**.** *Sea* $\{A_1, A_2, ..., A_n\}$ *un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de* 0*. Sea* B *un evento cualquiera del que se conocen las probabilidades condicionales,* $P(B|A_i)$*. Entonces, la probabilidad* $P(A_i|B)$ *viene dada por la expresión:*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

donde $P(A_i)$ *son las probabilidades a priori,* $P(B|A_i)$ *es la probabilidad de* B *en la hipótesis* A_i *y* $P(A_i|B)$ *son las probabilidades a posteriori.*

Con base en la definición de probabilidad condicionada se obtiene la Fórmula de Bayes, también conocida como Regla de Bayes:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

La regla de Bayes proporciona un modo natural de actualización de las creencias cuando aparece nueva información. Este proceso de aprendizaje inductivo por medio de este teorema es la base de la Inferencia Bayesiana. La diferencia fundamental entre la estadística clásica

(frecuentista) y la bayesiana es el concepto de probabilidad. Para la estadística clásica es un concepto objetivo, que se encuentra en la naturaleza, mientras que para la estadística bayesiana responde a la siguiente pregunta: ¿qué nos dicen los datos X acerca del parámetro θ? Ignora toda evidencia externa. Por el contrario, en el caso bayesiano, además de la muestra también juega un papel fundamental la información previa o externa que se posee en relación a los fenómenos que se tratan de modelar. En este caso, la pregunta que se responde es: ¿cómo cambian nuestros juicios originales acerca del valor de la cantidad desconocida θ a la luz de los datos X?

El objetivo es encontrar la distribución condicional de todas aquellas cantidades de interés cuyo valor desconocemos dado el valor conocido de las variables observadas.

Enfoque bayesiano

El problema que nos encontramos es el siguiente: tenemos datos X y queremos conocer el valor de ciertas cantidades θ que nos interesan, como parámetros del modelo, mediciones, etc. Como estadísticos, postulamos un modelo de probabilidad:

$$p(x|\theta)$$

Desde el punto de vista bayesiano, además, θ debe tener una distribución de probabilidad $p(\theta)$, que refleja nuestra incertidumbre inicial acerca del parámetro y se conoce como distribución a priori. Al incluir esta distribución, el análisis Bayesiano usa la probabilidad para representar la incertidumbre en todas las partes de un modelo estadístico. X, también es conocido, por lo que condicionamos en su valor observado x. Por lo tanto, nuestro conocimiento acerca del valor de θ queda descrito a través de su distribución final, $p(\theta|x)$, conocida como distribución a posteriori. El Teorema de Bayes nos dice como encontrarla y los pasos a seguir son los siguientes:

1. Especificación del modelo muestral, $p(x|\theta)$;
2. Especificación de una distribución inicial, $p(\theta)$;
3. Cálculo de la distribución final, $p(\theta|x)$, vía el Teorema de Bayes;
4. Resumen de la información contenida en $p(\theta|x)$ para hacer inferencias sobre las cantidades de interés.

Ejemplo teórico

Consideremos el primer ejemplo. Supongamos que sales positivo a un test para detectar el COVID-19. El doctor que te realizó la prueba te dijo que de cada 100 personas que tienen la enfermedad, el test puede detectar 99 casos. Ahora, ¿cuál es la probabilidad de que actualmente tengas la enfermedad, dado que el test fue positivo?. El Teorema de Bayes puede ser de gran ayuda. Definamos los siguientes eventos:

- H =Tener COVID-19,
- $\neg H$ = No tener COVID-19,
- E = Salir positivo en la prueba,
- $\neg E$ = Salir negativo en la prueba.

La distribución a priori $P(H)$ es difícil de determinar, pero en este caso un punto razonable para empezar es la frecuencia de la enfermedad en la población, supongamos que es 0.001. Por los demás datos del problema, $P(E|H) = 0.99$, $P(E|\neg H) = .01$, $P(\neg H) = 0.999$. Entonces tenemos lo siguiente:

$$P(H|E) = \frac{P(E|H)P(E)}{P(H)P(E|H) + P(\neg H)P(E|\neg H)} = \frac{.99(.001)}{.001(.99) + .999(.01)} = 9\,\%$$

Se obtiene una probabilidad de 9 % de realmente tener la enfermedad después de salir positivo y es un valor muy bajo. Parece magia, pero es sentido común aplicado a matemáticas. La formula de Bayes no está hecha para aplicarse una sola vez, fue creada para ser usada múltiples veces y cada vez ganando evidencia y actualizando la probabilidad de que algo es verdad. El primer ejemplo, cuando sales positivo a una prueba, qué pasaría si fueras al doctor y vuelves a aplicarte la prueba pero ahora por otro laboratorio independiente y ese test también vuelve a ser positivo. Ahora, ¿cuál es la probabilidad de que actualmente tengas la enfermedad? En lugar de usar $P(H) = 0.001$ como distribución a priori, usamos la distribución posterior obtenida en el ejemplo anterior, $P(H) = 0.09$. Calculamos la nueva probabilidad posterior:

$$P(H|E) = \frac{.99(.09)}{.09(.99) + .91(.01)} = 90.73\,\%$$

La nueva probabilidad basado en dos pruebas positivas, es del 91 %. Hay 91 % de probabilidad de tener la enfermedad, dos diferentes resultados de dos diferentes laboratorios incrementan las probabilidades pero aún no es tan alta como el nivel de precisión del test, que es del 99 %. El Teorema de Bayes es de gran importancia porque nos permite determinar la probabilidad de las causas a partir de los efectos que han podido ser observados. Esta idea también puede combinarse con métodos de simulación para analizar nuestros datos y resolver problemas, el método de Computación Bayesiana Aproximada es un ejemplo de ello.

Computación Bayesiana Aproximada (CBA)

Utilizaremos el método conocido como Computación Bayesiana Aproximada, que, aunque es ineficiente computacional-mente, es fácil de entender. El método CBA es un enfoque basado en simulación que no requiere una formulación explícita de la función de verosimilitud $p(x|\theta)$. En cambio, este método aproxima la función de verosimilitud mediante simulaciones, cuyos resultados se comparan con los datos observados. CBA utiliza un mecanismo de aceptar-rechazar para realizar el cálculo posterior.
Necesitamos tres cosas para definir nuestro proceso de estimación:

1. *Datos:* es el valor observado, nuestro conjunto de observaciones $x_1, ..., x_n$.
2. *Un modelo generador (MG):* es cualquier tipo de programa computacional, expresión matemática o conjunto de reglas, que recibe como argumento un conjunto de parámetros fijos y nos devuelve datos simulados. Su función principal es aproximar la función de verosimilitud mediante simulaciones.
3. *Distribución a priori:* ¿Qué información se tiene sobre el modelo, antes de observar los datos?.

Para entender mejor la aplicación de este método resolveremos un ejercicio conocido.

Problema de captura y recaptura

Un estadístico está interesado en el número N de peces que hay en un estanque. Él captura 250 peces, los marca y los regresa al estanque. Unos cuantos días después regresa y atrapa peces hasta que obtiene 50 peces marcados, en ese punto también tiene 124 peces no marcados (la muestra total es de 174 peces).
¿Cuál es la estimación de N?, ¿de qué manera se puede resolver este problema utilizando Inferencia Bayesiana?

Solución con CBA

Regresamos a la estructura de un *CBA*. En este problema en concreto los *datos* son la observación obtenida, x son los *peces no marcados* que se extraen después de obtener 50 peces marcados. El *modelo generador* corresponde al proceso de *marcar y recapturar* que se define más adelante. El parámetro a estimar es $\theta =$ *número de peces en el lago*. En este caso, lo que necesitamos estimar es el número de peces en el lago y sabemos que el número de peces no marcados es igual a 124.

Marcar y recapturar

Definiremos una función que realice el procedimiento de marcar y recapturar a los peces. Será nuestro MG que simularemos y nos ayudará a estimar el número de peces en el lago θ.

1. Capturar 250 peces;
2. Marcarlos y regresar los al lago;
3. Después de un tiempo, capturar peces hasta que se obtienen 50 peces marcados;
4. Contar, en los resultados del paso anterior, cuántos peces no están marcados (no contamos el número de peces marcados porque este siempre será igual a 50. La observación que nos interesa son los peces no marcados).

Estimación del modelo

Definimos el siguiente procedimiento para ajustar el modelo:

- Obtener una muestra aleatoria de tamaño n de la distribución a priori del parámetro $\theta_1, ..., \theta_n$;
- Aplicar el MG a cada observación obtenida de la muestra y así obtener un conjunto de observaciones $x_1, ..., x_n$;
- Nos quedaremos que las θ_i cuyos valores generados sean iguales a los observados (en el caso en particular, nos quedaremos con las θ_i que generan 124 peces no marcados);
- La distribución de aquellos parámetros que cumplen la condición representa la probabilidad de que la observación haya sido producida por cierto valor del parámetro.

A partir de la distribución obtenida podemos hacer estimaciones del valor desconocido.

Distribución a priori no informativa

Supongamos que la distribución a priori de $\theta \sim Unif\{250, ..., 1500\}$. Cuando la distribución a priori es la distribución uniforme la llamamos *distribución a priori no informativa* porque no contiene mucha información sobre el parámetro. Le estamos asignando el mismo peso de probabilidad a todos los valores. A continuación programamos el MG, extraemos la muestra de la distribución a priori y graficamos su histograma.

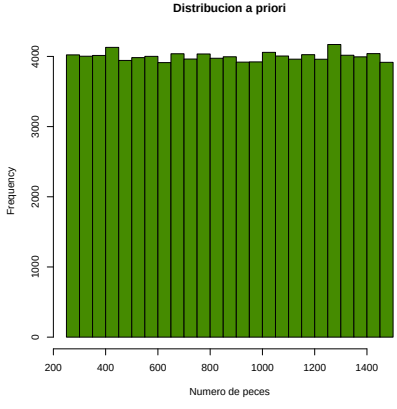


Figura 21: Distribución a priori

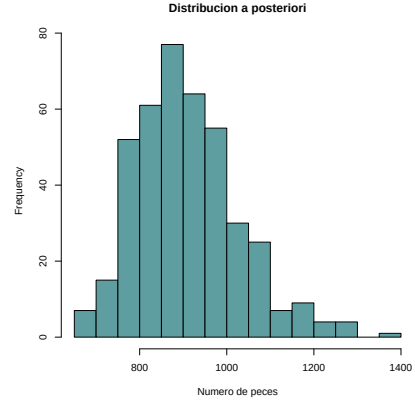


Figura 22: Distribución posterior

Aunque la distribución es uniforme discreta al principio, cuando observamos la distribución posterior del parámetro es distinto al inicial. No parece que la distribución posterior sea simétrica, está sesgada a la izquierda. La mayoría de las observaciones se encuentran entre 800 y 950. Utilizamos los datos observados después de aplicar el modelo generador de datos para actualizar nuestro conocimiento sobre el parámetro y obtener una mejor aproximación de la distribución del mismo. Por ejemplo, los valores menores a 600 o mayores 1400 nunca simularon datos iguales a los que nosotros observamos, por eso es que la probabilidad posterior es muy pequeña. La distribución posterior es lo que el modelo sabe sobre el parámetro

después de observar los datos. La distribución posterior contiene información del modelo y de los datos. A pesar de obtener como resultado final la distribución del parámetro, nos interesa conocer un valor puntual. Generalmente tomamos el valor que mayor probabilidad tiene de generar los datos (*MLE*) o podemos tomar el valor esperado de la distribución. El número de peces estimados, utilizando el método de *MLE*, es el siguiente: 869 peces

Distribución a priori informativa

Una ventaja de utilizar el enfoque bayesiano, es que puedes incluir fuentes de información además de los datos, por ejemplo, la opinión de un experto. Supongamos que el pescador más experimentado del lugar te comenta lo siguiente: "Siempre ha habido muchos peces en el lago. Alrededor de 1000, diría yo". Esta información se incluye en la distribución a priori del parámetro. Cambiamos la distribución de la uniforme a otra distribución más informativa. Elegimos como distribución a priori una distribución normal con media $\mu = 1,000$ y una desviación estándar de 100.

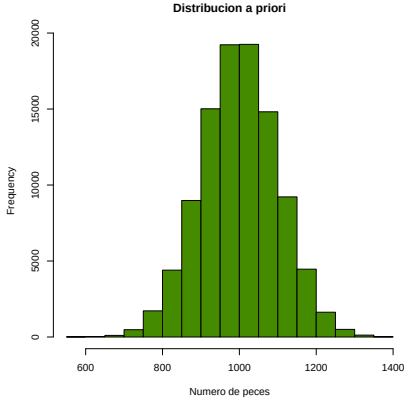


Figura 23: Distribución a priori

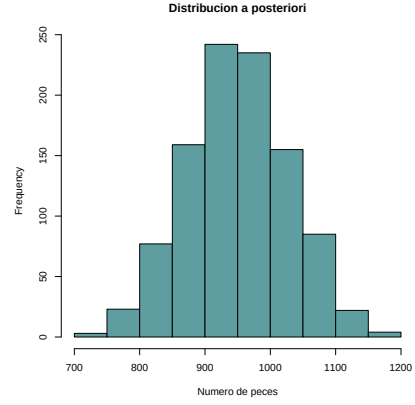


Figura 24: Distribución posterior

Parece ser que después de usar la información otorgada por el pescador experto y la información proveniente de los datos, es más probable que el número de peces esté entre 900 y 1,000. Además, la distribución a posterior ahora toma una forma simétrica y parecida a la normal, por influencia de la distribución a priori que elegimos en esta ocasión. El número de peces estimados, utilizando el método de *MLE*, es el: 955. Incluir la opinión del experto dentro del análisis hace que las estimaciones cambien y mejoren (suponiendo que el experto no nos engaña).

Bibliografía

[1] Conchi Ausín. *Introducción a la Inferencia Bayesiana.* `http://halweb.uc3m.es/esp/Personal/personas/causin/esp/2012-2013/SMB/Tema6.pdf` (accesado: 29.11.2020).

[2] Ash, Robert B., and Doléans-Dade Catherine. *Probability and Measure Theory.* Academic Press, 1999.

[3] Brown, Lawrence D., et al. "Interval Estimation for a Binomial Proportion." *Statistical Science*, Institute of Mathematical Statistics, 2001, `projecteuclid.org/euclid.ss/1009213286`.

[4] C. Dalitz. Çonstruction of Confidence Intervals". *Technical Report* No. 2017-01, 2017, `https://arxiv.org/pdf/1807.03582.pdf`.

[5] Grabinsky, Guillermo. *Teoría De La Medid*a. Las prensas de ciencias, 2000.

[6] Grazian, Clara, and Yanan Fan. "A Review of Approximate Bayesian Computation Methods via Density Estimation: Inference for Simulator-Models." *WIREs Computational Statistics*, vol. 12, no. 4, 2019, `https://arxiv.org/pdf/1909.02736.pdf`.

[7] Rick Wicklin. "Coverage Probability of Confidence Intervals: A Simulation Approach." The DO Loop, 8 Sept. 2016, `blogs.sas.com/content/iml/2016/09/08/coverage-probability-confidence-intervals.html`. (accesado: 07.11.2020)

[8] Tao, Terence. *An Introduction to Measure Theory.* American Mathematical Society, 2011.

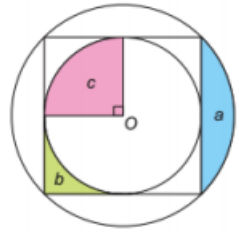
[9] Gútierrez Peña, Eduardo. *Estadística bayesiana: Teoría y Conceptos Básicos.* `http://www.dpye.iimas.unam.mx/soriano/BAYES/DOCUMENTOS/NOTAS/INFERENCIA%20BAYESIANA.pdf`. (accesado: 29.11.2020)

[10] Rohrer, Brandon. *How Bayes Theorem works.* `https://www.youtube.com/watch?v=5NMxiOGL39M`. (accesado: 27.11.2020).

Activa tus neuronas

Retos matemáticos

1. En la siguiente figura se resaltan tres regiones a, b y c las cuales están determinadas por un cuadrado con centro en O y por las circunferencias inscrita y circunscrita.



¿Cuál de las siguientes afirmaciones es correcta?

a) $c = a + b$.
b) $c = a - b$.
c) $c = 2a + b$.
d) $c = a + 2b$.
e) $c = 2a - b$.
f) $c = 2b - a$.

2. La mayor potencia de 2 que divide al producto $1 \times 2 \times 3 \times 4 \times \cdots \times 2027 \times 2028$ es 2^{2020}.

¿Cuál es la mayor potencia de dos que divide al producto $1 \times 2 \times 3 \times 4 \times \cdots \times 4047 \times 4048$?

3. En la siguiente igualdad, a, b y c son números enteros positivos, ¿cuál es el valor de c?

$$\frac{10}{7} = a + \frac{1}{b + \frac{1}{c}}.$$

4. Observe que en la igualdad $360 = 90 + 120 + 150$, el número 360 se escribió como la suma de tres números los cuales son proporción 3, 4 y 5 del mismo número, 30, respectivamente.

¿De cuántas maneras se puede escribir el número 360 como suma de tres sumandos enteros, en orden creciente y con proporción de tres números enteros positivos consecutivos?

Enigmas matemáticos

1. La abuela Sara quiere saber cuál de sus cinco nietas había hecho un dibujo en la pared de la sala de estar. Las nietas hicieron las siguientes declaraciones:

   - Nayeli: Yo no fui.
   - Allison: Yo tampoco fui.
   - Diana: Quien dibujó fue Yaressi o Andrea.
   - Yaressi: No fue Andrea, ni Roxanna.
   - Andrea: No fue Diana.
   - Roxanna: Diana no dice la verdad.

   Si solo una de sus nietas mintió, ¿quién hizo el dibujo?

2. En un lago hay 10 rocas sospechosamente alineadas y una ranita, la cual se encuentra en la primera roca. Sabemos que la rana puede saltar a cualquier roca que esté en frente, pero no puede retroceder. ¿De cuántas formas puede llegar a la décima roca?

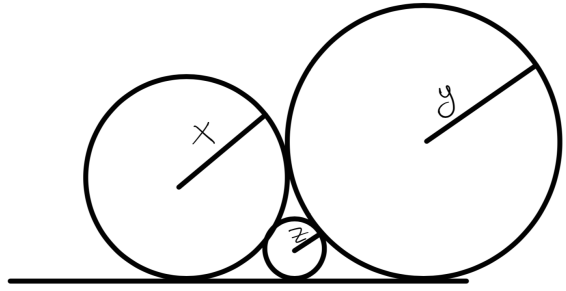
   Extra: ¿Cuál sería la respuesta si el lago tuviera n rocas?

-0.703467422498391652049818601859902130342928431034

Zona Olímpica

1. Encuentra todos los valores $x, y, z \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, tales que $13xy45z$ es divisible por 792. Donde las variables x, y, z representan dígitos del número de 7 cifras.

2. Sean 3 círculos con radios x, y, y z con una misma linea tangente. Estos círculos son tangentes entre si como se puede ver en la siguiente imagen. Encuentra la relación entre x, y, z



3. a) Muestra que la curva $x^3 + 3xy + y^3 = 1$ contiene un conjunto de puntos distintos A, B, C tales que ΔABC es equilátero.

   b) Encuentra $A(\Delta ABC)$

4. Sea $\{a_n\}_1^\infty$ una sucesión definida como $a_1 = 1$, $a_n = a_1 a_2 \cdots a_{n-1} + 1$. Encuentra $\sum_{n=1}^{\infty} \frac{1}{a_n}$

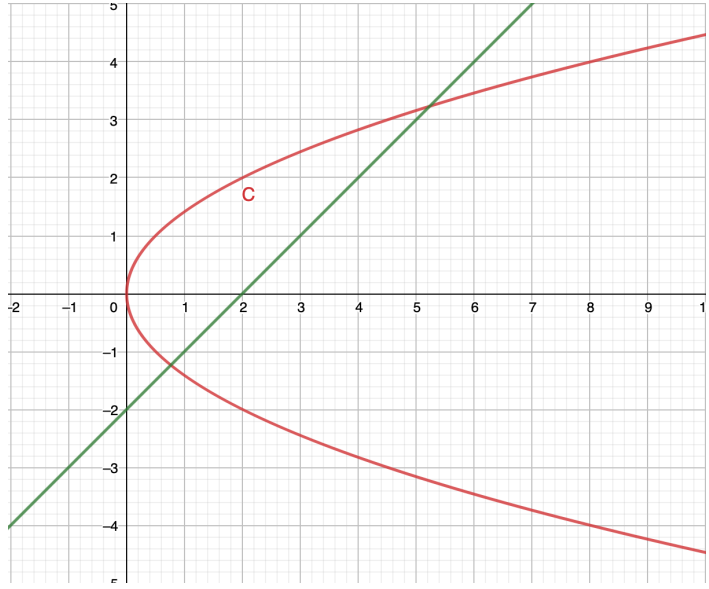
5. Para $a, b, c \in N$ resuelve:

$$\left(1 + \frac{1}{a}\right)\left(1 + \frac{1}{b}\right)\left(1 + \frac{1}{c}\right) = 2$$

6. Como se muestra en la figura de abajo, hay una recta $l : x - y - 2 = 0$ y una parábola $C : y^2 = 2px$ $(p > 0)$.

a) Si la recta l pasa por el foco de la parábola C, ¿Cuál es la ecuación de la parábola?

b) Hay dos puntos P, Q en C tales que P y Q son simétricos respecto a l. Pueba que el punto medio del segmento de recta PQ es $(2-p, -p)$.



Pregunta de Erdős

Sea $f : \mathcal{R} \to \mathcal{R}$ una función infinitamente diferenciable que satisface $f(0) = 0$, $f(1) = 1$, y $f(x) \geq 0 \ \forall x \in \mathcal{R}$. Prueba que existe un entero positivo n y un real x tal que $f^{(n)}(x) < 0$.

OMIC y sus matematiquitos

Teresa Calderón
Estudiante de Matemáticas Aplicadas
Montserrat Olvera
Estudiante de Actuaría y Finanzas
Javier Navarro
Estudiante de Actuaría
Andrés Villareal
Estudiante de Matemáticas Aplicadas

La Olimpiada de Matemáticas ITAM Construye (OMIC) nació hace cuatro años con el propósito de impartir clases de matemáticas para la preparación de alumnos cursando primaria y secundaria para su participación en la Olimpiada Metropolitana.

OMIC es una rama de "ITAM Construye", la cual es una organización estudiantil del Instituto Tecnológico Autónomo de México (ITAM), que busca generar un sentimiento de comunidad entre los alumnos universitarios y sus vecinos de la colonia Tizapán.

Los estudiantes del ITAM que participan en OMIC han creado un espacio en donde los alumnos se sienten cómodos y seguros para acercarse y explorar el mundo de las matemáticas a través de métodos didácticos y amigables, como lo mencionan en los pilares del programa. Se ha construido un lugar en donde se fomenta el aprendizaje más a fondo sobre temas que se ven en sus respectivas escuelas, al igual que la aclaración de dudas sobre temas matemáticos en general.

Estando conscientes que la única manera para seguir mejorando y creciendo el programa es preguntando a los alumnos su opinión, OMIC realizó una encuesta en donde sus estudiantes pudieron expresar su opinión sobre las matemáticas, sus planes a futuro en cuanto a qué van a estudiar en la universidad y su percepción del programa OMIC.

En el área de la encuesta sobre las matemáticas en general los alumnos respondieron que les agradan e incluso mencionaron que sus temas favoritos son las fracciones, las divisiones y el cálculo de áreas. También reconocieron que utilizan lo aprendido en su vida diaria, por lo que entienden la importancia de los temas que se ven en clase y su impacto en su día a día.

Sobre su futuro universitario las respuestas fueron muy diversas, pero hubo algunos que aunque no sabían qué estudiar específicamente, sabían que tendría que ser algo relacionado con matemáticas.

En cuanto a la opinión de los alumnos sobre OMIC fue positiva, pero muy diversa. Algunos mencionaron que era una herramienta complementaria de su escuela para seguir aprendiendo matemáticas. Otros definieron OMIC como el lugar donde realmente entienden los temas e incluso donde aprenden nuevos, lo que les genera una gran ventaja. Y por último, varios comentaron que OMIC ha sido el lugar donde han "perdido el miedo" por las matemáticas y que incluso han llegado a ser divertidas.

Aún con la pandemia, OMIC ha continuado con su compromiso de seguir apoyando y preparando a sus alumnos en su estudio y para la Olimpiada de Matemáticas, impartiendo sus clases de forma remota. Tras recibir los resultados de la tercera etapa de la Olimpiada Metropolitana, dos de sus alumnas de tercero de secundaria han sido invitadas a los entrenamientos virtuales los cuales comenzaron el pasado 23 de marzo.

Por parte de los profesores que forman OMIC, que nuestros alumnos participen en la Olimpiada Metropolitana, así como las opiniones expresadas por los alumnos en la encuesta, demuestra que nuestro trabajo, aún con el reto de educar a distancia, ha logrado su objetivo principal: compartirles nuestro gusto por las matemáticas y crear un ambiente donde ellos puedan disfrutarlas tanto como nosotros.

OMIC
OLIMPIADA DE MATEMÁTICAS
ITAM CONSTRUYE