



A VUELTAS CON LOS NÚMEROS PRIMOS Y LA CONJETURA DE GOLDBACH (1742)

ON PRIME NUMBERS AND GOLDBACH'S CONJECTURE (1742)

INDICE

	<u>Pág.</u>
Resumen / Resum / Abstract.....	3
PARTE I. NÚMEROS PRIMOS, ESTADÍSTICA Y SERIES NUMÉRICAS.....	4
1. Introducción. La Teoría de Números y los números primos	4
2. Clasificación y ampliaciones sucesivas del concepto de número.....	5
3. Amistad con un genio llamado Euler	10
4. La infinitud de los números primos.....	11
5. Convergencia y divergencia de series y recíprocos de los números primos	13
6. Estadística descriptiva de los números primos.....	19
6.1. Contenidos en los primeros 200 números naturales	19
6.2. Contenidos en los primeros 1000 números naturales	20
6.3. Contenidos en los primeros 10000 números naturales	23
6.3.1. Introducción.....	23
6.3.2. Valores centrales de la distribución de frecuencias.....	25
6.3.3. Medidas de dispersión.....	31
6.3.4. Asimetría o sesgo.....	34
6.3.5. Apuntamiento o curtosis	37
6.3.6. Medidas de concentración.....	37
6.3.7. Medida de independencia	41

	<u>Pág.</u>
7. Muestreo estratificado	44
7.1. Objeto.....	44
7.2. Afijación uniforme.....	44
7.3. Afijación proporcional	44
7.4. Afijación óptima	45
7.5. Afijación valoral	46
8. Mecanismo generador de números primos	46
 PARTE II. PLANTEAMIENTO Y DEMOSTRACIONES.....	 53
1. La conjetura fuerte o binaria de Goldbach (planteamiento).....	53
2. La conjetura débil o ternaria de Goldbach (planteamiento).....	54
3. Las diferentes medias de dos números primos (demostración)	55
4. Una aproximación geométrica (demostración).....	59
4.1. Conjetura fuerte o binaria	59
4.2. Conjetura débil o ternaria	60
5. Otro intento de demostración de la conjetura fuerte o binaria	62
6. Metáfora o parábola de las gotas de agua (demostración)	64
 PARTE III. UNAS REFLEXIONES FINALES.....	 67
 PARTE IV. REFERENCIAS BIBLIOGRÁFICAS Y RECOMENDACIONES.....	 71
 Símbolos, abreviaturas y siglas.....	 76
 Anexo de información complementaria	 73
1. Números primos comprendidos en los 10000 primeros números naturales	 78
2. Percentiles de la distribución χ^2 de Pearson.....	82
3. Distribución normal tipificada N (0,1)	83
 Relación de figuras / Relación de tablas	 84
 Índice onomástico de autores.....	 85

Josep M. Franquet Bernis
Dr. Ingeniero Agrónomo, EUR-ING
Dr. Ciencias Económicas y Empresariales
Diplomado en Investigación Operativa
Miembro de la Sociedad Catalana de Matemáticas
Miembro de la Real Sociedad Matemática Española
Académico Numerario de la Real Academia Europea de Doctores
Ex Director del Centro Asociado de Tortosa y del Campus Nordeste de la UNED

RESUMEN

Una vez presentadas diversas consideraciones acerca de los números primos, se observa que la conjetura fuerte o binaria de Goldbach proclama que todo número par suficientemente grande (mayor que 2) puede expresarse como suma de dos números primos, mientras que la conjetura débil o ternaria aduce que todo número impar suficientemente grande (mayor que 5) puede expresarse como suma de tres números primos, y ésta ya ha sido demostrada en el año 2013. El presente artículo concluye presentando un intento de demostración de la conjetura fuerte partiendo no solo del razonamiento matemático sino también del lógico-inductivo. El trabajo se inicia también con diversas consideraciones en relación a las series numéricas y a la estadística descriptiva de los números primos contenidos en los primeros 10000 números naturales.

Palabras clave: número primo, número natural, serie, término general, modelo, convergencia, infinitud, metáfora, inducción, distribución, concentración, muestreo, estrato.

RESUM

Un cop presentades diverses consideracions envers els nombres primers, s'observa que la conjectura forta o binària de Goldbach proclama que tot nombre parell suficientment gran (més gran que 2) pot expressar-se com a suma de dos nombres primers, mentre que la conjectura feble o ternària addueix que tot nombre imparell o senar prou gran (més gran que 5) pot expressar-se com a suma de tres nombres primers, i aquesta darrera ja ha estat demostrada a l'any 2013. El present article conclou, amb un intent de demostració de la conjectura forta partint no només del raonament matemàtic sinó també del lògic-inductiu. El treball comença també amb diverses consideracions en relació a les sèries numèriques i a l'estadística descriptiva dels nombres primers continguts en els primers 10000 nombres naturals.

Mots clau: nombre primer, nombre natural, sèrie, terme general, model, convergència, infinitud, metàfora, inducció, distribució, concentració, mostreig, estrat.

SUMMARY / ABSTRACT

Once various considerations about prime numbers have been presented, it is observed that Goldbach's strong or binary conjecture proclaims that every sufficiently large even number (greater than 2) can be expressed as the sum of two prime numbers, while the weak or ternary conjecture claims that every sufficiently large odd number (greater than 5) can be expressed as the sum of three prime numbers, and this has already been demonstrated in 2013. This article concludes by presenting an attempt at a demonstration of the strong conjecture based not only on mathematical reasoning but also on logical-inductive reasoning. The work also begins with various considerations in relation to the numerical series and the descriptive statistics of the prime numbers contained in the first 10,000 natural numbers.

Key words: prime number, natural number, series, general term, model, convergence, infinity, metaphor, induction, distribution, concentration, sampling, stratum.

PARTE I. NÚMEROS PRIMOS, ESTADÍSTICA Y SERIES NUMÉRICAS

1. INTRODUCCIÓN. LA TEORÍA DE NÚMEROS Y LOS NÚMEROS PRIMOS

Creo, junto con Alberto A. Enríquez Vanderkan, que los números primos ejercen sobre sus múltiples estudiosos una fascinación hipnótica. Existen tres razones básicas para ello. La primera es que cualquier número natural puede descomponerse en factores de números primos, aunque algunos de estos números primos se repitan, lo que constituye el Teorema Fundamental de la Aritmética¹ que implica que el 1 no puede ser primo al decir que: cualquier número entero mayor que 1 puede escribirse como un único producto de números primos. Si aceptáramos al 1 como primo habría infinitas formas de descomponer cada número (10 sería 2×5 , pero también $2 \times 5 \times 1$, $2 \times 5 \times 1 \times 1$, y así sucesivamente). La segunda tiene que ver con la conjetura débil de Goldbach que establece, como veremos a continuación, que todo número impar mayor que cinco se puede expresar como la suma de tres números primos. Esta conjetura ya fue demostrada el año 2013. La tercera razón es por la conjetura fuerte o binaria de Goldbach, que establece que todo número par mayor que 2 puede escribirse como suma de dos números primos. Esta conjetura, en cambio, no ha sido demostrada hasta la fecha y la trataremos aquí, con mayor especificidad, en la segunda parte de nuestro trabajo.

En general, la Teoría Analítica de los Números o, simplemente, la Teoría de Números es la rama de las Matemáticas que estudia las propiedades de los números (en particular los enteros) y concretamente los naturales (enteros positivos) como el mínimo conjunto que es inductivo. Contiene muchos problemas que pueden ser comprendidos por los “no matemáticos”, como creo es el caso del que aquí nos ocupa y que puede considerarse como el 8º de los problemas matemáticos del Milenio.

En el año 1912, Landau (1877-1938) estableció en el Quinto Congreso Internacional de Matemáticos² de Cambridge una lista de cuatro de los problemas sobre números primos, que se conocen como los “problemas de

¹ El teorema fundamental de la aritmética (o teorema de factorización única) establece que todo número natural tiene una representación única como producto de factores primos, salvo el orden. Un mismo factor primo puede aparecer varias veces. Ello implica que las funciones aritméticas aditivas y multiplicativas están completamente determinadas por sus valores en las potencias de los números primos. El resultado apareció ya en los *Elementos* de Euclides en el año 300 a.C. y alcanzó su formulación definitiva con K. F. Gauss (1777-1855) el año 1801. El 1 se representa entonces como un producto vacío. La importancia de este teorema es una de las razones para excluir el 1 del conjunto de los números primos. Si se admitiera el 1 como número primo, el enunciado del teorema requeriría efectuar aclaraciones adicionales.

² El **Congreso Internacional de Matemáticos** (*International Congress of Mathematicians*, **ICM**) es el más importante congreso en la comunidad matemática mundial. Se celebra cada cuatro años bajo los auspicios de la Unión Matemática Internacional (IMU). En el congreso de 1900, en París, David Hilbert anunció su famosa lista de 23 problemas abiertos en matemáticas, ahora llamados los “problemas de Hilbert”. Al congreso de 1998 asistieron 3.346 participantes. El programa constó de 21 horas de lecturas plenarias y 169 conferencias de 45 minutos sobre temas especializados impartidas por matemáticos distinguidos seleccionados por los organizadores del Congreso. Adicionalmente, cada participante podía presentar su trabajo en 15 minutos de exposición o bien a través de un cartel. Según la *American Mathematical Society*, más de 4500 participantes asistieron al congreso de 2006 que tuvo lugar en Madrid.

Landau". Ninguno de ellos está resuelto hasta la fecha. Se trata de la conjetura binaria de Goldbach que trataremos aquí, la de los números primos gemelos, la de Legendre (1798) y la de los primos de la forma $(n^2 + 1)$.

2. CLASIFICACIÓN Y AMPLIACIONES SUCESIVAS DEL CONCEPTO DE NÚMERO

Cabe destacar, en primer lugar, que el concepto de número natural ha sido bastante bien asimilado por los humanos en la mayoría de las civilizaciones, siendo profusamente utilizado³. Sin embargo, todas las ampliaciones del concepto de número han pasado al dominio del conocimiento de sólo unos pocos a lo largo de los años. Así podemos ver, por ejemplo, que la profesión de "cambista" ha revestido una gran importancia en todos los tiempos debido, principalmente, a las dificultades aritméticas que estas operaciones llevaban consigo por el manejo de números racionales. Otro punto a destacar es la posición del juez entre los árabes; la herencia, debido al código coránico, conlleva muchas dificultades en su división y es, por ello, que en los pueblos islámicos dichos jueces ostentaban una posición o rango especial por su facilidad de realizar este tipo de reparticiones (nuevamente aparece aquí la dificultad de trabajar con los números racionales).

Estas dificultades operativas de las diversas ampliaciones del concepto de número no han sido superadas hasta en época muy reciente debido, principalmente, a la sistematización de las sucesivas ampliaciones. Recordemos, como muestra de que estos problemas no son muy lejanos, la creación por J. Neper (1550-1617) de sus célebres *tablillas*, que son unas verdaderas anticipaciones de las reglas de cálculo y las calculadoras empleadas en los siglos XX y XXI, para que sus contemporáneos pudieran realizar, con relativa facilidad, los productos y divisiones de números, y todo esto en pleno siglo XVII. Por todo ello, los conceptos que vamos a exponer a continuación para ver las ampliaciones del concepto de número son, sin duda, problemas que, de un modo u otro, se le habrán presentado a muchas personas a lo largo de la historia y que, en algunas ocasiones, habrán dado lugar a una visión de los números de más amplio campo de acción.

Podemos ver la aparición de los números enteros al pensar en un gran almacén que empaqueta un cierto producto. En cada paquete, según se desvíe del peso exacto, en más o menos, se le etiqueta el signo + o el – para indicar el sentido de la desviación, además de un determinado número para indicar la desviación en magnitud. Sin embargo, la aparición de los números enteros podemos verla realmente en cualquier etapa del desarrollo del sistema de registro contable de dobles cuentas, a partir del monje franciscano renacentista

³ Un número natural es cualquiera de los números que se usan para contar los elementos de ciertos conjuntos. Los números naturales se representan con la notación $\{\mathbb{N}\} = \{1, 2, 3, 4, \dots\} = \{\mathbb{Z}^+\}$. De dos números vecinos, el que se encuentra a la derecha se llama "siguiente" o "sucesivo", por lo que el conjunto de los números naturales es ordenado e infinito. Algunas características de los números naturales son: a) todo número mayor que 1 va después de otro número natural, b) entre dos números naturales siempre hay un número finito de naturales (interpretación de conjunto no denso), c) dado un número natural cualquiera, siempre existe otro natural mayor que éste (interpretación de conjunto infinito), d) entre un número natural y su sucesor no existe ningún otro número natural.

italiano Luca Pacioli (1494), en su obra "*De computis et scripturis*" ("De las cuentas y las escrituras")⁴.

En cualquier momento de un registro contable puede ocurrir que, al hacer balance de dicho registro, debido a las malas o buenas gestiones productivas y comerciales, existen pérdidas o ganancias, lo que se refleja en números enteros. Esto nos lleva a construir el conjunto de los números enteros por los elementos: ... -3, -2, -1, 0, +1, +2, +3,... Aparece aquí la primera ampliación del conjunto de número natural. Efectivamente, este conjunto de números enteros es una ampliación, pues lo contiene como caso particular, ya que los números enteros pueden ser positivos y negativos y los números enteros positivos coinciden, precisamente, con los números naturales. Posteriormente, el conjunto de los números primos $\{P\}$ se constituye como un subconjunto de los números naturales $\{N\}$, de tal suerte que, mediante ampliaciones sucesivas (ver Fig. 1) se cumple que: $\{P\} \subset \{N\} \subset \{Z\} \subset \{Q\} \subset \{R\} \subset \{C\}$.

Los números primos absolutos, que serán objeto de tratamiento en el presente Artículo, son aquellos que sólo son divisibles entre ellos mismos y el 1, es decir, que si intentamos dividirlos por cualquier otro número, el resultado no es entero. Así, un número natural e impar p es primo si:

- 1) $p > 1$.
- 2) $\text{Div}(p) = \{1, p\}$.

Por consenso, el número 1 no se considera primo. El número 2 es el único número primo que a la vez es par. A los números naturales que no son primos se les denomina "números compuestos", que son los números naturales que tienen algún divisor natural aparte de sí mismos y del 1, y, por lo tanto, pueden factorizarse. El número 1 no se considera primo, ya que, en ese caso, la factorización de los números naturales (que constituye el teorema fundamental de la aritmética) no sería única, puesto que siempre se puede multiplicar por 1, pero tampoco se considera un número compuesto.

La propiedad de ser número primo se denomina "primalidad". En la teoría algebraica de números, los números primos se denominan *números racionales primos* para distinguirlos de los *números gaussianos primos*⁵. La primalidad no depende del sistema de numeración, pero sí de la estructura algebraica de

⁴ Se trata del Título Noveno, Tratado XI de su "*Summa de Arithmetica, Geometria, Proportioni et Proportionalita*", editado en Venecia, primer tratado que expone y divulga el sistema conocido por el nombre de "partida doble" sobre el que se basa el desarrollo de la contabilidad hasta nuestros días.

⁵ Los números primos gaussianos tienen una propiedad interesante: cada primo gaussiano es un divisor propio de algún primo entero. Es decir, si q es un primo gaussiano, entonces existe un entero primo p tal que q divide p (en los enteros gaussianos). En otras palabras, la "estructura aritmética" o "estructura teórica de números" de los enteros gaussianos es una versión más detallada de la misma estructura en los enteros.

“anillo”⁶ donde se estudia la primalidad. El 2 es un primo racional; sin embargo, posee factores como entero gaussiano.

La mayoría de los números se comportan acorde a reglas sencillas y claras. Por el contrario, los números primos son un auténtico incordio: aparecen donde quieren, sin previo aviso, de una forma aparentemente caótica, y sin seguir ningún tipo de pauta. Y lo peor del caso es que no se pueden ignorar, han sido objeto de multitud de estudios en la matemática y resultan fundamentales en la criptografía. El enigma de los números primos radica, básicamente, en que no se reconoce todavía un patrón (ley de formación o término general) en ellos y, en su consecuencia, no resulta posible saber cuál será el primo siguiente.

Estos números primos, como señala acertadamente A. Enríquez Vanderkan, poseen diversas particularidades, algunas de las cuales son:

1ª) Un número natural mayor que 3 es primo si su raíz cuadrada es un número decimal y si no es divisible por ningún número primo menor o igual al entero de la raíz cuadrada. Esto constituye la base del algoritmo para determinar si un número entero es o no primo.

2ª) Todos los números primos son impares a excepción del 2, por lo tanto, su último dígito es 1, 3, 7 ó 9. El 5 no está incluido en esta lista puesto que, aunque es impar y primo, nunca es la cifra final de otro número primo. De hecho, todos los números que terminan en 5 son múltiplos de éste y, en su consecuencia, no son primos. Todos los números naturales que terminan en 0 son divisibles entre 5 también, por lo que tampoco serán primos.

3ª) El par (2,3) está formado por los únicos números primos consecutivos.

4ª) Una “laguna” de números no primos es un segmento de los números naturales en donde no existe inserto ningún número primo. Se pueden construir lagunas de números no primos o compuestos entre números primos, de tamaños tan grandes como se desee.

5ª) Los divisores propios de un número natural son los números naturales que lo pueden dividir, resultando de cociente otro número natural y de resto 0, es decir, la división es exacta. Cada número tiene una cantidad concreta de divisores. Dos números naturales distintos son “amigos” si a cada uno de ellos se le obtiene sumando los divisores propios del otro.

6ª) Un número primo p no es amigo de ningún número m mayor que 1. Puesto que al ser p primo tiene exactamente un divisor propio, el 1 y, por tanto, la

⁶ En álgebra abstracta, un **anillo** es cualquier conjunto R no vacío con dos operaciones de composición interna, adición y multiplicación, siendo con la primera un grupo abeliano y la segunda operación asociativa y además distributiva respecto de la primera. Si se da la situación de que la segunda operación sea conmutativa se le nombra *anillo conmutativo* y si hubiera un elemento identidad o neutro para la segunda operación se llama *anillo unitario*. Para el caso, el anillo de los cuaternios (que son una extensión de los números reales similar a la de los números complejos) es un anillo unitario y no conmutativo y el anillo de los enteros pares es un anillo conmutativo sin unidad. El concepto de anillo surge en el siglo XIX y viene a generalizar las propiedades de los números enteros. La palabra “anillo” parece haber sido introducida por David Hilbert (1862 -1943).

suma de los divisores propios de p es 1 distinto de m . Los números primos, pues, no tienen amigos.

7ª) Los números primos de la forma $(4n + 1)$ son igual a la suma de dos cuadrados perfectos; por lo que no son números gaussianos primos. En tanto que los números primos de la forma $(4n + 3)$ sí son números gaussianos primos. Todo número racional primo es un número gaussiano entero, sin ser necesariamente número gaussiano primo.

8ª) Los números primos “gemelos” son aquellos a los que les separa una diferencia de número par, el número 2, que también es primo. Por ejemplo: (3,5), (5,7), (11,13). Estos dos números primos se diferencian en dos unidades. La tabla 1, que se expone a continuación, facilita la elaboración de un algoritmo para encontrar números primos gemelos. No se ha podido demostrar si hay infinitas parejas de primos gemelos o bien si éstas son finitas. En efecto, si observamos las 15 parejas de números primos gemelos contenidos en los primeros 200 números naturales, que aparecen en la tabla 1, se tiene:

Tabla 1. Números primos gemelos de los primeros 200 naturales.

Pareja de primos gemelos	Suma	Producto	Producto +1	Representación potencial del Producto +1
3 y 5	8	15	16	4^2
5 y 7	12	35	36	6^2
11 y 13	24	143	144	12^2
17 y 19	36	323	324	18^2
29 y 31	60	899	900	30^2
41 y 43	84	1763	1764	42^2
59 y 61	120	3599	3600	60^2
71 y 73	144	5183	5184	72^2
101 y 103	204	10403	10404	102^2
107 y 109	216	11663	11664	108^2
137 y 139	276	19043	19044	138^2
149 y 151	300	22499	22500	150^2
179 y 181	360	32399	32400	180^2
191 y 193	384	36863	36864	192^2
197 y 199	396	39203	39204	198^2

Curiosamente, si en la última columna de esta tabla multiplicamos la cifra del exponente por la base obtendremos la columna de la “Suma”, mientras que esa base es la media aritmética de la pareja de gemelos en cuestión. La columna “Producto + 1” continuaría así: 51984 (227 y 229), 57600 (239 y 241), 72900 (269 y 271), 79524 (281 y 283), 97344 (311 y 313), 121104 (347 y 349), 176400 (419 y 421), 186624 (431 y 433), 213444 (461 y 463), 272484 (521 y 523), 324900 (569 y 571), 360000 (599 y 601), 381924 (617 y 619), 412164 (641 y 643), 435600 (659 y 661), 656100 (809 y 811), 675684 (821 y 823), 685584 (827 y 829), 736164 (857 y 859), ..., cuyas raíces cuadradas permiten identificar sucesivamente las parejas de primos gemelos (entre paréntesis). Ver nota explicativa al final del presente capítulo (*).

9ª) Los números primos entre sí o números primos relativos son dos o más números que no tienen más divisor común que 1. El mayor divisor común o

máximo común divisor (m.c.d.) de varios números primos entre sí es 1. Así, 8 y 15 son primos entre sí o primos relativos porque su único factor común es 1, porque 8 es divisible por 2, pero 15 no, y 15 es divisible por 3 y por 5, pero 8 no. Los números 7, 12 y 15 son primos entre sí porque 7 no divide a 12 ni a 15; 2 divide a 12, pero no a 7 ni a 15; 3 divide a 12 y a 15, pero no a 7; 5 divide a 15, pero no a 7 ni a 12; luego, su único divisor común es 1. En cambio, 12, 14 y 18 no son primos entre sí porque 2 los divide a todos.

Obsérvese que para que dos o más números sean primos entre sí no es necesario que sean primos absolutos. Así, 8 no es primo, 15 tampoco, y sin embargo, son primos entre sí; 7 es primo, 12 no lo es y 25 tampoco y son primos entre sí. Ahora bien, si dos o más números son primos absolutos cada uno de ellos, evidentemente serán primos entre sí.

Una clasificación ilustrativa acerca de los diferentes tipos de números, donde se sitúan los primos que estamos tratando, puede verse a continuación:

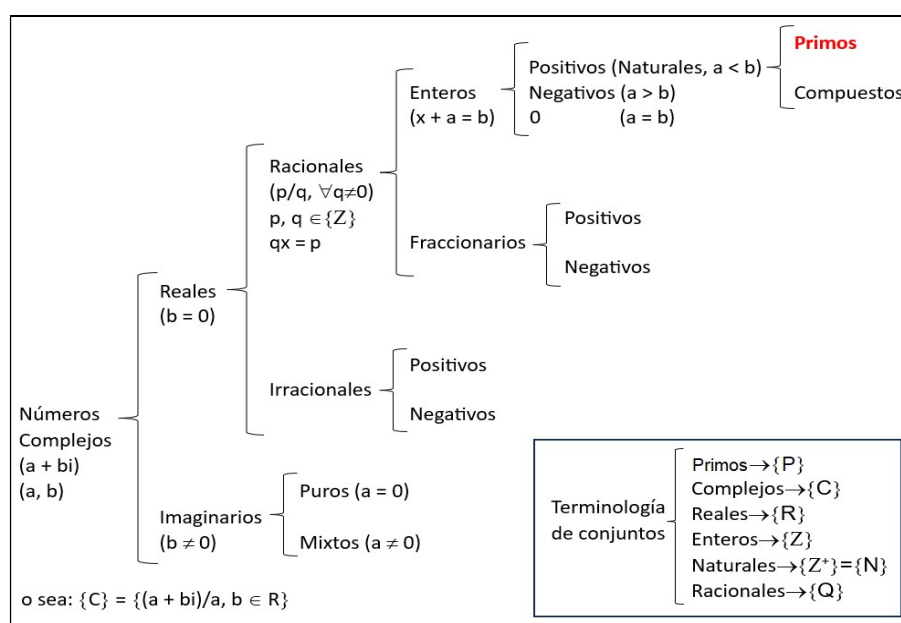


Fig. 1. Ampliaciones sucesivas del concepto de número.

Otro tema relacionado, que juzgamos de particular interés, hace referencia a los conceptos de “simulación” y “modelos matemáticos”, que se utilizan para el análisis de diferentes fenómenos económicos, meteorológicos, médicos y sociológicos. Descartes ya parecía vislumbrar, por primera vez, la noción general de isomorfismo⁷ (que él llamaba “semejanza”) y la posibilidad de identificar relaciones u operaciones isomorfas; puso como ejemplos los de la adición y la multiplicación. Sin embargo, estas ideas atrevidas no tuvieron

⁷ En matemáticas, un **isomorfismo** (del griego *iso-morfos*: igual forma) es un homomorfismo (o más generalmente un morfismo) que admite un inverso. El concepto matemático de **isomorfismo** pretende captar la idea de tener la misma estructura. Dos estructuras matemáticas entre las que existe una relación de isomorfismo se llaman **isomorfas**. El interés de los isomorfismos reside en el hecho de que dos objetos isomorfos tienen las mismas propiedades (excluyendo información adicional como la estructura o los nombres de los objetos). Así, las estructuras isomorfas no pueden distinguirse desde el punto de vista de la estructura únicamente, y pueden identificarse.

ningún eco entre sus contemporáneos y habría que esperar hasta el gran desarrollo del álgebra de mediados del siglo XIX para vislumbrar el comienzo de la materialización de los sueños cartesianos.

Es precisamente en este momento histórico cuando los modelos se multiplican y se acostumbra a pasar de una teoría a otra a través de un simple cambio de lenguaje. Mediante el descubrimiento de las coordenadas homogéneas⁸ en el año 1837, Möbius (1790-1868) no sólo pudo entender las nociones fundamentales de puntos impropios e imaginarios introducidos por Poncelet (1788-1867) y apreciar en todo su valor, junto con Gergonne (1771-1859), Plücker (1801-1868) y Chasles (1793-1880), el principio de la dualidad, sino que pudo dar un tratamiento completo del invariante fundamental de la geometría proyectiva: la “razón doble” de cuatro puntos alineados.

El uso, cada vez más extendido, de la noción de “modelo” permitiría, en el mismo siglo, llevar a cabo la unificación de las matemáticas soñada por los pitagóricos. Solamente quedaban a parte los números enteros como “productos exclusivos de nuestro espíritu”, tal y como los definía Gauss en 1832, oponiéndolos a la noción de espacio. No obstante, a finales del siglo XIX, los números enteros y los números primos en ellos contenidos pasan a ser el fundamento de todas las matemáticas clásicas. Los modelos basados en la aritmética van adquiriendo más importancia con la extensión del método axiomático y la concepción de los objetos matemáticos como creaciones libres, prodigiosas y admirables del espíritu humano.

3. AMISTAD CON UN GENIO LLAMADO EULER

El matemático prusiano Christian Goldbach (1690-1764), había nacido en Königsberg, Brandeburgo-Prusia, actual Kaliningrado, Rusia. Estudió leyes, idiomas y matemáticas. Realizó varios viajes a través de Europa y conoció a varios matemáticos famosos, como Gottfried Leibniz, Leonhard Euler y Daniel Bernoulli. En 1725, se convirtió en historiador y profesor de matemáticas en San Petersburgo. Tres años más tarde, se trasladó a Moscú para trabajar para el zar Pedro II de Rusia. Realizó importantes trabajos en el campo de las matemáticas, pero hoy en día es conocido por la llamada “conjetura de Goldbach” o “conjetura fuerte de Goldbach”. También estudió y demostró varios teoremas sobre potencias perfectas. La “conjetura fuerte de Goldbach” ha sido verificada para todos los números pares menores que 4×10^{14} por Joerg Richstein (1998).

La “conjetura débil de Goldbach” fue resuelta en el 2013 por el matemático peruano Harald Andrés Helfgott, quien consiguió demostrar que para todo número impar mayor que 10^{30} la conjetura es cierta. Luego, con el uso de una computadora, verificó que cada número impar menor a 10^{30} (o incluso menor a 8.875×10^{30}) podía expresarse como la suma de tres números primos.

⁸ En geometría proyectiva, las coordenadas homogéneas son un instrumento usado para describir un punto en el espacio proyectivo. También pueden usarse como un sistema alternativo de coordenadas para trabajar en el espacio euclídeo, pues éste puede verse como un subconjunto del espacio proyectivo. De ese modo, las coordenadas homogéneas son ampliamente usadas en infografía para la representación de escenas en tres dimensiones.

Goldbach planteó por carta, fechada el 7 de junio de 1742 (que incluye temas estrictamente matemáticos y algunos asuntos personales que involucraban a personajes de la cultura francesa, alemana, italiana y rusa) su famosa conjetura sobre los números primos a su amigo Euler⁹, algo más joven, que éste no la pudo resolver. Hay que tener en cuenta que L. Euler (1707-1783), probablemente uno de los tres mayores matemáticos de la Historia moderna (junto a Gauss y Riemann) es autor de la considerada, por muchos tratadistas, como “la más bella fórmula de las Matemáticas” (aunque ocho años antes también la mencionara Roger Cotes en su expresión logarítmica, denominada también como “nuestra joya” por el físico Richard Feynman), que establece la relación existente entre las funciones trigonométricas y la función exponencial compleja para cualquier número real x , a saber:

$$\begin{cases} e^{ix} = \cos x + i \cdot \sin x \\ e^{-ix} = \cos x - i \cdot \sin x \end{cases} \quad \text{que, confirma la identidad trigonométrica:}$$

$$1 = (\cos x + i \cdot \sin x) \times (\cos x - i \cdot \sin x) = \cos^2 x - i^2 \cdot \sin^2 x = \cos^2 x + \sin^2 x,$$

que, con $x = \pi$ radianes = 180° sexagesimales, se convierte en:

$$e^{i\pi} + 1 = 0 \text{ (“identidad de Euler”).}$$

4. LA INFINIDAD DE LOS NÚMEROS PRIMOS

El cardinal del conjunto de los números primos es infinito, por lo que el número de sumas binarias que se pueden realizar entre ellos también lo es, y el resultado será siempre un número par puesto que se trata de la adición de dos números impares. De ahí se deduce que se pueden formar infinitos números pares N pero, *a priori*, ello no implica que todos los N se pueden expresar necesariamente como suma de dos números primos y no se constituye aquí la locución latina *per se et essentialiter*, en la conjetura fuerte de Goldbach que estamos analizando.

Existen numerosas demostraciones acerca de la infinidad de los números primos. Así, en un interesante *paper* de Amo *et al.* (2013) se recogen muchas de ellas, con su autoría y desarrollo. Concretamente, por orden cronológico, las de Euclides (300, a.C., en el libro IX de su obra *Elementos*), Goldbach (1730-1742), Euler (1737), Kummer (1878), Perrot (1881), Stieltjes (1890), Thue (1897), Braun (1897), Auric (1915), Metrod (1917), Pólya (1938), Fürstenberg (1955), Harris (1956), Wunderlich (1965), Dressler (1975), Nair (1982) y otras varias más recientes de Hacks (1983), Pollack (2003), etc. En estas demostraciones se recogen, como apoyaturas, numerosos conceptos

⁹ Euler conoció a Goldbach en 1727 -en ese año, la Universidad de Basilea lo rechazó como docente- y Nicolás y Daniel Bernoulli lo invitaron a formar parte de la Academia de Ciencias de San Petersburgo que estaba bajo el mando y cuidado de Catalina I, quien murió poco antes del traslado de Euler a Rusia, en mayo de 1727. Por afinidad de intereses iniciaron una sólida amistad y relación profesional, constatada en la correspondencia que sostuvieron desde 1729 hasta 1764. Existen cerca de doscientas cartas —al menos las conocidas— con temas tan variados como teoría de los números, series, números complejos, teoría de funciones, cálculo, etcétera (en 1741 Euler comunicó a Goldbach su famosa identidad). En ellas se pueden encontrar las semillas de muchos de los resultados que posteriormente Euler publicaría.

matemáticos que no procede desarrollar aquí por comprensibles razones de espacio, tales como recurrencias, funciones de Euler o de Riemann (1859), evaluación euleriana de series diversas, números de Fermat y de Fibonacci, topologías, fracciones continuas, formulaciones de Legendre (1898) y Stirling, y otros.

Por lo que se refiere a esta última formulación (Amo *et al.*, 2013), veamos que combinando las fórmulas de Stirling, $n! \cong e^{n \ln n - n + \frac{1}{2} \ln(2\pi n)}$ y de Legendre, $n! = \prod_p p^{\left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \dots}$ tenemos lo siguiente:

$$n^n e^{-n} < n! = \prod_p p^{\left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \dots} \leq \prod_p p^{\frac{n}{p} + \frac{n}{p^2} + \frac{n}{p^3} + \dots} = \left(\prod_p p^{\frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \dots} \right)^n \leq C^n.$$

La desigualdad $n^n e^{-n} < C^n$ es equivalente a $\frac{n}{C} < e$. Esto último sólo es posible para un número finito de valores de n . Por ello, concluimos que el número de primos es infinito.

Otra demostración parecida tiene lugar utilizando la fórmula de Legendre, con lo que tenemos la expansión en serie de $(2n!)/(n!)^2$, en $n = 0$, mediante la serie de Taylor:

$$1 + \frac{\pi^2 n^2}{6} + n^3 \psi^{(2)}(1) + \frac{19 \pi^4 n^4}{360} + O(n^5)$$

donde $\psi^{(n)}(x)$ es la n -ésima derivada de la función digamma.

Y, por otra parte, en $n = \infty$:

$$4^n \left(\frac{\sqrt{\frac{1}{n}}}{\sqrt{\pi}} - \frac{\left(\frac{1}{n}\right)^{3/2}}{8\sqrt{\pi}} + \frac{\left(\frac{1}{n}\right)^{5/2}}{128\sqrt{\pi}} + \frac{5\left(\frac{1}{n}\right)^{7/2}}{1024\sqrt{\pi}} - \frac{21\left(\frac{1}{n}\right)^{9/2}}{32768\sqrt{\pi}} - \frac{399\left(\frac{1}{n}\right)^{11/2}}{262144\sqrt{\pi}} + O\left(\left(\frac{1}{n}\right)^6\right) \right)$$

Así mismo:

$$\begin{aligned} \binom{2n}{n} &= \frac{(2n)!}{n!n!} = \frac{\prod_p p^{\left[\frac{2n}{p}\right] + \left[\frac{2n}{p^2}\right] + \left[\frac{2n}{p^3}\right] + \dots}}{\prod_p p^{2\left[\frac{n}{p}\right] + 2\left[\frac{n}{p^2}\right] + 2\left[\frac{n}{p^3}\right] + \dots}} = \\ &= \prod_p p^{\left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right] + \left[\frac{2n}{p^2}\right] - 2\left[\frac{n}{p^2}\right] + \left[\frac{2n}{p^3}\right] - 2\left[\frac{n}{p^3}\right] + \dots} = \prod_p p^{\log_p(2n)} \leq \prod_p 2n = 2^r n^r \end{aligned}$$

Si aplicamos la fórmula de Stirling al primer término, tenemos que, para valores grandes de n , $2^n = O(n^r)$. Pero esto último sólo es posible para un número finito de valores de n . Deducimos, entonces, que el cardinal del conjunto de los números primos es infinito.

Otros matemáticos han demostrado la infinitud de los números primos con diversos métodos procedentes de diversas áreas de las matemáticas tales como al álgebra conmutativa y la topología. Algunas de estas demostraciones se basan en el uso de sucesiones infinitas con la propiedad de que cada uno de sus términos es coprimo con todos los demás, por lo que se crea una biyección (aplicación biyectiva, o sea, inyectiva y suprayectiva o sobreyectiva simultáneamente) entre los términos de la sucesión y un subconjunto (infinito) del conjunto de los primos.

Una sucesión que cumple dicha propiedad es la sucesión de Euclides-Mullin, que deriva de la demostración euclídea de la infinitud de los números primos, ya que cada uno de sus términos se define como el factor primo más pequeño de uno más el producto de todos los términos anteriores. La sucesión de Sylvester¹⁰ se define de forma similar, puesto que cada uno de sus términos es igual a 1 más el producto de todos los anteriores. Aunque los términos de esta última sucesión no son necesariamente todos primos, cada uno de ellos es coprimo con todos los demás, por lo que se puede escoger cualquiera de sus factores primos, por ejemplo, el menor de ellos, y el conjunto resultante será un conjunto infinito cuyos términos son todos primos.

5. CONVERGENCIA Y DIVERGENCIA DE SERIES Y RECÍPROCOS DE LOS NÚMEROS PRIMOS

Por cierto, en cuanto a la aplicabilidad de la formulación del matemático escocés J. Stirling (1692-1770) que referíamos en el punto anterior, de gran utilidad para el cálculo de límites de sucesiones de números reales y de la que se hacen eco algunas de las más recientes demostraciones acerca de la infinitud de los números primos, como hemos visto, vamos a ejemplificar el estudio de la convergencia o carácter de la serie de números enteros positivos de término general:

$$a_n = \frac{n!}{n^n}$$

La conocida fórmula de Stirling viene dada por la expresión:

$$\lim_{n \rightarrow \infty} \frac{n!}{e^{-n} \cdot n^n \cdot \sqrt{2\pi n}} = 1$$

de donde se obtiene la equivalencia, de gran utilidad:

$$n! \approx e^{-n} \cdot n^n \cdot \sqrt{2\pi n}.$$

¹⁰ La sucesión de números enteros de Sylvester se llama así en honor de James Joseph Sylvester (1814-1897), quien la investigó por primera vez en 1880. Sus términos crecen de forma doblemente exponencial, y la suma de sus inversos constituye una serie de fracciones unitarias que converge a 1 más rápidamente que ninguna otra serie de fracciones unitarias con la misma suma. La manera en que se define permite que sus términos se factoricen más fácilmente que otros números del mismo orden de magnitud, pero, debido al ritmo de crecimiento de los mismos, sólo se conoce la factorización completa en factores primos de unos pocos términos. Los términos de esta sucesión también han tenido usos en la representación finita de fracciones egipcias de suma 1, así como en las variedades sasakianas y las de Albert Einstein (1879-1955).

Más exactamente la fórmula es como sigue:

$$n! = \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} - \frac{1}{1680n^7} + \dots}$$

donde el último término del producto del 2º miembro (la exponencial) tiende a 1 cuando n tiende a infinito.

La lista de los numeradores es: 1, -1, 1, -1, 1, -691, 1, -3617, 43867, -174611, ..., alternativamente positivos y negativos.

La lista de los denominadores es: 12, 360, 1260, 1680, 1188, 360360, 156, 122400, 244188, 125400, ...

Desarrollando este último término también se puede reescribir la fórmula como:

$$n! = \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \frac{571}{2488320n^4} + \dots\right)$$

Una acotación de la fórmula es:

$$\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{12n}}$$

Por ejemplo, operando con el número primo $p = n = 29$, se tiene:

$$29! = 8\,841\,761\,993\,739\,701\,954\,543\,616$$

$$e^{\frac{1}{12 \cdot 29+1}} = 1.002869438\dots$$

$$e^{\frac{1}{12 \cdot 29}} = 1.002877696\dots$$

$$29! = \sqrt{2\pi 29} \cdot \left(\frac{29}{e}\right)^{29} \cdot 1.002877577\dots$$

Una expresión alternativa usando la función gamma de Euler-Legendre¹¹ es:

$$n! = \int_0^{+\infty} x^n \cdot e^{-x} \cdot dx,$$

que mediante el oportuno cambio de variable ($x = n \cdot y$) e integrando por partes conduce a:

¹¹ En matemáticas, la función gamma es una aplicación que extiende el concepto de factorial a los números reales y complejos. De hecho, la función gamma extiende el concepto de factorial a cualquier valor complejo de z . La función gamma aparece en varias funciones de distribución de probabilidad, por lo que es bastante usada tanto en probabilidad y estadística como en combinatoria.

$$\begin{aligned}
 n! &= \int_0^{+\infty} e^{n \cdot \ln x - x} \cdot dx = n \int_0^{+\infty} e^{n \cdot \ln ny - ny} \cdot dy = n \int_0^{+\infty} e^{n(\ln ny - y)} \cdot dy = \\
 &= n \int_0^{+\infty} e^{n(\ln n + \ln y - y)} \cdot dy = e^{n \cdot \ln n} \cdot n \int_0^{+\infty} e^{n(\ln y - y)} \cdot dy.
 \end{aligned}$$

Aplicando el método de Laplace se obtiene:

$$\int_0^{+\infty} e^{n(\ln y - y)} \cdot dy \approx \sqrt{\frac{2\pi}{n}} \cdot e^{-n}, \text{ de donde:}$$

$$n! \approx e^{n \cdot \ln n} \cdot n \cdot \sqrt{\frac{2\pi}{n}} \cdot e^{-n} = \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n, \text{ c.s.q.d.}$$

La fórmula en cuestión resulta útil en diversas áreas como la mecánica estadística, donde aparecen ecuaciones que contienen factoriales del número de partículas (en nuestro caso de gotitas de agua, como veremos posteriormente al final de la parte II). Puesto que los sistemas macroscópicos típicos tienen del orden del famoso número del científico italiano Avogadro (aproximadamente es $N_A = 6 \times 10^{23}$) de partículas¹², la fórmula de Stirling resulta, en la práctica, exacta. Además, la fórmula de Stirling es diferenciable, lo que permite el cálculo muy aproximado de los puntos críticos de la función correspondiente, como extremos absolutos y relativos o locales (máximos y mínimos), puntos de inflexión, intervalos de crecimiento, decrecimiento, concavidad, convexidad, etc., en expresiones que contienen factoriales.

Se trata, ahora, de estudiar, como curiosidad, la convergencia de la serie que conformaremos con los 5 primeros números primos de término general:

$a_n = \frac{n!}{n^n}$; para lo que: $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \frac{n!}{n^n} = 0$, con su correspondiente expansión en serie¹³:

¹² Desde la revisión efectuada en la 26ª Conferencia General de Pesas y Medidas (CGPM), celebrada en el año 2018, la **constante de Avogadro** (símbolos: **L**, **N_A**) tiene un valor exacto definido como $6.022\,140\,76 \times 10^{23} \text{ mol}^{-1}$. El **número de Avogadro** equivale a la cantidad de $6.022\,140\,76 \times 10^{23}$ expresada en notación científica, esto es, 602 214 076 000 000 000 000 000 (seiscientos dos mil doscientos catorce trillones setenta y seis mil billones de unidades, una cifra que aún resulta inferior al número de gotas de agua superficial del planeta Tierra o al número de astros del universo, como se explica posteriormente en este mismo artículo). La constante del eminente físico-químico Avogadro es el factor de proporcionalidad existente entre el número de partículas o entidades elementales y la cantidad de sustancia. Al dividir la cantidad de entidades elementales, cualesquiera que sean, entre la constante de Amedeo Avogadro (1776-1856) se obtiene, en definitiva, la cantidad de sustancia.

¹³ En matemáticas, una **expansión en serie** es un método para calcular una función que no puede ser expresada solamente mediante operadores elementales (suma, resta, multiplicación y división). La *serie* resultante a menudo puede limitarse a un número finito de términos, dando como resultado una aproximación de la función expandida. Cuantos menos términos de la secuencia se usen, más simple resultará esta aproximación. Con frecuencia, la inexactitud resultante (es decir, la serie de los términos omitidos) se puede describir mediante una ecuación que involucra una cota superior asintótica.

$$e^{-n} \left(\sqrt{2\pi} \sqrt{n} + \frac{1}{6} \sqrt{\frac{\pi}{2}} \sqrt{\frac{1}{n}} + \frac{1}{144} \sqrt{\frac{\pi}{2}} \left(\frac{1}{n}\right)^{3/2} - \frac{139 \sqrt{\frac{\pi}{2}} \left(\frac{1}{n}\right)^{5/2}}{25920} - \frac{571 \sqrt{\frac{\pi}{2}} \left(\frac{1}{n}\right)^{7/2}}{1244160} + \frac{163879 \sqrt{\frac{\pi}{2}} \left(\frac{1}{n}\right)^{9/2}}{104509440} + \frac{5246819 \sqrt{\frac{\pi}{2}} \left(\frac{1}{n}\right)^{11/2}}{37623398400} + O\left(\left(\frac{1}{n}\right)^{13/2}\right) \right)$$

Resulta una serie numérica descendente y cumple, como se ve, la condición necesaria, pero no suficiente, de convergencia.

Este mismo problema puede, también, resolverse por aplicación de la fórmula de Stirling, con lo que el término general de la serie quedará así:

$$a_n \approx \frac{e^{-n} \times n^n \times \sqrt{2\pi n}}{n^n} = \frac{\sqrt{2\pi n}}{e^n}.$$

Aplicando, ahora, el primer criterio de Cauchy, quedará:

$$\lim_{n \rightarrow \infty} \sqrt[n]{a_n} = \lim_{n \rightarrow \infty} \frac{\sqrt[2n]{2\pi n}}{e} = \frac{1}{e} \times \lim_{n \rightarrow \infty} \sqrt[2n]{2\pi n} =$$

(aplicando el criterio de Stolz de la raíz, pues es del tipo ∞^0 indet.)

$$\begin{aligned} &= \frac{1}{e} \times \lim_{n \rightarrow \infty} \frac{\sqrt[2n-2(n-1)]{2\pi n}}{\sqrt[2n-2(n-1)]{2\pi(n-1)}} = \\ &= \frac{1}{e} \times \lim_{n \rightarrow \infty} \sqrt{\frac{n}{n-1}} = \frac{1}{e} \times \sqrt{\lim_{n \rightarrow \infty} \frac{n}{n-1}} = \frac{1}{e} \times 1 < 1 \rightarrow \text{CONVERGENTE, c.s.q.d.} \end{aligned}$$

Alternativamente, aplicando el criterio de D'Alembert¹⁴ o del cociente, se tiene:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} &= \lim_{n \rightarrow \infty} \frac{\frac{(n+1)!}{n!}}{\frac{n^n}{n^n}} = \lim_{n \rightarrow \infty} \frac{(n+1)! \cdot n^n}{(n+1)^{n+1} \cdot n!} = \lim_{n \rightarrow \infty} \frac{n^n}{(n+1)^n} = \\ &= \lim_{n \rightarrow \infty} \left(\frac{n}{n+1} \right)^n = \lim_{n \rightarrow \infty} \left(\frac{1}{1 + \frac{1}{n}} \right)^n = \lim_{n \rightarrow \infty} \frac{1}{\left(1 + \frac{1}{n} \right)^n} = \frac{1}{e} < 1, \end{aligned}$$

¹⁴ Jean le Rond d'Alembert (1717- 1783) fue un matemático y filósofo francés. Se le considera uno de los máximos exponentes del movimiento ilustrado, concibiendo las Ciencias como un todo integrado y una herramienta para el progreso de la Humanidad. Abordó la Matemática a través de la Física, con el problema de los tres cuerpos. Esto le llevó a estudiar las ecuaciones diferenciales ordinarias y en derivadas parciales. También probó un criterio de convergencia para series que es el que justamente hemos aplicado aquí.

luego la serie reafirma su carácter CONVERGENTE. Lo mismo podría demostrarse por aplicación de otros diferentes criterios usuales, como el 2º criterio de Cauchy (o del logaritmo), el de Pringsheim, el de Gauss (por comparación de series), el de Raabe-Rolf o el de Cauchy $2^n \cdot a_2^n$ o “condensación” para series decrecientes. En todo caso, tratándose de una serie positiva resulta aplicable el criterio del “test-integral”, a saber:

$$a_n = f(n) = \frac{n!}{n^n}, \text{ y tomamos: } f(x) = \frac{x!}{x^x}.$$

En el intervalo $x > 1$, $f(x) > 0$, y la suma es decreciente. Tomando $\xi = 1$ como límite inferior del intervalo de integración, se tiene:

$$\int_1^{+\infty} f(x) \cdot dx = \lim_{u \rightarrow +\infty} \int_1^u \frac{x!}{x^x} dx = \left[x + \frac{1}{4} x^2 (-2 \ln(x) - 2\gamma + 1) + \right. \\ \left. + \frac{1}{108} x^3 (18 \ln^2(x) + 36\gamma \ln(x) - 12 \ln(x) + 3\pi^2 + 18\gamma^2 - 12\gamma + 4) + O(x^4) \right]_1^{+\infty} = 1.33057 \dots$$

Resolviéndolo como expansión en serie generalizada de Puiseux de la integral en $x = 0$, siendo γ la constante de Euler-Mascheroni, ver nota explicativa al final del presente capítulo (**). Luego, efectivamente, la serie es convergente.

Esto puede comprobarse también mediante la aproximación de Stirling:

$$n! = e^{-n} \cdot n^n \cdot \sqrt{2\pi n}; \text{ de donde se deduce que:}$$

$$f(n) = \frac{n!}{n^n} = \frac{\sqrt{2\pi n}}{e^n}, \text{ o sea: } f(x) = \frac{\sqrt{2\pi x}}{e^x}. \text{ Entonces:}$$

$$\int_1^{+\infty} f(x) \cdot dx = \sqrt{2\pi} \int_1^{+\infty} \frac{\sqrt{x}}{e^x} \cdot dx = \sqrt{2\pi} \left(\frac{1}{e} + 0.1394 \dots \right) = \\ = 2.51(0.3678 \dots + 0.1394 \dots) = 2.51 \times 0.50728 = 1.27327 \dots,$$

que no modifica la conclusión ya obtenida respecto al carácter de la serie estudiada pero que conforma un resultado próximo al anteriormente hallado, con un error absoluto de $E_A = 1.33057 - 1.27327 = 0.0573$, y relativo de:

$$E_R = \frac{0.0573 \times 2}{1.33057 + 1.27327} \times 100 = \frac{0.1146}{2.60384} \times 100 = 4.4\%.$$

Estas diferencias pueden deberse tanto a la de la propia aproximación de Stirling como a las operaciones realizadas del ajuste decimal. Determinando dicho error relativo por el método más tradicional, se tiene:

$$E_R = \frac{0.0573}{1.33057} \times 100 = 4.3\%.$$

Así mismo, utilizando ahora los 5 primeros números primos, se puede también formar la serie descendente: $\{p_i\} = \{p_1, p_2, p_3, p_4, p_5, \dots\}$, o sea:

$$\left\{ \begin{array}{l} p_1 = 2 \Rightarrow \frac{2!}{2^2} = \frac{1}{2} = 0.50000 \\ p_2 = 3 \Rightarrow \frac{3!}{3^3} = \frac{2}{9} = 0.22222 \\ p_3 = 5 \Rightarrow \frac{5!}{5^5} = \frac{120}{3125} = 0.03840 \\ p_4 = 7 \Rightarrow \frac{7!}{7^7} = \frac{5040}{823\,543} = 0.00612 \\ p_5 = 11 \Rightarrow \frac{11!}{11^{11}} = \frac{39\,916\,800}{285\,311\,670\,611} = 0.00014 \\ \dots \end{array} \right.$$

y resulta la serie numérica descendente de términos positivos:

$$0.5, 0.22222, 0.0384, 0.00612, 0.00014, 0.00002, \dots$$

También, en dicho trabajo (Amo *et al*, 2013), se presenta una completa disquisición acerca del carácter de la serie de los recíprocos o inversos de los números primos, que resulta ser divergente. En general, estos autores realizan una exposición recopilatoria de demostraciones debidas a otros investigadores.

De hecho, se tiene la serie formada por los recíprocos de los números primos:

$$\sum_{p_n} \frac{1}{p_n} = \sum_{p_n} \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \frac{1}{11}, \dots \right) = \ln \ln(+\infty) \rightarrow \infty,$$

y Euler llegó a la conclusión que es divergente. Aparte de la demostración que podríamos considerar “original” debida a Euler, existen otras diversas demostraciones que no vamos a reseñar aquí por su considerable extensión y prolijidad (Erdős [9], Mixon, Oresme, Edwards y Penney [8-bis], Dux, Bellman [3], Moser [22], Pinasco, Clarkson [5], ...), exceptuando quizás la basada en la desigualdad de Dusart (1998), que indica que $\forall n \geq 6$, se cumple para los números primos tal que: $p_n \in \{P\}$, $p_n < n \times \log n + n \times \log \log n$, que:

$$\sum_{n=1}^{\infty} \frac{1}{p_n} \geq \sum_{n=6}^{\infty} \frac{1}{p_n} \geq \sum_{n=6}^{\infty} \frac{1}{n \times \log n + n \times \log \log n} = \sum_{n=6}^{\infty} \frac{1}{2n \log n} = \infty, \text{ que diverge.}$$

Aplicando ahora el criterio del test-integral (Cauchy – McLaurin)¹⁵ a la serie de la izquierda, se obtiene: $f(x) = \frac{1}{x}$. En el intervalo $x > 1$, $f(x) > 0$, y la serie es

¹⁵ Esta prueba es del tipo de comparación aunque, en este caso, la comparación tiene lugar entre una integral y una serie, en vez de entre dos series. El criterio de la integral para series nos dice que si tenemos $f(x)$, que es una función continua, positiva y decreciente en el intervalo entre k e ∞ , y $f(n)$ es

decreciente. Entonces, tomando $\xi = 1$ como límite inferior del intervalo de integración, se tiene:

$$\int_1^{+\infty} f(x)dx = \lim_{u \rightarrow +\infty} \int_1^u \frac{dx}{x} = [\ln x]_1^{+\infty} = \infty - 0 = \infty,$$

luego la serie es divergente.

Por último, el criterio de Pringsheim (1850-1941) o del producto también resulta de aplicación al caso, puesto que ofrece:

$$\lim_{p_n \rightarrow \infty} (p_n)^\alpha \times \frac{1}{p_n} = (\text{con } \alpha = 1) = 1 \neq 0,$$

y como resulta que: $\alpha \leq 1$, la serie es divergente, c.s.q.d.

6. ESTADÍSTICA DESCRIPTIVA DE LOS NÚMEROS PRIMOS

6.1. Contenidos en los primeros 200 números naturales

El enfoque estadístico del problema planteado no resulta de aplicación habitual en la Teoría de Números, pero creemos que también puede resultar útil y aportar una visión alternativa o complementaria.

Se presenta y señala la tabla 2 de los 46 números primos contenidos en la muestra de los primeros 200 números naturales. Así:

Tabla 2. Números primos (1 a 200).

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200

igual a a_n ; entonces, primero, si la integral desde k hasta ∞ de $f(x)$ con respecto a x es convergente, la sumatoria desde n igual a k hasta ∞ de a_n , también lo es. Y segundo, si la integral desde k hasta ∞ de $f(x)$ con respecto a x es divergente, la sumatoria desde n igual a uno hasta ∞ de a_n , también lo es.

A continuación, se forman las sumas de 2 números primos para configurar los números pares naturales N del 1 al 200.

Tabla 3. Formación de pares naturales N por adición de 2 números primos (1 a 200).

N	p ₁ +p ₂	N	p ₁ +p ₂	N	p ₁ +p ₂	N	p ₁ +p ₂	N	p ₁ +p ₂
2	--	42	5+37	82	3+79	122	13+109	162	5+157
4	2+2	44	7+37	84	5+79	124	11+113	164	7+157
6	3+3	46	3+43	86	3+83	126	13+113	166	3+163
8	3+5	48	5+43	88	5+83	128	19+109	168	5+163
10	3+7	50	3+47	90	7+83	130	3+127	170	3+167
12	5+7	52	5+47	92	3+89	132	5+127	172	5+167
14	3+11	54	7+47	94	5+89	134	3+131	174	7+167
16	3+13	56	3+53	96	7+89	136	5+131	176	3+173
18	5+13	58	5+53	98	19+79	138	7+131	178	5+173
20	3+17	60	7+53	100	3+97	140	3+137	180	7+173
22	3+19	62	3+59	102	5+97	142	3+139	182	3+179
24	5+19	64	3+61	104	3+101	144	5+139	184	3+181
26	3+23	66	5+61	106	3+103	146	7+139	186	5+181
28	5+23	68	7+61	108	5+103	148	11+137	188	7+181
30	7+23	70	3+67	110	3+107	150	11+139	190	11+179
32	3+29	72	5+67	112	3+109	152	3+149	192	11+181
34	3+31	74	3+71	114	5+109	154	3+151	194	3+191
36	5+31	76	3+73	116	3+113	156	5+151	196	3+193
38	7+31	78	5+73	118	5+113	158	7+151	198	5+193
40	3+37	80	7+73	120	7+113	160	3+157	200	3+197

Los primeros números pares que acabamos de comprobar, no sólo son representables como suma de dos primos, sino que el número de representaciones de N como suma de dos primos parece crecer con N. Esto se debe, entre otras cosas, a que los primos son infinitos, como hemos visto en el apartado anterior 4 de este mismo trabajo.

6.2. Contenidos en los primeros 1000 números naturales

Es posible elaborar la siguiente tabla de los 168 números primos contenidos en la muestra de los primeros 1000 números naturales:

Tabla 4. Distribución de los primos contenidos en los 1000 primeros naturales.

L	p_i	$P_i \uparrow$	$P_i \downarrow$	f_i	$F_i \uparrow$	$F_i \downarrow$	x_i	$x_i \cdot p_i$	x_i^2	$x_i^2 \cdot p_i$
1-100	25	25	143	0.149	0.149	0.851	50	1250	2500	62500
100-200	21	46	122	0.125	0.274	0.726	150	3150	22500	472500
200-300	16	62	106	0.095	0.369	0.631	250	4000	62500	1000000
300-400	16	78	90	0.095	0.464	0.536	350	5600	122500	1960000
400-500	17	95	73	0.101	0.566	0.434	450	7650	202500	3442500
500-600	14	109	59	0.083	0.649	0.351	550	7700	302500	4235000
600-700	16	125	43	0.095	0.744	0.256	650	10400	422500	6760000
700-800	14	139	29	0.083	0.828	0.172	750	10500	562500	7875000
800-900	15	154	14	0.089	0.917	0.083	850	12750	722500	10837500
900-1000	14	168	0	0.083	1.000	0.000	950	13300	902500	12635000
TOTAL Σ	168	--	--	1.000	--	--	--	76300	--	49280000

Seguidamente, se deduce el histograma correspondiente de rectángulos yuxtapuestos en orden a su magnitud decreciente:

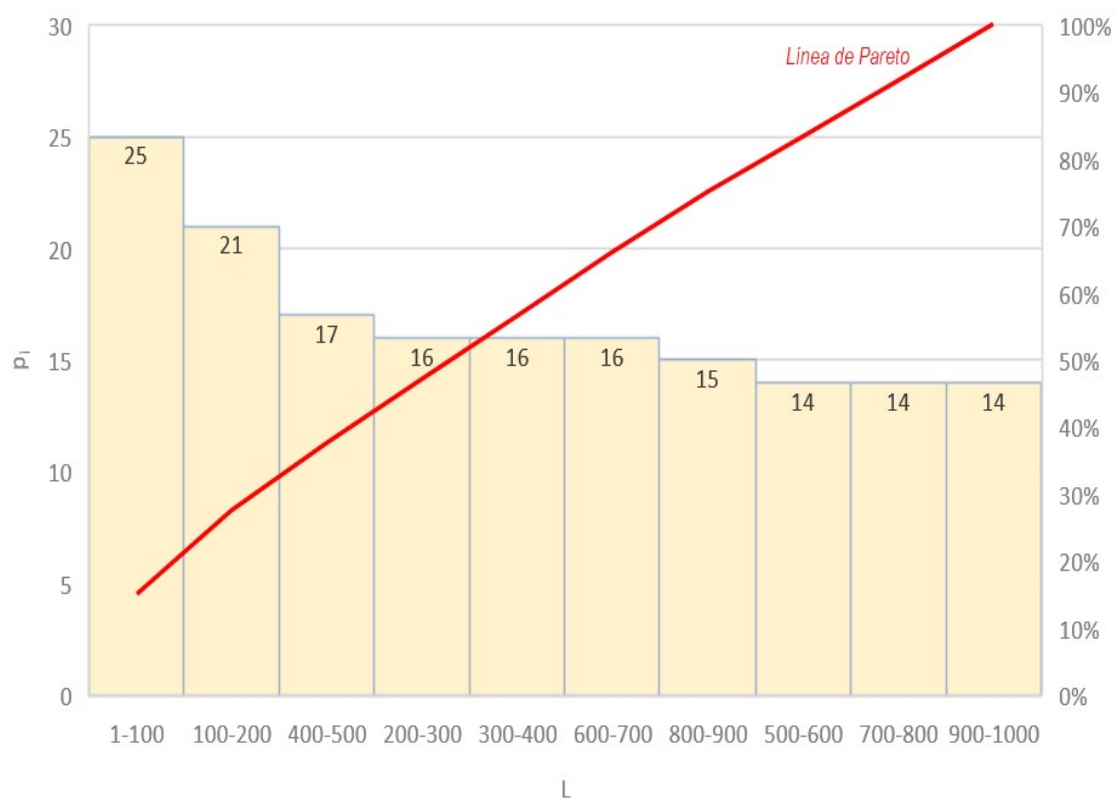


Fig. 2. Histograma de rectángulos yuxtapuestos (1-1000).

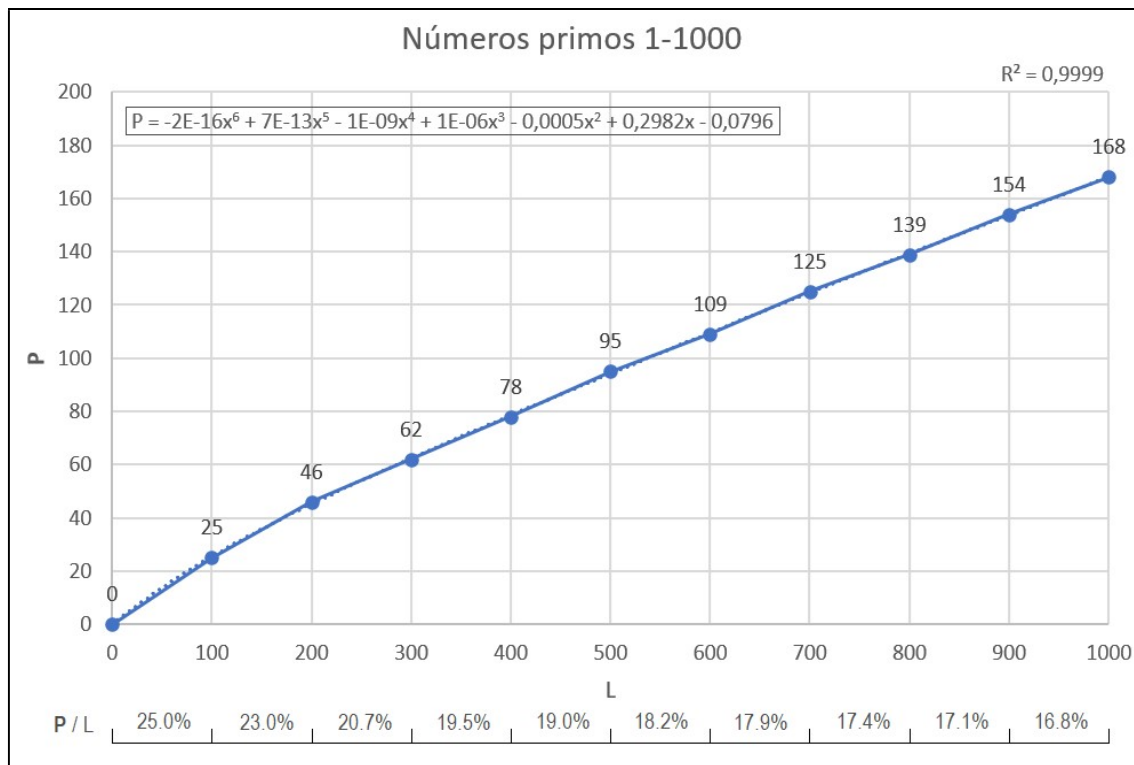


Fig. 3. Diagrama acumulativo ascendente (1-1000).

El ajuste MC a una parábola de 6º grado es prácticamente perfecto, pues el coeficiente de correlación no lineal es $r \cong 1$, aunque por la propia forma de la curva, en este caso, parece más interpretable una regresión lineal. Por otra parte, los coeficientes de las potencias de 2 a 6 son excesivamente pequeños. Esto se parece más a una interpolación (mediante los polinomios de Newton o Lagrange del cálculo numérico) que a la metodología de Estadística y Teoría de la Probabilidad.

La correlación presupone la linealidad entre las variables del problema planteado. Para comprobar la relación existente entre variables no lineales, podemos jerarquizar las variables y utilizar una prueba de correlación lineal. El coeficiente de correlación de Spearman sólo funciona si las variables tienen una relación monótona. Es preferible utilizar la regresión no lineal en lugar de la regresión de mínimos cuadrados ordinarios cuando no se pueda modelar adecuadamente la relación con parámetros lineales (los parámetros son lineales cuando cada término del modelo es aditivo y contiene solo un parámetro que multiplica el término).

De hecho, conviene utilizar Pearson cuando los datos sean continuos, se distribuyan normalmente y se está interesado en evaluar relaciones lineales. Por el contrario, conviene optar por Spearman cuando se trate de datos ordinales, distribuciones no normales o bien si se sospecha la existencia de una asociación no lineal.

6.3. Contenidos en los primeros 10000 números naturales

6.3.1. Introducción

Para configurar la correspondiente tabla de cálculo, tendremos en cuenta el criterio de Sturges (1926) para los 1229 números primos contenidos en la muestra de $M = 10000$ números naturales, con lo que se tendrá un número aproximado de intervalos de clase de:

$$k = 1 + (\log 10000 / \log 2) = 1 + (4 / 0.30103) = 14.29,$$

cifra que redondearemos por defecto a $k = 10$, a efectos operativos, con lo que la amplitud de cada intervalo de clase será de:

$$c = (L_0 - L_{10}) / k = 10000 / 10 = L_i - L_{i-1} = 1000 \text{ números naturales.}$$

Siendo:

- P_i = frecuencia acumulada de números primos.
- p_i = frecuencia simple de números primos.
- L = intervalos de clase de los números naturales.
- x_i = marca de clase.
- f_i = frecuencia relativa simple.
- F_i = frecuencia relativa acumulada.

El resultado viene expresado en la siguiente figura:

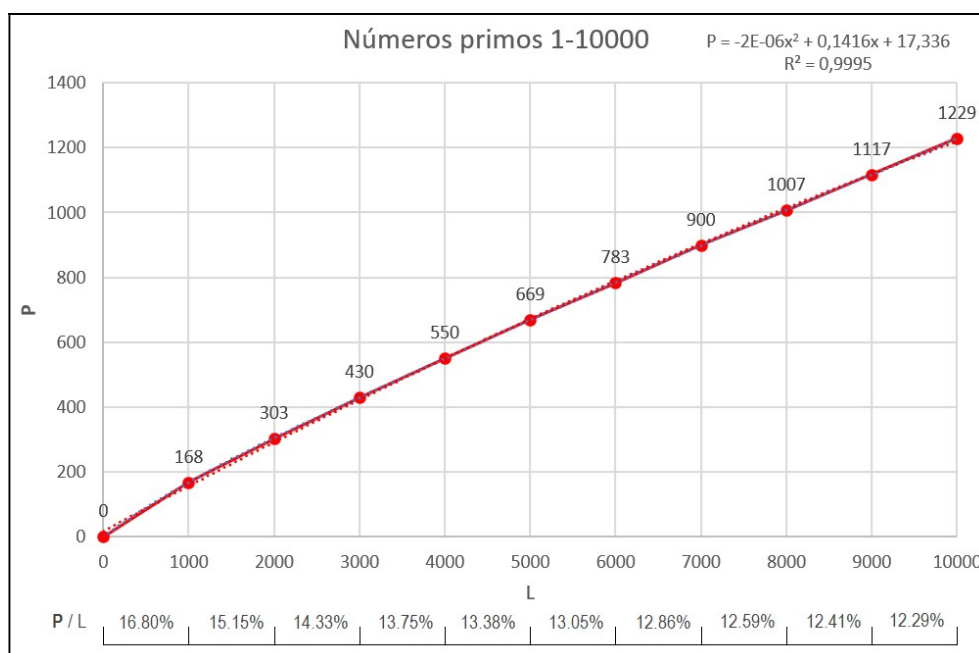


Fig. 4. Diagrama acumulativo ascendente (1-10000).

El ajuste MC a una parábola de 2º grado es prácticamente perfecto, pues el coeficiente de correlación no lineal es $r \cong 1$.

Del mismo modo, se elabora la tabla 5 de los 1229 números primos contenidos en los primeros 10000 números naturales y podríamos seguir elaborando la tabla correspondiente de los 1754 números primos contenidos en los primeros 15000 números naturales, así:

Tabla 5. Distribución de los primos contenidos en los 10000 primeros números naturales.

L	p_i	$p_i \uparrow$	$p_i \downarrow$	f_i	$F_i \uparrow$	$F_i \downarrow$	x_i	$x_i \cdot p_i$	x_i^2	$x_i^2 \cdot p_i$
1-1000	168	168	1061	0.137	0.137	0.863	500	84000	250000	42000000
1001-2000	135	303	926	0.110	0.247	0.753	1500	202500	2250000	303750000
2001-3000	127	430	799	0.103	0.350	0.650	2500	317500	6250000	793750000
3001-4000	120	550	679	0.098	0.448	0.552	3500	420000	12250000	1470000000
4001-5000	119	669	560	0.097	0.544	0.456	4500	535500	20250000	2409750000
5001-6000	114	783	446	0.093	0.637	0.363	5500	627000	30250000	3448500000
6001-7000	117	900	329	0.095	0.732	0.268	6500	760500	42250000	4943250000
7001-8000	107	1007	222	0.087	0.819	0.181	7500	802500	56250000	6018750000
8001-9000	110	1117	112	0.090	0.909	0.091	8500	935000	72250000	7947500000
9001-10000	112	1229	0	0.091	1.000	0.000	9500	1064000	90250000	10108000000
TOTAL Σ	1229	--	--	1.000	--	--	--	5748500	--	37485250000

A continuación, se puede elaborar el histograma correspondiente de rectángulos yuxtapuestos en orden a su magnitud decreciente:

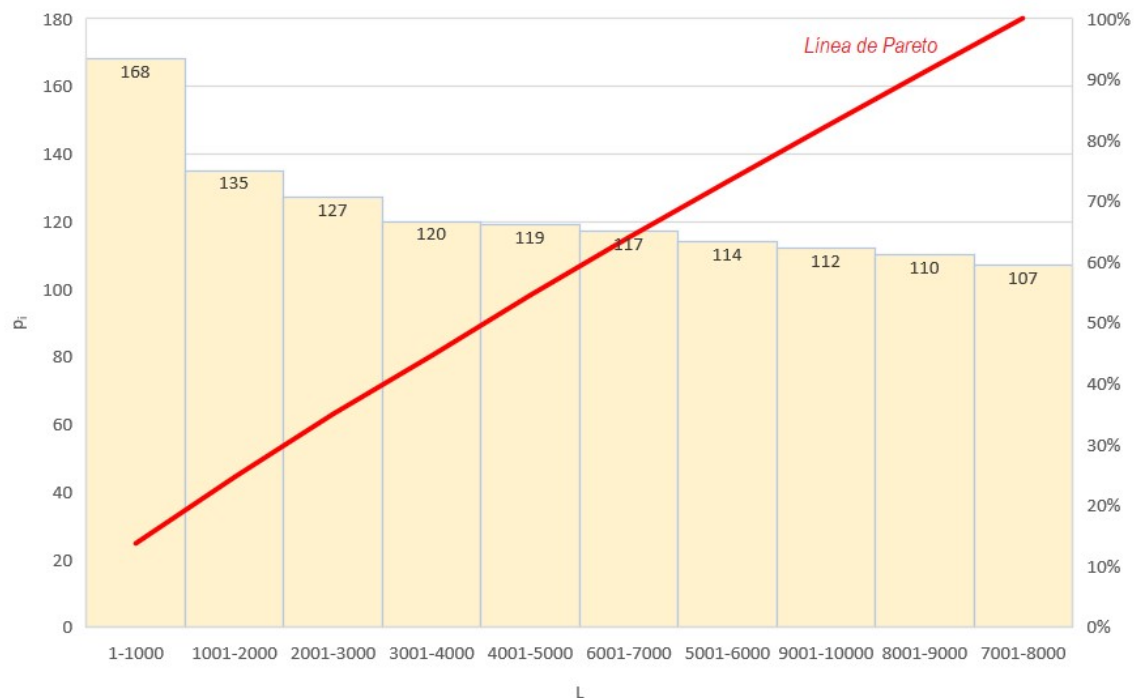


Fig. 5. Histograma de rectángulos yuxtapuestos (1-10000).

Precisamente, en el año 1855 se pudo probar la conjetura fuerte o binaria de Goldbach pero sólo analizando hasta el número natural 10000.

6.3.2. Valores centrales de la distribución de frecuencias

a) Conceptos previos:

Es necesario, con frecuencia, representar un conjunto de datos mediante un solo número que, de tal suerte, se vuelve descriptivo del conjunto entero. El tipo de número escogido dependerá de la característica particular del conjunto que se desee describir. En este apartado de nuestro estudio, se tratará de las denominadas *medidas de situación central*. Por lo que se refiere a la caracterización del valor central de la distribución de los números primos, veamos que G. Udny Yule¹⁶ (Alcaide, 1973), estadístico inglés, en su manual "Introducción a la Teoría de la Estadística", ha precisado las condiciones que debe cumplir también una buena caracterización del valor central de una serie de valores. En resumen, son las siguientes:

a) *La característica del valor central debe ser definida objetivamente a partir de los datos de la serie, sin que haya lugar a intervenir ninguna apreciación subjetiva del estadístico operador.*

b) *Debe depender de todas las observaciones de la serie, a ser posible.* Señalemos que, no obstante, hay veces que se plantea el problema de decidir si debe tenerse en cuenta una observación ("outlier") que es notablemente distinta de todas las demás de su conjunto o bien si puede ser rechazada por considerarse que tal observación tiene carácter excepcional debido a algún factor extraño a la serie como, por ejemplo, un error de observación o de medida. En este sentido, recomendamos la aplicación de los denominados "métodos robustos".

c) *Debe tener, en la medida de lo posible, un sentido concreto, sencillo y fácil de comprender.* Si se tiene en cuenta que muchos de los valores centrales de las series han de ser utilizados por personas poco familiarizadas con la Estadística (los "no matemáticos"), se comprende la preferencia que en la realidad se ha dado a la media aritmética como característica del valor central de que goza esta propiedad, de una interpretación sencilla.

d) *Debe ser de cálculo fácil y rápido,* aunque las modernas herramientas de cálculo que proporciona la Informática, aún con un número elevado de datos a procesar, hacen bastante más innecesaria esta condición.

e) *Debe ser poco sensible a las fluctuaciones del muestreo.* Frecuentemente las observaciones se efectúan, no sobre el conjunto completo o "universo" de elementos a estudiar, sino sobre una parte de éstos que recibe el nombre de "muestra" (en nuestro caso los números primos contenidos en los primeros

¹⁶ George Udny Yule (1871-1951), estadístico escocés con estudios de ingeniería y física, fue un colaborador de Pearson, que hizo algunos aportes a la obra de este último. Trabajó en correlación, y también en curvas asimétricas, como su predecesor. Colaboró en la publicación de Pearson, proporcionando un ejemplo de la aplicación de ajuste de una curva asimétrica a datos sobre distribución de la pobreza en Inglaterra y Gales. Pero luego se movió en direcciones independientes. Relacionó la regresión con el método de los mínimos cuadrados, proporcionando un gran conjunto de algoritmos que habían desarrollado los astrónomos, para la solución de las ecuaciones normales, asociadas al cálculo de la regresión. Los trabajos publicados por Yule cubren hasta la primera década del siglo XX.

10000 números naturales), que debe ser suficientemente representativa de dicho “universo” o “población” y tener el tamaño adecuado. La presente consideración resulta particularmente interesante en el caso del estudio de un número determinado de valores. Las observaciones hechas sobre los elementos componentes de la muestra constituyen la serie estadística de la cual se determina el valor central. Es evidente que “a priori” no resulta que el valor central correspondiente a la muestra adoptada coincida con el valor central que se obtendría si se hiciese una serie estadística que abarcase un conjunto más amplio o incluso todo el conjunto completo de los números primos a estudiar, ni que coincidan siquiera con los correspondientes a distintas muestras que se eligiesen al azar, puesto que conforme aumente el tamaño de la muestra en estudio (primeros 100000, 1000000, ..., números naturales) los resultados de los promedios serán crecientes y distintos. Conviene hacer notar, al respecto, que esta elección del valor central sólo será posible cuando se conozca la ley de distribución del fenómeno en estudio; la variación del valor central y de otros estadísticos en las distintas muestras de números primos entra de lleno en la parte de la Teoría Estadística conocida por la denominación de “Teoría de las Muestras”.

f) *Debe ser adecuada a los cálculos algebraicos posteriores.* Se comprende fácilmente la importancia de tal condición con sólo pensar en el caso muy frecuente de tratar de determinar el valor central que corresponde a una serie global resultado de reunir varias series estadísticas parciales.

De entre las cinco medias clásicas que se emplean (aritmética, geométrica, cuadrática, armónica y cúbica) se ve inmediatamente que la aritmética es la que mejor reúne las anteriores condiciones de Yule, si bien ni ella ni las otras cuatro proporcionan indicación alguna acerca de la repartición de los datos de las series o de sus posiciones respectivas ni sobre las desviaciones de unos respecto a otros. Se limitarán a condensar todos los datos de la serie en uno solo, *la media*, como síntesis de todos ellos.

En particular, debe tenerse en cuenta que las medias aritmética ($\bar{X}$), cuadrática (C) y cúbica (M_3) dan mucho relieve a los elementos grandes de la serie y, desde luego, la segunda y la tercera todavía más que la primera. Por el contrario, las medias geométrica (G), armónica (H) destacan la influencia de los valores pequeños y reducen la influencia de los valores grandes, lo que habrá que tener bien presente en los trabajos de este tipo. Ello se pone claramente de manifiesto al comparar los valores alcanzados por ellas.

Alternativamente, en vez de omitir los datos aislados a los que hemos hecho referencia y calcular una especie de “media modificada” para tratar de evitar la dificultad mencionada, se podría llevar a cabo el estudio mediante otra medida de situación (posición) o valor central llamada *mediana*, que no se “calcula” sino que se “localiza” (la aplicación de la fórmula correspondiente en una distribución de frecuencias agrupadas por intervalos de clase persigue la finalidad de encontrarla, no de calcularla), lo que la distingue conceptualmente de un parámetro. Por definición, en una distribución unitaria de frecuencias, la mediana de un conjunto de datos es el valor del elemento central (si el número de números primos es impar como resulta ser en nuestro caso: $P = 1229$) o

bien la semisuma de los dos elementos centrales (si dicho número es par) cuando dichos datos están ordenados, es decir, dispuestos en orden de magnitud creciente o decreciente.

En el caso de una distribución de frecuencias estructurada por intervalos de clase, como la que aquí planteamos, el cálculo de la mediana viene dado por la fórmula correspondiente cuyo origen no procede especificar aquí. A diferencia de la media aritmética y otras, pues, la mediana no se afecta fácilmente con los valores extremos (Freund & Williams, 1964-1972), por cuya razón se la considera como un estimador robusto de localización.

La mediana, como la media aritmética, siempre existe y es única para cada conjunto de datos. En su contra posee la farragosidad de lo que representa ordenar manualmente grandes conjuntos de datos (aunque el ordenador puede suplir la expresada carencia) y, lo que es peor desde el punto de vista de la inferencia estadística: una mediana de una muestra de números primos no es, en general, una estimación *tan fiable* de una media del conjunto como lo es la media aritmética o esperanza matemática de los mismos datos. De hecho, las medianas de muchas muestras tomadas de la misma población o universo suelen variar más ampliamente que las correspondientes medias de las muestras.

Recordemos, por último, que los promedios o valores centrales de la distribución de probabilidad deben calcularse a partir de datos homogéneos y numerosos, condiciones ambas inherentes a toda buena estadística.

Comenzaremos, como es habitual en estos casos, por el cálculo de los diferentes promedios o valores de la distribución de frecuencias de los números primos contenidos en los primeros 10000 números naturales. Para ello, y a partir de la tabla inicial 5, se obtienen las siguientes medidas de posición:

b) Media aritmética y esperanza matemática:

Hay entre ambas una ligera diferencia de concepto y se aplican en diferentes contextos. La esperanza matemática se calcula de antemano (*ex ante*), considerando la probabilidad de cada resultado, mientras que la media aritmética se obtiene después de observar los resultados obtenidos de la realización de un experimento aleatorio (*ex post*). En esencia, pues, la esperanza matemática es una media teórica (*a priori*), mientras que la media aritmética resulta ser una media empírica (*a posteriori*).

En nuestro caso, la suma de los 10000 primeros números naturales, que conforman una progresión aritmética de diferencia 1, es:

$$S_N = \frac{10000(1+10000)}{2} = 50\,005\,000 ,$$

y la media aritmética de dichos números naturales será, pues,

$$\bar{X} = S_N/10000 = 5000.5.$$

Además, la media aritmética de los $P = 1229$ números primos contenidos en estos primeros $M = 10000$ números naturales es:

$$\bar{X} = \frac{5736396}{1229} = 4667.5.$$

Sin embargo, el cálculo de esta media aritmética teniendo en cuenta los intervalos de clase definidos en la tabla 5 anterior, arrojaría alternativamente un valor de:

$$\bar{X} = \frac{\sum_{i=1}^{10} x_i \cdot p_i}{P} = \frac{5748500}{1229} = 4677.4,$$

con una pequeña diferencia en relación a la determinación anterior que es propia del agrupamiento en clases efectuado.

c) Media cuadrática (RMS):

Por otra parte, la media cuadrática de esta distribución de frecuencias viene dada por:

$$C = \sqrt{\frac{\sum_{i=1}^{10} x_i^2 \cdot p_i}{P}} = \sqrt{\frac{37485250000}{1229}} = 5522.7.$$

d) Moda:

Por lo que se refiere a la “moda” (unimodal), que se hallará contenida en el intervalo de clase de mayor frecuencia simple u ordinaria, se tiene:

$$Mo = L_{i-1} + \frac{p_{i+1}}{p_{i-1} + p_{i+1}} \times c_i = 0 + \frac{135}{0 + 135} \times 1000 = 1000.$$

e) Mediana o segundo cuartil:

Se hallará en el intervalo de clase que contiene a:

$$p_i^{\uparrow} = \frac{P}{2} = \frac{1229}{2} = 614.5.$$

$$Me = Q_2 = L_{i-1} + \frac{\frac{P}{2} - p_{i-1}^{\uparrow}}{p_i} \times c_i = 4001 + \frac{614.5 - 550}{119} \times 1000 = 4543,$$

que se halla situada en el punto de corte de los diagramas acumulativos ascendente y descendente, por definición. Atendiendo, no obstante, a la distribución unitaria de los números primos que incluimos en el Anexo del presente trabajo, le corresponde un valor primo $Me = 4523$, debiéndose la pequeña diferencia observada al hecho de haberse efectuado el cálculo

partiendo de una distribución conjunta de frecuencias (particionada por intervalos de clase).

Se cumple: $C = 5522.7 > \bar{X} = 4677.4 > Me = 4543 > Mo = 1000$.

f) Media geométrica:

$$G = \text{anti log} \left(\sum_{i=1}^{10} p_i \cdot \log x_i / P \right) = \text{anti log} \frac{4338.557}{1229} = 3389.6$$

g) Media armónica:

$$H = \frac{P}{\sum_{i=1}^{10} \frac{p_i}{x_i}} = \frac{1229}{0.615} = 1998.4$$

Desde luego, las cuatro medias clásicas aquí estudiadas, además de la media cúbica calculada posteriormente, que constituyen estimadores de posición o valores centrales de la distribución de frecuencias, como no podría ser de otra manera desde el punto de vista teórico, quedan ordenadas monotónicamente con arreglo a su magnitud, del modo siguiente:

armónica < geométrica < aritmética < cuadrática < cúbica

$(H = 1998.4) < (G = 3389.6) < (\bar{X} = 4677.4) < (C = 5522.7) < (M_3 = 6094.9)$

y puede observarse que dichos valores se compadecen en su magnitud con la construcción geométrica de la Fig. 11.

h) Primer cuartil:

Se hallará en el intervalo de clase que contiene a:

$$p_i^{\uparrow} = \frac{P}{4} = \frac{1229}{4} = 307.25$$

$$Q_1 = L_{i-1} + \frac{\frac{P}{4} - p_{i-1}^{\uparrow}}{p_i} \times c_i = 2001 + \frac{307.25 - 303}{127} \times 1000 = 2035,$$

aunque la distribución unitaria ofrece el valor del número primo 2029.

i) Tercer cuartil:

Se hallará en el intervalo de clase que contiene a:

$$p_i^{\uparrow} = \frac{3 \cdot P}{4} = \frac{3687}{4} = 921.75$$

$$Q_3 = L_{i-1} + \frac{\frac{3P}{4} - p_{i-1}^{\uparrow}}{p_i} \times c_i = 7001 + \frac{921.75 - 900}{107} \times 1000 = 7204 ,$$

aunque la distribución unitaria ofrece el valor del número primo 7219.

j) Trimedia (Tukey):

Las medidas robustas de tendencia central son estadísticos que tratan de paliar los problemas de estimación asociados a distribuciones anómalas. Cabe señalar, entre ellas, las denominadas medias “k-recortadas” o “k-winsorizadas” así como la “trimedia” que, en nuestro caso, vendrá dada por la expresión:

$$T = \frac{Q_1 + 2Q_2 + Q_3}{4} = \frac{2035 + 2 \times 4543 + 7204}{4} = 4581 ,$$

que en el caso de la distribución unitaria de los números primos ofrece un valor:

$$T = \frac{Q_1 + 2Q_2 + Q_3}{4} = \frac{2029 + 2 \times 4523 + 7219}{4} = 4573.5 ,$$

que varía ligeramente en relación al resultado anteriormente obtenido por la agrupación efectuada de las frecuencias en intervalos de clase.

La trimedia, en definitiva, es una aproximación a la media que se usa mucho en meta análisis, suponiendo que la población es finita.

k) 1º tercil:

Se hallará en el intervalo de clase que contiene a:

$$p_i^{\uparrow} = P / 3 = 1229 / 3 = 409.67$$

$$T_1 = L_{i-1} + \frac{\frac{P}{3} - p_{i-1}^{\uparrow}}{p_i} \times c_i = 2001 + \frac{409.67 - 303}{127} \times 1000 = 2841$$

l) 2º tercil:

Se hallará en el intervalo de clase que contiene a:

$$p_i^{\uparrow} = 2P / 3 = 819.33$$

$$T_2 = L_{i-1} + \frac{\frac{2P}{3} - p_{i-1}^{\uparrow}}{p_i} \times c_i = 6001 + \frac{819.33 - 783}{117} \times 1000 = 6311.5$$

De este modo, la distribución de los primeros 10000 números naturales y primos por sus cuantiles (cuartiles y terciles) es la siguiente:

	$Q_0 = T_0$	Q_1	T_1	$Q_2 = M_e$	T_2	Q_3	$Q_4 = T_3$	
1		2035	2841	4543	6311	7204	10000	⇒ Núm. natural
(2)		(2029)	(2819)	(4523)	(6301)	(7219)	(9973)	⇒ Núm. primo corresp.

Fig. 6. Distribución y valor de los cuantiles (1-10000).

m) Mediana de Gastwirth (1966):

Dentro de la Estadística robusta, cabe considerar aquí también este promedio (Franquet, 2019) que ofrece más información muestral que la mediana y es más resistente que la trimedia a la influencia de *outliers*, por lo que constituye una alternativa válida a ellas como parámetro de posición. Viene dada por la expresión:

$$M_{eg} = \frac{4M_e + 3T_1 + 3T_2}{10} = \frac{4 \times 4543 + 3 \times 2841 + 3 \times 6311.5}{10} = 4563,$$

cuyo valor apenas cambia tomando en consideración los valores de los parámetros dimanantes de la distribución unitaria de frecuencias como puede comprobarse en la tabla 1 del Anexo de información complementaria, o sea:

$$M_{eg} = \frac{4M_e + 3T_1 + 3T_2}{10} = \frac{4 \times 4523 + 3 \times 2819 + 3 \times 6301}{10} = 4545.2.$$

n) Media cúbica:

Vendrá dada por:

$$M_3 = \sqrt[3]{\frac{\sum_{i=1}^{10} x_i^3 \times p_i}{P}} = \sqrt[3]{\frac{2.78268 \times 10^{14}}{1229}} = 6094.9$$

6.3.3. Medidas de dispersión

Las medidas de dispersión, también llamadas “medidas de variabilidad”, muestran precisamente la variabilidad de una distribución indicando, por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de un promedio o no. Cuanto mayor sea ese valor, mayor será la variabilidad; por el contrario, cuanto menor sea, más homogénea será a la media. Así, se sabe si todos los casos son parecidos o bien varían mucho entre ellos.

Para calcular la variabilidad que la distribución de los números primos tiene respecto de su media, se calcula la media de las desviaciones de las puntuaciones respecto a la media aritmética. Pero la suma de esas desviaciones es siempre cero (por una conocida propiedad de la media

aritmética), así que se adoptan dos clases de estrategias para salvar este problema. Una de ellas es tomando las desviaciones en valor absoluto (obteniéndose la denominada “desviación media”) y la otra es tomando las desviaciones al cuadrado (obteniéndose la “varianza” y su raíz cuadrada, la “desviación típica”, “desviación standard” o “desviación cuadrática media”).

Por lo que se refiere a la desviación típica o “standard” σ como medida de dispersión absoluta de la distribución de los números primos contenidos en los 10000 primeros números naturales, veamos que su valor viene dado por la expresión:

$$\sigma = \sqrt{C^2 - \bar{X}^2} = \sqrt{5522.7^2 - 4677.4^2} = 2936.3$$

y procede realizar la corrección por agrupamientos en clases o intervalos de amplitud $c = 1000$.

En este sentido, la **corrección de Sheppard** es una corrección que se propone para el cálculo de momentos estadísticos muestrales cuando estos se calculan para datos agrupados en intervalos a partir de las marcas de clase o puntos medios de cada intervalo. La corrección de Sheppard más conocida es la que corresponde a la varianza o momento central de orden 2. Se utiliza en distribuciones continuas donde las colas "colas" van gradualmente a cero en ambas direcciones.

Aquí, la corrección Sheppard para la determinación de la σ_c más ajustada de los datos del problema arroja un valor de la varianza corregida de:

$$\begin{aligned}\sigma_c^2 &= \sigma^2 - \frac{c^2}{12} = 5522.7^2 - 4677.4^2 - \frac{1000^2}{12} = \\ &= 8\,622\,144.5 - 83\,333.3 = 8\,538\,811.2 ; \text{ de donde: } \sigma_c = 2922.1,\end{aligned}$$

que lógicamente resulta ser de alguna menor cuantía que en la primera determinación efectuada. Ello podría obligar a efectuar una ligera revisión de los cálculos anteriores en los que se haya hecho intervenir a la expresada medida de dispersión de los valores de la variable aleatoria estadística utilizada.

En su caso, podría también llevarse a cabo la comprobación de Charlier denominada así en homenaje a su inventor, el astrónomo sueco Carl Charlier (1862-1934), que es una comprobación de la realización correcta de los cálculos efectuados de la varianza muestral cuando ésta se ha calculado para un conjunto de datos agrupados en una distribución de frecuencias a través de la computación de una nueva columna de valores a partir de la correspondiente tabla de frecuencias.

Para ello, estableceremos la siguiente tabla auxiliar de cálculo:

Tabla 6. Comprobación de Charlier.

L	x_i	p_i	$x_i \cdot p_i$	x_i^2	$p_i \cdot x_i^2$	$(x_i + 1)^2$	$p_i \cdot (x_i + 1)^2$
1-1000	500	168	84000	250000	42000000	251001	42168168
1001-2000	1500	135	202500	2250000	303750000	2253001	304155135
2001-3000	2500	127	317500	6250000	793750000	6255001	794385127
3001-4000	3500	120	420000	12250000	1470000000	12257001	1470840120
4001-5000	4500	119	535500	20250000	2409750000	20259001	2410821119
5001-6000	5500	114	627000	30250000	3448500000	30261001	3449754114
6001-7000	6500	117	760500	42250000	4943250000	42263001	4944771117
7001-8000	7500	107	802500	56250000	6018750000	56265001	6020355107
8001-9000	8500	110	935000	72250000	7947500000	72267001	7949370110
9001-10000	9500	112	1064000	90250000	10108000000	90269001	10110128112
TOTAL Σ	50000	1229	5748500	--	37485250000	--	37496748229

Aplicamos, a continuación, la comprobación de Charlier (1931) para dar por buenas las sumas calculadas en las tablas precedentes:

$$\sum_{i=1}^{10} p(x_i + 1)^2 = \sum_{i=1}^{10} p_i \cdot x_i^2 + 2 \sum_{i=1}^{10} p_i \cdot x_i + \sum_{i=1}^{10} p_i =$$

$$= 37485250000 + 2 \times 5748500 + 1229 = 37496748229.$$

La igualdad es correcta, por lo que las sumas efectuadas para el cálculo de la varianza también lo son.

De ahí se deduce, como medida de dispersión relativa, un coeficiente de variación de Pearson:

$$CV = \frac{\sigma_c}{\bar{X}} \times 100 = \frac{2922.1}{4677.4} \times 100 = 62.47\%$$

y un coeficiente de uniformidad de: $100 (1 - CV) = 100 (1 - 0.6247) = 37.53 \%$.

Otras medidas de interés son las siguientes:

- Coeficiente de apertura: $C_{ap} = \frac{9973}{2} = 4986.5$.
- Recorrido o rango: entre los valores extremos de los números primos contenidos en el subconjunto de los números naturales del 1 al 10000, o sea:

$$R = 9973 - 2 = 9971.$$

- Recorrido relativo: $R' = \frac{R}{\bar{X}} = \frac{9971}{4677.4} = 2.13$.
- Recorrido intercuartílico: $R'' = Q_3 - Q_1 = 7204 - 2035 = 5169$.

- Recorrido semi-intercuartílico:

$$R''' = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{5169}{7204 + 2035} = 0.60.$$

Veamos, así mismo, que la tabla 7 auxiliar de cálculo, en su última columna, nos permitirá la determinación de otra medida de dispersión absoluta a la que ya nos hemos referido con anterioridad: la desviación media absoluta¹⁷ (DM) respecto a la media aritmética (que sería mínima con respecto a la mediana, sin tener en cuenta la moda), que resulta ser la media de las desviaciones de las observaciones con respecto a la media aritmética, donde la desviación debe tomarse en valor absoluto o módulo, a saber:

$$DM = \frac{\sum_{i=1}^{10} |x_i - \bar{X}| \cdot p_i}{p} = \frac{3139334}{1229} = 2.554.4 \approx 65\sigma / 7 = 6 \times 2922.1 / 7 = 2504.7$$

6.3.4. Asimetría o sesgo

Las medidas de *asimetría* o *sesgo* son indicadores que permiten establecer el *grado de simetría* (o asimetría) que presenta la distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica. Como eje de simetría consideramos una recta paralela al eje de ordenadas que pasa por la media aritmética de la distribución. Si una distribución es perfectamente simétrica, existe el mismo número de valores a la derecha que a la izquierda de la media o esperanza matemática, y por lo tanto, el mismo número de desviaciones con signo positivo que con signo negativo.

Decimos que hay asimetría positiva (o a la derecha) si la "cola" a la derecha de la media es más larga que la de la izquierda, es decir, si hay valores más separados de la media a la derecha. Por el contrario, diremos que hay asimetría negativa (o a la izquierda) si la "cola" a la izquierda de la media es más larga que la de la derecha, es decir, si hay valores más separados de la media a la izquierda.

Se podría pensar que definir la simetría usando la mediana para variables continuas y usando la media aritmética para variables discretas deviene una elección arbitraria. En realidad, esto no es así, pues si una variable es continua, coinciden ambos criterios de simetría (con respecto a la media aritmética y a la mediana). Es más, se tiene que media y mediana coinciden para las distribuciones continuas perfectamente simétricas. Por otro lado, en el caso de

¹⁷ Por cierto, que la DM puede usarse también para estimar la desviación típica σ , y es fácil de calcular en pequeñas muestras si las desviaciones se calculan con respecto a la mediana Me . Las eficiencias del recorrido R y de la DM son menores que en estimaciones más sencillas de calcular que la DM. Para muestras de los tamaños comprendidos entre 5 y 10, estas estimaciones utilizan la diferencia existente entre la suma de los dos valores mayores y la suma de los dos valores menores (Dixon y Massey, 1969). Para muestras de tamaños más grandes, como es el caso de la empleada en este trabajo sobre los números primos, se podrían usar tres o cuatro de los valores mayores y de los más pequeños.

variables discretas, la distribución es simétrica si el lado derecho del diagrama se obtiene por una imagen especular desde la media. En este caso coincide la media con la mediana si el número de observaciones es impar. Si la variable es continua simétrica y unimodal, coinciden también la media aritmética, la mediana y la moda.

Dentro de los diversos tipos de asimetría posible, vamos a destacar los dos fundamentales:

-Asimetría positiva:

Si las frecuencias más altas se encuentran en el lado izquierdo de la media, mientras que en el derecho hay frecuencias más pequeñas (*cola*).

-Asimetría negativa:

Cuando la cola está situada en el lado izquierdo, al contrario de lo que sucede en el caso anterior.

Cuando realizamos un estudio descriptivo como el que aquí se desarrolla en relación a la distribución de los números primos, resulta altamente improbable que la distribución de frecuencias sea *totalmente* simétrica. En la práctica diremos que la distribución de frecuencias es simétrica si lo es sólo de un modo aproximado. Por otro lado, observando cuidadosamente la gráfica de la distribución (Fig. 5), podemos ver claro de qué lado (izquierdo) están las frecuencias más altas. Conviene definir, entonces, unos estadísticos que ayuden a interpretar la asimetría, a los que llamaremos *índices de asimetría*.

A continuación, emplearemos, en el caso que estamos desarrollando en el presente trabajo, algunos de los índices de asimetría o sesgo más usuales, como son el índice basado en los tres cuartiles, el momento de tercer orden¹⁸ y la distancia existente entre la moda y la media o la media y la mediana.

Por lo que se refiere a las restantes características de la distribución de los números primos, veamos que una de la asimetría o sesgo la constituye el denominado "1r. coeficiente de asimetría de Pearson", dado por:

$$P_1 = \frac{\bar{X} - M_0}{\sigma_c} = \frac{4677.4 - 1000}{2922.1} = 1.26 > 0,$$

luego se trata de una asimetría a la derecha (positiva).

Así mismo, el 2º coeficiente de asimetría de Pearson ofrece:

$$P_2 = \frac{3(\bar{X} - M_e)}{\sigma_c} = \frac{3(4677.4 - 4543)}{2922.1} = 0.14 > 0,$$

¹⁸ El conocimiento de la simetría permite precisar, de alguna manera, la forma de una distribución de frecuencias, pero tal conocimiento puede mejorarse notoriamente al disponer de otras características de dicha distribución y, de modo general, mediante los denominados "momentos de la distribución", que pueden ser con respecto al origen (de los que son conocidos la media aritmética y el cuadrado de la media cuadrática C^2) o bien con respecto a la media (como la varianza σ^2).

que confirma el resultado anteriormente obtenido, o bien el “coeficiente de sesgo cuartílico”, también conocido como “coeficiente de asimetría de Yule-Bowley”, de valor:

$$P_3 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} =$$

$$= \frac{7204 - 2 \times 4543 + 2035}{5169} = 0.30 > 0,$$

que conduce, en todos los casos, a conclusiones similares.

Así mismo, a partir de los resultados obtenidos de la muestra se elaborará la siguiente tabla:

Tabla 7. Tabla de cálculos auxiliares (1-10000).

L	p _i	x _i	p _i /x _i	log x _i	p _i · log x _i	x _i -X	(x _i -X) ³ ·p _i	(x _i -X) ⁴ ·p _i	x _i -X · p _i
1-1000	168	500	0.336	2.699	453.427	-4177.4	-1.22468E+13	5.11594E+16	701799.8
1001-2000	135	1500	0.090	3.176	428.772	-3177.4	-4.33053E+12	1.37597E+16	428946.3
2001-3000	127	2500	0.051	3.398	431.538	-2177.4	-1.31101E+12	2.85457E+15	276527.3
3001-4000	120	3500	0.034	3.544	425.288	-1177.4	-1.95853E+11	2.30594E+14	141285.6
4001-5000	119	4500	0.026	3.653	434.732	-177.38	-664141950.2	1.17805E+11	21108.22
5001-6000	114	5500	0.021	3.740	426.401	822.62	63460376399	5.22038E+13	93778.68
6001-7000	117	6500	0.018	3.813	446.111	1822.6	7.08393E+11	1.29113E+15	213246.5
7001-8000	107	7500	0.014	3.875	414.632	2822.6	2.40625E+12	6.79193E+15	302020.3
8001-9000	110	8500	0.013	3.929	432.236	3822.6	6.14435E+12	2.34875E+16	420488.2
9001-10000	112	9500	0.012	3.978	445.505	4822.6	1.25622E+13	6.05829E+16	540133.4
TOTAL Σ	1229	50000	0.615	--	4338.643	3226.2(*)	3.79988E+12	1.6021E+17	3139334

(*) Este valor sería 0 de haber considerado una $\bar{X} = 50000/10 = 5000$ por la 1ª propiedad de la media aritmética.

De los resultados que se deducen de la tabla anterior, se infiere inmediatamente que el momento central (respecto a la media aritmética) de tercer orden es:

$$m_3 = \frac{\sum_{i=1}^{10} (x_i - \bar{X})^3 \cdot p_i}{P} = \frac{3.79988 \times 10^{12}}{1229} = 3\,091\,847\,030,$$

por lo que se tendrá un "coeficiente directo de asimetría" o "coeficiente de sesgo" de Fisher (Alcaide, 1973) de:

$$g_1 = \frac{m_3}{\sigma_c^3} = \frac{3091847030}{2922.1^3} = 0.12 > 0,$$

esto es, moderadamente asimétrica, lo que confirma los resultados anteriores.

6.3.5. Apuntamiento o curtosis

Por lo que se refiere al apuntamiento o “curtosis”, veamos que en la teoría de la probabilidad y estadística, la *curtosis* (o “kurtosis”) es una medida de la forma o apuntamiento de las distribuciones. Así, las medidas de curtosis (también llamadas de apuntamiento o de concentración central) tratan de estudiar la mayor o menor concentración de frecuencias existentes alrededor de la media aritmética y en la zona central de la distribución. De este modo, los índices de curtosis miden la mayor o menor concentración de los datos justamente alrededor de la media aritmética, partiendo como referencia de la distribución normal o curva campaniforme de Gauss. Aquí, utilizando el momento central o respecto al origen de 4º orden y a partir de los resultados obtenidos en la tabla 7 anterior, se tiene:

$$m_4 = \frac{\sum_{i=1}^{10} (x_i - \bar{X})^4 \cdot p_i}{P} = \frac{1.6021 \times 10^{17}}{1229} = 130\,358\,014\,646\,054$$

pese a que, aplicando la correspondiente corrección de Sheppard para la curtosis, este valor quedaría escasamente reducido a:

$$m_{4c} = m_4 - (c^2/2) \times m_2 + (7/240) \times c^4 = 126\,076\,252\,312\,721,$$

por lo que se tendrá, en definitiva, un coeficiente de curtosis de Fisher de:

$$g_2 = \frac{m_4}{\sigma_c^4} - 3 = \frac{130358014646054}{2922.1^4} - 3 = -1.21 < 0,$$

y lo mismo sucede considerando los nuevos valores de m_{4c} y m_{2c} una vez efectuada la corrección propugnada, luego se trata de una distribución algo platicúrtica o achatada.

Este coeficiente se aplica a distribuciones unimodales simétricas o moderadamente asimétricas, para medir la curtosis o apuntamiento de la distribución de frecuencias; más específicamente, de la función masa de probabilidad o la función de densidad. Mide el apuntamiento de la distribución de la variable aleatoria estadística en relación a una distribución patrón, denominada “normal estándar”, para la que el coeficiente es nulo. En el caso que nos ocupa, pues, tiene lugar un menor apuntamiento que el propio de la campana de Gauss.

6.3.6. Medidas de concentración

Por último, la curva de Lorenz es la línea poligonal quebrada que une los puntos $(0, 0)$, (q_1, p_1) , (q_2, p_2) , (q_3, p_3) , ..., (q_k, p_k) . En el caso planteado, tenemos que $k = 10$ (número de clases).

Otra forma de observar la curva de Lorenz es estimando el área de la superficie que se encuentra comprendida entre la curva y la diagonal del primer cuadrante del círculo (la recta $P_i = Q_i$). Esa superficie se denomina **área de**

concentración. El **índice de Gini** constituye un índice de concentración de los valores de la variable aleatoria estadística y equivale justamente al doble del área de concentración. Su valor estará siempre comprendido entre los valores cero y uno ($0 \leq I_G \leq 1$).

Se considera, en definitiva, que existe equidistribución de los números primos cuando $P_i = Q_i$, y en este caso, la expresada curva de Lorenz adoptaría la configuración gráfica siguiente:

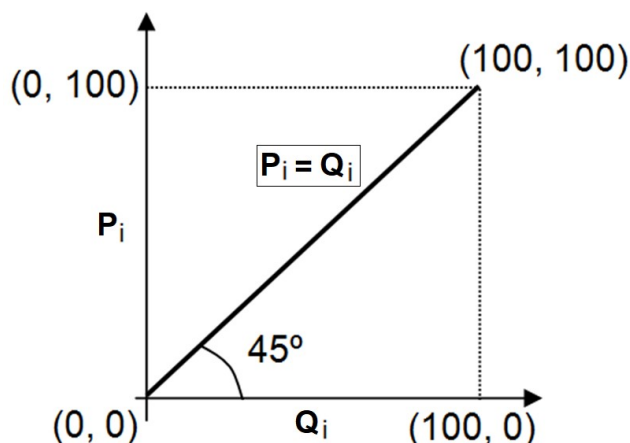


Fig. 7. Coordenadas cartesianas rectangulares de las variables p y q .

o sea, se trata de una recta que descansa teóricamente sobre la bisectriz del primer cuadrante del círculo ($P_i = Q_i$). En nuestro caso, la tabla 8 de cálculos auxiliares para la determinación del índice de Gini I_G será la siguiente:

Tabla 8. Auxiliar de cálculo del Índice de Gini.

L_i	p_i	$f_i = p_i / \sum p_i$	$f_i \cdot 100$	q_i	$(q_i / \sum q_i) \cdot 100$	P_i	Q_i	$P_i - Q_i$
1-1000	168	0.137	13.67	1000	10.00	13.67	10.00	3.67
1001-2000	135	0.110	10.98	1000	10.00	24.65	20.00	4.65
2001-3000	127	0.103	10.33	1000	10.00	34.99	30.00	4.99
3001-4000	120	0.098	9.76	1000	10.00	44.75	40.00	4.75
4001-5000	119	0.097	9.68	1000	10.00	54.43	50.00	4.43
5001-6000	114	0.093	9.28	1000	10.00	63.71	60.00	3.71
6001-7000	117	0.095	9.52	1000	10.00	73.23	70.00	3.23
7001-8000	107	0.087	8.71	1000	10.00	81.94	80.00	1.94
8001-9000	110	0.090	8.95	1000	10.00	90.89	90.00	0.89
9001-10000	112	0.091	9.11	1000	10.00	100.00	100.00	0.00
TOTAL Σ	1229	1	100.00	10000	100.00	582.26	550.00	32.26

Y se obtiene un índice de Gini, según la fórmula dada por Pulido (1971), de:

$$I_G = \frac{\sum_{i=1}^{k-1} (P_i - Q_i)}{P_i} = \frac{32.26}{482.26} = 0.06689 = 6.69\%$$

Con ello ya es posible el trazado de la curva poligonal de Lorenz, así:

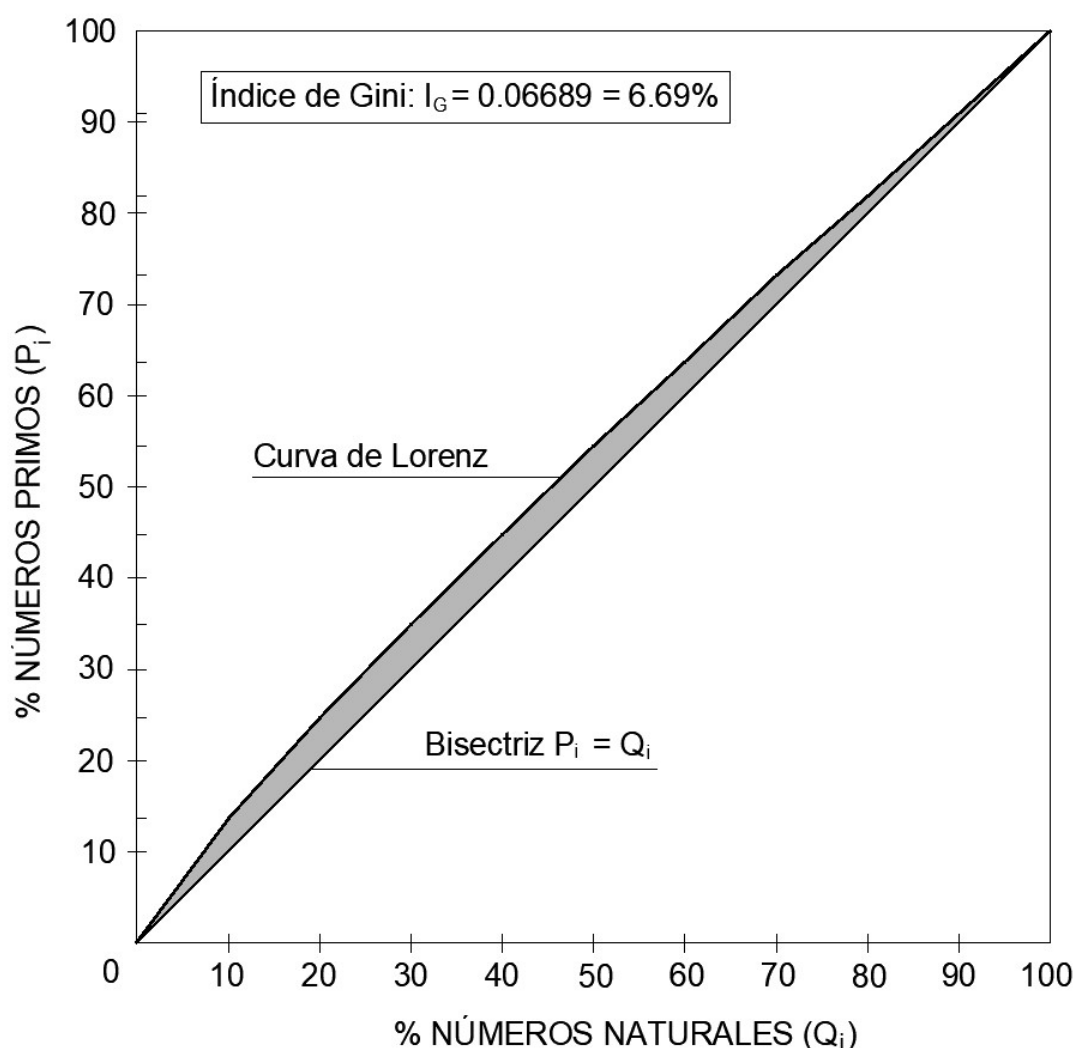


Fig. 8. Índice de Gini y curva de Lorenz.

Así pues, a medida que aumenta el porcentaje de números naturales investigados, el porcentaje de los números primos disminuye, tal como puede apreciarse en la Fig. 9 mediante un simple ajuste polinómico de 6º grado. Frente al 12.29% que hemos visto hasta ahora, en los primeros 100000 números naturales sólo hay 9592 números primos, es decir, el 9.59%. Si continuamos el proceso, encontraremos que de 1 000 000 de números naturales, sólo el 7.845% son primos. En 10 000 000 hay 664579 lo que disminuye el porcentaje al 6.65%. En los primeros 100 000 000 números naturales hay 5 761 455 primos (5.76%), en los primeros 10^9 naturales hay una cantidad de 50 847 534 primos (5.08%), en los primeros 10^{10} naturales hay una cantidad de 455 052 511 primos (4.55%) y en los primeros 10^{11} naturales sólo hay 4 118 054 813 primos (4.12%), o sea, con el porcentaje *in decrescendo*.

Si continuáramos buscando así indefinidamente el porcentaje de números primos se acercaría al 0% aunque nunca alcanzaría esta cifra, puesto que, como hemos visto en el anterior apartado 4, los números primos son infinitos.

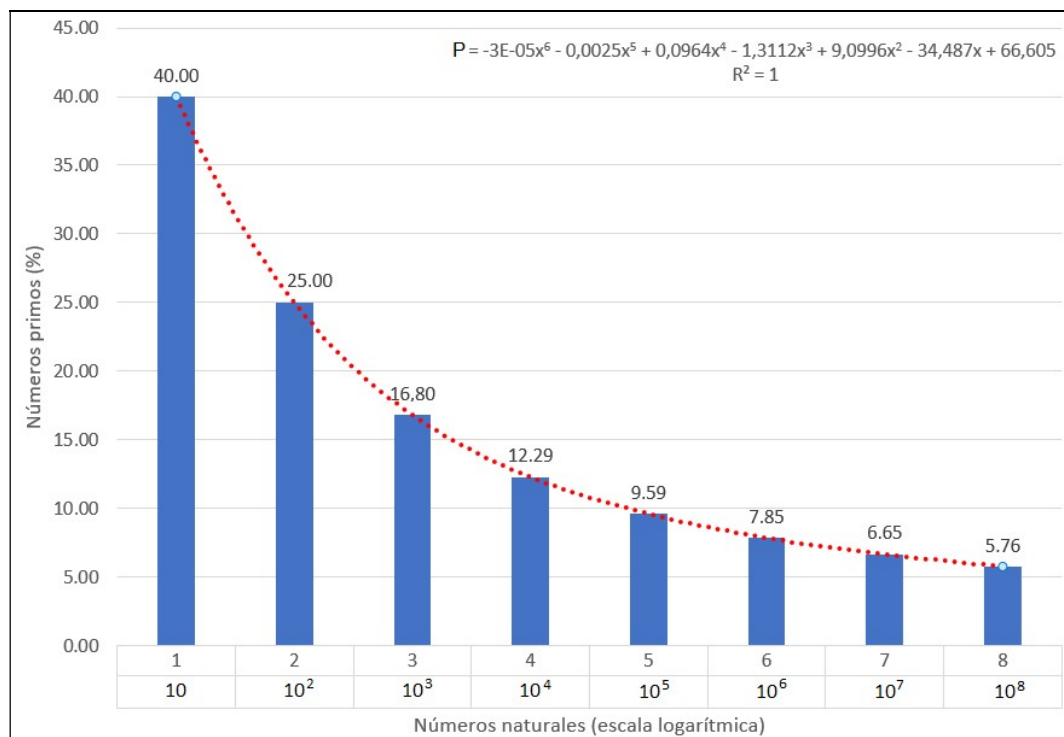


Fig. 9. Números primos contenidos en los primeros 10^8 números naturales.

Se observa, en definitiva, que el porcentaje acumulado de números primos, en la curva poligonal de Lorenz, siempre de halla por encima del porcentaje acumulado de los números naturales, lo que sucederá en todas las muestras que se quieran estudiar a partir del origen, cuestión ésta que confirma la ligera asimetría a la derecha (positiva) de la correspondiente distribución de frecuencias que ya se ha puesto de manifiesto en el epígrafe anterior.

La distribución de los números primos es un asunto reiterativo de investigación en la teoría de números: si se consideran números aisladamente, los primos parecieran estar distribuidos de modo probabilístico, pero la distribución «global» de los números primos se ajusta a leyes bien definidas¹⁹.

Otros índices que se pueden considerar aplicables al caso que nos ocupa son el de Williamson y el de concentración de Lorenz, que se obtiene (este último) mediante la aplicación de la fórmula basada en los porcentajes acumulados, empleados ambos comúnmente en los trabajos prácticos.

La verificación de todos estos resultados, que ha resultado plenamente satisfactoria, se ha realizado con la antigua aplicación estadística "R

¹⁹ Los números primos están bien distribuidos, en el sentido de que, para $n \geq 2$, existe al menos un número primo comprendido entre n y $2n$. Se ha demostrado que, en general, los números primos están más dispersos a medida que se tiende al infinito. Los números primos cercanos entre sí tienden a evitar la repetición de sus últimos dígitos, es decir, un número primo que termina en 1 es menos probable que sea seguido por otro que termina en 1 de lo que cabría esperar de una secuencia aleatoria. Así lo explica Kannan Soundararajan, de la Universidad de Stanford (California), que informó del descubrimiento con su colega Robert Lemke Oliver. A pesar de que los números primos se utilizan en una serie de aplicaciones, tales como la criptografía, este sesgo 'anti-igualdad' no tiene ningún uso práctico o incluso ninguna implicación más amplia de la teoría de números.

commander” y actual “r Studio”²⁰. Además, cuando se emplea esta última hay que señalar qué paquetes de la misma se han usado.

6.3.7. Medida de independencia

La prueba que se desarrolla a continuación se utiliza para determinar si existe una diferencia estadísticamente significativa entre la frecuencia esperada (q_i) y las frecuencias observadas (p_i) en una o más categorías de una tabla de contingencia. Para tablas de contingencia con tamaños de muestra más pequeños, se utiliza en su lugar una prueba exacta de Fisher.

Por otra parte, utilizando la prueba de Pearson χ^2 (o chi-cuadrado) con $k = 10$ categorías o intervalos de clase, cabe preguntarse si las observaciones de la proporción de números primos existentes para cada categoría (p_i) están o no de acuerdo con los valores teóricos correspondientes para cada una de ellas, o sea, la proporción de números naturales (q_i) de cada clase. Es decir, hay que contrastar la hipótesis de independencia que establece los valores de las frecuencias teóricas (q_i). Con ello, el estadístico usado es el siguiente:

$$\chi^2 = \sum_{i=1}^{10} \frac{(p_i - q_i)^2}{q_i} = 1.902,$$

que es el valor resultante de la tabla siguiente:

Tabla 9. Auxiliar de cálculo del estadígrafo χ^2 (1-10000).

L_i	Frecuencia relativa (%)		$d_i = p_i - q_i $	$(p_i - q_i)^2$	$(p_i - q_i)^2/q_i$	$(p_i - q_i - 0.5)^2$	$(p_i - q_i - 0.5)^2/q_i$
	p_i	q_i					
1-1000	13.670	10.00	3.670	13.466	1.347	10.047	1.005
1001-2000	10.985	10.00	0.985	0.969	0.097	0.235	0.023
2001-3000	10.334	10.00	0.334	0.111	0.011	0.028	0.003
3001-4000	9.764	10.00	0.236	0.056	0.006	0.070	0.007
4001-5000	9.683	10.00	0.317	0.101	0.010	0.033	0.003
5001-6000	9.276	10.00	0.724	0.524	0.052	0.050	0.005
6001-7000	9.520	10.00	0.480	0.230	0.023	0.000	0.000
7001-8000	8.706	10.00	1.294	1.674	0.167	0.630	0.063
8001-9000	8.950	10.00	1.050	1.102	0.110	0.302	0.030
9001-10000	9.113	10.00	0.887	0.787	0.079	0.150	0.015
TOTAL Σ	100	100	0.000	19.020	$\chi^2 = 1.902$	11.545	$\chi_c^2 = 1.154$

²⁰ Se trata de una interfaz gráfica de usuario para el lenguaje de programación R, liberada bajo licencia GNU GPL y desarrollada y mantenida por el Departamento de Sociología de la Universidad McMaster. Es quizás la alternativa más viable basada en R a los paquetes estadísticos usuales como SPSS. Con cada opción de menú utilizada presenta el código R que se ha ejecutado. También conocida como **Rcmdr**, esta aplicación puede ser instalada desde R como un paquete más de este lenguaje. Ha sido utilizada como un entorno de aprendizaje en numerosos cursos y libros de estadística para estudiantes, tecnólogos y científicos.

Alternativamente, también la χ^2 puede calcularse así:

$$\chi^2 = \sum_{i=1}^{10} \frac{p_i^2}{q_i} - N, \text{ siendo:}$$

$$\sum_{i=1}^{10} \frac{p_i^2}{q_i} = 18.69 + 12.07 + 10.68 + 9.53 + 9.38 +$$

$$+ 8.60 + 9.06 + 7.58 + 8.01 + 8.30 = 101.902,$$

con lo que: $\chi^2 = 101.902 - 100 = 1.902$, que confirma plenamente el resultado obtenido en la tabla 9.

Como puede verse, el estadístico χ^2 tiene aquí: $n = k - 1 = 10 - 1 = 9$ grados de libertad (g.l.)²¹. El valor crítico correspondiente, con el nivel de significación del 5% ($P = 0.950$)²², es de $\chi^2_{9,0.95} = 16.92 > 1.902$, según puede comprobarse en la tabla 2 del Anexo de Información Complementaria y, por lo tanto, no existe razón suficiente (al no ser el valor observado suficientemente grande en cuanto a sus diferencias o discrepancias con el valor teórico) como para rechazar la hipótesis teórica o alternativa (q_i), aceptándose la hipótesis nula, por lo menos en la muestra estudiada de los números primos contenidos en los primeros 10000 números naturales (tabla 1 del mencionado Anexo).

Por otra parte, cuando se aplican a datos discretos los resultados para distribuciones continuas debe practicarse la corrección de Yates y comparar los resultados obtenidos. Si ambos valores conducen a la misma conclusión para un cierto nivel de significación, raramente se presentan dificultades. Si conducen a conclusiones diferentes se puede, o bien incrementar los tamaños muestrales o, si esto no es posible, se pueden emplear métodos de probabilidad exactos de acuerdo con la distribución multinomial²³ (Spiegel, 1981).

Esta última se basa en que si los sucesos $E_1, E_2, \dots, E_k$, pueden ocurrir con probabilidades respectivas: $p_1, p_2, \dots, p_k$, entonces resulta la probabilidad de que: $E_1, E_2, \dots, E_k$, ocurran $X_1, X_2, \dots, X_k$, veces respectivamente, viene dada por la expresión:

²¹ El número de grados de libertad de un estadístico se define como el número de observaciones independientes de la muestra (tamaño muestral) menos el número de parámetros de la población o universo que deben estimarse. A mayor complejidad o amplitud del problema mayor será también el n° de g.l.

²² El nivel de significación es el límite para juzgar un resultado como estadísticamente significativo. Si el valor obtenido es menor que el nivel de significación, se considera que el resultado es estadísticamente significativo. El nivel de significación también se conoce como el nivel α .

²³ En teoría de la probabilidad, la distribución multinomial o distribución multinómica es una generalización de la distribución binomial. La distribución binomial es la probabilidad de un número de éxitos en N sucesos de Bernoulli independientes, con la misma probabilidad de éxito en cada suceso. En una distribución multinomial, el análogo a la distribución de Bernoulli es la distribución categórica.

$$(N! / X_1! \cdot X_2! \cdot \dots \cdot X_k!) \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k} = \frac{N!}{X_1! \cdot X_2! \cdot \dots \cdot X_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k} = \frac{N! \cdot \prod_{i=1}^k p_i^{x_i}}{\prod_{i=1}^k X_i!},$$

donde: $X_1 + X_2 + \dots + X_k = \sum_{i=1}^k X_i = N$.

Esta distribución teórica de probabilidad, que constituye una generalización de la conocida distribución binomial²⁴, se llama *distribución multinomial* que ha sido relacionada, ya que la expresión anterior es el término general del desarrollo multinomial: $(p_1 + p_2 + \dots + p_k)^N$. Los números teóricos a veces para que ocurran los sucesos: $E_1, E_2, \dots, E_k$, en N repeticiones, son los siguientes: $Np_1, Np_2, \dots, Np_k$, respectivamente.

El “coeficiente de contingencia” C de Pearson deriva del estadístico anterior, y mide el grado de interdependencia (relación, asociación) entre 2 atributos. A mayor valor de C mayor es el grado de asociación entre ambos. En este caso, la χ^2 corregida ofrece un valor según la tabla 9:

$$\chi^2 = \frac{(|13.67 - 10| - 0.5)^2}{10} + \frac{(|10.98 - 10| - 0.5)^2}{10} + \dots + \frac{(|9.11 - 10| - 0.5)^2}{10} = 1.154,$$

con un “coeficiente de contingencia” de: $C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{1.154}{1.154 + 100}} = 0.107$.

El campo de variabilidad de dicho coeficiente C vendrá dado por:

- Si $\chi^2 \rightarrow 0 \Rightarrow \lim_{\substack{d_i \rightarrow 0 \\ \chi^2 \rightarrow 0}} \sqrt{\frac{\chi^2}{\chi^2 + n}} = 0$; si $d_i \rightarrow 0 \Rightarrow$ existe independencia entre los atributos analizados.
- Si $\chi^2 \rightarrow \infty \Rightarrow \lim_{\substack{d_i \rightarrow \infty \\ \chi^2 \rightarrow \infty}} \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\lim_{\chi^2 \rightarrow \infty} \frac{1}{1 + \frac{n}{\chi^2}}} = 1$, en este caso $d_i \rightarrow \infty \Rightarrow$

existe dependencia entre los atributos, o sea, C : $0 \leq C \leq 1$, o bien, $C \in [0, 1]$.

d_i = son las discrepancias o diferencias existentes entre los valores reales y los teóricos o esperados, de tal modo que se pueden producir las situaciones extremas relacionadas.

²⁴ Una distribución binomial, en Estadística, es una distribución de probabilidad discreta (función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra) que describe el número de éxitos al realizar n experimentos o ensayos de independientes entre sí, acerca de una variable aleatoria. La distribución binomial se define como una serie de experimentos o ensayos en los que sólo podemos tener 2 posibles resultados (éxito o fracaso), siendo el éxito la variable aleatoria.

En cualquier caso, para poder excluir -además del efecto del tamaño de la muestra- la influencia de la dimensión de la tabla (el número de formas de la característica) sobre el límite superior del C y asegurar así la comparabilidad de los resultados, se utiliza el C corregido del siguiente modo:

$$C(\text{corregida}) = \sqrt{\frac{k}{k-1}} \times C = \sqrt{\frac{2}{1}} \times 0.107 = 0.151$$

cuyo valor indica que se trata de características independientes.

También pueden emplearse otros coeficientes como el de Cramers V (que es una medida simétrica para la intensidad de la relación entre dos o más variables de la escala nominal cuando, por lo menos, una de las dos variables tiene dos formas), el Phi ϕ (también llamado coeficiente de correlación de cuatro campos) o la contingencia cuadrática media $\chi^2/2$ (cuanto mayor es esta medida, tanto más intensa es la relación existente entre las dos características analizadas), conduciendo todos ellos a conclusiones similares.

7. MUESTREO ESTRATIFICADO

7.1. Objeto

Una manera alternativa de proceder consiste en realizar un muestreo estratificado sobre el conjunto de los números naturales, debiendo determinarse qué parte de la muestra ha de extraerse de cada estrato, lo que puede llevarse a efecto mediante cuatro procedimientos generales de afijación: uniforme, proporcional, óptima y valoral. Así, por ejemplo, en el caso que nos ocupa, con una población de $N = 10000$ números naturales que se divide a conveniencia en 3 subpoblaciones o estratos de tamaños $N_1 = 5000$, $N_2 = 3000$ y $N_3 = 2000$ elementos, se desea, una vez realizada la estimación pertinente, extraer una muestra del 10% ($n = 1000$ elementos o números naturales) dividida en tres submuestras de tamaños respectivos n_1 , n_2 y n_3 , lo que se lleva a cabo por los procedimientos indicados. Puede verse, al respecto, la Tabla 6 con la distinción colorimétrica de las tres submuestras inicialmente consideradas.

7.2. Afijación uniforme

Vendrá dada simplemente por:

$$n_1 = n_2 = n_3 = n/3 = 1000/3 \cong 333 \text{ números naturales.}$$

7.3. Afijación proporcional

Vendrá dada por:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{n_1 + n_2 + n_3}{N_1 + N_2 + N_3} = \frac{n}{N}; \text{ o sea:}$$

$$\frac{n_1}{5000} = \frac{n_2}{3000} = \frac{n_3}{2000} = \frac{1000}{10000}; \text{ de donde:}$$

$$n_1 = 500; \quad n_2 = 300; \quad n_3 = 200.$$

7.4. Afijación óptima

Vendrá dada por:

$$\frac{n_1}{N_1 \cdot \sigma_1} = \frac{n_2}{N_2 \cdot \sigma_2} = \frac{n_3}{N_3 \cdot \sigma_3} = \frac{n}{\sum_{i=1}^3 N_i \cdot \sigma_i},$$

para lo que hay que conocer previamente los valores de las desviaciones típicas σ_1 , σ_2 y σ_3 de cada estrato, lo cual puede realizarse mediante la extracción de muestras-piloto lo que tiende a compensar la posible mayor heterogeneidad de un estrato determinado mediante un incremento en el tamaño de la submuestra correspondiente. En este caso, suponiendo que:

$$\begin{cases} \sigma_1 = 1500 \text{ (con } CV_1 = 1500/2500 = 0.60) \\ \sigma_2 = 1000 \text{ (con } CV_2 = 1000/1500 = 0.66) \\ \sigma_3 = 1500 \text{ (con } CV_3 = 700/1000 = 0.70) \end{cases}$$

Resultarían entonces:

$$\frac{n_1}{5000 \times 1500} = \frac{n_2}{3000 \times 1000} = \frac{n_3}{2000 \times 1500} = \frac{1000}{13500000},$$

de donde: $n_1 = 556$, $n_2 = 222$ y $n_3 = 222$.

Se observa claramente que la mayor variabilidad (CV) del estrato N_3 obliga a incrementar el tamaño de la submuestra correspondiente n_3 en relación al caso de la afijación proporcional.

No obstante la determinación anterior basada en un simple supuesto, se han calculado específicamente las desviaciones típicas de tres submuestras o estratos más homogéneos en cuanto a su respectivo tamaño, o sea: $N_1 = 4000$, $N_2 = 3000$ y $N_3 = 3000$ elementos. Se desea, una vez realizada la estimación pertinente, extraer una muestra con un nivel de confianza del 95% ($z = 1.96$) y un margen de error del 2%, con $p = 0.12$ y $q = 0.88$. Ello es así porque la proporción de elementos que poseen -en la población- la característica de ser números primos es del 12.29% (ver tabla 8) y su complementario es del 87.71%. La muestra a obtener se hallará dividida en tres submuestras de tamaños respectivos n_1 , n_2 y n_3 , lo que se lleva a cabo por los procedimientos indicados.

Su tamaño necesario será, para el caso de una población finita:

$$n = \frac{N \times z^2 \times p \times q}{(N-1) \times e^2 + z^2 \times p \times q} = \frac{10000 \times 1.96^2 \times 0.12 \times 0.88}{9999 \times 0.02^2 + 1.96^2 \times 0.12 \times 0.88} = 921.$$

El tamaño requerido de la muestra, con estos condicionantes, es pues de 921 elementos, que redondeamos a 1000 a efectos prácticos y por razones de seguridad (aplicando el principio de prudencia), habiéndose obtenido los siguientes resultados de las desviaciones típicas y medias aritméticas de cada submuestra, que ofrecen una versión más real del problema planteado, y pueden observarse en la siguiente tabla con los resultados obtenidos de las 3 submuestras.

Tabla 10. Características de la distribución de los números primos (del 1 al 10000)

Submuestra	Desv. típica σ	Media aritm. $\bar{X}$	Coef. Variac. CV	Tamaño
1	1306	1862	0.70	n_1
2	3016	5494	0.55	n_2
3	4639	8515	0.55	n_3
Población	2922.1	4677.4	0.62	N

En este caso real deberá realizarse una afijación óptima, en relación a la variabilidad (CV) de cada submuestra. Al tratarse de medias aritméticas diferentes para cada submuestra, procede realizar dicha afijación óptima en base a los respectivos coeficientes de variación de Pearson, como medida relativa de la dispersión de la variable aleatoria estadística, con lo que:

$$\frac{n_1}{N_1 \times CV_1} = \frac{n_2}{N_2 \times CV_2} = \frac{n_3}{N_3 \times CV_3} = \frac{n}{\sum_{i=1}^3 N_i \times CV_i},$$

y entonces:

$$\frac{n_1}{4000 \times 0.7} = \frac{n_2}{3000 \times 0.55} = \frac{n_3}{3000 \times 0.55} = \frac{1000}{6100}, \text{ de donde resultan:}$$

$$n_1 = 459; \quad n_2 = 270; \quad n_3 = 271.$$

7.5. Afijación valoral

En algunos casos se podrían introducir -de conocerse o calcularse previamente- factores económicos de coste c_1 , c_2 y c_3 para la obtención de cada submuestra, cuestión ésta que entendemos no procede emplear en el supuesto aquí desarrollado.

8. MECANISMO GENERADOR DE NÚMEROS PRIMOS

Por lo que se refiere a las funciones generadoras (o mejor rastreadoras) de números primos, en general, veamos que Owens (2008) afirma que un número primo es un entero positivo el cual tiene exactamente dos divisores positivos distintos. Por otro lado Prieto (2013) añade que un número natural es primo si tiene exactamente dos divisores distintos: 1 y él mismo.

Según Rowland (2008), persiste la creencia de que los números primos están distribuidos de manera aleatoria y esa intuición ha llevado tradicionalmente a los matemáticos a despertar su interés sobre aquellas funciones que rastrean números primos de manera confiable. Xiao (2015), comenta que las evidencias del conocimiento sobre número primos se remontan a la civilización egipcia, y desde entonces se han propuesto diversas fórmulas que generan sólo algunos números primos. Ribenboim (1995) menciona que hay tres clases de funciones generadoras de números primos y que las podemos englobar en tres clasificaciones principales, a saber:

- Una función que produce al n -ésimo número primo $f(n) = P_n$.
- Una función $f(n)$ que genera siempre un primo, y $f(n) \neq f(m)$ para $m \neq n$.
- El conjunto de valores positivos de f es igual al conjunto de los números primos.

Si definimos ahora la denominada “función contador” $\pi(x)$ como el número de números primos menores o iguales que x , el teorema del número primo afirma que: $\pi(x) \sim x/\log x$. En otras palabras, la probabilidad de que un entero de tamaño n sea primo es aproximadamente $1/\log n$. Esta función cuenta el número de números primos menores o iguales a cierto número real x . Se denota mediante $\pi(x)$, no debiendo confundirse con el famoso número irracional π , y analíticamente se define como:

$$\pi(x) = \{p \in P / p \leq x\}.$$

Chen and Wang (1989) consiguieron substituir la imprecisa expresión “suficientemente grande” por “todo impar mayor que 3.33×10^{43000} ”. Sin embargo, este número es todavía demasiado grande como para comprobar si los impares anteriores son representables como suma de tres primos. J.M. Deshouillers, G. Effinger, H. Riele y D. Zinoviev, asumiendo la Hipótesis de Riemann generalizada, han conseguido rebajar este número hasta 2×10^{12} , que resulta más accesible a las técnicas de computación actuales (Cilleruelo, 2013).

Uno de los primeros intentos se plasmó en la fórmula de Pierre de Fermat (1601-1665) para rastrear números primos, a saber: $F_n = 2^{2^n} + 1$, con n natural (los cinco primeros términos, 3 ($n=0$), 5 ($n=1$), 17 ($n=2$), 257 ($n=3$) y 65537 ($n=4$) lo son), pero Euler probó que no era así en el año 1732. En efecto, al considerar $n = 5$ se obtiene un número compuesto: 4294967297, que es el número más pequeño que, siendo un número de Fermat, no es primo.

Una fórmula de los números primos sería aquella que generase los números primos, exactamente y sin excepción alguna. Otra gran cuestión es qué se considera como una «fórmula»²⁵ y lo que no. No existe, hoy en día, ninguna fórmula polinómica para obtener todos los números primos. Puede objetarse

²⁵ En la ciencia, se denominan “fórmulas” al conjunto de operaciones básicas que, mediante símbolos y valores numéricos diversos, permiten resolver operaciones de cálculo y obtener resultados concretos a resultados de su aplicación.

que el término «fórmula» se restringe solamente a los polinomios. ¿Podrían usarse sumatorias, factoriales y la función piso? Si así fuera, de hecho, sí existen fórmulas para obtener números primos. Una interpretación razonable de la palabra «fórmula» es la de «una máquina de Turing²⁶ que se detiene bajo todas las entradas». Bajo esta interpretación, ciertamente existen máquinas de Turing que se detienen, capaces de computar el enésimo número primo. Aún así, no se sabe cómo calcular el enésimo número primo en tiempo polinómico, esto es, todavía no se conoce alguna fórmula fácilmente computable.

Existen varias fórmulas polinómicas que generan o rastrean *todos* los números primos, y *solo* los números primos. En particular, la propuesta por Jones *et al.* en 1976 resulta, probablemente, el ejemplo más conocido. Por supuesto, la deducción de este polinomio es complicada. Se basa en un teorema de la década anterior que reza que cualquier conjunto definido de forma recursivamente numerable, como el de los números primos, se puede describir de forma polinomial. En su día, la publicación de un teorema de esta magnitud alcanzó una gran repercusión, pues cerraba siglos de investigación matemática.

No obstante, aunque la fórmula constituye un éxito de la Teoría de Números, no tiene utilidad práctica total, puesto que no permite saber, de antemano, si el resultado va a ser positivo (y primo) o negativo (e inútil), existiendo, por cierto, muchas más posibilidades de que salga negativo, aunque reconozcamos que, a nivel teórico, la fórmula resulte muy impactante. De esta forma, el problema de generar números primos sigue abierto.

Para números relativamente pequeños, es posible simplemente aplicar la división de prueba a cada número impar sucesivo. Los tamices primarios casi siempre son más rápidos. El tamizado de números primos es la forma más rápida conocida de enumerar, de manera determinista, los números primos. Existen algunas fórmulas conocidas que pueden calcular el siguiente número primo, pero no se conoce una forma de expresarlo en términos de los números primos anteriores. Además, no se conoce ninguna manipulación general efectiva y/o extensión de alguna expresión matemática (incluso aquellas que incluyan números primos posteriores) que calcule de manera determinista el siguiente número primo.

Si seguimos aplicando la teoría de los números algebraicos, se puede mostrar un resultado aún mayor: no existe una función polinómica no constante $P(n)$ que evalúe a un número primo para casi todos los enteros n . Así, el polinomio cuadrático de Euler: $P(n) = n^2 + n + 41$, devuelve algunos números primos para todos los enteros no negativos < 40 . Los números primos para $n = 0, 1, 2, 3, \dots$ son, respectivamente: 41, 43, 47, 53... Las diferencias entre los términos son:

²⁶ La máquina de Turing es un modelo matemático que simula el funcionamiento de un computador mediante una cinta infinita y una cabeza lectora-escritora. Es capaz de realizar cualquier operación computacional si se le da un conjunto de reglas, lo que la convierte en un concepto clave en la teoría de la computación. Una manera fácil de entenderlo es imaginar una cinta de papel (como las que usan las antiguas calculadoras) que puede moverse hacia la izquierda o la derecha, y sobre la cual se escriben y se leen símbolos. Este proceso está controlado por una serie de reglas que indican qué hacer en cada situación.

2, 4, 6, 8, 10,... Para $n = 40$, se produce un número cuadrado, 1681, el cual es igual a 41×41 , o sea, constituye el menor número compuesto para esta fórmula. De hecho, si 41 divide a n , también divide a $P(n)$. El expresado fenómeno se relaciona con la conocida “espiral de Ulam”²⁷, la cual también es implícitamente cuadrática. Hay muchas constantes enteras, b y c , tales que la función: $f(n) = 4n^2 + bn + c$ genera, a medida que crece n a lo largo de los números naturales $\{1, 2, 3, \dots\}$, una gran cantidad de números primos en comparación con la proporción de primos existente en números de magnitud similar. Este hallazgo fue tan célebre que la espiral de Ulam apareció en la portada de la revista *Scientific American* en marzo de 1964. Existen otras variantes de la espiral de Ulam, tales como la espiral de Sacks, que también muestran patrones sin explicación aparente.

Cierto es que mediante interpolación podemos construir un polinomio $P(n)$ que *genere* los números primos que queramos a partir de los valores que elijamos (por ejemplo, un polinomio que dé unos ciertos números primos concretos para n , desde 0 a 1000), pero posiblemente el grado del mismo nos resulte enorme y con unos coeficientes muy complicados. Lo más interesante del expresado *polinomio de Euler* es, sin duda, su bajo grado, 2, así como sus sencillos coeficientes.

Existen también otros polinomios, de expresión muy sencilla, que generan una aceptable cantidad de números primos. Uno de ellos, por ejemplo, es el de 2º grado: $P(n) = n^2 - n + 41$, que es prácticamente igual al anterior...y que en realidad no aporta mucha más información al asunto que ahora mismo nos ocupa. Ofrece primos para n de 0 a 40 (es decir, uno más que el polinomio anterior), pero de los naturales 1 a 40 salen los mismos que en el caso anterior y el que da de más es repetido. Concretamente es el 41, que aparece para n igual a 0 y a 1. También existen otros polinomios mucho más complejos que estos, tanto en grado como en coeficientes. Y cabe preguntarse al respecto, ¿habrá algún polinomio que dé siempre números primos? Pues no, no lo hay. El propio Goldbach demostró que, con coeficientes enteros, no es posible encontrar un polinomio que dé números primos para todo número natural, y más tarde Legendre demostró lo mismo para las funciones algebraicas racionales.

Volviendo al polinomio inicial de Euler, $n^2 + n + 41$, tiene cierto interés ver qué ocurre con dicho polinomio, en el sentido de *generar* primos, si cambiamos el término independiente 41 por otro número. ¿Para qué números tendremos un polinomio *parecido*? Concretamente, buscamos los números enteros positivos t tales que $n^2 + n + t$ genera números primos desde $n = 0$ hasta $n = t - 2$. Se sabe que eso solamente ocurre con 6 números enteros, que son precisamente los siguientes: 2, 3, 5, 11, 17, 41, que se denominan *números afortunados de Euler*

²⁷ La Espiral de Ulam, un descubrimiento realizado por el matemático polaco Stanislaw Marcin Ulam (1909-1984) en 1963, representa un fascinante puente entre la curiosidad humana y el poder de la visualización matemática. Este fenómeno surgió de un momento aparentemente trivial, mientras Ulam hacía garabatos durante una reunión. Sin embargo, lo que emergió de este acto casual fue una herramienta capaz de revelar patrones ocultos en los números primos, generando una nueva forma de percibir la distribución de estos números fundamentales que son objeto de nuestro estudio.

(Euler's lucky numbers). Todo esto se puede ver reflejado en la tabla 11 y figura 10 siguientes:

Tabla 11. Generación de números primos para $0 \leq n \leq 39$.

n	p	n	p	n	p	n	p	n	p	n	p
0	41		241		487		751		1031		1303
1	43	14	251		491		757	31	1033		1307
2	47		257		499		761		1039		1319
3	53		263	21	503		761		1049		1321
	59		269		509		773		1051		1327
4	61		271		521		787		1061		1361
	67		277		523	27	797		1063		1367
5	71	15	281		541		809		1069	36	1373
	73		283	22	547		811		1087		1381
	79		293		557		821		1091		1399
6	83		307		563		823		1093		1409
	89		311		569		827	32	1097		1423
7	97	16	313		571		829		1103		1427
	101		317		577		839		1109		1429
	103		331		587	28	853		1117		1433
	107		337	23	593		857		1123		1439
	109	17	347		599		859		1129	37	1447
8	113		349		601		863		1151		1451
	127		353		607		877		1153		1453
9	131		359		613		881	33	1163		1459
	137		367		617		883		1171		1471
	139		373		619		887		1181		1481
	149		379		631		907		1187		1483
10	151	18	383	24	641	29	911		1193		1487
	157		389		643		919		1201		1489
	163		397		647		929		1213		1493
	167		401		653		937		1217		1499
11	173		409		659		941		1223		1511
	179		419		661		947		1229	38	1523
	181	19	421		673		953	34	1231		1531
	191		431		677		967		1237		1543
	193		433		683	30	971		1249		1549
12	197		439	25	691		977		1259		1553
	199		443		701		983		1277		1559
	211		449		709		991		1279		1567
13	223		457		719		997		1283		1571
	227	20	461		727		1009		1289		1579
	229		463		733		1013		1291		1583
	233		467		739		1019		1297		1597
	239		479	26	743		1021	35	1301	39	1601

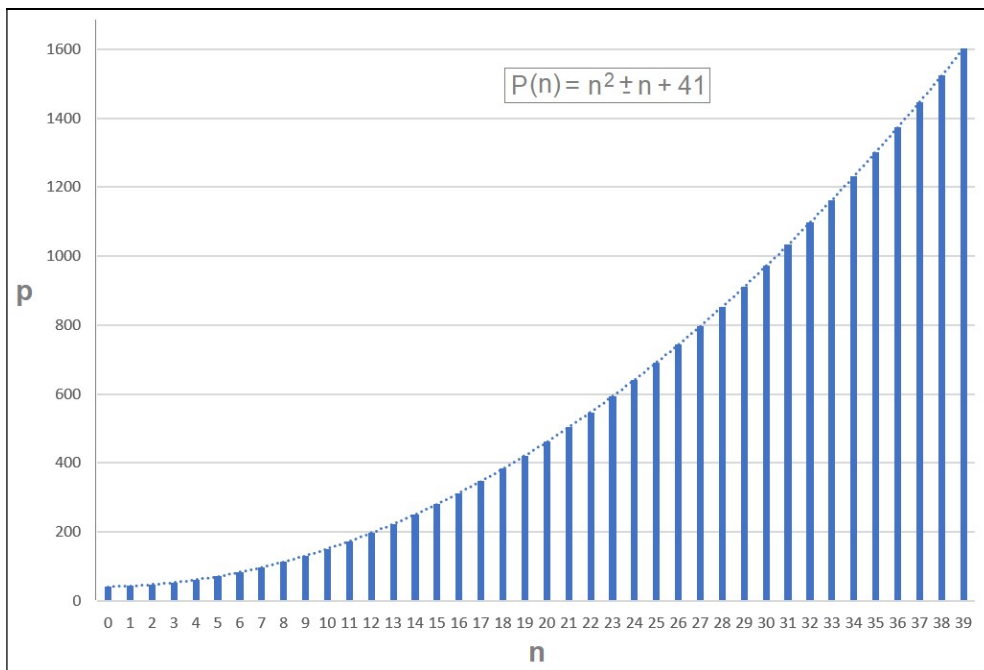


Fig. 10. Polinomio de Euler generador de números primos.

Camacho (2021) y otros autores han realizado diversas propuestas al respecto. Puede que la expresión más sencilla y famosa en este asunto sea el ya citado polinomio cuadrático de Euler, encontrado en el año 1772. Substituyendo el valor de n por cualquier número natural del 0 al 39, el resultado obtenido es un número primo. Por desgracia, la fórmula de Euler no tiene mucho más que ofrecer. En primer lugar, porque como se ha visto, al probar con $n = 40$ nos devuelve un número compuesto y pierde parte de su encanto, puesto que no *solo* genera números primos. Y después, porque no produce *todos* los números primos: solo algunos, y siempre mayores de 41. Es imposible, por ejemplo, que esta fórmula genere el 2, el 11 o el 17.

Basándonos ahora en el teorema de Dirichlet (1805-1859) sobre las progresiones aritméticas, se sabe que funciones lineales del tipo: $f(x) = ax + b$ producen infinitos números primos siempre y cuando a y b sean primos relativos (aunque tal función no asumirá valores primos para cualquier x). No se conoce si existe un polinomio invariable de, al menos, con su grado mayor que 2, que genere un número infinito de valores que son primos.

Por otra parte, un teorema general del matemático ruso Matiyasévich (1947-) dice que si un determinado conjunto se define como un conjunto de ecuaciones diofánticas, también puede ser definido como un sistema de ecuaciones diofánticas con sólo 9 variables. Por lo tanto, existe un polinomio que genera números primos como el anterior de tan sólo 10 variables. Sin embargo, el grado de dicho polinomio es muy grande (del orden de 10^{45}). Visto de otra manera, también podemos transformar dicho polinomio a otro de grado 4, pero esta vez con 58 variables explicativas.

En particular, citamos la extensísima propuesta por Jones *et al.*, en 1976, que hemos referido con anterioridad. Este polinomio, de grado 25, produce todos los números primos que existen. El aspecto es complejo, pero no resulta difícil

de utilizar: eligiendo un valor natural para cada una de sus 26 variables (por ejemplo, $a = 1$, $b = 2$, ..., $z = 26$) y substituyéndolo en el polinomio. Si el resultado es un número positivo, entonces será primo. Si es negativo, debe lógicamente descartarse. Así, según se van proporcionando sucesivamente valores a todas las variables, se van obteniendo *todos* los números primos que existen, sin excepciones.

Por supuesto, la deducción de este polinomio resultó ser harto complicada. Se basaba en un teorema de la década anterior que reza que “cualquier conjunto definido de forma recursivamente numerable, como el de los números primos, se puede describir de forma polinomial”. En su día, la publicación de un teorema de esta magnitud llamó extraordinariamente la atención, pues cerraba siglos de búsqueda matemática.

En cualquier caso, aunque esta última fórmula puede considerarse un éxito de la teoría de números, no posee una indiscutible utilidad práctica, puesto que no permite saber de antemano si el resultado va a ser positivo (y primo) o negativo (e inútil a estos efectos), existiendo, por cierto, muchas más posibilidades de que salga negativo. Aunque hay que reconocer que, a nivel teórico, el esfuerzo desplegado por sus autores fue notable, no es, pues, una fórmula práctica para encontrar todos los números primos.

También existen otros sencillos polinomios que generan números primos. A continuación, presentamos una recopilación de algunas fórmulas polinómicas, las que hemos considerado más interesantes, puesto que, como puede consultarse en la literatura especializada, existen cientos de ellas. Así:

- $n^3 + n^2 + 17$, que ofrece 11 primos distintos para n desde 0 hasta 10.
- $2n^2 + 11$, igual que el anterior.
- $2n^2 + 29$, que da 29 primos distintos para n desde 0 hasta 28.
- $36n^2 - 810n + 2753$, que ofrece 45 primos distintos para n desde 0 hasta 44.

En cualquier caso, en la teoría computacional de números, una notable variedad de algoritmos hacen posible generar números primos de manera eficiente. Se utilizan en diversas aplicaciones, como por ejemplo, hash²⁸, criptografía de clave pública y búsqueda de factores primos en grandes cantidades.

²⁸ Un 'hash' es el resultado de aplicar una función matemática que toma una entrada y la transforma en una cadena de caracteres, generalmente una representación alfanumérica de longitud fija de los datos de entrada. Es decir, es un **código formado por letras y números que resume o representa a un conjunto de datos determinados**. Por ejemplo, podría ser un código que represente un vídeo, una foto o un documento. La característica distintiva de una función 'hash' es que, partiendo de un determinado conjunto de datos de entrada, siempre se generará el mismo valor alfanumérico. Es decir, es un **código único para esos datos**. Adicionalmente, hay dos características importantes a considerar. Por un lado, la probabilidad de identificar dos conjuntos de datos que correspondan al mismo 'hash' es extremadamente baja. Por otra parte, un pequeño cambio en los datos de entrada resultará en un valor completamente diferente del 'hash'.

PARTE II. PLANTEAMIENTO Y DEMOSTRACIONES

1. LA CONJETURA FUERTE O BINARIA DE GOLDBACH (planteamiento)

Se entiende así:

“Todo número par suficientemente grande (mayor que 2) puede expresarse como suma de dos números primos”. El primero de ellos: $4 = 2 + 2$.

Estamos operando en el conjunto de los números enteros positivos o naturales $\{Z^+\} = \{N\}$, y partimos de las siguientes

HIPÓTESIS AXIOMÁTICAS SIMPLES:

- 1) Todos los números primos son impares a excepción del 2 (el recíproco no es cierto) y sumándoles o restándoles la unidad se obtiene un número compuesto par.
- 2) La suma de dos números pares o dos impares es otro número par (el recíproco no es cierto).
- 3) Si sumamos o restamos dos unidades a un número par se obtiene otro número par y lo mismo sucede con los números impares (se obtiene otro número impar).

TESIS:

En efecto, sea: $N = p_1 + p_2$ $\begin{cases} N = \text{número par} > 2 \\ p_1 \text{ y } p_2 \text{ son primos} \end{cases}$

y la suma de dos números primos (impares) es siempre una cantidad par, según la anterior hipótesis 2).

Sumando 2 unidades a ambos miembros de la igualdad anterior (al mismo resultado puede llegarse restando 2 unidades a dichos miembros) se tiene que:

$$\underbrace{N \pm 2}_{\text{par } 3} = \underbrace{(p_1 \pm 1)}_{\text{par } 1} + \underbrace{(p_2 \pm 1)}_{\text{par } 1} = p_1 + p_2 \pm 2.$$

Y así sucesivamente. Partiendo de:

$$\begin{aligned}
N \pm 0 &= \overbrace{(p_1 \pm 0)}^{\text{Impar}} + \overbrace{(p_2 \pm 0)}^{\text{Impar}} = p_1 + p_2 \\
N \pm 2 &= \overbrace{(p_1 \pm 1)}^{\text{P}} + \overbrace{(p_2 \pm 1)}^{\text{P}} = p_1 + p_2 \pm 2 \\
N \pm 4 &= \overbrace{(p_1 \pm 2)}^{\text{I}} + \overbrace{(p_2 \pm 2)}^{\text{I}} = p_1 + p_2 \pm 4 \\
N \pm 6 &= \overbrace{(p_1 \pm 3)}^{\text{P}} + \overbrace{(p_2 \pm 3)}^{\text{P}} = p_1 + p_2 \pm 6 \\
&\dots\dots\dots
\end{aligned}$$

Representando por: P = par e I = impar.

2. LA CONJETURA DÉBIL O TERNARIA DE GOLDBACH (planteamiento)

Se entiende así:

“Todo número impar suficientemente grande (mayor que 5) puede expresarse como suma de tres números primos”. El primero de ellos: $7 = 2 + 2 + 3$. El segundo sería: $9 = 2 + 2 + 5$. El tercero sería: $11 = 3 + 3 + 5$. El cuarto de ellos sería: $13 = 3 + 5 + 5$. Y así sucesivamente.

Partimos ahora de las siguientes

HIPÓTESIS AXIOMÁTICAS SIMPLES:

- 1) La suma de tres números pares es otro número par (el recíproco no es cierto) y la suma de tres números impares es otro número impar.
- 2) Si sumamos o restamos tres unidades a un número impar se obtiene un número par, puesto que sumándole o restándole dos unidades se obtiene un número impar y añadiéndole o restándole una unidad más se obtiene un número par.
- 3) Si sumamos o restamos una unidad a un número primo, a excepción del 2 que es par, se obtiene un número compuesto par.

TESIS:

En efecto, sea: $N = p_1 + p_2 + p_3 \begin{cases} N = \text{número impar} > 5 \\ p_1, p_2 \text{ y } p_3 \text{ son primos} \end{cases}$

y la suma de tres números primos (impares) es una cantidad impar según la hipótesis 1).

Sumando ahora 3 unidades a los dos miembros de la igualdad anterior (al mismo resultado puede llegarse restando 3 unidades a dichos miembros) se tiene que:

$$N \pm 3 = \overbrace{(p_1 \pm 1) + (p_2 \pm 1) + (p_3 \pm 1)}^{\text{par 1)}} = p_1 + p_2 + p_3 \pm 3$$

$\underbrace{\quad}_{\text{par 2)}} \quad \underbrace{\quad}_{\text{par 3)}} \quad \underbrace{\quad}_{\text{par 3)}} \quad \underbrace{\quad}_{\text{par 3)}}$

La conjetura débil de Goldbach es cierta, ya que:

$$\forall N > 5 \Rightarrow N = p_1 + p_2 + p_3, \text{ pues ya está demostrada (Helfgott, 2013).}$$

De hecho, esta conjetura débil o ternaria ya ha sido demostrada por procedimientos muy prolijos (más de 200 páginas) por H. Helfgott continuando los trabajos de Hardy-Wright (1997), Littlewood (1914), Vinogradov (1937), Lin, Wang (1984), Chen Jing-run (1973) y otros.

Las conjeturas siguientes pueden hacerse extensivas según el número de sumandos primos que intervienen. Así, podríamos enunciar que “todo número par suficientemente grande (mayor que 6) puede expresarse como suma de cuatro números primos”. El primero de ellos podría ser el $8 = 2 + 2 + 2 + 2$, seguido del $10 = 2 + 2 + 3 + 3$. El tercero sería: $12 = 2 + 2 + 3 + 5$. El cuarto sería: $14 = 2 + 2 + 5 + 5$. Y así sucesivamente.

Otro resultado interesante que se ha conseguido probar es que **todo número par puede escribirse como suma de, como mucho, 6 primos** (Olivier, 1991). Sería necesario, no obstante, analizar específicamente cada una de esas conjeturas.

Sumando algebraicamente ahora 4 unidades a los dos miembros de la igualdad anterior (al mismo resultado puede llegarse restando 4 unidades a dichos miembros) se tiene que:

$$N \pm 4 = \overbrace{(p_1 \pm 1) + (p_2 \pm 1) + (p_3 \pm 1) + (p_4 \pm 1)}^{\text{par 1)}} = p_1 + p_2 + p_3 + p_4 \pm 4$$

$\underbrace{\quad}_{\text{par 2)}} \quad \underbrace{\quad}_{\text{par 3)}} \quad \underbrace{\quad}_{\text{par 3)}} \quad \underbrace{\quad}_{\text{par 3)}} \quad \underbrace{\quad}_{\text{par 3)}}$

También la Teoría de Cribas ha reportado una apoyatura importante en el estudio del problema, partiendo de la antigua criba de Eratóstenes (276-194 a. C.). Una forma especial de la cual fue expuesta por Euler muchos siglos después.

3. LAS DIFERENTES MEDIAS DE DOS NÚMEROS PRIMOS (demostración)

A continuación, se presenta un gráfico representativo, en la Fig. 11, con la construcción geométrica para hallar las medias aritmética ($\bar{X}$), cuadrática (C), geométrica (G) y armónica (H) de dos números primos cualesquiera p_1 y p_2 cuya suma constituye el número par natural N.

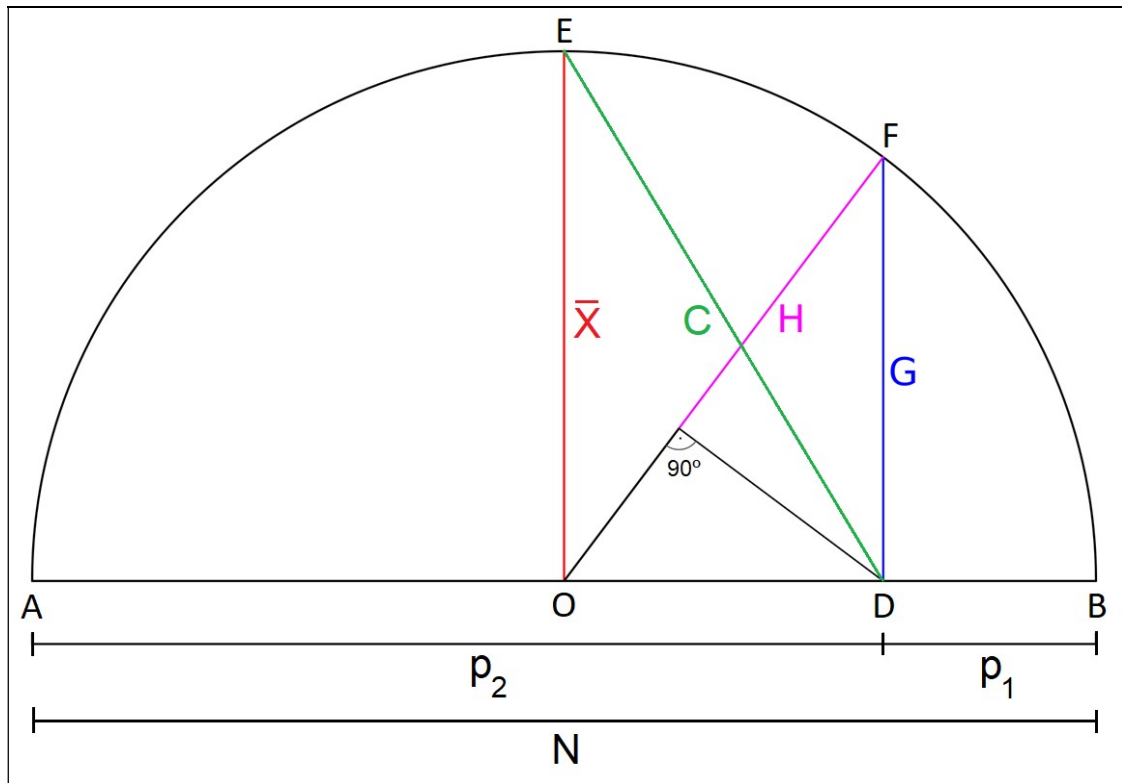


Fig. 11. Construcción geométrica de las cuatro medias empleadas.

Consideremos ahora que, en el caso de la conjetura fuerte de Goldbach que tratamos de demostrar, se suman 2 números primos p_1 (segmento $\overline{BD}$) y p_2 (segmento $\overline{AD}$) para obtener siempre un número par natural N (segmento $\overline{AB}$).

Las diferentes medias clásicas de estos dos números primos quedarán definidas así:

$$H = \frac{2}{\frac{1}{p_1} + \frac{1}{p_2}} = \frac{2}{\frac{p_1 + p_2}{p_1 p_2}} = \frac{2p_1 p_2}{p_1 + p_2}$$

$$G = \sqrt{p_1 p_2}; \quad \bar{X} = \frac{p_1 + p_2}{2}; \quad C = \sqrt{\frac{p_1^2 + p_2^2}{2}}$$

Vamos a demostrar, en primer lugar, que $H \leq G$, o sea:

$$\frac{2p_1 p_2}{p_1 + p_2} \leq \sqrt{p_1 p_2}$$

Elevando al cuadrado los miembros de la anterior desigualdad y operando adecuadamente, se obtiene:

$$4p_1^2p_2^2 \leq p_1p_2(p_1 + p_2)^2 \quad ; \quad 4p_1p_2 \leq (p_1 + p_2)^2$$

$$4p_1p_2 \leq p_1^2 + p_2^2 + 2p_1p_2 \quad ; \quad 0 \leq p_1^2 + p_2^2 - 2p_1p_2$$

$$0 \leq (p_1 - p_2)^2$$

Con esto queda demostrado que $H \leq G$. De otra parte, $G \leq \bar{X}$, ya que:

$$\sqrt{p_1p_2} \leq \frac{p_1 + p_2}{2} \quad ; \quad 4p_1p_2 \leq (p_1 + p_2)^2$$

Entonces se tendrá que: $0 \leq (p_1 - p_2)^2$, por tanto, queda demostrado que resulta $H \leq G \leq \bar{X}$. Por último, teniendo en cuenta la relación que liga la media cuadrática con la aritmética y la variancia, esto es: $C^2 = \bar{X}^2 + \sigma^2$, se deduce también que: $\bar{X} \leq C$ (RMS), como se quería demostrar.

Esta demostración, además, puede generalizarse para cualquier número de números primos, como por ejemplo en el caso de la conjetura débil o ternaria de Goldbach, con los números primos p_1 , p_2 y p_3 . En cualquier caso, la demostración de la última desigualdad también puede realizarse analíticamente a partir de la definición de las dos medias, esto es:

$$\bar{X} = \frac{p_1 + p_2}{2} \quad ; \quad C = \sqrt{\frac{p_1^2 + p_2^2}{2}} \quad . \text{Entonces:}$$

$$\bar{X}^2 = \frac{p_1^2 + p_2^2 + 2p_1p_2}{4} \quad ; \quad C^2 = \frac{p_1^2 + p_2^2}{2} = \frac{p_1^2 + p_2^2}{4} + \frac{p_1^2 + p_2^2}{4} \quad ;$$

$$\bar{X}^2 = \frac{p_1^2 + p_2^2}{4} + \frac{p_1p_2}{2} \quad ; \quad C^2 - \bar{X}^2 = \frac{p_1^2 + p_2^2}{4} - \frac{p_1p_2}{2} = \frac{(p_1 - p_2)^2}{4} = \sigma^2$$

Se ha de observar que esta diferencia de fracciones ha de ser necesariamente positiva o nula. En efecto, se trata de comparar las expresiones:

$\frac{p_1^2 + p_2^2}{4}$ y $\frac{2p_1p_2}{4}$, o más concretamente sus numeradores, o sea: $p_1^2 + p_2^2$ y $2p_1p_2$. Como se tiene que el cuadrado de una diferencia ofrece el valor: $(p_1 - p_2)^2 = p_1^2 + p_2^2 - 2p_1p_2 \geq 0 \Rightarrow p_1^2 + p_2^2 \geq 2p_1p_2$; también se cumple que: $C^2 \geq \bar{X}^2$, y además: $\bar{X} \leq C$ (RMS), c.s.q.d.

También otra medida del valor central de una distribución de frecuencias empleada en Estadística es la media cúbica que, para datos sin agrupar, viene dada por la expresión:

$$M_3 = \sqrt[3]{\frac{\sum p_i^3}{2}} = \sqrt[3]{\frac{p_1^3 + p_2^3}{2}}$$

que ofrecería el mayor valor central de los otros valores promedios anteriormente relacionados.

Si ahora ejemplificamos con los números primos $p_1 = 3$ y $p_2 = 7$, representados por sendos segmentos cuya suma $N = \overline{AB} = 10$ es el diámetro de la semicircunferencia de radio $\overline{X} = \overline{OA} = \overline{OE} = \overline{OF} = \overline{OB} = 5$ en la construcción geométrica de la Fig. 11, se tienen los siguientes valores de las medias usualmente empleadas:

$$\begin{aligned} H &= \frac{2p_1 p_2}{p_1 + p_2} = \frac{2 \times 3 \times 7}{3 + 7} = \frac{42}{10} = 4.20; \\ G &= \sqrt{p_1 p_2} = \sqrt{3 \times 7} = 4.58; \quad \overline{X} = \frac{p_1 + p_2}{2} = \frac{3 + 7}{2} = 5.00; \\ C &= \sqrt{\frac{p_1^2 + p_2^2}{2}} = \sqrt{\frac{3^2 + 7^2}{2}} = \sqrt{\frac{9 + 49}{2}} = 5.39 \\ M_3 &= \sqrt[3]{\frac{27 + 343}{2}} = \sqrt[3]{185} = 5.70, \end{aligned}$$

confirmándose la desigualdad ya demostrada:

$$(H = 4.20) < (G = 4.58) < (\overline{X} = 5.00) < (C = 5.39) < (M_3 = 5.70).$$

También debe tenerse en cuenta, por lo que se refiere a la varianza, que:

$$\sigma^2 = \frac{1}{2} \sum_{i=1}^2 (p_i - \overline{X})^2 = \left(\frac{1}{2} \sum_{i=1}^2 p_i^2 \right) - \overline{X}^2 = \frac{p_1^2 + p_2^2}{2} - \left(\frac{p_1 + p_2}{2} \right)^2 = \dots = \frac{(p_1 - p_2)^2}{4},$$

con lo que, en este ejemplo concreto, sucede que:

$$\sigma^2 = \frac{(p_1 - p_2)^2}{4} = \frac{(3 - 7)^2}{4} = 4, \Rightarrow \sigma = (|p_1 - p_2|)/2 = 2, \text{ y entonces:}$$

$$\begin{aligned} CV (\text{Pearson}) &= \frac{\sigma}{\overline{X}} \times 100 = \frac{2}{5} \times 100 = 40\%, \text{ y un coeficiente de uniformidad de:} \\ 100 (1 - CV) &= 100 (1 - 0.40) = 60\%. \end{aligned}$$

$$\text{Comprobación: } \sigma = (C^2 - \overline{X}^2)^{1/2} = (29 - 5^2)^{1/2} = 4^{1/2} = 2.$$

Por lo que se refiere a la media aritmética $\overline{X}$ de 2 números primos p_1 y p_2 (o mejor el punto medio), veamos que su resultado puede ser par o impar, y hay infinitas medias aritméticas que se pueden formar a partir de 2 números primos, cuyo número también es infinito como se ha visto. Consecuentemente, con dichas medias se pueden formar infinitos pares naturales N que se podrán expresar como suma de 2 números primos, lo que constituye la conjetura fuerte de Goldbach que se trata de demostrar. Así:

$$\bar{X} = \frac{p_1 + p_2}{2} = \frac{N}{2} \Rightarrow N = 2 \cdot \bar{X} = p_1 + p_2 ,$$

resultando que el duplo de la media aritmética de 2 números primos será, necesariamente, un número natural par expresado como suma de ambos números primos, lo que demuestra la mencionada conjetura.

CONCLUSIÓN:

La conjetura fuerte de Goldbach sería cierta, ya que se cumple que:

$$\forall N > 2 \Rightarrow N = p_1 + p_2$$

4. UNA APROXIMACIÓN GEOMÉTRICA (demostración)

4.1. Conjetura fuerte o binaria

Un segmento es la porción de recta limitada por dos puntos, llamados “extremos” o “finales”. A diferencia de una semirrecta, que tiene un determinado origen pero se extiende infinitamente en una dirección, el segmento está acotado en ambos extremos.

En nuestro caso, a la vista de la Fig. 11, asimilando dos segmentos concatenados, consecutivos y colineales con los números primos p_1 ($\overline{BD}$) y p_2 ($\overline{AD}$), su adición conforma otro segmento de longitud N ($\overline{AB}$) determinado por los extremos no comunes de los segmentos considerados.

El conjunto de los segmentos métricos constituye una magnitud, de la que los segmentos son cantidades que asimilaremos al cardinal de cada número primo p_1 y p_2 . Es posible determinar entre ellos relaciones y efectuar las operaciones definidas para los elementos de una magnitud.

Geométricamente, la suma de segmentos es otro segmento que se obtiene construyendo colinealmente segmentos ordenadamente congruentes con los dados. La suma de los dos segmentos p_1 y p_2 es otro segmento que tiene por inicio el origen del primer segmento y como final el final del segundo segmento. La longitud del segmento suma N es igual a la suma de las longitudes respectivas de los dos segmentos que lo conforman, p_1 y p_2 .

El concepto de infinito se utiliza para muchas cuestiones diferentes, y no debe confundirse. Un segmento siempre consta de infinitos puntos individuales²⁹.

²⁹ A finales del siglo XIX, el matemático ruso-alemán Georg Cantor (1845-1918) conmocionó al mundo de las matemáticas, rompiendo las creencias existentes hasta el momento sobre el concepto de infinito. Entre los revolucionarios resultados que demostró están que existe más de un infinito o que la cantidad de puntos de un segmento es la misma que la cantidad de puntos de un cuadrado. Fue el inventor, junto con el alemán Dedekind (1831-1916), de la Teoría de Conjuntos, que constituye la base de las matemáticas modernas. Merced a sus atrevidas investigaciones sobre los conjuntos infinitos, fue el primer matemático capaz de *formalizar* la noción de infinito bajo la forma de los números transfinitos (cardinales y ordinales).

Esto se puede demostrar rigurosamente y no resulta nada difícil. Básicamente, entre dos puntos distintos de una línea hay un tercer punto («entre» no debería significar necesariamente el punto medio de una trayectoria recta que los conecta). La longitud de la línea es un asunto diferente. Puede ser finita o infinita. La totalidad de los puntos que componen el segmento es, en cualquier caso, infinita. De hecho, se puede demostrar, con bastante facilidad, que la cardinalidad de los puntos de cualquier línea es siempre la misma, por lo que, en cierto sentido, todas las líneas tienen la misma «cantidad» de puntos, aunque la forma en que estos puntos están dispuestos puede conferirle cualidades geométricas totalmente diferentes.

En la geometría euclidiana, un segmento es una porción de línea recta que se extiende en una misma dirección; por lo tanto, tiene una sola dimensión y contiene un número infinito de puntos, por lo que no es posible asignarle un cardinal.

Cualquier segmento con al menos dos puntos tiene infinitos puntos, porque, intuitivamente, dados dos puntos distintos, hay un tercero, distinto de ambos, digamos, el punto medio. Esto es, se puede representar cualquier segmento N (número natural par mayor de 2) como adición de dos segmentos primos cualesquiera (exceptuando el número primo 2), que resulta una cantidad par y natural, y su número es infinito según hemos visto en el anterior punto 4 del presente trabajo.

El giro levógiro del radio $\overline{OB}$ sobre cada uno de los infinitos puntos de la semicircunferencia determina infinitas proyecciones verticales sobre el segmento $\overline{AB}$ (N) que lo particionan en los dos segmentos métricos p_1 ($\overline{BD}$) y p_2 ($\overline{AD}$), de longitudes iguales cuando $D = O$ o bien desiguales cuando $D \neq O$, cuya suma es igual a N , pudiendo existir infinitos pares que (p_1, p_2) que cumplen esta condición habida cuenta de la infinitud de los números primos.

Por otra parte, de la contemplación de la Fig. 11 se deduce también gráficamente que $N = 2 \cdot \overline{X}$, con $\overline{X} = \text{cte.}$, o sea, que un mismo número par y natural N se puede representar de infinitas maneras y combinaciones $(p_1 + p_2)$ sobre el segmento $\overline{AB}$ conforme se produce el giro de 180° del radio $\overline{OB}$ hasta $\overline{OA}$.

CONCLUSIÓN:

Ello confirma que la conjetura fuerte de Goldbach sería cierta, ya que se cumple geoméricamente que:

$$\forall N > 2 \Rightarrow N = p_1 + p_2$$

4.2. Conjetura débil o ternaria

Lo mismo puede verse en relación a la conjetura débil o ternaria de Goldbach, a partir de la siguiente composición geométrica:

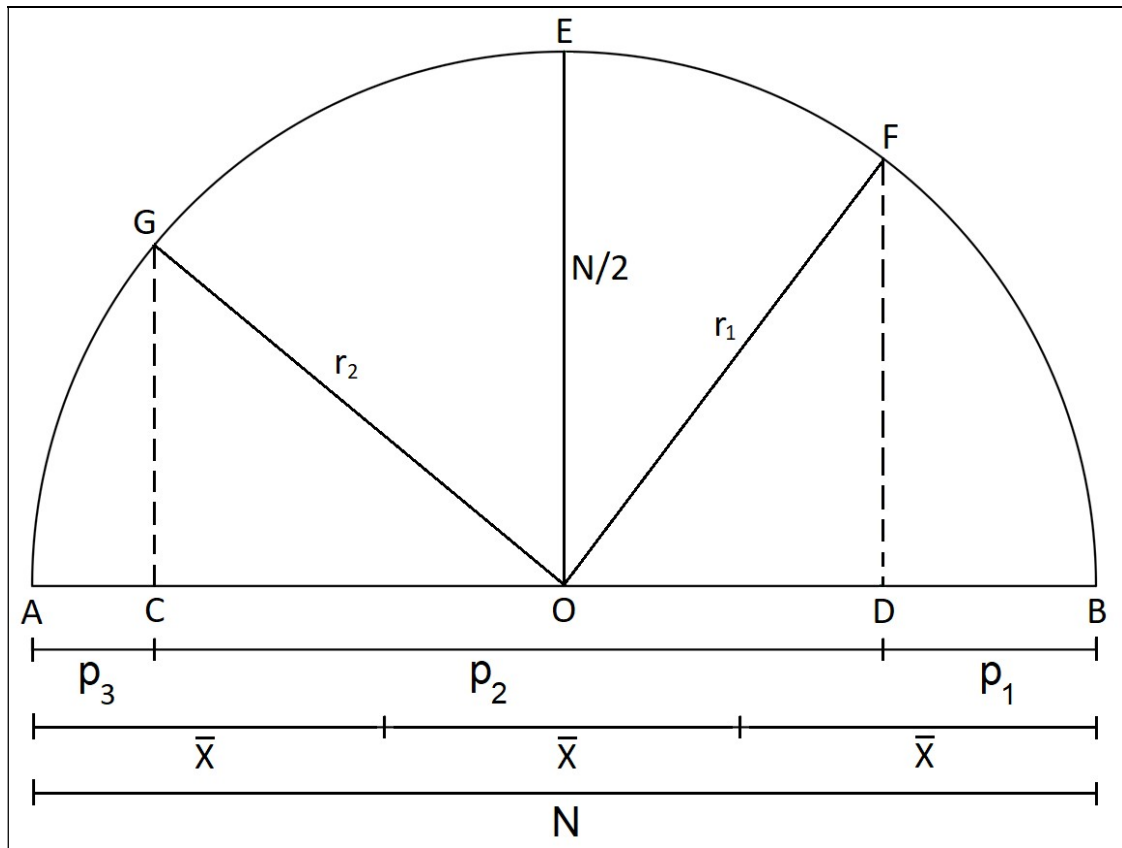


Fig. 12. Construcción geométrica de la conjetura ternaria de Goldbach (I).

Aquí, los radios r_1 y r_2 , de valor $N/2$, pueden girar indistintamente y en sentidos iguales u opuestos sobre la semicircunferencia AGEFB y las proyecciones verticales sobre el segmento o diámetro $\overline{AB}(N)$ le particionan en tres partes desiguales (p_1 , p_2 y p_3) o bien iguales (todas ellas iguales a $\bar{X}$). De hecho, el mismo resultado se obtiene haciendo girar indefinidamente $360^\circ = 2\pi$ radianes dichos radios para configurar la circunferencia completa.

En este caso, $N = p_1 + p_2 + p_3$, entonces:

$$\bar{X} = \frac{p_1 + p_2 + p_3}{3} = \frac{N}{3}, \text{ o sea, } N = 3\bar{X}.$$

El problema también puede plantearse geométricamente a partir de una elipse de semiejes a y b inscrita a la circunferencia de radio r (de hecho, la primera sección cónica se convierte en la segunda con sus semiejes iguales $a = b = r$ y sin excentricidad) y el radio r , de valor $N/2$, que gira indistintamente en ambos sentidos (levógiro y dextrógiro) con las proyecciones verticales sobre el segmento $\overline{AB}(N)$ de sus puntos de corte con la elipse y la circunferencia circunscrita (Fig. 13). Las separaciones existentes entre dichas proyecciones representan los números primos sumandos.

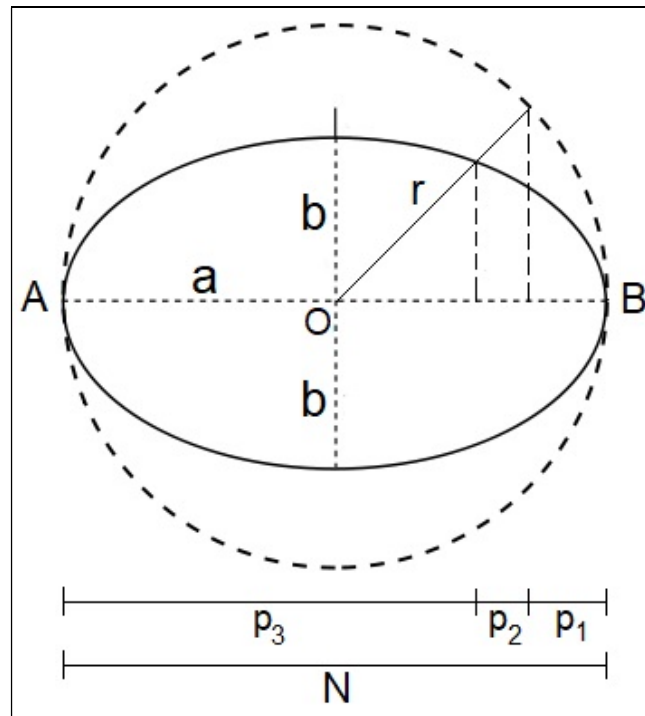


Fig. 13. Construcción geométrica de la conjetura ternaria de Goldbach (II).

Todo ello confirma también que la conjetura débil de Goldbach sería cierta, ya que se cumple geoméricamente que:

$$\forall N > 5 \Rightarrow N = p_1 + p_2 + p_3$$

De hecho, como sus nombres sugieren, la demostración de la conjetura fuerte ya implicaría la de la conjetura débil, lo que se consigue restando 3 a su número impar y expresando después $N - 3$ (par) como suma de dos números primos (Helfgott, 2013).

La resolución, en fin, de la conjetura débil de Goldbach implica un resultado relacionado con la conjetura fuerte, y es que **todo número par puede expresarse por tanto como suma de, como mucho, 4 números primos**.

5. OTRO INTENTO DE DEMOSTRACIÓN DE LA CONJETURA FUERTE O BINARIA

La “distancia o diferencia” (d) que separa dos números primos, consecutivos o no, es siempre un número natural par, a excepción de los primos 2 y 3, como puede comprobarse, por ejemplo, analizando los primos contenidos en los primeros 200 números naturales (véase tabla 13). De este modo:

$$N(d) = p_1 - p_2 = \Delta, \text{ ya que: } p_1 = p_2 + N(d),$$

luego puede afirmarse o conjeturarse que “Todo número par natural (N) puede representarse como diferencia de dos números primos”. Esto puede corroborarse en la tabla siguiente de los primeros 200 números pares naturales. Así:

Tabla 12. Formación de pares naturales N por sustracción de 2 números primos (1 a 200).

N	$p_1 - p_2$	N	$p_1 - p_2$	N	$p_1 - p_2$	N	$p_1 - p_2$	N	$p_1 - p_2$
2	--	42	47-5	82	89-7	122	127-5	162	167-5
4	7-3	44	47-3	84	89-5	124	127-3	164	167-3
6	9-3	46	53-7	86	89-3	126	131-5	166	173-7
8	11-3	48	53-5	88	101-13	128	131-3	168	173-5
10	13-3	50	53-3	90	97-7	130	137-7	170	173-3
12	17-5	52	59-7	92	97-5	132	137-5	172	179-7
14	17-3	54	59-5	94	97-3	134	137-3	174	179-5
16	19-3	56	59-3	96	101-5	136	139-3	176	179-3
18	23-5	58	61-3	98	101-3	138	149-11	178	181-3
20	23-3	60	67-7	100	103-3	140	151-11	180	191-11
22	29-7	62	67-5	102	107-5	142	149-7	182	193-11
24	29-5	64	67-3	104	107-3	144	149-5	184	191-7
26	29-3	66	71-5	106	109-3	146	149-3	186	191-5
28	31-3	68	71-3	108	113-5	148	151-3	188	191-3
30	37-7	70	73-3	110	113-3	150	157-7	190	193-3
32	37-5	72	79-7	112	131-19	152	157-5	192	197-5
34	37-3	74	79-5	114	127-13	154	157-3	194	197-3
36	41-5	76	79-3	116	127-11	156	163-7	196	199-3
38	41-3	78	83-5	118	131-13	158	163-5	198	211-13
40	43-3	80	83-3	120	127-7	160	163-3	200	211-11

Por otra parte, las “distancias o diferencias” que separan dos números primos consecutivos contenidos en los primeros 200 números naturales pueden verse en la siguiente tabla:

Tabla 13. Distancias existentes entre 2 números primos consecutivos (1 a 200).

Número Primo	Diferencia Δ	Número Primo	Diferencia Δ	Número Primo	Diferencia Δ	Número Primo	Diferencia Δ
2	1	3	2	5	2	7	4
11	2	13	4	17	2	19	4
23	6	29	2	31	6	37	4
41	2	43	4	47	6	53	6
59	2	61	6	67	4	71	2
73	6	79	4	83	6	89	8
97	4	101	2	103	4	107	2
109	4	113	14	127	4	131	6
137	2	139	10	149	2	151	6
157	6	163	4	167	6	173	6
179	2	181	10	191	2	193	4
197	2	199					

(Cuando $\Delta = 2$, se trata de primos gemelos)

A la vista de la tabla 12, escogiendo al azar los 6 números pares naturales: 10, 12, 14, 80, 94 y 192, y sumándoles un conveniente número par natural (6 u 8, en este caso), se forma lo siguiente:

$$\begin{array}{l}
 +6 \left\{ \begin{array}{l} 10 = 13 - 3 \\ \mathbf{16} = 13 - 3 + 6 = \mathbf{13 + 3} \end{array} \right. \\
 +8 \left\{ \begin{array}{l} 12 = 17 - 5 \\ \mathbf{20} = 17 - 5 + 8 = \mathbf{17 + 3} \end{array} \right. \\
 +6 \left\{ \begin{array}{l} 14 = 17 - 3 \\ \mathbf{20} = 17 - 3 + 6 = \mathbf{17 + 3} \end{array} \right. \\
 +6 \left\{ \begin{array}{l} 80 = 83 - 3 \\ \mathbf{86} = 83 - 3 + 6 = \mathbf{83 + 3} \end{array} \right. \\
 +6 \left\{ \begin{array}{l} 94 = 97 - 3 \\ \mathbf{100} = 97 - 3 + 6 = \mathbf{97 + 3} \end{array} \right. \\
 +8 \left\{ \begin{array}{l} 192 = 197 - 5 \\ \mathbf{200} = 197 - 5 + 8 = \mathbf{197 + 3} \end{array} \right.
 \end{array}$$

Y así podríamos obtener cualquier número par natural que se expresaría, siempre necesariamente, como suma de dos números primos (Goldbach).

6. METÁFORA O PARÁBOLA DE LAS GOTAS DE AGUA (demostración)

a) Una sencilla imagen que vislumbra toda el agua de los océanos, mares y aguas continentales (ríos, lagos, embalses, arroyos, ...) de nuestro planeta azul puede resultar relevante al respecto del problema que nos ocupa. Se estima que esta agua supone un volumen de 1386 millones de km^3 . Por otra parte, algunas experiencias realizadas coinciden en señalar que una sola gotita de agua cubica, por término medio, 0.33 ml, o sea, 330 mm^3 . Teniendo en cuenta que:

$$1 \text{ km}^3 = 10^9 \text{ m}^3 = 10^{18} \text{ mm}^3 ; 1 \text{ mm}^3 = 0.001 \text{ ml} ; 1 \text{ ml} = 1000 \text{ mm}^3$$

resultará que el agua total superficial de la Tierra, o sea:

$$1386 \times 10^6 \text{ km}^3 = 1386 \times 10^6 \times 10^{18} \text{ mm}^3 = 1386 \times 10^{24} \text{ mm}^3$$

está constituida por: $\frac{1386 \times 10^{24} \text{ mm}^3}{330 \text{ mm}^3 / \text{gotita}} = 4.2 \times 10^{24}$ gotitas, que equivalen a una

cantidad $N = 2.1 \times 10^{24}$ gotas, obtenidas cada una por la suma de dos gotitas o números primos p_i (en nuestro caso).

Así pues, dichas gotitas elementales, unidas con otras formando grupos o sumas de 2, constituyen un conjunto de cardinal: $N = 2.1 \times 10^{24}$ gotas (que asimilaremos a los números pares en número inmensamente grande o *quasi* infinito. Habría que recordar aquí que, aunque hubiera algunas excepciones k , según el concepto de infinito³⁰, también resultaría que: $\infty - k = \infty$).

Esta metáfora hídrica superaría ampliamente la comprobación realizada por potentes ordenadores hasta la fecha para todos los números pares $N < 10^{18}$. En efecto:

$$\frac{2.1 \times 10^{24}}{10^{18}} = 2.1 \times 10^6 = 2\,100\,000$$
 veces mayor que la efectuada con las mejores herramientas digitales actuales. En cualquier caso, sobre esta metáfora cabría remarcar que se trata de un recurso explicativo y no de un modelo formal.

b) Podríamos establecer otra metáfora basada en el número de estrellas del universo infinito. Si se utiliza la cifra más pesimista de estrellas del universo, se estima que hay unos 10^{25} astros (es decir, 10 cuatrillones). Del mismo modo, si se realiza la estimación más optimista, sobre los granitos de arena existentes en todas las playas y desiertos de la Tierra, se llega incluso a cifras muy superiores. Siempre podría haber 2 elementos (p_1 y p_2) de estos conjuntos naturales que se unieran para obtener una pareja N de tamaño superior. Y al ser infinitos, como se ha visto, los elementos (números primos p_i) que se pudieran adicionar, también lo serían sus sumas N (las gotas de agua).

Ello explicaría también la conjetura fuerte o binaria de Goldbach, ya que todo número par N puede perfectamente estar formado por una pareja de números primos (p_i, p_j) que se suman, cuestión ésta que también es posible intuir claramente por inducción completa. A este respecto, hay que considerar que, en el campo de la lógica y la filosofía, la inducción o razonamiento inductivo es un método de razonamiento lógico capaz de obtener conclusiones generales a partir de premisas particulares, siempre que estas últimas sugieran o sostengan la conclusión, sin llegar nunca a garantizarla.

Es, pues, un método de conocimiento que permite obtener, por generalización, un enunciado general a partir de enunciados que describen casos particulares. La inducción se considera completa cuando se han observado todos los casos particulares, y aquí ya hemos concluido que es cierta la infinitud de los números primos (apdo. 4 del presente trabajo) y sus correspondientes agrupaciones binarias, por lo que la generalización a la que da lugar se considera válida. En la mayoría de otros casos, no obstante, es imposible alcanzar una “inducción completa”, por lo que el enunciado general a que da lugar la aplicación de dicho método quedaría sometido a un cierto grado de

³⁰ Lo primero que debemos tener claro es que **el infinito no es un número real**, es, más bien... una idea. El concepto de **infinito** (de símbolo: ∞) aparece en varias ramas de la matemática, la filosofía y la astronomía, en referencia a una cantidad sin límite o sin final, contrapuesto al concepto de finitud. En matemáticas, el infinito aparece de diversas formas: en geometría, el punto al infinito en geometría proyectiva y el punto de fuga en geometría descriptiva; en análisis matemático, los límites infinitos; y en teoría de conjuntos como números transfinitos referentes a los números ordinales infinitos, que son mayores que cualquier número natural.

incertidumbre. Entonces hablaríamos de “inducción incompleta”, que no se puede considerar como un esquema de inferencia formalmente válido, ni se puede justificar empíricamente, por lo que nos ofrece “verdades” que gozan de un mayor o menor grado de probabilidad, pero que no nos pueden garantizar su certeza absoluta.

c) Y ahora, como una curiosidad geodésica, que nada tiene que ver con los números primos que estamos tratando, veamos que la longitud del meridiano terrestre, partiendo de la clásica definición del metro (unidad de medida en el sistema métrico decimal) cómo la “diez millonésima parte del cuadrante del meridiano terrestre”, supondría 40000 km, con lo que:

$$2\pi r = 40000 \text{ km} \Rightarrow r = \frac{40000}{6.2831854} = 6366 \text{ km},$$

aunque la medición del radio ecuatorial arroja un resultado algo mayor (6378 km) y el radio polar algo menor (6357 km) habida cuenta del ligero achatamiento o elipticidad de la esfera terráquea que se determina por:

$$c = (a - b)/a,$$

donde **c** es el achatamiento, **a** el eje mayor o ecuatorial, y **b** el eje menor o polar; esto es, donde $c = (12738 - 12441) / 12738 = 0.023316$, y por lo tanto, resulta: $a - b = 297 \text{ km}$, que se expresa como 1 : 297.

Simplificativamente, la superficie de dicha *quasi* esfera (en realidad es un elipsoide de revolución o esferoide), viene dada por:

$$S = 4\pi r^2 = 4\pi \times 6366^2 = 509\,264\,183 \text{ km}^2,$$

lo que implicaría que, de repartirse toda el agua superficial del planeta de un modo uniforme por toda la superficie planetaria, alcanzaría una altura media de:

$$h = \frac{1386 \times 10^6 \text{ km}^3}{509\,264\,183 \text{ km}^2} = 2.722 \text{ km} = 2722 \text{ m}.$$

CONCLUSIÓN:

La conjetura fuerte de Goldbach sería cierta, ya que se cumple que:

$$\forall N > 2 \Rightarrow N = p_1 + p_2$$

Espero que, todo ello, les resulte de utilidad.

PARTE III. UNAS REFLEXIONES FINALES

¿Y qué sucede con el número primo dos? Es un número entero y par, pero no es la suma de dos primos porque sólo se puede formar sumando dos unos, y el número uno no es primo. Resultaría entonces que para este número la conjetura fuerte de Goldbach no se cumple. ¿Es posible que hayamos resuelto aquí la conjetura fuerte de Goldbach? Evidentemente que no, pues para esto se impone la restricción $\forall N > 2$.

Otra cuestión que se puede reflexionar es acerca de la necesidad de plantearse y resolver elucubraciones fútiles. Por ejemplo, la siguiente conjetura de carácter antropomorfo: “todos los seres humanos tienen cabeza sobre los hombros”. *In pure*, no llegaremos a la demostración definitiva de esta aseveración hasta que no hayamos encontrado, al menos, una persona descabezada y viva, lo que refutaría el anterior aserto, por lo que se trata de ir recorriendo el mundo pacientemente hasta hallar con ella, o bien no habiendo hallado tan extraño ejemplar humanoide sobre la faz de la Tierra, después de haberlos contemplado todos, concluir gozosamente que la conjetura es cierta.

Se me ocurre, en relación a lo expuesto en el párrafo anterior, que la cabeza (parte superior del cuerpo humano en la que están situados el cerebro y los principales órganos sensoriales) es un órgano consubstancial con el ser humano, por lo cual resulta absolutamente precisa su presencia en el mismo, y sin la cual no existe dicho ser, por lo que es absurdo seguir buscando. En este sentido, habrá quien piense (con la cabeza) que sería más útil para la humanidad que algunas mentes bien pensantes sobre la materia dedicaran sus loables esfuerzos a crear y/o perfeccionar modelos matemáticos de predicción del clima, de mejora de la salud de las personas, de prospectiva económica o de análisis de los movimientos migratorios a escala mundial.

Algo así se puede comparar con la búsqueda afanosa de un número N , natural y par, que no pueda representarse como suma de dos números primos, cuyo número, por cierto, es infinito. Los números primos son elementos y parte consubstancial (subconjunto) de los números naturales y la suma simple de dos de ellos, al ser impares, siempre ofrecerá una cantidad par. Pues bien, inasequibles al desaliento, ya llevamos comprobada, mediante el apoyo de potentes sistemas informáticos, una monstruosidad de números naturales pares menores que 10^{18} sin detectar excepción alguna a la famosa conjetura. Y la fiesta continúa.

Parece absurdo, pues, seguir probando estoicamente muestras todavía mayores, puesto que si se cumple la conjetura, jamás se llegaría a encontrar la prueba negativa de la misma habida cuenta de la infinitud del universo a investigar. Pese a los denodados e ímprobos esfuerzos realizados hasta la fecha, aún quedan infinitos números por probar, y los matemáticos sospechan que, por mucho que busquen, no encontrarán un solo caso en que no se cumpla la conjetura de Goldbach. Sostienen, con razón, que a medida que los números naturales crecen, también aumentan las maneras en que pueden expresarse a partir de los números primos. Por ejemplo: el número diez puede hacerse sumando el cinco consigo mismo o bien el tres con el siete. El cien,

por ejemplo, puede construirse de seis formas diferentes, y por lo que hemos visto, el número de combinaciones posible se dispara a medida que estudiamos números cada vez más grandes. Por inducción podría deducirse o intuirse, pues, que la conjetura fuerte de Goldbach es cierta.

Se trata de un argumento muy simple: cuanto mayor es un número par N mayor es el número de formas en las que podemos expresarlo como suma de dos números naturales. Por tanto, también mayor es la probabilidad de que exista una forma de escribirlo en la que los dos números en cuestión sean primos, que también son elementos de $\{N\}$, puesto que $\{P\} \subset \{N\}$. Posiblemente no sirva ello como demostración “ad hoc”, pero puede servirnos de idea para enfocar nuestros estudios sobre el tema. En efecto, esta conjetura ha sido investigada por muchos teóricos de números y ha sido comprobada por ordenadores para todos los números pares menores que 4×10^{18} . Y en el caso de la conjetura débil o ternaria, con Helfgott (2013) se ha alcanzado la fabulosa cifra de 8.875×10^{30} números naturales rastreados.

En realidad, con la comprobación computacional actual de la conjetura fuerte o binaria de Goldbach, se está realizando un análisis censal o universal frente al análisis muestral, o sea, no se está empleando la Teoría de la Probabilidad y Estadística para determinar el tamaño de la muestra. Pues bien, si en base a criterios de máxima prudencia adoptamos, “a priori”, un nivel de confianza del 99.9% (ver Anexo de información complementaria, tabla 3) y un error muestral de estimación máximo permisible del 1%, con estos exigentes condicionantes, el tamaño de la muestra de números naturales a analizar en la búsqueda objetiva números primos puede determinarse empleando el muestreo aleatorio simple³¹, partiendo de que se trata de una población infinita de números naturales $n = \infty$, tomando: $z_{\alpha/2} = 3.09$ como valor de la función de Gauss acumulativa de la distribución normal para una determinada probabilidad, lo que exigiría el recuento de un número de naturales (tamaño necesario de la muestra para estimar la media poblacional en una población infinita) de:

$$n' = \frac{z_{\alpha/2}^2 \times \sigma_c^2}{e^2} = \frac{3.09^2 \times 2922.1^2}{0.01^2} \cong 8.15 \times 10^{11},$$

que resulta ser una cantidad enorme de elementos a analizar pero aún así muy inferior a 4×10^{14} (números pares menores comprobados por J. Richstein) y también de exigencia muy inferior a la que se viene manejando hasta la fecha (4×10^{18}), como ya hemos puesto de manifiesto ahora mismo y en otros apartados del presente trabajo. Y, pese a ello, como prueba demostrativa de la conjetura, sigue sin hallarse un número natural N par que no se pueda escribir o expresar como suma de dos números primos. Posiblemente, en este contexto sería mejor usar conceptos de la física cuántica que pudieran aportar más luz para la búsqueda de números primos.

³¹ El muestreo aleatorio simple es una técnica esencial en la investigación estadística que proporciona una muestra representativa de la población. Consiste en seleccionar un subconjunto aleatorio de individuos de la población objetivo para representar a todo el grupo. Es un procedimiento de muestreo probabilístico que da a cada elemento de la población objetivo y a cada posible muestra de un tamaño determinado, la misma probabilidad de ser seleccionado.

También se ha demostrado que la conjetura de Goldbach es **cierta para casi todos los números pares**. En este contexto matemático, esta expresión quiere decir que el límite del cociente entre el número de pares menores que m para los que se cumple la conjetura y el número de todos los pares menores que m , es igual a 1 (cuando m tiende a infinito). Intuitivamente, podríamos decir, que para números pares muy grandes, sería rarísimo que no se cumpliera la conjetura.

La mayor parte de los matemáticos creen, pues, que la conjetura es cierta, y se basan mayoritariamente en las consideraciones estadísticas sobre la distribución probabilística de los números primos en el conjunto de los números naturales: cuanto mayor sea el número entero par N , se hace más «probable» que pueda ser escrito como suma de dos números primos. Esas consideraciones estadísticas presentan evidencia informal a favor de la conjetura (tanto en su forma débil como en la fuerte) para números enteros suficientemente grandes: cuanto mayor es el entero, más formas hay disponibles de ese número para ser representado como la suma de otros dos o tres números, y lo más "probable" resulta ser que al menos una de estas representaciones consista enteramente a partir de números primos. Esto queda visible mediante la contemplación de la representación gráfica del denominado “cometa de Goldbach”:

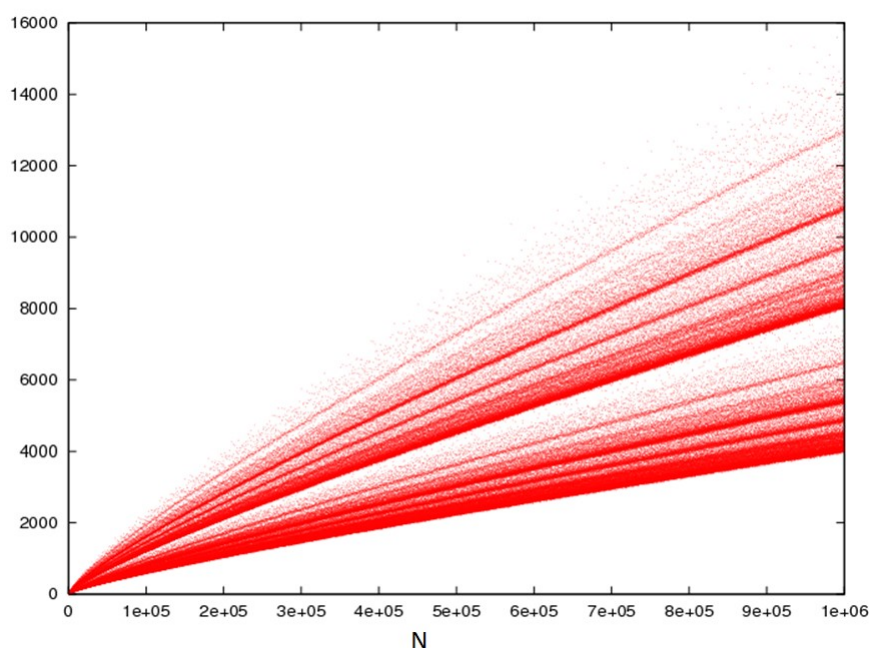


Fig. 14. Diferentes maneras de expresar N natural y par como suma de 2 números primos ($4 \leq N \leq 10^6$).

Ambas conjeturas de Goldbach (1690-1764), en definitiva, tanto la fuerte como la débil que hemos tratado aquí, de aceptarse por la comunidad científica nuestros razonamientos, ya podrían considerarse como teoremas.

Hemos partido de algún planteamiento previo para enfocar el tratamiento del problema que nos ocupa, acompañado de alguna reflexión sobre el estado de la cuestión y la constatación de la dificultad que, a lo largo de la historia, han

experimentado diferentes matemáticos a la hora de intentar su resolución. Debe diferenciarse, en nuestro trabajo, entre la argumentación lógico-inductiva y la demostración estrictamente formal.

Normalmente se clasifican tres tipos de razonamiento: el lógico, el creativo y el **heurístico**. Este último hace referencia a la modalidad de pensamiento que busca la solución a los problemas que se le presentan a través de conocimientos previos o de la propia experiencia, resolviéndolos de forma práctica.

Se conoce como heurística al **conjunto de técnicas o métodos para resolver un problema**. La palabra heurística es de origen griego *εὕρισκειν* que significa “hallar”. También, se afirma que la heurística se basa en la experiencia propia del individuo, y en la de otros para encontrar la solución más viable al problema aunque no por ello deba ser necesariamente óptima. Ello nos ha inspirado, sin duda.

La heurística es vista también como el arte de inventar. Consiste en la búsqueda de estrategias, métodos y criterios que permitan resolver problemas, a través de la creatividad y el pensamiento divergente o lateral. En general, se considera que la capacidad heurística es un rasgo característico de los humanos. Así, su enfoque no reside en crear un sistema totalmente lógico y óptimo para resolver el problema en cuestión, sino más bien proporcionarle una solución rápida o inmediata de una manera creativa e intuitiva.

Actualmente se han hecho adaptaciones al expresado término en diferentes áreas. De este modo, definen la heurística como un arte, técnica o procedimiento práctico o informal, para resolver un problema. Alternativamente, Imre Lakatos (1922-1974) la define como un conjunto de reglas metodológicas no necesariamente forzosas, positivas y negativas, que sugieren o establecen cómo proceder y qué problemas evitar a la hora de generar soluciones y elaborar hipótesis.

Según el matemático húngaro George Pólya (1887-1985), que por cierto realizó contribuciones fundamentales en la teoría de números y la probabilidad, la base de la heurística se halla en la experiencia de resolver problemas y en ver cómo otros lo hacen. Consecuentemente, se dice que hay búsquedas ciegas, búsquedas heurísticas (basadas en la experiencia) y búsquedas racionales.

También hemos creído útil efectuar aquí una aproximación a la heurística en el estudio del problema planteado. En las ciencias computacionales también se habla muchas veces de procesos heurísticos o estrategias heurísticas en diferentes sentidos, aunque en general se le entiende como un proceso o regla que permite alcanzar una solución práctica, aunque no necesariamente óptima o formal (Romanycia y Pelletier, 1985). En la literatura que versa sobre el razonamiento heurístico se ha asumido una distinción tajante y en muchos casos incluso una oposición entre el razonamiento formal y el razonamiento heurístico. Argumento en que la discusión se ha centrado en cuál de estos tipos de razonamientos debe marcar la pauta del buen razonamiento o cuál debe ser considerado como modelo de nuestro razonar. Kahneman (1996),

Gigerenzer (1996) y la tradición de heurística y sesgo asumen que el razonamiento formal debe ser el canon de buen razonamiento, suscribiendo lo que se conoce como “visión estándar de la racionalidad”; es decir, las reglas derivadas de los axiomas de la lógica, del cálculo de probabilidades, la estadística y de las teorías formales de la decisión racional.

Hemos pretendido realizar, en el presente artículo, una aproximación a una demostración mucho más simple que las emprendidas hasta la fecha, que fuera asequible también para los “no matemáticos”. De hecho, si resulta fácil la comprensión del enunciado de la conjetura, también podría resultar relativamente sencilla su demostración. Pero, como siempre, aceptamos cualquier otra opinión que será, sin duda alguna, mucho más autorizada que la nuestra.

Por cierto, y con alguna contraposición a lo expresado en párrafos anteriores, en la misma célebre carta de Goldbach a Euler, nuestro egregio personaje escribe una reflexión muy interesante y sabia sobre las matemáticas en sí mismas, que también quisiera aplicar a nuestro humilde trabajo:

“No creo que sea totalmente inútil plantear aquellas proposiciones que son muy probables aunque falte una verdadera demostración, pues aún cuando se descubra que son incorrectas, pueden conducir al descubrimiento de una nueva verdad.”

PARTE IV. REFERENCIAS BIBLIOGRÁFICAS Y RECOMENDACIONES

En cualquier caso, para saciar la curiosidad de los lectores más interesados en el tema, además de la bibliografía ya citada en el mismo cuerpo del trabajo, son recomendables los siguientes textos y artículos relacionados con esta famosa conjetura matemática de Goldbach tal como la hemos expuesto aquí. A saber:

[1] ALCAIDE, A. *Estadística económica*. Ed. SAETA. Madrid, 1973. 370 p. Proporciona una buena apoyatura sobre la estadística descriptiva.

[2] AMO, E. de, DÍAZ M. y FERNÁNDEZ J. *Demostraciones de la infinidad de los números primos*. Épsilon - Revista de Educación Matemática 2013, Vol. 30(2), nº 84, 69-88. ISSN: 2340-714X.

[3] BELLMAN, R. *A note on the divergence of a series*. *Amer. Math. Monthly*, 50, 318-319, 1943.

[3-bis] CAMACHO, J. “Una fórmula que genera números primos”. *Revista Digital: Matemática, Educación e Internet*, vol. 22, núm. 1, pp. 1-9, 2021. Instituto Tecnológico de Costa Rica.

[3-ter] CHARLIER, C. V. L. *Vorlesungen über die Grundzüge der mathematischen Statistik*. Lund (Suecia), 1931 (versión castellana del Dr. José González Galé, en Cuadernos de Trabajo del Instituto de Biometría, Buenos Aires, 1936).

- [4] CILLERUELO, J. y CÓRDOBA, A. *La Teoría de los Números*, Mondadori, 1992. Hay un capítulo dedicado a los métodos de criba donde se demuestra el teorema de Vigo Brun y el teorema de Schnirelmann.
- [4-bis] CHEN, J.R. «On the representation of a larger even integer as the sum of a prime and the product of at most two primes». 1973. *Sci. Sinica* **16**: 157-176.
- [5] CLARKSON, J.A. *On the series of prime reciprocals*. *Proc. Amer. Math. Soc.*, 17, 541, 1966.
- [6] DIXON, W.J. y MASSEY, F.J. *Introducción al análisis estadístico*. Ed.: Del Castillo, S.A. Madrid, 1969. 489 p. Proporciona una buena apoyatura sobre la estadística descriptiva.
- [7] DRESSLER, R.E. *A lower bound for $\pi(n)$* . *Amer. Math. Monthly*, 82, 151-152. 1975.
- [8] DUSART, P. *Autour de la fonction qui compte le nombre de nombres premiers*. Tesis doctoral para la Universidad de Limoges, 1998.
- [8-bis] EDWARDS, C. y PENNEY, D. *Cálculo con Geometría Analítica*. México, D.F.: Prentice Hall. 1997.
- [9] ERDÖS, P. *Über die Reihe $\sum 1/p$* . *Mathematica Zutphen B*, **7**, 1-2. 1938.
- [10] EULER, L. *Variae observationes circa series infinites*. 1737. *Comm. Acad. Petropolitanae*, 9, 160-188.
- [10-bis] FERNÁNDEZ, J., SÁNCHEZ, R. y ÚBEDA, M. “Fórmulas que generan números primos”. Universidad de Almería. Épsilon - Revista de Educación Matemática 2022, nº 110, 77-101.
- [11] FRANQUET, J.M. *Modelo estadístico de distribución espacial de las acequias en una zona regable*. Ed. Universidad Nacional de Educación a Distancia (UNED), Centro Asociado de Tortosa. Colección Cadup-Estudios, 2019. 146 p. Proporciona una buena apoyatura sobre la estadística descriptiva.
- [12] FREUND, J.E. y WILLIAMS, F.J. *Elementos modernos de estadística empresarial*. Ed. Prentice/Hall International, Inc. London, 1964-1972. 462 p. Proporciona una buena apoyatura sobre la estadística descriptiva.
- [13] FURSTENBERG, H. *On the infinitude of primes*. *Amer. Math. Monthly*, 62, 353. 1955.
- [13-bis] GIGERENZER, G. “How to make cognitive illusions disappear: Beyond ‘heuristics and biases’”, *European Review of Social Psychology*, núm. 2, pp. 83-115. 1996.
- [14] GOLDBACH, C. “Letter to Leonhard Euler” (Moscú, 7 de junio de 1742).

- [15] GUY, RICHARD. *Unsolved Problems in Number Theory*, 2nd edition, Springer-Verlag, New York, 1995. Extensa colección de problemas sin resolver, con muchas referencias.
- [15-bis] HACKS, J. *Über einige für Primzahlen charakteristische Beziehungen*. Acta Mathematica, 17(1), 205. 1983.
- [16] HARDY, G. H. y WRIGHT, E. *An introduction to the Theory of Numbers*, Oxford at the Clarendon Press, 1997. Un clásico de la teoría de los números.
- [17] HARRIS, V.C. *Another proof of the infinitude of primes*. Amer. Math. Monthly, 63, 711. 1956.
- [18] HELFGOTT, H.A. «Major arcs for Goldbach's theorem». 2013. [arXiv:1305.2897](https://arxiv.org/abs/1305.2897)
- [18-bis] JONES, J. P.; SATO, D.; WADA, H. y WIENS, D. "Diophantine representation of the set of prime numbers". American Mathematical Monthly, 83(6), 449–464, 1976.
- [18-ter] KAHNEMAN, D. "Prize Lecture: Maps of Bounded Rationality", disponible en: Kahneman, Daniel, 2002, "Prize Lecture: Maps of Bounded Rationality", disponible en: <http://www.Nobelprize.org> . Nobel Media AB 2013, (consultado el 30/1/2014).
- [19] LANDAU, E. *Vorlesungen über Zahlentheorie* (3 volum.). Leipzig: Hirzel. 1927.
- [20] LEGENDRE, A.M. *Essai sur la Theorie des Nombres*. Paris: Duprat. 1798.
- [21] LITTLEWOOD, J.E. *Distribution des nombres premiers*, C. R. Acad. Sci. Paris 158, pp. 1869-1872, 1914.
- [22] MOSER, L. *On the series*. Amer. Math. Monthly, 65, 104-105, 1958.
- [23] NAIR, M. *On Chebyshev-type inequalities for primes*. Amer. Math. Monthly, 89(2), 126-129. 1982.
- [24] NATHANSON, M. B. *Additive number theory: The Classical Bases*, Graduate Texts in Mathematics 165, Springer-Verlag, 1996. Un texto avanzado y bien escrito, que incluye las demostraciones del Teorema de Vinogradov y del Teorema de Chen Jing-run.
- [25] OWENS, M. "¿Is There a Formula that Generates Prime Generates Prime Numbers?". A Sonoma State Math Colloquium, 2008. Estados Unidos de Norteamérica. Recuperado el 15 de junio de 2020: https://web.sonoma.edu/math/colloq/primes_sonoma_state_9_24_08.pdf.

- [26] POLLACK, P. *Not Always Buried Deep. Selections from Analytic and Combinatorial Number Theory*, 2003. Recuperado de la dirección <http://www.princeton.edu/~ppollack/notes/notes.pdf>.
- [27] PRIETO, C. "Los Números Primos: Hechos y Conjeturas". Segundo Encuentro con los Números Antioquia Colombia, 2013. Recuperado el 15 de junio de 2020: <https://paginas.matem.unam.mx/cprieto/phocadownloadpap/presentaciones/pdf>.
- [28] PULIDO, A. *Estadística y técnicas de investigación social*. Ed. Anaya. Madrid, 1971. Proporciona una buena apoyatura sobre la estadística descriptiva.
- [29] RAMARÉ, O. «Contribution au problème de Goldbach: tout entier supérieur a 1 est somme d'au plus 13 nombres premiers» / Grenoble : Atelier national de reproduction des thèses, 1991.
- [30] RIBENBOIM, P. *The New Book of Prime Number Records*, 3rd edition Springer-Verlag, New York, 1995. Un libro de fácil lectura, muy recomendable.
- [31] RICHSTEIN, J. *Verifying the Goldbach Conjecture up to 4×10^{14}* , Mathematics of Computation, 1998.
- [32] RIEMANN, B. *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie, pp. 671-680, 1859.
- [33] ROMANYCIA, M. H. y PELLETIER, F.J. "What is a heuristic?", *Computational Intelligence*, núm. 1, pp. 47-58, 1985.
- [34] ROWLAND, E. "A Natural Prime – Generating Recurrence", 2008. Recuperado de ArXiv el 14 de junio de 2020: <https://arxiv.org/abs/0710.3217v3>.
- [35] SPIEGEL, M.R. *Teoría y problemas de probabilidad y estadística*. Ed. Mc Graw-Hill. México, 372 p. 1981. Proporciona una buena apoyatura sobre la estadística descriptiva.
- [36] STURGES, H. "The choice of a class-interval". *Journal of Amer. Statist. Assoc*, 21, pp. 65-66, 1926. Recuperado de: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0073-47212007000200010
- [37] VALLÉE POUSSIN, CH. J. DE LA. «Recherches analytiques sur la théorie des nombres premiers», Ann. Soc. Sci. Bruxelles, vol. 20, pp. 183-256, 281-297, 1896.
- [38] VINOGRADOV, I.M. "Representation of an odd number as a sum of three primes", Dokl. Akad. Nauk. SSR 15, 291–294, 1937.
- [39] WANG, Y. *Goldbach's conjecture World Scientific*, Publ., Singapore, 1984.

[40] WUNDERLICH, M. *Another proof of the infinite prime's theorem*. *Amer. Math. Monthly*, 72, 305. 1965.

[41] XIAO, K. "The Prime Number Formulas". 2015. Recuperado de ViXra el 15 de junio de 2020: <https://vixra.org/pdf/1501.0129v1.pdf>.

* * * * *

(*) Nota explicativa (p. 8):

Hay que tener en cuenta que, en este caso, con la pareja de números primos (p_1, p_2) , tal que: con $p_2 = p_1 + 2$, se cumple que:

$$\begin{aligned} A &= \sqrt{p_1 p_2 + 1} = \frac{p_1 + p_2}{2}, \text{ elevando al cuadrado se tiene:} \\ A^2 &= p_1 p_2 + 1 = p_1(p_1 + 2) + 1 = p_1^2 + 2p_1 + 1 = (p_1 + 1)^2. \text{ Así mismo:} \\ A^2 &= \frac{(p_1 + p_2)^2}{4} = \frac{p_1^2 + p_2^2 + 2p_1 p_2}{4} = \frac{p_1^2 + (p_1 + 2)^2 + 2p_1(p_1 + 2)}{4} = \\ &= \frac{p_1^2 + p_1^2 + 4 + 4p_1 + 2p_1^2 + 4p_1}{4} = \frac{4p_1^2 + 8p_1 + 4}{4} = p_1^2 + 2p_1 + 1 = (p_1 + 1)^2 \end{aligned}$$

De donde resulta la semisuma: $A = p_1 + 1$.

() Nota explicativa (p. 17):**

γ es la constante de Euler-Mascheroni. La **constante de Euler-Mascheroni**, (también conocida como *constante de Euler*), a la que ya nos hemos referido con anterioridad, es una constante matemática que aparece principalmente en la teoría de números, y se denota con la letra griega minúscula γ (gamma). Se define como el límite de la diferencia entre la serie armónica y el logaritmo natural o neperiano, a saber:

$$\gamma = \lim_{n \rightarrow \infty} \left[\sum_{k=1}^n \frac{1}{k} - \ln(n) \right] = \int_1^{\infty} \left(\frac{1}{[x]} - \frac{1}{x} \right) dx$$

Su valor aproximado es:

$$\gamma \approx 0.577\ 215\ 664\ 901\ 532\ 860\ 606 \dots$$

Esta constante apareció por primera vez en el año 1734, en un artículo escrito por Leonhard Euler, denominado *De Progressionibus harmonicis observationes*, calculando los 6 primeros dígitos para la constante y llamándola C. En 1781 calcularía otros 10 decimales más. En 1790, Lorenzo Mascheroni (1750-1800) calcularía los primeros 19 decimales y la denotaría como A. Ya más tarde se denotaría de la forma moderna como γ , debido a su conexión con la función gamma.

El número γ no se ha probado que sea algebraico o trascendente; de hecho, ni siquiera se conoce si γ es irracional o no. El análisis de fracciones continuas revela que, de ser racional, su denominador debe ser muy elevado (actualmente del orden de 10^{242080}). Debido a que está presente en un gran número de ecuaciones y relaciones, la racionalidad o irracionalidad de γ se halla, sin duda, entre los problemas abiertos más importantes de las matemáticas.

SÍMBOLOS, ABREVIATURAS Y SIGLAS

%	Porcentaje (tanto por cien)
...	Puntos suspensivos (etcétera)
↑, ↓	Ascendente, descendente
→	Tiende a ...
/	Dividido entre, tal que
a.C.	Antes de Cristo
art.	Artículo
c	Amplitud del intervalo de clase
C	Media cuadrática, coeficiente de contingencia
C _{ap}	Coeficiente de apertura
CGPM	Conferencia General de Pesas y Medidas
Coef.	Coeficiente
cos	Coseno
c.s.q.d.	Como se quería demostrar
CV	Coeficiente de variación de Pearson
DM	Desviación media absoluta
Dr.	Doctor
=	Igual a ...
≠	Desigual a ...
e	Base de los logaritmos neperianos o naturales, 2.7182818284...
Ed.	Editorial
et alt.	<i>Et altri</i>
etc.	Etcétera
≈, ≐, ≅	Aproximadamente igual, equivalente
≤	Menor o igual
≥	Mayor o igual
∃	Existe
⊂	Está contenido
∈	Pertenece
Fig.	Figura
!	Factorial
G	Media geométrica
g.l.	Grados de libertad
GNU GPL	<i>General Public License</i>
g ₁	Coeficiente de asimetría o sesgo (Fisher)
g ₂	Coeficiente de curtosis (Fisher)
H	Media armónica
⇒	Implica
∞	Infinito
ICM	<i>International Congress of Mathematicians</i>
I _G	Índice de Gini
IMU	Unión Matemática Internacional
k	número de intervalos de clase
km	Kilómetros
L	intervalo de clase
lim	Límite
ln	Logaritmo natural o neperiano
log	Logaritmo decimal

MC	Mínimo cuadrático
M_e	Mediana
M_{eg}	Mediana de Gastwirth
m_3	Momento orden 3 respecto a la media aritmética o central
M_3	Media cúbica
m_4	Momento orden 4 respecto a la media aritmética o central
ml	Mililitros
mm	Milímetros
M_o	Moda
n^o	Número
p. = pág.	Página
P, Pr	Probabilidad, coeficiente asimetría o sesgo
p_i, p_j	Números primos
pp. = págs.	Páginas
$\forall$	Para todo
Q_1	Primer cuartil
Q_2	Segundo cuartil (Mediana)
Q_3	Tercer cuartil
Q_4	Cuarto cuartil
$\{R\}$	Conjunto de los números reales
r	Coeficiente de correlación lineal
R^2	Recorrido o rango, coeficiente de determinación o crítico
R'	Recorrido relativo
Ref.	Referencia
RMS	<i>Root mean square</i>
s	Desviación típica muestral
s^2	Variancia muestral
sin	Seno
ss.	Siguientes
T	Trimedia (Tukey)
T_1, T_2	Terciles
UNED	Universidad Nacional de Educación a Distancia
V	Coeficiente de Cramers
v.gr.	<i>Verbi gratia</i>
x	Abscisa, marca de clase
$\bar{X}$	Media aritmética
y	Ordenada
Z	Transformación de Fisher

LETRAS GRIEGAS

α	Esperanza matemática (media aritmética), nivel de significación o confianza
χ^2	Distribución y estadígrafo Chi-cuadrado
π	Irracional 3.1415927...
σ	Desviación típica poblacional
σ^2	Varianza poblacional
σ_c	Desviación típica corregida
ϕ	Coeficiente de correlación de 4 campos
Σ	Sumatorio
Π	Productorio o multiplicatorio

ANEXO DE INFORMACIÓN COMPLEMENTARIA

1. NÚMEROS PRIMOS COMPRENDIDOS EN LOS 10000 PRIMEROS NÚMEROS NATURALES

1	2 3 5 7 11 13 17 19 23 29
2	31 37 41 43 47 53 59 61 67 71
3	73 79 83 89 97 101 103 107 109 113
4	127 131 137 139 149 151 157 163 167 173
5	179 181 191 193 197 199 211 223 227 229
6	233 239 241 251 257 263 269 271 277 281
7	283 293 307 311 313 317 331 337 347 349
8	353 359 367 373 379 383 389 397 401 409
9	419 421 431 433 439 443 449 457 461 463
10	467 479 487 491 499 503 509 521 523 541
11	547 557 563 569 571 577 587 593 599 601
12	607 613 617 619 631 641 643 647 653 659
13	661 673 677 683 691 701 709 719 727 733
14	739 743 751 757 761 769 773 787 797 809
15	811 821 823 827 829 839 853 857 859 863
16	877 881 883 887 907 911 919 929 937 941
17	947 953 967 971 977 983 991 997 1009 1013
18	1019 1021 1031 1033 1039 1049 1051 1061 1063 1069
19	1087 1091 1093 1097 1103 1109 1117 1123 1129 1151
20	1153 1163 1171 1181 1187 1193 1201 1213 1217 1223
21	1229 1231 1237 1249 1259 1277 1279 1283 1289 1291
22	1297 1301 1303 1307 1319 1321 1327 1361 1367 1373
23	1381 1399 1409 1423 1427 1429 1433 1439 1447 1451
24	1453 1459 1471 1481 1483 1487 1489 1493 1499 1511
25	1523 1531 1543 1549 1553 1559 1567 1571 1579 1583
26	1597 1601 1607 1609 1613 1619 1621 1627 1637 1657
27	1663 1667 1669 1693 1697 1699 1709 1721 1723 1733
28	1741 1747 1753 1759 1777 1783 1787 1789 1801 1811
29	1823 1831 1847 1861 1867 1871 1873 1877 1879 1889
30	1901 1907 1913 1931 1933 1949 1951 1973 1979 1987
31	1993 1997 1999 2003 2011 2017 2027 2029 2039 2053
32	2063 2069 2081 2083 2087 2089 2099 2111 2113 2129
33	2131 2137 2141 2143 2153 2161 2179 2203 2207 2213

34	2221	2237	2239	2243	2251	2267	2269	2273	2281	2287
35	2293	2297	2309	2311	2333	2339	2341	2347	2351	2357
36	2371	2377	2381	2383	2389	2393	2399	2411	2417	2423
37	2437	2441	2447	2459	2467	2473	2477	2503	2521	2531
38	2539	2543	2549	2551	2557	2579	2591	2593	2609	2617
39	2621	2633	2647	2657	2659	2663	2671	2677	2683	2687
40	2689	2693	2699	2707	2711	2713	2719	2729	2731	2741
41	2749	2753	2767	2777	2789	2791	2797	2801	2803	2819
42	2833	2837	2843	2851	2857	2861	2879	2887	2897	2903
43	2909	2917	2927	2939	2953	2957	2963	2969	2971	2999
44	3001	3011	3019	3023	3037	3041	3049	3061	3067	3079
45	3083	3089	3109	3119	3121	3137	3163	3167	3169	3181
46	3187	3191	3203	3209	3217	3221	3229	3251	3253	3257
47	3259	3271	3299	3301	3307	3313	3319	3323	3329	3331
48	3343	3347	3359	3361	3371	3373	3389	3391	3407	3413
49	3433	3449	3457	3461	3463	3467	3469	3491	3499	3511
50	3517	3527	3529	3533	3539	3541	3547	3557	3559	3571
51	3581	3583	3593	3607	3613	3617	3623	3631	3637	3643
52	3659	3671	3673	3677	3691	3697	3701	3709	3719	3727
53	3733	3739	3761	3767	3769	3779	3793	3797	3803	3821
54	3823	3833	3847	3851	3853	3863	3877	3881	3889	3907
55	3911	3917	3919	3923	3929	3931	3943	3947	3967	3989
56	4001	4003	4007	4013	4019	4021	4027	4049	4051	4057
57	4073	4079	4091	4093	4099	4111	4127	4129	4133	4139
58	4153	4157	4159	4177	4201	4211	4217	4219	4229	4231
59	4241	4243	4253	4259	4261	4271	4273	4283	4289	4297
60	4327	4337	4339	4349	4357	4363	4373	4391	4397	4409
61	4421	4423	4441	4447	4451	4457	4463	4481	4483	4493
62	4507	4513	4517	4519	4523	4547	4549	4561	4567	4583
63	4591	4597	4603	4621	4637	4639	4643	4649	4651	4657
64	4663	4673	4679	4691	4703	4721	4723	4729	4733	4751
65	4759	4783	4787	4789	4793	4799	4801	4813	4817	4831
66	4861	4871	4877	4889	4903	4909	4919	4931	4933	4937
67	4943	4951	4957	4967	4969	4973	4987	4993	4999	5003
68	5009	5011	5021	5023	5039	5051	5059	5077	5081	5087
69	5099	5101	5107	5113	5119	5147	5153	5167	5171	5179
70	5189	5197	5209	5227	5231	5233	5237	5261	5273	5279
71	5281	5297	5303	5309	5323	5333	5347	5351	5381	5387
72	5393	5399	5407	5413	5417	5419	5431	5437	5441	5443
73	5449	5471	5477	5479	5483	5501	5503	5507	5519	5521

74	5527	5531	5557	5563	5569	5573	5581	5591	5623	5639
75	5641	5647	5651	5653	5657	5659	5669	5683	5689	5693
76	5701	5711	5717	5737	5741	5743	5749	5779	5783	5791
77	5801	5807	5813	5821	5827	5839	5843	5849	5851	5857
78	5861	5867	5869	5879	5881	5897	5903	5923	5927	5939
79	5953	5981	5987	6007	6011	6029	6037	6043	6047	6053
80	6067	6073	6079	6089	6091	6101	6113	6121	6131	6133
81	6143	6151	6163	6173	6197	6199	6203	6211	6217	6221
82	6229	6247	6257	6263	6269	6271	6277	6287	6299	6301
83	6311	6317	6323	6329	6337	6343	6353	6359	6361	6367
84	6373	6379	6389	6397	6421	6427	6449	6451	6469	6473
85	6481	6491	6521	6529	6547	6551	6553	6563	6569	6571
86	6577	6581	6599	6607	6619	6637	6653	6659	6661	6673
87	6679	6689	6691	6701	6703	6709	6719	6733	6737	6761
88	6763	6779	6781	6791	6793	6803	6823	6827	6829	6833
89	6841	6857	6863	6869	6871	6883	6899	6907	6911	6917
90	6947	6949	6959	6961	6967	6971	6977	6983	6991	6997
91	7001	7013	7019	7027	7039	7043	7057	7069	7079	7103
92	7109	7121	7127	7129	7151	7159	7177	7187	7193	7207
93	7211	7213	7219	7229	7237	7243	7247	7253	7283	7297
94	7307	7309	7321	7331	7333	7349	7351	7369	7393	7411
95	7417	7433	7451	7457	7459	7477	7481	7487	7489	7499
96	7507	7517	7523	7529	7537	7541	7547	7549	7559	7561
97	7573	7577	7583	7589	7591	7603	7607	7621	7639	7643
98	7649	7669	7673	7681	7687	7691	7699	7703	7717	7723
99	7727	7741	7753	7757	7759	7789	7793	7817	7823	7829
100	7841	7853	7867	7873	7877	7879	7883	7901	7907	7919
101	7927	7933	7937	7949	7951	7963	7993	8009	8011	8017
102	8039	8053	8059	8069	8081	8087	8089	8093	8101	8111
103	8117	8123	8147	8161	8167	8171	8179	8191	8209	8219
104	8221	8231	8233	8237	8243	8263	8269	8273	8287	8291
105	8293	8297	8311	8317	8329	8353	8363	8369	8377	8387
106	8389	8419	8423	8429	8431	8443	8447	8461	8467	8501
107	8513	8521	8527	8537	8539	8543	8563	8573	8581	8597
108	8599	8609	8623	8627	8629	8641	8647	8663	8669	8677
109	8681	8689	8693	8699	8707	8713	8719	8731	8737	8741
110	8747	8753	8761	8779	8783	8803	8807	8819	8821	8831
111	8837	8839	8849	8861	8863	8867	8887	8893	8923	8929
112	8933	8941	8951	8963	8969	8971	8999	9001	9007	9011
113	9013	9029	9041	9043	9049	9059	9067	9091	9103	9109

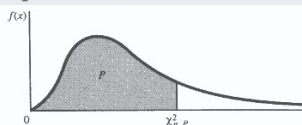
114	9127 9133 9137 9151 9157 9161 9173 9181 9187 9199
115	9203 9209 9221 9227 9239 9241 9257 9277 9281 9283
116	9293 9311 9319 9323 9337 9341 9343 9349 9371 9377
117	9391 9397 9403 9413 9419 9421 9431 9433 9437 9439
118	9461 9463 9467 9473 9479 9491 9497 9511 9521 9533
119	9539 9547 9551 9587 9601 9613 9619 9623 9629 9631
120	9643 9649 9661 9677 9679 9689 9697 9719 9721 9733
121	9739 9743 9749 9767 9769 9781 9787 9791 9803 9811
122	9817 9829 9833 9839 9851 9857 9859 9871 9883 9887
123	9901 9907 9923 9929 9931 9941 9949 9967 9973

Fuente: www.vaxasoftware.com

NOTA: Los cuartiles (Q_0 , Q_1 , $Q_2 = Me$, Q_3 , Q_4) se han señalado en color rojo, y los terciles (T_1 , T_2) en color azul.

2. PERCENTILES DE LA DISTRIBUCIÓN χ^2 DE PEARSON

Esta tabla proporciona los valores $\chi^2_{n, p}$, tales que

$$P = P(X \leq \chi^2_{n, p}) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} \int_0^{\chi^2_{n, p}} x^{\frac{n}{2}-1} \cdot e^{-\frac{1}{2}x} dx$$


$P \backslash n$	0,750	0,900	0,950	0,975	0,990	0,995	0,999
1	1,32	2,70	3,84	5,02	6,63	7,88	10,83
2	2,77	4,60	5,99	7,38	9,21	10,60	13,82
3	4,11	6,25	7,81	9,35	11,34	12,84	16,27
4	5,38	7,78	9,49	11,14	13,28	14,86	18,47
5	6,63	9,24	11,07	12,83	15,09	16,75	20,52
6	7,84	10,64	12,59	14,45	16,81	18,55	22,46
7	9,04	12,02	14,07	16,01	18,48	20,28	24,32
8	10,22	13,36	15,51	17,53	20,09	21,96	26,12
9	11,39	14,68	16,92	19,02	21,67	23,59	27,88
10	12,55	15,99	18,31	20,48	23,21	25,19	29,59
11	13,70	17,28	19,68	21,92	24,72	26,76	31,26
12	14,85	18,55	21,03	23,34	26,22	28,30	32,91
13	15,98	19,81	22,36	24,74	27,69	29,82	34,53
14	17,12	21,06	23,68	26,12	29,14	31,32	36,12
15	18,25	22,31	25,00	27,49	30,58	32,80	37,70
16	19,37	23,54	26,30	28,85	32,00	34,27	39,25
17	20,49	24,77	27,59	30,19	33,41	35,72	40,79
18	21,60	25,99	28,87	31,53	34,81	37,16	42,31
19	22,72	27,20	30,14	32,85	36,19	38,58	43,82
20	23,83	28,41	31,41	34,17	37,57	40,00	45,32
21	24,93	29,62	32,67	35,48	38,93	41,40	46,80
22	26,04	30,81	33,92	36,78	40,29	42,80	48,27
23	27,14	32,01	35,17	38,08	41,64	44,18	49,73
24	28,24	33,20	36,42	39,36	42,98	45,56	51,18
25	29,34	34,38	37,65	40,65	44,31	46,93	52,62
26	30,43	35,56	38,89	41,92	45,64	48,29	54,05
27	31,53	36,74	40,11	43,19	46,96	49,64	55,48
28	32,62	37,92	41,34	44,46	48,28	50,99	56,89
29	33,71	39,09	42,56	45,72	49,59	52,34	58,30
30	34,80	40,26	43,77	46,98	50,89	53,67	59,70
40	45,62	51,80	55,76	59,34	63,69	66,77	73,40
50	56,33	63,17	67,50	71,42	76,15	79,49	86,66
60	66,98	74,40	79,08	83,30	88,38	91,95	99,61
70	77,58	85,53	90,53	95,02	100,42	104,21	112,29
80	88,13	96,58	101,87	106,62	112,32	116,32	124,77
90	98,65	107,56	113,14	118,13	124,11	128,29	137,20
100	109,09	118,49	124,34	129,56	135,80	140,16	149,38

NOTA: Para valores grandes de los grados de libertad se puede utilizar la fórmula aproximada:

$$\chi^2_{\alpha} = n \left(1 - \frac{2}{9n} + Z_{\alpha} \sqrt{\frac{2}{9n}} \right)^3,$$

siendo Z_{α} la desviación normal y n el número de grados de libertad. Así, v. gr.:

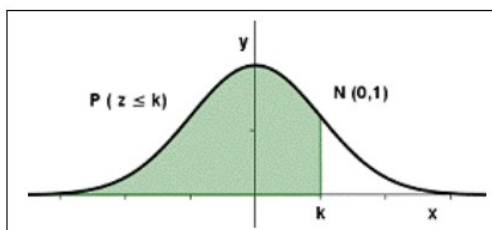
$$\chi^2_{99} = 60 \cdot \left(1 - 0,00370 + 2,326 \cdot 0,06086 \right)^3 = 60 \cdot (1,1379)^3 = 88,4$$

para el percentil 99 con 60 grados de libertad (Dixon y Massey, 1969).

Del mismo modo, se tendría que para el percentil 99 con 36 g.l.: $\chi^2_{99} = 54,324$.

3. DISTRIBUCIÓN NORMAL TIPIFICADA $N(0,1)$

Los valores de la tabla normal representan el área bajo la curva normal hasta un valor positivo de z .



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8930
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9561	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9934	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9901	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9954	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

RELACIÓN DE FIGURAS

- Fig. 1. Ampliaciones sucesivas del concepto de número.
Fig. 2. Histograma de rectángulos yuxtapuestos (1-1000).
Fig. 3. Diagrama acumulativo ascendente (1-1000).
Fig. 4. Diagrama acumulativo ascendente (1-10000).
Fig. 5. Histograma de rectángulos yuxtapuestos (1-10000).
Fig. 6. Distribución y valor de los cuantiles (1-10000).
Fig. 7. Coordenadas cartesianas rectangulares de las variables p y q .
Fig. 8. Índice de Gini y curva de Lorenz.
Fig. 9. Números primos contenidos en los primeros 10^8 números naturales.
Fig. 10. Polinomio de Euler generador de números primos.
Fig. 11. Construcción geométrica de las cuatro medias empleadas.
Fig. 12. Construcción geométrica de la conjetura ternaria de Goldbach (I).
Fig. 13. Construcción geométrica de la conjetura ternaria de Goldbach (II).
Fig. 14. Diferentes maneras de expresar N natural y par como suma de 2 números primos ($4 \leq N \leq 10^6$).

RELACIÓN DE TABLAS

- Tabla 1. Números primos gemelos de los primeros 200 naturales.
Tabla 2. Números primos (1 a 200).
Tabla 3. Formación de pares naturales N por adición de 2 números primos (1 a 200).
Tabla 4. Distribución de los primos contenidos en los 1000 primeros números naturales.
Tabla 5. Distribución de los primos contenidos en los 10000 primeros números naturales.
Tabla 6. Comprobación de Charlier.
Tabla 7. Tabla de cálculos auxiliares (1-10000).
Tabla 8. Auxiliar de cálculo del Índice de Gini.
Tabla 9. Auxiliar de cálculo del estadígrafo χ^2 (1-10000).
Tabla 10. Características de la distribución de los números primos (del 1 al 10000).
Tabla 11. Generación de números primos para $0 \leq n \leq 39$.
Tabla 12. Formación de pares naturales N por sustracción de 2 números primos (1 a 200).
Tabla 13. Distancias existentes entre 2 números primos consecutivos (1 a 200).



ÍNDICE ONOMÁSTICO DE AUTORES

ALCAIDE, A.: 25, 36, 71.

AMO, E. DE.: 11, 12, 18, 71.

AURIC: 11.

AVOGADRO, A.: 15.

BELLMAN, R.: 18, 71.

BERNOULLI, D.: 10, 11, 42.

BOWLEY, A. L.: 36.

BRAUN: 11.

CAMACHO, J.: 51, 71.

CANTOR, G.: 59.

CAUCHY, A. L.: 16, 17, 18.

CHARLIER, C.: 32, 33, 71, 84.

CHASLES, M. F.: 10.

CHEN, J.R.: 47, 55, 71, 72, 73.

CILLERUELO, J.: 47, 72.

CLARKSON, J. A.: 18, 72.

CÓRDOBA, A.: 72.

COTES, R.: 11.

D'ALEMBERT, J. R.: 16.

DEDEKIND, R.: 59.

DESCARTES, R.: 9.

DESHOILLERS, J. M.: 47.

DÍAZ, M.: 71.

DIRICHLET, P. G. L.: 51.

DIXON, W. J.: 34, 72, 82.

DRESSLER, R. E.: 11, 72.

DUSART, P.: 18, 72.

DUX, E.: 18.

EDWARDS, C.: 18, 85.

EFFINGER, G.: 47.

EINSTEIN, A.: 13.

ENRÍQUEZ VANDERKAN, A. A.: 4, 7.

ERDÖS, P.: 18, 72.

EUCLIDES: 4, 11, 13.

EULER, L.: 1, 10, 11, 12, 14, 17, 18, 48, 49, 51, 55, 71, 72, 75, 84.

FERMAT, P.: 12, 47.

FERNÁNDEZ, J.: 71, 72.

FEYNMAN, R.: 11.

FIBONACCI, L.: 12.

FISHER, R.: 36, 37, 41, 76, 77.

FRANQUET, J. M.: 2, 31, 72.
FREUND, J. E.: 27, 72.
FURSTENBERG, H.: 72.

GASTWIRTH, J. L.: 31, 77.
GAUSS, K. F.: 4, 10, 11, 17, 37, 68.
GERGONNE, J. D.: 10.
GIGERENZER, G.: 71, 72.
GINI, C.: 38, 39, 76, 84.
GOLDBACH, C.: 1, 2, 3, 4, 5, 10, 11, 24, 49, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 65, 66, 67, 68, 69, 71, 72, 73, 74, 84.
GUY, R.: 73.

HACKS, J.: 11, 73.
HARDY, G. H.: 55, 73.
HARRIS, V. C.: 11, 73.
HELFGOTT, H. A.: 10, 55, 62, 68, 73.
HILBERT, D.: 4, 7.

KAHNEMAN, D.: 70, 73.
KUMMER, E. E.: 11.

LANDAU, E.: 4, 5, 73.
LAPLACE, P. S.: 15.
LEGENDRE, A.M.: 5, 12, 14, 49, 73.
LEMKE OLIVER, R.: 40.
LITTLEWOOD, J.E.: 55, 73.
LORENZ, M. O.: 37, 38, 39, 40, 84.

MASCHERONI: 17, 75.
MASSEY, F.J.: 34, 72, 82.
MATIYASÉVICH, Y. V.: 51.
MCLAURIN, C.: 18.
METROD: 11.
MIXON: 18
MÖBIUS, A. F.: 10.
MOSER, L.: 18, 73.
MULLIN: 13.

NAIR, M.: 11, 73.
NATHANSON, M. B.: 73.
NEPER, J.: 5.

OLIVIER.: 55.
ORESME, N.: 18.
OWENS, M.: 46, 73.

PACIOLI, L.: 6.

PEARSON, K.: 2, 22, 25, 33, 35, 41, 43, 46, 58, 76, 82.
PELLETIER, F.J.: 70, 74.
PENNEY, D.: 18, 72.
PERROT, J.: 11.
PINASCO: 18.
PLÜCKER, J.: 10.
POLLACK, P.: 11, 74.
PÓLYA, G.: 11, 70.
PONCELET, J. V.: 10.
PRIETO, C.: 47, 74.
PRINGSHEIM, A.: 17, 19.
PUISEUX: 17.
PULIDO, A.: 38, 74.

RAABE, J. L.: 17.
RAMARÉ, O.: 74.
RIBENBOIM, P.: 47, 74.
RICHSTEIN, J.: 10, 68, 74.
RIELE, H.: 47.
RIEMANN, B.: 12, 47, 74.
ROLF: 17.
ROMANYCIA, M. H.: 70, 74.
ROWLAND, E.: 47, 74.

SACKS, R.: 49.
SHEPPARD, W. F.: 32, 37.
SOUNDARARAJAN, K.: 40.
SPIEGEL, M. R.: 42, 74.
STIELTJES, T. J.: 11.
STIRLING, J.: 12, 13, 15, 16, 17.
STOLZ, O.: 16.
STURGES, H.: 23, 74.
SYLVESTER, J. J.: 13.

THUE: 11.
TURING, A.: 48.

ULAM, S. M.: 49.

VALLÉE POUSSIN, CH. J. DE LA.: 74.
VINOGRADOV, I.M.: 55, 73, 74.

WANG, Y.: 47, 55, 74.
WILLIAMS, F.J.: 27, 72.
WILLIAMSON: 40.
WRIGHT, E.: 55, 73.
WUNDERLICH, M.: 11, 75.

XIAO, K.: 47, 75.

YATES, F.: 42.

YULE, G. U.: 25, 26, 36.

ZINOVIV, D.: 47.