CIENCIA

El primer

Luis Hernández Lamoneda
Centro de Investigación en Matemáticas, A.C.

El Dr. Luis Hernández Lamoneda obtuvo la Licenciatura en Matemáticas en la UNAM y su Doctorado en la Universidad de Utah, E.E.U.U. Es investigador del CIMAT y miembro del Sistema Nacional de Investigadores (Nivel II) dentro del Área de Física, Matemáticas y Ciencias de la Tierra. Es miembro de la Academia Mexicana de Ciencias.

Quiero contarles de un fenómeno curioso, descubierto en 1881 por el astrónomo y matemático Simon Newcomb y que trata sobre la distribución del primer dígito de una lista de números. En particular, quiero mostrarles el ejemplo más sencillo de una lista que satisface la distribución que Newcomb descubrió: la sucesión de potencias de 2. Esto, porque me parece un trozo muy bonito de matemáticas. Es bonito porque entrelaza ideas que surgen de sitios que aparentemente no tienen relación entre ellos: el primer dígito de un número, rotaciones por ángulo irracional, las propiedades elementales del logaritmo. También, porque casi todo lo que se necesita son matemáticas de bachillerato. Los dígitos son el 0, 1, 2, 3, 4, 5, 6, 7, 8 y 9. Todos, menos el cero, pueden aparecer al principio de un número. Así, el primer dígito del número 70368744177664, es el 7. Piensa en una lista grande (quizás infinita) de números. Por ejemplo, la lista que consta del número de habitantes de cada uno de los municipios del país. Hay 2,464 municipios. Los hay de todos tamaños; mientras que Santa Magdalena Jicotlán, Oaxaca, tiene 93 habitantes, en Iztapalapa viven 1,820,888 personas. De esta lista enorme de números, vamos a contar cuántos de ellos *comienzan* en 1, cuántos en 2, etc. Uno podría pensar que todos los dígitos del 1 al 9 aparecerían en el primer lugar, más o menos, en las mismas cantidades, pero esto no es así.

LA LEY DE BENFORD

Resulta que el 1 aparece como primer dígito en esta lista más que cualquier otro. Alrededor de un 30% de las veces. Seguido del 2 que aparece algo como 17.5% de las veces. Y así, sucesivamente disminuyendo su proporción, hasta el 9 que únicamente aparece, como primer dígito, un 4.5% de las veces. Estas proporciones, o frecuencias, satisfacen una ley logarítmica, la *Ley de Benford*. Se dice que una lista de números satisface la ley de Benford si la frecuencia con que el dígito d aparece como primer dígito está dada por $\log\left(\frac{d+1}{d}\right)$. Por ejemplo, esta ley predice que el 1 aparecerá con una frecuencia de *log 2≈0.301*, mientras que el 7 lo hará únicamente un *log (8/7) ≈0.058* de las veces. En la liga https://testingbenfordslaw.com/ se pueden encontrar ejemplos curiosos (como el de los municipios mexicanos) de listas que (aproximadamente) cumplen esta ley. Por ejemplo, la lista de las distancias a las estrellas o la lista del PIB de todos los países del mundo o el número de seguidores de Twitter. Es claro que hay listas que no satisfacen la Ley de Benford. Por ejemplo, la estatura de los humanos en centímetros o los números telefónicos de Morelos; así que saber si una lista la satisface puede ser ventajoso (por ejemplo, en [TH]) se explica cómo fue usada esta ley para descubrir fraudes cometidos por empresas neoyorquinas. Estas empresas reportaban pocos 1's y demasiados 6's). Luego es interesante saber qué propiedades garantizarían que una lista satisficiera la Ley de Benford. El artículo [TH] -que puede consultarse libremente en internet- además de reflexionar sobre propiedades generales que implican la ley, incluye otras aplicaciones interesantes en finanzas y cómputo. También la historia de su descubrimiento llama la atención. Los invito a leerlo. No es fácil imaginar como una lista de números puede satisfacer tan peculiar distribución en la cantidad de sus primeros dígitos. Así que en el resto de esta nota voy a explicar el ejemplo más simple -que además es muy bonito- de una lista (infinita, de hecho) con esta propiedad.

LAS POTENCIAS DE 2

La lista es la siguiente: 2, 4, 8, 16, 32, 64,128...; esto es, miramos a todas las potencias de 2: 2^1, 2^2, 2^3, 2^4, ... Mientras que el *último dígito* de esta sucesión es fácilmente predecible (se van repitiendo 2, 4, 8, 6, en ese orden), aparentemente nada es claro acerca del primer dígito. Por ejemplo, ni siquiera es fácil decidir si el número 7 aparece como primer dígito en esta lista (experimenta con tu celular o calculadora y verifica que 2^{46}=70368744177664). Uno puede hacer otros experimentos numéricos. Por ejemplo, con una hoja de cálculo contar las apariciones de cada dígito en el primer lugar de las primeras "muchas" potencias de 2. Para las primeras cien tenemos:

dígito d:	1	2	3	4	5	6	7	8	9
# veces que aparece d en el 1er lugar:	30	17	13	10	7	7	6	5	5

Cuadro 1: Primeras 100 potencias de 2.

Y para las primeras diez mil potencias:

dígito d:	1	2	3	4	5	6	7	8	9
Frecuencia:	30.10%	17.61%	12.49%	9.70%	7.91%	6.70%	5.79%	5.12%	4.58%

Cuadro 2: Frecuencia de d como primer dígito (primeras 10,000 potencias de 2).

Compara con la tabla de frecuencias logarítmicas dadas por la Ley de Benford.

dígito d:	1	2	3	4	5	6	7	8	9
$\log\left(\frac{d+1}{d}\right)$	30.10%	17.61%	12.49%	9.69%	7.92%	6.70%	5.80%	5.12%	4.58%

Cuadro 3: Ley de Benford

Así que *experimentalmente* vemos que la lista de potencias de 2 parece satisfacer la Ley de Benford Pero ¿cómo lo explicamos? y ¿cómo vemos que si tomamos una cantidad cada vez mayor, de potencias, entonces las frecuencias de los primeros dígitos se acercarán más y más a la Ley de Benford? De camino a entender esto, descubriremos la siguiente (¿sorprendente?) propiedad: escoge un número cualquiera, como por ejemplo, el número telefónico de la UAEM, 7773297900. Veremos que hay una infinidad de potencias de 2 que comienzan con este número. Y lo mismo sucederá con cualquier otro número que escojas (mil 7 seguidos, por ejemplo). Más aún, se puede decir que tan frecuentes son dentro de la lista (infinita) de todas las potencias de 2. Antes de dar una idea de por qué esto sucede, te muestro otro ejemplo que quizás te “mueva un poco más el tapete”.

Usando un círculo para explicar el fenómeno

Mira otra lista parecida a la anterior. Ahora consideramos a las potencias de 5: 51=5, 52=25, 53=125, 54=625.... Esta vez el último dígito es siempre el mismo, ¿pero y el primero? Veamos,

dígito d:	1	2	3	4	5	6
Frecuencia:	30.1%	17.61%	12.49%	9.69%	7.93%	6.69

Cuadro 4: Frecuencia de d como primer dígito en las primeras 10,000 potencias de 5.

Increíble, ¿verdad? Se obtienen prácticamente las mismas frecuencias que para las potencias de 2. De hecho, lo mismo sucederá con las potencias de 3, de 7 o de casi cualquier otro número (¡pero no de diez!): *a la larga* las frecuencias de cada dígito en el primer lugar de cualquier potencia *convergerán* a las dadas por la Ley de Benford. Para ello, el número que usemos para generar la lista de potencias no va a importar, sólo importa una propiedad que todos (el 2, el 5, etc.) tienen: su logaritmo (base diez) es un número *irracional*.
Vamos a ver qué tiene que ver la irracionalidad de *log2* con la distribución del primer dígito de las potencias de 2. Recuerda que un número es *irracional* si NO puede escribirse como una fracción *p/q*. Es fácil comprobar que *log2* es irracional. La prueba es por contradicción: Suponer que log *2=p/q* es equivalente a decir que log *2q=p*, que es lo mismo (por definición de logaritmo base diez) que 2q=10p. Ahora, el número de la izquierda termina en 2,4,8 o 6, mientras que el de la derecha termina en cero ¡Contradicción!
Considera un círculo, que llamaré *C*, cuyo perímetro mida 1 (así que su radio medirá $1/2\pi$). Marca el punto más a la derecha de *C* y llámalo E. Podemos pensar en todos los números como los puntos de una recta infinita. Toma esta recta y enróllala alrededor del círculo. Comienza poniendo el 0 sobre E y enrolla en la dirección contraria a las manecillas del reloj. Observa

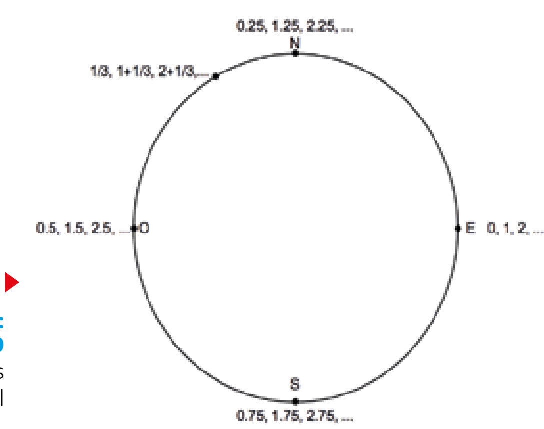


FIGURA 1: ENROLLANDO los números alrededor del círculo *C*.

que, al enroll
mismo punto
su diferencia
ros *cuya part*
fican con el n

Cada número
que hay que g
vueltas es un
mos nuevame
te decimal, e
y nos dice qu
ejemplo, el nú
que girar una
más de ella. A
llega después
(en la direcci

Por ejemplo,
medio círculo
tienen parte d
tienen parte d
Fíjate ahora e
< log 9 <
dividen a C
entre E=log 1
log 3 y así ha
Si tomamos u
y a éste le sac
es fácil ver
que su parte
mero log8.38
8388608=8.3
= 6 + log 8.3
mentales del
log8388608 '
palabras, com
ces su logarit

LOS L
C

¿Cómo usam
conocido cua
potencia de 2
logaritmo, lo
dígito comien
entre log 4 y
nuestro núme
4. Más aún, n
podemos dete
sino los prime
N cae -digam
log 4.15 y *lo*

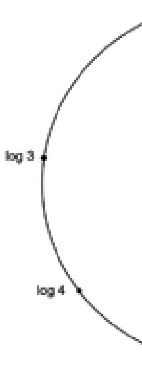


FIGURA 2: M

ACADEMIA DE CIENCIAS DE MORELOS, A.C.

ESTA PUBLICACIÓN FUE REVISADA POR EL COMITÉ EDITORIAL DE LA ACADEMIA DE CIENCIAS DE MORELOS

Ligas de interés
[TH] Ted Hill, *The*
publications/PAPE

¿Comentarios y sugerencias?, ¿P

dígito de un número

lar dos números caerán sobre el del círculo exactamente cuando sea un entero. Esto es, dos núme- *e decimal es la misma*, se identi- nismo punto de C. O piénsalo así.

	7	8	9
%	5.8%	5.11%	4.58%

o representa el *número de vueltas* girar el punto E. Si el número de entero, después de rotar acabare- nte en E. Si el número tiene par- ntonces únicamente ella importa ue ángulo habrá que girar E. Por imero 4/3=1+1/3 nos dice que hay vuelta entera y además un tercio Así que le toca el punto al que se s de rotar a E un tercio de vuelta ón contraria a las manecillas).

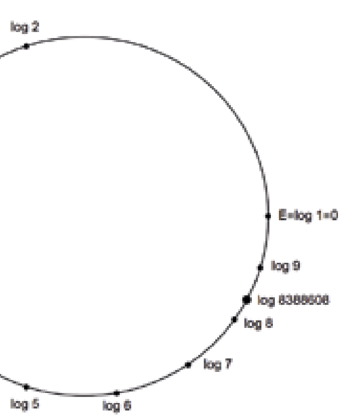
todos los enteros caen sobre E; a o de distancia (en O) están los que decimal 0.5; arriba (en N) los que decimal 0.25, etc.

en los números log 1= 0 < log 2< log10 =1 sobre el círculo C. Te en 9 "arquitos": el comprendido 1 y log 2, el de extremos log 2 y sta el último que va de log 9 a E. un número, por ejemplo 8388608, camos su logaritmo, log8388608,

decimal coincide con la del nú- 88608. Simplemente observa que 88608×106, luego log 8388608 388608, por las propiedades ele- logaritmo. Así que, sobre C, "cae" entre *log8* y *log9*. En otras no 8388608 comienza en 8, enton- mo cae en C entre log 8 y log 9.

LOGARITMOS SON LA LAVE AL ENIGMA

os esto? Si N es un número des- lquiera (en nuestro caso, alguna) entonces mirando donde cae su g N, en el círculo te dice con qué nza N. Por ejemplo, si log N cae log 5, entonces concluiremos que ero misterioso N comienza con un nirando intervalos más chiquitos, erminar no sólo el primer dígito, eros tantos como queramos: si log os- en el arquito con extremos en *g 4.16*, entonces podremos asegu-



IIRAR log N en el círculo te dice con qué dígito comienza N.

rar que los primeros tres dígitos de N son 415. Si queremos probar que existe una potencia de 2 que comience con el número telefónico de la UAEM, *basta con saber que hay uno cuyo logaritmo cae en el arquito entre log7.773297900 y log 7.773297901.*

Los números que nos interesan son los de la sucesión de potencias de 2: 2, 2^2, 2^3, 2^4, ...; si a cada uno de ellos le ``sacamos" su logaritmo, esta lista se transforma en la sucesión de múltiplos de *log2*: *log 2, 2log 2, 3log 2, 4log 2*, ... Lo anterior por la propiedad básica del logaritmo que asegura que log 2n=nlog2, para cualquier n. Pero esta sucesión de múltiplos, vista en el círculo C, no es otra cosa que la *órbita* del punto E bajo rotaciones sucesivas por un ángulo de *log2*-vueltas. Expliquemos esto un poco más. Vimos que, al enrollar la recta de los números, cada número corresponde a girar *C* tantas vueltas como el número indica. Cuando giramos *C*, *log2*-vueltas (poco menos de un tercio de vuelta) el punto E va a dar al punto marcado por log 2. Si volvemos a girar por el mismo ángulo, éste ahora irá a *2log 2*, etc.

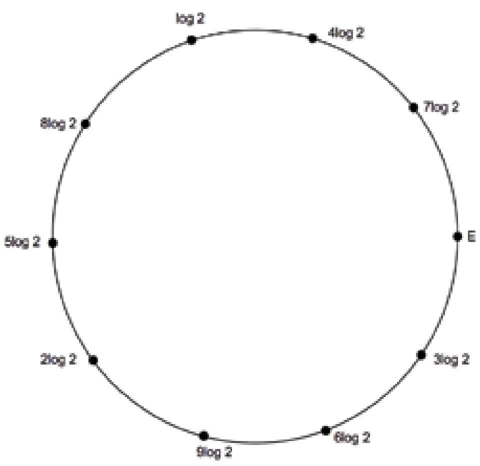


FIGURA 3: ROTANDO *log 2*-vueltas.

Es claro que, si rotamos a E por una fracción p/q de vuelta, después de aplicarla q veces, volveremos al punto inicial E. Por ejemplo, rotar 2/5 de vuelta lleva a E en el punto α que está a 144 grados de E; si aplicamos de nuevo la rotación caeremos en 2α y así sucesivamente hasta que después de 5 rotaciones volvemos a E. Sucesivas rotaciones sólo nos harán movernos sobre estos mismos 5 puntos.

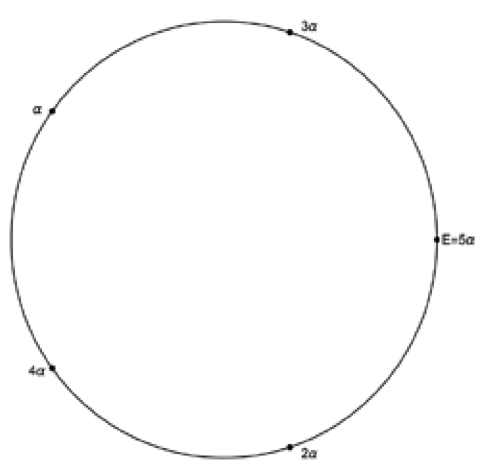


FIGURA 4: ÓRBITA de E bajo sucesivas rotaciones por un ángulo de 2/5 de vuelta

Pero recuerda que log2 es *irracional*. Es fácil ver que entonces no se van a repetir puntos sobre la órbita. Si *nlog* 2 y *mlog* 2 cayeran sobre el mismo punto en el círculo, eso querría decir que la diferencia *nlog* 2-*mlog* 2 sería un entero k, pero entonces log 2=*k/(n-m)*, contradiciendo su irracionalidad. Sin embargo, puede decirse más acerca de esta órbita.

Usando únicamente matemáticas que se estudian en bachillerato (y el Principio del Palomar) se puede dar una prueba sencilla y bonita de que la sucesión de múltiplos de cualquier irracional ξ: ξ, 2ξ,..., nξ,..., es *densa en el círculo C*. El Principio del Palomar te dice que si tienes *n+1* palomas dentro de n palomares, entonces en alguno de los palomares habrá (al menos) dos palomas. Esto significa que *cualquier arco* de *C* contendrá una infinidad de puntos de esta sucesión.

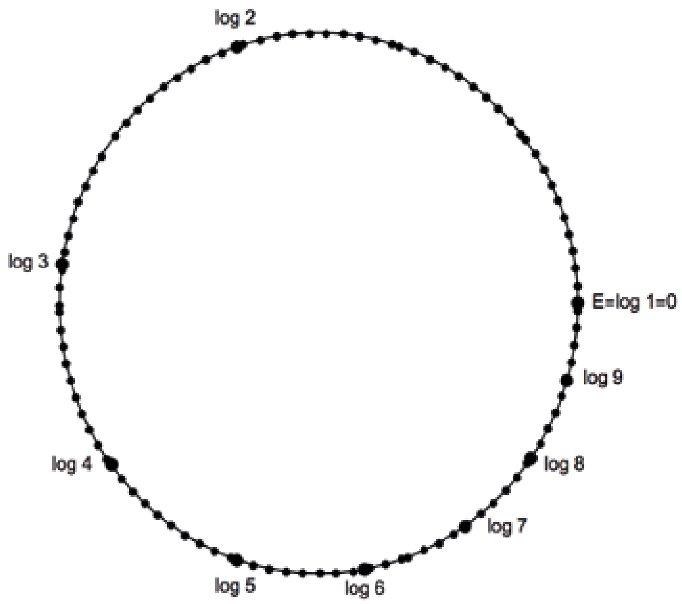


FIGURA 5: LOS primeros 100 múltiplos de *log 2*. Cuenta cuantos puntos caen en cada intervalo y compara con la tabla del Cuadro 1.

Como *log 2* es irracional, cualquier arco de círculo, contendrá una infinidad de múltiplos *nlog 2 (=log 2^n).* Pero recuerda que el arco en el que cae el logaritmo de un número te indica con qué dígito comienza dicho número (o más generalmente, con que lista de dígitos comienza el número). Queda claro entonces que si escogemos cualquier número N (por ejemplo, el teléfono de la UAEM) y miramos el arco de círculo entre *log N* y *log (N+1)* una infinidad de múltiplos de log 2 caerá en este intervalito. Esto es, una infinidad de potencias de 2 comenzará con N.

Esto nos dice mucho, pero no explica todavía por qué la lista de potencias de 2 satisface el patrón de frecuencias de la Ley de Benford. Para ello necesitamos un resultado de Hermann Weyl -ya no elemental- de hace 100 años.

Teorema [Weyl, 1917] Supón que ξ *es un número irracional y considera la sucesión de múltiplos* nξ *sobre el círculo C. Entonces, para cualquier arco **J** en C, la proporción del número de puntos nξ que caen dentro de **J** es, a la larga, exactamente la longitud de **J***. Más precisamente: si k es el número de elementos en {ξ,2ξ, …, nξ} que están dentro de J, entonces *k/n* → longitud(J), cuando n →∞.

Por ejemplo, sabemos que las potencias de 2 que comienzan en 7, son exactamente aquellas para las cuales su logaritmo está en el arco de círculo entre log 7 y log 8. Weyl nos dice que la proporción de tales potencias es, a la larga, la longitud de dicho arco, esto es

$$\log 8-\log 7=\log (8/7)\approx 0.0579919,$$

que es la frecuencia predicha por la Ley de Benford. En general, una potencia de 2 comenzará con el dígito d si su logaritmo cae entre log d y log (d+1). La longitud de tal arco es

$$\log (d+1)-\log d=\log\left(\frac{d+1}{d}\right),$$

y el Teorema de Weyl nos dice que esa será la frecuencia, a la larga, de las potencias que comiencen con el dígito *d*.

Igualmente, a la larga, la proporción de potencias de 2 que comenzarán con el número N será $\log\left(\frac{N+1}{N}\right)$. ¿Y las potencias de 5? Para ellas tendríamos que mirar la sucesión *nlog 5*, pero en vista de que log5 es también irracional, Weyl nos dice que sus múltiplos se *equidistribuyen* a lo largo del círculo. Esto es, a la larga, la proporción de ellos en cada intervalo es la longitud de dicho intervalo (igual que para las potencias de 2). Por eso, en nuestra tabla de frecuencias del primer dígito obtenemos -prácticamente- los mismos números. Y esto mismo sucederá también para las potencias de 3,4,6,7,8 y 9, pues el logaritmo base 10 de todos ellos es un número irracional -y eso es lo único que importó.

Resumiendo: ¿Por qué, entonces, la sucesión de potencias de 2 satisface la Ley de Benford? Esto es, ¿por qué el 1 aparece en primer sitio más veces que el 2 y éste más que el 3, etc.? La razón es que el primer dígito de cualquiera de ellos está determinado por la posición de su logaritmo en el círculo *C* -después de enrollar la recta de los números alrededor de *C*. Así, los que comienzan con d son aquellos cuyo logaritmo cae en el arco entre *log d* y *log (d+1)*. Weyl nos dice que, a la larga, habrá tantos que caigan allí como la longitud de dicho arco, esto es $\log (d+1)-\log d=\log\left(\frac{d+1}{d}\right)$. Pero esta es la Ley de Benford. El resultado de Weyl habla de rotaciones por un ángulo igual a un múltiplo irracional de vuelta. El logaritmo base diez que aprendemos en la prepa relaciona estas ideas geométricas de Weyl con la lista de números con que empezamos.

Echa un ojo al artículo de Ted Hill si quieres comenzar a adentrarte en la parte más probabilística de este fenómeno.

Esta columna se prepara y edita semana con semana, en conjunto con investigadores morelenses convencidos del valor del conocimiento científico para el desarrollo social y económico de Morelos. Desde la Academia de Ciencias de Morelos externamos nuestra preocupación por el vacío que genera la extinción de la Secretaría de Innovación, Ciencia y Tecnología dentro del ecosistema de innovación estatal que se debilita sin la participación del Gobierno del Estado.

first digit phenomenon, American Scientist, https://hill.math.gatech.edu/ ER%20PDFS/TheFirstDigitPhenomenonAmericanScientist1996.pdf