Comparativa, mediante la Ley de Benford, de las bases nitrogenadas existentes en la secuencia genómica del virus del ébola en huésped humano

Por **Juan Alberto Vera Herrera**

mr_javh@proton.me

Privada Tamaulipas #108, Colonia Valle de Leones, C.P. 88930

Recepción: 8 de Noviembre de 2023
Aceptación: 9 de Julio de 2024

Resumen

Las aplicaciones existentes de la ley de Benford al análisis de secuencias genómicas es casi exacta debido a la abundante cantidad de bases nitrogenadas presentes en el ADN o el ARN. Diseccionando los componentes que forman el ADN o el ARN, es decir, fosfatos, pentosas, puentes de hidrógeno y bases nitrogenadas fue posible llegar a la conclusión que la suma acumulada de diferencias (errores) para un mejor análisis es el grupo que se compone de tres (03) bases nitrogenadas y a lo que se conoce en la biología como codón. Esto podría, con nuevos análisis de virus y variantes genómicas, establecer un vínculo con la biología sintética para la manipulación específica de bases nitrogenadas en secuencias genómicas.

Abstract

The existing applications of Benford's law to genomic sequence analysis is nearly exact due to the plentiful amount of nitrogenous bases present in DNA or RNA. By dissecting the components that make up DNA or RNA, i.e. phosphates, pentoses, hydrogen bonds and nitrogenous bases, it was possible to reach the conclusion that the accumulated sum of differences (errors) for a better analysis is the group that is made up of three (03) nitrogenous bases and what is known in biology as a codon. This could, with further analysis of viruses and genomic variants, establish a link to synthetic biology for the specific manipulation of nitrogenous bases in genomic sequences.

PALABRAS CLAVE: Ley de Benford, nucleótidos, bases nitrogenadas, ADN, ARN, virus, ébola, genoma, probabilidades, distribución de dígitos, orden de magnitud.

1. INTRODUCCIÓN

Es importante enfatizar que el propósito de este artículo es utilizar uno de los tres niveles de aplicación de la ley de Benford: verificación de la aplicabilidad, detección de datos anómalos y aplicación cruzada con otros métodos. El análisis de datos al aplicar la ley de Benford permite clasificar los mismos en tres grupos de agrupaciones (clusters, del inglés), es decir, en aquellos que siguen la ley de Benford, en aquellos que no siguen la ley de Benford y en aquellos que tienen algún dato anómalo que no permite que se siga la ley. Y es en esos casos donde aparentemente no se sigue la ley de Benford, y donde a ciertas escalas o tamaños la aplicación de funciones de distribución de probabilidad como la distribución de Benford con Física Estadística a los ORFs en el estudio del genoma de células eucariotas y procariotas permite el estudio de L/S.

Así, esta investigación determinó cómo el análisis de la secuencia genómica de EBOV sigue la ley de Benford, agrupando los datos por partes de $'n'$ cantidad de datos; el siguiente paso es el análisis de datos anómalos con una extrapolación a la replicación del ADN en términos de mutaciones genéticas (como la 'crisis de información' en el estudio de los YORF), variantes en cepas o cánceres, que, aunque estos eventos biológicos no son más que producto del azar o el caos, deben ser analizados con el único fin de servir en futuras investigaciones.

Con base a los resultados obtenidos podría definirse que el estudio analítico de tres (03) bases nitrogenadas por grupo, lo que se denomina grupo 03, podría acelerar los análisis analíticos para poder determinar qué base nitrogenada debería de ser manipulada (positiva o negativamente) para fines de investigación en encontrar una vacuna o metodología por CRISPR/Cas9 en los codones.

La importancia de comenzar el análisis con el EBOV es debido a que no cuenta con una vacuna (totalmente) aprobada y a que puede tener algunas mutaciones el virus, lo que sucede en la mayoría de los organismos vivos o no vivos, diferentes expresiones génicas y la presentación de estructuras de tallo y hoja; el estudio del EBOV debido a la similitud con los rabdovirus y paramixovirus, así como la virtual similitud con el MARV es que se realiza el presente artículo con la finalidad de proveer una perspectiva diferente para poder comprender el funcionamiento de los errores genéticos o mutaciones. El estudio del EBOV es posible debido a que cuando se estudian las diferentes mutaciones genéticas o mutaciones del ARN entre los materiales y métodos en el proceso de replicación para las proteínas sea de la cápside o de alguna proteína en específico.

Contenido

2. LA LEY DE BENFORD Y SUS APLICACIONES GENERALES

Newcomb fue el primero en identificar que las hojas que contenían los primeros números de la mantisa de los libros de tablas de logaritmos eran las más buscadas y mostraban desgaste frente a los últimos números, que estaban más limpios y mostraban menos uso, mientras que Benford también observó este fenómeno. Esto condujo al uso de la ley de Benford en los procesos de contabilidad y auditoría para detectar fraudes en los libros contables, es decir, para identificar la aparición de datos anómalos (sospechosos) [1], [2].

En la actualidad con el uso de cantidades masivas de información es importante analizar la honradez y fiabilidad de la información, sobre todo de la que pertenece al campo de la ciencia, así como la identificación de datos anómalos e ir más allá en una posibilidad de aplicaciones a la biología, debido a que ya se aplica en diferentes campos de las ciencias sociales y ciencias naturales, además de las aplicaciones que se han dado en las ciencias financieras y económicas para la detección de fraude en las ganancias reportadas [1]–[7].

Para buscar una correcta aplicación interdisciplinaria de la ley de Benford, los esfuerzos que hoy se realizan para comprender su esencia, capacidad, aplicabilidad, entre otras cualidades deben de seguirse efectuando y evaluando su efectividad[3].

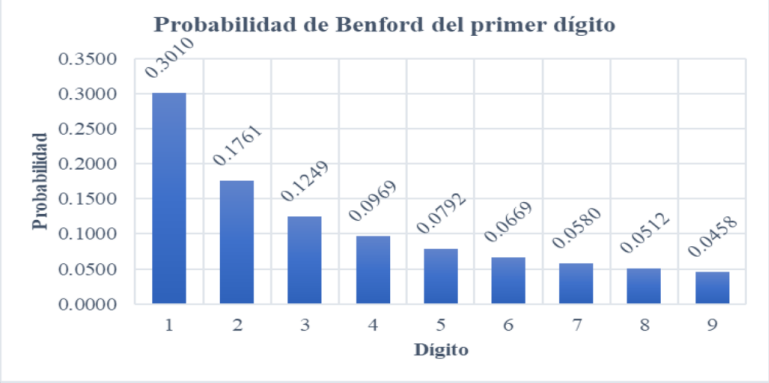


Gráfico de la distribución de probabilidades del primer dígito conforme a la ley de Benford.

Con la entrada del mundo a la Era del Zettabyte (ZB) se ha comenzado a tener una creciente preocupación por el manejo de datos, calidad, errores o falsificación de estos [2], [3]. Es por estos cambios en la información o datos que las aplicaciones de la ley de Benford y Distribución de Benford a la Física Estadística permite un avance en la Teoría de Información con las entropías de Gibbs y Shannon [5].

Contenido

Las aplicaciones que tiene la ley de Benford en la biología o ciencias de la vida (específicamente en la física) son muy importantes, algunas de ellas son: longitud del dominio de la proteína biológica, distribución de la vida, intensidad de la línea espectral, en la espectroscopía de transición atómica compleja, el periodo de vida del hadrón, relación de energía pérdida de la ralentización de la autorrotación del púlsar, entre otras investigaciones fuera del sistema solar en lo referente a estudios macros [1], [3]–[7].

Y aunque hasta el momento se pueda dar por asumido que la ley de Benford cuenta con menos "error" cuando se cuenta con una cantidad considerable de datos, que conforme a la experiencia de cada investigador y cada área puede comenzar desde 100 o 1000 datos para un buen análisis [2]; aunque es importante indicar que contar con esta cantidad de datos y la asimetría en la distribución de los datos o información no es totalmente indicativo de un comportamiento relacionado a la distribución de Benford [1].

Existe mucha información con respecto al manejo de datos informáticos, como se expuso con anterioridad, sin embargo, en la aplicación de la ley de Benford [3], [4] también existe aplicación puntualizada a estudiar el genoma con funciones de distribuciones de probabilidad de la ley de Benford tal es el caso de las aplicaciones de la física estadística (o mecánica estadística) en las YORF [5].

Por lo que, si se realiza una analogía a las condiciones en las que se reproduce el ADN y la cantidad de ácidos nucleicos que se encuentran "empaquetados" dentro del genoma humano, el cual consta de aproximadamente 3 mil millones de nucleótidos, por lo que sabemos que al tratarse de un proceso tan caótico que podría tener lugar en cuestión de minutos, dando lugar a la presencia de errores y que puede ser altamente probable en la replicación de las cadenas de ADN [5], [8], [9].

La información procesada en la ley de Benford debe por lo menos cumplir con las cuatro condiciones generales, las cuales son:

1. La cantidad de datos es lo suficientemente grande para representar todas las muestras.

2. La cantidad de datos es ilimitada.

3. Los datos se forman naturalmente, es decir, sin influencia de factores humanos o mínima influencia.

4. Los datos no pueden ser altamente acoplados y cohesivos.

Tres tipos de datos que se analizan con la ley de Benford son:

- Tamaño de la población de animales.
- Tamaño de la población de plantas.
- Cantidad de seguidores, me gusta o RT de un determinado influencer.

Además, otra aplicación de la ley de Benford con una integración de la física estadística es el análisis del genoma por codones, sean células procariotas o eucariotas, en donde la aplicación de las funciones de distribución de probabilidad se puede dar de forma simple o por superposición de dos funciones de distribución de probabilidad mutuamente excluyentes [5], [6], [10].

Por lo que las aplicaciones de la ley de Benford al estudio del genoma, ADN, ARN y proteínas, aunque no ha sido ampliamente estudiado si ha tenido algunos avances en al respecto, lo anterior significa grandes avances para las aplicaciones biofísicas [5]–[7]. Sin embargo, también es importante indicar que la ley de Benford no siempre se cumple en los casos del estudio genómico y puede deberse a mutaciones o alteraciones forzadas por agentes externos [1].

3. IMPACTO DEL ÉBOLA Y OTROS VIRUS EN DIVERSOS SECTORES PROFESIONALES Y EVOLUTIVOS

El funcionamiento básico, sin procesos extremadamente caóticos, son las células y estas se dividen en procariotas y eucariotas, las cuales se encuentran divididas en los diferentes reinos:

- Procariotas: Animalia, Fungi, Plantae y Protista
- Eucariotas: Bacteria y Archaea

Sin embargo, los virus no se pueden clasificar en lo anterior, debido a que no se consideran organismos vivos, pero necesitan de estos para poder replicarse[11], [12]. Por lo que los virus patogénicos con ARN genómico son un reto para el sistema de salud mundial por el potencial de convertirse, para los seres humanos, en epidemias o pandemias [9], [13], [14].

Aunque el virus SARS-CoV-2 es un coronavirus y el EBOV es un filovirus, la semejanza la tienen en que se basan en ARN y sobre todo en que han tenido un impacto en la salud de los seres humanos o han llegado a los humanos por efecto de la zoonosis [9]; además el EBOV también tiene similitudes con el MARV, con virus de Sudán y el virus de Bundibugyo [15].

En la actualidad, el impacto del cambio climático al sistema de salud es significativo, esto debido a las mutaciones que tienen los virus por el aumento de la temperatura en los virus de ARN altamente patógenos, por ejemplo, los virus de: influenza, zika, encefalitis y dengue [9]. El análisis de secuencia genética del EBOV se encuentra similarmente organizado a los rabdovirus y paramixovirus y su similitud es virtualmente muy estrecha al MARV , el cual produce reacciones en los humanos similares al EBOV [8], [13], [15].

También es importante indicar que debido a la más reciente pandemia a causa del virus SARS-CoV-2 que desarrolla la enfermedad CoVID-19 permitió relacionar con el virus del ébola estudios que permitieron observar las alteraciones al sistema inmunológico entre el EBOV y el SARS-CoV-2 en los cuerpos humanos, debido a que se tenía un avance en las terapias antirretrovirales y vacunas con el EBOV se logró algo contra el SARS-CoV-2, evitando así la EVD, y el CoVID-19 [9], [14].

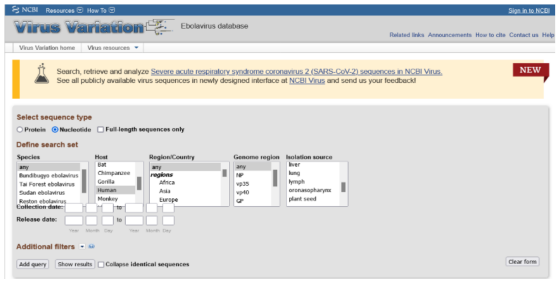
La importancia de identificar las mutaciones en los virus como el SARS-CoV-2 y el EBOV es para que el sistema inmunológico del ser humano pueda contar con los anticuerpos necesarios cuando se susciten las mutaciones y que en dado caso de que el cuerpo humano no pueda producir anticuerpos, entonces producir vacunas con base en el ARN [9], [13], [14].

La importancia de estudiar el EBOV se debe también a que es considerado una ITS debido a su prevalencia en los fluidos seminales o vaginales de las personas (hombre o mujeres), además de encontrarse en otros fluidos corporales, ejemplo: la leche materna [9]; además hasta el año 2022 la mayor parte de los tratamientos para el EBOV y el MARV son paliativos ya que no existen vacunas ni tratamientos autorizados disponibles para las infecciones por EBOV y MARV [15].

4. DESCRIPCIÓN DEL MÉTODO PARA EL ANÁLISIS DE LA LEY DE BENFORD EN LA SECUENCIA GENÓMICA DEL ÉBOLA

Las aplicaciones de la Ley de Benford se encuentran mayormente orientadas a los dígitos, es decir, y si bien esto es correcto por la naturaleza matemática de la ley, se hace una analogía a los datos cualitativos de las bases nitrogenadas en el ADN/ARN.

Se inicio con el análisis del genoma del ADN del EBOV, siendo que la base de datos *Ebolavirus database de Virus Variation de NCBI (National Center for Biotechnology Information, U.S. National Library of Medicine)* proporcionó una cantidad de 18,957 bases nitrogenadas. Es crucial indicar que es ADN y no ARN debido a que la base de datos proporciono entre las bases nitrogenadas a la timina (T) y no al uracilo (U) que es representativo de una cadena de ARN, esto se debe a que la secuencia genómica se obtuvo a partir de un huésped, en inglés, humano.



Base de datos del National Center for Biotechnology Information, U.S. National Library of Medicine.

La información se obtiene en un archivo txt con separación de grupos de 10 bases nitrogenadas y cada fila contiene 6 grupos, es decir, cada fila contiene 60 bases nitrogenadas.

Composición del archivo txt del EBOV descargado del portal del NCBI.

Contenido

Se procedió a filtrar la información para eliminar los caracteres no funcionales (espacios) para el análisis de la información mediante un algoritmo en Python, a través del cual se indicó la cantidad de bases nitrogenadas que debería de tener cada fila, desde 1 hasta 10 bases nitrogenadas. Esto es con base a la cantidad de nucleótidos que se dan por grupo que indican un exón y que es de 10 nucleótidos, los exones a su vez codifican proteínas. Es importante indicar que un nucleótido está conformado por una base nitrogenada, un fosfato y una pentosa.

Conforme a la ley de Benford, del primer dígito, se procede a solamente tomar la primer base nitrogenada de cada grupo, previamente se verificó el orden de las bases nitrogenadas para poder asignar el orden que le podría corresponder con base a los números cardinales para una posterior comparativa con la ley de Benford. Además, en las tablas se muestra la diferencia (decimal y porcentual) entre la probabilidad de encontrar dicho aspecto o base nitrogenada conforme al dígito que le correspondería en la ley de Benford.

ID	ASPECTO	FRECUENCIA	PROB	PROB, %	PROB BENFORD	PROB BENFORD, %	DIF	DIF, %	SSD
1	Fosfato	18957	0.2857	28.57%	0.3010	30.10%	0.0153	1.53%	0.0002
2	Pentosa	18957	0.2857	28.57%	0.1761	17.61%	0.1096	10.96%	0.0120
3	Puente H	9478	0.1429	14.29%	0.1249	12.49%	0.0179	1.79%	0.0003
4	a	6060	0.0913	9.13%	0.0969	9.69%	0.0056	0.56%	0.0000
5	t	5110	0.0770	7.70%	0.0792	7.92%	0.0022	0.22%	0.0000
6	c	4035	0.0608	6.08%	0.0669	6.69%	0.0061	0.61%	0.0000
7	g	3752	0.0565	5.65%	0.0580	5.80%	0.0014	0.14%	0.0000
SUMA		**66349**	**1.00**	**100.00%**	**0.90**	**90.31%**	**0.16**	**15.82%**	**0.0126**

Tabla I. Cuantificación de todos los aspectos que se encuentran en la cadena de ADN analizada.

ID	ASPECTO	FRECUENCIA	PROB	PROB, %	PROB BENFORD	PROB BENFORD, %	DIF	DIF, %	SSD
1	Fosfato	18957	0.3333	33.33%	0.3010	30.10%	0.0323	3.23%	0.0010
2	Pentosa	18957	0.3333	33.33%	0.1761	17.61%	0.1572	15.72%	0.0247
3	a	6060	0.1066	10.66%	0.1249	12.49%	0.0184	1.84%	0.0003
4	t	5110	0.0899	8.99%	0.0969	9.69%	0.0071	0.71%	0.0000
5	c	4035	0.0710	7.10%	0.0792	7.92%	0.0082	0.82%	0.0001
6	g	3752	0.0660	6.60%	0.0669	6.69%	0.0010	0.10%	0.0000
SUMA		**56871**	**1.00**	**100.00%**	**0.85**	**84.51%**	**0.22**	**22.42%**	**0.0262**

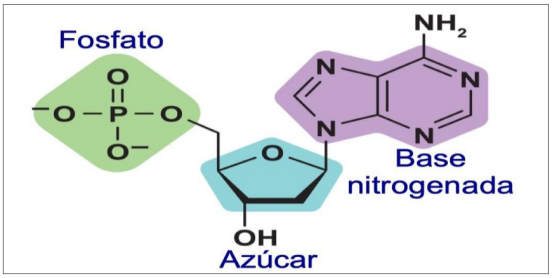
Tabla II. Cuantificación de los aspectos que se encuentran en la cadena de ADN analizada, a excepción de los puentes de hidrógeno.

ID	BASE	FRECUENCIA	PROB	PROB, %	PROB BENFORD	PROB BENFORD, %	DIF	DIF, %	SSD
1	a	6060	0.3197	31.97%	0.3010	30.10%	0.0186	1.86%	0.0003
2	t	5110	0.2696	26.96%	0.1761	17.61%	0.0935	9.35%	0.0087
3	c	4035	0.2129	21.29%	0.1249	12.49%	0.0879	8.79%	0.0077
4	g	3752	0.1979	19.79%	0.0969	9.69%	0.1010	10.10%	0.0102
SUMA		**18957**	**1.00**	**100.00%**	**0.70**	**69.90%**	**0.30**	**30.10%**	**0.0270**

Tabla III. Cuantificación de las bases nitrogenadas que se encuentran en la cadena de ADN analizada.

Contenido

La cantidad de fosfatos y pentosas (azúcar de 5 átomos de carbono) es por cada nucleótido y como se indicó con anterioridad, un nucleótido está conformado por un fosfato, una pentosa y una base nitrogenada. La adenina (A) le corresponde el dígito 01, la timina (T) le corresponde el dígito 02, la citosina (C) le corresponde el dígito 03, la guanina (G) le corresponde el dígito 04.



Conformación de un nucleótido [16].

Con base en el análisis anterior, se procedió a utilizar solamente las bases nitrogenadas para realizar una comparación de encontrar las bases nitrogenadas en la primer columna (analogía al primer dígito) por cada grupo y contrastarlo con la probabilidad del primer dígito de la ley de Benford. Para fines de ejemplificar, en este artículo se hará mención solamente al grupo 03 (03 bases nitrogenadas por fila), sin embargo, se mostrarán los resultados para todos los 10 grupos.

FILA	BASES NITROGENADAS		
1	g	g	a
2	c	a	c
3	a	c	a
4	a	a	a
5	a	g	a
. . .	. . .	. . .	. . .
6317	t	g	t
6318	g	t	g
6319	t	c	c

Tabla IV. Ejemplificación del acomodo de las bases nitrogenadas para el grupo 03.

GRUPO 03									
ID	BASE	FRECUENCIA	PROB	PROB, %	PROB BENFORD	PROB BENFORD, %	DIF	DIF, %	SSD
1	a	1984	0.2093	20.93%	0.3010	30.10%	0.0917	9.17%	0.0084
2	t	1669	0.1761	17.61%	0.1761	17.61%	0.0000	0.00%	0.0000
3	c	1385	0.1461	14.61%	0.1249	12.49%	0.0212	2.12%	0.0004
4	g	1281	0.1351	13.51%	0.0969	9.69%	0.0382	3.82%	0.0015
SUMA		**6319**	**0.67**	**66.66%**	**0.70**	**69.90%**	**0.15**	**15.11%**	**0.0103**

Tabla V. Ejemplificación del acomodo de las bases nitrogenadas para el grupo 03.

Contenido

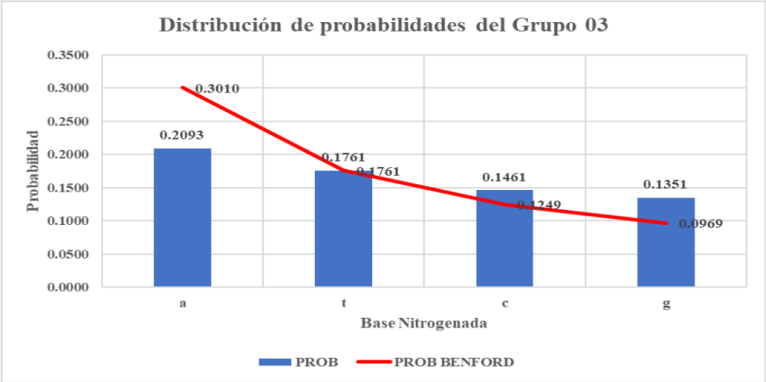


Gráfico comparativo de la distribución de probabilidades de las bases nitrogenadas en el grupo 03 versus la probabilidad del primer dígito de la ley de Benford.

Este mismo análisis fue realizado para los 09 grupos restantes (01, 02, 04, 05, 06, 07, 08, 09 y 10) por lo que se muestran los resultados para los 10 grupos en la tabla VI.

BASE NITROGENADA	PROBABILIDAD POR GRUPO										PROBABILIDAD BENFORD
	01	02	03	04	05	06	07	08	09	10	
a	0.3197	0.3184	0.2093	0.1572	0.1297	0.1074	0.0928	0.0787	0.0712	0.0646	0.3010
t	0.2696	0.2730	0.1761	0.1384	0.1053	0.0864	0.0743	0.0691	0.0565	0.0518	0.1761
c	0.2129	0.2101	0.1461	0.1075	0.0840	0.0735	0.0622	0.0532	0.0484	0.0405	0.1249
g	0.1979	0.1984	0.1351	0.0970	0.0811	0.0660	0.0564	0.0491	0.0461	0.0431	0.0969

Tabla VI. Compendio de las probabilidades por grupo versus la probabilidad de Benford asignada a cada base nitrogenada.

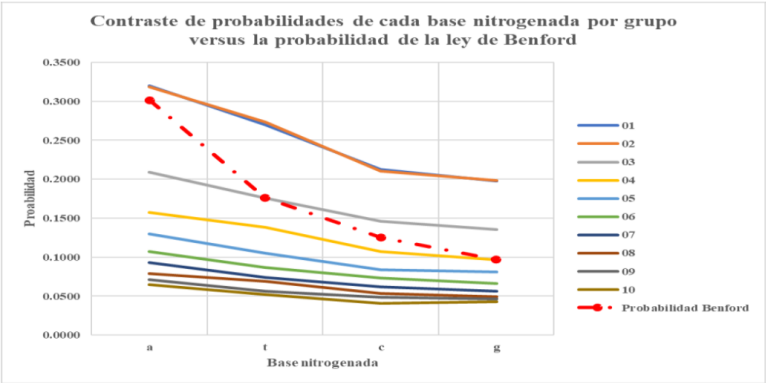


Gráfico comparativo de las probabilidades de las bases nitrogenadas por grupo (01 – 10) versus la probabilidad del primer dígito de la ley de Benford.

Se tomó la diferencia (error) de cada base nitrogenada en cada grupo y se realizó la siguiente tabla y gráfico para ver como evolucionaba la diferencia entre la probabilidad de cada base nitrogenada con respecto al valor de la probabilidad de la ley de Benford por el dígito asignado a cada base nitrogenada.

GRUPO	a	t	c	g
1	0.0186	0.0935	0.0879	0.1010
2	0.0174	0.0969	0.0852	0.1015
3	0.0917	0.0000	0.0212	0.0382
4	0.1438	0.0377	0.0174	0.0000
5	0.1714	0.0708	0.0410	0.0158
6	0.1936	0.0897	0.0514	0.0309
7	0.2082	0.1018	0.0627	0.0405
8	0.2223	0.1070	0.0718	0.0479
9	0.2298	0.1195	0.0765	0.0508
10	0.2365	0.1243	0.0844	0.0538

Tabla VII. Compendio de las diferencias (errores) de la probabilidad de cada base nitrogenada versus la probabilidad de la ley de Benford.

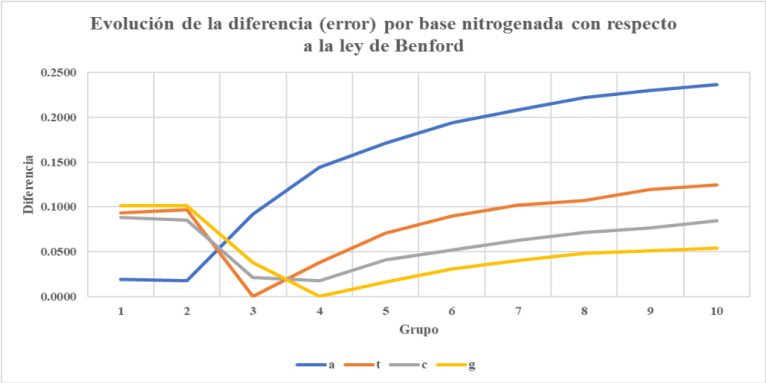


Gráfico de la evolución de la diferencia (error) de la probabilidad de cada base nitrogenada con respecto a la probabilidad de la ley de Benford.

Para una ejemplificación grupal se realizó la suma de las diferencias existentes entre cada base nitrogenada y la probabilidad de Benford para cada dígito asignado a cada base nitrogenada por grupo, es decir, para el grupo 01 se sumó la diferencia que se tuvo entre: la probabilidad de la adenina (A) y la probabilidad de Benford para el dígito 01 más la probabilidad de la timina (T) y la probabilidad de Benford para el dígito 02 más la probabilidad de la citosina (C) y la probabilidad de Benford para el dígito 03 más la probabilidad de la guanina (G) y la probabilidad de Benford para el dígito 04; lo anterior da una suma acumulada de errores de pronóstico debido a que el valor real es la probabilidad de la base nitrogenada y el valor de pronóstico es el valor de la ley de Benford.

Contenido

GRUPO	DIFF GRUPAL	SSD
1	0.3010	0.0270
2	0.3010	0.0273
3	0.1511	0.0103
4	0.1990	0.0224
5	0.2989	0.0363
6	0.3656	0.0491
7	0.4132	0.0593
8	0.4489	0.0683
9	0.4767	0.0755
10	0.4989	0.0814

Tabla VIII. Compendio de la suma acumulada de diferencias (errores) de la probabilidad de cada base nitrogenada versus la probabilidad de la ley de Benford, por grupo.

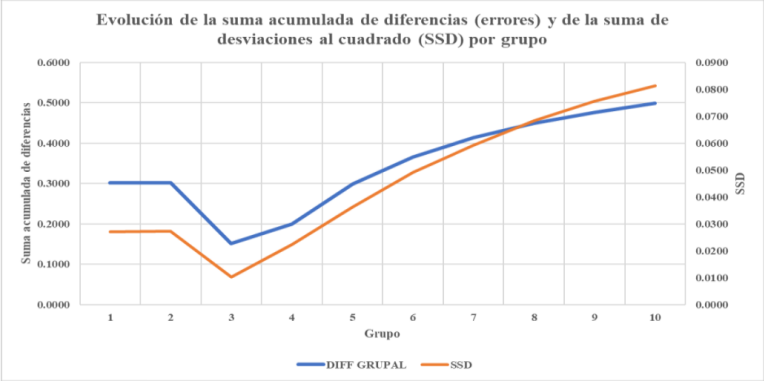


Gráfico de la evolución de la suma acumulada de diferencias (errores) y de la suma de desviaciones al cuadrado (SSD) de la probabilidad de cada base nitrogenada con respecto a la probabilidad de la ley de Benford, por grupo.

5. DISCUSIÓN DE LOS RESULTADOS

A través de las tablas I, II y III se pudo observar que la probabilidad si varía al contrastarle con los valores obtenidos para los dígitos en la ley de Benford, pero que en proporciones siguen siendo invariantes, además las mutaciones genéticas tienen más probabilidad de suscitarse en las bases nitrogenadas que en los fosfatos, las pentosas o en el todo (nucleótido). Lo que conlleva a recordar que el ciclo celular puede completarse en 12 horas, entonces, la replicación de cada nucleótido se da en 0.0000144 segundos/nucleótido (43,200/3,000,000,000) por lo tanto se comprende que es un proceso caótico y que puede conllevar a errores, lo que da presencia a mutaciones, cáncer, variantes, entre otros eventos.

En la tabla VI y su gráfico correspondiente se observa que las probabilidades para cada base nitrogenada de los grupos 01 y la 02 se encuentran por encima de los valores de probabilidad de la ley de Benford, mientras que las probabilidades para cada base nitrogenada de los grupos 04 a la 10 quedan por debajo de la ley de Benford, siendo que las probabilidades de cada base nitrogenada para el grupo 03 son las que tienen más similitud a las probabilidades de la ley de Benford.

De la tabla VII y su gráfico se puede observar que la diferencia de la probabilidad de la adenina (A) va en aumento conforme aumenta la cantidad de bases nitrogenadas, es decir, conforme aumenta el grupo. La timina (T) y citosina (C) tienen una mínima diferencia (error) en el grupo 03, mientras que la guanina (G) tiene su mínima diferencia en el grupo 04.

Se observo en la tabla VIII y su gráfico que la suma acumulada de diferencias (errores) y de la suma de desviaciones al cuadrado (SSD) es la misma para el grupo 01 y 02, mientras que existe un valor mínimo cuando la cantidad de bases nitrogenadas existentes en el grupo es de 03, es decir, un grupo 03; de ahí en adelante (grupos 04 al grupo 10) va en aumento la suma acumulada de diferencias (errores) y en la suma de desviaciones al cuadrado (SSD), siendo está última utilizada para determinar el valor de Chi-cuadrada.

6. CONCLUSIONES

Con base en esta investigación se puede apreciar en la tabla I y II que el orden que tomen el fosfato o la pentosa para su analogía con la ley de Benford podría decirse que es prácticamente indiferente debido a que el valor podría tener errores porcentuales menores al 5 % e indicaría conforme a la ley de Benford que es poco probable que se den variaciones (mutaciones) en estos aspectos del ADN. Se observo que tanto en la evolución de las diferencias (errores) por base nitrogenada (tabla VII) y en la evolución de la suma acumulada de diferencias (errores) por grupo y de la suma de desviaciones al cuadrado (SSD) (tablas VIII) que se tienen diferencias mínimas con respecto a los valores de probabilidad de la ley de Benford del primer dígito asignado a la base nitrogenada, lo que corresponde al grupo 03, lo anterior podría ser consistente a que tres (03) nucleótidos conforman un codón o un anticodón.

Se sabe que un codón es una secuencia de ADN o ARN compuesta por tres nucleótidos (fosfato, pentosa y base nitrogenada) que conforma una unidad de información genómica que codifica para un determinado aminoácido o señaliza la terminación de una proteína.

Con base en lo anterior podría indicarse que, aunque la suma acumulada de diferencias (errores) para el grupo 03 [tres (03) bases nitrogenadas y por lo tanto tres (03) nucleótidos] sea de 0.1511 (15.11 %) sea superior al 0.05 (5 %) que exhorta la ciencia para temas de investigación, aspecto que también se observa con los valores de la suma de desviaciones al cuadrado (SSD) 0.0103 para el grupo 03 y menores a este valor para cada una de las bases nitrogenadas y cuando los valores SSD son cercanos indican que los datos tienen un ajuste ideal.

Sin embargo, podría prescindirse de considerar a los fosfatos, pentosas y puentes de hidrógeno del análisis y solamente considerar a las bases nitrogenadas, sin embargo, al ser puramente una investigación absolutamente analítica y no teórica es esencial corroborar esta conclusión con otras mutaciones del virus, análisis genómico en otro huésped y otros virus para poder llegar a una conclusión generalizada.

7. RECOMENDACIONES Y/O SUGERENCIAS

Las aplicaciones que este artículo podría llevar a cabo en el estudio del ADN y ARN (ya sea de los diferentes reinos de seres vivos, virus o incluso priones) y tendrían aplicaciones prácticas, es decir, para reducir síntesis y/o procesos de laboratorio con un previo análisis fisicomatemático o analítico.

Debido a que en ocasiones se realizan diferentes procedimientos de laboratorio en la síntesis para obtener una mutación, la cual es al azar o "esperada"; pudiendo partir de analizar cada variante genética del virus solamente con la cantidad de bases nitrogenadas por grupo y la primer base nitrogenada de cada grupo.

Se sugiere contrastar todos los linajes de cada tipo (o especie) de virus y verificar la diferencia entre cada cadena de ADN/ARN y correlacionar las variables ambientales o animales (hospedero o reservorio) donde ocurrió la mutación para concluir por qué esta variación en la probabilidad de Benford fue causada por un factor externo que ocurrió y así poder hacer una analogía de que esta variación en la probabilidad de Benford no es un error, sino un efecto (un por qué) que tiene una causa.

Contenido

Se sugiere aquí que a través del análisis de diferentes secuencias genómicas en diferentes huéspedes y variantes del EBOV u otros virus, así como en diferentes regiones se podría llegar a determinar qué base nitrogenada sea más probable de manipular para poder tener una "corrección de un error" al funcionamiento y poder tener o un cese de la actividad del virus, pero esto podría ser anexando biología sintética para la síntesis y ensamblaje del ADN en lugares específicos en donde se sabe que hay más libertad de "manipulación", siendo una posible analogía a la manipulación de los datos numéricos que se hacen en los registros contables, si la probabilidad del dígito buscado se aleja (positiva o negativamente) del valor de probabilidad de la ley de Benford es que fue manipulado; con aplicaciones de biología sintética podría determinarse que en donde la ley de Benford no cumple se puedan modificar esos codones (que inician con adenina, citosina o guanina).

8. AGRADECIMIENTOS

Como autor principal del presente artículo externo mi más sincero agradecimiento al árbitro anónimo que verificó este manuscrito. Sus comentarios, sugerencias y recomendaciones fueron un gran aporte para potenciar la calidad del presente trabajo. Su experiencia en el ámbito matemático fortaleció sustancialmente este artículo científico. Con base en esto es que se agradece el tiempo y esfuerzo dedicados al proceso de revisión.

9. ABREVIACIONES

ADN: Ácido desoxirribonucleico.

ARN: Ácido ribonucleico.

cDNA: ADN complementario o ADN copia.

CoVID-19: Enfermedad "Coronavirus Infection Disease 2019"
provocada por SARS-CoV-2.

EBOV: Virus del Ébola.

EVD: Enfermedad del Virus del Ébola.

ITS: Infección de Transmisión Sexual.

L/S: Sistemas vivos.

MARV: Virus de Marburgo.

mRNA: ARN mensajero.

ncDNA: ADN no codificante.

ORF: Open Reading Frameworks.

pdf: Probability Distribution Function ó
Función de Distribución de Probabilidad.

pORF: Probabilidad de ORF.

pre-mRNA: pre ARN mensajero.

SARS-CoV-2: Coronavirus 2 del Síndrome Respiratorio Agudo Severo.

SSD: Suma de desviaciones al cuadrado.

yORF: Variable independiente (y) para ORF.

Contenido

Referencias

[1] A. E. Kossovsky, "On the Mistaken Use of the Chi-Square Test in Benford's Law", Stats (Basel), vol. 4, núm. 2, pp. 419–453, may 2021, doi:10.3390/stats4020027.

[2] W. Bernardino da Silva, S. K. de Melo Travassos, y J. I. de Freitas Costa, "Using the Newcomb-Benford Law as a Deviation Identification Method in Continuous Auditing Environments: A Proposal for Detecting Deviations over Time.", RevistaContabilidade&Finanças, vol. 28, núm. 73, pp. 11–26, 2017, doi:10.1592/1808- 057x201702690.

[3] F. Li, S. Han, H. Zhang, J. Ding, J. Zhang, y J. Wu, "Application of Benford's law in Data Analysis", en Journal of Physics: Conference Series, Institute of Physics Publishing, mar. 2019. doi:10.1088/1742-6596/1168/3/032133.

[4] M. Sambridge, H. Tkalčić, y A. Jackson, "Benford's law in the natural sciences", Geophys Res Lett, vol. 37, núm. 22, nov. 2010, doi:10.1029/2010GL044830.

[5] J. L. Friar, T. Goldman, y J. Pérez-Mercader, "Genome sizes and the Benford distribution", PLoS One, vol. 7, núm. 5, may 2012, doi:10.1371/journal.pone.0036624.

[6] C. G. Wohl, "Benford's Law", 2011.

[7] K. B. Lee, S. Han, y Y. Jeong, "COVID-19, flattening the curve, and Benford's law", Physica A: Statistical Mechanics and its Applications, vol. 559, dic. 2020, doi:10.1016/j.physa.2020.125090.

[8] A. Sanchez, M. P. Kiley, B. P. Holloway, y D. D. Auperin, "Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus", Virus Res, vol. 29, núm. 3, pp. 215–240, 1993, doi:10.1016/0168- 1702(93)90063-S.

[9] C. R. Dhanya et al., "RNA Viruses, Pregnancy and Vaccination: Emerging Lessons from COVID-19 and Ebola Virus Disease", Pathogens, vol.11, núm.7. MDPI, el 1 de julio de 2022. doi:10.3390/pathogens11070800.

[10] NIH, "Codón", National Human Genome Research Institute. Consultado: el 31 de octubre de 2023. [En línea]. Disponible en: https://www.genome.gov/es/genetics-glossary/Codon#

[11] M. Borrell y I. Gortazar, Eds., ¿Qué quieres saber de la ciencia? Barcelona: Océano, 1979.

Contenido

[12] R. Sohn, "Meet the 'exclusome': A mini-organ just discovered in cells that defends the genome from attack.", Live Science.

[13] S. Bach et al., "Identification and characterization of short leader and trailer RNAs synthesized by the Ebola virus RNA polymerase", PLoSPathog, vol. 17, núm. 10, oct. 2021, doi:10.1371/journal.ppat.1010002.

[14] D. G. Ithinjiet al., "Multivalent viral particles elicit safe and efficient immunoprotection against Nipah Hendra and Ebola viruses", NPJ Vaccines, vol. 7, núm. 1, dic. 2022, doi:10.1038/s41541-022-00588-5.

[15] M. H. Abir et al., "Pathogenicity and virulence of Marburg virus", Virulence, vol. 13, núm. 1. Taylor and Francis Ltd., pp. 609–633, 2022. doi:10.1080/21505594.2022.2054760.

[16] Significados.com, "Qué es un nucleótido". Consultado: el 31 de octubre de 2023. [En línea]. Disponible en: https://www.significados.com/nucleotido/