



Universidad Michoacana de San Nicolás de Hidalgo

**Facultad de Ciencias Físico-Matemáticas
"Mat. Luis Manuel Rivera Gutiérrez"**

**"Sobre la ley Benford del dígito
más significativo"**

TESIS

que para obtener el título de

Licenciado en Ciencias Físico-Matemáticas

presenta

José Alfredo Cervantes Guzmán

director de tesis

Dr. Eugenio P. Balanzario Gutiérrez

Diciembre 2011

**SOBRE LA LEY BENFORD DEL DÍGITO
MÁS SIGNIFICATIVO**

ÍNDICE ANALÍTICO

Capítulo 1

INTRODUCCIÓN

§1.1 El descubrimiento de la ley Benford	2
§1.2 Algunas explicaciones de la ley Benford	5
§1.3 Hipótesis clave	7
§1.4 Invarianza bajo la escala y bajo la base	10
§1.5. Aplicaciones	14

Capítulo 2

CONCEPTOS BÁSICOS

§2.1 Conceptos	17
----------------------	----

Capítulo 3

CONDICIONES SUFICIENTES PARA SATISFACER BENFORD

§3.1 Condiciones para satisfacer la ley Benford	27
§3.2 Benford de una manera aproximada	36
§3.3 Contraejemplo de la ley Benford	38

Capítulo 4

UN MODELO SIMPLE PARA LA LEY BENFORD

§4.1 Introducción	42
§4.2 Benford en el grupo $\mathbf{Z}_2 \times \mathbf{Z}_2$	43
§4.3 Benford en el grupo $\mathbf{R} = \mathbf{T} \times \mathbf{Z}$	47

Capítulo 5	
SIMULANDO FRAUDE	
§5.1 Prueba χ^2	55
§5.2 Simulaciones	59
APÉNDICE 1	
Simula1	66

Resumen

Si observamos un conjunto de datos numérico y nos preguntan ¿Cuál es la probabilidad de que el número 1 sea el primer dígito significativo? de manera casi inmediata contestaríamos que es $1/9$, pues teniendo nueve opciones, $\{1, \dots, 9\}$, nuestra intuición nos haría pensar que los nueve dígitos tienen la misma probabilidad. Pero esto es erróneo, en realidad es $\log_{10}(2)$. Estas probabilidades nos la proporciona la ley Benford, que en un principio fue considerada como un fenómeno matemático “curioso”, pero en la década de los 70's se empezó a utilizar en aplicaciones, principalmente en la computación, por lo que fue necesario obtener un mejor entendimiento de esta ley. No fue hasta los 90's que Hill, utilizando la teoría de Lebesgue, logro la justificación matemática de la ley Benford, pero no entendió bien la naturaleza de esta ley.

El objetivo de este trabajo es entender la naturaleza de este fenómeno, para esto nuestra herramienta de trabajo es el análisis de Fourier, con el que encontraremos condiciones suficientes para que una variable aleatoria satisfaga la ley Benford. Después proponemos un modelo bastante sencillo en el grupo $\mathbf{Z}_2 \times \mathbf{Z}_2$ para entender el comportamiento de la ley Benford en este grupo. Luego, generalizaremos la idea del modelo a los reales positivos, cumpliendo con nuestro objetivo. También hacemos algunas simulaciones para ver que tan útil puede ser la ley Benford en conjunto con la prueba χ^2 para detectar anomalías en conjuntos de datos.

Capítulo 1

Introducción

La ley Benford sobre el primer dígito significativo describe una propiedad que comparten muchos conjuntos de datos. En particular, de esta ley podemos deducir que la frecuencia de la aparición del primer dígito significativo no es uniforme, es decir, no es $1/9$ como esperaríamos.

Para entender lo que dice la ley Benford nos apoyaremos en las funciones $D_\beta^i(\xi)$, que al evaluarlas se obtiene el i -ésimo dígito significativo de ξ escrito en base β . Por ejemplo $D_{10}^1(0.00426) = 4$, $D_{10}^2(0.00426) = 2$, $D_5^1(421) = 3$ y $D_5^3(421) = 4$ ($421 = 3141$ en base 5).

Escribiremos $D^i = D_{10}^i$, para referirnos a las funciones del i -ésimo dígito en base 10. Note que $D_\beta^1 : \mathbf{R}^+ \rightarrow \{1, \dots, \beta - 1\}$ y $D_\beta^i : \mathbf{R}^+ \rightarrow \{0, 1, \dots, \beta - 1\}$, para $i \neq 1$, donde $\mathbf{R}^+ = (0, \infty)$.

Lo que dice la ley Benford es que la frecuencia de la aparición del primer dígito significativo está dada como sigue

$$\mathbf{P}\{D^1(\xi) = d\} = \log_{10} \left(\frac{d+1}{d} \right),$$

con $\xi \in \mathbf{R}^+$ y $d \in \{1, 2, \dots, 9\}$. Con ésto, la ley Benford se traduce en términos de las frecuencias del primer dígito significativo.

A continuación veremos un poco de la historia que ha girado en torno a la ley Benford, su descubrimiento y algunas de las ideas que han surgido através del

tiempo intentando, en su mayoría de manera errónea, explicar el comportamiento de la ley Benford. Encontraremos que la justificación matemática se hizo casi un siglo después de su descubrimiento introduciendo hipótesis como invarianza bajo la escala e invarianza bajo la base.

1.1 El descubrimiento de la ley Benford

En 1881 Simon Newcomb, astrónomo americano, publicó un artículo [24] en el que expresa lo siguiente: “Que los diez dígitos no ocurran con igual frecuencia debe ser evidente para cualquiera que haga uso de las tablas logarítmicas.” Newcomb informó que las primeras páginas de las tablas logarítmicas son más usadas que las últimas, con lo cual estableció: “La ley de probabilidad de la ocurrencia de los números es tal que la mantisa de sus logaritmos es igualmente probable.”

Newcomb concluyó que

$$\mathbf{P}\{D^1(\xi) = d\} = \log_{10} \left(\frac{d+1}{d} \right). \quad (1)$$

Es decir, la probabilidad de que un número real positivo ξ tenga como primer dígito significativo 1 (en base 10) es aproximadamente 0.301, véase la tabla 1.

d	$\mathbf{P}\{D^1(\xi) = d\}$
1	0.301
2	0.1761
3	0.1249
4	0.0969
5	0.0792
6	0.0669
7	0.0580
8	0.0512
9	0.0458

Tabla 1. Frecuencias del primer dígito significativo según la ley Benford.

Recordemos qué cosa es la mantisa del logaritmo de un número real positivo ξ en base 10. Si vemos a ξ como un producto de la forma $\xi = r \times 10^k$, donde $1 \leq r < 10$ y $k = \max\{\kappa \in \mathbf{Z} : 10^\kappa \leq \xi\}$, entonces

$$\log_{10} \xi = \log_{10} r + k.$$

La mantisa del logaritmo en base 10 de ξ es $\log_{10} r$. Esto se puede hacer para cualquier base β , con $\beta \geq 2$ entero. Por ejemplo consideremos los siguientes números:

$$1674, \quad 0.0784, \quad 354.7 \quad y \quad 0.0003547.$$

Tenemos que $1674 = 1.674 \times 10^3$, $0.0784 = 7.84 \times 10^{-2}$, $354.7 = 3.547 \times 10^2$ y $0.0003547 = 3.547 \times 10^{-4}$. Entonces, las respectivas mantisas de los logaritmos son las siguientes: $\log_{10}(1.674) = 0.224$, $\log_{10}(7.84) = 0.894$, $\log_{10}(3.547) = 0.55$ y $\log_{10}(3.547) = 0.55$.

Cabe hacer notar que Newcomb no ofreció prueba alguna de este fenómeno del primer dígito significativo, su descubrimiento fue intuitivo. Más tarde, 57 años después del descubrimiento de Newcomb, la ley fue redescubierta por el físico Frank Benford [5] ofreciendo evidencia empírica basada en frecuencias de los primeros dígitos significativos provenientes de 20 tablas diferentes. En estas tablas incluía áreas de superficies de ríos, calor específico de compuestos químicos, estadísticas de la liga americana de baseball, direcciones postales, etc. Diaconis y Freeman [9] ofrecieron evidencia de que Benford manipuló algunos errores de redondeo para obtener un mejor ajuste a la ley Benford, aunque los datos no manipulados son una buena aproximación. El artículo de Newcomb fue pasado por alto, por lo que el fenómeno del primer dígito significativo fue conocido como la ley Benford.

Newcomb también formuló la ley del segundo dígito significativo. Esta ley afirma que

$$\mathbf{P}\{D^2(\xi) = d\} = \sum_{k=1}^9 \log_{10} \left(\frac{(10 \times k + d) + 1}{10 \times k + d} \right) \quad (2)$$

con $d \in \{0, 1, \dots, 9\}$.

Aparte de la evidencia que ofreció Benford, se ha ofrecido bastante evidencia empírica adicional, como lo son algunas constantes físicas, la vida media de elementos radioactivos, datos contables, etc.

1.2 Algunas explicaciones de la ley Benford

Desde Benford, han surgido numerosas explicaciones acerca de las frecuencias que aparecen en la tabla 1. Tuvo que pasar casi un siglo desde lo que observó Newcomb hasta encontrar alguna explicación matemática satisfactoria, lograda por Hill en los 90's. Aunque con ésto Hill no pudo resolver el problema de entender bien la naturaleza de la ley Benford, en cambio sí pudo encontrar el espacio de probabilidad adecuado para la ley Benford y demostró que la única distribución del dígito significativo es la logarítmica, también conocida como distribución Benford.

Uno de los primeros intentos de explicar la ley Benford fue investigando el conjunto de los números naturales, intentando probar que la ley proviene naturalmente de nuestro sistema numérico. Como un comienzo se intentó demostrar que el conjunto de números naturales con primer dígito uno, $\{1, 10, 11, \dots, 100, 101, \dots\}$, tiene una probabilidad de $\log_{10} 2$ en los naturales. El problema con el que se encontraron es que este conjunto no tiene alguna densidad natural, es decir, el límite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \#\{i \in \mathbf{N} : i \leq n \text{ y } D^1(i) = 1\},$$

no existe, ya que oscila entre $1/9$ y $5/9$, como se puede observar en la figura 1.

Hay varias formas de definir un límite de este tipo, muchas de ellas nos conducen al deseado $\log_{10} 2$, pero el problema sigue siendo el mismo para esas explicaciones: Los métodos de suma e integración para el caso discreto (números naturales) y continuo (números reales positivos) no son únicos y muchos de ellos no nos conducen a la ley Benford.

Aunque probemos que el conjunto de números reales positivos con primer dígito significativo d tiene una densidad $\log_{10}(1+d^{-1})$, no se ha solucionado el problema de

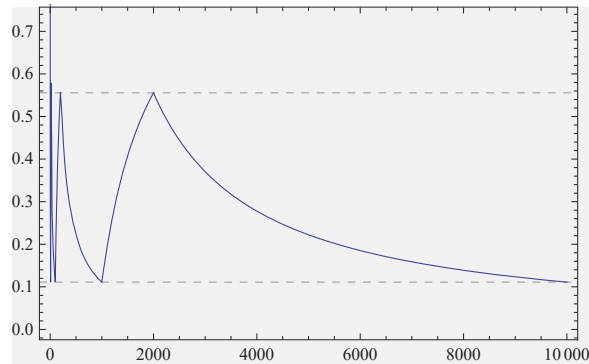


Figura 1.

entender el comportamiento de la ley Benford. No basta con medir esos conjuntos, por lo que siguieron surgiendo hipótesis. Las hipótesis subsecuentes se centraban en la suposición de que la ley Benford es una ley que se cumple en la naturaleza y por lo tanto es universal.

El deseo por entender el fenómeno del primer dígito significativo creció exponencialmente en los 70's debido, en su mayoría, al auge de las computadoras, ya que bajo la suposición de que la ley Benford es una ley universal, se podían obtener aplicaciones como el optimizar errores de redondeo al hacer operaciones, encontrar bases óptimas para el almacenaje de datos, etc.

Una de las aplicaciones que se le ha visto gran futuro es la detección de fraude, como lo mostró Nigrini [25] en su tesis doctoral. Actualmente se siguen proponiendo aplicaciones basadas en la ley Benford para la detección de fraude. Por eso es necesario entender la naturaleza de la ley Benford, ya que con su ayuda se pueden desarrollar aplicaciones en diversas áreas.

1.3 Hipótesis clave

Hubo un momento en que la ley Benford estuvo en un entorno misterioso, en el sentido de que no había algún argumento puramente matemático que pudiese explicar este fenómeno. Entonces ¿Qué es la ley Benford? y ¿Cuándo un conjunto de datos satisface la ley Benford? El día de hoy podemos afirmar que no hay misterio alguno y podemos atrevernos a decir que entendemos la ley Benford, pero para ello veamos las hipótesis que fueron la clave para abordar este fenómeno.

Quien dio la pauta para abordar la ley Benford fue Pinkham [26], con la siguiente hipótesis: “Si hay alguna ley universal en la naturaleza, entonces debería aparecer sin importar las unidades que se utilizan para contar.” Por ejemplo, pensemos en un conjunto de datos que satisface la ley Benford y supongamos que los datos son expresados con el sistema Inglés. Si los datos son expresados con el sistema MKS, el conjunto deberá satisfacer la ley Benford. El convertir unidades a otras se logra por medio de la multiplicación de un escalar. Con esto diremos que la ley Benford es invariante bajo la escala.

El principal problema de esta hipótesis, como lo mostro Knuth [20], es que no hay una medida de probabilidad que sea invariante bajo la escala en los subconjuntos Borel de \mathbf{R}^+ , ya que la medida del conjunto $(0,1)$ deberá de ser la misma que la medida de todo intervalo $(0,t)$, con $t \in \mathbf{R}^+$, lo cual contradice la aditividad contable en probabilidad, por lo que la medida debe ser 0. El asumir esta hipótesis como verdadera implicará invarianza bajo la base, es decir, si un conjunto de datos satisface la ley Benford no importa la base que utilizemos para escribir los datos, de igual manera satisfacerán la ley Benford. Invarianza bajo la base al igual que

invarianza bajo la escala son propiedades que caracterizan a la ley Benford.

Con esto en mente, el problema se reduce a encontrar el dominio apropiado para la ley del dígito significativo, lo cuál hizo Hill [17]. Pensemos en el conjunto de todos los números reales positivos con primer dígito significativo k en base 10, $\{D^1(\xi) = k\}$, entonces

$$\{D^1(\xi) = k\} = \bigcup_{n \in \mathbf{Z}} [k, k+1) \times 10^n,$$

con $k \in \{1, 2, \dots, 9\}$. Análogamente, consideremos los conjuntos con m -ésimo dígito significativo j , $\{D^m(\xi) = j\}$, con $m \in \mathbf{N} \setminus \{1\}$ y $j \in \{0, 1, 2, \dots, 9\}$. Consideremos como espacio muestral a \mathbf{R}^+ y sea la σ -álgebra de los eventos el σ -campo generado por $\{D^1, D^2, \dots\}$, la cual denotaremos por \mathcal{M} . Esta σ -álgebra es conocida como la σ -álgebra mantisa. Note que esta σ -álgebra es un sub- σ -campo de los conjuntos Borel, y que en efecto

$$S \in \mathcal{M} \quad \Longleftrightarrow \quad S = \bigcup_{n \in \mathbf{Z}} B \times 10^n,$$

para algún conjunto Borel $B \subseteq [1, 10)$.

La σ -álgebra \mathcal{M} tiene las siguientes propiedades:

- i. Todo conjunto no vacío en \mathcal{M} es infinito con puntos de acumulación en 0 y en ∞ .
- ii. \mathcal{M} es cerrada bajo multiplicación escalar, es decir,

$$s > 0 \text{ y } S \in \mathcal{M} \Rightarrow sS \in \mathcal{M}.$$

- iii. \mathcal{M} es cerrada bajo raíces naturales, pero no bajo exponentes naturales

$$m \in \mathbf{N} \text{ y } S \in \mathcal{M} \Rightarrow S^{1/m} \in \mathcal{M}.$$

iv. \mathcal{M} es auto-similar en el sentido de que si

$$S \in \mathcal{M} \quad \text{entonces} \quad 10^m S = S, \quad \text{para todo} \quad m \in \mathbf{Z}.$$

Las contradicciones antes mencionadas respecto a invarianza bajo la escala se han eliminado gracias a la propiedad (ii). La propiedad (iii) nos proporcionará la invarianza bajo la base, como lo veremos en la siguiente sección.

También podemos formular la ley general del dígito significativo (en base 10) de la siguiente forma

$$\mathbf{P}\{D^1(\xi) = d_1, \dots, D^j(\xi) = d_j\} = \log_{10} \left(1 + \left(\sum_{i=1}^j d_i \times 10^{j-i} \right)^{-1} \right), \quad (3)$$

con $d_1 \in \{1, 2, \dots, 9\}$ y $d_j \in \{0, 1, 2, \dots, 9\}$ para $j \neq 1$. Un corolario interesante de la ley general del dígito significativo como lo mostró Hill [16], es el siguiente:

Corolario 1. Los dígitos significativos son dependientes.

Note que de (2), la probabilidad incondicional de tener a 3 como segundo dígito significativo es aproximadamente 0.1043, mientras que por (3) la probabilidad condicional de tener a 3 como segundo dígito significativo, dado que el primer dígito es 1, es aproximadamente 0.1069.

En las figuras 2 a 5 se muestran las frecuencias para el primer, segundo, tercer y cuarto dígito según la ley Benford. Note que cuando n crece la diferencia entre las frecuencias tiende a ser menor, es decir, cuando $n \rightarrow \infty$ la distribución del n -ésimo dígito significativo tiende a ser la uniforme.

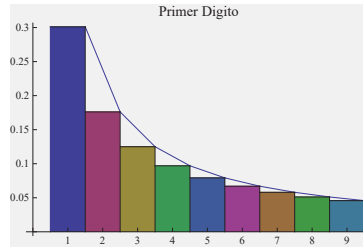


Figura 2. Frecuencias del primer dígito según Benford.

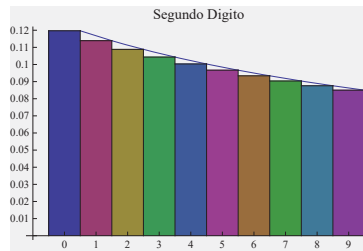


Figura 3. Frecuencias del segundo dígito según Benford.

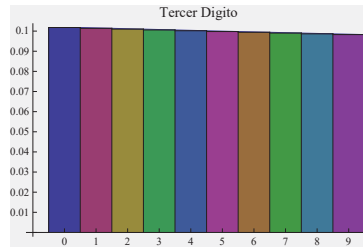


Figura 4. Frecuencias del tercer dígito según Benford.

1.4 Invarianza bajo la escala y bajo la base

Con el espacio de probabilidad apropiado, podemos ahora introducir la noción de invarianza bajo la escala.

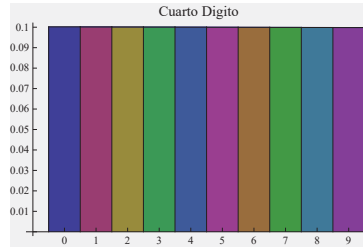


Figura 5. Frecuencias del cuarto dígito según Benford.

Definición 1. Una medida de probabilidad \mathbf{P} en $(\mathbf{R}^+, \mathcal{M})$ es invariante bajo la escala sí $\mathbf{P}(S) = \mathbf{P}(sS)$, para toda $s > 0$ y $S \in \mathcal{M}$.

Como veremos en el siguiente teorema, la invarianza bajo la escala es una caracterización de la ley Benford.

Teorema 1. Una medida de probabilidad \mathbf{P} en $(\mathbf{R}^+, \mathcal{M})$ es invariante bajo la escala sí y sólo sí satisface la ley Benford, es decir,

$$\mathbf{P} \left(\bigcup_{n \in \mathbf{Z}} [1, t) \times 10^n \right) = \log_{10} t, \quad \text{para toda } t \in [1, 10). \quad (4)$$

Para la demostración véase Hill [17].

Invarianza bajo la base es otra hipótesis que caracteriza a la ley Benford. Para dejar más clara esta idea pensemos en $\{D^1(\xi) = 1\}$, que como vimos anteriormente es el conjunto de números positivos con primer dígito significativo 1 (en base 10). Este conjunto puede ser reescrito de la siguiente forma

$$\begin{aligned} \{D_1 = 1\} &= \bigcup_{n \in \mathbf{Z}} [1, 2) \times 10^n \\ &= \bigcup_{n \in \mathbf{Z}} [1, 2) \times 100^n \cup \bigcup_{n \in \mathbf{Z}} [10, 20) \times 100^n, \end{aligned}$$

que resulta ser el conjunto de números positivos cuyo primer dígito significativo en base 10 esta en $\{1, 10, 11, \dots, 19\}$. Recordemos que el espacio de probabilidad apropiado \mathcal{M} lo encontramos en términos de la base 10. Ahora para generalizar esta idea a cualquier base $\beta \geq 2$, reemplazemos 10 por β y \log_{10} por \log_{β} . A este espacio lo representaremos como \mathcal{M}_{β} . Todas las propiedades, definiciones y teoremas son análogos al caso $\beta = 10$. Con ésto en mente podemos introducir la definición de invarianza bajo la base.

Definición 2. Una medida de probabilidad \mathbf{P} en $(\mathbf{R}^+, \mathcal{M}_{\beta})$ es invariante bajo la base sí

$$\mathbf{P}(S) = \mathbf{P}(S^{1/m}) \quad \text{para toda } m \in \mathbf{N} \quad \text{y para todo } S \in \mathcal{M}_{\beta}.$$

Esta definición no es lo suficientemente fuerte, ya que en un principio sólo se es invariante bajo potencias de la base inicial. Con el siguiente teorema veremos que ésto es suficiente para cualquier base. Sea δ_1 la delta de Dirac del conjunto $S_1 = \bigcup_{n \in \mathbf{Z}} \{1\} \times \beta^n$. Note que δ_1 está definida para todo $S \in \mathcal{M}_{\beta}$ por $\delta_1(S) = 1$, si $S_1 \subseteq S$, y 0 en otro caso. Sea \mathbf{P}_L la medida de probabilidad en (4), es decir,

$$\mathbf{P}_L \left(\bigcup_{n \in \mathbf{Z}} [1, t) \times \beta^n \right) = \log_{\beta}(t), \quad \text{para toda } t \in [1, \beta).$$

Ahora podemos formular el teorema 2, que al igual que el teorema 1 caracteriza a la ley Benford.

Teorema 2. Una medida de probabilidad \mathbf{P} en $(\mathbf{R}^+, \mathcal{M}_{\beta})$ es invariante bajo la base sí y sólo sí

$$\mathbf{P} = q\mathbf{P}_L + (1 - q)\delta_1,$$

para algún $q \in [0, 1]$.

Para la demostración véase Hill [17]. De los teoremas 1 y 2 se siguen los siguientes corolarios.

Corolario 2. La distribución logarítmica es la única distribución invariante bajo la base.

Corolario 3. Invarianza bajo la escala implica invarianza bajo la base.

Con este análisis podemos generalizar la ley Benford (1) para cualquier base $\beta \in \mathbf{Z}$, con $\beta \geq 2$, quedando de la siguiente forma

$$\mathbf{P}\{D_\beta^1(\xi) = d\} = \log_\beta \left(\frac{d+1}{d} \right), \quad (5)$$

con $d \in \{1, 2, \dots, \beta - 1\}$.

Análogamente, de (3) podemos formular la ley general del dígito significativo para cualquier base $\beta \geq 2$ entero, quedando de la siguiente forma

$$\mathbf{P}\{D_\beta^1(\xi) = d_1, D_\beta^2(\xi) = d_2, \dots, D_\beta^j(\xi) = d_j\} = \log_\beta \left(1 + \left(\sum_{i=1}^j d_i \times \beta^{j-i} \right)^{-1} \right), \quad (6)$$

con $d_1 \in \{1, 2, \dots, \beta - 1\}$ y $d_j \in \{0, 1, \dots, \beta - 1\}$, para $j \neq 1$.

Con ésto no hemos resuelto del todo el problema del fenómeno del dígito significativo, aún sigue el problema de entender la naturaleza de la ley Benford. En el capítulo 3 encontraremos condiciones suficientes para que una variable aleatoria satisfaga la ley Benford. Estas condiciones suficientes podrían hacernos dudar de la universalidad de la ley. En el capítulo 4 nos reconciliaremos con la idea de la universalidad de la ley Benford con la ayuda de un modelo bastante sencillo. Finalmente generalizaremos la idea del modelo simple a \mathbf{R}^+ , abarcando el caso general.

1.4 Aplicaciones

Desde que Newcomb descubrió este fenómeno del dígito significativo, la ley Benford parecía ser sólo un fenómeno matemático curioso, sin ninguna aplicación aparente. En los años 70's, con el auge de las computadoras empezaron a surgir aplicaciones muy interesantes. Cabe aclarar que son diversas las ramas en las que se puede aplicar la ley Benford. Por ejemplo en computación, teoría de números, ecuaciones diferenciales, estadística, termodinámica, etc. A continuación mencionaremos algunas de estas aplicaciones.

Hamming [14] ha propuesto aplicaciones de la ley Benford al problema de situar el punto decimal (binario) en el sistema numérico de una computadora, con el fin de minimizar el número de cambios de normalización después de calcular un producto, al problema de la estimación de la representación de error de números en base 2 y base 16, y al problema de la propagación de errores por redondeo. Schatte [29] expone que al escoger como base $b = 2^3$, se optimiza con respecto al uso de almacenamiento. Barlow y Bareiss [4] concluyeron que la computadora logarítmica tiene errores más pequeños en los intervalos de confianza por errores de redondeo que la computadora con punto flotante.

Varian [31] propuso usar a la ley Benford como una prueba de sensatez para datos, inspeccionando pronósticos de modelos matemáticos como prueba de bondad de ajuste a la ley Benford. El utilizó esta idea para inspeccionar modelos para la producción económica y pronósticar áreas de terrenos, entre otros usos. La idea fundamental en estas aplicaciones es: si datos de la vida real satisfacen la ley Benford, entonces buenos modelos matemáticos deberán satisfacerla.

Hamadeh [13] en su tesis de maestría propone utilizar a la ley Benford para su uso en la seguridad inalámbrica, para evitar la invasión de la red por intrusos y también para su utilización en la modelación de tráfico en la red.

Nigrini [25] en su tesis doctoral propone el uso de la ley Benford para detectar fraudes en datos contables. La hipótesis en la que se basa Nigrini es que datos financieros verdaderos (sin alterar) deberían aproximarse a la distribución Benford, mientras que datos alterados deberían desviarse considerablemente de la distribución Benford. Crowder [8] afirma que estafadores con frecuencia carecen de creatividad y tienden a inventar números que sean múltiplos de 5, 10 y 25. Se han desarrollado bastantes pruebas utilizando la ley Benford para detectar anomalías en datos contables. Algunas de ellas incluyen la prueba del primer dígito significativo, el segundo dígito significativo, los primeros dos dígitos significativos, los primeros tres dígitos significativos, duplicidad de números, redondeo y los últimos dos dígitos. Cada una de estas pruebas tienen ciertos usos. Por ejemplo, la prueba de los tres primeros dígitos significativos se utiliza para checar duplicidad de datos y para seleccionar muestras para hacer auditoría, mientras que la prueba de los últimos dos dígitos es utilizado para identificar datos inventados o redondeados. El problema de las pruebas del primer y segundo dígito significativo es que se necesitan muestras grandes para poder arrojar conclusiones confiables y sólo identificarán anomalías obvias. Varian [31] propone usar la ley Benford como una prueba de la veracidad de los supuestos datos científicos aleatorios. La ley Benford es sólo un indicador para detectar fraude y, aunque tiene sus limitaciones, utilizarla conjuntamente con otras herramientas para la detección de fraude puede resultar en una aplicación bastante útil y poderosa.

Capítulo 2

Conceptos Básicos

En este capítulo estudiaremos algunos conceptos básicos de probabilidad, lemas y teoremas que nos serán de gran utilidad para abordar el problema de entender la ley Benford.

Recordemos que una variable aleatoria X es una función $X : \Omega \rightarrow \mathbf{R}$, donde Ω es el espacio muestral y \mathbf{R} los números reales. Una variable aleatoria X es discreta si su rango es contable y si su función densidad $p : \mathbf{R} \rightarrow [0, 1]$, es tal que $p(k) = \mathbf{P}\{X = k\}$. La función distribución G y la función densidad p de X están relacionadas de la manera siguiente

$$G(x) = \mathbf{P}\{X \leq x\} = \sum_{i: x_i \leq x} p(x_i) \quad \text{y} \quad p(x) = G(x) - \lim_{y \rightarrow x^-} G(y).$$

Una variable aleatoria Y se dice que es continua si su función distribución F , donde $F(y) = \mathbf{P}\{Y \leq y\}$, puede escribirse como

$$F(y) = \int_{-\infty}^y f(t) dt,$$

para alguna función f integrable, $f : \mathbf{R} \rightarrow [0, \infty)$. La función f es llamada función densidad de la variable aleatoria continua Y . Las propiedades de la función distribución F son las siguientes:

- i. $F(-\infty) = \lim_{y \rightarrow -\infty} F(y) = 0$ y $F(\infty) = \lim_{y \rightarrow \infty} F(y) = 1$.
- ii. F es monótona creciente, es decir, $F(a) \leq F(b)$ para $a < b$.

iii. F es continua por la derecha, esto es,

$$\lim_{y \rightarrow a^+} F(y) = F(a).$$

Con esto en mente veamos las siguientes definiciones.

2.1 Conceptos

Definición 1. Sea X una variable aleatoria discreta con función de probabilidad p . La esperanza de X denotada como $E(X)$ (o media μ) está dada por

$$\mathbf{E}(X) = \mu = \sum_k k \cdot p(k).$$

De manera similar, si Y es una variable aleatoria continua con función densidad f , entonces

$$\mathbf{E}(Y) = \mu = \int_{-\infty}^{\infty} y \cdot f(y) dy.$$

Aquí hemos supuesto que la suma y la integral existen.

Definición 2. La varianza de una variable aleatoria es la esperanza del cuadrado de la desviación de μ . Si X es discreta con función de probabilidad p ,

$$\mathbf{V}(X) = \sigma^2 = \mathbf{E}[(X - \mu)^2] = \sum_k (k - \mu)^2 \cdot p(k).$$

Si Y es continua con función densidad f ,

$$\mathbf{V}(Y) = \sigma^2 = \mathbf{E}[(Y - \mu)^2] = \int_{-\infty}^{\infty} (y - \mu)^2 \cdot f(y) dy.$$

Si $\mathbf{E}(X^2)$ o $\mathbf{E}(Y^2)$ no son finitas, la varianza no está definida.

Definición 3. La función de distribución conjunta de X y Y es la función F , con $F : \mathbf{R}^2 \rightarrow [0, 1]$, que está dada por

$$F(x, y) = \mathbf{P}\{X \leq x, Y \leq y\}.$$

Definición 4. Las variables aleatorias X y Y son conjuntamente continuas con función densidad conjunta $f : \mathbf{R} \rightarrow [0, \infty)$, si

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv \quad \text{para cada } x, y \in \mathbf{R}.$$

Definición 5. La esperanza condicional de Y , dado que sucedió $X = x$ (para toda x tal que $f_X(x) > 0$), está definida por

$$\mathbf{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy,$$

donde

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy}.$$

Se tienen las siguientes propiedades:

- i. $\mathbf{E}[Y] = \int_{-\infty}^{\infty} \mathbf{E}[Y|X = x] f_X(x) dx$.
- ii. $\mathbf{E}[g(X)Y|X = x] = g(x)\mathbf{E}[Y|X = x]$, donde g es una función de valores reales uno a uno.

Definición 6. La varianza condicional de Y , dada X , con X y Y variables aleatorias conjuntamente continuas, está definida por

$$\mathbf{V}[Y|X = x] = \int_{-\infty}^{\infty} (y - \mathbf{E}[Y|X = x])^2 f_{Y|X}(y|x) dy.$$

suponiendo que $\mathbf{E}[Y^2] < \infty$.

La siguiente proposición nos será muy útil, ya que en el capítulo 4 la utilizaremos para entender el comportamiento de la ley Benford en el grupo $\mathbf{Z}_2 \times \mathbf{Z}_2$.

Proposición 1. Sean X y Y variables aleatorias conjuntamente continuas, entonces la varianza de Y es igual a

$$\mathbf{V}[Y] = \mathbf{E}[\mathbf{V}[Y|X]] + \mathbf{V}[\mathbf{E}[Y|X]].$$

Demostración. Supongamos que $\mathbf{E}[Y] < \infty$. Por la propiedad (i) y (ii) de la esperanza condicional, haciendo $g(x) = 1$ tenemos que

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y|X]].$$

Entonces se sigue que

$$\mathbf{V}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])^2] = \mathbf{E}[\mathbf{E}[(Y - \mathbf{E}[Y])^2|X]].$$

Note que para cualquier variable aleatoria Z tenemos que

$$\mathbf{E}[|Z - a|^2] = \mathbf{E}[|Z - \mathbf{E}[Z]|^2] + (\mathbf{E}[Z] - a)^2.$$

Entonces

$$\mathbf{E}[|Y - \mathbf{E}[Y]|^2|X] = \mathbf{E}[|Y - \mathbf{E}[Y|X]|^2|X] + (\mathbf{E}[Y|X] - \mathbf{E}[Y])^2.$$

Tomando esperanzas por ambos lados obtenemos

$$\begin{aligned} \mathbf{V}[Y] &= \mathbf{E}[\mathbf{E}[|Y - \mathbf{E}[Y]|^2|X]] \\ &= \mathbf{E}[\mathbf{E}[|Y - \mathbf{E}[Y|X]|^2|X]] + \mathbf{E}[(\mathbf{E}[Y|X] - \mathbf{E}[Y])^2] \\ &= \mathbf{E}[\mathbf{V}[Y|X]] + \mathbf{V}[\mathbf{E}[Y|X]]. \blacksquare \end{aligned}$$

Teorema 1. (Desigualdad de Chebyshev). Sea X una variable aleatoria con media μ y varianza σ^2 . Para $\epsilon > 0$,

$$\mathbf{P}\{|X - \mu| < \epsilon\} \geq 1 - \frac{\sigma^2}{\epsilon^2},$$

o de manera equivalente

$$\mathbf{P}\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}.$$

Demostración. Para el caso continuo, si $f(x)$ es la función densidad de X ,

$$\begin{aligned} \mathbf{V}[X] &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\ &= \int_{-\infty}^{\mu - \epsilon} (x - \mu)^2 f(x) dx + \int_{\mu - \epsilon}^{\mu + \epsilon} (x - \mu)^2 f(x) dx + \int_{\mu + \epsilon}^{\infty} (x - \mu)^2 f(x) dx. \end{aligned}$$

Omitiendo la integral no negativa de en medio, ya que $f(x) \geq 0$ para toda x , obtenemos la siguiente desigualdad,

$$\begin{aligned} \mathbf{V}[X] &\geq \int_{-\infty}^{\mu - \epsilon} (x - \mu)^2 f(x) dx + \int_{\mu + \epsilon}^{\infty} (x - \mu)^2 f(x) dx \\ &\geq \int_{|x - \mu| \geq \epsilon} (x - \mu)^2 f(x) dx \\ &\geq \int_{|x - \mu| \geq \epsilon} \epsilon^2 f(x) dx \\ &= \epsilon^2 \mathbf{P}\{|X - \mu| \geq \epsilon\}. \end{aligned}$$

Entonces

$$\mathbf{P}\{|X - \mu| \geq \epsilon\} \leq \frac{\mathbf{V}[X]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}.$$

Si la variable es discreta, reemplazamos las integrales por sumas. ■

Proposición 2. Si X y Y son variables aleatorias independientes que tienen como función densidad a $f_X(x)$ y $f_Y(x)$ respectivamente, entonces la variable aleatoria $Z = X \cdot Y$ tiene como función densidad a

$$f_Z(x) = \int_0^{\infty} f_X\left(\frac{x}{y}\right) f_Y(y) \frac{dy}{y}.$$

Demostración. Sea F la función distribución de Z , entonces

$$F(z) = \mathbf{P}\{Z \leq z\} = \mathbf{P}\{(X, Y) \in A_z\} = \int \int_{A_z} f(x, y) dx dy,$$

donde $A_z = \{(x, y) | xy \leq z\} = \{x < 0 | y \geq \frac{z}{x}\} \cup \{x > 0 | y \leq \frac{z}{x}\}$ y $f(x, y)$ es la densidad conjunta de X y Y . Supongamos que la integral existe, entonces por el teorema de Fubini podemos cambiar el orden de integración, así

$$F(z) = \int_{-\infty}^0 \left(\int_{z/x}^{\infty} f(x, y) dy \right) dx + \int_0^{\infty} \left(\int_{-\infty}^{z/x} f(x, y) dy \right) dx.$$

Haciendo $y = \frac{v}{x}$ ($dy = \frac{dv}{x}$), obtenemos

$$\begin{aligned} F(z) &= \int_{-\infty}^0 \left(\int_z^{-\infty} f(x, \frac{v}{x}) \frac{dv}{x} \right) dx + \int_0^{\infty} \left(\int_{-\infty}^z f(x, \frac{v}{x}) \frac{dv}{x} \right) dx \\ &= \int_{-\infty}^0 \left(\int_{-\infty}^z f(x, \frac{v}{x}) \frac{dv}{-x} \right) dx + \int_0^{\infty} \left(\int_{-\infty}^z f(x, \frac{v}{x}) \frac{dv}{x} \right) dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^z f(x, \frac{v}{x}) \frac{dv}{|x|} \right) dx. \end{aligned}$$

Si encontramos una función $g(x)$ tal que $F(z) = \int_{-\infty}^z g(x) dx$, diremos que $g(x)$ es la función densidad de Z . Intercambiando el orden de integración obtenemos

$$F(z) = \int_{-\infty}^z \left(\int_{-\infty}^{\infty} f(x, \frac{v}{x}) \frac{dx}{|x|} \right) dv.$$

Así la función densidad de Z es

$$f_Z(v) = \int_{-\infty}^{\infty} f(x, \frac{v}{x}) \frac{dx}{|x|}.$$

Como X y Y son variables aleatorias independientes, por definición tenemos

$$f(x, y) = f_X(x) f_Y(y).$$

Entonces la función densidad de Z la reescribiremos como

$$f_Z(v) = \int_{-\infty}^{\infty} f_X(x) f_Y\left(\frac{v}{x}\right) \frac{dx}{|x|}.$$

En caso de que X y Y sean variables aleatorias independientes positivas, tenemos $f_Z(z) = 0$ para $z \leq 0$ y

$$\begin{aligned} f_Z(v) &= \int_0^{\infty} f_X(x) f_Y\left(\frac{v}{x}\right) \frac{dx}{x} \\ &= \int_0^{\infty} f_X\left(\frac{v}{w}\right) f_Y(w) \frac{dw}{w}. \blacksquare \end{aligned}$$

La siguiente definición nos será de gran utilidad en el capítulo 3, ya que nos ayudará a encontrar funciones de densidad.

Definición 7. La transformada de Mellin de la función $g(x)$ en $s + 1$ denotada como $M_g(s)$, está definida por

$$M_g(s) = \int_0^{\infty} g(x) x^s dx.$$

El siguiente lema es muy interesante ya que con él podremos construir productos de variables aleatorias que satisfacen la ley Benford, dado que alguna de ellas la cumple.

Lema 1. Sean X y Y variables aleatorias independientes con función densidad g_1 y g_2 respectivamente, entonces la transformada de Mellin de g , donde g es la función densidad de la variable aleatoria $Z = X \cdot Y$, está dada por

$$M_g(s) = M_{g_1}(s) M_{g_2}(s).$$

Demostración. Sabemos por la proposición 2 que $g(x) = \int_0^\infty g_1(x/y)g_2(y)\frac{dy}{y}$.

Entonces

$$M_g(s) = \int_0^\infty x^s g(x) dx = \int_0^\infty x^s \left(\int_0^\infty g_1\left(\frac{x}{y}\right) g_2(y) \frac{dy}{y} \right) dx.$$

Supongamos que la transformada de Mellin existe. Por el teorema de Fubini podemos cambiar el orden de integración, entonces

$$M_g(s) = \int_0^\infty \left(\int_0^\infty x^s g_1\left(\frac{x}{y}\right) g_2(y) \frac{dx}{y} \right) dy.$$

Haciendo el cambio de variable $x = yz$ ($dx = ydz$), tenemos

$$\begin{aligned} M_g(s) &= \int_0^\infty \left(\int_0^\infty (yz)^s g_1(z) g_2(y) dz \right) dy \\ &= \int_0^\infty y^s g_2(y) \left(\int_0^\infty z^s g_1(z) dz \right) dy \\ &= \left(\int_0^\infty y^s g_2(y) dy \right) \left(\int_0^\infty z^s g_1(z) dz \right) \\ &= M_{g_1}(s) M_{g_2}(s). \blacksquare \end{aligned}$$

Como vimos en el capítulo 1, se ha utilizado la teoría de Lebesgue para abordar a la ley Benford. En nuestro caso utilizaremos el análisis de Fourier para entender esta ley. En el capítulo 3, usaremos el siguiente teorema del análisis de Fourier para encontrar condiciones suficientes para que una variable aleatoria satisfaga Benford.

Teorema 2. (Fórmula de la suma de Poisson). Si $f \in L^1(\mathbf{R})$, entonces

$$\sum_{n=-\infty}^{\infty} f(x+n) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n x},$$

donde $\hat{f}(n)$ es el n -ésimo coeficiente de Fourier, $\hat{f}(n) = \int_{-\infty}^{\infty} f(x) e^{2\pi i x n} dx$. En otras palabras, los coeficientes de Fourier de la periodización de f están dados precisamente por los valores de la transformada de Fourier de f en los enteros.

Demostración. Para ver ésto, es suficiente mostrar que ambos lados (las cuales son continuas) tienen los mismos coeficientes de Fourier (vistas como funciones en el círculo). Es claro que el m -ésimo coeficiente de Fourier de la parte derecha es $\hat{f}(m)$. Para la parte izquierda, haciendo $F(x) = \sum_{j=-\infty}^{\infty} f(x+j)$, es fácil notar que $F(x)$ es una función periódica de período 1. Entonces el m -ésimo coeficiente Fourier de $F(x)$ es

$$\begin{aligned}\hat{F}(m) &= \int_0^1 F(x)e^{2\pi i x m} dx \\ &= \int_0^1 \left[\sum_{j=-\infty}^{\infty} f(x+j) \right] e^{2\pi i x m} dx \\ &= \sum_{j=-\infty}^{\infty} \int_0^1 f(x+j)e^{2\pi i x m} dx.\end{aligned}$$

Haciendo $y = x + j$ obtenemos que

$$\begin{aligned}\hat{F}(m) &= \sum_{j=-\infty}^{\infty} \int_j^{j+1} f(y)e^{2\pi i y m + 2\pi i j m} dy \\ &= \int_{-\infty}^{\infty} f(y)e^{2\pi i y m} dy \\ &= \hat{f}(m). \blacksquare\end{aligned}$$

Teorema 3. (Lema de Riemann-Lebesgue). Sea φ una función continua y periódica de período 1, $\varphi : \mathbf{R} \rightarrow \mathbf{R}$. Suponga que $\int_0^1 \varphi(x) dx = 0$. Entonces para toda función integrable $f : [0, 1] \rightarrow \mathbf{R}$, se cumple que

$$\lim_{n \rightarrow \infty} \int_0^1 f(x)\varphi(nx) dx = 0.$$

Demostración. Supongamos que f es continua. Para toda $\epsilon > 0$ existe $n \in \mathbf{N}$ tal que para todo par $x, y \in [0, 1]$ tal que $|x - y| \leq 1/n$, se cumple que

$$\left| f\left(\frac{t}{n}\right) - f\left(\frac{j}{n}\right) \right| < \epsilon.$$

Sea $K = \int_0^1 |\varphi(x)| dx$. Para cada $1 \leq j \leq n$ tenemos $\int_{j-1}^j f(\frac{j}{n})\varphi(x) dx = 0$, y por lo tanto uniformemente en j ,

$$\left| \int_{j-1}^j f\left(\frac{x}{n}\right) \varphi(x) dx \right| \leq \int_{j-1}^j \left| f\left(\frac{x}{n}\right) - f\left(\frac{j}{n}\right) \right| |\varphi(x)| dx < K\epsilon.$$

Entonces

$$\left| \int_0^1 f(x)\varphi(nx) dx \right| \leq \frac{1}{n} \sum_{j=1}^n \left| \int_{j-1}^j f\left(\frac{x}{n}\right) \varphi(x) dx \right| < K\epsilon.$$

En el caso en el que f es solamente integrable, pero no necesariamente continua, sea h una función continua tal que $\int |f - h| < \epsilon$. Entonces

$$\left| \int_0^1 f(x)\varphi(nx) dx \right| \leq \max_{0 \leq x \leq 1} |\varphi(x)| \int_0^1 |f(x) - h(x)| dx + \left| \int_0^1 h(x)\varphi(nx) dt \right|$$

y la última integral tiende a cero cuando $n \rightarrow \infty$. ■

En el capítulo 3 aplicaremos el siguiente lema para producir variables aleatorias que satisfacen la ley Benford de manera aproximada.

Lema 2. Sean ξ una variable aleatoria con valores en los reales positivos y $f(x)$ su función densidad. Entonces

$$\left| \int_{-\infty}^{\infty} f(x)e^{-2\pi i j x} dx \right| < 1$$

para toda $j \in \mathbf{Z} \setminus \{0\}$.

Demostración. Por ser f función de densidad de ξ tenemos que

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad \text{y} \quad f(x) \geq 0.$$

Entonces

$$\begin{aligned} \left| \int_{-\infty}^{\infty} f(x) e^{-2\pi i j x} dx \right| &\leq \int_{-\infty}^{\infty} |f(x) e^{-2\pi i j x}| dx \\ &= \int_{-\infty}^{\infty} f(x) |e^{-2\pi i j x}| dx \leq 1. \end{aligned}$$

Supongamos que alguna de las igualdades se cumple para algún $j \in \mathbf{Z} \setminus \{0\}$, entonces

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x) e^{-2\pi i j x} dx \\ &= \int_{-\infty}^{\infty} f(x) (\cos(2\pi j x) - i \operatorname{sen}(2\pi j x)) dx \\ &= \int_{-\infty}^{\infty} f(x) \cos(2\pi j x) dx. \end{aligned}$$

Así

$$0 = \int_{-\infty}^{\infty} f(x) (1 - \cos(2\pi j x)) dx,$$

pero esto implica que $f(x) = 0$, ya que $1 - \cos(2\pi j x) \geq 0$, lo que es una contradicción dado que f es una función de densidad. Para $\int_{-\infty}^{\infty} f(x) e^{-2\pi i j x} dx = -1$ el razonamiento es análogo, por lo tanto la desigualdad es estricta. ■

Capítulo 3

Condiciones suficientes para satisfacer Benford

En este capítulo presentaremos dos teoremas que proporcionan condiciones suficientes para que una variable aleatoria absolutamente continua obedezca la ley Benford. Nuestra herramienta de trabajo es la fórmula de la suma de Poisson del análisis de Fourier. Con estas dos condiciones suficientes notaremos que la ley Benford no se observará siempre en conjuntos de datos numéricos. Esto parece contradecir la universalidad que conocemos acerca de esta ley que se observa en la naturaleza. Más adelante, en el capítulo 4, nos reconciliaremos con la idea de la universalidad de la ley Benford.

3.1 Condiciones para satisfacer la ley Benford

Sean $\beta \geq 2$ un entero y ξ una variable aleatoria con valores en \mathbf{R}^+ , con $g(x)$ como su función densidad. Sean $r = r_\xi \in \mathbf{R}$ y $k = k_\xi \in \mathbf{Z}$ tales que

$$\xi = r \times \beta^k,$$

con $k = \max\{\kappa : \beta^\kappa \leq \xi\}$ y $1 \leq r < \beta$. Entonces $\log_\beta \xi = \log_\beta r + k$, por lo que $\log_\beta \xi = \log_\beta r \pmod{1}$. Note que $D_\beta^1(\xi) = D_\beta^1(r)$, es decir, el primer dígito significativo de ξ es el mismo que el de r . Entonces tenemos que calcular la probabilidad $\mathbf{P}\{D_\beta^1(\xi) = d\}$, con $d \in \{1, 2, \dots, \beta - 1\}$, es equivalente a calcular $\mathbf{P}\{D_\beta^1(r) = d\}$, donde

$$\mathbf{P}\{D_\beta^1(r) = d\} = \mathbf{P}\{d \leq r < d + 1\} = \mathbf{P}\{\log_\beta(d) \leq \log_\beta r < \log_\beta(d + 1)\}.$$

Esto nos recuerda a lo que ya mencionó Newcomb: La ley de probabilidad de la ocurrencia de los números es tal que la mantisa de sus logaritmos es igualmente probable.

Sea $A = [\log_\beta d, \log_\beta(d+1))$, entonces

$$\mathbf{P}\{\log_\beta r \in A\} = \sum_{j \in \mathbf{Z}} \mathbf{P}\{\log_\beta \xi \in j + A\} = \int_A f(x) dx,$$

donde $f(x)$ es la función densidad de $\log_\beta r$. Como ξ tiene a $g(x)$ como función densidad entonces $\log_\beta \xi$ tiene a $f_0(x) = \log(\beta)\beta^x g(\beta^x)$ como función densidad, ya que

$$\begin{aligned} \mathbf{P}\{\log_\beta \xi \leq x\} &= \mathbf{P}\{\xi \leq \beta^x\} \\ &= \int_{-\infty}^{\beta^x} g(x) dx. \end{aligned} \tag{1}$$

Por el teorema fundamental del cálculo, derivando (1) obtenemos

$$\begin{aligned} \frac{d}{dx} \mathbf{P}\{\xi \leq \beta^x\} &= \frac{d}{dx} \int_{-\infty}^{\beta^x} g(x) dx \\ &= \log(\beta)\beta^x g(\beta^x) \\ &= f_0(x). \end{aligned}$$

Entonces

$$f(x) = \sum_{j \in \mathbf{Z}} f_0(x+j).$$

Como f es una función periódica, de período 1, $f(x+1) = f(x)$ para toda $x \in \mathbf{R}$.

Podemos utilizar la fórmula de la suma de Poisson y escribir

$$f(x) = \sum_{j \in \mathbf{Z}} e^{2\pi i j x} \int_{-\infty}^{\infty} f_0(x) e^{-2\pi i j x} dx. \tag{2}$$

Con este análisis podemos, ahora, enunciar el primer teorema.

Teorema 1. Sean $\eta > 0$ un número real fijo y ξ una variable aleatoria con valores en \mathbf{R}^+ , con $g(x)$ como su función densidad. Sea $f(x)$ la función densidad de la variable aleatoria $\eta \log_\beta \xi \pmod{1}$. Entonces

$$f(x) = \sum_{j \in \mathbf{Z}} e^{2\pi i j x} M_g \left\{ -\frac{2\pi j \eta}{\log \beta} \right\}.$$

Demostración. Por lo que obtuvimos con la fórmula de la suma de Poisson, de (2) sólo basta demostrar que

$$\int_{-\infty}^{\infty} f_0(x) e^{-2\pi i j x} dx = M_g \left\{ -\frac{2\pi j \eta}{\log \beta} \right\},$$

donde $f_0(x)$ es la función de densidad de $\eta \log_\beta \xi$ dada por $f_0(x) = \frac{\log \beta}{\eta} \beta^{\frac{x}{\eta}} g(\beta^{\frac{x}{\eta}})$.

Entonces

$$\int_{-\infty}^{\infty} f_0(x) e^{-2\pi i j x} dx = \int_{-\infty}^{\infty} \frac{\log \beta}{\eta} \beta^{\frac{x}{\eta}} g(\beta^{\frac{x}{\eta}}) e^{-2\pi i j x} dx.$$

Haciendo $y = \beta^{\frac{x}{\eta}}$, tenemos que

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log \beta}{\eta} \beta^{\frac{x}{\eta}} g(\beta^{\frac{x}{\eta}}) e^{-2\pi i j x} dx &= \int_0^{\infty} g(y) y^{-\frac{2\pi i j \eta}{\log \beta}} dy \\ &= M_g \left\{ -\frac{2\pi i j \eta}{\log \beta} \right\}. \blacksquare \end{aligned}$$

Corolario 1. Si $\sigma > 0$ y ξ tiene a $g(x/\sigma)/\sigma$ como su función densidad, entonces $\eta \log_\beta \xi \pmod{1}$ tiene como función densidad a

$$f(x) = \sum_{j \in \mathbf{Z}} \exp \left\{ 2\pi i j \left(x - \eta \frac{\log \sigma}{\log \beta} \right) \right\} M_g \left\{ -\frac{2\pi i j \eta}{\log \beta} \right\}.$$

Si la función densidad $g(x)$ satisface que $M_g\{-2\pi i j / \log \beta\} = 0$ para toda $j \in \mathbf{Z} \setminus \{0\}$, entonces $f(x) = 1$ para toda $x \in [0, 1]$. Por lo tanto, haciendo $A = [\log_\beta d, \log_\beta d + 1)$, se sigue que

$$\mathbf{P}\{d \leq r < d + 1\} = \mathbf{P}\{\log_\beta d \leq \log_\beta r < \log_\beta d + 1\} = \int_A dx = \log_\beta \left(\frac{d+1}{d} \right).$$

Este resultado es lo que predice la ley Benford para el primer dígito significativo de r (que es el mismo que de ξ) igual a d . Así que $M_g\{-2\pi ij/\log\beta\} = 0$, para toda $j \in \mathbf{Z} \setminus \{0\}$, es una condición suficiente para que la variable aleatoria ξ obedezca la ley Benford. Por el Lema 2.1 tenemos que si ξ_1 obedece la ley Benford, entonces la variable aleatoria $\xi = \xi_1 \xi_2$ también la obedecerá, ya que $M_g\{s\} = M_{g_1}\{s\}M_{g_2}\{s\}$.

Ejemplo 1. Sea X una variable aleatoria con función de densidad $f_0(x) = 1$ si $|x| \leq 1/2$ y $f_0(x) = 0$ si $|x| > 1/2$. Entonces

$$M_g \left\{ -\frac{2\pi ij}{\log\beta} \right\} = \int_{-1/2}^{1/2} e^{-2\pi ijx} dx = 0,$$

para toda $j \in \mathbf{Z} \setminus \{0\}$. Por lo tanto β^X satisface la ley Benford. De manera más general, si X esta distribuida uniformemente en un intervalo de la forma $[\nu, \nu + h]$, donde $\nu \in \mathbf{R}$ y $h \in \mathbf{N}$, entonces β^X satisface la ley Benford.

Con el siguiente teorema estableceremos algunas restricciones para que una variable aleatoria absolutamente continua satisfaga la ley Benford.

Teorema 2. Sea X una variable aleatoria absolutamente continua con función de densidad $f^0(x)$, donde f^0 satisface las siguientes condiciones

$$f^0(x) = \begin{cases} f^0(x) = f^0(-x) & \text{para toda } x, \\ f^0(x) = 0 & \text{si } |x| > 1, \\ f^0\left(\frac{1}{2} + x\right) = c - f^0\left(\frac{1}{2} - x\right) & \text{si } 0 \leq x \leq 1/2. \end{cases}$$

Donde $c = 2f^0(1/2)$. Sea $\beta \geq 2$ un entero, entonces $\xi = \beta^X$ satisface la ley Benford, es decir,

$$\mathbf{P}\{D_\beta^1(\xi) = d\} = \log_\beta \left(\frac{d+1}{d} \right).$$

Demostración. Tenemos que demostrar que para toda $j \in \mathbf{Z} \setminus \{0\}$ se tiene que $\hat{f}(j) = 0$. Para ésto note que

$$\begin{aligned}\hat{f}(j) &= \int_{-\infty}^{\infty} f^0(x) e^{-2\pi i j x} dx \\ &= \int_{-1}^1 f^0(x) e^{-2\pi i j x} dx \\ &= \int_{-1}^1 f^0(x) [\cos(-2\pi j x) + i \operatorname{sen}(-2\pi j x)] dx.\end{aligned}$$

Como $\operatorname{sen}(x) = -\operatorname{sen}(-x)$ y $f^0(-x) = f^0(x)$, tenemos que

$$\hat{f}(j) = \int_{-1}^1 f^0(x) \cos(-2\pi j x) dx.$$

Pero $\cos(-x) = \cos(x)$, así

$$\begin{aligned}\hat{f}(j) &= 2 \int_0^1 f^0(x) \cos(-2\pi j x) dx \\ &= 2 \left[\int_0^{1/2} f^0(x) \cos(-2\pi j x) dx + \int_{1/2}^1 f^0(x) \cos(-2\pi j x) dx \right].\end{aligned}$$

Para $\int_0^{1/2} f^0(x) \cos(-2\pi j x) dx$ hacemos $x = 1/2 - y$, entonces

$$\begin{aligned}\int_0^{1/2} f^0(x) \cos(-2\pi j x) dx &= \int_{1/2}^0 -f^0(1/2 - y) \cos(2\pi j y - \pi j) dy \\ &= (-1)^j \int_0^{1/2} f^0(1/2 - y) \cos(2\pi j y) dy,\end{aligned}$$

ya que $\cos(2\pi j y - \pi j) = \cos(2\pi j y) \cos(\pi j) + \operatorname{sen}(2\pi j y) \operatorname{sen}(\pi j) = (-1)^j \cos(2\pi j y)$.

Por otra parte para $\int_{1/2}^1 f^0(x) \cos(-2\pi j x) dx$ hacemos $x = 1/2 + y$, entonces

$$\begin{aligned}\int_{1/2}^1 f^0(x) \cos(-2\pi j x) dx &= \int_0^{1/2} f^0(1/2 + y) \cos(-2\pi j y - \pi j) dy \\ &= (-1)^j \int_0^{1/2} f^0(1/2 + y) \cos(2\pi j y) dy,\end{aligned}$$

ya que $\cos(-2\pi jy - \pi j) = \cos(2\pi jy) \cos(\pi j) - \sin(2\pi jy) \sin(\pi j) = (-1)^j \cos(2\pi jy)$.

Con lo que obtenemos

$$\begin{aligned} \hat{f}(j) &= 2(-1)^j \int_0^{1/2} \left[f^0\left(\frac{1}{2} - y\right) + f^0\left(\frac{1}{2} + y\right) \right] \cos(2\pi jy) dy \\ &= 2(-1)^j c \int_0^{1/2} \cos(2\pi jy) dy \\ &= 2(-1)^j c \left(\frac{\sin(2\pi jy)}{2\pi j} \Big|_0^{\frac{1}{2}} \right) = 0, \end{aligned}$$

ya que por hipótesis $[f^0(\frac{1}{2} - y) + f^0(\frac{1}{2} + y)] = c$. ■

Con el teorema anterior podemos ahora encontrar una infinidad de variables aleatorias que satisfacen la ley Benford, como lo muestra el siguiente ejemplo.

Ejemplo 2. Si $f^0(x)$ es como en el teorema 2, entonces está completamente determinada si conocemos sus valores en el intervalo $[0, 1/2]$. Por ejemplo tomamos

$$f^0(x) = \frac{1025 - 2048x^{12}}{1025 + 2048x}$$

para toda $x \in [0, 1/2]$ y $c = 2f^0(1/2)$. Si X tiene a $f^0(x)$ como su función densidad, entonces β^X satisface la ley Benford.

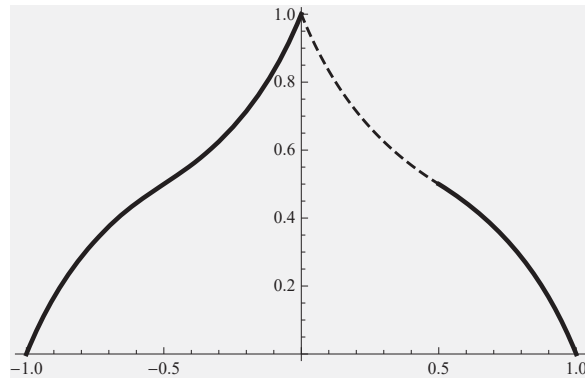


Figura 1.

Teorema 3. Sean $\beta \geq 2$ un entero y Y una variable aleatoria con valores reales, con $f^0(x)$ como su función densidad. Supongamos que β^Y satisface la ley Benford. Sea X una variable aleatoria con función densidad $f_0 = f^0 + f^1$, donde f^1 satisface las siguientes condiciones

$$f^1(x) = \begin{cases} f^1(x) = -f^1(-x) & \text{para toda } x, \\ f^1(x) = 0 & \text{si } |x| > 1, \\ f^1\left(\frac{1}{2} + x\right) = f^1\left(\frac{1}{2} - x\right) & \text{si } 0 \leq x \leq 1/2. \end{cases}$$

Entonces $\xi = \beta^X$ satisface la ley Benford.

Demostración. Calcularemos los coeficientes de Fourier de $f_0 = f^0 + f^1$,

$$\begin{aligned} \hat{f}_0(j) &= \int_{-\infty}^{\infty} f_0(x) e^{-2\pi i j x} dx \\ &= \int_{-\infty}^{\infty} [f^0(x) + f^1(x)] e^{-2\pi i j x} dx \\ &= \int_{-\infty}^{\infty} f^0(x) e^{-2\pi i j x} dx + \int_{-\infty}^{\infty} f^1(x) e^{-2\pi i j x} dx. \end{aligned}$$

Podemos hacer caso omiso de f^0 ya que por hipótesis β^Y satisface la ley Benford. Así, basta con ver que los coeficientes de Fourier de f^1 son iguales a cero. Para esto tenemos que para toda $j \in \mathbf{Z}$

$$\begin{aligned} \hat{f}^1(j) &= \int_{-\infty}^{\infty} f^1(x) e^{-2\pi i j x} dx \\ &= \int_{-1}^1 f^1(x) e^{-2\pi i j x} dx \\ &= \int_{-1}^1 f^1(x) [\cos(-2\pi j x) + i \operatorname{sen}(-2\pi j x)] dx. \end{aligned}$$

Como $\cos(x) = \cos(-x)$ y $f^1(x) = -f^1(-x)$, tenemos

$$\hat{f}^1(j) = \int_{-1}^1 i f^1(x) \operatorname{sen}(-2\pi j x) dx,$$

pero $\text{sen}(x) = -\text{sen}(-x)$, así

$$\begin{aligned}\hat{f}^1(j) &= 2i \int_0^1 f^1(x) \text{sen}(-2\pi jx) dx \\ &= 2i \left[\int_0^{1/2} f^1(x) \text{sen}(-2\pi jx) dx + \int_{1/2}^1 f^1(x) \text{sen}(-2\pi jx) dx \right].\end{aligned}$$

Para $\int_0^{1/2} f^1(x) \text{sen}(-2\pi jx) dx$ hacemos $x = 1/2 - y$, entonces

$$\begin{aligned}\int_0^{1/2} f^1(x) \text{sen}(-2\pi jx) dx &= \int_{1/2}^0 -f^1(1/2 - y) \text{sen}(2\pi jy - \pi j) dy \\ &= (-1)^{j+1} \int_0^{1/2} -f^0(1/2 - y) \text{sen}(2\pi jy) dy,\end{aligned}$$

ya que $\text{sen}(2\pi jy - \pi j) = (-1)^j \text{sen}(2\pi jy)$.

Por otra parte para $\int_{1/2}^1 f^1(x) \text{sen}(-2\pi jx) dx$ hacemos $x = 1/2 + y$, entonces

$$\begin{aligned}\int_{1/2}^1 f^1(x) \text{sen}(-2\pi jx) dx &= \int_0^{1/2} f^1(1/2 + y) \text{sen}(-2\pi jy - \pi j) dy \\ &= (-1)^{j+1} \int_0^{1/2} f^1(1/2 + y) \text{sen}(2\pi jy) dy,\end{aligned}$$

ya que $\text{sen}(-2\pi jy - \pi j) = (-1)^{j+1} \text{sen}(2\pi jy)$. Con lo que obtenemos

$$\begin{aligned}\hat{f}^1(j) &= 2i \int_0^1 f^1(x) \text{sen}(-2\pi jx) dx \\ &= 2i(-1)^{j+1} \int_0^{1/2} \left[f^1\left(\frac{1}{2} + y\right) - f^1\left(\frac{1}{2} - y\right) \right] \text{sen}(2\pi jy) dy \\ &= 0,\end{aligned}$$

ya que por hipótesis $[f^1(\frac{1}{2} + y) - f^1(\frac{1}{2} - y)] = 0$. ■

Los teoremas 2 y 3 nos proporcionan condiciones suficientes para encontrar familias de variables aleatorias que satisfacen la ley Benford.

Ejemplo 3. En la figura 2, podemos observar la densidad $f_0 = f^0 + f^1$, donde $f^0(x) = |x|$ tiene una forma triangular y f^1 está graficada como la línea punteada y está dada por

$$f^1(x) = \frac{y}{(1+y)^{11}},$$

si $y \in [0, 1/2]$. Y si $y \in [-1, 0) \cup (1/2, 1]$, f^1 se comporta como en el teorema 3. Como lo vimos en el ejemplo 1, $f^0(x) = |x|$ es la densidad de una variable aleatoria Y tal que B^Y satisface la ley Benford. La función f^1 satisface las tres condiciones del Teorema 3. Si X tiene a f_0 como su función densidad entonces β^X satisface la ley Benford.

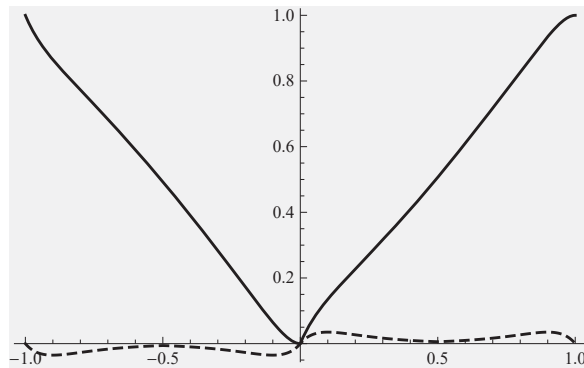


Figura 2.

Podemos utilizar los teoremas 2 y 3 para producir una infinidad de variables aleatorias que satisfacen la ley Benford. Los ejemplos así producidos parecerán artificiales porque los teoremas 2 y 3 implican restricciones de simetría en la función densidad. Aún así, ésto no contradice la naturaleza de la ley Benford, el de ser observada en la naturaleza.

3.2 Benford de una manera aproximada

Sea ξ una variable aleatoria que no obedece la ley Benford. Hay dos procesos por el cuales ξ puede estar cerca de obedecer la ley Benford. El primero de estos procesos es la multiplicación de variables aleatorias. Supongamos que $\xi = \xi_1 \xi_2$ es el producto de dos variables aleatorias. Para $\ell = 1, 2$ sea $g_\ell(x)$ la función densidad de ξ_ℓ . Entonces por el lema 2.2, tenemos

$$\left| M_{g_\ell} \left\{ -\frac{2\pi i j}{\log \beta} \right\} \right| < 1,$$

para toda $j \in \mathbf{Z} \setminus \{0\}$. Por lo tanto si $j \in \mathbf{Z} \setminus \{0\}$, entonces el j -ésimo coeficiente en la serie de Fourier del teorema 1, para la función densidad $f(x)$ de la variable aleatoria $\log_\beta \xi \pmod{1}$, es estrictamente más pequeño que los coeficientes de cualquiera de los dos: $\log_\beta \xi_1 \pmod{1}$ o $\log_\beta \xi_2 \pmod{1}$. La pequeñez de los coeficientes de Fourier de $f(x)$ de la variable aleatoria ξ contribuye a que ξ esté cerca de satisfacer la ley Benford. La ley Benford se satisface cuando el j -ésimo coeficiente de Fourier de $f(x)$ es igual a cero para toda $j \in \mathbf{Z} \setminus \{0\}$.

El otro proceso por el cual una variable aleatoria ξ puede aproximadamente seguir la ley Benford es la exponenciación. De hecho, si X tiene a $f_0(x)$ como su función densidad, entonces $\xi = \beta^X$ tiene a

$$g(y) = \frac{1}{y \log \beta} f_0(\log_\beta y),$$

como su función densidad. Sea $\eta > 0$. El j -ésimo coeficiente de Fourier de la función densidad $f(x)$ de la variable aleatoria $\log_\beta \xi^\eta \pmod{1}$ es

$$M_g \left\{ -\frac{2\pi i j \eta}{\log \beta} \right\} = \int_{-\infty}^{\infty} f_0(x) e^{-2\pi i j \eta x} dx.$$

Por el lema Riemann-Lebesgue (teorema 2.3) sabemos que

$$\lim_{\eta \rightarrow \infty} M_g \left\{ -\frac{2\pi i j \eta}{\log \beta} \right\} = 0,$$

para toda $j \in \mathbf{Z} \setminus \{0\}$. Por lo tanto ξ^η obedecerá aproximadamente la ley Benford cuando η sea suficientemente grande.

El corolario 1 es interesante, pues muestra una propiedad de invarianza respecto al parámetro σ . Esta propiedad de la invarianza es relevante cuando una variable aleatoria ξ no obedece la ley Benford. Sea $f(x)$ la función de densidad del corolario 1. Sea $\delta \in \mathbf{Z}$. Como

$$f(x, \sigma, \beta) = \sum_{j \in \mathbf{Z}} \exp \left\{ 2\pi i j \left(x - \eta \frac{\log \sigma}{\log \beta} \right) \right\} M_g \left\{ -\frac{2\pi i j \eta}{\log \beta} \right\},$$

entonces tenemos $f(x, \sigma, \beta) = f(x, \sigma \beta^\delta, \beta)$ para toda $x \in [0, 1]$. Esta propiedad de invarianza hace ver que no está justificado esperar que una variable aleatoria ξ , con función de densidad $g(x/\sigma)/\sigma$, satisfaga la ley Benford aproximadamente cuando σ se incrementa en magnitud. Esto es lo contrario de lo que uno podría esperar.

Ejemplo 4. Consideremos la variable aleatoria $|\xi|$, donde ξ tiene una distribución normal con media 0 y varianza σ^2 . Por el corolario 1, sabemos que $\log_\beta |\xi| \pmod{1}$ tiene a

$$f(x, \sigma, \beta) = \sum_{j \in \mathbf{Z}} \exp \left\{ 2\pi i j \left(x - \frac{\log \sigma}{\log \beta} - \frac{\log 2}{2 \log \beta} \right) \right\} \Gamma \left(\frac{1}{2} - \frac{\pi i j}{\log \beta} \right)$$

como su función densidad. Sea $\delta \in \mathbf{Z}$. Note que $f(x, \sigma, \beta) = f(x, \sigma \beta^\delta, \beta)$ para toda $x \in [0, 1]$. Si hacemos $\beta = 10$ podemos ver en la tabla 1 qué tan lejos puede estar

$|\xi|$ de satisfacer la ley Benford.

d	Benford	Normal
1	.301	.385
2	.176	.173
3	.124	.085
4	.096	.065
5	.079	.061
6	.066	.059
7	.057	.057
8	.051	.055
9	.045	.053

Tabla 1. Frecuencias esperadas por la ley Benford y las esperadas por $|\xi|$.

En la tercera columna de la tabla 1 se registran los valores de $\mathbf{P}\{d \leq r < d+1\}$ cuando $|\xi| = r \times \beta^k$ y ξ tiene una distribución normal con media 0 y varianza $\sigma_0^2 = (0.1328)^2$. La discrepancia registrada en la tabla 1 entre $\mathbf{P}\{d \leq r < d+1\}$ y Benford se observará cuando ξ tenga varianza σ^2 , donde $\sigma = \sigma_0 10^\delta$ y $\delta \in \mathbf{Z}$.

El corolario 1 también sugiere que la invarianza de $f(x)$ bajo transformaciones en σ ocurren sólo cuando $M_g(-2\pi i j / \log \beta) = 0$, para toda $j \in \mathbf{Z} \setminus \{0\}$. Así la independencia de $f(x)$ del parámetro σ parece implicar que la variable ξ seguirá la ley Benford de manera exacta.

3.3 Contraejemplo de la ley Benford

Utilizaremos un teorema del límite central para construir un contraejemplo de la ley Benford. Con ésto en mente, consideremos el caso de una variable aleatoria

β^X , donde $X = (Y_1 + \cdots + Y_n)/(\sigma\sqrt{n})$ es la suma de n variables aleatorias independientes e idénticamente distribuidas Y_ℓ y $\sigma \in (0, \infty)$. Los coeficientes de Fourier de la función densidad $f(x)$ de la variable aleatoria $X \pmod{1}$ dados por la fórmula de la suma de Poisson, son

$$\left\{ \hat{f}_Y \left(\frac{j}{\sigma\sqrt{n}} \right) \right\}^n \quad \text{donde} \quad \hat{f}_Y(\xi) = \int_{-\infty}^{\infty} f_Y(x) e^{-2\pi i \xi x} dx,$$

y $f_Y(x)$ es la función densidad de Y_ℓ , con $\ell = 1, \dots, 10$. Supongamos que $f_Y(x) \neq 0$ para toda $j \in \mathbf{N}$. Supongamos también que

$$0 = \int_{-\infty}^{\infty} x f_Y(x) dx, \quad 1 = \int_{-\infty}^{\infty} x^2 f_Y(x) dx \quad \text{y} \quad \int_{-\infty}^{\infty} |x|^3 f_Y(x) dx < \infty.$$

Recordemos que para toda $x \in \mathbf{R}$, tenemos $e^{ix} = 1 + ix - x^2/2 + O(|x|^3)$. Entonces

$$\hat{f}_Y(\xi) = \int_{-\infty}^{\infty} f_Y(x) [1 - 2\pi i \xi x - 2(\pi \xi x)^2 + O(|\xi x|^3)] dx = 1 - 2(\pi \xi)^2 + O(|\xi|^3).$$

Si $|x| \leq 1/2$, entonces $\log(1+x) = x + O(x^2)$. De ahí, $\log \hat{f}_Y(\xi) = -2(\pi \xi)^2 + O(|\xi|^3)$.

Así, cuando $\xi \rightarrow 0$

$$\lim_{n \rightarrow \infty} \left\{ \hat{f}_Y \left(\frac{j}{\sigma\sqrt{n}} \right) \right\}^n = \lim_{n \rightarrow \infty} \exp \left\{ -2 \left(\frac{\pi j}{\sigma} \right)^2 + O \left(\frac{|j|^3}{\sqrt{n}} \right) \right\} = \exp \left\{ -2 \left(\frac{\pi j}{\sigma} \right)^2 \right\}.$$

De igual manera si tomamos a $f_0(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$, entonces $f(x)$ la cual es la versión periódica de $f_0(x)$, tiene a $\exp\{-2(\pi j)^2\}$ como sus coeficientes de Fourier.

Para $\sigma > 0$, sea

$$\psi_\sigma(x) = \frac{\sigma}{\sqrt{2\pi}} \sum_{j \in \mathbf{Z}} \exp \left\{ -\frac{\sigma^2}{2} (x + j)^2 \right\}.$$

Sea $\xi = \beta^X$, entonces los cálculos abajo mostrados muestran que $\log_\beta \xi \pmod{1}$ tiene a ψ_σ como su distribución límite.

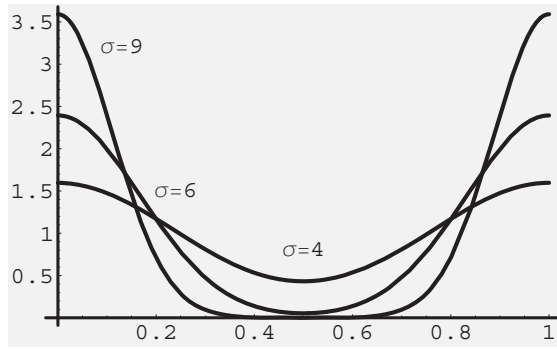


Figura 3.

En la figura 3 podemos ver la gráfica de ψ_σ , cuando $\sigma = 4, \sigma = 6$ y $\sigma = 9$. Observamos que ψ_σ se desvia considerablemente de la distribución uniforme cuando σ es grande.

Consideremos, por ejemplo, la variable aleatoria $\xi = 10^X$, donde

$$X = \frac{Y_1 + \cdots + Y_{10}}{9\sqrt{10}},$$

es la suma de diez variables aleatorias. Cada Y_ℓ está uniformemente distribuida en $[-1/2, 1/2]$.

d	Benford	ψ_9	Obs
1	.301	.496	.493
2	.176	.003	.002
3	.124	.000	.000
4	.096	.003	.002
5	.079	.019	.018
6	.066	.058	.059
7	.057	.109	.109
8	.051	.148	.155
9	.045	.159	.158

Tabla 2.

En la tercera columna de la tabla 2 podemos ver los valores de $\int_A \psi_9(x) dx$, donde $A = [\log_{10} d, \log_{10} d + 1)$ y $1 \leq d < 10$. En la cuarta columna de la misma tabla, podemos ver las frecuencias observadas de la ocurrencia para cada dígito d como el primer dígito significativo de ξ , cuando ξ fue observada un total de 10,000 veces. Los datos observados están cerca de ser predecidos por los valores en la tercera columna, pero muy lejos de seguir la ley Benford.

Capítulo 4

Un modelo simple para la ley Benford

Como ya vimos en el capítulo 3, los teoremas 3.2 y 3.3 nos proporcionan dos condiciones suficientes para que una variable aleatoria satisfaga la ley Benford, pero esas condiciones son bastante restrictivas. Entonces ¿Porqué observamos la ley Benford en la Naturaleza? En este capítulo consideramos un modelo bastante simple para la ley Benford en el grupo $\mathbf{Z}_2 \times \mathbf{Z}_2$. Observando el comportamiento de la ley Benford en este grupo, podemos reconciliarnos con la idea de la universalidad de la ley Benford. Después generalizaremos la descripción del modelo simple a los reales positivos, abarcando el caso general.

4.1 Introducción

Como vimos en el capítulo 3, el problema de si ξ satisface la ley Benford nos hace preguntarnos acerca de la distribución de $\log_\beta \xi \pmod{1}$. Si suponemos que la densidad f de $\log_\beta \xi$ es conocida, entonces queremos calcular la densidad \tilde{f} de $\log_\beta \xi \pmod{1}$. Este cálculo se lleva a cabo con la ayuda de la fórmula de la suma de Poisson como ya lo hemos hecho anteriormente.

En el caso general f tendrá como soporte a la recta real \mathbf{R} , mientras \tilde{f} tendrá como soporte el intervalo unitario $\mathbf{T} = [0, 1)$. Si pensamos a \mathbf{R} como producto de dos grupos \mathbf{Z} y \mathbf{T} , entonces veremos que la densidad f definida en $\mathbf{R} = \mathbf{Z} \times \mathbf{T}$ induce una densidad \tilde{f} en \mathbf{T} . La ley Benford se satisfecerá cuando la densidad inducida \tilde{f} en \mathbf{T} es la uniforme.

En la sección 2 de este capítulo consideraremos una distribución de probabilidad ρ definida en el grupo $\Omega = \Omega_1 \times \Omega_2$ y nos preguntamos acerca de la distribución de probabilidad inducida $\tilde{\rho}$ para alguno de sus factores, sin pérdida de generalidad pensemos en Ω_1 . Cuando $\tilde{\rho}$ sea la distribución uniforme en Ω_1 , entonces diremos que ρ obedece la ley Benford en Ω . Para propósitos de entender la naturaleza de la ley Benford consideraremos el caso simple del grupo $\Omega = \mathbf{Z}_2 \times \mathbf{Z}_2$. Veremos que en este sencillo caso la ley Benford no se satisfecerá puntualmente, pero se cumplirá en promedio. El significado exacto de esta afirmación quedará claro en la siguiente sección.

En la sección 3, generalizaremos la descripción del modelo simple en el grupo $\mathbf{Z}_2 \times \mathbf{Z}_2$, reformulándola para el caso general donde la distribución de probabilidad tiene como soporte a $\mathbf{R} = \mathbf{Z} \times \mathbf{T}$. Con esta reformulación evitamos tomar esperanzas sobre un espacio muestral multidimensional.

En el capítulo 3 hicimos ver que una variable aleatoria ξ no seguirá la ley Benford a menos que satisfaga algunas relaciones de simetría muy restrictivas. Por lo tanto, la ley Benford debería ser un suceso poco común. Esto está en contraste con lo que se ha observado empíricamente. El modelo simple en este capítulo nos hace volver a pensar en la idea de la universalidad de la ley Benford.

4.2 Benford en el grupo $\mathbf{Z}_2 \times \mathbf{Z}_2$

Sean α, β, γ y δ los elementos del grupo $\Omega = \mathbf{Z}_2 \times \mathbf{Z}_2$. También consideremos el tetraedro $\mathcal{T} = \{(x, y, z) \in \mathbf{R}^3 : 0 \leq x, 0 \leq y, 0 \leq z, x + y + z \leq 1\}$. Entonces, una distribución de probabilidad en Ω está determinada por un punto $(p_1, p_2, p_3) \in \mathcal{T}$

tal que

$$p_1 = \mathbf{P}\{\alpha\}, \quad p_2 = \mathbf{P}\{\beta\}, \quad p_3 = \mathbf{P}\{\gamma\} \quad \text{y} \quad p_4 = 1 - p_3 - p_2 - p_1 = \mathbf{P}\{\delta\}.$$

Pensemos que uno de los factores de Ω puede ser visto como $\Omega_1 = \{\{\alpha, \beta\}, \{\gamma, \delta\}\}$. Si la distribución de probabilidad $\rho = (p_1, p_2, p_3)$ está dada en Ω , entonces obtenemos una distribución de probabilidad $\tilde{\rho}$ en Ω_1 . En efecto, tenemos

$$\mathbf{P}\{\{\alpha, \beta\}\} = p_1 + p_2 \quad \text{y} \quad \mathbf{P}\{\{\gamma, \delta\}\} = p_3 + p_4.$$

Entonces ρ obedece la ley Benford sí y sólo sí $p_1 + p_2 = 1/2$. Ahora supongamos que los puntos (p_1, p_2, p_3) son aleatorios en \mathcal{T} y están uniformemente distribuidos en \mathcal{T} . Entonces es claro que una medida aleatoria $\rho = (p_1, p_2, p_3)$ en Ω satisficará la ley Benford con probabilidad 0. En efecto, la probabilidad de que $p_1 + p_2 = 1/2$ es igual a cero. Como se puede observar en la figura 1 el volumen del plano $p_1 + p_2 = 1/2$ en \mathcal{T} es cero.

Dada una medida de probabilidad fija $\rho = (p_1, p_2, p_3)$ en Ω , podemos construir una lista $L(\rho)$ de elementos observados en Ω . Esta lista puede ser construida con la ayuda de una variable aleatoria U , uniformemente distribuida en $[0, 1)$. De hecho, diremos que un elemento observado en $L(\rho)$ es igual a

$$\begin{cases} \alpha & \text{si } 0 \leq U < p_1, \\ \beta & \text{si } p_1 \leq U < p_1 + p_2, \\ \gamma & \text{si } p_1 + p_2 \leq U < p_1 + p_2 + p_3, \\ \delta & \text{si } p_1 + p_2 + p_3 \leq U < 1. \end{cases}$$

Dada una lista $L(\rho)$, sea m_1 el número de elementos en $L(\rho)$ que provienen de $\{\alpha, \beta\}$ y sea m_2 el número de elementos en $L(\rho)$ que provienen de $\{\gamma, \delta\}$. Si m

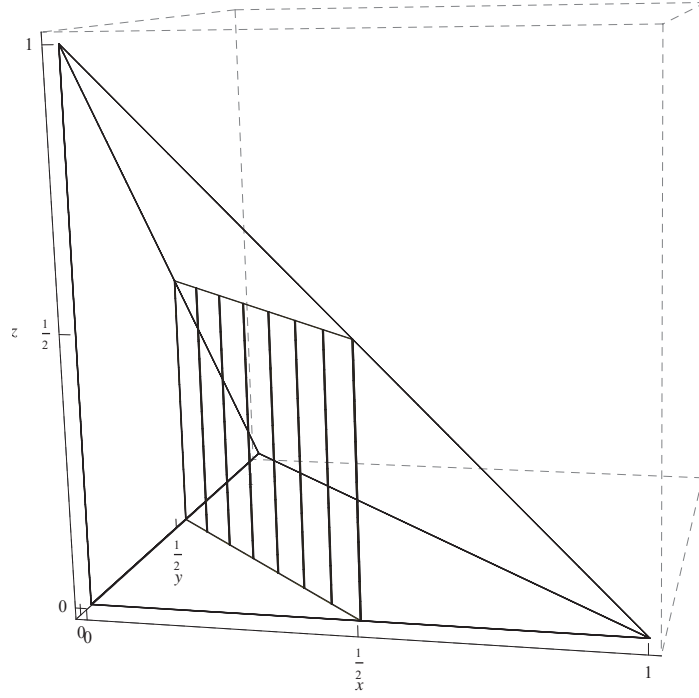


Figura 1. Tetraedro \mathcal{T} .

es el número de elementos en la lista $L(\rho)$, entonces la ley de los grandes números implica que

$$\lim_{m \rightarrow \infty} \frac{m_1}{m} = p_1 + p_2 \quad \text{y} \quad \lim_{m \rightarrow \infty} \frac{m_2}{m} = p_3 + p_4.$$

Un observador que examine la lista $L(\rho)$ sin conocer cómo fue construida dirá que ρ seguirá la ley Benford en Ω si m_1/m y m_2/m son aproximadamente $1/2$. Dadas q medidas $\rho_1, \rho_2, \dots, \rho_q$ en Ω , podemos construir una lista grande \mathcal{L} concatenando las listas $L(\rho_1), L(\rho_2), \dots, L(\rho_q)$. Si cada lista $L(\rho_i)$ tiene m elementos, entonces \mathcal{L} tendrá $n = mq$ elementos. Dada una lista \mathcal{L} construida de esta manera, debemos

decidir cuándo se observará la ley Benford. Para ésto, sea n_1 el número de elementos en \mathcal{L} los cuales pertenecen a $\{\alpha, \beta\}$ y sea n_2 el número de elementos en \mathcal{L} los cuales pertenecen a $\{\gamma, \delta\}$. Entonces la ley Benford se observará si

$$\lim_{n \rightarrow \infty} \frac{n_1}{n} = \frac{1}{2} \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{n_2}{n} = \frac{1}{2}. \quad (2)$$

Teorema 1. *La relación (2) se satisface para un número fijo m y cuando la convergencia es entendida como convergencia en probabilidad.*

Demostración. Para probar el teorema 1, sea X_j el número de elementos en $L(\rho_j)$ que pertenecen a $\{\alpha, \beta\}$. Entonces para cada $0 \leq k \leq m$,

$$\mathbf{P}\{X_j = k | \rho_j = (x, y, z)\} = \binom{m}{k} (x + y)^k (1 - x - y)^{m-k}.$$

Por lo tanto

$$\mathbf{P}\{X_j = k\} = 6 \int_0^1 \int_0^{1-x} \int_0^{1-x-y} \binom{m}{k} (x + y)^k (1 - x - y)^{m-k} dz dy dx.$$

Así que

$$\begin{aligned} \mathbf{E}(X_j) &= \sum_{k=0}^m k \mathbf{P}\{X_j = k\} \\ &= 6 \int_0^1 \int_0^{1-x} \int_0^{1-x-y} \sum_{k=0}^m k \binom{m}{k} (x + y)^k (1 - x - y)^{m-k} dz dy dx \\ &= 6 \int_0^1 \int_0^{1-x} \int_0^{1-x-y} m(x + y) dz dy dx \\ &= \frac{m}{2}. \end{aligned}$$

Para calcular la varianza $\mathbf{V}(X_j)$ note que

$$\mathbf{E}(X_j | \rho_j = (x, y, z)) = m(x + y) \quad \text{y} \quad \mathbf{V}(X_j | \rho_j = (x, y, z)) = m(x + y)(1 - x - y).$$

Note también que

$$\mathbf{E}[\mathbf{V}((X_j|\rho_j = (x, y, z)))] = 6 \int_0^1 \int_0^{1-x} \int_0^{1-x-y} m(x+y)(1-x-y) dz dy dx = \frac{m}{5}.$$

Además

$$\mathbf{V}[\mathbf{E}(X_j|\rho_j = (x, y, z))] = 6m^2 \int_0^1 \int_0^{1-x} \int_0^{1-x-y} (x+y)^2 dz dy dx - \left(\frac{m}{2}\right)^2 = \frac{m^2}{20}.$$

$$\text{Por lo tanto } \mathbf{V}(X_j) = \mathbf{E}[\mathbf{V}(X_j|\rho_j)] + \mathbf{V}[\mathbf{E}(X_j|\rho_j)] = \frac{m}{5} + \frac{m^2}{20}.$$

Ya que $n_1 = X_1 + \dots + X_q$, entonces el valor esperado de n_1/n es igual a $1/2$.

La varianza de n_1/n es igual a

$$\frac{1}{m^2 q^2} q \left(\frac{m^2}{20} + \frac{m}{5} \right) \leq \frac{1}{4q}.$$

De la desigualdad de Chebyshev obtenemos

$$\mathbf{P} \left\{ \left| \frac{n_1}{n} - \frac{1}{2} \right| \geq \epsilon \right\} \leq \frac{1}{4q\epsilon^2} \rightarrow 0.$$

cuando $q \rightarrow \infty$ (por lo tanto $n \rightarrow \infty$). ■

4.3 Benford en el grupo $\mathbf{R} = \mathbf{Z} \times \mathbf{T}$

Podemos obtener un mejor entendimiento del modelo simple que proponemos en la sección 2, notando que dada cualquier medida de probabilidad $\rho = (x, y, z) \in \mathcal{T}$, hay una medida $\rho' = (x', y', z') \in \mathcal{T}$ tal que la mezcla $(\rho + \rho')/2$ satisface la ley Benford. De hecho es suficiente tomar

$$(x', y', z') = \left(\frac{x}{x+y} - x, \frac{y}{x+y} - y, z' \right),$$

donde $0 \leq z' < x + y$. Si $(p_1, p_2, p_3) = (\rho + \rho')/2$, entonces $p_1 + p_2 = 1/2$. Así, para toda medida de probabilidad ρ en $\Omega = \mathbf{Z}_2 \times \mathbf{Z}_2$, hay otra medida ρ' en Ω tal que la mezcla satisface la ley Benford. Dicho con otras palabras, ρ' cancela la desviación de ρ respecto a la distribución Benford, es decir, la distribución Benford puede ser vista como “el centro de masa” del tetraedro \mathcal{T} .

Ahora, afirmamos que la misma cancelación de las desviaciones respecto a la distribución Benford toma lugar en $\mathbf{R} = \mathbf{Z} \times \mathbf{T}$. De hecho, dada una densidad de probabilidad f en \mathbf{R} , la densidad inducida en \mathbf{T} es

$$f^p(x) = \sum_{j \in \mathbf{Z}} f(x + j).$$

Note que escribimos f^p en lugar de \tilde{f} . Diremos que f^p es la versión periódica de f . Consideremos la familia paramétrica de las funciones densidad en \mathbf{R} dadas por

$$f_\theta(x) = f([x] + \langle x + \theta \rangle), \quad \text{con } 0 \leq \theta < 1,$$

donde $[x] = \sup\{n \in \mathbf{Z} : n \leq x\}$ y $\langle x \rangle = x - [x]$. Cada una de estas densidades en \mathbf{R} induce una densidad en \mathbf{T} dada por $(f_\theta)^p$. Note que $(f_\theta)^p = (f^p)_\theta$, ya que

$$(f_\theta)^p(x) = \sum_{j \in \mathbf{Z}} f([x] + \langle x + \theta \rangle + j)$$

y

$$(f^p)_\theta(x) = \sum_{j \in \mathbf{Z}} f([x + j] + \langle x + \theta + j \rangle).$$

Basta notar que $[x + j] = [x] + j$ y $\langle x + j + \theta \rangle = \langle x + \theta \rangle$ para obtener la igualdad. La observación decisiva para la reformulación de la idea que utilizamos en el grupo $\Omega = \mathbf{Z}_2 \times \mathbf{Z}_2$, es que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_{\frac{j}{n}}^p(x) = \int_0^1 f^p(x) = 1.$$

Note que la parte izquierda corresponde a una suma de Riemann y f^p es una densidad de probabilidad en \mathbf{T} . Por lo tanto, dada una densidad f en \mathbf{R} , siempre podemos encontrar densidades de probabilidad $f_{\frac{1}{n}}, f_{\frac{2}{n}}, \dots, f_{\frac{n-1}{n}}$ en \mathbf{R} tal que la mezcla (note que $f_0 = f$)

$$m_f = \frac{1}{n}(f + f_{\frac{1}{n}} + f_{\frac{2}{n}} + \dots + f_{\frac{n-1}{n}}),$$

satisface la ley Benford para cualquier grado de precisión. Así, dada una densidad f en \mathbf{R} , podemos encontrar densidades en \mathbf{R} que cancelarán la desviación de f con respecto a la distribución Benford.

Podemos medir la desviación de m_f respecto a la ley Benford. Dada una función $g : \mathbf{T} \rightarrow \mathbf{R}$, escribiremos $\|g\| = (\int_0^1 |g(x)|^2 dx)^{1/2}$. Del teorema de Parseval sabemos que

$$\|g\|^2 = \sum_{j \in \mathbf{Z}} |\hat{g}(j)|^2,$$

donde $\hat{g}(j) = \int_0^1 g(x) e^{-2\pi i j x} dx$ son los coeficientes de Fourier de g . Ahora calcularemos los coeficientes de Fourier de m_f^p , que es la versión periódica de m_f ,

$$\begin{aligned} \hat{m}_f^p(j) &= \int_0^1 m_f^p(x) e^{-2\pi i j x} dx \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \int_0^1 f^p\left(x + \frac{k}{n}\right) e^{-2\pi i j x} dx \\ &= \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i j \frac{k}{n}} \int_0^1 f^p(x) e^{-2\pi i j x} dx. \end{aligned}$$

Por lo tanto

$$\hat{m}_f^p(j) = \hat{f}^p(j) \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i j \frac{k}{n}} = \begin{cases} 0 & \text{si } j \not\equiv 0 \pmod{n}, \\ \hat{f}^p(j) & \text{si } j \equiv 0 \pmod{n}. \end{cases}$$

Sea $\mathbf{1} : \mathbf{T} \rightarrow \mathbf{R}$ la función constante 1, $\mathbf{1}(x) = 1$ para toda $x \in \mathbf{T}$. Entonces tenemos

$$\|\mathbf{1} - m_f^p\|^2 = \sum_{j \in \mathbf{Z} \setminus \{0\}} |\hat{f}^p(nj)|^2 \rightarrow 0,$$

cuando $n \rightarrow \infty$. Esto confirma nuestra afirmación acerca el hecho de que m_f satisface la ley Benford de manera aproximada cuando n es suficientemente grande.

La razón del porqué m_f satisface de manera aproximada la ley Benford es que los coeficientes de Fourier de la versión periódica de f_θ satisfacen la relación $\hat{f}_\theta^p(j) = e^{2\pi i j \theta} \hat{f}^p(j)$. Con ésto en mente, dada cualquier variable aleatoria absolutamente continua con valores en \mathbf{R}^+ , podemos dar una segunda construcción de una mezcla que seguirá de manera aproximada la ley Benford.

Sea ξ una variable aleatoria absolutamente continua con valores en \mathbf{R}^+ , con g como su función densidad. Para toda $\sigma > 0$, sea $f(x, \sigma)$ la densidad de $\log_\beta(\sigma\xi)$. Por el corolario 3.1 tenemos que la versión periódica de $f(x, \sigma)$ es

$$f^p(x, \sigma) = \sum_{j \in \mathbf{Z}} \exp \left\{ 2\pi i j \left(x - \frac{\log \sigma}{\log \beta} \right) \right\} M_g \left\{ -\frac{2\pi i j}{\log \beta} \right\}.$$

Se sigue inmediatamente que para los coeficientes de Fourier de $f^p(x, \sigma)$ se tiene

$$\hat{f}^p(j, \beta^\delta) = e^{-2\pi i j \delta} \hat{f}^p(j, 1),$$

para todo número δ real. De aquí, la misma relación se mantiene entre $\hat{f}^p(j, \beta^\delta)$ y $\hat{f}^p(j, 1)$ por un lado, y entre $\hat{f}_\theta^p(j)$ y $\hat{f}^p(j)$ por el otro. Así, tenemos de nuevo que la mezcla

$$\frac{1}{n} (f(x, 1) + f(x, \beta^{\frac{1}{n}}) + \cdots + f(x, \beta^{\frac{n-1}{n}})),$$

aproximadamente satisface la ley Benford.

A continuación presentaremos un ejemplo que podría ser representativo de lo que observamos en datos de la vida diaria. Sea $\xi \sim N(0, \sigma_0)$, es decir, ξ es una variable aleatoria normal con media 0 y varianza $\sigma_0^2 = (0.1328)^2$. Sean $\eta \sim N(0, 4.1995)$ y ζ la variable aleatoria cuya función densidad es la mezcla de las densidades de $|\xi|$ y $|\eta|$, así $g_\zeta = (g_{|\xi|} + g_{|\eta|})/2$.

d	Benford	ζ	$ \xi $	$ \eta $
1	.301	.300	.385	.216
2	.176	.176	.173	.179
3	.124	.120	.085	.155
4	.096	.096	.065	.128
5	.079	.081	.061	.101
6	.066	.069	.059	.078
7	.057	.058	.057	.059
8	.051	.050	.055	.045
9	.045	.044	.053	.035

Tabla 1. Frecuencias teóricas del primer dígito significativo de ζ , $|\xi|$ y $|\eta|$.

Note que $|\xi|$ y $|\eta|$ están bastante lejos de la distribución Benford, pero ζ está bastante cerca.

Con este análisis podemos decir que ya no hay misterio alguno en la ley del dígito significativo. Ahora sabemos porqué algunos conjuntos de datos naturales satisfacen la ley Benford. La idea es tener una lista de muestras aleatorias que provienen de distribuciones aleatorias.

Ahora lo que necesitamos es una manera de medir la desviación de un conjunto de datos respecto a la ley Benford, ésto con el fin de obtener aplicaciones óptimas. En el siguiente capítulo intentaremos simular una alteración en los datos (fraude)

de un conjunto que satisface Benford y utilizaremos la prueba χ^2 para medir la desviación respecto a la ley Benford. Con ésto veremos qué tan útil puede ser la prueba χ^2 en conjunto con la ley Benford para detectar datos manipulados.

Capítulo 5

Simulando fraude

En este capítulo simularemos una alteración en los datos de un conjunto, esto con el fin de intentar emular un fraude. Lo que haremos será generar una lista de datos que satisface la ley Benford, ya que como sabemos los datos contables representan una buena aproximación a la distribución Benford. Después, alteraremos la lista de manera aleatoria por medio de un parámetro, que representara el nivel de alteración de los datos en la lista. Es de esperar que la lista alterada presente una desviación significativa de la ley Benford. Entonces necesitamos una manera de medir la discrepancia entre lo práctico y lo teórico.

La mayoría de las veces el problema de medir la concordancia entre teoría y hechos se presenta de la siguiente forma: tenemos a nuestra disposición una muestra de n valores observados de una cierta variable aleatoria, y deseamos saber si puede considerarse como una variable aleatoria que tenga una distribución de probabilidad con ciertas propiedades. En algunos casos la distribución hipotética estará completamente especificada. En otros casos, dada una cierta clase de distribuciones, la pregunta es si la muestra podría haber sido obtenida de una variable aleatoria que tenga alguna distribución que pertenezca a la clase dada.

Consideremos el caso sencillo en que la distribución hipotética está completamente especificada, por ejemplo, por su función distribución $F(x)$. En tal caso, habremos de medir la hipótesis de que nuestra muestra procede de una variable con esta distribución.

Supongamos que la hipótesis que se va a medir es verdadera. Es de esperar que la función distribución $F^*(x)$ de la muestra sea una aproximación a la función distribución $F(x)$ hipotética. Trataremos de definir una medida no negativa de la desviación de F^* respecto a F . Esto puede hacerse de varias maneras, pero cualquier medida D de la desviación será función de los valores muestrales y tendrá una determinada distribución \mathcal{D} en el muestreo. Por medio de esta distribución \mathcal{D} podemos calcular la probabilidad $\mathbf{P}\{D > D_0\}$, que es la probabilidad de que la desviación D exceda a una cantidad dada D_0 . Esta probabilidad puede hacerse tan pequeña como queramos haciendo D_0 lo suficientemente grande.

Si la hipótesis es verdadera, cuando n sea lo suficientemente grande, la función distribución F^* de la muestra será aproximadamente igual a la función distribución F , por lo que es necesario utilizar alguna medida conveniente D de la desviación de F^* respecto a F . Mediante la distribución \mathcal{D} en el muestreo de D , hallaremos una cantidad D_0 tal que $\mathbf{P}\{D > D_0\} = \epsilon$. Si sucede un caso en el que un valor D sea tal que $D > D_0$, diremos que la desviación es significativa y consideraremos a la hipótesis desaprobada. Por el contrario, si $D \leq D_0$ se considera posible que la desviación se deba a fluctuaciones aleatorias, y los datos se consideran compatibles con la hipótesis. La probabilidad ϵ recibe el nombre de nivel de significación de la prueba y la escogemos dependiendo del nivel de exactitud que deseemos.

El problema que nos interesa es el caso sencillo en el que la medida se refiera a la concordancia entre la distribución F^* de un conjunto de valores muestrales y una distribución F teórica. Hablaremos entonces de una prueba de la bondad de ajuste.

Llamaremos hipótesis \mathcal{H} a la hipótesis de que nuestros datos constituyan una muestra de n valores de una variable aleatoria con función de probabilidad \mathcal{P} . Supongamos que \mathcal{P} está completamente especificada, de modo que en su expresión no aparecen parámetros desconocidos y que puede calcularse numéricamente la probabilidad $\mathcal{P}(S)$ para cualquier evento dado S . Se trata ahora de elaborar un método para medir si nuestros datos pueden considerarse conciliables con la hipótesis \mathcal{H} . Si la hipótesis es verdadera, la distribución en la muestra que es la sencilla distribución discreta F^* , que se obtiene al situar la masa $1/n$ en cada uno de los n puntos observados, puede considerarse como una aproximación a la distribución especificada por \mathcal{P} , llamémosle F . Debido a fluctuaciones aleatorias, lo general será que las dos distribuciones no coincidan, pero para valores grandes de n puede esperarse que la distribución F^* en la muestra forme una aproximación de la distribución F . Parece natural introducir alguna medida D de la desviación entre ambas distribuciones y basar nuestra prueba de la bondad de ajuste en las propiedades de la distribución \mathcal{D} en el muestreo de esa medida.

5.1 Prueba χ^2

Existen diversos modos de construir tales medidas de la desviación, siendo los más utilizados los relacionados con la χ^2 introducida por K. Pearson. Supongamos que el espacio de la variable aleatoria se divide en r partes S_1, \dots, S_r sin puntos comunes, y sean p_1, \dots, p_r los valores correspondientes de la función de probabilidad \mathcal{P} , de modo que

$$p_i = \mathcal{P}(S_i) \quad \text{y} \quad \sum_{i=1}^r p_i = 1,$$

donde $p_i > 0$ para toda i . Las r partes S_i pueden ser los r grupos en que se han dispuesto nuestros valores muestrales a efectos de tabulación. Sean ν_1, \dots, ν_r las respectivas frecuencias de estos grupos en la muestra, de manera que ν_i son los valores muestrales que pertenecen al conjunto S_i , y tales que $\sum_{i=1}^r \nu_i = n$, es decir, nuestra muestra es de n elementos.

Nuestro primer objetivo es encontrar una medida conveniente de la desviación entre la distribución de la muestra F^* y la hipotética F . Cada uno de los conjuntos S_i lleva como masa a ν_i/n en la distribución F^* y la masa p_i en F . Entonces, de acuerdo con el principio general de los mínimos cuadrados, adoptaremos como medida de la desviación una expresión de la forma $\sum_{i=1}^r c_i \left(\frac{\nu_i}{n} - p_i\right)^2$, donde los coeficientes c_i pueden elegirse de manera más o menos arbitraria. Pearson demostró que si tomamos $c_i = \frac{n}{p_i}$, obtenemos una medida de la desviación cuyas propiedades son esencialmente sencillas. De este modo tenemos la expresión

$$\chi^2 = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i} = \sum_{i=1}^r \frac{\nu_i^2}{np_i} - n. \quad (1)$$

Así χ^2 se expresa en función de las frecuencias observadas ν_i y de las frecuencias esperadas np_i de los r grupos. Note que escribimos χ^2 en lugar de D .

Pasaremos a investigar la distribución en el muestreo de χ^2 , suponiendo en todo lo que sigue que la hipótesis \mathcal{H} es verdadera. El siguiente teorema debido a K. Pearson afirma que cuando aumenta el tamaño de la muestra, la distribución de χ^2 en el muestreo tiende a una distribución límite completamente independiente de la función de probabilidad hipotética \mathcal{P} .

Teorema 1. *Cuando $n \rightarrow \infty$ la distribución de χ^2 en el muestreo tiende a la*

distribución definida por la función de densidad.

$$K_{r-1}(x) = \frac{1}{2^{\frac{r-1}{2}} \Gamma(\frac{r-1}{2})} x^{\frac{r-3}{2}} e^{-\frac{x}{2}} \quad (x > 0). \quad (2)$$

Diremos que en el límite, χ^2 se distribuye según una distribución χ^2 con $r-1$ grados de libertad.

Para la demostración véase [7].

Por medio de este teorema podemos introducir ahora una prueba de la bondad de ajuste para la hipótesis \mathcal{H} antes considerada. Lo primero que tenemos que hacer es calcular el valor de χ^2 por medio de (1), una vez hecho esto, necesitamos un nivel de confianza p (normalmente conocido como p -valor), lo cuál nos dará un valor crítico χ_0 , que anteriormente llamamos D_0 , que se relacionan por medio de

$$p = \int_{\chi_0}^{\infty} K_{r-1}(x) dx,$$

donde $K_{r-1}(x)$ es como en (2). Ahora queremos calcular la probabilidad de que la variable χ^2 exceda a χ_0^2 , que por el teorema 1 será aproximadamente p , es decir,

$$\mathbf{P}\{\chi^2 > \chi_0^2\} \approx \int_{\chi_0}^{\infty} K_{r-1}(x) dx = p.$$

Si la hipótesis \mathcal{H} es verdadera, queda prácticamente excluído que en una sola muestra podamos encontrar un valor de χ^2 mayor que χ_0^2 . Si en una muestra real hallamos un valor $\chi^2 > \chi_0^2$, diremos según ésto, que la muestra presenta una desviación significativa respecto a la hipótesis \mathcal{H} , y desaprobaremos esta hipótesis al menos hasta que no dispongamos de nuevos datos. La probabilidad de que ésto ocurra en el caso de ser efectivamente verdadera la hipótesis \mathcal{H} , ésto es, de que haya

sido desaprobada \mathcal{H} injustamente es precisamente la probabilidad $\mathbf{P}\{\chi^2 > \chi_0^2\}$, que es aproximadamente igual a p . Por otra parte, si hallamos un valor $\chi^2 \leq \chi_0^2$, consideraremos este resultado conciliable con la hipótesis \mathcal{H} . Evidentemente, un solo resultado de este tipo no puede considerarse como evidencia suficiente de la veracidad de la hipótesis. Si queremos obtener tal evidencia, tendremos que aplicar la prueba repetidamente a nuevos datos.

Análogamente, lo que podemos hacer para comprobar que la hipótesis es verdadera, es calcular χ^2 por medio de (1) y después calcular

$$\varrho = \int_{\chi^2}^{\infty} K_{r-1}(x) dx. \quad (3)$$

Si $\varrho > p$, entonces el resultado es conciliable con la hipótesis. Si por el contrario $\varrho \leq p$, queda desaprobada la hipótesis. Para entender mejor la prueba χ^2 veamos el siguiente ejemplo.

Ejemplo 1. Pensemos en un dado que se lanza 120 veces, si el dado no está “alterado” la probabilidad de que salga el número i , con $i \in \{1, \dots, 6\}$, en cada lanzamiento es $\mathbf{P}\{i\} = p_i = 1/6$. Supongamos que el experimento arrojó los resultados que se observan en la tabla 2, donde n es la longitud de la muestra, ν_i son las frecuencias observadas para cada S_i y np_i son las frecuencias esperadas según nuestra hipótesis.

i	1	2	3	4	5	6
ν_i	15	22	17	26	12	28
np_i	20	20	20	20	20	20

Tabla 1.

Para verificar que nuestra hipótesis es correcta, es decir, que el dado no está alterado utilizaremos la prueba χ^2 . Tenemos que $r - 1 = 5$ y que

$$\chi^2 = \sum_1^6 \frac{(\nu_i - np_i)^2}{np_i} = \frac{101}{10}.$$

Si consideramos a nuestro p -valor como 0.05, de (3) obtenemos que

$$\int_{\frac{101}{10}}^{\infty} K_5(x) dx = 0.07 > 0.05,$$

por lo que la hipótesis es aprobada y podemos considerar al dado sin alterar.

5.2 Simulaciones

A continuación expondremos un programa hecho en Mathematica, con el que intentamos simular lo que vendría siendo una alteración de datos (fraude). Esto con el fin de poder observar que tan útil puede ser la prueba χ^2 para detectar anomalías en datos que satisfacen Benford, con respecto esos mismos datos alterados que en teoría deberían de tener una desviación significativa con respecto a la distribución Benford. Para ver el código del programa que se utilizó en la simulación véase el código Simula1 en el apéndice.

El funcionamiento del programa lo explicaremos en los siguientes pasos:

- i. Generamos una gran lista \mathcal{L} conteniendo m sublistas, cada sublista contiene n elementos, ésto para simular datos contables que satisfacen Benford. Para generar las sublistas usamos una variable aleatoria de la forma $10^U \times N$, donde U es una variable aleatoria distribuida uniformemente en $(0, 1)$ y N es una variable aleatoria con distribución Normal.

ii. Fijamos un parámetro $\omega \in [0, 1]$, con el que alteraremos los datos en \mathcal{L} . Este parámetro representará el abuso en la falsificación de datos, obteniendo \mathcal{L}' . La forma de hacer ésto es con la ayuda de una variable aleatoria V uniformemente distribuida en $(0, 1)$. Para cada j -ésima sublista contenida en \mathcal{L} alteramos el i -ésimo elemento. Para ésto generamos V_i tal que si $V_i < \omega$ entonces cambiamos el primer dígito significativo del i -ésimo elemento de la j -ésima sublista de la siguiente forma:

$$\left\{ \begin{array}{ll} \text{si tiene a 1} & \text{cambiar por 2,} \\ \text{si tiene a 2} & \text{cambiar por 3,} \\ \text{si tiene a 3} & \text{cambiar por 4,} \\ \text{si tiene a 4} & \text{cambiar por 5,} \\ \text{si tiene a 5} & \text{cambiar por 6,} \\ \text{si tiene a 6} & \text{cambiar por 7,} \\ \text{si tiene a 7} & \text{cambiar por 8,} \\ \text{si tiene a 8} & \text{cambiar por 9,} \\ \text{si tiene a 9} & \text{cambiar por 1.} \end{array} \right.$$

Si por el contrario $V_i \geq \omega$, no alteramos el dato y pasamos con el $(i + 1)$ -ésimo elemento de la j -ésima sublista, esto hasta el n -ésimo elemento.

iii. Después de aplicar los pasos anteriormente descritos (i-ii) a la j -ésima sublista pasamos a la $(j + 1)$ -ésima sublista, ésto hasta llegar a la m -ésima sublista de \mathcal{L} . Esta alteración de datos es muy artificial en el sentido que una persona que falsifica datos normalmente no alteraría los datos de esta forma, pero para fines prácticos nos es suficiente, ya que podremos utilizar la prueba χ^2 para medir la desviación entre los datos de \mathcal{L} y \mathcal{L}' con respecto a la distribución Benford. Note que cuando $\omega = 0$, no se altera ningún dato ($\mathcal{L} = \mathcal{L}'$), mientras que cuando $\omega = 1$, alteramos todos los datos de la lista.

iv. Aplicamos la prueba χ^2 a las sublistas de cada lista \mathcal{L} y \mathcal{L}' , obteniendo un p -valor para cada una de las sublistas, digamos p_j y p'_j .

v. Fijamos un nivel de confianza ρ , tal que si $p \leq \rho$, diremos que representa una desviación significativa de la distribución Benford y por lo tanto puede ser sospecha de alguna alteración de datos. En el programa utilizamos $\rho = 0.05$. Esta sospecha no puede ser una afirmación de fraude debido a que aunque la lista grande \mathcal{L} fue construida de tal forma que satisfaga Benford, debido a fluctuaciones aleatorias es posible que en algunos casos \mathcal{L} se desvíe de la distribución Benford. Si por el contrario $p > \rho$, diremos que no hay sospecha alguna de falsificación de datos. Esto no afirma la veracidad de los datos pues debido a fluctuaciones aleatorias puede suceder que \mathcal{L}' tenga una desviación menor con respecto a la distribución Benford en comparación con \mathcal{L} .

vi. Comparamos todos los p_i y p'_i según (v) y calculamos el porcentaje en el que $p_j \leq \rho$ y $p'_j \leq \rho$. En teoría el porcentaje de $p_j \leq \rho$ debería de ser ρ . Así, podremos medir si hay alguna diferencia notable entre ambos porcentajes. Esto dependerá del parámetro ω utilizado.

En la tabla 2 se muestran los resultados obtenidos en la simulación de una lista con 2,000 sublistas cada una con 30 elementos, variando el parámetro ω .

ω	ρ	n	m	$\%p_j \leq \rho$	$\%p'_j \leq \rho$
0.1	0.05	30	2,000	0.0515	0.0625
0.2	0.05	30	2,000	0.0465	0.0675
0.3	0.05	30	2,000	0.0455	0.0960
0.4	0.05	30	2,000	0.0620	0.1245
0.5	0.05	30	2,000	0.0525	0.1910
0.6	0.05	30	2,000	0.0515	0.2685
0.7	0.05	30	2,000	0.0520	0.3355
0.8	0.05	30	2,000	0.0520	0.4680
0.9	0.05	30	2,000	0.0465	0.5955
1	0.05	30	2,000	0.0480	0.7400

Tabla 2.

En teoría el porcentaje de $p_j \leq \rho$ debe de ser 0.05. Como se puede ver en la quinta columna de la tabla 2 el porcentaje de $p_j \leq \rho$ oscila cerca de 0.05, que es normal debido a las fluctuaciones aleatorias y a que el número de elementos en las sublistas es pequeño.

Podemos ver que la discrepancia notable entre los porcentajes $p_j \leq \rho$ y $p'_j \leq \rho$ se da cuando $\omega = 0.3$, es decir, en este caso la prueba χ^2 detectó anomalías en aquellas listas en el que la alteración de datos es mayor o igual que el 30%, cuando $n = 30$. Uno de los problemas fundamentales es encontrar una medida adecuada con la que podremos detectar anomalías para parámetros pequeños, es decir, detectar fraudes sutiles.

Aunque ya existen distintas pruebas como la de los primeros dos dígitos y la de los últimos dos dígitos que tienen fines concretos, debemos desarrollar más herramientas (no necesariamente relacionadas con Benford) que en conjunto con las

aplicaciones actuales basadas en Benford puedan concluir en ser una herramienta poderosa para detectar fraude. Así, las aplicaciones de la ley Benford en detección de fraude son imprescindibles.

Ahora siguiendo los pasos descritos anteriormente hacemos la misma simulación variando el parámetro ω , salvo que ahora $n = 200$ y $m = 100$. Los resultados se pueden observar en la tabla 3.

ω	ρ	n	m	$\%p_j \leq \rho$	$\%p'_j \leq \rho$
0.1	0.05	200	100	0.03	0.03
0.2	0.05	200	100	0.01	0.22
0.3	0.05	200	100	0.04	0.40
0.4	0.05	200	100	0.08	0.70
0.5	0.05	200	100	0.02	0.92
0.6	0.05	200	100	0.05	1
0.7	0.05	200	100	0.06	1
0.8	0.05	200	100	0.02	1
0.9	0.05	200	100	0.06	1
1	0.05	200	100	0.03	1

Tabla 3.

Aquí la discrepancia entre $p_j \leq \rho$ y $p'_j \leq \rho$ se da a partir de $\omega = 0.2$. Es decir, cuando $n = 200$ la prueba χ^2 puede detectar anomalías cuando la alteración de datos es mayor o igual que el 20%. La principal razón de las diferencias entre la tabla 1 y 2 es que la prueba χ^2 es bastante confiable cuando las muestras tienden a ser grandes. En la tabla 2 tenemos $n = 30$, mientras que en la tabla 3 tenemos $n = 200$, es por eso que la diferencia entre los porcentajes p'_j tiende a ser mayor.

La siguiente tabla fue generada de la misma forma que la tabla 2, pero ahora

aplicamos los pasos (i)-(vi) 20 veces, generando 20 listas que contienen cada una 100 sublistas, cada sublista contiene 30 elementos. Después calculamos los porcentajes $p \leq \rho$ entre las 20 listas \mathcal{L} y \mathcal{L}' . En la tabla 4 se muestran los resultados.

ω	ρ	n	m	$\%p_j \leq \rho$	$\%p'_j \leq \rho$
0.1	0.05	30	100	0.0475	0.0575
0.2	0.05	30	100	0.0530	0.0830
0.3	0.05	30	100	0.0485	0.1015
0.4	0.05	30	100	0.0390	0.1325
0.5	0.05	30	100	0.0520	0.1870
0.6	0.05	30	100	0.0455	0.2470
0.7	0.05	30	100	0.0435	0.3590
0.8	0.05	30	100	0.0520	0.4670
0.9	0.05	30	100	0.0570	0.5925
1	0.05	30	100	0.0525	0.7395

Tabla 4.

De manera análoga, la siguiente tabla fue generada igual que la tabla 3, aplicando los pasos (i)-(vi) 10 veces, generando 10 listas \mathcal{L} y \mathcal{L}' , con $n = 200$ y $m = 10$, los resultados se pueden observar en la tabla 5.

ω	ρ	n	m	$\%p_j \leq \rho$	$\%p'_j \leq \rho$
0.1	0.05	200	10	0.06	0.06
0.2	0.05	200	10	0.05	0.18
0.3	0.05	200	10	0.03	0.32
0.4	0.05	200	10	0.03	0.73
0.5	0.05	200	10	0.04	0.92
0.6	0.05	200	10	0.05	0.98
0.7	0.05	200	10	0.07	1
0.8	0.05	200	10	0.03	1
0.9	0.05	200	10	0.08	1
1	0.05	200	10	0.04	1

Tabla 5.

Apéndice 1

A continuación veremos el código del programa Simula1 escrito en lenguaje de programación Mathematica 7. Enumeraremos las líneas del código para después dar una explicación del funcionamiento del programa.

```
1. Simula1[k1_, k2_, k3_] := Block[{i, j, l},
2.   A1 = A2 = A3 = A4 = A5 = A6 = A7 = A8 = T = TT = {};
3.   u1 = RandomReal[];
4.   u2 = RandomReal[];
5.   y1 = (-2 Log[u1])^(1/2)*Sin[2*Pi*u2];
6.   w = RandomReal[];
7.   y = RandomReal[];
8.   z = RandomReal[];
9.   Do[
10.    Do[
11.     AppendTo[A1, 10^w*y1]; w = RandomReal[]; u1 = RandomReal[];
12.     u2 = RandomReal[]; y1 = (-2 Log[u1])^(1/2)*Sin[2*Pi*u2], {k1}
13.    ]; AppendTo[A2, A1]; A1 = {}, {1, 1, k2}
14.   ];
15.   A3 = Abs[A2];
16.   Do[
17.    Do[
```

```

18.     AppendTo[A4,
19.         Catch[
20.             Do[
21.                 Do[
22.                     If[Characters[ToString[A3[[1, j]]]][[s]] == ToString[i],
23.                         Trow[i]
24.                     ], {i, 1, 9}
25.                 ], {s, 1, 100}
26.             ]
27.         ]
28.         ], {j, 1, k1}
29.     ]; AppendTo[A5, A4]; A4 = {}, {1, 1, k2}
30. ];
31. Do[
32.     Do[
33.         If[y < k3, AppendTo[A6, Which[
34.             A5[[dd, d]] == 1,2, A5[[dd, d]] == 2,3,
35.             A5[[dd, d]] == 3,4, A5[[dd, d]] == 4,5,
36.             A5[[dd, d]] == 5,6, A5[[dd, d]] == 6,7,
37.             A5[[dd, d]] == 7,8, A5[[dd, d]] == 8,9,
38.             A5[[dd, d]] == 9,1
39.         ]

```



```

40.           ], AppendTo[A6, A5[[dd, d]]]
41.       ]; y = RandomReal[], {d, 1, k1}
42.       ]; AppendTo[A7, A6]; A6 = {}, {dd, 1, k2}
43.       ];
44. ji2[chi_, n_] := 1 - (1/(2^(n/2)*Gamma[n/2]))*Integrate[t^((n/2)
45.           - 1)*Exp[-t/2], {t, 0, chi}];
46. b[i_, j_] := N[Count[A5[[i]], j]/k1];
47. bb[i_, j_] := N[Count[A7[[i]], j]/k1];
48. a[x_, y_] := Count[A5[[x]], y];
49. aa[x_, y_] := Count[A7[[x]], y];
50. ji22[i_] := N[k1*(Sum[((b[i, j]
51.           - Log10[1 + (1/j)])^2/Log10[1 + (1/j)]), {j, 1, 9}]]];
52. ji222[i_] := N[k1*(Sum[((bb[i, j]
53.           - Log10[1 + (1/j)])^2/Log10[1 + (1/j)]), {j, 1, 9}]]];
54. p[i_] := N[ji2[ji22[i], 8]];
55. pp[i_] := N[ji2[ji222[i], 8]];
56. Do[AppendTo[T, Which[p[i] >.05, 0, p[i] <= 0.05, 1]],
57.     {i, 1, k2}];
58. Do[AppendTo[TT, Which[pp[i] >.05, 0, pp[i]<= 0.05, 1]],
59.     {i, 1, k2}];
60. porc0 = N[Count[T, 1]/k2];
59. porc00 = N[Count[TT, 1]/k2];

```

```

60. Print["Tenemos una lista con ", k2," sublistas cada una con ",
61.     k1,"elementos, con un parametro w = ", k3," obtuvimos
62.     el numero de p y p' que son menor que el valor 0.05"];
63. Print["El porcentaje de p que son menor que 0.05 son ", porc0];
64. Print["El porcentaje de p' que son menor que 0.05 son ",
65.     porc00];
65. ]

```

Ahora explicaremos el funcionamiento del programa Simula1. La función Simula1, es una función que depende de tres variables $k1$, $k2$ y $k3$, donde $k1$ son el número de elementos que contendrá cada sublista, $k2$ es el número de sublistas que contendrá la gran lista \mathcal{L} y $k3$ es el parámetro con el que alteraremos los datos contenidos en la lista \mathcal{L} para obtener la lista alterada \mathcal{L}' .

Los pasos de las líneas 9-14 generan una lista \mathcal{L} con $k2$ sublistas, cada sublista con $k1$ elementos, tal que \mathcal{L} satisface Benford. Para ello, utilizamos la ayuda de una variable aleatoria que tiene la forma $10^U \times N$, donde U es una variable aleatoria distribuida uniformemente en $(0, 1)$ y N es una variable aleatoria con distribución Normal y media cero.

En la línea 15 sacamos el valor absoluto de los datos contenidos en \mathcal{L} . Después en las líneas 16-30, convertimos el número en cadenas de dígitos esto para poder identificar el primer dígito significativo de cada dato.

En las líneas 31-43 alteramos la lista \mathcal{L} por medio del parámetro $k3$ obteniendo \mathcal{L}' . En las líneas 44-45 definimos la función `ji2` con la que podremos calcular la

$\mathbf{P}\{\chi^2 > \chi_0^2\}$. En las líneas 46-47 definimos las funciones `b` y `bb`, con las que obtendremos el promedio de que el dígito i fue observado como primer dígito significativo en cada sublista de las listas \mathcal{L} y \mathcal{L}' , respectivamente. En las líneas 48-49 definimos las funciones `a` y `aa` con las que contamos el número de veces que fue observado el dígito i como primer dígito significativo en cada sublista de las listas \mathcal{L} y \mathcal{L}' , respectivamente.

En la línea 50 definimos la función `ji22` con la que obtendremos el valor χ^2 correspondiente a la lista \mathcal{L} con respecto a la distribución Benford, de manera similar en la línea 52 definimos `ji222` para calcular el valor χ^2 correspondiente a la lista \mathcal{L}' . Después en las líneas 54-55 definimos las funciones `p` y `pp` con las que calcularemos el p -valor de las sublistas contenidas en las listas \mathcal{L} y \mathcal{L}' , respectivamente. Después de las líneas 56-59 calcularemos el porcentaje de p_j y p'_j que fueron menor o igual que 0.05. Y por último en las líneas 60-65 presentamos los resultados obtenidos en la simulación.

Bibliografía

- [1] **Adhikari, A.K.; Sarkar, B.P.** *Distribution of most significant digit in certain functions whose arguments are random variables.* Sankhya Ser. B 30 1968 47-58.
- [2] **Ash, R.B.; Doléans-Dade, C..** *Probability and Measure Theory.* Academic Press, 2000.
- [3] **Balanzario, E.P.; Sánchez-Ortiz, J.** *Sufficient conditions for Benford's law.* Statist. Probab. Lett. 80 (2010), no. 23-24, 1713-1719.
- [4] **Barlow, J. L.; Bareiss, E. H.** *On roundoff error distributions in floating point and logarithmic arithmetic.* Computing 34 (1985), no. 4, 325-347.
- [5] **Benford, F.** *The law of anomalous numbers.* Proceedings of the American Philosophical Society 78 (1938), 288-291.
- [6] **Billingsley, P.** *Probability and Measure.* John Wiley & Sons, Inc., 1986.
- [7] **Cramér, H.** *Métodos Matemáticos de Estadística* Aguilar, 1963.
- [8] **Crowder, N.** *Fraud Detection Techniques.* Internal Auditor, 17-20 (1997).
- [9] **Diaconis, P.; Freedman, D.** *On rounding percentages.* J. Amer. Statist. Assoc. 74 (1979), no. 366, part 1, 359-364.
- [10] **Feller, W.** *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons, Inc., 1971.
- [11] **González, Heileman & Abdallah.** *A Generalization of Benford's law and its application to images.* Manuscrito.
www.gts.tsc.uvigo.es/gpsc/publications/fperez/ecc07.pdf

- [12] **Grimmett, G.R.; Stirzaker, D.R.** *Probability and Random Processes*. Oxford University Press, 2001.
- [13] **Hamadeh, N.** *Wireless Security and Traffic Modeling Using Benford's Law*. Tesis. B. Eng., American University of Beirut, 2002.
- [14] **Hamming, R. W.** *On the distribution of numbers*. Bell System Tech. J. 49 1970 1609-1625.
- [15] **Hill, T.P.** *A statistical derivation of the significant-digit law*. Statist. Sci. 10 (1995), no. 4, 354-363.
- [16] **Hill, T.P.** *The significant-digit phenomenon*. Amer. Math. Monthly 102 (1995), no. 4, 322-327.
- [17] **Hill, T.P.** *Base-invariance implies Benford's law*. Proc. Amer. Math. Soc. 123 (1995), no. 3, 887-895.
- [18] **Hoel, P.G.; Port, S.C.; Stone, C.J.** *Introduction to Probability Theory*. Houghton Mifflin Company, 1971.
- [19] **Jamain, A.** *Benford's Law*. Tesis. Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble, 2001.
- [20] **Knuth, D.E.** *The art of computer programming. Vol. 2: Seminumerical algorithms*. Addison-Wesley Publishing, Reading, MA, (1969) 219-229.
- [21] **Larsen, R.J.; Marx, M.L.** *An Introduction to Mathematical Statistics and Its Applications*. Pearson Education, Inc., 2006.
- [22] **Leemis, L.M.; Schmeiser, B.W.; Evans, D.L.** *Survival distributions satisfying Benford's law*. Amer. Statist. 54 (2000), no. 4, 236-241.

- [23] **de Marchi, S.; Hamilton, J.T.** *Assessing the Accuracy of Self-Reported Data: An Evaluation of the Toxics Release Inventory.* Journal of Risk and Uncertainty, Vol. 32, No. 1. (January 2006), pp. 57-76.
- [24] **Newcomb, S.** *Note on the frequency of use of the different digits in natural numbers.* American Journal of Mathematics 4 (1881), 52-67.
- [25] **Nigrini, M.** *The detection of income evasion through an analysis of digital distributions.* Ph. D. Thesis, Department of Accounting, University of Cincinnati (1992).
- [26] **Pinkham, R.S.** *On the distribution of first significant digits.* Ann. Math. Statist. 32 1961 1223-1230.
- [27] **Ross, S.M.** *Introduction to Probability Models.* Elsevier Inc. 2010.
- [28] **Roukema, B.** *Benford's Law Anomalies in the 2009 Iranian Presidential Election.* arXiv:0906.2789
- [29] **Schatte, P.** *On mantissa distributions in computing and Benford's law.* J. Inform. Process. Cybernet. 24 (1988), no. 9, 443-455.
- [30] **Stein, E.M.; Shakarchi, R.** *Fourier Analysis an Introduction.* Princeton University Press, 2003.
- [31] **Varian, H.R.** *Benford's law.* The American Statistician 26 (3), 65-66.
- [32] **Zetter, K.** *Crunching Iranian Election Numbers For Evidence of Fraud.*
<http://www.wired.com/threatlevel/2009/06/iran numbers>.