Análisis de los resultados electorales a partir de la Ley de Benford

R. Mansilla
CEIICH, UNAM

Resumen

En este trabajo se estudian ciertas propiedades estadísticas del número de votos obtenidos en cada una de las casillas por los dos principales candidatos a las recién concluidas elecciones presidenciales. Se obtienen para cada uno de ellos las distribuciones empíricas de la Ley de Benford, las cuales son comparadas con la distribución teórica, construyendo además los correspondientes intervalos de confianza.

Introducción.

El sistema de numeración de más frecuente uso en la actualidad es el sistema decimal. Si bien todas las computadoras actuales utilizan el sistema binario, la aplastante mayoría de las actividades humanas que requieren de mediciones cuantitativas se expresan en sistema decimal.

Habida cuenta de las disímiles magnitudes que en la vida cotidiana nos vemos obligados a cuantificar, parecería razonable que la frecuencia con que aparecen los diferentes dígitos decimales fuera uniforme en las mismas. Es sorprendente que esto no es así. Los dígitos aparecen en los números con diferentes frecuencias. Tal relación teórica es conocida como la Ley de Benford.

En 1881 el astrónomo y matemático inglés Simon Newcomb observó que las primeras páginas de las tablas de logaritmos que se encontraban en la biblioteca universitaria estaban más gastadas que las últimas. Como resulta obvio las primeras páginas están relacionadas con números cuyos primeros dígitos son pequeños (1, 2, etc.). No parecía evidente que los usuarios de estas tablas tuvieran preferencia por los números que comenzaban con cifras pequeñas y Newcomb concluyó correctamente que la frecuencia con que aparecían los diferentes dígitos era distinta. Ese mismo año publicó un trabajo sobre el tema[1].

Años mas tarde, este mismo resultado fue obtenido de manera independiente por el físico norteamericano Frank Benford[2]. A él debe esta regularidad su nombre. Benford recopiló más de 20,000 evidencias del cumplimiento de esta ley en bases de datos tan diferentes como áreas de ríos, pesos atómicos de los elementos químicos, números de las casa en una

---

[1] S. Newcomb, "Note of the frequency of use of the different digits in natural numbers", American Journal of Mathematics, **4**, pags. 39-40, 1881.

[2] F. Benford, "The law of anomalous numbers", Proceedings of the American Philosophical Society, **78** (4), pags. 551-572, 1938.

ciudad, etc. Más tarde P. Diaconis y D. Freedman[3] mostraron que en bases de datos donde existieran de manera consistente errores de redondeo la Ley de Benford se cumplía. Más aún, J. Boyle[4] recientemente ha demostrado que si los datos bajo estudio se construyen a partir de multiplicaciones y cocientes de variables aleatorias provenientes de diferentes fuentes, también satisfacen la antes mencionada ley.

Tal vez la aplicación más importante de la Ley de Benford sea en la detección de fraudes fiscales. La primera persona en sugerir tal aplicación fue el economista H. Varian[5], pero sin duda el trabajo más destacado en esta área es debido a M. J. Nigrini[6]. Muy recientemente, Cindy Durtschi *et al*.[7] desarrollaron un pormenorizado estudio de las aplicaciones de la Ley de Benford al proceso de auditorias. Además ofrecen en su trabajo un procedimiento estadístico mínimo para validar las hipótesis involucradas.

La ley de Benford.

Denotemos por $D = \{1, 2, \ldots, 9\}$ el conjunto de los dígitos decimales, excluyendo el cero. Sea $P(d)$ la probabilidad de que el dígito d aparezca en la primera posición de un número perteneciente a una base de datos. La ley de Benford afirma que:

$$P(d) = \frac{\log\left(1 + \frac{1}{d}\right)}{\log(10)}$$

---

[3] P. Diaconis, D. Freedman, "On rounding percentages", Journal of the American Statistical Association, **74**, pags. 359-364, 1979.

[4] J. Boyle, "An application of Fourier series to the most significant digit problem", American Mathematical Monthly, **101** (9), 1994.

[5] H. R. Varian, "Benford's Law", The American Statistician, **26**, pags. 65-66, 1972.

[6] M. J. Nigrini, "Taxpayer compliance application of Benford's law", Journal of American Taxation Association, **18** (1), pags. 72-92, 1996.
M. J. Nigrini; L. J. Mittenmaier, "The use of Benford law as an aid in analytical procedures", Auditing: A journal of practice and theory, **16** (2), pags. 52-67, 1997.
M. J. Nigrini, "Adding value with digital análisis", The Internal Auditor, **56** (1), pags. 21-23, 1999.

[7] C. Durtschi *et al*. "The effective use of Benford's Law to assit in detecting fraud in accounting data", Journal of Forensic Accounting, **5**, pags. 17-34, 2004.

Los valores obtenidos para cada dígito a partir de la relación anterior son:

d	$P(d)$
1	0.30103
2	0.17609
3	0.12494
4	0.09691
5	0.07918
6	0.06695
7	0.05799
8	0.05115
9	0.04576

En otras palabras, el dígito “1” debe aparecer en la primera posición de un conjunto de números aproximadamente el 30% de las veces, mientras que el dígito “9” solo lo aparece el 4.5% aproximadamente.

Resultados.

Utilizamos en nuestro análisis los resultados completos del PERP, casilla por casilla, los cuales pueden obtenerse en el siguiente sitio web:

http://em.fis.unam.mx/public/mochan/elecciones/fullprep.txt

De esta base de datos se tomaron los votos acumulados en una muestra de 65563 casillas por los candidatos Felipe Calderón y Andres M. Lopez Obrador. Con los mismos se calculó la distribución empírica de Benford para cada uno de los candidatos. En la Figura 1 aparecen estos resultados, así como los valores de la curva teórica de Benford. Nótese que la curva teórica ajusta perfectamente a la de AMLO para los dígitos $d = 8, 9$, mientras que para los dígitos pequeños existe muy poco acuerdo. Es conveniente señalar aquí que de la única manera en que el primer dígito de los votos obtenidos es 8 o 9 es en el caso de que la votación alcanzada este entre 70 y 90 puntos pues la cantidad máxima de boletas por casilla fue de 750.

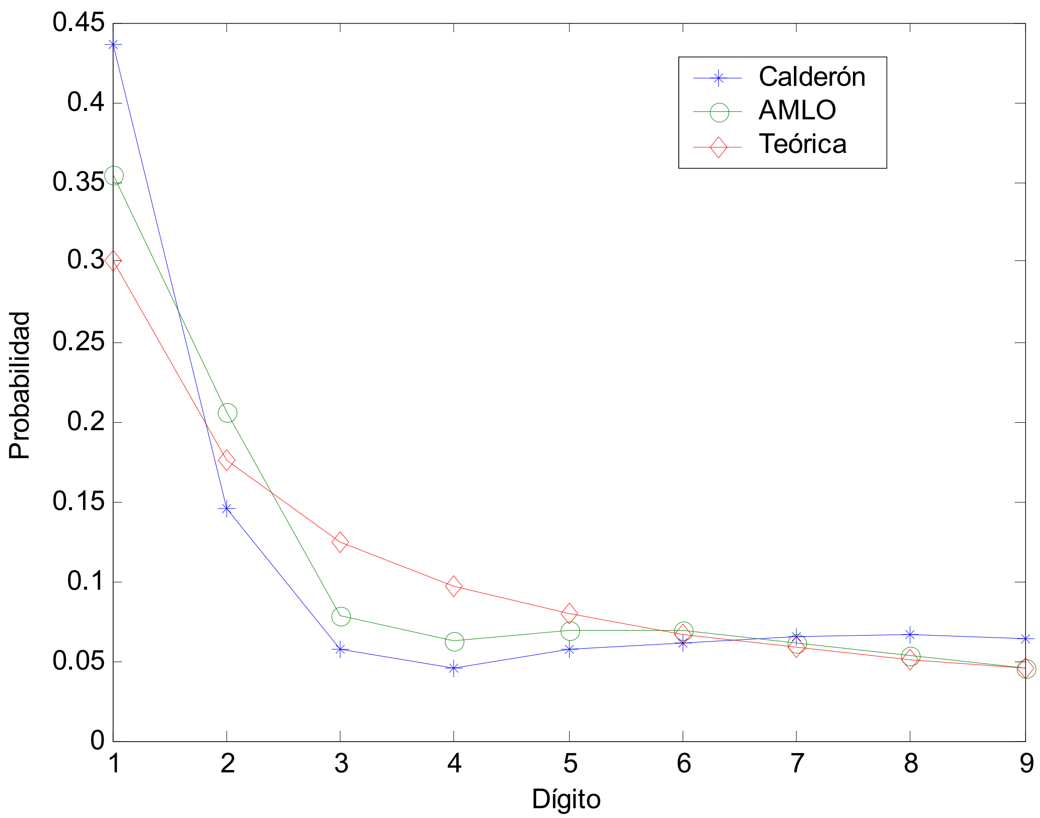

Fig. 1: Distribuciones empíricas y teóricas de la ley de Benford.

No obstante, una diferencia numérica no significa necesariamente un desacuerdo entre las magnitudes, pues los estimadores muestrales tienen fluctuaciones debidas a la finitud misma de las muestras. Para validar la igualdad es necesario hacer una prueba de hipótesis y calcular la zona de no rechazo de la misma. En la siguiente tabla aparecen las frecuencias de aparición de los dígitos para cada uno de los candidatos así como la zona de no rechazo para un 95% de confiabilidad de cada una de las frecuencias teóricas:

Dígito	Calderón	AMLO	Zona de no rechazo	
1	0.4362	0.3548	0.2975	0.3045
2	0.1450	0.2056	0.1732	0.1790
3	0.0574	0.0788	0.1224	0.1275
4	0.0459	0.0631	0.0946	0.0992
5	0.0572	0.0693	0.0771	0.0812
6	0.0616	0.0693	0.0650	0.0689
7	0.0654	0.0607	0.0562	0.0598
8	0.0667	0.0527	0.0495	0.0528
9	0.0645	0.0459	0.0442	0.0474

Nótese que las únicas frecuencias de los dígitos que caen en la zona de no rechazo son las frecuencias de AMLO para los dígitos $d = 8, 9$ como antes habíamos mencionado.

En la Figura 2 aparecen los resultados de la tabla anterior. Las pequeñas barras verticales en cada punto representan la zona de no rechazo. Nótese lo separados que están los valores reales de estos intervalos.

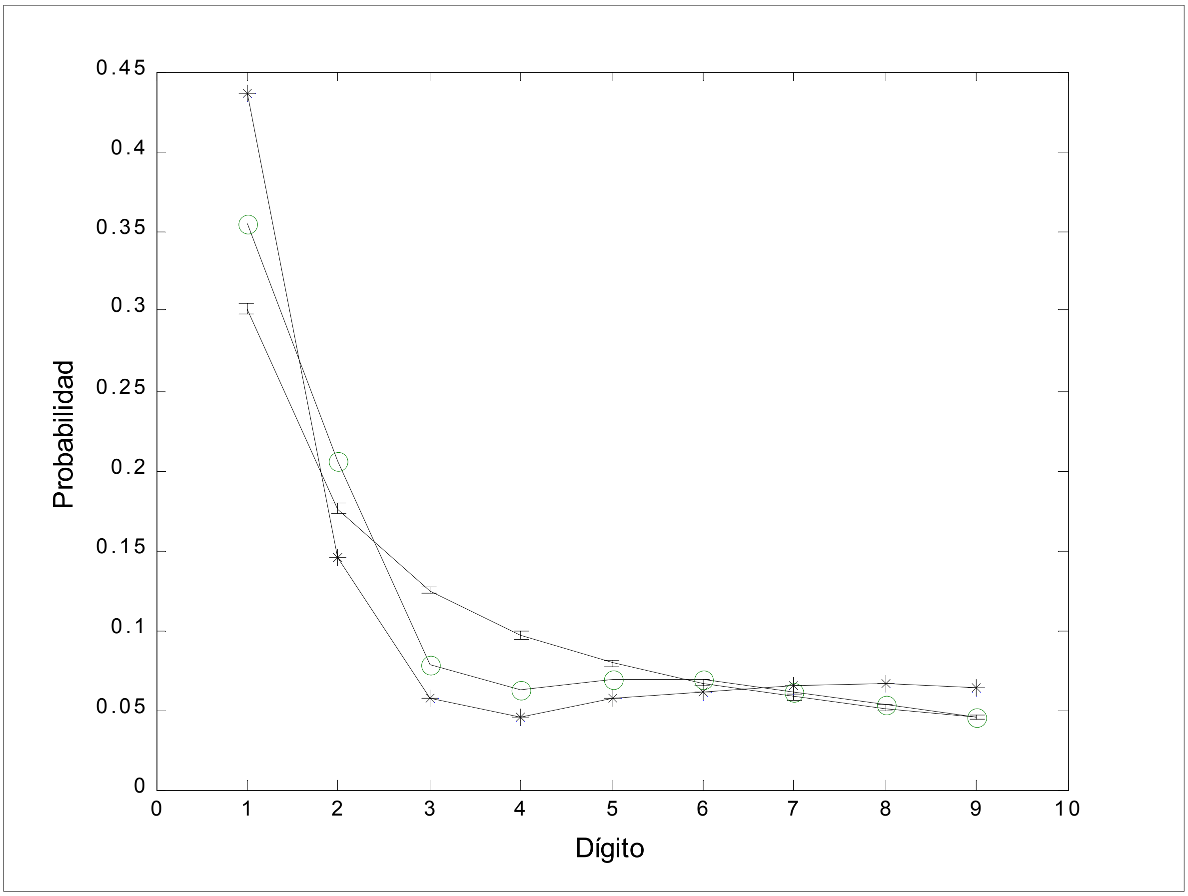

Fig. 2: Datos similares a los de la Fig. 1. En este caso se han agregado las zonas de no rechazo De la hipótesis nula para cada una de las proporciones teóricas.

Conclusiones.

Resulta muy difícil explicar el comportamiento de las distribuciones empíricas de los candidatos a la luz de los resultados teóricos antes expuestos. La ley de Benford es una regularidad bastante universal y toda divergencia de la misma debe ser observada con suspicacia.