

# Teoría estadística de errores aplicada al análisis de los resultados de la elección presidencial del año 2006 en México

Lúar Moreno Álvarez \*

**Resumen**—Los resultados de la elección presidencial mexicana celebrada en 2006 han sido debatidos ampliamente en la búsqueda por demostrar o refutar un posible fraude electoral. En este trabajo aplicamos la teoría estadística de errores de medición aleatorios a los errores aritméticos presentes en los cómputos de esa elección, y demostramos que dichos errores no se ajustan a la conducta estadística esperada de un conteo con equivocaciones cometidas al azar.

## I. INTRODUCCIÓN

Aunque existen avances importantes en la construcción de modelos de los mecanismos sociales generadores de la conducta estadística de los resultados en elecciones proporcionales, así como fuertes evidencias empíricas de su universalidad geográfica y temporal [1], hay relativamente poco consenso acerca de lo que ocurre en los procesos de selección de gobernantes en los países con votaciones absolutas. En el caso particular de las elecciones presidenciales de México en el año 2006, el estrecho margen porcentual de votación con que el candidato ganador fue determinado por el Tribunal Electoral de la Federación [2] ha dado origen a un extenso debate en la comunidad científica del país sobre la certidumbre y confiabilidad de los resultados electorales que se mantiene hasta el día de hoy. Ya sea que se mencione explícitamente o no en los trabajos que estudian el tema, el centro de este debate es la demostración de la existencia [3-5] o inexistencia [6-8] de un posible fraude electoral a partir del análisis estadístico del flujo de datos en el Programa de Resultados Electorales Preliminares (PREP) o de la información asentada en las actas de escrutinio y cómputo de la votación en cada casilla [9]. Sin embargo, la falta de un marco teórico general sobre la dinámica de este tipo de procesos electorales fomenta el uso, en ocasiones inadecuado, de las más diversas metodologías para demostrar las opiniones particulares de cada investigador, lo cual genera confusiones que impiden llegar a un acuerdo sobre la “legalidad” matemática de dicha elección. Por ese motivo, en este trabajo presentamos una propuesta de aproximación al problema de detección de resultados inverosímiles en los cómputos electorales de la elección presidencial mexicana del 2006, aplicando el análisis estadístico de los errores aleatorios de medición desarrollado por DeMoivre, Laplace y Gauss en el siglo XIX a los errores aritméticos encontrados en las actas de los resultados de la elección [10].

## II. ERRORES ARITMÉTICOS EN LOS RUBROS DE VOTACIÓN DE LAS ACTAS

De acuerdo con la legislación electoral vigente en el año 2006, las actas de escrutinio y cómputo de las casillas electorales contenían dos rubros de referencia: el número de boletas recibidas (BR) y el número de boletas sobrantes (BS), cuyo efecto conjunto puede incluirse en un rubro virtual que no aparece explícitamente en las actas, y al que llamaremos el rubro de las boletas entregadas ( $BE=BR-BS$ ); además de tres rubros fundamentales que dan cuenta de la votación total emitida en la casilla: el número de ciudadanos que votaron (CV), el número de boletas depositadas en la urna (BD), y la votación total contabilizada como la suma de los votos de todos los partidos (VT) [11].

En una casilla donde el conteo se efectúa sin errores debe cumplirse que  $BE=CV=BD=VT$ , de manera que si hay alguna diferencia en cualquiera de estos rubros, ésta constituirá un error aritmético cometido en esa casilla. Un análisis de la base de datos del PREP (que fue la única en la que aparecieron detallados estos rubros) indica que cerca de la mitad de las casillas presentaron errores aritméticos en estos rubros de sus actas de escrutinio y cómputo. Algunos autores señalan de manera equivocada que la magnitud del error absoluto promedio por casilla y la similitud del valor correspondiente en las casillas ganadas por uno u otro de los candidatos punteros es evidencia de que los errores están distribuidos de manera uniforme, y por tanto aleatoria, entre todas las casillas [6,8,12]. En realidad, la distribución probabilística que deben seguir los errores aleatorios está determinada por la teoría estadística del error, cuya fortaleza matemática y su universalidad son la base del tratamiento de las mediciones experimentales en todas las ciencias duras desde hace más de un siglo. En las secciones siguientes recordaremos las ecuaciones y conceptos elementales de esta teoría, y mostraremos cómo puede utilizarse para describir la aparición de errores al azar en el proceso de conteo de votos. Empleando esta teoría efectuaremos una simulación computacional que generará errores al azar en los rubros de cada una de las casillas de la base de datos analizada, pero manteniendo constante la votación para cada uno de los partidos políticos y, por tanto, el resultado final de la elección, a fin de comparar los resultados de dicha simulación con los de la elección real para verificar su aleatoriedad a través de las conductas estadísticas particulares que ésta exhiba.

\* Doctor en Física Aplicada por la Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla. Consultor Estadístico Independiente. Puebla, México. E-mail: luar.moreno@gmail.com

### III. MODELO ESTADÍSTICO DE ERRORES AL AZAR

Los errores aleatorios son intrínsecos a los aparatos de medición y ocurren invariablemente durante cualquier proceso de medida [10], ya sea que se trate de la determinación de una cierta cantidad física mediante un equipo de laboratorio sofisticado, o de la de las preferencias electorales de los ciudadanos asistentes a una cierta casilla a través del conteo de los votos emitidos para cada candidato.

Así, si la cantidad medida  $x$  tiene un valor real  $X$ , y cualquier error sistemático (o *bias*) se lleva a un nivel despreciable frente a cualquiera de las  $n$  fuentes de error aleatorio asociadas con la medida (que pueden incluir por ejemplo efectos de paralaje, tiempos de reacción, o conteos equivocados o un mal registro de los datos durante el proceso electoral), y que producen cada una de ellas un error de magnitud fija  $\varepsilon$ , de modo que el valor real  $X$  se encontrará dentro del intervalo  $[X-n\varepsilon, X+n\varepsilon]$ , entonces si hay  $v$  errores positivos con probabilidad de ocurrencia  $p$ , y  $n-v$  errores negativos con probabilidad  $1-p$ , entonces  $x = X + (2v - n)\varepsilon$ , y la probabilidad  $B(n,v)$  de obtener este resultado en particular está dada por la distribución binomial [10,13]:

$$B(n,v) = {}_n C_v p^v (1-p)^{n-v} \quad (1)$$

donde  ${}_n C_v$  es el *coeficiente binomial* de  $n$  en  $v$ . En el caso particular cuando  $p = 1-p = 1/2$ , tras ser escalada por el factor  $(1/2\varepsilon)$ , en el límite cuando  $\varepsilon \rightarrow 0$  y  $n \rightarrow \infty$  de modo que la desviación estandar  $\sigma_x = 2\varepsilon\sigma_v = \varepsilon n^{1/2}$  permanezca constante,  $B(n,v)$  tiende a la distribución normal con centro en  $X$  y anchura  $\sigma_x$ :

$$G(X, \sigma_x, x) = (2\pi \sigma_x^2)^{-1/2} \text{Exp}[-(x-X)^2/(2\sigma_x^2)] \quad (2)$$

Esto es, la distribución de los errores aleatorios en una medición siempre se aproximará a una curva gaussiana como la descrita por la ecuación (2), a menos que otros factores ajenos al mero azar estén involucrados en la aparición de estos errores. Si se efectúan los cambios de variable  $z = (x-X)/\sigma_x$  y  $\phi(0,1) = \sigma_x G(X, \sigma_x, x)$ , la ecuación (2) se transforma en la distribución normal canónica con centro en 0 y desviación estándar 1 [14].

En el caso de un conteo electoral se espera que la magnitud del error mínimo cometido sea  $\varepsilon = 1$  ya que cada voto puede ser contado de más o de menos con la misma probabilidad. Como cada voto es una fuente de error, entonces  $n = X$  representará a la votación total VT reportada para la casilla, mientras que  $x$  será el valor asentado en alguno de los rubros CV o BD, o el calculado para el rubro virtual BE. Puesto que no hay en la base de datos analizada ninguna casilla en la que la suma de los votos para cada partido sea distinta al valor VT reportado, la selección de VT como rubro de referencia resulta natural al no tener un error asociado en su determinación (lo que implica un cálculo computacional y no humano en la misma). El error  $e = 2v - VT$  para los rubros BE, CV o BD podrá tomar cualquier valor entre  $-VT$  y  $VT$ , lo cual considera los casos en que alguno de los rubros es dejado en blanco o reportado como 0 en el acta, o el caso

extremo en que cada uno de los votos es contado de más una vez (que podría ser cometido por ejemplo al efectuar el conteo con mucha prisa por un funcionario de casilla impaciente). De esta forma, la simulación computacional de los errores de cada casilla puede efectuarse generando 3 números al azar que sigan la distribución binomial (1) con los parámetros pertinentes.

En teoría, la distribución de los histogramas de errores para cada rubro de todas las casillas con la misma votación VT deberían seguir la distribución gaussiana (2) o su forma canónica después de efectuar las normalizaciones correspondientes. Sin embargo, el esmero con que muchos funcionarios de casilla realizaron la contabilidad de los votos consiguiendo que todos los rubros de sus actas de escrutinio y cómputo coincidieran constituye en sí mismo una fuente de alteración sistemática de la distribución que no se ve reflejada en un *bias* efectivo sobre la coordenada normalizada  $z$ , sino en la de la densidad de probabilidad canónica  $\phi(0,1)$ , como puede verse en la simulación de la figura 1 para los errores de las 455 casillas con  $VT = 350$ .

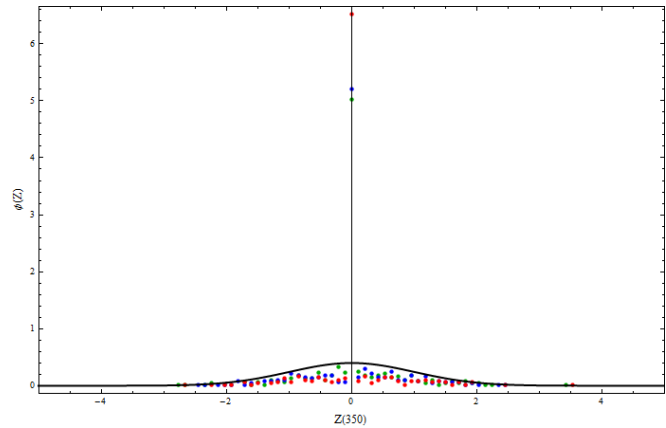


Figura 1. Histogramas de errores normalizados a las variables canónicas de la distribución normal de probabilidades. Los puntos por encima del valor máximo de 0.3989 de la distribución normal canónica graficada corresponden al *bias* generado por evitar sistemáticamente durante la simulación la aparición de errores al azar en las casillas bien contadas para alguno de los rubros en el acta. Verde – BE, Azul – CV, Rojo – BD.

Esta normalización a la distribución canónica nos permite comparar los resultados de todas las casillas en función de su desviación estándar normalizada, esperando como lo establece la estadística que el 99.8% de los errores se encuentre dentro del intervalo  $[-4,4]$ . De esta forma, los errores sistemáticos con un *bias* efectivo sobre la coordenada normalizada  $z$  se verán claramente reflejados al aparecer fuera de este intervalo de confianza.

### IV. RESULTADOS

#### A. Simulación de elección con errores aleatorios

Debido a que un número considerable de los rubros de las actas fue dejado en blanco, para descartar el sesgo del error debido a estas omisiones consideramos únicamente las

91,570 actas sin registros vacíos ni rubros con valores 0 o mayores a 760 (el número máximo de boletas permitidas por casilla). Ya que la determinación de los errores depende solamente del número de votos totales de la casilla, y a que suponemos que la aparición de dichos errores no depende de la localización geográfica -o geopolítica- de las mismas ni de su votación partidista particular (que mantenemos idéntica a la de la elección real), verificamos que la muestra de actas empleada fuera representativa de la distribución de VT de las 130,788 casillas contabilizadas en la elección, así como del número de personas en las listas nominales y en la votación por cada partido, encontrando desviaciones porcentuales menores a 0.06% en las casillas con menos de 600 votos, y sin desviaciones aparentes en la distribución para casillas de mayor votación. La figura 2 muestra la comparación de los errores por parejas de rubros, exhibiendo una conducta normal en dos dimensiones para los 3 grupos, mientras que la figura 3 presenta la relación entre los errores en BE, CV y BD y la votación total de la casilla VT, encontrándose como era de esperarse un mayor número de casillas bien contadas para las casillas de menor votación, y una mayor dispersión de los errores en la región de votación que concentra el mayor número de casillas, alrededor de VT = 350. La franja en el extremo derecho de la gráfica corresponde a las casillas especiales con VT cercana a 750.

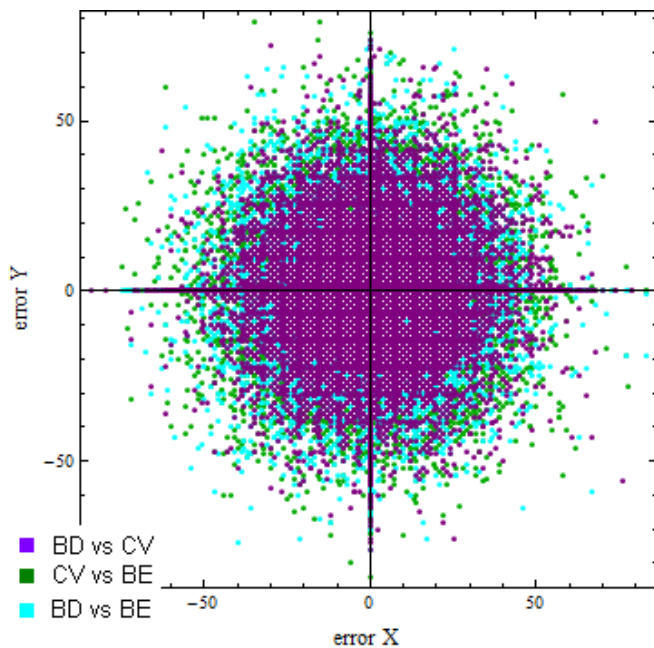


Figura 2. Errores simulados por parejas de rubros, exhibiendo una conducta normal centrada en cero en la región [-50,50] de cada eje.

En la figura 4 se presentan los errores de los distintos rubros para cada conjunto de casillas con la misma votación VT que se muestran en la figura 3, pero normalizados a sus desviaciones estándar canónicas. En ella puede notarse claramente la propiedad estadística de que para las votaciones totales consideradas es prácticamente imposible encontrar algún error aleatorio fuera de la banda de confianza de las 5 desviaciones estándar canónicas.

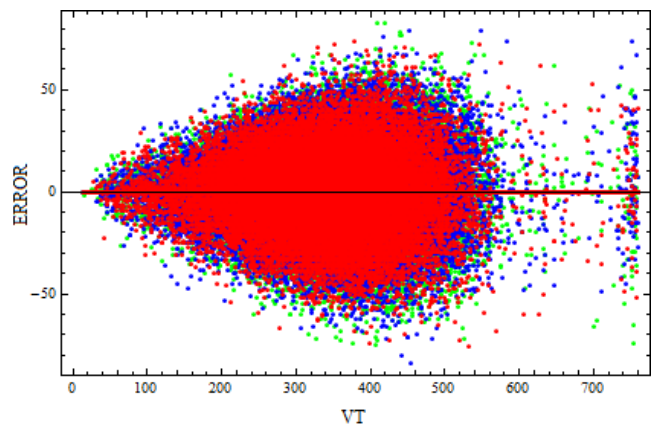


Figura 3. Errores en los rubros BE (verde), CV (azul) y BD (rojo) respecto de la votación total por casilla VT.

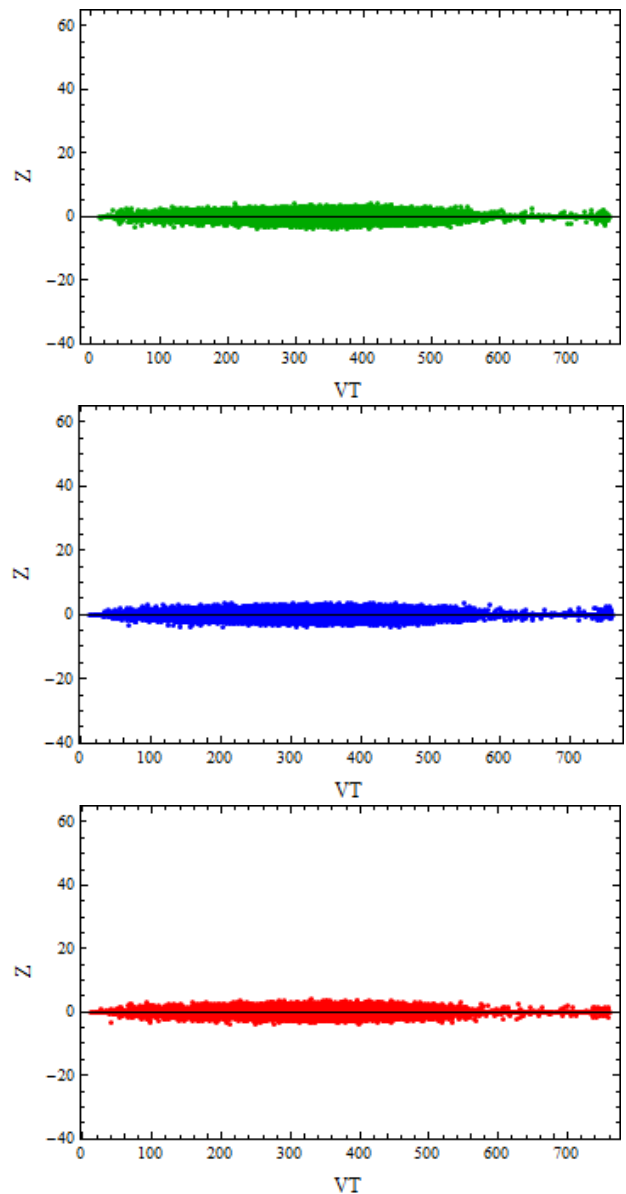


Figura 4. Errores en los rubros BE (arriba), CV (centro) y BD (abajo) normalizados a sus desviaciones estándar canónicas para cada conjunto de casillas con la misma VT.

Para determinar al ganador en una elección con errores al azar, es evidente que al aparecer sin ninguna predisposición entre las casillas con distintas VT, los conjuntos de casillas que comparten la misma magnitud del error en los distintos rubros constituyen muestras aleatorias del conjunto total de casillas, por lo que se espera que los histogramas (en realidad, polígonos de frecuencias) porcentuales de la votación para cada partido en función del número de errores reflejen aproximadamente los mismos porcentajes de votación de la elección. La figura 5 muestra estos histogramas para los errores aleatorios (arriba) y para sus expresiones normalizadas canónicamente (abajo) para el rubro CV. Ambas gráficas muestran como era de esperarse un comportamiento uniforme de los porcentajes de votación por partido respecto del resultado de la elección. La normalización canónica con cajas de tamaño 1 centradas en los valores enteros de Z tiene el efecto de suavizar aún más los histogramas.

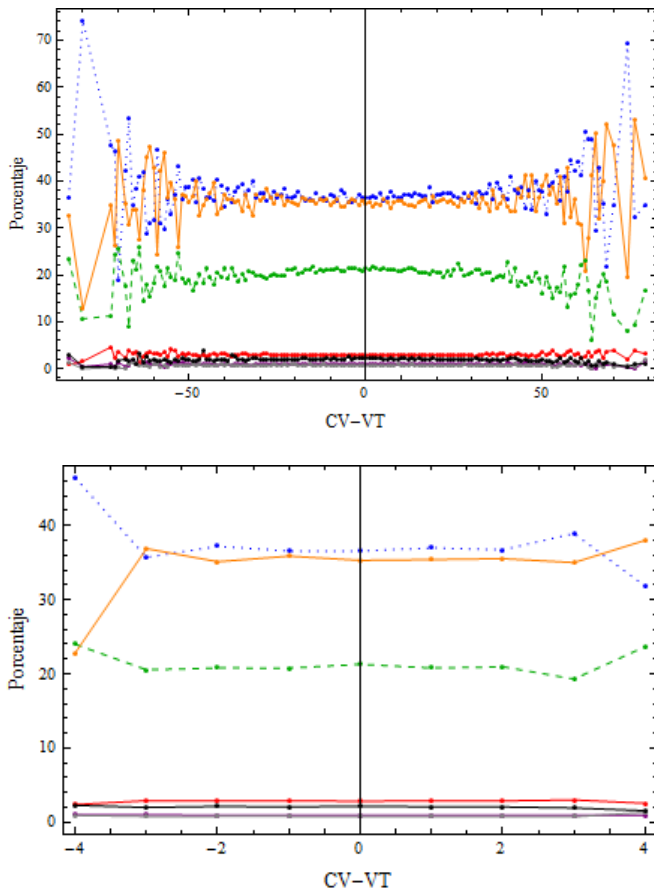


Figura 5. Histogramas (polígonos de frecuencias) porcentuales de la votación por partido respecto de los errores en los rubros CV (arriba) y su normalización canónica (abajo), mostrando claramente el ordenamiento uniforme siguiendo el resultado de la elección.

#### B. Análisis de los errores de la elección de 2006

Las figuras 6 a 9 muestran los mismos análisis realizados para la elección con errores aleatorios simulados presentados respectivamente en las figuras 2 a 5. Los gráficos en las figuras 6 y 7 muestran conductas muy distintas a las

esperadas para una elección con errores aleatorios. La figura 7 en particular exhibe una conducta extremadamente regular por fuera de la banda de confianza de  $\pm 50$  votos esperada de una conducta aleatoria.

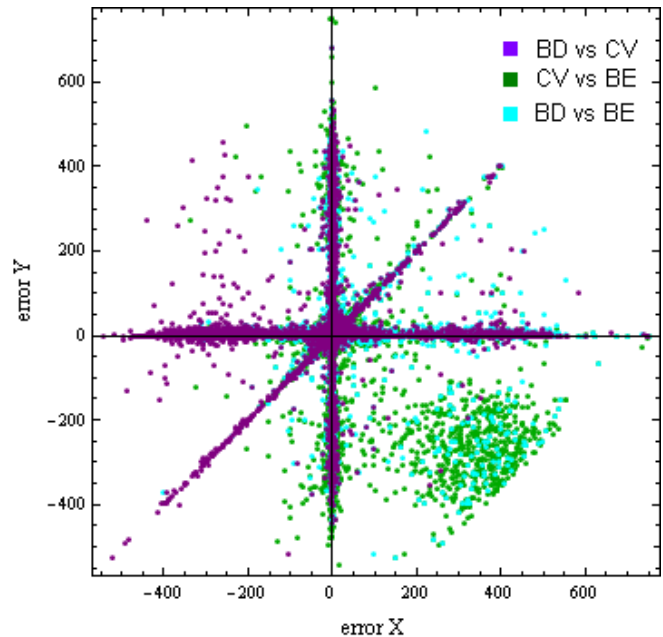


Figura 6. Errores por parejas de rubros de la elección de 2006, exhibiendo sólo para algunos puntos la conducta esperada centrada en cero en la región  $[-50,50]$  de cada eje.

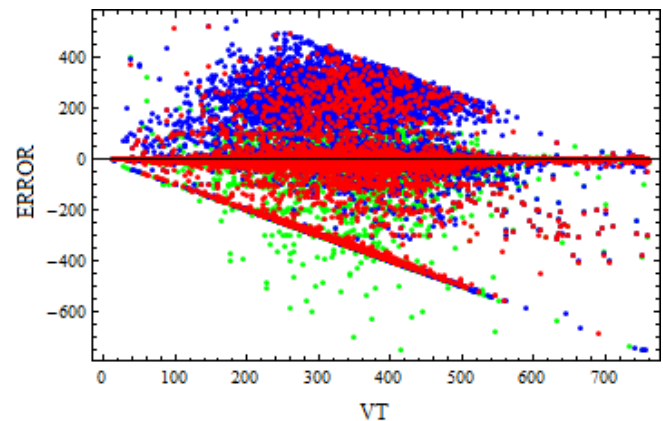


Figura 7. Errores en los rubros BE (verde), CV (azul) y BD (rojo) respecto de la votación total por casilla VT para la elección de 2006.

A diferencia de lo que se señala en varios trabajos [6,8,12] estas conductas no pueden considerarse como errores de omisión ya que en principio la muestra estadística analizada no contiene actas que presenten estos inconvenientes, además de que el modelo binomial de la teoría estadística de errores aleatorios considera este tipo de situaciones como parte de las posibilidades halladas en una casilla. La figura 8 respalda este argumento al probar que esos errores *sistemáticos* de alta regularidad se encuentran por fuera de la banda de confianza de  $\pm 5$  desviaciones estándar, donde es muy poco probable encontrar 3 puntos en 10,000,000.

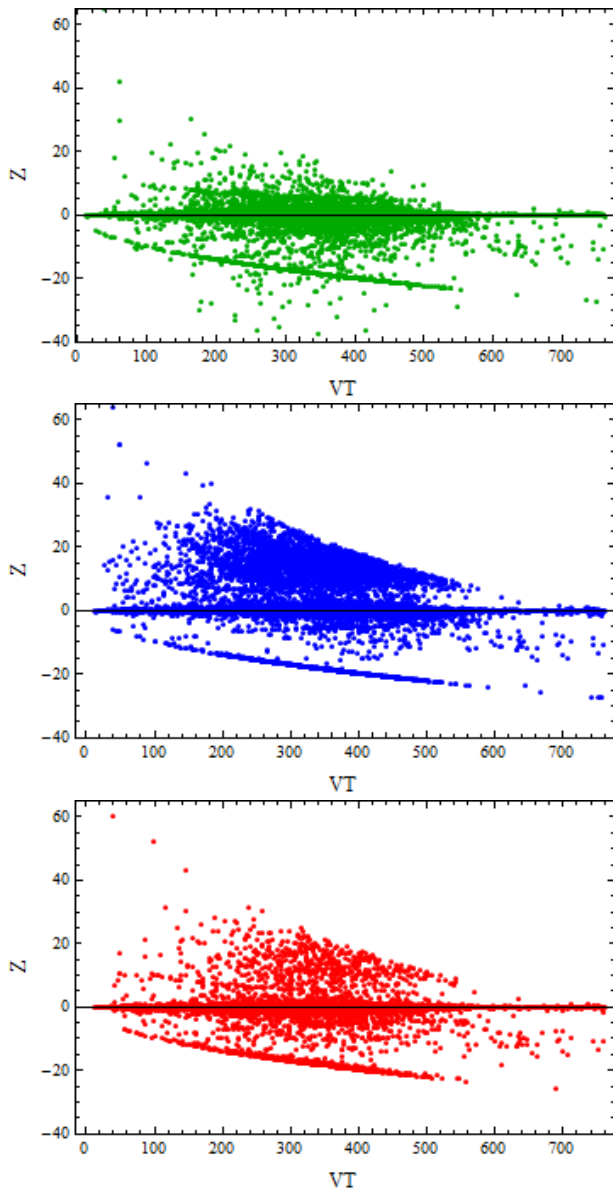


Figura 8. Errores en los rubros BE (arriba), CV (centro) y BD (abajo) normalizados a sus desviaciones estándar canónicas para cada conjunto de casillas con la misma VT en la elección de 2006.

Puesto que hay 3 rubros por cada una de las 91,570 actas en la muestra considerada, están graficados 274,710 puntos en cada imagen, por lo que resulta prácticamente imposible encontrar uno solo de ellos fuera de las 5 desviaciones estándar en una elección con errores aleatorios. Así mismo, las líneas regulares en función de la votación de la casilla que se muestran en las figuras 7 y 8 no son producto de ningún proceso aleatorio.

Los resultados para los polígonos de frecuencias porcentuales de la votación para cada partido en función de los errores en los rubros (arriba) y su normalización canónica (abajo), sobre todo en las bandas de confianza, revelan dos posibilidades igualmente preocupantes: al candidato en el segundo sitio fue al que más le perjudicaron los errores, o bien fue quien realmente ganó la elección.

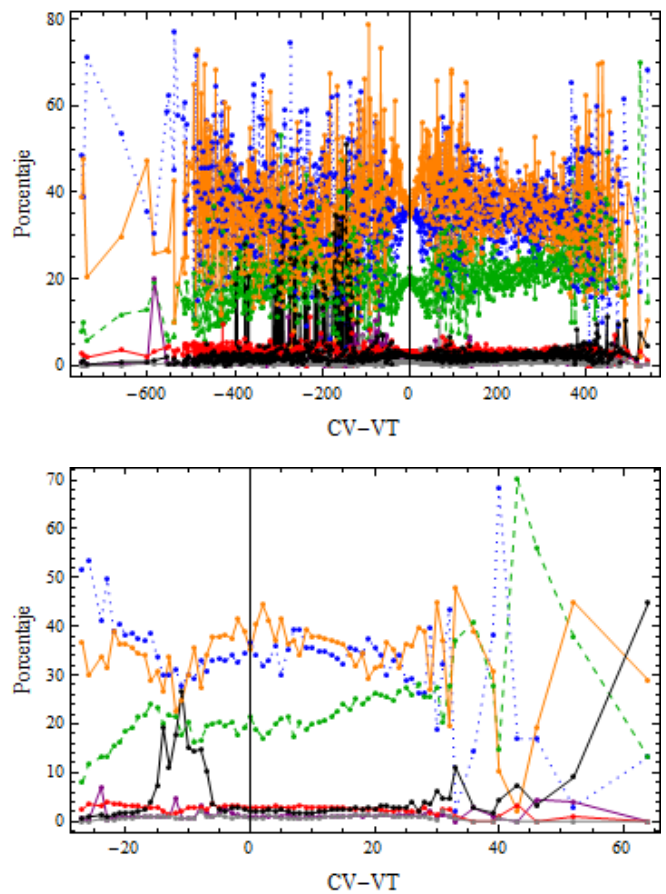


Figura 9. Histogramas (polígonos de frecuencias) porcentuales de la votación por partido respecto de los errores en los rubros CV (arriba) y su normalización canónica (abajo), mostrando un ordenamiento diferente al del resultado oficial de la elección.

Finalmente, el histograma del porcentaje de votación por partido respecto de la votación total de la casilla muestra otra inconsistencia: a pesar de la diferencia tan estrecha entre las votaciones de los candidatos punteros, el candidato en segundo lugar siempre lo estuvo en todos los grupos de casillas con el mismo número de votos totales VT, lo cual es extremadamente improbable.

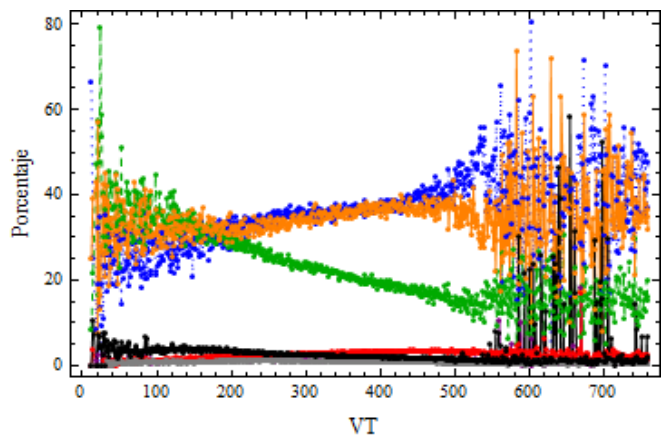


Figura 9. Histogramas (polígonos de frecuencias) porcentuales de la votación por partido en las casillas con la misma votación total VT, mostrando que el candidato en segundo lugar siempre se mantuvo en ese puesto respecto del rubro VT.

## V. CONCLUSIÓN

El análisis comparativo de la elección presidencial de México en el año 2006 con una simulación de errores aleatorios demuestra que los errores en la elección no fueron producto del azar, y que el partido afectado por dichos errores fue el designado oficialmente en el segundo puesto. Los resultados implican la aparición sistemática de errores para mantener a este candidato en el segundo lugar en función de la votación total aparente de la casilla, por lo que es necesario estudiar la naturaleza del proceso de aparición y el efecto específico de los mismos.

## AGRADECIMIENTOS

A Liz, Deb, Dany y Beto por su aliento para concluir este trabajo.

## REFERENCIAS

- [1] S. Fortunato, C. Castellano; Scaling and Universality in proportional elections. PRL 99, 138701, 1-4, (2007).
- [2] Dictamen relativo al Cómputo Final de la Elección de Presidente de los Estados Unidos Mexicanos, declaración de validez de la Elección y de Presidente Electo. Documento en Línea, revisado el 28/Junio/2012: [http://dof.gob.mx/nota\\_detalle.php?codigo=4930964&fecha=08/09/2006](http://dof.gob.mx/nota_detalle.php?codigo=4930964&fecha=08/09/2006)
- [3] J.A. López Gallardo, 2006 ¿Fraude Electoral?: Estudio de las Anomalías de la Elección Presidencial, Doble Hélice, México 2009.
- [4] Elecciones Presidenciales México 2006: ¿Anomalías en el PREP y el CD? Página Web y documentos internos de la misma. Revisada el 29/Junio/2012: <http://em.fis.unam.mx/~mochan/elecciones/>
- [5] Análisis Forense de Procesos Electorales. Página Web y documentos internos. Revisada el 29/Junio/2012: <http://www.fisica.unam.mx/octavio/>
- [6] J. Aparicio, Análisis estadístico de la elección presidencial de 2006 ¿fraude o errores aleatorios? Política y Gobierno, Vol. Temático 2009, 225-243 (2009).
- [7] G. Castañeda, I. Ibarra; Detección de fraude con modelos basados en agentes: las elecciones mexicanas de 2006. Perfiles Latinoamericanos 36, 43-69 (2010).
- [8] F. Pliego, El mito del fraude electoral, Pax, México 2007.
- [9] Base de Datos del Programa de Resultados Electorales Preliminares. Última revisión: Abril 2012. [http://www.ife.org.mx/documentos/proceso\\_2005-2006/prep2006/bd\\_prep2006/bd\\_prep2006.htm](http://www.ife.org.mx/documentos/proceso_2005-2006/prep2006/bd_prep2006/bd_prep2006.htm)
- [10] J.R. Taylor; An introduction to error analysis, University Science Books, USA 1997.
- [11] J.A. Crespo; 2006: hablan las actas: Las debilidades de la autoridad electoral mexicana, Debate-Random House Mondadori, México 2008.
- [12] Informe sobre errores aritméticos de las actas de escrutinio y cómputo encargado por el IFE a la empresa AC Nielsen, 2006. Última revisión: diciembre 2011. [http://www.ife.org.mx/docs/Internet/Docs\\_estaticos/Proceso2005\\_2006/informes\\_prep/errores\\_aritmeticos.pdf](http://www.ife.org.mx/docs/Internet/Docs_estaticos/Proceso2005_2006/informes_prep/errores_aritmeticos.pdf)
- [13] F. Reif; Fundamentals of Statistical and Thermal Physics, McGraw-Hill, USA 1965.
- [14] E. Kreyszig; Introductory Mathematical Statistics: Principles and Methods, John Wiley & Sons, USA 1970.