

Elecciones presidenciales en México, simulaciones computacionales y la ley de los grandes números

Luis Guillermo Cota Preciado
Departamento de Física
Facultad de Ciencias
UNAM

28th August 2006

Abstract

Las cifras oficiales de las recientes elecciones presidenciales en México han sido cuestionadas desde diversas ópticas. En este trabajo se presentan resultados de una simulación computacional que pretende esclarecer cómo *debió* haberse desarrollado la elección, y contrastarlos con las cifras oficiales.

1 Introducción

Desde los primeras aplicaciones de las computadoras en propósitos científicos, en las primeras décadas del siglo pasado, el uso de las computadoras como “laboratorios virtuales” no ha dejado de diversificarse y multiplicarse, de tal manera que hoy en día éstas son imprescindibles, por ejemplo, como laboratorios virtuales para desarrollo de nuevos fármacos, diseño aeroespacial, nuevos materiales, etc. Es imprescindible también el uso de la computación en el análisis de grandes cantidades de datos generados por procesos sociales, como el que aquí nos atañe: la elección presidencial del 2006 en México.

Una simulación no pretende y no puede reproducir todos los aspectos de un proceso real y por lo tanto, no es más que un modelo simplificado de éste. Sin embargo, como la simulación a menudo se alimenta con datos del proceso real, la simulación debe poder reproducir los principales resultados de éste. Sin embargo, la simulación a menudo arroja datos *finos* sobre el desarrollo del proceso y ayuda a comprenderlo mejor.

Dados los resultados oficiales del cómputo del Instituto Federal Electoral de México (IFE) sobre la elección presidencial del 2006, en forma del Programa de Resultados Electorales Preliminares (PREP) y, finalmente del Cómputo Distrital, se ha desatado una fuerte polémica entre quienes defienden estos resultados

y quienes, a la luz de argumentos estadísticos, los descalifican. Sin embargo, hasta donde este autor tiene noticia, las conclusiones a las que algunos han llegado han sido casi siempre sobre un análisis *post mortem* de los datos oficiales. Es el propósito de este trabajo *simular* la elección presidencial y ofrecer una visión de como *debió* haber transcurrido ésta dados los mismos resultados oficiales, y contrastar el proceso de votación simulado con el proceso real. El propósito de este estudio **no** es, por el contrario, reivindicar o defenestrar a ningún candidato o fuerza contendiente, y se declara expresamente que en la simulación no se ha introducido ningún artificio que sesgue intencionalmente los datos. El código empleado en la simulación está disponible para su eventual escrutinio y utilización.

2 Detalles técnicos

La generación de los datos se llevó a cabo mediante un programa en lenguaje *C*, escrito específicamente para este propósito. El análisis de datos se realizó mediante programas descartables hechos en lenguaje *AWK*.

Si bien no es rigurosamente necesario para aportar algunos resultados relevantes, a la simulación le fueron incorporados tantos *elementos de realidad* como fue posible:

- Mismo número total de votantes que en la votación real
- Mismo número de casillas
- Mismo número de votos por casilla
- Misma distribución porcentual final.

La *elección virtual* transcurrió, en contraste con la real, de acuerdo a las siguientes fases:

1. Votación. Cada voto fue generado aleatoriamente con respecto a las 6 posibilidades (voto nulo, voto por candidatos no registrados, voto por el PAN, voto por la Alianza por México, voto por la Alianza por el Bien de todos (en lo sucesivo, APBT), voto por Nueva Alianza, voto por Alternativa). Enseguida, una casilla se escogía al azar y el voto era “depositado” en ella siempre y cuando no se hubiera excedido el número real de votos reportados por cada casilla. El procedimiento se repitió hasta igualar el número real de votos reportado en cada casilla.
2. Cómputo distrital. Con todas las casillas con el número real de votos depositados desde el paso anterior, un valor aleatorio de tiempo fue generado para cada casilla, con el fin de simular la hora en la que los resultados de ésta fueron reportados (“subidos”) al IFE con el propósito de incorporar la variable *tiempo* a la simulación.

3. Cortes informativos cada 5 minutos. Con esto fue posible contabilizar el grado de avance de las elecciones.

El *corazón* de las diferentes etapas de la simulación es un generador de números aleatorios con distribución uniforme. Sin embargo, para poder simular razonablemente bien la distribución de los tiempos de recepción de resultados de cada casilla fue preciso implementar un generador de números aleatorios capaz de producir éstos en una distribución gamma, y parametrizarlo empíricamente. La elección de la función de distribución se hizo arbitrariamente, más allá de alguna consideración teórica, en base a la analogía que guarda la distribución observada de tiempos de respuesta con ésta.

Es pertinente mencionar que, si bien la elección virtual produjo como resultados los mismos porcentajes de votación (con variación de centésimas de punto porcentual) obtenidos por cada partido al final del Cómputo Distrital, éstos fueron introducidos como parámetros de la simulación. La justificación de esto reside en la presunción de que la intención de voto estaba plenamente definida *antes de la emisión del voto*. La simplificación que esto implica es que, a diferencia de las encuestas previas a la elección real, esta simulación considera que *el porcentaje de votantes indecisos fue 0*.

No está de más mencionar que hubo un *elemento de realidad* que se dejó intencionalmente fuera de la simulación: la distribución porcentual del voto *por casilla*. La incorporación de este criterio daría como resultado una copia de la elección, y no una elección *típica*, como se pretende simular. Una elección *típica* significa en este contexto una elección dominada por sucesos aleatorios *estadísticamente independientes* (el voto del elector X en la casilla A es esencialmente independiente del voto del elector Y en la casilla B . La casilla A hace su conteo y reporta sus resultados independientemente del proceso de conteo de la casilla B , etcétera).

Los resultados de la simulación fueron analizados bajo enfoques muy conservadores (es decir, se sospecha de los resultados de la simulación hasta que no existen dudas razonables sobre su validez), de tal manera que los aspectos en los que ésta es incapaz de describir satisfactoriamente un proceso real fueron descartados. Ejemplos de estos aspectos son la incapacidad de la simulación de reproducir satisfactoriamente preferencias electorales, y la incapacidad de reproducir un comportamiento acorde con la ley de Benford. Estas limitaciones –quizá temporales– de nuestro modelo no invalidan, sin embargo, los resultados que a continuación se presentan.

3 Resultados

3.1 Evolución temporal

3.1.1 Distribución de tiempos de recepción

Aquí se presentan los resultados de los cortes informativos simulados. La función de distribución de tiempos de recepción (distribución gamma) en la simulación

fue parametrizada por prueba y error hasta que las distribuciones producidas se aproximaron razonablemente a la forma de la distribución real de tiempos de recepción.

La figura 1 contrasta la distribución de tiempos de recepción de los resultados de cada casilla por el IFE con la distribución simulada de tiempos. La distribución simulada pretende ilustrar como *debió ser* lo que en este contexto denominamos una elección típica. La distribución simulada es siempre una curva suave y unimodal, formada por sucesos individuales, originados por la misma dinámica y *estadísticamente independientes*. La distribución real, por el contrario, presenta “zonas problemáticas” alrededor de los 37000, 48000 y 58000 segundos. Si bien se puede argüir que la distribución gamma pudiera no ser un buen modelo para representar la distribución real, la distribución real, por otro lado, parece no ser unimodal, es decir, parece ser la conjunción de por lo menos *dos* distribuciones unimodales, la primera con un pico alrededor de los 37000 segundos y la segunda con un pico alrededor de los 46000 segundos. El problema formal de esta interpretación reside en el hecho de que los tiempos de recepción debieron constituir, en su conjunto, *una* sola distribución, correspondiente a *una* sola dinámica, pues el proceso de cómputo, transmisión y difusión de resultados al IFE debió ser el mismo para cada casilla. Es posible observar, también, que, cualquier retraso artificial de la transmisión de los resultados en un cierto número de casillas en alguna zona o circunscripción hubiera significado otro “accidente” visiblemente manifiesto en la parte derecha de la curva. Lo que se observa, sin embargo, es lo contrario: el accidente visible en extremo izquierdo de la curva real acusa una interferencia artificial del proceso aleatorio durante la primera fase de la cadena *transmisión-recepción-distribución* de los resultados.

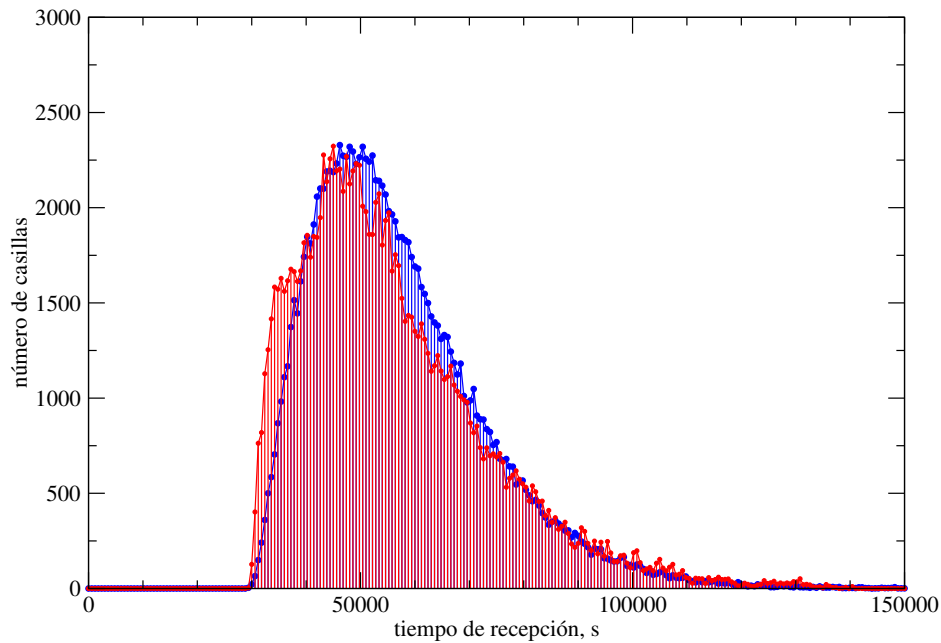


Figure 1: Histograma de tiempos de recepción de los datos de las casillas, en segundos, a partir del inicio del día 5 de julio de 2007. La distribución real de tiempos de recepción aparece en rojo. La distribución simulada aparece en azul.

3.1.2 Evolución temporal de la votación

Una vez obtenida una distribución razonable de tiempos de recepción es posible simular un proceso de votación típico, dominado por sucesos estadísticamente independientes, pero sujeto a las restricciones mencionadas en la sección 2. La figura 2 contrasta el proceso de votación real y el simulado. En principio es posible observar una buena concordancia global, lo cual indica que el modelo es razonablemente acertado. No hay anomalías aparentes en las curvas inferiores. Sin embargo, es notoria la intersección de las curvas superiores, correspondientes al PAN y a la APBT. Como esta gráfica representa los votos acumulados, una vez que se establece una tendencia -aquí, por ejemplo, por arriba de los 50,000 segundos- *cualquier* muestreo de las alturas relativas de las curvas debería contener, dentro de un pequeño margen de error, la distribución final del voto. La gráfica del proceso real representa una violación a la ley de los grandes números, lo cual no sólo es improbable, sino, en términos humanamente medibles, imposible.

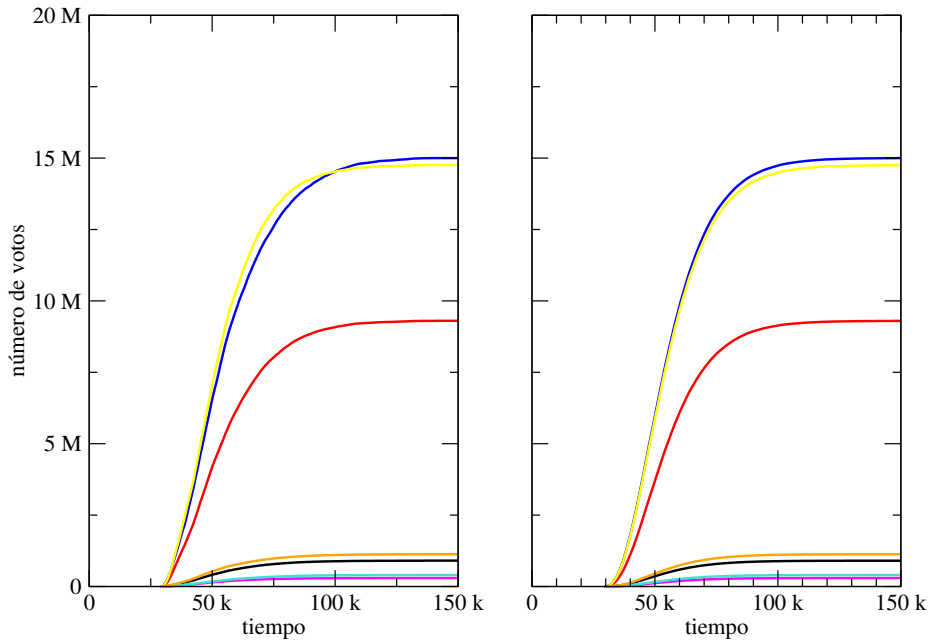


Figure 2: Evolución temporal del voto. El tiempo se expresa en segundos desde el inicio del día 5 de agosto de 2006. Las abreviaturas “k” y “M” representan aquí miles y millones, respectivamente. La gráfica de la izquierda se refiere a la votación real y la de la derecha a la votación simulada. La pautas de colores, vigente en lo sucesivo, es la siguiente: azul, PAN; amarillo, APBT; rojo, Alianza por México; anaranjado, Alternativa; negro, votos nulos; turquesa, Nueva Alianza; magenta, votos nulos.

Cuando se calcula la derivada con respecto al tiempo de la figura 2 se obtiene la figura 3, que representa las velocidades de recepción, en votos por segundo, para cada fuerza contendiente. Comparando las dos gráficas, resulta evidente la manera errática en la que se fue recibiendo (o transmitiendo, o difundiendo) la información del Cómputo Distrital. Esto podría ser indicativo algún factor de interferencia que afectó los tiempos de recepción (o de transmisión o de difusión) de múltiples casillas simultáneamente, de manera que resultó distorsionado el proceso de acopio de votos en conjunto. La suma instantánea de las contribuciones de sucesos aleatorios individuales (el reporte de cada uno de los resultados de cada casilla), *estadísticamente independientes* se presenta a la derecha, y representa cómo debió haber sido este proceso. No se pueden descartar como causas de estas fluctuaciones repetidos errores humanos operando *simultáneamente* a escala de decenas o cientos de casillas, o repetidos fallos en los sistemas de acopio (o de transmisión, o de difusión) de la información, aunque, dadas las circunstancias, se favorece la interpretación del factor de interferencia.

Por otro lado, la elección simulada presenta la velocidad de recepción de votos para el PAN como una envolvente que prácticamente *siempre* se sitúa

ligeramente arriba (con pequeñas fluctuaciones) de la envolvente de la APBT. Las curvas del proceso real muestran también el cruce entre las curvas del PAN y de la APBT. La interpretación del cruce en estas gráficas indica que la velocidad de recepción (o de transmisión, o de difusión) de votos del PAN *sufrió una aceleración* en la segunda mitad de la jornada del Cómputo Distrital. Se debe insistir en que, dada la ley de los grandes números, esta aceleración es *imposible* en una dinámica de sucesos estadísticamente independientes.

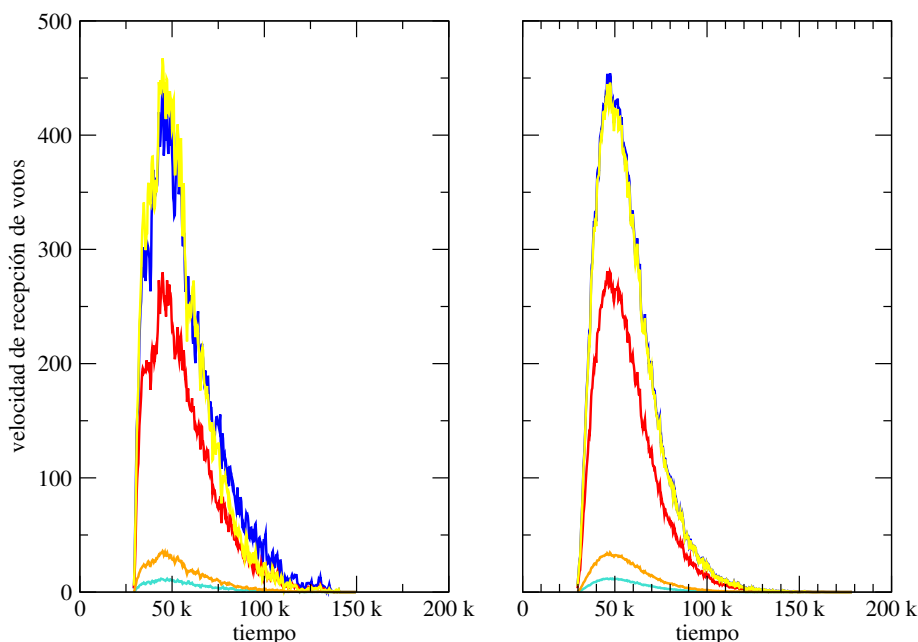


Figure 3: Velocidades de recepción de votos por el IFE durante el Cómputo Distrital. A la izquierda se representan las velocidades de la votación real y a la derecha las de la votación simulada.

4 Avance del Cómputo Distrital

4.1 Números absolutos

Mostramos aquí la acumulación de votos para cada fuerza contendiente en términos del número de casillas reportadas, figura 4. Es posible observar que, a esta escala, la acumulación de votos sigue una línea recta en todos los casos de la simulación, a la derecha, y en todos los casos del proceso real, *excepto* para el PAN y la APBT. La línea recta es la trayectoria esperada, pues habla de que, *en promedio*, cada fuerza electoral acumula una fracción aproximadamente constante de votos en cada casilla que se reporta. (Recuérdese que al promediar los resultados parciales de las regiones geográficas con distintas preferencias elec-

torales la distribución del voto tiende a uniformizarse rápidamente). Persiste, pues, bajo esta óptica, el “accidente” del cruce entre las trayectorias del PAN y de la APBT.

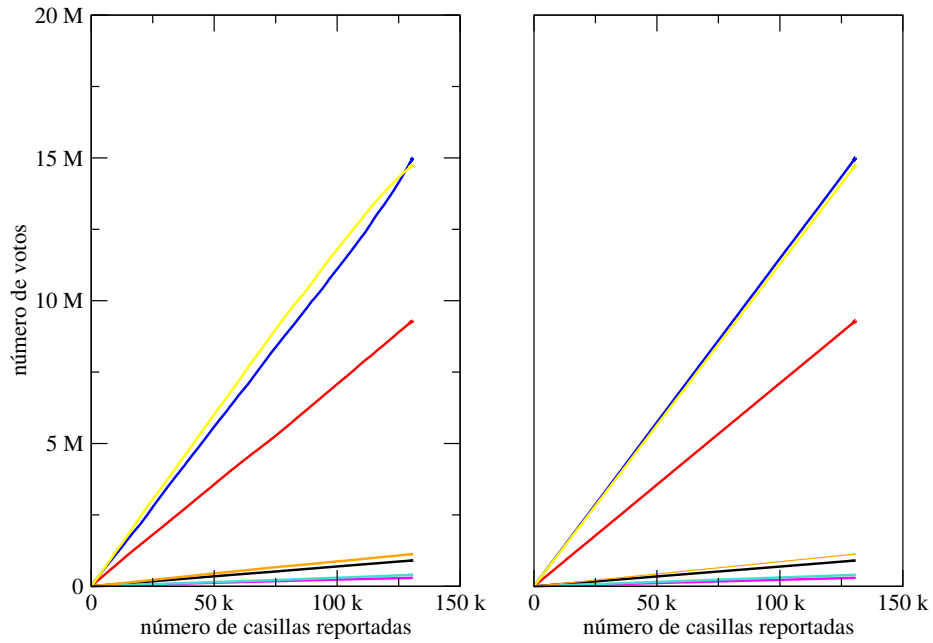


Figure 4: Acumulación de votos en términos del número de casillas reportadas durante los avances del Cómputo Distrital.

La derivada o tasa de cambio del voto con respecto a cada casilla nos proporciona una visión más puntual del proceso, como lo muestra la figura 5.

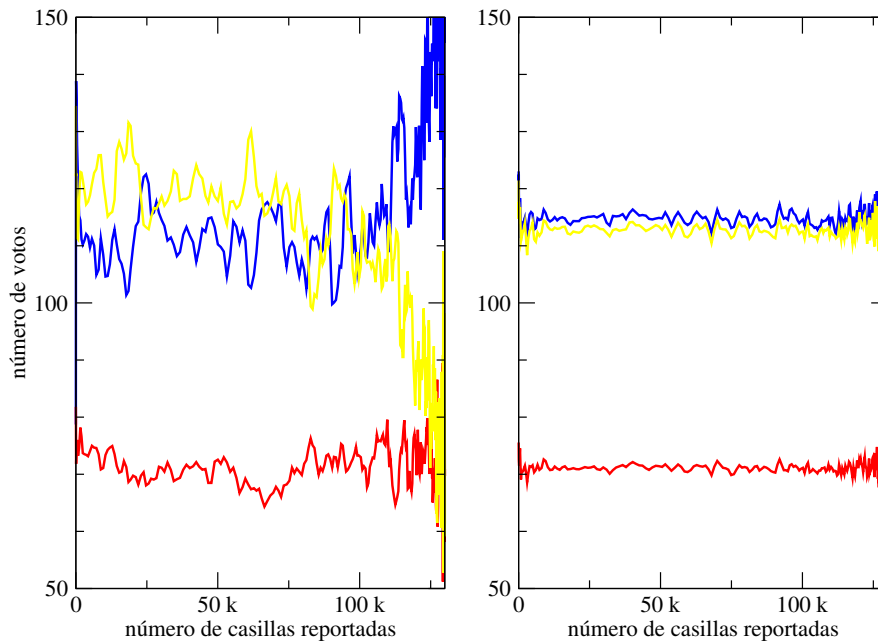


Figure 5: Comparación entre las tasas de cambio en el número de votos para cada fuerza contendiente (real, izquierda, y simulada, derecha) con respecto al avance de la votación en número de casillas reportadas. Sólo se muestran las curvas para las tres principales fuerzas contendientes.

La simulación muestra, como anteriormente, menores oscilaciones con respecto al proceso real. La explicación de esto podría deberse o bien a limitaciones de la propia simulación. Lo notorio de esta figura es otra vez el *cruce* entre las curvas de los dos principales contendientes. Sin embargo, la relevancia de ésta tiene que ver con que ofrece un indicio sobre la *naturaleza* de este cruce: dado que, a diferencia de la figura 3, la variable *tiempo* no aparece explícitamente, la tasa de cambio se refiere necesariamente al *número de votos por casilla*. La interpretación obligada es que la votación por casilla se mantuvo oscilante alrededor de una media para cada fuerza contendiente, aunque, a partir de un punto la votación media para el PAN se incrementó y la votación media para la APBT decreció. En otras palabras, *el número promedio de votos por casilla* se incrementó para el PAN y disminuyó para la APBT. Es importante insistir en que estas gráficas se refieren a números absolutos y no a fracciones o porcentajes sobre un total, por lo que el aumento en una de las cantidades *no necesariamente* tiene por qué producir una disminución en las otras. Un escenario en donde no se cumpliría esta interdependencia sería, por ejemplo, una casilla donde se hubiera dado la práctica del *carrusel* o el *taqueo* de votos. El que este crecimiento de la fracción de votos para el PAN se haya dado *en forma sistemática* hacia el final habla de una manipulación centralizada de la información.

El hecho, por otro lado, es que el aumento en el número de votos por casilla del PAN parece reflejar especularmente la pérdida de votos por la APBT. El conflicto con la ley de los grandes números, mencionado repetidamente antes, impide buscar en la estadística una explicación para este comportamiento. Y, debido a que las tendencias exhibidas son sistemáticas, se infiere que algún factor, posiblemente informático, de naturaleza central, interfirió en el proceso de transmisión-recepción-difusión de los resultados en la parte final del Cómputo Distrital, restando votos de la APBT y sumando votos al PAN.

Por otro lado, esta interpretación parece no sostenerse ante la información, ampliamente difundida el día del conteo, en el sentido de que los últimos estados en reportar sus resultados fueron mayoritariamente estados con preferencias panistas, dadas las reiteradas peticiones de recuento hechas por los representantes distritales de la APBT en esos estados. Si este fuera el caso, en efecto, la mecánica de sucesos aleatorios, estadísticamente independientes, se habría roto. Sin embargo, esto debería ser evidente en el histograma de tiempos de recepción, como se mencionó en la sección 3.1.1, y *además*, en la figura 5: si los últimos estados en reportar sus resultados fueron estados con preferencias panistas, los primeros debieron ser estados con preferencias por la APBT, y la figura debería acusar en el extremo izquierdo una tendencia opuesta: el número de votos por casilla obtenidos por la APBT debería disminuir sostenidamente de izquierda a derecha, mientras que el número de votos por casilla del PAN debería aumentar. Como esto no fue lo que ocurrió, se sostiene a nuestro juicio el argumento del factor informático de interferencia.

4.2 Grado porcentual de avance

En esta sección se muestra el avance de la votación como porcentaje de las preferencias electorales para cada fuerza contendiente en términos del porcentaje reportado del total de casillas. Esta es precisamente la forma condensada en que fue difundida la información por el PREP y durante el Cómputo Distrital, y la observación del comportamiento anómalo de esta gráfica a medida que se fue conformando fue lo que motivó, en principio, el análisis exhaustivo, *post mortem*, del conjunto total de los datos de la elección. El desglose de esta información en casillas sólo estuvo disponible hasta después de estos conteos, y es la que hemos utilizado en nuestros análisis.

Con la figura 6 se pretende mostrar, con detalle creciente, el desarrollo del Cómputo Distrital (real y simulado) en términos porcentuales. La primera fila muestra el proceso completo para todos los contendientes. La fila de enmedio muestra un enfoque sobre las tres principales fuerzas contendientes y la fila de abajo representa un enfoque aún más cerrado sobre el primer 10% de avance del proceso de conteo. A diferencia de la figura 5, aquí se muestran *porcentajes*, que son cantidades relativas de una fracción con respecto al total (que es la suma de las fracciones). Con esto se quiere decir que, cuando se habla de porcentajes, el crecimiento de una de las fracciones se da a expensas de las otras. Sin embargo, en el contexto de una dinámica de sucesos estadísticamente independientes, como son los reportes de los resultados de cada casilla, esto resulta

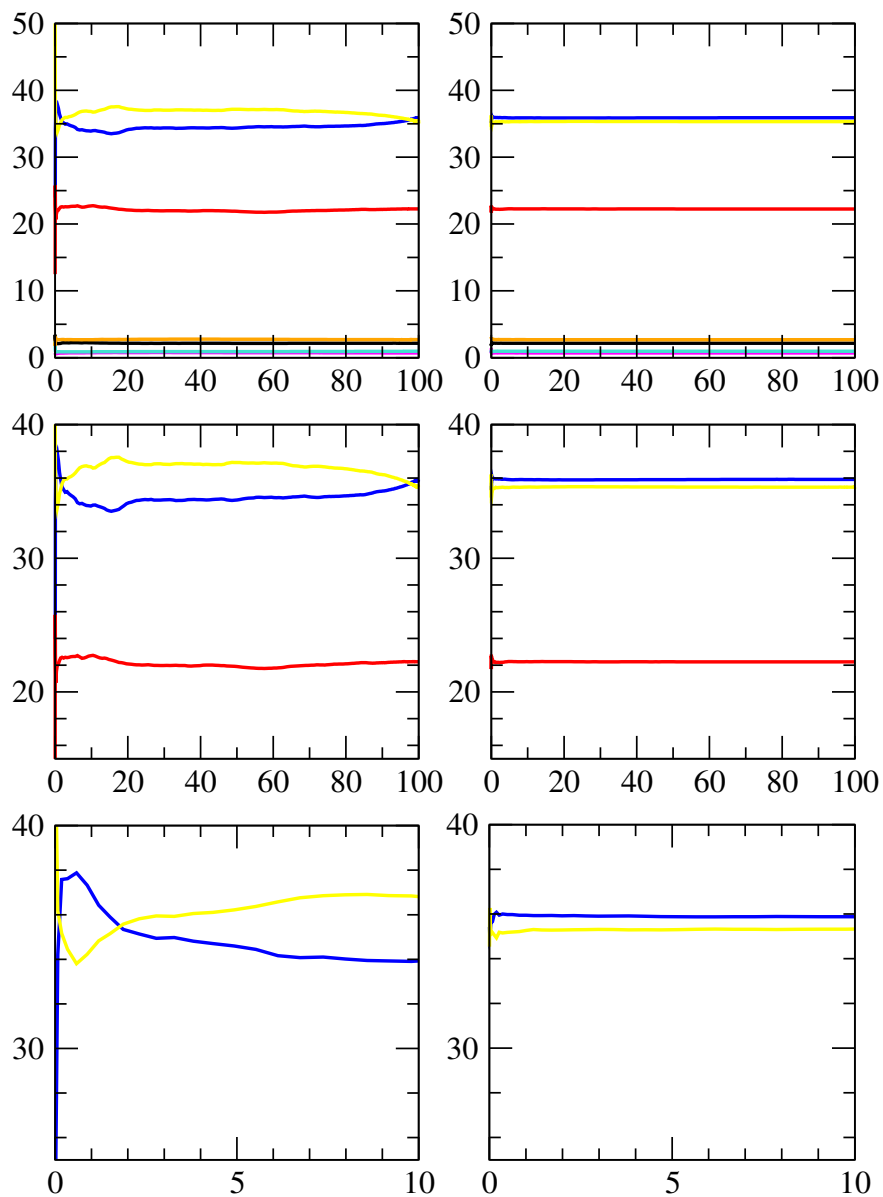


Figure 6: Porcentaje de votos *vs.* porcentaje de casillas computadas. El eje vertical representa el porcentaje de votación obtenido por cada fuerza contendiente y el eje horizontal representa el porcentaje de casillas computadas con respecto al máximo. Nótese el cambio de escala de arriba hacia abajo. A la izquierda se representan los resultados reales y a la derecha los resultados de la simulación.

imposible, como se ha sustanciado antes. Lo que se quiere enfatizar aquí es *cómo* se hubiera desarrollado un proceso de esta naturaleza (conteo simulado, gráficas del lado derecho). Aparentemente sin ninguna característica sobresaliente, estas gráficas representan, con mucha claridad, la manifestación de la ley de los grandes números: en un universo grande (de casillas), una muestra pequeña (como la obtenida con el primer 10% de la votación) ya representa, con un pequeño margen de error, la votación final. Las oscilaciones son esencialmente imperceptibles aún si se cambia de escala. Prácticamente todo el proceso se representa por líneas rectas *horizontales* (es decir, de pendiente cero). Se insiste, este *debió* ser el comportamiento para un conteo sin factores de interferencia.

Para descartar la posibilidad de que los datos de la simulación presentada en la figura 6 hubieran resultado casualmente demasiado “bien comportados”, se realizaron numerosas simulaciones adicionales, dos de las cuales se presentan en la figura 7.

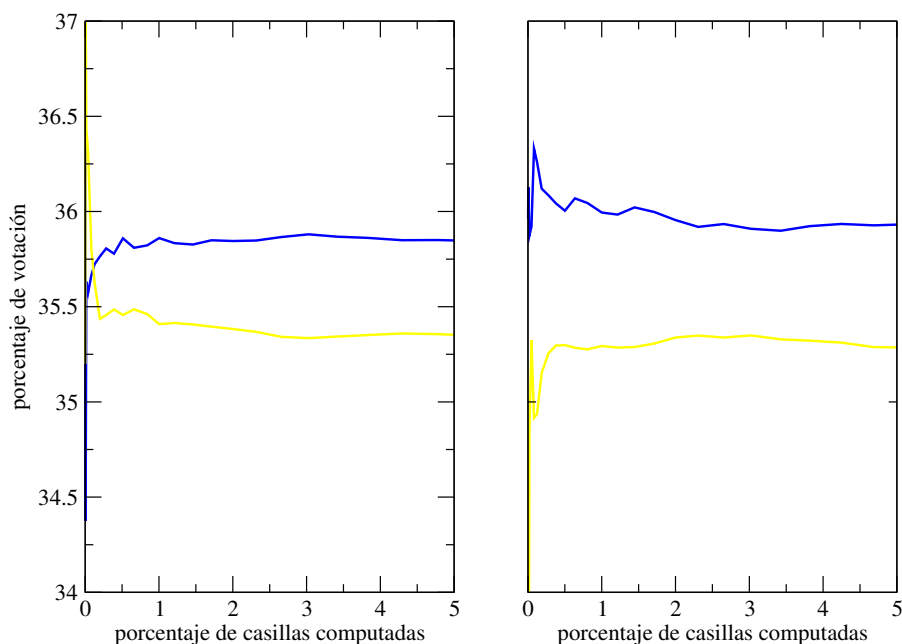


Figure 7: Primer 5% de avance de la votación para dos elecciones virtuales distintas. Nótese que en el caso de la izquierda existe un cruce, y en el caso de la derecha no. Ambas, sin embargo, coinciden en los resultados finales, con un margen de error de centésimas de punto porcentual.

Los resultados obtenidos en *todos* los casos arrojan escenarios de evolución del cómputo que sólo difieren en los estadios más tempranos, generalmente durante el primer 10% de avance del Cómputo Distrital. Los casos que se ilustran son, a la izquierda, un conteo que presenta un cruce, como aconteció en el conteo real (aunque en ésta el cruce se dio en un relativamente tardío 2% de

avance) y a la derecha un caso que no lo presenta. Ambos, sin embargo, son estadísticamente equivalentes. En otras palabras, *cualquiera* que hubiera sido la evolución inicial, hubiera desembocado en el mismo resultado final. Así pues, el comportamiento inicial de las cifras en el conteo no permite argumentar nada con respecto al resultado final.

5 Conclusiones

Este estudio contrasta los resultados de la elección presidencial de julio de 2006 en México con los resultados de una simulación computacional con el mismo número de electores, el mismo número de casillas y la misma distribución final del voto. Los resultados de la simulación se han usado para ilustrar *cómo* debió haber evolucionado el proceso de cómputo dados los resultados finales publicados por el IFE.

El análisis presentado aquí muestra que la distribución temporal del registro de casillas durante el Cómputo Distrital sufrió alteraciones que no sólo se limitaron al simple retraso en la cadena transmisión-recepción-distribución de los datos del Conteo Distrital, sino que también impactaron en el resultado electoral, incrementando artificialmente el número de votos por casilla para el PAN y disminuyéndolo artificialmente para la Alianza para el Bien de Todos.

6 Agradecimientos

El autor agradece a Pilar Alonso y a Pablo de la Mora el tiempo concedido generosamente durante la discusión y revisión de los conceptos y del contenido de este artículo. La responsabilidad del contenido recae, sin embargo, exclusivamente en el autor.