

Inhomogeneidades geopolíticas en la distribución de la lista nominal y su influencia en la distribución del cociente de participación: El caso mexicano

Hugo Hernández Saldaña

Universidad Autónoma Metropolitana en Azcapotzalco.
Av. San Pablo 180. Azcapotzalco. 02200 Cd. de México, México
e-mai: hhs@azc.uam.mx

11 de agosto de 2021

Resumen

Presentamos una actualización de dos estadísticas electorales pero considerando las inhomogeneidades en los diferentes distritos a través de la Lista Nominal para las últimas tres elecciones presidenciales. Encontramos que la distribución de la lista nominal, modulada por la geopolítica, a nivel casilla se encuentran agrupadas en tres zonas. Siendo solo un pequeño conjunto de casillas las que admiten alrededor de 750 votantes registrados. Haciendo efectiva esta estratificación se recupera la forma gaussiana para las zonas más pobladas la distribución de votantes por casilla. Para el cociente de participación la distribución no es gaussiana como podría esperarse en principio, sino que es compatible con una distribución de gaussiana sesgada o normal sesgada (*skew normal*). Se ofrece evidencia de que el sesgo es originado por la restricción dada por el límite de la lista nominal. Como un ejercicio se presentan los cálculos equivalentes, pero con otro nivel de agregación, para la elección presidencial y diputados de 1988. Aún cuando la desviación de distribuciones gaussianas es evidente la falta de datos a nivel casilla hace menos contundente el resultado.

Geopolitical Inhomogeneities in the Registered Voters' Distribution and Their Influence in the Voters' Participation Ratio Distribution: The Mexican Case.

We present an update of two electoral statistics but considering the inhomogeneities in the different districts through the Nominal List for the last three presidential elections. We find that the distribution of the nominal list, modulated by geopolitics, at the box level is grouped into three zones. Being just a small set of polls with around 750 accepted voters. By making this stratification effective, the Gaussian form is recovered for the most populated areas, the distribution of voters per square. For the participation quotient or ratio, the distribution is not Gaussian as might be expected in principle, but is compatible with a skewed normal or skew Gaussian distribution. Evidence is provided that the bias is caused by the restriction given by the nominal list limit. Equivalent calculations are presented as an exercise, but with another level of aggregation, for the 1988 presidential election and deputies. Even when the deviation from Gaussian distributions is evident, the lack of data at the box level makes the result less conclusive

1. Introducción

El estudio de las estadísticas electorales ha sido parte importante para entender algunos aspectos de los procesos modernos. Tanto para prever el ánimo de los electores como para entender cómo votan y explicar los resultados o para establecer de manera rápida los posibles resultados.

Un problema fundamental es la manera de organizar de manera igualitaria la distribución de las curules de acuerdo a la población. Éste problema es matemáticamente complicado y en países como Estados Unidos sigue siendo motivo de debates intensos. La distritación equitativa, que evite la selección geográfica de sectores que favorezcan cierto partido (la salamandrización o *gerrimandering*) sigue siendo un problema importante. Entre las razones se encuentra la distribución geográfica heterogénea de los electores y su no correspondencia con las fronteras políticas que la misma historia, la economía y la geografía han trazado. Aunado a ello, en sistemas políticos como el mexicano se decidió la conformación de una red de casillas ubicadas en lugares cercanos los votantes y con una cantidad máxima de votantes permitidos.

La estrategia propuesta para resolver éste problema es una estrategia de recosido simulado. Para la re-distribución de 2014, se usó un algoritmo de enjambre de abejas adicional Gutiérrez-Andrade et al. (2019). Como veremos más adelante, éste nuevo proceso mejoró la distribución de casillas.

Los resultados de ésta presentación corresponden a lo publicado en Hernández-Saldana (2021).

1.1. El problema

Como un ejercicio del cual surgen muchos números es natural la aplicación de estadísticas a los procesos electorales y más aún el mismo uso de estadística para diseñar las mismas votaciones. Así que es natural esperar que distribuciones estadísticas como el número de votos por casilla o la tasa, o radio, de participación cumplan con la ley de los grande números y el teorema del límite central. Esto hace que modelos simples como los de votos a la Bernoulli o más sofisticados como los modelos de voto como los resumidos en Castellano et al. (2009) sean razonables. Y pudiera resultar natural que desviaciones de éste comportamiento sean asociadas a alteraciones de las votaciones, incluida el no cumplimiento de la ley de Benford.

En el caso mexicano, la elección de 1988 es un parteaguas incluyendo este aspecto. El trabajo de Barberán et al. (1988) y las discusiones siguientes se concentran en la violación de la distribución de votos por casillas. En ese trabajo y en algunos subsecuentes, a raíz de las elecciones presidenciales de 2006, se argumentó que las estadísticas electorales en particular la distribución de votos por casilla, v_i , así como el radio o cociente de votos a lista nominal, $\tau_i = v_i/N_i$, o tasa de participación a nivel casilla, deberían cumplir con la ley de los grandes números y el teorema de límite central si hubieran sido confiables. Esto es, que las distribuciones tenderían a ser normales o gaussianas. Como no lo cumplieron se argumenta que dicha violación implica la existencia de fraude o irregularidades.

Sin embargo, dicha línea de argumentación es inconsistente pues pueden existir desviaciones a dicha ley y teorema y no hacer uso de la hipótesis de fraude para explicarlo. En la presente ponencia explicaremos lo encontrado en Hernández-Saldana (2021) sobre la distribución de voto y la tasa de participación a nivel casilla o radio de participación,

$$P_a(\tau_{a,i} = v_{a,i}/N_{a,i}), \quad (1)$$

para un grado de agregación a . En ambos casos, los resultados a nivel casilla presentan desviaciones de la gaussianidad. Es decir, no cumplen con el teorema de límite central.

Para el caso de la elección de 1988 dicha distribución puede ser adjudicada a un fraude, en las elecciones recientes no es el caso y a explicar dichas desviaciones en base a los propios datos dedicaremos el resto del trabajo.

1.2. Las elecciones de 1988

Como un ejercicio para entender el comportamiento de ambas distribuciones estudiamos las elecciones de diputados y presidencial de 1988. El autor no encontró disponibles datos a nivel casilla, por lo que se usan los datos a nivel de distrito. Para el caso de diputados los datos usados fueron los acopiados en la página de de Estudios de la Democracia y Elecciones (1991), donde se encuentra la votación total por distrito, así como la lista nominal correspondiente a ese grado de agregación. Para la elección a diputados se puede obtener el histograma de la distribución de voto y el radio de participación. En la figura 1 se muestra, en un histograma azul el resultado de este ejercicio. En línea quebrada negra aparece el ajuste gaussiano a la distribución. Como es apreciable, el ajuste es muy deficiente.

Como un *ejercicio* se puede hacer lo mismo para el caso presidencial si asumimos que la lista nominal es la misma para ambas elecciones. Note que esto no está directamente en las bases de datos consultadas y su actualización está pendiente. Para este ejercicio se muestra el histograma del radio de participación en el panel superior de la figura 1 en color negro. En línea roja quebrada se muestra el ajuste. De nuevo, el ajuste es muy malo y existen distritos con una altísima participación. No solo alejada del promedio, sino de el bulto de la distribución. Una característica interesante que se intentará abordar en cuanto se tenga acceso a registro de las actas de esa elección es tener una comparación de la asimetría que presentan éstas gráficas con las de elecciones recientes, como la de 2018, cuya tasa de participación puede ser modelada por una distribución normal sesgada. Véase la discusión más adelante y la figura 5. Ahora bien, para la conformación de el radio de participación se requiere de la distribución conjunta de los votos y de la lista nominal. Una manera simple de observar el comportamiento

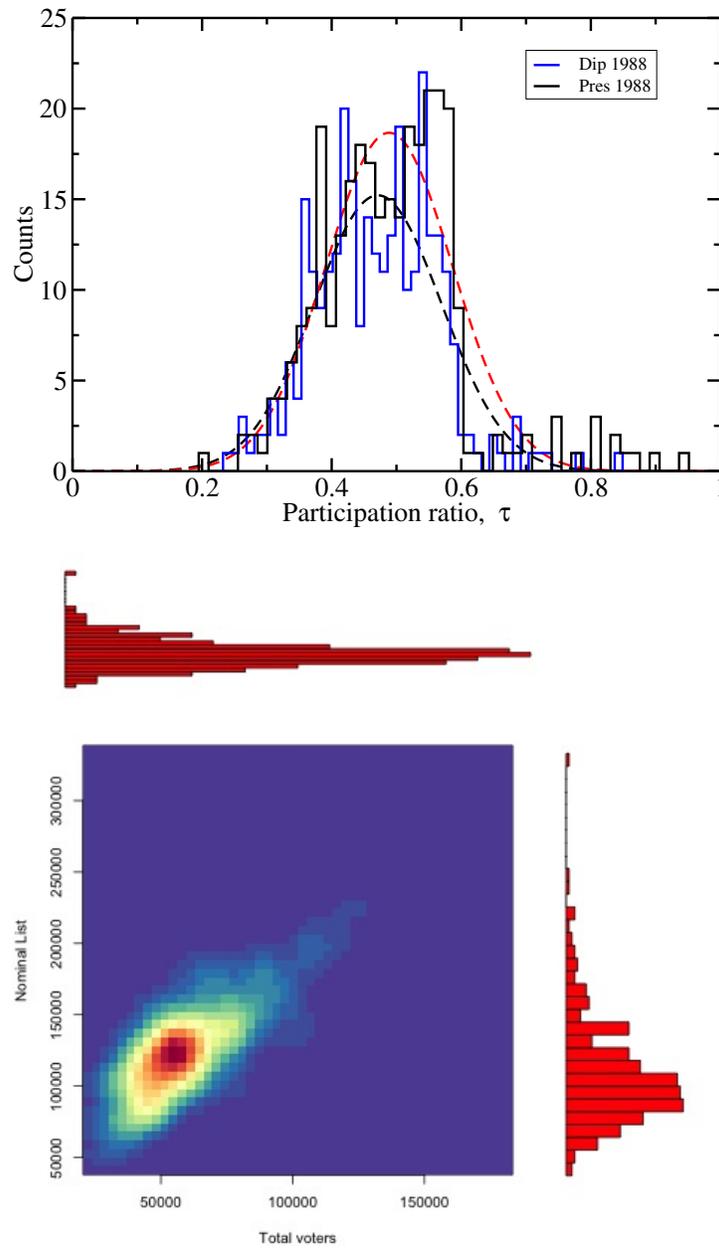


Figura 1: En el panel superior se presenta el radio de participación para la elección federal mexicana de 1988, τ , para la elección de diputados (el histograma azul) y presidencial (histograma negro). El nivel de agregación es el de distrito. En línea quebrada se dibuja el ajuste de una distribución gaussiana. En el panel inferior se muestra un mapa de calor la distribución de la lista nominal y para le número de votantes en la elección presidencial. Los histogramas correspondientes aparecen a la derecha y arriba. Gráfica elaborada por el autor usando los resultados publicados en del Colegio Electoral (1994) para el número de votantes en la elección presidencial, mientras que los de votos por diputados y la lista nominal corresponden a los datos de de Estudios de la Democracia y Elecciones (1991). *Note que la distribución presidencial es un ejercicio y que no fueron tomados de la misma base de datos.* Tomado de Hernández-Saldana (2021).

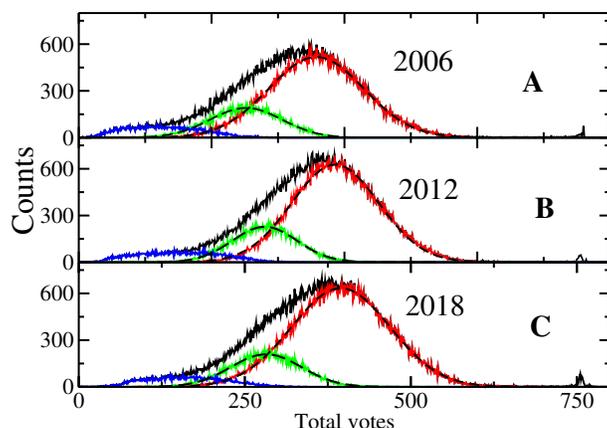


Figura 2: Histograma de frecuencias del total de votos para la elección electoral de 2006 (A), 2012 (B) y 2018 (C) a nivel casilla, considerando las contiguas como casillas independientes. En línea negra y para todos los paneles, el histograma corresponde a todos los datos disponibles. En rojo el histograma corresponde a las casillas cuya lista nominal tiene valores entre 501 y 750, que es la zona que denominamos (*I*) en la figura 3. En verde se muestran las casillas que corresponden a la zona (*II*) y en azul las casillas que tienen un límite menor a 375. En líneas quebradas, se presenta las gaussianas ajustadas a las zonas (*I*) y (*II*). Tomado de Hernández-Saldana (2021).

conjunto es el mapa de calor de las distribuciones (véase el panel inferior de la figura 1). En éste caso es irrelevante la escala del histograma conjunto, a este nivel de agregación lo relevante es que la inmensa mayoría de distritos se encuentran conjuntados en la zona roja. Existen algunos pocos que salen claramente del bulto. A diferencia de lo que observaremos para las elecciones recientes en las que la distribución se encuentra mucho más dispersa.

Concluamos este ejercicio comentando que es importante analizar los eventos pasados con las herramientas que se han ido construyendo.

2. La distribución de votos y la distribución de la lista nominal

Como se explicó anteriormente, lo que ingenuamente se puede esperar para el comportamiento de la distribución de votos y de la tasa de participación es que sigan una distribución normal.

Sin embargo las distribuciones obtenidas con los datos de las elec-

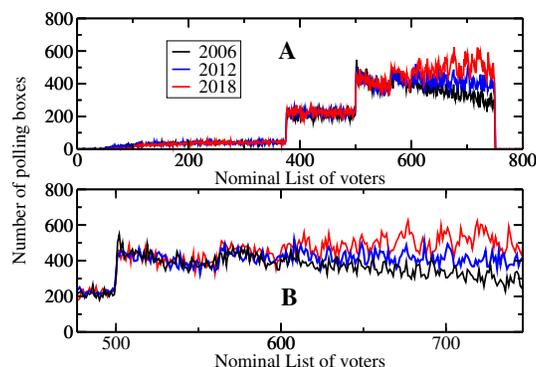


Figura 3: (A) En el eje vertical, el número de casillas con un cierto número de votantes de la lista nominal, NL, en el eje horizontal. Se muestran los datos para las elecciones presidenciales de los años, 2006, en línea negra; 2012 en azul y 2018 en rojo. Puede observarse que existen al menos tres zonas con saltos en 375 y 500. Las zonas fueron etiquetadas como (I), para la región con NL entre 501 y 750, (II) si el valor de NL esta entre 376 y 500 y (III) para los valores menores a 375. (B) Se presenta un acercamiento a la región (I) donde existe la posibilidad de que exista una nueva zona entre valores de 501 y 560. Gráfica de elaboración propia con datos obtenidos del INE/IFE. Tomado de Hernández-Saldana (2021).

ciones presidenciales recientes no lo cumplen. En todas ellas aparece un hombro para en los histogramas construidos con los datos a nivel casilla. Esto puede verse en líneas negras en la figura 2 para tres de las elecciones presidenciales recientes. El hombro aparece para casillas con una participación menor.

La explicación para éste comportamiento se encuentra en la distribución de la lista nominal. Aún cuando el número que recurrentemente aparece es de un máximo de 750 votantes por casilla, la distribución de número máximo de votantes por casilla es bastante diferente de una distribución normal. En la figura 3A se muestra para los tres procesos electorales bajo análisis. Ahí es claro que el proceso de distritación, complicado de por sí en un país heterogéneo y complejo como México, organizó la lista nominal en casillas en grupos de un número más o menos constante. Como se puede apreciar, el número de casillas con 750 votantes admisibles es aproximadamente el mismo que el de 740. Dentro de estas agrupaciones aparecen al menos tres escalones, a los que se les puso etiquetas, (I) para la zona donde las casillas tienen una lista nominal entre 501 y 750, (II) para aquellas con capacidad entre 376 y 500 y, finalmente, (III) para las casillas con capacidad

de hasta 375 votantes. No es claro si esta zonificación es resultado de los algoritmos de recosido simulado y estrategia de enjambre de abeja con los que fueron realizadas las distribuciones. El primer algoritmo se aplicó para las elecciones de 2006 y 2012 y los dos algoritmos para la elección de 2018. Sin embargo, dicha estructura se borra al hacer el análisis por distrito o municipio, es decir, ésta adjudicación de casillas se encuentra más o menos bien distribuida en la geografía y división política del país.

Un asunto distinto corresponde al comportamiento electoral en casillas con pocos electores aceptados, donde es común la existencia de irregularidades, véase por ejemplo Crespo (2008) y Klimek et al. (2012). En éste trabajo no abarcamos éste aspecto, pero queda pendiente.

¿Cuál es la consecuencia de ésta zonificación? Si se rehace la distribución de voto considerando la zona donde está clasificada la casilla obtenemos los histogramas adicionales que aparecen en la figura 3, en rojo la correspondiente a la zona (I) para las tres elecciones; en verde para la zona (II) y en azul para la (III). No hace falta ver los parámetros de ajuste. Es claro, que para las regiones populosas el ajuste de una distribución gaussiana es perfectamente razonable, aún sin mostrar otras pruebas o análisis. La zona despoblada tiene claramente no corresponde a una distribución normal, pero de la imagen es evidente que la conformación de hombro en la distribución total corresponde o es explicada por votos que provienen de ésta región.

3. Radio o cociente de participación

El radio o cociente de participación presenta una desviación de la gaussianidad si uno considera la totalidad de votos sin discriminación del número de votantes aceptado en la lista nominal. Sin embargo, para explicar la asimetría que se presenta en las distribuciones es útil considerar un ensemble de distribuciones, apropiadamente normalizadas y standardizadas, para cada una de las zonas. Es decir, al radio de participación

$$\tau_{i,k} = \frac{v_i}{k}, \quad (2)$$

de la casilla i con valor de la lista nominal k y la transformamos a

$$\eta_k = (\tau_{i,k} - \langle \tau \rangle_k) / \sigma_k. \quad (3)$$

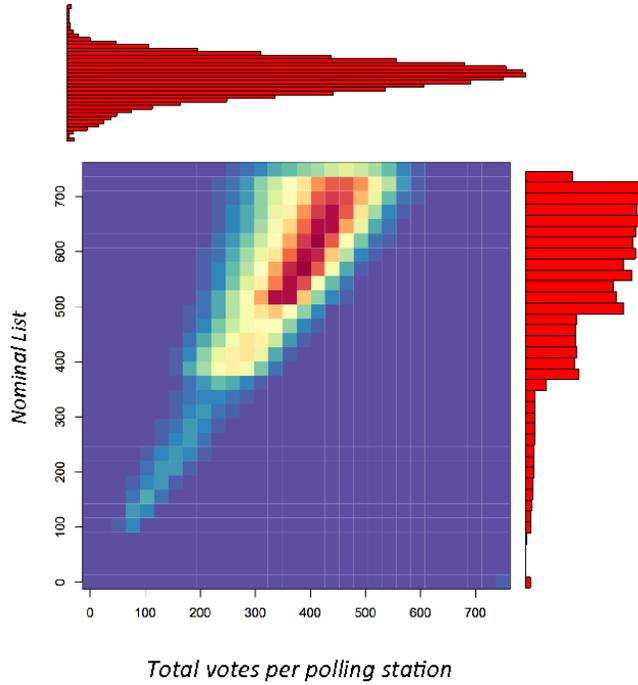


Figura 4: Mapa de calor del número de votos por casilla y la lista nominal para la elección presidencia de 2018. Los valores se van incrementando conforme se pasa del azul al rojo. EL histograma en la parte superior corresponde al total de votos y es la misma de la figura 2C en línea negra. A la derecha aparece el histograma de la lista nominal que, en detalle, se dibuja en la figura 3. Tomado de Hernández-Saldana (2021).

Aquí $\langle \tau \rangle_k$ es el promedio y σ_k la desviación standard. La distribución resultante para las zonas (I) y (II) resulta en una función asimétrica como la mostrada en la figura 5. En esta figura aparece el histograma para la variable η de la elección presidencial de 2018 y zona (I). Los otros casos son parecidos pero la asimetría es menos pronunciada. En línea negra se muestra el ajuste de una gaussiana sesgada dada por la expresión

$$\Pi(x) = \frac{2}{\sqrt{2\pi\omega}} \exp\left(-\frac{(x-\xi)^2}{2\omega^2}\right) \left(1 - \operatorname{erf}\left(\alpha \left(\frac{(x-\xi)}{\omega}\right)\right)\right), \quad (4)$$

donde $\operatorname{erf}(\cdot)$ es la función de error O'Hagan and Leonard (1976); Azzalini (1985).

La manera de generar éstas distribuciones es a través de un mecanismo de limitación. La idea es que la distribución de una variable

aleatoria, digamos z que está limitada por otra variable aleatoria w de tal manera que al tomar simultáneamente valores de ellas seleccionamos z sí y sólo si es menor al valor de w . Así seleccionada la distribución de z es una gaussiana sesgada. Este mecanismo es natural al caso que tenemos, pues los votos recibidos deben ser menores a los admitidos por la lista nominal. Los detalles matemáticos pueden verse en Hernández-Saldana (2021) o en las referencias antes citadas. Lo más interesante es que es posible escribir los parámetros de la distribución en términos de la correlación de las variables. Un ajuste con los parámetros obtenidos de esa manera es muy cercana al ajuste usando mínimos cuadrados, que es el mostrado en la figura 5.

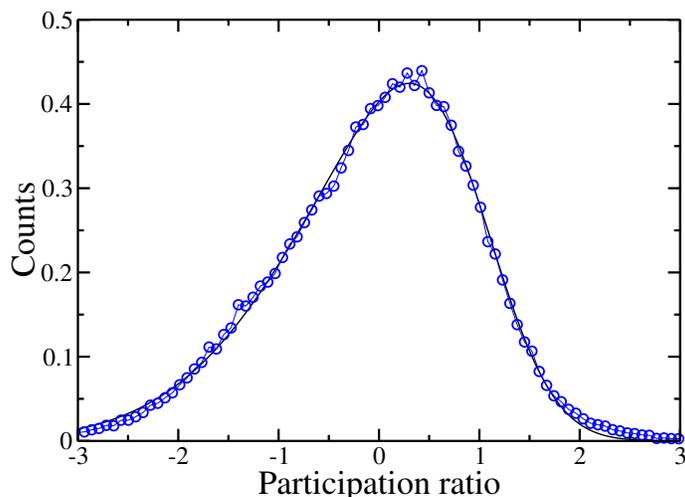


Figura 5: Distribución del radio de participación normalizado y standarizado , $\eta = (\tau_k - \langle \tau \rangle_k) / \sigma_k$ para la elección de 2018 y zona (I), en puntos y líneas azules. La línea negra corresponde al ajuste de datos para una gaussiana sesgada dada por la ecuación 4. Tomado de Hernández-Saldana (2021).

4. Bonus: La consulta de 2021

El primero de agosto de 2021 se llevó a cabo el primer ejercicio de democracia participativa en México. La organización fue realizada sin los recursos financieros solicitados por el INE por lo que las casillas fueron reconfiguradas. Con una participación de apenas 7.11 % del padrón electoral hubo alguna irregularidades. Aprovechando que los

datos están disponibles se puede hacer el mismo ejercicio y analizar algunas de las características del proceso. En la figura ?? se muestra el histograma del radio de participación para las poco más de 57 mil casillas. El estudio del bulto de la distribución se dejará para un análisis ulterior. Lo que resaltaremos en ésta ocasión son los eventos extremos, las casillas que tiene una participación de uno. Esto es extraño pues con la unión de varias casillas que en su mayoría tiene aproximadamente 2 mil votantes permitidos y en los . En la tabla 1 se dan los datos de las casillas con $\tau = 1$.

Cuadro 1: Tabla de casillas con cociente de participación igual a uno. TOTAL se refiere al total de opiniones recibidas y LN es al abreviación de Lista Nominal

ENTIDAD	DISTRITO FEDERAL	TOTAL	LN
CHIAPAS	PALENQUE	1986	1986
GUERRERO	TLAPA	1303	1303
GUERRERO	TLAPA	1302	1302
PUEBLA	HUAUCHINANGO	1965	1965

Al hacer la distribución de cociente de participación para los distritos de Palenque y Tlapa se encuentra que varias de las casillas se encuentran con radios de participación superiores a $\tau = 0,7$. Mientras que el caso de Huauchinango sí se trata de una casilla y el resto se condensan dentro de la distribución general.

5. Conclusiones

Aún cuando los procesos electorales recientes son procesos que generan miles de datos, el cumplimiento de las leyes estadísticas requiere de un entendimiento de la conformación de éstos datos, en particular a un grado de agregación de casilla. En el caso de la distribución de votos y de la tasa de participación o radio de participación las desviaciones de los comportamientos canónicos son debidos a la misma estructura del muestreo. El número de votos máximo por casilla está determinado por las constricciones de la distritación, dicha distribución está conformada por, al menos, tres zonas de agrupación distintas de la lista nominal, véase la figura 2. Al considerar esta zonificación las

distribuciones de votos con esta constricción recuperan la gaussianidad. Para el caso de la razón de participación, permite hacer evidente que su distribución no corresponde a una gaussiana, sino a una distribución asimétrica compatible con una gaussiana sesgada dada por la ecuación (4). Esto permite dar un mecanismo que explica la asimetría: corresponde a la constricción misma de tener un máximo de votos en la distribución, es decir la misma lista nominal.

Referencias

- Azzalini, A (1985), “A class of distributions which includes the normal ones.” *Scand. J. Statist.*, 12, 171–178.
- Barberán, José, Cuahutémoc Cárdenas, Adrain López Mojardín, and Jorge Zavala (1988), *Radiografía del fraude: Análisis de los resultados oficiales del 6 de julio*. Nuestro Tiempo, Mexico.
- Castellano, C, S Fortunato, and V Loreto (2009), “Statistical physics of social dynamics.” *Rev Mod Phys*, 81, 591?646.
- Crespo, José Antonio (2008), *2006: Hablan las actas.Las debilidades de la autoridad electoral mexicana*. Debate, Mexico.
- de Estudios de la Democracia y Elecciones, Centro (1991), “Cede laboratorio de análisis político y políticas públicas, base de datos.” URL <http://cede.izt.uam.mx:8089/cedeSys/faces/cede/dataBase.xhtml>.
- del Colegio Electoral, Dictamen (1994), “Declaratoria de presidente electo de los estados unidos mexicanos al licenciado carlos salinas de gortari para el periodo 1988-1994 emitida por la cámara de diputados de liv legislatura, erigida en colegio electoral.” In *Elecciones a Debate 1988 : las actas electorales perdidas* (Arturo Sánchez Gutiérrez, ed.), 208–230, Diana, Mexico.
- Gutiérrez-Andrade, Miguel Ángel, Eric Alfredo Rincón-García, Sergio Gerardo de-los Cobos-Silva, Pedro Lara-Velázquez, Roman Anselmo Mora-Gutiérrez, and Antonin Ponsich (2019), “Simulated annealing and artificial bee colony for the redistricting process in mexico.” *INFORMS Journal on Applied Analytics*, 49, 189–200, URL <https://doi.org/10.1287/inte.2019.0992>.

- Hernández-Saldana, H (2021), “Geopolitical inhomogeneities in the registered voters distribution and their influence in the voters participation ratio distribution: The mexican case.” *Frontiers in Applied Mathematics and Statistics*, 7, 518371.
- Klimek, P., Y. Yegorov, R. Hanel, and S. Thurner (2012), “Statistical detection of systematic election irregularities.” *Proceedings of the National Academy of Sciences USA*, 109, 6469?16473.
- O’Hagan, A and T Leonard (1976), “Bayes estimation subject to uncertainty about parameter constraints.” *Biometrika*, 63, 201–202.