LA FALACIA DEL EMPATE TÉCNICO ELECTORAL
The fallacy of technical tie in electoral polls

Arturo Erdely Ruiz[1]

Fecha de recepción: 29 de enero de 2018.
Fecha de aceptación: 12 de abril de 2018.

RESUMEN: En los meses previos al día de las elecciones, es una práctica usual que diversos actores políticos y ciudadanos soliciten la realización de cierto tipo de estudios con el propósito de conocer con anticipación cómo van perfilándose las preferencias electorales. Antes del día de la elección es posible recurrir a diversos tipos de encuestas consistentes en preguntar a un subconjunto (muestra) de los posibles votantes cuál es su preferencia electoral, realizar un análisis estadístico de la información recabada y hacer inferencias sobre el posible resultado de la elección. El día de la elección también es posible recurrir a otro tipo de ejercicios para inferir el resultado de la misma, antes de que se conozca el resultado del recuento total de votos emitidos varios días después: la encuesta de salida y el conteo rápido. En México y algunos otros países es común que empresas encuestadoras, políticos, analistas y hasta autoridades electorales hablen de que en un momento dado existe "empate técnico" entre dos candidatos punteros y, por tanto, no es estadísticamente posible inferir un ganador. El presente trabajo tiene por objetivo argumentar que el empleo de la expresión "empate técnico" no tiene sustento probabilístico, por lo que no debiera utilizarse en inferencias estadísticas derivadas de encuestas y conteos rápidos electorales, consecuentemente la incertidumbre sobre el posible resultado de una elección debiera expresarse mediante la estimación de la probabilidad de triunfo del candidato puntero.

1 Profesor de Carrera Titular "C" tiempo completo definitivo, Programa de Actuaría, Facultad de Estudios Superiores Acatlán – UNAM. Correo electrónico: arturo.erdely@comunidad.unam.mx.

Palabras clave: empate técnico, encuesta electoral, conteo rápido, probabilidad de triunfo.

ABSTRACT: In the months leading up to the day of the elections, it is common practice among political actors and citizens to request the execution of certain types of studies with the purpose of knowing in advance how the electoral preferences are shaping up. Before the day of the election it is possible to resort to various types of electoral surveys consisting of asking a subset (sample) of the possible voters what their electoral preference is, performing a statistical analysis of the information collected and making inferences about the possible outcome. On the day of the election it is also possible to resort to another type of exercises to infer the outcome, before several days later the result of the total count of votes cast is known: the exit poll and the quick count. In Mexico and some other countries, it is common for polling companies, politicians, analysts and even electoral authorities to speak of a "technical tie" between two leading candidates at a given moment and, therefore, it is not statistically possible to infer a winner. The present work aims to argue that the use of the expression "technical tie" has no probabilistic foundations and therefore should not be used in statistical inferences derived from polls and quick counts, and that uncertainty about the possible outcome of an election should be expressed by estimating the probability of victory of the leading candidate.

Keywords: technical tie, electoral poll, quick count, probability of victory.

Introducción

Un *evento* es una aseveración (o proposición lógica) que resultará verdadera o falsa una vez que se conozca el resultado del fenómeno (o experimento) al que se encuentra asociado. Si en un momento dado el resultado de dicho fenómeno es incierto también lo es la ocurrencia de (casi) cualquier evento asociado a él. La disciplina conocida como Estadística tiene por objetivo central el estudio de la incertidumbre (Lindley, 2000), particularmente la cuantificación y combinación de incertidumbres, auxiliándose de la *Teoría de la Probabilidad* y con base en información muestral:

> *... the statistician's role is to assist workers in other fields, the clients, who encounter uncertainty in their work. In practice, there is a restriction in that statistics is ordinarily associated with data; and it is the link between the uncertainty, or variability, in the data and that in the topic itself that has occupied statisticians [...] A scientific approach would mean the measurement of uncertainty; for, to follow Kelvin, it is only by associating numbers with any scientific concept that the concept can be properly understood* (Lindley, 2000).

Si E representa un evento de interés cuya ocurrencia es incierta, desde un punto de vista estadístico se busca cuantificar probabilísticamente el grado de incertidumbre que al respecto se tiene con base en la información disponible en un determinado momento en el tiempo. Lo usual es asociar a un evento de interés E un número en el intervalo [0,1] que se denota $P(E)$ y denomina *probabilidad del evento E*, bajo la convención de que valores más cercanos a 1 que a 0 representan mayor cercanía a la certidumbre de ocurrencia del evento, mientras que valores más cercanos a 0 que a 1 representan mayor cercanía a la certidumbre de no ocurrencia del evento, y por tanto $P(E)=0.5$ representa el mayor nivel de incertidumbre.

Al considerar, por ejemplo, la elección presidencial de México en el año 2018 como evento de interés:

E = "*el candidato López Obrador gana la elección presidencial*",

dependiendo de la información que en cada momento del proceso electoral se tenga es posible la asignación de distintas probabilidades a la ocurrencia de dicho evento. Antes del día de la elección es posible recurrir a diversos tipos de encuestas electorales consistentes en preguntar a un subconjunto (muestra) de los posibles votantes cuál es su preferencia electoral, realizar un análisis estadístico de la información recabada y hacer inferencias sobre el resultado de la elección. Dicho ejercicio involucra diversas fuentes de variabilidad que afectan la medición de la incertidumbre de interés: tamaño y representatividad de la muestra de votantes, la honestidad en sus respuestas y la no respuesta, así como los posibles cambios de opinión posteriores al levantamiento de la muestra, entre otros.

El día de la elección también es posible recurrir a otro tipo de ejercicios para inferir el resultado de la misma, antes de que varios días después se conozca el resultado del recuento total de votos emitidos: la encuesta de salida y el conteo rápido.

La primera consiste en preguntar a un subconjunto de las personas que ejercieron su voto por cuál candidato votaron justo al momento de salir de la casilla de votación, y en el análisis estadístico de la información recabada será nuevamente necesario considerar las fuentes de variabilidad ya mencionadas. El segundo consiste en realizar una inferencia estadística respecto al resultado de la elección con base en un subconjunto (muestra) del total de actas de escrutinio y cómputo, que son documentos llenados y firmados por los funcionarios de cada casilla electoral en donde asentaron el resultado del conteo de votos en la casilla correspondiente. Consecuentemente, se eliminan algunas fuentes importantes de variabilidad que tienen las encuestas electorales, como es el caso de la deshonestidad, la no respuesta o el cambio de opinión, pero persiste la proveniente del tamaño y representatividad de la muestra, entendiendo por esto último que la muestra analizada se "parezca mucho" en preferencias electorales al conjunto total de votación emitida, y esto es posible lograrlo con la adecuada aplicación de técnicas de muestreo y una eficiente implementación logística y tecnológica por parte de las autoridades electorales encargadas de dicho proceso.

En tiempos de procesos electorales y ejercicios estadísticos para realizar inferencias sobre los posibles resultados, en México y algunos otros países es común que empresas encuestadoras, políticos, analistas y hasta autoridades electorales hablen de que un momento dado existe "empate técnico" entre dos candidatos punteros y que por tanto no es estadísticamente posible inferir un ganador. El presente artículo tiene por objetivo argumentar que:

- El empleo de la expresión "empate técnico" no tiene sustento probabilístico y por tanto no debiera utilizarse en inferencias estadísticas derivadas de encuestas y conteos rápidos electorales; y
- La incertidumbre sobre el posible resultado de una elección debiera expresarse mediante la estimación de la probabilidad de triunfo del candidato puntero.

Primero se analizará lo que usualmente se entiende por "empate técnico" y se utilizarán argumentos intuitivos para criticar esta expresión. Después se abordarán algunos contraejemplos teóricos para demostrar que la probabilidad de triunfo de un candidato puede ser significativamente elevada a pesar de existir "empate técnico". Posteriormente se analizará el caso de la elección presidencial en México del año 2006 por tratarse de un caso en que (aparentemente) el resultado del conteo rápido no permitía una inferencia confiable.

I. ¿Qué es "empate técnico"?

Se denotará mediante la letra griega θ la proporción de votos que obtendrá un candidato participante en una elección entre dos o más candidatos, cantidad que en el lenguaje de la estadística se denomina *parámetro* de interés, cuyo valor es desconocido antes del recuento total de votos de dicha elección, y por tanto se desea recabar información que sea estadísticamente analizable para realizar algún tipo de *inferencia estadística* respecto a dicho parámetro.

Una *estimación puntual* de θ consiste en calcular un valor θ^* en el intervalo $[\,0, 1\,]$ (por tratarse de una proporción) con base en información recabada para tal fin, de modo que pueda considerarse una buena aproximación (en algún sentido estadístico) del valor desconocido θ.

Una *estimación por intervalo* de θ consiste en calcular un intervalo de $[\,a\,, b\,]$ de forma tal que el valor desconocido θ tenga una *aceptablemente elevada* probabilidad de estar dentro de dicho intervalo. Respecto a esto último es importante aclarar el uso de la expresión "aceptablemente elevada": una estimación por intervalo para θ que no requiere esfuerzo alguno y que además tiene la probabilidad máxima de 1 (es decir, 100 por ciento) de contener el verdadero valor de θ es justamente el intervalo $[\,0\,, 1\,]$, lo cual es tan cierto como inútil. Es posible construir intervalos de estimación de modo que su longitud $b - a$ sea menor, pero sacrificando la probabilidad de que dicho intervalo contenga al verdadero valor de θ ¿Cuánto sacrificar? Eso es totalmente arbitrario y su especificación ya no es un problema estadístico sino de criterio y decisión de quien aplica la técnica. Por ejemplo, el Instituto Nacional Electoral de México (INE, 2016) establece en su *Reglamento de Elecciones*, artículo 373, que para los conteos rápidos institucionales[2] debe utilizarse un nivel de 95 por ciento.

> ... d) La muestra deberá diseñarse con una *confianza*[3] de noventa y cinco por ciento, y con una precisión tal, que genere certidumbre estadística en el cumplimiento de los objetivos requeridos por el tipo de elección.

2 Por conteo rápido *institucional* se entiende aquél realizado por un Comité Técnico Asesor de los Conteos Rápidos de acuerdo con lo establecido en el artículo 362 del *Reglamento de Elecciones* (INE, 2016).

3 En estadística, los términos *confianza* y *probabilidad* no son equivalentes en un sentido estricto, aunque en la práctica se les interpreta como si lo fueran, situación que no es especialmente relevante discutir para los objetivos del presente artículo.

Respecto a la *precisión*, se denotará mediante una cantidad positiva ε, se refiere a un *margen de error* entre la cantidad estimada θ^* y la cantidad desconocida θ que se desea estimar, esto es, por ejemplo:

$$P(|\theta - \theta^*| \leq \varepsilon) = 0.95 \qquad (1)$$

donde el valor a utilizar para ε lo determina el usuario de la técnica estadística. En el citado artículo 373 del *Reglamento de Elecciones* (INE, 2016), por ejemplo, solo se pide un valor para ε «que genere certidumbre estadística en el cumplimiento de los objetivos requeridos por el tipo de elección», dejándolo abierto al criterio y decisión de quienes realicen el conteo rápido. Es importante destacar que el *tamaño de muestra* requerido para la encuesta o conteo rápido electoral, se denotará mediante un entero positivo n, depende del valor ε elegido, existiendo una *relación inversa* entre ellos: a menor/mayor valor de ε (esto es, a mayor/menor precisión) se requiere un mayor/menor tamaño de muestra n. La desigualdad en (1) es equivalente a:

$$\theta^* - \varepsilon \leq \theta \leq \theta^* + \varepsilon \qquad (2)$$

y permite la interpretación alternativa que, con probabilidad 95 por ciento, la proporción desconocida de votos θ estará en el intervalo:

$$[a, b] = [\theta^* - \varepsilon, \theta^* + \varepsilon] \qquad (3)$$

donde θ^* es una estimación puntual de θ con base en una muestra de tamaño n y una precisión (o margen de error) dado ε. Nótese que la longitud del intervalo resultante sería $b – a = 2\varepsilon$. Por ejemplo, si se desea $\varepsilon = 0.005$ (esto es un margen de error de 0.5 por ciento) entonces la longitud del intervalo resultante sería del doble, esto es de 0.01 (en porcentaje: 1 por ciento).

Considerando una contienda electoral en la que participan dos o más candidatos es posible realizar estimaciones por intervalo como en (3) de forma *individual* para cada candidato. Con toda intención se resalta la palabra *individual* ya que la técnica estadística para la construcción de dichos intervalos no considera la posible *interacción estadística* entre candidatos. Si se desea entrar en comparaciones y realizar una inferencia estadística respecto a qué candidato ganaría una elección, se debe recurrir a un procedimiento que incorpore además las posibles interacciones entre candidatos (ver discusión en la siguiente sección), y por tanto intervalos calculados de forma individual y que no reconocen dichas interacciones *no son intervalos estadísticamente comparables.*

Desafortunadamente, lo mencionado en el párrafo anterior es ignorado (con o sin intención) por muchos de quienes realizan, analizan y difunden encuestas y conteos rápidos electorales, recurriendo a la siguiente mala *praxis* estadística: si $[a_2, b_2]$ y $[a_1, b_1]$ son dos estimaciones por intervalo de dos candidatos punteros y se cumple la relación $a_2 < a_1 < b_2 < b_1$ (traslape de intervalos) entonces lo llaman "empate técnico", y en caso de que $b_2 < a_1$ (no traslape de intervalos) dicen que hay una "tendencia clara" que favorece al candidato del intervalo $[a_1, b_1]$.[4]

Aunque será en la siguiente sección donde se argumente formalmente por qué no tiene justificación probabilística la expresión "empate técnico", es posible adelantar algunas ideas intuitivas al respecto analizando posibles escenarios de traslape y no traslape de los intervalos de estimación para los dos candidatos punteros de una elección.

Considérese primero un escenario en el que existe un traslape de los intervalos estimados para los dos candidatos punteros (ver Tabla I). En tal caso por un traslape de tan solo una centésima de punto porcentual (0.01 por ciento) y existiendo una distancia entre las estimaciones puntuales de 0.69 por ciento se declararía un "empate técnico".

Tabla I: Ejemplo hipotético de "empate técnico" por existir traslape de intervalos.

Candidato	Intervalo estimado (%)	Estimación puntual (%)	Margen de error (%)
A	[35.75 , 36.45]	36.10	0.35
B	[35.06 , 35.76]	35.41	0.35

Fuente: elaboración propia.

Ahora considérese un escenario en el que no existe un traslape de los intervalos estimados (ver Tabla II). En este caso, con una distancia entre las estimaciones puntuales de 0.71 por ciento se declararía una "tendencia favorable" al candidato A, a pesar de que la diferencia entre las estimaciones puntuales es tan solo mayor en 2 centésimas de punto porcentual respecto al caso de

4 Aunque es poco probable, no sería imposible que ocurriera (por redondeo, por ejemplo) que $b_2 = a_1$ en cuyo caso sería interesante saber qué dirían los partidarios de esta mala praxis estadística: ¿sería o no "empate técnico"?

"empate técnico" cuando, como se analizará en la siguiente sección, la probabilidad de triunfo del candidato A sobre el candidato B es prácticamente la misma con o sin un "pequeño" traslape de intervalos.

Tabla ii: Ejemplo hipotético de "tendencia favorable" al candidato A por no existir traslape de intervalos.

Candidato	Intervalo estimado (%)	Estimación puntual (%)	Margen de error (%)
A	[35.75 , 36.45]	36.10	0.35
B	[35.04 , 35.74]	35.39	0.35

Fuente: elaboración propia.

El uso injustificado de la expresión "empate técnico" ya fue en alguna ocasión señalado:

> Si nos apegamos a la teoría estadística, no existe una definición —por lo menos al día en que se escribe este artículo, octubre de 2004— de lo que es un *empate técnico*. En México el término *empate técnico* hizo su aparición por primera vez en la televisión nacional en una encuesta de salida llevada a cabo en febrero de 1999, la jornada electoral resultó muy competida, pero un medio de comunicación a las 18:00 hrs., con un sentido más periodístico que técnico, se atrevió a declarar ganador para posteriormente conforme transcurría la tarde modificar su anuncio y salir a declarar *empate técnico*. Finalmente el triunfo fue para un partido distinto al que se había declarado [...] El uso de *empate técnico* parece ocultar al culpable de no poder dar resultados, el método. Los que lo usan parecen decir: *la elección está muy cerrada* cuando eso no es cierto siempre. Algunos incluso simplemente revisan el diseño muestral, ven que en la metodología dice 5 *por ciento de error* y de ahí cualquier distancia menor entre ganador y segundo lugar lo consideran empate. Ese es un error. Imagine simplemente una encuesta que diga que tiene 5 por ciento de error teórico y resulte en una distancia de 4 puntos entre primer y segundo lugar, alguien podrá pensar que es empate porque puede resultar en un ganador distinto, pero resulta que también podría

resultar en que la ventaja real fuera de 9 puntos, lo cual no refleja una contienda cerrada (Campos[5] y Penna, 2004).

El uso de la expresión "empate técnico" persiste por lo menos hasta el año en que se escribe el presente artículo, y como ejemplo basta citar, en el contexto de la elección presidencial de México en el año 2018, la nota periodística de la Figura I. Incluso autoridades electorales llegan a utilizar la expresión y la difunden en información dirigida a los electores, como es el caso del Instituto Electoral de Coahuila respecto a la elección de gobernador en el año 2017, que en su página de internet publicó lo que se muestra en la Figura II.

En la siguiente sección se aborda con rigor matemático que el empleo de la expresión "empate técnico" no tiene sustento probabilístico y por tanto no debiera utilizarse en inferencias estadísticas derivadas de encuestas y conteos rápidos electorales.

Figura I


EL FINANCIERO

HOME TLCAN BUSINESSWEEK ECONOMÍA EMPRESAS NACIONAL

NACIONAL

Meade en empate técnico con AMLO en preferencias electorales: Ochoa

El líder del PRI dijo que de acuerdo con la estrategia trazada junto con los partidos Verde y Nueva Alianza, el crecimiento de Meade en las preferencias será progresivo durante lo que resta de precampaña y en la campaña.

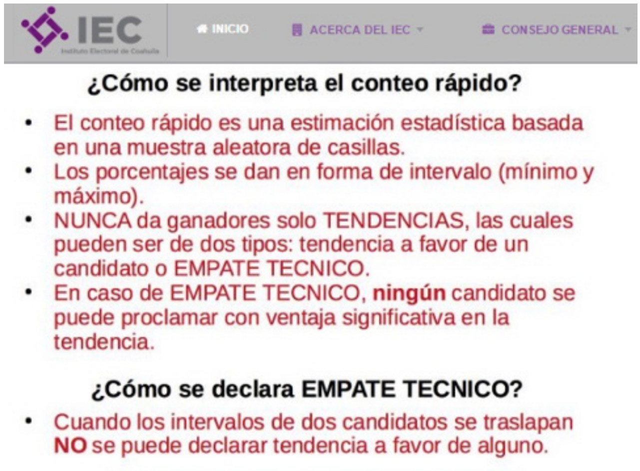
Magali Juárez 02.01.2018 Última actualización 02.01.2018

ETIQUETAS: aef, PRI, Andrés Manuel López Obrador, Enrique Ochoa Reza, TTV, José Antonio Meade Kuribreña,

Fuente: Periódico *El Financiero*. Disponible en http://www.elfinanciero.com.mx/nacional/meade-en-empate-tecnico-con-amlo-en-preferencias-electorales-ochoa.html. Consultado el 13 de enero de 2018.

5 Act. Roy Campos, presidente de Consulta Mitofsky, empresa especializada en generar, analizar y presentar información para el diseño de estrategias, estimaciones de proyección y evaluación de desempeño en México, Estados Unidos y Centroamérica. Disponible en www.consulta.mx. Consultado el 13 de enero de 2018.

FIGURA II



Fuente: disponible en www.iec.org.mx/v1/index.php/conteo-rapido. Consultado el 12 de junio de 2017.

II. Contraejemplos

Considérese una elección en la que hay más de dos candidatos contendientes y que es de interés calcular la probabilidad de triunfo entre los dos candidatos punteros. Las proporciones de votos de los dos candidatos punteros se representarán mediante las *variables aleatorias* X e Y mientras que en una tercera variable aleatoria Z se agregará la proporción restante de la votación emitida (otros candidatos registrados, votos anulados y votos por candidatos no registrados), de modo que debe cumplirse la condición:

$$X+Y+Z=1 \quad (4)$$

esto es, aunque antes del recuento total de votos las proporciones obtenidas mediante X, Y y Z son desconocidas, sea cual sea el resultado, su suma debe ser igual a 1 (es decir, equivalente al 100 por ciento de la votación emitida). De (4) se tiene que $Z=1-X-Y$ y por tanto una vez conocidos los valores para el vector aleatorio (X,Y) el valor de Z queda perfectamente determinado. Lo anterior implica que basta con tener un modelo probabilístico para el vector aleatorio (X,Y) y con ello sería suficiente para realizar cálculos de probabilidades para Z también.

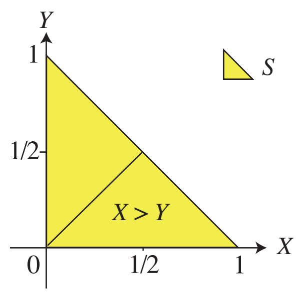
Como la proporción de votos a obtener por cada candidato es cualquier valor en el intervalo $[\,0\,,\,1\,]$ se considerará a (X,Y) como vector de variables aleatorias *continuas* sobre dicho intervalo (lo cual implica lo mismo para Z). Consecuencia de (4) se tiene también que $0 \leq X+Y \leq 1$ y por tanto $0 \leq X \leq 1-Y$, por lo que $P[(X,Y) \in S] = 1$, donde S es el *soporte* de (X,Y) dado por el conjunto:

$$S = \{(x,y) : 0 \leq y \leq 1,\ 0 \leq x \leq 1-y\} \qquad (5)$$

como se ilustra en la Figura III. Como (X,Y) representa las proporciones de votos de los dos candidatos punteros, es de interés calcular la probabilidad de triunfo de alguno de ellos, por ejemplo $P(X > Y)$ misma que se calcula mediante:

$$P(X > Y) = P[(X,Y) \in S_x] \text{ con } S_x = \{(x,y) : 0 \leq y < \tfrac{1}{2},\ y < x < 1-y\} \subset S \quad (6)$$

Figura III. Soporte S del vector aleatorio (X,Y) y la región donde $X > Y$.



Fuente: elaboración propia.

Toda la información para hacer cálculo de probabilidades respecto al vector aleatorio (X,Y) se encuentra en su función de distribución conjunta de probabilidades $F_{XY}(x,y) = P(X \leq x, Y \leq y)$ de donde se deducen además la funciones de distribución marginal (o individual) de las variables aleatorias X e Y, esto es $F_X(x) = P(X \leq x) = F_{XY}(x,+\infty)$ y $F_Y(y) = P(Y \leq y) = F_{XY}(+\infty, y)$. Como consecuencia del *Teorema de Sklar* (Sklar, 1959) se tiene, en este caso, que existe una relación funcional única entre la función de distribución conjunta F_{XY} y las funciones de distribución marginales F_X y F_Y de la forma siguiente:

$$F_{XY}(x,y) = \boldsymbol{C}(F_X(x),\ F_Y(y)), \qquad (x,y) \in \boldsymbol{R}^2 \quad (7)$$

donde $\boldsymbol{C}: [0,1]^2 \rightarrow [0,1]$ se denomina *función cópula* asociada al vector aleatorio (X,Y). Como las funciones de distribución marginales F_X y F_Y sólo explican el comportamiento probabilístico individual de las variables aleatorias X e Y *por separado*, entonces toda la información acerca de la *dependencia probabilística* entre dichas variables aleatorias se encuentra justamente en $\boldsymbol{C}$. Así, la función cópula que representa ausencia de dependencia (es decir, independencia) entre X e Y es $\boldsymbol{C}(u,v) = \Pi(u,v) = uv$, ya que es conocido resultado de la teoría de la probabilidad que X e Y son independientes si y sólo si $F_{XY}(x,y) = F_X(x)\,F_Y(y) = \Pi(F_X(x), F_Y(y))$.

También, como consecuencia de las *Cotas de Fréchet–Hoeffding* (Fréchet, 1951; Hoeffding, 1940) para funciones de distribución conjunta bivariadas, se tiene que:

$$\boldsymbol{W}\,(F_X(x), F_Y(y)) \leq F_{XY}(x,y) \leq \boldsymbol{M}\,(F_X(x), F_Y(y)) \qquad (8)$$

con las funciones cópula $\boldsymbol{W}(u,v) = máx\,\{u + v - 1, 0\}$ y $\boldsymbol{M}(u,v) = mín\,\{u,v\}$. Para un vector de variables aleatorias continuas (X, Y) ya se ha demostrado (Nelsen, 2006) que su función cópula $\boldsymbol{C} = \boldsymbol{M}$ si y sólo si Y es función estrictamente creciente de X (casi seguramente, en sentido probabilístico), y que $\boldsymbol{C} = \boldsymbol{W}$ si y sólo si Y es función estrictamente decreciente de X (casi seguramente, en sentido probabilístico).

Es inmediato verificar que para un vector de variables aleatorias continuas (X, Y) con soporte en S (ver Figura III) NO es posible la independencia entre dichas variables ya que de serlo debería cumplirse para todo $0 < x < 1$ que $P(Y > x) = P(Y > x \mid X = x)$ y claramente $P(Y > x) > 0$ mientras que la probabilidad $P(Y > x \mid X = x) = 0$ como consecuencia del soporte S, cuando $x > 0.5$.

¿Será posible que un vector aleatorio de variables aleatorias continuas (X, Y) con soporte en S alcance (o se acerque mucho a) las cotas de Fréchet–Hoeffding, es decir, que tenga por función cópula $\boldsymbol{C} = \boldsymbol{W}$ o bien $\boldsymbol{C} = \boldsymbol{M}$? La respuesta es en el sentido afirmativo: basta con que, por ejemplo, X tenga una distribución de probabilidad continua uniforme sobre el intervalo $[\,0\,,1\,]$ y definir la variable aleatoria $Y = g_\delta(X)$ donde la función $g_\delta : [0,1] \rightarrow [0\,, 1 - \delta]$ con parámetro $0 < \delta < 1$ está dada por:

$$g_\delta(x) = g_1(x)\,\mathbf{1}_{\{0 \leq x \leq \delta\}} + g_2(x)\,\mathbf{1}_{\{\delta < x \leq 1\}}$$

$$= \frac{1-\delta}{\delta}\,x\mathbf{1}_{\{0 \leq x \leq \delta\}} + (1-x)\,\mathbf{1}_{\{\delta < x \leq 1\}} \qquad (9)$$

Es relativamente sencillo demostrar[6] que la función de distribución de probabilidad marginal para Y resulta ser, en este caso, continua uniforme también, pero sobre el intervalo $[0,1-\delta]$ y que la función cópula correspondiente a (X, Y) en este caso resulta ser una familia paramétrica $\{\boldsymbol{C}_\delta\,(u,v)\!: 0 < \delta < 1\}$ como en el ejemplo 3.3 (Nelsen, 2006) donde $lím_{\delta\to 0^+}\boldsymbol{C}_\delta = \boldsymbol{W}$ y $lím_{\delta\to 1^-}\boldsymbol{C}_\delta = \boldsymbol{M}$. Por lo tanto, se analizarán los casos extremos de dependencia para un vector de variables aleatorias continuas (X, Y) con soporte en S.

Contraejemplo 1

Es posible tener $\boldsymbol{C} = \boldsymbol{W}$ si Y es función estrictamente decreciente de X. Por ejemplo, definiendo $Y = 1 - X$ se cumple con ello y además dicha relación funcional es posible dentro de S (ver Figura III), que de hecho coincide con parte de la *frontera* de S que une los puntos (0,1) y (1,0). En este caso, la probabilidad de que X triunfe sobre Y estaría dada por:

$$P(X>Y) = P(X>1-X) = P(X>½) = 1 - F_X(½) \qquad (10)$$

$$= P(1-Y>Y) = P(Y<½) = F_Y(½) \qquad (11)$$

donde

$$F_Y(y) = P(Y\leq y) = P(1-X\leq y) = P(X\geq 1-y) = 1 - F_X(1-y) \qquad (12)$$

De (10) se desprende que, para cualquier valor dado γ tal que $0 < \gamma < 1$, la probabilidad $P(X > Y)=1-\gamma$ si y sólo si $F_X(½)= \gamma$. De (11) se desprende que $P(X > Y) = 1-\gamma$ si y sólo si $F_Y(½) = 1-\gamma$. Esto implicaría que $[½\,,1]$ sea un intervalo de probabilidad $100(1-\gamma)\%$ para X y que $[0\,,½]$ sea un intervalo de probabilidad $100(1-\gamma)\%$ para Y, esto es:

$$F_X(1) - F_X(½) = 1 - \gamma = F_Y(½) - F_Y(0). \qquad (13)$$

Como $F_X(½) = \gamma$, si F_X es estrictamente creciente entonces por continuidad de F_X existe un valor $0 < x_{\gamma/2} < ½$ tal que $F_X(x_{\gamma/2}) = \gamma/2$ y análogamente también se tiene que existe un valor $½ < x_{1-\gamma/2} < 1$ tal que $F_X(x_{1-\gamma/2}) = 1 - \gamma/2$, y por tanto:

$$F_X(x_{1-\gamma/2}) - F_X(X_{\gamma/2}) = (1-\gamma/2) - \gamma/2 = 1-\gamma. \qquad (14)$$

Por argumentos análogos, también es posible encontrar $0 < y_{\gamma/2} < ½ < y_{1-\gamma/2} < X_{1-\gamma/2}$ tales que:

6 Se omiten los detalles por no ser especialmente relevantes para el objetivo del presente artículo.

$$F_Y(Y_{1-\gamma/2}) - F_Y(Y_{\gamma/2}) = (1-\gamma/2) - \gamma/2 = 1-\gamma \qquad (15)$$

por lo que de (14) y (15) se concluye que es posible construir para X un intervalo $[X_{\gamma/2}, X_{1-\gamma/2}]$de probabilidad 100 (1−γ)% y análogamente para Y un intervalo $[Y_{\gamma/2}, Y_{1-\gamma/2}]$ también de probabilidad 100 (1−γ)% que tengan intersección no vacía $[X_{\gamma/2}, Y_{1-\gamma/2}]$, y con probabilidad de triunfo para X sobre Y igual a $P(X>Y) = 1-\gamma$. Así, por ejemplo, es posible tomar un valor tan pequeño como $\gamma = 0.0001$ y con ello X tendría una probabilidad de triunfo sobre Y de 99.99 por ciento a pesar de existir traslape de los intervalos individualmente calculados.

Contraejemplo 2

Para que la función cópula ***C*** correspondiente al vector aleatorio (X, Y) fuese igual a ***M*** se requeriría que $Y=\phi(X)$ casi seguramente, donde ϕ es una función estrictamente creciente, lo cual no es posible sobre S ya que necesariamente existiría un valor $0 < x_0 < 1$ tal que $\phi(x_0)\ 1-x_0$, es decir, Y reportaría valores inadmisibles (fuera de S). Pero es posible lograr que ***C*** se aproxime tanto como se desee a ***M*** mediante $Y = g_\delta(X)$ con g_δ definida como en (9), esto es:

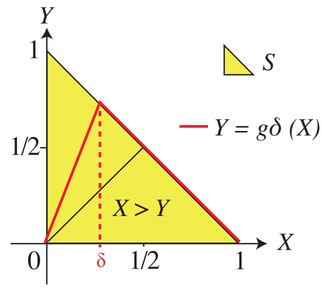
$$\begin{aligned} Y = g_\delta(X) &= g_1(X)\,\mathbf{1}_{\{0\le x\le\delta\}} + g_2(X)\,\mathbf{1}_{\{\delta<x\le 1\}} \\ &= \frac{1-\delta}{\delta}X\mathbf{1}_{\{0\le x\le\delta\}} + (1-X)\,\mathbf{1}_{\{\delta<x\le 1\}} \qquad (16) \end{aligned}$$

con parámetro $0 < \delta < 1$. El soporte de Y sería el intervalo $[0, 1-\delta]$ y su función de distribución para $0 \le y \le 1-\delta$ estaría dada por:

$$\begin{aligned} F_Y(y) = P(Y\le y) &= P(X \le g_1^{-1}(y)) + P(X \ge g_2^{-1}(y)) \\ &= F_X\left(\tfrac{\delta y}{1-\delta}\right) + 1 - F_X(1-y). \qquad (17) \end{aligned}$$

Con parámetro $\delta \ge ½$ la poligonal $Y = g_\delta(X)$ quedaría totalmente contenida en S_X, ver (6), y en tal caso $P(X>Y)=1$. Si se desea $P(X>Y) = 1-\gamma$ para algún valor determinado $0<\gamma<1$ sería entonces necesario utilizar algún valor $0<\delta<½$, ver Figura IV, en cuyo caso se tendría que $1-\gamma = P(X>Y) = P(X>½) = 1-F_X(½)$ y por lo tanto $F_X(½) = \gamma$.

Figura iv. En color rojo la gráfica de $Y = g_\delta(X)$



Fuente: elaboración propia.

Por otro lado, aplicando (17) se obtiene que:

$$F_Y(½) = F_X\left(\frac{\delta}{2(1-\delta)}\right) + 1 - \gamma = 1 - \gamma^* > 1 - \gamma \quad (18)$$

y aplicando argumentos análogos a los del *Contraejemplo 1* nuevamente es posible construir para X un intervalo $[x_{\gamma/2}, x_{1-\gamma 2}]$ de probabilidad $100(1-\gamma)\%$ y para Y un intervalo $[Y_{\gamma^*/2}, Y_{1-\gamma^*/2}]$ de probabilidad $100(1-\gamma^*)\% > 100(1-\gamma)\%$ que tengan traslape $[X_{\gamma/2}, Y_{1-\gamma^*/2}]$ y con probabilidad de triunfo para X sobre Y igual a $P(X > Y) = 1 - \gamma$. Así, por ejemplo, es posible tomar un valor tan pequeño como $\gamma = 0.0001$ y se tendría nuevamente que X tendría una probabilidad de triunfo sobre Y de 99.99 por ciento a pesar de existir traslape de los intervalos individualmente calculados.

Contraejemplo 3

En los dos contraejemplos anteriores se analizaron los casos extremos de dependencia que corresponden a las cotas inferior y superior de Fréchet–Hoeffding. Ahora se analizarán casos intermedios de dependencia para un vector de variables aleatorias continuas (X, Y) con soporte contenido en S. Considérese el caso de una elección en la que ninguno de los dos candidatos punteros alcanzaría mayoría absoluta de votos, en tal escenario se tendría que $P\big((X, Y) \in [0, ½]^2\big) = 1$. En este caso la probabilidad de triunfo de X sobre Y estaría dada por:

$$P(X > Y) = P\Big((X,Y) \in S_X \cap [0,\tfrac{1}{2}]^2\Big) = P(0 \leq X \leq \tfrac{1}{2}, 0 \leq Y < X). \quad (19)$$

Si B es una variable aleatoria continua con distribución de probabilidad *Beta* con parámetros $\alpha > 0$ y $\beta > 0$, definiendo $X = B/2$ se logra que $P(0 \leq X \leq \tfrac{1}{2}) = 1$, y la relación entre la función de distribución de X y la de B estaría dada por:

$$F_X(X \mid \alpha, \beta) = P(X \leq x) = P(B \leq 2x) = F_B(2x \mid \alpha, \beta). \quad (20)$$

De forma análoga a los dos contraejemplos anteriores, dados valores $0 < \gamma < 1$ así como $0 < 2\varepsilon < x_{1-\gamma/2} < \tfrac{1}{2}$ bastaría encontrar valores[7] $\alpha > 1$ y $\beta > 1$ tales que:

$$F_X(x_{1-\gamma/2} \mid \alpha, \beta) = 1 - \gamma/2 \ \text{ y } \ F_X(x_{\gamma/2} \mid \alpha, \beta =) \ \gamma/2 \quad (21)$$

donde $x_{\gamma/2} = x_{1-\gamma/2} - 2\varepsilon$. Lo anterior nos arroja un intervalo $[x_{\gamma/2}, x_{1-\gamma/2}]$ de longitud 2ε (esto es con margen de error ε) y de probabilidad $100(1-\gamma)\%$ para X. Con el intervalo para X se puede construir uno similar para Y que tenga un traslape deseado $0 < \tau < 2\varepsilon$: tómese $y_{1-\gamma/2} = x_{\gamma/2} + \tau$ y aplíquese el mismo procedimiento (21), con lo que se obtendrá un intervalo $[y_{\gamma/2}, y_{1-\gamma/2}]$ de longitud 2ε (esto es, con margen de error ε) y de probabilidad $100(1-\gamma)\%$ para Y que además tiene un traslape de longitud τ con el intervalo obtenido para X.

Los dos intervalos con traslape para X e Y se obtienen únicamente a partir de las distribuciones de probabilidad *marginales* de cada variable aleatoria, sin considerar aún la posible *dependencia* entre ellas, misma que se procederá a agregar ahora. Si $F_X(x \mid \alpha_1, \beta_1)$ y $F_Y(y \mid \alpha_2, \beta_2)$ son las funciones de distribución marginales obtenidas para lograr los intervalos traslapados anteriores, entonces por el *Teorema de Sklar* (Sklar, 1959) la función de distribución conjunta de (X,Y) es de la forma:

$$F_{XY}(x,y) = \boldsymbol{C}\,(F_X(x \mid \alpha_1, \beta_1), F_Y(y \mid \alpha_2, \beta_2)) \quad (22)$$

donde $\boldsymbol{C}$ es la función cópula que contiene toda la información sobre la dependencia entre las variables aleatorias X e Y. Para el propósito del presente contraejemplo se utilizará (entre muchas opciones disponibles) una función cópula perteneciente a la *familia Frank* (Frank, 1979) por ser ésta una familia paramétrica $\{\boldsymbol{C}_\delta : \delta \in\ ]-\infty, +\infty[\ \setminus \{0\}\}$ que abarca desde dependencias que se aproximan tanto como se desee a las cotas extremas de dependencia de Fréchet–Hoeffding ($\lim_{\delta \to -\infty} \boldsymbol{C}_\delta = \boldsymbol{W}$ y $\lim_{\delta \to +\infty} \boldsymbol{C}_\delta = \boldsymbol{M}$) hasta dependencias que pueden aproximarse a la independencia tanto como se desee

7 Las restricciones $\alpha > 1$ y $\beta > 1$ garantizan que la función de densidad de probabilidades *Beta* tenga un máximo y forma "acampanada".

($lím_{\delta \to 0}$ $\boldsymbol{C}_{\delta} = \Pi$). Bajo esta familia de cópulas, a mayor valor positivo del parámetro δ se obtiene una mayor dependencia positiva, a menor valor negativo se obtiene mayor dependencia negativa, y conforme el parámetro se acerque a cero se acercará más a la independencia. Una vez especificada (22) se calcula la probabilidad de triunfo de X sobre Y mediante (19).

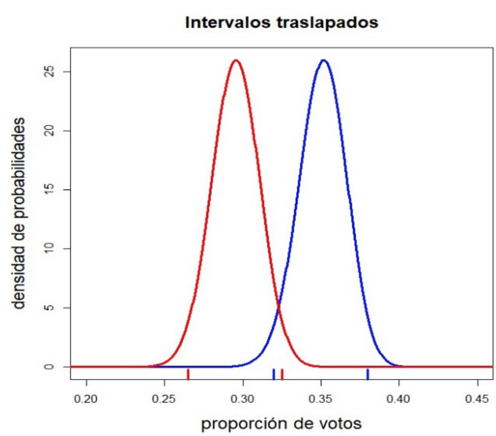
La solución del sistema de ecuaciones (21) tiene que aproximarse numéricamente, y el cálculo de la probabilidad de triunfo (19) se aproxima vía simulación. En este caso se hizo por medio del lenguaje de programación R (R Core Team, 2017) y el paquete *copula* (Hofert *et al.*, 2017). El código de programación se anexa en el Apéndice, así como un enlace para descarga del mismo, para fines de reproducibilidad de los cálculos.

Se utilizaron valores $\gamma = 0.05$, $x_{1-\gamma/2} = 0.38$, $\tau = 0.005$ y ε 0.03. Lo anterior genera intervalos de probabilidad 95 por ciento con margen de error de 3 por ciento y traslape de intervalos de 0.5 por ciento como sigue:

$$Y : [\,26.5\%, 32.5\%] \quad X: [32\%, 38.0\%] \qquad (23)$$

con funciones de densidad de probabilidad marginal como se ilustra en la Figura V.

Figura V. Funciones de densidad de probabilidad marginal para X (azul) e Y (rojo) junto con intervalos de probabilidad 95%, margen de error de 3% y traslape de 0.5%



Fuente: elaboración propia.

A las funciones de densidad de probabilidad marginal obtenidas, se agregó la función cópula Frank con distintos valores de *correlación de Spearman*[8] y se procedió a estimar vía simulación la probabilidad de triunfo de X sobre Y (ver Tabla III). El resultado de la simulación para el caso particular donde la correlación de Spearman es igual a –0.5 se ilustra en la Figura VI.

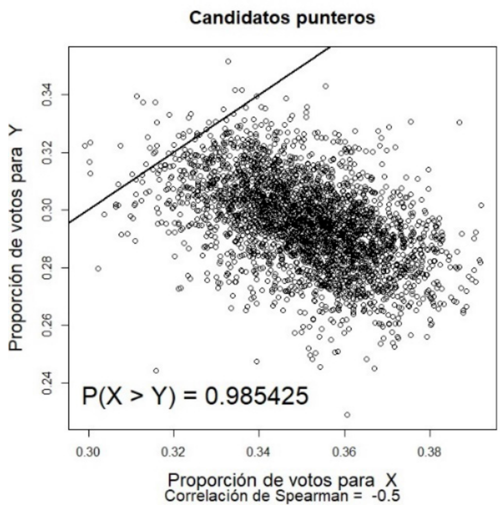
Tabla III: Probabilidad de triunfo de X sobre Y bajo distintos niveles de dependencia, con intervalos de probabilidad 95%, traslape de 0.5% y margen de error de 3%

Correlación de Spearman	Probabilidad de triunfo de X (en %)
–0.9	97.13
–0.5	98.54
–0.2	99.14
0.0	99.44
+0.3	99.77
+0.6	99.95
+0.9	99.99

Fuente: elaboración propia.

8 La *correlación de Spearman* es un valor en el intervalo [–1, +1] cuya interpretación es muy similar a la *correlación lineal* o *de Pearson*, con la ventaja de que solo depende de la cópula subyacente, no se contamina por el comportamiento de las distribuciones marginales y no depende de la existencia de momentos (Embrechts *et al*., 1999).

Figura VI. Simulaciones de (X, Y) con cópula Frank y correlación de Spearman igual a –0.5. La parte inferior de la línea recta corresponde a escenarios donde $X > Y$



Fuente: elaboración propia.

Como puede apreciarse en la Tabla III, a pesar de existir un traslape en los intervalos individuales para los candidatos punteros, bajo distintos niveles de dependencia (negativos, independencia, positivos) la probabilidad de triunfo de un candidato sobre el otro es siempre superior al nivel 95 por ciento con el que fueron construidos los intervalos.

Se verá ahora que la probabilidad de triunfo puede resultar casi la misma con un mínimo traslape de intervalos que sin él. Basta utilizar los valores $\tau_1 = +\,0.0001$ (traslape) y $\tau_2 = -0.0001$ (no traslape) con los demás valores igual que en lo anterior. Los resultados se resumen en la Tabla IV y deja claro que un mínimo traslape de intervalos *versus* una mínima separación no cambia significativamente la probabilidad de triunfo de un candidato sobre otro.

Tabla iv: Probabilidad de triunfo de X sobre Y con correlación de Spearman −0.5, intervalos de probabilidad 95% (traslapados en un caso y en otro no) y margen de error de 3%

Candidato Y	Candidato X	¿Traslape?	Probabilidad de triunfo de X (en %)
[26.01 , 32.01]	[32.00 , 38.00]	Sí	99.19
[25.99 , 31.99]	[32.00 , 38.00]	No	99.21

Fuente: elaboración propia.

En resumen, los tres contraejemplos anteriores demuestran que es posible tener elevadas probabilidades de triunfo de un candidato puntero a pesar de que exista algún traslape de los intervalos calculados por separado para los dos candidatos punteros, y que inclusive la probabilidad de triunfo puede resultar esencialmente la misma con un mínimo traslape de intervalos que sin él.

III. La elección presidencial de México en 2006

Se analiza este caso por tratarse de un ejemplo en el que, derivado de un conteo rápido a cargo de una autoridad electoral ésta decide, con asesoría de especialistas en estadística, no emitir una conclusión al respecto la noche del día de la elección. De acuerdo con Eslava (2006), quien formó parte del *Comité Técnico Asesor para la Realización de Conteos Rápidos* (ctarcr) en la elección presidencial de México el 2 de julio del año 2006:

> A pesar de que se obtuvieron estimaciones puntuales de los porcentajes, dos de los intervalos de confiabilidad asociados no se separaron por más de 0.006, margen acordado para poder proporcionar los resultados la misma noche. Por esta razón no se difundieron las cifras estimadas, sino solamente el aviso de que estadísticamente no era posible distinguir con un alto grado de confiabilidad a un candidato ganador [...] El conteo rápido, por ser de carácter institucional y cuyos resultados serían la base de un comunicado difundido a la ciudadanía la misma noche de la elección, estuvo sujeto a acuerdos previos y a restricciones como guardar la confidencialidad de la muestra; realizar estimaciones por intervalos de cuando menos una confiabilidad de 95%; que para identificar un partido o coalición ganadora, los intervalos asociados a

> los porcentajes estimados de votos para los partidos mayoritarios deberán distar en al menos 0.6% (0.006)...

En el *Informe sobre las actividades del Comité Técnico Asesor para la Realización de Conteos Rápidos* (CTARCR) del entonces Instituto Federal Electoral (IFE, 2006) se incluyen tres metodologías estadísticas para la estimación por intervalos (denominados métodos Robusto, Clásico y Bayesiano) así como la técnica de muestreo utilizada, entre otras cuestiones. Sin embargo, no aparece metodología o justificación estadística alguna para establecer como condición para identificar a un ganador que la distancia entre intervalos punteros fuese al menos de 0.6 por ciento. En la página 24 del citado informe del CTARCR simplemente se menciona respecto al *Proceso de Estimación* lo siguiente:

> Para el proceso de estimación de las proporciones de votos por partido y coalición, se especificó que se tendrían como reglas básicas las siguientes:
> 1. Los métodos de estimación utilizados deben llevar a conclusiones coherentes y comunes.
> 2. Realizar estimaciones por intervalos de cuando menos una confiabilidad del 95%.
> 3. Para poder identificar un ganador, los intervalos de las primeras dos fuerzas contendientes deberán distar en al menos 0.6%.

Estas "reglas básicas" que se adoptaron para la estimación del conteo rápido de dicha elección presentan diversos problemas. La *Regla 1* es inespecífica porque ¿cómo decidir si las conclusiones son "coherentes" y en qué sentido? ¿A qué se refiere con "conclusiones comunes"? El informe no lo aclara. La *Regla 2* no especifica el nivel de confiabilidad de las estimaciones, solo da un cota inferior para ello al decir «cuando menos una confiabilidad del 95 por ciento», lo cual dejó la puerta abierta a la utilización de distintos y arbitrarios niveles de confiabilidad por encima del 95 por ciento, pudiéndose llegar al absurdo de utilizar una confiabilidad del 100 por ciento y con ello entregar como intervalo de estimación [0%, 100%]. De hecho el informe tampoco aclara cuáles fueron los niveles de confiabilidad finalmente utilizados con cada uno de los tres métodos de estimación, ni si fue el mismo en los tres casos. De la *Regla 3* surge inmediatamente una pregunta: ¿a qué se refiere con que los intervalos estimados para los dos punteros "disten" en al menos 0.6 por ciento? Tampoco se incluye en dicho informe la fórmula que se utilizó para calcular "distancia entre intervalos", mucho menos un fundamento estadístico para ello ni para fijar esa "distancia" mínima en 0.6 por ciento, simplemente fue un "acuerdo" del CTARCR cuya fundamentación no fue documentada en el informe. Matemáticamente es posible definir alguna forma de calcular distancia entre intervalos (por ejemplo, utilizando la distancia de Hausdorff) pero para que esto pueda aplicarse en un procedimiento de inferencia estadística requeriría al menos una justificación técnica, que de haberla no fue documentada ni revelada.

Los intervalos estimados por el CTARCR la noche del 2 de julio de 2006 para los dos candidatos punteros (Felipe Calderón Hinojosa del PAN y Andrés Manuel López Obrador de la CPBT,[9] junto con el resultado del PREP[10] al cierre del mismo el 3 de julio[11] y el Cómputo Distrital[12] del 5 de julio se presentan en la Tabla V, y una gráfica comparando dichos intervalos bajo los tres métodos utilizados (Robusto, Clásico y Bayesiano) se presenta en la Figura VII.

Tabla V: Intervalos estimados la noche del 2 de julio de 2006 por los métodos estadísticos denominados por el CTARCR como Robusto, Clásico y Bayesiano, el resultado al cierre del PREP el 3 de julio, y el resultado del Cómputo Distrital el 5 de julio, para los dos candidatos punteros: Felipe Calderón Hinojosa (FCH) y Andrés Manuel López Obrador (AMLO)

Método	AMLO	Longitud intervalo	FCH	Longitud intervalo
Robusto	[34.24 , 36.38]	2.14	[35.25 , 37.40]	2.15
Clásico	[34.97 , 35.70]	0.73	[35.68 , 36.53]	0.85
Bayesiano	[35.07 , 35.63]	0.56	[35.77 , 36.40]	0.63
PREP	35.34	–	36.38	–
Cómputo Distrital	35.31	–	35.89	–

Fuente: elaboración propia.

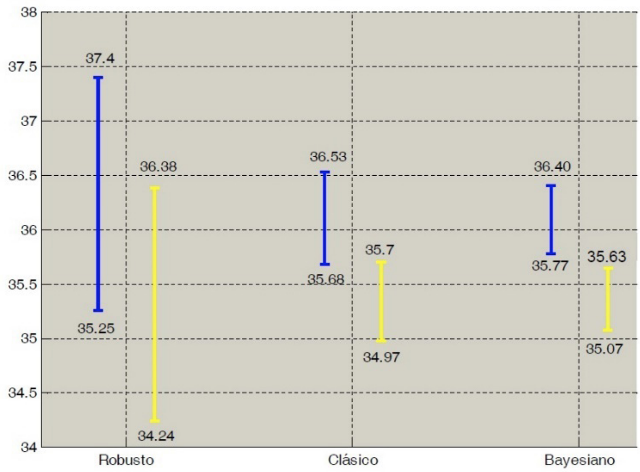
9 Las siglas CPBT corresponden a *Coalición por el Bien de Todos*. Las siglas PAN corresponden a Partido Acción Nacional.

10 PREP son las siglas del *Programa de Resultados Electorales Preliminares* que es un mecanismo de información que, al término de la jornada electoral, permite consultar, a través de internet, los resultados preliminares de las elecciones conforme los va recibiendo la autoridad electoral.

11 Fuente disponible en http://prep2006.ife.org.mx/PREP2006/prep2006.html. Consultado el 17 de enero de 2018.

12 Fuente disponible en http://portalanterior.ine.mx/documentos/Estadisticas2006/presidente/nac.html. Consultado el 17 de enero de 2018.

Figura vii. Intervalos estimados para Felipe Calderón Hinojosa (fch) en azul, y para Andrés Manuel López Obrador (amlo) en amarillo, bajo las metodologías denominadas Robusta, Clásica y Bayesiana



Fuente: IFE, 2006.

La *distancia de Hausdorff* (Munkres, 2002) para el caso de intervalos cerrados y acotados en el espacio métrico de los números reales bajo la métrica usual en dicho conjunto, resulta ser para un par de intervalos dados $I_1 = [a_1, b_1]$ e $I_2 = [a_2, b_2]$ como sigue:

$$d_H(I_1, I_2) = máx\{\,|a_1 - a_2|, |b_1 - b_2|\,\} \qquad (24)$$

En la Tabla vi se calcularon las distancias de Hausdorff (en puntos porcentuales) para los pares de intervalos estimados para los candidatos punteros bajo los tres métodos utilizados para el conteo rápido de 2006 (ver Tabla v), y se agregó, como simple referencia, la diferencia absoluta entre los puntos medios de dichos intervalos. Como se puede apreciar, bajo las tres metodologías estadísticas aplicadas, las distancias entre los pares de intervalos estimados para los dos candidatos punteros son en todos los casos superiores a 0.6 por ciento y, por tanto, de acuerdo con la *Regla 3* establecida por el propio Comité Técnico Asesor del Conteo Rápido, sí estuvieron en condiciones de identificar a un ganador de acuerdo con sus propias reglas.

Tabla vi: Distancias de Hausdorff y entre puntos medios de los intervalos (en puntos porcentuales) estimados la noche del 2 de julio de 2006 por los métodos denominados por el ctarcr como Robusto, Clásico y Bayesiano, para los dos candidatos punteros: Felipe Calderón Hinojosa (fch) y Andrés Manuel López Obrador (amlo)

Método	Distancia de Hausdorff	Diferencia entre puntos medios de los intervalos
Robusto	1.010	1.015
Clásico	0.710	0.770
Bayesiano	0.700	0.735

Fuente: elaboración propia.

Respecto a la *Regla 3*, Aparicio (2009) interpreta la "distancia" de 0.6 por ciento como *margen de error* al mencionar:

> [...] el margen de error del conteo rápido era de 0.6 por ciento pero como el resultado fue aún más cerrado (0.58%) no era posible declarar un ganador con los grados de confianza estadística comúnmente aceptados y que el mismo ife había anunciado previamente [...] Si bien el conteo rápido levantado en 7636 casillas la noche del 2 de julio no apuntaba a un claro ganador (es decir, fuera de los márgenes de error del instrumento), éste sí sugería una elección con un margen menor a 0.6 por ciento de los votos, al tiempo que daba a Felipe Calderón una mayor probabilidad de aventajar en el resultado final. Con este dato era claro que el resultado observado al cierre del prep la noche del 3 de julio, y que daba un margen de 1.04 por ciento a favor de Calderón, tenía que reducirse al llegar al cómputo distrital –tal como ocurrió: el cómputo distrital arrojó un margen de sólo 0.58 por ciento, el cual validó la estimación inicial del conteo rápido.

Consultando el informe entregado por el ctarcr (ife, 2006), en la página 8 se estableció lo siguiente:

> [...] el objetivo fundamental es obtener un diseño y tamaño de muestra que permitan estimar con un error aceptablemente pequeño, del 0.5% ...

y luego en la página 9:

> [...] el Comité evaluó el tamaño de muestra de 7500 casillas a la luz de los errores de estimación que arrojaría. Con ese tamaño de muestra se tiene un error de alrededor de 0.3% ...

De acuerdo con Eslava (2006), miembro de dicho CTARCR, con una muestra efectivamente recibida de 7236 casillas, el error observado estimado fue de 0.42 por ciento para el caso del PAN (candidato: Felipe Calderón Hinojosa) y de 0.36 por ciento para la CPBT (candidato: Andrés Manuel López Obrador), por lo que tampoco resulta aceptable la interpretación del 0.6 por ciento de la *Regla 3* como margen de error, tal cual se sugiere (Aparicio, 2009).

Mendoza y Nieto–Barajas (2016) presentan un enfoque bayesiano para las estimaciones del conteo rápido de la elección presidencial de México de 2006 justamente, y es importante aclarar, además, que Mendoza también formó parte del Comité Técnico Asesor para los Conteos Rápidos para dicha elección. El modelo bayesiano que publicaron, además de permitir las estimaciones por intervalo para las proporciones de votos que obtienen los candidatos, permite calcular la *probabilidad de triunfo*, y a las 10:15 p.m de la noche del día de la elección del 2 de julio de 2006 estimaron que la probabilidad de que el candidato del PAN (Felipe Calderón Hinojosa) fuese ganador era de ¡99.94 por ciento! Cálculo que contrasta fuertemente con la conclusión (Eslava, 2006) de que «estadísticamente no era posible distinguir con un alto grado de confiabilidad a un candidato ganador». Desafortunadamente la probabilidad de triunfo no se comunicó la noche de la elección, el comité asesor del conteo rápido simplemente decidió que no podía distinguir un ganador con base en un "acuerdo previo" sobre una determinada y arbitraria distancia interválica que nunca se documentó cómo se calcula ni qué justificación probabilística o estadística tiene. En palabras de Mendoza y Nieto–Barajas (2016):

> *Our Bayesian model produced disjoint 99% intervals, for the two leading parties, whose limits where really close to each other but separated. This fact is relevant because, non expert readers might think that the risk of a false winner call was large. This was not true and the evidence is provided precisely by the 22:15 dispersion diagram (bottom right panel in Fig. 4), where the probability of PAN being the winner is 0.9994. In fact, the marginal intervals could intersect and nonetheless the relevant information to call a winner is provided by the corresponding joint bivariate posterior distribution.*

Esto coincide con lo discutido en la sección anterior respecto a calcular la *probabilidad de triunfo* y que el asunto de que los intervalos de estimación se traslapen o no debería quedar fuera de toda consideración porque *no son probabilísticamente comparables*, y porque en un momento dado la probabilidad de triunfo puede resultar elevada a pesar de existir traslape entre dichos intervalos. Como ya se mencionó, el informe del CTARCR (IFE, 2006) no especificó cuáles fueron los niveles de confiabilidad finalmente utilizados con cada uno de los tres métodos de estimación, pero de Mendoza y Nieto–Barajas (2016) queda claro que al menos para el caso del método Bayesiano fue de 99 por ciento. Con esto último fue posible realizar un análisis similar al del *Contraejemplo 3* de la sección anterior (el código de programación se anexa en el Apéndice, así como un enlace para descarga del mismo, para fines de reproducibilidad) calculando bajo distintos niveles de dependencia tanto la probabilidad de triunfo de FCH (PAN) sobre AMLO (CPBT) como la probabilidad de que la diferencia entre ellos fuese menor a 0.6 por ciento. Los resultados se resumen en la Tabla VII, donde se aprecia que con probabilidades de triunfo muy elevadas en todos los casos, la probabilidad de que la diferencia fuese menor a 0.6 por ciento podía llegar a niveles hasta de 27 por ciento en un momento dado.

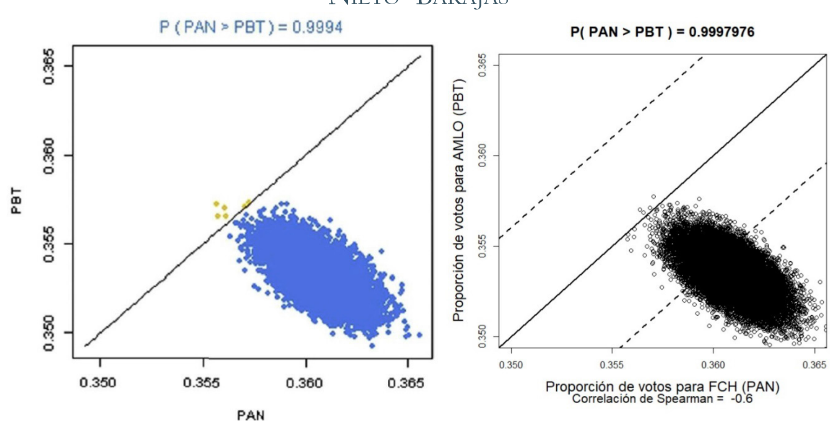
Tabla VII: Probabilidad de triunfo de FCH sobre AMLO y probabilidad de que la diferencia entre ambos fuese menor a 0.6% bajo distintos niveles de dependencia (correlación de Spearman y cópula Normal) y los intervalos de probabilidad 99% estimados bajo el método Bayesiano

Correlación de Spearman	P (FCH > AMLO) en %	P (| FCH – AMLO | < 0.6%) en %
–0.9	99.94	27.44
–0.6	99.98	25.76
–0.3	99.99+	23.52
0.0	99.99+	20.43
–0.3	99.99+	15.97
+0.6	99.99+	9.18
+0.9	99.99+	0.42

Fuente: elaboración propia.

Mendoza y Nieto–Barajas (2016) no incluyeron en su trabajo el grado de dependencia entre los dos candidatos punteros, pero en la Figura VIII se compara la gráfica de los autores donde simularon la distribución conjunta a posteriori de los candidatos punteros con la que estimaron la probabilidad de triunfo de FCH sobre AMLO (que a la vista implica una dependencia negativa) *versus* lo obtenido en la Tabla VII con una correlación de Spearman igual a –0.6 con cópula Normal[13] y los intervalos de probabilidad 99 por ciento estimados bajo el método Bayesiano. Esto nos da cierta idea de que aún con una probabilidad de triunfo de FCH sobre AMLO de casi 100 por ciento la probabilidad de que la diferencia en porcentaje de votos entre ambos candidatos fuese menor a 0.6 por ciento podía haber sido de aproximadamente 26 por ciento. En las gráficas de la Figura VIII, la parte inferior a la línea recta corresponde a escenarios donde FCH triunfa sobre AMLO, y la región encerrada entre las dos líneas punteadas corresponde al caso en que la diferencia en puntos porcentuales entre ambos candidatos punteros es menor a 0.6 por ciento.

Figura VIII. Izquierda: Simulaciones conjuntas a posteriori FCH/AMLO (PAN/PBT) en Mendoza y Nieto–Barajas (2016). Derecha: Simulaciones conjuntas con cópula Normal, correlación de Spearman –0.6 y utilizando los intervalos de probabilidad 99% estimados por Mendoza y Nieto–Barajas



Fuente: Mendoza y Nieto–Barajas, 2016.

13 Se utilizó una cópula *Normal* (o *Gaussiana*) porque el método Bayesiano utilizado en Mendoza y Nieto–Barajas (2016) se apoya en una distribución multivariada *Normal* –– *Wishart invertida*.

Si en lo anterior se repitieran los cálculos utilizando los intervalos estimados por el método Clásico (ver Tabla V) manteniendo cópula Normal con correlación de Spearman de −0.6 se obtendrían los resultados que se resumen en la Tabla VIII. Como puede apreciarse, la probabilidad de triunfo de FCH sobre AMLO es similarmente elevada con probabilidades no despreciables de que la diferencia entre los candidatos punteros fuese menor a 0.6 por ciento. Como el informe del Comité Técnico Asesor del conteo rápido no especificó el nivel de confianza utilizado para las estimaciones del método Clásico, se hizo el análisis con diversos niveles de confianza. Nótese que, a diferencia de los intervalos estimados por el método Bayesiano, los intervalos del método Clásico sí tienen traslape (ver Tabla V y Figura VII), y a pesar de ello la probabilidad de triunfo de FCH (PAN) es muy similar al caso Bayesiano, al igual que la probabilidad de que la diferencia entre candidatos fuese menor a 0.6 por ciento.

Tabla VIII: Probabilidad de triunfo de FCH sobre AMLO y probabilidad de que la diferencia entre ambos fuese menor a 0.6% con correlación de Spearman de −0.6, cópula Normal y los intervalos de probabilidad estimados bajo el método Clásico, suponiendo distintos niveles de confianza de dichos intervalos

Confianza del intervalo en %	P (FCH > AMLO) en %	P(| FCH − AMLO | < 0.6%) en %
95.0	98.30	31.91
99.0	99.74	26.82
99.5	99.88	25.03
99.9	99.99+	17.53

Fuente: elaboración propia.

En congruencia con el nivel de confianza utilizado para decidir el tamaño de muestra, el nivel de confianza de los intervalos estimados debiera ser también del 95 por ciento. En el caso de la elección presidencial de México en el año 2006, el informe final del comité técnico asesor para el conteo rápido (IFE, 2006) no especificó el nivel de confianza utilizado para los intervalos calculados bajo los tres métodos utilizados (denominados Robusto, Clásico y Bayesiano), pero de acuerdo a lo publicado al respecto por Mendoza y Nieto–Barajas (2016), al menos para el método Bayesiano fueron intervalos de probabilidad 99 por ciento (ver Tabla IX que corresponde a *Table 1* en Mendoza y Nieto–Barajas, 2016).

Tabla ix. Reproducción de *Table 1*

Table 1
Estimates of λ_j, $j=1,2,3$ and final results before refutations. Intervals are computed with 99% probability (confidence).

	PAN	APM	PBT
Final results	35.91	22.19	35.29
Bayesian	(35.73, 36.38)	(21.72, 22.22)	(35.05, 35.62)
Frequentist	(36.04, 36.19)	(21.91, 22.01)	(35.27, 35.39)

Fuente: Mendoza y Nieto–Barajas (2016).

Para comparar su método Bayesiano, Mendoza y Nieto–Barajas (2016) presentan los intervalos de confianza 99 por ciento que corresponderían al método Clásico (o Frecuentista). Lo que es de notarse es que las longitudes de los intervalos del método Clásico (o Frecuentista) resultaron de 0.15 (PAN), 0.10 (APM) y 0.12 (PBT), longitudes mucho menores que las de los intervalos reportados por el CTARCR (IFE, 2006) para el método Clásico (ver Tabla x) que fueron 0.85 (PAN), 0.60 (APM) y 0.73 (PBT), lo cual deja al descubierto que el nivel de confianza de los intervalos calculados con el método Clásico en el conteo rápido de 2006 fue aún superior al 99 por ciento, quizás algo así como 99.9 por ciento.

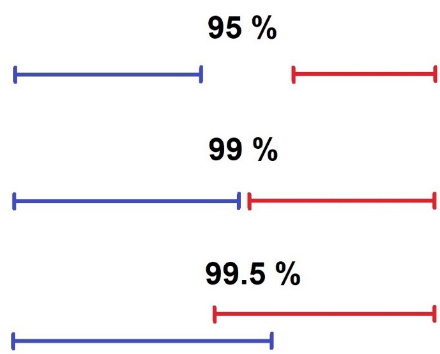
Tabla x.

	PAN	**APMx**	**CPBT**
PREP 3 de julio 10:36	**36.40**	**21.48**	**35.41**
ROBUSTO	**(35.25,37.4)**	**(20.85,22.7)**	**(34.24,36.38)**
CLÁSICO	**(35.68,36.53)**	**(21.66,22.26)**	**(34.97,35.7)**
BAYESIANO	**(35.77,36.40)**	**(21.72,22.24)**	**(35.07,35.63)**

Fuente: Informe del CTARCR (IFE, 2006).

La *longitud* de un intervalo de confianza es *función creciente* del nivel de confianza elegido, esto es, a mayor/menor nivel de confianza se obtiene una mayor/menor longitud del intervalo. En el informe del CTARCR (IFE, 2006) quedó asentado que dicho comité técnico acordó, entre otras cosas, que el nivel de confianza de los intervalos sería de *al menos* 95 por ciento. Pero al decir "al menos" dejó totalmente abierta la puerta a la discrecionalidad de quienes realizaron los cálculos de los intervalos, que ni siquiera utilizaron el mismo nivel de confiabilidad (99 por ciento para el método Bayesiano, 99.9 por ciento aproximadamente para el método Clásico). Y es que precisamente con discrecionalidad sobre el nivel de confianza a utilizar se puede lograr que los intervalos se separen o traslapen a placer, como se ilustra en la Figura IX.

Figura IX. Con la misma información muestral es posible lograr que intervalos estimados se separen o traslapen modificando el nivel de confianza para calcularlos.



Fuente: Elaboración propia.

Los intervalos estimados por el método denominado Robusto no son dignos de análisis alguno por la sencilla razón de que dicho método supone que la muestra ha sido seleccionada de acuerdo con un *esquema aleatorio simple*, lo cual resulta incongruente con el diseño de muestreo utilizado para dicho conteo rápido que fue *estratificado*. La idea de aplicar un *muestreo aleatorio estratificado* radica en el beneficio inferencial que se obtiene por establecer o escoger una estratificación tal que en cada estrato sea razonable suponer cierta *homogeneidad* en estas subpoblaciones, en contraste con la heterogeneidad que se tendría al obtener muestras directamente de la población total sin estratificar, en deterioro de la precisión de las inferencias que se desea realizar (Särndal *et al*., 1992). Es por ello que la *longitud* de los intervalos estimados por el método Robusto es considerablemente mayor que la de los métodos Bayesiano y Clásico (véase Figura VII). En el citado informe del CTARCR (IFE, 2006), respecto al método Robusto se afirma en la pág. 114:

> [...] si se logra tener una gran parte de la muestra prevista, los intervalos producidos deberán ser muy parecidos a los obtenidos con los métodos Bayesiano y clásico.

Como era de esperarse, no ocurrió así. A pesar de que se logró recabar 95.12 por ciento del total de la muestra prevista para el conteo rápido, 99.8 por ciento de los estratos y 100 por ciento de los distritos (IFE, 2006: 22) es evidente que los intervalos producidos por el método Robusto distan mucho de los obtenidos por los métodos Bayesiano y Clásico. En la Tabla XI se presenta una matriz de distancias de Hausdorff (24) entre los intervalos generados por los tres métodos para los dos candidatos punteros, correspondiendo la matriz triangular superior a los intervalos estimados para FCH (PAN) y la inferior lo correspondiente a AMLO (PBT). Mientras que las distancias de Hausdorff entre intervalos de los métodos Clásico y Bayesiano están en el rango 0.09 – 0.10 por ciento, la distancia del método Robusto respecto a los otros dos métodos oscila en 0.43 – 0.83 por ciento.

Tabla XI: Distancias de Hausdorff entre intervalos estimados bajos los métodos Robusto, Clásico y Bayesiano. Los valores de la matriz triangular superior corresponden a los intervalos estimados para FCH (PAN), y la inferior lo correspondiente a AMLO (PBT)

Distancia Hausdorff	Robusto	Clásico	Bayesiano
Robusto	–	0.43	0.52
Clásico	0.73	–	0.09
Bayesiano	0.83	0.10	–

Fuente: elaboración propia.

Conclusiones

Se demostró, tanto de forma teórica como mediante el análisis de un conteo rápido, que la *probabilidad de triunfo* electoral de un candidato puntero puede resultar muy cercana a 100 por ciento a pesar de que exista traslape de los intervalos estimados para la proporción de votos de los dos candidatos punteros. Declarar "empate técnico" porque hay traslape de intervalos no tiene fundamento probabilístico ni justificación estadística, y por tanto no es útil para interpretar encuestas o conteos rápidos electorales, ya que dichos intervalos se calculan de forma individual sin tomar en cuenta la *dependencia probabilística* entre candidatos, que resulta fundamental para contestar preguntas que involucran simultáneamente a dos o más de ellos; por ejemplo, que alguno triunfe sobre otro. En el caso particular de México, el *Reglamento de Elecciones* (INE, 2016) en su artículo 380, inciso 3, establece lo siguiente:

> Sea cual fuere la muestra recabada y los resultados obtenidos, el COTECORA[14] deberá presentar un reporte al Consejo General u órgano Superior de Dirección que corresponda, en el que indique, además, las condiciones bajo las cuales se obtuvieron los resultados, así como las conclusiones que de ellos puedan derivarse. Las estimaciones deberán presentarse en forma de intervalos de confianza para cada contendiente.

Esta redacción no establece que las conclusiones deban derivarse *exclusivamente* de los intervalos de confianza, sino de los *resultados obtenidos* –en un sentido más general– a partir de la muestra recabada, y dentro de ello cabe incluir el cálculo de la *probabilidad de triunfo* del candidato puntero con base en dicha información muestral, y no necesariamente a partir del criterio de que exista o no traslape de dichos intervalos, que se demostró no tiene sustento estadístico. Si, por ejemplo, se reportara que determinado candidato resultó con probabilidad 99 por ciento de ser triunfador de la elección, se entiende (y puede explicarse así) que existe una probabilidad de 1 por ciento de que no resulte finalmente así en el recuento total de votos (cómputo distrital).

La *probabilidad de triunfo* es, como se explicó en la introducción del presente artículo (Lindley, 2000), una medida del *grado de certidumbre/incertidumbre* sobre un evento de interés (el triunfo de un candidato en este caso) con base en la

14 COTECORA significa Comité Técnico para los Conteos Rápidos.

información disponible en un momento dado, y es así como podría explicarse a no expertos, porque es justamente lo que puede ofrecer la estadística con base en información parcial (muestral), ya que la certidumbre total solo es posible con acceso a la totalidad de la información.

Agradecimiento

El presente trabajo fue realizado con apoyo del programa PAPIIT – DGAPA – UNAM mediante el proyecto IN115817.

Referencias

Aparicio, Javier. 2009. "Análisis estadístico de la elección presidencial de 2006 ¿fraude o errores aleatorios?". En *Política y Gobierno (volumen Temático Elecciones en México)*, 2: 225–243.

Campos, Roy; Penna, Carlos. 2004. *¿"Empate técnico" o "too close to call"?* Disponible en www.consulta.mx. Consultado el 13 de enero de 2018.

Embrechts, Paul; McNeil, Alexander; Strauman, Daniel. 1999. "Correlation: pitfalls and alternatives". En *Risk Magazine*, 5: 69–71.

Eslava, Guillermina. 2006. En "Las elecciones de 2006, un análisis del conteo rápido". En *Ciencias*, 84: 30–37.

Frank, Maurice. 1979. "On the simultaneous associativity of F(x,y) and x+y–F(x,y)". En *Aequationes Mathematicae*, 19: 194–226.

Fréchet, Maurice. 1951. "Sur les tableaux de corrélation dont les marges sont données". En *Annales de l'Université de Lyon*, 14 (Sect. A Ser. 3): 53–77.

Hoeffding, Wassily. 1940. "Masstabinvariante Korrelationstheorie". En *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5: 179–223.

Hofert, Marius; Kojadinovic, Ivan; Maechler, Martin; Yan, Jun. 2017. *copula: Multivariate Dependence with Copulas. R package version 0.999–18*. Disponible en https://CRAN.R–project.org/package=copula. Consultado el 20 de abril de 2018.

IFE. Instituto Federal Electoral. 2006. *Informe sobre las actividades del Comité Técnico Asesor para la Realización de Conteos Rápidos*. Disponible en www.portalanterior.ine.mx. Consultado el 17 de diciembre de 2017.

INE. Instituto Nacional Electoral. 2016. *Reglamento de Elecciones (México)*. Disponible en www.ine.mx. Consultado el 12 de enero de 2018.

Lindley, Dennis. 2000. "The Philosophy of Statistics". En *Journal of the Royal Statistical Society. Series d (The Statistician)*, 49: 293–337.

Mendoza, Manuel; Nieto–Barajas, Luis Enrique. 2016. "Quick counts in the Mexican presidential elections: A Bayesian approach". En *Electoral Studies*, 43: 124–132.

Munkres, James. 2002. *Topología*. Madrid: Prentice Hall.

Nelsen, Roger. 2006. *An Introduction to Copulas.* Nueva York: Springer.

R Core Team. 2017. R: *A language and environment for statistical computing.* Vienna, Austria: r Foundation for Statistical Computing. Disponible en https://www.R–project.org/. Consultado el 20 de abril de 2018.

Särndal, Carl–Erik; Swensson, Bengt; Wretman, Jan. 1992. *Model Assisted Survey Sampling*. Nueva York: Springer.

Sklar, A. 1959. "Fonctions de répartition à n dimensions et leurs marges". En *Publications de l'Institut de statistique de l'Université de Paris*, 8: 229–231.

Apéndice

Código en lenguaje de programación r para los cálculos del Contraejemplo 3, también descargable de https://goo.gl/z5FYNg

```
# Construccio'n de intervalo para X
gama <- 0.05 # la probabilidad del intervalo sera' 1-gama
x.sup <- 0.38 # extremo superior del intervalo
epsilon.X <- 0.03 # margen de error
x.inf <- x.sup - 2*epsilon.X # extremo inferior del intervalo
# Estimacio'n de para'metros alfa y beta:
hX <- function(ab) (qbeta(1 - gama/2, ab[1], ab[2]) - 2*x.sup)^2 +
        (qbeta(gama/2, ab[1], ab[2]) - 2*x.inf)^2
(hX.optim <- nlm(hX, c(8, 1)))
# Comprobando probabilidad del intervalo:
pbeta(2*x.sup, hX.optim$estimate[1], hX.optim$estimate[2]) -
pbeta(2*x.inf, hX.optim$estimate[1], hX.optim$estimate[2])
```

```
# Graficar densidad de X y su intervalo:
x <- seq(0, 1/2, length.out = 1000)
plot(x, 2*dbeta(2*x, hX.optim$estimate[1], hX.optim$estimate[2]),
type = "l", xlab = "proporcio'n de votos", ylab = "densidad de probabilidades",
xlim = c(0.2, 0.45), col = "blue", lwd = 3, cex.lab = 1.5, cex.sub = 1.3)
rug(c(x.inf, x.sup), lwd = 3, col = "blue")
cat("Intervalo para X: [", x.inf, ",", x.sup, "]", "\n")
# Construccio'n de intervalo para Y
epsilon.Y <- 0.03 # margen de error
tau <- 0.005 # traslape
y.sup <- x.inf + tau # extremo superior del intervalo
y.inf <- y.sup - 2*epsilon.Y # extremo inferior del intervalo
# Estimacio'n de para'metros alfa y beta:
hY <- function(ab) (qbeta(1 - gama/2, ab[1], ab[2]) - 2*y.sup)^2 +
        (qbeta(gama/2, ab[1], ab[2]) - 2*y.inf)^2
(hY.optim <- nlm(hY, c(8, 1)))
# Comprobando probabilidad del intervalo:
pbeta(2*y.sup, hY.optim$estimate[1], hY.optim$estimate[2]) -
pbeta(2*y.inf, hY.optim$estimate[1], hY.optim$estimate[2])
# Graficar densidad de Y y su intervalo:
lines(x, 2*dbeta(2*x, hY.optim$estimate[1], hY.optim$estimate[2]), col = "red", lwd = 3)
rug(c(y.inf, y.sup), lwd = 3, col = "red")
title("Intervalos traslapados", cex.main = 1.5)
cat("Intervalo para Y: [", y.inf, ",", y.sup, "]", "\n")
# Dependencia para (X,Y): Co'pula Frank
library(copula) # Paquete <copula> previamente instalado
rho <- -0.5 # valor de correlacio'n de Spearman (en [-1,+1])
(th <- iRho(frankCopula(), rho)) # calculando para'metro
Copula.Frank <- frankCopula(th) # definiendo la co'pula
n <- 1000000 # tamanio de la simulacio'n
set.seed(666) # especificando semilla para reproducibilidad
uv.sim <- rCopula(n, Copula.Frank) # simulando pseudo-observaciones
# Agregando dependencia a (X,Y):
xy.sim <- matrix(nrow = n, ncol = 2)
xy.sim[ , 1] <- (1/2)*qbeta(uv.sim[ , 1], hX.optim$estimate[1], hX.optim$estimate[2])
xy.sim[ , 2] <- (1/2)*qbeta(uv.sim[ , 2], hY.optim$estimate[1], hY.optim$estimate[2])
cor(xy.sim, method = "spearman")[1, 2] # comprobando correlacio'n obtenida
# Comprobando estimaciones puntuales y por intervalo:
quantile(xy.sim[ , 1], probs = c(gama/2, 1/2, 1-gama/2))
```

```
quantile(xy.sim[ , 2], probs = c(gama/2, 1/2, 1−gama/2))
# Estimando probabilidad de triunfo de X sobre Y:
(P.triunfoX <− mean(xy.sim[ , 1] > xy.sim[ , 2]))
# Graficar parte de las simulaciones de (X,Y)
dev.new()
plot(xy.sim[1:3000, ], xlab = "Proporción de votos para X",
   ylab = "Proporcio'n de votos para Y", main = "Candidatos punteros",
  sub = paste("Correlacio'n de Spearman = ", rho), cex.lab = 1.5,
 cex.main = 1.5, cex.sub = 1.3, xlim = c(.3,.4), ylim = c(.23,.36))
text(0.325, 0.235, paste("P(X > Y) =", P.triunfoX), cex = 2)
abline(a = 0, b = 1, lwd = 2)
```

Código en lenguaje de programación r para los cálculos sobre la elección presidencial de 2006, también descargable de https://goo.gl/TKoctG

```
# Construccio'n de intervalo para FCH (PAN)
gama <− 0.01 # la probabilidad del intervalo sera' 1−gama
x.inf <− 0.3577 # extremo inferior del intervalo
x.sup <− 0.3640 # extremo superior del intervalo
epsilon.X <− (x.sup − x.inf)/2 # margen de error
# Estimacio'n de para'metros alfa y beta:
hX <− function(ab) (qbeta(1 − gama/2, ab[1], ab[2]) − 2*x.sup)^2 +
          (qbeta(gama/2, ab[1], ab[2]) − 2*x.inf)^2
(hX.optim <− nlm(hX, c(18, 18)))
# Comprobando probabilidad del intervalo:
pbeta(2*x.sup, hX.optim$estimate[1], hX.optim$estimate[2]) −
pbeta(2*x.inf, hX.optim$estimate[1], hX.optim$estimate[2])
# Graficar densidad de FCH y su intervalo:
x <− seq(0, 1/2, length.out = 10000)
plot(x, 2*dbeta(2*x, hX.optim$estimate[1], hX.optim$estimate[2]),
    type = "l", xlab = "proporcio'n de votos", ylab = "densidad de probabilidades",
    xlim = c(0.345, 0.365), col = "blue", lwd = 3, cex.lab = 1.5, cex.sub = 1.3,
    ylim = c(0, 400))
rug(c(x.inf, x.sup), lwd = 3, col = "blue")
cat("Intervalo FCH: [", x.inf, ",", x.sup, "]", "\n")
# Construccio'n de intervalo para AMLO (PBT)
y.inf <− 0.3507 # extremo inferior del intervalo
y.sup <− 0.3563 # extremo superior del intervalo
epsilon.Y <− (x.sup − x.inf)/2 # margen de error
# Estimacio'n de para'metros alfa y beta:
```

```
hY <– function(ab) (qbeta(1 – gama/2, ab[1], ab[2]) – 2*y.sup)^2 +
    (qbeta(gama/2, ab[1], ab[2]) – 2*y.inf)^2
(hY.optim <– nlm(hY, c(18, 18)))
# Comprobando probabilidad del intervalo:
pbeta(2*y.sup, hY.optim$estimate[1], hY.optim$estimate[2]) –
pbeta(2*y.inf, hY.optim$estimate[1], hY.optim$estimate[2])
# Graficar densidad de PBT y su intervalo:
lines(x, 2*dbeta(2*x, hY.optim$estimate[1], hY.optim$estimate[2]), col = “red”, lwd = 3)
rug(c(y.inf, y.sup), lwd = 3, col = “red”)
title(“AMLO (rojo) versus FCH (azul)”, cex.main = 1.5)
cat(“Intervalo para AMLO: [“, y.inf, “,”, y.sup, “]”, “\n”)
# Dependencia para (FCH,AMLO): Co’pula Normal
library(copula) # Paquete <copula> previamente instalado
rho <– –0.6 # valor de correlacio’n de Spearman (en [–1,+1])
(th <– iRho(normalCopula(), rho)) # calculando para’metro
Copula.Normal <– normalCopula(th) # definiendo la co’pula
n <– 10000000 # tamanio de la simulacio’n
set.seed(777) # especificando semilla para reproducibilidad
uv.sim <– rCopula(n, Copula.Normal) # simulando pseudo–observaciones
# Agregando dependencia a (X,Y):
xy.sim <– matrix(nrow = n, ncol = 2)
xy.sim[ , 1] <– (1/2)*qbeta(uv.sim[ , 1], hX.optim$estimate[1], hX.optim$estimate[2])
xy.sim[ , 2] <– (1/2)*qbeta(uv.sim[ , 2], hY.optim$estimate[1], hY.optim$estimate[2])
cor(xy.sim, method = “spearman”)[1, 2] # comprobando correlacio’n obtenida
# Comprobando estimaciones puntuales y por intervalo:
quantile(xy.sim[ , 1], probs = c(gama/2, 1/2, 1–gama/2))
quantile(xy.sim[ , 2], probs = c(gama/2, 1/2, 1–gama/2))
# Estimando probabilidad de triunfo de X sobre Y:
(P.triunfoX <– mean(xy.sim[ , 1] > xy.sim[ , 2]))
# Estimando probabilidad de diferencia menor a 0.6%
(P.diferencia <– mean(abs(xy.sim[ , 1] – xy.sim[ , 2]) < 0.006))
# Graficar parte de las simulaciones de (X,Y)
dev.new()
plot(xy.sim[1:30000, ], xlab = “Proporcio’n de votos para FCH (PAN)”,
    ylab = “Proporcio’n de votos para AMLO (PBT)”, main = paste(“P( PAN > PBT ) =”,
P.triunfoX),
    sub = paste(“Correlacio’n de Spearman = “, rho), cex.lab = 1.5,
    cex.main = 1.5, cex.sub = 1.3, xlim = c(.35,.365), ylim = c(.35,.365))
abline(a = 0, b = 1, lwd = 2)
abline(a = –0.006, b = 1, lwd = 2, lty = “dashed”)
abline(a = 0.006, b = 1, lwd = 2, lty = “dashed”)
```