CONTEO RÁPIDO ELECTORAL CON DEPENDENCIAS PROBABILÍSTICAS
Electoral quick count under probabilistic dependence

Resumen

Un conteo rápido electoral es un procedimiento estadístico cuyo principal objetivo es obtener una muestra representativa del total de casillas electorales, y medir la incertidumbre respecto al resultado final, antes de los cómputos distritales. Por lo general, se prefiere un muestreo aleatorio estratificado para reducir la variabilidad de las estimaciones. El presente trabajo demuestra que la dependencia entre estratos y entre candidatos debe ser tomada en cuenta para las inferencias estadísticas que se realicen. Se propone un modelo basado en funciones cópula, y se aplica a datos de las elecciones presidenciales en México de los años 2006, 2012 y 2018.

Palabras clave: conteo rápido, dependencia probabilística, función cópula.

Abstract

An electoral quick count is a statistical procedure whose main objective is to obtain a representative sample of all the polling stations in a certain election, and to measure the uncertainty about final results, before the total count of votes. A stratified sampling design is commonly preferred to reduce estimation variability. The present work shows that dependence among strata and among candidates should be taken into consideration for statistical inferences therein, and a copula-based model is proposed and applied to Mexico's presidential elections in the years 2006, 2012, and 2018.

Keywords: quick count, probabilistic dependence, copula function.

Introducción

En procesos electorales donde habrá de elegirse una persona para un determinado cargo (presidente, gobernador o presidente municipal, por ejemplo) por medio del voto directo de los ciudadanos con derecho a voto, es de interés particular medir la incertidumbre respecto al resultado, antes de los cómputos distritales, y esto es posible lograrlo por medio del procedimiento estadístico denominado *conteo rápido*, alimentado por medio de una *muestra* de resultados de casillas electorales la misma noche de la jornada electoral. El *Reglamento de Elecciones* del Instituto Nacional Electoral de México (INE, 2016) define y describe dicho procedimiento de la siguiente manera:

“Artículo 356. 1. Los conteos rápidos son el procedimiento estadístico diseñado con la finalidad de estimar con oportunidad las tendencias de los resultados finales de una elección, a partir de una muestra probabilística de resultados de actas de escrutinio y cómputo de las casillas electorales, cuyo tamaño y composición se establecen previamente, de acuerdo a un esquema de selección específico de una elección determinada, y cuyas conclusiones se presentan la noche de la jornada electoral...”

En este tipo de elecciones las principales preguntas de interés que pretende contestar un conteo rápido, antes de conocer el resultado de los cómputos distritales, son las siguientes:

1. ¿Qué porcentaje de votación obtendrá cada candidato?
2. ¿Quién resultará ganador?

Para contestar la primera pregunta, la ciencia estadística cuenta con herramientas para realizar *estimaciones puntuales y por intervalo* de los porcentajes de votación obtenidos por los candidatos participantes, con base en una muestra del total de casillas. Por lo general, se recurre un *muestreo aleatorio estratificado*, para reducir la variabilidad de las estimaciones.

Respecto a la segunda pregunta, una forma de contestarla es mediante una estimación de la *probabilidad de victoria* de cada candidato, ya que con información parcial (muestra de casillas) no es posible tener total certidumbre, ésta solo es posible hasta que se concluyen los cómputos distritales varios días después (totalidad de casillas). En el caso de México, dicha probabilidad no ha sido reportada en los conteos rápidos oficiales, y en ocasiones se pretende distinguir un posible candidato ganador comparando entre sí los intervalos estimados para los dos candidatos punteros; en caso de que se traslapen los intervalos lo llaman "empate técnico" y se dice que no es posible distinguir un posible ganador. Al respecto, Erdely (2018) ya ha demostrado que dicho criterio de "empate técnico" carece de fundamento estadístico, y que es posible tener elevadas probabilidades de victoria a pesar de que exista traslape de los intervalos estimados.

En las siguientes secciones, después de introducir notación apropiada y el muestreo aleatorio estratificado, se demuestra la necesidad de considerar en las estimaciones por intervalo la *dependencia probabilística* entre estratos, y también la necesidad de considerar la dependencia probabilística entre candidatos para el caso de la probabilidad de victoria del candidato puntero. Se propone además un modelo de estimación basado en *funciones cópula*, mismo que se aplica a los datos de las elecciones presidenciales en México en los años 2006, 2012 y 2018, y cuyos resultados se contrastan con los que se obtienen mediante el modelo propuesto por Mendoza y Nieto Barajas (2016).

Muestreo aleatorio estratificado

Se adoptará y/o adaptará la notación utilizada en Mendoza y Nieto-Barajas (2016), para facilitar la comparación del modelo que se propone versus el de dichos autores. El *marco muestral* a considerar consiste en K casillas electorales para un total de n posibles votantes (tamaño de la lista nominal de electores) que pueden expresar su decisión electoral en alguna de las siguientes formas:

- votar por alguno de los candidatos registrados;
- votar por algún candidato no registrado;
- anular su voto;
- abstenerse de acudir a votar.

Las posibilidades anteriores se denotarán mediante las categorías $1, 2, \ldots, J$ donde se reservará la categoría $J-2$ para candidatos no registrados, $J-1$ para votos anulados y J para las abstenciones, de modo que las categorías $1, \ldots, J-3$ sean únicamente para candidatos registrados, implicando con ello $J \geq 5$ (al menos dos candidatos registrados). En lo subsecuente, cuando se haga referencia a un candidato $j \in \{1, \ldots, J\}$ por simplicidad deberá entenderse en un sentido más amplio, esto es considerando también como "candidatos" a las categorías de: candidatos no registrados, votos nulos y abstenciones.

Se considera el caso en que el total de K casillas electorales se distribuyen en N subconjuntos disjuntos o *estratos* (distritos electorales, por ejemplo). Sea K_i el número de casillas electorales en el estrato $i \in \{1, \ldots, N\}$ y sea n_i el número de posibles votantes en el estrato i de modo que $\sum_{i=1}^{N} n_i = n$. Después del cierre de casillas el día de la jornada electoral, en cada casilla los votos depositados en las urnas son clasificados en las categorías $\{1, \ldots, J-1\}$.

Se representará mediante las variables aleatorias $X_{i,j}^{(k)}$ el número de votos que obtendrá el candidato $j \in \{1, \ldots, J\}$ en el estrato $i \in \{1, \ldots, N\}$ y casilla $k \in \{1, \ldots, K_i\}$. Con lo anterior, se definen ahora las variables aleatorias $X_{i,j} = \sum_{k=1}^{K_i} X_{i,j}^{(k)}$ que representan el número total de votos del candidato j en el estrato i, y las variables aleatorias $\Theta_{i,j} = X_{i,j}/n_i$ que representan la proporción de votos a favor del candidato j en el estrato i. Por consecuencia se tiene que $\sum_{j=1}^{J} X_{i,j} = n_i$ y $\sum_{j=1}^{J} \Theta_{i,j} = 1$ para cada estrato i.

Ahora se procede a definir las variables aleatorias X_j como el número total de votos para el candidato j y, por lo tanto, $X_j = \sum_{i=1}^{N} X_{i,j} = \sum_{i=1}^{N} n_i \Theta_{i,j}$. Finalmente, la proporción total de votos para cada candidato j se obtiene mediante las variables aleatorias:

$$\Theta_j = \frac{X_j}{n} = \sum_{i=1}^{N} \frac{n_i}{n} \Theta_{i,j} , \quad j \in \{1, \dots, J\} \qquad (1)$$

de donde es inmediato verificar que $\sum_{j=1}^{J} X_j = n$ y, consecuentemente, $\sum_{j=1}^{J} \Theta_j = 1$. La proporción de *participación efectiva* en la elección se calcula como el complemento del abstencionismo, esto es $1 - \Theta_J$.

Se considera también el caso usual en el que, para realizar el conteo rápido electoral, se utiliza un *muestreo aleatorio estratificado*, con muestreo aleatorio simple dentro de cada estrato, proporcional al número de posibles votantes (lista nominal) en cada estrato. Esto es, si en el estrato i hay K_i casillas electorales, de entre ellas se obtiene una muestra aleatoria simple de $c_i < K_i$ casillas, de modo que $c = \sum_{i=1}^{N} c_i$ represente el número total de casillas en la muestra estratificada y $c_i \approx \frac{n_i}{n} c$.

El tamaño de muestra c depende de la *precisión* que se desee alcanzar con los estimadores, misma que puede ser controlada estableciendo un *margen de error* $\varepsilon > 0$ y un *nivel de probabilidad* $100\alpha\%$ (donde $0 < \alpha < 1$). Si θ_j es la proporción de votos (desconocida al momento del conteo rápido, por supuesto) que obtiene el candidato j, y θ_j^* es un estimador puntual de θ_j entonces el tamaño de muestra c podría escogerse de tal magnitud que:

$$P\left(\left|\theta_j - \theta_j^*\right| \leq \varepsilon \right) = \alpha , \qquad (2)$$

pero para cada j la distribución de probabilidad del estimador puntual θ_j^* depende de parámetros poblacionales desconocidos (los que se pretende estimar, justamente), por lo que se puede recurrir a datos de una elección anterior similar, y simular muestras a partir de ello, pero aun procediendo de esta forma, para cada j normalmente se obtienen distintos valores de c que cumplen con la condición (2), por lo que se toma de entre ellos el mayor valor de c que permita a las categorías $j \in \{1, \dots, J-1\}$ que (2) se cumpla con un nivel de probabilidad de al menos $100\alpha\%$, esto es $P\left(\left|\theta_j - \theta_j^*\right| \leq \varepsilon \right) \geq \alpha$.

La desigualdad en (2) es equivalente a $\theta_j^* - \varepsilon \leq \theta_j \leq \theta_j^* + \varepsilon$, lo que permite la interpretación de que el valor desconocido θ_j pertenecerá al intervalo $[\, \theta_j^* - \varepsilon \,, \theta_j^* + \varepsilon \,]$ con (al menos) un nivel de probabilidad de $100\alpha\%$. Como consecuencia del Resultado 3.7.2 en Särndal *et al*. (1992) el siguiente es un estimador puntual para θ_j :

$$\theta_j^* = \frac{\sum_{i=1}^{N} K_i y_{i,j}}{\sum_{l=1}^{J} \sum_{i=1}^{N} K_i y_{i,l}} \quad , \quad j \in \{1, \dots, J\}\,, \tag{3}$$

donde $y_{i,j}$ es el número promedio de votos obtenidos por el candidato j en el estrato i.

Estimación por intervalo con estratos dependientes

La proporción total de votos a favor del candidato j en (1) es una combinación lineal convexa $\Theta_j = \sum_{i=1}^{N} \beta_i \Theta_{i,j}$ donde $\beta_i = n_i/n > 0$ y por lo tanto $\sum_{i=1}^{N} \beta_i = 1$. Definiendo $\mu_{i,j} = E(\Theta_{i,j})$ y $\sigma_{i,j}^2 = V(\Theta_{i,j})$, como consecuencia de la propiedad de linealidad de la esperanza es inmediato obtener:

$$\mu_j = E(\Theta_j) = \sum_{i=1}^{N} \beta_i \mu_{i,j} \quad , \quad j \in \{1, \dots, J\}\,, \tag{4}$$

de donde se desprende que la esperanza de Θ_j solo depende de las esperanzas individuales de las variables aleatorias $\{\Theta_{i,j} : i = 1, \dots, N\}$, sin necesidad de consideración alguna sobre la posible dependencia entre ellas. En contraste, en el cálculo de la varianza de Θ_j la posible dependencia no puede ignorarse ya que para su cálculo se requieren ciertas covarianzas:

$$\sigma_j^2 = V(\Theta_j) = \sum_{i=1}^{N} \beta_i^2 \sigma_{i,j}^2 + 2 \sum_{1 \leq i < r \leq N} \beta_i \beta_r Cov(\Theta_{i,j}, \Theta_{r,j}) \tag{5}$$

$$= \sum_{i=1}^{N} \sum_{r=1}^{N} \beta_i \beta_r Cov(\Theta_{i,j}, \Theta_{r,j}) \leq \sum_{i=1}^{N} \sum_{r=1}^{N} \beta_i \beta_r \sigma_{i,j} \sigma_{r,j} = \left(\sum_{i=1}^{N} \beta_i \sigma_{i,j} \right)^2 \tag{6}$$

donde (6) es consecuencia inmediata de la *Desigualdad de Cauchy-Schwarz*, véase por ejemplo Casella y Berger (2002), y proporciona el máximo valor posible para la varianza

σ_j^2. Sea $I_j(\alpha)$ un intervalo de probabilidad $0 < \alpha < 1$ para Θ_j , esto es, un intervalo tal que $P\left[\Theta_j \in I_j(\alpha)\right] = \alpha$. Sea z_α un valor tal que:

$$I_j(\alpha) = \left[\, \mu_j - z_\alpha \sigma_j \ ,\ \mu_j + z_\alpha \sigma_j \,\right] \qquad (7)$$

y lo análogo para $\Theta_{i,j}$ por medio de $I_{i,j}(\alpha) = \left[\, \mu_{i,j} - z_\alpha \sigma_{i,j} \,, \mu_j + z_\alpha \sigma_{i,j} \,\right]$. Si el intervalo $I_j^{\top}(\alpha)$ es un combinación lineal convexa de los intervalos $I_{i,j}(\alpha)$ con coeficientes $\beta_i = n_i/n$ entonces:

$$I_j^{\top}(\alpha) = \sum_{i=1}^{N} \beta_i \, I_{i,j}(\alpha) = \left[\, \sum_{i=1}^{N} \beta_i \mu_{i,j} - z_\alpha \sum_{i=1}^{N} \beta_i \sigma_{i,j} \ ,\ \sum_{i=1}^{N} \beta_i \mu_{i,j} + z_\alpha \sum_{i=1}^{N} \beta_i \sigma_{i,j} \,\right]$$

$$= \left[\, \mu_j - z_\alpha \sigma_j^{\top} \ ,\ \mu_j + z_\alpha \sigma_j^{\top} \,\right] \qquad (8)$$

donde la desviación estándar $\sigma_j^{\top} = \sum_{i=1}^{N} \beta_i \sigma_{i,j}$ es el máximo valor posible para σ_j de acuerdo con (6). En caso de que el supuesto de independencia entre estratos fuese correcta se tendría entonces que $Cov\left(\Theta_{i,j} \,, \Theta_{r,j}\right) = 0$ para todo $i \neq r$ y, como consecuencia de (5), la desviación estándar sería $\sigma_j^{\perp} = \sqrt{\sum_{i=1}^{N} \beta_i^2 \sigma_{i,j}^2}$, un valor menor que $\sigma_j^{\top}$ y que por lo tanto arroja un intervalo de menor longitud: $I_j^{\perp}(\alpha) = \left[\, \mu_j - z_\alpha \sigma_j^{\perp} \ ,\ \mu_j + z_\alpha \sigma_j^{\perp} \,\right] \subset I_j^{\top}(\alpha)$.

Pero si el supuesto de independencia entre estratos no es correcto, el intervalo $I_j^{\perp}(\alpha)$ tendría una longitud menor a la que le corresponde, y por tanto sería de probabilidad menor al nivel α que se desea. En otro extremo, utilizar el intervalo $I_j^{\top}(\alpha)$ suponiendo covarianzas máximas entre estratos típicamente arrojaría intervalos de mayor longitud a la que correspondería a un nivel de probabilidad deseado α, esto es: $I_j^{\perp}(\alpha) \subset I_j(\alpha) \subset I_j^{\top}(\alpha)$.

Lo ideal sería obtener los intervalos $I_j(\alpha)$ como en (7) mediante el valor de σ_j calculado por medio de (5), pero esto implicaría estimar $N(N-1)/2$ covarianzas que a su vez requerirían información que no está disponible en este caso particular (procesos electorales). Suponiendo que σ_j es un valor en el intervalo $[\, \sigma_j^{\perp} \,,\ \sigma_j^{\top} \,]$ entonces debe existir un valor $0 \leq \delta_j \leq 1$ tal que:

$$\sigma_j = \left(1 - \delta_j\right) \sigma_j^{\perp} + \delta_j \sigma_j^{\top} \qquad (9)$$

y en tal caso un intervalo de probabilidad α para Θ_j resulta como sigue:

$$I_j(\alpha) = \mu_j \pm z_\alpha\left[\left(1-\delta_j\right)\sigma_j^{\perp} + \delta_j\sigma_j^{\top}\right] = \left(1-\delta_j\right)\mu_j + \delta_j\mu_j \pm z_\alpha\left(1-\delta_j\right)\sigma_j^{\perp} \pm z_\alpha\delta_j\sigma_j^{\top}$$

$$= \left(1-\delta_j\right)I_j^{\perp}(\alpha) + \delta_j I_j^{\top}(\alpha)\,. \qquad (10)$$

En (10) se enfrenta ahora el problema de estimar δ_j para cada candidato $j \in \{1,\dots,J\}$. En este caso sí es factible estimar valores δ_j a partir de elecciones anteriores o similares por medio del siguiente:

Algoritmo 1

1. Estimar los intervalos $I_j^{\perp}(\alpha)$ e $I_j^{\top}(\alpha)$ para cada candidato $j \in \{1,\dots,J\}$.
2. Realizar un número elevado de simulaciones de muestras aleatorias estratificadas a partir de resultados de un cómputo distrital reciente.
3. Determinar cada δ_j tal que el intervalo (10) contenga el verdadero valor θ_j (del cómputo distrital) en $100\alpha\%$ de las simulaciones del paso anterior (probabilidad de cobertura α).

En resumen, el principal objetivo de esta sección fue demostrar que una incorrecta estimación de la (posible) dependencia entre estratos puede resultar en estimaciones por intervalo de menor o mayor longitud, que a su vez podrían tener una probabilidad de cobertura distinta a la deseada, y es por ello que algún tipo de ajuste debiera hacerse al respecto. En el caso particular de que los estratos sean distritos electorales, podría anticiparse de forma intuitiva que bien podría existir cierto grado de *dependencia positiva*, esto es que un mejor/peor desempeño de un candidato se deba posiblemente a un mejor/peor desempeño en varios distritos simultáneamente.

Un modelo basado en funciones cópula

La *probabilidad de victoria* para cada candidato registrado involucra cálculos por medio de una transformación del vector aleatorio $\left(\Theta_1,\dots,\Theta_J\right)$. Si W_j representa el evento de que el candidato registrado $j \in \{1,\dots,J-3\}$ triunfe en la elección, entonces:

$$P\left(W_j\right) = P\left(\text{máx}\left\{\Theta_1,\dots,\Theta_{J-3}\right\} = \Theta_j\right), \qquad (11)$$

donde $0 < \Theta_j < 1$ y $\sum_{j=1}^{J}\Theta_j$ como consecuencia de (1), lo que a su vez implica que $\Theta_J = 1 - \sum_{j=1}^{J-1}\Theta_j$ y, por lo tanto, basta conocer la función de distribución conjunta de probabilidades del vector aleatorio $(\mathit{\Theta}_1, \dots, \mathit{\Theta}_{J-1})$, cuyo soporte es un conjunto S conocido como *símplex*, en este caso de dimensión $J-1$:

$$S = \{(\theta_1, \dots, \theta_{J-1}) : \theta_1 > 0, \dots, \theta_{J-1} > 0; \theta_1 + \cdots + \theta_{J-1} < 1\}. \tag{12}$$

Si $\mathit{\Theta}_j$ es una variable aleatoria continua, con función de distribución marginal F_j y soporte en el intervalo abierto $]0,1[$, entonces como consecuencia del *Teorema de Sklar* (1959), véase Nelsen (2006), existe una única relación funcional C (conocida como *función cópula*) entre la función de distribución conjunta de probabilidades H del vector aleatorio $(\mathit{\Theta}_1, \dots, \mathit{\Theta}_{J-1})$ y sus funciones de distribución marginales, esto es:

$$H(\theta_1, \dots, \theta_{J-1}) = C\left(F_1(\theta_1), \dots, F_{J-1}(\theta_{J-1})\right). \tag{13}$$

Para el objetivo que interesa, es suficiente con estimar las funciones de distribución conjunta bivariadas (por pares de candidatos):

$$H_{j,l}(x, y) = C_{j,l}\left(F_j(x), F_l(y)\right), \quad j \neq l, \tag{14}$$

y calcular las probabilidades $P(\mathit{\Theta}_j > \mathit{\Theta}_l)$ para $j \neq l$ con $j, l \in \{1, \dots, J-3\}$. Para lo anterior, es necesario estimar las funciones de distribución marginales F_j y las funciones cópula bivariadas $C_{j,l}$ con base en la muestra aleatoria estratificada ya descrita en una sección anterior.

Para cada candidato j en un estrato i habrá información de c_i casillas electorales de la muestra estratificada, con valores observados $\left\{x_{i,j}^{(1)}, \dots, x_{i,j}^{(c_i)}\right\}$, a partir de los cuales es posible calcular las proporciones de votos obtenidas en cada casilla electoral de dicha muestra:

$$\theta_{i,j}^{(k)} = \frac{x_{i,j}^{(k)}}{n_i^{(k)}}, \qquad k \in \{1, \dots, c_i\}, \tag{15}$$

donde $n_i^{(k)}$ es el número de posibles votantes (lista nominal) en la casilla k. A partir de las proporciones obtenidas en (15) es posible estimar (de forma paramétrica o no paramétrica) cada $F_{i,j}$, esto es la función de distribución marginal de $\mathit{\Theta}_{i,j}$, y a partir de esto último calcular $I_{i,j}(\alpha)$, $\mu_{i,j}$ y $\sigma_{i,j}^2$. Más aún, como consecuencia de (4), es posible estimar directamente para cada candidato j:

$$\mu_j = E(\Theta_j) = \sum_{i=1}^{N} \frac{n_i}{n} \mu_{i,j} \ , \tag{16}$$

pero para los intervalos $I_j(\alpha)$ y la probabilidad de victoria se requiere calcular varianzas y covarianzas asociadas a la dependencia entre estratos y entre candidatos.

Como consecuencia de lo discutido en la sección anterior, bajo los valores cero y máximo de las covarianzas $Cov(\Theta_{i,j}\ , \Theta_{r,j})$ entre pares de estratos $(i \neq r)$, las varianzas para cada candidato serían $\sigma_j^{\perp}$ y $\sigma_j^{\top}$, respectivamente:

$$\sigma_j^{\perp} \ = \ \sqrt{\sum\nolimits_{i=1}^{N} \left(\frac{n_i}{n}\right)^2 \sigma_{i,j}^2} \ < \ \sum\nolimits_{i=1}^{N} \frac{n_i}{n} \sigma_{i,j} \ = \ \sigma_j^{\top} \ . \tag{17}$$

Por medio de (17) y de los valores δ_j obtenidos mediante el Algoritmo 1 de la sección anterior es posible calcular la desviación estándar para cada candidato:

$$\sigma_j = (1 - \delta_j)\sigma_j^{\perp} + \delta_j^{\top} \ . \tag{18}$$

Las funciones de distribución marginales F_j pueden estimarse de forma paramétrica como se explica a continuación. Considérese a $X_{i,j}^{(k)}$ como una variable aleatoria *Binomial* con parámetro conocido $n_i^{(k)}$ y parámetro desconocido $\theta_{i,j}$. Entonces el número total de votos para el candidato j en el estrato i en una muestra de tamaño c_i se calcula mediante $X_{i,j} = \sum_{k=1}^{c_i} X_{i,j}^{(k)}$, que también resulta ser una variable aleatoria Binomial pero con parámetro conocido $\sum_{k=1}^{c_i} n_i^{(k)}$ y parámetro desconocido $\theta_{i,j}$. Dados los valores observados $\left\{x_{i,j}^{(1)}, \dots, x_{i,j}^{(c_i)}\right\}$ y utilizando estimación bayesiana con una distribución *Beta* a priori no informativa, véase Bernardo y Smith (1994) por ejemplo, se obtiene para $\theta_{i,j}$ una distribución a posteriori Beta pero con parámetros:

$$\alpha_{i,j} \ = \ \frac{1}{2} + \sum_{k=1}^{c_i} x_{i,j}^{(k)} \ , \qquad \beta_{i,j} \ = \ \frac{1}{2} \ + \sum_{k=1}^{c_i} [n_i^{(k)} - x_{i,j}^{(k)}] \tag{19}$$

lo cual especifica las distribuciones marginales $F_{i,j}$, y con ello medias y varianzas:

$$\mu_{i,j} \;=\; \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}} \;=\; \frac{\frac{1}{2}+\sum_{k=1}^{c_i} x_{i,j}^{(k)}}{1+\sum_{k=1}^{c_i} n_i^{(k)}} \;, \quad \sigma_{i,j}^2 \;=\; \frac{\alpha_{i,j}\beta_{i,j}}{\left(\alpha_{i,j}+\beta_{i,j}\right)^2(\alpha_{i,j}+\beta_{i,j}+1)} \;. \qquad (20)$$

Finalmente, cada distribución marginal F_j puede aproximarse por medio de una distribución Beta con media μ_j calculada como en (16) y varianza σ_j^2 a partir de (18), y ya con ello obtener los intervalos $I_j(\alpha)$ es inmediato.

Para calcular las probabilidades $P\left(\Theta_j > \Theta_l\right)$ las distribuciones marginales F_j anteriores son necesarias, mas no suficientes para tal propósito ya que de acuerdo con (14) se requiere estimar las funciones cópula subyacentes. Las covarianzas por pares de candidatos $j \neq l$ se obtienen mediante:

$$\begin{aligned} Cov\left(\Theta_j\,,\Theta_l\right) \;&=\; Cov\left(\sum_{i=1}^{N}\frac{n_i}{n}\,\Theta_{i,j}\;,\sum_{r=1}^{N}\frac{n_r}{n}\,\Theta_{r,l}\right) \\ &=\; \frac{1}{n^2}\sum_{i=1}^{N}\sum_{r=1}^{N} n_i n_r Cov\left(\Theta_{i,j}\,,\Theta_{r,l}\right)\,, \end{aligned} \qquad (21)$$

$$\begin{aligned} Cov\left(\Theta_j\,,\Theta_l\right) \;&=\; \frac{1}{n^2}\sum_{i=1}^{N} n_i^2 Cov(\Theta_{i,j}\,,\Theta_{i,l}) \\ &\quad+\; \frac{1}{n^2}\sum\sum_{i\neq r}\; n_i n_r Cov(\Theta_{i,j}\,,\Theta_{r,l})\,, \end{aligned} \qquad (22)$$

donde las covarianzas entre candidatos dentro de un mismo estrato $Cov(\Theta_{i,j}\,,\Theta_{i,l})$ en (22) pueden estimarse a partir de los pares de proporciones observadas $\left\{\left(\theta_{i,j}^{(k)},\theta_{i,l}^{(k)}\right) : k = 1,\dots,c_i\right\}$ de (15), pero para las covarianzas entre candidatos en estratos distintos no hay forma de tener información a partir de la cual estimar, por lo que se hará el siguiente supuesto: $Cov(\Theta_{i,j}\,,\Theta_{r,l}) \approx Cov(\Theta_{i,j}\,,\Theta_{i,l})$ con lo cual sustituyendo en (21) se obtiene:

$$Cov\left(\Theta_j\,,\Theta_l\right) = \sum_{i=1}^{N}\frac{n_i}{n} Cov\left(\Theta_{i,j}\,,\Theta_{i,l}\right), \qquad (23)$$

y por lo tanto las correlaciones de Pearson por pares de candidatos resultan ser:

$$\rho_{j,l} \;=\; \frac{Cov(\Theta_j\,,\Theta_l)}{\sqrt{V\left(\Theta_j\right)V(\Theta_l)}} \;=\; \sum_{i=1}^{N}\frac{n_i}{n} corr(\Theta_{i,j}\,,\Theta_{i,l})\,. \qquad (24)$$

Una distribución de probabilidad bivariada en donde la dependencia queda expresada en términos del coeficiente de correlación de Pearson es la distribución *Normal* (o *Gaussiana*) bivariada. Sin embargo, en este caso particular dicha distribución presenta el inconveniente de que sus distribuciones marginales univariadas también tienen distribución Normal (con soporte de $-\infty$ a $+\infty$), y lo que se busca modelizar a nivel marginal son proporciones $0 < \Theta_j < 1$, así que dicho inconveniente puede solucionarse utilizando la función cópula Gaussiana con parámetro de correlación $-1 < \rho < +1$, véase por ejemplo Salvadori *et al.* (2007), y combinarla con las marginales F_j obtenidas anteriormente, aplicando el Teorema de Sklar (14). La función cópula Gaussiana carece de una fórmula explícita, por lo que las probabilidades $P(\Theta_j > \Theta_l)$ para $j \neq l$ con $j, l \in \{1, \ldots, J-3\}$ se estiman vía simulación probabilística, de acuerdo con el siguiente:

Algoritmo 2

1. Simular un número elevado m de observaciones bivariadas $(u_t\,, v_t)$ a partir de una función cópula Gaussiana con parámetro de correlación $\rho_{j,l}$ obtenido mediante (24), para $j, l \in \{1, \ldots, J-3\}$, $j < l$.
2. Calcular $\left\{\left(\theta_j^{*(t)}\,, \theta_l^{*(t)}\right) = \left(F_j^{-1}(u_t), F_l^{-1}(v_t)\right) :\ t = 1, \ldots, m\right\}$.
3. Estimar $P\left(\Theta_j > \Theta_l\right) \approx \frac{1}{m} \sum_{t=1}^{m} 1_{\{\theta_j^{*(t)} > \theta_l^{*(t)}\}}$.

La proporción de votos Θ_j es respecto al total de n posibles votantes (tamaño de la lista nominal de electores), pero usualmente la proporción de interés es aquella respecto a la votación efectivamente emitida en las urnas electorales (esto es, sin considerar abstenciones), la cual puede obtenerse mediante:

$$\Lambda_j = \frac{X_j}{\sum_{l=1}^{J-1} X_l} = \frac{\Theta_j}{\sum_{l=1}^{J-1} \Theta_l} = \frac{\Theta_j}{1 - \Theta_J} \ , \quad j \in \{1, \ldots, J-1\}\,, \qquad (25)$$

en donde $1 - \Theta_J$ es la proporción de participación efectiva en la elección. La transformación (25) no tiene impacto en las probabilidades de victoria ya que claramente $P\left(\Theta_j > \Theta_l\right) = P\left(\Lambda_j > \Lambda_l\right)$, pero sí lo tiene en las estimaciones puntuales y por intervalo para cada candidato: A partir de la distribución conjunta de probabilidades del vector aleatorio $(\Theta_j\,, 1 - \Theta_J)$, obtenida a partir de (14) para cada $j \in \{1, \ldots, J-1\}$, de forma análoga al Algoritmo 2 es necesario calcular en el tercer paso $\lambda_j^{*(t)} = \frac{\theta_j^{*(t)}}{1-\theta_J^{*(t)}}$, y a partir de estas simulaciones es posible estimar la distribución de probabilidad F_{Λ_j} de cada variable aleatoria Λ_j , y con ello obtener estimaciones puntual y por intervalo:

Algoritmo 3

1. Simular un número elevado m de observaciones bivariadas (u_t , v_t) a partir de una función cópula Gaussiana con parámetro de correlación $\rho_{j,J}$ obtenido mediante (24), para $j \in \{1, \dots , J-1\}$.
2. Calcular $\left\{\left(\theta_j^{*(t)} , \theta_J^{*(t)}\right) = \left(F_j^{-1}(u_t), F_J^{-1}(v_t)\right) :\ t = 1, \dots , m\right\}$.
3. Calcular $\left\{\lambda_j^{*(t)} = \frac{\theta_j^{*(t)}}{1-\theta_J^{*(t)}} :\ t = 1, \dots , m\right\}$.
4. Obtener estimaciones por intervalo a partir de los cuantiles empíricos de $\left\{\lambda_j^{*(t)} : t = 1, \dots , m\right\}$: $[a_j , b_j]$ tales que $F_{\Lambda_j}\left(b_j\right) - F_{\Lambda_j}\left(a_j\right) = \alpha$ y minimizando la longitud $b_j - a_j$ de dicho intervalo, para un nivel de probabilidad deseado $0 < \alpha < 1$.

En resumen, el principal objetivo de esta sección fue obtener estimaciones por intervalo $I_j(\alpha)$ para la proporción de votos que obtendría cada candidato, que tomen en consideración la dependencia entre estratos, así como proporcionar estimaciones de las probabilidades de victoria por pares $P\left(\Theta_j > \Theta_l\right)$, mediante la aplicación de funciones cópula para poder tomar en consideración tanto la dependencia entre estratos como entre candidatos. También se obtienen estimaciones puntuales y por intervalo para la proporción de votos sobre la votación efectivamente emitida, mismas que también requieren tomar en consideración dependencia entre estratos y entre candidatos.

Elecciones presidenciales en México en los años 2006, 2012 y 2018

El modelo propuesto en la sección anterior se aplicó *ex post* a las elecciones presidenciales de México en los años 2006, 2012 y 2018, utilizando información del Instituto Nacional Electoral (INE: 2006, 2012, 2018). En los conteos rápidos realizados por el INE, la base para la estratificación utilizada fueron los 300 distritos electorales federales del país, con un mayor refinamiento en el caso de las elecciones de 2006 y 2012 a más de 480 estratos considerando además la clasificación urbana/no-urbana de las casillas electorales, y con 350 estratos en el caso de la elección de 2018, sin refinamiento urbano/no-urbano, pero con más de 300 estratos debido a que en el caso de 10 (de 32) entidades federativas donde además hubo elección de gobernador, se utilizó la misma estratificación de los conteos rápidos locales (basada en los distritos electorales locales), por cuestiones de eficiencia en elecciones concurrentes. El estudio que se presenta a continuación, para efectos de una mayor comparabilidad, se utilizaron los 300 distritos electorales como estratos para las tres elecciones presidenciales citadas.

Los valores δ_j que se requieren en (9) y (10) reflejan el efecto de la dependencia entre estratos a la hora de obtener estimaciones por intervalo. Dichos valores pueden obtenerse *ex post* por medio del Algoritmo 1, véase la Tabla 1 donde los valores fueron estimados para los tres candidatos punteros en las tres elecciones en cuestión: PAN (*Partido Acción Nacional*, compitió sin alianzas en las elecciones de 2006 y 2012, y liderando una coalición de partidos políticos en 2018), PRI (*Partido Revolucionario Institucional*, liderando coaliciones de partidos políticos en las tres elecciones), y AMLO (*Andrés Manuel López Obrador*, candidato en 2006 y 2012 por una coalición de partidos políticos liderados por el PRD –*Partido de la Revolución Democrática*– y candidato en 2018 por el partido de reciente creación MORENA –*Movimiento de Regeneración Nacional*).

Aunque los tamaños de muestra planeados para realizar los conteos rápidos fueron similares para las tres elecciones, los tamaños de muestra efectivamente recibidos para cada elección fueron un tanto distintos: 7263, 6260 y 5254 casillas paras las elecciones de 2006, 2012 y 2018, respectivamente. En el presente estudio de simulación se utilizaron estos tamaños de muestra.

Tabla 1: Valores δ_j estimados por medio del Algoritmo 1 para una probabilidad de cobertura de 95% en las elecciones presidenciales de México en los años 2006, 2012 y 2018, generando 10 mil muestras estratificadas del tamaño obtenido en cada elección (para la proporción de votos respecto al tamaño de la lista nominal). Fuente: Elaboración propia con datos del INE (2006, 2012, 2018).

δ	2006	2012	2018
PAN	0.2238	0.1891	0.2366
PRI	0.1652	0.1507	0.2199
AMLO	0.1594	0.1595	0.1992

Las estimaciones obtenidas mediante el modelo propuesto (basado en funciones cópula) en la sección anterior se contrastan versus lo que se obtiene aplicando el modelo de Mendoza y Nieto-Barajas (2016), mismo que se resume a continuación:

$$X_{i,j}^{(k)} \mid \theta_{i,j}, \tau_{i,j} \sim Normal\left(n_i^{(k)}\theta_{i,j}, \frac{\tau_{i,j}}{n_i^{(k)}}\right) \qquad (26)$$

siendo $\frac{\tau_{i,j}}{n_i^{(k)}}$ el parámetro de precisión, como se estila bajo el enfoque bayesiano (esto es, el inverso de la varianza). Para cada candidato sus autores adoptan el supuesto de que el parámetro $\tau_{i,j}$ es constante dentro de un mismo estrato, y no relacionado con $\theta_{i,j}$. Más aún, consideran que la variable aleatoria $X_{i,j}^{(k)}$ es independiente de $X_{i,r}^{(k)}$ para todo $j \neq r$. Respecto a esto último, sus autores mencionan: «Esto es posiblemente un supuesto aún más fuerte, sin embargo, análisis previos con un modelo más complejo que consideró dependencia entre estas variables mostró que la dependencia era demasiado débil y podría entonces ser desestimada.» Argumentan, además, que la dependencia entre los parámetros de interés (λ_j) podría ser recuperada simulando valores a posteriori para $\theta_{i,j}$ y transformando dichos valores mediante las fórmulas (1) y (25).

Después de combinar una distribución a priori no informativa para $(\theta_{i,j}, \tau_{i,j})$ con información de la muestra estratificada recibida, dichos autores obtienen una distribución a posteriori condicional en los datos $\{x_{i,j}^{(k)}\}$, que es proporcional al producto de una distribución normal truncada para $\theta_{i,j}$, condicional en $\tau_{i,j}$, y una distribución gamma para $\tau_{i,j}$, esto es:

$$p\left(\theta_{i,j}, \tau_{i,j} \mid \left\{x_{i,j}^{(k)}\right\}\right) \propto Normal\left(\theta_{i,j} \;\middle|\; \frac{\sum_{k=1}^{c_i} x_{i,j}^{(k)}}{\sum_{k=1}^{c_i} n_i^{(k)}} \;, \tau_{i,j} \sum_{k=1}^{c_i} n_i^{(k)}\right) 1_{\{0 < \theta_{i,j} < 1\}}$$
$$\times\, Gamma\left(\tau_{i,j} \;\middle|\; \frac{c_i - 1}{2} \;,\; \frac{1}{2}\left\{\sum_{k=1}^{c_i} \frac{\left(x_{i,j}^{(k)}\right)^2}{n_i^{(k)}} - \frac{\left(\sum_{k=1}^{c_i} x_{i,j}^{(k)}\right)^2}{\sum_{k=1}^{c_i} n_i^{(k)}}\right\}\right) \qquad (27)$$

Mendoza y Nieto-Barajas (2016) aclaran que en (27) la distribución a posteriori para $\tau_{i,j}$ es *propia* sólo si el número de casillas c_i de la muestra correspondiente al estrato i es igual o mayor que 2. Sin embargo, es de aclararse que también se requiere que $x_{i,j} \neq 0$ ya que de no ser así entonces $x_{i,j}^{(k)} = 0$ para todo valor k y, con ello, el segundo parámetro en la distribución gamma de (27) sería igual a cero, lo cual resulta inadmisible para tal distribución. Las restricciones anteriores (típicamente no cruciales) no son requeridas bajo el modelo basado en funciones cópula.

Para cada una de las elecciones presidenciales analizadas (2006, 2012 y 2018), bajo el modelo que se propone se estimaron probabilidades de victoria por medio del Algoritmo 2, calculando en cada caso para el candidato puntero la probabilidad de victoria sobre el segundo lugar, y para los demás la probabilidad de triunfar sobre el candidato puntero. Se obtuvieron estimaciones por intervalo para las proporciones de votos (sobre la votación efectiva) obtenidas por los tres candidatos punteros en cada elección, por medio del Algoritmo 3, simulando 10 mil muestras estratificadas a partir de los cómputos distritales, y se estimó la probabilidad de cobertura de dichos intervalos por medio de la proporción de simulaciones en que el intervalo estimado contuvo al valor que pretendía estimar (el que resulta de los cómputos distritales). Estos resultados se contrastan versus los que se obtienen bajo las mismas simulaciones aplicando el modelo de Mendoza y Nieto-Barajas (2016). Todos los cálculos y simulaciones fueron realizados en el lenguaje de programación de R Core Team (2018), y en particular los cálculos que requieren funciones cópula mediante el paquete en R denominado «copula» de Hofert *et al.* (2017).

En la Tabla 2 se resumen los resultados obtenidos bajo ambos métodos: El promedio de los intervalos de probabilidad 95% obtenidos en las simulaciones, junto con la probabilidad de victoria estimada para cada candidato, como se describió anteriormente, y los resultados de los cómputos distritales. Todas las cantidades se presentan en puntos porcentuales. La última columna es el porcentaje de muestras estratificadas simuladas en que cada candidato obtuvo la mayor cantidad de votos de acuerdo con la fórmula (3), para ser comparada con las probabilidades de victoria estimadas por cada modelo. Bajo ambos modelos, el promedio de las estimaciones por intervalo resultan razonablemente centradas alrededor del valor que pretenden estimar de los cómputos distritales, pero en todos los casos las longitudes de dichos intervalos son mayores bajo el modelo que se propone que bajo el de Mendoza y Nieto-Barajas (2016).

Tabla 2: Promedio de 10 mil estimaciones puntuales y por intervalo para los porcentajes de votos sobre la votación efectiva, y estimaciones de probabilidad de victoria, para cada candidato y elección presidencial. Fuente: Elaboración propia con datos del INE (2006, 2012, 2018).

2006	**Modelo Mendoza - Nieto**					**Modelo con función cópula**					**2006**	
Candidato	Intervalo probabilidad 95%			Estimación	Probabilidad	Intervalo probabilidad 95%			Estimación	Probabilidad	**Cómputo**	% victoria
presidencial	inferior	superior	longitud	puntual	de victoria	inferior	superior	longitud	puntual	de victoria	**final**	muestreo
PAN	35.65	36.13	0.48	35.89	99.75	35.64	36.15	0.51	35.90	99.70	**35.89**	99.49
PRI	22.08	22.46	0.38	22.27	0.00	22.06	22.48	0.42	22.27	0.00	**22.26**	0.00
AMLO	35.09	35.52	0.43	35.30	0.25	35.08	35.53	0.45	35.31	0.30	**35.31**	0.51

2012	**Modelo Mendoza - Nieto**					**Modelo con función cópula**					**2012**	
Candidato	Intervalo probabilidad 95%			Estimación	Probabilidad	Intervalo probabilidad 95%			Estimación	Probabilidad	**Cómputo**	% victoria
presidencial	inferior	superior	longitud	puntual	de victoria	inferior	superior	longitud	puntual	de victoria	**final**	muestreo
PAN	25.20	25.62	0.42	25.41	0.00	25.20	25.63	0.43	25.41	0.00	**25.41**	0.00
PRI	38.01	38.43	0.42	38.22	100.00	37.99	38.45	0.46	38.22	100.00	**38.21**	100.00
AMLO	31.37	31.79	0.42	31.58	0.00	31.37	31.80	0.43	31.58	0.00	**31.59**	0.00

2018	**Modelo Mendoza - Nieto**					**Modelo con función cópula**					**2018**	
Candidato	Intervalo probabilidad 95%			Estimación	Probabilidad	Intervalo probabilidad 95%			Estimación	Probabilidad	**Cómputo**	% victoria
presidencial	inferior	superior	longitud	puntual	de victoria	inferior	superior	longitud	puntual	de victoria	**final**	muestreo
PAN	22.02	22.54	0.52	22.28	0.00	22.02	22.55	0.53	22.28	0.00	**22.28**	0.00
PRI	16.23	16.67	0.44	16.45	0.00	16.22	16.69	0.47	16.45	0.00	**16.41**	0.00
AMLO	52.87	53.42	0.55	53.15	100.00	52.82	53.49	0.67	53.15	100.00	**53.19**	100.00

Para decidir cuál modelo podría estar sobre o subestimando la variabilidad de las estimaciones, con la cual se relaciona directamente la longitud de los intervalos estimados, se estimó la probabilidad de cobertura para los mismos, la cual debiera ser, idealmente, de 95% precisamente, así que entre más cerca a este valor, mejor, véase la Tabla 3.

Tabla 3: Probabilidad de cobertura (en porcentaje) de los intervalos de probabilidad 95% estimados para el porcentaje de votos sobre la votación efectiva. Fuente: Elaboración propia con datos del INE (2006, 2012, 2018).

Probabilidad de cobertura	Modelo Mendoza - Nieto			Modelo función cópula		
	2006	2012	2018	2006	2012	2018
PAN	93.72	94.85	95.48	95.37	95.16	95.61
PRI	93.74	93.94	94.45	95.84	95.90	95.07
AMLO	93.97	92.91	90.07	95.05	94.11	95.27

En todos los casos, la probabilidad de cobertura bajo el modelo de Mendoza y Nieto-Barajas (2016) es menor a la obtenida por el modelo que se propone, en la mayoría de casos por debajo del ideal 95%, lo cual es congruente con intervalos de menor longitud como se aprecia en la Tabla 2, y esto es evidencia empírica de lo que se demostró de forma teórica en la sección tercera del presente trabajo: no considerar la dependencia entre estratos puede conducir a probabilidades de cobertura menores a las deseadas.

En cuanto a las probabilidades de victoria, en las elecciones de los años 2012 y 2018 la diferencia entre los candidatos en primer y segundo lugar fue tal (6.62% y 30.91%, respectivamente) que bajo ninguno de los escenarios simulados fue posible observar algo distinto. Pero en el caso de la elección del año 2006, donde la diferencia fue de apenas 0.58%, resulta suficientemente pequeña para que en algunos pocos de los escenarios simulados pudieran llegar a observarse victorias distintas a lo que resultó en el cómputo distrital, pero bajo ambos métodos se llega a probabilidades de victoria similares (véase la Tabla 2).

Recuérdese que en la sección cuarta del presente trabajo se argumentó, de forma teórica, que la dependencia entre candidatos forma parte del cálculo de la probabilidad de victoria, y que Mendoza y Nieto-Barajas (2016) mencionan que bajo un estudio previo que realizaron (sin dar detalles) con un modelo que tomaba en cuenta dependencia entre candidatos, dicha dependencia resultaba muy débil y podía ser desestimada. Este parece ser el caso de la elección presidencial del año 2006, pero la posibilidad de que no sea así en alguna otra elección está abierta, y el modelo basado en funciones cópula que se propone está preparado para ello.

Conclusiones

La principal contribución del presente trabajo fue demostrar que es necesario considerar la posible dependencia probabilística entre estratos y entre candidatos para las estimaciones por intervalo de un conteo rápido electoral, ya que de no hacerlo los intervalos estimados pudieran resultar con una probabilidad de cobertura distinta al nivel de probabilidad con que se calcularon. Más aún, se propuso un modelo basado en funciones cópula para tal fin, y para estimar también la probabilidad de victoria entre candidatos contendientes, que a su vez también requieren considerar ambos tipos de dependencia.

Se realizó un estudio de simulación para comparar los resultados que se obtienen con el modelo que se propone versus los del modelo de Mendoza y Nieto-Barajas (2016), utilizando datos oficiales de los cómputos distritales de las elecciones presidenciales en México en los años 2006, 2012 y 2018. Para estas elecciones, los resultados mostraron que la dependencia entre estratos y entre candidatos sí era relevante para que los intervalos estimados lograran una probabilidad de cobertura más cercana el nivel de probabilidad 95% deseado. Para el caso de la probabilidad de victoria, aunque teóricamente quedó demostrado que la dependencia entre candidatos también forma parte de su cálculo, en dichas elecciones en particular no resultó significativamente relevante para tal propósito, pero el modelo propuesto está preparado para el caso en que llegara a serlo.

Agradecimiento

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

Referencias

Bernardo, José Miguel; Smith, Adrian F.M. 1994. *Bayesian Theory*. Nueva York: Wiley.

Casella, George; Berger, Roger. 2002. *Statistical Inference*. Pacific Grove: Duxbury.

Erdely, Arturo. 2018. “La falacia del empate técnico electoral.” En: *Revista Mexicana de Estudios Electorales*. 20: 11 – 47.

Hofert, Marius; Kojadinovic, Ivan; Maechler, Martin; Yan, Jun. 2017. *copula: Multivariate Dependence with Copulas. R package version 0.999-18.* Descargable de https://CRAN.R-project.org/package=copula

INE. Instituto Nacional Electoral. 2006, 2012. *Bases de Datos de los Procesos Electorales Federales en México.* Disponible en http://siceef.ine.mx/downloadDB.html consultado el 3 de noviembre de 2018.

INE. Instituto Nacional Electoral. 2016. *Reglamento de Elecciones (México)*. Disponible en https://www.ine.mx/wp-content/uploads/2017/04/Reglamento_Elecciones.pdf consultado el 3 de noviembre de 2018.

INE. Instituto Nacional Electoral. 2018. *Conteos Rápidos: Procesos Electorales Federal y Locales 2017-2018*. Disponible en https://www.ine.mx/conteos-rapidos-procesos-electorales-federal-locales-2017-2018 consultado el 3 de noviembre de 2018.

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

Nelsen, Roger. 2006. *An Introduction to Copulas*. 2ª ed. Nueva York: Springer.

R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Descargable de https://www.R-project.org/

Salvadori, Gianfausto; De Michele, Carlo; Kottegoda, Nathabandu T.; Rosso, Renzo. 2007. *Extremes in Nature. An Approach Using Copulas.* Nueva York: Springer.

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.