UN MODELO DE SIMULACION ESTOCASTICA PARA LA ESTIMACION DE ESCAÑOS

Félix Aparicio

1. *Introducción*

El objeto del presente artículo es desarrollar un modelo teórico que permita realizar estimaciones de los escaños obtenidos por los distintos partidos y agrupaciones políticos en los procesos electorales. Estas estimaciones se basan en los resultados de encuestas preelectorales, y éstas no son objeto de estudio aquí, es decir, suponemos dados los resultados de una encuesta fiable a nivel provincial.

La principal ventaja del modelo que se expondrá a continuación es que se basa en una simulación estocástica y permite, por tanto, todo tipo de cálculos probabilísticos, tales como intervalos de confianza, probabilidades condicionadas, etc., tal y como se verá en los ejemplos del apartado tercero.

En cambio, no se puede esperar del modelo que supla las deficiencias de la encuesta preelectoral en la que se basa. Si las estimaciones de voto de los partidos dentro de cada provincia o división electoral son poco exactas, los resultados no serán fiables.

2. *Planteamiento del modelo*

Se considera un conjunto de I divisiones electorales. En el caso español $I=52$ (50 provincias más Ceuta y Melilla).

Supongamos ahora una de esas divisiones electorales, la número m. En esta división electoral se presentarán a las elecciones n_m partidos políticos. Sea π_i un estimador insesgado de la proporción de votos que obtendrá el partido i, $i=1,2, \ldots\ldots, n_m$. Sea N^* un estimador insesgado del número de votos que se emitirán en esa división electoral.

Se trata ahora de establecer una hipótesis razonable sobre la distribución conjunta del número de votos de cada partido.

En una primera aproximación se podría suponer normalidad en la distribución de los estimadores π_i. La media y varianza de estas distribuciones normales serían estimables. Multiplicando por el número de votantes (suponiendo éste constante) tendríamos la distribución marginal del número de votos a cada partido, ahora bien, no parece fácil hallar su distribución conjunta.

Otro enfoque, que es el que utilizaremos aquí como hipótesis de trabajo, es que, dados los estimadores π_i, la distribución del número de votos a cada partido es multinomial, $M_{n_m}(N^*, \pi_1, \ldots, \pi_i, \ldots, \pi_{n_m})$. Una posible justificación teórica es que se supone conocida la proporción de votos a cada partido, pero el número total de votos puede fluctuar debido a causas aleatorias «semejantes» a las que influyen en la extracción con reemplazamiento de bolas de distintos colores en una urna cuando la proporción de bolas de cada color es fija y conocida. Este esquema de urna da lugar a la distribución multinominal.

Si llamamos ξ a esta distribución multinomial, su función de probabilidad será:

$$P(\xi_1=h_1, \ldots, \xi_{n_m}=h_{n_m}) = \begin{cases} \dfrac{N^*!}{h_1! \cdot \ldots\ldots \cdot h_{n_m}!} \cdot \pi_1^{h_1} \cdot \ldots \cdot \pi_{n_m}^{h_{n_m}} & \text{Si } h_1+\ldots+h_{n_m}=N^* \\ 0 & \text{En otro caso.} \end{cases}$$

Por consiguiente, su media y su matriz de covarianzas serán:

$$E[\xi] = \begin{bmatrix} N^* \cdot \pi_1 \\ \cdot \\ N^* \cdot \pi_i \\ \cdot \\ N^* \cdot \pi_{n_m} \end{bmatrix}$$

$$COV[\xi] = \begin{bmatrix} N^*\pi_i(1-\pi_1) & & N^*\pi_1\pi_{n_m} \\ & N^*\pi_i(1-\pi_i) & \\ S & & \\ & & N^*\pi_{n_m}(1-\pi_{n_m}) \end{bmatrix}$$

Ahora bien, como es sabido (véase Cramer, pp. 479-480), la distribución multinomial corregida converge en ley a una normal multidimensional en los términos siguientes:

$$\begin{bmatrix} \dfrac{\xi_1 - N^* \cdot \pi_1}{\sqrt{N^* \cdot \pi_1}} \\ \cdots\cdots\cdots \\ \dfrac{\xi_{n_m} - N^* \pi_{n_m}}{\sqrt{N^* \cdot \pi_{n_m}}} \end{bmatrix} \xrightarrow{L} \eta$$

Donde η es una variable aleatoria normal multivariante con los siguientes parámetros:

$$\eta \frown N_{n_m} [(0, \ldots, 0)^T \wedge = \mathrm{I}_{n_m} - \pi \cdot \pi^T]$$

Donde

$$\pi^T = (\pi_1^{\frac{1}{2}}, \ldots, \pi_{n_m}^{\frac{1}{2}})$$

Por I_{n_m} representamos la matriz identidad de orden n_m.

La aproximación es buena, pues en cualquier división electoral N^* es un número elevado.

Se trata ahora de generar vectores aleatorios que sigan la distribución de η; para cada uno de estos vectores aleatorios haremos la repartición de escaños correspondiente (en el caso de España esto se haría aplicando la ley D'Hont); si, por ejemplo, el vector es el $(v_1, \ldots, v_i, \ldots, v_{n_m})^T$, repartimos los escaños como si el partido 1 hubiera obtenido v_1 votos, etc.; el n_m hubiera obtenido v_{n_m} votos.

Supongamos que en esa misma provincia o división electoral generamos p_m de esos vectores aleatorios, tendremos entonces p_m reparticiones de escaños, pero, como algunas se repetirán, tan sólo tendremos $k_m \leqslant p_m$ reparticiones distintas, cada una con su frecuencia relativa de aparición.

Distribución de escaños	*Frecuencia relativa*
D_1^m	f_1^m
$\vdots$	$\vdots$
$D_{k_m}^m$	$f_{k_m}^m$

Con

$$\sum_{j=1}^{k_m} f_j^m = 1$$

Se toman entonces estas f_j^m como probabilidades de obtener la correspondiente repartición de escaños en la provincia m.

El problema ahora es cómo generar vectores aleatorios que sigan la distribución normal multivariante η. En el apéndice 1 se explica la forma de hacer esto.

Una vez generados estos vectores se procede así:

Para cada vector $u=(u_1, \ldots, u_{n_m})^T$ que siga la distribución normal

$$u \frown N_{n_m}(0, \mathrm{I}_{n_m} - \pi \cdot \pi^T)$$

se despeja

$$\frac{v_i - N^* \cdot \pi_1}{\sqrt{N^* \cdot \pi_i}} = u_i \qquad i=1, \ldots, n_m$$

De donde

$$v_i = u_i \cdot \sqrt{N^* \cdot \pi_i} + N^* \cdot \pi_i \qquad i=1, \ldots, n_m$$

Veremos ahora cómo obtener la estimación de escaños a nivel nacional. Sean las I divisiones electorales y sean D^m las distribuciones de escaños para cada división electoral, con frecuencias relativas correspondientes F^m, es decir:

$$D^m = \begin{bmatrix} D_1^m \\ \vdots \\ D_{k_m}^m \end{bmatrix} \quad F^m = \begin{bmatrix} f_1^m \\ \vdots \\ f_{k_m}^m \end{bmatrix} \quad m=1, \ldots, \mathrm{I}$$

Establecemos la hipótesis de que las distribuciones de escaños en las distintas divisiones electorales son estadísticamente independientes entre sí; esta hipótesis parece más que razonable dado que los escaños en cada división electoral se distribuyen exclusivamente en base a los votos de esa división.

Se procede como sigue: se generan p vectores aleatorios, $w^1, \ldots, w^p$, cada uno de ellos de dimensión I.

$$w^j = (w_1^j, \ldots, w_\mathrm{I}^j)^T \qquad j=1,2, \ldots, p$$

Las componentes de cada uno de estos vectores han de ser independientes entre sí y w_i^j seguirá la distribución F^i de D^i, $i=1, \ldots, \mathrm{I}$. Es decir, en cada división electoral i se simula la distribución de escaños D^i con probabilidades

F^i (en el apéndice 2 se explica cómo hacer esto); se obtiene, así, para cada vector aleatorio w^j y para cada división electoral I, una distribución de escaños. Sumando en las divisiones electorales obtendremos, para cada vector aleatorio w^j, $j=1, \ldots, p$, una distribución de escaños a nivel nacional; se obtendrán, por tanto, p distribuciones de escaños a nivel nacional, pero habrán algunas que se repitan, con lo que sólo tendremos $h \leqslant p$ distribuciones de escaños a nivel nacional, a las que llamaremos $\overline{D}^1, \ldots, \overline{D}^h$; cada una tendrá su frecuencia relativa de aparición $\overline{f}_1, \ldots, \overline{f}_h$.

Distri. esc. nacional	$\overline{D}^1, \ldots\ldots, \overline{D}^h$
Frecuencias relativas	$\overline{f}_1, \ldots\ldots, \overline{f}_h$

Se puede tomar como distribución de escaños estimada a nivel nacional a las más probable de las D_j, es decir, a la moda de la anterior tabla o, en otras palabras, a

$$\overline{D}_q \,/\, \overline{f}_q \geqslant \overline{f}_p \qquad p=1, \ldots, h$$

Las ventajas de este método sobre los clásicos son las mismas ventajas de la simulación estocástica, es decir, que se trabaja con las distribuciones *a posteriori,* en vez de tomar medias *a priori,* pues esto último introduce sesgos notables y resta todo valor probabilístico a los resultados.

En el apartado 3 de ejemplos se ven algunas de las muchas posibilidades del modelo.

Este procedimiento aún no ha sido programado en ordenador, debido a que los programas serían muy laboriosos de realizar no por dificultades técnicas, sino por la gran cantidad de detalles a tener en cuenta.

Conviene insistir en el hecho de que este modelo no puede suplir las deficiencias de una encuesta que proporcione malas estimaciones de voto, tan sólo mejora las conclusiones que se pueden obtener sobre el reparto final de escaños, añadiendo a éstas todo tipo de cálculos probabilísticos.

Se sugiere a cualquier persona interesada en programar el método que no programe sino hasta la obtención de los $\overline{f}_j$ y los $\overline{D}_j$, pues su posterior explotación (cálculo de medias, tablas, etc.) se puede llevar a cabo con ayuda de cualquier paquete de programas de uso estadístico, tales como el BMDP, el SPSS, el SAS o cualquier otro de los muchos disponibles en el mercado.

3. *Ejemplos*

Como el modelo no ha sido aún programado en ordenador, los ejemplos que vienen a continuación son puramente teóricos y sirven tan sólo para aclarar qué tipo de conclusiones se pueden obtener.

Supongamos que, en unas elecciones a las que concurren cuatro partidos políticos, *A, B, C* y *D,* se reparten 350 escaños y que, al aplicar el modelo descrito, la distribución *a posteriori* es como sigue:

Distribución	$\overline{D}_1$ $\overline{D}_2$ $\overline{D}_3$ $\overline{D}_4$ $\overline{D}_5$ $\overline{D}_6$ $\overline{D}_7$ $\overline{D}_8$
Frec. relativa	1/8, ..., 1/8

Por simplicidad en los cálculos hemos supuesto que todos los $f_i = 1/8$, en la práctica esto no sería así.

A continuación ponemos cada $\overline{D}_i$.

$\overline{D}_1$	
A	176
B	100
C	50
D	24

$\overline{D}_2$	
A	178
B	95
C	52
D	25

$\overline{D}_3$	
A	175
B	98
C	49
D	29

$\overline{D}_4$	
A	175
B	97
C	50
D	28

$\overline{D}_5$	
A	177
B	96
C	50
D	27

$\overline{D}_6$	
A	176
B	99
C	51
D	24

$\overline{D}_7$	
A	174
B	101
C	48
D	26

$\overline{D}_8$	
A	174
B	102
C	51
D	23

En un caso práctico habría más de ocho distribuciones finales de escaños, lo que permitiría afinar más los resultados.

Veamos, en primer lugar, cuáles serían las distribuciones marginales de escaños de cada partido.

Directamente de las tablas anteriores:

Partido	*Escaños*	*Probabilidad*
A	174	$2/8 = P(\bar{D}_3) + P(\bar{D}_8)$
	175	$2/8 = P(D_4) + P(D_7)$
	176	2/8 etc.
	177	1/8 ·
	178	1/8 ·
C	48	1/8 ·
	49	1/8 ·
	50	3/8 ·
	51	2/8 ·
	52	1/8 ·
B	95	1/8 ·
	96	1/8 ·
	97	1/8 ·
	98	1/8 ·
	99	1/8 ·
	100	1/8 ·
	101	1/8 ·
	102	1/8 ·
D	23	1/8 ·
	24	2/8 ·
	25	1/8 ·
	26	1/8 ·
	27	1/8 ·
	28	1/8 ·
	29	1/8 ·

Podemos estimar los escaños que obtendrá cada partido por la media *a posteriori:*

$$\text{Escaños } (A) = 174 \cdot 2/8 + 175 \cdot 2/8 + \ldots\ldots + 178 \cdot 1/8 = 175{,}6$$

Análogamente,

$$\text{Escaños } (B) = 98{,}5$$
$$\text{Escaños } (C) = 50{,}1$$
$$\text{Escaños } (D) = 25{,}7$$

La anterior estimación de escaños de *A* se redondea a 176, que constituyen mayoría absoluta, pero ésta parece un poco ajustada; así, pues, podemos preguntarnos cuál es la probabilidad de que A obtenga esta mayoría absoluta. La respuesta es fácil:

$$P\,[\text{Escaños }(A)>175]=P(\overline{D}_1)+P(\overline{D}_2)+P(\overline{D}_5)+P(\overline{D}_6)=4/8=1/2$$

O sea, que esta mayoría absoluta no está tan clara. También podemos obtener un intervalo de confianza para los escaños de A, como

$$P\,[\text{Escaños }(A)\in\{175,176,177\}]=5/8=0{,}625$$

resulta que el intervalo [175,177] es un intervalo de confianza para los escaños de A al 62,5 por 100. Naturalmente, en la práctica tendríamos más distribuciones y podríamos construir intervalos de confianza al 90 por 100, al 95 por 100, etc.

También podemos obtener distribuciones bidimensionales de escaños; por ejemplo:

B \ A	$\leqslant 175$	$\geqslant 176$	
$\leqslant 98$	2/8	3/8	5/8
$\geqslant 99$	2/8	1/8	3/8
	4/8	4/8	

A partir de esta tabla se deduce, por ejemplo, que es poco probable (1/8) que ambos partidos obtengan simultáneamente muchos escaños; éste es un resultado lógico, pero el modelo *nos permite cuantificarlo.*

También podríamos, a partir de una tabla como la anterior, contrastar la hipótesis de que los escaños de un determinado partido X son independientes de los de otro partido Y.

Otra posibilidad es el cálculo de probabilidades condicionadas; por ejemplo, podemos calcular la probabilidad de que A obtenga mayoría absoluta condicionado por que B obtenga 100 o más escaños; esta probabilidad será:

$$P\,[\text{Escaños }(A)>175\ /\ \text{Escaños }(B)\geqslant 100]=$$
$$=\frac{P\,[\{\text{Escaños }(A)>175\}\cap\{\text{Escaños }(B)\geqslant 100\}]}{P\,[\text{Escaños }(B)\geqslant 100]}=\frac{1/8}{3/8}=1/3$$

es decir, una probabilidad reducida, como es lógico.

Las posibilidades son mucho mayores; se pueden hacer todo tipo de cálculos probabilísticos, los que hemos visto sólo pretenden servir de ejemplo.

APENDICE 1

Antes de comenzar la descripción de cómo se generan los vectores normales conviene hacer notar que la distribución multinomial es singular. En efecto, $\xi=(\xi_1, \ldots, \xi_{n_m})^T$ es tal que

$$\sum_{j=1}^{n_m} \xi_j = N^*$$

con probabilidad 1.

Además, η, la normal que aproxima a ξ, también es singular (véase Cramer, p. 481); por tanto, como nos va a hacer falta que la matriz de covarianzas de la normal multivariante sea definida positiva, trabajaremos sólo con las $n_m - 1$ primeras componentes de ξ. Esto no supone ningún problema, pues una vez obtenido el vector aleatorio que sigue la distribución de $(\xi_1, \ldots, \xi_{n_m-1})^T$, al que llamamos $(v_1, \ldots, v_{n_m-1})^T$, la última componente v_{n_m} se obtiene así:

$$v_1 + \ldots + v_{n_m-1} + v_{n_m} = N^*$$

luego

$$v_{n_m} = N^* - v_1 - \ldots - v_{n_m-1}$$

Por tanto, el problema es generar vectores aleatorios normales $u=(u_1, \ldots, u_{n_m-1})^T$ que sigan la distribución:

$$u \frown N_{n_m-1}(0^T, I_{n_m-1} - \pi \cdot \pi^T)$$

con

$$\pi = [\pi_1^{\frac{1}{2}}, \ldots, (\pi_{n_m-1})^{\frac{1}{2}}]^T$$

Esta normal ya no es singular.

Para generar vectores aleatorios según una distribución normal multivariante no singular se procede así:

Sea la normal $\zeta \frown N_r(\mu, \Sigma)$:

1. Por ser Σ simétrica y definida positiva, existe T matriz triangular inferior tal que $\Sigma = T \cdot T^T$.

2. Sea $\Phi \frown N_r(0, I_r)$, la distribución de $\Theta = T \cdot \Phi$ es normal, por ser función lineal de una normal, su media es 0 y su matriz de covarianzas es

$$COV(\Theta)=T \cdot COV(\Phi) \cdot T^T=T \cdot \mathrm{I} \cdot T^T=T \cdot T^T=\Sigma$$

Es decir, basta con generar r números aleatorios normales $N(0,1)$ independientes, y al vector Φ formado por ellos se le premultiplica por T, con ello habremos generado de una normal $N_r(0, \Sigma)$; si ahora sumamos μ tendremos la variable ζ.

Ahora bien, nos quedan dos puntos por aclarar:

A) La generación de los números normales.

B) La descomposición $\Sigma=T \cdot T^T$.

La primera de estas cosas es sencilla; existen muchos métodos para generar números normales (véase Rubinstein, pp. 86-91), uno de los más simples y eficientes es el siguiente:

Se generan 12 números aleatorios uniformes en el intervalo $[0,1]$ e independientes, les llamaremos $y_1, \ldots, y_{12}$. En virtud del teorema central del límite (TCL), la suma de estos 12 números tipificada se parecerá a una variable aleatoria $N(0,1)$; se comprueba en la práctica que la aproximación es razonablemente buena para 12 números.

En concreto, tenemos:

$$\frac{\sum_{i=1}^{12} y_i - 12 \cdot \frac{1}{2}}{D[\sum_{i=1}^{12} y_i]} = \frac{\sum_{i=1}^{12} y_i - 12 \cdot \frac{1}{2}}{\sqrt{12 \cdot (1-0)/12}} = \sum_{i=1}^{12} y_i - 6 \simeq N(0,1)$$

Así, pues, basta con sumar los 12 números uniformes y a su suma restarle la constante 6.

A su vez, para generar números uniformes en el intervalo $[0,1]$ se puede emplear cualquier método; la mayoría de ordenadores disponen de subrutinas del sistema a este efecto; de todas formas, en el apéndice 3 está escrita una subrutina Fortran que lo hace.

La respuesta a la parte B) es que se pueden emplear diversos métodos, de ellos expongo a continuación el algoritmo de la raíz cuadrada (véase Naylor, p. 117, o Rubinstein, pp. 65-67).

Sea $\Sigma=(\sigma_{ij})_{rxr}$, Σ simétrica y definida positiva, hacer:

1.º)

$$t_{i1}=\frac{\sigma_{i1}}{\sqrt{\sigma_{11}}} \qquad 1 \leqslant i \leqslant r$$

2.°)

$$t_{jj} = \sqrt{\sigma_{jj} - \sum_{k=1}^{j-1} t_{jk}^2} \qquad 1 < j \leqslant r$$

y, para cada j fijo

$$t_{ij} = \frac{\sigma_{ij} - \sum_{k=1}^{j-1} t_{ik} \cdot t_{jk}}{t_{jj}} \qquad j < i \leqslant r$$

APENDICE 2

Se trata ahora de ver cómo se simula la distribución de escaños D^i.

$$D^i = \begin{bmatrix} D_1{}^i \\ \vdots \\ D_{k_m}{}^i \end{bmatrix} \text{ con probabilidad } F^i = \begin{bmatrix} f_1{}^i \\ \vdots \\ f_{k_m}{}^i \end{bmatrix}$$

Sean

$$S_j = \sum_{k=1}^{j} f_k{}^i$$

es decir,

$$S_1 = f_1{}^i$$
$$S_2 = f_1{}^i + f_2{}^i$$
$$\vdots$$
$$S_{k_m} = f_1{}^i + \ldots\ldots + f_{k_m}{}^i = 1$$

Basta con generar un nuevo aleatorio uniforme en [0,1] (véase apéndice 3), sea y este número.

Se toma entonces $D_p{}^i$ como distribución de escaños simulada en la provincia i si se verifica:

$$S_p < y \leqslant S_{p+1}$$

APENDICE 3

La siguiente subrutina Fortran es un método congruencial multiplicativo para generar números uniformes en el intervalo [0,1].

```
FUNCTION RAND (IX)
INTEGER*4 A,P,IX,B15,B16,XHI,XALO,LEFTLO,FHI,K
DATA A/16807/,B15/32768/,B16/65536/,P/2147483647/
XHI=IX/B16
XALO=(IX-XHI*B16)*A
LEFTLO=XALO/B16
FHI=XHI*LEFTLO
K=FHI/B15
IX=((XALO-LEFTLO)*B16)-P)+(FHI-K*B15)*B16)+K
IF (IX.LT.0) IX=IX+P
RAND=FLOAT (IX) *1.525878906E-5
RETURN
END
```

Si se trabaja en 32 *bits,* la constante 1.525878906E−5 ($=1/2^{16}$) ha de ser sustituida por 2.328306437E−10 ($=1/2^{32}$).

BIBLIOGRAFIA

CRAMER, H.: *Métodos matemáticos de estadística,* Aguilar, Madrid, 1953.
RUBINSTEIN, Reuven Y.: *Simulation and the Monte Carlo Method,* Wiley, 1981.
NAYLOR y otros: *Técnicas de Simulación en Computadoras,* Limusa, México D. F., 1977.

CRITICA DE LIBROS