Sobre Patrones Repetitivos en los Resultados de un Referendum

José H. Nieto

1. Introducción

En el referendum realizado el 15 de agosto de 2004 en Venezuela cada elector tuvo la posibilidad de expresar si deseaba o no revocar el mandato al presidente de la República. Para ello se utilizaron máquinas de votación electrónicas, en las cuales cada elector podía pulsar SI o NO. La máquina registraba el voto y además imprimía un comprobante que el votante debía depositar en una urna, para efectos de control.

La votación se realizó en 4766 centros de votación automatizados y 3628 centros manuales (ver [1]). En cada centro automatizado había un número variable de máquinas (dependiendo del número de electores correspondientes al centro), para un total de 19664 máquinas.

La asignación de los electores a las máquinas se reslizaba en función de los dos últimos dígitos de la cédula de identidad del elector, procurando una distribución uniforme de los electores en las máquinas disponibles. Por ejemplo en los centros con 10 máquinas numeradas de 1 a 10, el elector con cédula c votaba en la máquina $\lfloor (c \bmod 100)/10 \rfloor + 1$.

Luego de finalizado el proceso algunas personas observaron coincidencias entre los votos obtenidos por el SI en diferentes máquinas de un mismo centro de votación. Por ejemplo en una máquina el NO obtuvo 172 votos y el SI 126, y en otra máquina el NO obtuvo 176 votos y el SI 126.

Varios representantes de la oposición (que respaldaban la opción SI) consideraron que coincidencias como la mencionada eran extremadamente improbables, y que su existencia era un indicio de fraude.

El propósito de este trabajo es calcular la probabilidad de algunas de esas coincidencias para ver si su aparición es o no razonable, y comparar los resultados con los obtenidos en simulaciones computarizadas del proceso eleccionario.

2. El modelo

En esta sección desarrollaremos un modelo probabilístico de un centro de votación. Supongamos que el centro consta de m máquinas de votación numeradas de 1 a m y que en él votan v personas. Supondremos que la distribución de los votantes en las mesas es uniforme, es decir que la probabilidad de que un votante tomado al azar vote en la mesa k es $1/m$ para cualquier $k = 1 \dots m$. Esto es razonable dado el sistema (basado en los dos últimos dígitos de la cédula de identidad) utilizado para asignar los votantes a las máquinas. Supondremos además que hay una probabilidad p $(0 \leq p \leq 1)$ de que un votante tomado al azar vote por el SI y la probabilidad complementaria $q = 1 - p$ de que vote por el NO. Esta suposición es razonable ya que los votantes asignados a cada máquina son una muestra representativa del universo de votantes del centro y a que obviamente la preferencia política de cada elector es independiente de los últimos dígitos de su cédula.

Sean S_i y N_i el número de votos SI y NO, respectivamente, obtenidos en la máquina i. Entonces la distribución conjunta de estas variables es

$$\mathrm{Prob}(S_i = s_i, N_i = n_i, i = 1, \dots, m) = \frac{v!}{s_1! s_2! \cdots s_m! n_1! \cdots n_m!} \left(\frac{p}{m}\right)^{s_1 + \cdots + s_m} \left(\frac{q}{m}\right)^{n_1 + \cdots + n_m}$$

Sea ahora $u(v, m, p, j)$ la probabilidad de que los resultados del SI en las primeras j mesas sean iguales, es decir

$$u(v, m, p, j) = \mathrm{Prob}(S_1 = S_2 = \cdots = S_j).$$

Entonces es fácil ver que

$$u(v, m, p, j) = \sum_{k \geq 0} \frac{v!}{(k!)^j (v - jk)!} \left(\frac{p}{m}\right)^{jk} \left((m - j)\frac{p}{m} + q\right)^{v - jk}$$

y en particular si $p = q = 1/2$

$$u(v, m, 1/2, j) = \left(\frac{2m - j}{2m}\right)^v \sum_{k \geq 0} \frac{v!}{(k!)^j (v - jk)! (2m - j)^{jk}}$$

3. Estudio de algunos casos

Examinaremos a continuación algunos resultados numéricos en situaciones típicas presentadas en el referendum mencionado en la introducción. Los cálculos de $u(v,m,p,j)$ fueron realizados con Maple mediante el siguiente procedimiento:

```
u := proc(v, m, p, j)
local t, s, k;
    t := 1;
    s := 1;
    for k to iquo(v, m) do
        t := t*product(i, i = v - j*k + 1 .. v - j*k + j)*
            (p/(m - j*p))^j/k^j;
        s := s + t
    end do;
    evalf(s*(1 - j*p/m)^v)
end proc
```

Por razones de eficiencia la sumatoria en k se realizó solamente hasta $\lfloor v/m \rfloor$ en vez de hasta $\lfloor v/j \rfloor$, con lo cual en todo caso se obtienen resultados inferiores a los exactos, pero en realidad la contribución de los términos con $k > \lfloor v/m \rfloor$ es insignificante.

3.1. Dos máquinas

Consideremos en primer lugar un centro con sólo dos máquinas (o equivalentemente una mesa con dos máquinas) en el cual hayan votado 1000 personas. La probabilidad de que en ambas el SI obtenga el mismo número de votos, suponiendo $p = 1/2$, es

$$u(1000, 2, 1/2, 2) = \left(\frac{1}{2}\right)^{1000} \sum_{k\geq 0} \frac{1000!}{(k!)^2(1000-2k)!(2)^{2k}} \approx 0{,}0178.$$

Si en cambio $p = 0{,}3$ entonces la probabilidad se incrementa a $u(1000, 2, 3/10, 2) \approx 0{,}023$. Esto significa que, en promedio, este suceso puede ocurrir en aproximadamente 2 de cada 100 centros (o mesas) con estas características.

3.2. Tres máquinas

Consideremos ahora un centro con tres máquinas (o bien una mesa con tres máquinas) donde votan 1200 personas. Para $p = 1/2$ la probabilidad de obtener el mismo resultado para el SI en las tres es

$$u(1200, 3, 1/2, 3) \approx 0{,}000459$$

y la de obtener resultados iguales en dos de las tres es

$$3u(1200, 3, 1/2, 2) - 2u(1200, 3, 1/2, 3) \approx 0{,}0589$$

(el término que se resta es una corrección para no contar tres veces los casos en que los resultados de las tres máquinas coinciden).

Los resultados análogos para $p = 0{,}3$ son:

$$u(1200, 3, 3/10, 3) \approx 0{,}0007658$$

$$3u(1200, 3, 3/10, 2) - 2u(1200, 3, 3/10, 3) \approx 0{,}0757$$

Aunque desconozco cuántas mesas con tres máquinas había estimo que fue un número cercano a 5000. En ese caso es razonable esperar alrededor de tres mesas con tres resultados idénticos para el SI y alrededor de 350 mesas con dos resultados iguales.

3.3. Cuatro y más máquinas

Consideremos ahora un centro con cuatro máquinas donde votan 1800 personas. Pueden presentarse cinco patrones diferentes para los votos del SI, con probabilidades p_1,...,p_5:

p_1 = Probabilidad de 4 resultados diferentes (a,b,c,d).
p_2 = Probabilidad de 2 resultados iguales y 2 diferentes (a,a,b,c).
p_{22} = Probabilidad de 2 pares de resultados iguales (a,a,b,b).
p_3 = Probabilidad de 3 resultados iguales y uno diferente (a,a,a,b).
p_4 = Probabilidad de 4 resultados iguales (a,a,a,a).

Entonces

$$p_4 = u(2000, 4, 1/2, 4) \approx 8{,}03 \cdot 10^{-6},$$

$$p_3 = 4u(2000, 4, 1/2, 3) - 3u(2000, 4, 1/2, 4) \approx 0{,}0014,$$

$$p_{22} \approx 3u(2000, 4, 1/2, 2)u(1000, 2, 1/2, 2) \approx 0{,}00095,$$

$$p_2 = \binom{4}{2} u(2000, 4, 1/2, 2) - 2p_3 - p_{22} + 6p_4 \approx 0{,}10.$$

En resumen, en promedio habrá dos resultados iguales en el 10 % de los casos, 3 resultados iguales en el 0.14 % de los casos y dos pares de resultados iguales en casi el 1 por mil de los casos.

Al aumentar el número de máquinas de votación la frecuencia de las repeticiones aumenta. Por ejemplo si hay 10 máquinas, es más probable que haya al menos dos resultados iguales a que sean todos diferentes. En efecto, la primera probabilidad es aproximadamente

$$\binom{10}{2}u(4000,10,1/2,2)-2\binom{10}{3}u(4000,10,1/2,3)$$
$$-\frac{1}{2}\binom{10}{2}\binom{8}{2}u(4000,10,1/2,2)u(3200,8,1/2,2)+6u(4000,10,1/2,4)\approx 0{,}59.$$

La probabilidad de obtener resultados iguales en tres mesas es aproximadamente

$$\binom{10}{3}u(4000,10,1/2,3)-3\binom{10}{4}u(4000,10,1/2,4)\approx 0{,}105.$$

y la de obtener resultados iguales en cuatro mesas es aproximadamente

$$\binom{10}{4}u(4000,10,1/2,4)-4\binom{10}{5}u(4000,10,1/2,5)\approx 0{,}002.$$

4. Simulación

Una manera alternativa de estudiar los patrones de repetición consiste en realizar una simulación computarizada, utilizando un generador de números pseudoaleatorios para simular el proceso de votación. El siguiente programa escrito en Turbo Pascal 7.0 simula un centro de votación:

```
Program Simul1;
var i,j,k,n,m: integer;
    p: real;
    si,no,a,b: array[0..9] of integer;
begin
n:= 4000;  (* numero de votantes *)
p:= 0.50;  (* probabilidad de votar SI *)
Randomize;
for i:=0 to 9 do begin si[i]:= 0; no[i]:= 0 end;
for k:=1 to n do begin
  j:= random(10);  (* seleccionar maquina del 0 al 9 *)
  if random < p then no[j]:= no[j] + 1  (* votar *)
              else si[j]:= si[j] + 1;
end;
write('NO:  ');
for i:=0 to 9 do write (no[i],' '); writeln;
```

```
write('SI:  ');
for i:=0 to 9 do write (si[i],' '); writeln;
write('Tot: ');
for i:=0 to 9 do write (no[i]+si[i],' '); writeln;
writeln;
end.
```

Al correr este programa se obtiene como salida el número de votos por el NO y por el SI en cada máquina de votación.

Naturalmente cada vez que se corre se obtienen resultados diferentes, pero si anáde un bucle que simule la votación, digamos, 10000 veces y se analizan los resultados obtenidos en busca de repeticiones se puede tener una idea aproximada del comportamiento de las mismas. Nosotros hicimos esto con semilla 123 (RandSeed:=123) para el generador y obtuvimos los siguientes resultados:

casos sin repeticiones: 3852
casos con dos resultados iguales: 5669
casos con tres resultados iguales: 462
casos con cuatro resultados iguales: 15
casos con cinco resultados iguales: 2
casos con seis o más resultados iguales: 0.

que son compatibles con los resultados teóricos de la sección anterior.

La simulación para 5000 mesas de tres máquinas con $p = 0{,}4$ y $n = 1200$ arrojó los resultados siguientes:

mesas con dos resultados iguales: 273.
mesas con tres resultados iguales: 3.

No sería difícil hacer una simulación bastante realista del proceso en su conjunto. Para ello hace falta conocer el número de centros, el número de votantes y de máquinas en cada centro y los resultados.

5. Uniforme vs. Binomial

Los patrones de repetición son bien conocidos para la distribución uniforme. Si una bolsa contiene bolillas numeradas de 1 a n y se extraen k al azar, con reposición, la probabilidad de obtener al menos dos números iguales es

$$p(n, k) = 1 - \frac{n(n-1)\cdots(n-k+1)}{n^k}$$

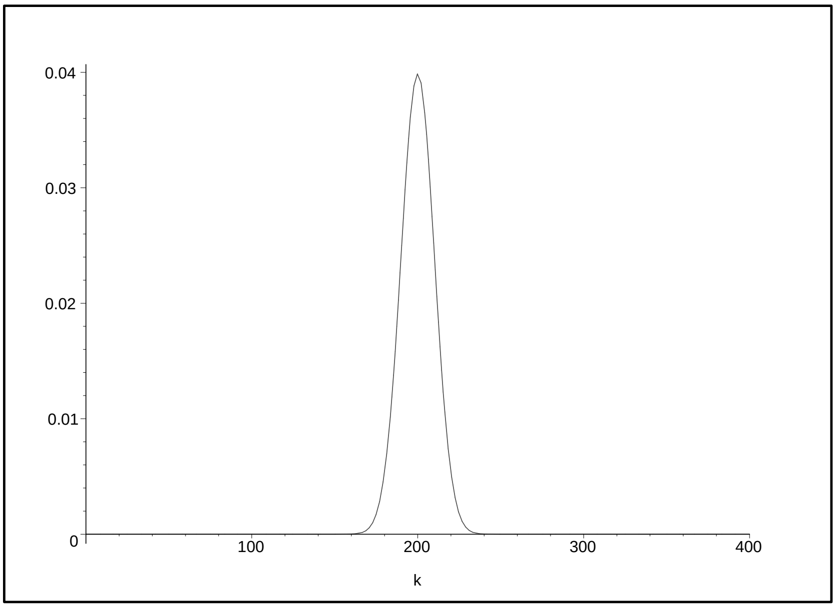

Figura 1: Distribución binomial con $n = 400$, $p = 1/2$.

El famoso *problema de los cumpleaños* (ver [2], p.49) consiste en averiguar si en un grupo de 23 personas seleccionadas al azar es más probable que haya 2 personas que cumplan anós en la misma fecha o que no las haya. Puesto que $p(365, 23) = 0{,}507$ lo más probable es que las haya, lo cual para muchos parece contrario a la intuición. En un grupo de 50 personas la probabilidad de una coincidencia se eleva a un sorprendente 97 %.

Ahora bien, si la distribución no es uniforme sino binomial la probabilidad de obtener resultados repetidos se eleva. Supongamos, por ejemplo, que se lanza una moneda 400 veces y se cuenta el número de caras. La probabilidad de obtener exactamente k caras es $2^{-n}\binom{400}{k}$, que es máximo para $k = 200$. Los siguientes valores más probables son 201 y 199, luego siguen 202 y 198, etc.

La siguiente figura ilustra la forma de la distribución binomial. Como se puede observar, prácticamente todos los valores estarán comprendidos entre 160 y 240, y puede probarse que el 99,77 % estarán comprendidos entre 170 y 230. Al tomar 10 valores en ese rango, si estuviesen distribuidos uniformemente la probabilidad de una repetición sería $f(61, 10) = 0{,}54 > 1/2$ (pero en realidad es mayor, ya que los valores se concentran alrededor del 200).

Otra forma de examinar la situación consiste en calcular la desviación estándar, a saber $\sigma = \sqrt{npq} = \sqrt{100} = 10$.

Si las preferencias de los votantes no están equilibradas sino que se inclinan hacia una de las dos opciones, los patrones repetitivos serán más frecuentes. Por ejemplo para una distribución binomial con $n = 400$ y $p = 0{,}70$ la desviación estándar es $\sqrt{npq} = \sqrt{400 \cdot 0{,}70 \cdot 0{,}30} = \sqrt{84} \approx 9{,}165$ y para $p = 0{,}80$ la desviación estándar es 8.

6. Conclusiones

La repetición del número de votos por una de las dos opciones en diferentes máquinas de una misma mesa o centro de votación, en un proceso en el cual había miles de mesas y centros automatizados y cerca de 20000 máquinas, es un fenómeno absolutamente normal aunque parezca contrario a la intuición.

La explicación de esta aparente paradoja consiste en que las distribuciones probabilísticas del numero de votos por una de las opciones en cada máquina de un mismo centro son binomiales y con parámetros semejantes, por lo cual los resultados varían en un rango bastante pequeño. Esto hace que se presente el mismo fenómeno que en el "problema de los cumpleaños", agudizado por la tendencia central y poco dispersa de la distribución binomial.

Referencias

[1] *Infraestructura Electoral Referendo Revocatorio Presidencial*, `http://www.cne.gov.ve/estadisticas/infraestructura2004_02.pdf`

[2] W. Feller, *Introducción a la Teoría de Probabilidades y sus Aplicaciones*, vol. 1, Limusa 1983.