Probabilidad condicional: La real y la de libro

Agustín Alvarez *

Enero de 2014

Resumen

En este trabajo se muestra como un cambio aparentemente insignificante en la formulación de un problema de probabilidad condicional puede cambiar la solución de manera drástica. Se muestra como pueden estar involucradas en la resolución del problema variables aleatorias "tácitas" que muchas veces obviamos e inccurrimos en una resolución errada del problema.

1. Introducción

El problema de los dos hijos, como lo llamó Martin Gardner, que data de al menos 1954 [1], traducido al español, es el siguiente:

a) El señor Lopez tiene dos hijos. La mayor es una nena. ¿Cuál es la probabilidad de que ambas sean nenas?

*Instituo de Cálculo, FCEN, Universidad de Buenos Aires

b) El señor García tiene dos hijos. Al menos uno de ellos es un nene. ¿Cuál es la probabilidad de que ambos sean nenes?

Cualquiera que realice un primer curso de probabilidad y aprenda probabilidad condicional estará en condiciones de llegar a las respuestas correctas a estos problemas que son 1/2 para el problema *a*) y 1/3 para el *b*).

Como docentes de materias básicas de probabilidad muchas veces nos vemos tentados a formular problemas de probabilidad condicional de una manera más natural o más real, contando un acontecimiento cotidiano y luego planteando la pregunta de modo que el problema resulte más atractivo. Podemos ver como ejemplo el problema propuesto por Gary Foshee en las conferencias de Gadering 4 Gardner [2]. Adrián Paenza cuenta el problema y la solución planteada por Foshee en la contratapa del diario página 12 [4]. El problema, Foshee lo planteó del siguiente modo:

- *Tengo dos hijos, uno es varón y nació un martes. ¿Cuál es la probabilidad de que ambos sean varones?*

La respuesta anti intuitiva a la que llegó es 13/27. En [3] se analiza con gran tino la validez de esta respuesta, notando en realidad que la solución no es la correcta con dicho planteo.

Otro ejemplo es el siguiente problema con el que me topé hace unos días que pretende formular el problema *b*) de Gardner ambientándolo en una situación real:

(1) *Conocemos a una mujer charlatana y en la conversación nos dice que tiene 2 hijos. Y además nos dice el nombre de uno de ellos: Felipe. Todo esto sin que nosotros le preguntemos nada. Cuál es la probabilidad de que ambos sean varones?*

En la siguiente sección resolveremos este problema.

El objetivo de este trabajo es ver como un cambio que puede parecer insignificante en la formuclación de un problema de probabilidad condicional puede cambiar totalmente la solución al problema. Y prevenir entonces sobre el cuidado que debe tenerse tanto como docente que pretende proponer con poética y romanticismo problemas más entretenidos, o como investigador que puede llegar a mal interpretar una probabilidad condicional.

Se puede ver un enfoque similar al propuesto en este escrito en [3]. Sin embargo creo que es importante el enfoque propuesto en este trabajo que pone de manifiesto como al cambiar un enunciado pueden aparecer variables aleatorias "tácitas" que cambian el problema y la solución.

2. Desarrollo del problema

2.1. La solución errada del problema

Cualquiera que haya aprendido los conceptos básicos de probabilidad condiconal no dudará en decir que para resolver (1) hay que calcular la probabilidad de que ambos hijos sean varones dado que uno de los hijos es varón. Y entonces hacer el clásico razonamiento de libro para calcular la probabilidad condicional. Este sería:

Pensemos en toda la población de madres con 2 hijos. Están aquellas que tuvieron en orden cronológico, primero un varón y segundo un varón (VV). Las que tuvieron primero un varón y segundo una mujer (VM). Las que tuvieron primero una mujer y segundo un varón (MV)

y las que tuvieron primero una mujer y segundo una mujer (MM). Partimos de la suposición de que estos 4 grupos aparecen en proporciones practicamente iguales en la población. O, sea, de las madres de 2 hijos, hay $\frac{1}{4}$ en cada uno de los 4 grupos mencionados. Una vez que Sabemos que Felipe es un varón, en realidad quedan tres grupos posibles a los que puede pertenecer la madre, que son todos menos el grupo de las madres que tuvieron 2 mujeres. Esos 3 grupos suman $\frac{3}{4}$ de la población de madres con 2 hijos y el grupo de 2 varones (VV) es $\frac{1}{4}$ de dicha población. Luego, la probabilidad de que ambos sean varones sería $\frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$. Si esa probabilidad hubiera estado bien calculada significa que yo le puedo apostar a la mujer que acabo de conocer a que adivino el sexo de su otro hijo, el/la hermano/a de Felipe. Y si arriesgo que es una mujer entonces tengo $\frac{2}{3}$ de probabilidades de acertar.

2.2. Heurística de que la solución está mal

Vamos a intentar convencernos de que hay algo mal en ese cálculo de probablidades condicional "de libro" que acabamos de hacer para resolver este problema de la "vida real". Una madre que se nos acerque a hablar, y nos diga que tiene 2 hijos y nos diga el nombre de uno de ellos, también nos podría haber dado el nombre de una nena, digamos Julieta. Entonces deduciríamos analogamente que la probabilidad de que Julieta tenga una hermana es $\frac{1}{3}$. Pero entonces, pensemos en la situación de que todas las madres de 2 hijos del mundo son charlatanas, y se nos acercan a hablar y nos dicen el nombre de uno de sus hijos. Es razonable pensar que la mitad de ellas aproximadamente nos dirán un nombre de varón y la otra mitad nos dirán un nombre de mujer. Entonces, ahora nos preguntamos que proporción de las madres del mundo con 2 hijos tienen ambos hijos varones. Bueno, de la mitad que nos dijeron el nombre de una nena, sabemos que ninguna tiene dos varones. Y de la otra mitad, supuestamente si la probabilidad del comienzo fue bien calculada, aproximadamente $\frac{1}{3}$ de esas madres tendrán ambos hijos varones. Pero entonces, del total de madres con 2 hijos, tenemos que $\frac{1}{3}$ de la mitad son las que tienen 2 varones. O sea, $\frac{1}{6}$ del total de la población de madres de 2 hijos tiene 2 varones. Pero habíamos acordado al comienzo que eran $\frac{1}{4}$ de estas madres las que tenían ambos varones. O sea, algo está mal!

Lo que está mal es no tener en cuenta que la madre está eligiendo con igual probabilidad de cuál de sus 2 hijos hablar. Y si tenemos eso en cuenta, eso cambia las probabilidades a posteriori. Y $P(VV|\text{habló de varón})$ pasa a ser $\frac{1}{2}$ y no $\frac{1}{3}$ como calculamos originalmente. Para entenderlo, pensémoslo primero intuitivamente. Supongamos que esta madre vino y nos habló de Felipe. ¿Les parece igual de probable que esta madre, haya tenido en orden VV o VM? Y, parecería que VV es más probable porque Felipe podría ser el mayor como el menor, mientras que en VM tenemos que tener la suerte de que ella haya elegido hablar del más grande. Lo planteamos más formalmente:

2.3. La solución correcta

Podemos pensar el problema de la siguiente manera: El experimento aleatorio consiste en primero elegir al azar una madre entre todas las madres que tienen 2 hijos y luego esta madre elige al azar sobre cuál de sus hijos va a hablar. Sean las variables aleatorias X e Y definidas del siguiente modo:

$$X = \begin{cases} 1 & \text{si la madre que elegimos tiene dos varones } (VV) \\ 2 & \text{si la madre que elegimos tuvo primero un nene y luego una nena } (VM) \\ 3 & \text{si la madre que elegimos tuvo primero una nena y luego un nene } (MV) \\ 4 & \text{si la madre que elegimos tiene dos nenas } (MM) \end{cases}$$

$$Y = \begin{cases} 0 & \text{si ella elige decir el nombre del mayor} \\ 1 & \text{si ella elige decir el nombre del menor} \end{cases}$$

Asumimos que estas variables aleatorias son independientes y tienen la siguiente función de probabilidad:

$P(X = i) = \frac{1}{4} \quad \forall i \in \{1, 2, 3, 4\}$ y $P(Y = 0) = P(Y = 1) = \frac{1}{2}$

También podemos pensar en la siguiente variable aleatoria:

$$Z = \begin{cases} 0 & \text{si ella nos dice un nombre de varon} \\ 1 & \text{si ella nos dice un nombre de mujer} \end{cases}$$

Habiendo definido estas variables aleatorias, el evento en que la madre nos dice el nombre Felipe, puede ser pensado como $\{Z = 0\}$. Y queremos calcular la probabilidad de que tenga 2 varones dado que nos nombró a Felipe. Bueno, esto sería: $P(X = 1|Z = 0)$.

$P(X = 1|Z = 0) = \frac{P(X=1,Z=0)}{P(Z=0)} = \frac{P(X=1)}{P(Z=0)} = \frac{\frac{1}{4}}{P(Z=0)}$

Descomponiendo a $\{Z = 0\}$ podemos calcular $P(Z = 0)$:

$$\begin{aligned}
\{Z=0\} &= \{X=1\} \cup (\{X=2\} \cap \{Y=0\}) \cup (\{X=3\} \cap \{Y=1\}) \\
P(Z=0) &= \tfrac{1}{4} + \tfrac{1}{4} \cdot \tfrac{1}{2} + \tfrac{1}{4} \cdot \tfrac{1}{2} \\
P(Z=0) &= \tfrac{1}{2}
\end{aligned}$$

Luego, concluimos que

$$P(X=1|Z=0) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

Luego, la idea de ganar dinero fácil apostando en el sexo del hermano/a de Felipe desaparece.

Lo asombroso es que el problema famoso y que se puede encontrar en muchos libros:

- Si una madre tiene 2 hijos. Cuál es la probabilidad de que ambos sean varones si al menos uno de ellos es varón

es un problema bien formulado, y está bien resolverlo como planteamos al comienzo diciendo que dicha probabilidad es $\frac{1}{3}$ pero es un problema de libro. Y en la vida real a veces nos podemos enterar que cierta madre tiene 2 hijos y que al menos uno es varón de un modo donde hay involucrada alguna otra variable aleatoria que cambia la probabilidad de que esta madre tenga dos varones.

2.4. Una formulación "real" que se condice con la probabilidad condicional de "libro"

Si por ejemplo hacemos el siguiente experimento: Vamos por ahí entrevistando mujeres. A aquellas que tienen 2 hijos varones les preguntamos si alguno de ellos es varón. En ese caso, de las madres que nos respondan que sí, es de esperar que $\frac{1}{3}$ de ellas tengan dos varones.

Pero en este caso no las dejamos elegir hablar de una mujer, si tenían un nene y una nena. Si no que las forzamos a responder nuestra pregunta sobre si tenían al menos un varón. Este experimento se corresponde con el problema que se resuelve comunmente en los libros. Y es bueno ver para quedarse tranquilos, que aquí $\frac{3}{4}$ de las mujeres responderán afirmativamente a nuestra respuesta. Y si entre ellas hay $\frac{1}{3}$ que tienen 2 varones, resulta que $\frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$ de las mujeres tienen 2 varones que no se contradice con lo que acordamos que ocurre.

Sin embargo, si el experimento es: Andar por ahí entrevistando mujeres, y una vez que nos dicen que tienen 2 hijos les pedimos que nos digan el nombre de uno de ellos. El hecho de que nos digan un nombre de varón nos asegura que tienen al menos un varón, pero eso no significa que la probabilidad de que ambos sean varones será $\frac{1}{3}$ sino que será $\frac{1}{2}$.

2.5. Moraleja

Hay que tener cuidado con el uso de la probabilidad condicional en la vida real, porque el modo en como nos enteramos del condicional puede cambiar la probabilidad condicional y no ser la clásica que aprendemos al comenzar a estudiar probabilidades.

2.6. Un ejemplo análogo

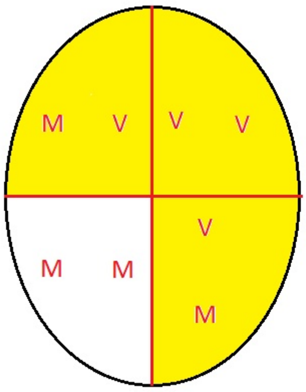
Otra formulación análoga al problema (1) y tal vez más sencillo de contar es:

- Vamos a la casa de un matrimonio conocido. Y sabemos que tienen 2 hijos pero no sabemos el sexo de ninguno de ellos. Cuando tocamos la puerta, nos sale a atender un varón que indudablemente es uno de los hijos. Cuál es la probabilidad en este caso de que ambos sean varones?

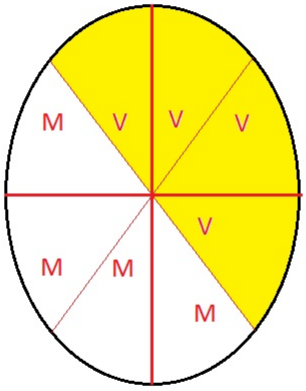
También en este caso será $\frac{1}{2}$ y no $\frac{1}{3}$ porque está presente la variable aleatoria de cuál de los 2 chicos sale a atender la puerta.

2.7. Gráficos que ayudan a entender

Veamos un par de gráficos de torta que ayudan a entender los 2 problemas, el de libro y alguna de las situaciones reales planteadas. Vemos primero un gráfico para el problema "de libro": Cada porción de la torta representa $\frac{1}{4}$ de las posibles familias de 2 hijos con un orden dependiendo qué nació primero, si varón o mujer. Y si nos preguntan la probabilidad de que sean ambos varones dado que al menos uno es varón, vemos en la torta que de las 3 partes pintadas, equiprobables a priori, sólo una corresponde al caso de Varón-Varón.



Ahora, si pensamos el problema de la "vida real":la mujer charlatana que nos va a decir el nombre de uno de sus 2 hijos. Podemos dividir la torta en 8 porciones a priori equiprobables, donde cada porción por un lado nos dice el orden de sexos de sus dos hijos si miramos a qué cuarto pertenece de la torta. Y la porción en si, nos dice de quien nos habló la mujer. Una vez que nos habla de un varón puede corresponder a cualquiera de los cuatro casos pintados en la siguiente torta. De los cuales la mitad corresponden al caso en que tiene dos varones, con lo cual la probabilidad de que sean dos varones dado que nos habló de un varón es 1/2 como habíamos visto antes.



Agradecimientos

Este trabajo fue apoyado por el subsidio W276 de la Universidad de Buenos Aires. Agradezco también a mis compañeros del Instituto de Cálculo con quienes compartimos discusiones sobre este problema en más de una oportunidad y me proveyeron por ejemplo la resolución gráfica del problema.

Referencias

[1] Gardner Martin. The Second Scientific American Book of Mathematical Puzzles and Diversions. Simon and Schuster(1954).

[2] G4G9: Ninth Gathering 4 Gardner Conference, March 2010 (http:// www.g4g4.com/).

[3] Lynch Peter, The Two-Child Paradox: Dicothomy and ambiguity, Irish Math. Soc. Bulletin 67 (2011), 67-73

[4] Paenza Adrian. El niño que nació un martes, nota de contratapa del diario Página 12 del 5 de octubre de 2011. URL: http://www.pagina12.com.ar/diario/contratapa/13-178241-2011-10-05.html