CONSIDERACIONES AL UTILIZAR LA SIMULACIÓN PARA APROXIMAR LA PROBABILIDAD DE UN EVENTO

Sanabria Brenes, Giovanni
gsanabriab@yahoo.com
Instituto Tecnológico de Costa Rica (Costa Rica)
Universidad de Costa Rica (Costa Rica)

RESUMEN

El auge que han tenido los computadores ha contribuido en la evolución del concepto frecuencial de la Probabilidad. Así la simulación de experiencias aleatorias en la computadora se ha convertido en una herramienta poderosa para resolver problemas. Esto ha influido notablemente en la Didáctica de la Probabilidad, lo cual se refleja en la presencia de propuestas, en congresos y encuentros de Didáctica, que hacen uso de la simulación para introducir la enseñanza de la Probabilidad o para ver algunas de sus aplicaciones. Sin embargo, ¿se tiene conciencia de las limitaciones de esta herramienta? ¿Se controla y evalúa conscientemente su aplicación? Este trabajo propone que la enseñanza del concepto frecuencial no se aborde aisladamente sino dialécticamente con los otros significados de Probabilidad.

PALABRAS CLAVE

Simulación, Probabilidad frecuencial, Probabilidad teórica, Didáctica de la Probabilidad.

INTRODUCCIÓN

Dado un fenómeno aleatorio con distintos resultados conocidos, la probabilidad de que ocurra un evento del fenómeno (subconjunto de resultados) se entiende como la medida de la posibilidad de que suceda el evento. Batanero (2005) brinda los distintos significados históricos de la Probabilidad: intuitivo, laplaciano, frecuencial, subjetivo y teórico. El significado frecuencial de la Probabilidad, que le llamaremos probabilidad frecuencial, se basa en la Ley de los Grandes Números.

Hace unas tres décadas era difícil ver que la probabilidad frecuencial tuviera un uso práctico. Sin embargo, con el auge de los computadores, este uso ha evolucionado enormemente. Así, la simulación de experiencias aleatorias con software es una herramienta que ha permitido la solución de problemas desde el enfoque frecuencial.

La Didáctica de la Probabilidad también le ha dado un lugar privilegiado al enfoque frecuencial. Es común ver en eventos y congresos de didáctica propuestas sobre la enseñanza de la Probabilidad con simulación de experiencias aleatorias en Excel, Geogebra o R, principalmente. Sin embargo, quizás se está dando un uso abusivo de la simulación con software. ¿Se tiene conciencia de las limitaciones de esta herramienta? ¿Se controla y evalúa conscientemente la aplicación de esta herramienta? El presente trabajo pretende dar respuestas a estos interrogantes y hacer conciencia sobre el uso correcto de la simulación de experiencias aleatorias.

MARCO DE REFERENCIA

Se pretende analizar una situación problema cuya solución se puede obtener utilizando la simulación computacional para hallar la probabilidad frecuencial de un evento. Para dicho análisis, matemáticamente se utilizó la Teoría General de Probabilidades en especial la teoría sobre variables y vectores aleatorios. Seguidamente, se indican los aspectos teóricos a considerar para realizar el análisis desde el punto de vista didáctico.

Sobre los significados de Probabilidad, citados en la introducción, Batanero (2005) señala:

> … su enseñanza no puede limitarse a una de estas diferentes perspectivas, en razón de que están ligadas dialécticamente. La probabilidad puede contemplarse como razón de posibilidad a favor y en contra, como evidencia proporcionada por los datos, como grado de creencia personal y como modelo matemático que ayuda a comprender la realidad (Batanero, 2005, p. 260).

Por lo tanto, una buena propuesta sobre la enseñanza de la probabilidad con simulación computacional debe integrar los otros significados de probabilidad, y no aislar el significado frecuencial en un mundo aparte.

Así, es importante que el estudiante confronte la probabilidad frecuencial obtenida por medio de simulación computacional con el valor intuitivo que él tenía de dicha probabilidad y el valor teórico que obtiene al calcular formalmente la probabilidad. Esto se puede lograr no solo a nivel universitario, sino desde los primeros niveles de educación.

Pero ¿cuál sería el objetivo de resolver problemas con ambos modelos: teórico y frecuencial? Se puede caer en el error que el estudiante no perciba la utilidad de uno de los modelos por obviar el otro. Sin embargo, al inicio el objetivo es distinto, es que logre aprehender la compatibilidad entre significados. Más adelante, es posible mostrar al alumno que uno de los modelos (teórico o frecuencial) es insuficiente para resolver algún problema por lo que se debe recurrir al otro modelo para resolverlo, lo ideal sería que luego el estudiante pueda discriminar sobre cual modelo utilizar.

Por otro lado, sobre la resolución problemas, Polya (1965) señala la necesidad de educar la intuición propia para desarrollar una heurística o arte para resolver los problemas. Sin embargo, Schoenfeld (1985) expone que además de las heurísticas planteadas por Polya es necesario agregar tres dimensiones para tener éxito en la resolución de problemas: recursos (conocimientos previos), control (habilidad para monitorear y evaluar el proceso de resolución del problema) y el sistema de creencias (creencias sobre lo que es conocer y hacer matemática).

ASPECTOS METODOLÓGICOS

Se tomó como base un problema presente en uno de los ejercicios del libro Comprendiendo las Probabilidades (Sanabria, 2012). Primero se indica la solución (S1) teórica usual a ese problema, luego se presenta la solución (S2) utilizando simulación computacional dada usualmente por los estudiantes cuando se les ha asignado el problema como tarea. La incompatibilidad de (S1) y (S2) conduce a realizar un análisis riguroso para hallar la falla. A la luz de los aspectos teóricos se revelan detalles sobre la redacción del problema. Finalmente, producto del análisis se obtienen algunas consideraciones sobre la resolución de problemas utilizando simulación computacional.

DESARROLLO

En los problemas que se resuelven con simulación computacional se pueden tener los recursos: matemáticos (conocimientos previos) y computacionales (como el manejo de Excel). Además por medio de ejemplos y resolviendo problemas de esta índole, los estudiantes pueden adquirir cierta heurística para simular las experiencias aleatorias en un software. Sin embargo, las otras dimensiones (control y sistema de creencias) es posible que no se aborden.

La simulación computacional fácilmente puede motivar al alumno al ver una herramienta que rápidamente le simula muchas veces una experiencia aleatoria. Sin embargo su sistema de creencias puede ser decisivo en la resolución correcta de problemas. Dentro de este sistema de creencias se encuentra: su simpatía con la computadora, su concepción de probabilidad como parte de la matemática, la correspondencia entre la experiencia real y simulada (por ejemplo, ¿es consciente de que cada reglón de Excel representa una experiencia simulada o hace las cosas de forma algorítmica sin entender?), su concepción de azar.

Por otro lado, el control que se deje ejercer en la resolución de problemas con simulación computacional se ve comprometido por creencias o suposiciones sobre la distribución de las variables aleatorias involucradas. En estos problemas es esencial primeramente la existencia de una experiencia aleatoria a simular y asociarle, a esta experiencia, variables aleatorias que me permitan simularla. Sin embargo, en algunas ocasiones se asumen distribuciones sobre las variables involucradas sin saberlo.

EJEMPLO

Un segmento de un metro es cortado en dos puntos al azar formando tres segmentos ¿Cuál es la probabilidad de que con ellos se pueda formar un triángulo?

Usualmente, desde la experiencia personal la solución a este problema es propuesta a los estudiantes por medio de simulación en Excel.

SOLUCIÓN DETERMINANDO LA PROBABILIDAD TEÓRICA.

Sean $x, y \in [0,1]$ las medidas de los primeros dos segmentos, el tercer segmento mide $1 - x - y$:

Figura 1. Representación del segmento
Fuente: Elaboración propia

Así el conjunto de posibles resultados al cortar el segmento en dos puntos al azar es:

$$\Omega = \{(x, y) \in [0,1] \times [0,1] | 1 - x - y > 0\}$$

Por otro lado, considere el evento A: los tres segmentos obtenidos al cortar forman un triángulo. Para que estos segmentos formen un triángulo deben cumplir la desigualdad triangular, así:

$$\begin{aligned} A &= \{(x, y) \in \Omega | x + y > 1 - x - y \wedge x + 1 - x - y > y \wedge y + 1 - x - y > x\} \\ &= \left\{(x, y) \in \Omega \middle| x + y > \frac{1}{2} \wedge y < \frac{1}{2} \wedge x < \frac{1}{2}\right\} \end{aligned}$$

La representación de Ω y A en el plano cartesiano es:

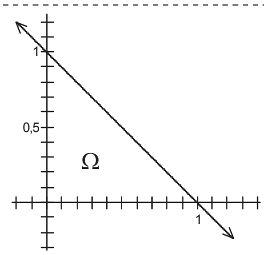

Figura 2. Espacio Muestral
Fuente: Elaboración propia

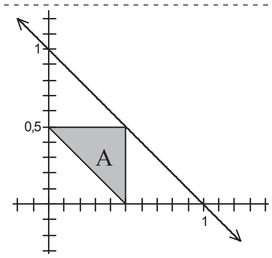

Figura 3. Evento A
Fuente: Elaboración propia

Así, se tiene que $P(A) = \frac{\text{área}(A)}{\text{área}(\Omega)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$.

SOLUCIÓN CON SIMULACIÓN

Es muy usual que los estudiantes resuelvan el problema de la siguiente forma utilizando simulación en Excel. ¿Cómo simular la experiencia aleatoria de cortar un segmento de 1 metro en dos puntos al azar? Tomando N como un número aleatorio entre 0 y 1, N será la medida del primer segmento. Esto quiere decir que los otros dos segmentos miden juntos $1 - N$ metros. Para dividir $1 - N$, se tomará otro número aleatorio M entre 0 y 1, que indicará el porcentaje del segmento $1 - N$ que corresponde al segundo segmento. Así, el segundo segmento medirá $M(1 - N)$. Así, se escribe en Excel:

	A	B	C	D	E
1	N	M	segmento 1	segmento 2	segmento 3
2	=ALEATORIO()	=ALEATORIO()	=A2	=B2*(1-A2)	=1-C2-D2

Figura 4. Primera simulación en Excel
Fuente: Elaboración propia

Luego, se escribe en la celda F1 la etiqueta: ¿cumplen la desigualdad triangular? Para responder esta pregunta se escribe en la celda F2:

=SI(MAX(C2:E2)<SUMA(C2:E2)- MAX(C2:E2); "SI"; "NO")

Para simular esta experiencia 500 veces, se seleccionan las celdas escritas de la fila 2 y con el mouse se arrastran estas fórmulas hasta la fila 501. Sin embargo el valor obtenido dista mucho de 0.25 que es el valor de la probabilidad teórica. Si se simula más veces, el valor andará cerca de 0.19 pero no de 0.25. ¿Que está fallando?

En el ejemplo anterior, no se está simulando correctamente la experiencia de cortar en dos puntos al azar el segmento. Tal parece que el control que se debe tener en la resolución del problema con simulación computacional se ve limitada por la intuición que se tiene de creer que todo está correcto. Pero el control de estas situaciones problema no es fácil, está dado por la probabilidad teórica. Surgen las dudas: ¿qué está mal en la simulación del ejemplo anterior? ¿Qué se está simulando?

Se procede a analizar y llevar un control del proceso de simulación del ejemplo anterior. Note que N y M siguen una distribución uniforme en [0,1], es decir $M, N \sim U[0,1]$. El segmento 1 mide N, el segmento 2 mide $Z = M(1-N)$ y el segmento 3 mide $1 - N - Z$. Note que Z no sigue una distribución uniforme (esto se retoma más adelante). Cada resultado (forma particular de cortar el segmento) queda determinado por los valores de N y M. Por lo tanto:

$$\Omega = \{(N, M) \in [0,1] \times [0,1]\}$$

El evento A, que los segmentos formen un triángulo, es:

$$\mathrm{A} = \{(\mathrm{N}, \mathrm{M}) \in \Omega | \mathrm{N} \leq \mathrm{Z} + 1 - \mathrm{N} - \mathrm{Z} \wedge \mathrm{Z} \leq \mathrm{N} + 1 - \mathrm{N} - \mathrm{Z} \wedge 1 - \mathrm{N} - \mathrm{Z} \leq \mathrm{Z} + \mathrm{N}\}$$

Simplificando y sustituyendo $Z = M(1-N)$ se obtiene:

$$\mathrm{A} = \left\{(\mathrm{N}, \mathrm{M}) \in \Omega \middle| \mathrm{N} \leq \frac{1}{2} \wedge \mathrm{M}(1-\mathrm{N}) \leq \frac{1}{2} \wedge \frac{1}{2} \leq \mathrm{M}(1-\mathrm{N}) + \mathrm{N}\right\}$$

Es decir,

$$\mathrm{A} = \left\{(\mathrm{N},\mathrm{M}) \in \Omega \middle| \mathrm{N} \leq \frac{1}{2} \wedge \frac{1-2\mathrm{N}}{2(1-\mathrm{N})} \leq \mathrm{M} \leq \frac{1}{2(1-\mathrm{N})}\right\}$$

Por lo tanto, para hallar $P(A)$ debemos utilizar la distribución conjunta de las variables independientes N y M. La función de densidad conjunta es:

$$f(n,m) = f_N(n)f_M(m) = 1 \text{ con } 0 \leq n \leq 1 \text{ y } 0 \leq m \leq 1$$

Por lo tanto,

$$\mathrm{P(A)} = \int_0^{\frac{1}{2}} \int_{\frac{1-2\mathrm{N}}{2(1-\mathrm{N})}}^{\frac{1}{2(1-\mathrm{N})}} \mathrm{dM\,dN} = \int_0^{\frac{1}{2}} \frac{\mathrm{N}}{(1-\mathrm{N})} \mathrm{dN} = \ln 2 - \frac{1}{2} \approx 0.193147$$

Vea que este valor calza con la simulación dada en el ejemplo anterior.

Resumiendo los ejemplos anteriores: se vio que la probabilidad teórica de que los segmentos formen un triángulo es de 0.25. Luego, por simulación computacional se obtuvo probabilidades cercanas a 0.19. Después, se realizó un control teórico de la simulación, obteniendo que la probabilidad teórica debería ser alrededor de 0.193147. Resultando dos probabilidades teóricas distintas, ¿dónde está el error?

Si se observa el análisis anterior se ve que Z no puede seguir una distribuición uniforme, pues:

$$F_Z(k) = P(Z \leq k) = P(M(1-N) \leq k) = P\left(M \leq \frac{k}{1-N}\right) = \int_0^{1-k} \int_{\frac{k}{1-N}}^{1} dMdN$$

Lo cual no dará una distribución uniforme. Esto quiere decir, que la longitud del segmento 2 no se escoge de forma uniforme de un conjunto de valores. Es más, su longitud no depende solo de M sino de la longitud del segmento 1, que es N. Así, la forma de realizar la simulación anterior implica una forma de cortar al azar los segmentos. El error se encuentra en la redacción del problema. En realidad las probabilidades teóricas no se contradicen, más bien corresponden a experiencias aleatorias distintas.

Así, la probabilidad teórica de que los tres segmentos formen un triángulo es de 0.25, cuando la experiencia aleatoria es:

E1: un segmento de un metro es cortado por una persona en dos puntos al azar formando tres segmentos, de la siguiente manera: la persona cierra los ojos elige dos puntos al azar del segmento, abre los ojos y corta en los puntos elegidos. Suponga que los cortes son independientes.

Además, la probabilidad teórica de que los tres segmentos formen un triángulo es aproximadamente de 0.193147, cuando la experiencia aleatoria es:

E2: un segmento de un metro es cortado por una persona en dos puntos al azar formando tres segmentos, de manera secuencial: primero corta el primer segmento al azar (cierra los ojos para elegir el punto donde cortar) y luego el resto lo divide al azar en dos segmentos.

Note que lo que diferencia las experiencias E1 y E2 son la independencia o no en los cortes. En la primera hay independencia de cortes y en la segunda no.

El control ejercido por la probabilidad teórica sobre la simulación permite observar que hay dos formas de interpretar la experiencia de cortar un segmento de un metro en dos puntos. Depende de la interpretación que se le haga, se obtienen resultados distintos. Esto señala la importancia que debe tener el definir la experiencia aleatoria y delimitarla correctamente cuando se intenta resolver un problema con probabilidad.

Por otro lado, lo que se simuló anteriormente es E1, ¿cómo simular correctamente E2?

SOLUCIÓN DEL EJEMPLO CON SIMULACIÓN

Dado que los cortes son independientes, para simularlos se requieren de dos números aleatorios M y N entre 0 y 1. Note que M y N son independientes. Estos representarán los cortes realizados del segmento de un metro. Por lo tanto, el segmento 1 mide el $MIN(N,M)$, el segmento 2 mide $|M-N|$ y el segmento 3 mide $1-MAX(M,N)$. Esto se puede simular en Excel como se hizo anteriormente.

Finalmente, se ha logrado resolver el problema planteado desde dos ópticas que dependen de cómo se interprete la experiencia aleatoria.

CONCLUSIONES

El trabajo ha permitido evidenciar la necesidad de que la probabilidad frecuencial no se aborde en solitario. Tal como lo dice Batanero (2015), debe haber una dialéctica entre los distintos significados de probabilidad.

Lo anterior no quiere decir que no se pueda utilizar el significado frecuencial para introducir a los estudiantes en el estudio de la probabilidad. Sin embargo, el profesor debe ser cuidadoso, y seleccionar solo aquellas simulaciones que el estudiante pueda justificar con sus conocimientos previos. Pues, dado que no se ha desarrollado el significado teórico, no puede llevar un control y justificar ciertas simulaciones. Si el estudiante las realiza se basará en suposiciones e intuiciones que le restan valor a la simulación.

Lo anterior concuerda con lo señalado con Brousseau (1986) sobre la paradoja de inadaptación a la exactitud: hay saberes para los cuales no se cuenta con situaciones problema adecuadas al nivel del alumno que permitan su enseñanza, por lo que si enseña el saber total este va a carecer de significado, o se opta por la enseñanza bajo la comprensión pero solo se enseñará un saber parcial. Así, pretender abordar el estudio de la probabilidad frecuencial antes de la probabilidad teórica no es lo mejor. Se recomienda un abordaje parcial de la probabilidad frecuencial con simulaciones al alcance del estudiante, después iniciar la enseñanza de la probabilidad teórica y luego retomar el significado frecuencial, profundizar su estudio y relacionarlo con el significado teórico.

Como se observó en los ejemplos y de acuerdo con Schoenfeld (1985), para la resolución de un problema con simulación computacional es necesario considerar:

1. La heurística o arte para resolver el problema. Por ejemplo, el abordaje de un problema de simulación con Excel implica desarrollar cierta habilidad para coordinar las funciones disponibles en Excel y los conocimientos.
2. Los recursos necesarios para obtener éxito en la simulación. Por ejemplo, los conocimientos previos y el manejo del lenguaje del software involucrado.
3. El control y monitoreo del proceso de simulación de la experiencia aleatoria. Como se evidenció, para tener una evaluación rigurosa de la simulación, para ciertas experiencias aleatorias, es posible recurrir a la probabilidad teórica y determinar si lo que se simula es realmente lo que se quiere simular.
4. Las creencias que se tengan. Por ejemplo, las creencias sobre qué es el azar, qué es la Probabilidad y sobre su relación con la matemática, que una experiencia es azarosa solo si su distribución es uniforme.

En particular, se comprueba en los ejemplos cómo un control de la simulación favorece una dialéctica entre significados de Probabilidad (teórico y frecuencial). Ese control es necesario para desmentir o esclarecer ciertos mitos e intuiciones. Sin embargo, los estudiantes deben ser conscientes de que no siempre es posible hacer un control de la simulación. De hecho, la esencia y el poder del modelo frecuencial es lograr resolver problemas en los cuales el modelo teórico se ve limitado.

Otro aspecto a considerar no solo en la resolución de problemas con simulación computacional sino en todo problema resuelto con un modelo probabilístico, es la delimitación y descripción de la experiencia aleatoria en el problema. Pocas veces se acostumbra a los estudiantes a describir la experiencia aleatoria involucrada antes de resolver el problema.

Como se observó en los ejemplos, el cómo se interprete la experiencia aleatoria involucrada influye en la resolución del problema. Esto permite plantear dos consideraciones importantes con respecto a la enseñanza de la Probabilidad:

1. Al inicio es importante que el profesor delimite bien la experiencia aleatoria involucrada en el problema por medio de su acertada redacción y que el estudiante, como parte de su compresión del problema, describa dicha experiencia. La delimitación le corresponde al profesor y la descripción al alumno.
2. Posteriormente parte del problema es que el alumno delimite y describa la experiencia aleatoria involucrada. Utilizar la probabilidad como un modelo para resolver un problema implica ver el problema en términos de una experiencia

aleatoria que debe ser bien definida, pues su definición influye en la solución del problema.

Sin duda la simulación computacional ha sido toda una novedad en la enseñanza de la Probabilidad. Pero se debe tener cuidado a la hora de implementar su enseñanza y tomar en cuenta consideraciones como las anteriormente expuestas.

REFERENCIAS

Batanero, C. (2005). Significados de la probabilidad en la educación secundaria. En R. Farfán y cols. (Eds.). *Relime, 8*(3), 247-263.

Brousseau, G. (1986). Fundamentos y métodos de la Didáctica de la Matemática. *Recherches en didactique des mathematiques*, *7*(2), 33-115.

Polya, G. (1965). *Cómo plantear y resolver problemas*. México: Trillas.

Sanabria, G. (2012). *Comprendiendo las Probabilidades*. Costa Rica: Editorial Tecnológica de Costa Rica.

Schoenfeld, A. (1985). *Mathematical Problem Solving*. Orlando: Academic Press.