TAMBIÉN EXISTEN BUENAS INTUICIONES EN PROBABILIDAD

Gabriel Yáñez
Universidad Industrial de Santander
Bucaramanga, Colombia

La literatura de la educación en probabilidad está llena de estudios que reportan las erradas intuiciones de los estudiantes y de la dificultad que implica cambiarlas por nuevas y acertadas. Ahora bien, de las buenas intuiciones poco se ha dicho, cuando se tiene claro que si se poseen es mucho el camino que se ahorra para lograr del estudiante aprehensiones conceptuales. En esta nueva dirección de rescatar las buenas intuiciones de los estudiantes, presento en este trabajo dos experiencias vividas en clase, que muestran que cuando el profesor adopta una actitud investigativa de escuchar a sus estudiantes puede descubrir ideas que, incluso, puede aprovechar para realizar desarrollos que nunca planeó, y descubrir ideas que le pueden permitir diseñar estrategias de enseñanza en temas considerados de mayor dificultad.

INTRODUCCIÓN

Si se quisiera caracterizar los temas que han ocupado la atención de los investigadores en educación estadística en los últimos treinta años, el estudio de las malas concepciones y/o intuiciones de los estudiantes respecto a los temas elementales de la probabilidad, sin duda ocuparía un lugar destacado. Desde los estudios sicológicos en la década de los 70 (Kahneman y cols. 1982) hasta los estudios propiamente didácticos en las décadas posteriores se observa esa línea de trabajo. Dado este estado de cosas, la pregunta obligada es: ¿qué se ha dicho respecto a las buenas intuiciones? Aunque existen trabajos destacables como los de Fischbein (1975) alrededor de la intuición de las frecuencias relativas que poseen las personas desde muy temprana edad, en general, no son muchos los reportes de investigaciones en este sentido.

Precisamente, este trabajo va en esa dirección ya que reporta dos experiencias vividas en el salón de clases que muestran las buenas intuiciones de los estudiantes en temas que, desde el punto de vista de su proceso educativo, son nuevos para ellos.
La primera situación se relaciona con una actividad realizada en los comienzos mismos de la teoría de la probabilidad relacionada con la probabilidad clásica y con estrategias de conteo. Esta actividad la presentamos en el siguiente apartado. La segunda se relaciona con temas de probabilidad condicional y conduce al diseño de una estrategia de solución de problemas relacionados con estos temas y que bautizamos como Enfoque Frecuencial Finito. Esta actividad la presentamos en el numeral 2. Finalmente realizamos algunas conclusiones relacionados con las experiencias realizadas.

SI LANZO TRES VECES TENGO EL TRIPLE DE PROBABILIDADES DE GANAR

En un curso de introducción a la probabilidad con profesores de educación media en ejercicio, se les propuso la siguiente situación: Imaginen que Colombia y Brasil juegan un partido de fútbol en el Campeonato Mundial de Fútbol y que se trata de decidir qué equipo inicia el partido. Con este propósito, el árbitro lanza un dado. Se trata de asignar resultados a los dos equipos que satisfagan las siguientes condiciones:

1. Que la selección sea justa.
2. Que la probabilidad de empezar Colombia sea el doble que la de Brasil.
3. Que la probabilidad de empezar Colombia sea el triple que la de Brasil.

Para las dos primeras situaciones los estudiantes dieron una variedad de válidas y esperadas respuestas. Para la tercera pregunta, de por sí la más interesante, se dieron varias e interesantes respuestas que presentamos en tres bloques distintos:

1. Una respuesta que coincide con el objetivo pensado a la hora de diseñar la pregunta: *Si se lanza el dado una vez, sea* x *el número de resultados a favor de Brasil y* $3x$ *el número de resultados a favor de Colombia. Como el total de resultados posibles es 6, se tiene que* $4x = 6$, *que no tiene soluciones enteras lo que indica que no es posible diseñar un juego con el requisito exigido.*

2. Respuestas asociadas a varios lanzamientos del dado caracterizados por un sesgo de equiprobabilidad de todos los resultados posibles, como el siguiente razonamiento:
"Se lanza el dado dos veces y se suman los resultados de cada dado. Como el total de posibilidades son 11 es imposible realizar la división requerida. El argumento de imposibilidad que usaron y que no explicitaron para el único lanzamiento de un dado, no lo explicitaron algebraicamente como en el caso anterior. Tal vez el hecho de que 11 es un número primo fue suficiente razón para argumentar la imposibilidad de la división.

3. Los intentos por resolver el problema con "sentido común" que generan la aprobación de los demás compañeros con expresiones como "lógico", "pues claro" y que muchas veces no son tan claras ni para aceptarlas ni para negarlas, pero que pueden generar discusión porque rompen el estado natural de las cosas en el salón de clase. La siguiente solución es de este tipo de respuestas: *El capitán de los brasileros lanza el dado una vez; después el capitán de Colombia lanza el dado tres veces y toma el mayor valor obtenido. El partido lo inicia el equipo que obtenga el mayor resultado. La solución se justifica por el hecho de que quien lanza tres veces tiene el triple de probabilidad de ganar que el que lanzó solamente una vez.*

Cuando se dio esta respuesta, el profesor y los estudiantes, excepto Ludwig, aceptaron la lógica de la respuesta. Ludwig dijo:

Ludwig: *Lo siento, pero yo no estoy de acuerdo con ustedes.*
Profesor: ¿Por qué*?*
Ludwig*: Yo no sé, pero yo no estoy tan seguro como ustedes.*
Profesor: *Está bien muchachos, tenemos que probarlo para convencer a Ludwig.*

Con la experiencia ganada al resolver los casos previos, el primer intento de solución es enumerar todos los posibles casos que se obtienen cuando Colombia lanza tres veces el dado, se calcula el resultado máximo y se compara con el resultado obtenido por Brasil. Luego se calculan los resultados a favor de Colombia y los favorables a Brasil y se calcula la razón entre ellos. Como el número de resultados posibles es $6^3 = 1296$ no era problema fácil contar los resultados enunciados, lo que obligó a estudiar variadas estrategias propuestas por los estudiantes. Después de muchos intentos se llegó a la siguiente estrategia de conteo:

1. El número de empates se corresponde con el número posible de resultados de Colombia: $6^3 = 1296$, ya que Brasil puede empatar si obtiene el mismo número que el máximo obtenido por Colombia.

2. El número de victorias de Brasil se obtiene considerando uno por uno sus posibles resultados y contando los resultados en que Colombia pierde.

Los cálculos se muestran en la Tabla 1.

Resultados de los Lanzamientos de Brasil	Número de veces que Colombia pierde
1	0
2	1^3
3	2^3
4	3^3
5	4^3
6	5^3
Total victorias de Brazil	$1^3 + 2^3 + \ldots + 5^3$

Tabla 1. Número de Empates entre Colombia y Brasil

3. Para encontrar el número de posibilidades a favor de Colombia, es suficiente con realizar la resta:

$$\text{RFC} = \text{Total} - \text{RFB} - \text{E} \qquad \text{donde}$$

RFC : Resultados a favor de Colombia
RFB : Resultados a favor de Brasil
E : Empates

Así, RFB = 225, RFC = 855 y la razón requerida es $\frac{855}{225} = 3{,}8$

Habiéndole dado la razón a Ludwig, era natural preguntarse por la razón de posibilidades si la razón entre lanzamientos es diferente a 3. La respuesta aparece en la Tabla 3.

Razón entre Lanzamientos	Razón entre probabilidades
2	2.27
3	3.8
4	5.62
5	7.79
6	10.04
7	13.5
8	17.1
9	21.5
10	26.8

Tabla 2. Razón entre probabilidades para razones dadas entre lanzamientos

La tabla muestra como la relación entre las probabilidades se hace mayor cuando la razón entre los lanzamientos crece.

Aclarada la inquietud inicial, y como producto de la discusión se analizó la situación desde un nuevo punto de vista: aumentar el número de caras del dado. ¿Será posible que con un dado que tenga más, o tal vez menos, caras que seis, la razón de los lanzamientos sea similar a la razón de las posibilidades?

Es decir, se trata de responder la siguiente pregunta:
Supongamos que existe un dado con cualquier número n de caras igualmente probables. Igualmente supongamos que dos jugadores, A y B, aceptan realizar un juego donde el que obtenga el mayor número al lanzar el dado es el ganador. El jugador A lanza el dado m veces y se queda con el número mayor, y el jugador B solamente una vez. ¿Cuál es la relación que existe entre las probabilidades de ganar de cada jugador cuando n, el número de caras del dado, aumenta?

Para ganar intuición acerca del resultados se desarrolló un programa en Mathematica que permitió calcular la razón entre los resultados a favor de cada jugador para diferentes valores de n (número de caras del dado) y m (número de lanzamientos del jugador A). Los resultados para los casos particulares m = 2, 3, 5, 10 y diferentes valores de n, se muestran en las siguientes gráficas.

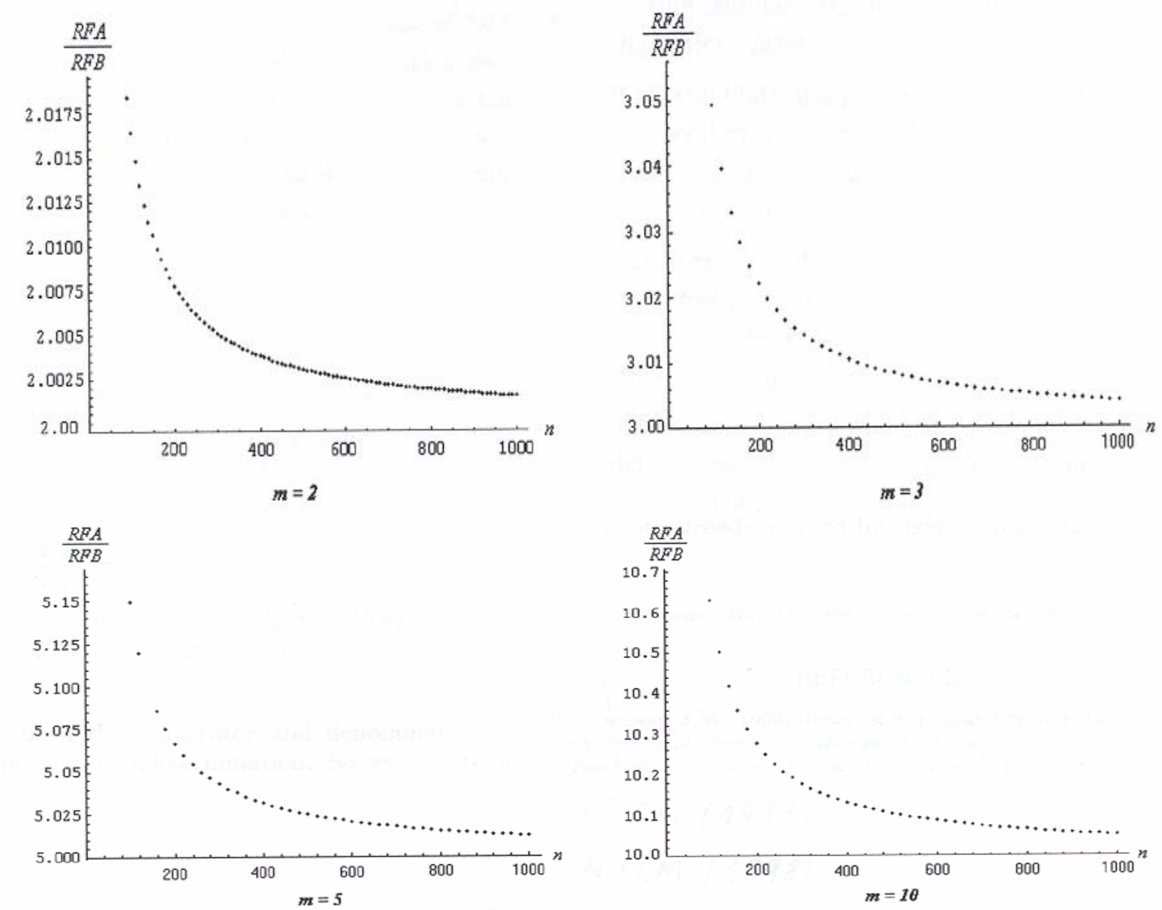

Gráfico 1. Razón entre posibilidades para diferente número de lanzamientos

De los gráficos se pueden conjeturar los siguientes resultados:

1. Para una razón dada entre lanzamientos, la razón entre las respectivas probabilidades de los dos jugadores decrece con el incremento del número de caras del dado.

2. La razón entre las probabilidades es más similar a la razón entre los lanzamientos cuando aumenta el número de caras del dado.

Llegados a este punto, se debía tomar la decisión de dejar las conjeturas y aceptarlas con base a los gráficos realizados, o intentar una demostración formal de estos hechos. En el caso particular que relatamos los estudiantes eran licenciados en matemáticas lo que garantizaba que la demostración no iba a causarles mayores dificultades. En todos los casos, abordar la demostración es un asunto relativo al tipo de estudiantes con los cuales se viva la experiencia.

El resultado que debemos probar formalmente lo podemos expresar de la siguiente forma:

$$\lim_{n\to\infty} \frac{RFA}{RFB} = m \quad (1)$$

donde RFA son los resultados a favor del jugador A, RFB los resultados a favor del jugador B, *n* el número de caras y *m* el número de lanzamientos del jugador A.
Con base en los resultados obtenidos en el caso de los tres lanzamientos, se sabe que

$$RFB = 1^m + 2^m + \ldots + (n-1)^m \quad (2)$$

y que

$$RFA = n^{m+1} - n^m - RFB \qquad (3)$$

Por lo tanto, el cálculo del límite (1) se reduce a calcular la suma

$$1^m + 2^m + \ldots + (n-1)^m \qquad (4)$$

para cualesquiera *m* and *n* números naturales.

Para los casos particulares $m = 1{,}2$, se sabe que

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

Calculando los límites correspondientes para estos dos casos se observa que se obtiene los límites 1 y 2 como se había conjeturado.

Aunque al principio el objetivo era encontrar una fórmula general para la suma de la expresión (4) este intento se dejó de lado por su dificultad. Consecuentemente, y después de analizar varios casos particulares, se conjeturó la siguiente expresión que luego se demostró por inducción para cualquier exponente *m* y cualquier número natural *n*

$$1^m + 2^m + \ldots + n^m = \frac{n^{m+1}}{m+1} + P(n) \qquad (5)$$

donde P(*n*) es un polinomio en *n* cuyo grado es menor o igual que *m.*

Remplazando (2), (3) and (5) in (1) se obtiene

$$\lim_{n\to\infty} \frac{RFA}{RFB} = \lim_{n\to\infty} \frac{n^{m+1} - n^m - \dfrac{n^{m+1}}{m+1} - P(n)}{\dfrac{n^{m+1}}{m+1} + P(n)} = m$$

después de dividir tanto el numerador como el denominador por n^{m+1} para suprimir la indeterminación. De esta forma llegamos al resultado esperado

Vale la pena notar, finalmente, el hecho de que el profesor no conocía la respuesta a la pregunta de los estudiantes lo que permitió que realizara un trabajo matemático real realizando varios intentos y cometiendo errores en frente de sus estudiantes e intentando diferentes estrategias.

LA RIFA DE LAS PIÑATAS

La segunda experiencia que presentamos se relaciona con las rifas que habitualmente se realizan en las fiestas de los niños. Concretamente, la situación es la siguiente:

En una piñata, a la cual asisten los padres de familia, los anfitriones realizan una rifa para las mamás presentes. La señora de la casa selecciona un número en cierto intervalo que escoge de acuerdo al número de participantes, y luego va preguntando, una por una, a las mamás por el número que han escogido. La mamá que adivine el número es la ganadora del regalo. La pregunta que se plantea es la siguiente: ¿Es esta rifa verdaderamente justa? ¿O depende del orden en que las personas vayan diciendo sus números?

Para simplificar el análisis supongamos que los asistentes son solo 5 y después de que la señora ha escrito en una papel el número ganador, entre 1 y 5, va preguntando a los invitados por su número en un cierto orden. ¿Cuál es la probabilidad de ganar que tiene cada uno de los 5 invitados?

Al preguntar a muchas personas por este problema, estudiantes universitarios y de colegio, personas de la calle, se obtuvieron, entre otras, las siguientes respuestas:

"Hay cinco opciones para cinco niños, la relación es 1 a 1, lo que implica que cada niño tiene la misma opción de ganar".
"Igual, puesto que hay una posibilidad para cada uno y depende de la suerte que tenga".
"*Depende si el primero falla, el segundo tiene mayor posibilidad; pero si el primero y el segundo fallan, la tendrá el tercero y así cada participante. Pero si el primero lo dice de una vez todos quedarán sin posibilidad".*

Si bien las dos primeras respuestas se limitan a establecer la relación biunívoca entre los participantes y los números posibles como razón única para aprobar la justicia del juego, la tercera respuesta ya deja entrever el juego de efectos contrarios que la situación plantea: los que escogen en los primeros lugares tienen más chance de poder participar, pero los que escogen al final tienen más probabilidad de acertar al quedar menos números por escoger. Esos efectos encontrados hicieron que algunos estudiantes no creyeran en la justicia del juego y escogieran el primer jugador o el segundo, o el tercero porque está en la mitad de los dos efectos, o el quinto porque si se llega hasta él es el seguro ganador.

Presentamos a continuación una respuesta dada por varios estudiantes de colegio sin estudios de probabilidad y que, como mostramos más adelante, encierra toda una estrategia general para resolver este tipo de problemas condicionales:

*"Sea 100% es el total de las posibilidades a repartir. De todos estas posibilidades le corresponden al primer jugador el 1/5*100%=20%. Para el segundo jugador solo quedan 80% de posibilidades, lo que hace que sus posibilidades de ganar sean ¼*80% = 20%. Cuando el tercer participante va a decir su número solo restan 60% de posibilidades por lo que las posibilidades que tiene son 1/3*60%=20%. Al ir a escoger el cuarto participante solo hay 40% de chances de ganar lo que hace que tenga ½*40%=20%. Las restantes 20% son las posibilidades del último participante.*

De esta respuesta basada en la intuición de los estudiantes, se pueden hacer varios comentarios:

- Los estudiantes adoptan la notación porcentual para definir un total de posibilidades que se deben distribuir entre los participantes. Si se piensa desde el punto de vista frecuencial, en esencia lo que se tiene es un total de 100 repeticiones de la rifa y se pretende calcular cuántas veces gana cada jugador.
- Estas posibilidades se distribuyen de acuerdo a un producto, siendo los factores una razón y un cierto número de posibilidades que varían de acuerdo al orden en que extraen.
- La razón es una probabilidad clásica que se calcula para cada jugador de acuerdo a la cantidad de números que quedan en el momento que él tiene que escoger. Así, para el primer jugador esta probabilidad es 1/5, para el segundo es 1/4, para el tercero es 1/3, para el cuarto es 1/2 y para el quinto es 1.
- Como bien se observa, estas probabilidades, tal y como está propuesta la rifa, son condicionadas a los resultados anteriores. Más concretamente, están condicionadas a que los participantes anteriores no hayan ganado, porque si alguno de ellos lo ha hecho sus probabilidades de ganar serían nulas.

¿Cómo se calcula el número de posibilidades que se le presentan a cada jugador? La forma que adoptan los estudiantes es hacer la diferencia entre las posibilidades iniciales y las que se ya se le asignaron a los jugadores anteriores. De esta forma, como al primer jugador se le asignaron 20% (20 de cada 100) cuando le toca el turno al segundo solo quedan 80%; igualmente como al segundo se le asignan 20 posibilidades, solo restan 60 al momento de entrar en acción el tercer participante. En igual forma para el cuarto y quinto participante.

En términos de probabilidad se puede considerar la probabilidad de perder del jugador que jugó antes y multiplicarla por las posibilidades que se le presentaron en el momento que él debía escoger su número. En el caso del primer participante, su probabilidad de perder es 4/5, como las posibilidades en ese momento son 100, las posibilidades en las que él no gana son las que se le presentan al segundo jugador: 4/5.100 = 80. La probabilidad de perder del segundo jugador es ¾, como las posibilidades son 80, las posibilidades en las que él no gana son ¾.80 = 60 y así sucesivamente.

Con el ánimo de respaldar teóricamente el procedimiento utilizado por los estudiantes, hablemos un poco del enfoque frecuencial de la probabilidad.

El enfoque frecuencial de la probabilidad supone que el experimento aleatorio se puede realizar repetidas veces bajo las mismas condiciones y que las repeticiones son independientes. La probabilidad de un evento A se calcula por la expresión

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N} \qquad (6)$$

donde N hace referencia al número de veces que se repite el experimento y N_A al número de veces en que el resultado es favorable al evento A. Identificamos el valor de esta probabilidad como *probabilidad frecuencial*.

En la práctica, el número de repeticiones es finito, lo que significa que realizado un cierto número *N* de ensayos, la probabilidad del evento A está dada por la expresión

$$P(A) \approx \frac{N_A}{N} \qquad (7)$$

donde el símbolo $\approx$ refleja que es una aproximación relativa a una serie *N* de ensayos del experimento aleatorio. Esta aproximación permite un manejo operativo con fracciones previo al cálculo del límite que define la probabilidad.

Al despejar N_A en (7) se obtiene una estimación del número de casos que son favorables a un cierto evento del que se conoce su probabilidad cuando se han realizado *N* repeticiones del experimento:

$$N_A \approx N \cdot P(A) \quad (8)$$

La integración de las expresiones (7) y (8) junto con el cálculo de *P(A)* asumiendo el enfoque clásico, es el fundamento de la solución implementada por los colegiales.
Para ejemplificar el método resolvamos el siguiente problema.

Se tiene una urna que contiene 2 bolas blancas y 2 bolas negras y se realizan sin sustitución 2 extracciones.

(i) ¿Cuál es la probabilidad de obtener una bola blanca en ambas extracciones?

(ii) ¿Cuál es la probabilidad de obtener una bola blanca en la segunda extracción?

Asumimos que se repite el experimento *N* veces. Según (8), dado que la probabilidad de extraer una bola blanca en la primera extracción es ½, existen aproximadamente ½ *N* casos donde la primera bola extraída es blanca; de éstos, 1/3 de este total son casos donde la segunda bola extraída también es blanca. Luego el total de casos donde ambas bolas son blancas es aproximadamente igual a $\left[\left(\frac{1}{2}\right)\cdot N\right]\times\left(\frac{1}{3}\right)=\frac{1}{6}N$. Al dividir por N obtenemos que la probabilidad de esta intersección es igual a $\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)=\frac{1}{6}$ que no es otra cosa que la regla del producto:

$$P(B_1\, y\, B_2) = P(B_1)\,P(B_2 \mid B_1)$$

Para interpretar los resultados que se obtienen al responder la segunda pregunta, realizamos un conteo de casos favorables al evento de que la segunda bola extraída es Blanca, contando los casos asociados con la primera extracción. Es decir, el total de casos donde la Segunda es blanca es igual a la suma de casos donde la primera y la segunda extracción son blancas, más los casos donde la primera es negra y la segunda es blanca. Considerando como antes un número N de casos generados, de acuerdo a (8) y con un análisis semejante al realizado para responder la primera pregunta se tiene que el número de casos es aproximadamente igual a $\left[\left(\frac{1}{2}\right)\cdot N\right]\times\left(\frac{1}{3}\right)+\left[\left(\frac{1}{2}\right)\cdot N\right]\times\left(\frac{2}{3}\right)=\frac{1}{2}N$. Por lo tanto, al dividir por N se obtiene la probabilidad de que la segunda bola extraída sea blanca. El resultado obtenido responde al Teorema de Probabilidad Total que para este caso particular tiene la forma:

$$P(B_2) = P(B_1\ y\ B_2) + P(N_1\ y\ B_2) = P(B_2 \mid B_1)P(B_1) + P(B_2 \mid N_1)P(N_1)$$

Vale la pena resaltar que el valor asumido para N no tiene ninguna importancia desde el punto de vista operativo (desaparece al calcular las frecuencias relativas), no obstante que la exactitud de los resultados simulados aumenta en la medida en que se aumenta el valor de N.

EL ENFOQUE FRECUENCIAL FINITO

Pensando en términos prácticos, proponemos la siguiente estrategia, que hemos llamado Enfoque Frecuencial Finito (EFF), para resolver problemas de probabilidad condicional:

Asuma un valor N de repeticiones del experimento aleatorio; aplicando la expresión (8) y el enfoque clásico de la probabilidad obtenga los valores necesarios que le permitan llenar la tabla de frecuencias asociada al problema dado.

Para mostrar como se utiliza la estrategia propuesta, resolvemos a continuación el problema del cáncer de mama que hicieron famoso Gigerenzer y Hoffrage (1995) en sus investigaciones sobre las ventajas de la información dada en frecuencias absolutas para resolver problemas de probabilidad condicional.

La probabilidad de cáncer de pecho es 1% en una mujer de cuarenta años que participa en un estudio de rutina. Si una mujer tiene cáncer de pecho, la probabilidad de que obtenga una mamografía positiva es 80%. Si no lo tiene, la probabilidad de que también obtenga un resultado positivo es 10%. Una mujer de un grupo de esta edad tuvo una mamografía positiva en un estudio rutinario. ¿Cuál es la probabilidad de que realmente tenga cáncer de pecho? _____%

Solución

Tomamos como *N* el valor 1000 y vamos calculando la cantidad de mujeres que forman parte de cada uno de los eventos marginales o de intersecciones de acuerdo a la información suministrada. Para organizar mejor las informaciones que se van obteniendo recurrimos a un diagrama de árbol:

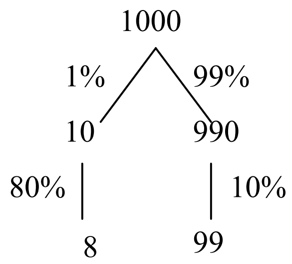

Gráfico 2. Diagrama de árbol asociado al problema del cáncer de pecho.

Con la información suministrada por el árbol podemos llenar la tabla con un poco de sentido común (entiéndase la versión de conteo del teorema de probabilidad total)

	Test Positivo	***Test Negativo***	***Total***
Con Cáncer	*8*	*2*	*10*
Sin Cáncer	*99*	*891*	*990*
Total	*107*	*893*	*1000*

Tabla 3. Tabla de frecuencias del problema del cáncer de pecho

Para responder la pregunta formulada tenemos que : P(Con Cáncer | Test Positivo) = 8/107 = 7.48%.

CONCLUSIONES

Las dos experiencias relatadas contienen algunos elementos que bien valen la pena destacar:

1. Definitivamente las voces de los estudiantes tienen muchas cosas que enseñarnos, no solamente para conocer sus malas intuiciones y/o concepciones, sino también para conocer las buenas y, a partir de ellas, diseñar actividades que los acerquen más fácilmente a los temas que pretendemos enseñarles.
2. Una de las tareas que debemos realizar los docentes se relaciona con la búsqueda de las razones que se esconden detrás de las intuiciones no explicadas de los estudiantes. Los dos ejemplos muestran que cuando el profesor asume ese trabajo, puede encontrar no solamente que lo que en principio le pareció desacertado, puede esconder algunas ideas que bien desarrolladas pueden conducir a resultados interesantes.
3. En el caso concreto del Enfoque Frecuencial Finito, se recomienda su adopción como estrategia de solución después de un trabajo intenso con el enfoque frecuencial que permita que los estudiantes diferencien entre el concepto de probabilidad asumida como el valor límite de las frecuencias relativas y el enfoque operatorio que se asume para deducir los resultados de la teoría de probabilidades y para resolver problemas.

REFERENCIAS

Fischbein (1975). The Intuitive sources of Probability Thinking in Children. Dordrecht: Reidel.

Gigerenzer, G., Hoffrage, U., (1995). How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review, 102, 684-704.

Kahneman, D., Slovic, ..Tversky, A. (1982). Reasoning under uncertainty: Heuristics and biases. Cambridge University Press.