

LA PROBABILIDAD Y LA RESOLUCIÓN DE PROBLEMAS

Agnelli, Héctor – Peparelli, Susana

Universidad Nacional de Río Cuarto – Argentina

Email: hagnelli@exa.unrc.edu.ar

Fundamentación

La resolución de problemas de probabilidades brinda la posibilidad de que el alumno comience a familiarizarse con la idea del modelado matemático de situaciones reales. Desde este punto de vista parece recomendable para el tratamiento del tema en clase trabajar con un conjunto de problemas prototípicos que permitan promover la discusión de los elementos básicos de la naturaleza aleatoria del fenómeno a modelar, la manera de construir el modelo, las herramientas a utilizar para hallar la solución, la interpretación que esta merece, y a la luz de los resultados obtenidos la confrontación del modelo con situaciones reales. Todas estas etapas pueden ser concebidas como componentes de un plan de trabajo similar al que se realiza cuando se lleva a cabo una tarea investigativa. Después de presentar cada uno de estos problemas sí se pueden dar para su resolución un listado de problemas genéricos que se resuelvan con el modelo adoptado.

Como ocurre ante una situación problemática de cualquier índole es necesario, para abordar su resolución, lograr la cabal comprensión del contexto en el que esta se origina. Los llamados problemas de probabilidad con expresiones tales como “calcular la probabilidad de...” son enunciados referidos a un fenómeno aleatorio que ocurre en el mundo real. Estos enunciados llevan implícito otro subyacente, “asuma que el fenómeno aleatorio tiene las propiedades postuladas en los axiomas de la Teoría de las Probabilidades y que por lo tanto, le son aplicables las conclusiones y teoremas deducidos en el sistema axiomático”. En la resolución de los problemas de probabilidades se utilizan modelos matemáticos que presuponen el establecimiento de supuestos que, en general no forman parte de los análisis previos a la resolución. Así es primordial efectuar el análisis de la situación aleatoria generadora del problema y la definición del mismo.

Los métodos formales de resolución de problemas de probabilidad dejan de lado las nociones intuitivas que el estudiante tiene acerca del problema, tanto en lo que se refiere a su planteamiento como al resultado esperado. Una vez obtenido el resultado analítico éste puede resultar sorprendente al estudiante si el hallazgo realizado contraría su intuición. Es conveniente seguir trabajando el problema desde distintos abordajes para reafirmar la convicción acerca del resultado hallado. Remover el pensamiento intuitivo ingenuo o al menos estar precavido de que el mismo puede llevar a conjeturas incorrectas es un muy buen aprendizaje, en el sentido que ante una nueva situación lo intuitivo será sometido a validación mediante el empleo de procedimientos formales.

La suposición básica más frecuente en la resolución de los problemas elementales de probabilidad es la de equiprobabilidad, además de la finitud de resultados posibles, pues estos supuestos permiten la aplicación de la fórmula de Laplace para arribar al cálculo de un resultado analítico. Pero la suposición de equiprobabilidad puede ser puesta en duda o al menos puede ser confrontada con “situaciones reales” en las que se presenta el fenómeno aleatorio bajo estudio. Por otra parte, una vez obtenida la solución del problema por métodos analíticos, la implementación de simulaciones y el análisis de estudios observacionales facilitarían la comprensión de la naturaleza aleatoria del resultado obtenido y la interpretación del concepto de probabilidad.

En este trabajo se analizan algunas de estas cuestiones, se propone una metodología para la resolución de problemas de probabilidades, ejemplificada con un problema típico de probabilidad: el problema de los cumpleaños.

El problema de los cumpleaños

¿Cuántas personas son necesarias para que sea mayor que $1/2$ la probabilidad de que al menos dos personas de n , que se encuentran en una habitación, cumplan años el mismo día?

Plan de trabajo

El plan de trabajo que proponemos contempla distintas etapas: analizar el escenario del problema, hacer una “lista” de ideas previas, definir formalmente el problema, realizar suposiciones para la elección del modelo, elegir las herramientas o fórmulas, efectuar la resolución, presentar las soluciones, hacer simulaciones, comparar con situaciones reales, hacer críticas al modelo.

1. Analizar el escenario del problema

En este problema se pide hallar un n , más precisamente el tamaño de un conjunto de personas que verifiquen una condición pero establecida ésta en términos de probabilidad. La condición hace referencia a la coincidencia en la fecha de cumpleaños entre las distintas personas que conforman un grupo, partiendo de la base que esta coincidencia no necesariamente ocurre a menos que se impongan condiciones de integración de las personas al grupo que hagan obvias las coincidencias, por ejemplo una reunión de hermanos mellizos. Por lo que, sin imponer condiciones constitutivas obvias podrían o no existir coincidencias en el grupo examinado. Y es interesante observar que podrían existir distintos tipos de coincidencias: pares, ternas, etc., de personas con igual fecha de nacimiento. Pueden por lo tanto ocurrir resultados diversos y de allí la naturaleza aleatoria del problema. En este caso se pide probabilidad mayor que $1/2$, es decir que sea más probable que ocurran las coincidencias a que todas las personas reunidas hayan nacido en días distintos. Pero es importante tener en cuenta que una vez hallado n , no necesariamente cada vez que se encuentre otra muestra de ese tamaño n hallado ocurrirán coincidencias, nuestra probabilidad sólo nos indica la posibilidad de ocurrencia a la larga. Si bien el cálculo del n se hará mediante un proceso de cálculo determinístico el fenómeno modelado es aleatorio.

2. Hacer una “lista” ideas previas

Los estudiantes tienen habitualmente intuiciones acerca del problema; o ideas acerca de cómo resolverlo. Estas ideas “listadas” serán después aceptadas o rechazadas cuando se alcance la solución del problema. Acerca de las intuiciones iniciales es oportuno mencionar lo señalado por Laplace: “La mente, como el sentido de la vista, tiene sus ilusiones, y así como el tacto corrige estas últimas, el pensamiento y los cálculos corrigen las primeras... Una de las grandes ventajas del cálculo de probabilidades es que nos enseña a desconfiar de nuestras primeras impresiones.”

En el problema de los cumpleaños, es posible que el alumno procediendo de acuerdo a su intuición pueda antes de calcular y formular formalmente el problema dar un valor aproximado de la respuesta buscada. En nuestra experiencia y la experiencia de otros, los resultados anticipados con este problema difieren fuertemente de los que luego se obtendrán con el cálculo. Aparece como natural pensar que a mayor cantidad de

personas en el grupo mayor será la probabilidad de ocurrencia del evento analizado y en general inicialmente se tiende a sobrestimar el valor de n .

3. Definir formalmente el problema

El problema consiste en hallar n tal que a partir de él se verifique

$$P(A_n) \geq \frac{1}{2}$$

Siendo A_n un evento de algún espacio muestral S que represente la fecha de cumpleaños de n personas y en A_n estén los elementos que representen al menos que dos personas, del grupo de n , cumplan años el mismo día

4. Suposiciones para la elección del modelo

El modelo a utilizar tiene dos partes constitutivas fundamentales: a) la definición de los resultados posibles es decir la configuración del *espacio muestral* y b) *la asignación de probabilidades* a estos resultados.

- Suposición 1.- (*construcción del espacio muestral*)

En ausencia de información que imponga alguna restricción en lo concerniente a la composición del grupo con relación a sus fechas de nacimientos, se puede asumir que los diferentes días de cumpleaños son 365 (se excluyen los años bisiestos).

Para facilitar la descripción se representará cada fecha de cumpleaños como un número de 1 a 365 y entonces cada resultado posible será una n -upla cuya i -ésima componente indica el día de cumpleaños de la i -ésima persona en el grupo

$$S = \{(s_1, \dots, s_n) / s_i = 1, \dots, 365, 1 \leq i \leq n\}$$

De aquí se desprende que el cardinal de S es

$$\#S = (365)^n$$

- Suposición 2.- (*asignación de probabilidades*).

Prescindiendo de las razones que puedan hacer que, en una sociedad particular, algunas épocas tengan preferencias por sobre otras para generar nacimientos, consideraremos todas las descripciones muestrales equiprobables.

5. Elección de herramientas o fórmulas

Entre las herramientas matemáticas más utilizadas en el abordaje de problemas probabilísticos encontramos la combinatoria y los conjuntos. Respecto de la primera Moore sostiene que “la combinatoria es un tema diferente y de mayor dificultad que la probabilidad”. Por lo que la enseñanza de la probabilidad centrada en la resolución de problemas de naturaleza claramente combinatoria no favorece la construcción del significado del objeto matemático “probabilidad”. En lo que respecta a los conjuntos, es claro que la notación conjuntista permite describir tanto el espacio muestral asociado a un experimento aleatorio como los eventos de interés, pero este aporte no es solo notacional ya que las operaciones entre conjuntos, aplicadas a la ocurrencia de eventos, permiten adoptar estrategias resolutivas que simplifican la primaria complejidad del problema enumerativo.

En nuestro problema, elegido el modelo se tiene que el espacio muestral es finito y los resultados son equiprobables por lo tanto se verifican los requisitos para aplicar la

fórmula de cálculo de probabilidades de Laplace es decir casos favorables sobre casos posibles, ahora la resolución del problema queda reducida a encontrar el cardinal del evento de interés, que formalmente puede ser expresado como:

$$A_n = \{(s_1, \dots, s_n) \in S / \exists k \geq 2, \{j_1, \dots, j_k\} \subset \{1, \dots, n\} \wedge s_{j_1} = \dots = s_{j_k}\}$$

6. Resolución

El computo de la probabilidad del evento, que existan coincidencias, implica computar todas las configuraciones del espacio muestral en las que al menos dos de sus componentes sean iguales, es decir dos, tres, hasta n coincidencias. Veamos por ejemplo como se haría el cálculo para un n conocido por ejemplo $n = 4$. En este caso tendríamos que contar el número de casos en que coinciden las cuatro personas, el número de casos en las que coinciden tres y el número de casos en las que coinciden dos de las personas, éste a su vez, lo podemos subdividir en dos situaciones: a) aquellas donde exactamente dos de las personas coinciden y las otras dos cumplen años en días diferentes a las personas que coinciden y diferente a la vez entre ellas, y b) hay dos grupos de coincidencias. Para efectivizar esta enumeración podemos asumir que el “juego” se realiza en diferentes etapas primero elegir la cantidad de grupos de coincidencias, después elegir el día en el que se presenta la coincidencia y luego elegir entre las personas restantes, si las hay, días de cumpleaños distintos.

Así, para cuatro personas que cumplan años el mismo día las posibilidades son 365

Para que tres personas cumplan años el mismo día

$$\binom{4}{3} 365 \cdot 364$$

Para que exactamente dos personas cumplan años el mismo día

$$\binom{4}{2} \cdot 365 \cdot 364 \cdot 363$$

Para que dos grupos de dos personas cumplan años el mismo día

$$3 \cdot 365 \cdot 364$$

De esta forma si llamamos A al evento de que exista al menos una coincidencia

$$P(A) = \#A / (365)^4 = 0.01635591247$$

Como se ve el cálculo resulta muy laborioso y existe una seria posibilidad de cometer errores en la aplicación del proceso combinatorio, notando además como la enumeración anterior se tornaría compleja si trabajamos con n indeterminado, como es nuestro caso. Por lo tanto parece adecuado elegir otra estrategia de resolución. En lugar de pensar en el evento “directo” pensar en el complementario, es decir, que todas las personas en el grupo cumplan años en días distintos y valernos de la fórmula

$$P(A) = 1 - P(A^c).$$

Particularizando en $n = 4$ y teniendo en cuenta que la no existencia de coincidencias implica que $\#A^c = 365 \cdot 364 \cdot 363 \cdot 362 = (365)_4$, queda

$$P(A) = 1 - (365)_4 / (365)^4 = 0.01635591247.$$

En general entonces resulta más fácil calcular el cardinal del evento complemento

$$A^c = \{s \in S / \forall i: 1 \leq i \leq n, \forall j: 1 \leq j \leq n, s_i \neq s_j \text{ para } i \neq j\}$$

Para calcular el cardinal de este conjunto se tiene en cuenta que la primera componente tiene 365 posibilidades, la segunda 364, y así sucesivamente, por lo tanto

$$\#A_n^c = 365 \cdot 364 \cdot 363 \dots (365 - (n - 1)) = 365_n$$

De esta manera

$$P(A^c) = \frac{365 \cdot 364 \dots (365 - n + 1)}{365^n} = \frac{365_n}{365^n}$$

La probabilidad buscada es

$$P(A_n) = 1 - P(A_n^c) = 1 - \frac{365_n}{365^n}$$

A continuación se presenta una tabla en la que se han calculado distintos valores de probabilidad en función del número de personas

n	10	21	22	23	24	25	30	80
$P(A^c)$.883	.556	.524	.493	.461	.431	.294	.00008
$P(A)$.117	.412	.476	.507	.539	.569	.706	.9999

De la tabla se desprende que basta con 23 personas para tener una probabilidad mayor a 1/2 para que exista alguna coincidencia.

Otra manera de solucionar el problema

Se siguen haciendo las dos suposiciones anteriores.

Sea A^c_i el evento que describe que el cumpleaños de la persona i no coincide con el de las anteriores $1 \leq i \leq n$, los tres primeros eventos serían

$$A_1^c = S$$

$$A_2^c = \{s \in S / s_2 \neq s_1\}$$

$$A_3^c = \{s \in S / s_3 \neq s_1 \text{ y } s_3 \neq s_2\}$$

Entonces

$$A^c = A_1^c \cap A_2^c \cap \dots \cap A_n^c \Rightarrow$$

$$P(A^c) = P(A_1^c) P(A_2^c / A_1^c) P(A_3^c / A_1^c \cap A_2^c) \dots P(A_n^c / \bigcap_{i=1}^{n-1} A_i^c)$$

$$P(A_1^c) = 1; P(A_2^c / A_1^c) = \frac{P(A_1^c \cap A_2^c)}{P(A_1^c)} = \frac{364 \cdot 365^{n-1}}{365^n} = \frac{364}{365}$$

$$P(A_3^c / A_1^c \cap A_2^c) = \frac{P(A_1^c \cap A_2^c \cap A_3^c)}{P(A_1^c \cap A_2^c)} = \frac{363}{365}$$

Entonces se tendrá igual que antes

$$P(A_n^c) = \frac{365_n}{365^n} \Rightarrow P(A_n) = 1 - \frac{365_n}{365^n}$$

Como se ve con este tipo de herramientas el uso de la notación y las operaciones con conjuntos es aún más necesaria para encontrar la solución del problema.

7. Reporte de la resolución

Es conveniente para afianzar la comprensión de lo realizado que el alumno presente un reporte que muestre los resultados, las suposiciones realizadas para alcanzarlos y la interpretación de los mismos. Como así también cualquier otro aspecto que el considere relevante y que haya sido utilizado en la resolución del problema.

8. Simulaciones

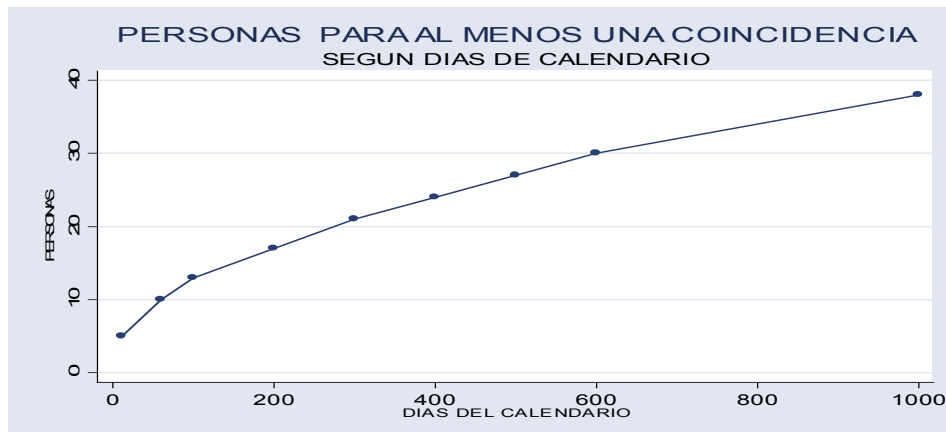
En este tipo de problemas también se puede reforzar la idea del vínculo entre frecuencia relativa y probabilidad realizando simulaciones. Estas simulaciones se utilizan como instrumento reafirmatorio de lo analítico y no necesariamente como supletorio.

Utilizando tablas de dígitos aleatorios (en una segunda etapa si se dispone se podrían hacer simulaciones con un computador) se simulan una o más muestras por alumno de tamaño 23. Una vez obtenidas las muestras y establecidas la frecuencia relativa de aparición del fenómeno se comparan estos valores con las probabilidades calculadas a priori. De hecho es posible que existan discrepancias notorias entre la frecuencia relativa y la probabilidad, dependiendo del muestreo realizado (cantidad de muestras). En cada ensayo (cada muestra obtenida por los estudiantes) se puede apreciar la presencia o ausencia del éxito (al menos una coincidencia en los cumpleaños), notando que la probabilidad de aparición del éxito es cercana a un $1/2$, aunque mayor, y hace que este tipo de repetición de ensayos se pueda asimilar a la de tiradas independientes de una moneda “legal”. La aproximación de la frecuencia relativa (convergencia en probabilidad) al valor calculado a priori para la probabilidad será mayor cuanto mayor sea el número de ensayos (ley de los grandes números). Se aprecia en algunos estudiantes una tendencia a transformar los resultados probabilísticas, desde el punto de vista de su interpretación, en resultados determinísticos del tipo 0-1 (0 no ocurre el evento, 1 el evento ocurre) pensando que un valor de la probabilidad mayor que $1/2$ implica ocurrencia “segura”. Así, existirá sorpresa cuando no aparezcan en una o más muestras, del tamaño calculado, ninguna coincidencia (ausencia de éxito). Por lo tanto la simulación aparece aquí como una herramienta o medio muy adecuado para comprender el significado operacional de la probabilidad en términos de frecuencia relativa.

9.- Generalización y comparación con situaciones reales

Con las dos suposiciones hechas el modelo construido es equivalente al que corresponde a la extracción con reemplazo de n bolillas, de igual forma, de una urna compuesta por 365 bolillas numeradas de 1 al 365. De esta manera la solución de nuestro problema es aplicable a cualquier otro fenómeno aleatorio que se pueda asimilar a este esquema. Se puede también generalizar para una urna de composición arbitraria en cuanto al tamaño, digamos N , y sacar entonces muestras de tamaño n ($n < N$). Por ejemplo buscar coincidencias en los meses de nacimiento ($N = 12$) o coincidencias en días pero en “años” de distinta duración ($N \neq 365$). En la figura se muestra, a manera de ejemplo, un gráfico con las personas necesarias para al menos una coincidencia, con probabilidad mayor que $1/2$, en el día del cumpleaños pero para “años” de distinta duración.

Por otro lado, resulta conveniente observar la frecuencia relativa del evento de interés situaciones problemáticas reales. A manera de ejemplo analizamos la fecha de cumpleaños de los 23 jugadores de cada uno de los 32 equipos que participaron en el último campeonato mundial de fútbol. Estas fechas fueron ubicadas en los registros de la página web de la Asociación Internacional de Fútbol. Se observaron 18 equipos en los que se presentaron coincidencias y 14 sin coincidencias, es decir la frecuencia relativa de al menos una coincidencia fue de 0.56.



10.- Críticas al modelo.

El supuesto básico de este modelo fue la equiprobabilidad, para poner a prueba este supuesto se puede realizar un estudio de los registros de nacimiento durante algunos años en los archivos de algún ente oficial. La no verificación del supuesto llevaría a construir modelos más refinados y por lo tanto de cálculo mucho más complejo y esto daría lugar a comentar la necesidad de la búsqueda de un compromiso entre fidelidad de representación y complejidad en la resolución al plantear modelos matemáticos de situaciones reales.

Conclusiones

Estimamos que la resolución de una gran número de problemas en los que se aplique rutinariamente metodología formal no contribuye a poner en evidencia las características esenciales del modelado matemático de los fenómenos aleatorios. Por el contrario trabajar problemas típicos, adecuadamente seleccionados, permite arrojar más luz sobre el cálculo y significado de la probabilidad.

Bibliografía

- Batanero, C. (2001) Aleatoriedad, Modelización, Simulación. X Jornadas de Aprendizaje y la Enseñanza de las Matemáticas, Zaragoza.
- Godino, J. D., Batanero, C. y Cañizares, M. J. (1987). Azar y probabilidad. Fundamentos didácticos y propuestas curriculares. Madrid: Síntesis.
- Leviatan, T. (2005) Teaching mathematics via problem solving- The case of probability. ICMI-EARCOME. Shangai, China
- Moore, D. (2003). La enseñanza agradable de las Matemáticas. Ed. Lynn Arthur Steen.
- Parzen, E. (1973) Teoría moderna de probabilidad y sus Aplicaciones. Limusa