Probabilidad

Miguel Molina

> *Vemos... que la teoría de la probabilidad es en realidad únicamente el sentido común reducido a cálculo; nos hace apreciar con exactitud lo que las mentes razonadoras sienten por una especie de instinto, sin ser muchas veces capaces de explicarlo...*
> *Es notable que [esta] ciencia, que nació al estudiar los juegos de azar, haya venido a constituir el objeto más importante del conocimiento humano.*
>
> *Laplace.*

Desde el punto de vista matemático, una *probabilidad* es una medida, es decir, una función que asigna números reales no negativos a determinados conjuntos y cumple algunas propiedades específicas. Estas propiedades –en el caso de la probabilidad- están reguladas por los axiomas de Kolmogorov y, si le preguntamos a un matemático qué cosa es una probabilidad, seguramente obtengamos como respuesta "todo lo que verifique los axiomas de Kolmogorov". En estas notas veremos una introducción elemental a la teoría matemática de la probabilidad, necesaria para comprender su utilización al evaluar la "fuerza" de algunos argumentos inductivos. Para esto último necesitaremos hacer algunas traducciones y reformulaciones de la teoría estándar de la probabilidad.

Espacio muestral.

Podemos pensar que la utilización de la probabilidad es adecuada cuando nos enfrentamos a la posibilidad de obtener diferentes resultados o sucesos, e ignoramos cuál de ellos se obtendrá. La probabilidad aparecería como un número asociado a cada suceso entre los que estamos considerando. Los ejemplos clásicos de estas situaciones son extraídos de los juegos de azar. Si tiramos un dado, podemos obtener como resultado 1, 2, 3, 4, 5, 6. Si tiramos una moneda podemos obtener como resultado "cara" o "número". Pero por supuesto, no son estas las únicas ocasiones en las que parece adecuada la utilización de la probabilidad. Podemos preguntarnos por el estado meteorológico del día de mañana, en cuyo caso, según nuestro interés, podríamos obtener como resultado "lluvioso" o "no lluvioso". Podemos también preguntarnos si una determinada persona, actualmente viva, lo seguirá estando dentro de diez años, y los resultados posibles serán "seguirá viva", "habrá muerto". Siempre que se habla de probabilidad entonces, estará determinado un conjunto llamado espacio muestral, cuyos elementos son los resultados posibles que caen bajo nuestro interés. El espacio muestral siempre es notado con la letra Ω.

Por ejemplo:

Si tiramos un dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$

Si tiramos una moneda, $\Omega = \{c, n\}$

Si tiramos una moneda, la recogemos y la volvemos a tirar, $\Omega = \{cc, cn, nc, nn\}$

Si elegimos un estudiante de facultad y le preguntamos la edad, $\Omega = \{x \in \mathbb{N} : x \geq 17\}$.

Observe en este último caso que tenemos un espacio muestral infinito en el cual figura como elemento el número 150 y todos los naturales mayores. Esto es posible, aunque sabemos perfectamente que no vamos a encontrar estudiantes de Facultad con más de 100 años. Por eso podríamos tomar como espacio muestral el siguiente conjunto $\Omega=\{x\in\mathbb{N}: 17\leq x \leq 100\}$. En general, en estas notas trabajaremos con espacios muestrales finitos haciendo anotaciones referentes a lo que sucede con espacios muestrales infinitos marcadas con una tipografía especial, pero estas anotaciones no serán necesarias para la comprensión del tema.

Podría parecer que el problema de la probabilidad quedaría resuelto una vez que se le asignara un número (su probabilidad) a cada uno de los elementos del espacio muestral. Por ejemplo, en el caso del dado, podríamos asignarle 1/6 a cada uno de los resultados y con eso no habría más que decir. Si esto fuese así, la probabilidad sería una función con dominio Ω y codominio los reales no negativos. **Pero esto no es así**, y es importante entender el porqué. Si definiésemos la probabilidad de ese modo, lo único que tendría probabilidad serían los elementos del espacio muestral. Pero frente a la tirada de un dado nos importa responder cosas como ¿cuál es la probabilidad de que salga un número par? ¿Cuál es la probabilidad de que salga un número primo? Ninguna de estas preguntas tendría respuesta si lo único que tiene probabilidad es que salga 1, que salga 2, …, que salga 6. Para responder nuestras preguntas necesitamos que la probabilidad esté definida sobre *subconjuntos* de Ω. Es decir, responder a la pregunta de cuál es la probabilidad de que salga un número par al tirar un dado es responder qué probabilidad se le asigna al conjunto {2, 4, 6}. De esta manera, y recordando que el conjunto $\mathscr{P}(\Omega)$ –llamado "partes de Ω"- es el conjunto cuyos elementos son los subconjuntos de Ω, vemos que lo correcto es definir la probabilidad como una función de $\mathscr{P}(\Omega)$ en los reales no negativos. **A los elementos de $\mathscr{P}(\Omega)$ se les llama *sucesos*** (son, salvo el vacío, conjuntos cuyos elementos son elementos de Ω). Además, si w es un elemento de Ω, llamaremos **suceso elemental** al conjunto {w}, o sea, **los sucesos elementales son los elementos unitarios de $\mathscr{P}(\Omega)$**.

Nota: *Esto que acabamos de decir no es válido en general. El lector con estudios de probabilidad sabe que en el caso finito, el dominio de la probabilidad puede ser cualquier álgebra de Ω, y en el caso infinito cualquier σ-álgebra, y que en general esta σ-álgebra se encuentra estrictamente incluida en $\mathscr{P}(\Omega)$.*

Por lo tanto diremos que una probabilidad sobre Ω es una función P cuyo dominio es $\mathscr{P}(\Omega)$, su codominio los números reales y cumple los axiomas de Kolmogorov:

1. $0 \leq P(A)$ para todo suceso A.
   (Toda probabilidad es no negativa, no existe ningún suceso cuya probabilidad sea, por ejemplo, -4. Nada obsta para que existan sucesos con probabilidad 0, y de hecho, veremos que es necesario que existan).

2. $P(\Omega)=1$.
   (La probabilidad del espacio muestral es 1. Como veremos en breve, esta es la máxima probabilidad que un suceso puede tener. El espacio muestral es el suceso conformado por todos los resultados posibles, de manera que le asignamos probabilidad 1 a "que suceda alguna de las cosas que pueden suceder".)

3. Si $A\cap B=\emptyset$, entonces $P(A\cup B) = P(A) + P(B)$ para sucesos A y B cualesquiera.
   (Si dos sucesos son disjuntos, es decir, ningún suceso elemental está en ambos a la vez, la probabilidad de la unión, o sea, de que suceda uno **u** otro es la suma de sus probabilidades. Por otra parte, por inducción se demuestra que siempre que se tenga una unión finita de conjuntos disjuntos, su probabilidad es la suma de las probabilidades de los conjuntos que se unen.)

Nota*: Obviamente esto solo funciona en el caso finito. En general, cuando tenemos definida la probabilidad sobre una* σ*-álgebra, el axioma 3 afirma que si* $\{A_i\}$ *es una familia numerable de conjuntos disjuntos dos a dos, entonces* $P(\cup A_i)=\sum_1^{\infty} P(Ai)$.

A partir de estos axiomas podemos demostrar todo lo que nos interesa acerca de la probabilidad. Veamos algunos teoremas:

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

(Esto es, la probabilidad de que **no suceda** A es igual a 1 menos la probabilidad de que **suceda** A).

Demostración:

Tenemos que $A^c\cap A=\emptyset$, ya que por definición, ningún elemento de A^c es elemento de A.

Por otro lado, tenemos que $A^c\cup A=\Omega$, ya que todo elemento de Ω que no es elemento de A es elemento de A^c. Por lo tanto

$$1 = P(\Omega) = P(A^c\cup A) = P(A^c)+P(A)$$

La primera igualdad se justifica por axioma 2, la segunda por lo que dijimos de la unión de A y A^c, y la tercera por axioma 3 teniendo en cuenta que estos dos conjuntos son disjuntos, como mostramos. Como ambas probabilidades suman 1, tenemos

$$\boxed{P(A^c) = 1 - P(A)}$$

En particular, como $\Omega^c=\emptyset$, tenemos que $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$

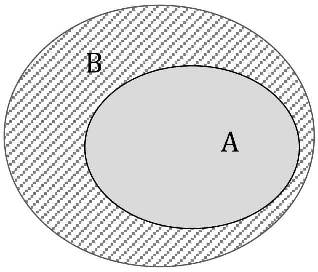
Ejemplo: Si la probabilidad que tiene un niño al nacer de llegar vivo a los 20 años es de 0,99; entonces la probabilidad que tiene de morir antes de esa edad es de 0,01.

Teorema 2. Si $A\subseteq B$, entonces $P(A)\leq P(B)$.

(Esto es, la probabilidad de un suceso incluido en otro es siempre menor o igual que la probabilidad del suceso que lo incluye. Obviamente, un suceso A está incluido en un suceso B si todos los sucesos elementales de A también lo son de B).

Demostración:

Como $A \subseteq B$, tenemos que $B = A \cup (B \setminus A)$, y esa unión es disjunta.



$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$, y como $P(B \setminus A)$ es no negativo, se obtiene el resultado buscado.

Comentario: Esto muestra que desde el punto de vista del cálculo de probabilidades, nunca puede ser mayor la probabilidad de algo como "Pía es cajera y hace yoga" que la de "Pía es cajera". A lo sumo podrán ser iguales. Veremos esto con mayor detalle al estudiar cómo trabajar con probabilidades de proposiciones.

Además, como se cumple que $A \subseteq \Omega$ para todo suceso A, y $P(\Omega) = 1$, se tiene que $P(A) \leq 1$ para todo suceso A.

Teorema 3 (Solo válido para el caso en que Ω es finito)

a) La suma de las probabilidades de los sucesos elementales es 1.

b) Si se conocen las probabilidades de todos los sucesos elementales menos uno de ellos, se conoce la probabilidad de cualquier suceso.

Demostración:

a) Sea $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$.

Entonces $\Omega = \{\omega_1\} \cup \{\omega_2\} \cup \ldots \cup \{\omega_n\}$ y la unión es disjunta.

Tenemos

$1 = P(\Omega) = P(\{\omega_1\}) + P(\{\omega_2\}) + \ldots + P(\{\omega_n\})$

Nota: Lo recién demostrado es que la suma de las probabilidades de los sucesos elementales debe ser 1, y nada más que eso. En particular, no mostramos que la probabilidad debe ser la misma para todos los sucesos elementales. Junto con lo que sigue, queda demostrado que cualquier asignación de probabilidad a los sucesos elementales que sume 1 sobre estos permite construir en forma única una probabilidad sobre $\mathcal{P}(\Omega)$. ***Ahora bien, todos sabemos que si preguntamos a alguien la***

probabilidad de que salga cara al lanzar una moneda, casi seguramente responderá "0,5", y si le preguntamos por la probabilidad de que salga 4 al lanzar un dado, casi seguramente responderá "1/6" (y hubiésemos obtenido la misma respuesta si la pregunta hubiera referido a la probabilidad de que salga otro número). Quien así responde está haciendo algo más que respetar los axiomas de la probabilidad. Está aplicando un principio llamado "de indiferencia" o "de razón insuficiente". Este principio indica que cuando no tenemos motivos o razones suficientes para considerar que un suceso elemental es más probable que otro u otros, debemos ser indiferentes ante ellos y considerar que todos tienen la misma probabilidad. Surge inmediatamente que al aplicar principio de razón insuficiente cuando Ω tiene n elementos, hay n sucesos elementales a cada uno de los cuales le asignamos la misma probabilidad, y la suma de esas probabilidades debe ser 1. Es necesario entonces que en esas condiciones tengamos

$P(\{\omega_i\})=1/n$ para todo suceso elemental $\{\omega_i\}$

Sin embargo, no hay nada en los axiomas de Kolmogorov que nos obligue a esta opción. El principio de indiferencia, si bien es el que usamos normalmente en una cantidad de situaciones, es talón de Aquiles de las teorías filosóficas de la probabilidad que hacen un fuerte uso de él. Cuando apliquemos principio de indiferencia sobre todo Ω, diremos que estamos aplicando "probabilidad clásica".

b) Por lo anterior, si conocemos la probabilidad de todos los sucesos elementales menos uno, en realidad conocemos la probabilidad de todos, ya que la probabilidad del restante se calcula como 1 menos la suma de las probabilidades de los conocidos. Entonces solo se debe demostrar que conociendo las probabilidades de todos los sucesos elementales es posible calcular la de cualquier suceso dado.

Sea A un suceso cualquiera. Si $A=\emptyset$, ya sabemos que $P(A)=0$.

Si $A\neq\emptyset$, entonces A tiene como elementos algunos sucesos elementales que llamaremos $\{\omega_{A1}\}, \{\omega_{A2}\}, ..., \{\omega_{Ak}\}$. Por lo tanto

$A= \{\omega_{A1}\} \cup \{\omega_{A2}\} \cup ... \cup \{\omega_{Ak}\}$ y esta unión es disjunta.

Entonces

$P(A)=P(\{\omega_{A1}\} \cup \{\omega_{A2}\} \cup ... \cup \{\omega_{Ak}\})= P(\{\omega_{A1}\}) + P(\{\omega_{A2}\})+ ... +P(\{\omega_{Ak}\})$.

Observación: En los casos en que aplicamos probabilidad clásica, es decir, cuando por aplicación del principio de razón insuficiente atribuimos la misma probabilidad a cada suceso elemental, estos resultados nos llevan a una expresión muy simple de la probabilidad de cualquier suceso. En particular, supongamos que Ω tiene n elementos, o sea $\Omega=\{\omega_1, \omega_2, ..., \omega_n\}$. En probabilidad clásica tenemos $P(\{\omega_i\})=1/n$ para todo suceso elemental $\{\omega_i\}$. Si A tiene k sucesos elementales $\{\omega_{A1}\}, \{\omega_{A2}\}, ..., \{\omega_{Ak}\}$ se cumple

$$P(A) = P(\{\omega_{A1}\}) + P(\{\omega_{A2}\}) + \ldots + P(\{\omega_{Ak}\}) = \underbrace{1/n + 1/n + \ldots + 1/n}_{\text{k sumandos}} = k/n$$

Observe que k es la cantidad de elementos de A y n la cantidad de elementos de Ω. Tenemos entonces que en probabilidad clásica se cumple lo siguiente:

Sea #X la cantidad de elementos del conjunto X (se le llama "cardinal de X"). Entonces, la probabilidad clásica indica:

$$P(A) = \frac{\#A}{\#\Omega}$$

Esto resume una célebre definición laplaceana de la probabilidad (clásica, por supuesto), que dice que la probabilidad de un suceso A es "la cantidad de casos favorables a A dividida entre la cantidad de casos posibles".

Ejemplo i.

Supongamos que alguien nos dice que tiene un dado trucado, de manera que sale el 6 aproximadamente la mitad de las veces que se tira. Además, el 1 no sale nunca, el 2 sale 1/10 de las veces que se tira, al igual que el 3 y el 4. Supongamos que se va a tirar ese dado, y que deseamos calibrar la probabilidad con las frecuencias conocidas. Observe que la calibración de las probabilidades de los sucesos elementales con sus frecuencias relativas siempre es posible, porque estas deben sumar 1. Queremos saber qué probabilidad debemos asignar a que salga un número primo.

En este caso tenemos Ω={1,2,3,4,5,6}, y deseamos averiguar la probabilidad del suceso A={2,3,5}

Según lo que se nos informó, atribuimos:

P(1) = 0 (Escribimos P(1) en vez de P({1}) por comodidad, aunque es un abuso notacional. La probabilidad se asigna a subconjuntos de Ω, no a elementos de ese conjunto).

P(2) = P(3) = P(4) = 1/10

P(6)=1/2.

Averiguamos P(5):

Como la suma de las probabilidades de todos los sucesos elementales debe ser 1, tenemos que

0 + 1/10 + 1/10 + 1/10 + 1/2 + P(5) = 1, por lo que P(5) = 2/10 = 1/5.

Entonces P(A) = P(2) + P(3) + P(5) = 1/10 + 1/10 + 2/10 = 4/10 = 2/5.

Ejemplo ii.

Supongamos ahora que también queremos calcular la probabilidad de que salga un número primo al arrojar un dado, pero no nos dicen nada sobre las características de este, y por eso nos parece adecuado aplicar el principio de razón insuficiente y por lo tanto probabilidad clásica. Tenemos

Ω={1,2,3,4,5,6},   #Ω=6

A={2,3,5}   #A=3

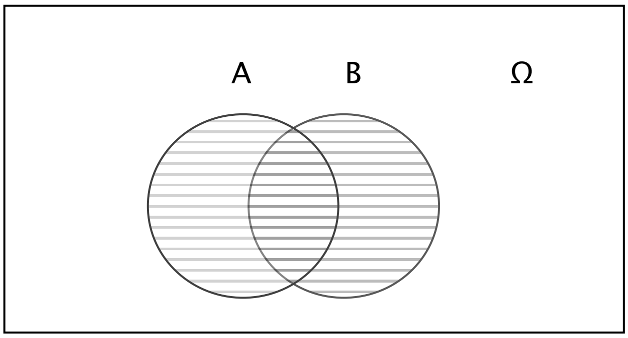
$$P(A) = \frac{\#A}{\#\Omega} = \frac{3}{6} = \frac{1}{2}$$

En este punto alguien puede preguntarse: Pues bien, ¿cuál es **la** probabilidad de A? La respuesta es que desde el punto de vista matemático, los dos resultados que hemos obtenido son correctos. Para decidir más allá de esto se necesita recurrir a la filosofía de la probabilidad, tema que será abordado posteriormente en el seminario.

Teorema 4.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Este teorema nos permite calcular la probabilidad de la unión de dos sucesos, siempre que conozcamos la probabilidad de cada uno de ellos y la de su intersección.

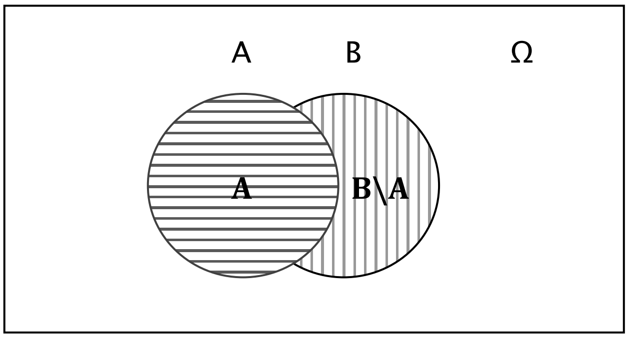


Nos interesa hallar la probabilidad de lo representado por la zona rayada, que tiene forma de "ocho acostado". Para hacerlo, no tenemos otro medio más que la utilización del Axioma 3, pero para ello es necesario que separemos esa zona en partes disjuntas.

Tenemos:

$A \cup B = A \cup (B \setminus A)$

El diagrama de Venn siguiente muestra que esto es así y también que la unión es disjunta. Ambos hechos son fácilmente justificables en forma sentencial, ya que si algo pertenece a la unión de A y B, entonces pertenece a A (o sea, pertenece a A y no a B o pertenece a A y a B) o pertenece a B y no a A, y recíprocamente. La unión es disjunta porque si un elemento pertenece a B\A, no puede pertenecer a A.

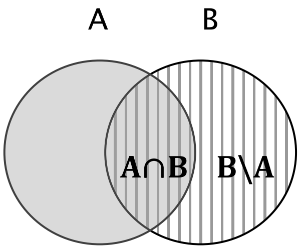


Por lo tanto, tenemos, aplicando el axioma 3

P(A∪B) = P(A) + P(B\A)   (1)

Concentrémonos en hallar la probabilidad de B\A. El diagrama siguiente muestra que podemos escribir

B = (A∩B) ∪ (B\A)   y que esta unión es disjunta



Por lo tanto, tenemos, aplicando el axioma 3

P(B) = P(A∩B) + P(B\A), de donde

P(B\A) = P(B) – P(A∩B)   (2)

Sustituyendo (2) en (1) obtenemos:

$$\boxed{P(A\cup B) = P(A) + P(B) - P(A\cap B)}$$

Ejemplo: Supongamos que de un mazo de cartas españolas (sin comodines) bien barajado extraemos una carta. ¿Cuál es la probabilidad de que la carta extraída sea impar o de copa?

La expresión “bien barajado” nos induce a aplicar principio de razón insuficiente y por lo tanto probabilidad clásica. El conjunto Ω tiene 48 elementos (uno por cada carta del mazo), el conjunto C (de las cartas de copa) tiene 12 elementos, y el conjunto I (de las cartas impares) tiene 24 elementos. Debemos calcular P(I∪C). Según lo que acabamos de ver

P(I∪C) = P(I) + P(C) – P(I∩C)

P(I) y P(C) se calculan inmediatamente a partir de los cardinales que tenemos. Por otro lado,

#P(I∩C) = 6, ya que el conjunto de referencia está formado por el 1, el 3, el 5, el 7, el 9 y el 11 de copa. Teniendo esto y aplicando probabilidad clásica obtenemos:

P(I∪C) = P(I) + P(C) – P(I∩C) = $\frac{\#I}{\#\Omega} + \frac{\#C}{\#\Omega} - \frac{\#(I \cap C)}{\#\Omega} = \frac{24}{48} + \frac{12}{48} - \frac{6}{48} = \frac{30}{48} = \frac{5}{8}$

Probabilidad condicional

La probabilidad condicional es un concepto de extrema importancia (es absolutamente fundamental para los intentos de dar respuesta al argumento de Hume sobre la inducción a partir de las interpretaciones subjetivas de la probabilidad, como veremos) que captura en cierta manera la idea de que unos sucesos dependen de otros. Comencemos a estudiarlo a través de un ejemplo.

Supongamos que se han tirado dos dados, uno rojo y uno verde. Se desea calcular la probabilidad de que la suma de los puntos obtenidos en cada dado sea 6. Parece adecuado aplicar principio de indiferencia, y por eso calculamos los cardinales de Ω y de (S=6) ("(S=6)" es el nombre del suceso "la suma de los puntos obtenidos es 6").

Ω es un conjunto de pares ordenados, en los que la primera componente representará los puntos obtenidos en el dado rojo, y la segunda, los obtenidos en el dado verde.

Ω={(1,1), (1,2),…, (1,6), (2,1), (2,2), …, (2,6), (3,1), (3,2), …, (3,6), (4,1), (4,2), …, (4,6), (5,1), (5,2), …, (5,6), (6,1), (6,2), …, (6,6)}

Es fácil ver que #Ω = 36.

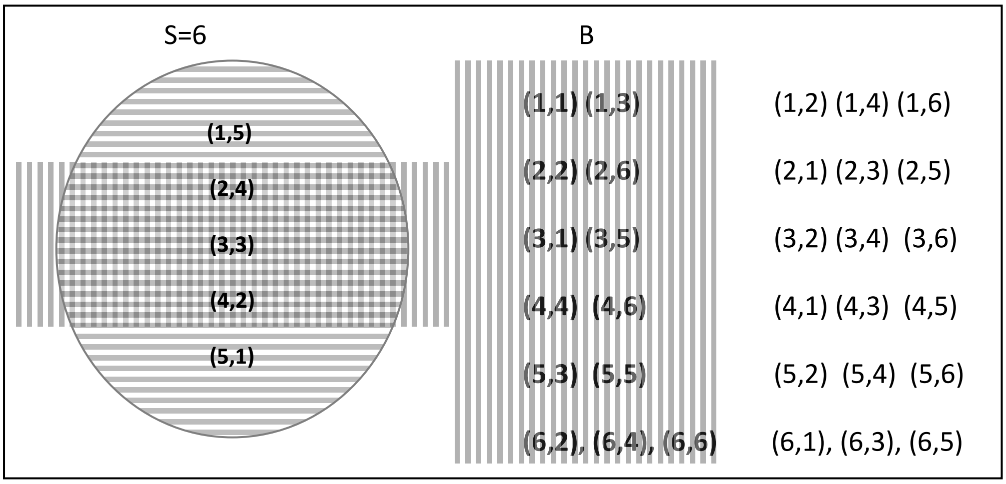
Por otro lado tenemos

(S=6) = {(1,5), (2,4), (3,3), (4,2), (5,1)}

De modo que #(S=6) = 5.

Entonces P (S=6) = $\frac{\#(S=6)}{\#\Omega} = \frac{5}{36}$

Supongamos ahora que se nos agrega la siguiente información: El puntaje que salió en ambos dados tenía la misma paridad (o sea que ambos son pares o ambos impares) y no salió 1 en un dado a la vez que 5 en el otro. ¿Cómo cambiamos el cálculo de probabilidad frente a la nueva información?



Llamemos B al suceso “ambos números tienen la misma paridad y no son (1,5) ni (5,1)”. Ese suceso está representado en el diagrama anterior por la zona rayada verticalmente. Ahora bien, si B ocurrió, son sus elementos los únicos que quedan como “posibles”. Es decir, nuestro “nuevo espacio muestral” pasa a ser B.

B={(1,1), (1,3), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), (4,2), (4,4), (4,6), (5,3), (5,5), (6,2), (6,4), (6,6)}

Estamos interesados en la probabilidad de que haya salido una suma igual a 6, **bajo la condición** de que haya sucedido B. Por lo tanto, consideramos los casos en que esto sucede, que son los doblemente rayados en el diagrama, o sea, los que conforman el conjunto (S=6)∩B.

(S=6)∩B = {(2,4), (3,3), (4,2)}

De manera que tenemos 16 casos posibles de los cuales 3 dan suma 6. Como seguimos bajo el principio de indiferencia, consideramos los 16 sucesos elementales igualmente probables (pero al ser los únicos posibles ahora cada uno de ellos tiene probabilidad 1/16), y consideramos la probabilidad de que la suma sea 6 **dado B** o **condicionada a B** como 3/16.

Si escribimos P((S=6)/B) para significar la probabilidad de (S=6) dado B, observe que lo que hicimos se puede representar como:

P((S=6)/B) = $\frac{P((S=6)\cap B)}{P(B)}$

Esto justifica la siguiente

Definición

Si P(B)≠0, *la probabilidad de A condicionada a B*, o *la probabilidad de A dado B* (notación: P(A/B)) es

P(A/B) = $\frac{P(A\cap B)}{P(B)}$

Se puede demostrar que P(./B) es una probabilidad para todo B de probabilidad no nula. Es decir, si recalculamos todas las probabilidades condicionadas a un suceso fijo, lo que obtenemos es una nueva asignación de números a los subconjuntos de Ω que verifica los axiomas de Kolmogorov.

Ejemplo:

Supongamos que en una clase de facultad averiguamos dos características de los presentes: el sexo y si fuman o no. Los resultados se resumen en la siguiente tabla:

	Hombres	Mujeres
Fuman	12	14
No fuman	17	22

A continuación, mediante un proceso aleatorizador (por ejemplo, escribir en papeles los nombres de cada uno de los estudiantes, colocar los papeles en una bolsa, revolver y extraer uno) se selecciona un estudiante. Nosotros no vemos el papel que salió, pero quien lo extrajo anuncia "Es una mujer". ¿Cuál es la probabilidad de que fume?

Lo que debemos calcular es la probabilidad de que fume **dado que es mujer**, no la mera probabilidad de que sea un fumador. Dando a los sucesos los nombres obvios, tenemos:

P(F/M)=$\frac{P(F\cap M)}{P(M)}$

Es claro que utilizando probabilidad clásica obtenemos

P(F∩M)=14/65, P(M)=36/65, de donde

P(F/M)=$\frac{P(F\cap M)}{P(M)} = \frac{14}{36} = \frac{7}{18}$

Observe que si hubiésemos calculado la probabilidad de seleccionar un fumador (sin condicionalizar) el resultado hubiera sido 26/65.

Independencia

Supongamos que en este momento un amigo me muestra una moneda y me dice que la va a tirar. Me pregunta qué probabilidad le asigno al suceso de que salga cara. Yo, que no sé nada sobre la moneda, aplicando principio de indiferencia le respondo que le asigno ½. Él, antes de lanzarla, me dice "Hoy es jueves. ¿Qué probabilidad le asignás a que salga cara dado que hoy es jueves?" Parece muy evidente que sería absurdo que yo cambiara mi respuesta anterior, dado que considero que el resultado de la tirada de la moneda es **independiente** del día de la semana en que se realice el lanzamiento.

Por otra parte, supongamos que estamos discutiendo con un amigo sobre si en tres días más pasarán más de 2000 personas por la cuadra de Tristán Narvaja entre 18 y Colonia desde las 13:00 hasta las 13:30. Yo, sin pensar mucho y considerando el movimiento de librerías, bares y facultades, le asigno una probabilidad muy baja, digamos 0,1. Pero entonces mi amigo me dice "Hoy es jueves". En este caso, estamos hablando del flujo peatonal en esa cuadra en domingo, y como todos sabemos, ese día y a esa hora está la feria, con lo que parece razonable reconsiderar drásticamente la probabilidad. Eso es porque consideramos que el flujo de peatones en esa cuadra es **dependiente** del día de la semana.

Veamos cómo se recoge esta idea a través de la probabilidad. Parece claro que si A es independiente de B, conocer que B sucede no afecta el cálculo previo de probabilidad de A. O sea, que tendríamos

P(A/B)=P(A).

Ahora bien, esto solo tiene sentido si P(B) no es cero, y además, no estamos seguros de que tengamos una condición simétrica (o sea, de que si definimos de esa manera la independencia, el hecho de que A sea independiente de B implique que B sea independiente de A, como evidentemente es deseable). Supongamos que ni P(A) ni P(B) son 0 y desarrollemos la condición impuesta arriba:

$P(A/B)=\frac{P(A\cap B)}{P(B)} = P(A)$

$P(B/A)=\frac{P(B\cap A)}{P(A)} = \frac{P(A\cap B)}{P(A)} = P(B)$

Se observa lo siguiente: Ambas condiciones implican $P(A\cap B)=P(A)P(B)$ y recíprocamente, si las probabilidades condicionales están definidas, la condición $P(A\cap B)=P(A)P(B)$ implica tanto P(A/B)=P(A) como P(B/A)=P(B).

De esta manera se define que dos sucesos A y B **cualesquiera** (no importa su probabilidad) son **independientes** si y solo si se cumple

$P(A\cap B)=P(A)P(B)$

Importante: la idea de dependencia no tiene nada que ver en principio con la de causalidad. Retomemos el ejemplo último, en el que se consideraba el sexo y si eran fumadores o no en un grupo de alumnos de facultad.

Consideremos los sucesos: "El alumno seleccionado es hombre" y "El alumno seleccionado fuma". Tenemos:

P(H)=29/65

P(F)= 26/65

$P(H\cap F)= 12/65\cong 0{,}1846$

$P(H)P(F)=754/4225\cong 0{,}1784$

Estos sucesos no son independientes y ni fumar es causa de ser hombre ni ser hombre es causa de fumar. Es verdad que dado que 0,1846 y 0,1784 son números muy cercanos, alguien podría pensar que los sucesos son "casi" independientes. Precisar esa idea requiere de nociones que al menos por ahora no veremos.

A veces postulamos la independencia de sucesos, y a partir de allí hacemos los cálculos pertinentes. Consideremos el siguiente problema:

Tres tiradores A, B y C apuntan a un blanco y disparan simultáneamente. Sabemos que A acierta el 80% de los tiros que dispara, B el 50% y C el 30%. ¿Cuál es la probabilidad de que al menos uno de ellos acierte en el blanco?

Si consideramos que hay dependencia entre los aciertos o yerros de los tiradores el problema se torna muy complicado. Pero considerando que hay independencia (como parece razonable, la probabilidad de que A acierte no parece cambiar si B y C aciertan, etc.), el problema admite una solución sencilla:

La probabilidad de que al menos uno de los tiradores dé en el blanco es 1 menos la probabilidad de que ninguno dé en él.

La probabilidad de que todos erren, suponiendo independencia, es

$P(A \text{ erre} \cap B \text{ erre} \cap C \text{ erre}) = P(A \text{ erre}) P(B \text{ erre}) P(C \text{ erre}) =$

$= (1-P(A \text{ acierte}))(1- P(B \text{ acierte})) (1-P(C \text{ acierte})) = 0{,}2 \,.\, 0{,}5.\, 0{,}7 = 0{,}07$

De manera que la probabilidad de que al menos uno acierte es

$1- P(A \text{ erre} \cap B \text{ erre} \cap C \text{ erre}) = 0{,}93$

Teorema de la probabilidad total

Intentemos ahora resolver este problema: Tenemos dos cajas, etiquetadas como A y B. En la caja A hay 5 bolas rojas y 3 negras. En la caja B hay 7 bolas negras y 2 rojas. Tiramos un dado. Si en el dado sale 1 o 2, extraemos una bola de la caja A revolviendo y sin mirar. Si en el dado sale 3, 4, 5 o 6, extraemos una bola de la caja B revolviendo y sin mirar. ¿Cuál es la probabilidad de extraer una bola roja?

Observemos que el suceso en el que estamos interesados (que salga una bola roja) se puede expresar como $R= (A\cap R) \cup (A^c\cap R)$ y que esta unión es disjunta. Es claro que por el planteo del problema, tenemos $A^c = B$.

Entonces es inmediato que

$P(R) = P(A\cap R) + P(A^c\cap R)$

Ahora bien, por definición de probabilidad condicional, podemos expresar la probabilidad de una intersección de la siguiente manera, siempre que la probabilidad del suceso condicionante no sea nula:

$P (A\cap R) = P(A)P(R/A)$ $\quad\quad$ $P (A^c\cap R)= P(A^c)P(R/A^c)$

Así obtenemos inmediatamente

$P(R) = P(A)P(R/A) + P(A^c)P(R/A^c)$

En nuestro caso, aplicando probabilidad clásica obtenemos:

$P(A) = 2/6 = 1/3$

$P(A^c) = 2/3$

$P(R/A) = 5/8$

$P(R/A^c) = 2/9$

por lo que

$P(R) = (1/3)(5/8) + (2/3)(2/9) = (1/3)(5/8 + 4/9) = (1/3)(45/72 + 32/72) = (1/3)(77/72) = 77/216$

Obviamente esto admite una generalización. Para plantearla vamos antes qué es una partición:

Definición: una partición de Ω es un conjunto de subconjuntos de Ω tales que

1. Ninguno es vacío
2. Son disjuntos dos a dos
3. Su unión es Ω

Teorema de la probabilidad total:

Sea $\{B_1, B_2, \ldots, B_n\}$ una partición de Ω tal que $P(B_i) \neq 0$ para todo i.

Entonces, siendo A un suceso cualquiera

$$\boxed{P(A)= P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \ldots + P(B_n)P(A/B_n)}$$

La demostración es totalmente análoga al razonamiento que se hizo para resolver el último problema.

Ejemplo:

En una imprenta, se imprimen 10000 volantes utilizando 4 impresoras, que llamaremos A, B, C y D. La impresora A mancha el 1% de las hojas que imprime, la B el 2%, la C el 3% y la D el 4%. De los 10000 volantes, se imprimen 3000 con la impresora A, 2000 con la B, 1000 con la C y 4000 con la D. Los volantes se empaquetan y se tiran todos juntos sobre los asistentes a un recital. Usted está en el recital y recoge uno. ¿Cuál es la probabilidad de que esté manchado?

Podemos aplicar la fórmula directamente, pero preferimos repasar el procedimiento. Los nombres de los sucesos son obvios:

{A, B, C, D} es una partición de Ω. Entonces

$M = (A \cap M) \cup (B \cap M) \cup (C \cap M) \cup (D \cap M)$ y la unión es disjunta

$P(M) = P(A \cap M) + P(B \cap M) + P(C \cap M) + P(D \cap M)$

Las probabilidades de A, B, C y D no son nulas de modo que podemos expresar las probabilidades de las intersecciones de la última fórmula de la siguiente manera:

$P(M) = P(A)P(M/A) + P(B)P(M/B) + P(C)P(M/C) + P(D)P(M/D)=$

$0{,}3 \cdot 0{,}01 + 0{,}2 \cdot 0{,}02 + 0{,}1 \cdot 0{,}03 + 0{,}4 \cdot 0{,}04 = 0{,}03$

Teorema de Bayes

El teorema de Bayes es un resultado muy simple de la teoría de la probabilidad, y como resultado matemático que es, se encuentra más allá de toda controversia. Sin embargo, ocupa un lugar absolutamente central en los intentos de responder al desafío de Hume a partir de una interpretación subjetiva de la probabilidad. Es decir, el teorema, informado con una determinada interpretación, forma el núcleo del bayesianismo, una de las corrientes epistemológicas con mayores pretensiones de justificar la inferencia no deductiva. El éxito de este programa sí es cuestión de controversia. Por supuesto, para comprender a cabalidad el bayesianismo, es imprescindible conocer la teoría matemática que le da fundamento.

Supongamos que tenemos dos sucesos A y B con probabilidades no nulas. Entonces se cumple:

$P(A/B) = \frac{P(A \cap B)}{P(B)}$, por lo que

$P(A \cap B) = P(B)P(A/B) \qquad (1)$

Además

$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \qquad (2)$

Sustituyendo (1) en (2)

$$\boxed{P(B/A) = \frac{P(B)P(A/B)}{P(A)}}$$

Y este sencillísimo resultado es el teorema de Bayes. Obviamente no se trata de nada matemáticamente recóndito, pero tiene un potencial enorme. ¿Por qué? Porque permite de alguna manera invertir una condicionalización, y eso parece de enorme importancia a la hora de evaluar hipótesis. Más explícitamente, muchas veces conocemos la probabilidad de que suceda A si sucede B, que es P(A/B). El teorema de Bayes nos permite –bajo ciertas condiciones- evaluar la probabilidad de que suceda B si es que sabemos que sucede A. Por ejemplo, supongamos que sabemos que la probabilidad de que un paciente tenga los síntomas S si tiene la enfermedad F es p. El teorema de Bayes nos permite evaluar, bajo ciertas condiciones, cuál es la probabilidad de que el paciente tenga F si presenta S. Por aquí se ve la punta del ovillo que guiaría hacia una justificación del razonamiento bajo incertidumbre: la evidencia serían los síntomas, y la utilización del teorema justificaría en un determinado grado la conclusión "el paciente tiene F". Por supuesto que es bastante más complejo y sutil que esto, pero la idea básica es esa.

Antes de ofrecer ejemplos, presentemos el teorema de otra forma:

Por razones que se harán evidentes más adelante, renombremos A y B como E (por evidencia) y H (por hipótesis) respectivamente.

La fórmula quedaría

$$\boxed{P(H/E) = \frac{P(H)P(E/H)}{P(E)}}$$

P(H) es la probabilidad *a priori* de H, P(E) la probabilidad *a priori* de E y P(H/E), calculada con la fórmula anterior, es la probabilidad *a posteriori* de H.

Hay más. Supongamos que tenemos una partición de hipótesis, es decir, un conjunto de hipótesis incompatibles (disjuntas) y exhaustivas (es decir, cuya unión es Ω).

Supongamos que $\{H_1, H_2, \ldots, H_n\}$ es una partición de Ω.

Entonces, por el teorema de la probabilidad total, tenemos que

$P(E) = P(H_1)P(E/H_1) + P(H_2)P(E/H_2) + \ldots + P(H_n)P(E/H_n)$

Entonces, tomando como H en la última fórmula recuadrada una cualquiera de las hipótesis de la partición H_i, obtenemos

$$P(H_i/E) = \frac{P(H_i)P(E/H_i)}{P(H_1)P(E/H_1) + P(H_2)P(E/H_2) + \ldots + P(H_n)P(E/H_n)}$$

De esta manera podemos evaluar la probabilidad de hipótesis alternativas a la luz de la evidencia E. Dejaremos la discusión de los problemas que esto plantea para el momento en que estudiemos filosofía de la probabilidad. Ahora veamos algunos ejemplos.

Ejemplo 1. Volvamos a un problema ya visto:

"Tenemos dos cajas, etiquetadas como A y B. En la caja A hay 5 bolas rojas y 3 negras. En la caja B hay 7 bolas negras y 2 rojas. Tiramos un dado. Si en el dado sale 1 o 2, extraemos una bola de la caja A revolviendo y sin mirar. Si en el dado sale 3, 4, 5 o 6, extraemos una bola de la caja B revolviendo y sin mirar." Habíamos averiguado la probabilidad de extraer una bola roja. Supongamos ahora que extraemos una bola, la miramos, y es roja. ¿Cuál es la probabilidad de que provenga de la caja A?

En este caso, tenemos que calcular P(A/R). Por Bayes:

$$P(A/R)=\frac{P(A)P(R/A)}{P(R)}$$

P(A) y P(R/A) se calculan directamente (y ya lo hicimos):

P(A)=1/3

P(R/A)=5/8

P(R) se calcula por teorema de la probabilidad total (también lo hicimos en el problema anterior). Tenemos

P(R) = 77/216.

Haciendo las cuentas, queda

P(A/R)=45/77

Ejemplo 2

Recordemos uno de los problemas del cuestionario:

"Usted es Juez en una ciudad en la cual el 85% de los taxis son verdes y el resto azules. En una noche con neblina, un taxi atropella a un peatón y huye del lugar. Muchas personas atestiguan que el peatón fue atropellado por un taxi pero afirman no haber podido distinguir el color. Solo un testigo dice que el taxi era de color azul y nadie dice que el taxi fuese verde. El testigo es sometido a pruebas bajo las mismas condiciones que las del día del accidente y resulta que reporta el color correctamente en el 80% de los casos; esto es con independencia del color del auto que le es presentado, afirma que el auto es del color correcto el 80% de las veces. Usted considera, a partir de esta información:

a) La probabilidad de que el taxi que protagonizó el accidente sea azul es 0,8.
b) Es más probable que el taxi que protagonizó el accidente sea azul que el que sea verde, pero la probabilidad es menor que 0,8.
c) Es tan probable que el taxi que protagonizó el accidente sea azul como que sea verde.
d) Es más probable que el taxi que protagonizó el accidente sea verde que el que sea azul."

Llamemos V al suceso "el taxi era verde", A al suceso "el taxi era azul", TV al suceso "el testigo dice que el taxi era verde" y TA al suceso "El testigo dice que el taxi era azul". Calculemos la probabilidad de que el taxi fuera azul a la luz de la evidencia, es decir, calculemos P(A/TA).

Por Bayes tenemos que

P(A/TA)= $\frac{P(A)P(TA/A)}{P(TA)}$

P(A) es la probabilidad a priori de que el taxi fuese azul. Como en esa ciudad el 15% de los taxis son azules le asignamos 0,15.

P(TA/A) es la probabilidad de que el testigo haya reportado correctamente que vio un taxi azul. Esa probabilidad es 0,8, ya que el testigo reporta correctamente el 80% de los casos con independencia del color que se le presenta.

P(TA) es la probabilidad a priori de que el testigo reporte haber visto un taxi azul. La podemos calcular usando probabilidad total, teniendo en cuenta que {A, V} es una partición de Ω.

P(TA) = P(A)P(TA/A) + P(V)P(TA/V)

P(V) es 1-P(A) = 0,85.

P(TA/V) es la probabilidad de que el testigo reporte haber visto azul cuando el auto era verde. Como el testigo falla el 20% de las veces, esta probabilidad es 0,2.

Tenemos entonces: P(A/TA)= $\frac{(0,15)(0,8)}{(0,15).(0,8)+(0,85).(0,2)} = \frac{12}{29} < 1/2$

De donde P(V/TA) = 1-(12/29) = 17/29 > 1/2 .

Por lo tanto, a la luz de la evidencia **y del conocimiento previo**, se encuentra que es más probable que el taxi fuera verde.

Otras cuestiones relativas a la condicionalización y a la utilización del teorema de Bayes.

A partir de la fórmula es evidente que al condicionalizar un suceso seguro (con probabilidad a priori 1) o imposible (con probabilidad a priori 0)[1], se obtienen probabilidades a posteriori idénticas a las que se tenía:

Supongamos que P(A)=0. Entonces

P(A/B)= $\frac{P(A\cap B)}{P(B)}$, pero como (A∩B)⊆A, se tiene que 0≤P(A∩B)≤P(A)=0, por lo que P(A∩B) = 0 y por lo tanto

P(A/B)=0.

Supongamos ahora que P(A)=1. Como P(./B) es una probabilidad, tenemos que

P(A/B) = 1 – P(A^c/B), y como

P(A^c)=0, por lo anterior, obtenemos P(A/B) = 1 – P(A^c/B) = 1 – 1 = 0.

Del mismo modo aunque en forma más evidente aun, es claro que si queremos evaluar a través del teorema de Bayes la probabilidad de una hipótesis dada una evidencia, si la hipótesis tiene probabilidad a priori extrema, la que se obtiene a posteriori es la misma que se tenía.

De esta manera se muestra que si utilizamos la condicionalización como método de ajuste sobre grados de creencia (una interpretación de la probabilidad), solo resulta en algo significativo cuando las probabilidades a priori no son extremas. Es decir, si hemos asignado una probabilidad 0 o 1 a un suceso o hipótesis, ninguna carga de evidencia modificará esa probabilidad por este método. **(Cuestión para pensar: ¿qué relación tiene esto con la monotonía de las inferencias deductivas?)**

[1] En espacios muestrales finitos lo es correcto decir que un suceso de probabilidad 1 es seguro y uno de probabilidad 0 imposible. Si el espacio muestral es infinito, esto no es así.

Veamos ahora una posibilidad de aplicar el teorema de Bayes para esgrimir un argumento escéptico acerca de la ocurrencia de milagros. Para la Real Academia Española, un milagro es un "*hecho no explicable por las leyes naturales y que se atribuye a intervención sobrenatural de origen divino*". Tomemos como ejemplo un milagro en el que creen millones de personas, la resurrección de Jesús, y hagamos algunos cálculos. Llamemos al suceso de que Jesús resucitó –algo claramente no explicable mediante las leyes naturales y que se atribuye a intervención sobrenatural de origen divino- RJ. ¿Cómo podemos evaluar la probabilidad de RJ? Bueno, obviamente, siendo que RJ, en caso de ser verdadero, es un hecho histórico, no una verdad de la lógica o de la matemática, tenemos que buscar la evidencia disponible. En nuestras condiciones, no parece que tengamos otra cosa que testimonios de escritores que relatan que Jesús resucitó. En particular, los autores de los cuatro evangelios están de acuerdo en el punto, según todos ellos, es el caso que RJ.

Evangelio según Mateo:

"28:1 Pasado el día de reposo, al amanecer del primer día de la semana, vinieron María Magdalena y la otra María, a ver el sepulcro.
28:2 Y hubo un gran terremoto; porque un ángel del Señor, descendiendo del cielo y llegando, removió la piedra, y se sentó sobre ella.
28:3 Su aspecto era como un relámpago, y su vestido blanco como la nieve.
28:4 Y de miedo de él los guardas temblaron y se quedaron como muertos.
28:5 Mas el ángel, respondiendo, dijo a las mujeres: No temáis vosotras; porque yo sé que buscáis a Jesús, el que fue crucificado.
28:6 No está aquí, pues ha resucitado, como dijo. Venid, ved el lugar donde fue puesto el Señor.
28:7 E id pronto y decid a sus discípulos que ha resucitado de los muertos, y he aquí va delante de vosotros a Galilea; allí le veréis. He aquí, os lo he dicho.
28:8 Entonces ellas, saliendo del sepulcro con temor y gran gozo, fueron corriendo a dar las nuevas a sus discípulos. Y mientras iban a dar las nuevas a los discípulos,
28:9 he aquí, Jesús les salió al encuentro, diciendo: ¡Salve! Y ellas, acercándose, abrazaron sus pies, y le adoraron."

Evangelio según Marcos:

"16:9 Habiendo, pues, resucitado Jesús por la mañana, el primer día de la semana, apareció primeramente a María Magdalena, de quien había echado siete demonios."

Evangelio según Lucas:

Según Lucas, quienes fueron a visitar el sepulcro lo hallaron vacío, y se encontraron con unos ángeles que les anunciaron que RJ era verdadero. Jesús mismo aparece después, a unos discípulos muy desanimados por la muerte de su maestro que se encaminan a Emaús. El extenso pasaje referido va de Lucas 24:1 a 24:36.

Evangelio según Juan:

"20:11 Pero María [Magdalena] estaba fuera llorando junto al sepulcro; y mientras lloraba, se inclinó para mirar dentro del sepulcro;
20:12 y vio a dos ángeles con vestiduras blancas, que estaban sentados el uno a la cabecera, y el otro a los pies, donde el cuerpo de Jesús había sido puesto.
20:13 Y le dijeron: Mujer, ¿por qué lloras? Les dijo: Porque se han llevado a mi Señor, y no

sé dónde le han puesto.
20:14 Cuando había dicho esto, se volvió, y vio a Jesús que estaba allí; mas no sabía que era Jesús."

Esta parece ser toda la evidencia de que disponemos a favor de RJ. (Estamos imponiendo el criterio de que la evidencia debe ser pública. Alguien puede decir que tiene evidencia privada de RJ, por ejemplo, puede decir que Jesús se le apareció y le dijo que había resucitado. Para quienes no poseen ese tipo de evidencia y, dada la innegable correlación entre testimonios de ese tipo y desórdenes mentales, tienden a no considerarlos, aparentemente toda la evidencia reposa en el testimonio trasmitido por los evangelistas.)

Llamemos a la evidencia disponible TE (testimonio de los evangelistas). Deseamos calcular entonces la probabilidad de RJ dado TE. El teorema de Bayes nos dice que

$$P(RJ/TE)=\frac{P(RJ)P(TE/RJ)}{P(TE)}$$

Aquí empezamos a ver las dificultades de la aplicación del teorema de Bayes a casos como este. Para aplicarlo, tenemos que aproximar razonablemente P(RJ), P(TE/RJ) y P(TE).

Veamos cada uno de esos factores:

1- P(RJ)
   Esta es la probabilidad *a priori* que le asignamos a que Jesús haya resucitado. Es la probabilidad que le atribuimos sin tomar en cuenta la evidencia. Alguien podría decir que sin evidencia a favor, le asigna simplemente 0. Pero en ese caso, lo que hacemos no serviría de nada, ya que el teorema de Bayes arrojaría que P(RJ/TE) también es cero. Además, suponer P(RJ)=0 es considerar el caso más extremo de rechazo a la hipótesis, significa estar absolutamente seguro de su negación. Supongamos que no adherimos a esto *a priori*, sino que aceptamos *la posibilidad* de que JR sea verdadero. Esto implica asignarle una probabilidad positiva. Puede ser muy pequeña, pero no puede ser 0. Una posibilidad razonable parece ser la siguiente: si le preguntamos a un cristiano cuántas personas resucitaron, seguramente nos responda que como máximo tres: Jesús, Lázaro, y la hija de Jairo, a quien Jesús levanta con las *ipsissima verba* "talita kumi". Bien, aceptemos entonces lo siguiente: la resurrección es algo tan raro que cada ser humano que ha existido ha tenido una probabilidad de resucitar igual a tres dividido el total de personas que han existido. ¿Cuántas personas han existido, entonces? Obviamente no se sabe con exactitud, pero el PRB (Population Reference Bureau) ha calculado que hasta 2002 habían nacido 106.456.367.669 personas[2]. Favoreciendo un poco la probabilidad de RJ pero simplificando las cuentas, aceptemos que han vivido 10^{11} personas. Parece razonable entonces asignarle una probabilidad *a priori* a RJ igual a 3.10^{-11}. No se asombre el lector por la pequeñez del número, es simplemente el reflejo de que la resurrección, caso de haber ocurrido, es efectivamente algo muy apartado de lo probable. Viene a ser algo apenas cien mil veces más improbable

2 http://www.prb.org/SpanishContent/Articles/2002/CualeselnumerototaldepersonasquehanvividoenlaTierra.aspx

que sacar el cinco de oro con un número, cosa con la que creo podemos estar de acuerdo.

2- P(TE/RJ)

Este factor parece más fácil de calcular, es la probabilidad de que tengamos el testimonio si la resurrección ocurrió. Cuanto más alto lo hagamos, más alta resultará la probabilidad de RJ dado TE. Parece razonable decir que esta probabilidad es casi 1, dado que la resurrección sería un hecho tan asombroso que difícilmente podría alguien presenciarla sin dejar testimonio de ella. Sin más comentarios, hagámosla 1.

3- P(TE)

Esta es la probabilidad que asignamos *a priori* a tener el testimonio de la resurrección. Para ayudarnos en su cálculo, apliquemos el teorema de la probabilidad total:

$P(TE) = P(TE/RJ)P(RJ) + P(TE/RJ^c)P(RJ^c)$

Observamos que tenemos todo lo necesario para calcular el primer sumando (que es $1.3.10^{-11} = 3.10^{-11}$, y que en el segundo sumando tenemos el factor

$P(RJ^c) = 1 – P(RJ) = 1 – 3.10^{-11}$.

Falta estimar $P(TE/RJ^c)$, que es la probabilidad que asignamos a tener el testimonio si no ocurrió la resurrección, o sea, la probabilidad que asignamos a que los evangelistas se hayan equivocado al trasmitir su testimonio o hayan mentido deliberadamente. Por lo tanto, está estrechamente relacionado con la **desconfianza** en el testimonio. Observemos que si asignamos

$P(TE/RJ^c)=0$

(O sea, si suponemos que es la probabilidad de tener el testimonio en caso de no haber ocurrido la resurrección es nula, lo que representa un grado máximo de confianza en el testimonio evangélico), el teorema de Bayes nos daría:

$$P(RJ/TE)=\frac{3.10^{-11}.1}{3.10^{-11}+0.(1-3.10^{-11})}=1$$

como es obvio, ya que si asignamos confianza **total** a un testimonio, debemos creer absolutamente lo que el testimonio dice.

Por otro lado, si asignamos

$P(TE/RJ^c)=1$

(O sea, si suponemos que es **seguro** que los evangelistas dan un testimonio que no merece ninguna consideración, el grado mínimo en la confianza en el testimonio evangélico), el teorema de Bayes nos diría:

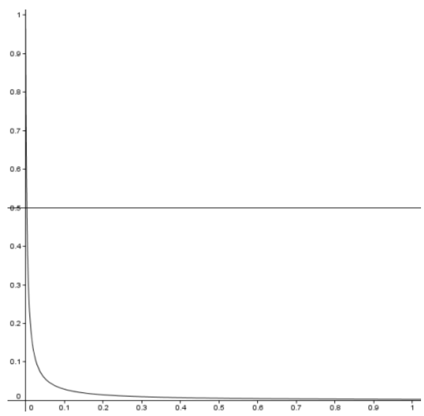
$$P(RJ/TE)=\frac{3.10^{-11}.1}{3.10^{-11}+1.(1-3.10^{-11})}=\frac{3.10^{-11}.1}{1}=3.10^{-11}$$

Como también es obvio, el testimonio de alguien que no merece confianza alguna no puede cambiar nuestra consideración previa de la probabilidad de un suceso.
Entonces, ¿qué grado de confianza asignamos a los evangelistas? Hay algo que parece claro: no es razonable asignarle cero ni uno. No podemos asignarle 0 porque los evangelistas narran algunos hechos que no solamente son posibles, sino que además los tenemos por verdaderos. Por ejemplo, sabemos que Poncio Pilato fue efectivamente prefecto en Judea, y esto por fuentes independientes a los evangelios. Tampoco podemos asignarle 1, porque es muy claro que los evangelios son escritos contradictorios. Si miramos, por ejemplo, las narraciones trascriptas más arriba sobre la resurrección, veremos que cuentan cosas diferentes relativas a la primera aparición de Jesús. Pero entre cero y uno hay infinitos valores, y sería imposible llegar a un acuerdo acerca de qué grado de confianza debe atribuirse al testimonio. En vez de eso, lo que podemos hacer es lo siguiente: Preguntémonos qué valor debería tener P(RJ/TE) para que considerásemos racional la creencia en RJ a partir de la evidencia. Nuevamente, frente a esta pregunta, puede haber matices subjetivos. Quienes tengan patrones altos, preferirán decir que es racional "creer" algo si su probabilidad es al menos 0,8; mientras que otros, más arriesgados, estarán conformes con que sea al menos 0,7. Pero en algo podemos estar de acuerdo: para que sea racional sostener una creencia, su probabilidad debe ser al menos ½, ya que si su probabilidad es menor que ½, la probabilidad de la negación de lo que estamos creyendo es mayor que la de lo que creemos. Entonces podemos preguntarnos cómo debería ser $P(TE/RJ^c)$ (el último factor que nos falta calcular) para que P(RJ/TE) sea **al menos** ½.

Por teorema de Bayes tenemos

$$P(RJ/TE)=\frac{3.10^{-11}.1}{3.10^{-11}+(1-3.10^{-11})P(TE/RJ^c)}$$

Vemos que P(RJ/TE) es **función** de $P(TE/RJ^c)$, o sea, para cada valor de $P(TE/RJ^c)$ entre 0 y 1, obtenemos un valor de P(RJ/TE). Si graficamos esta función, colocando en el eje horizontal $P(TE/RJ^c)$ y en el vertical P(RJ/TE), obtenemos una figura similar a la que se muestra (los números que aparecen no deben tomarse en cuenta, en el cálculo real aparecen números tan pequeños que las computadoras no manejan, sino que directamente igualan a cero, excepto cuando se usan programas especiales).



En la gráfica está marcada con una línea horizontal P(RJ/TE)=1/2. Vemos claramente que si averiguamos para qué valor de $P(TE/RJ^c)$ tenemos P(RJ/TE)=1/2, para todos los valores menores obtendremos una probabilidad de la resurrección mayor.

Impongamos entonces

$$P(RJ/TE)=\frac{3.10^{-11}.1}{3.10^{-11}+(1-3.10^{-11})P(TE/RJ^c)}=\frac{1}{2}$$

Esta ecuación tiene como solución

$$P(TE/RJ^c)=\frac{3.10^{-11}}{1-3.10^{-11}}=0{,}0000000000300000000000900\ldots$$

Por lo tanto, el grado de desconfianza debe ser totalmente minúsculo para que creamos en la resurrección con probabilidad ½. ¿Podemos asignar un grado de confianza tan alto a un testimonio? ¿A qué testimonio le asignamos tanta confianza? (Obsérvese que para tener probabilidad ½, llegamos a la conclusión que un error o mentira de los evangelistas en este caso debe ser tan improbable como la propia resurrección nos parecía a priori). Podemos decir que en general, la confiabilidad de un testimonio está relacionada con la coherencia y con la plausibilidad de lo que cuenta. Tomando en consideración que nadie creería que el grado de confianza en el historiador más escrupuloso, aunque use los métodos más modernos, puede ser tan alto como para decir algo así como que se equivoca solo una vez en cien mil millones de casos, y ese grado de confianza es el requerido en el testimonio evangélico para tener probabilidad de la resurrección ½, además de que los evangelios son incoherentes y además nos cuentan otra cantidad de sucesos muy improbables a priori, como otras resurrecciones, un nacimiento virginal, caminatas sobre el agua, conversiones de agua en vino, multiplicación de alimentos, etc., y teniendo en cuenta además que algunas de las cosas contadas (como el nacimiento virginal) no habían sido profetizadas pero los evangelistas sí creían que habían sido profetizadas[3], por lo que es altamente probable que hayan forjado la historia de acuerdo con ellas, de ninguna manera podemos

[3] "Todo esto aconteció para que se cumpliera lo que dijo el Señor por medio del profeta: «Una virgen concebirá y dará a luz un hijo y le pondrás por nombre Emanuel» (que significa: «Dios con nosotros»)" – Mt 1,22-23. Sin embargo, el texto referido "Por tanto, el Señor mismo os dará señal: La virgen concebirá y dará a luz un hijo, y le pondrá por nombre Emanuel" – Is 7,14, no dice "virgen" en el original hebreo, sino "alma", que significa "joven" (y además es obvio porque en el pasaje Isaías se está refiriendo a una mujer casada en el momento en que habla). Cuando la Biblia se traduce al griego, originando la Septuaginta, el pasaje se vierte como "La (parthenos) concebirá y dará a luz un hijo…", y "parthenos" quiere decir virgen. O sea, no se trataba de una profecía sino de esas cosas que pasan cuando se traduce. Aunque por supuesto alguien puede aducir que Dios, en su presciencia, sabía que se traduciría así de modo que efectivamente, estaba diciendo "joven" en hebreo para que se dijera "virgen" en griego y ser finalmente entendido. Pero en ese caso, ¿por qué el original hebreo no dice "virgen" (betula)?

razonablemente asignar un altísimo grado de confianza al testimonio evangélico, y por lo tanto, podemos decir que es totalmente irracional creer en la resurrección de Jesús a partir de la evidencia que se encuentra disponible.

Ejercicio: Estudie el artículo "A probabilidade da resurreição de Jesus" de Richard Swinburne (Revista *Episteme*, Porto Alegre, Núm. 18, 2004) y refute o al menos intente debilitar el argumento presentado en estas notas.