Teorema de Bayes

Adriana F. Chávez De la Peña
adrifelcha@gmail.com
-
Universidad Nacional Autónoma de México
Facultad de Psicología
Laboratorio 25
Proyecto PAPIME

1. Introducción.

El **Teorema de Bayes** (también conocido como Regla de Bayes) constituye una herramienta útil, tan flexible como poderosa, para <u>estimar la probabilidad de que un determinado evento ocurra dada la observación de cierta evidencia</u>.

La Regla de Bayes funciona a partir del cómputo de probabilidades. Como ya se discutió en el capítulo anterior, toda **Probabilidad** ($p(x)$) se define como un número real entre 0 y 1 que representa el grado de certidumbre que se tiene sobre la ocurrencia de un evento (x).

Ejemplos:

- La probabilidad de obtener un águila al tirar una moneda es de 0.5, ya que sólo hay dos resultados posibles, mutuamente excluyentes, que asumimos tienen la misma probabilidad de ocurrir.

$$p(Obtener\ águila\ en\ un\ volado) = 0.5$$

- La probabilidad de morir de un infarto durante el siguiente año en México es de 0.00028. Llegamos a este estimado ya que en México se reportan anualmente 35 mil muertes por infarto. La probabilidad de que cualquier ciudadano muera de un infarto se puede estimar dividiendo el número de casos reportados sobre el total de habitantes del país (122 millones).

$$p(Morir\ de\ un\ infarto\ durante\ el\ próximo\ año) = 0.00028$$

La certidumbre que se tiene respecto a la ocurrencia de un evento A ($p(A)$) puede cambiar dependiendo la información a la que se tiene acceso. Por ejemplo, si observamos que ha ocurrido un evento B, este puede proporcionar –o no- información relevante que cambie la certidumbre que se tenía inicialmente respecto al evento A, dependiendo de si dichos eventos están relacionados. A la probabilidad de que ocurra un evento A *dada (|)* la observación de cierta evidencia B, $p(A|B)$, se le conoce como **probabilidad condicional**.

Ejemplos:

- La probabilidad de que la mamá de Ana la deje salir de fiesta es mayor si antes de pedir permiso ordena su habitación, que si lo pide sin hacer nada.

$$p(Permiso) < p(Permiso|Quehacer)$$

- La probabilidad de que repruebe un examen de física avanzada es menor si he estudiado y leído al respecto, que si lo presento sin prepararme antes.

$$p(Reprobar) > \ p(Reprobar|He\ estado\ leyendo\ y\ viendo\ documentales\ sobre\ el\ tema)$$

Una vez establecida la noción de que cada evento en el mundo tiene una probabilidad de ocurrencia propia ($p(x)$) que puede, o no, variar tras observar otros eventos

($p(x|otros\ eventos)$), podemos distinguir formalmente entre eventos independientes y eventos no independientes:

- Los *eventos independientes* son aquellos cuya probabilidad de ocurrencia no tiene relación alguna y por tanto, permanece inalterada sin importar si ocurren o no simultáneamente: Si sabemos que los eventos A y B son independientes, observar que ocurre B no añade información sobre la posible ocurrencia de A.

$$p(A|B) = p(A)$$

> Ejemplo:
>
> La probabilidad de que llueva dado que la chica que se sentó junto a mí en el metro lleva puesta una minifalda, se mantiene igual al estimado que yo habría hecho de no haberse sentado junto a mí.
>
> $$p(\ Lluvia|\ La\ chica\ lleva\ minifalda) = p(Lluvia)$$

- Los *eventos no independientes* son aquellos cuya ocurrencia está relacionada en alguna medida; si observamos un evento B, que sabemos guarda relación con un evento A, su ocurrencia nos proporciona información para replantearnos la probabilidad de A.

$$p(A|B) < p(A)$$

ó

$$p(A|B) > p(A)$$

> Ejemplo:
>
> La probabilidad de que me asalten si estoy en un deportivo de Iztapalapa, es mayor que si estuviera en cualquier otro deportivo de la Ciudad de México,
>
> $$p(Asalto|\ Estoy\ en\ Iztapalapa) > p(Asalto)$$
>
> La probabilidad de que mi perro se enferme dado que he sido un dueño responsable y le he puesto toda sus vacunas es menor que si no lo hubiera hecho.
>
> $$p(Rabia|\ Vacunas\ completas) < p(Rabia)$$

Para calcular la Probabilidad Condicional de un evento A dada cierta evidencia B, de tal forma que podamos saber si nuestra certidumbre ha incrementado o decrecido, y cuánto, tenemos la siguiente ecuación:

$$(1) \qquad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

La Ecuación 1 captura un razonamiento bastante intuitivo: Si queremos saber qué tan probable es el evento A dado que estoy usando la ocurrencia de B como referencia, primero hay que considerar la probabilidad de que ambos eventos aparezcan simultáneamente en el mundo ($p(A \cap B)$) y sopesar dicha probabilidad en relación a la probabilidad de que el evento B aparezca en el mundo ($p(B)$), independientemente de su relación con A.

El diagrama de Venn que se muestra en la Figura 1 presenta el problema al que nos enfrentamos de manera gráfica: Imaginemos que estamos interesados en saber qué tan probable es el evento A una vez que hemos observado que ocurrió el evento B. Antes de observar B, tenemos una idea de cuál es la probabilidad de que ocurra A, ($p(A)$, el área del círculo amarillo). Ahora, lo que nos interesa es utilizar la ocurrencia de B como una referencia para actualizar nuestro estimado inicial, ($p(A|B)$). Si sabemos que los eventos A y B suelen presentarse juntos en el mundo con cierta probabilidad, ($p(A \cap B)$, el área verde donde ambos círculos se intersectan), necesitamos saber cuál es la probabilidad de que el evento observado B pertenezca a dicha área de intersección (en cuyo caso, anunciaría la ocurrencia simultánea de A), en relación a su **probabilidad marginal** ($p(B)$, el área total del círculo azul). Para ello, tenemos que obtener la razón entre el área de B que intersecta a A (la probabilidad conjunta de A y B) y el área total de B (la probabilidad marginal de B). La Ecuación 2 captura este razonamiento, llevándonos de vuelta a la Ecuación 1 previamente expuesta.

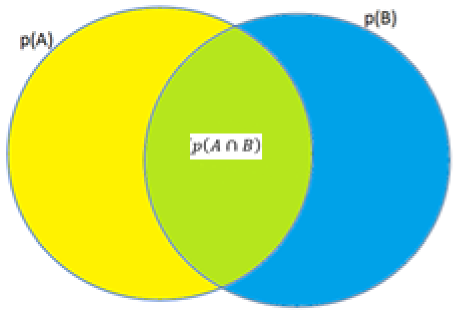

Figura 1. Diagrama de Venn representando la diferencia entre la probabilidad marginal de dos eventos posibles $p(A)$ y $p(B)$ y la probabilidad conjunta de los mismos $p(A \cap B)$

$$(2) \qquad P(A|B) = \frac{Probabilidad\ conjunta\ A\ y\ B}{Probabilidad\ marginal\ B} = \frac{P(A \cap B)}{P(B)}$$

De acuerdo a las Leyes de la Probabilidad, tenemos dos formas de estimar la **probabilidad conjunta** de dos eventos A y B, dependiendo de si A y B son, o no, independientes.

- Si A y B son eventos independientes, tenemos que:

$$(3)\ p(A \cap B) = p(A) \cdot p(B)$$

- Sin embargo, si los eventos A y B suelen estar relacionados en su ocurrencia, tenemos:

$$(4)\ p(A \cap B) = p(A) \cdot p(B|A)$$

Si sustituimos las ecuaciones 3 y 4, en la ecuación 1, podemos estimar la probabilidad condicional del evento A dada cierta evidencia B.

$$(5) \quad p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B)}{p(B)} = \frac{p(A) \cdot p(B)}{p(B)} = p(A)$$

$$(6) \quad p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B|A)}{p(B)}$$

La Ecuación 5 refiere a la Probabilidad de A dada la observación de un evento B, cuando se asume que ambos eventos son independientes. La probabilidad conjunta se sustituye por la Ecuación 3 y esto nos lleva directamente a la cancelación del término $p(B)$, demostrando formalmente que el estimado inicial ($p(A)$) permanece intacto.

La Ecuación 6, donde se asume que los eventos A y B están relacionados, nos permite utilizar la ocurrencia de B como información de referencia para calcular la probabilidad del evento A. **La Ecuación 6 es el Teorema de Bayes.**

Una vez desglosada la estructura y función de la Regla de Bayes, conviene familiarizarnos con la nomenclatura utilizada para referirnos a cada uno de los términos incluidos en el Teorema.

1) La probabilidad prior - $p(A)$

Al estimado inicial que tenemos sobre la probabilidad de ocurrencia del evento de interés A, antes de ver cualquier evidencia, se le conoce como **probabilidad prior**.

2) La verosimilitud - $p(B|A)$

En el numerador del Teorema de Bayes, encontramos dos términos que al multiplicarse representan la probabilidad conjunta de ocurrencia del evento de interés A y la evidencia B. Uno de ellos corresponde a la probabilidad prior; el segundo, se expresa como la probabilidad condicional de observar la evidencia B cuando el evento de interés A de hecho ocurre. A esta segunda probabilidad condicional, se le conoce como **verosimilitud.**

Tal como su nombre sugiere, la verosimilitud nos proporciona un indicador de la relación entre la evidencia B y el evento A, a partir de un razonamiento inverso: Cuando A de hecho ocurre, ¿con qué probabilidad aparece en compañía de B?

Es importante enfatizar que, contrario a lo que la intuición podría sugerir, las dos probabilidades condicionales implicadas en el Teorema de Bayes no son equivalentes. Es decir:

$$p(A|B) \neq p(B|A)$$

Esta relación suele quedar más clara si lo pensamos en términos de un ejemplo concreto: Imaginemos que estamos interesados en estimar la probabilidad de que llueva hoy, dado que vemos el cielo nublado. Hacernos esta pregunta tiene sentido, porque existe incertidumbre respecto a la posibilidad de que llueva, que puede reducirse sabiendo si el cielo está o no nublado. Sin embargo, sería extraño que nos preguntáramos si el cielo está nublado dado que escuchamos la lluvia caer, ya que sabemos que siempre que llueve es porque hay nubes en el cielo descargando agua. De tal forma que:

$$p(Lluvia|Cielo\ Nublado) \neq p(Cielo\ nublado|Lluvia)$$

A primera vista, puede parecer confuso que el cálculo de una probabilidad condicional dependa de nuestro conocimiento de una segunda probabilidad condicional. Sin embargo, estimar la verosimilitud no es tan complicado una vez que se conoce la estructura del entorno en que situamos nuestra predicción. Por ejemplo, regresando al ejemplo anterior, sabemos que la verosimilitud que relaciona el cielo nublado con los días lluviosos tiene un valor de 1.

$$p(Cielo\ nublado|Lluvia) = 1$$

3) La verosimilitud Marginal - $p(B)$

A la probabilidad marginal de observar el evento B que estamos tomando como referencia para estimar la probabilidad de ocurrencia de A, la conocemos como **verosimilitud marginal**.

La verosimilitud marginal se define formalmente de la siguiente forma:

$$(7)\ p(B) = \sum_i p(A_i) \cdot p(B|A_i) = \sum_i p(A_i \cap B)$$

La Ecuación 7 define a la verosimilitud marginal como la sumatoria de las probabilidades conjuntas de ocurrencia de la evidencia observada y todos los posibles eventos con que pudiera estar asociada. El subíndice 'i' corresponde al número de eventos posibles contemplados.

Matemáticamente hablando, sumar la probabilidad conjunta de la evidencia B con todos los eventos posibles A_i, garantiza que la probabilidad de ocurrencia de los distintos eventos posibles sea complementaria y que sumen a 1.

Por ejemplo, si interesa estimar la probabilidad de un solo evento A, identificamos la no ocurrencia de A ($\bar{A}$) como la única alternativa posible. Entonces, la verosimilitud marginal se obtiene sumando la probabilidad conjunta de los eventos A y B ($B \cap A$), y la probabilidad conjunta de que se observe la evidencia B en ausencia de A ($B \cap \bar{A}$):

$$(8)\ p(B) = \left(p(A) \cdot p(B|A) \right) + \left(p(\bar{A}) \cdot p(B|\bar{A}) \right)$$

4) La probabilidad posterior - $p(A|B)$

Al estimado final de la probabilidad de ocurrencia del evento A, una vez habiendo ponderado la información aportada por la evidencia B, se le conoce como **probabilidad posterior**.

$$p(A|B) = \frac{p(A)\, p(B|A)}{p(B)}$$

$$probabilidad\ posterior = \frac{prior\ \cdot verosimilitud}{verosimilitud\ marginal}$$

* * *

Una vez entendiendo para qué sirve el Teorema de Bayes, cómo funciona y qué reglas de la probabilidad sigue, y habiendo identificado las 'etiquetas' dadas a cada uno de los términos que le componen, ¡Felicidades! Ya estás hablando en 'Bayes',

2. Ejercicios:

2.a. Un simple problema de contenedores

> Imagina que eres invitado al gran banquete que el rey dará en su castillo. El rey es conocido por ser un tirano despiadado, que se enriquece a costa de la pobreza de su pueblo y tú quieres poner fin a ello, envenenándolo. Te has enterado que en todas sus reuniones el rey siempre manda hacer para sus invitados tartas de fresa y vainilla, por lo que decides preparar 50 tartas de vainilla y 50 tartas de fresa, cargadas de un veneno mortal.
>
> Cuando llegas al banquete y te infiltras en la cocina para dejar tus tartas envenenadas, te das cuenta de que el rey ya había mandado hacer 170 tartas de vainilla y 30 tartas de fresa. Mantienes esta información en mente mientras revuelves tus tartas envenenadas con las tartas hechas en el castillo.
>
> Regresas a tomar tu lugar en el banquete. Los vasallos leales al rey te dan la bienvenida, aunque parecen renuentes a confiar demasiado en ti durante tu primera visita. Uno de ellos notó tu visita a la cocina y te interroga al respecto. Dices que fuiste a depositar cien tartas que tú mismo preparaste, de vainilla y fresa por ser los sabores preferidos del rey, como muestra de gratitud.

> El vasallo suspicaz huele tus malas intenciones y frente a todos, saca una tarta aleatoria del montón revuelto que hay en la cocina y te pide que la comas. La tarta es de vainilla. ¿Cuál es la probabilidad de que mueras al comerla?

Para comenzar a resolver el problema, lo primero que tenemos que hacer es definir la probabilidad que nos interesa calcular y la ubicamos como una instancia del Teorema de Bayes

$$p(La\ tarta\ está\ envenenada|La\ tarta\ es\ de\ vainilla)$$

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

$$p(Veneno|Vainilla) = \frac{p(Veneno)p(Vainilla|Veneno)}{p(Vainilla)}$$

Una vez identificando la *probabilidad posterior* que nos interesa calcular, sólo queda definir qué valores toma cada término de la ecuación.

Empezamos con la *probabilidad prior* de que al extraer una tarta aleatoriamente, esta pertenezca al lote envenenado. Sabemos que en la cocina del castillo hay un total de 300 tartas, de las cuales 100 están envenenadas. Por lo tanto, la probabilidad de que al sacar una tarta del montón, al azar, esta pertenezca al grupo de las tartas envenenadas es de $\frac{1}{3}$. Es decir:

- $p(Veneno) = 0.333$

La *verosimilitud* nos proporciona una medida de correspondencia entre la evidencia que tenemos y el evento cuya probabilidad queremos inferir a partir de su ocurrencia. Es decir, corresponde a la probabilidad de sacar una tarta de vainilla, asumiendo que la tarta sacada aleatoriamente salió del lote envenenado. Como sabemos que el lote envenenado estaba compuesto por 100 tartas, mitad de vainilla y mitad de fresa, decimos que:

$$p(Vainilla|Veneno) = 0.5$$

Por último, falta definir la *verosimilitud marginal*. Es decir, saber qué tan probable es que la tarta que se extrajo aleatoriamente del montón total de tartas fuera de vainilla, con independencia de si está o no envenenada. Podemos calcular la probabilidad marginal directamente, obteniendo la razón entre el número de tartas de vainilla y el número total de tartas en el castillo, es decir:

$$p(Vainilla) = \frac{220\ tartas\ de\ vainilla}{300\ tartas\ en\ el\ castillo} = .7333$$

Una manera más formal de calcular la *verosimilitud marginal*, (Ver Ecuación 7), implica sumar los productos de las probabilidades prior de cada uno de los eventos posibles ($p(Veneno)$ y $p(No\ Veneno)$) con la respectiva verosimilitud que relaciona la evidencia observada con cada uno ($p(Vainilla|Veneno)$ y $p(Vainilla|No\ veneno)$). Como vemos, el resultado es el mismo:

$$p(Vainilla) = \left(\, p(Veneno) \cdot p(Vainilla|Veneno)\right) + \left(\, p(No\ veneno) \cdot p(Vainilla|No\ veneno)\right)$$
$$p(Vainilla) = \left(\, (0.333) \cdot (0.5)\right) + \left(\, (0.666) \cdot (.85)\right)$$
$$p(Vainilla) = (.1666) + (.5666)$$
$$p(Vainilla) = .7333$$

Finalmente, tenemos todo lo necesario para saber cuál es la probabilidad de que la tarta de vainilla esté envenenada:

$$p(Veneno|Vainilla) = \frac{p(Veneno)p(Vainilla|Veneno)}{p(Vainilla)}$$
$$p(Veneno|Vainilla) = \frac{(0.333)(0.5)}{(0.7333)}$$

$$p(Veneno|Vainilla) = 0.2272$$

Es decir, que después de todo parece ser que puedes comerte la tarta de vainilla con cierta tranquilidad, ya que la probabilidad de que mueras después de comerla es relativamente baja.

Un buen ejercicio para corroborar que nuestros cálculos estén representando adecuadamente la distribución de la probabilidad entre los distintos eventos contemplados, definidos como mutuamente excluyentes, es calcular la probabilidad de que la tarta de vainilla no esté envenenada. Si hemos hecho todo bien, deberíamos esperar que el resultado de aplicar la misma fórmula a la alternativa opuesta, sea el complemento de la probabilidad de que la tarta de vainilla esté envenenada. Comprobémoslo:

$$p(No\ veneno|Vainilla) = \frac{p(No\ veneno)p(Vainilla|No\ veneno)}{p(Vainilla)}$$
$$p(No\ veneno|Vainilla) = \frac{(0.666)(0.85)}{(0.7333)}$$
$$p(No\ veneno|Vainilla) = 0.7726$$

Al sumar las probabilidades posteriores obtenidas para cada uno de los escenarios posibles (La tarta está, o no, envenenada), vemos que efectivamente se trata de eventos mutuamente excluyentes, que suman a 1.

$$p(Veneno|Vainilla) + p(No\ veneno|Vainilla) = 1$$
$$(0.2272) + (0.7726) = 1$$

2.a.1. Ejercicios de repaso:

a) ¿Por qué sustituimos $p(Vainilla|No\ veneno) = 0.85$? ¿De dónde salió ese valor?

b) Imagina que, en vez de vainilla, la tarta extraída aleatoriamente fue de fresa. ¿Cuál es la probabilidad de que mueras ahora?

2.b. Un problema de asesinatos

> Hay un asesino en serie suelto por la ciudad. Nadie está seguro de cuál puede ser su identidad; sin embargo, la víctima de su último ataque consiguió escapar e informar a la policía de que, pese a llevar cubierto el rostro, pudo notar que se trataba de un individuo pelirrojo y que además sostenía el arma homicida con la mano izquierda.
>
> La policía inicia una búsqueda exhaustiva inmediatamente. El jefe de policía inicia una investigación, y confía en que no les cueste trabajo dar con el asesino dado a que solamente el 3% de los 30,000 ciudadanos que conforman la población, tiene el cabello rojo, y solamente el 15% es zurda. El jefe de policía recuerda sus clases de probabilidad de la prepa, y sabe que, por tratarse de eventos independientes, la probabilidad conjunta de que un individuo cualquiera sea simultáneamente zurdo y pelirrojo es de:

$$p(Zurdo\ \cap Pelirrojo) = p(Zurdo) \cdot p(Pelirrojo)$$
$$p(Zurdo\ \cap Pelirrojo) = (0.03) \cdot (0.15)$$
$$p(Zurdo\ \cap Pelirrojo) = 0.0045$$

> Cuando tres días después de iniciada la búsqueda, aparece un sujeto con dichas características en un motel de paso en los límites de la ciudad, el jefe de policía está muy entusiasmado y seguro de haber dado con el asesino, cerrando el caso para siempre.
>
> El sospechoso retenido, resulta ser un aclamado profesor de universidad, con altos conocimientos en estadística. El evento se vuelve una sensación mediática y la opinión pública está dividida entre quienes están convencidos de que es el asesino, y quienes no.
>
> El día del juicio, el juez encargado de resolver el caso es un egresado de una de las universidades más prestigiosas del mundo y conoce de estadística. Por lo que le es posible evaluar qué tanto aporta el aspecto físico del sospechoso como una posible evidencia de su involucramiento en los asesinatos reportados. El sospechoso defiende su inocencia apelando al Teorema de Bayes. ¿El sospechoso es, o no, aprehendido al final de su enjuiciamiento?

De acuerdo a la tarea que se nos presentan, existen solamente dos posibles conclusiones, cuya probabilidad se puede estimar aplicando el Teorema de Bayes:

- Escenario 1: El sospechoso es culpable (Culpable).
$$p(Culpable|Pelirrojo\ Zurdo) = \frac{p(Culpable)p(Pelirrojo\ Zurdo|Culpable)}{p(Pelirrojo\ Zurdo)}$$

- Escenario 2: El sospechoso es inocente (Inocente).

$$p(Inocente|Pelirrojo\ Zurdo) = \frac{p(Inocente)p(Pelirrojo\ Zurdo|Inocente)}{p(Pelirrojo\ Zurdo)}$$

Antes de conocer el testimonio de la última víctima, se creía que el asesino podía ser cualquiera de los 30,000 habitantes de la ciudad. Es decir, antes de contar con información, cada uno de estos tendría una probabilidad inicial de $\frac{1}{30,000}$ de ser el asesino. Por tanto tenemos que:

- Escenario 1:
$$p(Culpable) = \frac{1}{30,000} = 0.00003$$
- Escenario 2:
$$p(Inocente) = \frac{29,999}{30,000} = 0.99996$$

Partiendo del supuesto de que inicialmente cualquier ciudadano tenía la misma probabilidad de ser el asesino, la probabilidad de que específicamente el individuo que hemos identificado como sospechoso haya cometido los asesinatos, es muy baja. Por otro lado, la probabilidad de que el asesino se encuentre entre los habitantes restantes, y no en el caso específico que hemos elegido, es mucho más grande.

Continuando con la definición de los términos del Teorema de Bayes, definimos la verosimilitud como la probabilidad que relaciona el hecho de que nuestro sospechoso tenga un aspecto idéntico al reportado por el testigo, para evaluar su culpabilidad, y con la probabilidad de que no se trate del asesino, sino únicamente sea un ciudadano aleatorio que sea zurdo y pelirrojo.

- Escenario 1:

$$p(Pelirrojo\ Zurdo\ |Culpable) = 1$$

Si asumimos que el sospechoso retenido es de hecho el asesino, la probabilidad de que su aspecto físico coincida con el aspecto físico del asesino reportado por el testigo, es de 1.

- Escenario 2:

$$p(Pelirrojo\ Zurdo\ |Inocente) = 0.0045$$

Si asumimos que el sospechoso es inocente, esto sólo puede querer decir que el asesino es algún otro habitante de la ciudad. La probabilidad de que cualquier otro habitante de la ciudad sea al mismo tiempo pelirrojo y zurdo, es de 0.0045. Justo como reflejaba el razonamiento del jefe de policía expuesto en el ejemplo.

Finalmente, calculamos la verosimilitud marginal aplicando la Ecuación 7.

$$p(Pelirrojo\ Zurdo) = \big(\, p(Culpable) \cdot p(Pelirrojo\ Zurdo|Culpable)\big) + \big(\, p(Inocente) \cdot p(Pelirrojo\ Zurdo|Inocente)\big)$$
$$p(Pelirrojo\ Zurdo) = \big(\,(0.00003) \cdot (1)\big) + \big(\,(0.99996) \cdot (.0045)\big)$$
$$p(Pelirrojo\ Zurdo) = (0.00003) + (0.00449)$$
$$p(Pelirrojo\ Zurdo) = \ 0.004529$$

Nótese que una vez más, el resultado de aplicar la sumatoria presentada en la Ecuación 7 coincide con la probabilidad estimada de que un individuo cualquiera presente al mismo tiempo el cabello pelirrojo y sea zurdo.

Una vez identificados todos los términos, estamos listos para aplicar la Regla de Bayes:

- Escenario 1:

$$p(Culpable|Pelirrojo\ Zurdo) = \frac{p(Culpable)p(Pelirrojo\ Zurdo|Culpable)}{p(Pelirrojo\ Zurdo)}$$

$$p(Culpable\ |Pelirrojo\ Zurdo) = \frac{(0.00003)(1)}{(0.004529)} = \frac{0.00003}{0.004529}$$
$$p(Culpable|Pelirrojo\ Zurdo) = 0.0066$$

- Escenario 2:

$$p(Inocente|Pelirrojo\ Zurdo) = \frac{p(Inocente)p(Pelirrojo\ Zurdo|Inocente)}{p(Pelirrojo\ Zurdo)}$$
$$p(Inocente|Pelirrojo\ Zurdo) = \frac{(0.99996)(0.0045)}{(0.004529)} = \frac{0.004499}{0.004529}$$
$$p(Inocente|Pelirrojo\ Zurdo) = 0.9935$$

Contrario a lo que la intuición del jefe de policía sugería, parece ser que el hecho de que los rasgos del asesino sean muy poco comunes en la ciudad no proporciona evidencia suficiente para incriminar a un sospechoso con dichas características.

Al sumar las probabilidades posteriores obtenidas para cada escenario posible, vemos que se cumple con el supuesto de complementariedad de las probabilidades (Es decir, los escenarios probados son de hecho mutuamente excluyentes).

$$p(Culpable|Pelirrojo\ Zurdo) + p(No\ culpable|Pelirrojo\ Zurdo) = 1$$
$$(0.0066) + (0.99359) = 1$$

2.a.1. Ejercicios de repaso:

a) ¿De qué manera el total de habitantes de la ciudad repercute en la inferencia? Imagina que en vez de 30,000 personas, se tratara de un pueblo con sólo 100 habitantes, ¿Cómo afectaría esto a las conclusiones?
b) ¿De qué manera cambian las conclusiones finales, si en vez de dos piezas de información tuvieras sólo una? Imagina que no sabes que el asesino es pelirrojo, ¿cómo cambian las conclusiones?

3. Lecturas recomendadas:

Material amigable para principiantes:

- Downey, A. (2012) Think Bayes: Bayesian statistics made simple. Chapter 1: Bayes' theorem. Green Tea Press. Massachusets.
- Oxford Press (2015) Handbook of Computational and Mathematical Psychology.

Material un poco más formal y especializado:

- Mackay, D. () Information theory, inference and learning algorithms.