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 estimar la probabilidad de que un determinado evento ocurra, a la luz de cierta evidencia observada.

La Regla de Bayes funciona a partir del cómputo de probabilidades. Como ya se discutió en el capítulo anterior, toda **Probabilidad** ($p(x)$) puede definirse como un número 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.

$$p(Obtener\ \acute{a}guila\ en\ un\ volado) = 0.5$$

- La probabilidad de morir de un infarto durante el siguiente año en México es de 0.00028. Llegamos a este estimado ya que anualmente se reportan en México 35 mil muertes por infarto, en un país con alrededor de 122 millones de habitantes. 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.

$$p(Morir\ de\ un\ infarto\ durante\ el\ pr\acute{o}ximo\ a\tilde{n}o) = 0.00028$$

Como es de esperar, la certidumbre que tenemos respecto a la ocurrencia de un evento A ($p(A)$) varía dependiendo la información de la que disponemos. Es decir, si observamos que ciertos eventos ya se han presentado, su ocurrencia puede proveernos información que altere –o no- la certidumbre que se tenía inicialmente respecto a la ocurrencia de A, dependiendo de si el evento observado guarda o no, relación con A. 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 llueva hoy, dado que ha estado lloviendo toda la semana es mayor que la probabilidad de que llueva en cualquier otro día del año.

$$p(Lluvia) < p(Lluvia|\ Ha\ llovido\ toda\ la\ semana)$$

- 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 posibles en el mundo tiene una probabilidad de ocurrencia propia ($p(x)$) que puede, o no, variar tras observar la ocurrencia de otros eventos

$(p(x|eventos\ en\ el\ mundo)$, podemos distinguir 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. De tal forma que 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 que ocurre un evento B, que sabemos guarda relación con un evento A, su ocurrencia nos proporciona información para juzgar la probabilidad de A.

$$p(A|B) < p(A)$$
ó
$$p(A|B) > p(A)$$

> Ejemplo:
>
> La probabilidad de que me vayan a asaltar es mucho mayor si estoy en un deportivo de Iztapalapa, que si anduviera en cualquier otro punto aleatorio 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.
>
> $$p(Rabia|\ Vacunas\ completas) < p(Rabia)$$

Ahora bien, lo primero que deberíamos preguntarnos llegados a este punto es cómo estimar 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. Para ello, tenemos la siguiente ecuación:

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

El planteamiento de la Ecuación 1 captura un razonamiento bastante intuitivo: Si yo quiero saber qué tan probable es el evento A dado que estoy observando B, tengo que evaluar la probabilidad con que ambos eventos aparecen de manera simultánea en el mundo $(p(A \cap B))$ y sopesar dicha probabilidad con base en la probabilidad de que el evento B que quiero usar como evidencia, aparezca en el mundo $(p(B))$, sin importar su correspondencia 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.

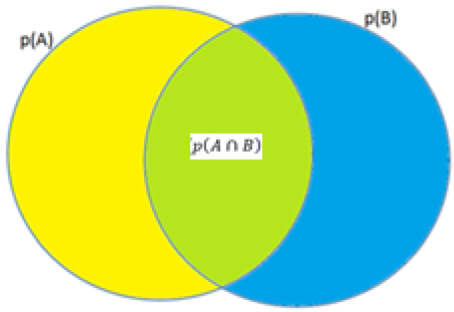

Figura 1. Diagrama de Venn representando la diferencia entre la probabilidad marginal de dos eventos posibles $p(A)$ y $p(B)$ y la probabilidad conjunta de los mismos $p(A \cap B)$

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

De acuerdo a las Leyes de la Probabilidad, tenemos dos formas de estimar la **probabilidad conjunta** de dos eventos A y B (i.e. la probabilidad de que dichos eventos ocurran simultáneamente, $A \cap B$), 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.

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

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

La Ecuación 5 corresponde a la Probabilidad de A tras observar un evento B, bajo el supuesto de que ambos eventos son independientes. La inserción de la Ecuación 3 en el lugar correspondiente a la Probabilidad Conjunta nos sirve para demostrar formalmente que el estimado inicial ($p(A)$) permanece inalterado, tras la cancelación de los términos ($p(B)$) en el denominador y numerador de la ecuación.

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. El primero de estos términos 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**.

Típicamente, la verosimilitud marginal se define formalmente de la siguiente forma:

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

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**.

* * *

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 qye le componen, ¡Felicidades! Ya estás hablando en 'Bayes',

2. Ejercicios:

(NOTA: Falta desarrollar el código, eso puede tomarme unos días más. Sin embargo, estos son los problemas que pretendo traducir a un código, con sus respectivas grafiquitas. Le agradecería mucho que los checaran en términos de qué tan claros y pertinentes son, y sobretodo, que estén bien hechos! (Yo los inventé))

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 para caballos.

Cuando llegas al banquete y te filtras 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:

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

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 galleta 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

Imagina que eres un detective. Acaba de ocurrir un asesinato en una fiesta con 20 asistentes (la mitad son hombres y la otra mitad, mujeres). Hay dos sujetos con actitudes sospechosas: Jessie y James. Tu tarea consiste en determinar, a partir de una inferencia bayesiana, si se les arresta o no.

Te dan acceso a la cámara de seguridad, en ella puedes ver el momento en que un sujeto, con el rostro cubierto, entra la baño inmediatamente después de la víctima, donde permanece por unos cuantos minutos antes de salir con claras muestras de ansiedad. El sujeto lleva puesto un saco.

Al llegar a la sala donde se encuentran reunidos todos los invitados, te das cuenta de que el 75% de los caballeros traen puesto un saco similar al del video, James es uno de ellos. ¿Cuál es la probabilidad de que Jessie o James, sean culpables?

Empezamos por definir los escenarios posibles, siendo todos mutuamente excluyentes. Existen tres probabilidades posteriores que se puede computar a partir de la evidencia:

- Escenario 1: Jessie es la culpable (Jessie)

$$p(Jessie|Camara\ muestra\ sujeto\ con\ saco) = \frac{p(Jessie)p(Camara\ muestra\ sujeto\ con\ saco|Jessie)}{p(Camara\ muestra\ sujeto\ con\ saco)}$$

- Escenario 2: James es el culpable (James)

$$p(James|Camara\ muestra\ sujeto\ con\ saco) = \frac{p(James)p(Camarra\ muestra\ sujeto\ con\ saco|James)}{p(Camara\ muestra\ sujeto\ con\ saco)}$$

- Escenario 3: Otro fue el culpable (Otro)

$$p(Otro|Camara\ muestra\ sujeto\ con\ saco) = \frac{p(Otro)p(Camara\ muestra\ sujeto\ con\ saco|Otro)}{p(Camara\ muestra\ sujeto\ con\ saco)}$$

Comenzamos a identificar los términos necesarios para aplicar el Teorema de Bayes.

Sabemos que había 20 personas en la fiesta. Antes de ver la cámara de vigilancia, no teníamos razón para asumir que cualquiera de los asistentes tuviera una mayor probabilidad prior de ser el asesino, es decir, toda persona en la fiesta tiene una probabilidad prior de $\frac{1}{20}$ de ser el asesino:

- $p(Jessie) = \frac{1}{20}$
- $p(James) = \frac{1}{20}$
- $p(Otro) = \frac{18}{20}$

También sabemos que un 75% de los hombres llevaban puesto el mismo saco que se ve en el video de vigilancia, y que en la fiesta habían 10 hombres y 10 mujeres. Con esta información podemos empezar a definir la verosimilitud que relaciona lo mostrado por la cámara con los distintos escenarios posibles:

- La probabilidad de que la cámara mostrara a un sujeto usando saco, asumiendo que Jessie fue la asesina, es de 0. Ya que sólo los hombres usan saco. De haber sido Jessie la culpable, la cámara no podría haber mostrado a alguien con saco.
$$p(Camara\ muestra\ sujeto\ con\ saco|Jessie) = 0$$

- La probabilidad de que la cámara mostrara a un sujeto usando saco, asumiendo que James fuera el asesino, es de 1. Es decir, dado que James de hecho está usando un saco oscuro, de haber sido el asesino no habría forma de que la cámara mostrara algo diferente.
$$p(Camara\ muestra\ sujeto\ con\ saco|James) = 1$$

- La probabilidad de que la cámara mostrara un sujeto usando saco, asumiendo que el culpable fuera algún otro asistente a la fiesta, es de .3375. Los otros 18 asistentes, excluyendo a Jessie y James, representan un .9 de la proporción total de invitados. Sólo la mitad de éstos eran hombres (0.45). Calculamos la proporción de hombres usando saco, omitiendo a James, al multiplicar por la probabilidad .75.

$$p(Camara\ muestra\ sujeto\ con\ saco|Otro) = (0.45)(0.75) = .3375$$

Nuevamente, tenemos dos formas de estimar la verosimilitud marginal. La primera, implica hacer uso directo de la información que tenemos sobre la tarea. Si sabemos que dentro de los 20 invitados, la mitad eran hombres y el 75% de ellos llevaba puesto un saco oscuro, tenemos que:

$$(0.5)(0.75) = .375$$

***Según yo, este razonamiento es válido. Pero por alguna razón, lleva a un resultado diferente al estimado real de la verosimilitud marginal yyyyy, sustituyéndolo en la fórmula de Bayes, no cumple el requisito de complementariedad

Finalmente, computamos la verosimilitud marginal aplicando una vez más la Ecuación 7:

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

$$p(Sujeto\ con\ saco) = (((p(Jessie) \cdot p(Saco|Jessie)) + ((p(James) \cdot p(Saco|James)) + (p(Otro) \cdot p(Saco|Otro)))$$

$$p(Sujeto\ con\ saco) = (((.05) \cdot (0)) + ((.05) \cdot (1)) + ((.9) \cdot (.3375)))$$

$$p(Sujeto\ con\ saco) = ((0) + (.05) + (.30375))$$

$$p(Sujeto\ con\ saco) = 0.35375$$

Una vez que tenemos todos los elementos necesarios para aplicar el Teorema de Bayes, sustituimos:

- Escenario 1: Jessie es la culpable (Jessie)

$$p(Jessie|Camara\ muestra\ sujeto\ con\ saco) = \frac{p(Jessie)p(Camara\ muestra\ sujeto\ con\ saco|Jessie)}{p(Camara\ muestra\ sujeto\ con\ saco)}$$

$$p(Jessie|Camara\ muestra\ sujeto\ con\ saco) = \frac{(0.05)(0)}{(0.35375)} = \frac{0}{(0.35375)}$$

$$p(Jessie|Camara\ muestra\ sujeto\ con\ saco) = 0$$

- Escenario 2: James es el culpable (James)

$$p(James|Camara\ muestra\ sujeto\ con\ saco) = \frac{p(James)p(Camarra\ muestra\ sujeto\ con\ saco|James)}{p(Camara\ muestra\ sujeto\ con\ saco)}$$

$$p(James|Camara\ muestra\ sujeto\ con\ saco) = \frac{(0.05)(1)}{(0.35375)}$$

$$p(James|Camara\ muestra\ sujeto\ con\ saco) = \frac{(0.05)}{(0.35375)}$$

$$p(James|Camara\ muestra\ sujeto\ con\ saco) = .1413$$

- Escenario 3: Otro fue el culpable (Otro)

$$p(Otro|Camara\ muestra\ sujeto\ con\ saco) = \frac{p(Otro)p(Camara\ muestra\ sujeto\ con\ saco|Otro)}{p(Camara\ muestra\ sujeto\ con\ saco)}$$

$$p(Otro|Camara\ muestra\ sujeto\ con\ saco) = \frac{(0.9)(0.3375)}{(0.35375)} = \frac{.30375}{.35375}$$

$$p(Otro|Camara\ muestra\ sujeto\ con\ saco) = 0.8586$$

Por último, para asegurarnos de que hicimos bien todos los cálculos, comprobamos la complementariedad de las probabilidades obtenidas:

$$p(Jessie|Saco) + p(James|Saco) + p(Otros|Saco) = 1$$

$$(0) + (0.1413) + (0.8586)$$

$$= 1$$

De tal forma que, de acuerdo a la evidencia arrojada por la cámara de seguridad, es imposible que Jessie esté involucrada en el asesinato. Esta primera conclusión, intuitivamente lógica, queda formalmente demostrada por la verosimilitud adquiriendo un valor de 0. Por otro lado, parece ser que el hecho de que James lleve puesto un saco como el que se ve en el video de seguridad no es evidencia suficiente para juzgarlo como culpable, ya que se trata de un saco utilizado por una buena proporción de los asistentes varones.

> Pero, ¡alto ahí! Decides echar un segundo vistazo al video y descubres que el sujeto del video abre la puerta del baño con la mano izquierda. Sabes que sólo 8% de la población mundial es zurdo, ¿Podría ser?
>
> Regresas a hablar con James y le ofreces un cigarrillo. James lo toma con la mano izquierda; ¡es zurdo!
>
> ¿Cuál es la probabilidad de que James sea el asesino a la luz de la nueva evidencia?

Definimos las probabilidades que se nos solicita estimar:

- Escenario 1: James es culpable (James)

$$p(James|Camara\ muestra\ a\ zurdo) = \frac{p(James)p(Camara\ muestra\ zurdo\ |James)}{p(Camara\ muestra\ a\ zurdo)}$$

- Escenario 2: Fue alguien más. (Otro)

$$p(Otro|Camara\ muestra\ a\ zurdo) = \frac{p(Otro)p(Camara\ muestra\ zurdo|Otro)}{p(Camara\ muestra\ zurdo)}$$

Identificamos las probabilidades prior. Esta vez, consideramos sólo a los asistentes varones.

- Escenario 1:

$$p(James) = \frac{1}{10} = 0.1$$

- Escenario 2:

$$p(Otro) = \frac{9}{10} = 0.9$$

Definimos la verosimilitud con que se relaciona la evidencia de la cámara y cada escenario:

- Escenario 1:

$$p(Camara\ muestra\ zurdo|James) = 1$$

- Escenario 2:

$$p(Camara\ muestra\ zurdo|Otro) = 0.08$$

Computamos nuestra verosimilitud marginal:

$$p(B) = \sum_i p(A_i) \cdot p(B|A_i)$$
$$p(Sujeto\ zurdo) = ((p(James) \cdot p(Zurdo|James)) + (p(Otro) \cdot p(Zurdo|Otro)))$$
$$p(Sujeto\ zurdo) = \left(((.1) \cdot (1)) + ((.9) \cdot (0.08))\right)$$
$$p(Sujeto\ zurdo) = ((0.1) + (0.072))$$
$$p(Sujeto\ con\ saco) = 0.172$$

Sustituimos en el Teorema de Bayes:

- Escenario 1:

$$p(James|Zurdo) = \frac{p(James)p(Zurdo|Jaime)}{p(Zurdo)}$$
$$p(James|Zurdo) = \frac{(0.1)(1)}{(0.172)} = \frac{0.1}{0.172}$$
$$p(James|Zurdo) = 0.5813$$

- Escenario 2:

$$p(Otro|Zurdo) = \frac{p(Otro)p(Zurdo|Otro)}{p(Zurdo)}$$
$$p(Otro|Zurdo) = \frac{(0.9)(0.08)}{(0.172)} = \frac{0.072}{0.172}$$
$$p(Otro|Zurdo) = 0.4186$$

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.