

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 dada la observación de cierta evidencia.

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). En su definición más simple, la probabilidad puede computarse a partir de la razón entre el número de casos que incluyen al evento x y el total de casos que es posible de observar.

$$p(x) = \frac{\text{Casos contemplados en } X}{\text{Totalidad de casos posibles}}$$

Ejemplos:

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

$$p(\text{Obtener águila en un volado}) = \frac{1}{2} = 0.5$$

- La probabilidad de morir de un infarto durante el siguiente año en México es de 0.059, dado que 35 mil muertes de entre las 590,693 reportadas anualmente en México, son a causa de un infarto.

$$p(\text{Morir de un infarto durante el próximo año}) = \frac{35,000}{590,693} = 0.059$$

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.

$$p(\text{Permiso}) < p(\text{Permiso}|\text{Quehacer})$$

- 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.

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

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

- Los *eventos independientes*, como su nombre sugiere, son aquellos cuya ocurrencia es independiente entre sí, por lo que la presentación de uno de ellos no altera la probabilidad de un segundo evento: Si sabemos que los eventos A y B son independientes, observar que ocurre B no añade información sobre la posible ocurrencia de A.

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

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

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.

$$p(\text{Lluvia} | \text{La chica lleva minifalda}) = p(\text{Lluvia})$$

- 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,

$$p(\text{Asalto} | \text{Estoy en Iztapalapa}) > p(\text{Asalto})$$

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.

$$p(\text{Rabia} | \text{Vacunas completas}) < p(\text{Rabia})$$

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:

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

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 $p(A \cap B)$

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

De acuerdo a las Leyes de la Probabilidad, la **probabilidad conjunta** de dos eventos A y B se puede computar con la siguiente ecuación:

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

Si insertamos la Ecuación 3 en el numerador de la Ecuación 2, obtenemos una ecuación que nos permite calcular la probabilidad de un evento A dada la observación de cierta evidencia B. La ecuación resultante **es el Teorema de Bayes**.

$$(4) \quad p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B|A)}{p(B)}$$

Las Ecuaciones 3 y 4 permiten enfatizar una diferencia importante en términos de la información que podemos extraer sobre la posible ocurrencia de A tras la observación de B, dependiendo de si A y B están o no definidos como eventos independientes. De acuerdo a la relación previamente especificada en términos de la igualdad entre la probabilidad condicional y la probabilidad marginal, tendríamos que:

Si al tratarse de eventos independientes $p(B|A) = p(B)$ y $p(A|B) = p(A)$, entonces:

$$\begin{aligned} p(A \cap B) &= p(A) \cdot p(B|A) \\ &= \\ (5) \quad p(A \cap B) &= p(A) \cdot p(B) \end{aligned}$$

La Ecuación 5, que no es más que otra forma de expresar la ecuación general que define las probabilidades conjuntas para el caso específico en que los eventos a evaluar son independientes, es de hecho identificada dentro de las leyes de la probabilidad como aquella que describe 'la probabilidad conjunta de eventos independientes'.

Finalmente, si sustituimos la Ecuación 5 en el Teorema de Bayes, reafirmamos formalmente que, dada su independencia, la observación de un evento A o B, no proporciona información adicional sobre la ocurrencia del otro:

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

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:

$$p(\text{Lluvia}|\text{Cielo Nublado}) \neq p(\text{Cielo nublado}|\text{Lluvia})$$

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.

$$p(\text{Cielo nublado}|\text{Lluvia}) = 1$$

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:

$$(6) \quad p(B) = \sum_i p(A_i) \cdot p(B|A_i) = \sum_i p(A_i \cap B)$$

La Ecuación 6 define la verosimilitud marginal como una sumatoria de probabilidades conjuntas. Tal y como lo hemos manejado hasta ahora, A corresponde al Evento cuya probabilidad interesa calcular y B a la evidencia que estamos observando y con base en la cual hacemos el cómputo. El subíndice i, indica que la sumatoria puede extenderse a tanta probabilidades conjuntas como eventos nos interese estimar ($A_1, A_2, \dots A_n$).

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 ($B \cap \bar{A}$):

$$(7) \quad p(B) = (p(A) \cdot p(B|A)) + (p(\bar{A}) \cdot p(B|\bar{A}))$$

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 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

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

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

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

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:

- $p(\text{Veneno}) = 0.333$

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:

$$p(\text{Vainilla} | \text{Veneno}) = 0.5$$

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:

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

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 ($p(\text{Veneno})$ y $p(\text{No Veneno})$) con la respectiva verosimilitud que relaciona la evidencia observada con cada uno ($p(\text{Vainilla} | \text{Veneno})$ y $p(\text{Vainilla} | \text{No veneno})$). Como vemos, el resultado es el mismo:

$$p(\text{Vainilla}) = (p(\text{Veneno}) \cdot p(\text{Vainilla} | \text{Veneno})) + (p(\text{No veneno}) \cdot p(\text{Vainilla} | \text{No veneno}))$$

$$\begin{aligned}
 p(\text{Vainilla}) &= (0.333) \cdot (0.5) + (0.666) \cdot (0.85) \\
 p(\text{Vainilla}) &= 0.1666 + 0.5666 \\
 p(\text{Vainilla}) &= 0.7333
 \end{aligned}$$

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

$$\begin{aligned}
 p(\text{Veneno}|\text{Vainilla}) &= \frac{p(\text{Veneno})p(\text{Vainilla}|\text{Veneno})}{p(\text{Vainilla})} \\
 p(\text{Veneno}|\text{Vainilla}) &= \frac{(0.333)(0.5)}{(0.7333)} \\
 p(\text{Veneno}|\text{Vainilla}) &= 0.2272
 \end{aligned}$$

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:

$$\begin{aligned}
 p(\text{No veneno}|\text{Vainilla}) &= \frac{p(\text{No veneno})p(\text{Vainilla}|\text{No veneno})}{p(\text{Vainilla})} \\
 p(\text{No veneno}|\text{Vainilla}) &= \frac{(0.666)(0.85)}{(0.7333)} \\
 p(\text{No veneno}|\text{Vainilla}) &= 0.7726
 \end{aligned}$$

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.

$$\begin{aligned}
 p(\text{Veneno}|\text{Vainilla}) + p(\text{No veneno}|\text{Vainilla}) &= 1 \\
 (0.2272) + (0.7726) &= 1
 \end{aligned}$$

2.a.1. Código y ejercicios de repaso:

El archivo 'FelBayes_Ejercicio1.py' contiene el código que desarrolla, resuelve e ilustra el problema previamente expuesto. Es recomendable abrir el archivo desde un editor de texto o IDE de Python para tener acceso al código y los comentarios añadidos para facilitar su relación con el problema. Revisa y modifica, de ser necesario, el código para dar respuesta a las siguientes preguntas:

- ¿Por qué sustituimos $p(\text{Vainilla}|\text{No veneno}) = 0.85$? ¿De dónde salió ese valor?
- 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:

$$\begin{aligned} p(\text{Zurdo} \cap \text{Pelirrojo}) &= p(\text{Zurdo}) \cdot p(\text{Pelirrojo}) \\ p(\text{Zurdo} \cap \text{Pelirrojo}) &= (0.03) \cdot (0.15) \\ p(\text{Zurdo} \cap \text{Pelirrojo}) &= 0.0045 \end{aligned}$$

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.

Sin embargo, 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 evidencia de su involucramiento en los asesinatos reportados. ¿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 (Es asesino).

$$p(\text{Es el Asesino} | \text{Pelirrojo Zurdo}) = \frac{p(\text{Es el asesino})p(\text{Pelirrojo Zurdo} | \text{Es el asesino})}{p(\text{Pelirrojo Zurdo})}$$

- Escenario 2: El sospechoso es inocente (No es asesino).

$$p(\text{No es el Asesino} | \text{Pelirrojo Zurdo}) = \frac{p(\text{No es el asesino})p(\text{Pelirrojo Zurdo} | \text{No es el asesino})}{p(\text{Pelirrojo Zurdo})}$$

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 habitante tendría una probabilidad inicial de $\frac{1}{30,000}$ de ser el asesino. Por tanto tenemos que:

- Escenario 1:

$$p(\text{Es el asesino}) = \frac{1}{30,000} = 0.00003$$

- Escenario 2:

$$p(\text{No es el asesino}) = \frac{29,999}{30,000} = 0.99996$$

Definimos la verosimilitud como la probabilidad que relaciona la evidencia con cada posible interpretación. Es decir, implica comparar la probabilidad de que, de ser el asesino, el sospechoso

tenga un aspecto idéntico al reportado por el testigo, contra la probabilidad de que el sospechoso sea un ciudadano inocente con características físicas idénticas a las del asesino.

- Escenario 1:

$$p(\text{Pelirrojo Zurdo} | \text{Es el asesino}) = 1$$

Si asumimos que el sospechoso retenido es de hecho el asesino, la probabilidad de que su aspecto físico coincida con la descripción reportada por el testigo, es de 1. En otras palabras, sería imposible que el sospechoso se viera diferente que el asesino, si se tratara del mismo individuo (la probabilidad complementaria, $p(\text{No pelirrojo ni zurdo} | \text{Es el asesino}) = 0$).

- Escenario 2:

$$p(\text{Pelirrojo Zurdo} | \text{No es el asesino}) = 0.0045$$

A pesar de la total certidumbre que se tiene sobre la probabilidad de que el aspecto físico del sospechoso corresponda con el del asesino asumiendo que se trata de la misma persona, existe la posibilidad de que el sospechoso capturado sea simplemente uno de los pocos pelirrojos zurdos que hay en la ciudad. Como se mencionó previamente, la probabilidad de que un ciudadano aleatorio sea simultáneamente zurdo y pelirrojo es igual al producto de las probabilidades de que un ciudadano sea zurdo o pelirrojo. Es decir, aún asumiendo que el sospechoso no es el asesino, sigue existiendo una probabilidad de 0.0045 de que se vea como él.

Finalmente, calculamos la verosimilitud marginal aplicando la Ecuación 7.

$$\begin{aligned} p(\text{Pelirrojo Zurdo}) &= (p(\text{Culpable}) \cdot p(\text{Pelirrojo Zurdo} | \text{Culpable})) + (p(\text{Inocente}) \cdot p(\text{Pelirrojo Zurdo} | \text{Inocente})) \\ p(\text{Pelirrojo Zurdo}) &= ((0.00003) \cdot (1)) + ((0.99996) \cdot (0.0045)) \\ p(\text{Pelirrojo Zurdo}) &= (0.00003) + (0.00449) \\ p(\text{Pelirrojo Zurdo}) &= 0.004529 \end{aligned}$$

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:

$$p(\text{Es el asesino} | \text{Pelirrojo Zurdo}) = \frac{p(\text{Es el asesino})p(\text{Pelirrojo Zurdo} | \text{Es el asesino})}{p(\text{Pelirrojo Zurdo})}$$

$$\begin{aligned} p(\text{Es el asesino} | \text{Pelirrojo Zurdo}) &= \frac{(0.00003)(1)}{(0.004529)} = \frac{0.00003}{0.004529} \\ p(\text{Es el asesino} | \text{Pelirrojo Zurdo}) &= 0.0066 \end{aligned}$$

- Escenario 2:

$$\begin{aligned} p(\text{No es el asesino} | \text{Pelirrojo Zurdo}) &= \frac{p(\text{No es el asesino})p(\text{Pelirrojo Zurdo} | \text{No es el asesino})}{p(\text{Pelirrojo Zurdo})} \\ p(\text{No es el asesino} | \text{Pelirrojo Zurdo}) &= \frac{(0.99996)(0.0045)}{(0.004529)} = \frac{0.004499}{0.004529} \\ p(\text{No es el asesino} | \text{Pelirrojo Zurdo}) &= 0.9935 \end{aligned}$$

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).

$$p(\text{Es el asesino}|\text{Pelirrojo Zurdo}) + p(\text{No es el asesino}|\text{Pelirrojo Zurdo}) = 1 \\ (0.0066) + (0.99359) = 1$$

2.a.1. Código y ejercicios de repaso:

El archivo 'FelBayes_Ejercicio2.py' contiene el código que desarrolla, resuelve e ilustra el problema tal y como se acaba de exponer. El archivo, abierto desde el editor de texto o IDE de Python, contiene los comentarios pertinentes para facilitar su relación con el problema a tratar. Modifica las líneas necesarias para dar respuesta a los siguientes problemas:

- a) Los resultados previamente expuesto demuestran que el aspecto físico del sospechoso no arroja evidencia suficiente para declararlo culpable contra el resto de la población, sin embargo, no dice nada respecto a su inocencia. Lo único que sabemos hasta ahora, es que es mucho más probable que el individuo en cuestión pertenezca al grupo de pelirrojos zurdos en la ciudad, a que sea el único habitante asesino de entre los 30, 000 habitantes. Aún podemos estar lidiando con el asesino entre el grupo de pelirrojos zurdos en la ciudad.
 - a.1) ¿Cuántos habitantes pelirrojos y zurdos hay en total, si sabemos que la probabilidad de ser simultáneamente pelirrojo y zurdo es de 0.0045 y que hay 30000 habitantes en total?
 - a.2) Sabiendo el número total de habitantes pelirrojos en la ciudad, ¿la prior previamente usada tiene sentido? De no tenerlo, ¿cuál debería ser su valor?
 - a.3) ¿La nueva prior cambia la estimación final sobre la probabilidad posterior de que el sospechoso en cuestión sea, o no, el asesino? ¿Por qué ocurre esto?

3.- Simulador del Teorema de Bayes

El archivo 'FelBayes_TeoremaBayes_Simulador.py' corresponde al simulador interactivo del Teorema de Bayes. Se puede acceder al código, con los comentarios que explican su estructura y construcción, desde cualquier editor de texto o IDE de Python. Sin embargo, para poder hacer uso de su cualidad interactiva, se le debe ejecutar desde la consola. Una vez hecho esto, se desplegará una interfaz interactiva como la que se presenta a continuación.

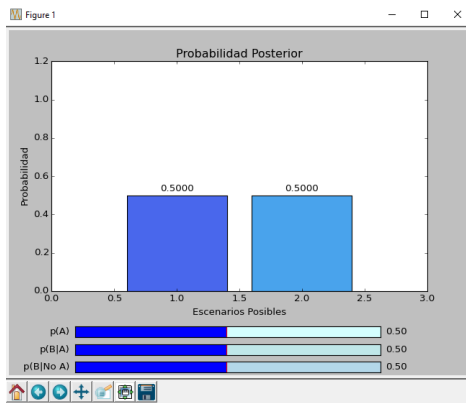


Fig. 1. Interfaz interactiva del Simulador del Teorema de Bayes.

El simulador consta de dos partes: En la parte superior, se muestran los histogramas correspondientes a la probabilidad posterior computada para cada uno de dos posibles escenarios en el mundo (A, y No A). Esto corresponde a una de las inferencias de probabilidad más simples que podría interesarnos computar donde nos interesa saber la probabilidad de que el evento A ocurra, contra su no ocurrencia. Los escenarios posibles contemplados son necesariamente mutuamente excluyentes, por lo que la suma de sus probabilidades debe darnos a 1.

En la parte inferior del simulador, encontraremos tres barras interactivas que permiten al usuario modificar los valores de cada uno de los elementos contenidos en la Regla de Bayes:

- $p(A)$: La primera barra interactiva controla el valor asignado a la probabilidad prior del Evento A.
- $p(B|A)$: La segunda barra interactiva asigna la verosimilitud que relaciona los datos observados con el Evento A. (i.e. ¿Qué tan probables es observar la evidencia, dado que asume que el Evento A ocurre?)
- $p(B|No A)$: La tercera barra corresponde a la verosimilitud que relaciona los datos observados con la ausencia del Evento A (definida como Evento No A). Esta verosimilitud alterna impacta en el cálculo de la Probabilidad Posterior a partir de los cambios que impone en el cómputo de la verosimilitud marginal.

3.a Ejercicios con el simulador

a) Cambia el valor de la probabilidad prior a 1, ¿Qué ocurre?

a.1) Con una prior de 1, cambia el valor de las verosimilitudes, ¿Cómo cambia el estimado de la posterior?, ¿Por qué?

b) Da el valor más bajo posible a la probabilidad prior, ¿Qué ocurre?

a.1) Con la prior en su valor mínimo, cambia el valor de las verosimilitudes, ¿Cómo cambia el estimado de la posterior?, ¿Por qué?

4. Lecturas recomendadas:

Material amigable para principiantes:

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

Material un poco más formal y especializado:

- Mackay, D. (2005) Information theory, inference and learning algorithms.