Clase 5: Teorema de Bayes

Matías Carrasco

16 de marzo de 2020

Resumen Comenzamos con la fórmula de la probabilidad total, que permite organizar los cálculos de manera sencilla en situaciones en donde es útil dividir el problema en casos. Luego seguimos con el Teorema de Bayes que permite invertir el orden de las probabilidades condicionales.

Índice

Fórmula de la probabilidad total

La fórmula de la probabilidad total junta la Axioma 3 de los axiomas de Kolmogorov con la regla del producto para calcular probabilidades. Comencemos con una situación simple, en la que el espacio muestral está dividido en dos casos mutuamente excluyentes C_1 y C_2. Nuestro interés es calcular la probabilidad de un evento A cualquiera.

Como $\Omega = C_1 \cup C_2$, y esta unión es disjunta, podemos particionar el evento A como $A = (C_1 \cap A) \cup (C_2 \cap A)$. Luego, por el Axioma 3, la probabilidad de A se descompone como

$$P(A) = P(C_1 \cap A) + P(C_2 \cap A).$$

Por la regla del producto para probabilidades, podemos reescribirla como

$$P(A) = P(A|C_1)\,P(C_1) + P(A|C_2)\,P(C_2). \tag{1}$$

Esta ecuación representa un caso particular de la fórmula de la probabilidad total.

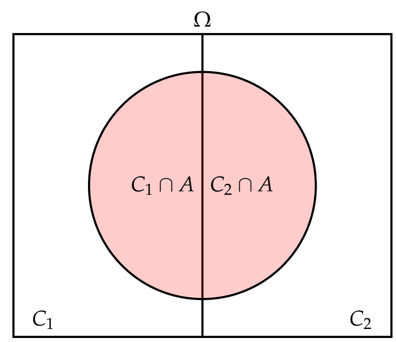


Figura 1: Diagrama de Venn de la fórmula de la probabilidad total.

Ejemplo 1 Una urna contiene 5 bolas rojas y 2 bolas verdes. Dos bolas se extraen de la urna, una a la vez. ¿Cuál es la probabilidad de que la segunda bola sea roja?

Vamos a resolver el problema de dos formas distintas. Si bien las bolas son idénticas entre sí, excepto por el color, podemos numerarlas para distinguirlas sin afectar las probabilidades de extracción. La urna consiste entonces de las bolas

$$\text{Urna} = \boxed{r_1 \quad r_2 \quad r_3 \quad r_4 \quad r_5 \quad v_1 \quad v_2}$$

El espacio muestral es entonces

$$\Omega = \{(x, y) : x \neq y \in \mathrm{U}\}.$$

Sean los eventos

- R_1 = "la primera bola es roja"
- V_1 = "la primera bola es verde"

- $R_2 =$ "la segunda bola es roja"
- $V_2 =$ "la segunda bola es verde".

Se nos pide calcular $P(R_2)$.

La forma rápida de calcular esto es con el siguiente razonamiento: cada bola tiene igual chance de ser la segunda bola. Como 5 de las 7 bolas son rojas, $P(R_2) = 5/7$. La probabilidad $P(R_2) = 5/7$ puede parecer sorprendente ya que el color de la primera bola afecta ciertamente las probabilidades para la segunda bola. Para convencernos hagamos el conteo directo: la cantidad de pares de Ω en los cuales y es roja es $6 \cdot 5$, pues hay 5 posibilidades rojas para la segunda bola, y por cada una de ellas hay 6 posibilidades para la primera (recordar que no importa el orden en el que llenamos los casilleros en la regla del producto). El cardinal de Ω es $6 \cdot 7$, por lo que obtenemos la misma probabilidad de $(6 \cdot 5)/(6 \cdot 7) = 5/7$. Otra forma de ver esto es: si no se nos da el color de la primera bola, entonces debemos considerar todas las posibilidades para la segunda bola.

Calculemos este mismo valor usando la fórmula de la probabilidad total (1). Primero, encontraremos las probabilidades condicionales. Este es un ejercicio de conteo simple:

$$P(R_2|R_1) = 4/6, \quad P(R_2|V_1) = 5/6.$$

Como R_1 y V_1 particionan el espacio muestral (juegan el papel de C_1 y C_2 en 1), tenemos que

$$P(R_2) = P(R_2|R_1)\,P(R_1) + P(R_2|V_1)\,P(V_1) = \frac{4}{6} \cdot \frac{5}{7} + \frac{5}{6} \cdot \frac{2}{7} = \frac{30}{42} = \frac{5}{7}. \tag{2}$$

■

No se necesita mucho para hacer un ejemplo donde (1) sea realmente la mejor manera de calcular la probabilidad. He aquí un juego con reglas un poco más complicadas.

Ejemplo 2 Una urna contiene 5 bolas rojas y 2 bolas verdes. Se extrae una bola. Si es verde, se agrega una bola roja a la urna y si es roja se agrega una bola verde a la urna. (La bola original no se vuelve a poner en la urna). Luego, se extrae una segunda bola. ¿Cuál es la probabilidad de que la segunda bola sea roja?

La fórmula de probabilidad total dice que $P(R_2)$ se puede calcular utilizando la expresión en la ecuación (2). Solo los valores para las probabilidades condicionales cambiarán. Tenemos

$$P(R_2|R_1) = 4/7, \quad P(R_2|V_1) = 6/7.$$

Por lo tanto

$$P(R_2) = P(R_2|R_1)\,P(R_1) + P(R_2|V_1)\,P(V_1) = \frac{4}{7} \cdot \frac{5}{7} + \frac{6}{7} \cdot \frac{2}{7} = \frac{32}{49}.$$

Urnas en probabilidad:

El Ejemplo 1 de urnas es un clásico y su uso se remonta a los orígenes de la probabilidad como disciplina. Es un modelo de juguete con muchísimas aplicaciones a situaciones reales.

Citamos de Wikipedia (`http://en.wikipedia.org/wiki/Urn_problem`):

> En probabilidad y estadística, un problema de urna es un ejercicio mental idealizado en el que algunos objetos de interés real (como átomos, personas, automóviles, etc.) se representan como bolas de colores en una urna u otro recipiente. Uno extrae una o más bolas de la urna y el objetivo es determinar la probabilidad de extraer un color u otro, o algunas otras propiedades.

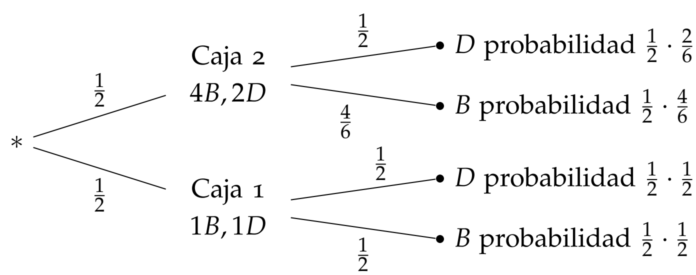
Es en este tipo de ejemplo en donde se ve claramente el potencial de (1). ■

Experimentos secuenciales y árboles

Los experimentos secuenciales son aquellos que constan de varias etapas, en donde el resultado de la i-ésima etapa depende de las etapas anteriores. En los experimentos secuenciales resulta útil dibujar árboles para llevar un registro de las probabilidades en cada etapa. Veamos algunos ejemplos sencillos para ver cómo funciona esta idea.

Ejemplo 3 Dos cajas tienen productos de una cierta industria. Una caja contienen un producto bueno y uno defectuoso. La otra caja contiene 4 productos buenos y 2 defectuosos. Se elige al azar una caja, de la cual también al azar se extrae un producto. Calcular la probabilidad de que el producto extraído resulte bueno.

Para empezar, construimos un árbol en el cual los nodos del primer nivel representan las cajas y los del segundo la calidad del producto.



Las aristas contienen etiquetas que representan la probabilidad del evento determinado por el nodo del árbol. Cuando seguimos un camino desde la raíz del árbol ($*$) hasta un nodo terminal, obtenemos una realización particular de un determinado evento. Si multiplicamos las probabilidades que aparecen como etiquetas de las aristas del camino obtenemos la probabilidad de dicho evento.

Así, es fácil calcular la probabilidad del evento

$$A = \{\text{el producto es bueno}\}.$$

Basta sumar las probabilidades de todos los caminos que terminan en nodos con una B:

$$\frac{1}{2}\left(\frac{1}{2} + \frac{4}{6}\right) = \frac{7}{12}.$$

¿Cómo podemos justificar este cálculo? Podemos considerar los eventos

$$C_1 = \{\text{se elige la caja 1}\} \text{ y } C_2 = \{\text{se elige la caja 2}\}.$$

Éstos forman una partición del espacio muestral. Notar que no hemos tenido necesidad de definir el espacio muestral, esta es la gran ventaja de (1). Las probabilidades condicionales son

$$P(A|C_1) = \frac{1}{2}, \; P(A|C_2) = \frac{4}{6}, \; \text{ y } \; P(C_i) = \frac{1}{2},$$

y por la fórmula de la probabilidad total resulta

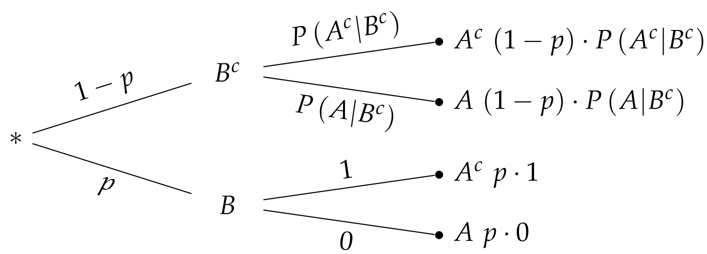
$$P(A) = P(A|C_1)\,P(C_1) + P(A|C_2)\,P(C_2) = \frac{1}{2}\left(\frac{1}{2} + \frac{4}{6}\right) = \frac{7}{12}.$$

Ejemplo 4 Una moneda sesgada (con probabilidad de obtener cara igual a $p > 0$) se lanza repetidamente hasta que salga cara. Calcular la probabilidad de que la primer cara aparezca en un número par de intentos.

Este ejemplo ya lo hemos visto, pero ahora lo resolveremos usando árboles. Sean

- $A =$ "se lanza la moneda un número par de veces"
- $B =$ "sale cara en el primer lanzamiento".

Construyamos un árbol como en los ejemplos anteriores:



El dato que nos falta es $P(A|B^c)$. Sin embargo, notar que el experimento continua si el primer lanzamiento resulta en cruz. Como lo que ocurre después del primer lanzamiento es independiente del mismo, es como si el experimento comenzara nuevamente. Lo que cambia es que para que A ocurra debe salir cara en una cantidad impar de lanzamientos (contando a partir del segundo). Luego $P(A|B^c) = P(A^c) = 1 - P(A)$. Entonces

$$\begin{aligned} P(A) &= P(A|B)\,P(B) + P(A|B^c)\,P(B^c) = P(A|B^c)\,P(B^c) \\ &= (1 - P(A))P(B^c) = (1 - P(A))(1-p) \end{aligned}$$

De aquí podemos despejar $P(A)$ para obtener

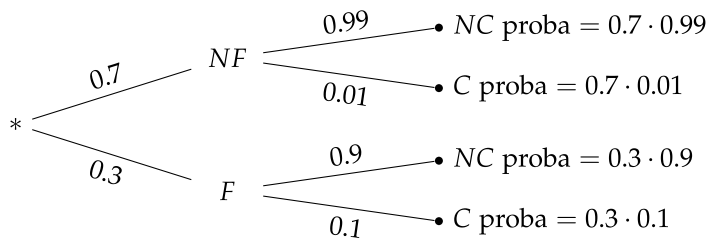
$$P(A) = \frac{1-p}{2-p}.$$

Notar que si $p = 1/2$ entonces $P(A) = 1/3$ como habíamos calculado antes.

Aunque el experimento no sea secuencial, podemos usar árboles para organizar los cálculos y que éstos resulten más sencillos.

Ejemplo 5 En cierta población hay un 30 % de fumadores. Se sabe que la probabilidad de enfermarse de cáncer de pulmón es igual a 0.1 para los fumadores y 0.01 para los no fumadores. Calcular la probabilidad de que una persona elegida al azar en la población se enferme de cáncer de pulmón.

Denotemos por F el evento ser fumador, por NF el de no ser fumador, y por C el de enfermarse de cáncer. Lo mejor es hacer un árbol:



Entonces

$$\begin{aligned} P(C) &= P(C|F)\,P(F) + P(C|NF)\,P(NF) \\ &= 0.1 \cdot 0.3 + 0.01 \cdot 0.7 = 0.037. \end{aligned}$$

Para resumir, enunciemos la fórmula de la probabilidad total en su versión general.

Fórmula general de la probabilidad total

Sean $C_1, C_2, \ldots,$ una partición numerable de Ω cuyos eventos tienen probabilidades positivas. Sea A un evento cualquiera. Entonces

$$P(A) = \sum_{i=1}^{\infty} P(A|C_i)\,P(C_i). \tag{3}$$

Demostración. La prueba sigue el mismo razonamiento que usamos para (1). Notar primero que $\Omega = \bigcup_{i=1}^{\infty} C_i$. Luego, podemos descomponer el evento A como

$$A = A \cap \Omega = \bigcup_{i=1}^{\infty} (A \cap C_i).$$

Al aplicar probabilidades, como los eventos C_i's son disjuntos dos a dos, obtenemos

$$P(A) = P\left(\bigcup_{i=1}^{\infty} (A \cap C_i)\right) = \sum_{i=1}^{\infty} P(A \cap C_i) = \sum_{i=1}^{\infty} P(A|C_i)\,P(C_i).$$

En la última igualdad hemos usado la regla del producto $P(A \cap C_i) = P(A|C_i)\,P(C_i)$. □

Invirtiendo probabilidades condicionales

Ya sabemos que las probabilidades condicionales no son simétricas, esto es $P(A|B) \neq P(B|A)$. Sin embargo, existe una relación entre ambas que aunque es simple, es muy importante y se conoce como Teorema de Bayes.

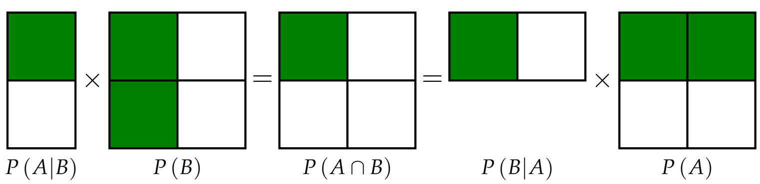
Supongamos que A y B son dos eventos con probabilidades positivas. Reescribiendo la definición de probabilidad condicional tenemos que

$$P(A \cap B) = P(A|B)\,P(B) \quad \text{y} \quad P(A \cap B) = P(B|A)\,P(A).$$

Como los miembros de la izquierda son iguales en ambos casos, obtenemos la igualdad siguiente:

$$P(A|B)\,P(B) = P(B|A)\,P(A). \tag{4}$$

Esta relación permite escribir una probabilidad condicional en función de la otra. Podemos visualizarla de la siguiente manera:



Esta relación junto con la fórmula de la probabilidad total (3) dan como resultado lo que se conoce como la fórmula de Bayes.

Fórmula de Bayes

Sean $C_1, C_2, \ldots,$ una partición numerable de Ω cuyos eventos tienen probabilidades positivas. Sea A un evento con probabilidad positiva. Entonces

$$P(C_k|A) = \frac{P(C_k)\,P(A|C_k)}{\sum_{i=1}^{\infty} P(C_i)\,P(A|C_i)}. \tag{5}$$

Demostración. Usando la fórmula de inversión de probabilidades condicionales (4) tenemos que

$$P(C_k|A) = \frac{P(A|C_k)\,P(C_k)}{P(A)}.$$

De la fórmula de la probabilidad total sabemos que

$$P(A) = \sum_{i=1}^{\infty} P(A|C_i)\, P(C_i).$$

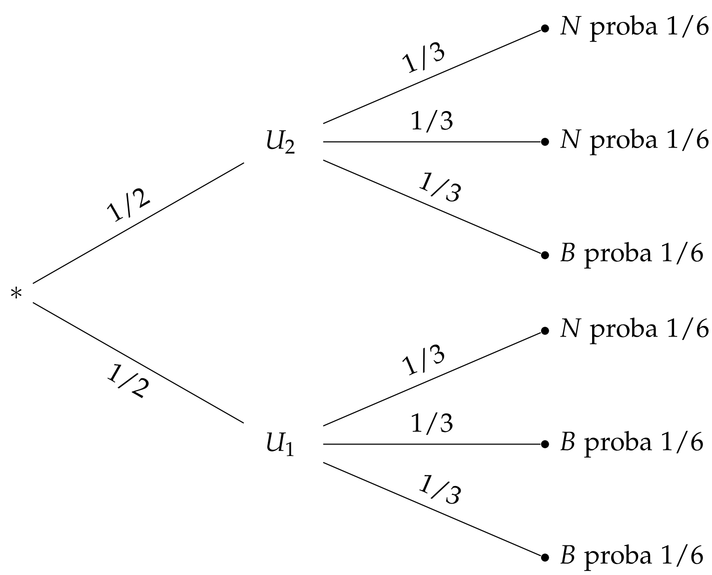
Reemplazando esta última en la ecuación anterior obtenemos (5). $\square$

Ejemplo 6 En una primera urna se tienen 2 bolas blancas y 1 negra, y en una segunda, 2 negras y 1 blanca. Se elige al azar una urna, y de ella también al azar se extrae una bola. ¿Cuál es la probabilidad de que la urna elegida haya sido la segunda, dado que la bola extraída es blanca?

Denotemos por U_i el evento "se elige la urna i" para $i = 1$ y 2. Entonces

$$\begin{aligned} P(U_2|\text{blanca}) &= \frac{P(\text{blanca}|U_2)\, P(U_2)}{P(\text{blanca}|U_1)\, P(U_1) + P(\text{blanca}|U_2)\, P(U_2)} \\ &= \frac{1/3 \cdot 1/2}{2/3 \cdot 1/2 + 1/3 \cdot 1/2} = 1/3. \end{aligned}$$

En este caso, en dónde las urnas y las bolas son equiprobales, podemos entender mejor el cálculo usando un árbol de posibilidades.



Notar que de las 3 blancas, solo hay una que proviene de la urna 2, por eso la probabilidad condicional es $1/3$. En general, el método del árbol de posibilidades funciona igual, pero debemos ponderar con las respectivas probabilidades condicionales. $\blacksquare$

Una falacia conocida en probabilidad, la *falacia de la frecuencia base*, muestra que es fácil confundir el significado de $P(B|A)$ y $P(A|B)$ cuando una situación se describe con palabras. Este es uno de los ejemplos clave de probabilidad condicional, al punto que será básico para la correcta interpretación de conceptos estadísticos que veremos más adelante. Es importante entenderlo a fondo.

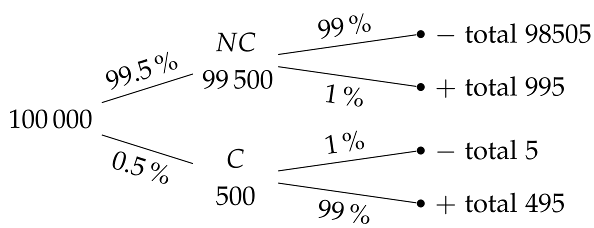
Ejemplo 7 Un test de drogas produce 99 % verdaderos resultados positivos para los consumidores de una determinada droga y 99 % verdaderos resultados negativos para los no consumidores. Supongamos que el 0.5 % de la población son consumidores de la droga. Si una persona elegida al azar resulta positivo, ¿cuál es la probabilidad de que sea un consumidor?

Denotemos por $+$ y $-$ los eventos "resultado positivo" y "resultado negativo" respectivamente, y por C y NC los eventos "es consumidor" y "no es consumidor". Entonces

$$\begin{aligned} P(C|+) &= \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|NC)\,P(NC)} \\ &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} = 0.332 \approx 33\,\%. \end{aligned}$$

Notar que aunque los tests son muy efectivos, pues producen falsos resultados con muy baja probabilidad, la probabilidad de ser consumidor dado que el test es positivo es más bien baja. Este resultado es un poco paradójico, y se explica porque hay poca población que es consumidora de la droga.

Una buena forma de visualizar esto es usando árboles como habíamos hecho antes. Imaginemos una población de 100 000 personas. En ésta, 0.5 % serán consumidores de la droga:



Completando el árbol vemos que el total de personas que esperamos sean positivas en el test es $495 + 995 = 1490$. De estos, solamente

$$\frac{495}{1490} \approx 0.33$$

son consumidores de la droga.

¿Qué pasa si variamos la proporción de gente que consume la droga? Supongamos que en lugar de 0.5 % el porcentaje de consumidores de la droga es $p \times 100\,\%$. En este caso la probabilidad resulta

$$P(C|+) = \frac{0.99 \cdot p}{0.99 \cdot p + 0.01 \cdot (1-p)}.$$

En el gráfico de la derecha vemos $P(C|+)$ en función de p. Notar cómo decae rápidamente a medida que p se hace cada vez menor.

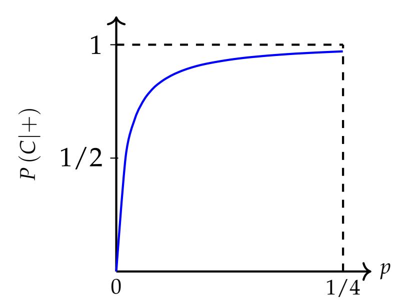


Figura 2: Probabilidad condicional de consumir dado que el test es positivo en función de p.

Para pensar: ¿Cuanto debe ser la proporción de consumidores de la droga para que la probabilidad $P(C|+) = 0.99$?

Esto se conoce como la falacia de la frecuencia base porque la frecuencia base de consumidores en la población es tan baja que la gran mayoría de las personas que toman la prueba no lo son, e incluso con una prueba tan precisa, la mayoría de los positivos serán personas que no consumen.

Para resumir la falacia de la frecuencia base con números específicos:

> el 99 % de todas las pruebas son correctas, no implica que el 99 % de las pruebas positivas sean correctas.

Nos referiremos a este ejemplo bastante seguido. Este y otros ejemplos similares están en el corazón de muchos malentendidos estadísticos.

Ejemplo 8 Otro truco que es útil para calcular probabilidades es hacer una tabla. Vamos a rehacer el ejemplo anterior utilizando una tabla construida con 100000 personas totales dividida de acuerdo con las probabilidades del ejemplo.

Construimos la tabla de la siguiente manera. Las 10000 personas forman el total general en la esquina inferior derecha. Utilizando $P(C) = 0.05$, calculamos que 500 de las 100000 personas son consumidoras. Asimismo, 99950 personas no lo son. En este punto la tabla se ve como:

	C	NC	total
+			
-			
total	500	99500	100000

Usando $P(+|C) = 0.99$ podemos calcular que el número de consumidores con resultado positivo es el 99 % de 500 o 495. Las otras entradas son similares. En este punto, la tabla se ve como

	C	NC	total
+	495	995	
-	5	98505	
total	500	99500	100000

Finalmente, sumamos las filas $+$ y $-$ para obtener la tabla completa

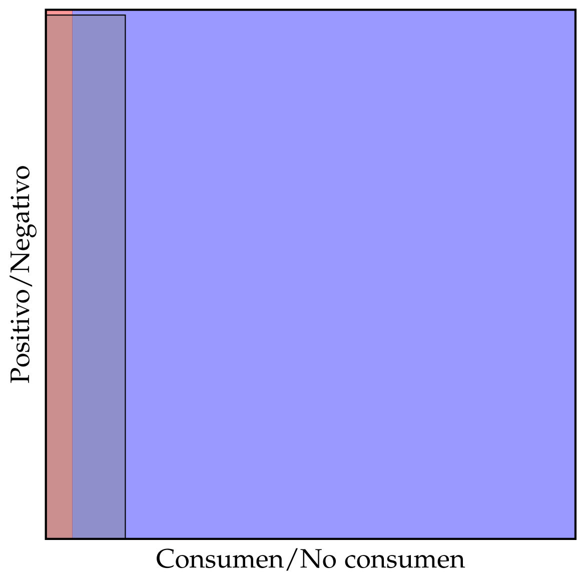
	C	NC	total
+	495	995	1490
-	5	98505	98510
total	500	99500	100000

Usando la tabla completa calculamos

$$P(C|+) = \frac{|C \cap +|}{|+|} = \frac{495}{1490} = 33\,\%$$

■

La siguiente figura ilustra la falacia de la frecuencia base. La gran zona azul representa a todas las personas no consumidoras. El área roja mucho más pequeña representa a los consumidores. El rectángulo sombreado representa a las personas que dan positivo. El área sombreada cubre la mayor parte del área roja y solo una pequeña parte del área azul. Aun así, la mayor parte del área sombreada es sobre el azul. Es decir, la mayoría de las pruebas positivas son de personas no consumidoras.



Un casino Bayesiano

Carla y Walter están jugando a un juego en el que la primera persona que consigue 6 puntos gana. La forma en que cada punto se decide es un poco extraña.

El Casino tiene 17 urnas que Carla y Walter no pueden ver pues se encuentran escondidas en un depósito. Las urnas contienen bolas blancas y negras, en diferentes proporciones. Si imaginamos que las urnas están numeradas del 0 al 16, la i-ésima urna tiene i bolas blancas y $16 - i$ bolas negras. Así, la urna 0 tiene todas las bolas negras, mientras que la urna 16 tiene todas las bolas blancas, y el resto de las urnas tiene cantidades intermedias de bolas blancas y negras. A excepción de la proporción de bolas blancas y negras, las urnas son idénticas entre sí.

Antes de que empiece el juego el Casino elige una de las urnas al azar. Luego, cada punto es decidido al azar de la siguiente manera: se extrae una bola de la urna, si la bola es blanca, Carla gana el punto; si es negra, Walter gana el punto. Luego la bola se vuelve a poner en la urna y se extrae otra bola, y así sucesivamente.

Claramente, la probabilidad de que Carla gane un punto es igual a la proporción de bolas blancas en la urna. Llamemos a esta probabilidad p, por lo que la probabilidad de que Walter gane un punto es $1-p$. Debido a que el Casino eligió al azar la urna con la cual jugar, cada valor de p es igualmente probable. La urna solo se elige al principio del juego, por lo que p es el mismo para cada punto.

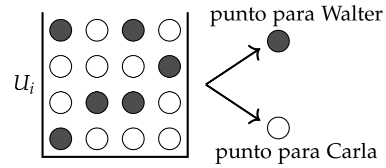


Figura 3: Si la bola es negra es un punto para Walter, si es blanca es un punto para Carla.

Supongamos que Carla ya está ganando 5 puntos a 3. ¿Cuál es la probabilidad de que Carla gane?

La proporción p de bolas blancas en la urna es una variable aleatoria, y puede tomar cualquiera de los valores

$$p_i = \frac{i}{16} \quad \text{para } i = 0, 1, \ldots, 16.$$

La respuesta es fácil si conocemos el valor de p. De hecho, para un valor de p determinado, la probabilidad de que Walter gane el juego es $(1-p)^3$ ya que si los siguientes tres puntos no son para él, entonces Carla gana el juego. Por tanto la probabilidad de que Carla gane es

$$1 - (1-p)^3.$$

Sin embargo, el verdadero valor de p no lo sabremos nunca con certeza. Uno está tentado a usar la información dada para adivinar un valor de p. Sin embargo, esto sería erróneo, y podemos calcular la probabilidad de que Carla (o Walter que es más fácil) gane el juego, sin adivinar directamente el valor de p.

Sabemos que se han jugado 8 turnos, y que Carla tiene 5 puntos y Walter 3. En otras palabras, sabemos que el evento

$$D = \{\text{se juegan al menos 8 turnos, Carla tiene 5 puntos y Walter 3}\}$$

ha ocurrido.

Comencemos por calcular la probabilidad de D. Usando la fórmula de la probabilidad total, tenemos que

$$P(D) = \sum_{i=0}^{16} P\left(D \middle| p = p_i\right) P\left(p = p_i\right).$$

Si $p = p_i$ la probabilidad de que Carla gane 5 puntos y Walter 3 es

$$P\left(D \middle| p = p_i\right) = \binom{8}{5} p_i^5 (1-p_i)^3.$$

Esta fórmula la podemos obtener razonando de la siguiente manera: el juego puede transcurrir de varias formas, pero en total sabemos que debe darse una serie de resultados del tipo (W, C, C, W, C, W, C, C), en donde la letra C significa que el punto fue para Carla y la letra W que fue para Walter. Como la probabilidad de que el punto sea para Carla es p_i y la probabilidad de que el punto sea para Walter es $1 - p_i$ (estamos asumiendo que $p = p_i$), vemos que cada una de estas secuencias tiene probabilidad $p_i^5(1-p_i)^3$. Resta entonces contar cuántas secuencias distintas hay. Sabemos que tiene largo 8, 5 letras C y 3 letras W. Sin embargo, eligiendo los lugares de las letras C quedan determinados los lugares de las letras W. Esto se puede hacer de $\binom{8}{5}$ formas distintas.

Además sabemos que el Casino elige la urna al azar, por lo que $P(p = p_i) = 1/17$. Juntando ambas cosas obtenemos

$$P(D) = \frac{1}{17}\binom{8}{5}\sum_{i=0}^{16} p_i^5(1-p_i)^3 = \frac{1}{17}\frac{1}{16^8}\binom{8}{5}\sum_{i=0}^{16} i^5(16-i)^3.$$

Esta fórmula es un poco asustadora pero una computadora la puede calcular sin problemas.

Ahora podemos usar el teorema de Bayes para calcular la probabilidad de que p sea igual a p_j, dados los puntos de Carla y Walter (esto es dado D). Aplicando la fórmula obtenemos:

$$P\left(p = p_j | D\right) = \frac{P\left(D | p = p_j\right) P\left(p = p_j\right)}{P(D)}$$

Sustituyendo los valores que hemos calculado más arriba, la probabilidad de que p sea igual a p_j queda

$$P\left(p = p_j | D\right) = \frac{j^5(16-j)^3}{\sum_{i=0}^{16} i^5(16-i)^3}.$$

Podemos graficar el lado derecho en función de j para ver cuál es el valor más probable de p, ver la Figura 4.

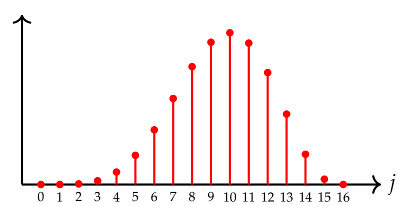


Figura 4: Probabilidad $P\left(p = p_j | D\right)$ en función de j.

Notar que el máximo de $P\left(p = p_j | D\right)$ se da en $j = 10$, lo cual corresponde a $p = p_{10} = 5/8$. Este sería el valor más probable para p dado los datos D del juego que tenemos hasta el momento, y si la pregunta hubiera sido "adivinar con qué urna están jugando" la mejor respuesta sería "con la urna 10". Podríamos estimar así la probabilidad de que Carla gane el juego como

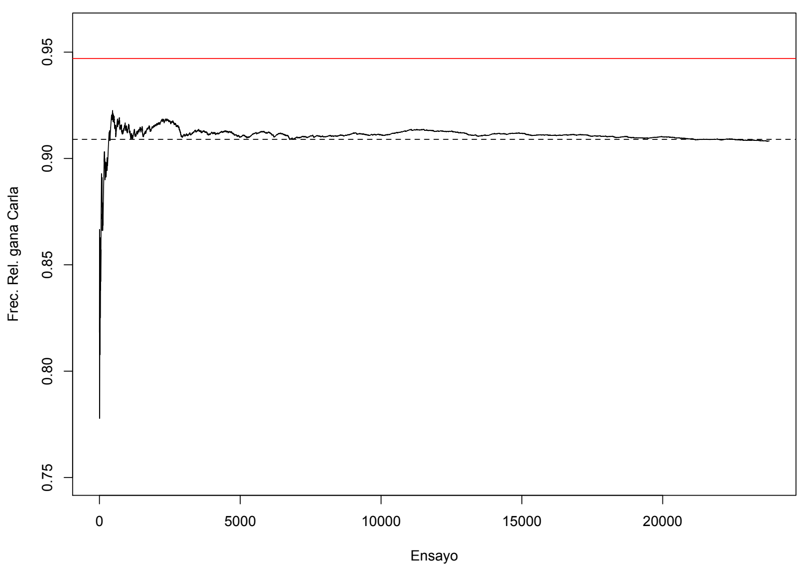
$$1 - (1 - 5/8)^3 \approx 0.947.$$

Pero esta estimación está un poco por arriba del verdadero valor. Para calcularla correctamente usaremos de nuevo la fórmula de la probabilidad total:

$$\begin{aligned} P(\text{gane C}|D) &= \sum_{j=0}^{16} P\left(\text{gane C} | p = p_j, D\right) P\left(p = p_j | D\right) \\ &= \sum_{j=0}^{16} \left(1 - (1-p_j)^3\right) P\left(p = p_j | D\right). \end{aligned}$$

De nuevo, aunque las cuentas son asustadoras, una computadora las hace en menos de un segundo. El resultado es $P(\text{gane C}|D) = 0.909$.

Para aquellos que no confían mucho en las cuentas que hicimos, este es un buen ejemplo en el cual la simulación por computadora nos puede ayudar. Para forzar la ocurrencia del evento D simplemente debemos olvidarnos de aquellas veces en las cuales D no ocurre. Para cada simulación en la cual D sí ocurre, vemos si Carla efectivamente gana el juego. Así, contamos el total de veces que Carla gana el juego entre aquellas veces en que D ha ocurrido, y la frecuencia relativa aproximará, si repetimos muchas veces el juego, la probabilidad de que Carla gane (dado D).



En la figura de arriba hicimos la simulación del juego 200 000 veces. Entre estas, 23 792 veces el evento D ocurrió. La frecuencia relativa de veces, entre estas 23 792, en las cuales Carla ganó el juego fue de 0.908. El gráfico muestra cómo fue cambiando la frecuencia relativa a medida que la computadora iba jugando. La linea horizontal roja es el valor más probable de p que calculamos anteriormente (0.947). La linea punteada a la cual convergen las frecuencias relativas es 0.909.