4. Paradigma probabilístico

1. Considere el siguiente problema: En una fábrica hay dos maquinas. La primera realiza el 60% de la producción total y la segunda realiza el 40% restante. Pero de la producción total ya sabemos que el 3% de la producción defectuosa viene de la primer máquina mientras que el 5% viene de la segunda. Si hemos encontrado un producto defectuoso ¿con qué probabilidad viene de la segunda máquina?

   Solución. Consideremos el espacio muestral $\Omega = M_1 \cup M_2$ y al evento $D \subseteq \Omega$, en donde:

   - M_1: Producto de la máquina 1.
   - M_2: Producto de la máquina 2.
   - D: Producto defectuoso.

   Por lo cual tenemos la siguiente información:

   - $P(M_1) = 0.6$
   - $P(M_2) = 0.4$
   - $P(D \mid M_1) = 0.03$
   - $P(D \mid M_2) = 0.05$

   La probabilidad de que un producto haya sido fabricado por la segunda máquina dado que ya sabíamos que el producto está defectuoso, está dada por:

   $$P(M_2 \mid D) = \frac{P(M_2 \cap D)}{P(D)}$$

   Sabemos que

   $$P(D \mid M_2) = \frac{P(D \cap M_2)}{P(M_2)}$$

   Al combinar las dos expresiones tenemos

   $$P(M_2 \mid D) = \frac{P(M_2)P(D \mid M_2)}{P(D)}$$

   En donde

   $$P(D) = P(M_1)P(D \mid M_1) + P(M_2)P(D \mid M_2) = 0.038$$

   Por lo tanto

   $$P(M_2 \mid D) = \frac{P(M_2)P(D \mid M_2)}{P(D)} = \frac{0.4(0.05)}{0.038} \approx 0.5263$$

2. Suponga que tenemos una prueba para detectar una enfermedad muy rara en la población. Sobre la eficacia podemos decir lo siguiente:

   Sea N el evento de que la prueba resulte negativa y E el evento de que el paciente tenga esta enfermedad. Sabemos que:

   $$P(N^c \mid E) = 0.95$$

   $$P(N \mid E^c) = 0.96$$

   $$P(E) = 0.01$$

   Es fácil comprobar que la probabilidad de que el paciente tenga la enfermedad dado que la prueba resultó negativa es de:

   $$P(E \mid N) = 0.000526$$

   Esto es lo que conocemos como un falso negativo, ya que el evento es N pero estamos considerando que el paciente efectivamente tiene la enfermedad E.

   Calcula las probabilidades para un verdadero positivo $P(E \mid N^c)$, las de un falso positivo $P(E^c \mid N^c)$ y la de un verdadero negativo $P(E^c \mid N)$.

   Solución. Consideremos el espacio muestral $\Omega = E \cup E^c$ y al evento $N \subseteq \Omega$ como se muestra en la figura 1, y en donde:

   - E: Tener la enfermedad.
   - N: Dar negativo en la prueba.

   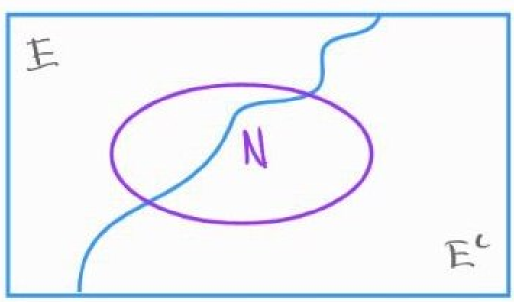

   Figure 1: Espacio muestral del ejercicio 4.2

   También entendemos a E^c como no tener la enfermedad y a N^c como dar positivo en la prueba. Tenemos que

   $$P(E^c) = 1 - P(E) = 0.99$$

   Para calcular las probabilidades que nos piden, necesitamos lo siguiente:

   - $P(N \cap E)$

- $P(N \cap E^c)$
- $P(N^c \cap E)$
- $P(N^c \cap E^c)$
- $P(N)$

Las segunda y tercera salen directamente de la información con la que contamos, entonces

$$P(N \cap E^c) = P(E^c)P(N \mid E^c) = 0.99(0.96) = 0.9504$$
$$P(N^c \cap E) = P(E)P(N^c \mid E) = 0.01(0.95) = 0.0095$$

Ahora calculemos $P(N)$, por un lado sabemos que

$$P(N) = P(E)P(N \mid E) + P(E^c)P(N \mid E^c)$$

Por otro lado también sabemos que

$$P(N)P(E \mid N) = P(E)P(N \mid E)$$

Al combinar las dos expresiones tenemos

$$P(N) = \frac{P(E^c)P(N \mid E^c)}{1 - P(E \mid N)} = \frac{0.99(0.96)}{1 - 0.000526} \approx 0.9509$$

Vamos a calcular la primera y lo cual ya es directo

$$P(N \cap E) = P(N)P(E \mid N) = 0.9509(0.000526) \approx 0.0005$$

Para calcular la cuarta, notemos que cada una de las intersecciones representa un conjunto en la figura 1, es decir, tenemos lo siguiente

$$\Omega = (N \cap E) \cup (N \cap E^c) \cup (N^c \cap E) \cup (N^c \cap E^c)$$

También notemos que cualesquiera dos uniendos son ajenos a pares, por lo cual

$$1 = P(\Omega) = P(N \cap E) + P(N \cap E^c) + P(N^c \cap E) + P(N^c \cap E^c)$$

Entonces

$$P(N^c \cap E^c) = 1 - (P(N \cap E) + P(N \cap E^c) + P(N^c \cap E)) = 1 - (0.0005 + 0.9504 + 0.0095) = 0.0396$$

Finalmente calculemos las tres probabilidades que nos piden. Verdadero positivo:

$$P(E \mid N^c) = \frac{P(E \cap N^c)}{P(N^c)} = \frac{P(E \cap N^c)}{1 - P(N)} = \frac{0.0095}{1 - 0.9509} \approx 0.1935$$

Falso positivo:

$$P(E^c \mid N^c) = \frac{P(E^c \cap N^c)}{P(N^c)} = \frac{P(E^c \cap N^c)}{1 - P(N)} = \frac{0.0396}{1 - 0.9509} \approx 0.8065$$

Verdadero negativo:

$$P(E^c \mid N) = \frac{P(E^c \cap N)}{P(N)} = \frac{0.9504}{0.9509} \approx 0.9995$$

3. Considerando que la Entropía de Shannon se calcula como

$$H(X) = -\sum_{i}^{n} p(x) \cdot \log_2(p(x))$$

Calcula la entropıa de las siguientes cadenas:

(b) 010100110001110101001100011101010011000111

1.0

(c) 000000000000000000000000000000000000000000

0

(d) 0101010101011011011101110010101010010111010

0.9934472383802027

(e) ¿Cuál de estas cadenas dirías que fue generada al azar y cuál tiene un patrón computable?

Diría que la primera es un patrón computable y la 3 fue elegida al azar, pues al probar la función con una cadena que yo elegí al azar obtuve un resultado muy similar al de la tercera cadena.