MSEL
Modelling in Science Education and Learning
Modelling in Science Education and Learning
Volume 12 (2), 2019 DOI: 10.4995/msel.2019.10830
Instituto Universitario de Matemática Pura y Aplicada
Universitat Politècnica de València

Redes bayesianas y diagnóstico médico. Una forma diferente de aprender probabilidades condicionadas

Bayesian nets and medical diagnosis. A different way to learn conditional probabilities

Vicente D. Estruch, Francisco J. Boigues, Anna Vidal, José I. Pastor
Universitat Politècnica de València
vdestruc@mat.upv.es, fraboipl@mat.upv.es, avidal@mat.upv.es, jpastogi@mat.upv.es

Abstract

Las redes bayesianas constituyen una herramienta formal que permite modelar procesos caracterizados por la incertidumbre, lo cual es propio de muchos problemas reales. Con una red bayesiana se puede establecer un modelo completo sobre un conjunto de variables aleatorias y sus relaciones. Dicho modelo se puede utilizar para estimar probabilidades de ciertas variables de la red, que se denominan variables de estado, cuando son fijadas otras variables, llamadas variables de evidencia. El proceso de obtener la distribución de probabilidad de las variables de estado dadas las evidencias se denomina inferencia probabilística bayesiana. En este trabajo, tras introducir las redes bayesianas, se expone cómo utilizarlas en el aula analizando problemas de diagnóstico médico. Con esto se persigue un aprendizaje más significativo de los conceptos de independencia y de probabilidad condicional, que son esenciales para una correcta aplicación de numerosos métodos probabilísticos y estadísticos.

Bayesian networks are a formal tool that allows to model processes characterized by uncertainty, which is typical of many real problems. A Bayesian network can establish a comprehensive model on a set of random variables and their relationships. This model can be used to estimate probabilities of certain variables of the network, which are called state variables, when other variables, named evidence variables, are fixed. The process of obtaining the probability distribution of the state variables, when the evidences are fixed, is named Bayesian probabilistic inference. In this work, after introducing the Bayesian networks, it is exposed how to use them in the classroom analysing medical diagnosis problems. This leads to a more meaningful learning of the concepts of independence and conditional probability, which are essential for the correct application of numerous probabilistic and statistical methods.

Palabras clave: Redes bayesianas, Teorema de Bayes, modelización matemática, diagnóstico médico.
Keywords: Bayesian networks, Bayes theorem, mathematical modeling, medical diagnosis.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

1. Introducción

La probabilidad es un concepto básico en la estadística inferencial por ser fundamento de muchas otras nociones (Boigues y Estruch, 2017). Es obvio que hablar de la probabilidad es hablar de la regla de Laplace, pero también de la perspectiva frecuentista que asocia a cada suceso la frecuencia relativa del mismo después de experimentar muchas veces. Otro enfoque interesante de la probabilidad es el enfoque bayesiano, donde la probabilidad de un suceso puede variar en función de evidencias que modifiquen las circunstancias iniciales. Si queremos que los estudiantes adquieran fundamentos sólidos en el aprendizaje de la probabilidad, tenemos que conseguir que incorporen a su esquema probabilístico las diferentes perspectivas de la probabilidad.

Cuando se estudia el Teorema de Bayes, algunos conceptos relacionados con el mismo, como por ejemplo las nociones de probabilidad condicionada o la probabilidad conjunta, suelen presentar problemas serios de comprensión. De hecho, algunos libros recientes optan por ni tan siquiera presentar el Teorema de Bayes como elemento a aprender (Peck, 2014; Brase y Brase, 2016). Por otra parte, existen trabajos didácticos que indican la existencia de sesgos en el razonamiento con la probabilidad condicional (Batanero, Ortiz y Serrano, 2016). Comprender la independencia y la probabilidad condicional es esencial para, en su caso, una correcta aplicación del teorema de Bayes en muchos procesos de modelización matemática. Sin embargo, desde un punto de vista psicológico y didáctico, estos conceptos presentan, además de dificultades de aprendizaje, problemas relacionados con su aplicación en la resolución de problemas y en la toma de decisiones (Díaz y de la Fuente, 2005). Por lo tanto, es importante encontrar dispositivos de enseñanza que puedan ayudar a los estudiantes a comprender mejor estos conceptos (Behar y Grima 2001, Díaz, Batanero y Contreras, 2010).

Las redes bayesianas constituyen una herramienta que permite modelar procesos caracterizados por la incertidumbre, lo cual es propio de infinidad de problemas reales. Uno de los primeros campos de aplicación de las redes bayesianas fue el de la medicina, más en concreto el diagnóstico de enfermedades (Leyva-Vázquez et al. 2013). En medicina no suelen haber síntomas que sean estrictamente propios de una enfermedad concreta, por lo tanto, no se puede afirmar que una persona tiene una enfermedad solamente con saber que se tiene un síntoma de esta. Si añadimos más información a la conocida, la complejidad del problema a manejar puede aumentar considerablemente, pero se puede alcanzar un mejor diagnóstico. Las redes bayesianas, permiten manejar dicha información a partir de relaciones entre origen y consecuencias de la enfermedad.

Las redes bayesianas constituyen también un campo de estudio muy importante en el área de la inteligencia artificial. Una red bayesiana proporciona un modelo completo sobre un conjunto de variables aleatorias y sus relaciones. Dicho modelo se puede utilizar para estimar probabilidades de ciertas variables de la red (variables de estado) cuando otras variables (variables de evidencia) son fijadas.

El objetivo de este trabajo es presentar una herramienta de modelización, las redes bayesianas, que ayuden a entender desde otra perspectiva el teorema de Bayes, y consecuentemente mejorar el significado de la probabilidad condicionada, la probabilidad conjunta o la noción de independencia.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

2. La probabilidad condicionada y el enfoque bayesiano de la probabilidad

Con el objetivo de comprender mejor el enfoque bayesiano de la probabilidad, planteamos el siguiente ejemplo.

Si lanzamos un dado de seis caras, aplicando la regla de Laplace la probabilidad de obtener un seis sería $1/6$. Formalmente el espacio muestral sería $E = \{1, 2, 3, 4, 5, 6\}$ y si A es el suceso “sacar un seis” tenemos que $P(A) = 1/6$, que se denomina, en el enfoque bayesiano, la probabilidad “a priori” del suceso A.

En cambio, si nos dicen que en el lanzamiento del dado se sabe que ha salido un número par, que, denotaremos como el suceso B, entonces la probabilidad de A condicionada a B sería $1/3$, ya que el espacio muestral se ha restringido a $E = \{2, 4, 6\}$. Dicha probabilidad se suele denotar como $P(A|B) = 1/3$. Es decir, la probabilidad “a priori” del suceso A ha aumentado ante la evidencia que supone conocer el suceso B. Esta nueva probabilidad $P(A|B)$ suele denominarse probabilidad “a posteriori” del suceso A, dada la evidencia B.

La probabilidad “a posteriori” o condicionada se puede obtener a partir de la igualdad descrita en la Ecuación 1:

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

que, aplicada al ejemplo del dado, lleva a otra manera de calcular la probabilidad “a posteriori”:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

Por otra parte, si despejamos la probabilidad de la intersección en la Ecuación 1, obtenemos

$$P(A \cap B) = P(A|B)P(B). \tag{2}$$

La probabilidad condicionada nos lleva, de forma natural, a introducir la noción de sucesos independientes: *Dos sucesos A y B son independientes si, y solo si,* $P(A|B) = P(A)$ y $P(B|A) = P(B)$. Atendiendo a la Ecuación 2, obtenemos una expresión para la caracterización de los sucesos independientes:

$$P(A \cap B) = P(A)P(B). \tag{3}$$

En determinados problemas de probabilidad, se suele suponer lo que se denomina hipótesis de independencia condicional: *Se dice que los sucesos X e Y son independientes dado el suceso Z si, y solo si,*

$$P\big((X, Y)|Z\big) = P\big((X \cap Y)|Z\big) = P(X|Z)P(Y|Z). \tag{4}$$

La probabilidad de la intersección de sucesos, $P(A \cap B)$ se denomina también “probabilidad conjunta” de los sucesos A y B, que también denotaremos como $P(A, B)$.

Puesto que los sucesos se comportan como conjuntos, se tiene que $A \cap B = B \cap A$ y, por lo tanto:

$$P(A \cap B) = P(B \cap A) = P(B|A)P(A). \tag{5}$$

A partir de las Ecuaciones 1 y 5 se deduce lo que se conoce como Teorema de Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{6}$$

El *Teorema de Bayes* se puede generalizar atendiendo a lo que se denomina una partición del espacio muestral E: *Un conjunto de sucesos* $\{A_1, A_2, \ldots, A_n\}$, *tales que*

- $A_i \cap A_j = \emptyset$ *para todo* $i, j \in \{1, 2, \ldots, n\}$ *siempre que* $i \neq j$, *y*
- $E = \bigcup_{i=1}^{n} A_i$,

se dice que es una partición de E. Entonces se tiene que:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)}. \tag{7}$$

Para obtener la expresión de $P(B)$ se ha tenido en cuenta que:

$$B = B \cap E = B \cap \left(\bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n} (B \cap A_i) = \bigcup_{i=1}^{n} (B, A_i).$$

Observemos que $\{B \cap A_1, B \cap A_2, \ldots, B \cap A_n\}$, cumplen que $(B \cap A_i) \cap (B \cap A_j) = \emptyset$ para todo $i, j \in \{1, 2, \ldots, n\}$ con $i \neq j$, con lo que:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B, A_i). \tag{8}$$

La expresión del *Teorema de Bayes* (Ecuaciones 6 y 7) permite identificar la relación existente entre la probabilidad a priori (esto es $P(A)$ ó $P(A_i)$) y la probabilidad a posteriori (es decir $P(A|B)$ ó $P(A_i|B)$), respectivamente.

Cuando se estudian los fundamentos de la teoría de la probabilidad, parece casi ineludible el apartado dedicado al Teorema de Bayes y sus aplicaciones. El orden seguido en los textos (Newbold, Carlson y Thorne, 2010) y en clase, muy a menudo es el siguiente:

1. Definición de la probabilidad condicionada y de los sucesos independientes.
2. Teorema de la probabilidad total
3. Teorema de Bayes
4. Problemas de aplicación del Teorema de Bayes

Los ejemplos de aplicación del *Teorema de Bayes* que se plantean en el aula suelen ser bastante característicos, y hasta recurrentes. Es típico el problema relacionado con la producción de piezas defectuosas en cierta fábrica.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Problema 1. *El 60 % de los tornillos producidos por una fábrica proceden de la máquina A y el 40 % de la máquina B. El porcentaje de tornillos defectuosos producidos por la máquina A es 1 % y en cambio por la máquina B es 5 %. ¿Cuál es la probabilidad de que, sabiendo que un tornillo es defectuoso, proceda de la máquina A?*

Solución clásica al problema 1 con el Teorema de Bayes.
Definimos los sucesos

$$D = \{\textit{ser un tornillo defectuoso}\}$$

$$A = \{\textit{ser producido por la máquina } A\} \qquad B = \{\textit{ser producido por la máquina } B\}.$$

El problema nos pide que hallemos $P(A|D)$ Dado que $\{A, B\}$ constituye una partición del espacio muestral podemos aplicar la Ecuación 7:

$$P(A|D) = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B)} = \frac{0.01 \cdot 0.6}{0.01 \cdot 0.6 + 0.05 \cdot 0.4} = \frac{0.006}{0.026} = 0.2308.$$

Veremos que este problema también se puede resolver modelándolo mediante una red bayesiana, lo cual puede ayudar a los estudiantes a enriquecer su esquema de probabilidad, proporcionándole, además, una potente herramienta para la modelización matemática de procesos reales.

3. ¿Qué es una red bayesina?

Para definir correctamente una red bayesiana necesitamos introducir unos cuantos conceptos de teoría de grafos:

- Un **nodo** es cualquier elemento de un conjunto dado V.
- Un **arco** es todo par ordenado (A, B) de nodos de V. Gráficamente estaríamos hablando de una flecha que une dos nodos $(A \to B)$.
- Un **grafo dirigido** es un conjunto de nodos conectados entre ellos por arcos, que indican un sentido obligatorio para la transición entre nodos. En la Figura 1 tenemos un grafo donde es posible la transición de A hacia B (no de B hacia A), de B hacia D y de A hacia C y de C hacia D.
- Un **camino** en un grafo es cualquier conjunto ordenado de nodos $\{N_1, N_2, \ldots, N_r\}$ tal que dados dos nodos consecutivos, N_j y N_{j+1}, o bien $N_j \to N_{j+1}$ o $N_{j+1} \to N_j$. Es decir, es cualquier recorrido que se puede realizar sobre el grafo, pasando de unos nodos a otros siempre que estén conectados mediante una flecha y sin tener en cuenta el sentido de esta. En el grafo de la Figura 1 un camino posible sería (D, B, A, C).
- Un **camino dirigido** es todo camino tal que dados dos nodos consecutivos N_j y N_{j+1}, se cumple necesariamente que $N_j \to N_{j+1}$. Es decir, un camino dirigido es todo recorrido sobre el grafo, en el cual se tiene en cuenta el sentido indicado por la flecha, con lo cual podemos pasar de un nodo N_j a un nodo N_{j+1} siempre y cuando haya una flecha que salga de N_j y llegue a N_{j+1}. En el grafo de la Figura 1, un camino dirigido sería (A, B, D) ya que $A \to B \to D$.
- Un grafo es **conexo** si, para cualquier par de nodos del grafo, existe un camino que los une.

- Un **ciclo** es un camino dirigido del grafo que empieza y termina en el mismo nodo.
- Un grafo es **acíclico** si no contiene ciclos.
- Un **nodo** Y **es padre o causa de un nodo** X si existe un arco (Y, X) en el grafo. El conjunto de padres de un nodo X, que denotaremos $P_a(X)$, estaría formado por todos los nodos de los cuales salen arcos que llegan directamente a X. En el caso del grafo representado en la Figura 1 se tiene que:

$$Pa(A) = \emptyset, \quad P_a(B) = A, \quad P_a(C) = A, \quad P_a(D) = \{B, C\}.$$

- Los **descendientes** de un nodo X del grafo, son los nodos a los que llegan arcos que salen directamente de X. Denotaremos dicho conjunto como, $D_s(X)$. En el grafo de la Figura 1 se tiene

$$D_s(A) = \{B, C\}, \quad D_s(B) = D, \quad D_s(C) = D.$$

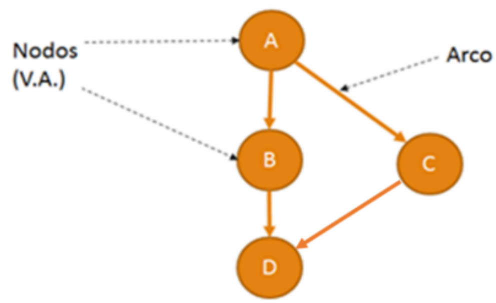

Figura 1: Ejemplo de grafo conexo acíclico.

Una red bayesiana consiste en

- Un conjunto de variables aleatorias, $V = \{X_1, X_2, \ldots, X_r\}$.
- Un conjunto de relaciones entre las variables aleatorias de V que se puede representar mediante un grafo conexo y acíclico, G, en el cual los nodos son las variables de V.
- Una distribución de probabilidad conjunta, $P(X_1, X_2, \ldots, X_r)$, sobre las variables de V.

En una red bayesiana, puesto que cada nodo es una variable aleatoria, los padres de un nodo X, constituirán también una variable aleatoria, en concreto la variable aleatoria conjunta (intersección) formada a partir de todos los nodos de los cuales sale un arco que llega directamente a X.

En la red bayesiana representada de la Figura 1, los nodos (A, B, C, D) representarían variables aleatorias y los arcos representan relaciones de dependencia probabilística directa entre las variables. Si un par de nodos están unidos mediante un arco, indica que la variable a la que apunta el arco es dependiente de la que está en el origen de éste. En la Figura 1 se representa una red bayesiana en la que "D depende de $B \cap C$", "B depende de A" y "C depende de A". La notación utilizada en este caso sería

$$P_a(A) = \emptyset, \quad P_a(B) = A, \quad P_a(C) = A, \quad P_a(D) = B \cap C = (B, C).$$

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

La estructura del grafo también informa sobre la independencia condicional de una variable (o conjunto de variables) dada otra variable (u otras variables). Por ejemplo, en la red de la Figura 1, el suceso "D dado $P_a(D)$" ("D dado $B \cap C$"), o escrito de otra forma, $D|(B, C)$), es condicionalmente independiente de A.

Si para cada nodo de la red (X) se conoce la distribución de $X|P_a(X)$, se puede deducir de forma inmediata la probabilidad conjunta de todos los nodos de la red. A partir de dicha probabilidad conjunta, se podrá deducir cualquier otra probabilidad que relacione las variables de la red. En el caso de que los nodos o variables de la red bayesiana sean $X_1, X_2, \ldots, X_n$, las relaciones de dependencia e independencia condicionada en la red permiten demostrar que la probabilidad conjunta (que equivale a la de la intersección) viene dada por:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i, P_a(X_i)). \tag{9}$$

Por lo tanto, podemos decir que tenemos información completa de la red bayesiana si conocemos la estructura del grafo y las distribuciones de $X|P_a(X)$ para cada nodo, X, de la red.

Por ejemplo, para la red bayesiana de la Figura 1, la aplicación de la Ecuación 9 nos lleva a:

$$\begin{aligned} P(A = A_1, B = B_1, C = C_1, D = D_1) &= P(A_1, B_1, C_1, D_1) \\ &= P(D_1|(B_1, C_1))P(C_1|A_1)P(B_1|A_1)P(A_1) \end{aligned} \tag{10}$$

Como se puede observar en la Ecuación 10, aplicar la igualdad dada por la Ecuación 9 es muy sencillo observando el grafo de la red bayesiana.

4. Resolver una red bayesiana

Los problemas de diagnóstico médico proporcionan contextos adecuados para ensayar la modelización mediante redes bayesianas. Al modelar un problema de diagnóstico médico simple mediante una red bayesiana es conveniente tener en cuenta las siguientes variables:

- Variable principal, X, "Tener la enfermedad", que, generalmente, tendrá dos modalidades, S (Si) y N (No).
- Variables de evidencia, E_i, $i = 1, 2, \ldots, r$, "la evidencia i apoya que se tenga (o que no se tenga) la enfermedad".

Para resolver la red en este caso, se procederá por pasos de la siguiente forma:

1. Fijamos la variable principal y la modalidad de interés para la variable principal (por ejemplo, variable principal X y modalidad de interés S, o de forma resumida, $X = S$).
2. Determinamos la probabilidad a priori del suceso de interés, $P(X = S) = P(S)$ ó $P(S|(E_{s_1}, E_{s_2}, \ldots, E_{s_k}))$ siendo $P_a(X) = (E_{s_1}, E_{s_2}, \ldots, E_{s_k})$.
3. Fijamos los sucesos considerados evidencias (descendientes o padres) del principal, $(\{E_1, E_2, \ldots, E_r\})$.
4. Calculamos la distribución conjunta, es decir, $P(A \cap E_1 \cap \ldots \cap E_r) = P(A, E_1, \ldots, E_r)$.
5. Calculamos la probabilidad a posteriori, es decir, $P(A|E_1 \cap \ldots \cap E_r) = P(A|E_1, \ldots, E_r)$.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Solución al problema 1 utilizando redes bayesianas.

Aunque el problema 1 no es de diagnóstico médico, veamos que también se puede resolver, sin referencia alguna al Teorema de Bayes, modelándolo mediante una red bayesiana, en concreto la red bayesiana más simple.

En el problema 1 sólo hay involucradas dos variables aleatorias: "Máquinas que producen tornillos" (M) y el "estado de un tornillo" (Et). Además, el que un tornillo elegido al azar esté en determinado estado, obviamente está condicionado por la máquina que lo haya podido producir. Este simple razonamiento nos lleva de forma natural a una red bayesiana cuyo gráfico se muestra en la Figura 2. Dicha red es, obviamente, la red bayesiana de estructura más simple (dos nodos y un arco que los une).

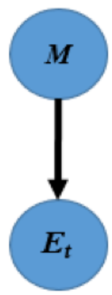

Figura 2: Grafo de la red bayesiana asociada al problema 1.

La estructura que describe el grafo de la Figura 2 se establece sobre la idea de que "según la máquina que fabrique un tornillo (causa), ésta tendrá más o menos posibilidades de ser defectuoso (efecto)". Los nodos en la red son, por una parte, M, con los estados posibles: máquina A y máquina B, y Et, con las modalidades posibles, defectuoso (D) y no defectuoso (ND). El enunciado del problema permite completar la red bayesiana mediante las siguientes probabilidades:

- $M \in \{A, B\}$,
  $P(M = A) = P(A) = 0.6, P(M = B) = P(B) = 0.4,$
- $Et/M \in \{D|A, D|B, ND|A, ND|B\}$
  $P(D|A) = 0.01, P(ND|A) = 0.99$ *y* $P(D|B) = 0.05, P(ND|B) = 0.95.$

Pasemos a resolver el problema:

1. *La variable principal sería M y el suceso de interés sería A*
2. *La probabilidad a priori del suceso de interés sería $P(A) = 0.6$*
3. *La evidencia sería el estado, Et, en concreto, $Et = D$*
4. *La distribución conjunta se describe en la Tabla 1.*

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Interacción	Población A	Población B
	D	**ND**
A	$P(A,D) = P(D\|A)P(A) =$ $= 0.01 \cdot 0.6 = 0.006$	$P(A,ND) = P(ND\|A)P(A) =$ $= 0.99 \cdot 0.6 = 0.594$
B	$P(B,D) = P(D\|B)P(B) =$ $= 0.05 \cdot 0.4 = 0.02$	$P(B,ND) = P(ND\|B)P(B) =$ $= 0.95 \cdot 0.4 = 0.38$

Tabla 1: Probabilidad conjunta de las variables M y Et.

La probabilidad a posteriori sería $P(A|D)$ que se obtiene a partir de la expresión

$$P(A|D) = \frac{P(A,D)}{P(D)} = \frac{0.006}{0.006 + 0.02} = \frac{0.006}{0.026} = 0.2308,$$

puesto que $P(D) = P(A,D) + P(B,D)$.

Como hemos visto anteriormente, la solución clásica se fundamenta en el conocimiento (memorístico) del Teorema de Bayes y del Teorema de la probabilidad total, mientras que la solución basada en la red bayesiana involucra un proceso de modelización que se cimienta en construir y resolver la red bayesiana y se concreta en obtener la probabilidad conjunta.

5. Modelando con redes bayesianas problemas de diagnóstico médico

5.1. Redes bayesianas simples

Veamos un ejemplo con la red bayesiana más simple. El problema 2 es una adaptación del que aparece en Morales Medina (2013), que a su vez está tomado del libro *El hombre anumérico*, de John Allen Paulos (Tusquets Editores, 1990).

Problema 2. *Supongamos que un paciente acude a una consulta médica porque cree padecer cierta enfermedad. El facultativo, durante el proceso del diagnóstico, le aplica un test. Supongamos que el test diagnóstico es fiable al 95 %, lo cual significa que el test da positivo el 95 % de las veces que se aplica a alguien enfermo (sensibilidad del test) y que da negativo el 98 % de las veces que se lo aplicamos a alguien sano (especificidad del test). Supongamos también que el porcentaje de personas que padecen la enfermedad en realidad es el 0.5 %, es decir, cinco de cada mil personas tienen realmente dicha enfermedad (prevalencia de la enfermedad). Tras aplicar el test al paciente, el resultado es positivo. ¿Cuál es la probabilidad de que dicho paciente padezca realmente la enfermedad?*

El problema 2 es muy parecido en su formalización al problema 1. En dicho problema sólo se consideran dos variables aleatorias:

$$E = \text{“tener o no la enfermedad”}$$

$$T = \text{“resultado al aplicar el test diagnóstico”}.$$

El hecho de que el resultado del test dependa directamente de si éste se aplica a una persona sana o enferma nos lleva, de forma natural, a una red bayesiana cuyo gráfico se muestra en la Figura 3.

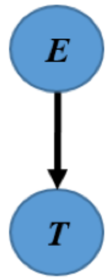

Figura 3: Grafo de la red bayesiana asociada al problema 2.

La variable E (Estado) tiene dos posibles modalidades:

$$S = \{\text{estar enfermo}\} \qquad N = \{\text{no estar enfermo}\},$$

con $P(S) = 0.005$ ($P(N) = 0.995$). Igualmente, la variable T también admite dos modalidades:

$$+ = \{\text{dar positivo}\} \qquad - = \{\text{dar negativo}\}.$$

De estos sucesos conocemos las probabilidades que completan la información que caracteriza la red: $P(+|S) = 0.95$, (consecuentemente tenemos que $P(-|S) = 1 - 0.95 = 0.05$) y, además, $P(-|N) = 0.98$ (por tanto $P(+|N) = 1 - 0.98 = 0.02$). Las probabilidades anteriores nos permiten averiguar las probabilidades conjuntas de las dos variables aplicando la Ecuación 7 (véase Tabla 2):

	$+$	$-$
S	$P(S,+) = P(+\|S)P(S) =$ $0.95 \cdot 0.005 = 0.00475$	$P(S,-) = P(-\|S)P(S) =$ $= 0.05 \cdot 0.005 = 0.00025$
S	$P(N,+) = P(+\|N)P(N) =$ $= 0.02 \cdot 0.995 = 0.0199$	$P(N,-) = P(-\|N)P(N) =$ $= 0.98 \cdot 0.995 = 0.9751$

Tabla 2: Probabilidad conjunta de las variables del problema 2.

En el problema se nos pide $P(S|+)$. Pasemos a resolverlo:

1. La variable principal sería E y el suceso de interés sería $E = S$
2. La probabilidad a priori del suceso de interés sería $P(S) = 0.005$
3. La evidencia sería el resultado del test, T, en nuestro caso $T = +$
4. La distribución conjunta se describe en la Tabla 2.
5. La probabilidad a posteriori sería $P(S|+)$ que se obtiene a partir de la expresión

$$P(S|+) = \frac{P(S,+)}{P(+)},$$

puesto que, $\{S, N\}$ constituyen una partición del espacio muestral, E, se tiene que

$$P(+) = P(S,+) + P(N,+) = 0.00475 + 0.0199 = 0.02465,$$

con lo que

$$P(S|+) = \frac{P(S,+)}{P(+)} = \frac{0.00475}{0.02465} = 0.1927.$$

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Como era de esperar, la probabilidad de que el paciente padezca le enfermedad ha pasado del 0.5 % (probabilidad a priori) al 19 % (probabilidad a posteriori).

También podríamos preguntarnos qué pasaría si el test hubiera dado negativo. La red bayesiana sigue siendo la misma, pero en su resolución hay ciertas variaciones.

1. La variable principal sería E y el suceso de interés sigue siendo $E = S$
2. La probabilidad a priori del suceso de interés seguiría siendo $P(S) = 0.005$
3. La evidencia ahora sería $T = -$
4. La distribución conjunta se describe en la Tabla 2
5. La probabilidad a posteriori sería $P(S|-)$ que se obtendría de:

$$P(S|-) = \frac{P(S,-)}{P(-)}.$$

Puesto que, $P(-) = P(S,-) + P(N,-) = 0.00025 + 0.9751 = 0.97535$, entonces

$$P(S|-) = \frac{P(S,-)}{P(-)} = \frac{0.00025}{0.97535} = 0.000256.$$

En este caso la probabilidad de tener la enfermedad se reduce significativamente respecto de la probabilidad a priori. Se ha pasado del 0.5 % al 0.0256 %, es decir una reducción hasta la veinteava parte, aproximadamente.

5.2. Redes bayesianas con tres nodos en secuencia

Añadir nueva información, haciendo más complejo el modelo, puede abordarse de forma bastante simple completando la red bayesiana. Veamos el problema 3.

Problema 3. *Supongamos las condiciones del problema 2 pero añadiendo información nueva: La enfermedad pueden transmitirla las mascotas. El paciente no tiene mascota, pero un amigo de este tiene un gato y el paciente ha estado en contacto con el animal. Se ha encontrado en artículos científicos que, en la población de personas en contacto con mascotas, la prevalencia de la enfermedad es 0.03.*

En este problema aparece una nueva variable o nodo (M) corresponde a la variable "Contacto previo con mascotas", la cual no tiene por qué ser aleatoria. De hecho, en este caso no lo es. Dicha variable tendría dos modalidades: "Si Contacto" (C) y "No Contacto" (NC). El valor de M influirá, obviamente, en la distribución de la variable E. Vamos a ensayar una nueva red bayesiana (Figura 4).

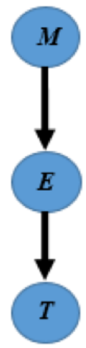

Figura 4: Grafo de la red bayesiana asociada al problema 3.

La resolución de esta nueva red es análoga a la descrita para el problema 2 ya que lo único que vamos a modificar será la probabilidad a priori del suceso principal. Ahora, la $P(S)$ no sería $0,005$ sino $P(S) = 0.03$ y, consecuentemente, la tabla de las probabilidades conjuntas (Tabla 3) variará respecto a la Tabla 2.

	$+$	$-$
S	$P(S,+) = P(+\|S)P(S) =$ $0.95 \cdot 0.03 = 0.0285$	$P(S,-) = P(-\|S)P(S) =$ $= 0.05 \cdot 0.03 = 0.0015$
S	$P(N,+) = P(+\|N)P(N) =$ $= 0.02 \cdot 0.97 = 0.0194$	$P(N,-) = P(-\|N)P(N) =$ $= 0.98 \cdot 0.97 = 0.9506$

Tabla 3: Probabilidad conjunta de las variables del problema 3.

Pasemos a resolver la red bayesiana:

1. La variable principal sería E y el suceso de interés sigue siendo $S = E$
2. La probabilidad a priori del suceso de interés es $P(S) = 0.03$
3. La evidencia sería el suceso $T = +$
4. La probabilidad conjunta se describe en la Tabla 3
5. La probabilidad a posteriori sería $P(S|+)$ que se obtiene a partir de

$$P(S|+) = \frac{P(S,+)}{P(+)}.$$

Puesto que, $P(+) = P(S,+) + P(N,+) = 0.0285 + 0.0194 = 0.0479$, entonces

$$P(S|+) = \frac{P(S,+)}{P(+)} = \frac{0.0285}{0.0479} = 0.595.$$

Con las dos evidencias, se observa un incremento notable de la probabilidad a posteriori de estar enfermo, que alcanza casi el 60 %. Con solo una evidencia (la que ofrece el test) la probabilidad de estar enfermo era aproximadamente del 19 %.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

5.3. Redes bayesianas con tres nodos divergentes

Estudiemos una nueva variación sobre la red inicial (Figura 5) a la que se le añade una nueva variable asociada al síntoma "tener fiebre".

Problema 4. *Supongamos las condiciones del problema 3 pero añadimos una nueva información: El paciente presenta un nuevo síntoma, tener fiebre que representaremos como St. La nueva variable admite una especificidad del 96 % y una sensibilidad del 99 %. ¿Qué probabilidad tiene el paciente de padecer la enfermedad?*

En este problema aparece un nuevo nodo (S_t) que corresponde a la variable "Tener fiebre" Dicha variable tendría dos modalidades: "Tiene fiebre" (F) y "No tiene fiebre" (NF). Además, conocemos las siguientes probabilidades: $P(F|S) = 0.99$ y por tanto $P(NF|S) = 0.01$. También sabemos que $P(NF|N) = 0.96$, y consecuentemente $P(F/N) = 0.04$.

En esta nueva red (Figura 5), se tiene que los nodos T y S_t son descendientes de E, ya que, evidentemente, el resultado del test y la presencia o no de fiebre vendrá motivado por el hecho de tener o no la enfermedad.

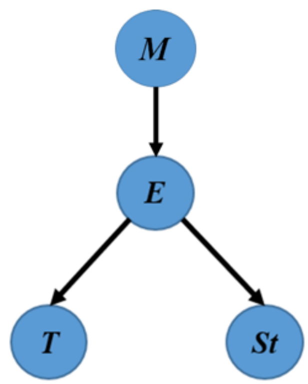

Figura 5: Grafo de la red bayesiana asociada al problema 4.

Para resolver esta nueva red bayesiana seguimos los pasos usuales:

1. La variable principal sería E y el suceso de interés sigue siendo $E = S$
2. La probabilidad a priori sería $P(S) = 0.03$
3. La evidencia sería el suceso $T = +$ en el test y tener fiebre, $S_t = F$, que denotamos $(+, F)$
4. La probabilidad conjunta sería la correspondiente a (E, T, S_t)
5. La probabilidad a posteriori de interés sería $P(S|+, F)$ que se obtiene de

$$P(S|+, F) = \frac{P(S, +, F)}{P(+, F)}.$$

Pasemos a realizar los cálculos:

$$P(S, +, F) = P(S)P(+|S)P(F|S) = 0.03 \cdot 0.95 \cdot 0.99 = 0.028215$$
$$P(N, +, F) = P(N)P(+|N)P(F|N) = 0.97 \cdot 0.02 \cdot 0.04 = 0.000776.$$

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

En estas últimas probabilidades, se ha tenido en cuenta que $P_a(T) = \{E\}$ y $P_a(S_t) = \{E\}$ y se ha aplicado la ecuación la Ecuación 10. Consecuentemente nos encontramos que

$$P(+, F) = P(S, +, F) + P(N, +, F) = 0.28215 + 0.000776 = 0.282926.$$

Por tanto, la probabilidad a posteriori será

$$P(S|+, F) = \frac{P(S, +, F)}{P(+, F)} = \frac{0.028215}{0.028991} \approx 0.97.$$

Como era de esperar, la nueva evidencia (presencia de fiebre) supone que la probabilidad de estar enfermo aumente (hasta más del 97 %) respecto a la que se obtenía considerando sólo la evidencia del resultado del test (19 %).

5.4. Clasificadores bayesianos

Desde un punto de vista formal, un proceso de clasificación consistirá en asignar una clase c_i, perteneciente a un conjunto de clases $C = \{c_1, \ldots, c_k\}$ a cierto objeto, representado por un vector de características o atributos $X_j = (x_{j1}, \ldots, x_{jm})$, para $j \in \{1, 2, \ldots, n\}$ y $k \leq j$. Hay dos tipos de clasificadores:

- **No supervisado o de agrupamiento**: En este caso las clases de C son desconocidas, y el problema consiste en dividir un conjunto de n objetos, X_j, en k clases, de forma que a objetos similares le será asignada una misma clase. Éste sería el caso del análisis de conglomerados o análisis cluster.
- **Supervisado**: En este caso, las k clases de C se conocen de antemano, y la dificultad consiste en encontrar una función que asigne cada objeto a su clase correspondiente.

En el caso de la clasificación supervisada, el problema de clasificación consiste en encontrar una función, f, sobre el conjunto de características o atributos, A_t, e imagen en el conjunto de clases, C, de forma que, para cada elemento $X_j \in A_t$, $f(X_j) = c_i$.

Desde un enfoque bayesiano, el problema de la clasificación supervisada consiste en asignar un objeto descrito por un conjunto de atributos o características, $X_j = (x_{j1}, \ldots, x_{jm})$, a una de las clases posibles, $c_1, c_2, \ldots, c_k$, de C, de forma que la probabilidad de pertenecer a la clase, dados los atributos, sea máxima, es decir:

$$f(X_j) = c_i \text{ si } \max\Big(P(C|X_1, X_2, ..., X_m)\Big) = P(c_i|X_1, X_2, \ldots, X_m).$$

El clasificador bayesiano simple (CBS o *Naive Bayes*) se basa en la suposición de que todos los atributos son independientes dada la clase, es decir, cada atributo X_i es condicionalmente independiente de los demás atributos dada la clase: $P(X_i|X_j, C) = P(X_i|C)$, para cualquier j, con $j \neq i$. Este planteamiento nos lleva, de manera natural a una red bayesiana con la estructura que se muestra en la Figura 6:

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

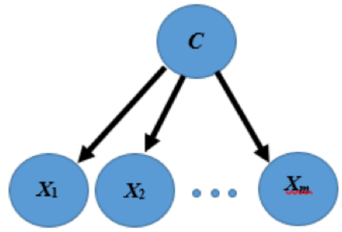

Figura 6: Grafo de un clasificador bayesiano simple.

Resolveremos el problema de clasificación obteniendo la distribución $P(C|X_1, X_2, \ldots, X_m)$, de forma análoga a como se ha realizado en los problemas 2, 3 y 4. En concreto el problema 3 se resolvería con un clasificador bayesiano simple en base a la Figura 7:

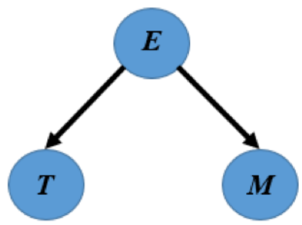

Figura 7: Grafo del problema 3 siguiendo un clasificador bayesiano.

Al formular el grafo como en la Figura 7 se ha perdido en gran medida la intuición de que los arcos se establecen siguiendo una lógica causa-efecto. En este caso M (contacto con mascotas) y T (resultado del test) pasan a ser evidencias que permiten establecer la clase más probable (tener la enfermedad o no tener la enfermedad).

6. Conclusiones

Los problemas 2, 3 y 4 son casos sencillos, en el contexto de diagnóstico en medicina, de un problema más general que podemos esquematizar mediante el grafo de la Figura 8, donde $O \in \{O_1, O_2, \ldots O_n\}$ son posibles orígenes de la enfermedad (*diagnóstico sindrómico*), $S \in \{S_1, S_2, \ldots, S_k\}$ son síntomas de la enfermedad y $P \in \{P_1, P_2, \ldots, P_m\}$ son resultados de pruebas. La Figura 8 muestra una estructura estándar de entre las posibles, que obviamente son muchas.

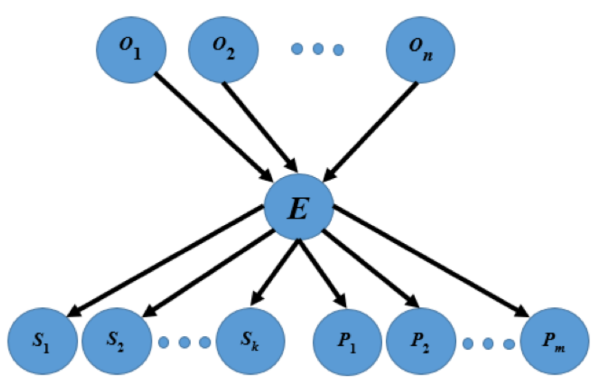

Figura 8: Grafo usual para red bayesiana asociada a un problema general de diagnóstico en medicina.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

En el problema general descrito en la Figura 8, una de las distribuciones que definen la red sería $P(E|O_1, O_2, \ldots, O_n)$. Desde la perspectiva de la formulación de la red, el número de orígenes de la enfermedad, n, debe ser relativamente pequeño para no complicar la determinación de dicha distribución. Si n es grande, puede ser preferible tratar el problema desde el enfoque de los clasificadores bayesianos.

Es inmediato el hecho de que los cálculos para resolver redes bayesianas se van complicando a medida que se incrementa el número de arcos y nodos. Para casos más complejos, se puede recurrir a software específico que ayuda a plantear y resolver la red. Utilizar dicho software puede ser interesante también en docencia ya que esto permitirá centrar la enseñanza en la modelización, dejando los cálculos tediosos a la máquina.

Uno de los programas más intuitivos para trabajar con redes bayesianas es Netica (Norsys Software Corp., 2018). Para trabajar con dicho programa, lo primero es introducir el grafo (nodos y arcos). Posteriormente, para cada nodo, se establecen las modalidades y las probabilidades condicionadas a los padres. El siguiente paso es compilar los datos. Tras la compilación, se llega a un grafo en cuyos nodos se tienen las distribuciones individuales para cada variable aleatoria. La Figura 9 reproduce una captura de pantalla tras la compilación para el problema 4, utilizando la versión v6.04 del programa Netica.

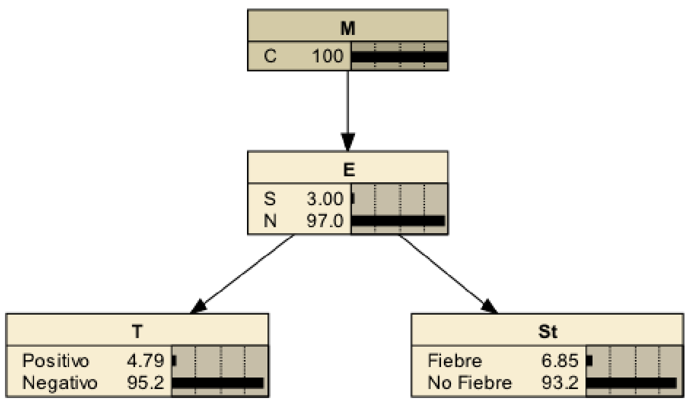

Figura 9: Red bayesiana y probabilidades individuales tras la compilación para el problema 4 obtenidas con el programa Netica.

Observemos, en la Figura 9, que se tiene que

$$P(+) = 0.0479, P(-) = 0.952, P(F) = 0.0685, P(NF) = 0.932.$$

Para obtener las probabilidades condicionadas, se establece la probabilidad 1 para las evidencias correspondientes. En Netica esto se hace con el ratón, corriendo hasta el 100 la barra de la modalidad elegida, en el nodo correspondiente a cada evidencia. De esta forma, en el nodo de la variable principal, aparecerá la distribución a posteriori de dicha variable (ver Figura 10). A la vista del resultado se tiene que $P(S|+, F) = 0.973$ y $P(N|+, F) = 0.0268$.

Las redes bayesianas aplicadas a problemas de diagnóstico médico, entendidas como herramientas de aprendizaje, permiten afrontar la resolución de problemas de probabilidad condicionada y probabilidad conjunta desde la modelización gráfica. Los problemas clásicos de aplicación del Teorema de Bayes pueden ser resueltos, de forma sencilla, sin tener tan siquiera que recurrir explícitamente a enunciar y aplicar dicho teorema. Pero las redes bayesianas permiten ir más lejos, abriendo innumerables posibilidades a la creatividad de los alumnos, que pueden trabajar sobre una red inicial afrontando un mismo problema básico desde diversos

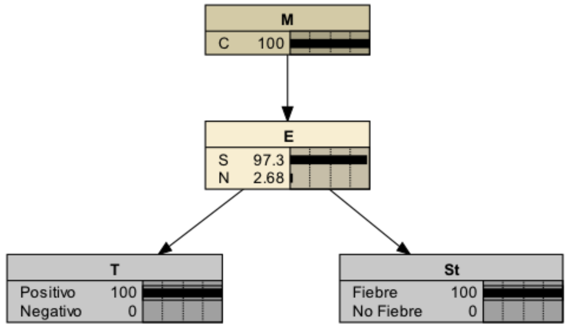

Figura 10: Red bayesiana y probabilidades a posteriori para el problema 4, obtenidas con el programa NETICA, tras establecer como evidencias $T = +$ y $S_t = F$.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

enfoques o incorporando al problema otros elementos que acerquen la situación estudiada a un contexto más real.

Referencias

Batanero, C., Ortiz, J.J., Serrano, L., Albanese, V. (2016).
Razonamiento sobre probabilidad condicional en situaciones de riesgo.
Suma, 83, pp 73–80.
http://www.ugr.es/~batanero/documentos/SUMA-Riesgo.pdf

Behar, R., Grima, P. (2001).
Mil y una dimensiones del aprendizaje de la estadística.
Estadística Española, 43(148), 189–207.

Boigues, F.J. y Estruch, V.D. (2017).
Aproximación frecuencialista de la Probabilidad.
Colección artículo docente de la UPV.
http://hdl.handle.net/10251/82991

Brase C.H. y Brase C.P. (2016).
Understanding Basic Statistics.
7ª edición, Metric version. Ed. Cengage Learning.

Díaz, C., Batanero, C., Contreras, J.M. (2010).
Teaching independence and conditional probability.
Boletín de Estadística e Investigación Operativa, nº. 26 (2), 149–162.

Díaz, C., de la Fuente, I. (2005).
Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística.
Epsilon, 59, 245–260.
http://www.ugr.es/~batanero/pages/ARTICULOS/epsilon_condicional.pdf
https://www.researchgate.net/publication/265369926_Teaching_independence_and_conditional_probability

Leyva-Vázquez, M., Pérez-Teruel, K., Febles Estrada, A., Gulín-González, J. (2013).
Técnicas para la representación del conocimiento causal: un estudio de caso en Informática Médica.
Revista Cubana de información en ciencias de la salud, 24(1), pp 73–83.
http://scielo.sld.cu/pdf/ics/v24n1/ics06113.pdf

Morales Medina, M.A. (2013).
Bayes y las pruebas de detección de enfermedades.
Blog Gausianos
www.gaussianos.com/bayes-y-las-pruebas-de-deteccion-de-enfermedades/

Newbold, P., Carlson, W.L., Thorne, B. M (2010).
Statistics for business and economics. Upper Saddle River etc.
Pearson Education International. 7th ed

Norsys Software Corp. (2018).
Netica v.6.04 64 bit for Ms Windows 7 to 10.
https://www.norsys.com/

Peck, R. (2014).
Statistics: Learning from Data. Preliminary Edition.
Ed. Cengage Learning.

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL