Revista
Educación
y Desarrollo Social

HERMILDA SUSANA RONDÓN TRONCOSO*
LUIS ALEJANDRO LADINO GASPAR**
PATRICIA ORDUZ CAMACHO***

ARTÍCULO DE INVESTIGACIÓN
FECHA DE RECEPCIÓN: 17 DE SEPTIEMBRE DE 2014
FECHA DE EVALUACIÓN: 9 DE OCTUBRE DE 2014

ACERCA DE LA ENSEÑANZA DEL TEOREMA DE BAYES

About teaching Bayes Theorem

Sobre a ensinança do teorema de Bayes

* Licenciada en matemáticas y física Universidad del Tolima. Especialista en estadística (Universidad Nacional de Colombia), Maestría en Tecnología Educativa Tecnológico de Monterrey (México). Profesor Asistente. Coordinadora de Probabilidad. Correo electrónico: Susana.rondon@escuelaing.edu.co

** Físico de la Universidad Nacional de Colombia. Maestría en Física Universidad Nacional de Colombia. Doctor en Física Universidad de Kentucky (USA). Profesor Asociado. Coordinador de Laboratorios de Física en Escuela Colombiana de Ingeniería Julio Garavito. Correo electrónico: luis.ladino@escuelaing.edu.co

*** Ingeniera Electricista de la Escuela Colombia de Ingeniería Julio Garavito. Especialista en distribución de energía de la Universidad Nacional de Colombia. Profesor Asistente, Coordinadora de Laboratorios de Física en Escuela Colombiana de Ingeniería Julio Garavito Correo: patricia.orduz@escuelaing.edu.co

ACERCA DE LA ENSEÑANZA DEL TEOREMA DE BAYES

Hermilda Susana Rondón Troncoso
Luis Alejandro Ladino Gaspar
Patricia Orduz Camacho

Referencia: Rondón, H.S., Ladino, L.A. y Orduz, P. (2015). Acerca de la enseñanza del teorema de Bayes. *Revista Educación y Desarrollo Social*. 9(1), 144-159.

RESUMEN

En la mayoría de los textos de probabilidad para estudiantes de pregrado, los ejercicios sobre el Teorema de Bayes se resuelven aplicándolo directamente. Esta forma árida de presentar el tema lo hace difícil de comprender y poco atractivo y, en lugar de captar el interés de los estudiantes, se logra una desmotivación. Como consecuencia no habrá preocupación por conocer más de lo que puede ofrecer este teorema en la solución de problemas de la vida real. El objetivo de este artículo es presentar una forma alternativa y efectiva para introducir el teorema de Bayes. Para ello, se comienza usando los diagramas de árbol, que generan de manera natural el teorema de Bayes. Adicionalmente, mediante el uso de las tablas de doble entrada el estudiante puede verificar inmediatamente la respuesta al problema, por cuanto se generan los mismos resultados. A modo de ilustración de la metodología propuesta, se soluciona el problema de determinar la probabilidad de que una persona contagiada del virus del SIDA lo haya adquirido de una de las siguientes tres posibles formas: transmisión sexual, transfusiones de sangre y uso de jeringas en drogas, cuando se conocen las respectivas probabilidades de contagio. Finalmente, los resultados de esta metodología mostraron un gran impacto positivo en la respuesta de los estudiantes, por cuanto desarrollaron mayor habilidad, interés y motivación en el tema. Además, se logró introducir el primer paso hacia el aprendizaje de los árboles de decisión, que son una aplicación del Teorema de Bayes.

Palabras Clave: árboles de decisión, diagramas de árbol, tablas de doble entrada, Teorema de Bayes.

ABSTRACT

In most probability texts for undergraduates, exercises on Bayes theorem are solved by applying directly. This arid way of presenting the subject makes unatractive and difficult to understand, and instead of capturing the interest of the students, a lack of motivation is the result. As a result there will be no interest on knowing what this theorem can offer in solving real life problems. The aim of this paper is to present an alternative and effective way to introduce Bayes theorem. To this end, we first use tree diagrams, which naturally generate Bayes theorem. Additionally, through the use of two-way tables the student can immediately verify the answer to the problem, since the same results are generated. To illustrate the proposed methodology, we solve the problem of determining the probability that a person infected with AIDS got the virus through one of three possible ways: sexual transmission, blood transfusions or needles in drug use, when the respective infection probabilities are known. Finally, the results of this methodology showed a positive impact on student response, because they developed greater ability, interest and motivation in the subject. Furthermore, it was possible to introduce the first step towards learning decision trees, which are an application of Bayes' Theorem.

Keywords: decision trees, tree diagrams, two-way contingency tables, Bayes' theorem

RESUMO

Na maioria dos textos de probabilidade para estudantes de pre-graduação, os exercícios sobre o Teorema de Bayes resolvem-se aplicando diretamente o Teorema; esta forma árida de apresentar o assunto torna-o mais difícil para a sua compreensão, sem atrativos, e em vez capturar o interesse dos alunos consegue é uma desmotivação e como consequência não haverá preocupação por saber além do que pode oferecer este teorema na solução de problemas da vida real. O objetivo deste artigo é apresentar uma forma alternativa e eficaz da introdução do Teorema de Bayes. Para fazer isso, começa-se usando os diagramas da árvore, que geram de maneira natural o Teorema de Bayes. Além disso, através do uso de tabelas de duas vias o aluno pode conferir imediatamente a resposta para o problema; por quanto se geram os mesmos resultados. Para ilustrar a metodologia proposta, soluciona-se o problema de determinar a probabilidade de que uma pessoa infectada com o vírus tenha sido adquirido a partir de uma das seguintes três possíveis formas: transmissão sexual, transfusões de sangue e o uso de seringas em drogas quando são conhecidas as respectivas probabilidades de contágio. Finalmente, os resultados desta metodologia mostraram um grande impacto positivo nas respostas dos alunos, porque eles desenvolveram maior capacidade, interesse e motivação no assunto. Além disso, foi possível introduzir o primeiro passo para a aprendizagem das árvores de decisão, que são uma aplicação do Teorema de Bayes.

Palavras-chave: Árvores de decisão, diagramas de árvores, tabelas de duas vias, Teorema de Bayes.

La clasificación bayesiana ha sido muy utilizada en diversos campos tales como sistemas (Sucar, L.E. 2003) en clasificación de información, en física (Mora, C. 2007) en procesos de imágenes, y en medicina en selección embrionaria (Morales, 2008) y en el campo del cáncer de mama (Quiñones, 2010), entre otros. Un tema preliminar a la clasificación bayesiana es el de los árboles de decisión que podría ser visto después del Teorema de Bayes.

Tradicionalmente los ejemplos resueltos sobre el Teorema de Bayes, que vienen en los libros de probabilidad y estadística son resueltos: por el Teorema como en los libros de Mendenhall (2010), Montgomery (2004) y Sheldon (2000); o por el Teorema y diagramas de árbol como los de en Devore (2012), Navidi (2006), Seymor (1991) y Walpole (2012); o por el Teorema y tablas de doble entrada como en el de Levin (1996); y por los tres métodos (Teorema, diagramas de árbol y tablas de doble entrada) como en los de Anderson (2008), Freund & Garay (1992), Lind, William Robert (2004), Nieves & Federico, (2010) y Kazmier & Díaz (1993).

Cualquiera que sea el método presentado tiene como desventaja que no presenta un ejercicio resuelto por los tres métodos que conduzca a dar respuesta a la misma pregunta; de igual forma no permite que los estudiantes establezcan comparaciones entre las metodologías empleadas, considerando que son estudiantes que hasta ahora están familiarizándose con el tema y que muchos de los ejercicios tiene cierto grado de dificultad.

A lo largo de los años de experiencia docente se ha podido observar que los estudiantes logran una mejor comprensión sobre el Teorema de Bayes cuando se inicia esta temática resolviendo los ejercicios mediante los diagramas de árbol. Luego sí conviene hacer uso del Teorema que ahora será simplemente cuestión de reemplazar los valores del diagrama del árbol y finalmente utilizar las tablas de doble entrada o de contingencia. Esta metodología facilita el aprendizaje para los estudiantes y les confirma si el ejercicio que hicieron está bien o no, al lograr la misma respuesta por cualquiera de los tres métodos y sin importar a qué programa de estudio pertenezca el estudiante.

El presente artículo plantea la posibilidad de resolver un ejercicio empleando la metodología propuesta en el párrafo anterior, con base en los buenos resultados que se han logrado desde el punto de vista didáctico y de motivación hacia el estudio de esta temática. Se intenta también ir más allá, al introducir el tema sobre árboles de decisión una vez visto el Teorema de Bayes. Los diagramas de árbol permiten dar el primer paso hacia el aprendizaje de los árboles de decisión. Además de ser una aplicación del Teorema de Bayes, estos últimos son la primera forma de clasificación bayesiana y son de gran aplicabilidad en diversas áreas del conocimiento.

EL TEOREMA DE BAYES

En su forma algebraica más simple, el Teorema de Bayes se refiere al cálculo de la probabilidad condicional del evento A dado que ha ocurrido el evento B. La forma general del Teorema de Bayes es (Walpole, 2012, p. 72-80):

Si $A_1, A_2, \ldots\ldots, A_n$ son eventos exhaustivos y exclusivos tales que $P(A_i) > 0$,

$\forall$ i = 1, 2,....n entonces para cualquier evento A tal que P(A) > 0, se cumple:

Dos eventos A y B son excluyentes si $P(A \cap B) = 0$, Una familia de eventos: $A_1, A_2, \ldots\ldots, A_n$ es exhaustiva si $\cup A_i = \Omega$

$$p(A_i / A) = \frac{P(A / A_i)P(A_i)}{\sum_{i=1}^{n} P(A / A_i)P(A_i)}$$

Caso General (1)

$$P(A / B) = \frac{P(A)P(B / A)}{P(A)P(B / A) + P(A')P(B / A)'}$$

Para dos eventos excluyentes. (2)

El Teorema de Bayes es válido en todas las aplicaciones que tengan que ver con teoría de la probabilidad. Sin embargo, existe una gran controversia sobre el tipo de probabilidades que emplea. Concretamente, los seguidores de la estadística tradicional únicamente admiten probabilidades basadas en experimentos repetibles y que tengan una confirmación empírica. Entre tanto, los seguidores de Bayes aceptan probabilidades subjetivas, según lo expresado por Silva & Muñoz (2007) en el sentido de que la teoría bayesiana está demostrando ser de gran utilidad en ciertas estimaciones basadas en el conocimiento subjetivo a priori. Además, el hecho de permitir hacer una revisión de esas estimaciones en función de la evidencia empírica es lo que está abriendo nuevas formas de hacer conocimiento. Una aplicación de esto son los clasificadores bayesianos que son frecuentemente usados en medicina (Silva & Muñoz, 2007).

A continuación se planteará y resolverá un problema inicial sobre el Teorema de Bayes, por tres métodos: En primer lugar, el diagrama de árbol, en segunda instancia aplicando el Teorema y finalmente por tabla de doble entrada o de contingencia.

PROBLEMA

De los enfermos de SIDA de un hospital, el 60% adquirió el virus por transmisión sexual, el 20% por transfusiones de sangre y el 20% por uso de jeringas en drogas. Entre los primeros, el 10% está en período terminal, entre los segundos el 20%, y entre los terceros el 5%. Se elige al azar un paciente de SIDA y resulta estar en estado terminal. ¿Cuál es la probabilidad de que haya adquirido el virus por transfusión de sangre?

SOLUCIÓN

Se plantearán los tres eventos simples del problema y el cuarto evento que es el condicional.

Los tres eventos simples son:

T_x: Transmisión sexual

T_s: Transfusión de sangre

T_j: Transmisión por jeringas con drogas

El evento condicional y su complemento serán:

P_t: Está en periodo terminal; NPT: No está en periodo terminal

Probabilidades marginales o a priori:

$$P(T_x) = 0.60\ ;\quad P(T_s) = 0.20\ ;\quad P(T_j) = 0.20;$$

$$\textstyle\sum_{i=1}^{n} P_i \ = \ \sum P(T_x) + P(T_s) + P\left(T_g\right) = 1 \quad (3)$$

$$Probabilidades\ condicionales \begin{Bmatrix} P(P_t/T_x) = \ 0.10 \\ P(P_t/T_s) = \ 0.20 \\ P(P_t/T_j) = \ 0.05 \end{Bmatrix}$$

SOLUCIÓN POR DIAGRAMA DE ÁRBOL

En la solución por diagrama de árbol se pueden apreciar las probabilidades marginales (a priori), las condicionales y las conjuntas. De esta manera el estudiante podrá identificarlas y utilizarlas en contexto en un problema de este estilo. La solución lograda de la Figura 1 es:

$$P(T_s/P_t) = \frac{(0{,}20)(0{,}20)}{(0{,}60)(0{,}10) + (0{,}20)(0{,}20) + (0{,}20)(0{,}05)}$$

$= 0{,}3636 = 36{,}36\ \%$

PROBABILIDADES MARGINALES O A PRIORI	PROBABILIDADES CONDICIONALES	PROBABILIDADES CONJUNTAS
$P(T_x) = 0.60$	$P(P_g/T_x) = 0.10$	$P(T_x \cap P_t) = 0.60 \times 0.10$
	$P(NPT) = 0.90$	
$P(T_s) = 0.20$	$P(P_t/T_s) = 0.20$	$P(T_s \cap P_t) = 0.20 \times 0.20$
	$P(NPT) = 0.80$	
$P(T_j) = 0.20$	$P(P_t/T_j) = 0.05$	$P(T_j \cap P_t) = 0.20 \times 0.05$
	$P(NPT) = 0.95$	

Figura 1. Solución por Diagrama de Árbol

> El Teorema de Bayes es válido en todas las aplicaciones que tengan que ver con teoría de la probabilidad. Sin embargo, existe una gran controversia sobre el tipo de probabilidades que emplea. Concretamente, los seguidores de la estadística tradicional únicamente admiten probabilidades basadas en experimentos repetibles y que tengan una confirmación empírica.

Una vez resuelto el problema por diagrama de árbol, se pasa a dar solución por el segundo método (aplicando el Teorema). Este paso será ahora más sencillo porque se pueden apreciar en el diagrama de árbol de la Figura 1 las probabilidades que son requeridas en la utilización del Teorema, para este segundo método.

SOLUCIÓN UTILIZANDO EL TEOREMA DE BAYES

La probabilidad pedida es: ¿Cuál es la probabilidad de que el paciente haya adquirido el virus por transfusión de sangre?, dado que la persona está en estado terminal. Ver Ecuación 4.

$$P(T_s/P_t) = \frac{P(T_s \cap P_t)}{P(P_t)} \tag{4}$$

Se puede apreciar que la probabilidad conjunta es reemplazada por las probabilidades marginales (a priori) y las probabilidades condicionales, que se encuentran en la Figura 1. Ver Ecuación 5.

$$P(T_s/P_t) = \frac{P(T_s)P(P_t/T_s)}{P(T_x)(P_t/T_x)+P(T_s)(P_t/T_s)+P(T_g)\,(P_t/T_g)} \tag{5}$$

En el numerador de la Ecuación 5, van las probabilidades del caso dos, es decir de los contagiados por el virus a través de transfusión sanguínea (el evento de interés); en el denominador deben considerarse todos los casos (los tres:

por transmisión sexual, transfusión de sangre y por uso de jeringas, estos tres eventos conforman omega, el espacio muestral).

Ahora, reemplazando los valores de todas las probabilidades indicadas en la Ecuación 5, se obtiene:

$$P(T_s/P_t) = \frac{(0,20)(0,20)}{(0,60)(0,10)+(0,20)(0,20)+(0,20)(0,05)}$$

$$= 0,3636 = 36,36\ \%$$

A continuación se presentará la solución del problema por el tercer método: Tabla de doble entrada o de contingencia.

SOLUCIÓN UTILIZANDO TABLA DE DOBLE ENTRADA O DE CONTINGENCIA

Recordemos que lo que se pide es la probabilidad de que el enfermo haya adquirido el virus por transfusión de sangre, dado que la persona está en estado terminal. De la Tabla 1, se tiene que:

$$P(T_s/P_t) = \frac{0.04}{0.11} = 0,3636 = 36,36\%$$

El valor 0.11 corresponde al total de personas que están en período terminal (la condición), lo que va en el denominador, luego se mira por esa fila (los que están en periodo terminal) y se ubica el valor 0.04 que son los casos que corresponden a los contagiados por transfusión de sangre, valor que se escribe en el numerador, que da como resultado 36.36 %.

Recordemos que lo que se pide es la probabilidad de que el enfermo haya adquirido el virus por transfusión de sangre, dado que la persona está en estado terminal.

Tabla 1: Solución del Problema por Tabla de Doble Entrada o de Contigencia

FORMAS DE CONTAGIO / FASE	Tx "TRANSMISIÓN SEXUAL"	Ts "TRANSMISIÓN DE SANGRE"	Ts "TRANSMISIÓN POR JERINGAS CON DROGAS"	TOTALES
PERÍODO TERMINAL	0.06 (0.60)(0.10)	0.04 (0.20)(0.20)	0.01 (0.20)(0.05)	0.11
NO PERÍODO TERMINAL	0.54 (0.60)(0.90)	0.16 (0.20)(0.80)	0.19 (0.20)(0.95)	0.89
TOTALES	0.60	0.20	0.20	1

CONCLUSIÓN FINAL

La probabilidad de haber adquirido el virus por transfusión de sangre, dado que el paciente está en periodo terminal, es del 36.36 %. Se puede apreciar que por cualquiera de los tres métodos la respuesta es la misma 36,36%.

Lo que se quiere hacer notar es que casi siempre en un curso regular de probabilidad, la solución que se da a un problema sobre el Teorema de Bayes es aplicando directamente la regla de Bayes (el Teorema) pero para el estudiante éste puede resultar el método menos sencillo de comprender. Se sugiere entonces empezar este tema utilizando los diagramas de árbol, para luego hacer ver de dónde sale la regla (el Teorema).

Se asume que para los estudiantes resulta más sencillo comprender el Teorema partiendo de los diagramas de árbol lo que hace más motivante su aprendizaje. Si además se utilizan las tablas de doble entrada o de contingencia, los estudiantes descubrirán que utilizando cualquiera de los métodos el resultado es el mismo y sabrán con certeza si lo que hicieron está bien o no al comparar las metodologías empleadas en la solución del problema. Se considera que esto puede ser enriquecedor desde el punto de vista didáctico.

El hecho de emplear las tablas de doble entrada o de contingencia les permite además darse cuenta de que lo aprendido en temas anteriores como probabilidades a priori o marginales, condicionales, conjuntas e independencia estadística, entre otros, en donde resolvieron ejercicios que hacían uso de estas tablas, también les es útil en esta nueva temática: el Teorema de Bayes. Bien es sabido que retomar temas vistos en otros apartados y usarlos permite que los estudiantes se apropien de los conceptos y adquieran dominio sobre éstos.

A continuación se describirá un aspecto relevante de aplicación del Teorema de Bayes como lo son los árboles de decisión, para el cual se mostrará un ejemplo mediante la utilización del programa WinQSB, en el que se presentarán los pantallazos logrados mediante la utilización de este software así como la interpretación de los resultados. Lo anterior con el propósito de motivar de alguna manera al docente para que pueda encontrarle utilidad a la temática de los árboles de decisión, por considerar que es una aplicación sobre el Teorema de Bayes y de gran relevancia por cuanto es muy utilizada en situaciones reales y en muchos campos del conocimiento como se menciónó en la introducción.

Se quiere mostrar mediante un ejemplo sencillo, y haciendo uso de software, que es posible lograr buenos resultados en el aula de clase, teniendo en cuenta que el estudiante de hoy en día es más amigo del uso de la computadora para resolver ejercicios la cual se ha convertido en su principal fuente de aprendizaje y estará más motivado hacia el estudio de una nueva temática si hace uso de la tecnología. Además, teniendo en cuenta que muchos estudiantes de las carreras de Ingeniería Industrial, Eléctrica y Administración se verán enfrentados con problemas de mayor complejidad sobre árboles de decisión, si ellos alcanzan a tener sus primeras bases en el curso de probabilidad, podrán abordar sin tantos inconvenientes las asignaturas en donde verán esta temática.

ARBOLES DE DECISIÓN

El árbol de decisión es un algoritmo de minería de datos de tipo clasificación. Éste conforma una estructura gráfica en la cual cada nodo representa un factor, los nodos van apareciendo en orden jerárquico y se derivan de un conjunto de datos; son fáciles de interpretar y permiten utilizar atributos de tipo discreto y continuo (Sánchez, Miranda & Cerda, 2004).

Los árboles de decisión permiten realizar análisis predictivos. Para ello nos apoyaremos en el programa WinQSB.

EJERCICIO SOBRE ÁRBOL DE DECISIÓN

Juan, un jugador de futbol sobresaliente, tiene dos posibilidades para seguir su carrera futbolística; una de ellas es continuar jugando en el Mónaco o ingresar al Real Madrid. De continuar en el Mónaco tiene una probabilidad del 75 % de recibir en promedio 14 mil euros diarios; si no continúa en el Mónaco y deseara volver algún día ganaría en promedio 6 mil euros diarios con una probabilidad del 25%. Juan se ha destacado como un buen jugador y tiene la posibilidad de ser admitido o no en el Real Madrid; si es admitido existen dos posibilidades: Una es ser un jugador activo, caso en el cual ganará en promedio 18 mil euros diarios y la otra es permanecer en el banco por si hay necesidad de un relevo y ganaría en promedio 10 mil euros diarios. Indique cuál es la mejor decisión que puede tomar Juan. Un ejercicio similar podrá ser encontrado en García (2012).

SOLUCIÓN

Para realizar el árbol Juan tiene que tomar una decisión (Nodo 1 de decisión o D) del que se desprenden otros dos nodos (los nodos 2 y 3 que son nodos probabilísticos o C de chance). El nodo 2 corresponde a la posibilidad de ingresar al Real Madrid y el nodo 3 representa seguir jugando en el Mónaco. Si Juan decide ingresar al Real Madrid (una de las alternativas), entonces tiene el nodo 4 que representa ser admitido en este equipo con una probabilidad del 75%, si es admitido, entonces Juan podría ser un jugador activo o permanecer en el banco (que representan los nodos 5 y 6, que son nodos terminales); el nodo 7 representa la otra posibilidad de no ser admitido en el Real Madrid. Los nodos finales 8 y 9 representan las dos posibilidades que tiene Juan de continuar o no continuar en el Mónaco; cada una de estas ramas tiene los valores esperados y las probabilidades en algunos casos se deben calcular a mano respetando algunos criterios propios de los árboles de decisión.

Teniendo toda esta información se procede a utilizar el software para finalmente lograr el árbol de decisión. Inicialmente este árbol de decisión se calcula a mano. En él se ubica toda la información que posteriormente se organizará en una matriz de datos, para finalmente ser ingresada en el programa, que para este ejemplo será WinQSB.

A continuación se pasa toda la anterior información en la matriz de datos en el siguiente orden: Los nodos (número de nodos que en este caso son 9), los nombres de estos nodos, el tipo de nodo, los nodos que son sucesores, los valores esperados y finalmente los valores de probabilidad. Ver Tabla 2. Algunos valores esperados se deben calcular de la siguiente manera:

Recordemos que los valores esperados se calculan teniendo en cuenta el siguiente algoritmo:

$$\mu(x) = E(x) = \sum_{i=1}^{n} V_i P_i \qquad (6)$$

donde V_i, corresponde a los valores de los nodos terminales y P_i, corresponde a los valores de probabilidad. Tener presente que la suma de todas las probabilidades debe ser uno, así:

$$\sum_{i=1}^{n} P_i = 1$$

El valor esperado para el nodo 4 es de 18 porque se toma el mismo valor esperado del nodo 5 que fue de 14, debido a que es el mayor valor esperado entre los nodos 5 y 6 (el del nodo 6 fue de 10, valores que dio el enunciado del problema).

18 x 0.75 + 4 x 0.25 = 14.5 (corresponde al valor esperado para el nodo 2)

14 x 0.75 + 6 x 0.25 = 12 (corresponde al valor esperado para el nodo 3).

Finalmente el valor esperado para el nodo inicial uno es también de 14,5. Por la misma razón anterior, fue el valor esperado mayor entre los valores esperados de los nodos 2 y 3 que fueron respectivamente 14.5 y 12. Ver Figura 2.

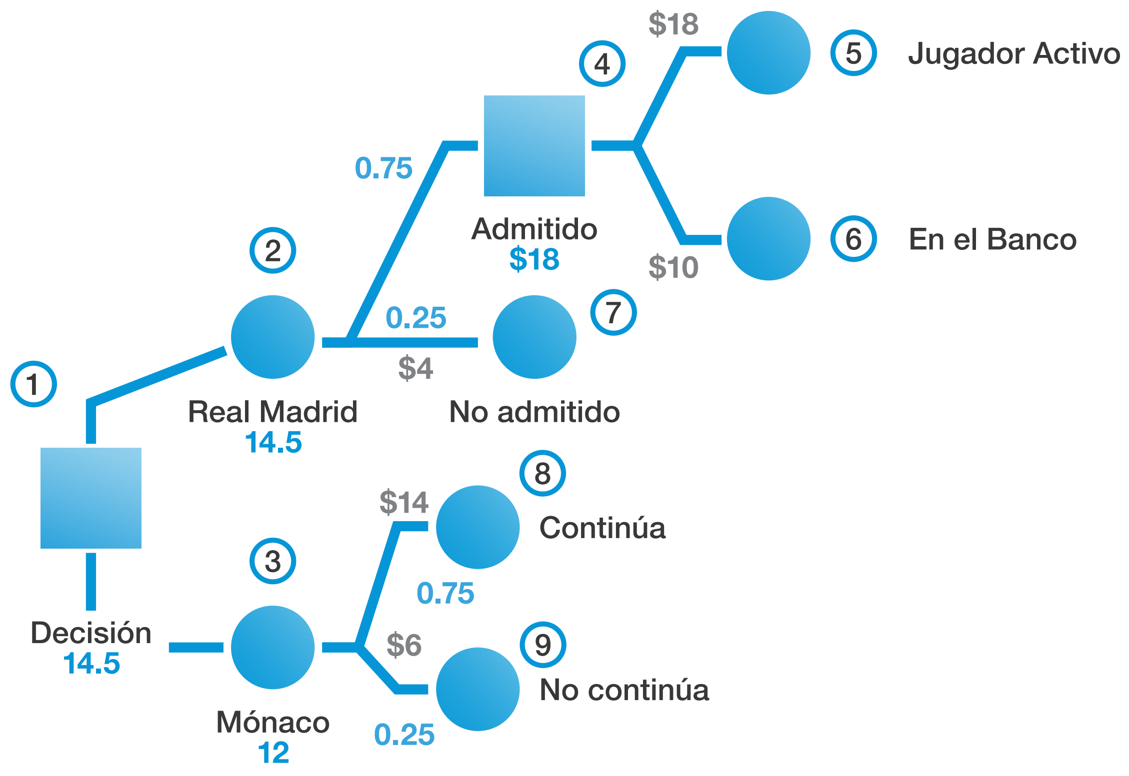


Figura 2. Árbol de Decisión Elaborado a Mano con los Valores

Tabla 2: Matriz con la Información Lista para ser Procesada en WinQSB

NODOS	NOMBRE	TIPO	SUCESORES	VALOR ESPERADO	PROBABILIDAD
1	Decisión	D	2,3		
2	Real Madrid	C	4,7		
3	Mónaco	C	8,9		
4	Admitido	D	5,6		0,75
5	Jugador Activo			18	
6	En el Banco			10	
7	No Admitido			4	0,25
8	Continua			14	0,75
9	No continua			6	0,25

A continuación se presenta la Tabla 2 con toda la información extraída del problema.

Ahora se podrá apreciar el ejemplo en WinQSB.

PROCEDIMIENTO

Se da inicio a WinQSB, se ubica el menú decisión Analysis y estando allí se hace clic en File, New Problem; se selecciona la opción Decision Tree Analysis, a continuación se da el título del ejercicio, que para este caso fue "El problema". Luego se marca el número de nodos (hay que recordar que el árbol tenía 9 nodos incluyendo los terminales) y luego se da clic en OK. Ver Figura 3.

A continuación aparece la matriz en donde se introduce la información que está en la Tabla 2. Es importante tener en cuenta que los nodos sucesores (columna 4) de la Figura 4 se escriben separados por comas, mientras que los valores de probabilidad (columna 6) de la misma figura se escriben con punto.

Al tener toda la información en la matriz, se da inicio a la solución. Hay que recordar hacer clic en el muñequito del recuadro rojo en donde aparece la información de la Figura 5.

La interpretación de la Figura 5 es la siguiente. La Figura dice que el valor esperado es de 14.50 miles de euros, que Juan debe aceptar ingresar en el Real Madrid, y lo mejor

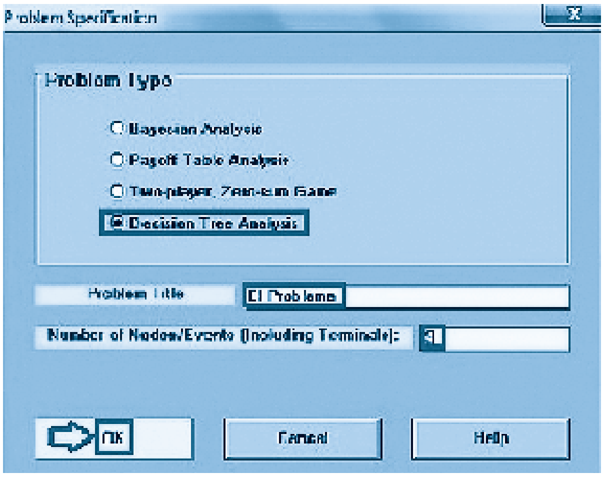


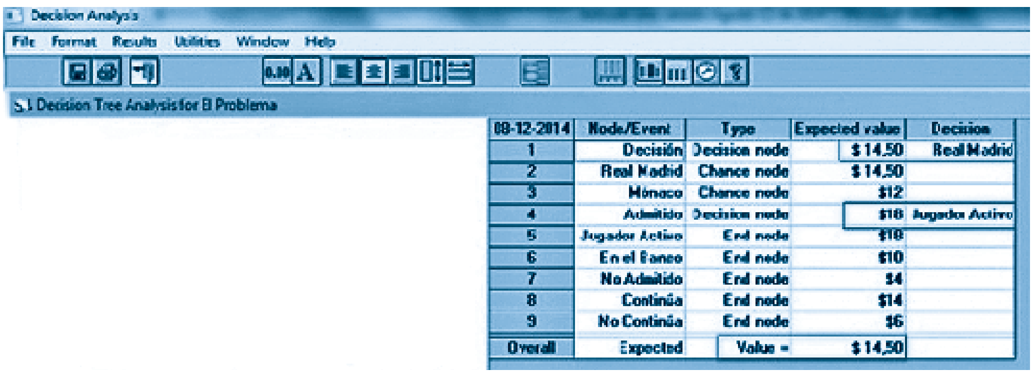
Figura 3. Pantallazo en WinQSB para el análisis de decisión

es tener la posibilidad de ser un jugador activo, el valor esperado es de 14.50 miles de euros que le pagará el Real Madrid.

En la Figura 6 se presenta la imagen en el cual se da la información que aparecen en los círculos rojos, para finalmente lograr el árbol de decisión. No olvidar dar clic en el árbol. Se agranda un poco el recuadro rojo para una mejor visualización, luego se seleccionan los valores esperados para cada nodo y que cada nodo tenga el número y el nombre. A continuación se podrá apreciar el árbol de decisión quedando la solución del problema propuesto (Figura 7).

Node/Event Number	Node Name or Description	Node Type (enter D or C)	Immediate Following Node (numbers separated by ',')	Node Payoff (+ profit, - cost)	Probability (if available)
1	Decisión	D	2,3		
2	Real Madrid	C	4,7		
3	Mónaco	C	8,9		
4	Admitido	D	5,6		0.75
5	Jugador Activo			10	
6	En el Banco			10	
7	No Admitido			4	0.25
8	Continúa			14	0.75
9	No Continúa			6	0.25

Figura 4. Pantallazo en WinQSB de la Matriz de Datos de la Tabla 2



09-12-2014	Node/Event	Type	Expected value	Decision
1	Decisión	Decision node	$14.50	Real Madrid
2	Real Madrid	Chance node	$14.50	
3	Mónaco	Chance node	$12	
4	Admitido	Decision node	$10	Jugador Activo
5	Jugador Activo	End node	$10	
6	En el Banco	End node	$10	
7	No Admitido	End node	$4	
8	Continúa	End node	$14	
9	No Continúa	End node	$6	
Overall	Expected	Value =	$14,50	

Figura 5. Pantallazo de Resultado del Análisis de Decisión Arrojado por el Programa WinQSB

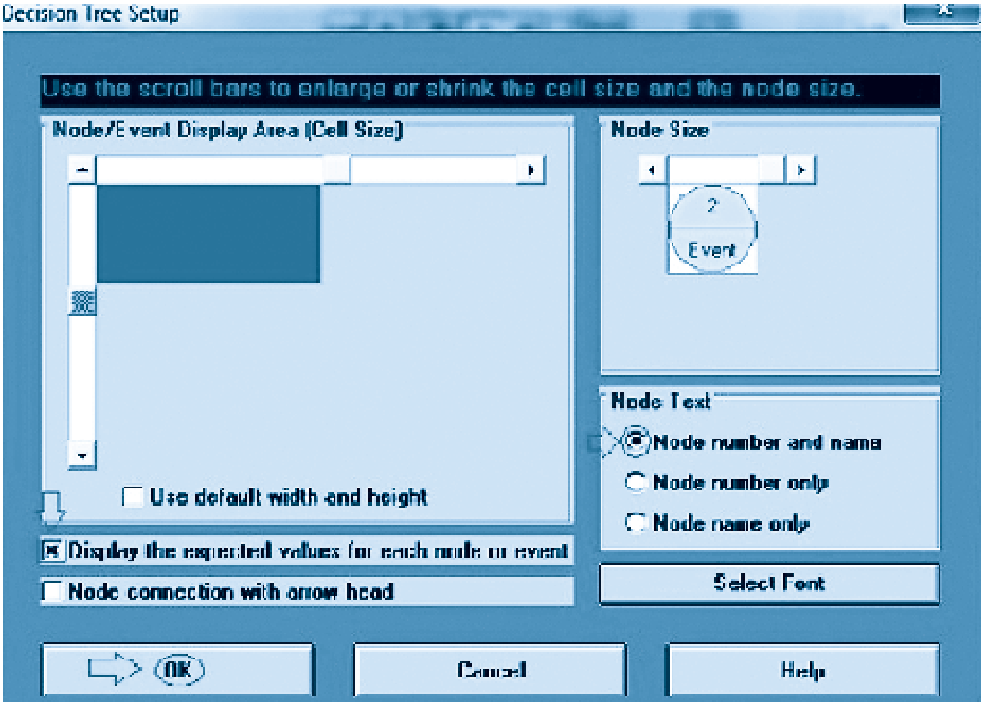


Figura 6. Pantallazo preliminar al árbol de decisión arrojado por WinQSB

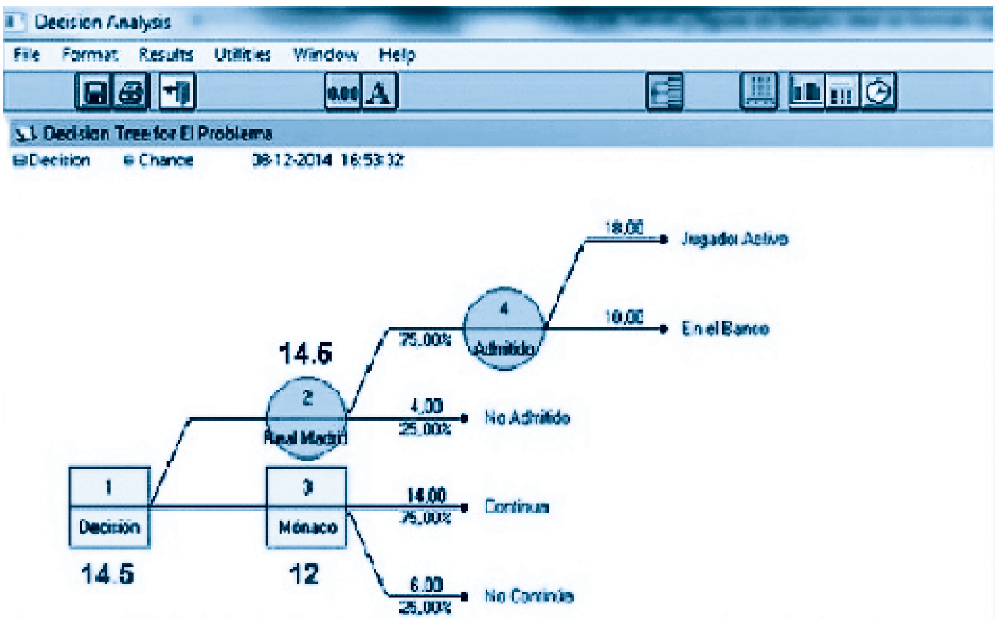


Figura 7. Pantallazo del Árbol de Decisión Logrado con el Software WinQSB

El árbol de decisión es un algoritmo de minería de datos de tipo clasificación. Éste conforma una estructura gráfica en la cual cada nodo representa un factor, los nodos van apareciendo en orden jerárquico y se derivan de un conjunto de datos.

CONCLUSIONES

La realización de problemas en el aula de clase suele ser más productivo si se presentan varios métodos de solución lo cual le permite al estudiante establecer comparaciones entre uno y otro método, generando discusión, que es lo que se busca finalmente cuando se enfatiza sobre algún tema; además se facilita el aprendizaje de los estudiantes.

Podría ser de gran utilidad que el estudiante conociera algunas de las aplicaciones que tiene el Teorema de Bayes en muchos campos del conocimiento, como lo son los árboles de decisión tema previo a la clasificación bayesiana.

Los cálculos matemáticos que resultan al elaborar los árboles de decisión pueden ser más sencillos si se hace uso de algún software, en este caso WinQSB que tiene la ventaja de ser de libre distribución, lo que permite ver de forma rápida y precisa los resultados que se desprenden de este tipo de análisis.

El trabajar un ejemplo que trate de algún tema que pueda llamar la atención del estudiante, y que además se apoye en el uso de alguna herramienta tecnológica, podría lograr una mayor motivación hacia el estudio de cualquier asignatura y en particular en probabilidad que, en la mayoría de ocasiones, se tiene como una asignatura poco atractiva por su denso contenido temático.

Siempre se discute sobre si los cursos de tipo teórico como lo es la asignatura de probabilidad no tienen nada que ver con el manejo de algún tipo de software. Se dice que es todo lo contrario, que el Teorema de Bayes va mucho más allá de lo que comúnmente se trabaja en clase, pero debido a las exigencias del currículo y la falta de tiempo resulta complicado seguir profundizando en este interesante tema.

REFERENCIAS

- Anderson, S. W. (2008). *Estadística para administración y economía.* CENGAGE Learning.
- Devore, J. L. (2012). *Probabilidad y estadística para ingeniería y ciencias.* CENGAGE Learnig.
- Freund, J. E & Simon, G. A. (1992). Estadística elemental. Prentice Hall.
- García. (2012). *Cómo realizar un árbol de decisión en WinQSB. [Video]*. Disponible en: https://www.youtube.com/watch?v=zyWrC0v0jtw
- Kazmier, L. (1993). *Estadística aplicada a la administración y a la economía.* McGraw Hill.
- Levin, R. R. (1996). *Estadística para administradores Schaum.* Prentice Hall.
- Lind, D. M., William, M. & Robert, D. M. (2004). *Estadística para administración y economía*. Alfaomega.
- Mendenhall, W., Robert, j. B. & Barbara, M. B. (2010). *Introducción a la probabilidad estadística*. CENGAGE Learnig.
- Montgomery, R. (2004). *Probabilidad y estadística aplicada a la ingeniería*. Limusa Wiley.
- Mora, C. (2007). *Latin American journal of physics education.* Vol. 1, No. 1, recuperado el 15 de enero de 2014 de www.journal.lapen.org.mx
- Morales, V. D. (2008). *Clasificadores bayesianos en la selección embrionaria en tratamientos de reproducción asistida*. Tesis doctoral. Departamento de Ciencias de la Computación e Inteligencia Artificial. Del país vasco. Donostia. Recuperado el 29 de diciembre de 2013 de: http://www.sc.ehu.es/ccwbayes/members/dinora/archivos/tesisDinora.pdf
- Navidi, W. (2006). *Estadística para ingenieros y científicos.* Mc Graw Hill.
- Nieves, A. D. & Federico, C. D. (2010). *Probabilidad y estadística para ingeniería un enfoque moderno.* Mc Graw Hill.
- Quiñones, F. (2010). *Diagnóstico de cáncer de mama mediante redes bayesianas.* Recuperado el 29 de diciembre del 2013 de: http://fdquinones.wordpress.com/2010/06/01/diagnostico-de-cancer-de-mama-mediante-redes-bayesianas/
- Sánchez, D.; Miranda, M. y Cerda, L. (2004). *Reglas de asociación aplicadas a la detección de fraude con tarjetas de crédito*. XII Congreso Español sobre Tecnologías y Lógica Fuzzy.
- Seymour, L. (1991). *Probabilidad.* Schaum. Mc Graw Hill.
- Silva, L.C., Muñoz, V. (2007). XVII Reunión Científica SEE. Santiago de Compostela. Debate sobre métodos frecuentistas *vs* bayesianos. Recuperado en diciembre 15 de 2013 de http://lcsilva.sbhac.net/Debate%20frec-bayes.pdf
- Sheldon, M. R. (2000). *Prbabilidad y estadística para ingenieros.* Mc Graw Hill.
- Sucar, L.E. (2003). *Redes bayesianas: principios y aplicaciones.* Puebla. Recuperado el 15 de abril de 2014 de http://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf.
- Walpole, R., Myers, R., & Myers, S. (2012). *Probabilidad y estadística para ingeniería y ciencias*. Pearson Education.