

Métodos probabilistas

Práctica del tema 1: Fundamentos de redes
bayesianas

Gabriel Vázquez Torres

Métodos probabilistas

Práctica del tema 1: Fundamentos de redes bayesianas

Contenido

Descripción de la práctica	2
Ejercicio 1.1.....	3
Ejercicio 1.2.....	7
Apartado 1:.....	7
Apartado 2.....	8
Ejercicio 1.3.....	9
Apartado 1.....	10
Apartado 2.....	10
Ejercicio 1.4.....	12
Ejercicio 1.5.....	13
Ejercicio 1.6.....	14
Ejercicio 1.7.....	15
Apartado 1.....	15
Apartado 2.....	15
Apartado 3.....	16
Apartado 4 y 5	16

Descripción de la práctica

Realizar diferentes ejercicios que se mostrarán en los siguientes puntos.

NOTA: Comentar que esta práctica la he realizado con WORD por falta de tiempo y por desconocimiento de Latex y Lyx. En las siguientes prácticas trataré de usar Latex.

Ejercicio 1.1

Ejercicio 1.1.

Sea una distribución de probabilidad P dada por la siguiente tabla.

a	b	c	$P(a, b, c)$
$+a$	$+b$	$+c$	0,03
$+a$	$+b$	$\neg c$	0,12
$+a$	$\neg b$	$+c$	0,07
$+a$	$\neg b$	$\neg c$	0,28
$\neg a$	$+b$	$+c$	0,03
$\neg a$	$+b$	$\neg c$	0,27
$\neg a$	$\neg b$	$+c$	0,02
$\neg a$	$\neg b$	$\neg c$	0,18

Señale cuáles de estas relaciones son verdaderas y cuáles son falsas. Para la primera y la quinta, indique detalladamente los cálculos que ha realizado.

1. $I_P(A, B)$
2. $I_P(A, C)$
3. $I_P(B, C)$
4. $I_P(A, B|C)$
5. $I_P(B, C|A)$

Ilustración 1: Ejercicio 1

En primer lugar vamos a calcular las probabilidades marginales de las siguientes combinaciones:

-----	bc	$\neg bc$	$\neg b\neg c$	$b\neg c$	$P(x_i)$
$+a$	0,03	0,07	0,28	0,12	0,5
$\neg a$	0,03	0,02	0,18	0,27	0,5
$P(x, y)$	0,06	0,09	0,46	0,39	1,00

Tabla 1: Probabilidad marginal para "a"

-----	ac	¬ac	¬a¬c	a¬c	P(xi)
+b	0,03	0,03	0,27	0,12	0,45
¬b	0,07	0,02	0,18	0,28	0,55
P(x,y)	0,10	0,05	0,45	0,40	1,00

Tabla 2: Probabilidad marginal para "b"

-----	ab	¬ab	¬a¬b	a¬b	P(xi)
+c	0,03	0,03	0,02	0,07	0,45
¬c	0,12	0,27	0,18	0,28	0,55
P(x,y)	0,15	0,30	0,20	0,35	1,00

Tabla 3: Probabilidad marginal para "c"

Para todos los casos que no son condicionales, tenemos que:

$$I_p(X, Y) \Rightarrow P(x, y) = P(x) \cdot P(y) \quad \forall x, \forall y$$

Ilustración 2: Cálculo de la independencia probabilística

Para 1:

Como para que dos características sean independientes entre sí hay que realizar la operación anterior para todas las variables, hemos encontrado una en la que no se verifica la igualdad:

$$R = P(+a) = 0,5$$

$$M = P(+b) = 0,45$$

$$R \cdot M = 0,225$$

$$P(+a, +b) = 0,15$$

Como 0,15 no es igual a 0,225, la relación "1." es Falsa por la fórmula anterior.

Para 2: Es Falsa.

Para 3. Es Falsa.

Teniendo en cuenta para la 4 y 5:

$$I_p(X, Y | Z) \Rightarrow P(x, y | z) = P(x | z) \cdot P(y | z) \quad \forall x, \forall y, \forall z$$

$$P(x, y | z) = \frac{P(x, y, z)}{P(z)}$$

Ilustración 3: Cálculo de independencia probabilística con condición

Para 4 es Falsa

Para 5:

Teniendo en cuenta las fórmulas, cogemos un valor para A,B y C. En este caso: +a,+b,+c

En primer lugar calculamos $P(+b,+c | +a) = P(+b,+c,+a)/P(+a) \rightarrow$

$P(+b,+c,+a)$ nos lo dan en la tabla: \rightarrow

$$(0.03)/0,5 = 0,06$$

En segundo lugar calculamos $P(+b | +a) = P(+b,+a)/P(+a) \rightarrow$

$$0,15/0,5 = 0,3$$

En tercer lugar, calculamos: $P(+c | +a) = P(+c,+a)/P(+a) \rightarrow$

$$0,1/0,5 = 0,2$$

En cuarto lugar, comprobamos la igualdad:

$$0,06 = 0,3 * 0,2 \rightarrow 0,06 = 0,06$$

Para este caso sí se cumple la igualdad, probaremos con otro caso para llegar intentar llegar a un caso en el que no se cumpla. Si se cumple para todos, entonces la independencia probabilística es CIERTA.

Teniendo en cuenta las fórmulas, cogemos un valor para A,B y C. En este caso: $\neg a,+b,+c$

En primer lugar calculamos $P(+b,+c | \neg a) = P(+b,+c,\neg a)/P(\neg a) \rightarrow$

$P(+b,+c,\neg a)$ nos lo dan en la tabla: \rightarrow

$$(0.03)/0,5 = 0,06$$

En segundo lugar calculamos $P(+b | \neg a) = P(+b,\neg a)/P(\neg a) \rightarrow$

$$0,30/0,5 = 0,6$$

En tercer lugar, calculamos: $P(+c|\neg a) = P(+c, \neg a)/P(\neg a) \rightarrow$

$$0,05/0,5 = 0,1$$

En cuarto lugar, comprobamos la igualdad:

$$0,06 = 0,6 * 0,1 \rightarrow 0,06 = 0,06$$

Para este caso sí se cumple la igualdad, probaremos con otro caso para llegar intentar llegar a un caso en el que no se cumpla. Si se cumple para todos, entonces la independencia probabilística es CIERTA.

Siguiendo los mismos cálculos para todos los casos posibles, verificamos que en todos los casos la igualdad es cierta. Por lo tanto, la relación del punto 5 es CIERTA.

Ejercicio 1.2

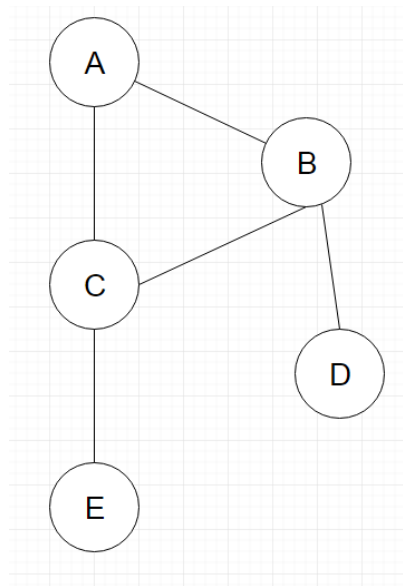


Ilustración 4: Representación del grafo del ejercicio 1.2

Apartado 1:

Tenemos en cuenta:

Definición 1.82 (Camino activo en un grafo no dirigido) Sea un grafo no dirigido \mathcal{G} , dos nodos A y B de \mathcal{G} y un subconjunto C de nodos de \mathcal{G} tal que ni A ni B pertenecen a C .

- Un camino de dos nodos, es decir, $A-B$ (un solo enlace), siempre está *activo*.
- Un camino de n nodos, es decir $A-X_1-\dots-X_{n-2}-B$ está *activo* cuando ningún nodo entre A y B pertenece a C , es decir, $\{X_1, \dots, X_{n-2}\} \cap C = \emptyset$; si algún X_i pertenece a C se dice que el camino está *inactivo*.

Ilustración 5: Definición de camino activo en grafo no dirigido.

Tenemos en cuenta la definición para grafo conectado y separado:

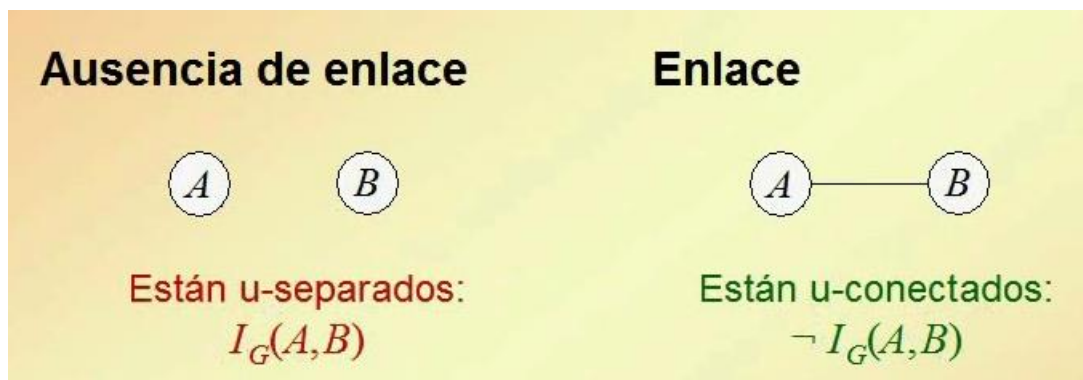


Ilustración 6: Representación de grafo separado o conectado (imagen obtenida del video “separación de grafos”).

Tenemos en cuenta que nos están pidiendo en todos los apartados, al no aparecer el signo “¬” delante de la “I” que significa “separados” o “independientes”, la AFIRMACIÓN de la separación entre sendos nodos de cada apartado.

- a) Es FALSA ya que existen dos caminos conectados entre a y b: $\{A \rightarrow B\}$, $\{A \rightarrow C \rightarrow B\}$.
- b) Es FALSA ya que existe un camino conectado entre b y d: $\{B \rightarrow D\}$
- c) Es FALSA ya que existen dos caminos conectados entre a y d: $\{A \rightarrow B \rightarrow D\}$, $\{A \rightarrow C \rightarrow B \rightarrow D\}$
- d) Es FALSA ya que existen dos caminos conectados entre d y e: $\{D \rightarrow B \rightarrow C \rightarrow E\}$, $\{D \rightarrow B \rightarrow A \rightarrow C \rightarrow E\}$

Apartado 2

Teniendo en cuenta la definición:

Definición 1.84 (Nodos conectados) Sea un grafo dirigido o no dirigido \mathcal{G} y un subconjunto C de \mathcal{G} . Dos nodos de \mathcal{G} (no pertenecientes a C) están *conectados dado C* cuando existe al menos un camino activo entre ellos. En otro caso se dice que están *separados dado C*. La notación $I_{\mathcal{G}}(A, B|C)$ indica que A y B están separados y por tanto $\neg I_{\mathcal{G}}(A, B|C)$ indica que están conectados.

Ilustración 7: Definición de nodos conectados

Y la Ilustración 66:

- a) Es FALSA ya que existe un camino posible entre a y b: $\{A \rightarrow B\}$. El camino $\{A \rightarrow C \rightarrow B\}$ ha quedado bloqueado por C.
- b) Es VERDADERA ya que no existe ningún camino posible si el nodo B es el condicionante entre a y d. El camino $\{A \rightarrow C \rightarrow B \rightarrow D\}$ ha quedado bloqueado por B.
- c) Es VERDADERA ya que no existe ningún camino entre d y e si el nodo C es el condicionante. Los caminos $\{D \rightarrow B \rightarrow C \rightarrow E\}$ y $\{D \rightarrow B \rightarrow A \rightarrow C \rightarrow E\}$ han quedado bloqueados por C.
- d) Es FALSA ya que existe un camino entre d y e: $\{D \rightarrow B \rightarrow C \rightarrow E\}$. El camino $\{D \rightarrow B \rightarrow A \rightarrow C \rightarrow E\}$ ha quedado bloqueado por A.

Ejercicio 1.3

Teniendo en cuenta la siguiente definición:

Definición 1.83 (Camino activo en un grafo dirigido) Sea un grafo dirigido \mathcal{G} , dos nodos A y B de \mathcal{G} y un subconjunto C de \mathcal{G} tal que ni A ni B pertenecen a C .

- Un camino de dos nodos, es decir, $A \rightarrow B$ o $B \rightarrow A$ (un solo enlace), siempre está *activo*.
- Un camino de tres nodos puede ser de varios tipos: $A \rightarrow X \rightarrow B$, $B \rightarrow X \rightarrow A$, $A \leftarrow X \rightarrow B$ o $A \rightarrow X \leftarrow B$.
- Un camino del tipo $A \rightarrow X \rightarrow B$ o $A \leftarrow X \rightarrow B$ está *activo* si X no pertenece a C . Cuando X pertenece a C se dice que el camino está inactivo porque ha sido *bloqueado* por X .
- Un camino del tipo $A \rightarrow X \leftarrow B$ está *activo* si X o alguno de sus descendientes pertenece a C . En este caso se dice que el camino ha sido *activado* por el nodo que pertenece a C . Cuando ni X ni sus descendientes pertenecen a C , el camino está inactivo.
- Un camino de n nodos está activo si cada par de enlaces consecutivos forman un camino activo.

Ilustración 8: Definición de camino activo en un grafo dirigido.

Y que $I_G(A, B)$ representa un grafo que no está conectado entre los nodos A y B .

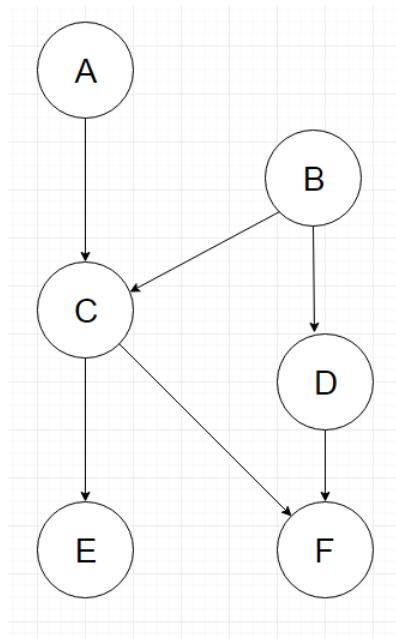


Ilustración 9: Representación del grafo dirigido del ejercicio 1.3

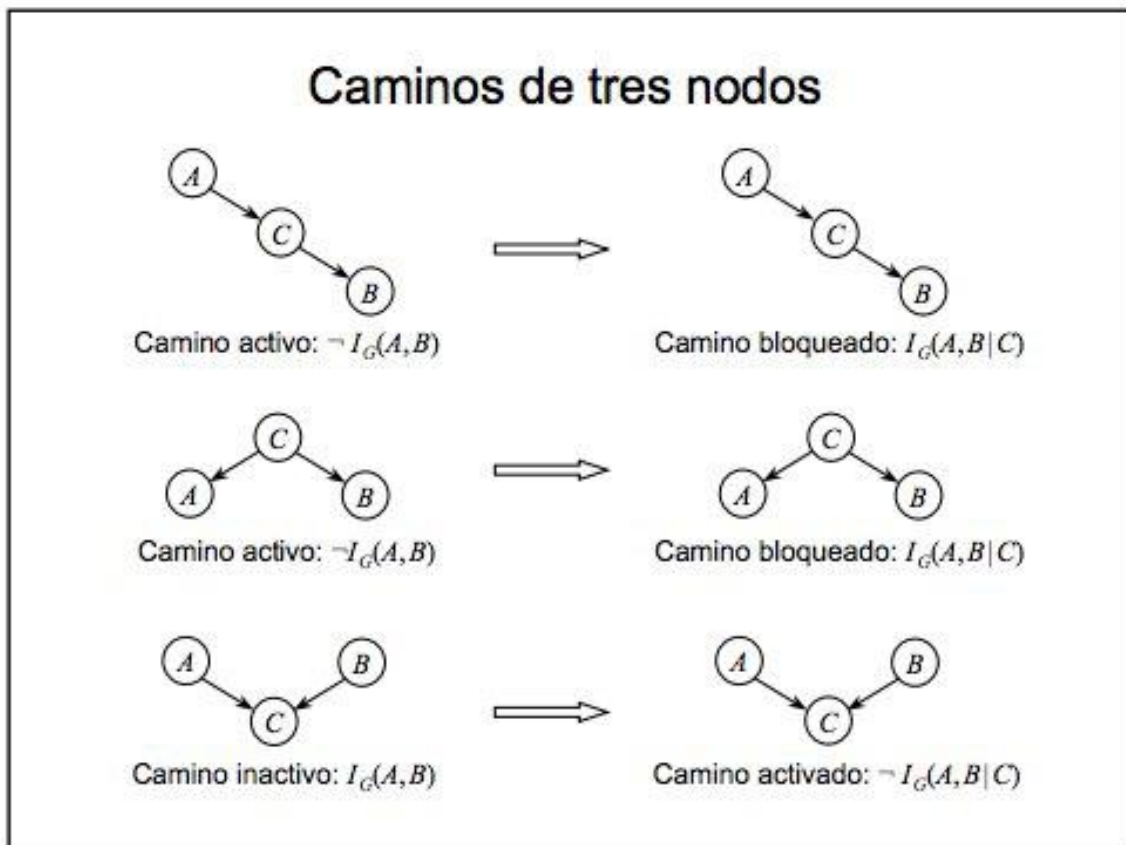


Ilustración 10: Camino de tres nodos en grafo dirigido

Apartado 1

- Es CIERTO ya que es un camino inactivo.
- Es FALSO ya que hay un camino activo $\{A \rightarrow C \rightarrow E\}$. El camino $\{B \rightarrow D \rightarrow F \rightarrow C \rightarrow E\}$ sería inactivo.
- Es CIERTO ya que todos los caminos son inactivos. El camino $\{B \rightarrow D\}$ sería un camino activo.
- Es FALSO ya que existe el camino activo $\{A \rightarrow C \rightarrow F\}$. Existe otros activos como: $\{D \rightarrow F\}$, $\{B \rightarrow D \rightarrow F\}$, $\{C \rightarrow F\}$ y $\{B \rightarrow C \rightarrow F\}$
- Es FALSO ya que gracias a B, existe un camino directo entre D y E. Los caminos $\{B \rightarrow C \rightarrow E\}$, $\{B \rightarrow D\}$ son directos.
- Es FALSO ya que existe el camino activo $\{E \rightarrow C \rightarrow F\}$

Apartado 2

- Es FALSO ya que existe un camino activado $\{A \rightarrow C \rightarrow B\}$ y $\{B \rightarrow C \rightarrow A\}$ entre A y B. Los caminos $\{A \rightarrow C\}$, $\{B \rightarrow C\}$, $\{A \rightarrow C \rightarrow E\}$, $\{B \rightarrow C \rightarrow E\}$ y $\{B \rightarrow C \rightarrow F\}$ se han desactivado.
- Es FALSO ya que el descendiente E ha activado el camino A,C,B. Los caminos $\{A, C, E\}$ y $\{B, C, E\}$ han sido desactivados.

- c) Es FALSO ya que el descendiente D ha activado el camino $\{A \rightarrow C \rightarrow B\}$ y $\{B \rightarrow C \rightarrow A\}$. Se han desactivado los caminos $\{A, C, F\}$, $\{B, C, F\}$, $\{B, D, F\}$, $\{D, F\}$.
- d) Es FALSO ya que el camino $\{A \rightarrow C \rightarrow B \rightarrow D\}$ ha sido activado por C. Los caminos $\{A \rightarrow C \rightarrow B\}$ y $\{B \rightarrow C \rightarrow A\}$ se han activado. Los caminos $\{A \rightarrow C\}$, $\{B \rightarrow C\}$, $\{A \rightarrow C \rightarrow E\}$, $\{B \rightarrow C \rightarrow E\}$ y $\{B \rightarrow C \rightarrow F\}$ se han desactivado.
- e) Es FALSO ya que el camino $\{A \rightarrow C \rightarrow D\}$ ha sido activado por F. Los caminos $\{C \rightarrow D\}$ han quedado activados.
- f) Es FALSO ya que existe el camino $\{A \rightarrow C \rightarrow B \rightarrow D \rightarrow F\}$. Los caminos $\{A \rightarrow C \rightarrow B\}$ y $\{B \rightarrow C \rightarrow A\}$ se han activado. Los caminos $\{A \rightarrow C\}$, $\{B \rightarrow C\}$, $\{A \rightarrow C \rightarrow E\}$, $\{B \rightarrow C \rightarrow E\}$ y $\{B \rightarrow C \rightarrow F\}$ se han desactivado.
- g) Es CIERTO ya que no existen ningún camino entre A y F. Los caminos $\{A, C, E\}$, $\{B, C, F\}$ y $\{B, C, E\}$ han sido desactivados.
- h) Es CIERTO ya que no existe ningún camino activado entre D y E. Se ha desactivado el camino $\{C, D\}$.
- i) Es CIERTO ya que no existe ningún camino entre E y F por la condición de C. Los caminos $\{A \rightarrow C \rightarrow B\}$ y $\{B \rightarrow C \rightarrow A\}$ se han activado. Los caminos $\{A \rightarrow C\}$, $\{B \rightarrow C\}$, $\{A \rightarrow C \rightarrow E\}$, $\{B \rightarrow C \rightarrow E\}$ y $\{B \rightarrow C \rightarrow F\}$ se han desactivado.

Ejercicio 1.4

Para los ejercicios 1.4, 1.5 y 1.6 se podría tener en cuenta el “conjunto vacío” que es un subconjunto de la distribución P . Todos los grafos cubren el conjunto vacío al menos en dos variables. El conjunto vacío se representa mediante “ \emptyset ”.

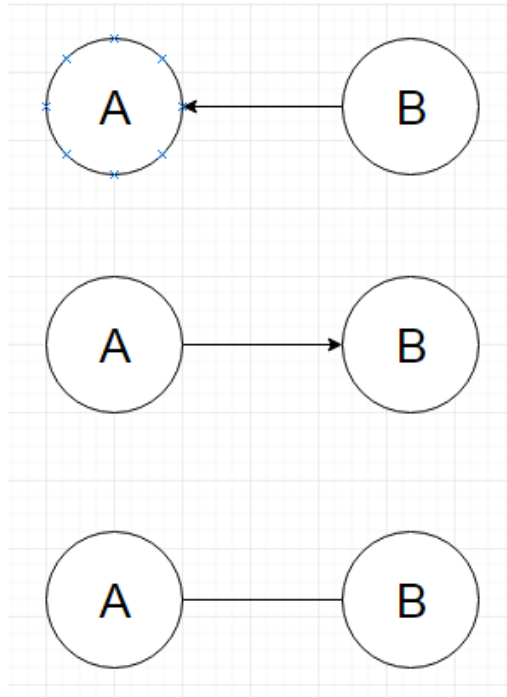


Ilustración 11: Respuesta al ejercicio 1.4

Ejercicio 1.5

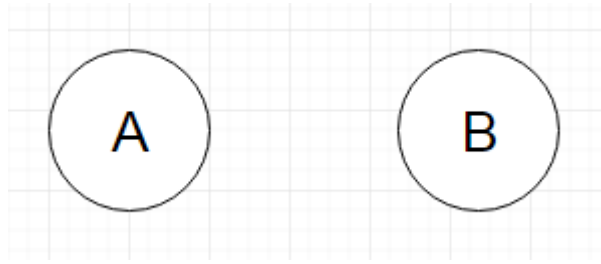


Ilustración 12: Respuesta al ejercicio 1.5

Ejercicio 1.6

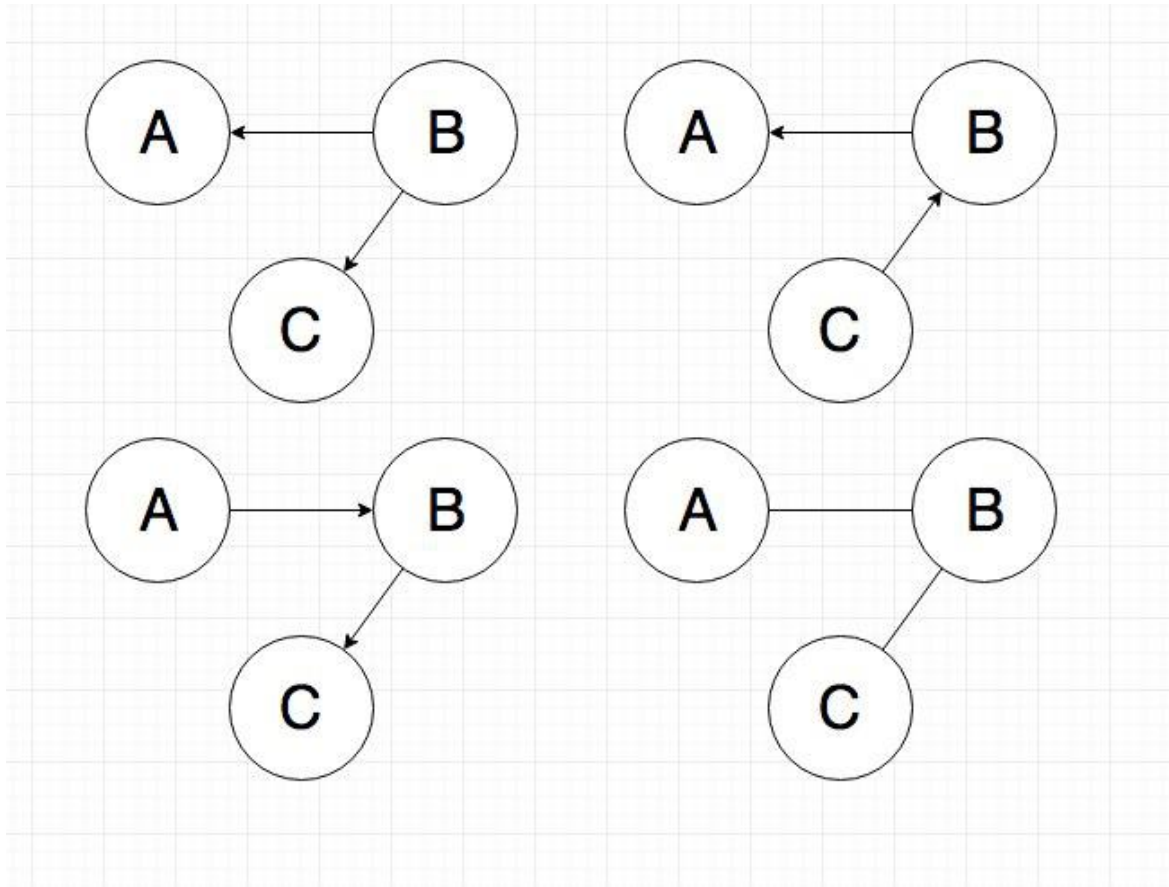


Ilustración 13: Respuesta al ejercicio 1.6

Ejercicio 1.7

En cierto país, la prevalencia de la fiebre tifoidea es el 0'001 y la de la tuberculosis 0'01. La fiebre tifoidea produce fiebre siempre, y bradicardia (ritmo cardíaco lento) en el 40 % de los casos. La tuberculosis produce fiebre en el 60 % de los casos y taquicardia (ritmo cardíaco más rápido de lo normal) en el 58 %. La prevalencia de la fiebre en pacientes que no sufren ninguna de estas dos enfermedades es el 1'5 %, la de bradicardia el 0'05 % y la de taquicardia el 1'3 %.

Ilustración 14: Ejercicio 1.7

Apartado 1

Tenemos dos variables independientes: las enfermedades. En este caso son, la fiebre tifoidea y la tuberculosis. Introducimos la hipótesis de que los diagnósticos son exclusivos debido a que es muy improbable que se den los dos casos ($0,001 \cdot 0,01 = 0,00001$).

También es cierto que estos dos diagnósticos no son exhaustivos porque es posible que no haya ninguna de las enfermedades. Por ello, la variable "E" (Enfermedad) puede tomar 3 posibles valores: Ef (Enfermedad fiebre tifoidea), Et (Enfermedad tuberculosis), En (Enfermedad ninguna).

Las probabilidades a priori para E son: $P(E_f) = 0,001$, $P(E_t) = 0,01$ y $P(E_n) = 0,989$

Los síntomas que se pueden tener son: fiebre (f), taquicardia (t) y bradicardia (b).

Los valores están representados en el apartado 3.

Apartado 2

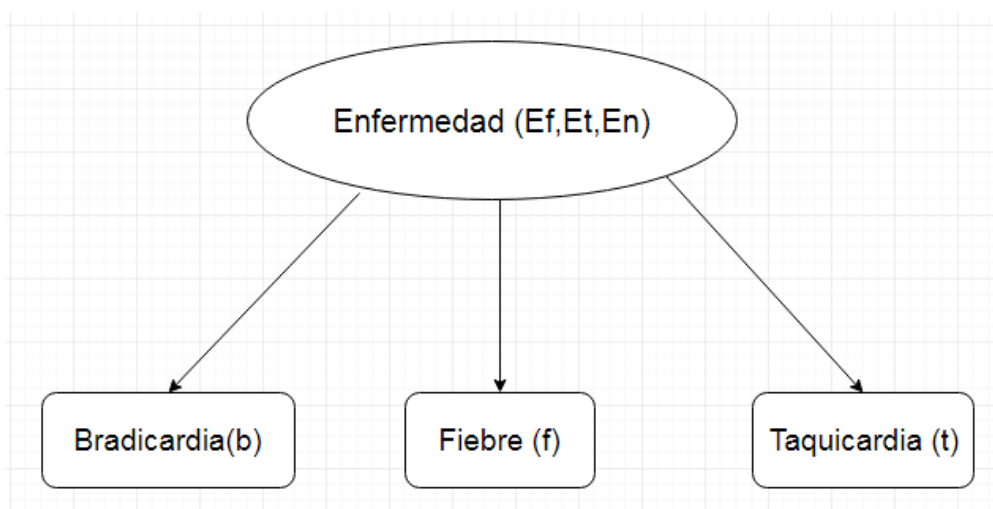


Ilustración 15: Diagrama correspondiente al método bayesiano ingenuo para el problema.

Apartado 3

$P(f E)$	E_f	E_t	E_n
+f	1	0.6	0.015
¬f	0	0,4	0,985

Tabla 4: Probabilidad condicional para el síntoma fiebre según enfermedad.

$P(t E)$	E_f	E_t	E_n
+t	0	0.58	0.0005
¬t	1	0,42	0,9995

Tabla 5: Probabilidad condicional para el síntoma taquicardia según enfermedad.

$P(b E)$	E_f	E_t	E_n
+b	0,4	0	0,013
¬b	0,6	1	0,987

Tabla 6: Probabilidad condicional para el síntoma braquicardia según enfermedad.

Apartado 4 y 5

Las hipótesis son: bradicardia y taquicardia son totalmente independientes.

Otra hipótesis es que si tenemos fiebre, podemos tener cualquier enfermedad (o ambas, pero este caso lo hemos despreciado por su poca posibilidad)

Así, podemos estar seguros de: $P(E|b,t) = \alpha * P(E) * P(b|E) * P(t|E)$

Añadiendo la fiebre, tenemos que: $P(E|b,t,f) = \alpha * P(E) * P(b|E) * P(t|E) * P(f|E) \rightarrow T1$

Podríamos calcular α para todas nuestras variables. Debemos tener en cuenta que la fiebre no es independiente a cada enfermedad. Así pues, debemos calcular T1 para todas las variables que son independientes. Así tenemos un valor de α para cada posibilidad de la enfermedad (E_f , E_t , E_n), siendo α el porcentaje real de que ocurra una u ocurra otra teniendo en cuenta los hallazgos.

Podríamos analizar qué enfermedad es más probable según la fiebre. Para ello podemos utilizar la fórmula:

$$P(E_f|+f,+b)/P(E_t|+f,+t) = (P(E_f) * P(+b|E_f) * P(+f|E_f))/(P(E_t) * P(+t|E_t) * P(+f|E_t))$$



Según el resultado, si es mayor a 1, la enfermedad Ef para fiebre es más probable. Si es inferior a 1, entonces, si tenemos fiebre, Et es más probable.