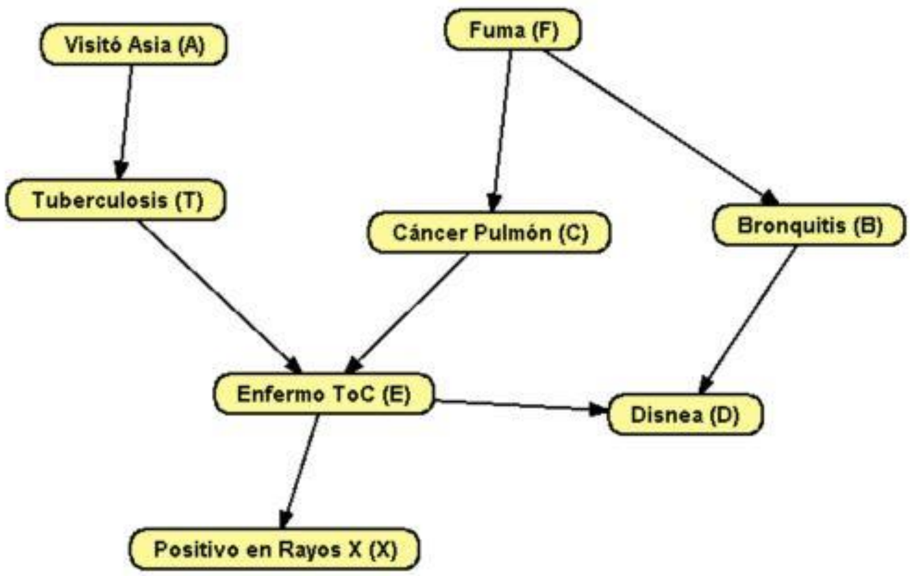
Práctica 4 Inteligencia Artificial

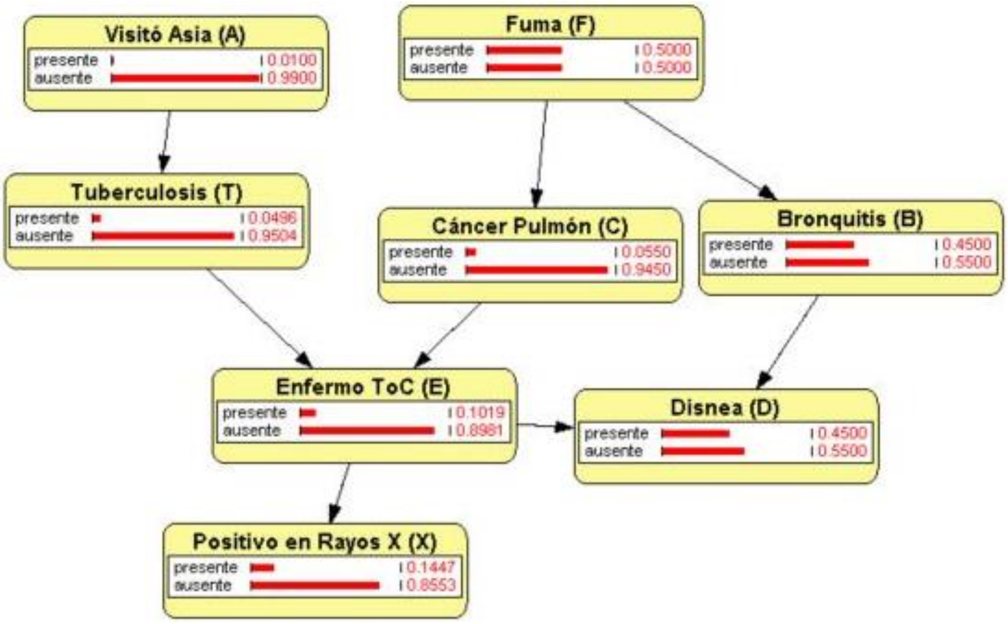
Alumno: Luis García Garcés

NIA:739202

1.Crea la red Bayesiana para las enfermedades del pulmón (red Asia), e introduce sus tablas de probabilidades.

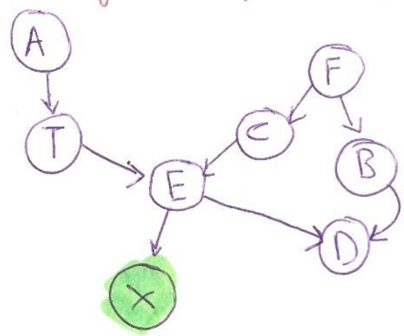


Red bayesiana obtenida en openmarkov.



Estudio previo del apartado a).

a) ¿Son A y F independientes dado X?



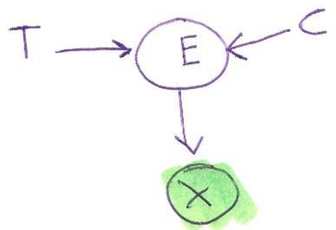
Serán independientes cuando no haya ningún camino activo de A a F.

Hay que comprobar si hay un camino activo.

Los caminos activos se componen de tripletas activos.

El primer triplete activo que encontramos es el de A ⟶ T ⟶ E. Siendo estas variables dependientes entre sí.

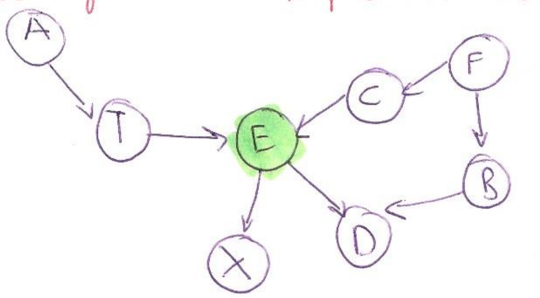
F ⟶ C ⟶ E también es un triplete activo.



Es un triplete activo ya que X es la evidencia y E es su antecesor.

[Al haber encontrado un camino activo compuesto por los tripletes anteriores, podemos deducir que son dependientes. No son independientes dado X.]

2) ¿A y B son independientes dado E?



A → T → E es un triplete activo.

T → E ← C es un triplete activo

C ← F → B es un triplete activo.

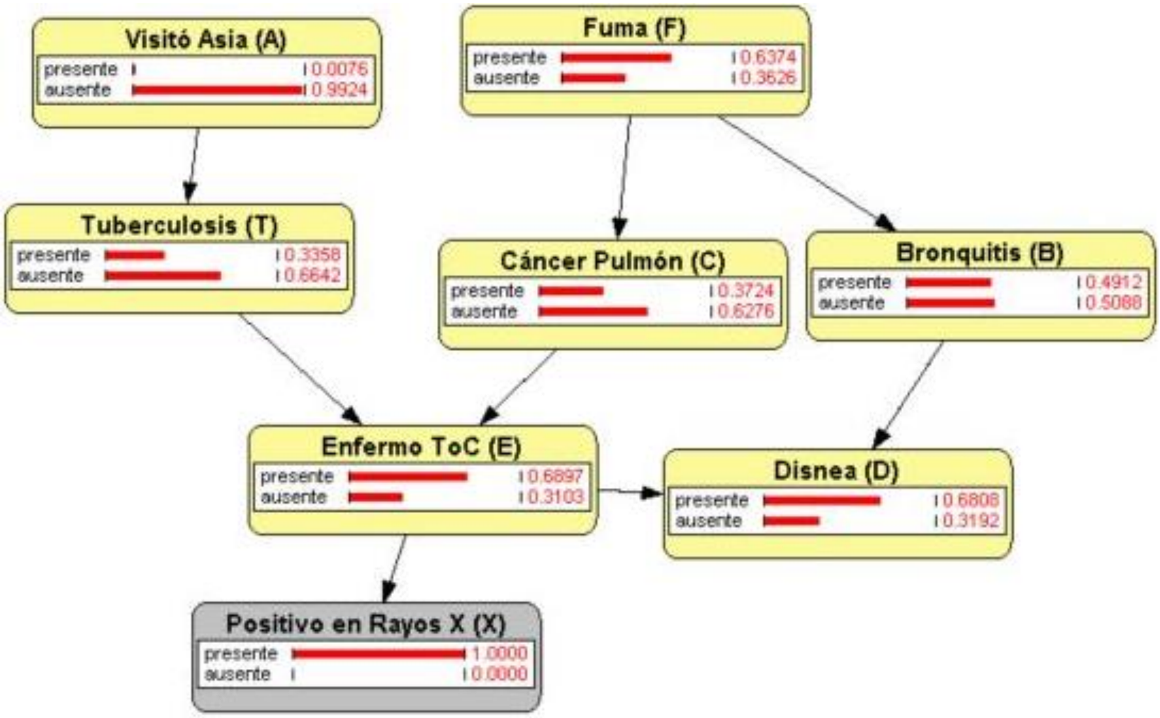
~~F → B → D~~

F → C → E es un triplete activo.

Como podemos comprobar hay un camino activo. Por lo tanto no son independientes, son dependientes.

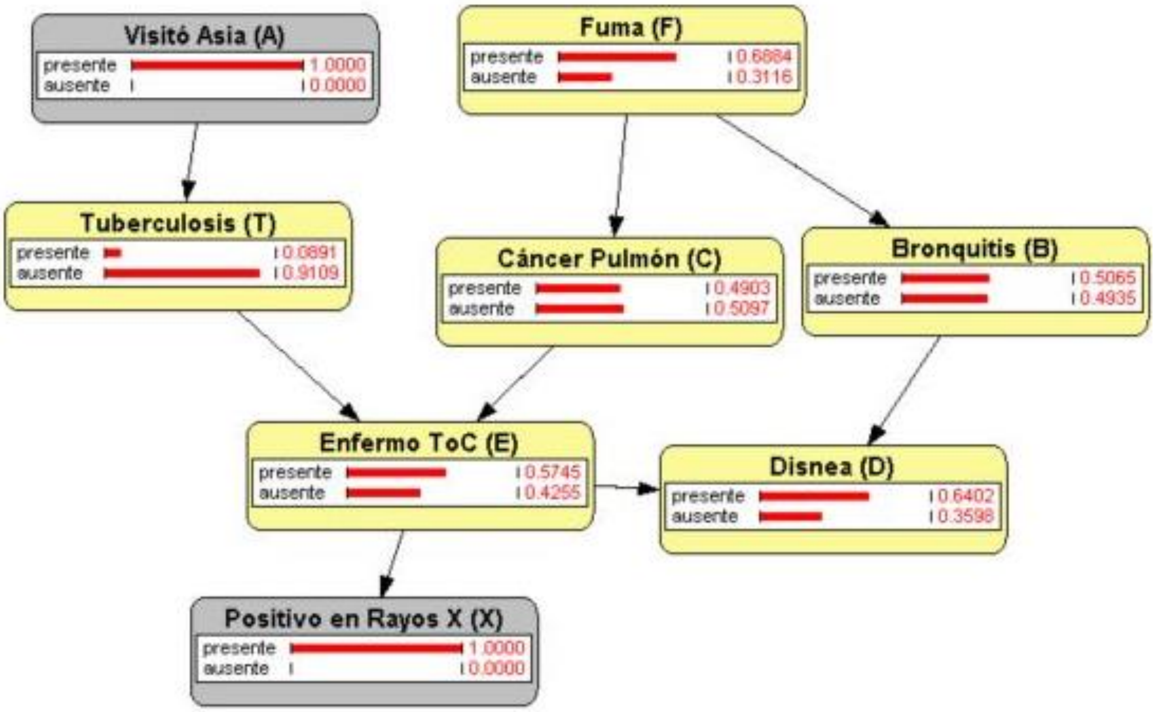
a) Haz las pruebas necesarias para comprobar si son ciertas las dos propiedades de independencia condicional del apartado 1.a) del estudio previo.

Para verificar que estas afirmaciones son correctas, se comprueba gráficamente.



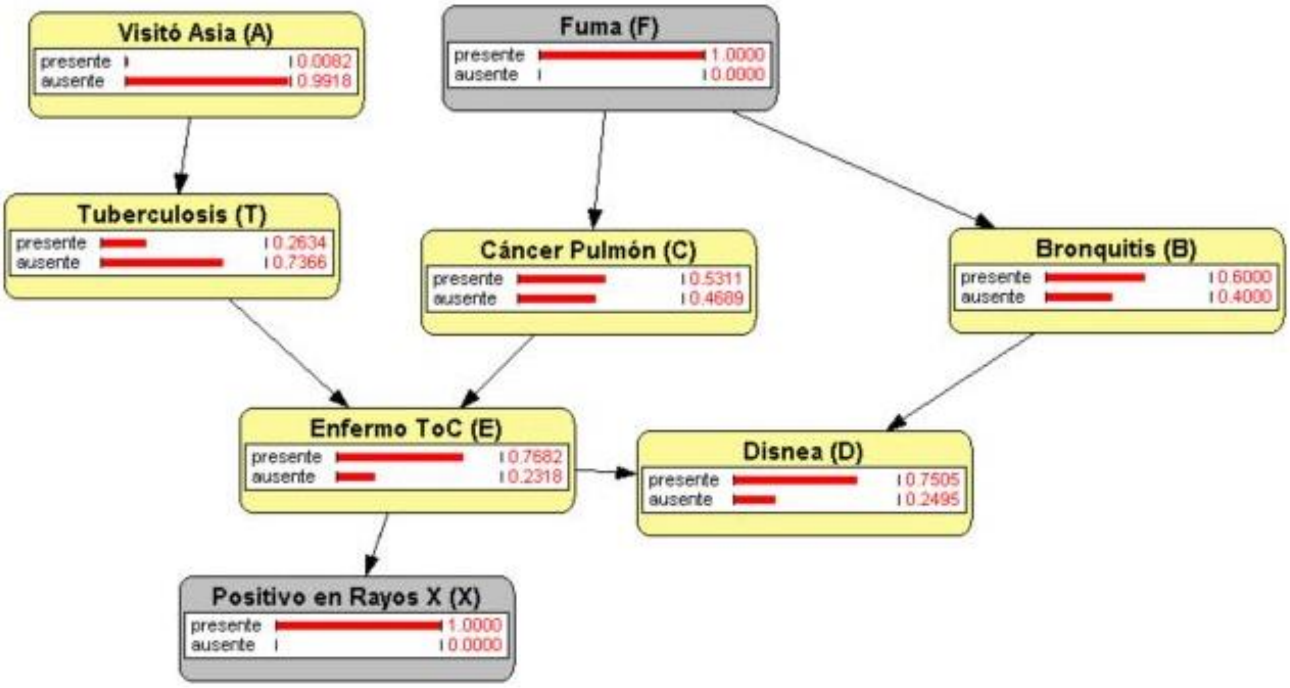
Valores iniciales de la red bayesiana dado X.

Comprobamos que A y F son dependientes modificando el valor de A.



Como podemos contemplar los valores de las probabilidades de F han cambiado y no solo los de F, sino los del resto de variables. Inicialmente la probabilidad de F presente es de 0,6374 y ahora al modificar A ha pasado a 0,6884.

Realizamos la misma comprobación, pero modificando el valor de F y comprobamos que las probabilidades de A han cambiado.



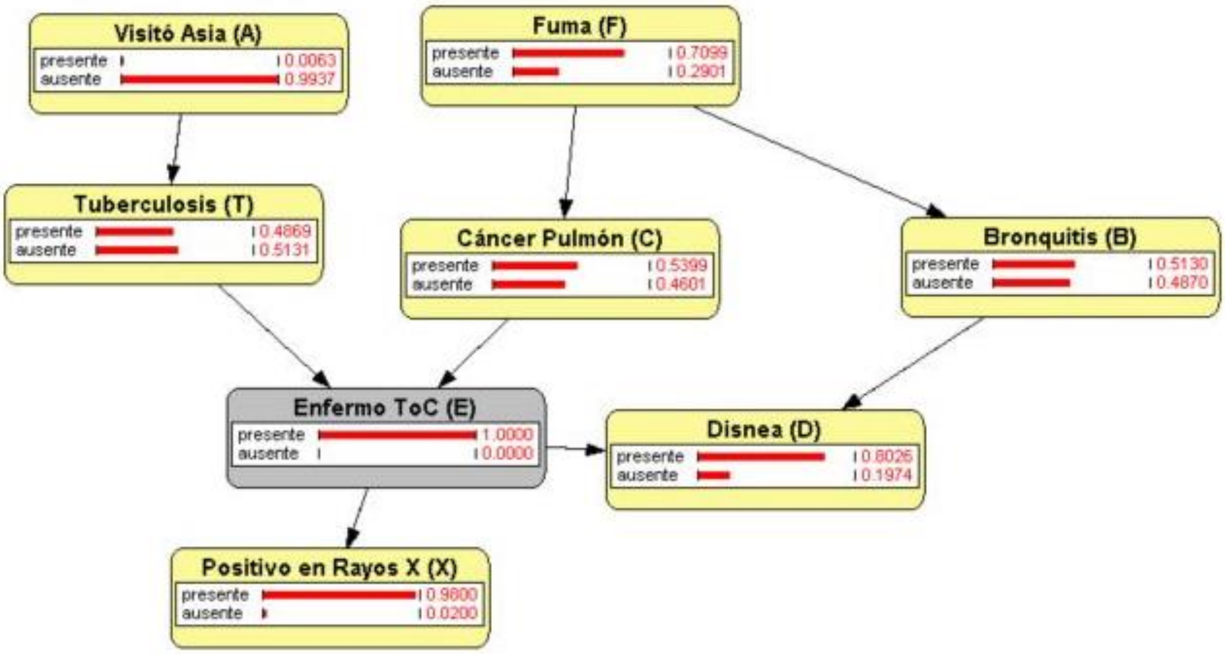
Observamos que la probabilidad de A presente a pasado de 0,0076 a 0,0082.

Esto indica que podemos afirmar que Ay F son dependientes dado X.

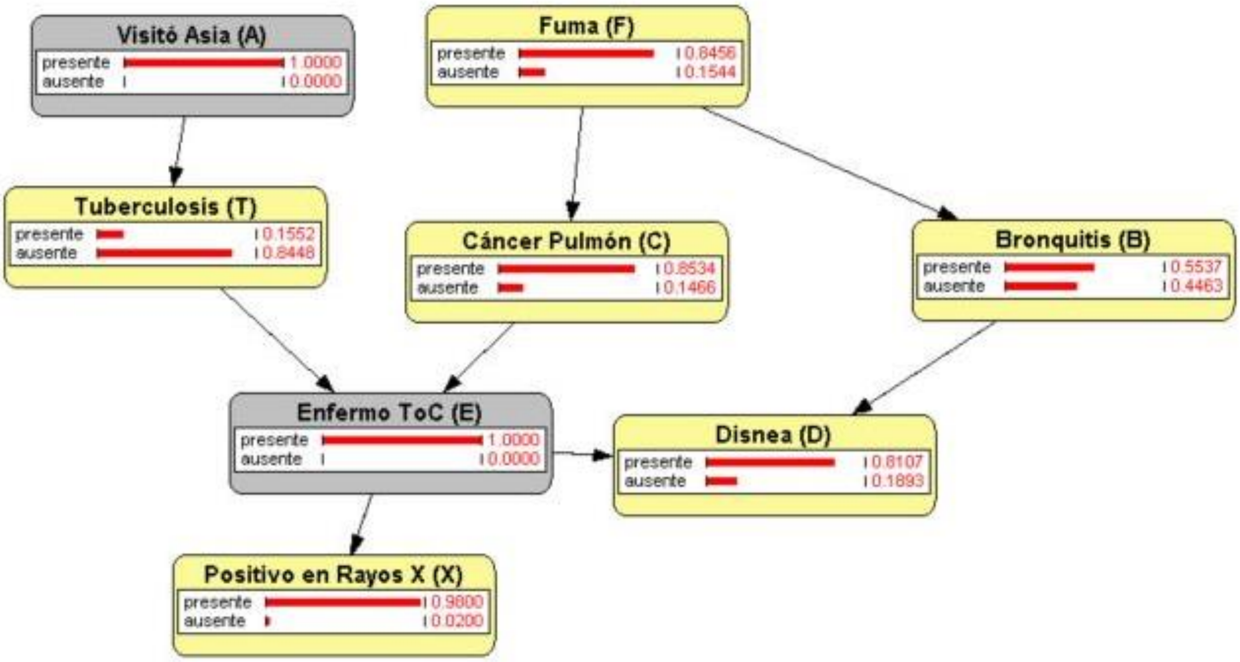
Confirmado el resultado de la primera pregunta. Toca verificar el resultado de la siguiente.

¿Son A y B independientes dado E?

En el estudio previo hemos concluido que no lo son, ahora lo confirmaremos gráficamente.

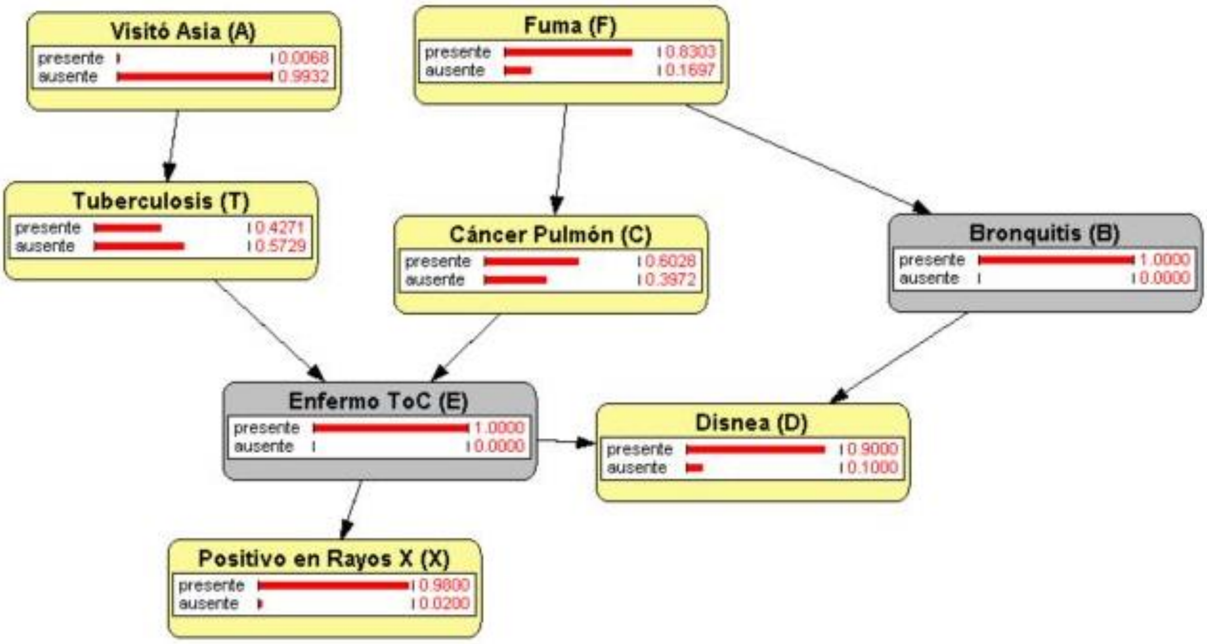


Inicialmente modificamos A, para comprobar que el valor de las probabilidades de F cambia.



La probabilidad de F presente pasa a ser de 0,7099 a 0,8456.

Para finalizar modificaremos la probabilidad de B.



Levemente, pero la probabilidad de A presente se ha visto modificada al actualizar el valor de B.

Por lo tanto, podemos confirmar que A y B son dependientes dado E.

Estudio previo del apartado b).

b] calcular $P(C \mid +x, +d, -f -t)$

Mediante eliminación de variables

1] Ignorar.

Las variables que no son ni evidencia ni query son A, B y E.

No podemos ignorar ni B ni E ya que son antecesores de evidencias y se encuentran en caminos activos.

Podemos ignorar A, ya que, aunque sea antecesor de T no influye sobre el resto de variables cuando $P(T) = 1$, no forma parte de un camino activo.

2] Instanciar

Queremos obtener $P(C \mid +x, +d, -f, -t)$

Tenemos : $P(+D \mid E, B)$ $\quad P(X \mid E)$

$P(+B \mid F)$ $\quad P(-F)$

$P(C \mid F)$ $\quad P(-T)$

$P(E \mid C, T)$

E	B	$P(d \mid E, B)$
+	+	0,9
+	−	0,7
−	+	0,8
−	−	0,1

X	E	$P(X \mid E)$
+	+	0,98
+	−	0,05
−	+	0,02
−	−	0,95

C	T	$P(+e \mid C, T)$
+	+	1
+	−	1
−	+	1
−	−	0

1

C	F	P(C | F)
+	+	0,1
+	−	0,01
−	+	0,9
−	−	0,99

B	F	P(B | F)
+	+	0,6
+	−	0,3
−	+	0,4
−	−	0,7

3) Eliminar variables que no nos interesan

Tenemos que eliminar B y E.

1) Eliminamos B.

E	B	$P(+d \mid E,B) \cdot P(B \mid -f) = P(B,D \mid -f, E)$
+	+	$0,9 \cdot 0,3 = 0,27$
+	−	$0,7 \cdot 0,7 = 0,49$
−	+	$0,8 \cdot 0,3 = 0,24$
−	−	$0,1 \cdot 0,7 = 0,07$

$0,27 + 0,49 \rightarrow 0,76$

$0,24 + 0,07 \rightarrow 0,31$

Hemos eliminado B $\rightarrow$ $P(+d \mid -f, E) = 0,76$

$P(+d \mid -f, -E) = 0,31$

2) Eliminamos E

C	E	$P(+d \mid -f, E) \cdot P(+e \mid C, -t) \cdot P(+x \mid E)$
+	+	$0,76 \cdot 1 \cdot 0,98 = 0,7448$
+	−	$0,31 \cdot 0 \cdot 0,05 = 0$
−	+	$0,76 \cdot 0 \cdot 0,98 = 0$
−	−	$0,31 \cdot 1 \cdot 0,05 = 0,0155$

$\rightarrow 0,7448$

$\rightarrow 0,0155$

$P(+x, +d \mid -f, +c, -t) = 0,7448$

$P(+x, +d \mid -f, -c, -t) = 0,0155$

4) juntar todos los factores restantes y normalizar.

$P(+c \mid -f) = 0{,}01$ $\quad P(-c \mid -f) = 0{,}99$

$P(-f) = 1$ $\quad P(-t) = 1$

$P(+x, +d \mid -f, +c, -t) = 0{,}7448$

$P(+x, +d \mid -f, -c, -t) = 0{,}0155$

$P(+x, +d \mid -f, C, -t) \cdot P(-f) = P(+x, +d, -f \mid C, -t)$

$P(+x, +d, -f \mid C, -t) \cdot P(-t) = P(+x, +d, -f, -t \mid C)$

C	$P(C, -f) \cdot P(+x, +d, -f, -t \mid C)$	$= P(+x, +d, -f, -f, C)$
+	$0{,}01 \cdot 0{,}7448$	$0{,}007448$.
−	$0{,}99 \cdot 0{,}0155$	$0{,}015345$

Normalizamos

$$\frac{0{,}007448}{(0{,}007448 + 0{,}015345)} = 0{,}326766989$$

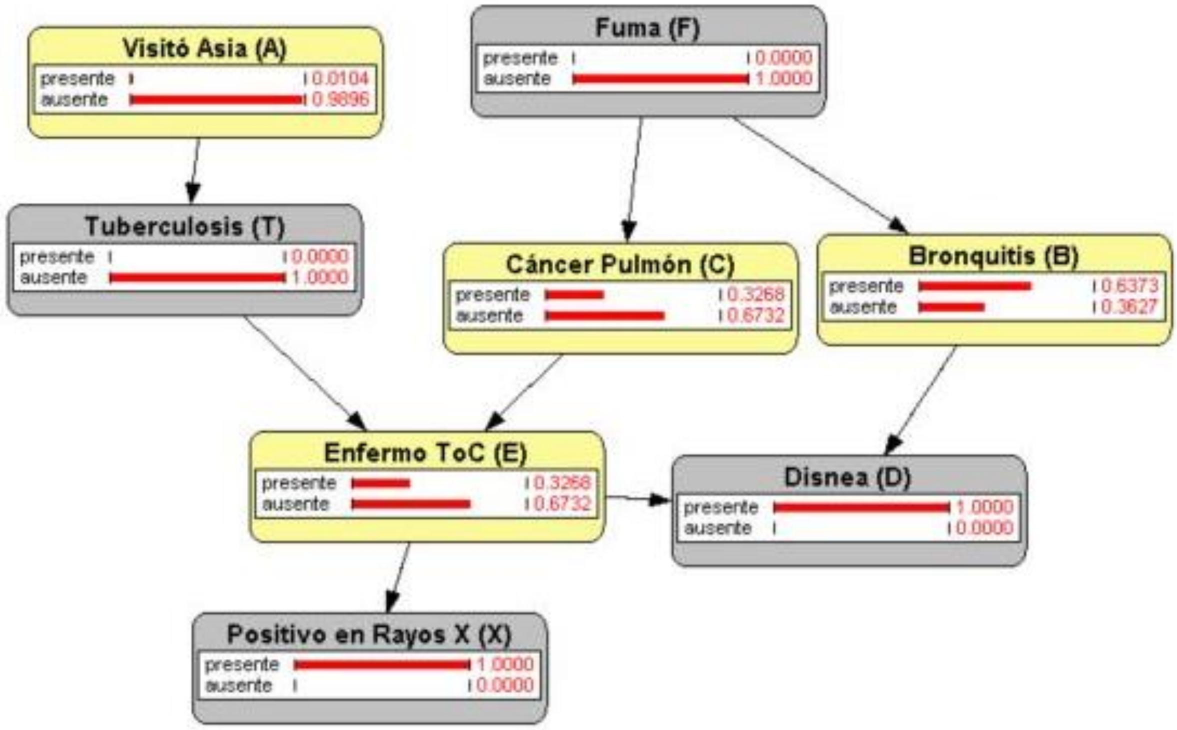
$$\frac{0{,}015345}{(0{,}007448 + 0{,}015345)} = 0{,}67323301$$

Resultado → La probabilidad de tener cáncer dando positivo en rayos X, teniendo disnea, no fumando y no teniendo tuberculosis es de 0,326766989.

En el estudio previo llegamos a la conclusión de que la probabilidad de que un paciente tenga cáncer, dando positivo en rayos X, teniendo disnea, no fumando y no teniendo tuberculosis es de 0,326766989.

b) Calcula la probabilidad pedida en el apartado 1.b) del estudio previo. Comprueba que las variables que pensabas que se pueden ignorar, no influyen en el resultado.

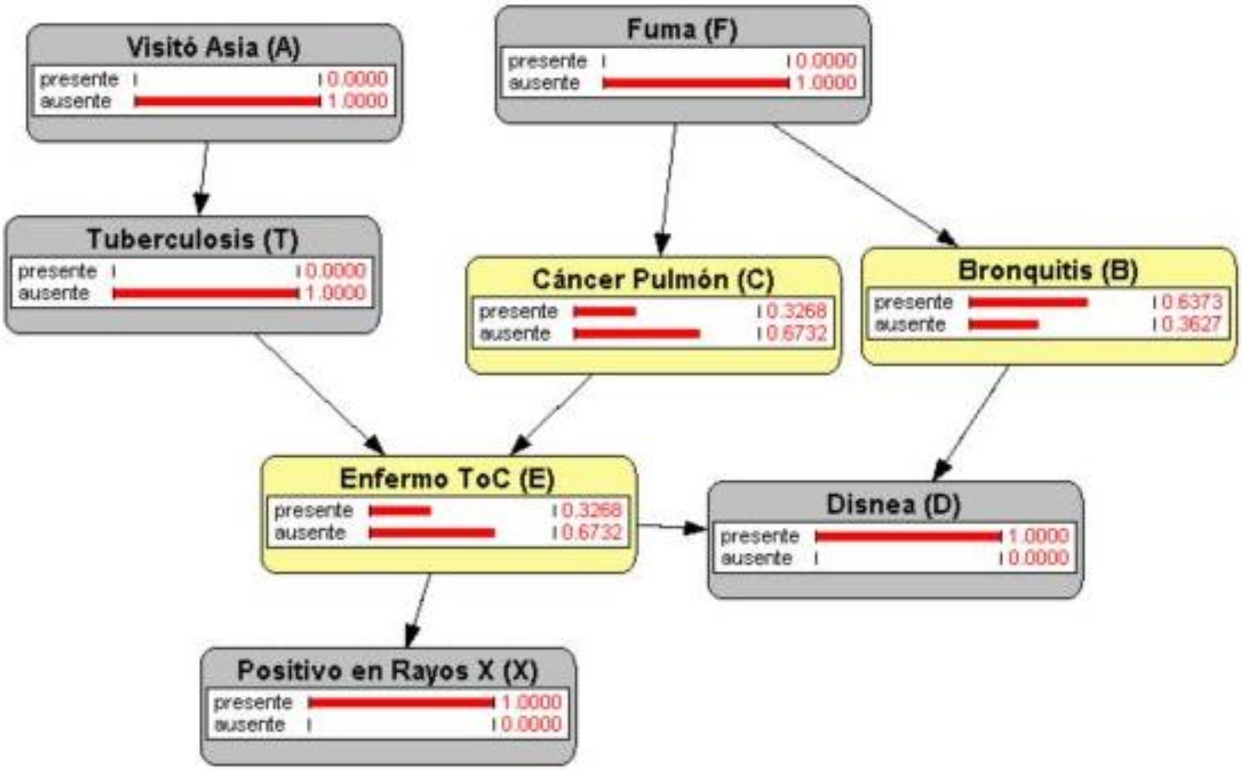
Para comprobar si este resultado es correcto creamos la red Bayesiana en OpenMarkov y completamos las tablas con las probabilidades dadas en el enunciado.



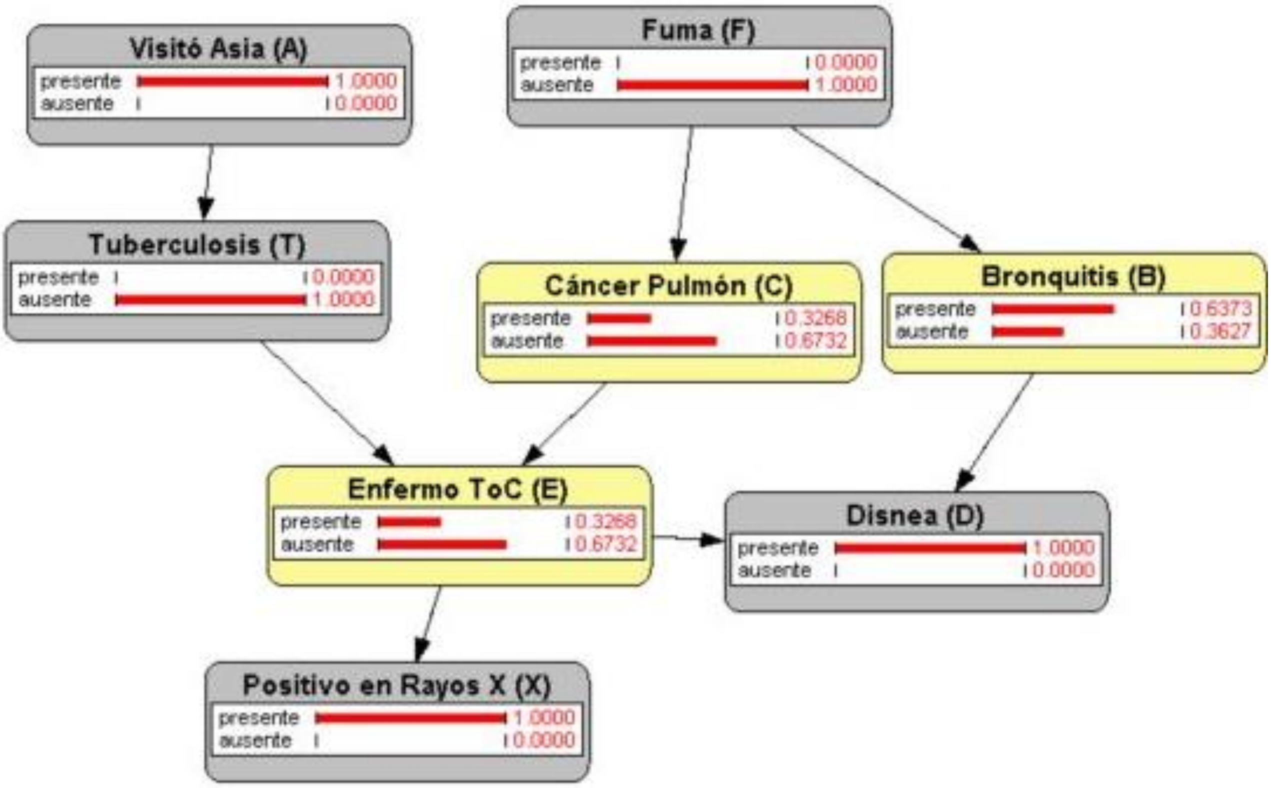
Obtenemos que el resultado es muy similar 0, 32688.Podemos atribuirle esta diferencia al número de decimales utilizados para realizar los cálculos. Aún podemos afirmar que el resultado es correcto.

En el primer paso del algoritmo de eliminación de variables hemos ignorado A-

A continuación, se muestra que el valor de A con P(-f) =1 no influye sobre la red Bayesiana.



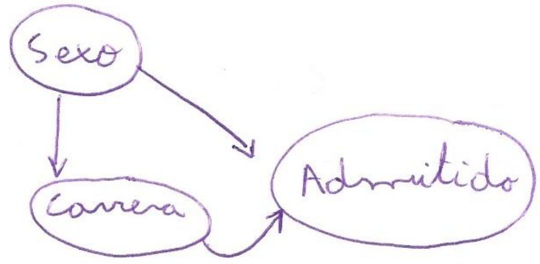
Visitó Asia (A)
presente 0.0000
ausente 1.0000
Fuma (F)
presente 0.0000
ausente 1.0000
Tuberculosis (T)
presente 0.0000
ausente 1.0000
Cáncer Pulmón (C)
presente 0.3268
ausente 0.6732
Bronquitis (B)
presente 0.6373
ausente 0.3627
Enfermo ToC (E)
presente 0.3268
ausente 0.6732
Disnea (D)
presente 1.0000
ausente 0.0000
Positivo en Rayos X (X)
presente 1.0000
ausente 0.0000



Visitó Asia (A)
presente 1.0000
ausente 0.0000
Fuma (F)
presente 0.0000
ausente 1.0000
Tuberculosis (T)
presente 0.0000
ausente 1.0000
Cáncer Pulmón (C)
presente 0.3268
ausente 0.6732
Bronquitis (B)
presente 0.6373
ausente 0.3627
Enfermo ToC (E)
presente 0.3268
ausente 0.6732
Disnea (D)
presente 1.0000
ausente 0.0000
Positivo en Rayos X (X)
presente 1.0000
ausente 0.0000

2.El fichero DatosBerkeley.xls contiene datos simulados de admisión de alumnos en Berkeley. Prueba a aprender la Red Bayesiana con OpenMarkov a partir de estos datos(menú Herramientas◊ Aprendizaje + Botón de opción “Aprendizaje Automático”)

A] Red Bayesiana



B] Tenemos que obtener:
$P(S)$, $P(A \mid S, C)$ y $P(C \mid S)$

Hombres = 2127
Mujeres = 1101
Total = 3228

1]

S	P(S)
M	2127/3228 = 0,658922
F	1101/3228 = 0,341078

2]

C	S	P(C|S)
A	M	825/2127 = 0,38787
B	M	560/2127 = 0,263281617
C	M	325/2127 = 0,1528
D	M	417/2127 = 0,196
A	F	108/1101 = 0,098
B	F	25/1101 = 0,0227
C	F	593/1101 = 0,5386
D	F	373/1101 = 0,338783

Probabilidad del sexo y probabilidad de la carrera dado el sexo.

3) P(A | S,C)

C	S	A	P(A | S,C)
A	M	+	0,62
A	M	−	0,38
A	F	+	0,824
A	F	−	0,176
B	M	+	0,63
B	M	−	0,37
B	F	+	0,68
B	F	−	0,32
C	M	+	0,36923
C	M	−	0,63077
C	F	+	0,34
C	F	−	0,66
D	M	+	0,3309
D	M	−	0,6691
D	F	+	0,345
D	F	−	0,655

Probabilidad de admisión dado el sexo y la carrera.

C	M	$P(C \mid S) \cdot P(S)$	$P(C, S)$
A	M	0,38787 · 0,658922	0,255376
B	M	0,263281617 · 0,658922	0,1735
C	M	0,1328 · 0,658922	0,1007
D	M	0,196 · 0,658922	0,129
A	F	0,098 · 0,341078	0,0334
B	F	0,0227 · 0,341078	0,0077424
C	F	0,5386 · 0,341078	0,1837
D	F	0,338783 · 0,341078	0,11555

C	P(C)
A	0,288976
B	0,18124
C	0,2844
D	0,24455

Probabilidad de cada carrera.

A	C	S	P(A | S,C) · P(C | S) = P(A,C | S)	P(A | S)
+	A	F	0,824 · 0,098 = 0,080752	
+	B	F	0,68 · 0,0227 = 0,015436	0,396192
+	C	F	0,34 · 0,5386 = 0,183124	
+	D	F	0,345 · 0,338783 = 0,11688	
−	A	~~M~~ F	0,176 · 0,098 = 0,017248	0,601888
−	B	~~M~~ F	0,32 · 0,0227 = 0,007264	
−	C	~~M~~ F	0,66 · 0,5386 = 0,355476	
−	D	~~M~~ F	0,655 · 0,338783 = 0,2219	
+	A	M	0,62 · 0,38787 = 0,24	
+	B	M	0,63 · 0,263281617 = 0,16587	0,3273
+	C	M	0,36923 · 0,1528 = 0,05642	
+	D	M	0,3309 · 0,196 = 0,065	
−	A	M	0,38 · 0,38787 = 0,1474	
−	B	M	0,37 · 0,263281617 = 0,097414	
−	C	M	0,63077 · 0,1528 = 0,0964	0,46961
−	D	M	0,655 · 0,196 = 0,1284	

A	S	P(A | S) · P(S) = P(A,S)	P(A)
+	M	0,5273 · 0,658922	0,48258
+	F	0,396192 · 0,341078	
−	M	0,46961 · 0,688922	0,5174
−	F	0,601888 · 0,341078	

Probabilidad de admisión.

c) Podemos apreciar que en las carreras B, C y D las probabilidades de admisión de hombres y mujeres son muy similares. Podemos deducir que en estas carreras no ha habido discriminación.

En cambio en la carrera A la probabilidad de aceptar mujeres es superior a la de aceptar hombres (0,824 > 0,62). En esta carrera si que podríamos hablar de un caso de discriminación de género.

d) calcular la probabilidad de admisión para una mujer

P(A | f)

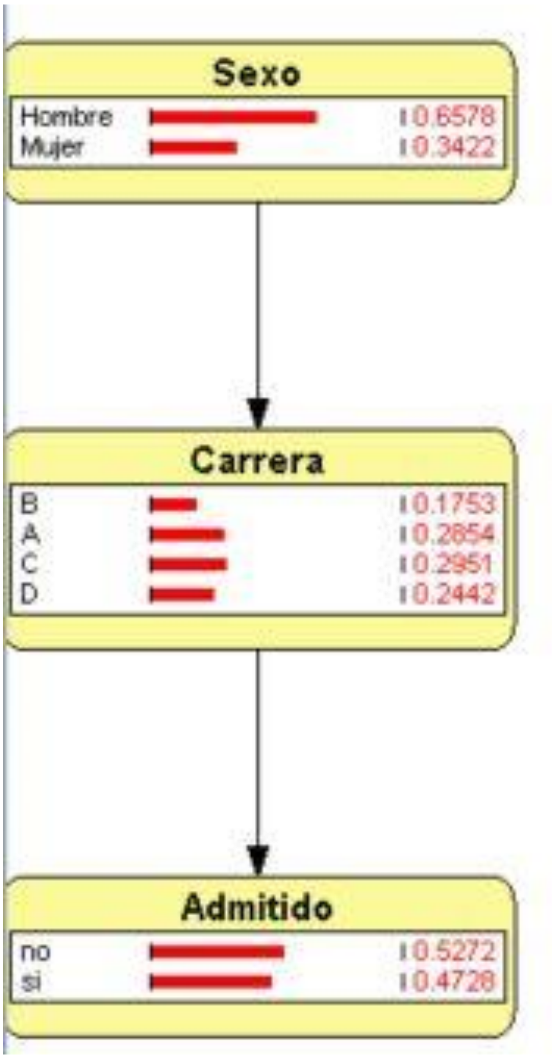
Tenemos P(A, C | S), P(C | S) y P(S).

Eliminamos C, solo nos interesan las P de mujeres.

A	C	S	P(A | f, C)	· P(C | f)	= P(A, C | f)	P(A | f)
+	A	f	0,824	· 0,098	= 0,080752	
+	B	f	0,68	· 0,0227	= 0,015436	
+	C	f	0,34	· 0,5386	= 0,183124	+ → 0,396192
+	D	f	0,345	· 0,338783	= 0,11688	
−	A	f	0,176	· 0,098	= 0,017248	
−	B	f	0,32	· 0,0227	= 0,007264	
−	C	f	0,66	· 0,5386	= 0,355476	+ → 0,601888
−	D	f	0,655	· 0,338783	= 0,2219	

Resultado → La probabilidad de admisión siendo mujer es de 0,396192.

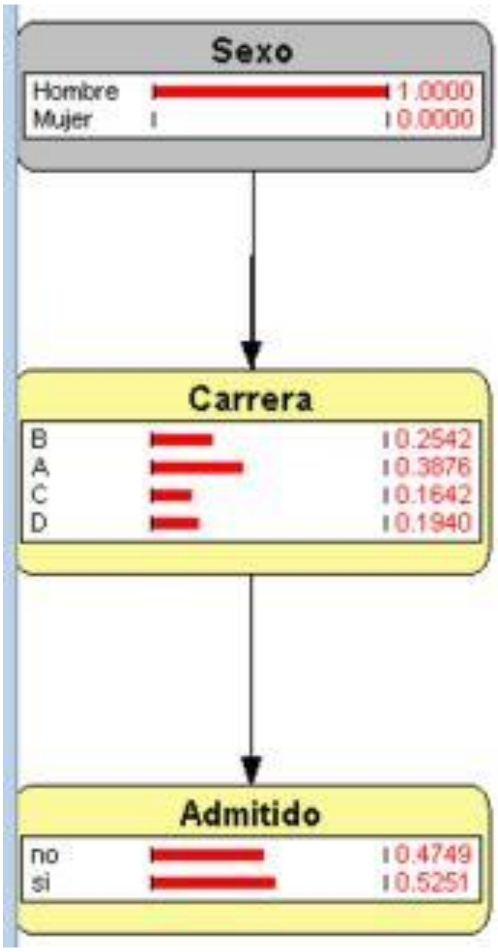
a) Comprueba si la red y las tablas de probabilidad son como habías previsto. Si no lo son, discute las diferencias.

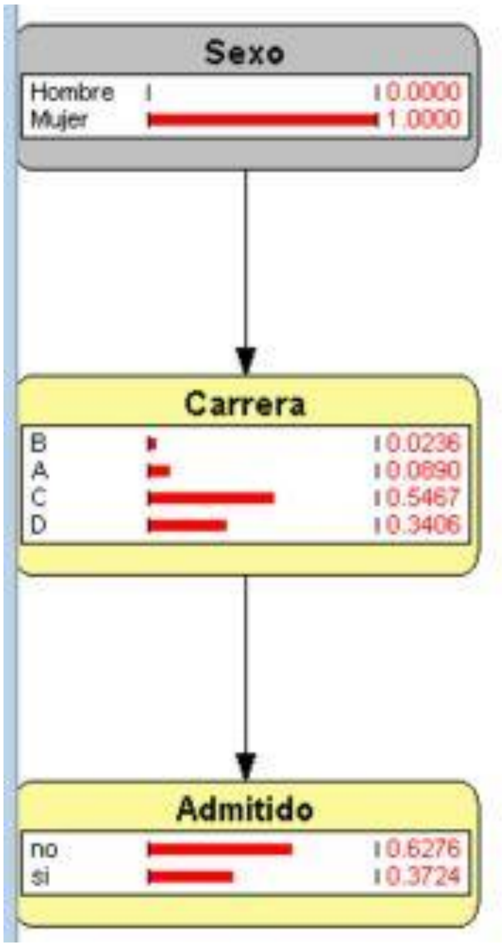


Las tablas de obtenidas son similares, aunque las probabilidades no son exactas ya que los datos de aprendizaje son mayores. En los cálculos hechos a manos había un total de 3228 solicitudes, en cambio en los datos de aprendizaje hay 5001 solicitudes.

b) Calcula la probabilidad de admisión para un hombre y para una mujer.

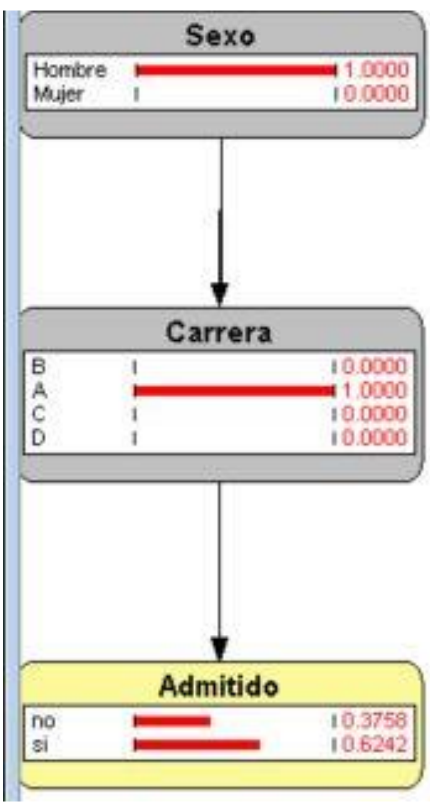
La probabilidad de admisión de una mujer y de un hombre es de 0,5251, para el hombre, y 0,3724 para el caso de la mujer.

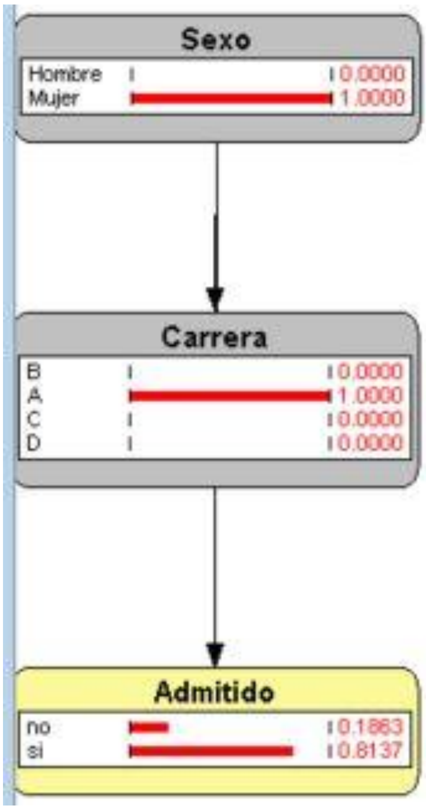




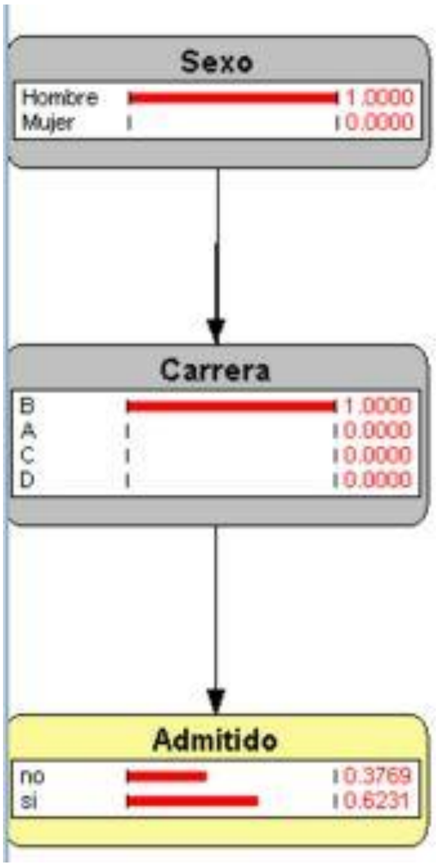
c)Si son distintas, ¿es porque hubo discriminación de género, o puede haber algún otro motivo?

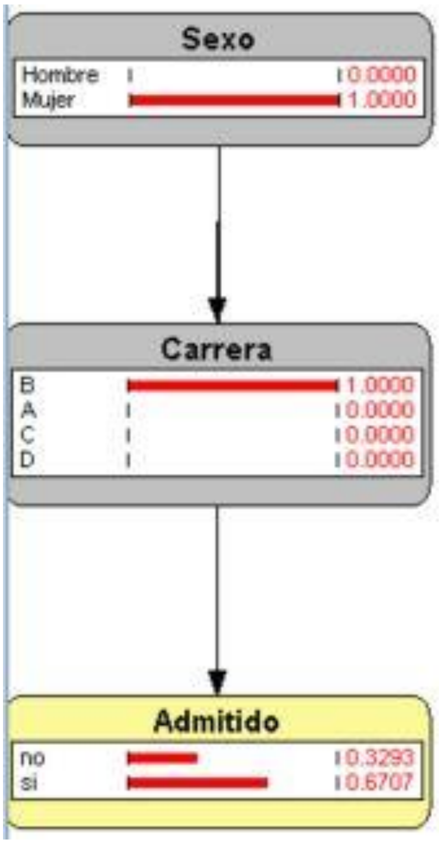
Dado estos resultados podríamos intuir que ha habido discriminación de género, pero si comprobamos la admisión en cada una de las carreras la situación cambia.

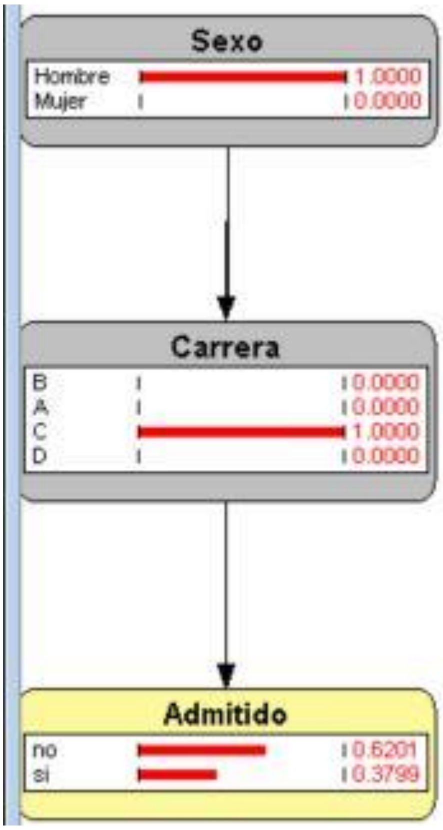


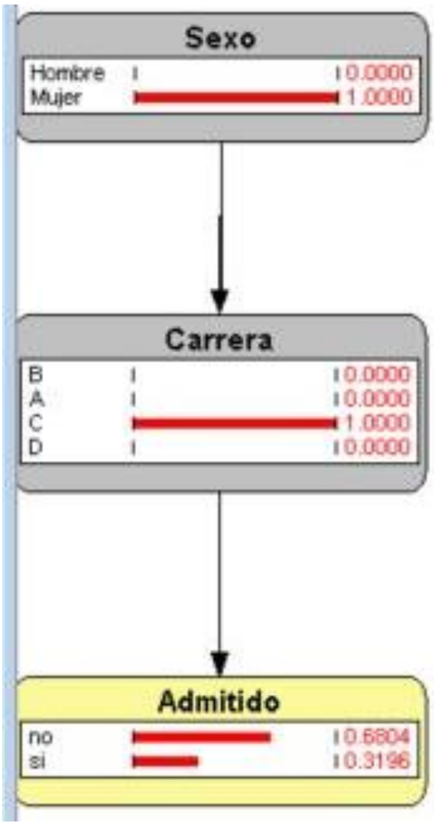


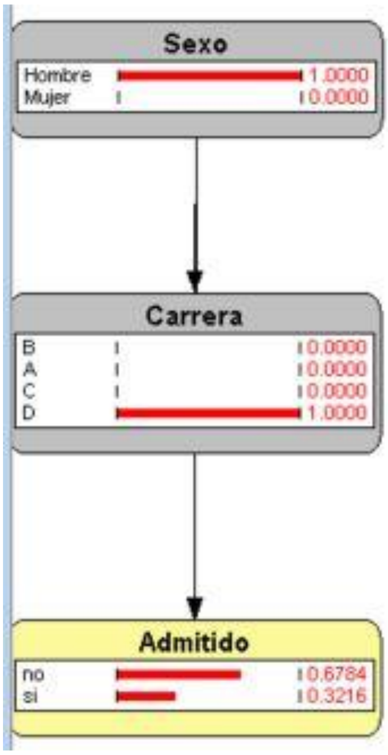
En la carrera A la probabilidad de admisión de las mujeres es superior a la de los hombres.

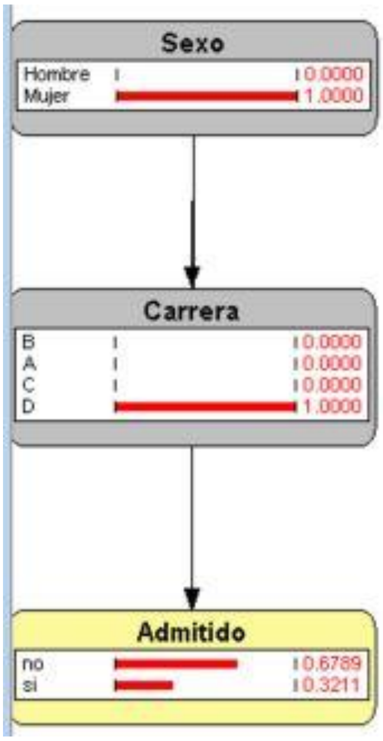












En el resto de carreras las probabilidades de admisión son muy similares.

Por lo tanto, podemos concluir que el hecho de que las probabilidades de admisión de hombres y mujeres sean distintas no es por discriminación de género.