

Tarea 4: Gráficos probabilísticos

Saul Ivan Rivas Vega

Aprendizaje Automatizado

8 de mayo de 2020

1. Ejercicio 1

En una planta nuclear hay una alarma que se activa cuando un indicador de temperatura excede un umbral. El indicador mide la temperatura del núcleo. Considera las variables booleanas A (alarma), D_A (alarma defectuosa), D_I (indicador defectuoso) y las variables enteras I (lectura del indicador) y T (temperatura real del núcleo).

- Dibuja 3 redes bayesianas válidas diferentes que capturen el comportamiento del proceso, entre ellas incluye aquella que capture el mayor número de independencias condicionales con el menor número de arcos. Discute el modelo representado por cada una de las redes.
- De las redes dibujadas, escribe sus distribuciones conjuntas en términos de sus probabilidades condicionales.
- Asumiendo que las variables enteras I y T pueden tomar un máximo de 100 valores, ¿Cuál sería el número de valores necesarios en cada nodo y el total en cada una de las redes?

1.1.

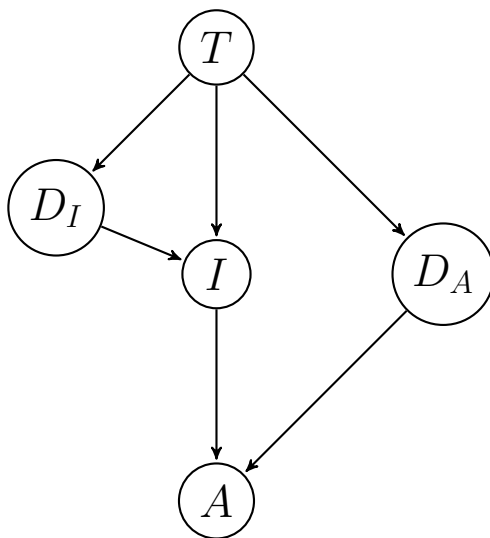


Figura 1: Red 1.

En esta red se interpreta que si la temperatura tenga cierto valor, influye en la probabilidad de que tanto el indicador como la alarma sean defectuosos esto es por la intuición de que quizá ambos funcionen perfectamente hasta que la temperatura del núcleo este en valores muy altos. De igual forma la temperatura influye directamente en el valor que este reportando el indicador. El indicador se ve influenciado naturalmente por la probabilidad de que este sea defectuoso. Para que la alarma este activa influye si es defectuosa y también

el valor que este reportando el indicador.

Su distribución conjunta en términos de sus probabilidades condicionales es:

$$P(T, D_I, I, D_A, A) = P(T)P(D_I|T)P(I|D_I, T)P(D_A|T)P(A|I, D_A) \quad (1)$$

$P(T)$					
0	1	2	3	...	99
$P(T = 0)$	$P(T = 1)$	$P(T = 2)$	$P(T = 3)$...	$P(T = 99)$

Cuadro 1: Tabla de valores para la variable \mathbf{T} , son **100**.

0

T	$P(D_I = 0 T)$	$P(D_I = 1 T)$
0	$P(D_I = 0 T = 0)$	$P(D_I = 1 T = 0)$
1	$P(D_I = 0 T = 1)$	$P(D_I = 1 T = 1)$
2	$P(D_I = 0 T = 2)$	$P(D_I = 1 T = 2)$
...
99	$P(D_I = 0 T = 99)$	$P(D_I = 1 T = 99)$

Cuadro 2: Tabla de valores para la variable $\mathbf{D_I}$, son **200** ($D_I \times T = 2 \times 100$).

T	$P(D_A = 0 T)$	$P(D_A = 1 T)$
0	$P(D_A = 0 T = 0)$	$P(D_A = 1 T = 0)$
1	$P(D_A = 0 T = 1)$	$P(D_A = 1 T = 1)$
2	$P(D_A = 0 T = 2)$	$P(D_A = 1 T = 2)$
...
99	$P(D_A = 0 T = 99)$	$P(D_A = 1 T = 99)$

Cuadro 3: Tabla de valores para la variable $\mathbf{D_A}$, son **200** ($D_A \times T = 2 \times 100$).

D_I	T	$P(I = 0 T, D_I)$	$P(I = 1 T, D_I)$...	$P(I = 99 T, D_I)$
0	0	$P(I = 0 T = 0, D_I = 0)$	$P(I = 1 T = 0, D_I = 0)$...	$P(I = 99 T = 0, D_I = 0)$
0	1	$P(I = 0 T = 1, D_I = 0)$	$P(I = 1 T = 1, D_I = 0)$...	$P(I = 99 T = 1, D_I = 0)$
0	2	$P(I = 0 T = 2, D_I = 0)$	$P(I = 1 T = 2, D_I = 0)$...	$P(I = 99 T = 2, D_I = 0)$
0
0	99	$P(I = 0 T = 99, D_I = 0)$	$P(I = 1 T = 99, D_I = 0)$...	$P(I = 99 T = 99, D_I = 0)$
1	0	$P(I = 0 T = 0, D_I = 1)$	$P(I = 1 T = 0, D_I = 1)$...	$P(I = 99 T = 0, D_I = 1)$
1	1	$P(I = 0 T = 1, D_I = 1)$	$P(I = 1 T = 1, D_I = 1)$...	$P(I = 99 T = 1, D_I = 1)$
1	2	$P(I = 0 T = 2, D_I = 1)$	$P(I = 1 T = 2, D_I = 1)$...	$P(I = 99 T = 2, D_I = 1)$
1
1	99	$P(I = 0 T = 99, D_I = 1)$	$P(I = 1 T = 99, D_I = 1)$...	$P(I = 99 T = 99, D_I = 1)$

Cuadro 4: Tabla de valores para la variable \mathbf{I} , son **20000** ($D_I \times T \times I = 2 \times 100 \times 100$).

D_A	I	$P(A = 0 I, D_A)$	$P(A = 1 I, D_A)$
0	0	$P(A = 0 I = 0, D_A = 0)$	$P(A = 1 I = 0, D_A = 0)$
0	1	$P(A = 0 I = 1, D_A = 0)$	$P(A = 1 I = 1, D_A = 0)$
0	2	$P(A = 0 I = 2, D_A = 0)$	$P(A = 1 I = 2, D_A = 0)$
0
0	99	$P(A = 0 I = 99, D_A = 0)$	$P(A = 1 I = 99, D_A = 0)$
1	0	$P(A = 0 I = 0, D_A = 1)$	$P(A = 1 I = 0, D_A = 1)$
1	1	$P(A = 0 I = 1, D_A = 1)$	$P(A = 1 I = 1, D_A = 1)$
1	2	$P(A = 0 I = 2, D_A = 1)$	$P(A = 1 I = 2, D_A = 1)$
1
1	99	$P(A = 0 I = 99, D_A = 1)$	$P(A = 1 I = 99, D_A = 1)$

Cuadro 5: Tabla de valores para la variable \mathbf{A} , son **400** ($A \times I \times D_A = 2 \times 100 \times 2$).

Finalmente tomando en cuenta los valores obtenidos de las 5 tablas obtenemos el total de valores en la red:

$$100 + 200 + 200 + 20000 + 400 = 20900 \quad (2)$$

1.2.

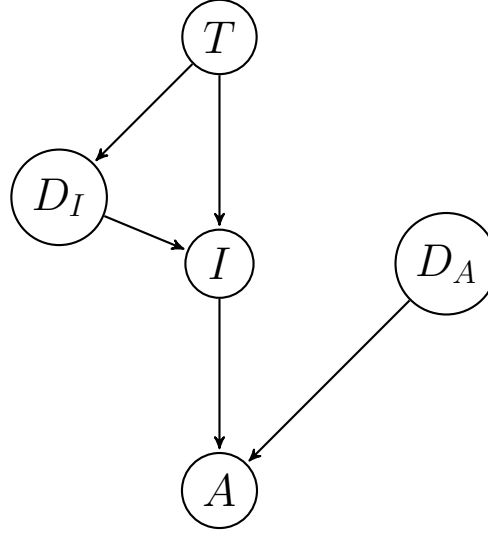


Figura 2: Red 2.

En esta red se interpreta prácticamente lo mismo que en el anterior sin embargo se busca introducir mas independencias condicionales y minimizar los arcos sin interferir demasiado en la intuición de la red original. Aquí la temperatura deja de influir para el evento de que la alarma sea defectuosa. Un ejemplo de independencia condicional que se agrega es $D_I \perp\!\!\!\perp D_A | I$, porque ahora los caminos de D_I a D_A son: $\{D_I, T, I, A, D_A\}$ y $\{D_I, I, A, D_A\}$ y tienen las tripletas $[\{D_I, T, I\}, \{T, I, A\}, \{I, A, D_A\}]$ y $[\{D_I, I, A\}, \{I, A, D_A\}]$ respectivamente. Y dado I la tripleta del primer camino $\{T, I, A\}$ y la tripleta del segundo camino $\{D_I, I, A\}$ son inactivas ya que ambas son cadenas causales con la variable observada en medio, entonces ningún camino es activo y por lo tanto D_I es condicionalmente independiente de D_A dado I .

Su distribución conjunta en términos de sus probabilidades condicionales es:

$$P(T, D_I, I, D_A, A) = P(T)P(D_A)P(D_I|T)P(I|D_I, T)P(A|I, D_A) \quad (3)$$

$P(T)$					
0	1	2	3	...	99
$P(T = 0)$	$P(T = 1)$	$P(T = 2)$	$P(T = 3)$...	$P(T = 99)$

Cuadro 6: Tabla de valores para la variable T , son **100**.

T	$P(D_I = 0 T)$	$P(D_I = 1 T)$
0	$P(D_I = 0 T = 0)$	$P(D_I = 1 T = 0)$
1	$P(D_I = 0 T = 1)$	$P(D_I = 1 T = 1)$
2	$P(D_I = 0 T = 2)$	$P(D_I = 1 T = 2)$
...
99	$P(D_I = 0 T = 99)$	$P(D_I = 1 T = 99)$

Cuadro 7: Tabla de valores para la variable D_I , son **200** ($D_I \times T = 2 \times 100$).

$P(D_A)$	
0	1
$P(D_A = 0)$	$P(D_A = 1)$

Cuadro 8: Tabla de valores para la variable D_A , son **2**.

D_I	T	$P(I = 0 T, D_I)$	$P(I = 1 T, D_I)$...	$P(I = 99 T, D_I)$
0	0	$P(I = 0 T = 0, D_I = 0)$	$P(I = 1 T = 0, D_I = 0)$...	$P(I = 99 T = 0, D_I = 0)$
0	1	$P(I = 0 T = 1, D_I = 0)$	$P(I = 1 T = 1, D_I = 0)$...	$P(I = 99 T = 1, D_I = 0)$
0	2	$P(I = 0 T = 2, D_I = 0)$	$P(I = 1 T = 2, D_I = 0)$...	$P(I = 99 T = 2, D_I = 0)$
0
0	99	$P(I = 0 T = 99, D_I = 0)$	$P(I = 1 T = 99, D_I = 0)$...	$P(I = 99 T = 99, D_I = 0)$
1	0	$P(I = 0 T = 0, D_I = 1)$	$P(I = 1 T = 0, D_I = 1)$...	$P(I = 99 T = 0, D_I = 1)$
1	1	$P(I = 0 T = 1, D_I = 1)$	$P(I = 1 T = 1, D_I = 1)$...	$P(I = 99 T = 1, D_I = 1)$
1	2	$P(I = 0 T = 2, D_I = 1)$	$P(I = 1 T = 2, D_I = 1)$...	$P(I = 99 T = 2, D_I = 1)$
1
1	99	$P(I = 0 T = 99, D_I = 1)$	$P(I = 1 T = 99, D_I = 1)$...	$P(I = 99 T = 99, D_I = 1)$

Cuadro 9: Tabla de valores para la variable I , son **20000** ($D_I \times T \times I = 2 \times 100 \times 100$).

Finalmente tomando en cuenta los valores obtenidos de las 5 tablas obtenemos el total de valores en la red:

$$100 + 200 + 2 + 20000 + 400 = 20702 \quad (4)$$

D_A	I	$P(A = 0 I, D_A)$	$P(A = 1 I, D_A)$
0	0	$P(A = 0 I = 0, D_A = 0)$	$P(A = 1 I = 0, D_A = 0)$
0	1	$P(A = 0 I = 1, D_A = 0)$	$P(A = 1 I = 1, D_A = 0)$
0	2	$P(A = 0 I = 2, D_A = 0)$	$P(A = 1 I = 2, D_A = 0)$
0
0	99	$P(A = 0 I = 99, D_A = 0)$	$P(A = 1 I = 99, D_A = 0)$
1	0	$P(A = 0 I = 0, D_A = 1)$	$P(A = 1 I = 0, D_A = 1)$
1	1	$P(A = 0 I = 1, D_A = 1)$	$P(A = 1 I = 1, D_A = 1)$
1	2	$P(A = 0 I = 2, D_A = 1)$	$P(A = 1 I = 2, D_A = 1)$
1
1	99	$P(A = 0 I = 99, D_A = 1)$	$P(A = 1 I = 99, D_A = 1)$

Cuadro 10: Tabla de valores para la variable \mathbf{A} , son **400** ($A \times I \times D_A = 2 \times 100 \times 2$).

1.3.

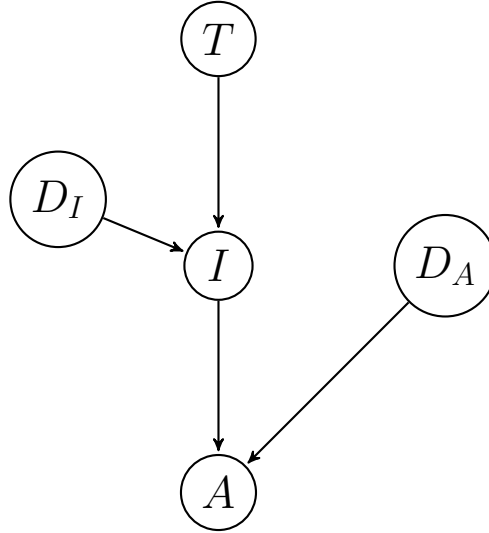


Figura 3: Red 3.

Finalmente para esta tercer red llegamos a una de las posibilidades de la red inicial con menor numero de arcos y mayor numero de independencias condicionales al quitar la influencia de la temperatura en el evento de que el indicador sea defectuoso. Un ejemplo de independencia condicional que se agrega es $T \perp\!\!\!\perp A|I$, porque ahora el camino de T a A es uno inactivo dado I ya que es una cadena causal con la variable observada en medio, y por ser el único todos los caminos de T a A están inactivos y por lo tanto T es condicionalmente independiente de A dado I .

Su distribución conjunta en términos de sus probabilidades condicionales es:

$$P(T, D_I, I, D_A, A) = P(T)P(D_A)P(D_I)P(I|D_I, T)P(A|I, D_A) \quad (5)$$

$P(T)$					
0	1	2	3	...	99
$P(T = 0)$	$P(T = 1)$	$P(T = 2)$	$P(T = 3)$...	$P(T = 99)$

Cuadro 11: Tabla de valores para la variable \mathbf{T} , son **100**.
0

$P(D_I)$	
0	1
$P(D_I = 0)$	$P(D_I = 1)$

Cuadro 12: Tabla de valores para la variable $\mathbf{D_I}$, son **2** ($D_I \times T = 2 \times 100$).

$P(D_A)$	
0	1
$P(D_A = 0)$	$P(D_A = 1)$

Cuadro 13: Tabla de valores para la variable $\mathbf{D_A}$, son **2**.

D_I	T	$P(I = 0 T, D_I)$	$P(I = 1 T, D_I)$...	$P(I = 99 T, D_I)$
0	0	$P(I = 0 T = 0, D_I = 0)$	$P(I = 1 T = 0, D_I = 0)$...	$P(I = 99 T = 0, D_I = 0)$
0	1	$P(I = 0 T = 1, D_I = 0)$	$P(I = 1 T = 1, D_I = 0)$...	$P(I = 99 T = 1, D_I = 0)$
0	2	$P(I = 0 T = 2, D_I = 0)$	$P(I = 1 T = 2, D_I = 0)$...	$P(I = 99 T = 2, D_I = 0)$
0
0	99	$P(I = 0 T = 99, D_I = 0)$	$P(I = 1 T = 99, D_I = 0)$...	$P(I = 99 T = 99, D_I = 0)$
1	0	$P(I = 0 T = 0, D_I = 1)$	$P(I = 1 T = 0, D_I = 1)$...	$P(I = 99 T = 0, D_I = 1)$
1	1	$P(I = 0 T = 1, D_I = 1)$	$P(I = 1 T = 1, D_I = 1)$...	$P(I = 99 T = 1, D_I = 1)$
1	2	$P(I = 0 T = 2, D_I = 1)$	$P(I = 1 T = 2, D_I = 1)$...	$P(I = 99 T = 2, D_I = 1)$
1
1	99	$P(I = 0 T = 99, D_I = 1)$	$P(I = 1 T = 99, D_I = 1)$...	$P(I = 99 T = 99, D_I = 1)$

Cuadro 14: Tabla de valores para la variable \mathbf{I} , son **20000** ($D_I \times T \times I = 2 \times 100 \times 100$).

D_A	I	$P(A = 0 I, D_A)$	$P(A = 1 I, D_A)$
0	0	$P(A = 0 I = 0, D_A = 0)$	$P(A = 1 I = 0, D_A = 0)$
0	1	$P(A = 0 I = 1, D_A = 0)$	$P(A = 1 I = 1, D_A = 0)$
0	2	$P(A = 0 I = 2, D_A = 0)$	$P(A = 1 I = 2, D_A = 0)$
0
0	99	$P(A = 0 I = 99, D_A = 0)$	$P(A = 1 I = 99, D_A = 0)$
1	0	$P(A = 0 I = 0, D_A = 1)$	$P(A = 1 I = 0, D_A = 1)$
1	1	$P(A = 0 I = 1, D_A = 1)$	$P(A = 1 I = 1, D_A = 1)$
1	2	$P(A = 0 I = 2, D_A = 1)$	$P(A = 1 I = 2, D_A = 1)$
1
1	99	$P(A = 0 I = 99, D_A = 1)$	$P(A = 1 I = 99, D_A = 1)$

Cuadro 15: Tabla de valores para la variable **A**, son **400** ($A \times I \times D_A = 2 \times 100 \times 2$).

Finalmente tomando en cuenta los valores obtenidos de las 5 tablas obtenemos el total de valores en la red:

$$100 + 2 + 2 + 20000 + 400 = 20504 \quad (6)$$

2. Ejercicio 2

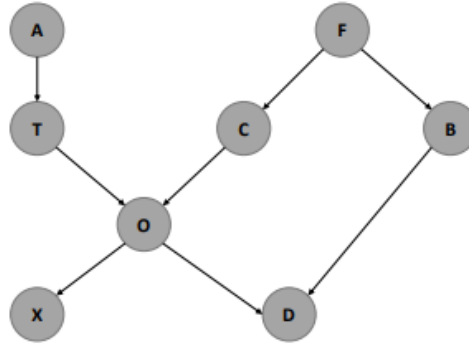


Figura 4: Modelo gráfico de diagnóstico de enfermedades de pulmón.

A una clínica le concierne el diagnóstico de enfermedades de pulmón. Como se puede ver en el modelo de la Figura 4, una visita a Asia (A) hace que la probabilidad de tener tuberculosis (T) aumente. Los nodos en la gráfica tienen el siguiente significado: Di si las

Etiqueta	Significado
A	Visita a Asia
T	Tuberculosis
C	Cáncer de pulmón
O	Tuberculosis o Cáncer de pulmón
F	Fumador
X	Prueba de rayos X positiva
B	Bronquitis
D	Disnea

siguientes relaciones de independencia condicional son verdaderas o falsas y explica por qué:

2.1. $T \perp\!\!\!\perp F|D$

De T a F existen 2 caminos de nodos:

1. $\{T, O, C, F\}$
2. $\{T, O, D, B, F\}$

En el Camino 1 existen las siguientes tripletas $[\{T, O, C\}, \{O, C, F\}]$, la tripleta $\{T, O, C\}$ esta activa ya que es un efecto común donde uno de los descendientes es la variable observada (en este caso nuestra variable observada D es descendiente de O). La otra tripleta $\{O, C, F\}$ es una cadena causal pero ninguna de sus variables es observada, así decimos que esta activa. Como las 2 tripletas del camino 1 fueron activas, el camino 1 es activo.

Como hay al menos un camino activo entre T y F dado D decimos que no se garantiza la independencia condicional y por lo tanto es **falso** afirmar que $T \perp\!\!\!\perp F|D$ solo con los datos de la gráfica.

2.2. $C \perp\!\!\!\perp B|F$

De C a B existen 2 caminos de nodos:

1. $\{C, F, B\}$
2. $\{C, O, D, B\}$

En el Camino 1 solo existe la tripleta $\{C, F, B\}$, la cual es inactiva ya que es una causa común donde la variable de en medio F si es observada. Como no todas las tripletas del camino 1 fueron activas decimos que el camino es inactivo. En el Camino 2 existen las siguientes tripletas $[\{C, O, D\}, \{O, D, B\}]$, la tripleta $\{C, O, D\}$ esta activa ya que es una cadena causal donde ninguna variable es observada. La otra tripleta $\{O, D, B\}$ es un efecto común pero ninguna de sus variables es observada, así decimos que es inactiva. Como no todas las tripletas del camino 1 fueron activas decimos que el camino es inactivo.

Finalmente como no hay un solo camino activo entre C y B dado F se garantiza la independencia condicional y por lo tanto es **verdadero** afirmar que $C \perp\!\!\!\perp B|F$ solo con los datos de la gráfica.

2.3. $A \perp\!\!\!\perp F|C$

De A a F existen 2 caminos de nodos:

1. $\{A, T, O, D, B, F\}$
2. $\{A, T, O, C, F\}$

En el Camino 1 existen las siguientes tripletas $[\{A, T, O\}, \{O, D, B\}, \{D, B, F\}]$. La tripleta $\{O, D, B\}$ esta inactiva ya que es un efecto común donde ninguna variable es observada. Como no todas las tripletas del camino 1 son activas decimos que el camino es inactivo.

En el Camino 2 existen las siguientes tripletas $[\{A, T, O\}, \{O, C, F\}]$. La tripleta $\{O, C, F\}$ esta inactiva ya que es una cadena causal donde la variable de en medio C es observada. Como no todas las tripletas del camino 2 son activas decimos que el camino es inactivo.

Finalmente como no hay un solo camino activo entre A y F dado C se garantiza la independencia condicional y por lo tanto es **verdadero** afirmar que $A \perp\!\!\!\perp F|C$ solo con los datos de la gráfica.

2.4. $A \perp\!\!\!\perp F|C, D$

De A a F existen 2 caminos de nodos:

1. $\{A, T, O, D, B, F\}$
2. $\{A, T, O, C, F\}$

En el Camino 1 existen las siguientes tripletas $[\{A, T, O\}, \{O, D, B\}, \{D, B, F\}]$. La tripleta $\{A, T, O\}$ es activa ya que es una cadena causal donde ninguna variable es observada. La tripleta $\{T, O, D\}$ es activa ya que es una cadena causal donde ninguna variable es observada. La tripleta $\{O, D, B\}$ es activa ya que es un efecto común donde la variable de en medio D es observada. La tripleta $\{D, B, F\}$ es activa ya que es una cadena causal que a pesar de que D es observada, esta no es la variable de en medio, en este caso la variable de en medio es B y no es observada. Como todas las tripletas del camino 1 son activas decimos que el camino es activo.

Como hay al menos un camino activo entre A y F dado C, D decimos que no se garantiza la independencia condicional y por lo tanto es **falso** afirmar que $A \perp\!\!\!\perp F | C, D$ solo con los datos de la gráfica.

3. Ejercicio 3

Imagina una clínica que ayuda a pacientes con ébola en un área afectada por el virus. La red de la Figura 5 intenta capturar la dinámica por la cual las personas que sufren los síntomas pueden llegar a esta clínica y ver a un especialista. Existe la posibilidad de que alguien con ébola ($E = \text{verdadero}$) muestre síntomas, por ejemplo sangrado ($S = \text{verdadero}$), fiebre ($F = \text{verdadero}$) y visite la clínica ($V = \text{verdadero}$). El sangrado incrementa el riesgo de complicaciones ($C = \text{verdadero}$) y la persona puede ser llevada a ver un doctor especialista ($D = \text{verdadero}$).

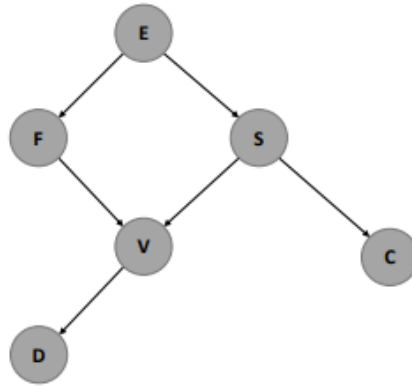


Figura 5: Modelo gráfico de detección de ébola.

- $P(E = \text{verdadero}) = 0.01$
- $P(F = \text{verdadero} | E = \text{verdadero}) = 0.6$
- $P(F = \text{verdadero} | E = \text{falso}) = 0.1$
- $P(S = \text{verdadero} | E = \text{verdadero}) = 0.8$
- $P(S = \text{verdadero} | E = \text{falso}) = 0.05$
- $P(V = \text{verdadero} | F = \text{verdadero}, S = \text{verdadero}) = 0.8$
- $P(V = \text{verdadero} | F = \text{verdadero}, S = \text{falso}) = 0.5$
- $P(V = \text{verdadero} | F = \text{falso}, S = \text{verdadero}) = 0.7$
- $P(V = \text{verdadero} | F = \text{falso}, S = \text{falso}) = 0.0$
- $P(C = \text{verdadero} | S = \text{verdadero}) = 0.75$
- $P(C = \text{verdadero} | S = \text{falso}) = 0.1$
- $P(D = \text{verdadero} | V = \text{verdadero}) = 0.6$
- $P(D = \text{verdadero} | V = \text{falso}) = 0.0$

Con estas probabilidades y el modelo gráfico de la Figura 5, realiza lo siguiente:

3.1.

Escribe la distribución conjunta de la red bayesiana en función de las probabilidades condicionales.

3.2.

Si un paciente es llevado al doctor ($D = \text{verdadero}$), usando un paquete de software calcula la probabilidad de que no tenga ébola ($P(E = \text{falso} | D = \text{verdadero})$).

3.3.

Convierte a la red bayesiana en un modelo gráfico no dirigido (campo aleatorio de Markov) y dibújalo. Captura tantas relaciones de independencia condicional como sea posible.

3.4.

Debido a una campaña de concientización de la salud, las personas son alentadas a visitar la clínica en caso de que tengan fiebre. Esto incrementa la cantidad de visitas de personas con fiebre sin importar el estado de cualquier otra variable.

3.4.1.

¿Qué probabilidades condicionales en la red se modifican debido a este cambio y en qué sentido?

3.4.2.

Describe cualquier efecto que esto tenga en la proporción de personas con complicaciones que visiten la clínica. Menciona exactamente qué probabilidades condicionales usaste para llegar a tu conclusión.

3.5.

Asume que alguien que no tiene fiebre va al doctor, ¿qué relación de independencia condicional existe en la distribución que no puede ser descubierta a través del grafo solamente?

3.6. Opcional

Calcula $P(V|E = \text{verdadero})$ usando el método de eliminación de variables, describiendo de forma detallada el procedimiento que seguiste (2 puntos extras).