

Razonamiento con Incertidumbre

Javier Béjar

Inteligencia Artificial - 2021/2022 1Q

CS - GEI- FIB



Razonamiento con incertidumbre

- ⊙ Todos los mecanismos de representación de conocimiento vistos están basados en la lógica bajo estos supuestos:
 - Todo hecho sobre el que razonemos debe poder ser evaluado como cierto o falso
 - Para poder razonar necesitamos tener todos los hechos a nuestra disposición

- ⊙ Pero en la práctica nos encontramos con estos problemas
 - Representar el conocimiento para cubrir todos los hechos que son relevantes para un problema es difícil
 - Existen dominios en los que se desconocen todos los hechos y reglas necesarias para resolver el problema
 - Existen problemas en los que aún teniendo las reglas para resolverlos no disponemos de toda la información necesaria o no tenemos confianza absoluta en ellas
 - En otros problemas la reglas no se aplican siempre o su confianza cambia con la confianza que tenemos en los hechos

Modelos Probabilistas

- ⊙ Los modelos probabilistas se basan en la **teoría de la probabilidad**
- ⊙ Las probabilidades se utilizan para **modelizar nuestra creencia** sobre los posibles valores que pueden tomar los hechos
- ⊙ Cada **hecho tendrá una distribución de probabilidad** asociada que nos permitirá tomar decisiones
- ⊙ La probabilidad de un hecho podrá ser **modificada por nuestra creencia** en otros hechos que estén relacionados

¿Cogerás un paraguas mañana?

Vivo en Barcelona, no llueve nunca

La previsión es que haya nubes

Igual sí, igual no

Hoy el suelo esta mojado

Pues mejor que lo coja

¿Cogerás un paraguas mañana?

Vivo en Barcelona, no llueve nunca

La previsión es que haya nubes

Igual sí, igual no

Hoy el suelo esta mojado

Pues mejor que lo coja

¿Cogerás un paraguas mañana?

Vivo en Barcelona, no llueve nunca

La previsión es que haya nubes

Igual sí, igual no

Hoy el suelo esta mojado

Pues mejor que lo coja

¿Cogerás un paraguas mañana?

Vivo en Barcelona, no llueve nunca

La previsión es que haya nubes

Igual sí, igual no

Hoy el suelo esta mojado

Pues mejor que lo coja

¿Cogerás un paraguas mañana?

Vivo en Barcelona, no llueve nunca

La previsión es que haya nubes

Igual sí, igual no

Hoy el suelo esta mojado

Pues mejor que lo coja

¿Cogerás un paraguas mañana?

Vivo en Barcelona, no llueve nunca

La previsión es que haya nubes

Igual sí, igual no

Hoy el suelo esta mojado

Pues mejor que lo coja

- ⊙ El elemento básico de teoría de probabilidades es la **variable aleatoria**
- ⊙ Una variable aleatoria tiene un dominio de valores, podemos tener variables aleatorias **booleanas**, **discretas** o **continuas**.
- ⊙ Definiremos una **proposición lógica** como cualquier fórmula en lógica de enunciados o predicados
- ⊙ Una **proposición lógica** **tendrá asociada** una **variable aleatoria** que indicará nuestro grado de creencia en ella

- ⊙ Una variable aleatoria tendrá asociada una **distribución de probabilidad**
- ⊙ La forma de expresar esta distribución de probabilidad dependerá del tipo de variable aleatoria (Discretas: Binomial, Multinomial, ..., Continuas: Normal, χ^2 , ...)
- ⊙ Nosotros trabajaremos sólo con variables aleatorias discretas
- ⊙ La unión de variables aleatorias se puede describir mediante una **distribución de probabilidad conjunta**

- ⊙ Denotaremos como $P(a)$ la probabilidad de que la proposición (variable aleatoria) A tenga el valor a .

Por ejemplo, la proposición $Fumar$ puede tener los valores $\{fumar, \neg fumar\}$, $P(\neg fumar)$ es la probabilidad de la proposición $Fumar = \neg fumar$

- ⊙ Denotaremos como $P(A)$ al **vector de probabilidades** de todos los posibles valores de la proposición A

- Definiremos como **probabilidad a priori** ($P(a)$) asociada a una proposición como el **grado de creencia en ella a falta de otra información**
- Definiremos como **probabilidad a posteriori** o **condicional** ($P(a|b)$) como el **grado de creencia** en una proposición **tras la observación de proposiciones asociadas** a ella
- La probabilidad a posteriori se puede definir a partir de probabilidades a priori como:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

- Esta fórmula transforma en la **regla del producto**:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

Los axiomas de la probabilidad serán el marco que restringirá las cosas que podremos creer y deducir

- ⊙ Toda probabilidad esta en el intervalo $[0, 1]$

$$0 \leq P(a) \leq 1$$

- ⊙ La proposición *cierto* tiene probabilidad 1 y la proposición *falso* tiene probabilidad 0

$$P(\text{cierto}) = 1 \quad P(\text{falso}) = 0$$

- ⊙ La probabilidad de la disyunción se obtiene mediante la fórmula

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

- ⊙ **Marginalización:** Probabilidad de una proposición atómica con independencia de los valores del resto de proposiciones

$$P(Y) = \sum_z P(Y, z)$$

- ⊙ **Probabilidades condicionadas:** Probabilidad de una proposición dados unos valores para algunas proposiciones e independiente del resto de proposiciones (a partir de la regla del producto)

$$P(X|e) = \alpha \sum_y P(X, e, y)$$

El valor α es un factor de normalización que corresponde a factores comunes que hacen que las probabilidades sumen 1

Consideremos un problema en el que intervengan las proposiciones:

$$Fumador = \{fumador, \neg fumador\}$$

$$Sexo = \{varon, mujer\}$$

$$Enfisema = \{enfisema, \neg enfisema\}$$

	<i>enfisema</i>		$\neg enfisema$	
	<i>varon</i>	<i>mujer</i>	<i>varon</i>	<i>mujer</i>
<i>fumador</i>	0.2	0.1	0.05	0.05
$\neg fumador$	0.02	0.02	0.23	0.33

$$P(enfisema \wedge varon) = 0,2 + 0,02$$

$$P(fumador \vee mujer) = 0,2 + 0,1 + 0,05 + 0,05 + 0,02 + 0,33$$

$$\begin{aligned} P(Fumador|enfisema) &= \langle P(fumador, enfisema, varon) \\ &\quad + P(fumador, enfisema, mujer), \\ &\quad P(\neg fumador, enfisema, varon) \\ &\quad + P(\neg fumador, enfisema, mujer) \rangle \\ &= \alpha \langle 0,3, 0,04 \rangle \\ &= \langle 0,88, 0,12 \rangle \end{aligned}$$

- ⊙ Hacer estos procesos de inferencia requiere almacenar y recorrer la distribución de probabilidad conjunta de todas las proposiciones
- ⊙ Suponiendo proposiciones binarias el coste en espacio y tiempo es $O(2^n)$ siendo n el número de proposiciones
- ⊙ Para cualquier problema real estas condiciones son impracticables
- ⊙ Necesitamos mecanismos que nos simplifiquen el coste del razonamiento

- ⊙ Por lo general no todas las proposiciones que aparecen en un problema están relacionadas entre si
- ⊙ Muestran la propiedad que denominaremos **independencia probabilística**
- ⊙ Esto quiere decir que unas proposiciones no influyen en las otras y por lo tanto podemos reescribir sus probabilidades como:

$$P(X|Y) = P(X); \quad P(Y|X) = P(Y); \quad P(X, Y) = P(X)P(Y)$$

- ⊙ Dadas estas propiedades podremos reescribir las probabilidades conjuntas de manera mas compacta reduciendo la complejidad

- ⊙ Hemos enunciado la regla del producto como:

$$P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$$

- ⊙ Esto nos lleva a lo que denominaremos la **regla de Bayes**

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- ⊙ Esta regla y la propiedad de independencia serán el fundamento del razonamiento probabilístico y nos permitirá relacionar las probabilidades de unas evidencias con otras

- ⊙ Suponiendo que podemos estimar exhaustivamente todas las probabilidades que involucran todos los valores de la variable Y podemos reescribir la formula de Bayes como:

$$P(Y|X) = \alpha P(X|Y)P(Y)$$

- ⊙ Suponiendo independencia condicional entre dos variables podremos escribir:

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

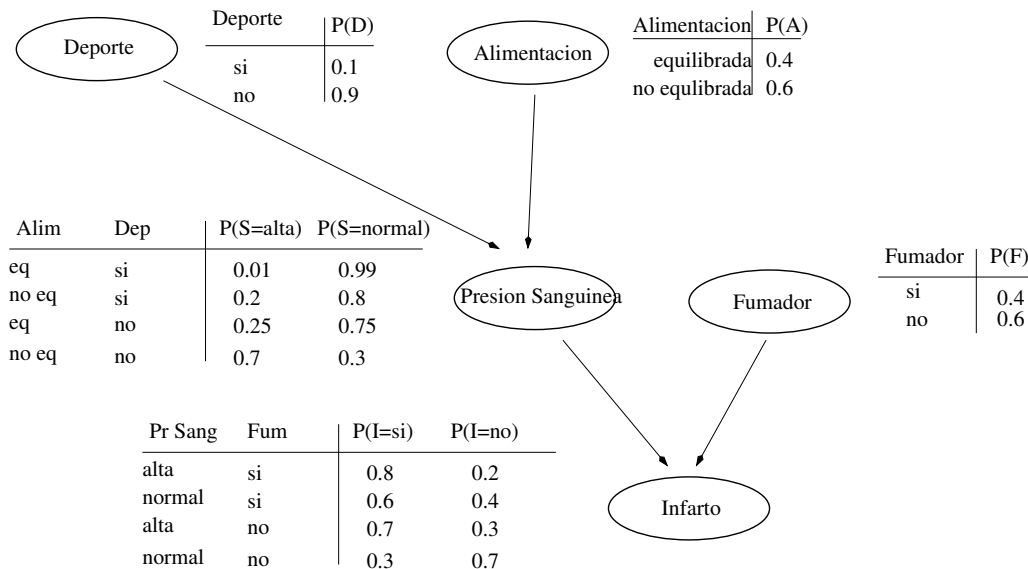
- ⊙ De manera que:

$$P(Z|X, Y) = \alpha P(X, Y|Z)P(Z) = \alpha P(X|Z)P(Y|Z)P(Z)$$

Redes Bayesianas

- ⊙ Determinando la independencia entre variables podemos simplificar el cálculo de la combinación de sus probabilidades y su representación
- ⊙ Las **redes bayesianas** permiten la representación de las relaciones de independencia entre variable aleatorias
- ⊙ Una red bayesiana es un **grafo dirigido acíclico** que tiene información probabilística en sus nodos indicando cual es la influencia de sus padres en el grafo sobre el nodo ($P(X_i | \text{padres}(X_i))$)
- ⊙ El significado intuitivo de un enlace entre dos nodos X e Y es que la variable X tiene influencia sobre Y
- ⊙ El conjunto de probabilidades representadas en la red describe la distribución de probabilidad conjunta de todas las variables

- ⊙ Determinar la probabilidad de que una persona sufra un infarto
- ⊙ Su probabilidad está determinada por cuatro variables: practicar **deporte**, **alimentación** equilibrada, **presión sanguínea** y ser **fumador**
- ⊙ Sabemos que la presión sanguínea depende directamente del deporte y la alimentación, que son variables independientes, y que el ser fumador es independiente del resto de variables
- ⊙ Saber esto nos permite crear la **red de dependencias** entre las variables
- ⊙ Nuestro conocimiento del dominio nos permite estimar la probabilidad que corresponde a cada una de las variables independientes y su influencia entre las variables dependientes



- ⊙ En cada nodo de la red aparece la distribución de probabilidad del nodo respecto a sus padres
- ⊙ Esto permite factorizar la distribución de probabilidad conjunta, convirtiéndose en el producto de probabilidades condicionales independientes

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{padres}(x_i))$$

$$\begin{aligned} &P(\text{Infarto} = \text{si} \wedge \text{Presion} = \text{alta} \wedge \text{Fumador} = \text{si} \\ &\wedge \text{Deporte} = \text{si} \wedge \text{Alimentacion} = \text{equil}) \\ &= \\ &P(\text{Infarto} = \text{si} | \text{Presion} = \text{alta}, \text{Fumador} = \text{si}) \\ &P(\text{Presion} = \text{alta} | \text{Deporte} = \text{si}, \text{Alimentacion} = \text{equil}) \\ &P(\text{Fumador} = \text{si})P(\text{Deporte} = \text{si})P(\text{Alimentacion} = \text{equil}) \\ &= 0,8 \times 0,01 \times 0,4 \times 0,1 \times 0,4 \\ &= 0,000128 \end{aligned}$$

- ⊙ Las propiedades de las redes bayesianas nos dan ciertas ideas sobre como construirlas. Si consideramos que (regla del producto):

$$P(x_1, x_2, \dots, x_n) = P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1}, \dots, x_1)$$

- ⊙ Iterando el proceso tenemos que:

$$\begin{aligned} P(x_1, \dots, x_n) &= P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \\ &\quad \cdots P(x_2 | x_1) P(x_1) \\ &= \prod_{i=1}^n P(x_i | x_{i-1}, \dots, x_1) \end{aligned}$$

- ⊙ Esta es la llamada **regla de la cadena**

- ⊙ Dadas estas propiedades, podemos afirmar que si $padres(X_i) \subseteq \{X_{i-1}, \dots, X_1\}$, entonces:

$$P(X_i | X_{i-1}, \dots, X_1) = P(X_i | padres(X_i))$$

- ⊙ Esto quiere decir que una red bayesiana es una representación correcta de un dominio sólo si cada nodo es condicionalmente independiente de sus predecesores en orden, dados sus padres
- ⊙ Para lograr esto, los padres de una variable X_i deben ser aquellos de entre las variables X_1, \dots, X_{i-1} que influyan directamente en X_i

- ⊙ Como comentamos, el coste de representar la distribución de probabilidad conjunta de n variables binarias es $O(2^n)$
- ⊙ La representación de **redes bayesianas** nos permite una **representación mas compacta** gracias a la **factorización** de la distribución conjunta
- ⊙ Suponiendo que cada nodo de la red tenga como máximo k padres ($k \ll n$), un nodo necesitará 2^k para representar la influencia de sus padres, por lo tanto el espacio necesario será $O(n2^k)$.
- ⊙ Por ejemplo, con 10 variables y suponiendo 3 padres como máximo tenemos 80 frente a 1024, con 100 variables y suponiendo 5 padres tenemos 3200 frente a aproximadamente 10^{30}

- ⊙ La **inferencia probabilística** calcula la **distribución de probabilidad a posteriori** de un conjunto de variables dada la observación de un evento (valores observados para un subconjunto de variables)
- ⊙ Dada X , la variable sobre la que queremos conocer la distribución
- ⊙ Dado \mathbf{E} , conjunto de variables de las que conocemos su valor E_1, \dots, E_n
- ⊙ Dado \mathbf{e} , conjunto de valores observados para las variables de E
- ⊙ \mathbf{Y} será el conjunto de variables ocultas (no observadas)
- ⊙ $\mathbf{X} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$ será el conjunto completo de variables
- ⊙ Nos plantearemos el cálculo de $P(X|\mathbf{e})$

- ⊙ **Inferencia por enumeración**: Cualquier probabilidad condicionada se puede calcular como la suma de todos los posibles casos a partir de la distribución de probabilidad conjunta.

$$P(X|\mathbf{e}) = \alpha P(X, \mathbf{e}) = \alpha \sum_y P(X, \mathbf{e}, \mathbf{y})$$

- ⊙ La **red bayesiana** nos permite **factorizar la distribución de probabilidad conjunta** y obtener una expresión mas fácil de evaluar
- ⊙ Por ejemplo, podemos calcular la probabilidad de **ser fumador si se ha tenido un infarto y no se hace deporte**

$$P(Fumador|Infarto = si, Deporte = no)$$

La distribución de probabilidad conjunta de la red sería:

$$P(D, A, S, F, I) = P(I|S, F)P(F)P(S|D, A)P(D)P(A)$$

Debemos calcular $P(F|I = si, D = no)$, por lo tanto tenemos

$$\begin{aligned} P(F|I = s, D = n) &= \alpha P(F, I = s, D = n) \\ &= \alpha \sum_{A \in \{e, \neg e\}} \sum_{S \in \{a, n\}} P(D = n, A, S, F, I = s) \\ &= \alpha P(D = n)P(F) \sum_{A \in \{e, \neg e\}} P(A) \sum_{S \in \{a, n\}} P(S|D = n, A)P(I = s|S, F) \end{aligned}$$

Si enumeramos todas las posibilidades y las sumamos de acuerdo con la distribución de probabilidad conjunta tenemos que:

$$\begin{aligned}
 &P(Fumador|Infarto = si, Deporte = no) \\
 &= \alpha \langle 0,9 \cdot 0,4 \cdot (0,4 \cdot (0,25 \cdot 0,8 + 0,75 \cdot 0,6)) + 0,6 \cdot (0,7 \cdot 0,8 + 0,3 \cdot 0,6)), \\
 &\quad 0,9 \cdot 0,6 \cdot (0,4 \cdot (0,25 \cdot 0,7 + 0,75 \cdot 0,3) + 0,6 \cdot (0,7 \cdot 0,7 + 0,3 \cdot 0,3)) \rangle \\
 &= \alpha \langle 0,253, 0,274 \rangle \\
 &= \langle 0,48, 0,52 \rangle
 \end{aligned}$$

- ⊙ El **algoritmo de eliminación de variables** intenta evitar la repetición de cálculos que realiza la inferencia por enumeración
- ⊙ Se usan técnicas de programación dinámica guardando cálculos intermedios para cada variable y así reutilizarlos (*factores*)
- ⊙ El cálculo se realiza evaluando la expresión de la distribución de probabilidad conjunta de izquierda a derecha
- ⊙ Los **factores** correspondientes a cada variable se van acumulando según se necesita
- ⊙ La ventaja de este algoritmo es que las variables no relevantes desaparecen al ser factores constantes

Function: Elimination de Variables (X, e, rb)factores $\leftarrow []$ vars $\leftarrow \text{REVERSE}(\text{VARS}(rb))$ **foreach** $var \in vars$ **do** factores $\leftarrow \text{concatena}(\text{factores}, \text{CALCULA-FACTOR}(var, e))$ **if** var es variable oculta **then** factores $\leftarrow \text{PRODUCTO-Y-SUMA}(var, \text{factores})$ **return** $\text{NORMALIZA}(\text{PRODUCTO}(\text{factores}))$

- ⊙ CALCULA-FACTOR genera el factor correspondiente a la variable en la función de distribución de probabilidad conjunta
- ⊙ PRODUCTO-Y-SUMA multiplica los factores y suma respecto a la variable oculta
- ⊙ PRODUCTO multiplica un conjunto de factores

- ⊙ Un **factor** corresponde a la **distribución de probabilidad de un conjunto de variables dadas las variables ocultas**
- ⊙ Se representa por una tabla que para cada combinación de variables ocultas da la probabilidad de las variables del factor

	Y	Z	
	C	C	0.2
$f_X(Y, Z)=$	C	F	0.4
	F	C	0.8
	F	F	0.6

- ⊙ Los factores tienen **dos operaciones**: **suma y producto**

- ⊙ La **suma** se aplica a **un factor** y sobre **una variable oculta del factor**. Como resultado obtenemos una matriz reducida en la que las filas del mismo valor se han acumulado

$$f_{X\bar{Z}}(Y) = \sum_Z f_X(Y, Z) = \begin{array}{c|c} Y & \\ \hline C & 0.6 \\ F & 1.4 \end{array}$$

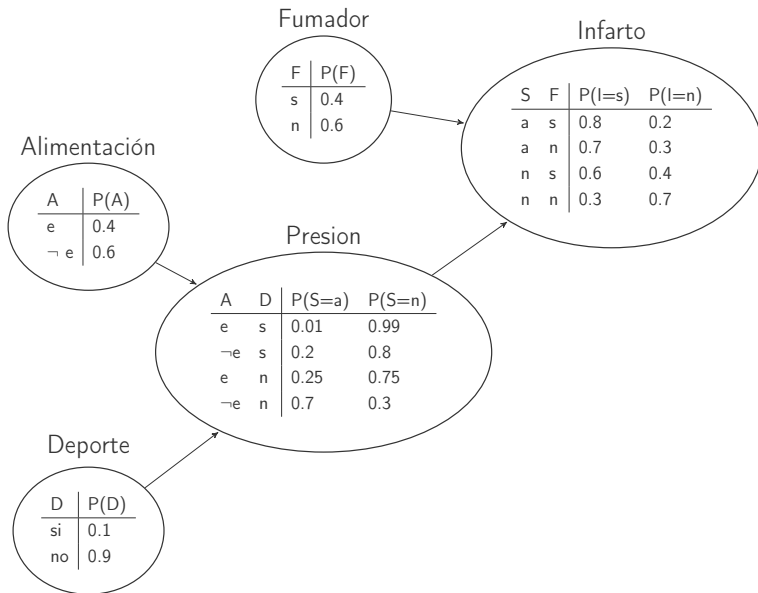
- ⊙ Es igual que una operación de agregación sobre una columna en bases de datos

- ⊙ El **producto** de factores permite **juntar** varios **factores** utilizando las **variables ocultas comunes**

$$f_{X_1 X_2}(Y, W, Z) = f_{X_1}(Y, Z) \times f_{X_2}(Z, W) =$$

Y	Z		Z	W		Y	Z	W	
C	C	0.2	C	C	0.3	C	C	C	$0,2 \times 0,3$
C	F	0.8	C	F	0.7	C	C	F	$0,2 \times 0,7$
F	C	0.4	F	C	0.1	C	F	C	$0,8 \times 0,1$
F	F	0.6	F	F	0.9	C	F	F	$0,8 \times 0,9$
						F	C	C	$0,4 \times 0,3$
						F	C	F	$0,4 \times 0,7$
						F	F	C	$0,6 \times 0,1$
						F	F	F	$0,6 \times 0,9$

- ⊙ Es igual que una operación de join en una base de datos multiplicando los valores de las columnas de datos



Volveremos a calcular $P(\text{Fumador} | \text{Infarto} = \text{si}, \text{Deporte} = \text{no})$ a partir de la distribución de probabilidad conjunta:

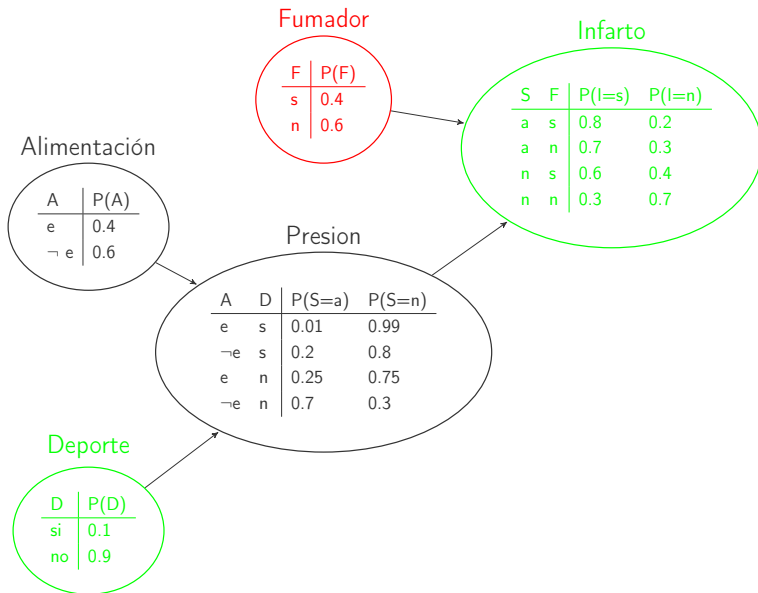
$$P(D, A, S, F, I) = P(I | S, F) P(F) P(S | D, A) P(D) P(A)$$

Debemos calcular $P(F | I = \text{si}, D = \text{no})$, por lo tanto tenemos

$$\begin{aligned} P(F | I = s, D = n) &= \alpha P(I = s, F, D = n) \\ &= \alpha \sum_{A \in \{e, \neg e\}} \sum_{S \in \{a, n\}} P(D = n, A, S, F, I = s) \end{aligned}$$

En esta ocasión no sacamos factores comunes para seguir el algoritmo

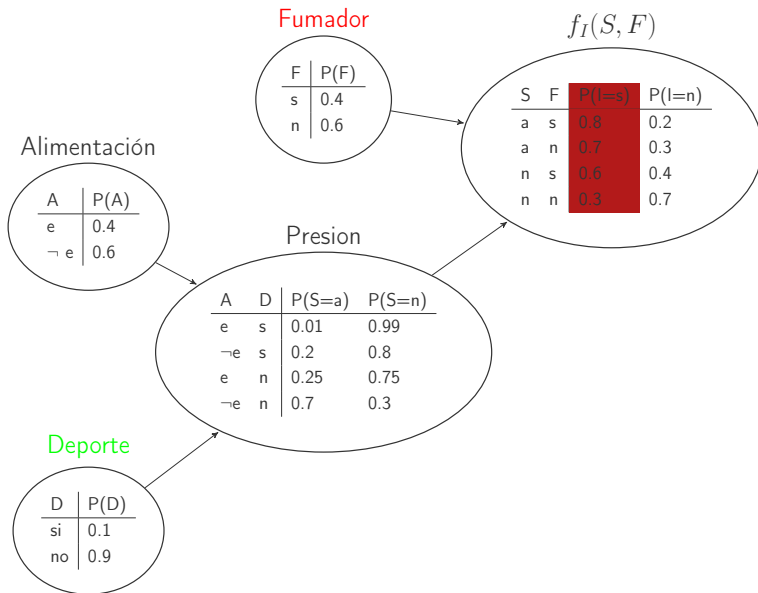
$$\alpha P(D = n) \sum_{A \in \{e, \neg e\}} P(A) \sum_{S \in \{a, n\}} P(S | D = n, A) P(F) P(I = s | S, F)$$



El algoritmo empieza calculando el factor para la variable *Infarto* ($P(I = s|S, F)$), ésta tiene fijo su valor a si, depende de las variables *Presión Sanguinea* y *Fumador*

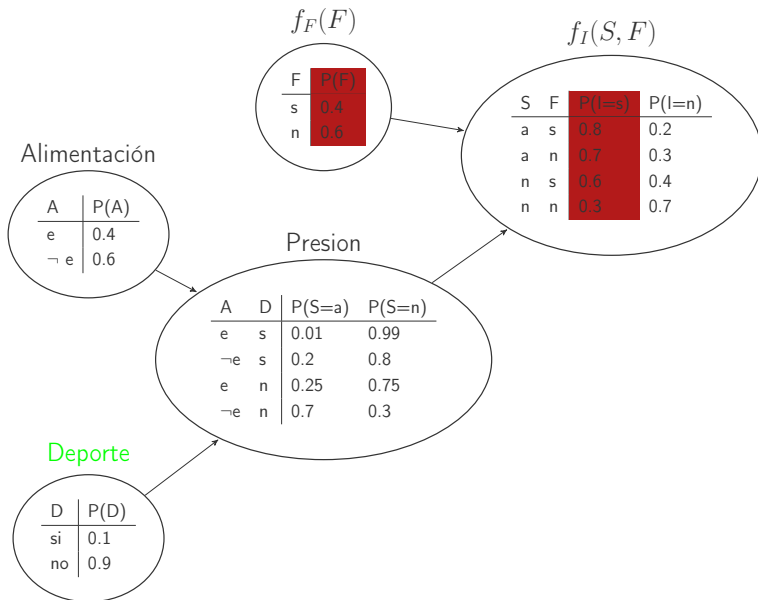
			S	F	
			a	s	0.8
			a	n	0.7
			n	s	0.6
			n	n	0.3

 $f_I(S, F) =$



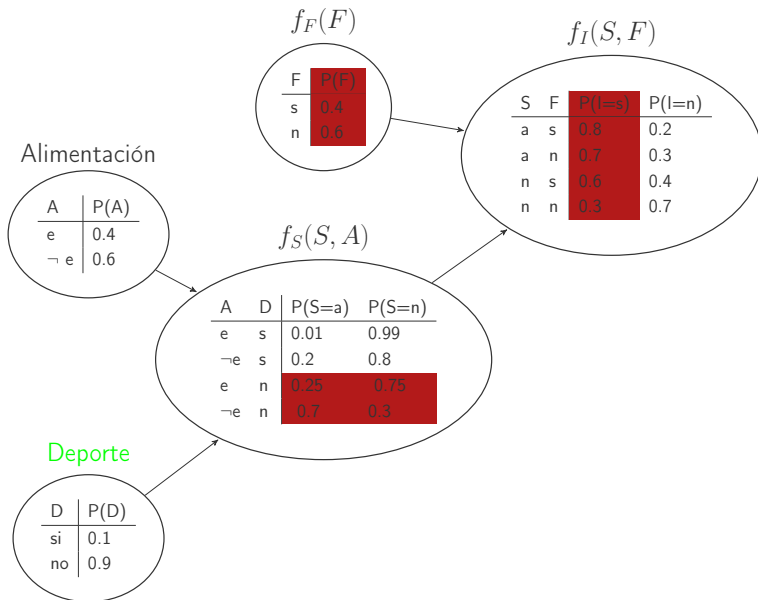
La variable fumador ($P(F)$) no depende de ninguna otra variable, al ser la variable que preguntamos el factor incluye todos los valores

$$f_F(F) = \begin{array}{c|c} F & \\ \hline s & 0.4 \\ n & 0.6 \end{array}$$



La variable *Presión Sanguinea* ($P(S|D = n, A)$), depende de las variable *Deporte* que tiene fijo su valor a *no* y *Alimentación*. Esta es una variable oculta, por lo que se debe calcular para todos sus valores

		S	A	
$f_S(S, A) =$	a	e		0.25
	a	$\neg e$		0.7
	n	e		0.75
	n	$\neg e$		0.3



Al ser la variable *Presión Sanguinea* una variable oculta debemos acumular todos los factores que hemos calculado

$$f_S(S, A) \times f_F(F) \times f_I(S, F)$$

$$f_{FI}(S, F) = f_F(F) \times f_I(S, F) =$$

S	F	
a	s	0.8×0.4
a	n	0.7×0.6
n	s	0.6×0.4
n	n	0.3×0.6

$$f_{FIS}(S, F, A) = f_{FI}(S, F) \times f_S(S, A) =$$

S	F	A	
a	s	e	$0.8 \times 0.4 \times 0.25$
a	s	$\neg e$	$0.8 \times 0.4 \times 0.7$
a	n	e	$0.7 \times 0.6 \times 0.25$
a	n	$\neg e$	$0.7 \times 0.6 \times 0.7$
n	s	e	$0.6 \times 0.4 \times 0.75$
n	s	$\neg e$	$0.6 \times 0.4 \times 0.3$
n	n	e	$0.3 \times 0.6 \times 0.75$
n	n	$\neg e$	$0.3 \times 0.6 \times 0.3$

Y ahora sumamos sobre todos los valores de la variable S para obtener el factor correspondiente a la variable *Presión Sanguinea*

$$f_{FIS}(F, A) = \sum_{S \in \{a, n\}} f_{FIS}(S, F, A) =$$

F	A	
s	e	$0.8 \times 0.4 \times 0.25 + 0.6 \times 0.4 \times 0.75 = 0.26$
s	$\neg e$	$0.8 \times 0.4 \times 0.7 + 0.6 \times 0.4 \times 0.3 = 0.296$
n	e	$0.7 \times 0.6 \times 0.25 + 0.3 \times 0.6 \times 0.75 = 0.24$
n	$\neg e$	$0.7 \times 0.6 \times 0.7 + 0.3 \times 0.6 \times 0.3 = 0.348$

Alimentación

A	P(A)
e	0.4
$\neg e$	0.6

$f_{FIS}(F, A)$

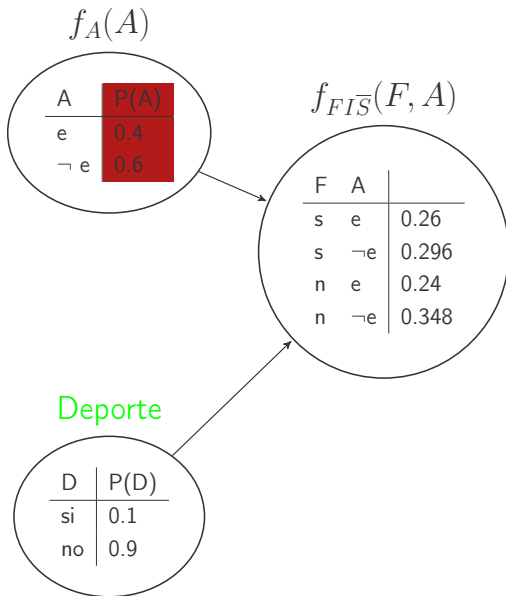
F	A	
s	e	0.26
s	$\neg e$	0.296
n	e	0.24
n	$\neg e$	0.348

Deporte

D	P(D)
si	0.1
no	0.9

El factor de la variable *Alimentación* ($P(A)$) no depende de ninguna variable, al ser una variable oculta generamos todas las posibilidades

$f_A(A)=$	A	
	e	0.4
	$\neg e$	0.6



Ahora debemos acumular todos los factores calculados

			F	A	
			s	e	$0.26 \times 0.4 = 0.104$
			s	$\neg e$	$0.296 \times 0.6 = 0.177$
			n	e	$0.24 \times 0.4 = 0.096$
			n	$\neg e$	$0.348 \times 0.6 = 0.208$

$$f_{AFI\bar{S}}(F, A) = f_A(A) \times f_{FI\bar{S}}(F, A) =$$

Y ahora sumamos sobre todos los valores de la variable A para obtener el factor correspondiente a la variable *Alimentación*

$$f_{\overline{AFIS}}(F) = \sum_{A \in \{e, \neg e\}} f_{AFIS}(F, A) =$$

F	
s	$0.104 + 0.177 = 0.281$
n	$0.096 + 0.208 = 0.304$

Deporte

D	P(D)
si	0.1
no	0.9

 $f_{AFIS}(F)$

F	
s	0.281
n	0.304

Y por último la variable *Deporte* ($P(D = n)$) tiene el valor fijado a *no* y dado que no depende de la variable *fumador* se puede obviar, ya que es un factor constante.

Ahora, si normalizamos a 1

$$P(F|I = s, D = n) = \begin{array}{c|c} F & \\ \hline s & 0.48 \\ n & 0.52 \end{array}$$

- ⊙ La complejidad del algoritmo de eliminación de variables depende del tamaño del mayor factor, que depende del orden en el que se evalúan las variables y la topología de la red
- ⊙ El orden de evaluación que escogeremos será el topológico según el grafo
- ⊙ La complejidad de la inferencia exacta es NP-hard en el caso general
- ⊙ Si la red bayesiana cumple que para cada par de nodos hay un único camino no dirigido (**poliárbol**) entonces se puede calcular en tiempo lineal
- ⊙ Para calcular el caso general se recurre a algoritmos aproximados basados en técnicas de muestreo