INFERENCIA Y REDES DE BAYES

PARA INTELIGENCIA ARTIFICIAL

AGENDA

- Teorema de Bayes
- Independencia de variables
- Redes Bayesianas
- D-Separation
- Inferencia por Enumeracion en una red bayesiana
- Eliminacion de variables

TEOREMA DE BAYES

Bayes

Thomas Bayes, matemático británico, estudió el problema de la determinación de la probabilidad de las causas a través de los efectos observados, su teorema se resuelve el problema conocido como **de la probabilidad inversa**

El teorema de Bayes tiene muchas aplicaciones, incluyendo **Aprendizaje Automático**, se usa en modelos de clasificación como el Naïve Bayes.

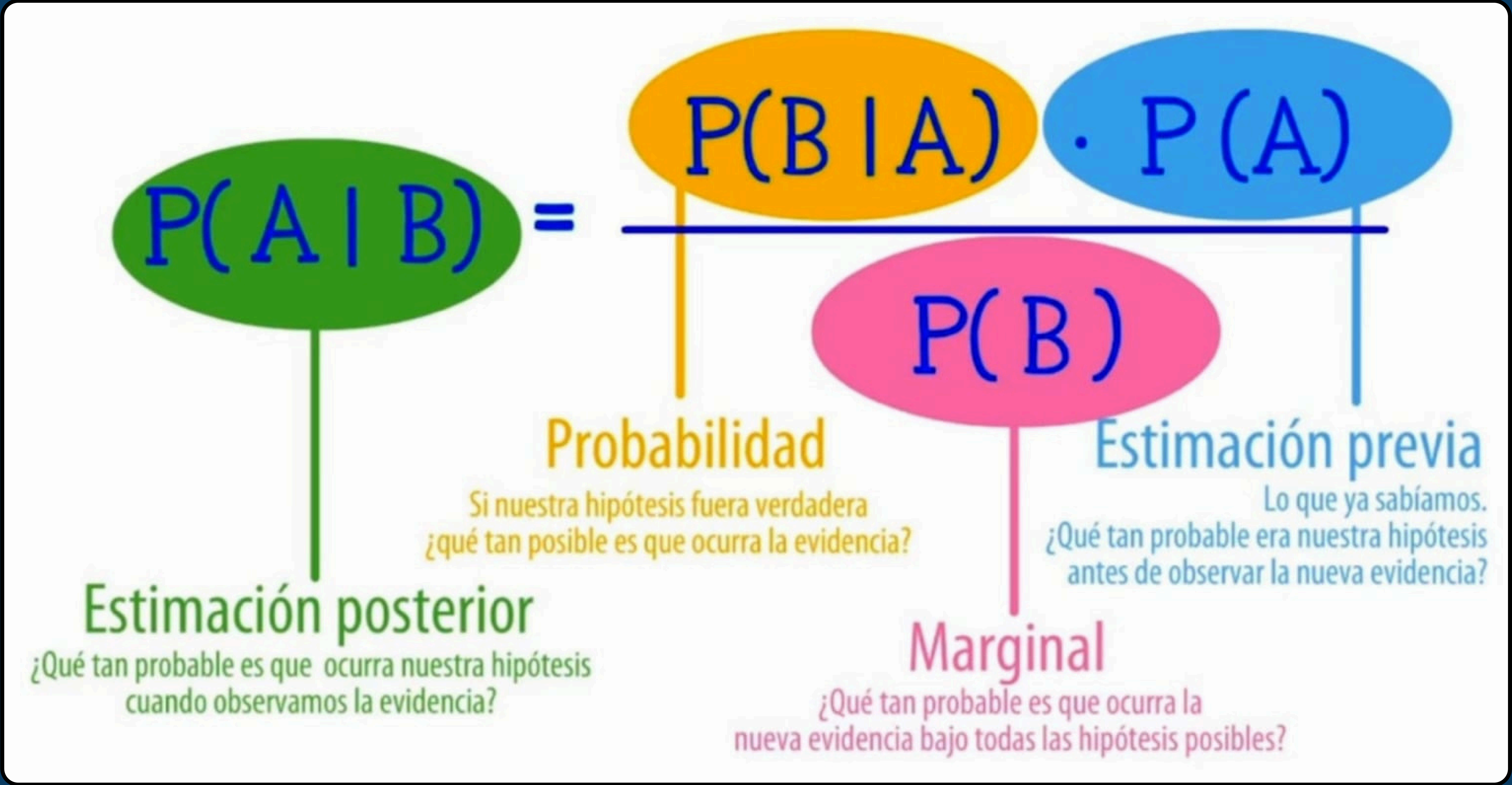
Teorema de Bayes

Es una regla matemática que nos dice cómo actualizar nuestras creencias cuando obtenemos nueva información.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(Causa|Efecto) = \frac{P(Efecto|Causa)P(Causa)}{P(Efecto)}$$

Teorema de Bayes



Inferencia con Teorema de Bayes

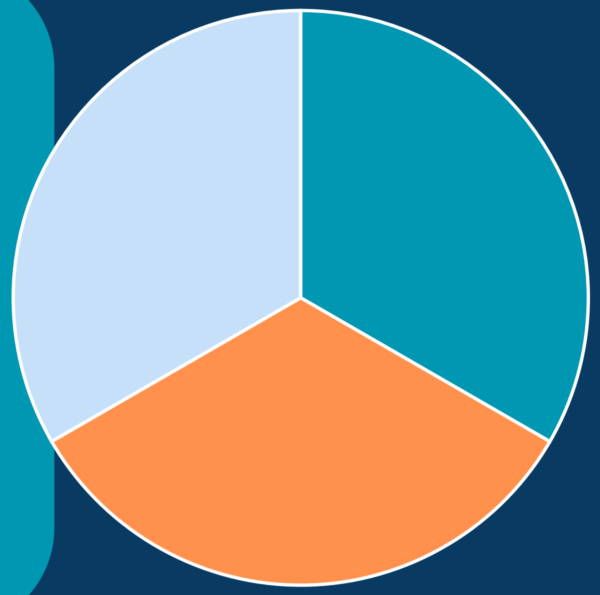
Segun datos del MSPAS durante 2016, 8 de cada 10 pacientes con Chinkunguya presentaban dolor de cuerpo entre sus sintomas.

En una muestra aleatoria de 10,000 Guatemaltecos se determinó que 1 de cada 10 presentaba síntomas de dolor de cuerpo, y existió unicamente un caso confirmado de Chinkunguya.

¿Cual es la probabilidad de tener Chinkunguya si tengo dolor de cuerpo?

Variables

- Dolor de Cuerpo = D (+d,−d)
- Chinkunguya = C (+c,−c)



Probabilidad de tener dolor de cuerpo (*P(+d)*)

que 1 de cada 10 presentaba síntomas de dolor de cuerpo

$$P(+d) = \frac{1}{10} = 0.10$$

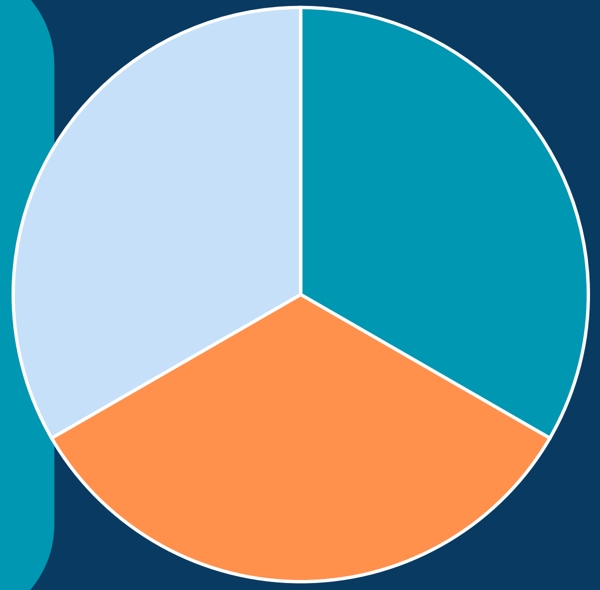
Probabilidad de tener Chinkunguya (*P(+c)*)

En una muestra aleatoria de 10,000 Guatemaltecos y existió unicamente un caso confirmado de Chinkunguya

$$P(+c) = \frac{1}{10,000} = 0.0001$$

Variables

- Dolor de Cuerpo = D (+d,−d)
- Chinkunguya = C (+c,−c)



Probabilidad de tener dolor de cuerpo dado que se tiene Chinkunguya (*P(+d | +c)*)

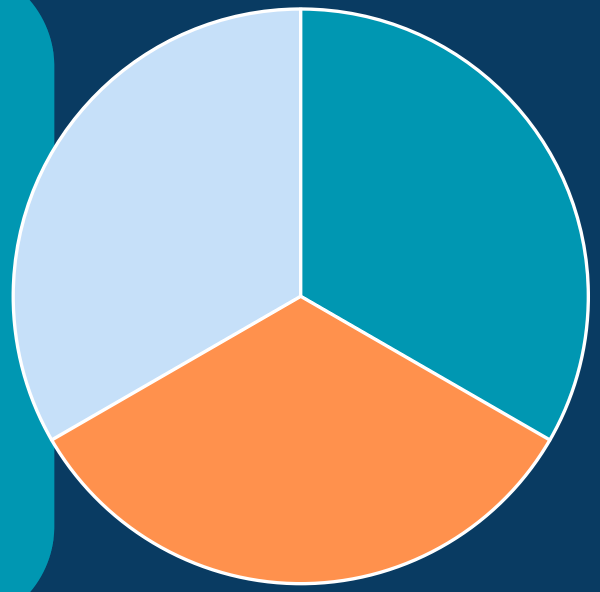
8 de cada 10 pacientes con Chinkunguya presentaban dolor de cuerpo entre sus síntomas

$$P(+d|+c) = \frac{8}{10} = 0.8$$

Entonces si tengo dolor de cuerpo, **¿Cuál es la probabilidad de tener Chinkunguya** (*P(+c | +d)*) ?

$$P(+c|+d) = \frac{P(+d|+c)P(+c)}{P(+d)} = \frac{0.8 * 0.0001}{0.1} = 0.0008$$

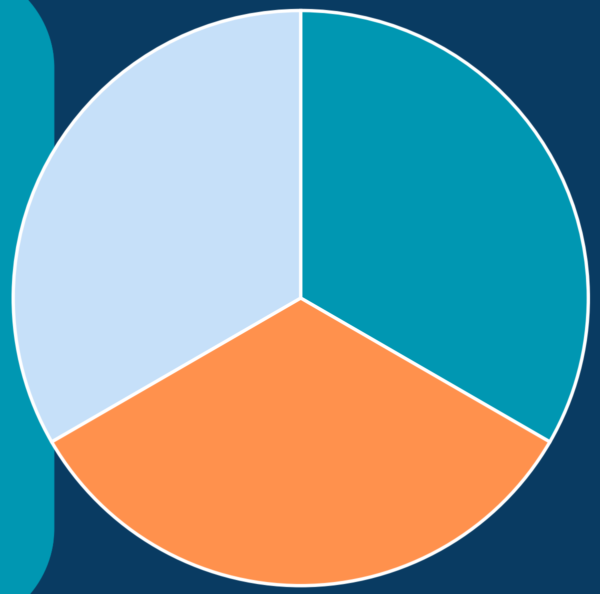
¿Que puede hacer el agente con los modelos?



Modelos probabilisticos describen como es el mundo, por tanto el agente puede:

- Razonar acerca de variables dada cierta evidencia **(inferir)**
- Explicación **(razonamiento diagnostico)**
- Predicción **(razonamiento causal)**

Ejercicio de inferencia con teorema de Bayes



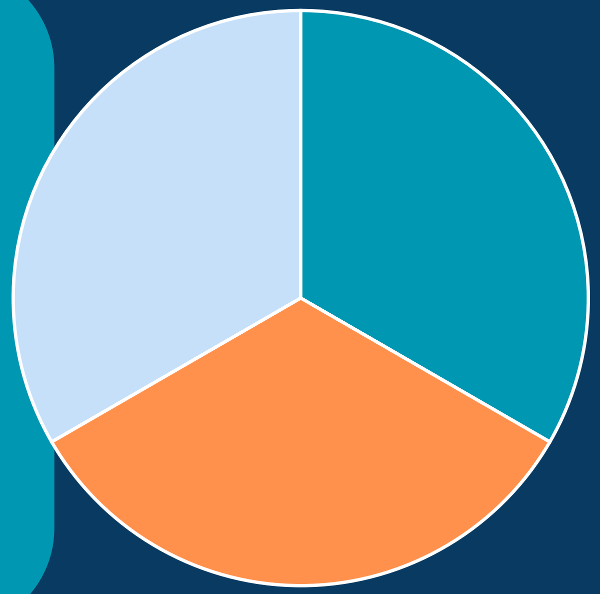
Clima	Prob.
Soleado	0.8
Lluvioso	0.2

D	C	Prob.
Mojado	Soleado	0.1
Seco	Soleado	0.9
Mojado	Lluvioso	0.7
Seco	Lluvioso	0.3

Distrubucion conjunta
o distribucion condicional?

¿P(C|D=seco)?

Si el suelo está seco, ¿qué tan probable es que el clima haya sido soleado o lluvioso?



Tenemos:

2 tipos de clima: Soleado y Lluvioso.

2 estados del suelo: Seco y Mojado.

Probabilidad total de suelo seco

Combinamos ambas opciones (soleado y lluvioso): $P(D = seco) = (0.9 \cdot 0.8) + (0.3 \cdot 0.2) = 0.72 + 0.06 = 0.78$

Aplicamos el Teorema de Bayes

Clima soleado dado que el suelo está seco $$P(C = soleado | D = seco) = \frac{0.9 \cdot 0.8}{0.78} = \frac{0.72}{0.78} \approx 0.9231$$

Clima lluvioso dado que el suelo está seco $$P(C = soleado | D = seco) = \frac{0.3 \cdot 0.2}{0.78} = \frac{0.06}{0.78} \approx 0.0769$$

Conclusión

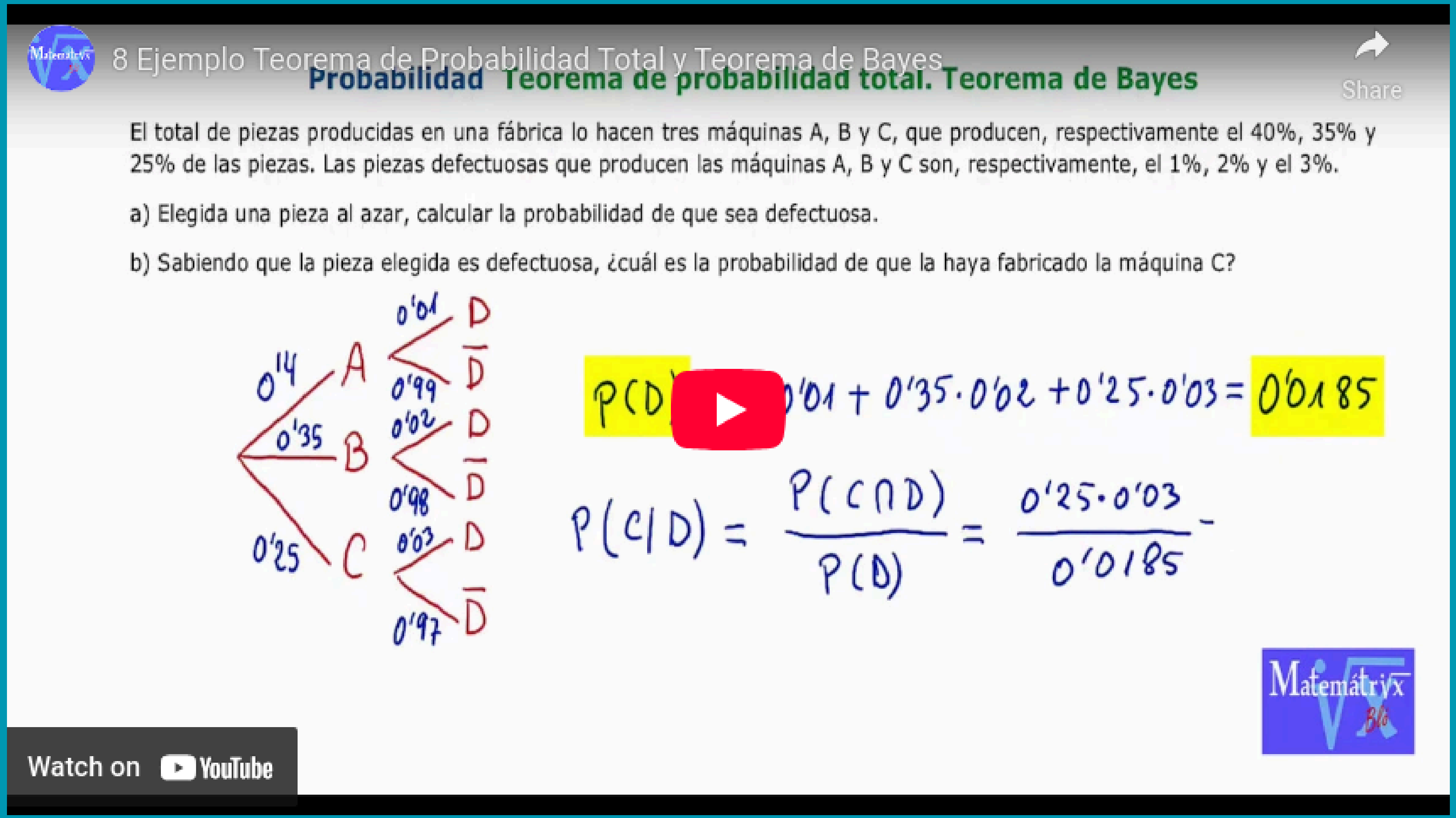
Si el suelo está seco, hay:

- 92.3% de probabilidad de que el clima sea soleado.
- 7.7% de probabilidad de que el clima sea lluvioso.

Teorema de probabilidad total

Es una herramienta que nos permite calcular la probabilidad de un evento considerando todas las maneras en que ese evento puede ocurrir, a través de una partición del espacio muestral.

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)(A_i)}$$




Probabilidad condicional explicada de manera visual (Teorema de Bayes) | Khan Acade...
Share
Watch on YouTube

Resuelve...

En un sorteo recibe un premio si sacas una canica roja de un saco de 100 canicas, donde 20 canicas son rojas y el resto azules. De las canicas rojas 15 son chicas y 5 grandes, mientras que de las azules 70 son chicas y 10 grandes. Si puedes sentir el tamaño, ¿Qué tamaño te conviene sacar?

Datos del problema

Total de canicas: 100

Canicas rojas: 20

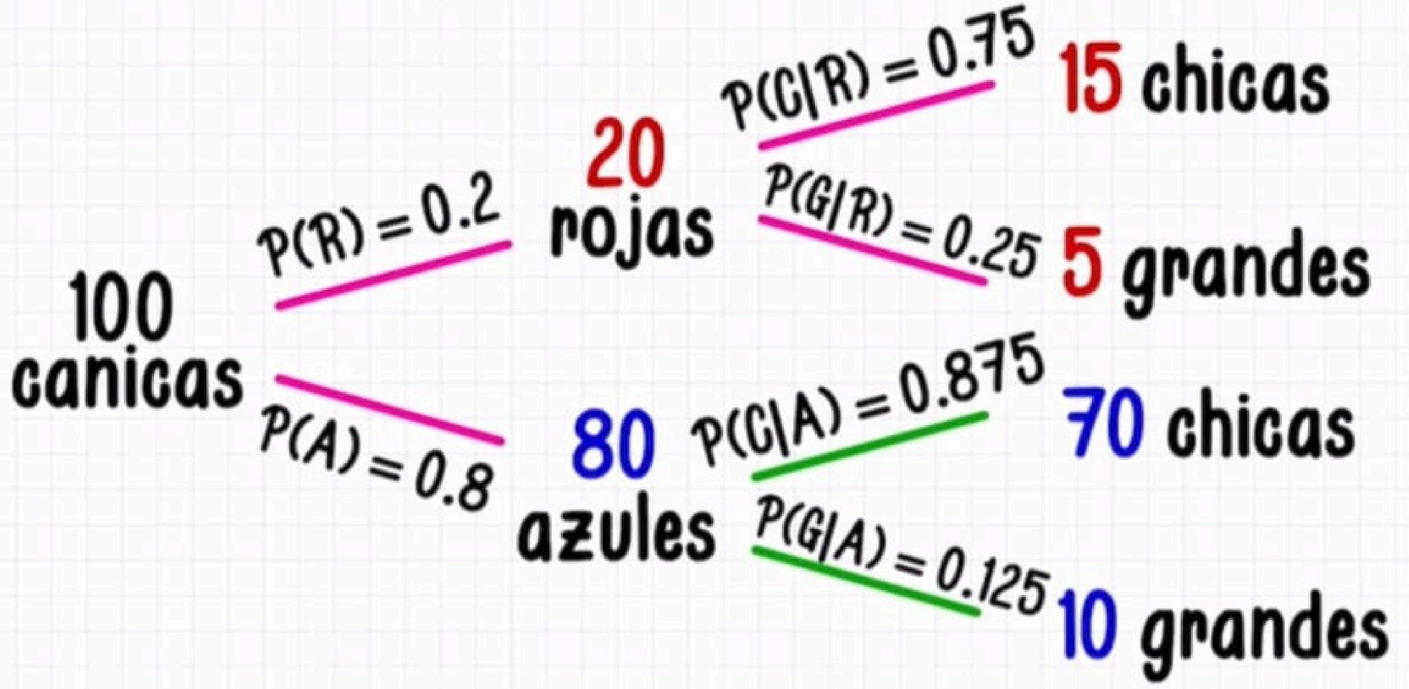
- Chicas: 15
- Grandes: 5

Canicas azules: 80

- Chicas: 70
- Grandes: 10

Resuelve...

Teorema de Bayes



$$\boxed{P(A|B)=\frac{P(B|A)\cdot P(A)}{P(B)}}$$

$$P(C)=(0.2)(0.75)+(0.8)(0.875)$$
$$=0.15+0.7=0.85$$

$$P(G)=(0.2)(0.25)+(0.8)(0.125)$$
$$=0.05+0.1=0.15$$

$$P(R|C)=\frac{P(C|R)\cdot P(R)}{P(C)}=\frac{(0.75)(0.2)}{0.85}=\frac{0.15}{0.85}=0.1765 \; \} \; 17.65\%$$

$$P(R|G)=\frac{P(G|R)\cdot P(R)}{P(G)}=\frac{(0.25)(0.2)}{0.15}=\frac{0.05}{0.15}=0.3333 \; \} \; 33.33\%$$

INDEPENDENCIA DE VARIABLES

Independencia de Variables

Se refiere a que el comportamiento de una variable no afecta en absoluto al comportamiento de la otra.

Es decir, conocer el valor de una variable no proporciona ninguna información sobre el valor de la otra.

Decimos que son independientes si, para cualquier par de valores **x** e **y**, se cumple:

$$\forall x, y \;\; P(x, y) = P(x)P(y) \dashrightarrow \;\; X \perp\!\!\!\perp Y$$

La independencia es una suposición fundamental en muchos modelos, ya que simplifica el análisis y el cálculo de probabilidades.

Independencia de Variables

Ejemplo: Lanzamiento de dos dados

Variables

- X = Resultado del primer dado (valores posibles: 1, 2, 3, 4, 5, 6)
- Y = Resultado del segundo dado (valores posibles: 1, 2, 3, 4, 5, 6)

Como es un dado justo:

$P(X=x) = 1/6$ para $x=1,2,3,4,5,6$

$P(Y=y) = 1/6$ para $y=1,2,3,4,5,6$

Es decir, cada número tiene la misma probabilidad de salir en cada dado.

Independencia de Variables

Como son dos dados independientes, la probabilidad de que el primer dado sea un 3 y el segundo dado sea un 5 es:

$$P(X=3 \text{ y } Y=5)=P(X=3)\cdot P(Y=5)$$

$$P(X=3 \text{ y } Y=5)=1/6\cdot 1/6=1/36$$

Esta es la fórmula de independencia: el producto de las probabilidades individuales (marginales) nos da la probabilidad conjunta.

El resultado del primer dado no afecta al resultado del segundo dado.
Por eso, saber el valor de X (primer dado) no cambia las probabilidades de Y (segundo dado).
Esto es independencia.

En este caso, tirar un dado no afecta al otro, por lo que:

$$P(X,Y)=P(X)P(Y) \Rightarrow X \perp Y$$

Independencia Condicional

Ocurre cuando dos variables son independientes entre sí al condicionar (fijar) una tercera variable

Si conocemos el valor de esa tercera variable, el conocimiento de una de las dos primeras no aporta información adicional sobre la otra.

$$\forall x, y, z \;\; P(x, y|z) = P(x|z)P(y|z) \dashrightarrow X \perp\!\!\!\perp Y|Z$$

Independencia Condicional

Ejemplo: Clima, paraguas la calle mojada

Variables

- X = La calle está mojada. (sí/no)
- Y = La gente lleva paraguas (sí/no)
- Z = El clima (lluvioso o soleado)

Si vemos a alguien con paraguas, es razonable pensar que probablemente la calle esté mojada (porque quizás llovió), así que:

$$P(X \mid Y) \neq P(X)$$

X y Y no son independientes, ya que ver un paraguas nos da información sobre la calle mojada.

Independencia Condicional

Si sabemos que el clima es lluvioso, entonces:

- Saber que alguien lleva paraguas ya no añade mucha información extra sobre si hay charcos, porque el clima (Z) ya explica eso.

Entonces:

- $P(X \mid Y, Z=\text{lluvioso})=P(X \mid Z=\text{lluvioso})$
- X y Y son condicionalmente independientes dado Z.

Sin saber el clima, X y Y están relacionados.
Sabiendo el clima (Z), X y Y ya no dependen uno del otro. Toda la información sobre charcos y paraguas está explicada por el clima.

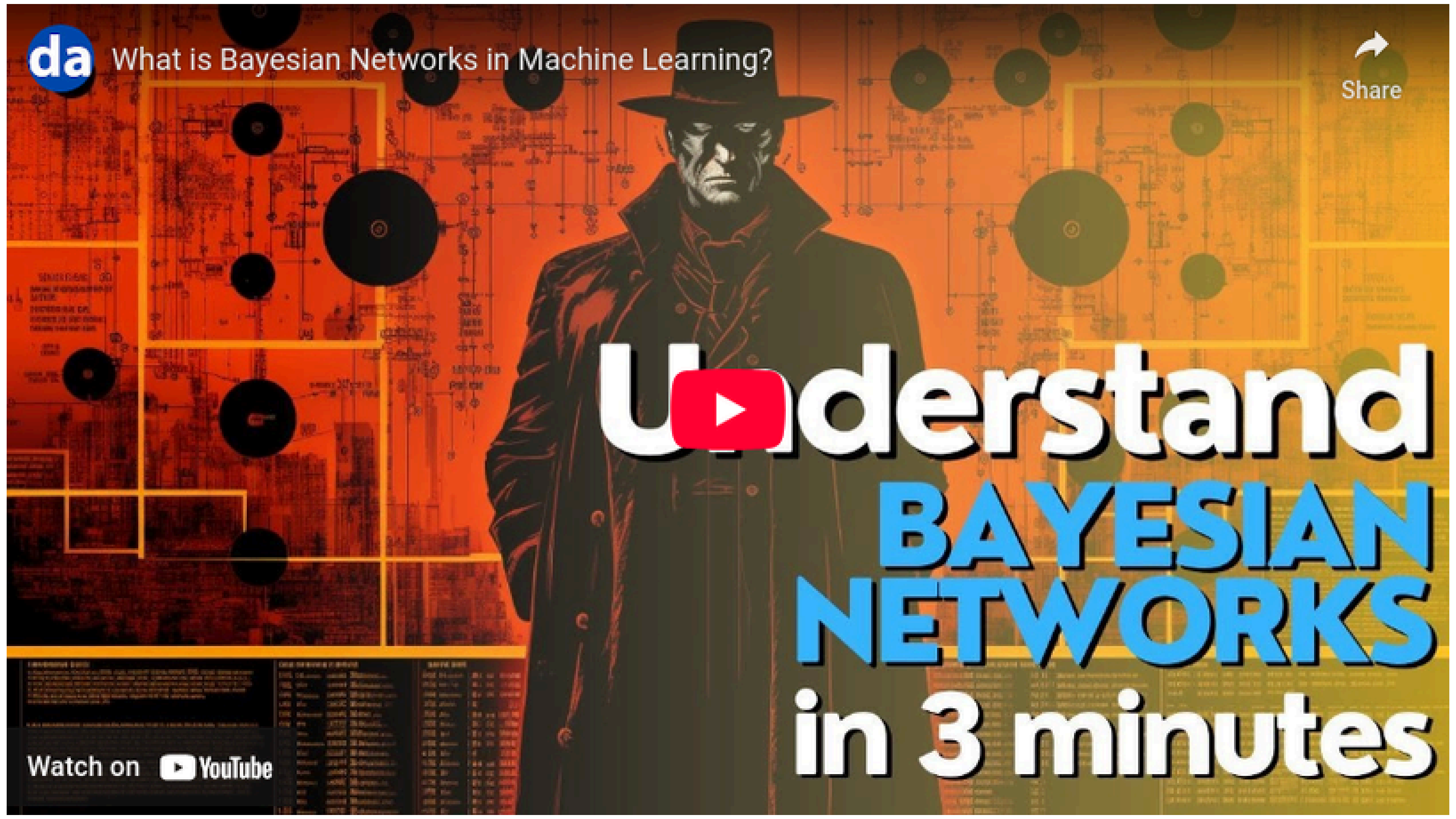
Diferencias Clave

Independencia	Independencia Condicional
$P(X, Y) = P(X)P(Y)$	$P(X, Y \mid Z)=P(X \mid Z) \cdot P(Y \mid Z)$
No depende de ninguna tercera variable	Involucra las variables X, Y y una tercera Z.
Ejemplo: Lanzar dos dados. El resultado de un dado no afecta al otro.	Ejemplo: El uso de paraguas (Y) y que el suelo esté mojado (X) son independientes si sabes que está lloviendo (Z).

- **Independencia: X y Y no se afectan.**
- **Independencia condicional: X y Y parecen relacionados, pero esa relación desaparece cuando conoces Z.**

REDES BAYESIANAS

Redes Bayesianas



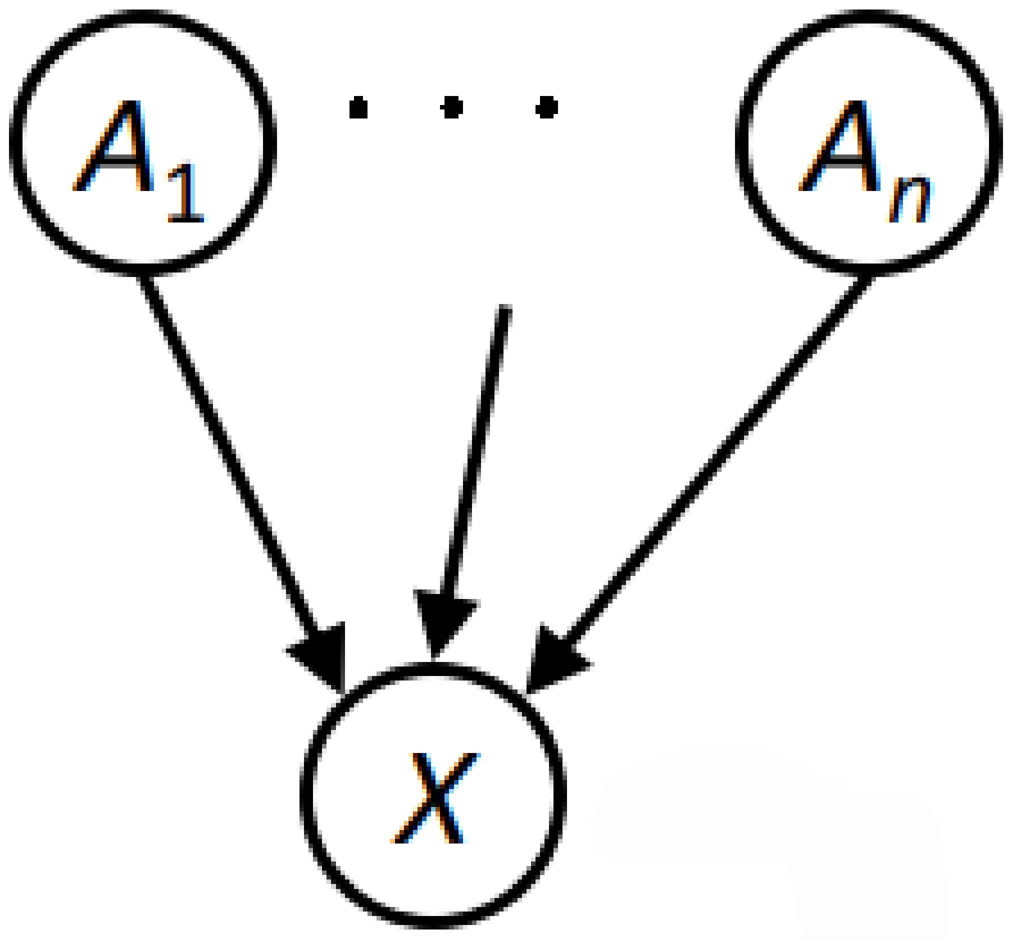
Se conoce también como:

- Red probabilística (Probabilistic Network)
- Red Causal (Causal Network)
- Red de Creencias (Belief Network)
- Mapa de Conocimiento (Knowledge Map)

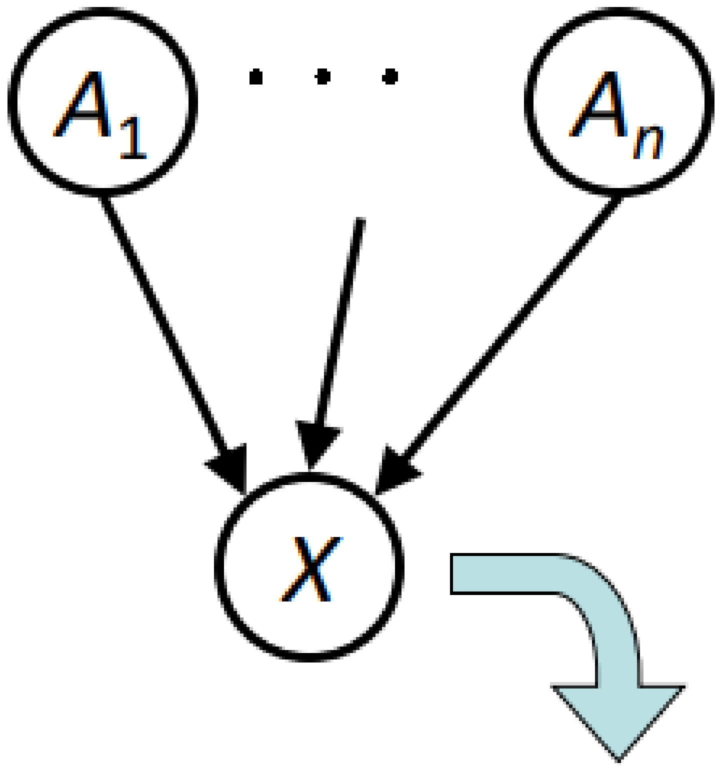
Redes Bayesianas

Es un modelo probabilístico gráfico que representa un conjunto de variables y las relaciones de dependencia condicional entre ellas, utilizando un grafo dirigido acíclico:

- Cada nodo representa una variable aleatoria
- Las flechas o conexiones entre nodos indican la dirección de la influencia o dependencia, muchas veces interpretada como una relación causal



Redes Bayesianas



$$P(X|A_1, \ldots A_n)$$

La red bayesiana utiliza el teorema de Bayes para actualizar las probabilidades de los eventos a medida que se incorpora nueva evidencia

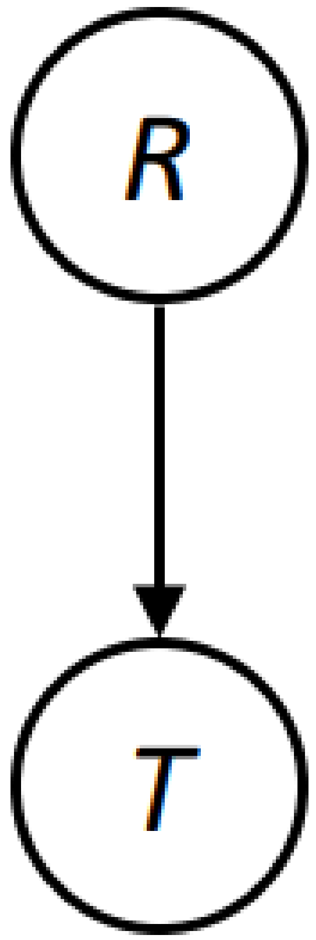
Describe distribuciones conjuntas complejas (modelos), usando distribuciones locales (probabilidades condicionales).

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

Ejemplo Redes Bayesianas - Tráfico y Lluvia

En una ciudad, se ha observado que la probabilidad de que llueva en un día cualquiera es de 1/4. Además, si llueve, la probabilidad de que haya tráfico es de 3/4, mientras que si no llueve, la probabilidad de tráfico es de 1/2.

Utilizando una red bayesiana, calcula la probabilidad de que llueva y no haya tráfico en un día determinado.



$P(R)$

+r	1/4
-r	3/4

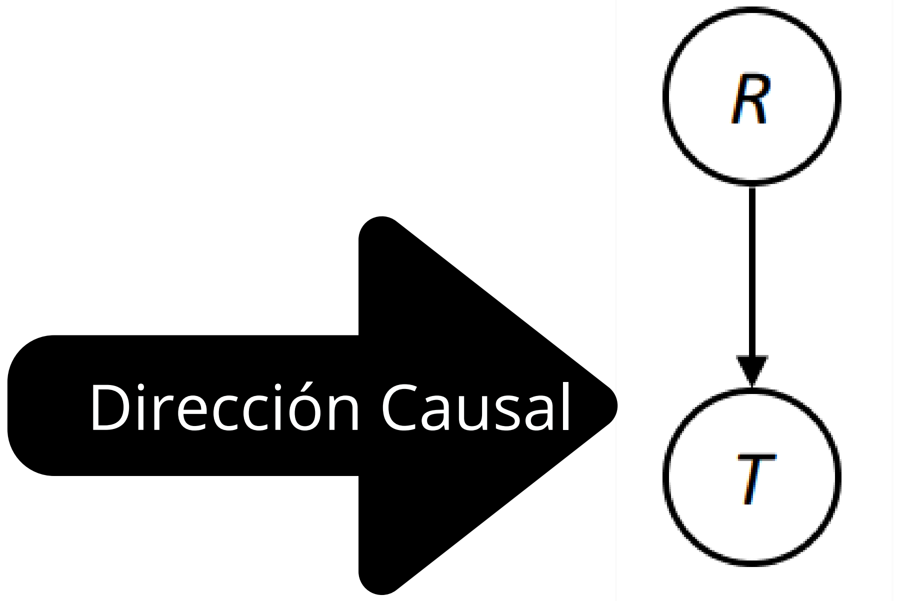
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

Ejemplo Redes Bayesianas – Tráfico y Lluvia

Calcula la probabilidad de que llueva y no haya tráfico en un día determinado.

Utiliza la fórmula de la probabilidad conjunta para calcular P(+r,-t).



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

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

$$P(+r, -t) = P(+r)P(-t|+r)$$

$$= \frac{1}{4} * \frac{1}{4} = \frac{1}{16}$$

Ejemplo Redes Bayesianas - Robo

Estás en el trabajo y recibes una llamada de tu vecino John, que te dice que la alarma de tu casa está sonando. Sin embargo, tu otra vecina Mary no te llama. A veces, la alarma se activa por pequeños terremotos.

La pregunta es: *¿Hay un ladrón?*

Variables

La red bayesiana modela estas variables:

- Burglar: ¿Hay un ladrón?
- Earthquake: ¿Hubo un terremoto?
- Alarm: ¿Está sonando la alarma?
- JohnCalls: ¿John te llama?
- MaryCalls: ¿Mary te llama?

Relación causal

La red refleja relaciones causales, es decir, cómo unas variables causan otras:

- Si hay un ladrón, puede activar la alarma.
- Si hay un terremoto, también puede activar la alarma.
- Si la alarma suena, eso puede hacer que:
    - John te llame.
    - Mary te llame.

Ejemplo Redes Bayesianas - Robo

B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

¿Qué te permite esta red?

- Calcular probabilidades conjuntas de todas las variables.
- Estimar, dado que John te llama pero Mary no, cuál es la probabilidad de que haya un ladrón.
- Actualizar tus creencias conforme obtienes evidencia (por ejemplo, si después te enteras que hubo un terremoto, ajustas la probabilidad de que haya un ladrón).

Ejemplo Redes Bayesianas - Robo

B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

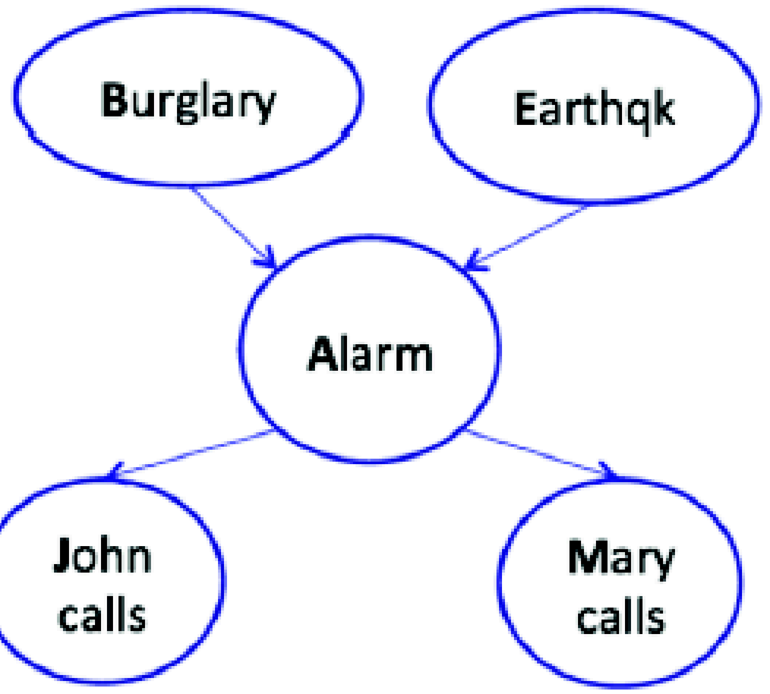
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Para este problema en particular, las variables son:

- Burglary (B): Robo.
- Earthquake (E): Terremoto.
- Alarm (A): Alarma.
- John calls (J): Llamada de John.
- Mary calls (M): Llamada de Mary.

Ejemplo Redes Bayesianas - Robo

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

El orden debe respetar la estructura de la red, es decir, si X es padre de Y, entonces X debe aparecer antes que Y en el orden:

- Burglary (B) y Earthquake (E) son variables independientes.
- Alarm (A) depende de B y E.
- John calls (J) y Mary calls (M) dependen de A.

Un orden válido es: B,E,A,J,M

Ejemplo Redes Bayesianas – Robo

B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

¿Cuál es la probabilidad que se active la alarma, es debido a un robo y se llama únicamente a Mary?

- B: Robo (Burglary)
  - B=+b significa que hubo un robo
  - B=-b significa que no hubo un robo
- E: Terremoto (Earthquake)
  - E=+e terremoto
  - E=-e sin terremoto
- ...
- M: Mary llama (Mary calls)
  - M=+m Mary llama
  - M=-m Mary no llama

Ejemplo Redes Bayesianas - Robo

B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

Queremos encontrar la probabilidad de que la alarma se active debido a un robo y que solo Mary llame.

Esto se puede expresar como:

$$\underbrace{\mathbf{P(+b, -e, +a, -j, +m)}}$$

Probabilidad de que la alarma suene, haya robo, no haya terremoto, Mary llame, y John no

Ejemplo Redes Bayesianas - Robo

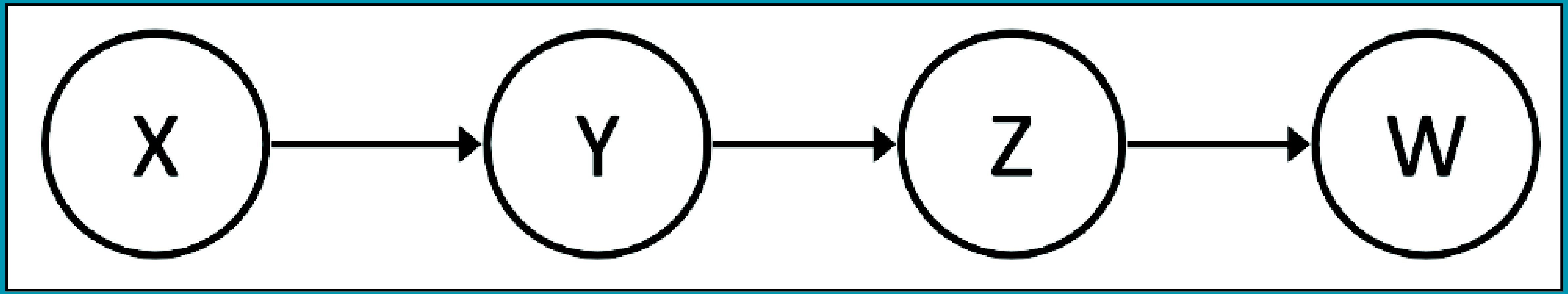
$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a| + b, -e)P(-j| + a)P(+m| + a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

- La regla de la cadena codifica distribuciones conjuntas en secuencia de variables
- Las redes de Bayes asumen que la influencia directa son los padres
- Es comprobable a través del grafo

$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

D-SEPARATION

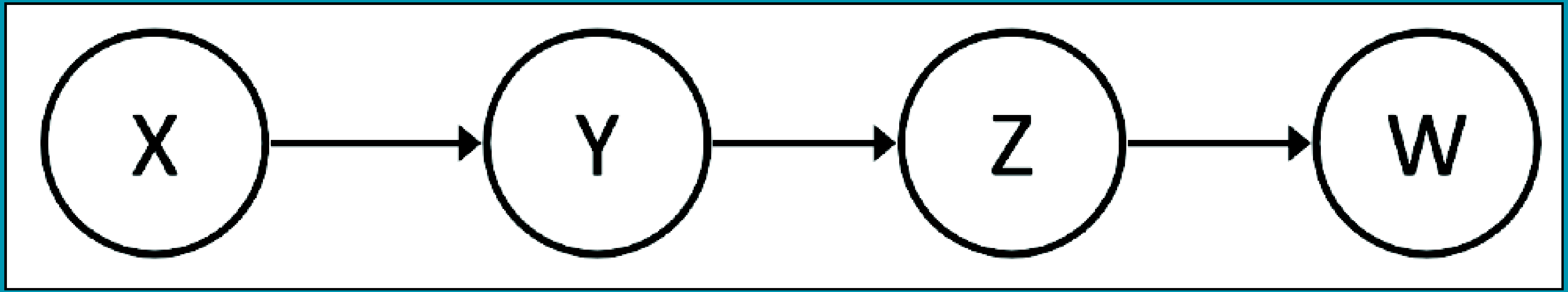
¿Son dos nodos independientes dada la evidencia?



Para determinar si dos nodos son independientes dado otro nodo, debemos ver si la influencia de uno sobre el otro se ve interrumpida por la evidencia.

¿Son dos nodos independientes dada la evidencia?

Digamos que queremos saber si "Estudiar" (X) y "Pasar el Examen" (W) son independientes, dado el conocimiento de "Confianza" (Z)



Estudio: Si el estudiante estudia o no.

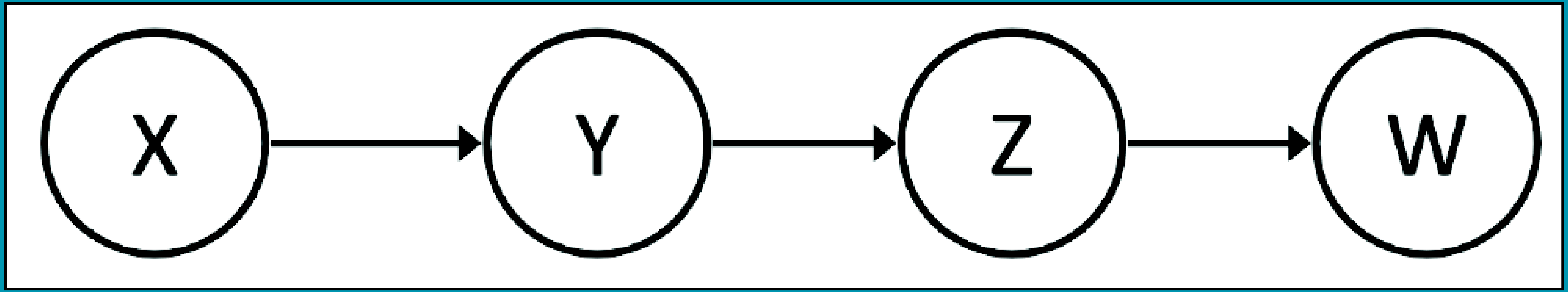
Dormir Bien: Si el estudiante durmió bien la noche anterior al examen.

Confianza: El nivel de confianza del estudiante al entrar al examen.

Pasar el Examen: Si el estudiante pasa o no el examen.

¿Son dos nodos independientes dada la evidencia?

Conociendo Z...



Si sabemos el nivel de confianza (Z), entonces la influencia de estudiar (X) en pasar el examen (W) se modula a través de la confianza (Z).

En este caso, estudiar (X) y pasar el examen (W) son independientes dado que conocemos la confianza (Z).

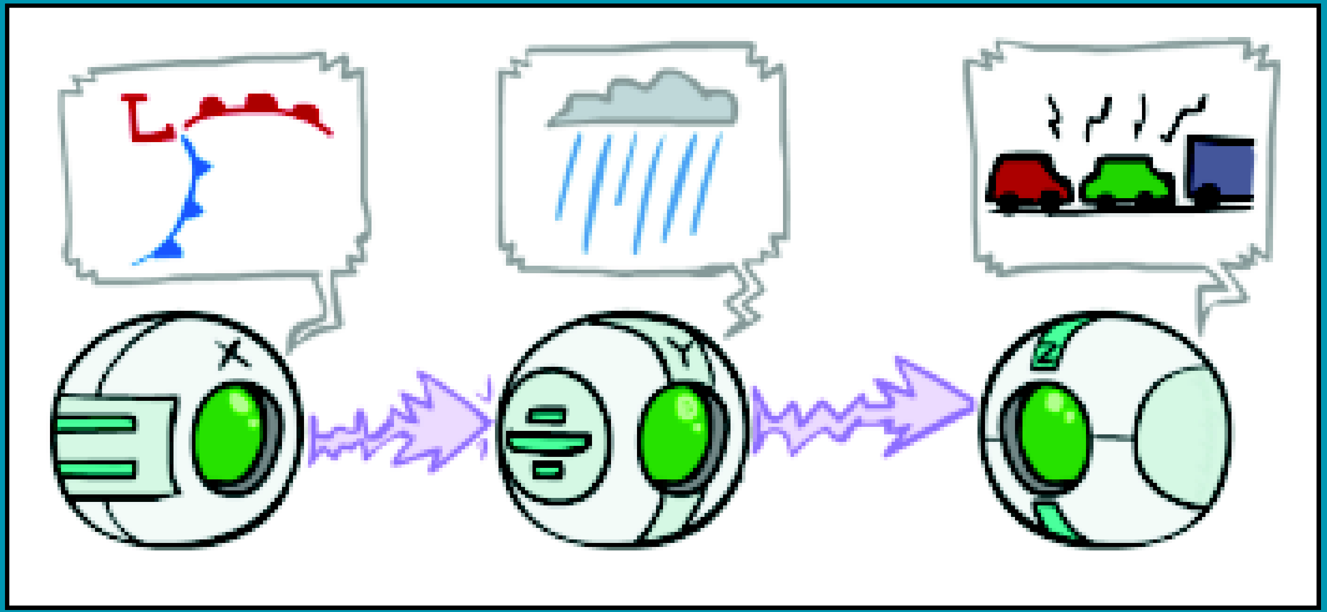
D-Separation

En una red bayesiana, dos nodos A y B están d-separados (es decir, son condicionalmente independientes) dado un conjunto de evidencia E, si todos los caminos entre A y B quedan "bloqueados" por E.

Existen tres patrones básicos a tener en cuenta:

- **Casual Chain**
- **Common Case**
- **Common Effect**

D-Separation - Casual Chain

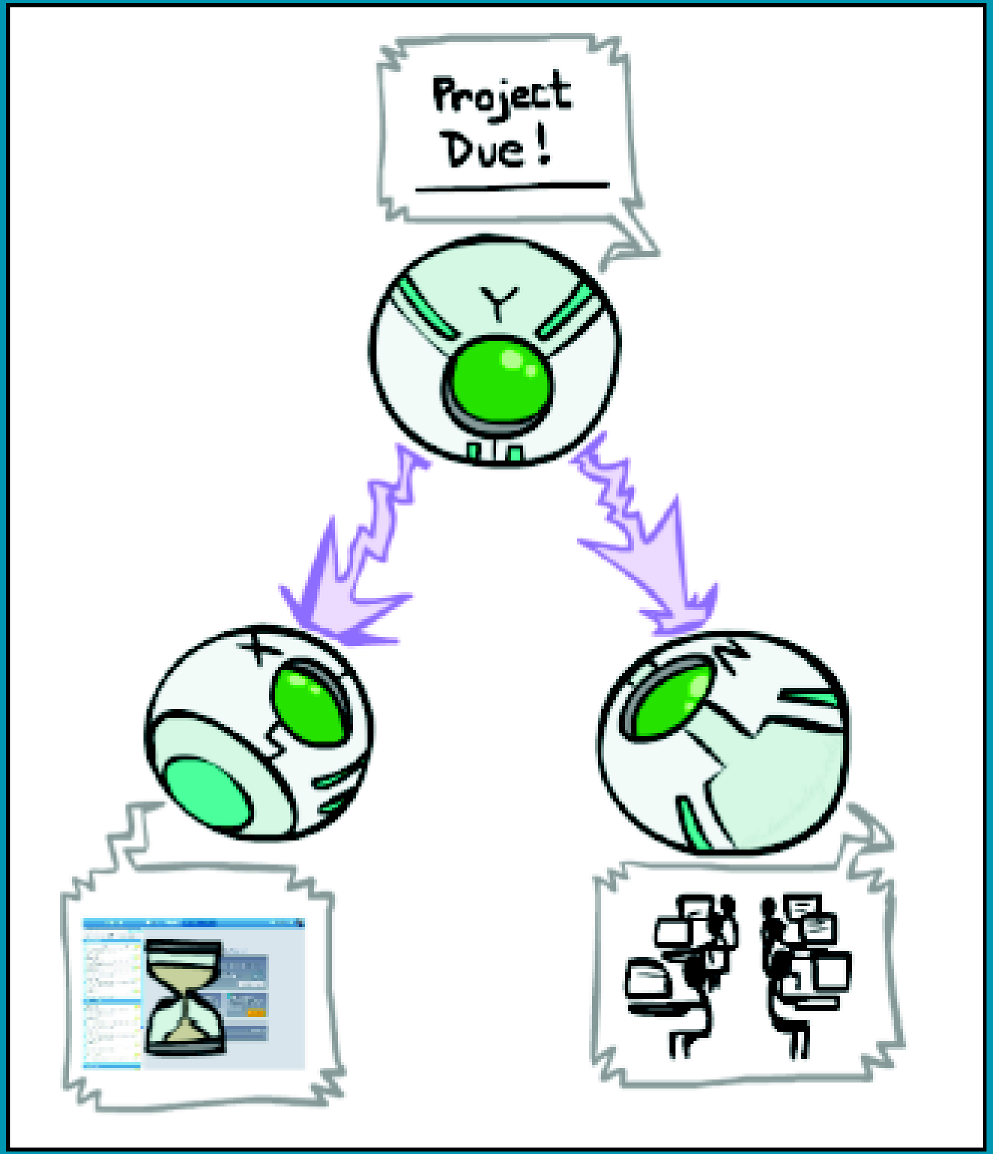


$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

$$\begin{aligned} P(z|x,y) &= \frac{P(x,y,z)}{P(x,y)} \\ &= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} \\ &= P(z|y) \end{aligned}$$

Sin condicionar en Y, el camino está abierto (dependencia).
Si condicionas en Y, bloqueas el camino (independencia).

D-Separation – Common Cause



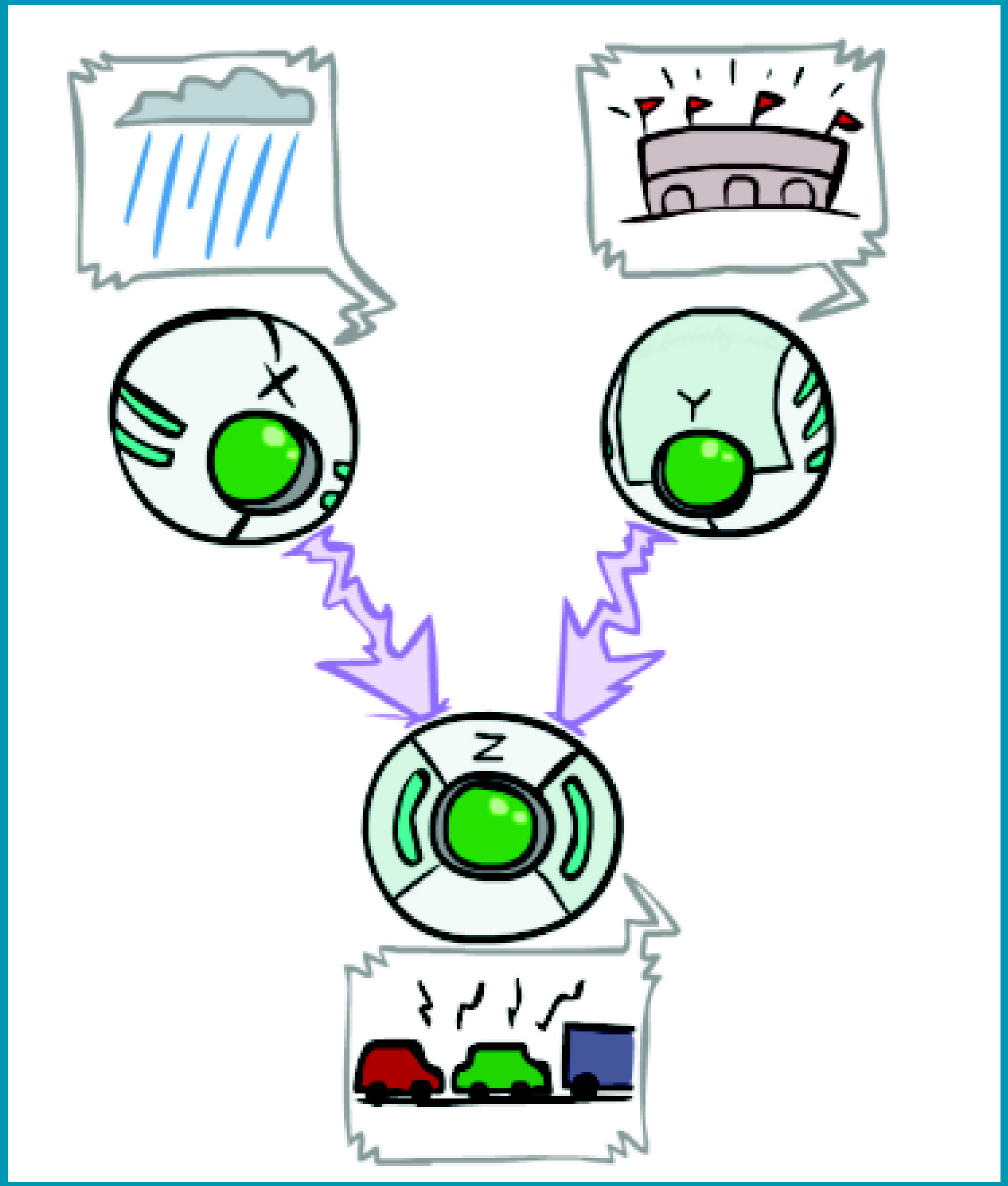
$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)}$$

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

$$= P(z|y)$$

Sin condicionar en Y, el camino está abierto (dependencia).
Si condicionas en Y, bloqueas el camino (independencia).

D-Separation – Common Effect

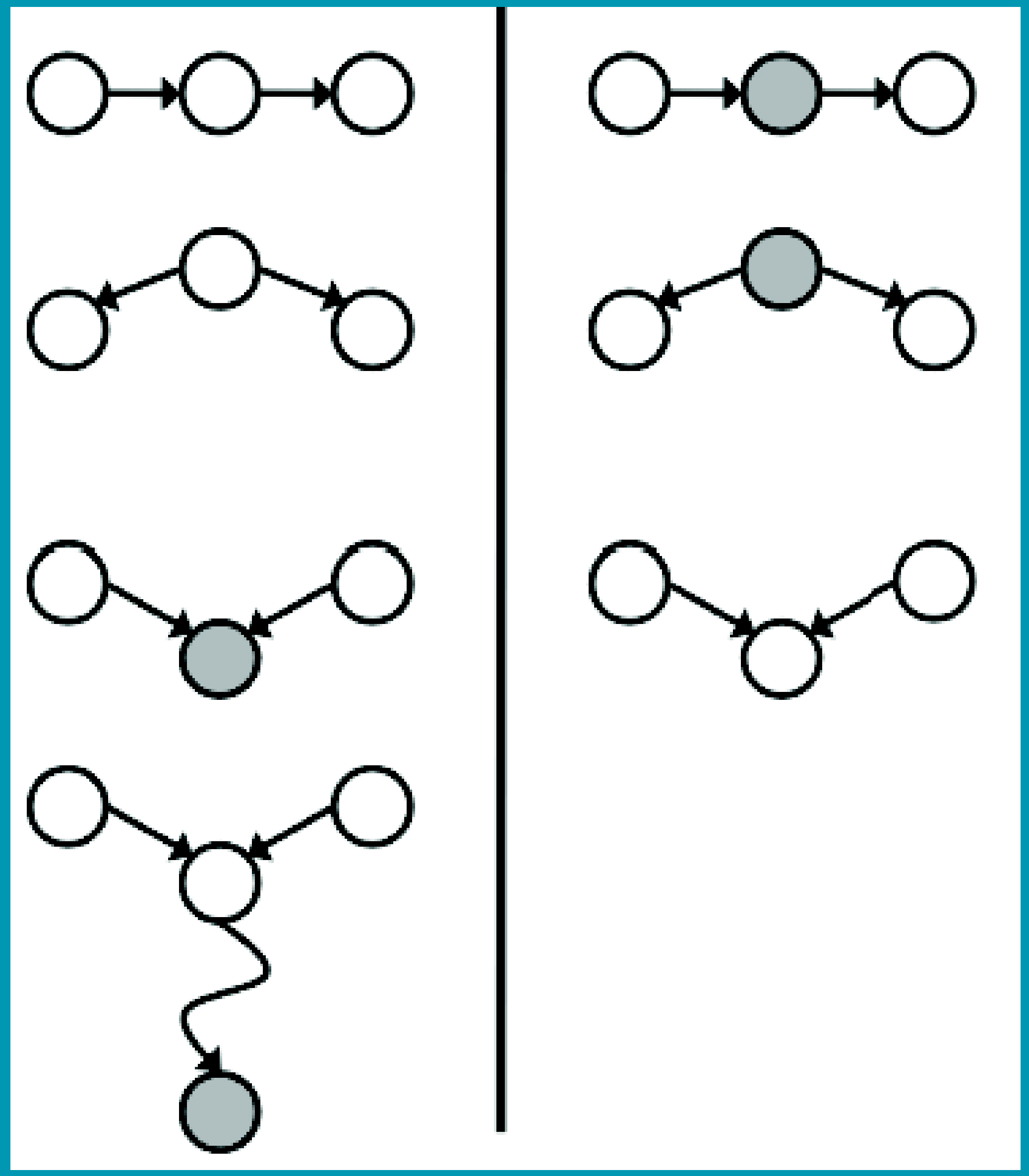


- Sin condicionar en Z, el camino está bloqueado (independencia).
- Si condicionas en Z (o en un descendiente de Z), se abre el camino (dependencia).

D-Separation

- Debe existir camino entre las variables
- Una vez se establece el(los) camino(s) se analizara por triadas
- Un camino es activo si todas las triadas son activas
- Si hay un camino activo, no hay independencia

Triada Activa **Triada Inactiva**



INFERENCIA

Dada una Red Bayesiana, ¿Cuál es $P(X \mid e)$?

Inferencia

Probabilidad posterior

$$P(Q|E_1 = e_1, \dots E_k = e_k)$$

$$\left.\begin{array}{ll} \text{Evidence variables:} & E_1 \dots E_k = e_1 \dots e_k \\ \text{Query* variable:} & Q \\ \text{Hidden variables:} & H_1 \dots H_r \end{array}\right\} \begin{array}{c} X_1, X_2, \dots X_n \\ \textit{All variables} \end{array}$$

Inferencia por Enumeracion

La inferencia por enumeración en redes bayesianas consiste en calcular la probabilidad conjunta de todas las variables relevantes mediante la suma (o marginalización) de todas las combinaciones posibles de valores para las variables no observadas.

Inferencia por Enumeracion

Supongamos que queremos resolver:

$$P(+b \mid +j, +m)$$

Entonces:

$$P(+b \mid +j, +m) = \frac{P(+b, +j, +m)}{P(+j, +m)}$$

$$P(+b, +j, +m) = \sum_{a} \sum_{e} P(+b)P(e)P(a \mid b, e)P(j \mid a)P(m \mid a)$$

Inferencia por Enumeracion

$$
\begin{aligned}
P(+b,+j,+m) = \quad & P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+ \\
& P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+ \\
& P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+ \\
& P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a)
\end{aligned}
$$

Calculamos toda (o casi toda la distribución conjunta) y luego enumeramos variables.

Inferencia por Enumeracion

Este enfoque es conceptualmente sencillo, presenta algunas desventajas importantes:

Complejidad Exponencial: La cantidad de combinaciones posibles crece exponencialmente con el número de variables.

Ineficiencia Computacional: Evaluar todas las combinaciones resulta impracticable para redes grandes.

Alto Consumo de Memoria: Guardar todas las combinaciones posibles consume una gran cantidad de memoria.

Redundancia de Cálculo: Muchas combinaciones son redundantes, ya que se vuelven a calcular probabilidades parciales repetidamente.

ELIMINACION DE VARIABLES

Eliminacion de Variables

La técnica de eliminación de variables aborda estas desventajas al evitar la enumeración completa de todas las combinaciones.

Agrupación de Sumas: Se agrupan y simplifican sumas parciales antes de realizar la marginalización.

Factorización Inteligente: Se explota la estructura de independencia de la red para calcular solo las combinaciones necesarias.

Uso de Factores: Se representan los factores condicionales en lugar de trabajar directamente con la distribución conjunta completa.

Eliminacion de Variables - Factor 1

Podemos identificar los diferentes factores o elementos que estan involucrados dentro de una red bayesiana

Distribución conjunta

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Distribución conjunta parcial (selected) P(x, Y)

$P(cold, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

Eliminacion de Variables - Factor 2

Condicional simple
(suma 1) P(Y |**x**)

$P(W|cold)$

T	W	P
cold	sun	0.4
cold	rain	0.6

Familia de condicionales P(Y |X)

$P(W|T)$

T	W	P
hot	sun	0.8
hot	rain	0.2
cold	sun	0.4
cold	rain	0.6

$P(W|hot)$

$P(W|cold)$

Eliminacion de Variables - Factor 3

Familia especificada (consecuencias) P(y|X)

$P(rain|T)$

T	W	P	
hot	rain	0.2	$P(rain|hot)$
cold	rain	0.6	$P(rain|cold)$

Una vez identificados los elementos o factores dentro de la red bayesiana podemos empezar el proceso de eliminacion con el siguiente ejemplo

Ejemplo – Eliminacion de Variables

Hoy juegan los rojos, puede que haya trafico y llegue tarde a clase.

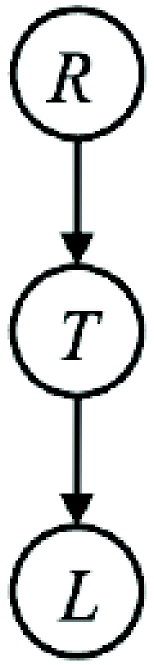
Variables

Juegan los rojos **(R)**
Trafico **(T)**
Llegue tarde a clase **(L)**

¿Llego tarde a clase?

Ejemplo - Eliminacion de Variables

Dada la red bayesiana



$P(R)$

+r	0.1
-r	0.9

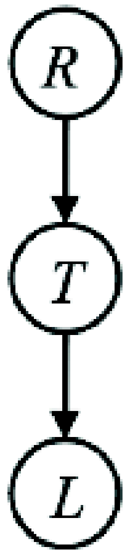
$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

EV - Join



$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

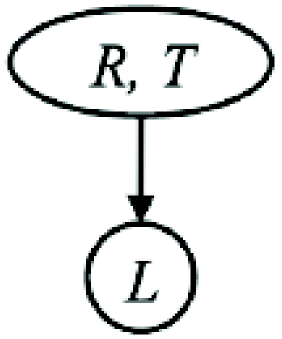
+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

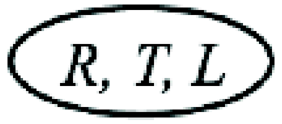
$P(R,T)$

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9





$P(R,T,L)$

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

EV – Eliminacion – Suma

R, T, L → *T, L* → *L*

$P(R,T,L)$

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

$P(T,L)$

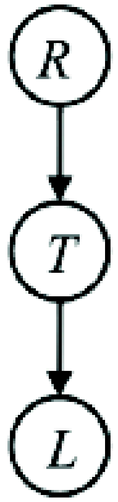
+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

$P(L)$

+l	0.134
-l	0.886

EV - Alternativa

Join → **Eliminacion** → **Join** → **Eliminacion** →



$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

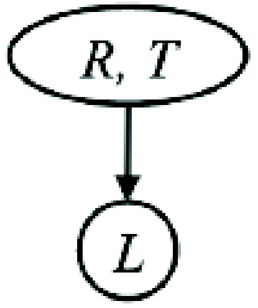
+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

$P(R,T)$

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

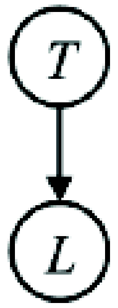


$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

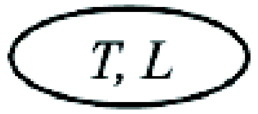
$P(T)$

+t	0.17
-t	0.83



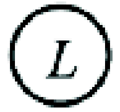
$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9



$P(T,L)$

+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

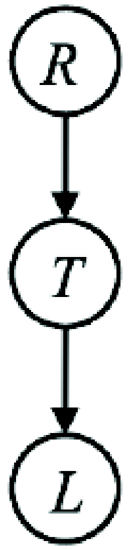


$P(L)$

+l	0.134
-l	0.886

Ejemplo – Eliminacion de Variables

Dada la red bayesiana



$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

Consultar

$$P(L|+r)$$

Ejemplo - Eliminacion de Variables

Aplicamos condicion simple P(L |+r), eliminamos -r

$P(+r)$

+r	0.1

$P(T|+r)$

+r	+t	0.8
+r	-t	0.2

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

Ejemplo – Eliminacion de Variables

Join

$P(+r)$

+r	0.1

$P(T|+r)$

+r	+t	0.8
+r	-t	0.2

P(T,+r)

+r	+t	0.08
+r	-t	0.02

Ejemplo - Eliminacion de Variables

Join

P(T,+r)

+r	+t	0.08
+r	-t	0.02

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

P(L,T,+r)

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.02
+r	-t	-l	0.018

Ejemplo – Eliminacion de Variables

Eliminacion T

$$P(+r, L)$$

+r	+l	0.026
+r	-l	0.074

Ejemplo – Eliminacion de Variables

Normalizar

$P(+r, L)$

+r	+l	0.026
+r	-l	0.074

$$P(L \mid +r) = \frac{P(L,+r)}{P(+r)}$$

$P(L| + r)$

+l	0.26
-l	0.74

$$P(+r) = P(+r,+l) + P(+r,-l) = 0.026 + 0.074 = 0.1$$

P(+r) es la probabilidad marginal de +r, que se obtiene sumando P(L,+r) sobre todos los valores posibles de L. Así, los valores de la tabla se “reescalan” de modo que sumen 1, convirtiéndose en una dist. condicional

PROCEDIMIENTO GENERAL

1) Obtenemos una consulta

2) Iniciamos con los factores (probabilidades conjuntas c/ evidencia)

3) Mientras existan variables ocultas se hace join y eliminaciones

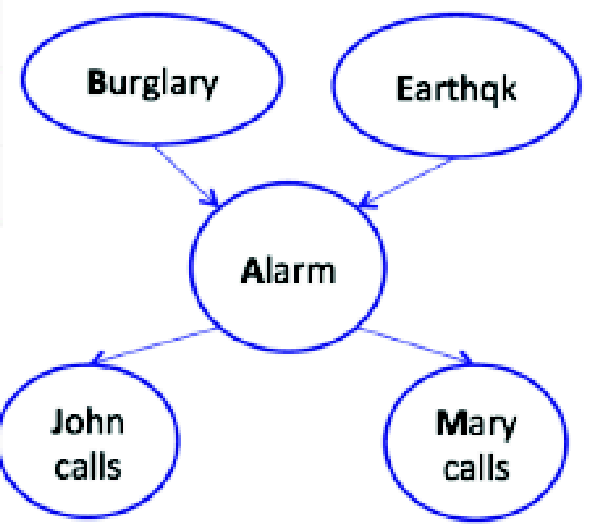
4) Normalizamos

Ejemplo – EV en Red Bayesiana – Robo

$P(B|j,m)$

Consulta

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

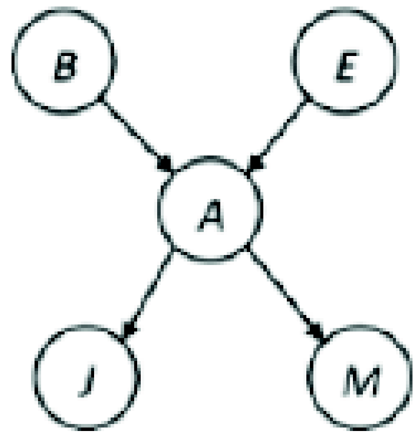
Ejemplo – EV en Red Bayesiana – Robo

Identificamos Variables

Consulta: B

Evidencia: j,m

Ocultas: E, A



Identificamos involucradas en la red segun variables

$$P(B|j,m) \propto P(B,j,m)$$

$$\boxed{P(B) \quad P(E) \quad P(A|B,E) \quad P(j|A) \quad P(m|A)}$$

Ejemplo – EV en Red Bayesiana – Robo

Buscamos eliminar variables ocultas, empezando con A

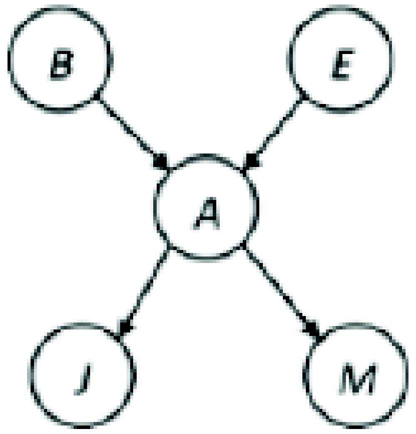
A

Join **Eliminacion**

$$
\begin{matrix} P(A|B,E) \\ P(j|A) \\ P(m|A) \end{matrix} \xrightarrow{\times} P(j,m,A|B,E) \xrightarrow{\sum} P(j,m|B,E)
$$

entonces

$$
\boxed{P(B) \qquad P(E) \qquad P(j,m|B,E)}
$$



Ejemplo - EV en Red Bayesiana - Robo

Continuamos eliminando E, que es una variable oculta

despues de remover A

$P(B) \quad P(E) \quad P(j, m|B, E)$

E

$P(E)$

$P(j, m|B, E)$

Join

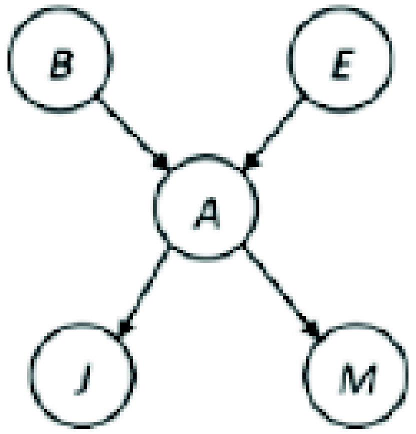
$\xrightarrow{\times}$ $P(j, m, E|B)$

Eliminacion

$\xrightarrow{\Sigma}$ $P(j, m|B)$

despues de remover E

$P(B) \quad P(j, m|B)$



Ejemplo – EV en Red Bayesiana – Robo

despues de remover E

$P(B) \quad P(j, m|B)$

B

Join

$$P(B),\ P(j,m|B) \xrightarrow{\times} P(j,m,B) \xrightarrow{\text{Norm}} P(B|j,m)$$

Ya tenemos las variables, solo queda hacer un ultimo Join y normalizar

Tipos de Inferencia en Redes Bayesianas

- **Inferencia Exacta**
  - Propósito y Función
    - Es el cálculo preciso de la probabilidad de una variable de interés dada un conjunto de evidencias.
    - Se basa en la teoría de la probabilidad y utiliza técnicas específicas para obtener resultados sin aproximaciones.
  - Métodos de Inferencia Exacta
    - Enumeración Completa: Se calcula la probabilidad marginalizando sobre todas las combinaciones de valores de las variables ocultas. Es efectivo en redes pequeñas, pero su complejidad crece exponencialmente con el número de variables.
    - Eliminación de Variables (Variable Elimination): Aplica la regla de Bayes y elimina variables innecesarias de manera eficiente, optimizando el cálculo en comparación con la enumeración completa.
    - Propagación en Árboles de Juntas (Junction Tree Algorithm): Convierte la red en una estructura manejable que permite inferencias eficientes, siendo el método más adecuado para redes con estructura de árbol.
  - Cuándo Usar Inferencia Exacta
    - ✔ Cuando la red tiene pocas variables o una estructura sencilla, como en forma de árbol.
    - ✔ Cuando es necesario obtener resultados precisos sin recurrir a aproximaciones.
- **Inferencia Aproximada**
  - Propósito y Función
    - Se utiliza cuando la inferencia exacta es computacionalmente inviable, como en redes con miles de nodos.
    - Los métodos de aproximación permiten estimar la probabilidad de interés sin necesidad de cálculos exactos.
  - Métodos de Inferencia Aproximada
    - Muestreo de Monte Carlo: Genera muestras aleatorias de la distribución de la red y utiliza estadísticas para estimar probabilidades. Incluye técnicas como Muestreo de Gibbs y Muestreo por Importancia.
    - Métodos de Propagación de Creencias (Belief Propagation - BP): Aproxima las probabilidades mediante el envío de mensajes entre nodos. Es eficaz en redes con estructura de árbol, pero no es exacto en redes con ciclos.
    - Aproximaciones Variacionales: Se basa en el uso de una distribución más simple para aproximar la verdadera distribución de la red, siendo común en modelos de Machine Learning.
  - Cuándo Usar Inferencia Aproximada
    - ✔ Cuando la red bayesiana es grande y la inferencia exacta resulta demasiado costosa.
    - ✔ Cuando se necesitan respuestas rápidas y una estimación es suficiente.