

2.1: Razonamiento Impreciso. Lógica Difusa.

- Imprecisión de la Información. Lógica Difusa. Conjuntos Difusos. Operaciones.
- Proceso Inferencial: Fusificación, Defusificación, Inferencia Difusa.
- Entornos y Aplicaciones

2.2: Incertidumbre. Razonamiento Probabilístico.

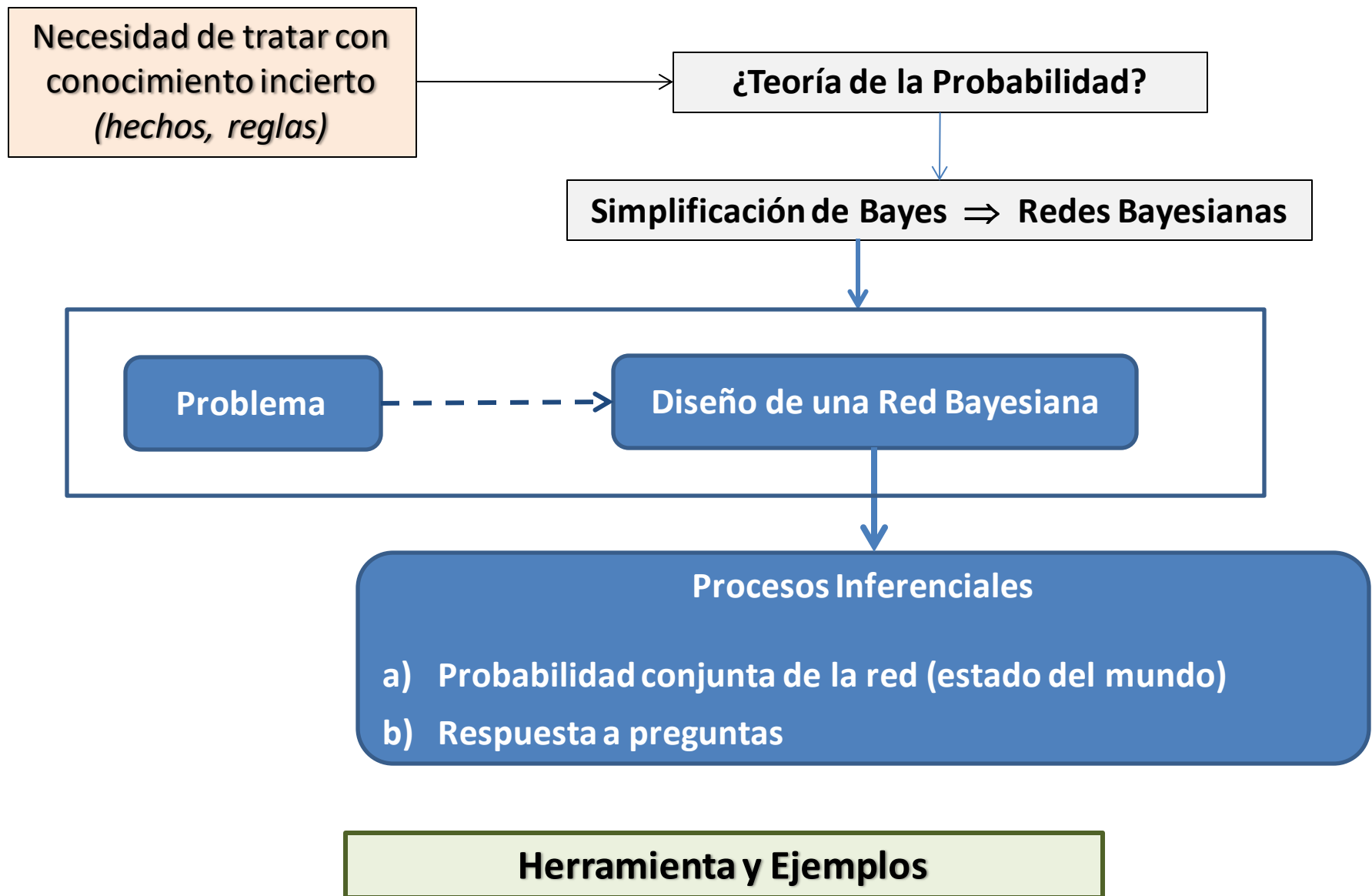
- Conceptos básicos. Aplicación Teoría de la Probabilidad. Modelos simples.
- Redes Bayesianas

Bibliografía

- Inteligencia Artificial: Un enfoque moderno. Rusell, Norvig. Prentice Hall, 2004. Cap. 13, 14, 16
- Inteligencia Artificial. Técnicas, métodos y aplicaciones. Varios autores. McGraw Hill (2008) Cap.6-7
- Inteligencia Artificial. Una nueva síntesis. N. Nilsson McGraw Hill (2000).

Otras referencias:

- FuzzyClips (desarrollado por National Research Council , Canada)
- European Centre for Soft Computing (<http://www.softcomputing.es/>)



Trabajo Académico: Diseño y evaluación de una red bayesiana

Incertidumbre

En un **contexto de incertidumbre**, una **decisión racional** (un plan, diagnóstico, conclusión, respuesta, etc.) depende de los grados de creencia de los distintos **datos y relaciones** que la soportan:

Planifico un viaje de 30' para llegar al aeropuerto.

- No debo inferir llegar al aeropuerto a tiempo, sino: *“espero llegar a tiempo si el coche no se avería, si no se queda sin gasolina, si no hay un gran atasco, etc.”*
- Hay probabilidad de retrasos, tal que un mayor margen de tiempo daría más seguridad, pero también aumentaría la posibilidad de una espera improductiva.

Un paciente refiere un dolor de muelas.

- Ello puede ser debido a una “caries” o a una “infección”.
- En base a ciertas evidencias, causalidad e inferencia probabilística podré confirmar con mayor o menor probabilidad el diagnóstico.

En general, los sistemas no representan todo el conocimiento, ni tienen acceso a todo el conocimiento sobre el entorno:

- **Simplificaciones:** Se obvian antecedentes no relevantes o de ocurrencia infrecuente.
 - **Ignorancia teórica:** No es habitual tener un conocimiento teórico completo (al 100%) del dominio.
 - **Ignorancia práctica:** Fallo de datos, pruebas, medidas, fiabilidad observaciones, etc.
 - **Mundo real no determinista:** consecuencias inesperadas, incidencias externas, etc.
- Hay incertidumbre en los hechos, en las condiciones o en las consecuencias de una acción.
- Una regla **“antecedente → consecuente”** no es una consecuencia lógica, sino un **grado de creencia** sobre la consecuencia, dado un antecedente (quizás con su propio **grado de creencia**).

La herramienta básica para tratar con el razonamiento incierto es la Teoría de la Probabilidad

Esquema de Razonamiento Incierto

Sistema Lógico	Sistema con Incertidumbre
e : antecedente, h : consecuente Implicación : $e \rightarrow h$	e : Evidencias, h : Hipótesis. Regla Incierta: $e \rightarrow h$, con una creencia asociada

Evidencia: Razones (datos) por las que las conclusiones (hipótesis) son creídas.

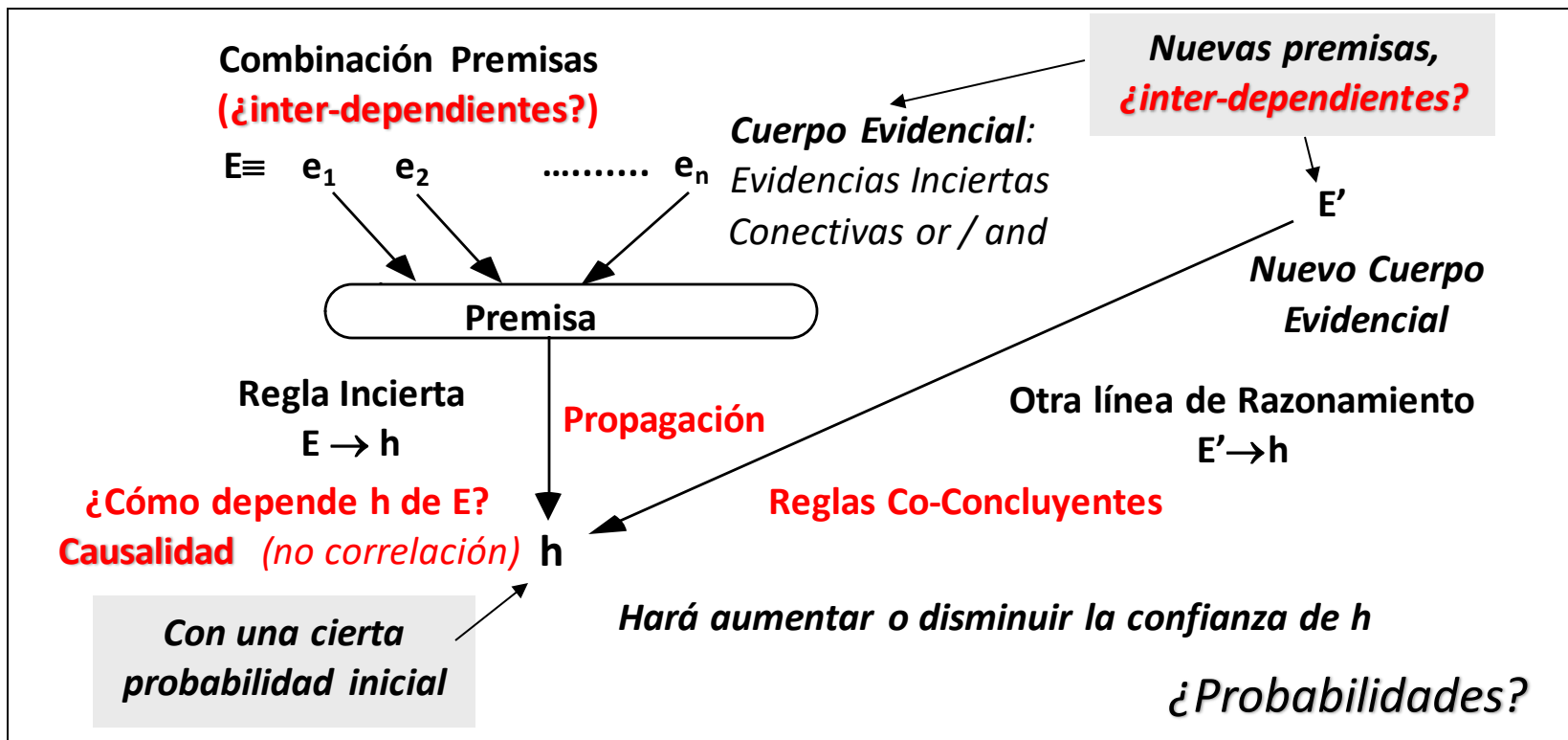
Las evidencias son observaciones (datos) con *grados de creencia*.

Regla incierta:

La *presencia/ausencia* de evidencias soportan/rebajan la creencia sobre hipótesis: $(e \rightarrow h)$, $(\neg e \rightarrow \neg h)$

- Se extiende la noción de 'implicación' para representar el grado con el que la hipótesis h es confirmada (*o rebajada*) por la presencia (*o ausencia*) de la evidencia e .
- La certeza sobre una hipótesis se obtendrá en base a una acumulación de observaciones competitivas (evidencias).
- Nuevas y discriminantes evidencias deben incrementar o decrementar la creencia sobre la hipótesis.

$$e_1 \rightarrow h, \quad e_2 \rightarrow h, \quad e_3 \rightarrow h, \quad \dots, \quad e_n \rightarrow h$$



Ha habido cierta discusión sobre si la **Teoría de la Probabilidad (formal)** es adecuada para representar la incertidumbre en un sistema inteligente.

Principales argumentos en contra:

- Probabilidad NO es lo mismo que creencia. Dificultad para modelar el conocimiento.
- Complejidad del razonamiento. Interdependencia de premisas y/o conclusiones.
 - Dos variables X, Y son independientes sii $\forall x \in X, \forall y \in Y: P(x, y) = P(x) * P(y)$
 - Dos variables están correlacionadas sii no son independientes.
- Causalidad implica correlación, ¡pero no a la inversa! Ej. ¿comer más helados implica más ahogados en piscina?

INCONVENIENTE-1: “Creencia” no es lo mismo que “Probabilidad”

- Un grado de creencia 0.8 *no significa que “es cierto en un 80% de los casos”*, sino una alta expectativa (80%) de que se cumpla.
- Además, *la creencia suele ser subjetiva*: El humano maneja creencias (subjetivo), que se deben cuantificar (probabilidad), y ello tiene una implicaciones matemáticas no previstas.

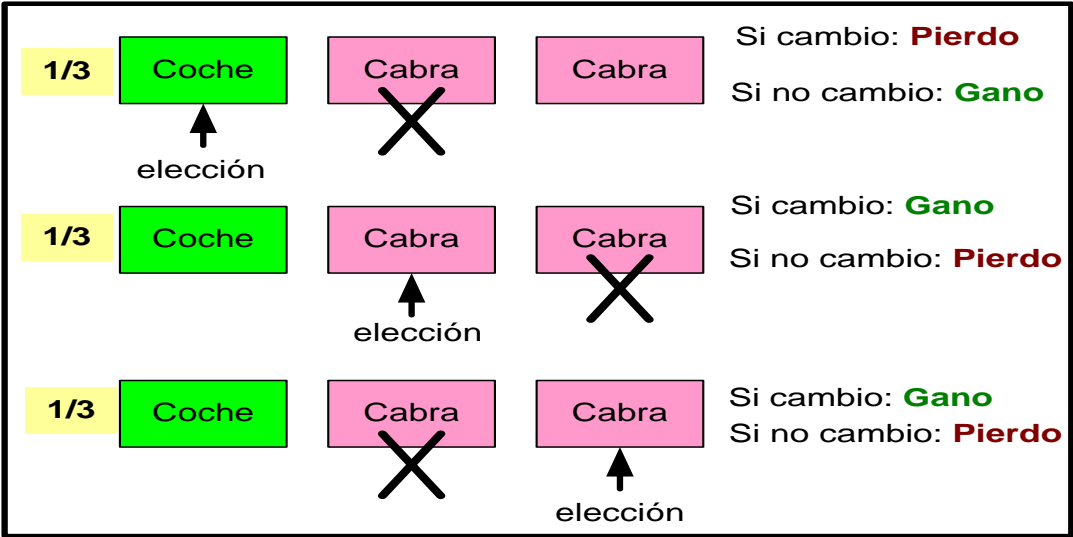
Supongamos que, en un concurso, podemos escoger entre tres puertas: detrás de una de ellas hay un coche, y detrás de las otras dos, una cabra. Escogemos una puerta y, el presentador, que sabe lo que hay detrás de las puertas, abre otra que contiene una cabra (*Paradoja de Monty Hall*)

[VIDEO](#)

Entonces pregunta: “¿Quiere cambiar de puerta?” ¿Es mejor cambiar la elección?

Una suposición errónea es que, una vez sólo quedan dos puertas, ambas tienen la misma probabilidad (50%) de contener el coche.

Es falso ya que el presentador abre la puerta después de la elección de jugador. La probabilidad que había de que el coche estuviera en la puerta que abre se acumula en la tercera puerta.



Una adecuada decisión debe basarse en una adecuada representación del conocimiento.

La probabilidad de ganar el coche si cambiamos de puerta es de 2/3 (66 %),

si no cambiamos, es de 1/3 (33 %)

NOCIONES BÁSICAS PROBABILIDAD

- Probabilidad de un hecho: $P(x)$ *(Se asimilaría a la Creencia en x , como un hecho individual)*

Los grados de creencia, representados como probabilidades, deben cumplir los axiomas de la teoría de la probabilidad.

$$0 \leq P(a) \leq 1, \quad P(a) + P(\neg a) = 1, \quad P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

- *Probabilidad de un estado concreto del mundo*: Se requiere la probabilidad conjunta de n -hechos interdependientes (no podemos asegurar que sean independientes): $P(x_1, x_2, \dots, x_n)$.

Con variables *discretas*, la *probabilidad conjunta* \Rightarrow una tabla de $|x_1| * |x_2| * \dots * |x_n|$ entradas, donde el sumatorio de todas las probabilidades es 1: $\sum_{i=1,n} P(x_i) = 1$

Ejemplo:
 $x_1, x_2 \in \{T, F\}$

	$x_1=T$	$x_1=F$
$x_2=T$	0.4	0.3
$x_2=F$	0.1	0.2

Por ejemplo: $P(x_1=T, x_2=T) = 0,4$

La **complejidad resulta $|x_i|^n$** , luego es **muy complejo** representar la probabilidad conjunta de un conjunto de datos.

Inferencia: Podemos utilizar la Probabilidad Condicional. ¿Probabilidad b , dado a ?: $P(b/a) = P(a \wedge b) / P(a)$

Por ejemplo: Síntoma: x_1 : dolor, Diagnóstico: x_2 : gripe ¿ $P(x_2:\text{gripe}=T / x_1:\text{dolor}=T)$?

$$P(x_2:\text{gripe}=T / x_1:\text{dolor}=T) = P(x_2:\text{gripe}=T \wedge x_1:\text{dolor}=T) / P(x_1:\text{dolor}=T) = 0.4 / (0.4 + 0.1) = 0.8$$

La aplicación de la probabilidad condicional para realizar inferencias requiere disponer de la tabla que representa la probabilidad conjunta. Es inviable en problemas realistas: Complejidad $|x_i|^n$

Para realizar inferencias podemos aplicar la Regla de Bayes:

- Dado (a) y una regla ($a \rightarrow b$)
- Regla de Bayes: $P(b/a) = P(a/b) P(b) / P(a)$

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

Aplicando e
igualando miembros

Donde :

$P(b/a)$: Probabilidad de b, dada a

$P(a/b)$: probabilidad de observar a, dada b

$P(b)$: Probabilidad a priori de b

$P(a)$: Probabilidad a priori de a

Por ejemplo:

Conocemos: ***Dolor_de_cabeza*** \rightarrow ***Gripe***

Observamos que una persona le duele la cabeza y sabemos que:

Probabilidad a priori de dolor: 0.05

Probabilidad a priori de gripe: 0.00002

Probabilidad de tener dolor, si tiene gripe: 0.5

*Con una regla clásica,
podríamos deducir 'Gripe'
mediante el hecho
'Dolor_de_Cabeza'*

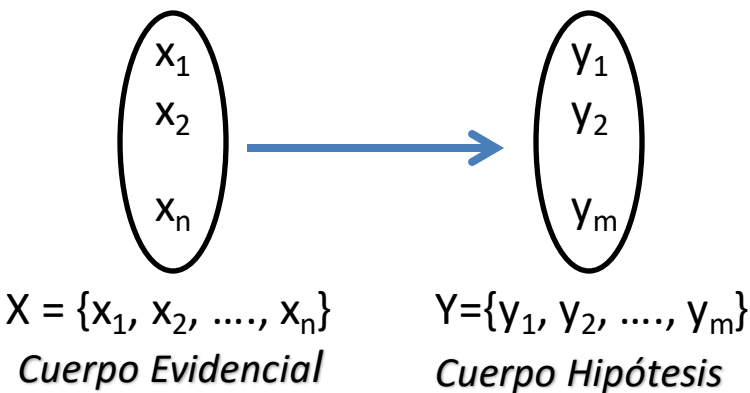
$$P(\text{gripe/dolor}) = P(\text{dolor/gripe}) P(\text{gripe}) / P(\text{dolor}) = 0.5 * 0.00002 / 0.05 = 0.0002$$

El razonamiento para obtener $P(b/a)$ requiere
conocer las probabilidades a priori, $P(a)$ y $P(b)$, y la probabilidad dependiente $P(a/b)$

Regla de Bayes General

En general disponemos de un

- **conjunto de datos** $X=\{x_1, x_2, ..., x_n\}$,
que permiten deducir un
- **conjunto de resultados** $Y=\{y_1, y_2, ..., y_m\}$:



Entonces, $P(Y/X) = P(X/Y) * P(Y) / P(X) =$

$$= \frac{P(\{x_1, x_2, ..., x_n\} / \{y_1, y_2, ..., y_m\})}{P(\{x_1, x_2, ..., x_n\})} * \frac{P(y_1, y_2, ..., y_m)}{P(\{y_1, y_2, ..., y_m\})}$$

La aplicación de esta regla requiere:

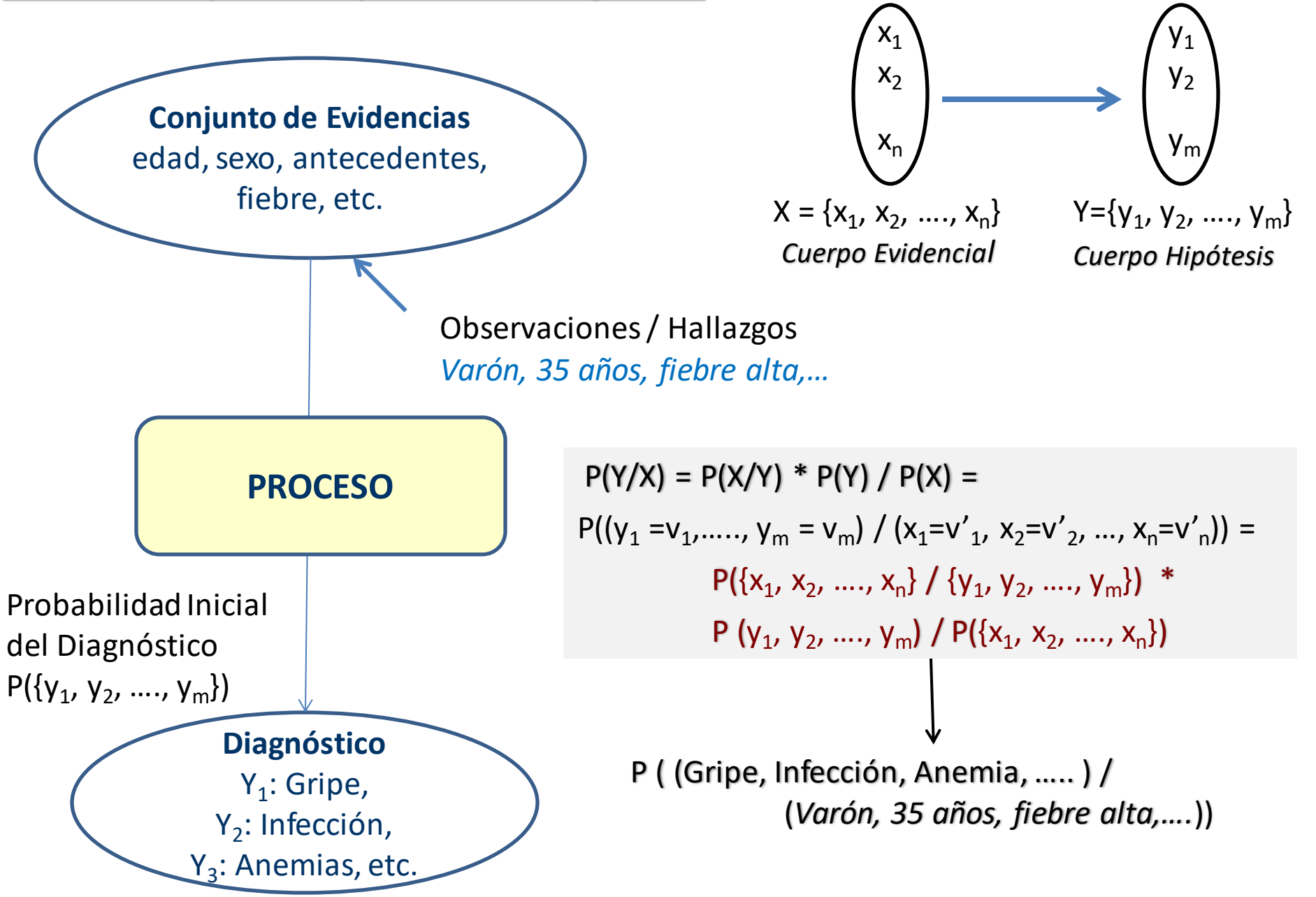
- **Representar la interdependencia de las variables X:** Probabilidad conjunta de X; de cada combinación del valor de las variables $\{x_1, x_2, ..., x_n\}$: Tabla con una complejidad $|x_i|^n$.
- **Representar la interdependencia de las variables Y:** Probabilidad conjunta de Y, de cada combinación del valor de las variables $\{y_1, y_2, ..., y_m\}$: Tabla con una complejidad $|y_i|^m$.
- **Representar la relación causal entre las variables X e Y:** Probabilidades condicionales $P(X/Y)$, *observaciones de X dadas Y*, para cada combinación posible de las variables de X con cada combinación posible de las variables de Y. Complejidad: $|x_i|^n * |y_i|^m$

Además, suele requerirse **encadenamiento**: puede ocurrir que se tenga la observación de un conjunto de evidencias $E' \{e_1, e_2, ..., e_n\}$, que permita obtener la probabilidad sobre los antecedentes X, etc.

$$E \Rightarrow X \Rightarrow Y \quad \text{debiéndose calcular: } P(Y/X,E) = P(X/Y, E) * P(Y/E) / P(X/E)$$

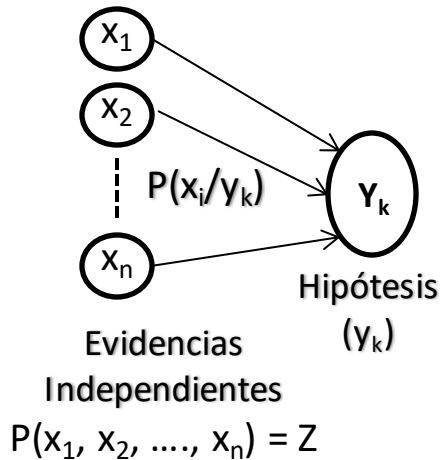
La aplicación formal de Bayes es inviable en la práctica (computacionalmente costosa)

Formulacion Típica de un problema de Diagnóstico



Simplificación: Modelo Simple de Bayes. Asume que:

- (i) las variables evidencia $X=\{x_1, x_2, \dots, x_n\}$ son independientes entre sí, e independientemente observables, y
- (ii) las variables conclusión $\{y_k\}$ son excluyentes entre sí.



Ahora, $P(Y/X) = P(X/Y) * P(Y) / P(X)$ resulta para cada hipótesis:

$$P(y_k, (x_1, x_2, \dots, x_n)) = P(y_k) * \prod_i (P(x_i/y_k)) / Z$$

*Aplicado en
clasificadores
simples bayesianos*

Su aplicación requiere:

- Independencia entre las variables X
- Exclusión mutua entre las variables Y
- Tabla de dependencia condicional $P(x_i/y_k)$

⇒ **Redes Bayesianas**

Otras simplificaciones: se han propuesto diferentes modelos para tratar la incertidumbre, más simples, pero informalmente basados en Bayes:

- **Modelo de Probabilidades Subjetivas:** Factores de Suficiencia (LS) y Necesidad (LN),
- **Teoría Evidencial;** Factores de Credibilidad (MD) e Incredibilidad (MB),
- **Factores de Certeza** (Mycin). Simplificadamente: $a (cf1), a \rightarrow b (cf2): b (cf1 * cf2)$

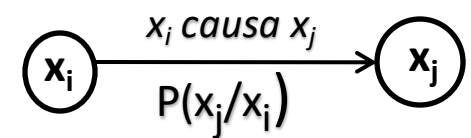
La aplicación de estos modelos simples conlleva muchos riesgos de obtener resultados no adecuados.

- Término acuñado por Judea Pearl (1985)
- **Modelo probabilista** inspirado en la **causalidad**.
- Una Red Bayesiana proporciona una descripción completa de las dependencias (**relaciones de causalidad**) en el dominio.
- Asociado a un **modelo gráfico**, cuyos nodos representan variables y los arcos las relaciones causales
- **Simplifica** enormemente el razonamiento probabilístico (a cambio de unas guías de diseño de la red bayesiana): **Gran desarrollo y aplicabilidad**
- Permite **actualizar la información** que se concluye (nodos-objetivo) a partir de una información de entrada (nodos-evidencia), posiblemente subjetiva, y de las relaciones causales de la red
- Inferencias de tipo **abductivo (causas) y predictivo (efectos)**.
- Desarrollo de **algoritmos eficientes**, modelos de **diagnóstico** y entornos de aplicación

Redes Bayesianas (modelo gráfico probabilístico)

Grafo acíclico que representa en sus arcos las dependencias condicionales (probabilísticas) entre las variables del dominio:

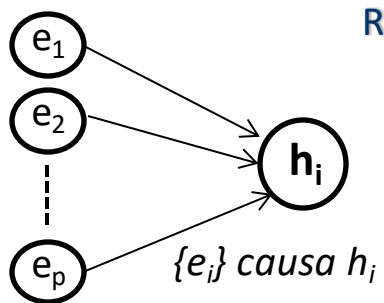
- **Nodos:** Variables aleatorias (datos o conclusiones). Asumimos {T, F}.
- **Arcos:** Dependencia condicional entre variables: $x_i \rightarrow x_j$ (x_i padre de x_j)



x_i	$P(x_j=T/x_i)$	$P(x_j=F/x_i)$
T		
F		

Las variables pueden ser dependientes (conclusiones) o independientes (datos o evidencias):

- Variables independientes, $\{e_i\}$, tienen probabilidades a priori: $P(e_1), P(e_2), \dots, P(e_k)$, o bien son observadas y tienen un valor dado $P(e_i)=1$
- Variables dependientes (h_i) tienen una distribución de probabilidad condicionada:



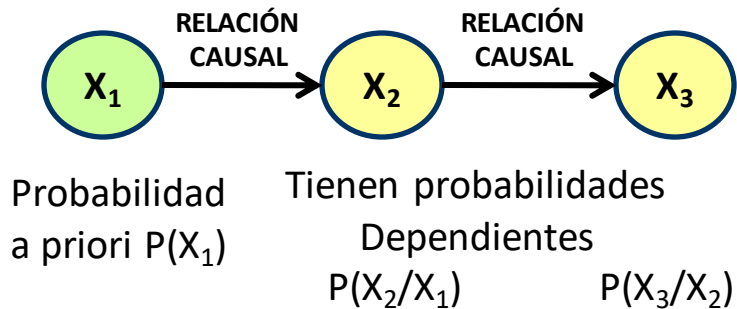
Representada mediante una **Tabla de Probabilidad Condicional**

e_1, e_2, \dots, e_p	$P(h_i=T/(e_1, e_2, \dots, e_p))$
Si $e_i \in \{T, F\}$, 2^p combinaciones	

$P(h_i=F/(e_1, e_2, \dots, e_p))$
es el complementario

La Tabla de Probabilidad Condicional representa la dependencia causal que existe entre los valores que toma la variable h_i respecto a los valores que toman su variables padres (e_1, e_2, \dots, e_p)

Ejemplos de Redes Bayesianas



Ejemplo de probabilidad priori

$P(X_1=T)$	$P(X_1)=F$
0.8	0.2

*La relación de dependencia causal **incrementa/decrementa** la probabilidad de la conclusión dependiendo del **cumplimiento o no** de las premisas.*

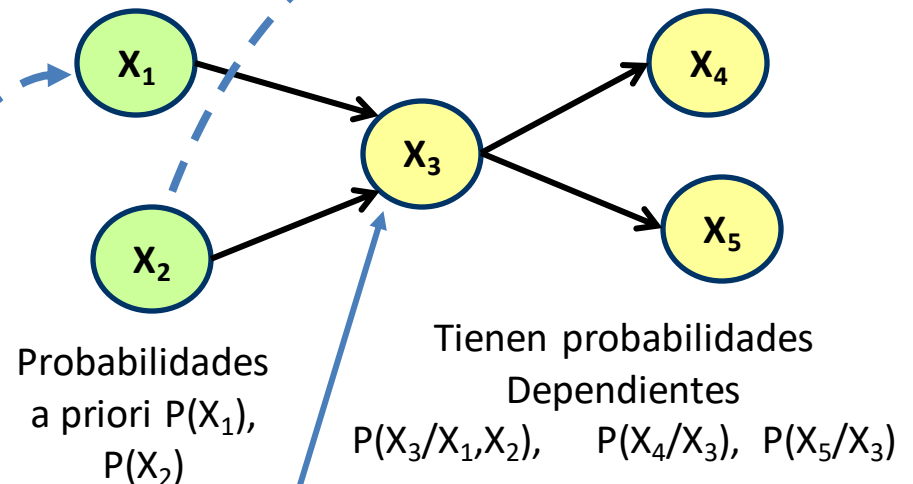
$X_i \rightarrow X_j$

\Rightarrow

Dependiendo de los valores de la tabla.

- La observación de x_i aumenta/disminuye la probabilidad de x_j
- La no-observación de x_i disminuye/aumenta la probabilidad de x_j

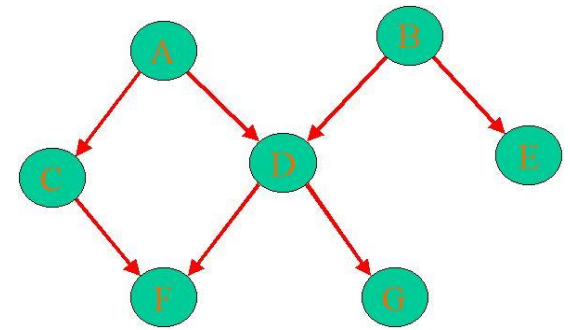
Ejemplo Variable Observada
 $P(X_2=T)=1$



Ejemplo de probabilidad condicional

X_1	X_2	$P(X_3=T)$	$P(X_3=F)$
$X_1=T$	$X_2=T$	0.6	0.4
$X_1=T$	$X_2=F$	0.2	0.8
$X_1=F$	$X_2=T$	0.8	0.2
$X_1=F$	$X_2=F$	0.7	0.3

Una red bayesiana es una representación correcta del dominio si “cada nodo es condicionalmente independiente (de forma directa) del resto de variables (excepto de sus padres).” *¡Problema de diseño!*

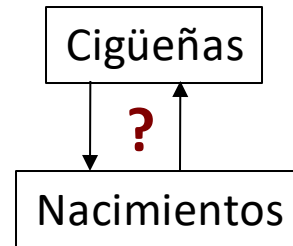


También hay que distinguir entre **causalidad** y mera **correlación**
(¡La causalidad implica correlación, pero no a la inversa!)

Ejemplo (Díez, UNED):

Un estudio llevado a cabo en Inglaterra demostró que había una fuerte correlación entre el número de cigüeñas de cada localidad y el número de nacimientos.

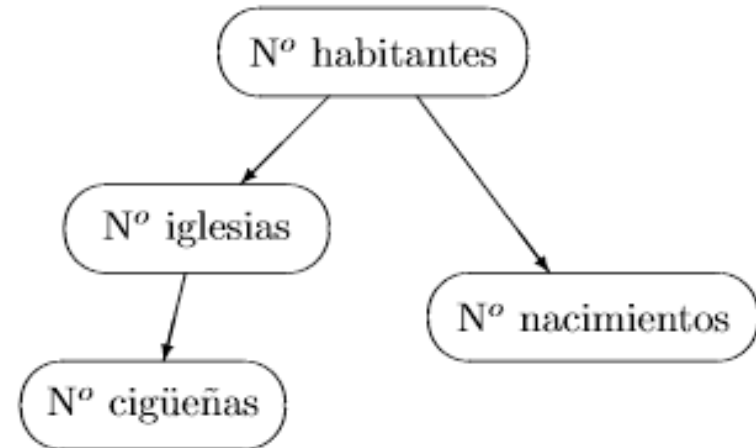
- ¿Podría utilizarse este hallazgo para afirmar que las cigüeñas traen a los niños?
- ¿O es la presencia de niños lo que causa, o atrae, a las cigüeñas?



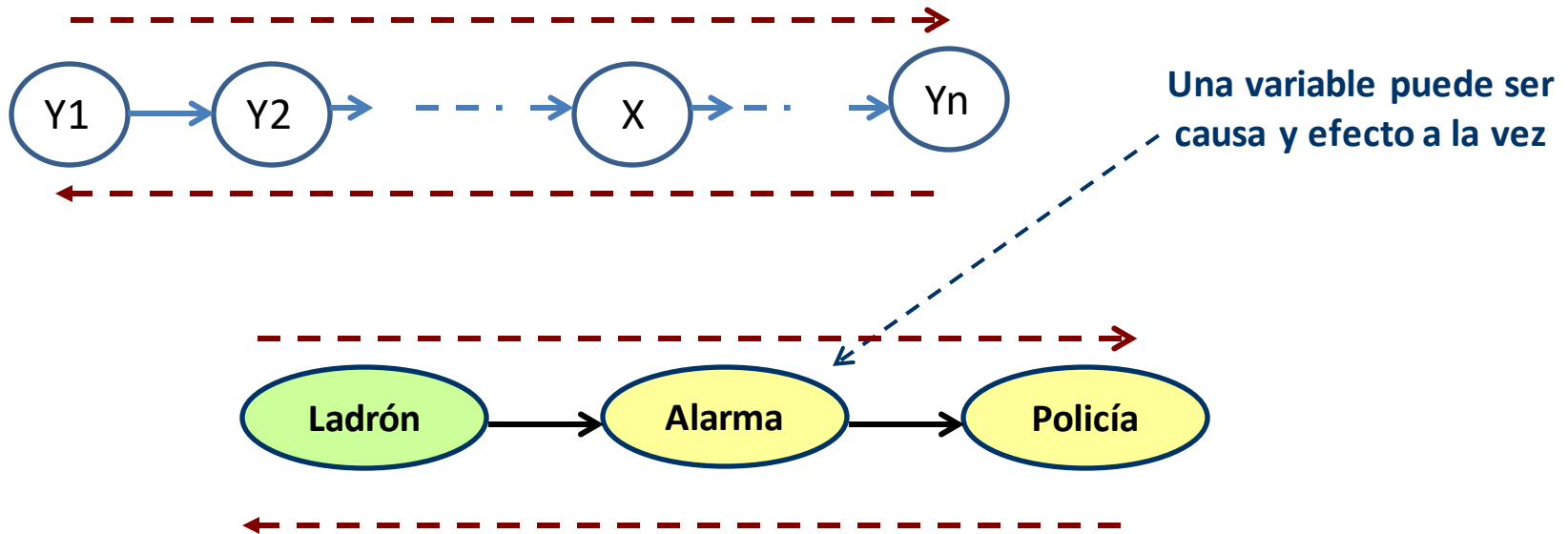
La explicación más simple es que a más habitantes, más iglesias y más campanarios donde las cigüeñas pueden poner los nidos.

Hay una cierta correlación entre habitantes, cigüeñas y nacimientos.

La correlación entre número de cigüeñas y nacimientos no implica causalidad.



1) Conexión Serie: Causa → Efecto

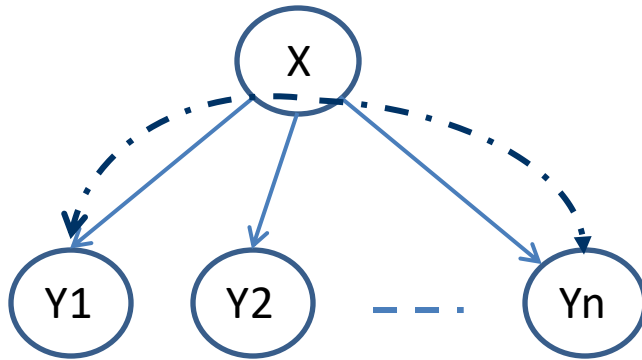


La evidencia se transmite de un extremo a otro, salvo que una variable intermedia esté instanciada:

- **Inducción:** Y_1 condiciona los resultados de Y_2 , que es causa de Y_3 , y así sucesivamente hasta Y_n .
- **Abducción:** Una evidencia sobre Y_n transmite certidumbre hasta Y_1 .

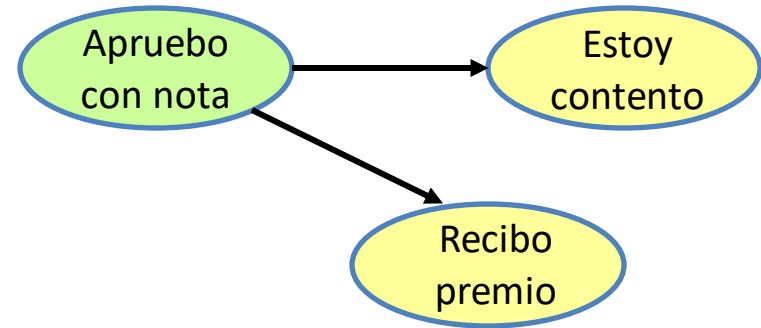
Pero si una variable intermedia X es conocida (está instanciada), las variables Y_1 e Y_n dejan de influirse mutuamente (pasan a ser independientes o separadas).

2) Conexión Divergente: Causa \rightarrow {Efectos}



Una causa X produce diversos efectos Y_i

(una enfermedad X puede manifestarse a través de todos los síntomas Y_i)



Si estoy contento, modifico mi certidumbre (\uparrow) sobre haber aprobado con nota y , a su vez, de recibir un premio (\uparrow)

La evidencia se transmite entre los hijos de una conexión divergente, salvo que la Variable-Padre esté instanciada:

Una información sobre una variable Y_i influye en cualquier otra variable Y_j a través de X .

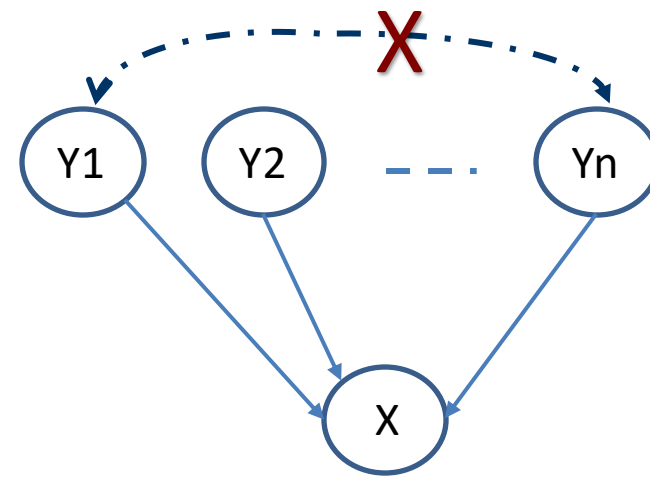
- *Ejemplo: Una evidencia sobre Y_1 modificará nuestra certidumbre en X , y ésta nuestra credibilidad en Y_2, Y_3, \dots, Y_n .*

Pero si X es conocida (*está instanciada*), se bloquea el flujo de información entre Y_i e Y_j , y pasan a ser independientes.

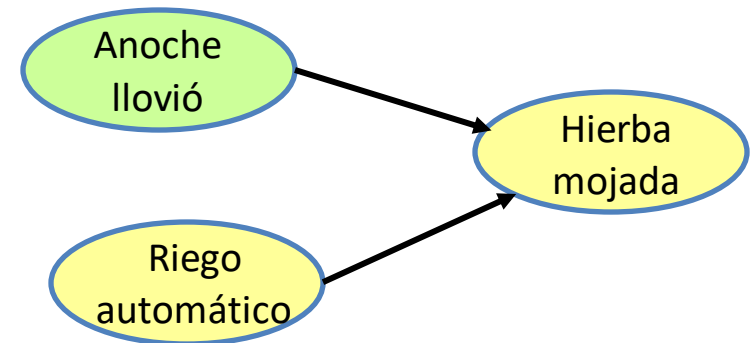
3) Conexión Convergente: {Causas} → Efecto

La evidencia no se transmite entre los padres $\{Y_2, Y_3, \dots, Y_n\}$ de una conexión convergente, salvo que la variable X que las conecta esté instanciada (o uno de sus descendientes sea conocido).

- **Si el efecto X es desconocido**, cada una de sus posibles causas Y_i no aporta información sobre las otras posibles causas: las variables Y_i son independientes.
- **Si el efecto X es conocido** (o una de sus consecuencias/descendientes), la evidencia de una causa Y_i puede producir un cambio de credibilidad en el resto de las causas Y_j .



**Distintas causas Y_i
pueden producir un mismo efecto X**



Si la hierba está mojada, conocer (o no) si anoche llovió puede afectar a saber si funcionó el riego automático.

Ejemplo-1 Estando trabajando, mi vecino Adán me llama diciendo que la alarma de mi casa está sonando. Pero mi vecina Eva no me llama. Sé que algunas veces, a causa de pequeños temblores, la alarma se dispara. ¿Hay un ladrón?

conocimiento \Rightarrow diseño

DISEÑO RED BAYESIANA

Se debe determinar :

- Variables independientes (datos, evidencias)
- Variables dependientes (conclusiones)
- Arcos: dependencias causales entre variables (reglas)

ESQUEMAS:

Causa \rightarrow Efecto
Causa \rightarrow {Efectos}
{Causas} \rightarrow Efecto

Y :

- Probabilidades iniciales de las variables independientes (datos y evidencias)
- Tabla probabilidades condicionales (arcos a las variables dependientes)

Variables (instanciadas sobre {T, F}):

Independientes: Ladrón, Temblor

Dependientes: Alarma, Llamada-Adán, Llamada-Eva

Conocimiento Causal (Red Bayesiana):

- | | | |
|---------------------------------------|---------------|---------------------------------|
| Un ladrón puede activar la alarma | \Rightarrow | Si ladrón entonces alarma |
| Un temblor puede activar la alarma | \Rightarrow | Si temblor entonces alarma |
| La alarma puede causar que Adán llame | \Rightarrow | Si alarma entonces Llamada-Adán |
| La alarma puede causar que Eva llame | \Rightarrow | Si alarma entonces Llamada-Eva |

Ejemplo-1 Estando trabajando, mi vecino Adán me llama diciendo que la alarma de mi casa está sonando. Pero mi vecina Eva no me llama. Sé que algunas veces, a causa de pequeños temblores, la alarma se dispara.

Variables (instanciadas sobre {T, F}):
Independientes: Ladrón, Temblor *Dependientes:* Alarma, Llamada-Adán, Llamada-Eva

Conocimiento Causal (Red Bayesiana):

- Un ladrón puede activar la alarma

Un temblor puede activar la alarma

La alarma puede causar que Adán llame

La alarma puede causar que Eva llame
- ⇒ Si ladrón entonces alarma

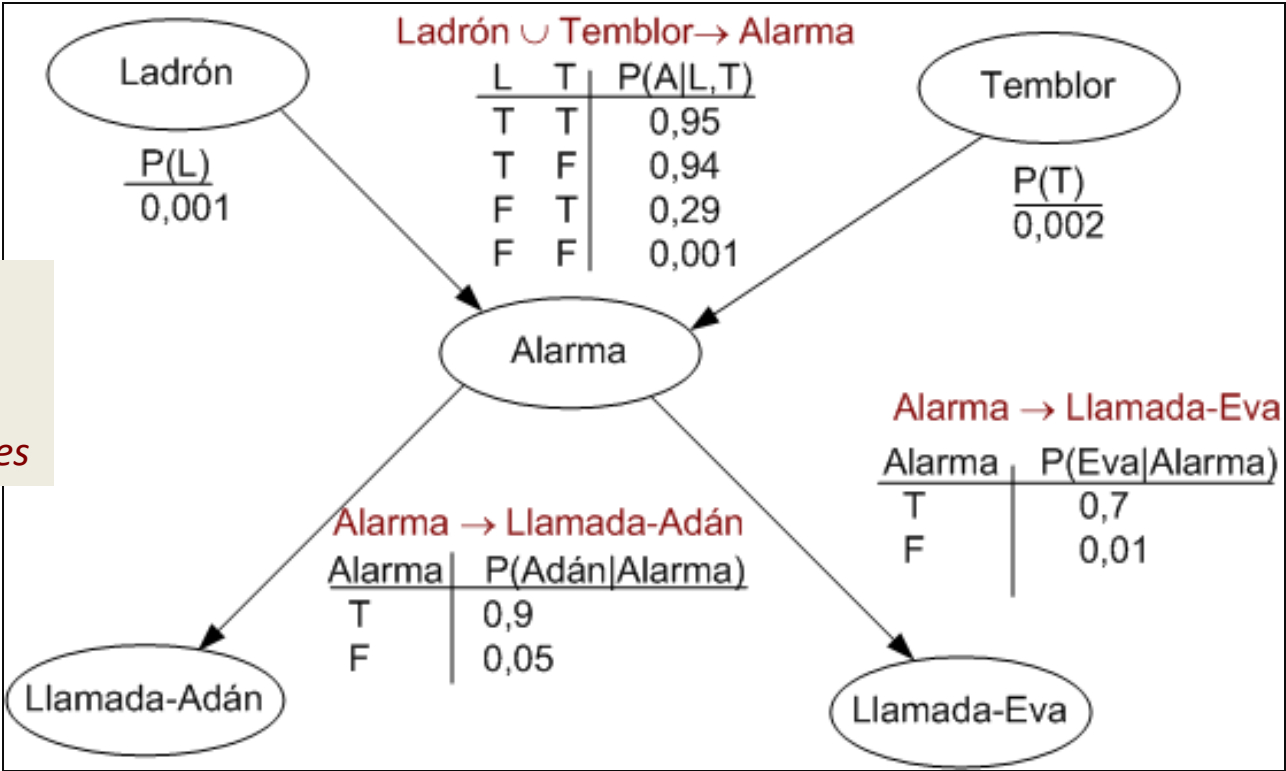
⇒ Si temblor entonces alarma

⇒ Si alarma entonces Llamada-Adán

⇒ Si alarma entonces Llamada-Eva

Diseño de la Red Bayesiana

- 1. Variables y Arcos
- 2. Probabilidades a priori
- 3. Probabilidades condicionales



Ejemplo-2

Un alumno podría obtener su título de carrera (CARRERA) si aprueba una asignatura-A (APROBAR).

También tiene más posibilidades de trabajo (TRABAJO) si acredita conocimientos de la asignatura-A (APROBAR).

El alumno puede superar (APROBAR) la asignatura-A estudiando (ESTUDIA) o copiando (COPIA).

¿Variables, Relaciones Causales?

¿Red Bayesiana?

EJEMPLO-3 (E. Millán)

Juan está **estornudando**. Las causas posibles son que se ha **resfriado** o que tiene **rinitis**. La rinitis puede estar causada porque sus amigos tienen un **gato** y que Juan sea **alérgico** a los gatos. Un **gato** hace que los muebles puedan tener **arañazos**.

¿Por qué estornuda Juan?

Variables Observación (o Evidencia) *(Observables directamente)*

Arañazos
Estornudar

Variables Objetivo *(No observables directamente)*

Juan-Alergia
Juan-Resfriado

Variables Factores, Auxiliares *(Ayudan al modelado)*

Rinitis
Gato

Las Variables Factores pueden ser:

- **Promotoras:** La variable afectada es más probable cuando están presentes.
- **Inhibidoras:** La variable afectada es menos probable cuando están presentes.
- **Requeridas:** Si no presentes, no ocurre la variable afectada.
- **Preventivas:** si están presentes, no ocurre la variable afectada.

Analizando el enunciado:

- *La rinitis puede producir estornudos*
- *El resfriado puede producir estornudos*
- *La existencia de un gato puede producir rinitis si se es alérgico*
- *Un gato puede hacer arañazos*

Ejemplo-4 (Charniak)

Supongamos que quiero saber si mi mujer está en casa, basándome en la siguiente información:

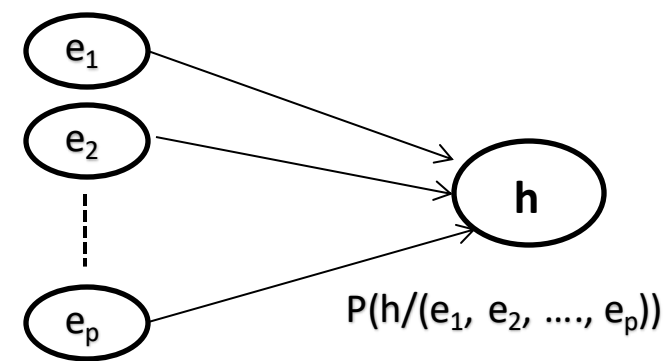
- Si mi esposa sale de casa (está *fuera*), usualmente (pero no siempre) enciende la *luz* de la entrada
- Hay otras ocasiones en las que también enciende la luz de la entrada
- Si *no hay nadie en casa*, el *perro está fuera*
- Si el perro tiene *problemas intestinales*, también se *deja fuera*
- Si el *perro está fuera*, oigo sus *ladridos*
- Podría oír ladridos y pensar que son de mi perro aunque no fuera así

Variables:

- Fuera (mujer no en casa) → Perro, Luz,
- Luz (luz en la entrada),
- Perro (perro fuera) → Ladra
- Inst (problemas intestinales) → Perro
- Oigo (oigo al perro ladrar)

En el diseño debe tener en cuenta que:

En una red bayesiana, la **Tabla de Probabilidad Condicional** representa la **dependencia causal** que existe entre los valores que toma la variable h respecto a los valores que toman su variables padres **independientes** ($e_1, e_2, ..., e_p$):



- **Dependencia entre las variables $\{e_i\}$:** ¿hay evidencias dependientes entre sí, tal que no deben acumular probabilidad sobre h ?

Ejemplo:	Si alguien ve todas las finales de fútbol, Si alguien ve la final de la Copa del Rey, Si alguien es de un equipo de fútbol,	entonces el fútbol es su afición favorita entonces el fútbol es su afición favorita entonces el fútbol es su afición favorita
-----------------	---	---

1. Si conocemos que '*Pedro vio todas las finales de fútbol*', podemos deducir que es aficionado al fútbol.
2. Pero la evidencia de que '*Pedro vio la final de la Copa del Rey*' es totalmente dependiente y no debe incrementar la probabilidad de la hipótesis, ya que no añade nueva información.
3. En cambio, si conocemos que '*Pedro es de un equipo de fútbol*', entonces sí incrementa su probabilidad.

- Distinguir entre evidencias independientes, pero mutuamente excluyentes sobre h (*la creencia de h será debida solo a una de ellas: o toma veneno OREX se dispara*):

$$\text{Toma-veneno} \vee \text{Se-Dispara} \rightarrow \text{Muere}$$

Y evidencias independientes pero no excluyentes sobre h , tal que su existencia acumula creencia sobre h . (*La creencia de h será debida a la acumulación de todas sus evidencias*):

$\text{Mamífero} \rightarrow \text{Humano}$	$\text{Omnívoro} \rightarrow \text{Humano}$
---	---

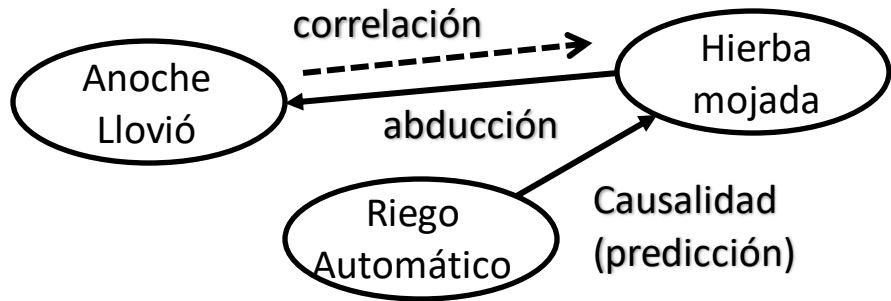
Esto se debe tener en cuenta en la tabla de probabilidades condicionales.

- Si una red bayesiana (*grafo dirigido acíclico*) representa relaciones causales, no debe mezclarse con relaciones abductivas (*ciclos implícitos*).

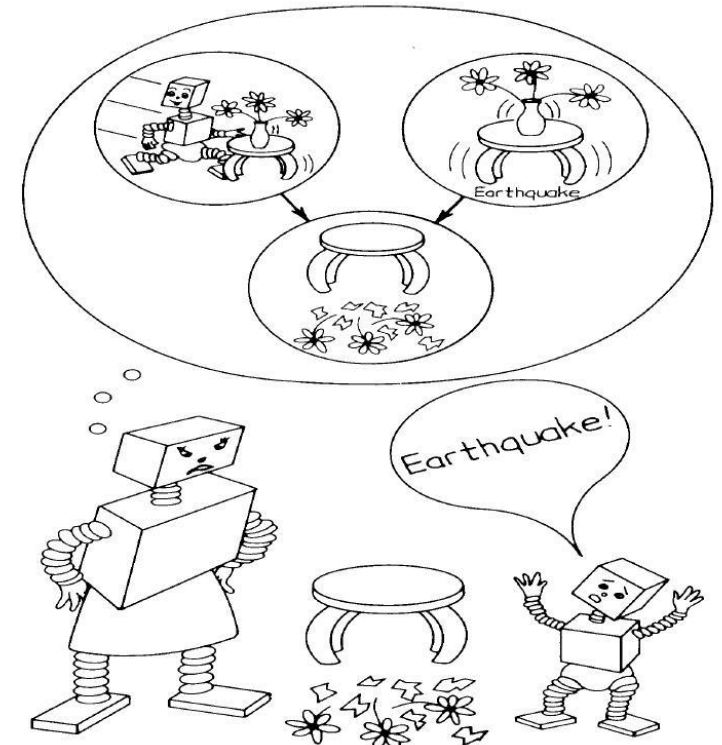
R1: Anoche funcionó el riego automático → Hierba mojada por la mañana

R2: Hierba mojada por la mañana → Anoche llovió

¿Si constatamos que anoche funcionó el riego, podemos llegar a deducir que anoche llovió?



Causalidad implica correlación, pero no a la inversa.



Causa, Causa→Efecto

Efecto

Deducción

Efecto, Causa→Efecto

Causa

Abducción

La mezcla de causales 'causales (deductivas)' (R1) con relaciones de 'diagnóstico (abductivas)' (R2) suele dar resultados catastróficos:

Probabilistic Reasoning in a Causal Network

ESTADO en una Red Bayesiana: Distribución de Probabilidad Conjunta

Distribución de Probabilidad Conjunta de una Red Bayesiana:

Distribución de probabilidad para todas las variables $\{X_i\}$:

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

donde, si x_i es una variable independiente $P(x_i / \text{padres}(x_i)) = P(x_i)$.

La distribución de probabilidad conjunta tiene una entrada para cada asignación posible a las variables.

Si $x_i \in \{T, F\}$, 2^n entradas.

La Distribución de Probabilidad Conjunta puede usarse para responder a cualquier pregunta en el dominio. Por ejemplo:

- Probabilidad de un **estado conjunto de toda la red** (todas las variables): una entrada de $P(x_1, x_2, \dots, x_n)$
- Probabilidad de un **estado parcial de la red** (probabilidad de que algunas variables tomen un valor).
Por ejemplo, la variable x_k no se considera: $P(x_1, x_2, \dots, x_k=T, \dots, x_n) + P(x_1, x_2, \dots, x_k=F, \dots, x_n)$
- Probabilidad de que **alguna variable** tome un valor: $P(x_k=T) / (P(x_k=T) + P(x_k=F))$

En la aplicación de una red bayesiana:

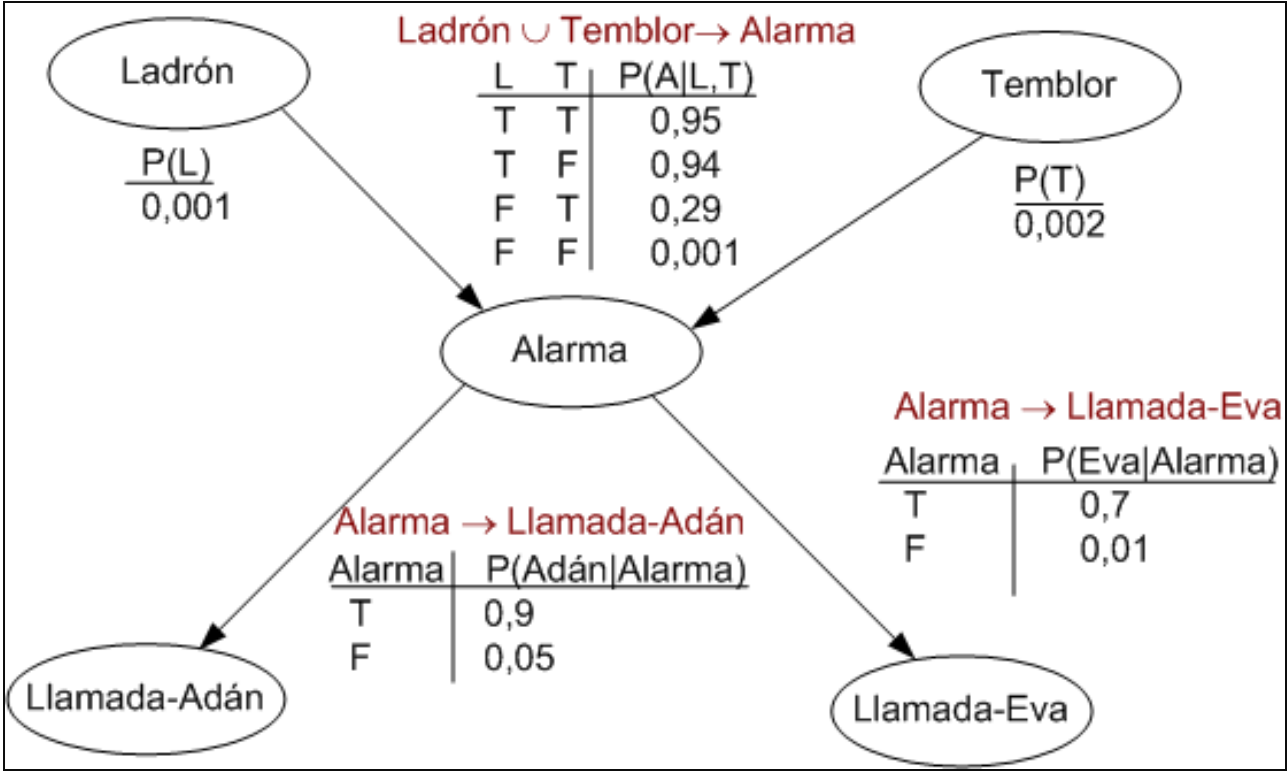
- Las variables independientes pueden tener un valor observado, $P(X_i=x_i)=1$, o probabilidades a priori.
- Existen algoritmos típicos para la propagación de probabilidades en redes bayesianas.

Ejemplo Estando trabajando, mi vecino Adán me llama diciendo que la alarma de mi casa está sonando. Pero mi vecina Eva no me llama. Sé que algunas veces, a causa de pequeños temblores, la alarma se dispara.

Variables (instanciadas sobre {T, F}):
Independientes: Ladrón, Temblor *Dependientes:* Alarma, Llamada-Adán, Llamada-Eva

Conocimiento Causal (Red Bayesiana):

- Un ladrón puede activar la alarma
- Un temblor puede activar la alarma
- La alarma puede causar que Adán llame
- La alarma puede causar que Eva llame
- ⇒ Si ladrón entonces alarma
- ⇒ Si temblor entonces alarma
- ⇒ Si alarma entonces Llamada-Adán
- ⇒ Si alarma entonces Llamada-Eva

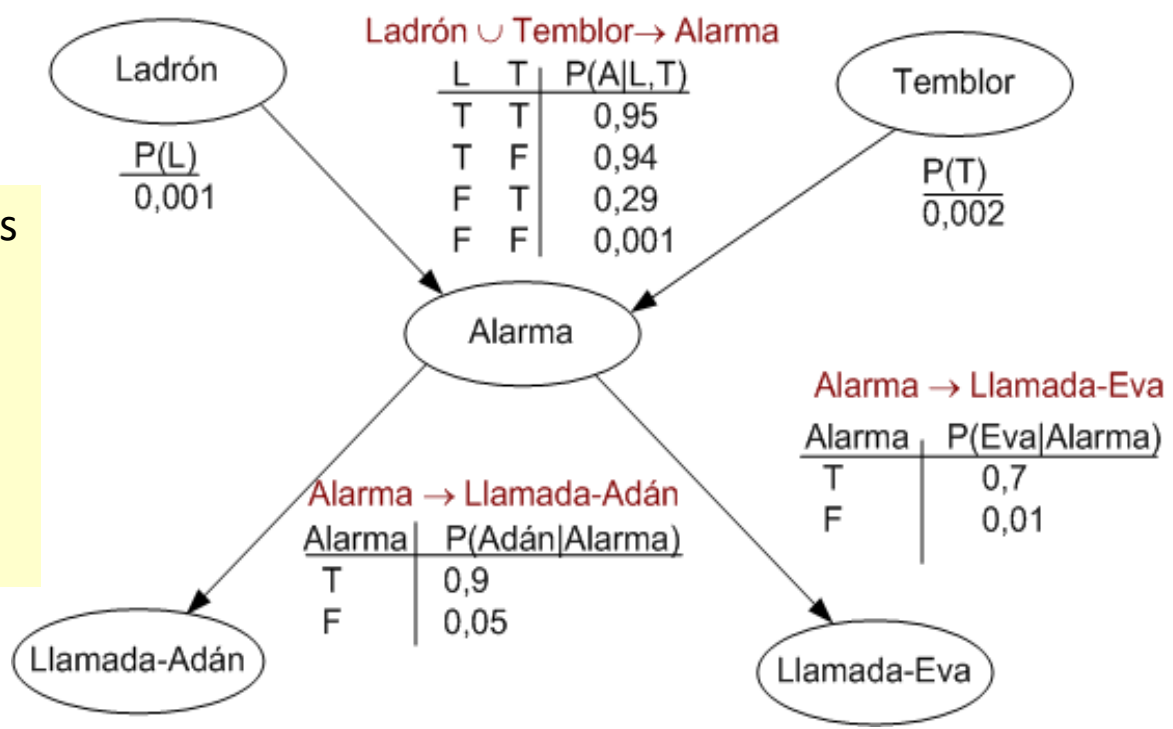


Aplicación: Probabilidad Conjunta de todas las variables

Probabilidad Conjunta de todas las variables $\{x_i\}$:

$$\prod_{i=1,n} P(x_i \mid \text{padres}(x_i))$$

donde,
si x_i es independiente:

$$P(x_i \mid \text{padres}(x_i)) = P(x_i)$$


Mi vecino Adán me llama diciendo que la alarma de mi casa está sonando. Pero mi vecina Eva no me llama. ¿Cuál es la Probabilidad Conjunta de que no sea un temblor y sea un ladrón?

$P(\text{Ladrón}, \neg\text{Temblor}, \text{Alarma}, \text{Adán}, \neg\text{Eva}) =$

$$P(\text{Ladrón}) * P(\neg\text{Temblor}) * P(\text{Alarma} | (\text{Ladrón}, \neg\text{Temblor})) * P(\text{Adán} | \text{Alarma}) * P(\neg\text{Eva} | \text{Alarma}) =$$
$$0.001 * (1 - 0.002) * 0.94 * 0.9 * (1 - 0.7) = \mathbf{0.000253}$$

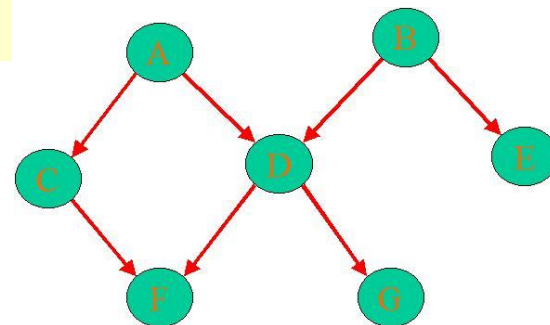
¿Y la probabilidad de que la alarma suene, Adán y Eva me llamen, pero no sea un temblor ni un ladrón?

Respuesta a preguntas (tras la observación de eventos)

Dada una Red Bayesiana, donde se tienen:

E: Conjunto de variables evidencias (*posiblemente observables*)

Y: Conjunto de variables no-evidencia (*no observables*)



Problema: Se realiza la observación de algunas variables-evidencia: $e \subset E$, $P(e) = 1$

Y se requiere una información sobre una variable x : $\hat{P}(x=v)$?

El objetivo es obtener $P(x=v/e)$:

$$P(x=v/e) = \alpha \sum_Y P(x=v, e, y).$$

Sumatorio sobre todas las combinaciones de valores posibles de las variables Y (no observables), dada la observancia de las variables e

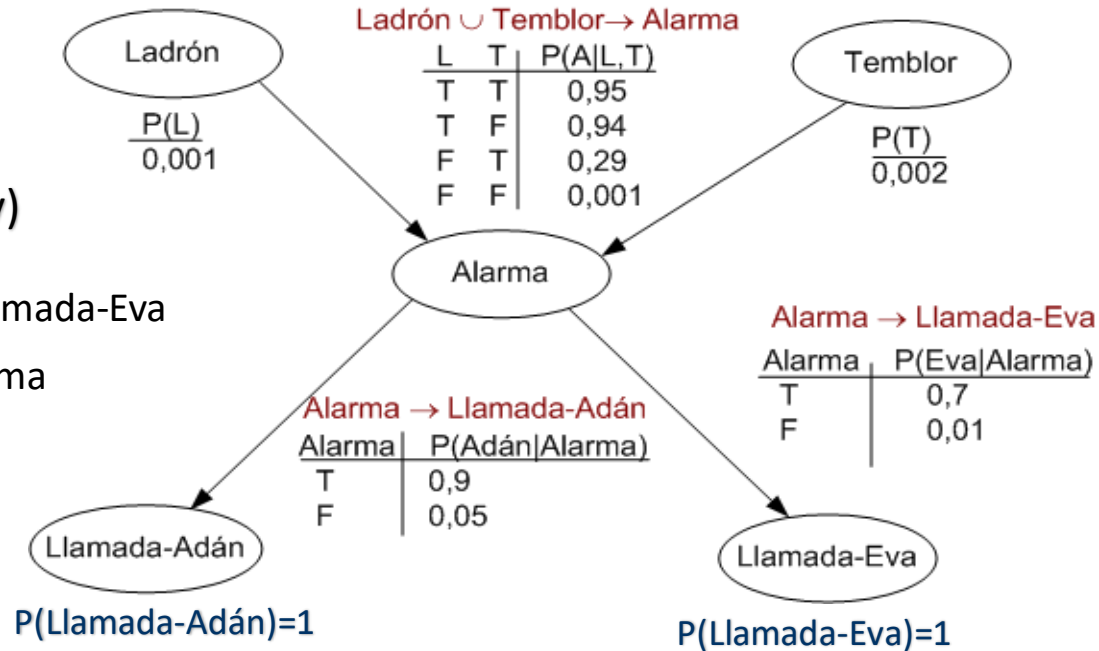
Un factor de normalización α debe asegurar que $P(x=v/e) + P(\neg(x=v)/e) = 1$

- La inferencia probabilística es computacionalmente intratable en el peor de los casos.
La complejidad depende de si los eventos observados son variables dependientes o independientes.
- Existen algoritmos para realizar inferencias aproximadas, aplicables en casos reales cuando las inferencias exactas son imposibles.

Pregunta: ¿Hay un ladrón?

$$P(x=v/e) = \alpha \sum_y P(x=v, e, y)$$

Variables no observadas: Alarma



$$P(\text{Ladrón}=T \mid (\text{Llamada-Adán}=T, \text{Llamada-Eva}=T)) =$$

$$P(\text{Ladrón}=T / (\text{Alarma}=T, \text{Llamada-Adán}=T)) +$$

$$P(\text{Ladrón}=T / (\text{Alarma}=F, \text{Llamada-Adán}=T)) +$$

$$P(\text{Ladrón}=T / (\text{Alarma}=T, \text{Llamada-Eva}=T)) +$$

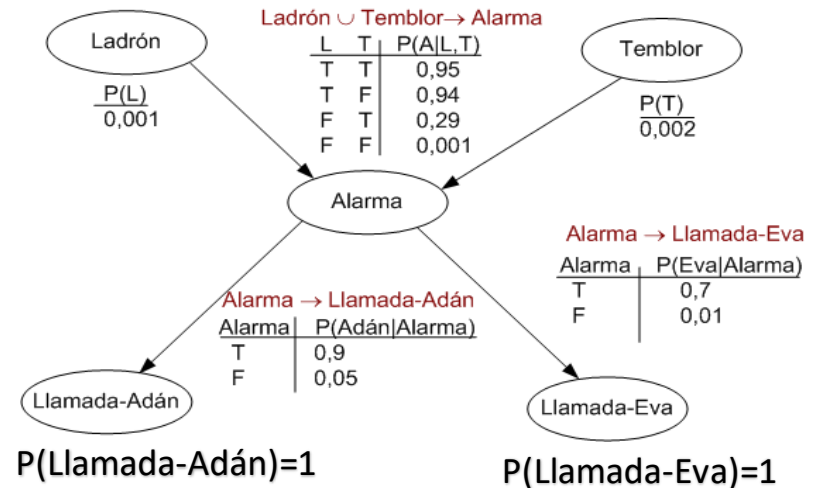
$$P(\text{Ladrón}=T \mid (\text{Alarma}=F, \text{Llamada-Eva}=T))$$

Y posterior normalización:

$$P(\text{Ladrón}=T) / (P(\text{Llamada-Adán}=T, \text{Llamada-Eva}=T) +$$

$$P(\text{Ladrón}=F) / (\text{Llamada-Adán}=T, \text{Llamada-Eva}=T) = 1$$

Adán y Eva me llaman, ¿Hay un ladrón?



Inspecting Factors

Click on a value to see its derivation or select a new factor to inspect it.

Reorder Variable Columns Inspecting Factor: **f11(Ladron)**

Ladron	Value
T	0.28417
F	0.71583

Derivation

This factor was derived by normalizing the remaining factor:
 $f11(\text{Ladron})$

Summing probabilities over all domain values:
 Normalization constant = $f11(\text{Ladron}=T) + f11(\text{Ladron}=F)$
 $0,00208 = 0,00059 + 0,00149$

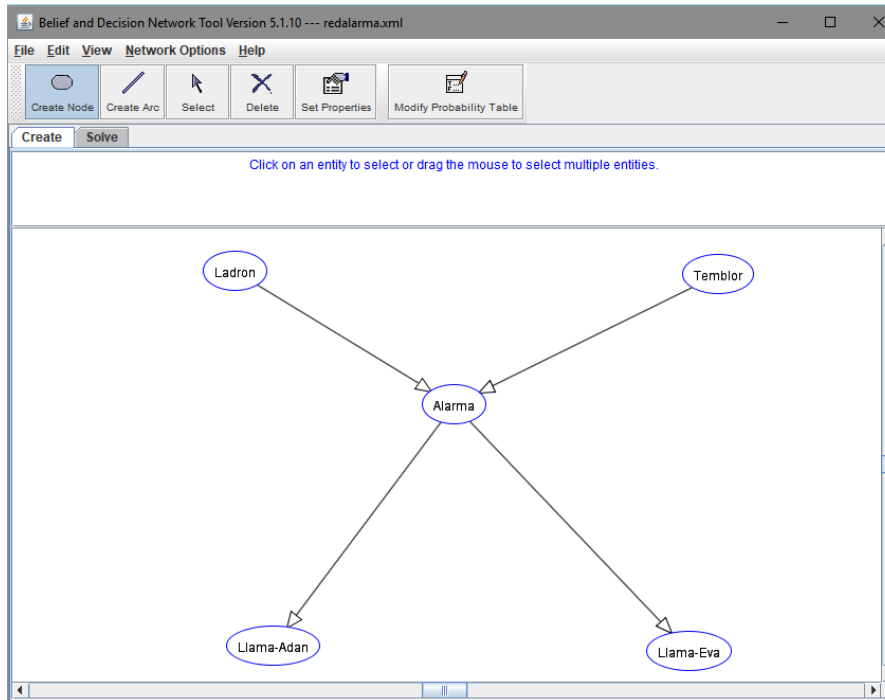
Normalizing the resultant probabilities:
 $f11(\text{Ladron}) = (0,00059 / 0,00208) = 0,28417$

Llama-A Observed
 Llama-Eva Observed Val: T

También podría obtenerse la probabilidad del Ladrón, dadas unas probabilidades en las evidencias (Llamada-Adán, Llamada-Eva)

Herramienta de Aplicación:

- Aplicación JavaWeb (y ayuda): <http://www.aispace.org/bayes/index.shtml>
<http://www.aispace.org/bayes/bayes.jnlp>
- Executable Jar File en Poliformat



Permite definir una red Bayesiana y realizar inferencias probabilísticas en base al conocimiento causal

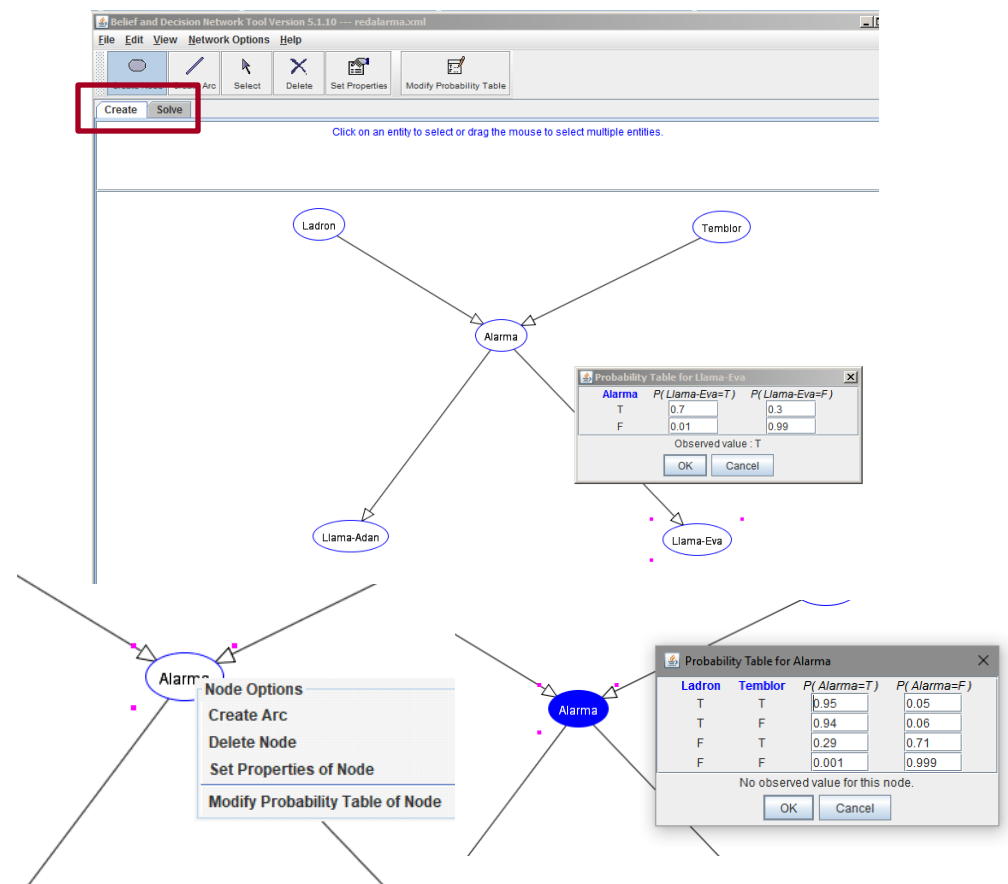
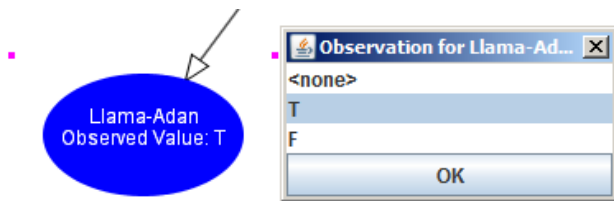
Otros: MSBNx (<http://research.microsoft.com/adapt/MSBNx>)
Norsys (<http://www.norsys.com/netica.html>), etc.

ETAPAS

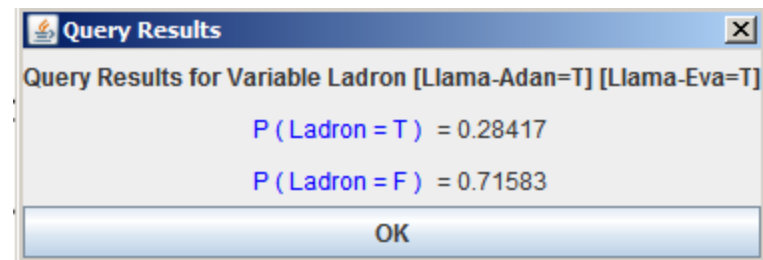
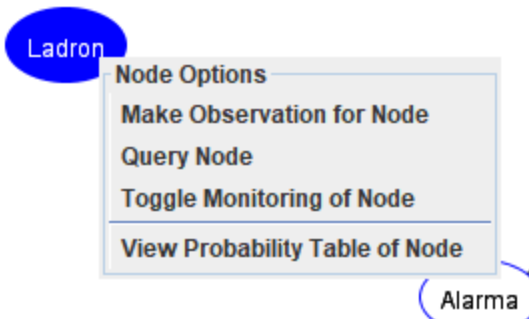
1) Crear la red (Create): nodos (variables), arcos (relaciones causales), probabilidades a priori, y probabilidades dependientes/condicionales

2) Resolver la red (Solve)

Dar valor a las variables observadas



3) Interrogar variable pregunta (Query)



Ejemplo



Prob. a priori:

$$P(\text{Ladrón}=T) = 0.1$$

$$P(\text{Ladrón}=F) = 0.9$$

Prob. Dependiente (Alarma/Ladrón):

Ladrón	P(Alarma=T)	P(Alarma=F)
T	0.8	0.2
F	0.1	0.9

Prob. Dependiente (Policía/Alarma):

Alarma	P(Policía=T)	P(Policía=F)
T	0.7	0.3
F	0.01	0.99

Caso-1: ¿Qué probabilidad hay de que haya un ladrón, suene la alarma y acuda la policía?

$$P(\text{Ladrón}) * P(\text{Alarma}=T/\text{Ladrón}=T) * P(\text{Policía}=T/\text{Alarma}=T) = 0.1 * 0.8 * 0.7 = 0.056 \quad (\text{Estado de la Red: Prob. Conjunta})$$

Caso-2: ¿Qué probabilidad hay de que haya un ladrón? $P(\text{Ladrón}=T) = 0.1$ (Probabilidad a priori)

Caso-3: ¿Qué probabilidad hay de que suene la alarma? (¿ $P(\text{Alarma}=T)$?, sin observaciones)

$$P(\text{Alarma}=T) = P(\text{Ladrón}=T) * P(\text{Alarma}=T/\text{Ladrón}=T) + P(\text{Ladrón}=F) * P(\text{Alarma}=T/\text{Ladrón}=F) =$$

$$= (0.1 * 0.8) + (0.9 * 0.1) = 0.17$$

Nota: de aquí sale que $P(\text{Alarma}=F) = 1 - 0.17 = 0.83$

Caso-4: ¿Qué probabilidad hay de que acuda la policía? (¿ $P(\text{Alarma}=T)$?, sin observaciones)

$$P(\text{Policía}=T) = P(\text{Alarma}=T) * P(\text{Policía}=T/\text{Alarma}=T) + P(\text{Alarma}=F) * P(\text{Policía}=T/\text{Alarma}=F) =$$

$$= (0.17 * 0.7) + (0.83 * 0.01) = 0.1273$$



Prob. a priori:

$$P(\text{Ladrón}=T) = 0.1$$

$$P(\text{Ladrón}=F) = 0.9$$

Prob. Dependiente:

Ladrón	P(Alarma=T)	P(Alarma=F)
T	0.8	0.2
F	0.1	0.9

Prob. Dependiente:

Alarma	P(Policía=T)	P(Policía=F)
T	0.7	0.3
F	0.01	0.99

Caso-5: Hay un ladrón (observado, lo que quiere decir que $P(\text{Ladrón}=T)=1$). ¿Qué probabilidad hay de que acuda la policía: $P(\text{Policía}=T)$?

$$\begin{aligned}
 P(\text{Policía}=T) &= P(\text{Policía}=T/\text{Alarma}=T) * P(\text{Alarma}=T/\text{Ladrón}=T) + \\
 &P(\text{Policía}=T/\text{Alarma}=F) * P(\text{Alarma}=F/\text{Ladrón}=T) = \\
 &= 0.7 * 0.8 + 0.01 * 0.2 = 0.562
 \end{aligned}$$

Caso-6: ¿Qué probabilidad hay de que haya un ladrón y acuda la policía?

$$\begin{aligned}
 &P(\text{Ladrón}) * P(\text{Alarma}=T/\text{Ladrón}=T) * P(\text{Policía}=T/\text{Alarma}=T) + \\
 &+ P(\text{Ladrón}) * P(\text{Alarma}=F/\text{Ladrón}=T) * P(\text{Policía}=T/\text{Alarma}=F) = \\
 &= (0.1 * 0.8 * 0.7) + (0.1 * 0.2 * 0.01) = 0.056 + 0.0002 = 0.0562
 \end{aligned}$$



Prob. a priori:

$$P(\text{Ladrón}=T) = 0.1$$

$$P(\text{Ladrón}=F) = 0.9$$

Prob. Dependiente:

Ladrón	P(Alarma=T)	P(Alarma=F)
T	0.8	0.2
F	0.1	0.9

Prob. Dependiente:

Alarma	P(Policía=T)	P(Policía=F)
T	0.7	0.3
F	0.01	0.99

Caso-7: La policía ha acudido: observado $P(\text{Policía}=T)=1$. ¿Qué probabilidad hay de que haya sonado la alarma?

$$P(\text{Alarma}=T) / [(P(\text{Alarma}=T) + P(\text{Alarma}=F))] = 0.119 / [0.119 + 0.0083] = 0.9348 \quad (*normalización*)$$

Dado que:

$$P(\text{Alarma}=T) = P(\text{Policía}=T/\text{Alarma}=T) * [P(\text{Alarma}=T/\text{Ladrón}=T) * P(\text{Ladrón}=T) + P(\text{Alarma}=T/\text{Ladrón}=F) * P(\text{Ladrón}=F)] = 0.7 * [0.8 * 0.1 + 0.1 * 0.9] = 0.119$$

$$P(\text{Alarma}=F) = P(\text{Policía}=T/\text{Alarma}=F) * [P(\text{Alarma}=F/\text{Ladrón}=T) * P(\text{Ladrón}=T) + P(\text{Alarma}=F/\text{Ladrón}=F) * P(\text{Ladrón}=F)] = 0.01 * [0.2 * 0.1 + 0.9 * 0.9] = 0.0083$$

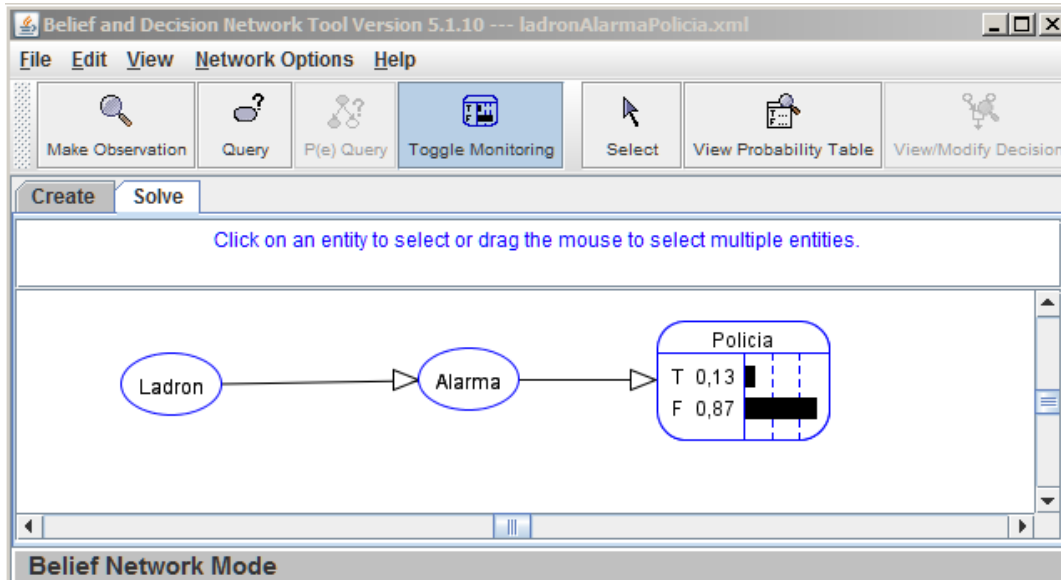
Caso-8: La policía ha acudido: $P(\text{Policía}=T)=1$. ¿Qué probabilidad hay de que haya un ladrón?

$$P(\text{Ladrón}=T) / [(P(\text{Ladrón}=T) + P(\text{Ladrón}=F))] = 0.0562 / [0.0562 + 0.0711] = 0.44148 \quad (*normalización*)$$

Dado que:

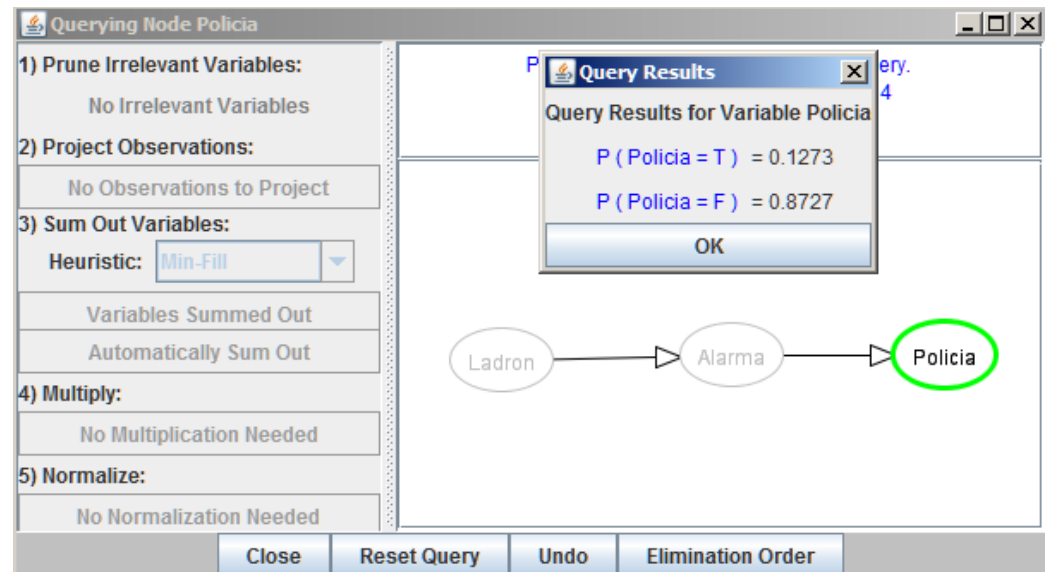
$$P(\text{Ladrón}=T) = P(\text{Ladrón}=T) * [P(\text{Ladrón}=T/\text{Alarma}=T) * P(\text{Alarma}=T/\text{Policía}=T) + P(\text{Ladrón}=T/\text{Alarma}=F) * P(\text{Alarma}=F/\text{Policía}=T)] = 0.1 * [0.8 * 0.7 + 0.2 * 0.01] = 0.1 * 0.562 = 0.0562$$

$$P(\text{Ladrón}=F) = \dots = 0.0711$$



Caso-4

¿Qué probabilidad hay de que acuda la policía?



Querying Node Policía

Click on a factor to inspect it

Current Factors:

f3()
f5(Policia)

Eliminated Factors:

f0(Ladrón)
f1(Ladrón, Alarma)
f2(Alarma, Policía)
f4(Alarma)

1) Prune Irrelevant Variables:
No Irrelevant Variables

2) Project Observations:
Observations Projected

3) Sum Out Variables:
Heuristic: Min-Fill

Variables Summed Out:
Automatically Sum Out

4) Multiply:
Multiply Final Factors

5) Normalize:
Normalize Final Factors

Multiply final factors by pressing "Multiply Final Factors".

Caso-5

Se hace la observación $P(\text{Ladrón})=T$ y se obtiene el valor de $P(\text{Policía}=T)=0.562$

Inspecting Factors

Click on a value

Reorder Variable Columns

Policia

T
F

This factor was derived by eliminating "Alarma"

Close Reset Query Un

Belief and Decision Network Tool Version 5.1.10 --- untitled.xml

File Edit View Network Options Help

Make Observation Query P(e) Query Toggle Monitoring Select View Probability Table View/Modify Decision Add No-forgetting Arcs Optimize Decisions

Create Solve

Click on a regular node to view its probability table.

Caso-7 y Caso-8

Se hace la observación $P(\text{Policía})=T$ y se obtienen los valores de $P(\text{Alarma}=T)=0.93$ y $P(\text{Ladrón}=T)=0.44$

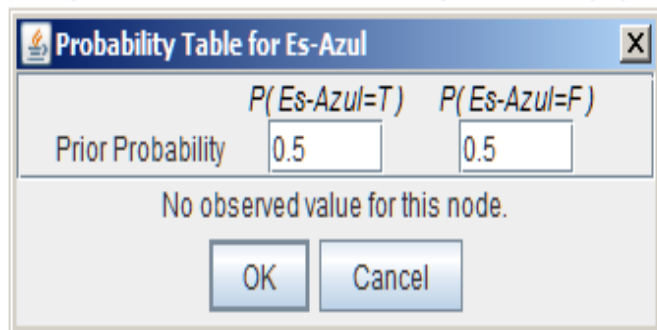
Belief Network Mode

EJEMPLO (Russell, Norvig):

El testigo de un accidente nocturno, en el que un taxi implicado se da a la fuga, asegura que el taxi era azul. Todos los taxis de la ciudad son verdes o azules.

Sin embargo, exhaustivas experimentaciones posteriores demuestran que, bajo condiciones de poca iluminación, la distinción entre azul y verde es fiable un 75%. ¿Es posible calcular la credibilidad del color azul del taxi?

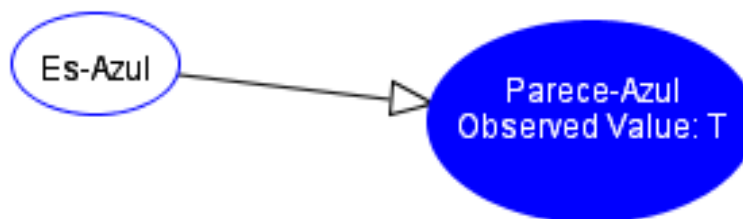
Principio de Indiferencia de Laplace $\Rightarrow P(A)=0,5$



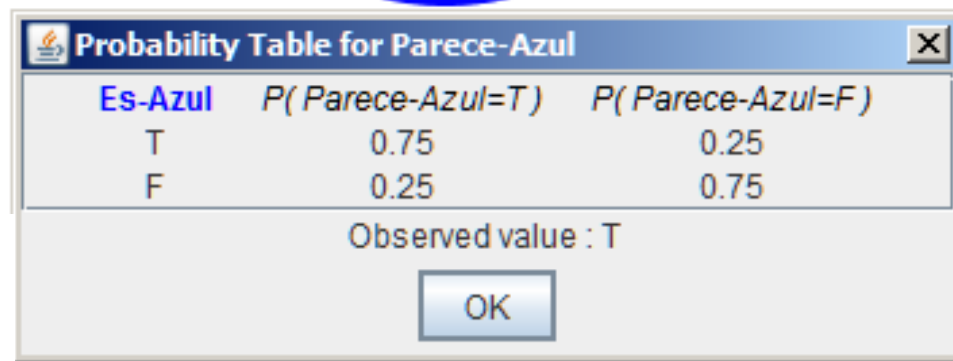
	$P(\text{Es-Azul}=T)$	$P(\text{Es-Azul}=F)$
Prior Probability	0.5	0.5

No observed value for this node.

OK Cancel



$P(PA/A)=0,75$

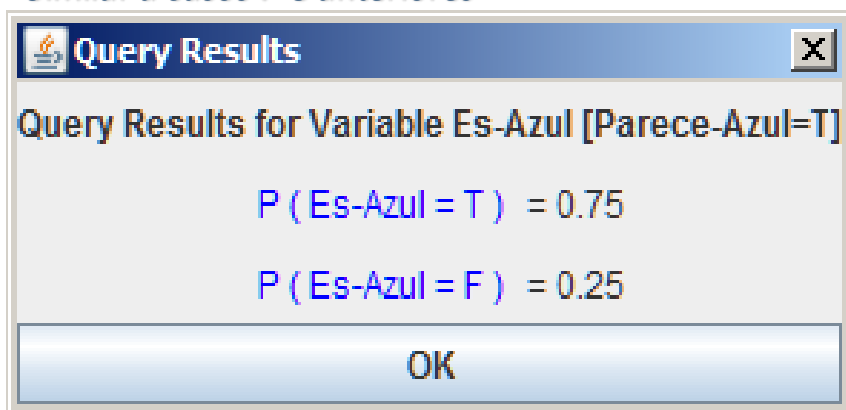


Es-Azul	$P(\text{Parece-Azul}=T)$	$P(\text{Parece-Azul}=F)$
T	0.75	0.25
F	0.25	0.75

Observed value : T

OK

Similar a casos 7-8 anteriores



Query Results for Variable Es-Azul [Parece-Azul=T]

$P(\text{Es-Azul} = T) = 0.75$

$P(\text{Es-Azul} = F) = 0.25$

OK

¿Y si solo 1 de cada 10 taxis son azules?

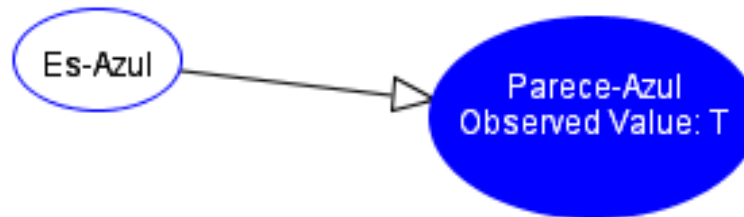
¿Y si solo 1 de cada 10 taxis son azules?

Probability Table for Es-Azul

	$P(\text{Es-Azul}=T)$	$P(\text{Es-Azul}=F)$
Prior Probability	0.1	0.9

No observed value for this node.

OK Cancel



Probability Table for Parece-Azul

Es-Azul	$P(\text{Parece-Azul}=T)$	$P(\text{Parece-Azul}=F)$
T	0.75	0.25
F	0.25	0.75

Observed value : T

OK

Query Results

Query Results for Variable Es-Azul [Parece-Azul=T]

$P(\text{Es-Azul} = T) = 0.25$

$P(\text{Es-Azul} = F) = 0.75$

OK

Ejercicio propuesto (Trabajo Académico)

Entrega en Tarea Poliformat

(último día 31 de Octubre, 23:55h)

Un juez tiene el criterio de juzgar la culpabilidad de un acusado en base a si se prueba que tiene sus huellas en el arma, tiene un motivo, y no tiene una coartada.

Huellas	Motivo	No-Coartada	Culpable
T	T	T	0.9
T	T	F	0.7
T	F	T	0.5
T	F	F	0.3
F	T	T	0.8
F	T	F	0.8
F	F	T	0.5
F	F	F	0.001

La policía detiene al sospechoso con estas pruebas:

Se encuentran huellas en el arma (Creencia: 0.9), posiblemente debido a otros factores.

El acusado tiene un motivo (Creencia: 0.5), posiblemente debido a otros factores.

El acusado tiene una coartada (Creencia: 0.7), posiblemente debido a otros factores.

Ejercicio propuesto (Trabajo Académico)

Entrega en Tarea Poliformat (último día 31 de Octubre, 23:55h)

Se pide, con una pequeña memoria en PDF:

- Utilizando en el entorno anterior, diseñad la red bayesiana y responded: ¿con qué probabilidad el sospechoso es culpable?
- Incluid variaciones en las creencias de las pruebas que aporta la policía. Obtened la respuesta con eventos observados $P(e)=1$.
- Introducid nueva información (datos) que permitan deducir las probabilidades sobre huellas, motivos o coartadas del acusado (a través de la probabilidad de esos datos y de la probabilidad condicional de 'huellas', 'motivos' o 'coartadas' respecto a dichos nuevos datos).

Posibles extensiones a valorar:

- Se encuentran confirmaciones de la propiedad del arma utilizada, aunque con una fiabilidad del 80%. Ello también incrementa la confianza en que sea culpable.
- Un amigo del acusado confirma su coartada exculpatoria, aunque su credibilidad es solo del 70%.
- Diversas evidencias sitúan al acusado en la escena del crimen, lo que hace también modificar la creencia de su coartada. Por ej., un testigo afirma ver a alguien parecido al acusado en la escena del crimen, aunque debido a la poca luz su confianza es del 60%; existe una cámara con una confianza en el reconocimiento del 80%, etc.

Este tipo de razonamiento se utiliza para resolver investigaciones criminales en el mundo real

https://elpais.com/sociedad/2019/11/21/actualidad/1574348607_829222.html

CASO DIANA QUER >

La hipótesis de un edema enfrenta a los forenses del 'caso Diana Quer'

Dos grupos de médicos coinciden en que El Chicle usó la brida para estrangular a la joven pero discrepan sobre si una hinchazón en el área genital se debe a una violación o al agua del pozo



SILVIA R. PONTEVEDRA

Santiago de Compostela · 22 NOV 2019 · 18:00 CET

≡ EL PAÍS

"tenido lugar en un acto sexual con maniobras intempestivas". Hasta esta mañana nadie sabía nada de esto. El propio médico no lo había recogido en su informe judicial. Y el magistrado, Ángel Pantín, le ha exigido molesto: "Esto merece una explicación, señor Pampín".

Por un lado, en la sala, se sentaban los tres médicos que se responsabilizaron de la auténtica autopsia del cadáver en el Imelga (Instituto de Medicina Legal de Galicia), entre ellos el que está considerado como uno de los mejores antropólogos forenses de España, Fernando Serrulla. Por otro, tres especialistas y un profesor de Matemáticas que trabajaron bajo el mando de José Blanco Pampín, jefe de Patología Forense del Imelga, autor de autopsias tan sonadas como la de la niña Asunta Basterra y la de Ramón Sampedro. En el caso de la muerte de Diana Quer, este equipo había llevado a cabo la "revisión" sobre el papel de la primera autopsia. Y también, con ayuda del matemático, un estudio estadístico sobre crímenes sexuales basado en el Teorema de Bayes que concluye que El Chicle violó a su víctima "con un 99,99939% de probabilidad".

- Las Redes Bayesianas son una forma natural (y gráfica) de representar las relaciones causales condicionales.

Es el método actualmente más aceptado para tratar la incertidumbre en sistemas de IA y hacer inferencias probabilísticas.

Etapas: (i) Diseño de la Red Bayesiana (conocimiento causal dominio),
(ii) Observación de Evidencias (problema),
(iii) Respuesta a preguntas (Inferencia Bayesiana)

- Los algoritmos de propagación de las probabilidades dependen de la topología de la red bayesiana (árbol, poliárboles, redes multiconectadas, etc.)

Existen diversos tipos y modelos de redes bayesianas, adaptados a cada problemática:

- Redes Bayesianas Dinámicas: cambian con el tiempo, y lo ocurrido en (t) tiene relación con lo que suceda en $(t+1)$.
- Redes de Markov: subconjunto de las Redes Bayesianas.
- Redes Gaussianas (variables continuas, distribución Gaussiana).
- Redes Bayesianas Híbridas (contiene variables discretas y continuas), etc.