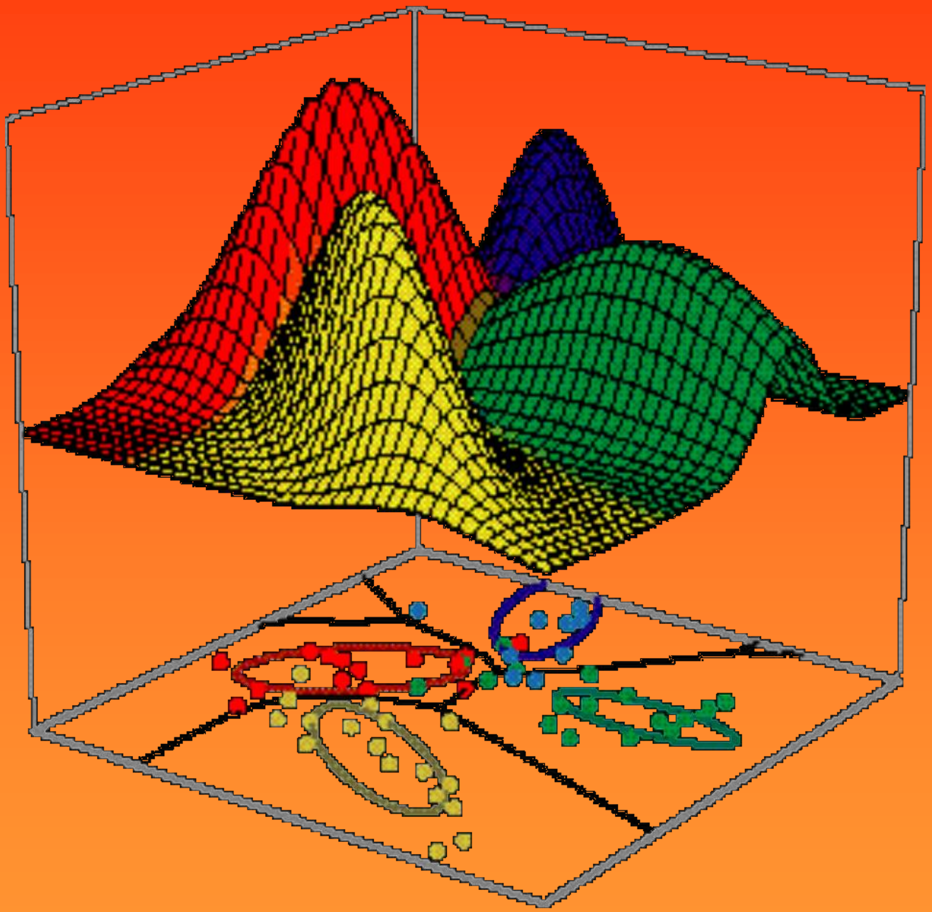


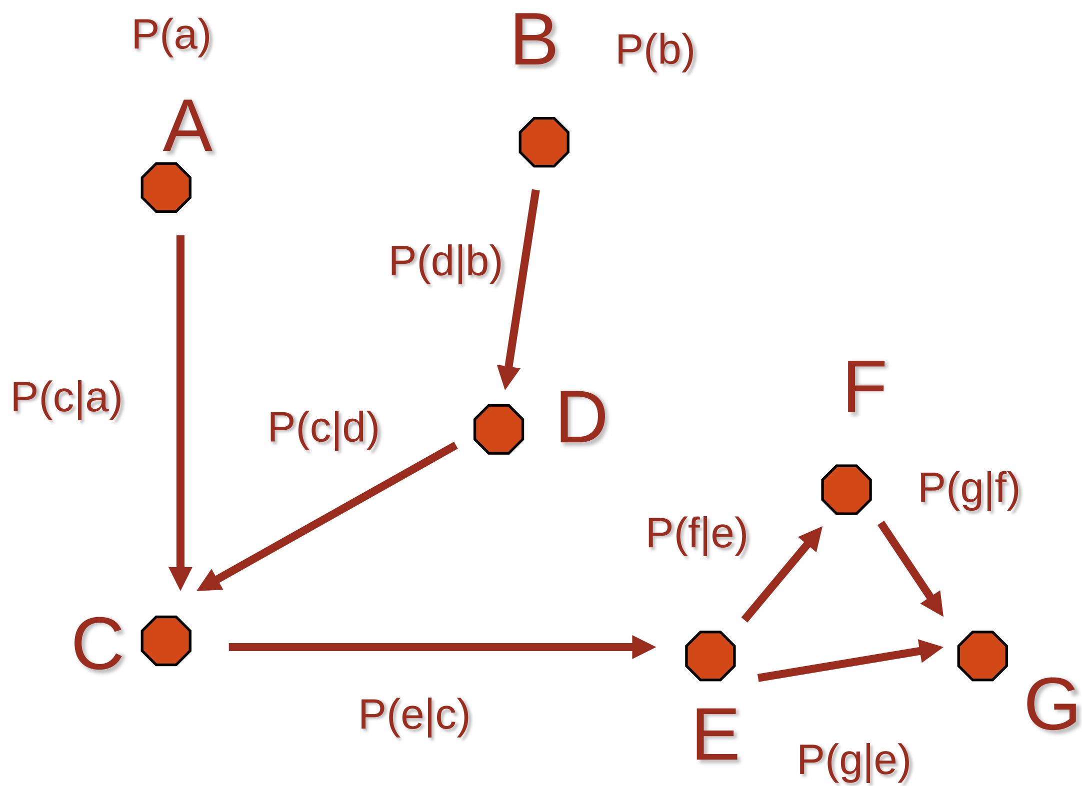
Pattern Classification

All materials in these slides were taken from
Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000
with the permission of the authors and the publisher

3.9 Bayesian Belief networks

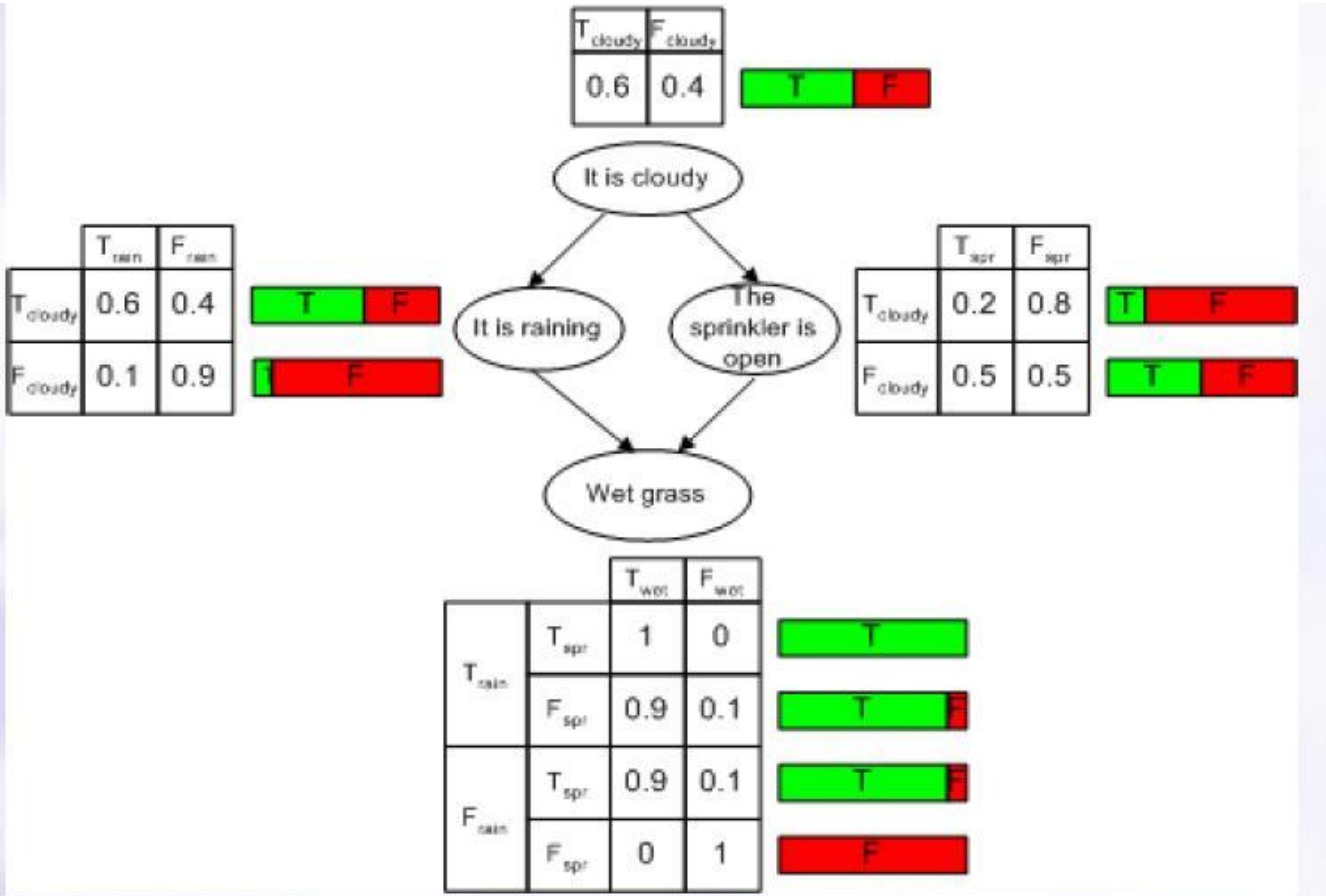
- Los metodos descriptos hasta el momento suponen que las distribuciones por clase pueden ser parametrizadas por un vector θ
- Si existe algun conocimiento acerca de θ tambien pudo ser usado
- A veces, existe informacion sobre las dependencias entre las componentes del vector de caracteristicas
- Estas relaciones pueden representarse gráficamente como redes bayesianas.

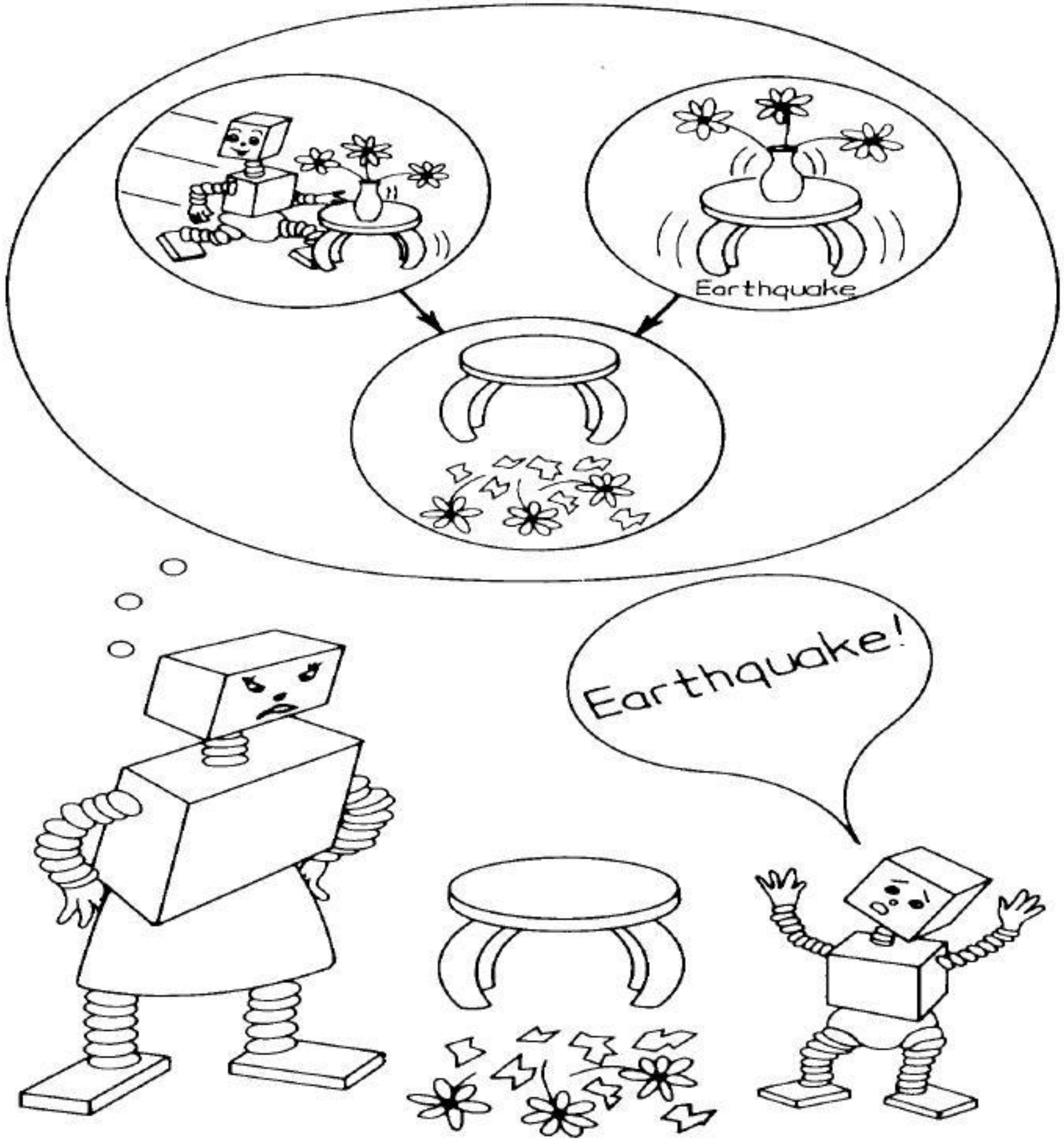
Red Bayesiana



Una red bayesiana consiste de nodos (mayusculas) y sus estados asociados (minusculas). Por lo cual nodo A tiene estados $a_1,a_2,...$ y nodo B tiene estados $b_1,b_2,\ldots P(c|a)$ puede ser una matriz con entradas $P(c_i|a_j)$

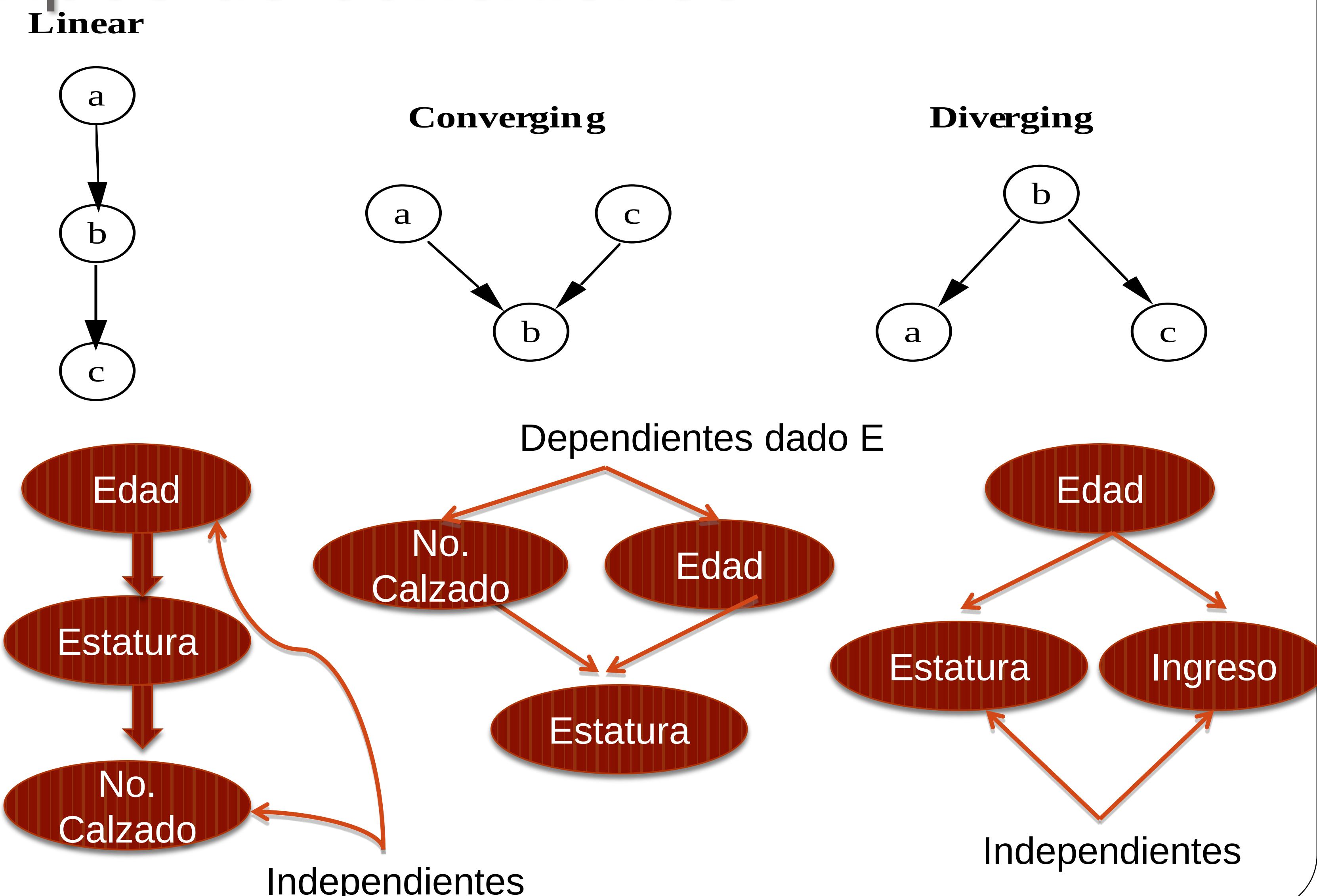
Definicion: grafo dirigido acíclico conexo, más una distribución de probabilidad sobre sus variables

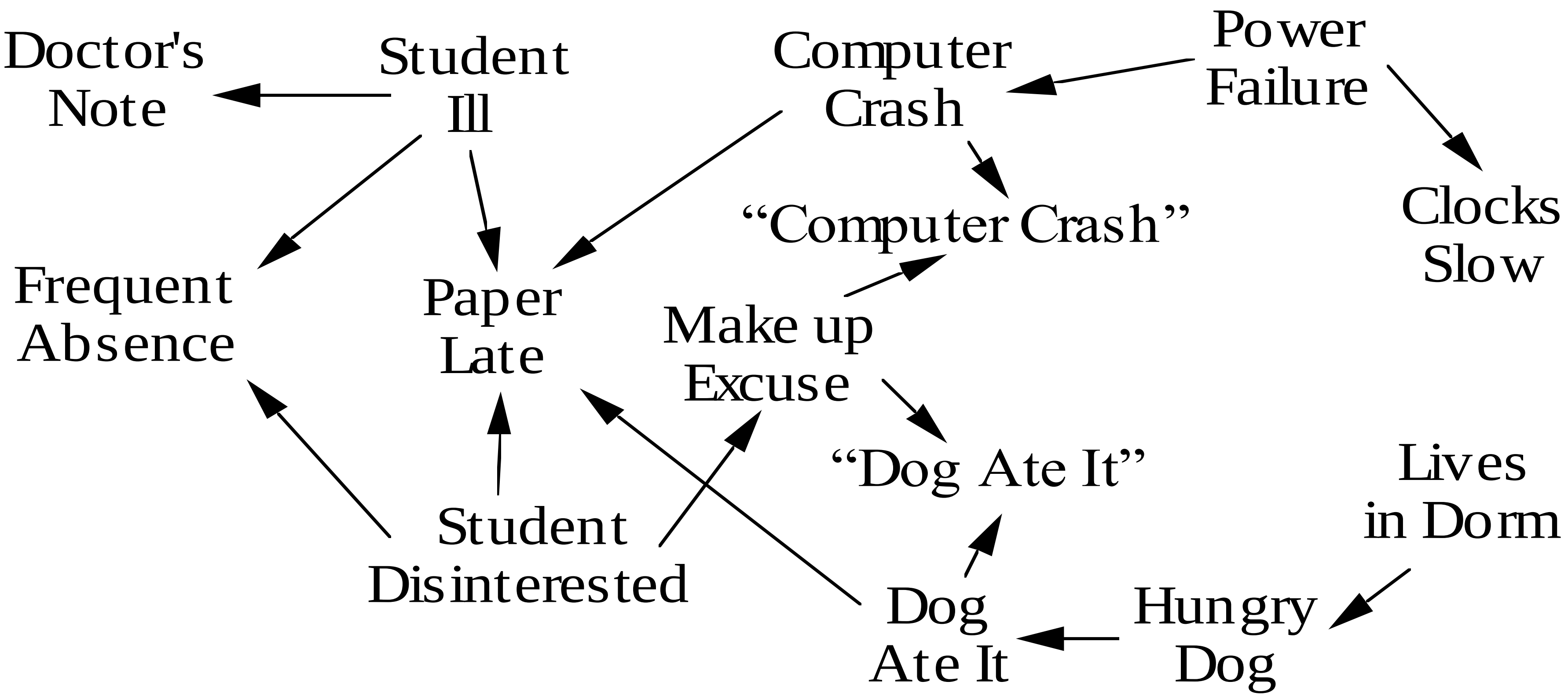




Probabilistic Reasoning in a Causal Network

Tipos de conexiones



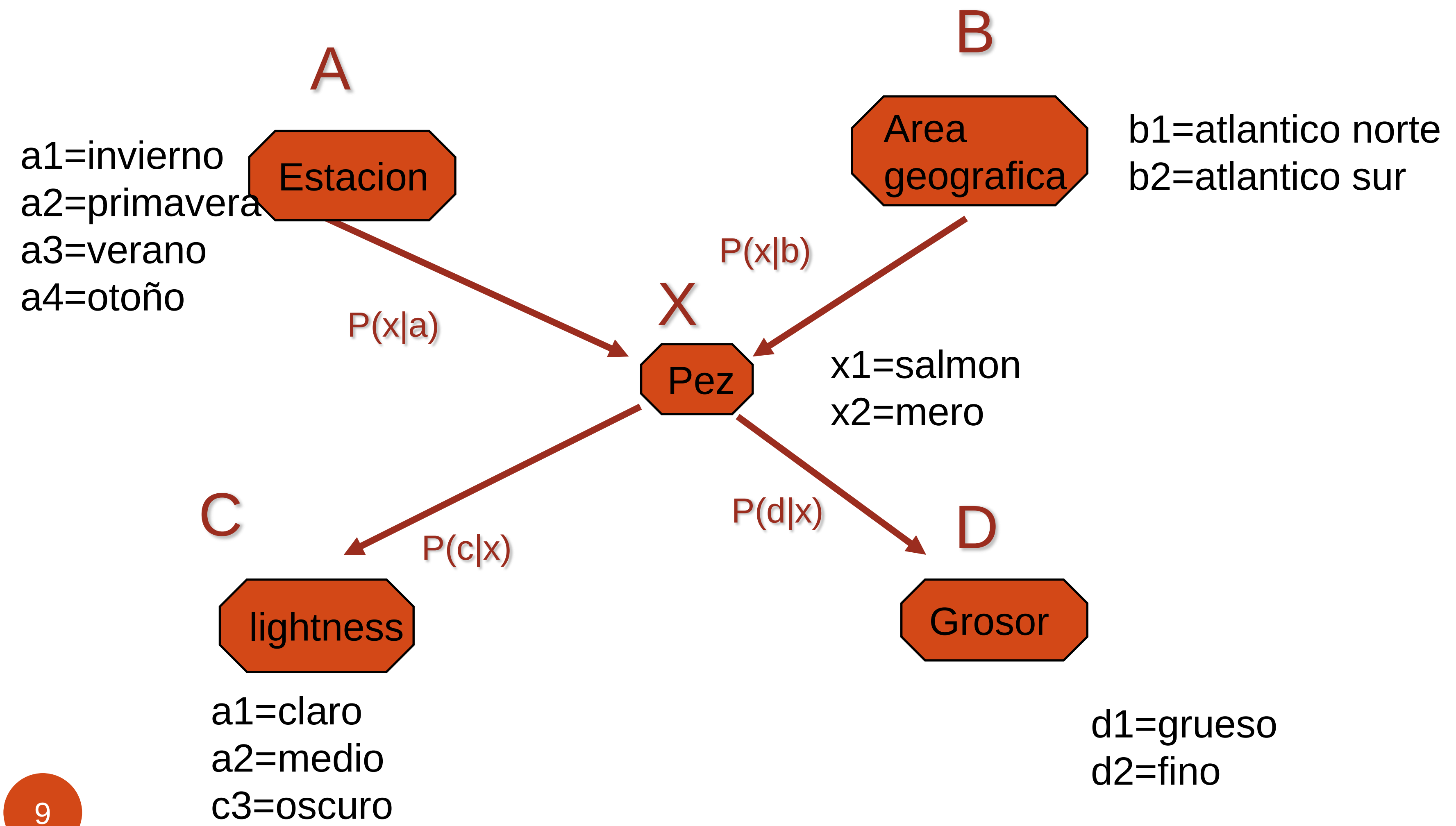


Diagnósticos?
Predicciones?

Clasificacion del mero y el salmon

- A representa estación, con cuatro valores equiprobables a1=invierno, a2=primavera, a3=verano, a4=otoño
- El nodo B representa el area geografica donde el pez ha sido capturado.: b1=Atlantico norte, b2 Atlantico Sur, y son equiprobables.
- A y B son padres del nodo X, el cual representa el pez capturado, y tiene dos posibles valores, salmon o mero.
- Los nodos hijos representan lightness, C, con c1=claro, c2=medio, c3=oscuro, asi como D representa la grosor, con d1=grueso y d2=fino.

Red Bayesiana para el problema del salmon



- La estacion y el area geografica de pesca no dependen del pez, pero si dependen las variables lightness y grosor.
- $P(a_j)$ =0.25, $P(b_j)$=0.5
- Un experto genera las siguientes probabilidades:

$P(x_i \mid a_j)$	salmon	mero
Invierno	.9	.1
Primavera	.3	.7
Verano	.4	.6
Otoño	.8	.2

$P(x_i \mid b_j)$	salmon	mero
Norte	.65	.35
Primavera	.95	.05

- El experto tambien genera probabilidades condicionales para los eventos conjuntos y las variables hijos

$P(x_1 \mid a_i b_j)$	Norte	Sur
Invierno	.9	.8
Primavera	.3	.2
Verano	.4	.1
Otoño	.7	.6

$P(c_i \mid x_j)$	salmon	mero
Claro	.33	.8
Medio	.33	.1
Oscuro	.34	.1

$P(d_i \mid x_j)$	salmon	mero
grueso	.4	.95
fino	.6	.05

Verifiquemos que sin evidencia $P(x_1)=P(x_2)$

$$P(x_1) = \sum_{i,j,k,l} P(x_1, a_i, b_j, c_k, d_l)$$

$$= \sum_{i,j,k,l} P(a_i)P(b_j)P(x_1 \mid a_i, b_j)P(c_k \mid x_1)P(d_l \mid x_1)$$

$$= \sum_{i,j} P(a_i)P(b_j)P(x_1 \mid a_i, b_j)$$

$$= (0.25)(0.5)\sum_{i,j} P(x_1 \mid a_i, b_j)$$

$$= (0.25)(0.5)[0.9 + 0.3 + 0.4 + 0.7 + 0.8 + 0.2 + 0.1 + 0.6]$$

$$= 0.5$$

Evidencia en el pez

- Recolectemos la evidencia, *$\{e_A, e_B, e_C, e_D\}$, asumiendo que son independientes unos de los otros .*
- Supongamos que es invierno, i.e., $P(a_1|e_A) = 1$ *y* $P(a_i|e_A) = 0$ *for* $i = 2, 3, 4$.
- *Supongamos que no sabemos de que area vino el bote, pero que una cierta tripulacion le gusta pescar al sur* ; por lo cual asumimos que $P(b_1|e_B) = 0.2$ *y* $P(b_2|e_B) = 0.8$.
- *Medimos el pez y encontramos que es liviano y decidimos* $P(e_C|c_1) = 1$, $P(e_C|c_2) = 0.5$, *y* $P(e_C|c_3) = 0$.
- *Supongamos que existe oclusion y que no podemos medir el ancho del pez, por lo cual* $P(e_D|d_1) = P(e_D|d_2)$.

Padres e hijos

- $P_P(x_1)$ probabilidad de salmon producida por sus padres
- $P_C(x_1)$ probabilidad de salmon producida por sus hijos

$$
\begin{aligned}
&P_P(x_i) = P(x_i \mid e^P) \\
&\propto P(x_i \mid a_1, b_1)P(a_1 \mid e)P(b_1 \mid e) + P(x_i \mid a_1, b_2)P(a_1 \mid e)P(b_2 \mid e) \\
&+ P(x_i \mid a_2, b_1)P(a_2 \mid e)P(b_1 \mid e) + P(x_i \mid a_2, b_2)P(a_2 \mid e)P(b_2 \mid e) \\
&+ P(x_i \mid a_3, b_1)P(a_3 \mid e)P(b_1 \mid e) + P(x_i \mid a_3, b_2)P(a_3 \mid e)P(b_2 \mid e) \\
&+ P(x_i \mid a_4, b_1)P(a_4 \mid e)P(b_1 \mid e) + P(x_i \mid a_4, b_2)P(a_4 \mid e)P(b_2 \mid e) \\
&= P(x_i \mid a_1, b_1)P(a_1 \mid e)P(b_1 \mid e) + P(x_i \mid a_1, b_2)P(a_1 \mid e)P(b_2 \mid e)
\end{aligned}
$$

$$P_P(x_1) = 0.82, P_P(x_2) = 1 - 0.82 = 0.18$$

Calculamos ahora $P_C(x_1)$

$$P_c(x_1) = P(e^c \mid x_1) = P(e_C \mid x_1)P(e_D \mid x_1) =$$
$$= [P(e_C \mid c_1)P(c_1 \mid x_1) + P(e_C \mid c_2)P(c_2 \mid x_2) + P(e_C \mid c_3)P(c_3 \mid x_3)]$$
$$\times [P(e_D \mid d_1)P(d_1 \mid x_1) + P(e_D \mid d_2)P(d_2 \mid x_1)]$$
$$= [(1.0)(0.33 + (0.5)(0.33) + 0(.34)] \times [1.0(0.4) + (1.0)(0.6)]$$
$$= 0.495$$

- Un calculo similar da $P(x_2) \propto 0.85.$

$$P(x_i|e) \propto P_c(x_i)P_P(x_i)$$

- *Renormalizando (i.e., dividiedo por su suma), resulta*

$$P(x_1 \mid e) = \frac{(0.82)(0.495)}{(0.82)(0.495) + (0.18)(0.85)} = 0.726$$
$$P(x_2 \mid e) = \frac{(0.18)(0.85)}{(0.82)(0.495) + (0.18)(0.85)} = 0.274$$

Respuesta al ejemplo

- Dada toda la evidencia recolectada en la red, el resultado mas probable es x1=salmon

- Con esta red se pueden dar distintas respuestas, sobre otras variables presentes en la red.

Dependencias desconocidas

- Cuando las dependencias son desconocidas, podemos suponer la menor de las hipótesis, la independencia condicional dada la categoria

$$p(\omega_k \mid x) \propto \prod_{i=1}^{d} p(x \mid \omega_k)$$

- Este es la llamada regla naïve de bayes, que a menudo tiene mucho exito en la practica, y que puede se expresada como una red muy simple.

Procesos estocásticos

- Un sistema informático complejo se caracteriza por demandas de carácter aleatorio y por ser dinámico
- Necesitamos una herramienta que modele procesos aleatorios en el tiempo, y para ello usaremos los *procesos estocásticos*
- Un proceso estocástico es una familia de variables aleatorias parametrizadas por el tiempo

Procesos estocásticos

- Un proceso estocástico es una familia de variables aleatorias definida sobre un espacio de probabilidad. Es decir:

$$\left\{ X_t : \Omega \rightarrow \Re, \quad t \in T \right\}$$

$$\omega \rightarrow X_t(\omega) = X(\omega, t)$$

Procesos estocásticos

- Tendremos que *X* es una función de dos argumentos. Fijado $\omega=\omega_0$, obtenemos una función determinística (no aleatoria), llamada trayectoria:

$$X(\cdot,\omega_0): T \to \Re$$

$$t \to X(t,\omega_0)$$

Procesos estocásticos

- Asimismo, fijado $t=t_0$, obtenemos una de las variables aleatorias de la familia:

$$X(t_0,\cdot):\Omega \to \Re$$

$$\omega \to X(t_0,\omega)$$

Procesos estocásticos

- El espacio de estados S de un proceso estocástico es el conjunto de todos los posibles valores que puede tomar dicho proceso:

$$S = \left\{ X_t(\omega) \mid t \in T \wedge \omega \in \Omega \right\}$$

Ejemplo de proceso estocástico

- Lanzamos una moneda al aire 6 veces. El jugador gana 1 peso cada vez que sale cara (C), y pierde 1 peso cada vez que sale sello (F).
- X_i = estado de cuentas del jugador después de la i-ésima jugada
- La familia de variables aleatorias $\{X_1, X_2, \ldots, X_6\}$ constituye un proceso estocástico

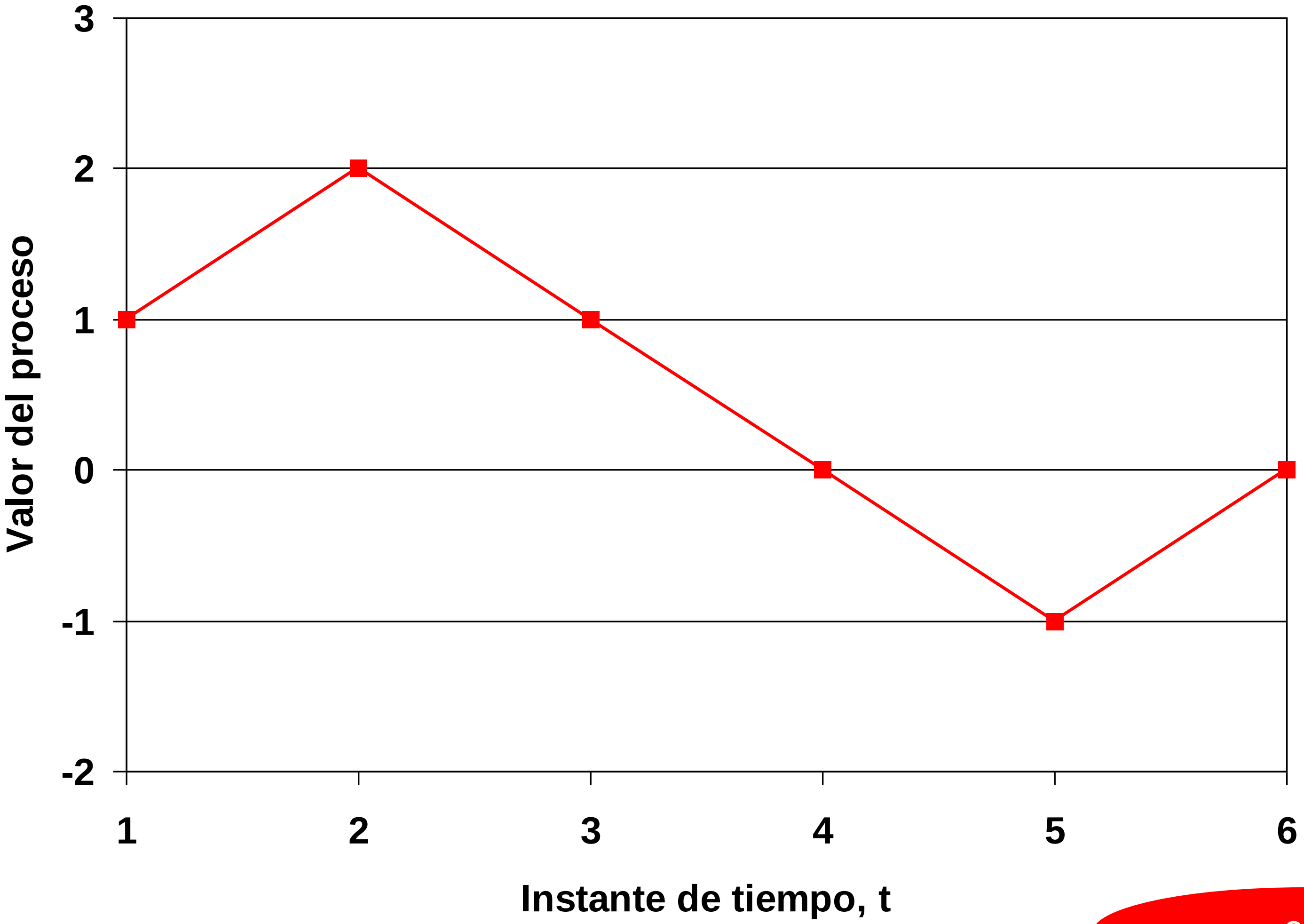
Ejemplo de proceso estocástico

- Ω={CCCCCC,CCCCCF,…}
- card(Ω) = 2^6 = 64
- $P(\omega)=1/64 \quad \forall\ \omega \in \Omega$
- T={1, 2, 3, 4, 5, 6}
- S={–6, –5, …, –1, 0, 1, 2, …, 5, 6}
- $X_1(\Omega)=\{-1, 1\}$
- $X_2(\Omega)=\{-2, 0, 2\}$

Ejemplo de proceso estocástico

- Si fijo ω, por ejemplo ω_0=CCFFFC, obtengo una secuencia de valores completamente determinista:
- $X_1(\omega_0)=1$, $X_2(\omega_0)=2$, $X_3(\omega_0)=1$, $X_4(\omega_0)=0$, $X_5(\omega_0)=-1$, $X_6(\omega_0)=0$
- Puedo dibujar con estos valores la *trayectoria del proceso*:

Ejemplo de proceso estocástico



Ejemplo de proceso estocástico

- Si fijo t, por ejemplo $t_0=3$, obtengo una de las variables aleatorias del proceso:

$$X_3 : \Omega \to \Re$$

$$\omega \to X_3(\omega)$$

- Los posibles valores que puede tomar el proceso en $t_0=3$ son: $X_3(\Omega)=\{-3, -1, 1, 3\}$

Ejemplo de proceso estocástico

- Podemos hallar la probabilidad de que el proceso tome uno de estos valores:

$$P[X_3(\omega)=1]=P[CFC]+P[CCF]+P[FCC]=3\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{3}{8}$$

$$P[X_3(\omega)=3]=P[CCC]=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$$

$$P[X_3(\omega)=-1]=P[FCF]+P[FFC]+P[CFF]=3\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{3}{8}$$

$$P[X_3(\omega)=-3]=P[FFF]=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$$

Clasificación de los procesos estocásticos

	S discreto	**S continuo**
T discreto	Cadena	Sucesión de variables aleatorias continuas
T continuo	Proceso puntual	Proceso continuo

Ejemplos de los tipos de procesos estocásticos

- Cadena: Ejemplo anterior
- Sucesión de variables aleatorias continuas: cantidad de lluvia caída cada mes
- Proceso puntual: Número de clientes esperando en la cola de un supermercado
- Proceso continuo: velocidad del viento

Cadenas de Markov

- Las cadenas de Markov y los procesos de Markov son un tipo especial de procesos estocásticos que poseen la siguiente propiedad:
- **Propiedad de Markov:** Conocido el estado del proceso en un momento dado, su comportamiento futuro no depende del pasado. Dicho de otro modo, “dado el presente, el futuro es independiente del pasado”

Cadenas de Markov

- Sólo estudiaremos las cadenas de Markov, con lo cual tendremos espacios de estados S discretos y conjuntos de instantes de tiempo T también discretos, T={t_0, t_1, t_2,…}
- Una cadena de Markov (CM) es una sucesión de variables aleatorias X_i, $i \in \mathbf{N}$, tal que:

$$P\left[X_{t+1} = j \big/ X_0, X_1, \ldots, X_t \right] = P\left[X_{t+1} = j \big/ X_t \right]$$

que es la expresión algebraica de la propiedad de Markov para T discreto.

Probabilidades de transición

- Las CM están completamente caracterizadas por las probabilidades de transición en una etapa,

$$P\left[X_{t+1} = j \big/ X_t = i\right], \quad i, j \in S, t \in T$$

- Sólo trabajaremos con CM homogéneas en el tiempo, que son aquellas en las que

$$\forall i, j \in S \ \forall \, t \in T, P\left[X_{t+1} = j \big/ X_t = i\right] = q_{ij}$$

donde q_{ij} se llama probabilidad de transición en una etapa desde el estado i hasta el estado j

Matriz de transición

- Los q_{ij} se agrupan en la denominada matriz de transición de la CM:

$$Q = \begin{pmatrix} q_{00} & q_{01} & q_{02} & \cdots \\ q_{10} & q_{11} & q_{12} & \cdots \\ q_{20} & q_{21} & q_{22} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} = \left(q_{ij}\right)_{i,j \in S}$$

Propiedades de la matriz de transición

- Por ser los q_{ij} probabilidades,

$$\forall i, j \in S, \quad q_{ij} \in [0,1]$$

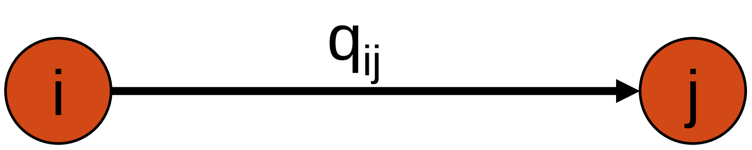
- Por ser 1 la probabilidad del suceso seguro, cada fila ha de sumar 1, es decir,

$$\forall i \in S, \quad \sum_{j \in S} q_{ij} = 1$$

- Una matriz que cumpla estas dos propiedades se llama matriz estocástica

Diagrama de transición de estados

- El diagrama de transición de estados (DTE) de una CM es un grafo dirigido cuyos nodos son los estados de la CM y cuyos arcos se etiquetan con la probabilidad de transición entre los estados que unen. Si dicha probabilidad es nula, no se pone arco.

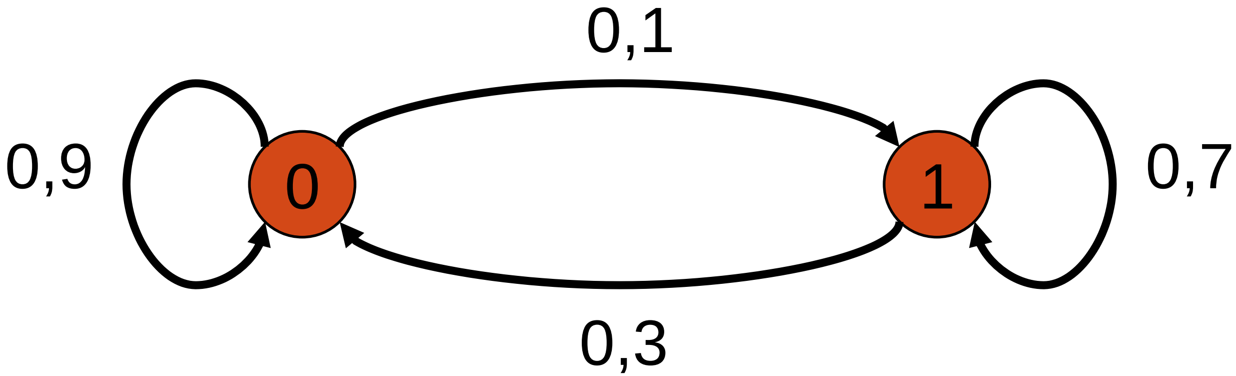


Ejemplo: línea telefónica

- Sea una línea telefónica de estados ocupado=1 y desocupado=0.
- Si en el instante t está ocupada, en el instante t+1 estará ocupada con probabilidad 0,7 y desocupada con probabilidad 0,3.
- Si en el instante t está desocupada, en el t+1 estará ocupada con probabilidad 0,1 y desocupada con probabilidad 0,9.

Ejemplo: línea telefónica

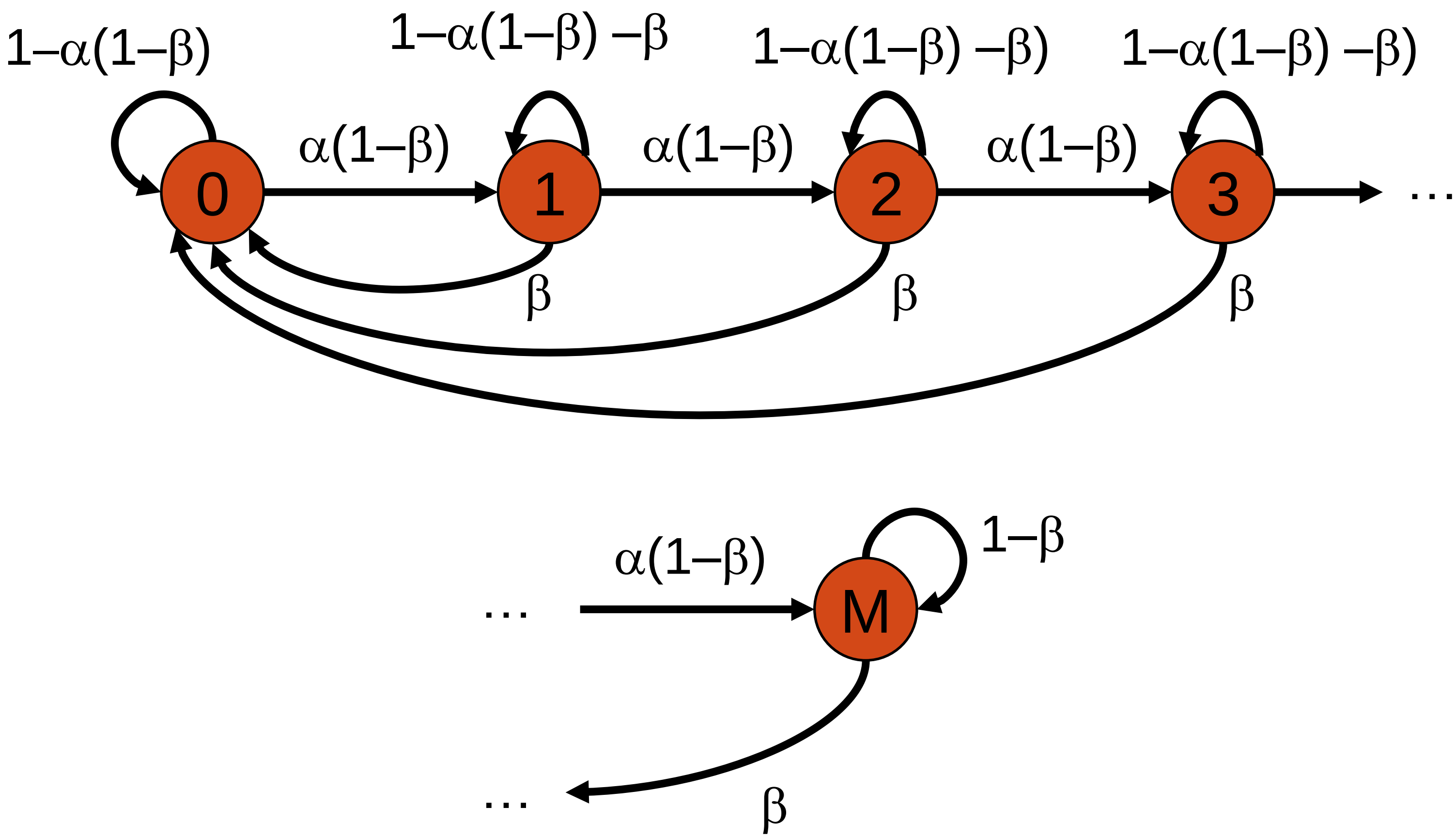
$$Q = \begin{pmatrix} 0{,}9 & 0{,}1 \\ 0{,}3 & 0{,}7 \end{pmatrix}$$



Ejemplo: *buffer* de E/S

- Supongamos que un *buffer* de E/S tiene espacio para M paquetes.
  - En cualquier instante de tiempo podemos insertar un paquete en el *buffer* con probabilidad α o bien el *buffer* puede vaciarse con probabilidad β. Si ambos casos se dan en el mismo instante, primero se inserta y luego se vacía.
- Sea X_t=nº de paquetes en el *buffer* en el instante t.
- Suponiendo que las inserciones y vaciados son independientes entre sí e independientes de la historia pasada, $\{ X_t \}$ es una CM, donde S={0, 1, 2, …, M}

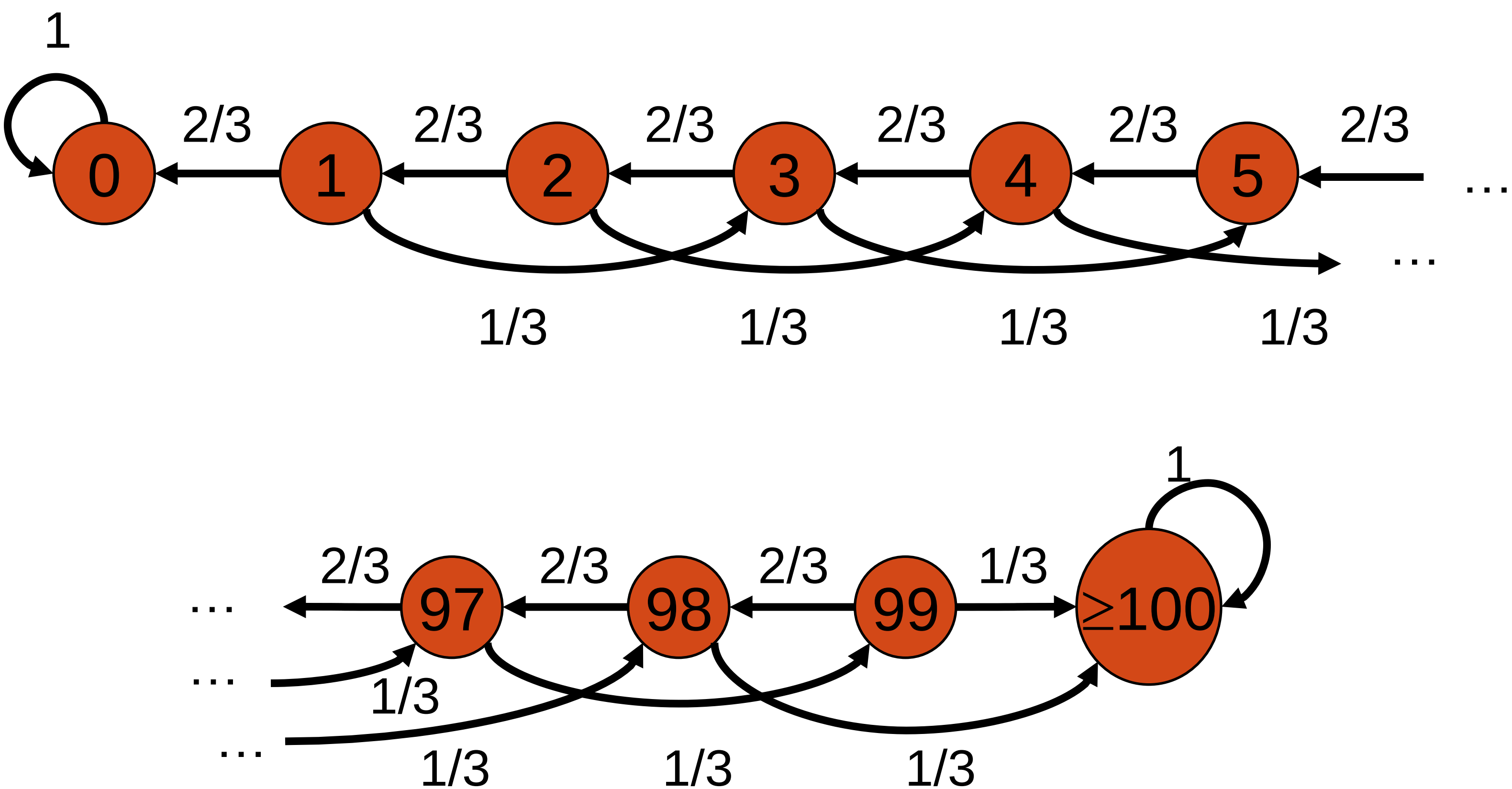
Ejemplo: *buffer* de E/S



Ejemplo: Lanzamiento de un dado

- Se lanza un dado repetidas veces. Cada vez que sale menor que 5 se pierde 1 peso, y cada vez que sale 5 ó 6 se gana 1 peso. El juego acaba cuando se tienen 0 peso ó 100 pesos.
- Sea X_t=estado de cuentas en el instante t. Tenemos que $\{ X_t \}$ es una CM
- S={0, 1, 2, …, 100}

Ejemplo: Lanzamiento de un dado



Ejemplo: organismos unicelulares

- Se tiene una población de organismos unicelulares que evoluciona así:
  - cada organismo se duplica con probabilidad 1–p o muere con probabilidad p.
- Sea X_n el nº de organismos en el instante n. La CM $\{ X_n \}$ tendrá S = { 0, 1, 2, 3, … } = **N**
- Si hay i organismos en el instante n, en el instante n+1 tendremos k organismos que se dupliquen e i–k que mueran, con lo que habrá 2k organismos.

Ejemplo: organismos unicelulares

- Mediante la distribución binomial podemos hallar las probabilidades de transición $q_{i,2k}$ (el resto de probabilidades son nulas):

$$\forall k \in \{0,1,2,\ldots,i\}, \quad q_{i,2k} = \binom{i}{k}(1-p)^k p^{i-k}$$

Hidden Markov models

- Belief nets son una forma muy poderosa de representar dependencias
- Si las dependencias tienen temporalidad, es decir se observa un proceso que se desarrolla en el tiempo, pueden observarse estados que estan muy influenciados por el estado anterior
- A pesar de que generan modelos mas complejos, las mismas ideas reaparecen
- HMM tienen una cantidad de parametros cuyos valores explican los patrones de cada categoria.
- Luego, un nuevo patron puede ser clasificado por el modelo por el estado que mejor explica el patron.

Definición de un modelo oculto de Markov

- Una Cadena de Markov Oculta es un proceso binario (doblemente estocástico) compuesto de
  - una cadena de Markov homogénea con espacio de estados finito
  - un proceso estocástico cuya distribución condicional solo depende de la cadena anterior

En cualquier estado particular, la observación puede ser generada, de acuerdo a la distribución de probabilidades de emisión.

Sólo el resultado observable, no el estado, es visible a un observador externo por lo que los estados están “ocultos”.

Las transiciones entre estados “ocultos”están dadas por un conjunto de probabilidades de transición.

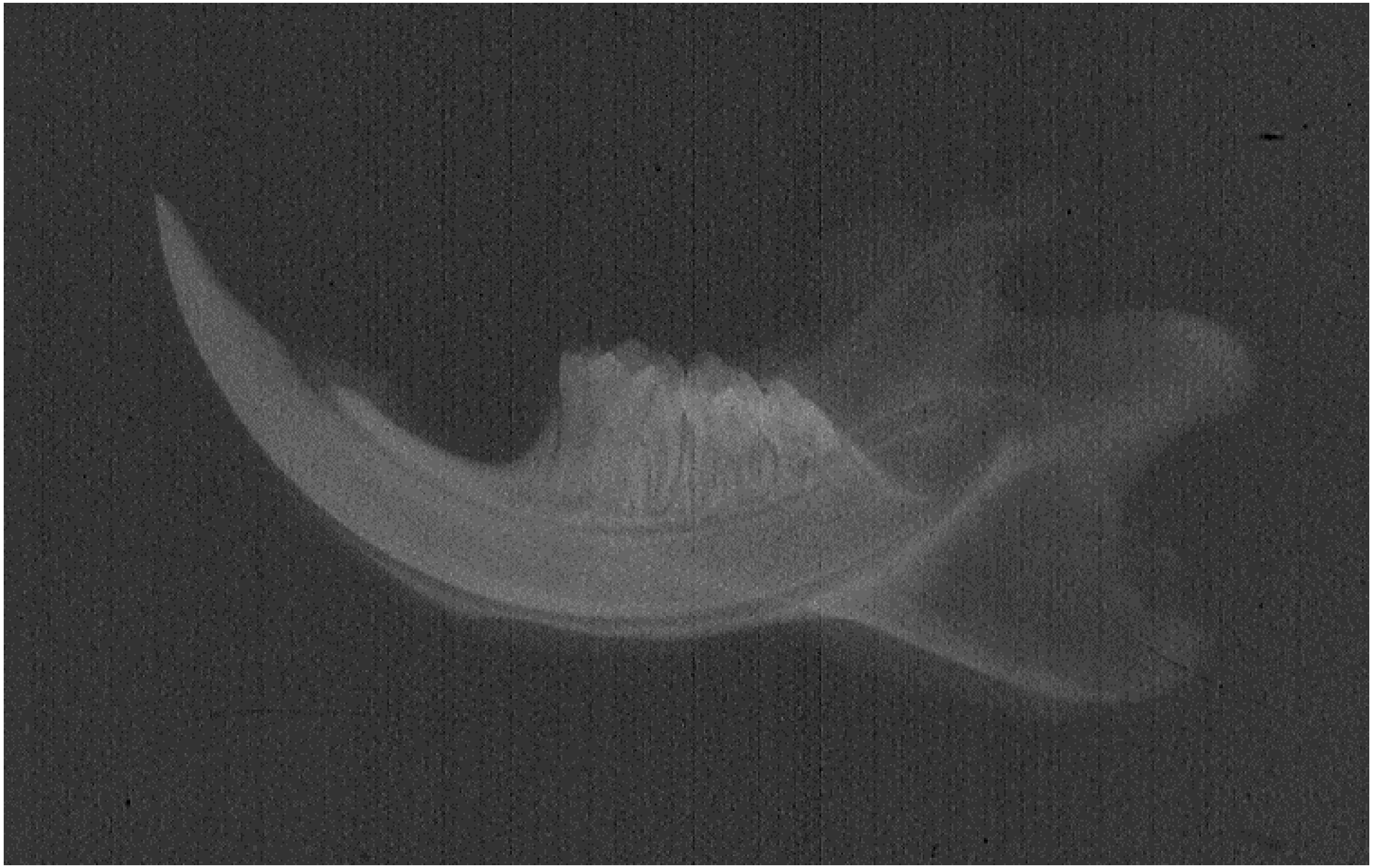
El humor del cocinero

- Supongamos que la comida que sirve el comedor es usualmente catalogada como buena regular o mala.
- Supongamos que la comida servida depende del humor del cocinero, que puede ser bueno o malo.
- Solo se observa la calidad de la comida, el humor del cocinero esta “oculto”.

Lo oculto es lo que interesa

- Supongamos entonces que estamos tratando de modelar clases visibles en secuencia, por ejemplo,
  - recorriendo una imagen de izquierda a derecha, como un vector, observamos las etiquetas de tres clases, hueso, fondo y cartilago, y nos on independientes
    - Si estamos viendo cartilago, lo veremos varios pixeles seguidos
    - Si estamos viendo fondo, lo veremos mas veces en forma continua
    - Si estamos viendo hueso, se ve menos veces que el fondo, peor mas que el cartilago

Imagen con cuatro clases ocultas



Cadena Oculta es Markoviana

- Vamos a denotar el proceso oculto por $\omega(t)$, una sucesion particular de largo T se denota por

$$\omega^T = \{\omega(1), \omega(2), ..., \omega(T)\}$$

- El modelo considera probabilidades de transicion entre los estados

$$a_{ij} = P(\omega_j(t+1) \mid \omega_i(t))$$

- Las probabilidades de transicion son independientes del tiempo y no nulas solo a un paso

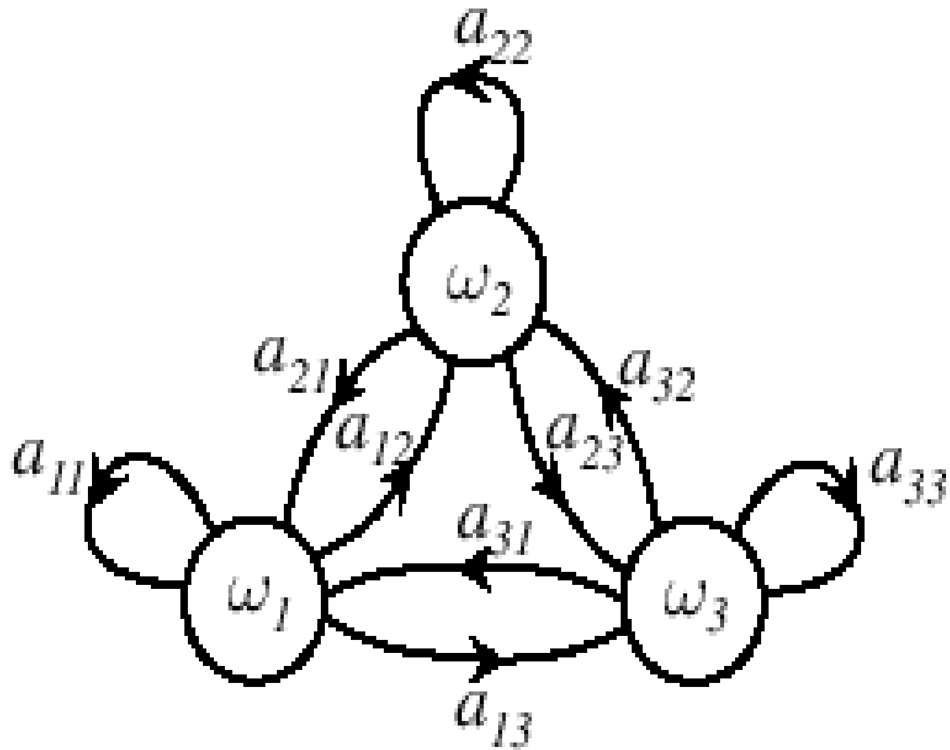


FIGURE 3.8. The discrete states, ω_i, in a basic Markov model are represented by nodes, and the transition probabilities, a_{ij}, are represented by links. In a first-order discrete-time Markov model, at any step t the full system is in a particular state $\omega(t)$. The state at step $t+1$ is a random function that depends solely on the state at step t and the transition probabilities. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Proceso visible: necesitamos ayuda!!

- La interaccion de los estados visibles $v_j(t)$ con los estados ocultos $\omega_i(t)$ esta dada por

$$b_{ij} = P(v_j(t) \mid \omega_i(t))$$

- Hay tres problemas importantes relacionados con este modelo
  - Problema de Evaluacion
  - Problema de decodificacion
  - Problema de aprendizaje

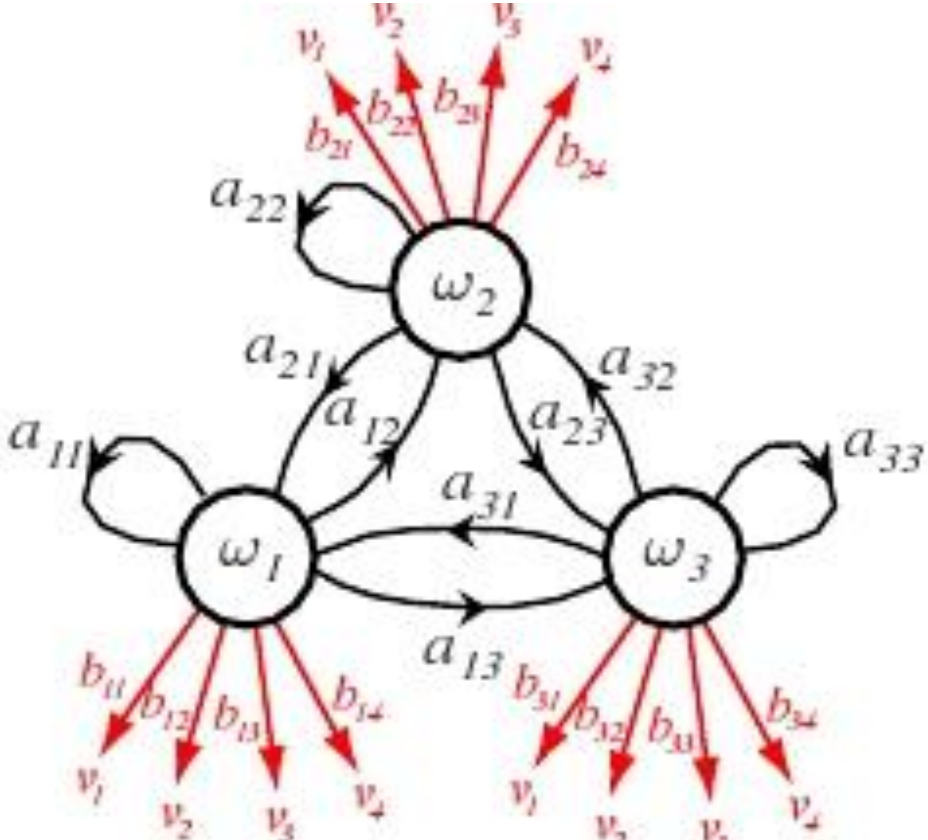


FIGURE 3.9. Three hidden units in an HMM and the transitions between them are shown in black while the visible states and the emission probabilities of visible states are shown in red. This model shows all transitions as being possible; in other HMMs, some such candidate transitions are not allowed. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Problema de evaluacion

Calcular la probabilidad de que el modelo produzca la secuencia V^T de estados visibles. Esto es:

$$P(V^T) = \sum_{r=1}^{r_{max}} P(V^T \mid \omega_r^T) P(\omega_r^T)$$

donde cada r indexa una secuencia particular de T estados ocutos

$$\omega_r^T = \{\omega(1), \omega(2), \ldots, \omega(T)\}$$

$$(1) \qquad P(V^T \mid \omega_r^T) = \prod_{t=1}^{t=T} P(v(t) \mid \omega(t))$$

$$(2) \qquad P(\omega_r^T) = \prod_{t=1}^{t=T} P(\omega(t) \mid \omega(t-1))$$

Usando las ecuaciones (1) and (2), se puede escribir:

$$P(V^T) = \sum_{r=1}^{r_{max}} \prod_{t=1}^{t=T} P(v(t) \mid \omega(t))\, P(\omega(t) \mid \omega(t-1))$$

Interpretacion:

La probabilidad de que se observe una secuencia particular de T estados visibles, V^T es igual a la suma sobre todos las r_{max} posibles secuencias de estados ocultos, de las probabilidades condicionales de que el sistema haya realizado esa transicion particular, multiplicada por la probabilidad de que haya emitido el estado visible de la secuencia observada

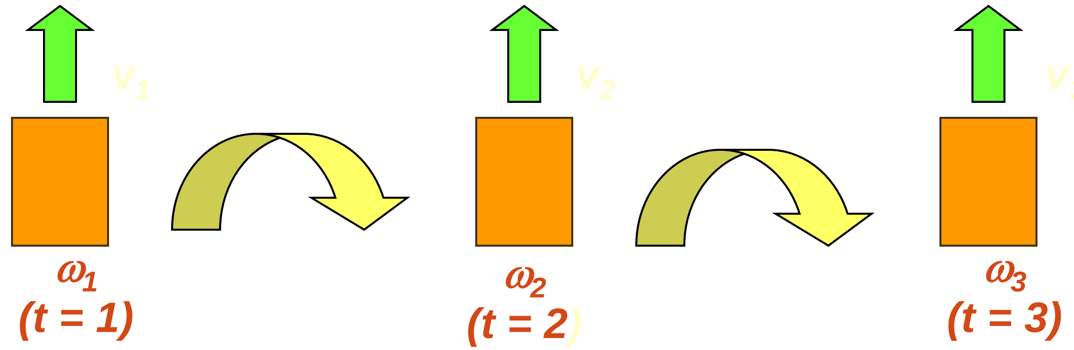
Ejemplo:

Sea $\omega_1, \omega_2, \omega_3$ los estados ocultos; v_1, v_2, v_3 los estados visibles y $V^3 = \{v_1, v_2, v_3\}$ las secuencias de estados visibles

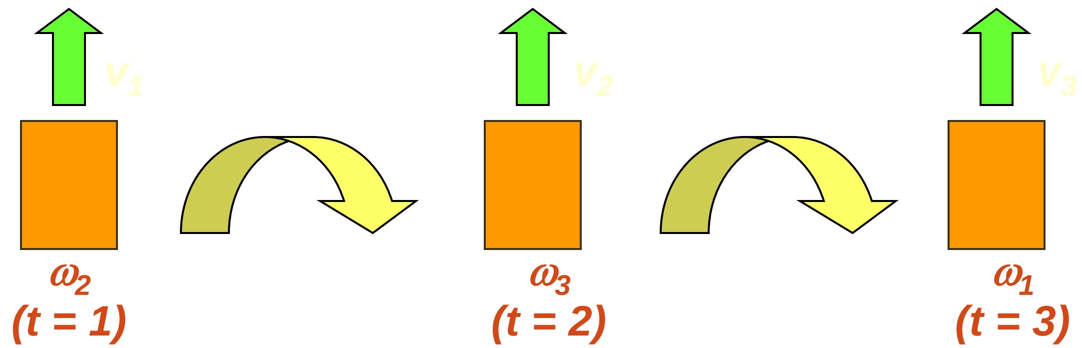
$$P(\{v_1, v_2, v_3\}) = P(\omega_1).P(v_1 \mid \omega_1).P(\omega_2 \mid \omega_1).P(v_2 \mid \omega_2).P(\omega_3 \mid \omega_2).P(v_3 \mid \omega_3)$$

+…+ (terminos posibles en la suma= todos los posibles(3^3= 27) cases !)

Primera posibilidad:



Segunda posibilidad:



$P(\{v_1, v_2, v_3\}) = P(\omega_2).P(v_1 \mid \omega_2).P(\omega_3 \mid \omega_2).P(v_2 \mid \omega_3).P(\omega_1 \mid \omega_3).P(v_3 \mid \omega_1) + \ldots +$

Por lo cual:

$$P(\{v_1, v_2, v_3\}) = \sum_{\substack{\text{sequencias} \\ \text{posibles}}} \prod_{t=1}^{t=3} P(v(t) \mid \omega(t)).P(\omega(t) \mid \omega(t-1))$$

Problema de decodificacion: secuencia optima

Dada una secuencia de estados visibles V^T, el problema de decodificacion es encontrar la secuencia mas probable de estados ocultos .

$$Encontrar \quad \hat{\omega}(1), \hat{\omega}(2), ..., \hat{\omega}(T) \;\; tal \;\; que:$$

$$\hat{\omega}(1), \hat{\omega}(2), ..., \hat{\omega}(T) = \underset{\omega(1), \omega(2), ..., \omega(T)}{\arg\max} \; P\left[\omega(1), \omega(2), ..., \omega(T), v(1), v(2), ..., V(T) \mid \lambda\right]$$

Notemos que la sumatoria desaparecio, siendo que queremos encontrar un unico caso

Si el modelo se escribe como: $\lambda = [\pi, A, B]$

$\pi = P(\omega(1) = \omega)$ *(Probabilidades iniciales de los estados)*

$A = a_{ij} = P(\omega(t+1) = j \mid \omega(t) = i)$ *(transiciones ocultas)*

$B = b_{jk} = P(v(t) = k \mid \omega(t) = j)$ *(relaciones condicionales)*

En el ejemplo anterior el computo corresponde a la seleccion del mejor camino entre los siguientes:

$$\{\omega_1(t=1), \omega_2(t=2), \omega_3(t=3)\},\ \{\omega_2(t=1), \omega_3(t=2), \omega_1(t=3)\}$$
$$\{\omega_3(t=1), \omega_1(t=2), \omega_2(t=3)\},\ \{\omega_3(t=1), \omega_2(t=2), \omega_1(t=3)\}$$
$$\{\omega_2(t=1), \omega_1(t=2), \omega_3(t=3)\}$$

Problema de aprendizaje

- El tercer problema consiste en determinar el metodo de ajustar los parametros del modelo $\lambda = [\pi, A, B]$ que satisface un problema de optimizacion. Se busca el mejor modelo

$$\hat{\lambda} = [\hat{\pi}, \hat{A}, \hat{B}]$$

que maximice la probabilidad de observar la secuencia

$$\underset{\lambda}{Max}\, P(V^T \mid \lambda)$$

Se usan procedimientos iterativos para obtener la optimizacion