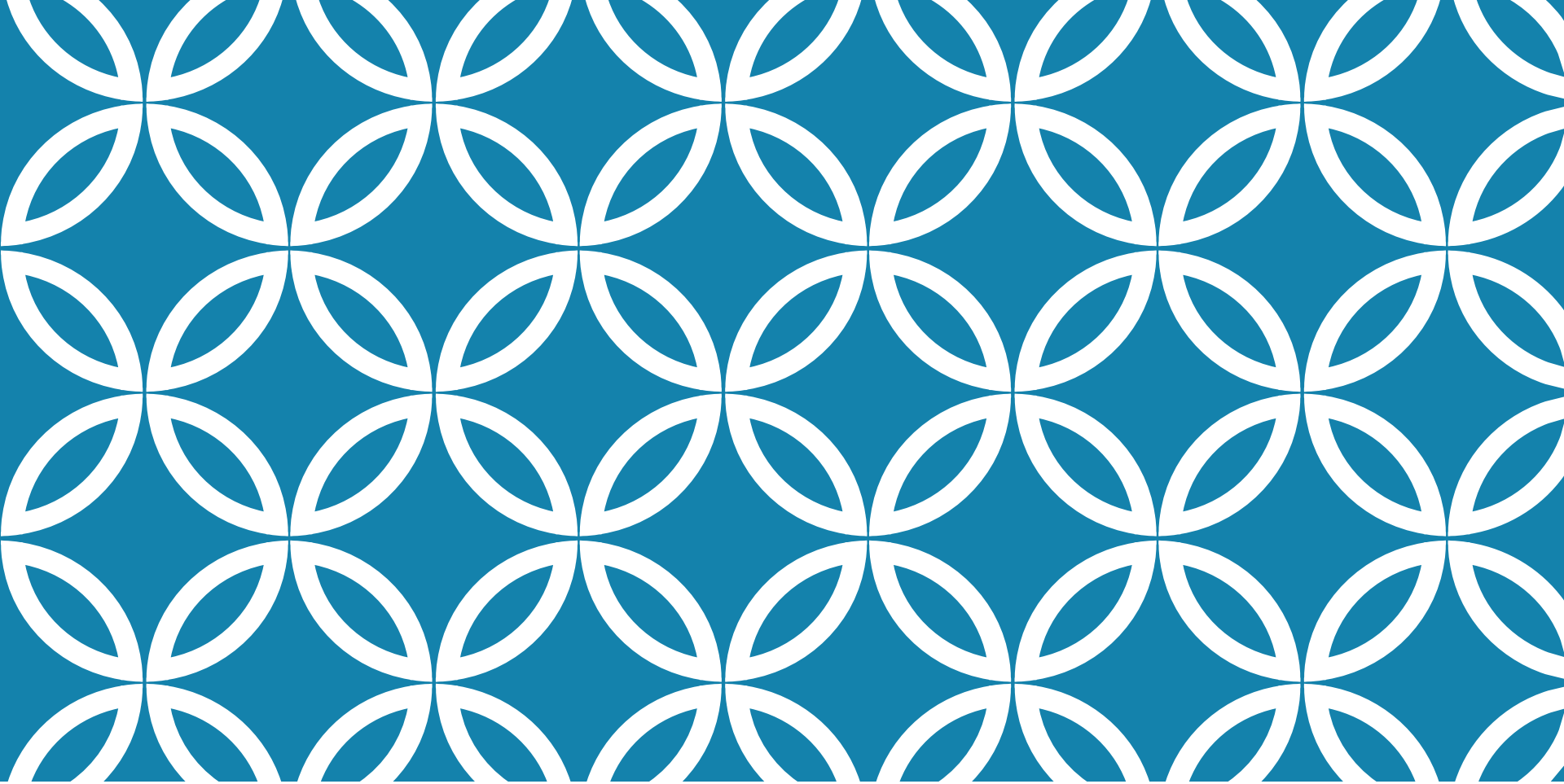


RAZONAMIENTO PROBABILÍSTICO EN EL TIEMPO. MODELOS DE MARKOV

José Manuel Molina López
Grupo de Inteligencia Artificial Aplicada

CONTENIDO

Probabilidad en el tiempo. Procesos Estocásticos

Modelos de Markov

Modelos de Markov Ocultos (HMMs)

Procesos de Decisión de Markov (MDPs)

Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

MOTIVACIÓN

¿Qué sabemos? Razonamiento probabilístico en mundos estáticos

El mundo es dinámico ¿Cómo podemos considerar la influencia del tiempo?

Por ejemplo, cómo afecta el hecho de que lloviera ayer ($Llueve_t$) al hecho de que llueva hoy ($Llueve_{t+1}$)

Modelos de Markov (MM) (también denominados Procesos de Markov o cadenas de Markov) y Modelos de Markov ocultos (HMM)

¿Dónde se pueden utilizar? Los algoritmos de búsqueda proporcionan una secuencia de acciones para resolver el problema, pero ¿Qué pasa si las acciones no son deterministas? ¿Cómo podemos elegir las acciones que tienen más probabilidad de llevarnos al estado meta?

Procesos de Decisión de Markov (MDPs) y Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

PROCESOS ESTOCÁSTICOS

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

Un proceso estocástico es una familia de variables aleatorias definida sobre un espacio de probabilidad.

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

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

PROCESOS ESTOCÁSTICOS

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

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

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

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:

$$S = \{X_t(\omega) | \, t \in T \wedge \omega \in \Omega\}$$

EJEMPLO DE PROCESO ESTOCÁSTICO

Lanzamos una moneda al aire 6 veces. El jugador gana 1 € cada vez que sale cara (C), y pierde 1 € cada vez que sale cruz (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(ω)=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}

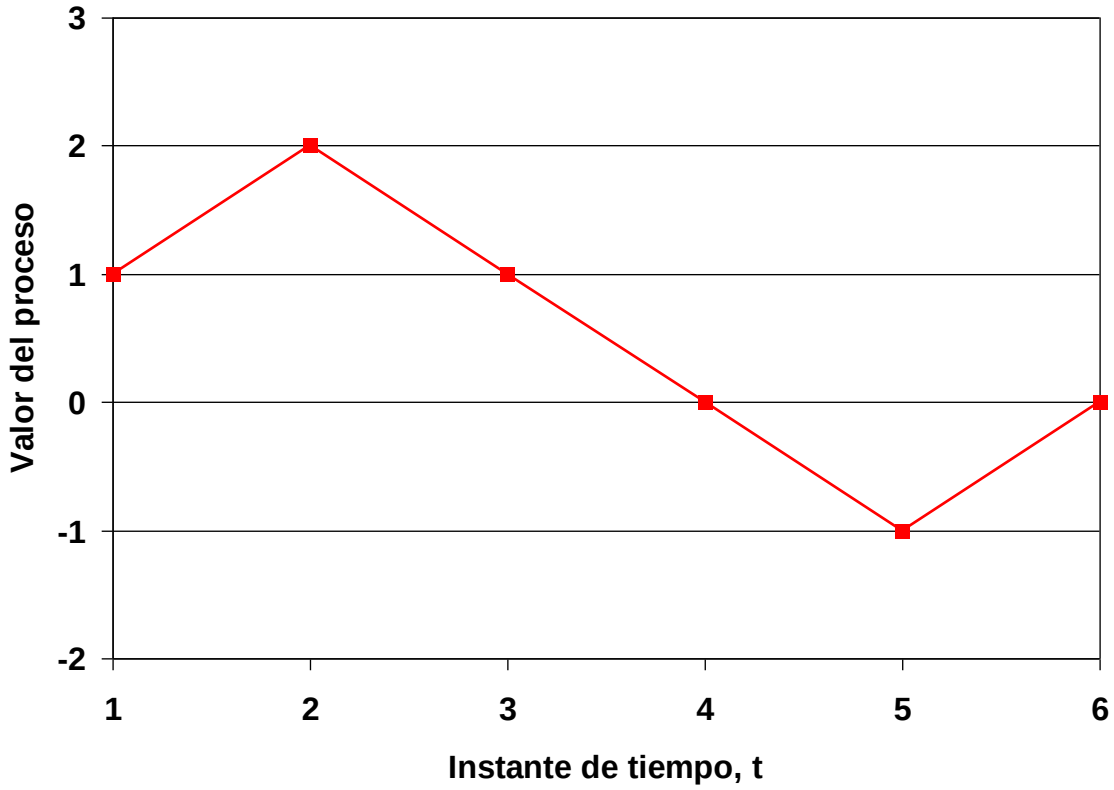
EJEMPLO DE PROCESO ESTOCÁSTICO

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:

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

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

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

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

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

CLASIFICACIÓN DE LOS PROCESOS ESTOCÁSTICOS

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

MODELOS, CADENAS Y PROCESOS 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, $T=\{t_0, t_1, t_2,\ldots\}$

Una cadena de Markov (CM) es una sucesión de variables aleatorias X_i, $i \in \mathbf{N}$, tal que:

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

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,

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

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

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

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,

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

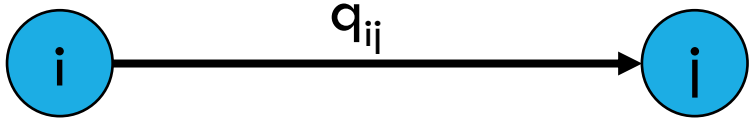
- Por ser 1 la probabilidad del suceso seguro, cada fila ha de sumar 1, es decir,

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

- 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.

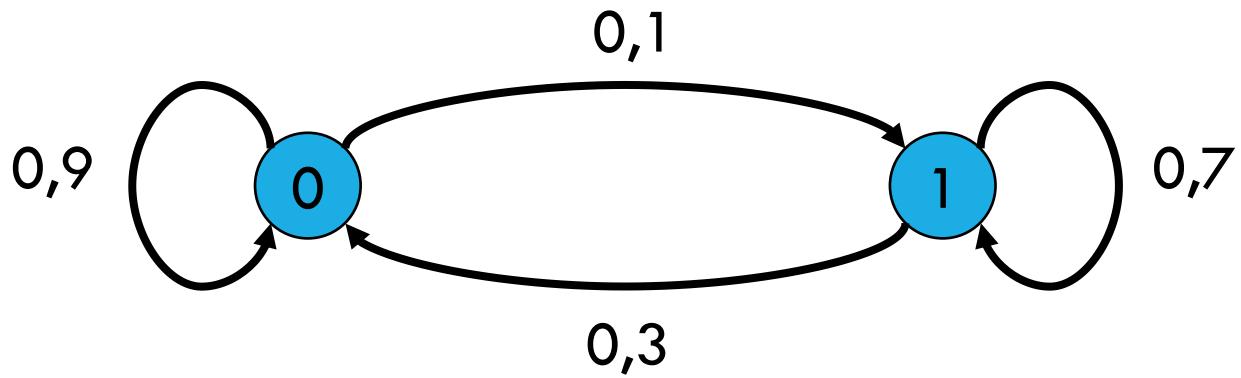


EJEMPLO: LÍNEA TELEFÓNICA

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

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

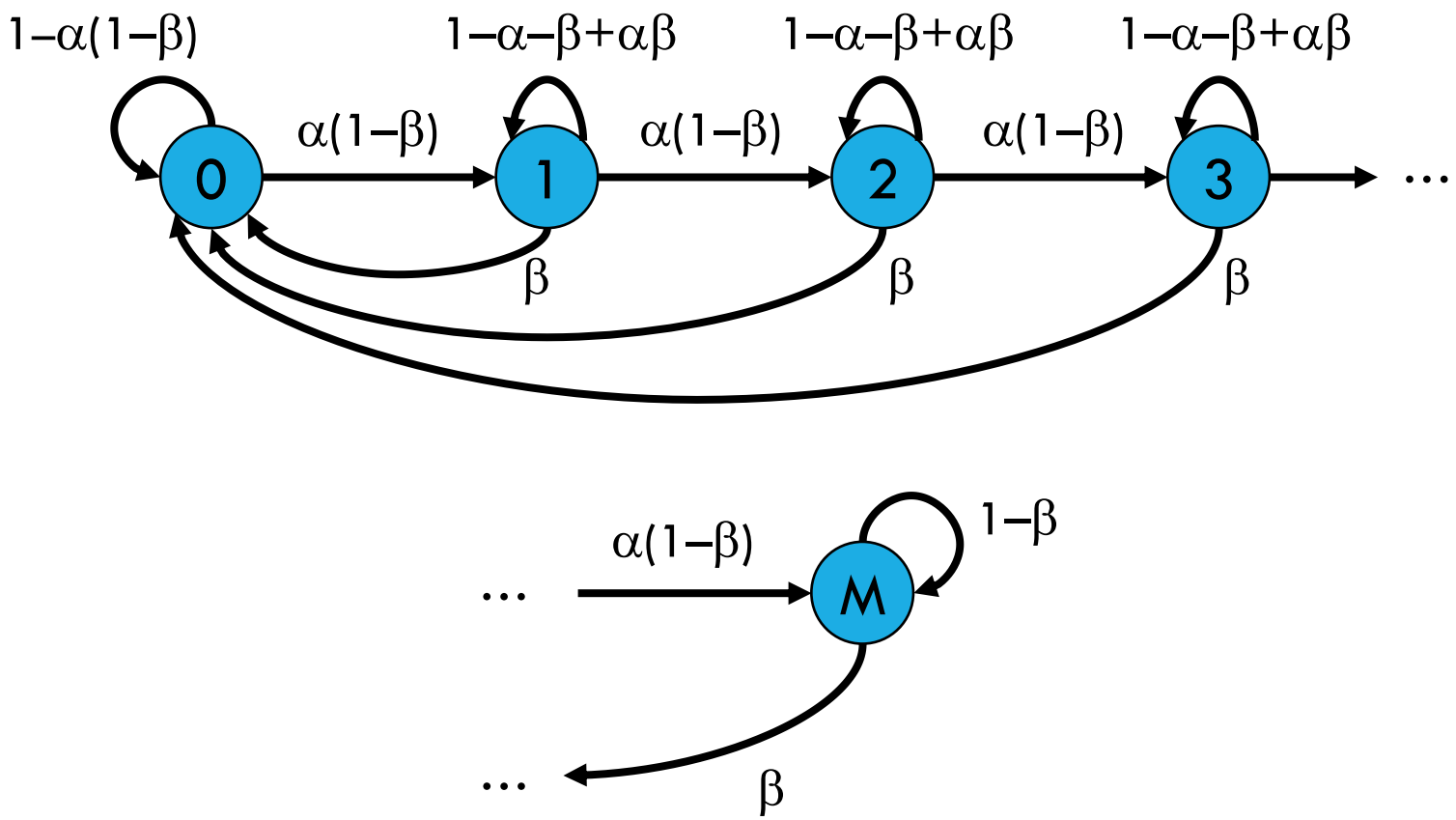


EJEMPLO: *BUFFER* DE E/S

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=n° 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



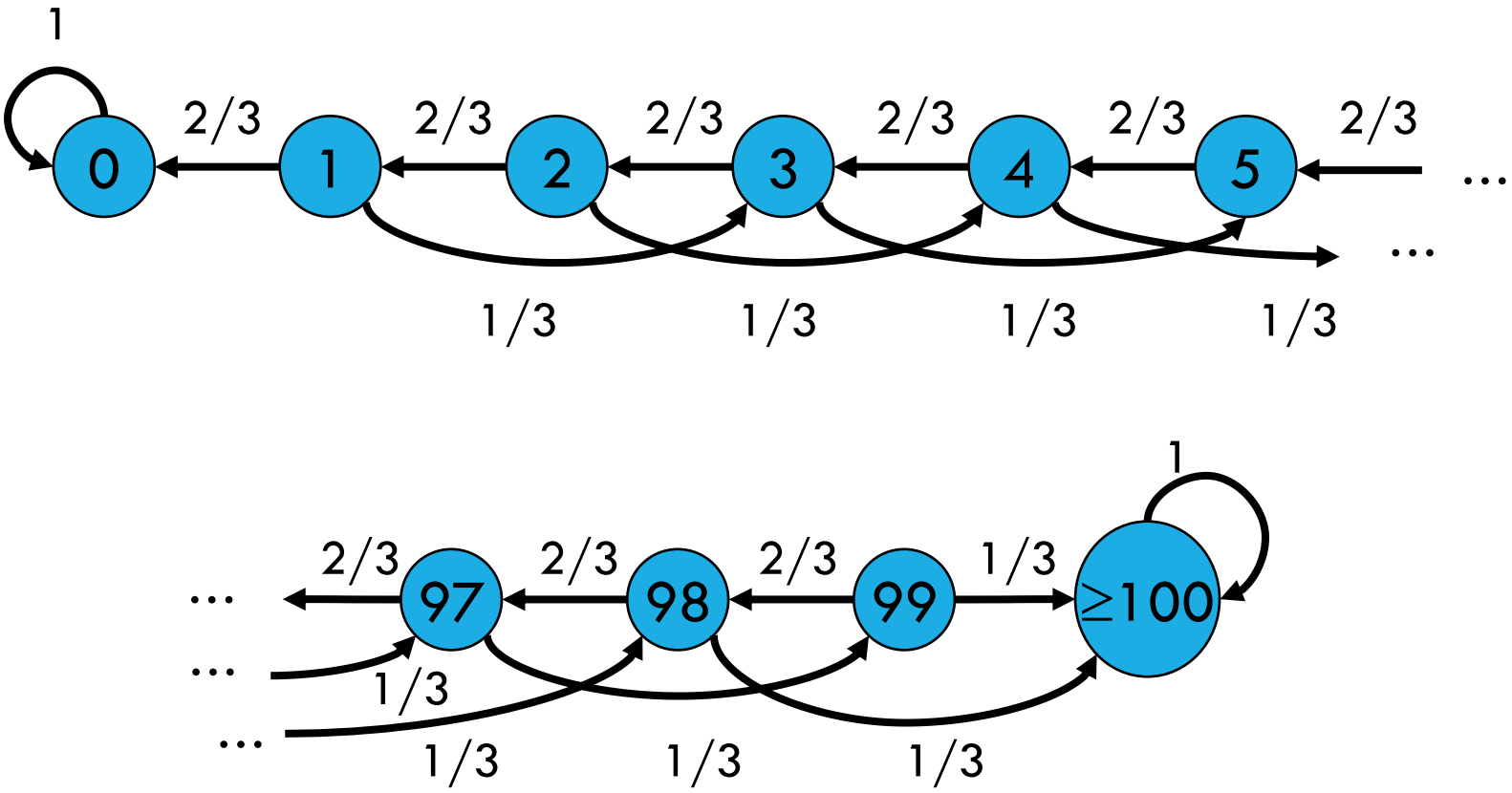
EJEMPLO: LANZAMIENTO DE UN DADO

Se lanza un dado repetidas veces. Cada vez que sale menor que 5 se pierde 1 €, y cada vez que sale 5 ó 6 se gana 1 €. El juego acaba cuando se tienen 0 € ó 100 €.

Sea X_t=estado de cuentas en el instante t. Tenemos que $\{ X_t \}$ es una CM

S={0, 1, 2, …, 100}

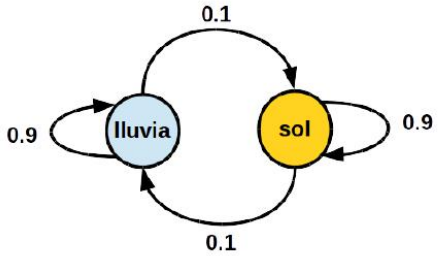
EJEMPLO: LANZAMIENTO DE UN DADO



EJEMPLO: SOL Y LLUVIA

Ejemplo pequeño

- Estados: $E_t = \{Tiempo_t\}$
- Transiciones: $P(E_t/E_{t-1})$



- Inicialmente, $P(E_0 = sol) = 1$, $P(E_0 = lluvia) = 0$
- ¿Cuál es la probabilidad de sol al día siguiente, E_1?

$$P(E_1 = sol) = P(E_1 = sol/E_0 = sol)P(E_0 = sol) + P(E_1 = sol/E_0 = lluvia)P(E_0 = lluvia)$$
$$= 0.9 \times 1.0 + 0.1 \times 0.0 = 0.9$$

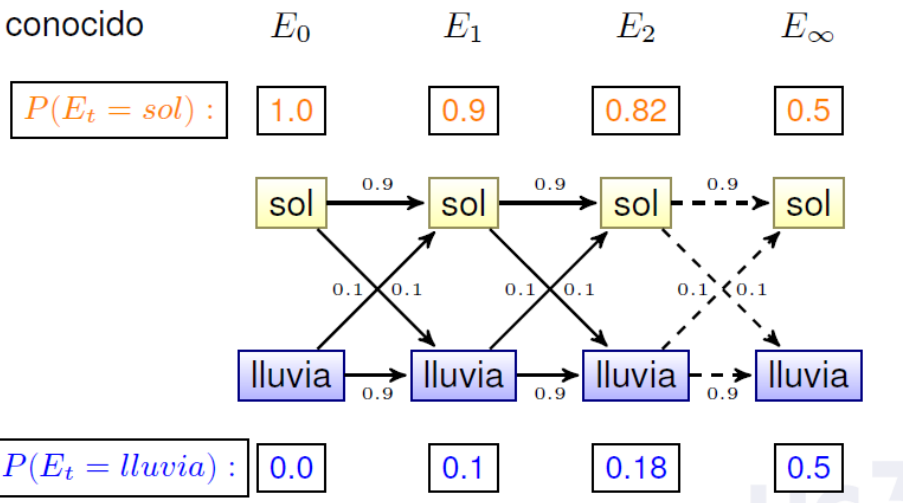
EJEMPLO: SOL Y LLUVIA

Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido



DOS CUESTIONES RELEVANTES DE LOS MODELOS DE MARKOV

Se pueden representar mediante una red bayesiana que represente la distribución de la probabilidad conjunta:

- Nodos: E_0 E_1 E_2 E_3 … E_n
- Probabilidades de transición: $p(E_{t+1} / E_t)$

¿Cómo podemos razonar para n muy grande?

ECUACIONES DE CHAPMAN-KOLMOGOROV

Teorema: Las probabilidades de transición en n etapas vienen dadas por la matriz Q^n:

$$\forall i, j \in S, P\left[X_{t+n} = j \,/\, X_t = i \right] = q_{ij}^{(n)}$$

- **Demostración:** Por inducción sobre n
  - Caso base (n=1). Se sigue de la definición de q_{ij}

ECUACIONES DE CHAPMAN-KOLMOGOROV

- Hipótesis de inducción. Para cierto n, suponemos cierta la conclusión del teorema.
- Paso inductivo (n+1). Para cualesquiera i,j∈S,

$$P\left[X_{t+n+1} = j \middle/ X_t = i \right] = \sum_{k \in S} P\left[\left(X_{t+n} = k \right) \wedge \left(X_{t+n+1} = j \right) \middle/ X_t = i \right] =$$

$$= \sum_{k \in S} P\left[X_{t+n} = k \middle/ X_t = i \right] P\left[X_{t+n+1} = j \middle/ X_{t+n} = k \right] = \{H.I.\} =$$

$$= \sum_{k \in S} q_{ik}^{(n)} P\left[X_{t+n+1} = j \middle/ X_{t+n} = k \right] = \sum_{k \in S} q_{ik}^{(n)} q_{kj} = q_{ij}^{(n+1)}$$

ECUACIONES DE CHAPMAN-KOLMOGOROV

Por este teorema sabemos que la probabilidad de transitar de i hasta j en n pasos es el elemento (i,j) de Q^n. Para evitar computaciones de potencias elevadas de matrices, se intenta averiguar el comportamiento del sistema en el límite cuando $n \to \infty$, llamado también comportamiento a largo plazo

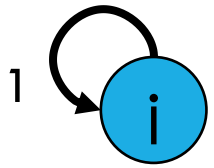
A continuación estudiaremos esta cuestión

CLASIFICACIÓN DE ESTADOS

Probabilidad de alcanzar un estado:

$$\forall i, j \in S, \quad \nu_{ij} = P\left[X_n = j \text{ para algún } n > 0 \,/\, X_0 = i \right]$$

- Diremos que un estado $j \in S$ es alcanzable desde el estado $i \in S$ sii $\nu_{ij} \neq 0$. Esto significa que existe una sucesión de arcos (camino) en el DTE que van desde i hasta j.
- Un estado $j \in S$ es absorbente sii $q_{jj}=1$. En el DTE,



- Además los estados pueden ser:
  - Efímero. Ningún estado conduce a él.
  - Transitorio. Tras pasar por él, al cabo de cierto número de etapas, la cadena de Markov ya no regresa a él.
  - Recurrente. Si no es transitorio, esto es, si tras pasar por él, la cadena de Markov siempre regresa a él. Se puede definir un “periodo”.

DISTRIBUCIÓN ESTACIONARIA Y COMPORTAMIENTO LÍMITE

Definición. Л es una distribución estacionaria sobre E si Л P= Л .

1. Las distribuciones estacionarias otorgan probabilidad cero a los estados transitorios.
2. Cada grupo de estados recurrentes intercomunicados tiene una única distribución estacionaria.
3. Cuando el número de etapas converge a infinito,

$$P^t \longrightarrow S \qquad \text{y} \qquad P^{(t)} = P^{(0)}P^t \longrightarrow P^{(0)}S$$

4. Si R_t es el número de veces que la cadena pasa por el estado E_i en las t primeras etapas, cuando t tiende a infinito,

$$\frac{R_i}{t} \longrightarrow P^{(0)}S \qquad \text{casi seguro.}$$

ESTIMACIÓN DE LOS PARÁMETROS DE UNA CADENA DE MARKOV

A partir de una realización de la cadena de Markov, se pueden estimar las probabilidades de transición mediante las siguientes proporciones observadas:

$$\hat{p}_{ij} = \frac{\text{Numero de transiciones observadas de } E_i \text{ a } E_j}{\text{Numero de transiciones observadas desde } E_i}$$

Esto presenta limitaciones dependiendo de cómo haya evolucionado la realización observada. Además, no permite estimar las probabilidades iniciales.

Por estos motivos es conveniente disponer de varias realizaciones de la cadena de Markov.

MODELOS OCULTOS DE MARKOV

En el Modelo de Markov, el estado en el que nos encontramos es conocido. Pero cómo podemos manejar una situación en la que recibimos una observación del sistema pero no sabemos en qué estado está el sistema y tenemos que inferirlo.

Si cada estado de un modelo de Markov emite una observación con una cierta incertidumbre

¿Podemos saber en qué estado se encuentra el sistema al recibir la observación?

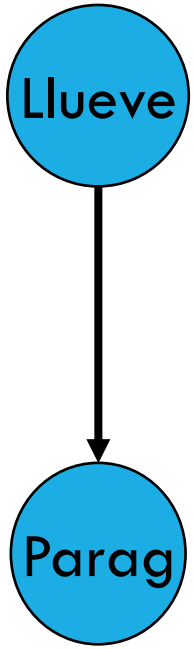
UN EJEMPLO

Un alumno estudia día tras día en su estudio sin ventanas, y quiere saber si llueve o no. Hay una sola persona a la que ve entrar cada día, su compañero de piso, y puede ver si lleva o no un paraguas.

Variables aleatorias:

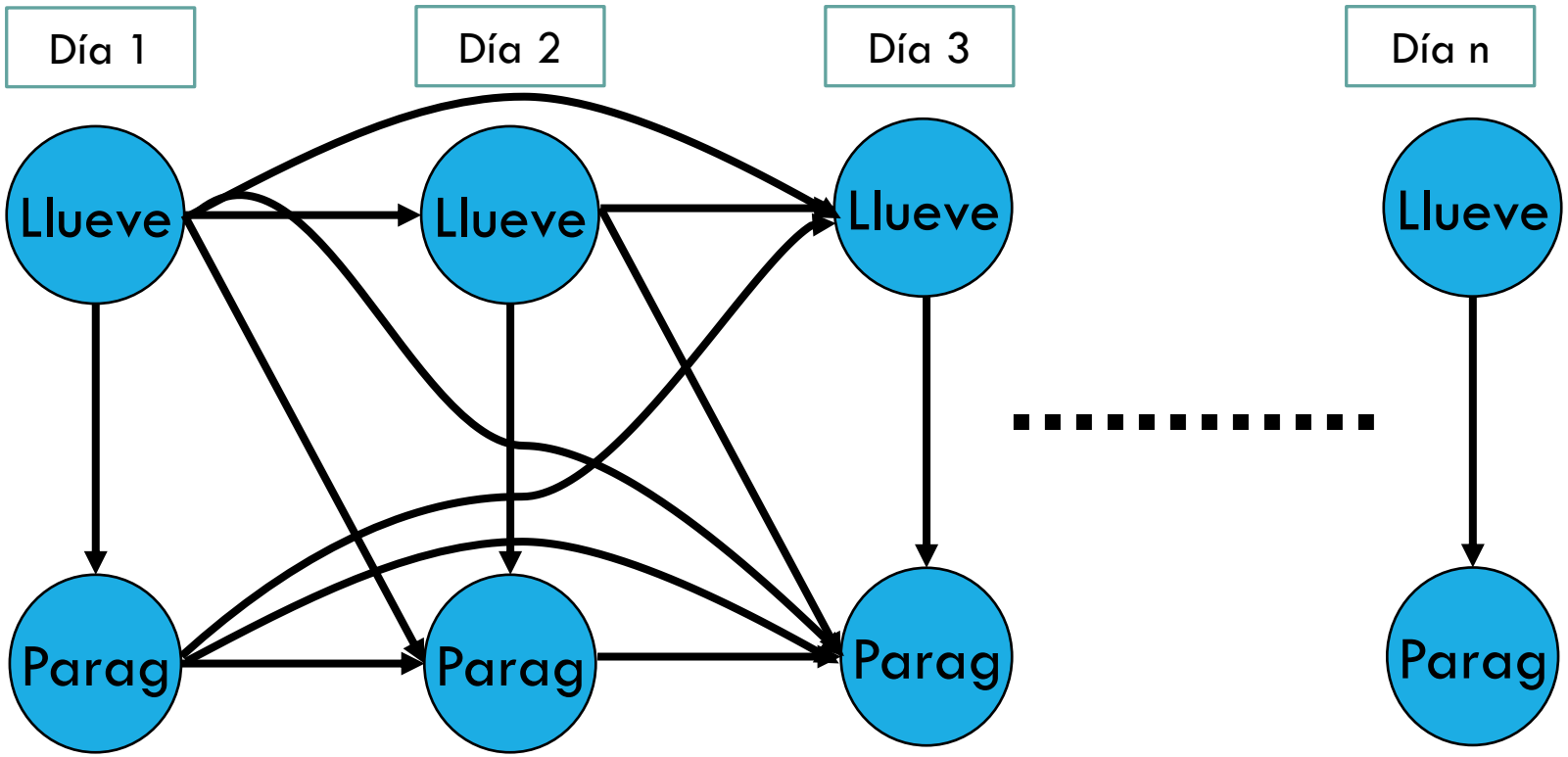
- L: Llueve. (**Estado oculto**)
- P: La persona lleva un paraguas. (**Observación**)

Para representarlo, podemos construir la siguiente Red Bayesiana:



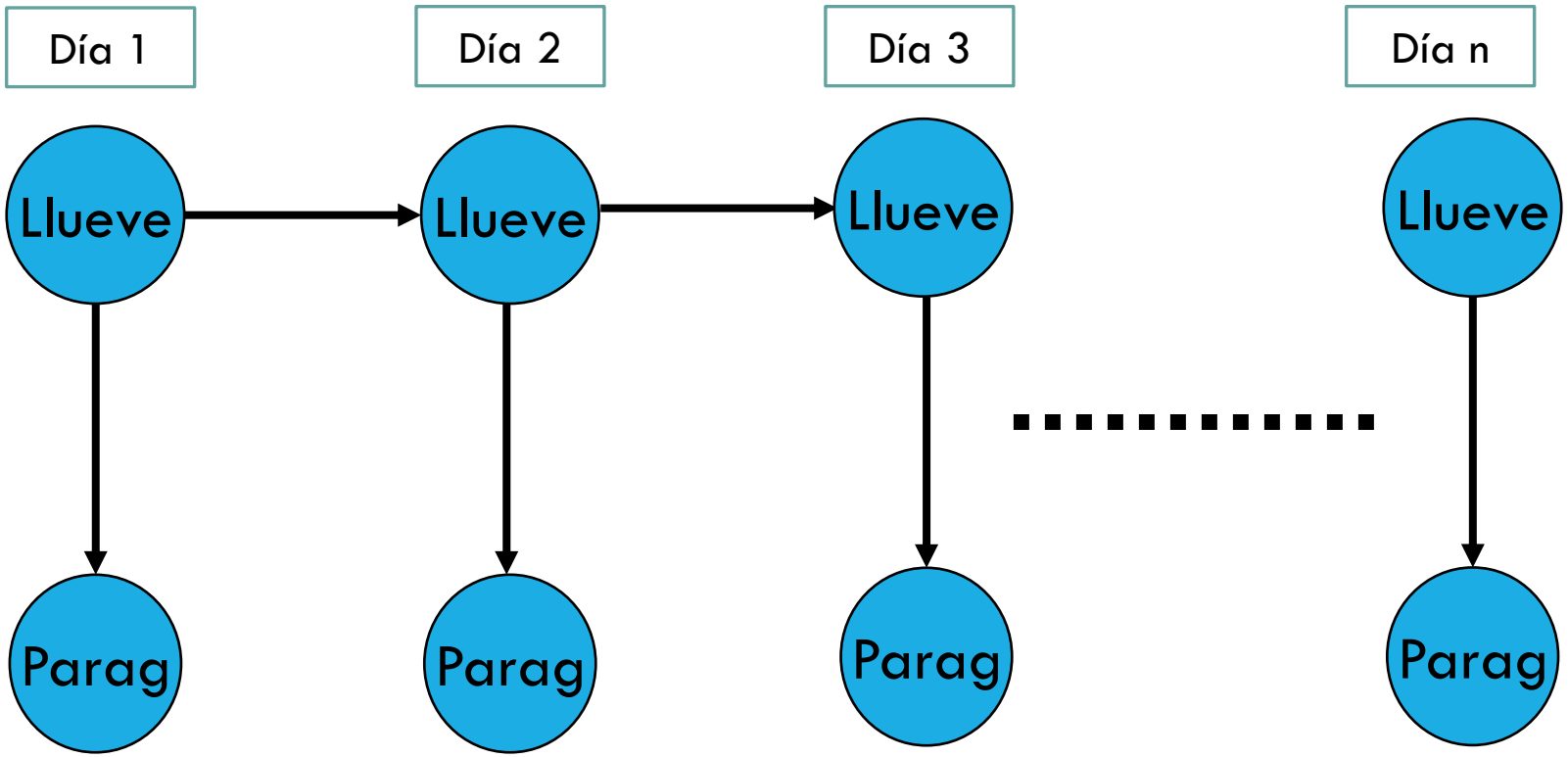
UN EJEMPLO

Si introducimos el tiempo se complica:



UN EJEMPLO

Si lo hacemos Markoviano, HMM:



INFERENCIA EN HMM

Inferencia

- Filtrado o monitorización. Distribución de probabilidad de estado actual dada evidencia previa
- Predicción. Futuros estados, dada evidencia previa
- Suavizado. Para estimar los estados pasados, dado evidencia actual

Explicación más probable

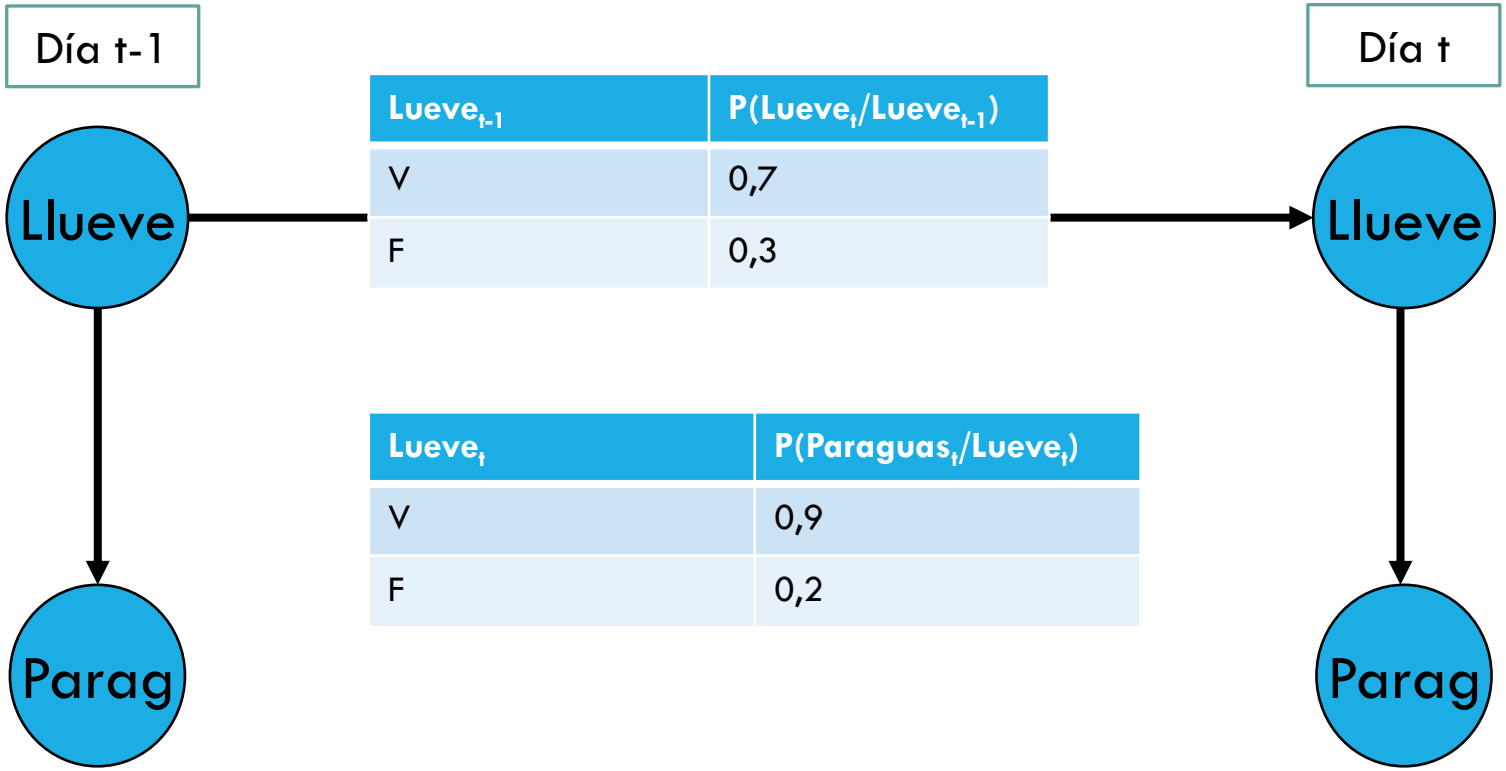
- Secuencia de estados que maximiza una secuencia de observables

Estimación de los parámetros del HMM.

- Aquellos que minimizan una secuencia de observables de una secuencia de estados

UN EJEMPLO DE FILTRADO

Si trajo paraguas los dos primeros días seguidos, ¿cuál es la probabilidad de que el segundo esté lloviendo?

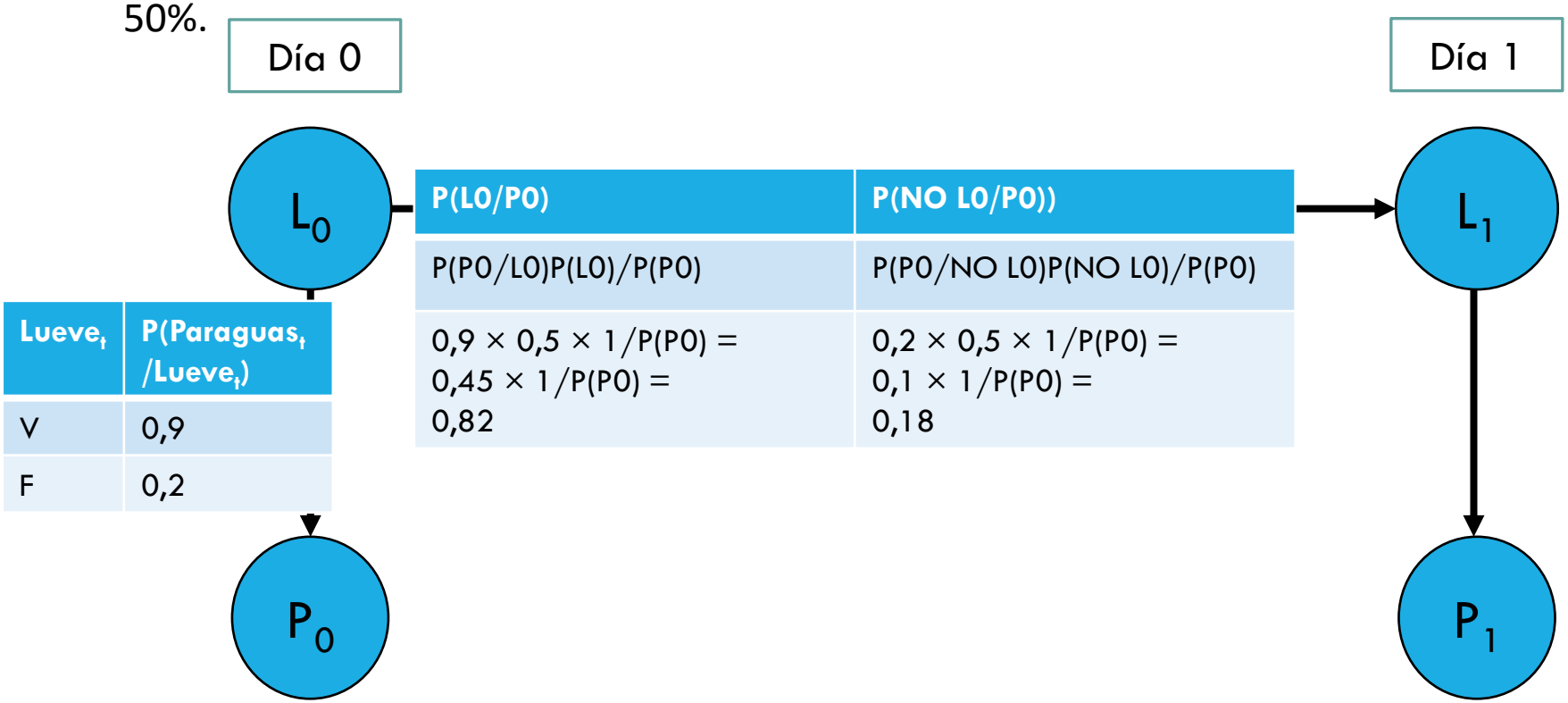


$Lueve_{t-1}$	$P(Lueve_t/Lueve_{t-1})$
V	0,7
F	0,3

$Lueve_t$	$P(Paraguas_t/Lueve_t)$
V	0,9
F	0,2

UN EJEMPLO DE FILTRADO

Si trajo paraguas los dos primeros días seguidos, ¿cuál es la probabilidad de que el segundo esté lloviendo? Siendo la probabilidad a priori de lluvia el 50%.

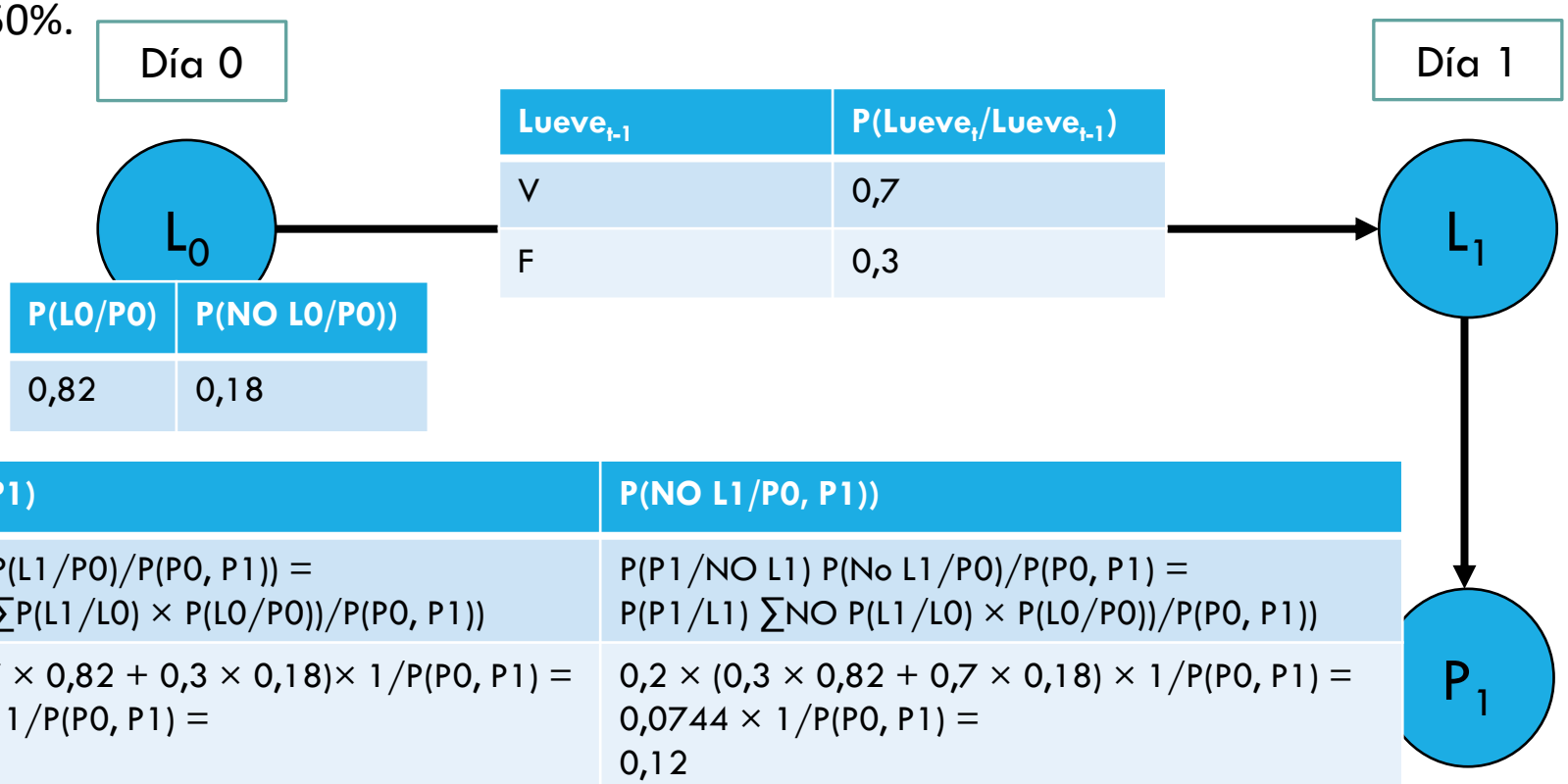


$Lueve_t$	$P(Paraguas_t/Lueve_t)$
V	0,9
F	0,2

P(L0/P0)	P(NO L0/P0))
P(P0/L0)P(L0)/P(P0)	P(P0/NO L0)P(NO L0)/P(P0)
0,9 × 0,5 × 1/P(P0) = 0,45 × 1/P(P0) = 0,82	0,2 × 0,5 × 1/P(P0) = 0,1 × 1/P(P0) = 0,18

UN EJEMPLO DE FILTRADO

Si trajo paraguas los dos primeros días seguidos, ¿cuál es la probabilidad de que el segundo esté lloviendo? Siendo la probabilidad a priori de lluvia el 50%.



$Lueve_{t-1}$	$P(Lueve_t/Lueve_{t-1})$
V	0,7
F	0,3

P(L0/P0)	P(NO L0/P0))
0,82	0,18

P(L1/P0, P1)	P(NO L1/P0, P1))
P(P1/L1) P(L1/P0)/P(P0, P1)) = P(P1/L1) ∑P(L1/L0) × P(L0/P0))/P(P0, P1))	P(P1/NO L1) P(No L1/P0)/P(P0, P1) = P(P1/L1) ∑NO P(L1/L0) × P(L0/P0))/P(P0, P1))
0,9 × (0,7 × 0,82 + 0,3 × 0,18)× 1/P(P0, P1) = 0,5652 × 1/P(P0, P1) = 0,88	0,2 × (0,3 × 0,82 + 0,7 × 0,18) × 1/P(P0, P1) = 0,0744 × 1/P(P0, P1) = 0,12

RESUMEN DE INFERENCIA EN CADENAS DE MARKOV OCULTAS

En lugar de observar los estados de la cadena de Markov, observamos otros elementos, bajo ciertas probabilidades:

Elementos de una cadena de Markov oculta.

Espacio de estados: $E = \{E_1, E_2, \ldots, E_s\}$

Matriz de transición: $P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{pmatrix}$ siendo $p_{ij} = P\left(X_{t+1} = E_j / X_t = E_i\right)$

Alfabeto de símbolos observables: $A = \{a_1, \ldots, a_m\}$

Probabilidades de emisión: $B = \left(b_i(a)\right)$ siendo $b_i(a) = P(E_i \text{ emita el simbolo } a)$

Distribución inicial: $P^{(0)} = \left(p_1^{(0)}, p_2^{(0)}, \ldots, p_s^{(0)}\right)$ siendo $p_i^{(0)} = P\left(X_0 = E_i\right)$

TRES PROBLEMAS

Llamemos λ al conjunto de parámetros del modelo de Markov oculto, y

$$O = (O_1, \ldots, O_T)$$

a una realización de la cadena de Markov oculta.

Problema 1. Calcular $P(O / \lambda)$

Problema 2. Encontrar la secuencia de estados

$$X = (X_1, \ldots, X_T)$$

que mejor se corresponda con la secuencia observada O, bajo el modelo λ .

Problema 3. Estimar los parámetros del modelo. Lo haremos buscando λ que haga máxima $P(O / \lambda)$.

PRIMER PROBLEMA

Si supiéramos cuál ha sido la sucesión de estados, entonces

$$P(O/X,\lambda) = b_{x_1}(O_1)\cdot b_{x_2}(O_2)\cdots b_{x_T}(O_T)$$

La probabilidad de una sucesión de estados es

$$P(X/\lambda) = p_{x_1}^{(0)}\cdot p_{x_1x_2}\cdot p_{x_2x_3}\cdots p_{x_{T-1}x_T}$$

Entonces, por la ley de probabilidades totales

$$P(O/\lambda) = \sum_X P(X/\lambda)P(O/X,\lambda)$$

$$= \sum_X p_{x_1}^{(0)}\cdot p_{x_1x_2}\cdot p_{x_2x_3}\cdots p_{x_{T-1}x_T}\cdot b_{x_1}(O_1)\cdot b_{x_2}(O_2)\cdots b_{x_T}(O_T)$$

PROCEDIMIENTO ADELANTE/ATRÁS (INDUCCIÓN)

Definimos las funciones adelante así:

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_T, X_t = E_i / \lambda)$$

Las funciones adelante se pueden calcular por inducción así:

Paso inicial

$$\alpha_1(i) = p_i^{(0)} \cdot b_i(O_1)$$

Inducción

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{s} \alpha_t(j)\, p_{ji} \right] \cdot b_j(O_{t+1})$$

Paso final

$$P(O/\lambda) = \sum_{i=1}^{s} \alpha_T(i)$$

PROCEDIMIENTO ADELANTE/ATRÁS (INDUCCIÓN)

Definimos las funciones atrás así:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T / X_t = E_i, \lambda)$$

Las funciones atrás se pueden calcular por inducción así:

Paso inicial

$$\beta_T(i) = 1$$

Inducción

$$\beta_t(i) = \sum_{j=1}^{s} p_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$$

Paso final

$$P(O/\lambda) = \sum_{i=1}^{s} \alpha_T(i)$$

SEGUNDO PROBLEMA: ALGORITMO DE VITERBI

Buscamos la cadena de estados que mejor se corresponda con la secuencia observada (problema 2).
Formalizamos esto en el objetivo siguiente:

$$\max_X P(O, X/\lambda)$$

Definimos las funciones:

$$\delta_t(i) = \max_{x_1, x_2, \ldots, x_{t-1}} P(x_1, x_2, \ldots, x_{t-1}, x_t = E_i, O_1, O_2, \ldots, O_t/\lambda)$$

Estas funciones y los argumentos donde se alcanza el máximo se pueden calcular por inducción así:

Paso inicial $\delta_1(i) = p_i^{(0)} \cdot b_i(O_1)$ $\psi_1(i) = 0$

Inducción $\delta_t(i) = \max_{j \in \{1,\ldots,s\}} \left[\delta_{t-1}(j)\, p_{ij}\right] \cdot b_j(O_t)$ $\psi_t(i) = \arg\max_{j \in \{1,\ldots,s\}} \left[\delta_{t-1}(j)\, p_{ij}\right]$

Paso final $P^* = \max_i \delta_T(i)$ $x_T^* = \arg\max_i \delta_T(i)$

Secuencia de estados $x_t^* = \psi_{t+1}\left(x_{t+1}^*\right)$

TERCER PROBLEMA: ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Lo haremos por máxima verosimilitud y aplicaremos un método de tipo EM.

$$\max_{\lambda} P(O/\lambda)$$

Definimos las funciones:

$$\xi_t(i,j) = P\left(x_t = E_i, x_{t+1} = E_j / O, \lambda\right)$$

Se pueden calcular a partir de las funciones adelante y atrás así:

$$\xi_t(i,j) = \frac{\alpha_t(i)\, p_{ij} b_j(O_{t+1})\, \beta_{t+1}(j)}{P(O/\lambda)}$$

Además podemos considerar todas las transiciones que parten de un estado:

$$\gamma_t(i) = \sum_{j=1}^{S} \xi_t(i,j)$$

Los parámetros estimados se actualizan de la siguiente manera:

$$\hat{p}_i^{(0)} = \gamma_1(i)$$

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\hat{b}_i(k) = \frac{\sum_{t:O_t=a_k}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}$$

PROCESO DE DECISIÓN DE MARKOV

Acciones con efectos probabilísticos:

- La distribución de probabilidad de un estado depende de la acción "escogida" o "ejecutada".
- Objetivos diversos: encontrar qué acciones ejecutar para llegar a cierta meta con alta probabilidad, con menor coste, a la mejor meta posible, etc.

Forma de atacar el problema de toma de decisiones en condiciones de incertidumbre.

Se definen con una tupla MDP =< S; A; P;C >

- S: estados (incluyendo estado inicial y posiblemente meta(s))
- A: acciones
- P: tabla de transiciones P(s'|s,a)
- R: refuerzo R(s,a,s') o C: coste C(a) caso más simple
- Factor de descuento = 1 caso más simple

Estado observable, acciones con coste fijo, acciones no determinísticas, estado meta, minimización de coste, análogo a BÚSQUEDA (en escalada).

SOLUCIÓN DE UN MDP

En búsqueda definíamos un plan: secuencia de acciones

En MDP definimos una política: forma de decidir que acción ejecutar para cualquier estado

- Maximiza el refuerzo o minimiza el coste esperados
- Garantiza llegar a una meta

Para cada MDP existe una política óptima

Una política es más que un plan, porque tolera fallos

PROPIEDADES DE LOS MDP

Suponemos que los estados son observables (si no, sería un POMDP).

El proceso es markoviano: el resultado s' de aplicar la acción a sólo depende del estado s.

El proceso es estacionario: para el mismo estado s y la misma acción a, la distribución de los estados resultantes es siempre Pa(s'/s).

Para cada estado y acción, hay una distribución de probabilidad tal que:

$\sum_{S'} P(S'/S) = 1.$

EJEMPLO TRANSPORTE (DETERMINISTA)

Tenemos que ir desde casa a la Universidad pero tenemos dos opciones: metro y tren. Para coger el metro hay que estar en la Estación. Para coger el tren, en la estación de Cercanías. Tenemos acciones que nos llevan a cada una de las paradas. Cada acción tiene asociado un coste en tiempo

Estado: Localización L $\in$ {Casa, Universidad, Estación, Cercanías }

Estado inicial: L = Casa.

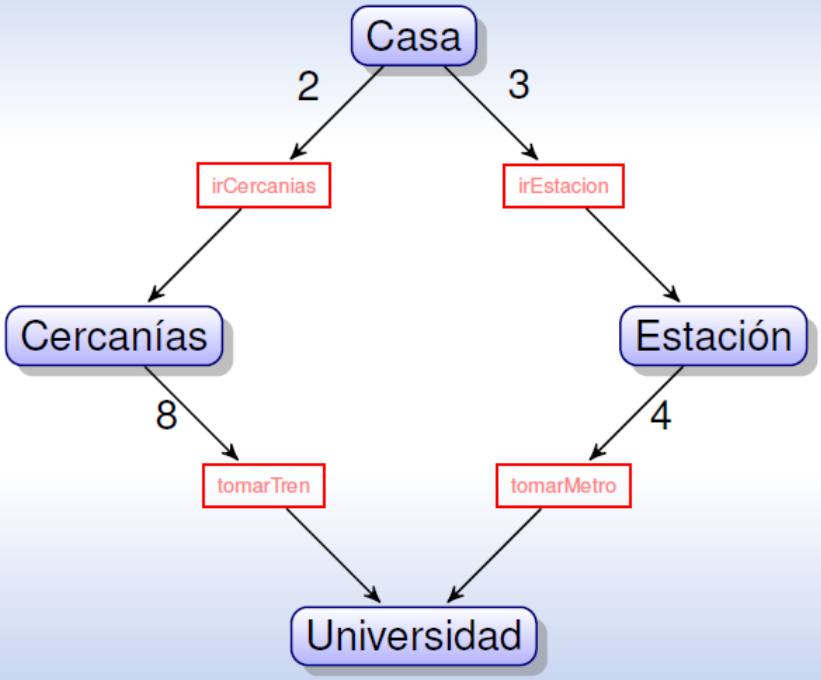
Meta: L=Universidad

Acciones:

- irCercanias: Aplicable en Casa, lleva a Cercanías,c(irCercanias) = 2.
- irEstacion: Aplicable en Casa, lleva a Estación, c(irEstacion) = 3.
- tomarTren: Aplicable en Cercanías, lleva a Universidad, c(tomarTren) = 8.
- tomarMetro: Aplicable en Estación, lleva a Universidad, c(tomarMetro) = 4.

EJEMPLO TRANSPORTE (DETERMINISTA)

- Podemos modelarlo con un grafo de búsqueda:



EJEMPLO TRANSPORTE (INCERTIDUMBRE)

Como hay huelga de trabajadores del metro , estimo que hay un 30% de posibilidades de encontrar una manifestación en el camino (Obstáculo). En ese caso hay que dar un rodeo que nos lleva 15 minutos.

Estado: Localización L $\in$ {Casa, Universidad, Estación, Obstáculo, Cercanías }

Estado inicial: L = Casa.

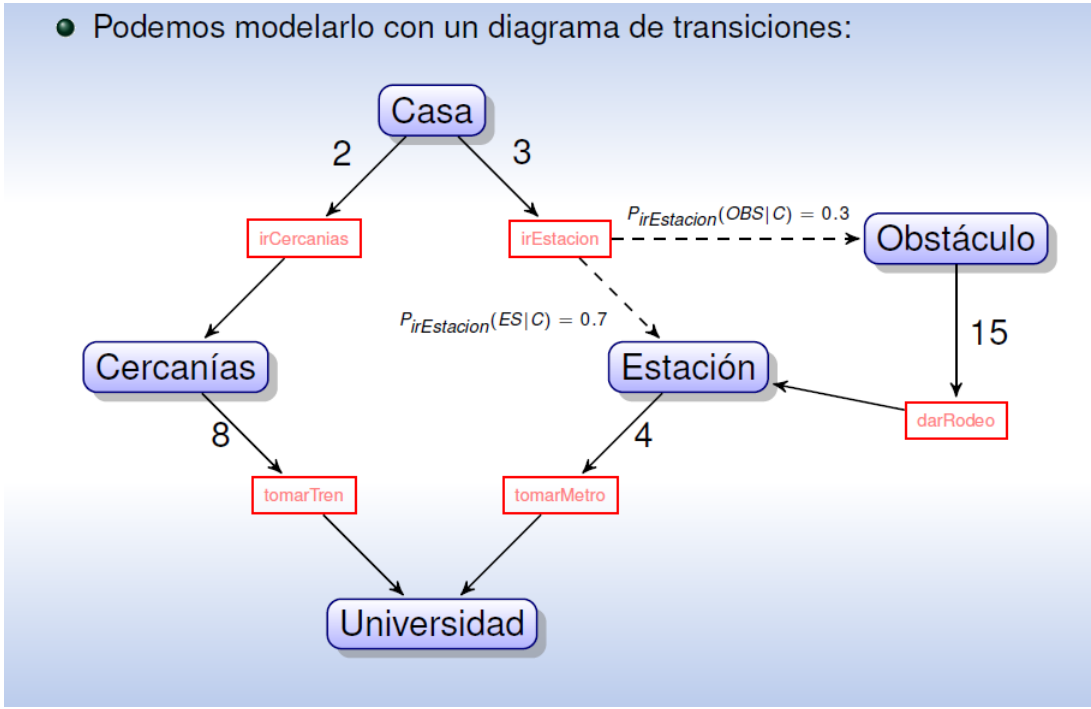
Meta: L=Universidad

Acciones:

- irCercanias: Aplicable en Casa, lleva a Cercanías,c(irCercanias) = 2.
- irEstacion: Aplicable en Casa, lleva a Estación, c(irEstacion) = 3.
- darRodeo: Aplicable en Obstáculo, lleva a Estación, c(darRodeo) = 15.
- tomarTren: Aplicable en Cercanías, lleva a Universidad, c(tomarTren) = 8.
- tomarMetro: Aplicable en Estación, lleva a Universidad, c(tomarMetro) = 4.

EJEMPLO TRANSPORTE (INCERTIDUMBRE)

- Podemos modelarlo con un diagrama de transiciones:



POLÍTICA EN UN MDP

Política π: función que toma un estado s y retorna una acción $a = \pi(s)$, que se ejecuta en el estado s.

Si la política π es completa, el agente siempre sabrá que hacer, sin importar que efecto producen las acciones.

Para definir la política óptima calcularemos unos valores para cada estado y los usaremos para escoger la mejor acción.

VALOR DE UN ESTADO

V(s): valor de s es el coste que esperamos pagar para llegar desde s a la meta si tomamos la mejor decisión posible en cada estado

En Búsqueda (determinista):

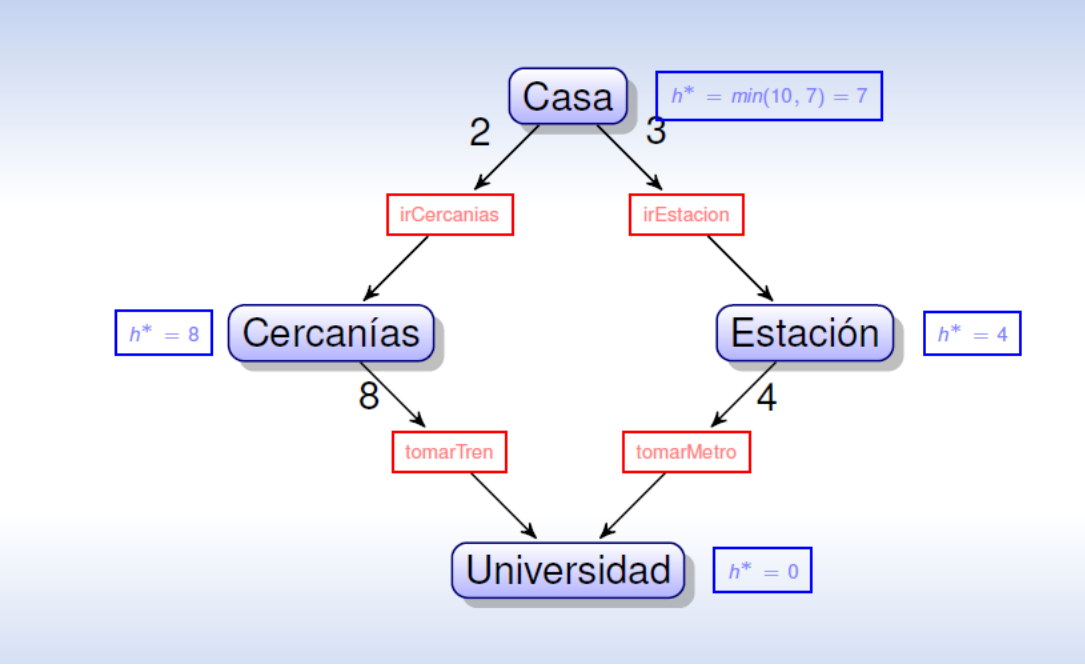
- Búsqueda con Dikjstra o A* encuentran la mejor solución posible.
- También el algoritmo de Escalada, si pudiésemos usar como heurística la heurística perfecta h*(s): coste real del mejor camino desde s hasta la meta

En este caso V(s) = h*(s).

En MDP:

- V(s) = coste esperado de la estrategia óptima para ir desde s a la meta.
- Si tengo V(s), podría calcular política óptima π^*.

DECISIONES ÓPTIMAS EN EL CASO DETERMINISTA



RESOLUCIÓN DE UN MDP

Resolver un MDP: Encontrar la Política Óptima para llegar del estado inicial s' a un estado meta sg.

Política óptima π^*: La de menor coste esperado.

Si aplicar la acción a1 lleva siempre a s', el coste es:

$c(a1) + costeDesde(s')$

Si aplicar la acción a1 tiene efecto probabilístico, el coste esperado es:

$c(a1) + \Sigma P_a(s'/s) \times costeDesde(s')$

Llamamos Valor de un estado s a V(s) = costeDesde(s) y lo podemos ver como un análogo de h* para el caso no determinista

RESOLUCIÓN DE UN MDP

Por lo tanto, encontrar la política óptima pasa por dos fases:

1 Calcular V(s) , valor de cada uno de los estados

2 La política óptima en cada estado s consiste en escoger la acción con mínimo coste esperado.

Para cada estado se calcula Q(s; a): el coste esperado si se toma la acción a en el estado s. Depende de los valores que tengan los estados sucesores de s , s'.

$$Q(s; a) = c(a) + \Sigma P_a(s'/s) \times V(s')$$

El valor de s será el mínimo para las acciones posibles en s (A(s)):

$$V(s) = \min\{Q(s; a)\}$$

La Política Óptima es la que escoge siempre precisamente esa a que minimiza Q(s; a):

$$\pi^*(s) = \text{argmin } \{Q(s; a)\}$$

ECUACIÓN DE BELLMAN PARA MDPS CON COSTE

Para MDPs

El coste esperado de aplicar la acción a, y actuar perfectamente desde allí es

$$Q(s; a) = c(a) + \Sigma\, P_a(s'/s) \times V(s')$$

Así, para MDPs -> Ecuación de Bellman

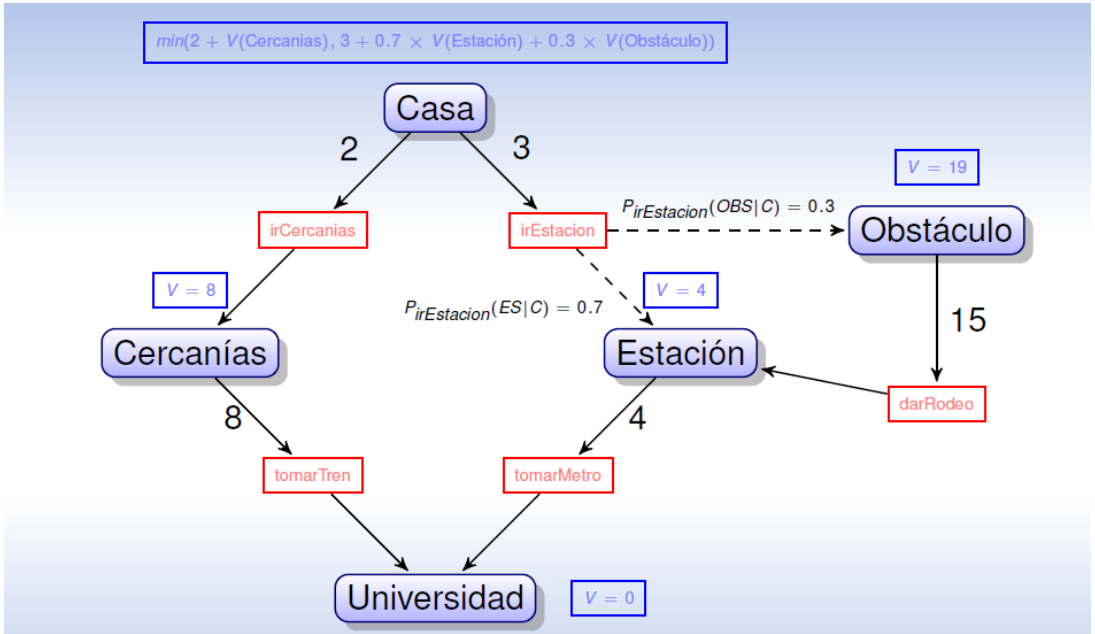
$V(s) = 0$ si s es una meta.

Si no,

$$V(s) = \min_a [c(a) + \Sigma\, P_a(s'/s) \times V(s')]$$

NOTA: Obsérvese que V(s) es un número (min), y $\pi^*(s)$ es una acción (argmin)

EJEMPLO DEL TRANSPORTE, VALORES



- Pero cuando las acciones tienen más de un resultado hay que calcular el coste esperado

ALGORITMO DE ITERACIÓN DE VALORES

Problema: la ecuación de Bellman no es lineal. Se resuelve mediante una técnica de programación dinámica.

Algoritmo: actualizar V(s) sobre todos los estados s, hasta que V(s) CASI no varíe con las iteraciones (punto fijo).

Algoritmo de Iteración de Valores.

1 Inicialmente $V(s) = 0$ para todos los estados.

2 Para cada posible estado $s \in S$, distinto de la meta:

$V(s) := \min_a [c(a) + \Sigma P_a(s'/s) \times V(s')]$

3 Si cambió algún valor V(s'), para algún estado s', volver a (2).

4 Si no, terminar.

Iteración de valores termina cuando alcanza punto fijo

CORRECCIÓN DEL ALGORITMO DE ITERACIÓN DE VALORES

Las ecuaciones de Bellman con costes, y el algoritmo de iteración de valores básicos son versiones simplificadas.

Estas son las condiciones en que dicho algoritmo converge:

- Costes de acción > 0
- Existe una meta.
- La meta es alcanzable desde cualquier estado. Es decir, que desde cada estado existe una secuencia de acciones que llega a la meta con probabilidad > 0.

PROCESO DE DECISIÓN DE MARKOV PARCIALMENTE OBSERVABLES

Proceso de Markov + observabilidad parcial + acciones =

HMM + acciones =

MDP + observabilidad parcial =

POMDP