

TEMA 4. ANÁLISIS DE DECISIONES

PLANIFICACIÓN DE PROYECTOS Y ANÁLISIS DE RIESGOS

Grado en Ingeniería Informática. Mención Tecnologías de la Información

Dpto. Lenguajes y Ciencias de la Computación

UNIVERSIDAD DE MÁLAGA

4. Análisis de decisiones

- Toma de decisiones sin experimentación
- Toma de decisiones con experimentación
- Árboles de decisión
- Teoría de decisión markoviana

LECTURA RECOMENDADA:

Hillier , Lieberman. *Introduction to Operation Research*. Cap. 15

Taha. *Investigación de Operaciones*. Cap. 19

Análisis de decisiones

- Estudio de decisiones que deben tomarse en un ambiente de gran incertidumbre
- ¿Es el momento adecuado para tomar la decisión?
- ¿Es preferible realizar antes alguna prueba?
- EJEMPLO: Goferbroke Co

Estado del Terreno	Pago (miles €)	
Alternativa	Petróleo	Seco
Perforar	700	-100
Vender	90	90
Probabilidad de estado	1/4	3/4

Antes de decidir puede llevar a cabo una exploración sísmológica para obtener una mejor estimación de la probabilidad de encontrar petróleo.

La compañía opera con poco capital. La pérdida de 100000 euros sería grave

Toma de decisiones sin experimentación

- El tomador de decisiones debe elegir una opción entre un conjunto de varias acciones posibles (alternativas factibles)
- El resultado se verá afectado por factores aleatorios fuera del control del tomador de decisiones (estados de la naturaleza)
- Para cada combinación alternativa factible / estado de la naturaleza el tomador de decisiones sabe cuál sería el pago resultante: medida cuantitativa del valor de las consecuencias del resultado. Puede ser la ganancia monetaria neta (utilidad), un valor esperado en el sentido estadístico... Se usa una tabla de pagos

Toma de decisiones sin experimentación

El análisis de decisiones puede verse como un juego de dos jugadores con ganancia 0.

Jugadores: tomador de decisiones y naturaleza

Las opciones alternativas y los estados de la naturaleza son las estrategias disponibles para cada jugador

1. El tomador de decisiones debe elegir una de las opciones posibles
2. La naturaleza “elige” entre los estados posibles
3. Cada combinación resulta en el pago reflejado en la tabla de pagos
4. Esta tabla debe usarse para encontrar una opción óptima para el tomador de decisiones según un criterio adecuado

En la teoría de juegos se supone ambos jugadores racionales y eligiendo estrategias para su propio beneficio. Este criterio no puede aplicarse a la naturaleza: su decisión es aleatoria: probabilidades a priori

Toma de decisiones sin experimentación

CRITERIOS PARA ENCONTRAR LA OPCIÓN ÓPTIMA

1. Pago máximo-mínimo: Para cada opción posible se busca el pago mínimo sobre todos los estados de la naturaleza posibles. Después se encuentra el máximo de estos mínimos y se elige esa opción. TOMADOR MUY PRECAVIDO: Se pone en el peor de los casos posibles
2. Máxima probabilidad: Se identifica el estado más probable de la naturaleza. Para este estado se encuentra la opción con máximo pago y se elige esta opción. DESVENTAJA: ignora mucha información relevante
3. Regla de decisión de Bayes: se utilizan la mejores estimaciones disponibles de las probabilidades de los estados de la naturaleza para calcular el valor esperado de pago para cada opción. Se elige la opción de mayor pago esperado. VENTAJA: incorpora toda la información disponible. INCONVENIENTE: estimaciones subjetivas, poco fiables

Toma de decisiones sin experimentación

ANÁLISIS DE SENSIBILIDAD

p =probabilidad de encontrar petróleo

Pago esperado opción 1:

$$700p - 100(1-p) = 800p - 100$$

Pago esperado opción 2:

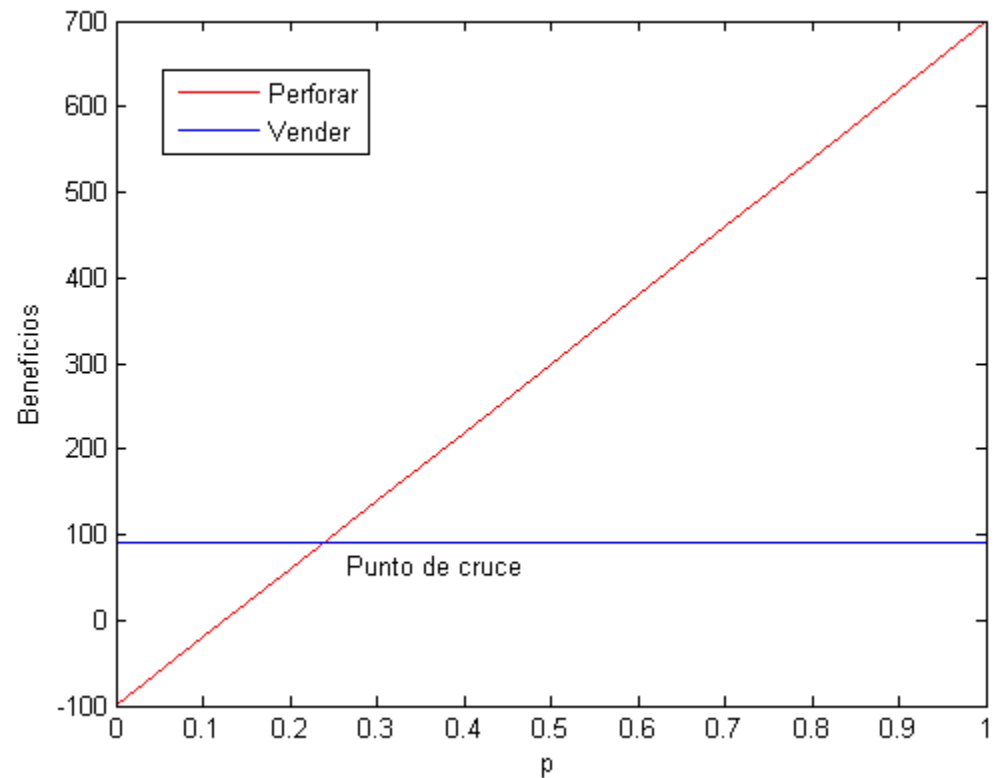
$$90$$

Punto de cruce (crossover point):

$$800p - 100 = 90,$$

$$p = 19/80$$

La mejor opción es perforar si $p > 19/80$



Toma de decisiones con experimentación

Probabilidades a posteriori

$N = n^{\circ}$ posible de estados de la naturaleza

$P(\text{Estado}=i)$ = probabilidad a priori de que el estado de la naturaleza sea i ,
 $i=1,2,\dots,N$

Resultado = j resultado de la experimentación (variable aleatoria)

j = valor posible del resultado

$P(\text{Estado}=i \mid \text{Resultado}=j)$ = probabilidad a posteriori de que el estado de la naturaleza sea i dado que el resultado de la experimentación sea j

Dados $P(\text{Estado}=i)$ y $P(\text{Resultado}=j \mid \text{Estado} = i)$ para $i=1,2,\dots,N$ ¿Cuál es el valor de $P(\text{Estado}=i \mid \text{Resultado} =j)$

Toma de decisiones con experimentación

EJEMPLO: Goferbroke Co

Probabilidades a priori

Estados de la naturaleza: N= 2: SECO, PETRÓLEO

$P(\text{Estado}=\text{SECO})= 3/4$

$P(\text{Estado}=\text{PETRÓLEO})= 1/4$

ESTUDIO PARA MEJORAR LAS ESTIMACIONES: COSTE 30000 EUROS

Resultado= FAVORABLE, DESFAVORABLE

PROBABILIDADES RESULTADO/ESTADO	ESTADOS	
RESULTADO	PETRÓLEO	SECO
FAVORABLE	0.6	0.2
DESFAVORABLE	0.4	0.8

¿Probabilidades a posteriori? ÁRBOL DE PROBABILIDADES

Toma de decisiones con experimentación

TEOREMA DE LA PROBABILIDAD TOTAL

$$P(R = j) = \sum_{k=1}^N P(R = j|E = k)P(E = k)$$

TEOREMA DE BAYES

$$P(E = i|R = j) = \frac{P(E = i, R = j)}{P(R = j)} = \frac{P(R = j|E = i)P(E = i)}{\sum_{k=1}^N P(R = j|E = k)P(E = k)}$$

Toma de decisiones con experimentación

EJEMPLO: Goferbroke Co

PROBABILIDADES CONOCIDAS

$$P(E=S) = \frac{3}{4}; \quad P(R=F | E=S) = 0.2; \quad P(R=D | E=S) = 0.8$$

$$P(E=P) = \frac{1}{4}; \quad P(R=F | E=P) = 0.6; \quad P(R=D | E=P) = 0.4$$

PROBABILIDADES CONJUNTAS

$$P(R=F, E=S) = 0.2 * (3/4) = 0.15; \quad P(R=D, E=S) = 0.8 * (3/4) = 0.6$$

$$P(R=F, E=P) = 0.6 * (1/4) = 0.15; \quad P(R=D, E=P) = 0.4 * (1/4) = 0.1$$

PROBABILIDADES TOTALES

$$P(R=F) = 0.15 + 0.15 = 0.3$$

$$P(R=D) = 0.6 + 0.1 = 0.7$$

PROBABILIDADES A POSTERIORI

$$P(E=S | R=F) = 0.15 / 0.3 = 0.5; \quad P(E=S | R=D) = 0.6 / 0.7 = 6/7$$

$$P(E=P | R=F) = 0.15 / 0.3 = 0.5; \quad P(E=P | R=D) = 0.1 / 0.7 = 1/7$$

Toma de decisiones con experimentación

EL VALOR DE LA EXPERIMENTACIÓN

Antes de realizar cualquier experimento debe determinarse su valor potencial:

1. Valor esperado de la información perfecta

Supuesto que el experimento determina el estado verdadero de la naturaleza se elige la opción con pago máximo para ese estado. Se calcula el pago esperado ponderando cada estado con la probabilidad a priori del estado

$VEIP = \text{pago esperado de la información perfecta} - \text{pago esperado sin experimentación}$

2. Valor esperado de la experimentación

Se calcula el pago esperado de la información perfecta para cada resultado posible. Cada pago se pondera con la probabilidad del resultado y se resta el coste de la experimentación

Toma de decisiones con experimentación

EJEMPLO: Goferbroke Co

TABLA DE PAGOS CON LA INFORMACIÓN PERFECTA

Estado del Terreno	Pago (miles €)	
Alternativa	Petróleo	Seco
Perforar	700	-100
Vender	90	90
Pago máximo	700	90
Probabilidad de estado	1/4	3/4
Pago esperado información perfecta	$700/4 + 270/4 = 242.5$	

$$VEIP = 242.5 - 100 = 142.5$$

Toma de decisiones con experimentación

EJEMPLO: Goferbroke Co

PAGOS ESPERADOS CONDICIONADOS AL RESULTADO (en miles de euros)

PAGO SI RESULTADO DESFAVORABLE: $90 \cdot (1/7) + 90 \cdot (6/7) = 90$

PAGO SI RESULTADO FAVORABLE: $700 \cdot (1/2) - 100 \cdot (1/2) = 300$

PAGO ESPERADO: $90 \cdot 0.7 + 300 \cdot 0.3 = 153$

VEE: VALOR ESPERADO DE LA EXPERIMENTACIÓN: $153 - 100 = 53$

Árboles de decisiones

Permiten visualizar el problema y organizar el trabajo computacional antes descrito

Son especialmente útiles cuando hay que tomar varias decisiones en secuencias

□ Indican que en ese punto hay que tomar una decisión

○ Indican que en ese punto ocurre un evento aleatorio

En cada rama se introducen los números:

- Si no están entre paréntesis indican un flujo de dinero (ganancias o pérdidas)
- Si están entre paréntesis indica la probabilidad de que suceda ese evento
- Se suman los flujos de dinero en cada rama hacia la derecha y se escribe el valor total

Árboles de decisiones

Una vez construido el árbol se analiza el problema:

1. Se empieza por la derecha y se mueve hacia la izquierda una columna cada vez
2. En los nodos de evento se calcula el pago esperado multiplicando los pagos esperados de cada rama por su probabilidad y sumando los productos. El resultado se escribe sobre el nodo y designa el pago esperado por ese nodo
3. En los nodos de decisión se comparan los pagos esperados en cada rama y se elige la alternativa más prometedora. En la otra elección se colocan unas barras verticales como barrera para rechazar la rama

Una vez completo el árbol, el tomador de decisiones se mueve de izquierda a derecha. La regla de decisión de Bayes dice que se debe seguir sólo las ramas abiertas para lograr el mayor beneficio esperado

Proceso de decisión markoviano

PROCESOS ESTOCÁSTICOS

- Un proceso estocástico de tiempo discreto es una descripción de la relación entre las variables aleatorias X_0, X_1, \dots que representan alguna característica de un sistema en puntos discretos en el tiempo.
 - Ejemplo: ruina del jugador: inicialmente tengo 2€, en los tiempos 1, 2, ... participo en un juego en el que apuesto 1€ que gano con probabilidad p y pierdo con probabilidad $1-p$. Dejo de jugar cuando mi capital es 4€ o he perdido todo mi capital. Si X_i es la cantidad de dinero que tengo en el tiempo i , X_0, X_1, \dots es un proceso estocástico.
- Un proceso estocástico de tiempo continuo es un proceso estocástico en el que el estado del tiempo se puede examinar en cualquier momento.
 - Ejemplo: número de personas en un supermercado a los t minutos de abrir

Proceso de decisión markoviano

CADENAS DE MARKOV

- Cadena de Markov: proceso estocástico de tiempo discreto que para $t=0,1,2,\dots$ y todos los estados verifica

$$P(X_{t+1}=i_{t+1} \mid X_t=i_t, X_{t-1}=i_{t-1}, \dots, X_1=i_1, X_0=i_0)=P(X_{t+1}=i_{t+1} \mid X_t=i_t)$$

- Hipótesis de estabilidad: $P(X_{t+1}=j \mid X_t=i)=p_{ij}$ (no depende de t)

- Probabilidades de transición: p_{ij}

- Matriz de probabilidades de transición:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix}$$

- Se debe verificar: $\sum_{j=1}^s p_{ij} = 1$

- Las cadenas de Markov que cumplen la hipótesis de estabilidad se llaman cadenas estacionarias de Markov.
- Distribución inicial de probabilidad de una cadena de Markov: $q=[q_1,\dots,q_s]$ donde $q_i=P(X_0=i)$

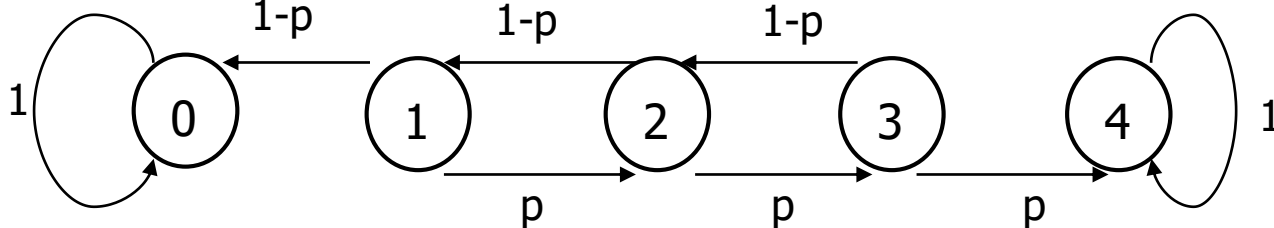
Proceso de decisión markoviano

Ejemplo: la ruina del jugador es una cadena de Markov estacionaria

- Estados: 0, 1, 2, 3, 4
- Matriz de transición

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- La matriz de transición se puede representar con un grafo en el que cada nodo representa un estado y cada arco la probabilidad de transición entre estados.



Proceso de decisión markoviano

PROBABILIDADES DESPUÉS DE N PASOS

- Si una cadena de Markov estacionaria está en el estado i en el tiempo m , ¿cuál es la probabilidad de que n períodos después la cadena esté en el estado j ?

$$P(X_{m+n}=j | X_m=i) = P(X_n=j | X_0=i) = P_{ij}(n)$$

- $P_{ij}(n)$ es la probabilidad en la etapa n de una transición del estado i al estado j

$$P_{ij}(1) = p_{ij}, \quad P_{ij}(2) = \sum_{k=1}^s p_{ik} p_{kj}, \quad P_{ij}(n) = \text{elemento } ij\text{-ésimo de } P^n$$

- Probabilidad de estar en el estado j en el tiempo $n = \sum_{i=1}^s q_i P_{ij}(n)$

Proceso de decisión markoviano

PROBABILIDADES EN ESTADO ESTACIONARIO

- Si P es la matriz de transición de una cadena ergódica de s estados entonces existe un vector $\pi = [\pi_1 \ \pi_2 \ \dots \ \pi_s]$ tal que

$$\lim_{n \rightarrow \infty} P^n = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_s \\ \pi_1 & \pi_2 & \cdots & \pi_s \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_s \end{bmatrix}$$

- Es decir, $\lim_{n \rightarrow \infty} P_{ij}(n) = \pi_j$
- A π se le llama distribución de estado estable o de equilibrio para la cadena de Markov

Proceso de decisión markoviano

CLASIFICACIÓN DE ESTADOS EN UNA CADENA DE MARKOV

- Dados dos estados i y j , una trayectoria de i a j es una sucesión de transiciones que comienza en i y termina en j , de forma que cada transición de la secuencia tenga probabilidad positiva.
- Un estado j es alcanzable desde un estado i si hay una trayectoria de i a j .
- Dos estados i y j se comunican si i es alcanzable desde j y j es alcanzable desde i .
- Un conjunto de estados S en una cadena de Markov es cerrado (constituyen una clase de la cadena) si ningún estado fuera de S es alcanzable desde un estado en S .
- Un estado i es absorbente si $p_{ii}=1$

Proceso de decisión markoviano

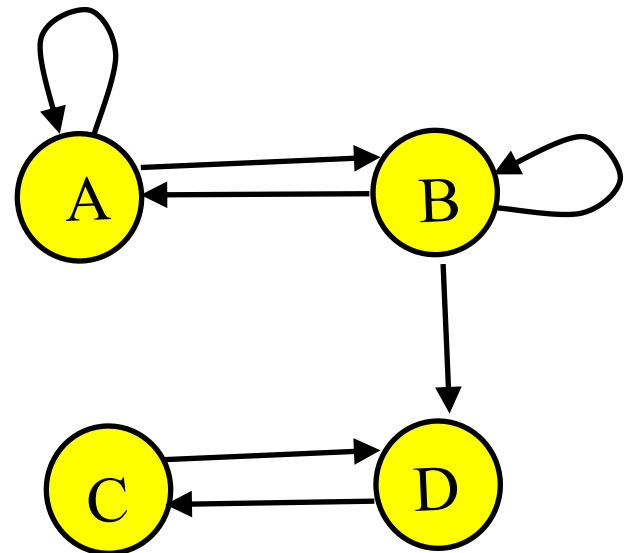
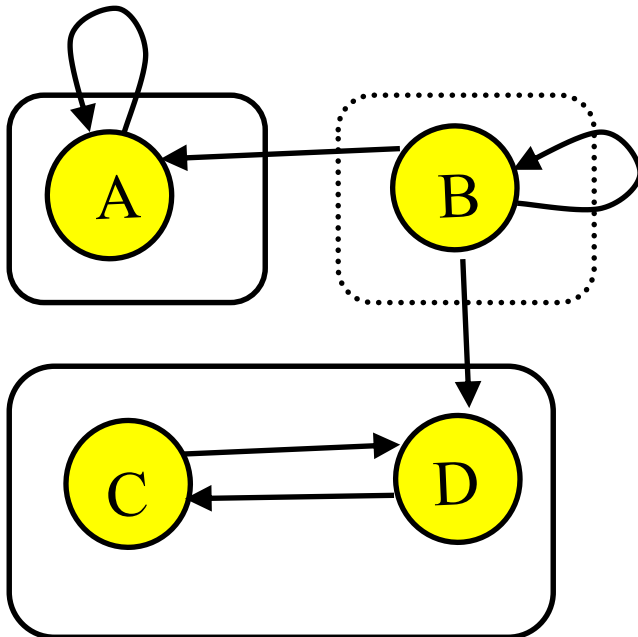
CLASIFICACIÓN DE ESTADOS EN UNA CADENA DE MARKOV

- Un estado i es transitorio si hay un estado j alcanzable desde i , pero el estado i no es alcanzable desde j .
- Un estado es recurrente si no es transitorio.
- Un estado i es periódico con periodo $k > 1$ si k es el menor número tal que todas las trayectorias que parten del estado i y regresan al estado i tienen una longitud múltiplo de k .
- Si un estado recurrente no es periódico es aperiódico.
- Si todos los estados de una cadena son recurrentes, aperiódicos y se comunican entre sí, la cadena es ergódica.

Proceso de decisión markoviano

CLASIFICACIÓN DE ESTADOS EN UNA CADENA DE MARKOV

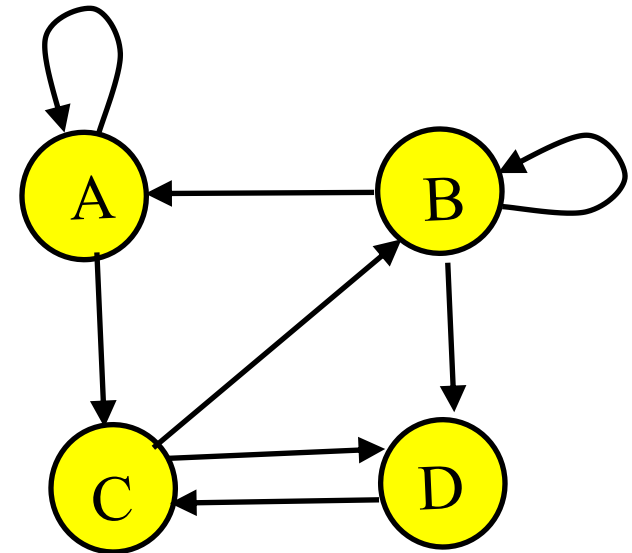
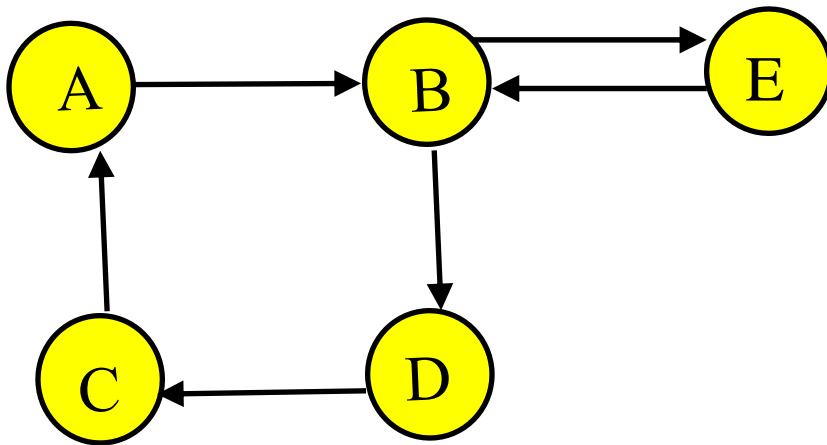
- ¿Recurrentes?
- ¿Transitorios?



Proceso de decisión markoviano

CLASIFICACIÓN DE ESTADOS EN UNA CADENA DE MARKOV

- ¿Periodo de A?
- ¿Cadena periódica?
- ¿Cadena ergódica?



Proceso de decisión markoviano

PROBABILIDADES EN ESTADO ESTACIONARIO

- π se puede determinar a partir de la ecuación: $\pi_j = \sum_{k=1}^s \pi_k p_{kj}$
- En forma matricial $\pi = \pi P$
- Este sistema tiene un número infinito de soluciones porque el rango de P siempre resulta ser menor o igual que s-1
- También se debe verificar:

$$\pi_1 + \pi_2 + \dots + \pi_s = 1$$

- Probabilidad de que una transición determinada deje el estado j = probabilidad de que una transición determinada entre al estado j.

$$\pi_j (1 - p_{jj}) = \sum_{k \neq j} \pi_k p_{kj}$$

Proceso de decisión markoviano

INTERPRETACIÓN INTUITIVA DE LAS PROBABILIDADES DE ESTADO ESTABLE

- Probabilidad de que una transición determinada deje el estado $j = \pi_j (1 - p_{jj})$
- Probabilidad de que una transición determinada entre al estado $j = \sum_{k \neq j} \pi_k p_{kj}$
- En el estado estable el flujo de probabilidad hacia cada estado debe ser igual al flujo de probabilidad que sale de cada estado: probabilidades de equilibrio
- El comportamiento de una cadena de Markov antes de alcanzar el estado estable se llama comportamiento transitorio.
- Para su estudio se utilizan las fórmulas dadas anteriormente para $P_{ij}(n)$.

Proceso de decisión markoviano

- Aplicación de la programación dinámica a un proceso de decisión estocástico
- Las probabilidades de transición entre estado están descritas por una cadena de Markov.
- La estructura de recompensas del proceso está descrita por una matriz cuyos elementos individuales son el coste o el beneficio de moverse de un estado a otro.
- Las matrices de transición y de recompensas dependen de las alternativas de decisión.
- Objetivo: determinar la política óptima que maximice el ingreso esperado en un número finito o infinito de etapas.

Proceso de decisión markoviano

PROBLEMA DEL JARDINERO (TAHA)

- ESTADOS:

1 BUENO

2 REGULAR

3 MALO

- DECISIONES:

1 NO ABONAR

$$P_1 = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix} \quad R_1 = \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

2 ABONAR

$$P_2 = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{bmatrix} \quad R_2 = \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix}$$

Proceso de decisión markoviano

MODELO DE ETAPAS FINITAS

- Objetivo: optimizar ingreso esperado al final de un período de tamaño N
- $P^k=[p_{ij}^k]$ y $R^k=[r_{ij}^k]$ matrices de transición y recompensa para la alternativa k
- $f_n(i)$ = ingreso esperado óptimo de las etapas $n, n+1, \dots, N$ si el estado del sistema al inicio de la etapa n es i

$$f_n(i) = \max_k \left\{ \sum_{j=1}^m p_{ij}^k [r_{ij}^k + f_{n+1}(j)] \right\}, \quad n = 1, 2, \dots, N,$$
$$f_{N+1}(j) = 0, \quad j = 1, 2, \dots, m$$

Proceso de decisión markoviano

MODELO DE ETAPAS FINITAS

- N = numero total de etapas
- $f_n(i)$ = ingreso esperado óptimo de las etapas $n, n+1, \dots, N$

i	v_i^1	v_i^2
1	5.3	4.7
2	3	3.1
3	-1	0.4

$$v_i^k = \sum_{j=1}^m p_{ij}^k r_{ij}^k$$

$$f_n(i) = \max_k \left\{ \sum_{j=1}^m p_{ij}^k [r_{ij}^k + f_{n+1}(j)] \right\}, n = 1, 2, \dots, N$$

- Etapa 3: $f_3(i)$

i	v_i^1	v_i^2	$f_3(i)$	k
1	5.3	4.7	5.3	1
2	3	3.1	3.1	2
3	-1	0.4	0.4	2

Proceso de decisión markoviano

MODELO DE ETAPAS FINITAS

Etapa 2			
i	K=1	K=2	f ₂ (i) k
1	8.03	8.19	8.19 2
2	4.75	5.61	5.61 2
3	-0.6	2.13	2.13 2

Etapa 1			
I	K=1	K=2	f ₁ (i) k
1	10.38	10.74	10.74 2
2	6.87	7.92	7.92 2
3	1.13	4.23	4.23 2

$$f_n(i) = \max_k \left\{ \sum_{j=1}^m p_{ij}^k [r_{ij}^k + f_{n+1}(j)] \right\}, n = 1, 2, \dots, N$$

Proceso de decisión markoviano

MODELO DE ETAPAS INFINITAS

- Nos interesan políticas para las que existan soluciones de estado estable
- Métodos:
 - Enumeración exhaustiva: se evalúan todas las políticas estacionarias posibles del problema de decisión
 - Iteración de política: determina la política óptima de forma iterativa

Proceso de decisión markoviano

ENUMERACIÓN EXHAUSTIVA

- Problema de decisión con S políticas estacionarias
- Pasos del método
 - Calcular el ingreso de una etapa esperado de la política s dado el estado i, $i=1,2,\dots,m$:

$$v_i^s = \sum_{j=1}^m p_{ij}^s r_{ij}^s$$

- Calcular las probabilidades estacionarias de largo plazo de la matriz de transición asociada a la política s
- Determinar el ingreso esperado de la política s por paso de transición:

$$E^s = \sum_{i=1}^m \pi_i^s v_i^s$$

- La política óptima s^* se determina de forma que $E^{s^*} = \max_s \{E^s\}$

Proceso de decisión markoviano

MODELO DE ETAPAS INFINITAS

- POLITICAS ESTACIONARIAS:
 - I. NO ABONAR NUNCA
 - II. ABONAR SIEMPRE
 - III. ABONAR SI 3, NO ABONAR SI 1 Ó 2
 - IV. ABONAR SI 2, NO ABONAR SI 1 Ó 3
 - V. ABONAR SI 1, NO ABONAR SI 2 Ó 3
 - VI. ABONAR SI 1 Ó 2, NO ABONAR SI 3
 - VII. ABONAR SI 1 Ó 3, NO ABONAR SI 2
 - VIII. ABONAR SI 2 Ó 3, NO ABONAR SI 1

Proceso de decisión markoviano

ITERACIÓN DE POLÍTICAS

- Problema de decisión con S políticas estacionarias
- Para una política específica:
 - Rendimiento total esperado en la etapa n :

$$f_n(i) = v_i + \sum_{j=1}^m p_{ij} f_{n+1}(j), \quad i = 1, 2, \dots, m$$

- η : Número de etapas que faltan por considerar

$$f_\eta(i) = v_i + \sum_{j=1}^m p_{ij} f_{\eta-1}(j), \quad i = 1, 2, \dots, m$$

Proceso de decisión markoviano

ITERACIÓN DE POLÍTICAS

- Ingreso esperado por etapa: $E = \pi_1 v_1 + \pi_2 v_2 + \dots + \pi_m v_m$
- Para η grande $f_\eta(i) = \eta E + f(i)$ donde $f(i)$ es un término constante que representa el efecto sobre el ingreso de comenzar en el estado i .
- Sustituyendo en la ecuación recursiva y simplificando

$$E = v_i + \sum_{j=1}^m p_{ij} f(j) - f(i), \quad i = 1, 2, \dots, m$$

que es un sistema de m ecuaciones y $m+1$ incógnitas: $E, f(1), \dots, f(m)$.

Proceso de decisión markoviano

ITERACIÓN DE POLÍTICAS

- Para determinar el valor máximo de E se sigue un proceso iterativo que termina cuando dos políticas sucesivas son idénticas:

- Paso de determinación del valor: se elige una política arbitraria s. Suponiendo $f_s(m)=0$ se resuelven las ecuaciones:

$$E^s = v_i^s + \sum_{j=1}^m p_{ij}^s f^s(j) - f^s(i), \quad i = 1, 2, \dots, m$$

- Paso de mejoramiento de política: Para cada estado i determina la política k que produce

$$\max_k \left\{ v_i^k + \sum_{j=1}^m p_{ij}^k f^s(j) \right\}, \quad i = 1, 2, \dots, m$$

- Las decisiones óptimas que resultan para los estados 1,2,...,m constituyen la nueva política t. Si s y t son idénticas, t es óptima. Si no es así, se repite el proceso con s=t.

Proceso de decisión markoviano

ITERACIÓN DE POLÍTICAS

- Política inicial I: no abonar nunca

- Iteración del valor

$$E = 5.3 + 0.2f(1) + 0.5f(2) + 0.3f(3) - f(1)$$

$$E = 3 + 0.5(f(2) + 0.5f(3)) - f(2)$$

$$E = -1 + f(3) - f(3)$$

$$f(3) = 0$$

$$\rightarrow \begin{aligned} f(1) &= 12.88 \\ f(2) &= 8 \\ f(3) &= 0 \end{aligned}$$

$$P_1 = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix} \quad R_1 = \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

- Mejoramiento de política

	$v_i^k + \sum_{j=1}^3 p_{ij}^k f(j)$			
I	K=1	K=2	max	k
1	11.875	13.36	13.36	2
2	7	9.19	9.19	2
3	-1	4.24	4.24	2

Proceso de decisión markoviano

ITERACIÓN DE POLÍTICAS

- Nueva politica II: abonar siempre

$$P_2 = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{bmatrix} \quad R_2 = \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix}$$

- Iteración del valor

$$E = 4.7 + 0.3f(1) + 0.6f(2) + 0.1f(3) - f(1)$$

$$E = 3.1 + 0.1f(1) + 0.6f(2) + 0.3f(3) - f(2)$$

$$E = 0.4 + 0.05f(1) + 0.4f(2) + 0.55f(3) - f(3)$$

$$f(3) = 0$$

→

$$f(1) = 6.75$$

$$f(2) = 3.79$$

$$f(3) = 0$$

- Mejoramiento de política

	$v_i^k + \sum_{j=1}^3 p_{ij}^k f(j)$			
I	K=1	K=2	max	k
1	8.54	8.99	8.99	2
2	4.89	6.05	6.05	2
3	-1	2.25	2.25	2