Universidad de Carabobo
Facultad Experimental de Ciencias y Tecnología
Departamento de Matemáticas
Investigación de Operaciones IV

Procesos de Decisión Markovianos

Estefania
Profesora Nathylyn Mendoza

Concepto

En lugar de describir cómo evoluciona un proceso como lo hacíamos anteriormente con las cadenas de Markov, al estudiar los procesos de decisión markovianos nos enfocaremos en el diseño del proceso para optimizar el desempeño de una cadena de Markov de tiempo discreto.

Para cada estado posible, tendremos que decidir cuál será la transición del proceso, el próximo estado. Al hacer esta decisión, buscaremos optimizar el costo inmediato de la transición y los costos subsecuentes de realizar las próximas transiciones. También se afectarán las probabilidades de transición futuras del proceso. La meta es optimizar este costo para tomar las decisiones óptimas durante el proceso. A esto lo llamamos un proceso de decisión markoviano.

Descripción detallada

El proceso de decisión markoviano se puede describir de la siguiente forma:

- Observamos el estado de la cadena de Markov de tiempo discreto **después** de cada transición. Podemos denotar el estado con la letra i.
- Después de cada observación (de realizar la transición), se debe tomar otra decisión para la próxima transición del proceso. Esta decisión k se toma de un conjunto de posibles decisiones (k = 1, 2, 3, …, K). Dependiendo de la estructura de la cadena de Markov y de las posibles transiciones, cada estado puede tener un conjunto distinto de decisiones relevantes.
- Si se toma la decisión $d_i = k$ estando en el estado i, se tiene un costo inmediato por realizar la transición denotado por C_{ik}.
- Dicha decisión va a determinar las probabilidades de transición para la siguiente transición. Estas probabilidades de transición desde el estado i al tomar la decisión k se denotan por $p_{ij}(k)$.

Política para el proceso de decisión markoviano

Una política para el proceso de decisión markoviano es una especificación de las decisiones que tomarían para cada estado específico. Es decir, la transición que realizaría el proceso de decisión markoviano si la cadena de Markov se encuentra en un estado específico. Por ejemplo: $(d_0, d_1, d_2, ..., d_M)$ para M posibles estados.

Objetivo

El objetivo es encontrar una política óptima para el proceso en cuanto a los costos (o beneficios) inmediatos y subsecuentes del proceso. Por ejemplo, minimizar el costo promedio esperado por unidad de tiempo.

Propiedades de las Políticas

- **Política estacionaria:** siempre que el proceso se encuentre en un estado específico, la decisión será la misma (la indicada por la política).
- **Política determinística:** siempre se tomará una decisión específica de acuerdo a la política cuando el proceso esté en un estado específico No será una decisión aleatoria basada en probabilidades.

Ejemplo

“Un fabricante opera una máquina clave en el núcleo de uno de sus procesos. Debido a que se le da un uso pesado la máquina se deteriora con rapidez, lo que afecta tanto la calidad como la cantidad de producción que ella genera. Por lo tanto, al final de cada semana se realiza una inspección exhaustiva cuyo resultado es la clasificación de las condiciones de la máquina en uno de cuatro estados posibles:”

Estado	Condición
0	Tan buena como nueva
1	Operable: deterioro menor
2	Operable: deterioro mayor
3	Inoperable: producción de calidad inaceptable

Se recolectan datos sobre el estado de la máquina en cada inspección y se genera la siguiente **matriz de transición de un paso** con la **frecuencia relativa** de cada transición del estado representado por cada fila al estado representado por cada columna. Esta frecuencia relativa representa la probabilidad de transición.

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	0	0	0	1

Interpretando la matriz de transición:

- No se puede realizar una transición hacia el estado 0 ya que la máquina nunca volverá a estar tan buena como nueva luego de realizar jornadas de trabajo.
- No se puede volver a un estado inferior, a un estado que represente mejores condiciones de la máquina.
- Una vez que la máquina llega al estado 3, es decir, cuando se vuelve inoperable, la probabilidad de mantenerse en dicho estado es 1. Es un **estado absorbente**. No volverá a un estado más productivo. En su lugar, será reemplazada por una máquina nueva.

Según la descripción del problema, el análisis estadístico realizado mostró que el proceso posee la **propiedad markoviana** de falta de memoria. Es decir, las probabilidades de transición no dependen de estados previos del proceso.

Costos

- La empresa incurre en un costo cuando la máquina produce productos defectuosos. Estos costos se describen en la siguiente tabla:

Estado	Costo esperado debido a artículos defectuosos, $
0	0
1	1 000
2	3 000

- Cuando la máquina llega al estado 3 (inoperable), debe ser reemplazada por una máquina nueva.
    - Para realizar este cambio se requiere 1 semana, durante la cual se tiene una pérdida de $2,000 por la producción perdida.
    - La máquina nueva tiene un costo de $4,000 dólares.

- Por lo tanto, el costo total de llegar al estado 3 es de **$6,000**.

Sucesión de Máquinas

Con la política de sucesión de máquinas igualmente se tiene una cadena de Markov con la siguiente matriz de transición:

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

Costo Promedio Esperado por Unidad de Tiempo (a largo plazo)

Este costo se usa como una medida de desempeño en los procesos de decisión markovianos.

- Primero se calculan las probabilidades de estado estable de la cadena de Markov con el sistema de ecuaciones $\pi P = \pi$ más la condición de que las probabilidades de estado estable deben sumar 1.

Para el ejemplo, se resolvería el siguiente sistema de ecuaciones:

$$\pi_0 = \pi_3,$$

$$\pi_1 = \frac{7}{8}\pi_0 + \frac{3}{4}\pi_1,$$

$$\pi_2 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2,$$

$$\pi_3 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2,$$

$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3.$$

Despejando el sistema para igualar las primera cuatro ecuaciones a 0, se obtiene el siguiente sistema:

$0 \ = \ -\pi_0 + \pi_3$

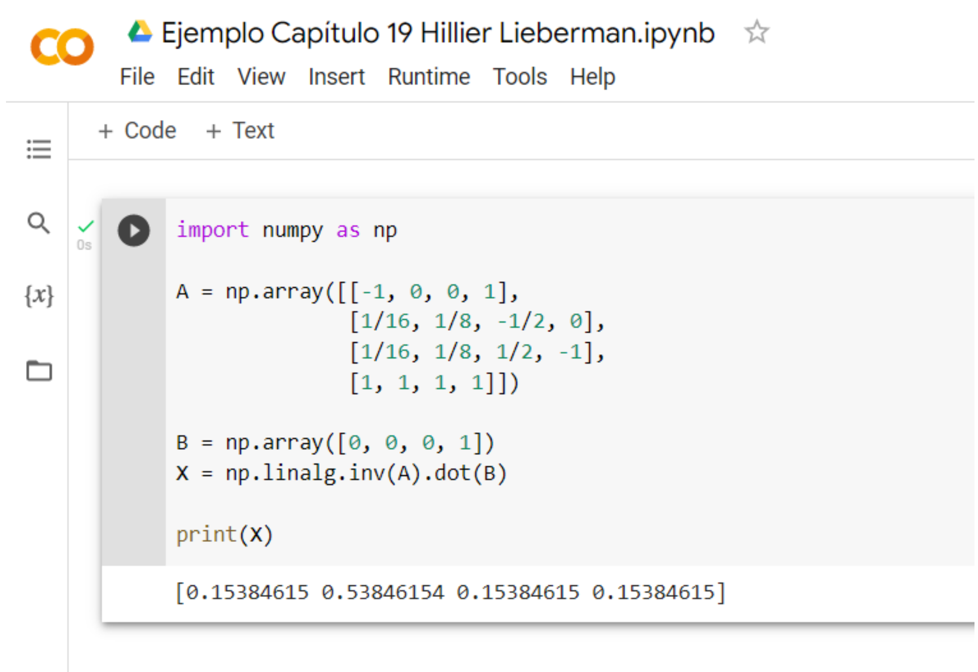
$0 \ = \ 7/8\pi_0 \ - \ 1/4\pi_1$

$0 \ = \ 1/16\pi_0 + \ 1/8\pi_1 - 1/2\pi_2$

$0 \ = \ 1/16\pi_0 + \ 1/8\pi_1 + 1/2\pi_2 - \pi_3$

$1 \ = \ \pi_0 + \ \pi_1 + \pi_2 \ + \pi_3$

El cual puede ser resuelto en Google Colab al remover una ecuación redundante:




```
import numpy as np

A = np.array([[-1, 0, 0, 1],
              [1/16, 1/8, -1/2, 0],
              [1/16, 1/8, 1/2, -1],
              [1, 1, 1, 1]])

B = np.array([0, 0, 0, 1])
X = np.linalg.inv(A).dot(B)

print(X)
```

```
[0.15384615 0.53846154 0.15384615 0.15384615]
```

Los valores corresponden a:

$$\pi_0 = \frac{2}{13}, \qquad \pi_1 = \frac{7}{13}, \qquad \pi_2 = \frac{2}{13}, \qquad \pi_3 = \frac{2}{13}.$$

Costo Promedio

Para calcular el costo promedio estimado.

$$0 \ * \ \pi_1 + \ 1000 \ * \pi_1 \ + \ 3000 \ * \pi_2 \ + \ 6000 \ * \pi_3$$

$$25\,000 / 13$$

$$\$1\,923.08$$

Políticas de Mantenimiento

Las decisiones después de cada inspección son las siguientes:

Decisión	Acción	Estados relevantes
1	No hacer nada	0, 1, 2
2	Reparación general (el sistema regresa al estado 1)	2
3	Reemplazo (el sistema regresa al estado 0)	1, 2, 3

En este tabla se resumen los costos del ejemplo:

Decisión	Estado	Costo esperado por producir artículos defectuosos, $	Costo de mantenimiento, $	Costo (ganancia perdida) por producción perdida, $	Costo total por semana, $
1. No hacer nada	0	0	0	0	0
	1	1 000	0	0	1 000
	2	3 000	0	0	3 000
2. Repar. gral.	2	0	2 000	2 000	4 000
3. Reemplazar	1, 2, 3	0	4 000	2 000	6 000

Solución del ejemplo por enumeración exhaustiva

La siguiente tabla enumera las políticas relevantes para el ejemplo prototipo:

Política	Descripción verbal	$d_0(R)$	$d_1(R)$	$d_2(R)$	$d_3(R)$
R_a	Reemplazo en el estado 3	1	1	1	3
R_b	Reemplazo en el estado 3, reparación general en el estado 2	1	1	2	3
R_c	Reemplazo en los estados 2 y 3	1	1	3	3
R_d	Reemplazo en los estados 1, 2 y 3	1	3	3	3

Nota: los números de esta tabla corresponden a los números asignados a cada decisión en la tabla anterior.

- 1: No hacer nada.
- 2: Reparación general
- 3: Reemplazar

Cada política tiene una matriz de transición específica. Veamos:

Para la política R_a:

- La probabilidad de transición entre el estado 3 (inoperable) y el estado 0 (como nueva) es 1 porque la máquina se reemplaza con una nueva.

	R_a			
Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

Para la política R_b:

- La probabilidad de transición del estado 2 al estado 1 es 1 porque se realiza una reparación general cuando la máquina llega al estado 2.
- La probabilidad de transición entre el estado 3 (inoperable) y el estado 0 (como nueva) es 1 porque la máquina se reemplaza con una nueva.
- La probabilidad de transición del estado 2 al estado 2 y al estado 3 es 0 porque al llegar al estado 2 se realiza una reparación general, así que la máquina no puede continuar en el mismo estado o dañándose.

Estado	R_b			
	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	1	0	0
3	1	0	0	0

Para la política R_c:

- La probabilidad de transición entre el estado 2 y el estado 0 es 1 porque al llegar al estado 2 se reemplaza la máquina por una nueva.
- La probabilidad de transición entre el estado 3 (inoperable) y el estado 0 (tan buena como nueva) es 1 porque la máquina se reemplaza con una nueva.
- La máquina no puede retornar a un estado inferior porque en esta política no se repara la máquina sino que se reemplaza por una nueva.

Estado	R_c			
	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	1	0	0	0
3	1	0	0	0

Para la política R_d:

- La probabilidad de transición de los estados 1, 2 y 3 al estado 0 (tan buena como nueva) es 1 porque la política establece que se realizará un reemplazo de la máquina en los estados 1, 2 y 3.
- La máquina no puede retornar a un estado anterior ya que esta política no contempla realizar mantenimiento sino un reemplazo total de la máquina.

Estado	R_d 0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	1	0	0	0
2	1	0	0	0
3	1	0	0	0

A partir de la última columna de esta tabla:

Decisión	Estado	Costo esperado por producir artículos defectuosos, $	Costo de mantenimiento, $	Costo (ganancia perdida) por producción perdida, $	Costo total por semana, $
1. No hacer nada	0	0	0	0	0
	1	1 000	0	0	1 000
	2	3 000	0	0	3 000
2. Repar. gral.	2	0	2 000	2 000	4 000
3. Reemplazar	1, 2, 3	0	4 000	2 000	6 000

Podemos obtener el costo C_{ik} de tomar una decisión k si el proceso está en un estado i específico:

Estado i \ Decisión k	C_{ik} (en miles de dólares) 1	2	3
0	0	—	—
1	1	—	6
2	3	4	6
3	—	—	6

Para la decisión 1:

- Si el proceso está en el estado 0 y se toma la decisión 1 (No hacer nada), el costo es 0.
- Si el proceso está en el estado 1 y se toma la decisión 1 (No hacer nada), el costo es de $1 000 por fabricar artículos defectuosos.
- Si el proceso está en el estado 2 y se toma la decisión 1 (No hacer nada), el costo es de $3 000 por fabricar artículos defectuosos.

Para la decisión 2:

- Si el proceso está en el estado 2 y se toma la decisión 2 (Reparación General), el costo es de $4 000.

Para la decisión 3:

- Si el proceso está en el estado 1 y se toma la decisión 3 (Reemplazar), el costo es de $6 000.
- Si el proceso está en el estado 2 y se toma la decisión 3 (Reemplazar), el costo es de $6 000.
- Si el proceso está en el estado 3 y se toma la decisión 3 (Reemplazar), el costo es de $6 000.

Para calcular el costo promedio esperado (a largo plazo) por unidad de tiempo, usamos la siguiente fórmula:

$$E(C) = \sum_{i=0}^{M} C_{ik}\pi_i$$

Donde:

- π_i representa la probabilidad de estado estable.
- k representa la decisión tomada según la política implementada y el estado actual del sistema i.

Calcular Costo Promedio Esperado a Largo Plazo para Cada Política

Anteriormente calculamos los valores de π_i para la política R_a:

$$\pi_0 = \frac{2}{13}, \qquad \pi_1 = \frac{7}{13}, \qquad \pi_2 = \frac{2}{13}, \qquad \pi_3 = \frac{2}{13}.$$

Ahora, calcularemos estos valores para las políticas R_b, R_c, R_d con Google Colab.

Para la política R_b:

$$\pi P = \pi$$

más la ecuación $1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$

Sistema de ecuaciones:

$\pi_0 = \pi_3$

$\pi_1 = 7/8\pi_0 + 3/4\pi_1 + \pi_2$

$\pi_2 = 1/16\pi_0 + 1/8\pi_1$

$\pi_3 = 1/16\pi_0 + 1/8\pi_1$

$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$

Al despejar para igualar el lado izquierdo de las primeras cuatro ecuaciones a cero, se obtiene:

$0 = -\pi_0 + \pi_3$

$0 = 7/8\pi_0 - 1/4\pi_1 + \pi_2$

$0 = 1/16\pi_0 + 1/8\pi_1 - \pi_2$

$0 = 1/16\pi_0 + 1/8\pi_1 - \pi_3$

$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$

Resolviendo en Google Colab:

Política R_b: Costo Promedio Estimado a Largo Plazo

[5]

```
import numpy as np

A = np.array([[-1, 0, 0, 1],
              [7/8, -1/4, 1, 0],
              [1/16, 1/8, -1, 0],
              [1, 1, 1, 1]])

B = np.array([0, 0, 0, 1])
X = np.linalg.inv(A).dot(B)

print(X)
```

```
[0.0952381  0.71428571 0.0952381  0.0952381 ]
```

Es decir:

$\pi_0 = 2/21$

$\pi_1 = 5/7$

$\pi_2 = 2/21$

$\pi_3 = 2/21$

Para la política R_c:

Sistema de ecuaciones:

$\pi_0 = \pi_2 + \pi_3$

$\pi_1 = 7/8\pi_0 + 3/4\pi_1$

$\pi_2 = 1/16\pi_0 + 1/8\pi_1$

$\pi_3 = 1/16\pi_0 + 1/8\pi_1$

$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$

Al despejar para igualar el lado izquierdo de las primeras cuatro ecuaciones a cero, se obtiene:

$0 = -\pi_0 + \pi_2 + \pi_3$

$0 = 7/8\pi_0 - 1/4\pi_1$

$0 = 1/16\pi_0 + 1/8\pi_1 - \pi_2$

$0 = 1/16\pi_0 + 1/8\pi_1 - \pi_3$

$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$

Resolviendo en Google Colab:

Política R_c: Costo Promedio Estimado a Largo Plazo

```
import numpy as np

A = np.array([[-1, 0, 1, 1],
              [7/8, -1/4, 0, 0],
              [1/16, 1/8, -1, 0],
              [1, 1, 1, 1]])

B = np.array([0, 0, 0, 1])
X = np.linalg.inv(A).dot(B)

print(X)
```

```
[0.18181818 0.63636364 0.09090909 0.09090909]
```

Es decir:

$\pi_0 = 2/11$

$\pi_1 = 7/11$

$\pi_2 = 1/11$

$\pi_3 = 1/11$

Para la política R_d**:**

Sistema de ecuaciones:

$\pi_0 = \pi_1 + \pi_2 + \pi_3$

$\pi_1 = 7/8\pi_0$

$\pi_2 = 1/16\pi_0$

$\pi_3 = 1/16\pi_0$

$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$

Al despejar para igualar el lado izquierdo de las primeras cuatro ecuaciones a cero, se obtiene:

$$0 = -\pi_0 + \pi_1 + \pi_2 + \pi_3$$

$$0 = 7/8\pi_0 - \pi_1$$

$$0 = 1/16\pi_0 - \pi_1$$

$$0 = 1/16\pi_0 - \pi_2$$

$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3$$

Resolviendo en Google Colab:

Política R_d: Costo Promedio Estimado a Largo Plazo

```
import numpy as np

A = np.array([[-1, 1, 1, 1],
              [7/8, -1, 0, 0],
              [1/16, 0, -1, 0],
              [1, 1, 1, 1]])

B = np.array([0, 0, 0, 1])
X = np.linalg.inv(A).dot(B)

print(X)
```

```
[0.5     0.4375   0.03125 0.03125]
```

Es decir:

$$\pi_0 = 1/2$$

$$\pi_1 = 7/16$$

$$\pi_2 = 1/32$$

$$\pi_3 = 1/32$$

Ahora usaremos este resultado para calcular el costo promedio esperado a largo plazo del ejemplo:

Política	$(\pi_0, \pi_1, \pi_2, \pi_3)$	$E(C)$, en miles de dólares
R_a	$\left(\frac{2}{13}, \frac{7}{13}, \frac{2}{13}, \frac{2}{13}\right)$	$\frac{1}{13}[2(0) + 7(1) + 2(3) + 2(6)] = \frac{25}{13} = \$1\,923$
R_b	$\left(\frac{2}{21}, \frac{5}{7}, \frac{2}{21}, \frac{2}{21}\right)$	$\frac{1}{21}[2(0) + 15(1) + 2(4) + 2(6)] = \frac{35}{21} = \$1\,667 \leftarrow$ Mínimo
R_c	$\left(\frac{2}{11}, \frac{7}{11}, \frac{1}{11}, \frac{1}{11}\right)$	$\frac{1}{11}[2(0) + 7(1) + 1(6) + 1(6)] = \frac{19}{11} = \$1\,727$
R_d	$\left(\frac{1}{2}, \frac{7}{16}, \frac{1}{32}, \frac{1}{32}\right)$	$\frac{1}{32}[16(0) + 14(6) + 1(6) + 1(6)] = \frac{96}{32} = \$3\,000$

En base a estos resultados, la política óptima es R_b, la cual consiste en reemplazar la máquina cuando se encuentre en el estado 3 y hacer una reparación cuando se encuentre en el estado 2. En este caso, el costo total esperado a largo plazo por semana es de $1 667.

Tipos de Políticas

Existen dos tipos de políticas:

- **Política determinística:**
  - Tipo de política usada para determinar las decisiones que se deben tomar en un proceso de decisión Markoviano en base al estado actual del proceso. En dicha política, se asigna el valor 0 o 1 para reflejar si la decisión D_{ik} debe tomarse o no cuando el proceso está en el estado i. Cada fila debe contener un solo valor 1 y el resto debe ser 0 para especificar la decisión que se debe tomar.

Por ejemplo:

$$\begin{array}{cc} & \text{Decisión } k \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \text{Estado } i \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \end{array}$$

- **Política aleatorizada:**
    - Tipo de política que usa distribución de probabilidades. En este caso, D_{ik} refleja la probabilidad de tomar la decisión k si el proceso se encuentra en el estado i. Cada fila suma 1 ya que se debe tomar una decisión al realizar la transición.

$$\begin{array}{cc} & \text{Decisión } k \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \text{Estado } i \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}. \end{array}$$

Al permitir políticas aleatorizadas, de forma que D_{ik} sean variables continuas, se puede formular un modelo de programación lineal para encontrar la política óptima.

Formulación de un Modelo de Programación Lineal

El modelo de programación lineal busca minimizar las variables de decisión y_{ik}, la probabilidad incondicional de estado estable de que el sistema se encuentre en el estado i y se tome la decisión k.

$$\text{Minimizar} \qquad Z = \sum_{i=0}^{M} \sum_{k=1}^{K} C_{ik} y_{ik},$$

sujeta a las restricciones

$$(1) \qquad \sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} = 1.$$

$$(2) \qquad \sum_{k=1}^{K} y_{jk} - \sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} p_{ij}(k) = 0, \qquad \text{para } j = 0, 1, \ldots, M.$$

$$(3) \qquad y_{ik} \geq 0, \qquad \text{para } i = 0, 1, \ldots, M; \; k = 1, 2, \ldots, K.$$

Programación Lineal para el Ejemplo

Variables de decisión: y_{01}, y_{11}, y_{13}, y_{21}, y_{22}, y_{23}, y_{33} donde y_{ik} representa tomar la decisión k si el proceso se encuentra en el estado i.

Función Objetivo:

$$\text{Minimizar} \qquad Z = 1\,000 y_{11} + 6\,000 y_{13} + 3\,000 y_{21} + 4\,000 y_{22} + 6\,000 y_{23} + 6\,000 y_{33},$$

Sujeto a:

$$
\begin{aligned}
y_{01} + y_{11} + y_{13} + y_{21} + y_{22} + y_{23} + y_{33} &= 1 \\
y_{01} - (y_{13} + y_{23} + y_{33}) &= 0 \\
y_{11} + y_{13} - \left(\frac{7}{8}y_{01} + \frac{3}{4}y_{11} + y_{22}\right) &= 0 \\
y_{21} + y_{22} + y_{23} - \left(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}\right) &= 0 \\
y_{33} - \left(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}\right) &= 0
\end{aligned}
$$

Además:

$$\text{todas las } y_{ik} \geq 0.$$

Al aplicar el método simplex, se obtienen los resultados:

$$y_{01} = \frac{2}{21}, \qquad (y_{11}, y_{13}) = \left(\frac{5}{7}, 0\right), \qquad (y_{21}, y_{22}, y_{23}) = \left(0, \frac{2}{21}, 0\right), \qquad y_{33} = \frac{2}{21},$$

Y para calcular D_{ik}:

$$D_{ik} = \frac{y_{ik}}{\sum_{k=1}^{K} y_{ik}}.$$

Se obtiene:

$$D_{01} = 1, \qquad (D_{11}, D_{13}) = (1, 0), \qquad (D_{21}, D_{22}, D_{23}) = (0, 1, 0), \qquad D_{33} = 1.$$

Según esta política:

- Cuando se está en el estado 0 o 1 se debe tomar la decisión 1, no hacer nada.
- Cuando se está en el estado 2 se debe realizar una reparación general.
- Cuando se está en el estado 3 se debe realizar un reemplazo total de la máquina.

El mismo resultado que se obtuvo con la enumeración exhaustiva de todas las políticas posibles.

Algoritmo de mejoramiento de políticas

Este algoritmo es eficiente porque nos permite llegar (casi siempre) a una solución óptima en un número relativamente pequeño de iteraciones.

Inicia con la *determinación del valor* escogiendo una política arbitraria R_1 y se resuelve el sistema de ecuaciones para encontrar su solución. Luego, se procede a mejorar la política. Esta política mejorada se denota con R_2. Ambos pasos representan una iteración del algoritmo. El proceso continúa hasta que dos sucesiones consecutivas tengan el mismo resultado.

Solución del Ejemplo con el Algoritmo de Mejoramiento de Políticas

Paso inicial: de forma arbitraria se escoge la política que establece que se reemplace la máquina (decisión 3) cuando el proceso se encuentre en el estado 3 y que no se haga nada (decisión 1) en otros estados.

Este es el resumen de la política inicial:

Política R_1

Estado	Decisión
0	1
1	1
2	1
3	3

Esta es la matriz de transición para esa política:

Matriz de transición

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

Estos son los costos:

Costos

Estado	C_{ik}
0	0
1	1 000
2	3 000
3	6 000

Para la iteración 1 se resuelve el siguiente sistema de ecuaciones:

$$g(R_1) = \qquad + \frac{7}{8}v_1(R_1) + \frac{1}{16}v_2(R_1) - v_0(R_1).$$

$$g(R_1) = 1\,000 \qquad + \frac{3}{4}v_1(R_1) + \frac{1}{8}v_2(R_1) - v_1(R_1).$$

$$g(R_1) = 3\,000 \qquad + \frac{1}{2}v_2(R_1) - v_2(R_1).$$

$$g(R_1) = 6\,000 + v_0(R_1).$$

La solución simultánea es:

$$g(R_1) = \frac{25\,000}{13} = 1\,923$$

$$v_0(R_1) = -\frac{53\,000}{13} = -4\,077$$

$$v_1(R_1) = -\frac{34\,000}{13} = -2\,615$$

$$v_2(R_1) = \frac{28\,000}{13} = 2\,154.$$

Ahora se realiza el mejoramiento de la política. Deseamos minimizar estas expresiones para cada estado i posible tomando la decisión k óptima.

$$\begin{array}{ll} \text{Estado 0:} & C_{0k} - p_{00}(k)(4\,077) - p_{01}(k)(2\,615) + p_{02}(k)(2\,154) + 4\,077 \\ \text{Estado 1:} & C_{1k} - p_{10}(k)(4\,077) - p_{11}(k)(2\,615) + p_{12}(k)(2\,154) + 2\,615 \\ \text{Estado 2:} & C_{2k} - p_{20}(k)(4\,077) - p_{21}(k)(2\,615) + p_{22}(k)(2\,154) - 2\,154 \\ \text{Estado 3:} & C_{3k} - p_{30}(k)(4\,077) - p_{31}(k)(2\,615) + p_{32}(k)(2\,154). \end{array}$$

En el estado 1:

	Estado 1					
Decisión	C_{1k}	$p_{10}(k)$	$p_{11}(k)$	$p_{12}(k)$	$p_{13}(k)$	**Valor de la expresión**
1	1 000	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	1 923 ← Mínimo
3	6 000	1	0	0	0	4 538

Por lo tanto, se elige la decisión 1 si se está en el estado 1.

En el estado 2:

	Estado 2					
Decisión	C_{2k}	$p_{20}(k)$	$p_{21}(k)$	$p_{22}(k)$	$p_{23}(k)$	**Valor de la expresión**
1	3 000	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1 923
2	4 000	0	1	0	0	−769 ← Mínimo
3	6 000	1	0	0	0	−231

Por lo tanto, se elige la decisión 2 como la decisión que se debe tomar si se está en el estado 2.

Se obtiene la siguiente política R_2:

Política R_2

Estado	Decisión
0	1
1	1
2	2
3	3

Matriz de transición

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	1	0	0
3	1	0	0	0

Costos

Estado	C_{ik}
0	0
1	1 000
2	4 000
3	6 000

Como esta política no es idéntica a R_1, se continúa el proceso para encontrar R_3.

Se resuelve el sistema de ecuaciones:

$$g(R_2) = \quad + \frac{7}{8}v_1(R_2) + \frac{1}{16}v_2(R_2) - v_0(R_2).$$

$$g(R_2) = 1\,000 + \frac{3}{4}v_1(R_2) + \frac{1}{8}v_2(R_2) - v_1(R_2).$$

$$g(R_2) = 4\,000 + v_1(R_2) - v_2(R_2).$$

$$g(R_2) = 6\,000 + v_0(R_2).$$

Se encuentra la solución:

$$g(R_2) = \frac{5\,000}{3} = 1\,667$$

$$v_0(R_2) = -\frac{13\,000}{3} = -4\,333$$

$$v_1(R_2) = -3\,000$$

$$v_2(R_2) = -\frac{2\,000}{3} = -667.$$

Se mejora la política al minimizar las siguientes expresiones:

$$\text{Estado 1:} \qquad C_{1k} - p_{10}(k)(4\,333) - p_{11}(k)(3\,000) - p_{12}(k)(667) + 3\,000$$
$$\text{Estado 2:} \qquad C_{2k} - p_{20}(k)(4\,333) - p_{21}(k)(3\,000) - p_{22}(k)(667) + 667.$$

Decisión	Valor del estado 1	Valor del estado 2
1	1 667	3 333
2	—	1 667
3	4 667	2 334

Por lo tanto, la política resultante es:

Política R_3

Estado	Decisión
0	1
1	1
2	2
3	3

Esta política es idéntica a la anterior, así que se encontró la política óptima para este proceso y el algoritmo finaliza su ejecución.

Criterio del costo descontado

Este criterio usa una medida alternativa de desempeño: el costo descontado total esperado. Esta medida usa un factor de descuento α con $0 < \alpha < 1$ y es un criterio preferido cuando el periodo de duración de la cadena de Markov es lo suficientemente largo para que el valor del dinero en el tiempo se deba tomar en cuenta. Otra ventaja es que se puede adaptar a cadenas de Markov con tiempos distintos.

Se puede describir como: i / $(1 + i)$ donde i es la tasa de interés actual por periodo.

Método de Aproximaciones Sucesivas

Este método se usa para encontrar rápidamente una aproximación a la solución óptima cuando solo quedan n periodos de observación (asumiendo que se está trabajando con una cadena de Markov de tiempo discreto). Se busca encontrar una política óptima para la decisión que se toma en el primer periodo cuando quedan n periodos antes de terminar el proceso.
A medida que n crece, las políticas óptimas convergen hacia la solución óptima para un proceso infinito.

Solución del ejemplo por el método de aproximaciones sucesivas
Se usará el factor de descuento $\alpha = 0.9$.

Se define:

V_i^n = costo descontado total esperado por seguir una política óptima, dado que el proceso comienza en el estado *i* y le quedan sólo *n* periodos de operación.[5]

Para la primera iteración, se obtienen los valores para cada V_i^1 junto con el valor de k que minimiza, entre paréntesis:

$$V_0^1 = \min_{k=1} \{C_{0k}\} = 0 \qquad (k = 1)$$
$$V_1^1 = \min_{k=1,3} \{C_{1k}\} = 1\,000 \qquad (k = 1)$$
$$V_2^1 = \min_{k=1,2,3} \{C_{2k}\} = 3\,000 \qquad (k = 1)$$
$$V_3^1 = \min_{k=3} \{C_{3k}\} = 6\,000 \qquad (k = 3)$$

Para la segunda iteración:

$$V_0^2 = 0 + 0.9\left[\frac{7}{8}(1\,000) + \frac{1}{16}(3\,000) + \frac{1}{16}(6\,000)\right] = 1\,294 \qquad (k = 1)$$
$$V_1^2 = \min\left\{1\,000 + 0.9\left[\frac{3}{4}(1\,000) + \frac{1}{8}(3\,000) + \frac{1}{8}(6\,000)\right],\ 6\,000 + 0.9[1(0)]\right\} = 2\,688 \qquad (k = 1)$$
$$V_2^2 = \min\left\{3\,000 + 0.9\left[\frac{1}{2}(3\,000) + \frac{1}{2}(6\,000)\right],\ 4\,000 + 0.9[1(1\,000)],\ 6\,000 + 0.9[1(0)]\right\} = 4\,900 \qquad (k = 2)$$
$$V_3^2 = 6\,000 + 0.9[1(0)] = 6\,000 \qquad (k = 3).$$

Para la tercera iteración:

$$V_0^3 = 0 + 0.9\left[\frac{7}{8}(2\,688) + \frac{1}{16}(4\,900) + \frac{1}{16}(6\,000)\right] \qquad = 2\,730 \qquad (k = 1)$$

$$V_1^3 = \min\left\{1\,000 + 0.9\left[\frac{3}{4}(2\,688) + \frac{1}{8}(4\,900) + \frac{1}{8}(6\,000)\right],\right.$$
$$\left. 6\,000 + 0.9[1(1\,294)]\right\} = 4\,041 \qquad (k = 1)$$

$$V_2^3 = \min\left\{3\,000 + 0.9\left[\frac{1}{2}(4\,900) + \frac{1}{2}(6\,000)\right],\right.$$
$$\left. 4\,000 + 0.9[1(2\,688)],\ 6\,000 + 0.9[1(1\,294)]\right\} = 6\,419 \qquad (k = 2)$$

$$V_3^3 = \qquad 6\,000 + 0.9[1(1\,294)] \quad = 7\,165 \qquad (k = 3).$$

Se obtiene la política óptima para el problema de periodos infinitos y los costos son similares.

Referencia

Hillier, F. & Lieberman, G. (2010). Introducción a la Investigación de Operaciones.