90
2EJ

UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

METODOS DE SOLUCION DE PROCESOS MARKOVIANOS DE DECISION EN TIEMPO DISCRETO

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A :

P R E S E N T A :

EVA MARIA VERDE OSORIO

DIRECTOR DE TESIS:
M. en C. BEATRIZ E. RODRIGUEZ FERNANDEZ

1995

VNIVERSDAD NACIONAL
AVFN°MA DE
MEXICO

M. EN C. VIRGINIA ABRIN BATULE
Jefe de la División de Estudios Profesionales
Facultad de Ciencias
P r e s e n t e

Los abajo firmantes, comunicamos a Usted, que habiendo revisado el trabajo de Tesis que realiz(ó)ron la pasante(s) EVA MARIA VERDE OSORIO

con número de cuenta 8852644-3 con el Título:

"METODOS DE SOLUCION DE PROCESOS MARKOVIANOS DE DECISION EN TIEMPO DISCRETO".

Otorgamos nuestro Voto Aprobatorio y consideramos que a la brevedad deberá presentar su Examen Profesional para obtener el título de ACTUARIO

GRADO	NOMBRE(S)	APELLIDOS COMPLETOS	FIRMA
M. en C.	Beatriz Eugenia	Rodríguez Fernández	Beatriz Rodríguez F.
Director de Tesis			
M. en I.O.	María del Carmen	Hernández Ayuso	Ma del Carmen Hdez Ayuso
Mat.	Jesús Agustín	Cano Garcés	
Mat.	Jorge Francisco	De la Vega Góngora	
Suplente			
Act.	María Esther	Caamaño Sierra	
Suplente			

A mi pequeño Daniel, fuerza inspiradora en la terminación de este trabajo.

Agradecimientos

Quiero agradecer a Carmen Hernández y a Agustín Cano, así como a Ma. Esther Caamaño y a Jorge de la Vega haber aceptado ser parte de mi jurado, con lo cual dedicaron varias horas de su valioso tiempo a revisarlo.

Muy especialmente agradezco a Beatriz Rodríguez, directora de esta tesis, las certeras observaciones que hizo a la misma las que en todo momento contribuyeron a su mejoramiento. Por sobre todo ello agradezco su amistad, sin la cual el tiempo que pasamos corrigiendo este trabajo no hubiera transcurrido tan agradablemente.

INDICE

CAPITULO IV EJEMPLOS

INTRODUCCION

Dentro de los procesos estocásticos que, como es conocido, involucran dos características esenciales: la aleatoriedad y el tiempo, encontramos un tipo especial de proceso en el que interviene una característica adicional: para cada estado en el que se encuentra el proceso es posible tomar una decisión, elegible dentro de un conjunto finito de ellas, y a la cual se encuentra asociada una ganancia. A este tipo de proceso se le denomina Proceso Markoviano de Decisión.

El presente trabajo está orientado básicamente a estudiar tres de los métodos más importantes para la resolución de un Proceso Markoviano de Decisión, es decir, para la obtención de una política que maximice las ganancias a que nos referíamos en el párrafo anterior.

El capítulo I puede considerarse un capítulo introductorio, en el cual lo que se persigue es estudiar de una forma general los conceptos más importantes de una cadena de Markov.

Con los conceptos revisados en el capítulo I , resulta más fácil entrar directamente al estudio de los Procesos Markovianos de Decisión, tema que abarca precisamente el capítulo II.

Habiéndose hablado de lo que es un Proceso Markoviano de Decisión, el capítulo III está orientado al estudio de tres de los métodos más comunmente estudiados, y que permiten obtener una política óptima para tales procesos.

Finalmente, en el capítulo IV se ve ilustrada la aplicación práctica de los tres métodos estudiados, mediante la resolución de dos ejemplos.

CAPITULO I

CADENAS DE MARKOV

INTRODUCCION

El estudio del comportamiento a través del tiempo de un gran número de fenómenos ha llevado a científicos de distintas ramas de la ciencia a interesarse cada vez más en los modelos de procesos aleatorios dinámicos. Por ejemplo: los biólogos están interesados en el crecimiento y extinción de poblaciones animales y humanas; los físicos estudian el movimiento de pequeñas partículas; los sicólogos observan los distintos procesos de aprendizaje; todos estos fenómenos pueden ser modelados como procesos aleatorios dinámicos o más comúnmente llamados procesos estocásticos.

Los procesos estocásticos proporcionan también un método para el estudio cuantitativo de operaciones administrativas y consecuentemente juegan un papel muy importante en las disciplinas modernas de administración e investigación de operaciones (control de inventario, teoría de colas, etc.) .

En términos generales podemos decir que un fenómeno que surge como un proceso el cual se desarrolla en el tiempo de forma controlada por leyes de probabilidad es llamado un proceso estocástico.

Definición 1.1 PROCESO ESTOCASTICO

Desde el punto de vista de la teoría matemática de la probabilidad, un proceso estocástico está definido como una colección de variables aleatorias $\{X(t), t \in T\}$ donde el conjunto T es llamado el conjunto de índices del proceso.

No hay ninguna restricción sobre la naturaleza de T, sin embargo dos casos importantes son cuando:

i) T es discreto, por ejemplo: $T=\{0,\pm1,\pm2,....\}$ ó $T=\{0,1,2,....\}$ y en este caso se dice que el proceso estocástico es un proceso en tiempo discreto o con parámetro discreto.

ii) T es continuo $T=\{t \mid -\infty < t < \infty\}$ ó $T=\{t \geq 0\}$ en cuyo caso se dice que se trata de un proceso en tiempo continuo.

Frecuentemente t representa el tiempo, aunque en algunas situaciones puede representar algo diferente, por ejemplo, la distancia a un origen arbitrario, y entonces la variable aleatoria (v.a.) X(t) puede contar el número de defectos en el intervalo (0,t) a lo largo de un cable o el número de automóviles en el intervalo (0,t) a lo largo de una carretera.

Un proceso estocástico está determinado por tres características esenciales: por el espacio de estados S, que es el rango de posibles valores para la v.a. X(t), por el conjunto de índices T y por la relación de dependencia entre las variables aleatorias.

Precisamente por cumplir con una relación de dependencia muy especial entre las variables aleatorias, dentro de los procesos estocásticos se destacan los procesos de Markov.

Definición 1.2 PROCESO DE MARKOV

Un proceso de Markov $\{X(t)\}$ es un proceso estocástico con la propiedad de que dado el valor de X(t), los valores de X(s) para $s > t$ no están influidos por los valores de X(u) para $u < t$; en otras palabras, la probabilidad con la cual se da cualquier comportamiento futuro del proceso dado que se conoce con exactitud el estado actual, no se ve alterada por el conocimiento adicional que se tenga del comportamiento del proceso en el pasado. A esta propiedad se le denomina Propiedad de Markov.

Los procesos de Markov se clasifican:

i) De acuerdo a la naturaleza del conjunto de índices del proceso: T puede tener parámetro t discreto o continuo, y

ii) De acuerdo a la naturaleza del espacio de estados: S puede tener un número finito o numerable de elementos y en este caso se dice que es discreto, o bien contener un número infinito de ellos y en este caso decimos que es continuo.

Un proceso de Markov cuyo espacio de estados es discreto, es llamado una cadena de Markov. En este estudio sólo nos avocaremos al estudio de cadenas de Markov cuyo espacio de estados sea un conjunto finito.

Definición 1.3 CADENA DE MARKOV EN TIEMPO DISCRETO O CON PARAMETRO DISCRETO.

Una cadena de Markov $\{X(t),\ t=0,1,2,....\}$ para la cual el conjunto de índices es $T=(0,1,2....)$ es decir es un conjunto discreto, es llamada una cadena de Markov en tiempo discreto o con parámetro discreto. En este caso la propiedad de Markov, en términos de probabilidad quedaría expresada como:

$$P(X(t+1)=j \mid X(0)=i_0,\ X(1)=i_1,\ .....\ X(t-1)=i_{t-1},\ X(t)=i)$$
$$=P(X(t+1)=j \mid X(t)=i)$$

A partir de este momento sólo nos avocaremos al estudio de cadenas de Markov en tiempo discreto y nos referiremos a ellas simplemente como cadenas de Markov.

Cuando se estudia un conjunto de variables aleatorias uno de los objetivos más importantes, sino es que el más importante, es determinar su distribución conjunta.

En el caso de las cadenas de Markov, es posible lograr este objetivo al conocer la distribución inicial del proceso y las probabilidades de transición en un paso. Definiremos primero cada uno de estos conceptos y algunos otros que nos permitirán más adelante demostrar este resultado.

Definición 1.4 PROBABILIDAD DE TRANSICION EN UN PASO

La probabilidad de que en el próximo periodo $X(t+1)$ se encuentre en el estado j, dado que en periodo t actual se encuentra en el estado i, se le conoce como probabilidad de transición del estado i al estado j en un paso y se denota:

$$P_{t,t+1}(j \mid i)=P(X(t+1)=j \mid X(t)=i)$$

Como se observa en la notación, las probabilidades de transición son funciones que dependen no solamente de los estados inicial y final, sino también del tiempo en que ocurre la transición.

Definición 1.5 PROBABILIDADES DE TRANSICION ESTACIONARIAS

Cuando las probabilidades de transición en un paso son independientes del periodo en el cual se encuentra el proceso cuando ocurre la transición, es decir, son independientes del parámetro t, decimos que la cadena de Markov tiene probabilidades de transición estacionarias y denotamos: $P_{t,t+1}(j \mid i) = P(j \mid i)$.

A menos que se especifique lo contrario, a partir de este momento sólo consideraremos probabilidades de transición estacionarias, es decir:

$$P_{t,t+1}(j \mid i) = P(j \mid i) \qquad \forall\ t \in T$$

Para cada pareja de estados (i,j) i,j $\in$ S, existe la probabilidad de transición en un paso $P(j \mid i)$, las cuales por ser probabilidades satisfacen que:

i) $P(j \mid i) \geq 0 \quad i,j \in S$

ii) $\sum_{j \in S} P(j \mid i) = 1 \quad i,j \in S$

La condición ii) establece que en cada periodo ocurre una transición, así que por conveniencia diremos que ha ocurrido una transición aún cuando el proceso no cambie de estado con respecto al estado en el que se encontraba al inicio del periodo anterior.

Definición 1.6 DISTRIBUCION INICIAL DEL PROCESO

La probabilidad de que X(0)= i, $i \in S$ denotada por $P(X(0) = i) = \Pi_0(i)$ es llamada la distribución inicial del proceso y cumple que:

i) $\Pi_0(i) \geq 0 \quad i \in S$

ii) $\sum_{i \in S} \Pi_0(i) = 1$

Definición 1.7 MATRIZ DE PROBABILIDADES DE TRANSICION EN UN PASO

La matriz de probabilidades de transición en un paso o simplemente matriz de probabilidades de transición, es un arreglo cuadrado cuya dimensión es el número de estados del proceso, en el cual se agrupan las probabilidades de transición en un paso y se denota como $P = \| P(j \mid i) \|$.

Así, el renglón asociado al estado i contiene la distribución condicional de probabilidad de $X(t+1)$ dado que $X(t)= i$.

Como habíamos mencionado anteriormente, en una cadena de Markov $\{X(t)\}$ es posible determinar la distribución conjunta de un grupo de variables aleatorias, al conocer únicamente su distribución inicial y las probabilidades de transición estacionarias en un paso. Supongamos que nos interesa conocer la distribución conjunta de $(X(1),X(2),X(3))$, para este caso ilustraremos este hecho.

Por la ley Multiplicativa de la Probabilidad:

$$P(X(0)= i_0, X(1)= i_1, X(2)= i_2) =$$

$$P(X(0)= i_0)\, P(X(1)= i_1 \mid X(0)= i_0)\, P(X(2)= i_2 \mid X(0)= i_0 , X(1)= i_1) =$$

$$\Pi_0 (i_0)\, P(i_1 \mid i_0)\, P(X(2)= i_2 \mid X(0)= i_0 , X(1)= i_1) \quad .... (1.1)$$

Usando que $\{X(t)\}$ es una cadena de Markov con probabilidades de transición estacionarias, tenemos que:

$$P(X(2)= i_2 \mid X(0)= i_0 , X(1)= i_1) =$$

$$P(X(2)= i_2 \mid X(1)= i_1) = P(i_2 \mid i_1)$$

Sustituyendo este resultado en la ecuación 1.1, tenemos que:

$$P(X(0)= i_0, X(1)= i_1, X(2)= i_2) =$$

$$\Pi_0 (i_0)\, P(i_1 \mid i_0)\, P(i_2 \mid i_1)$$

Usando inducción es fácil demostrar que:

$$P(X(0)= i_0, X(1)= i_1, ...X(t)= i_t) =$$

$$\Pi_0 (i_0)\, P(i_1 \mid i_0)\, P(i_2 \mid i_1) ...... P(i_t \mid i_{t-1}) \quad .... (1.2)$$

Como conclusión a lo anterior podemos decir que una cadena de Markov está completamente definida si se conocen su matriz de probabilidades de transición y la distribución de probabilidad de la cadena al tiempo 0 .

El estudio de una cadena de Markov busca entre otras cosas determinar las probabilidades de posibles realizaciones futuras del proceso, es decir, determinar con qué probabilidad el proceso se encontrará en cierto estado $j \in S$, dado que han transcurrido n periodos.

Definición 1.8 MATRIZ DE PROBABILIDADES DE TRANSICION EN N-PASOS

$P^n = \| P^n(j|i) \|$ es llamada la matriz de probabilidades de transición en n-pasos y agrupa las probabilidades de transición con las que el proceso pasa de un estado inicial i a un estado j en n-pasos. Formalmente:

$$P^n(j|i) = P(X(t+n)=j \mid X(t)=i) \quad .... (1.3)$$

Como ya se había hecho notar anteriormente, estamos refiriéndonos a probabilidades de transición estacionarias, por lo que el lado izquierdo de la ecuación 1.3 no depende del periodo inicial t.

Una relación de recurrencia que permite calcular P^n está dada en el siguiente teorema. Este resultado es conocido como ecuación de Chapman-Kolmogorov en tiempo discreto.

TEOREMA 1.1

Sea $P = \| P(j|i) \|$ la matriz de probabilidades de transición de la cadena de Markov $\{X(t), t \in T\}$, N el número de estados y r y s enteros no negativos tales que r+s=n, entonces:

$$P^n(j|i) = \sum_{k=0}^{N} P^r(k|i)P^s(j|k) \quad .... (1.4)$$

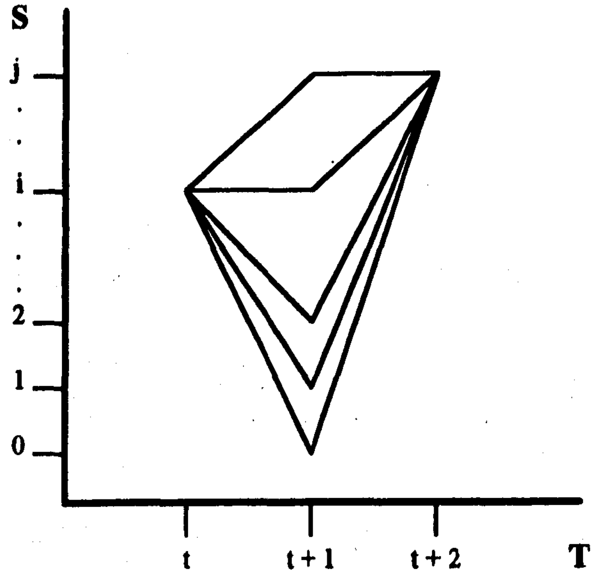
Demostración:

Definimos:

$$P^0(j|i) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Cuando n=1 es inmediato con r=1 y s=0 ó r=0 y s=1.

Para el caso n=2 la ecuación (1.4) puede entenderse intuitivamente observando la siguiente figura:



El evento de ir del estado i al tiempo t al estado j al tiempo t+2 puede realizarse de N formas mutuamente excluyentes al desplazarse primero a un estado intermedio k al tiempo t+1 y pasar de ahí al estado j al tiempo t+2.

En otras palabras, el evento {X(t)= i , X(t+2)= j} es la unión de los N eventos mutuamente excluyentes {X(t)= i, X(t+1)= k, X(t+2)= j} k= 0,1,2,....N, esto es:

$$P(X(t)=i, X(t+2)=j)=\sum_{k=0}^{N} P(X(t)=i, X(t+1)=k, X(t+2)=j) \ldots \quad (1.5)$$

Usando la ecuación (1.1) y el hecho de que se trata de probabilidades de transición estacionarias tenemos que:

$P(X(t)=i, X(t+1)=k, X(t+2)=j)=$

$P(X(0)=i, X(1)=k, X(2)=j)=$

$\Pi_0(i)\, P(k \mid i)\, P(j \mid k) \ldots \quad (1.6)$

Sustituyendo este resultado en la ecuación (1.5) tenemos que:

$$P(X(t)=i, X(t+2)=j)=\sum_{k=0}^{N} P(X(t)=i, X(t+1)=k, X(t+2)=j)$$

$$=\sum_{k=0}^{N} \Pi_0(i)\, P(k \mid i)\, P(j \mid k)$$

Por ley de probabilidades totales sabemos que:

$$P^2(j \mid i) = P(X(t+2)=j \mid X(t)=i) = \frac{P(X(t)=i,\ X(t+2)=j)}{P(X(t)=i)}$$

Debido a que se trata de probabilidades de transición estacionarias, tenemos que:

$$= \frac{P(X(0)=i,\ X(2)=j)}{P(X(0)=i)}$$

$$= \frac{P(X(0)=i,\ X(2)=j)}{\Pi_0(i)}$$

Usando las ecuaciones (1.5) y (1.6) el numerador del cociente anterior puede ser expresado como:

$$= \Pi_0(i) \sum_{k=0}^{N} P(k \mid i)\, P(j \mid k)$$

Por lo que finalmente $P^2 = \sum_{k=0}^{N} P(k \mid i)\, P(j \mid k)$.

Para el caso general el argumento que se emplearía para realizar la demostración sería el mismo. ■

Resumiendo hemos llegado a que:

$$P^n = \sum_{k=0}^{N} P^r(k \mid i)\, P^s(j \mid k) \qquad \text{con } r+s=n$$

Usando la teoría de matrices observamos que esta relación no es sino la fórmula de multiplicación de matrices, en otras palabras, las probabilidades de transición $P^n(j \mid i)$ pueden obtenerse como las entradas de la matriz que resulta de elevar la matriz P (matriz de probabilidades de transición en un paso) a la potencia n.

Siguiendo con la idea del análisis del comportamiento del proceso a través del tiempo, revisaremos ahora algunos conceptos relativos a la clasificación de estados en una cadena de Markov.

Definición 1.9 ESTADO ACCESIBLE

Se dice que un estado $j \in S$ es accesible de otro estado $i \in S$, si existe algún entero $n \geq 0$ para el que $P^n(j \mid i) > 0$; en otras palabras, el estado j es accesible del estado i si existe la probabilidad positiva de que en un número finito de transiciones el estado j pueda ser "visitado" dado que el estado inicial de la cadena es i.

Si dos estados i y j son accesibles uno del otro se dice que i y j están comunicados y se denota $i \leftrightarrow j$.

El concepto de comunicación es una relación de equivalencia, es decir, cumple con las siguientes propiedades:

i) La reflexividad: $i \leftrightarrow i$

Es una consecuencia de la definición dada en el Teorema 1.1:

$$P^0(j \mid i) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

ii) La simetría: si $i \leftrightarrow j$ entonces $j \leftrightarrow i$

Esta propiedad se deduce de la definición de comunicación entre estados.

iii) La transitividad: si $i \leftrightarrow j$ y $j \leftrightarrow k$ entonces $i \leftrightarrow k$

Si $i \leftrightarrow j$ y $j \leftrightarrow k$ entonces existen enteros n y m tales que $P^n(j \mid i) > 0$ y $P^m(k \mid j) > 0$.

Usando la ecuación de Chapman-Kolmogorov en tiempo discreto, y la hipótesis de que $P^n(j \mid i) > 0$ y $P^m(k \mid j) > 0$ tenemos que:

$$P^{m+n}(k \mid i) = \sum_{r=0}^{N} P^m(r \mid i)\, P^n(k \mid r)$$

$$\geq P^m(j \mid i)\, P^n(k \mid j) > 0$$

$$\Rightarrow P^{m+n}(k \mid i) > 0$$

$\therefore$ Existe un entero $(m+n)$ tal que la probabilidad de ir del estado i al estado k es estrictamente positiva, es decir, k es accesible de i.

Para demostrar que existe un entero $z \geq 0$ tal que $P^z(i \mid k) > 0$, es decir que i es accesible de k se emplea un argumento análogo y de esta forma queda demostrado que si $i \leftrightarrow j$ y $j \leftrightarrow k$ entonces $i \leftrightarrow k$.

Definición 1.10 CLASE COMUNICANTE

Dado $j \in S$, su clase comunicante, denotada por $C(j)$ se define como el conjunto de todos los estados $k \in S$ que están comunicados con j, formalmente:

$k \in C(j)$ si y sólo si $k \leftrightarrow j$.

Una característica importante que cumplen las clases comunicantes es la siguiente:

Si C_1 y C_2 son clases comunicantes entonces:

i) $C_1 = C_2$ ó
ii) $C_1 \cap C_2 = \varnothing$

Probaremos la afirmación anterior partiendo de que existe un elemento en la intersección de C_1 y C_2.

Sean $C_1 = C(j)$ y $C_2 = C(k)$. Supongamos que existe $h \in S$, tal que $h \in C(j)$ y $h \in C(k) \Rightarrow h \leftrightarrow j$ y $h \leftrightarrow k$.

Sea $g \in C(j) \Rightarrow g \leftrightarrow j$, pero como $h \leftrightarrow j$ entonces por la transitividad de la relación de equivalencia $g \leftrightarrow h$; por otro lado $h \leftrightarrow k$ entonces, nuevamente por transitividad $g \leftrightarrow k$ es decir $g \in C(k)$.

$\therefore C(j) \subset C(k)$.

La prueba de que $C(k) \subset C(j)$ es análoga, comenzando en este caso al suponer que existe $g \in S$ tal que $g \in C(k)$.

$\therefore C(j) = C(k)$ con lo cual queda demostrada la afirmación anterior.

Definición 1.11 CADENA DE MARKOV IRREDUCIBLE

Una cadena de Markov $\{X(t), t=0,1,2,....\}$ es irreducible si la relación de equivalencia (dada en la definición anterior) induce sólo una clase, en otras palabras, un proceso es irreducible si $\forall\, i, j \in S$ $i \leftrightarrow j$ .Cuando dicha relación de equivalencia induce más de una clase se dice que la cadena es reducible.

Para ilustrar el concepto de reducibilidad de una cadena de Markov consideremos la siguiente matriz de probabilidades de transición:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \end{bmatrix} \end{array}$$

Para la cadena de Markov $\{X(t)\}$ cuya matriz de probabilidades de transición es P, aplicando la definición de clase comunicante determinamos que:

1) En C(0) sólo está el 0, ya que el estado 4 es accesible del estado 0, pero el estado 0 resulta no ser accesible del estado 4.

2) En C(1) sólo está el estado 1, ya que aunque existe la probabilidad positiva de ir del estado 1 al estado 2, el estado 2 resulta no ser accesible del estado 1 y por lo tanto no puede estar en su clase.

Siguiendo con el mismo razonamiento obtenemos que:

$C(2)=\{2,3,4,5\}$
$C(3)=\{2,3,4,5\}$
$C(4)=\{2,3,4,5\}$
$C(5)=\{2,3,4,5\}$

$\Rightarrow C_0=C(0), C_1=C(1),\ C_2= C(2)= C(3)=C(4)=C(5)$
$\Rightarrow C_0 \cap C_1 \cap C_2= \varnothing$

Por lo que $\{X(t)\}$ no es irreducible, ya que existen tres clases comunicantes C_0, C_1 y C_2.

Definición 1.12 PERIODICIDAD DE UNA CADENA DE MARKOV

Se define el periodo de un estado i, denotado por d(i) como el máximo común divisor (m.c.d.) de todos los enteros $n\geq 1$ para los que se cumple $P^n(i \mid i) > 0$. Si $P^n(i \mid i) = 0$ para todo $n\geq 1$ se define $d(i)= 0$.

Para ilustrar el concepto de periodicidad de una cadena de Markov veamos el ejemplo de la caminata aleatoria.

Una caminata aleatoria es una cadena de Markov cuyo espacio de estados es un subconjunto finito o infinito de enteros, que se caracteriza porque al encontrarse el proceso en un estado i, en una transición simple solo puede pasar a cualquiera de sus estados vecinos i-1 ó i+1 o bien permanecer en el estado i en el que se encuentra.

Supongamos que se tiene una caminata aleatoria cuyo espacio de estados es el subconjunto de estados 0,1,2,3 es decir, S={0,1,2,3}. En términos generales la matriz asociada a este proceso es de la forma:

$$P=\begin{bmatrix} r & p & 0 & 0 \\ q & r & p & 0 \\ 0 & q & r & p \\ 0 & 0 & q & r \end{bmatrix}$$

Donde p,q>0, r≥0 y p+q+r=1.

Las probabilidades de transición quedan expresadas como:

$$P(j\mid i)=\underset{\substack{i=0,1,2,3\\ j=0,1,2,3}}{P(X=j\mid X=i)}=\begin{cases} q \text{ si } j=i-1 \\ r \text{ si } j=i \\ p \text{ si } j=i+1 \\ 0 \text{ e.o.c.} \end{cases}$$

Si r=0 entonces en una transición simple el proceso no puede permanecer en el mismo estado en el que se encontraba, así que tiene que moverse a cualquiera de sus estados vecinos inmediatos; la matriz P se reescribiría como:

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

Bajo esta condición (r=0) cada estado tiene periodo 2 ya que sólo cuando n es par existe una probabilidad positiva de regresar al estado inicial i , por lo que el m.c.d. del conjunto {n| n es par} es 2, es decir, para una caminata aleatoria con r= 0 el periodo de cualquiera de los estados es 2.

Cuando en una cadena de Markov cada uno de sus estados tiene periodo 1 se dice que es aperiódica.

Otro de los conceptos importantes dentro de la clasificación de estados es la recurrencia. Veamos algunos conceptos que nos permitirán más adelante definirla.

Consideremos un estado i arbitrario pero fijo. Para cada entero $n\geq 0$ definimos:

$f_{i,i}^{(0)} = 0$
$f_{i,i}^{(n)} = P\,(X(n)= i,\ X(v)\neq i,\ v= 1,2,\ldots n-1 \mid X(0)= i)\quad n\geq 1.$

En otras palabras $f_{i,i}^{(n)}$ es la probabilidad de que empezando en el estado i el primer retorno a dicho estado ocurra exactamente en la n-ésima transición.

$f_{i,i}^{(n)}$ puede calcularse recursivamente de acuerdo a la siguiente relación:

$$P^n(i \mid i) = \sum_{k=0}^{N} f_{i,i}^{(k)} P^{(n-k)}(i \mid i) \qquad n \geq 1 \ldots\ (1.7)$$

La ecuación (1.7) se obtiene al calcular $P^n(i \mid i)$ usando el concepto de primer retorno al estado i .Así, consideremos todas las posibles realizaciones del proceso para el cual $X(0)= i$, $X(n)=i$ y el primer retorno al estado i ocurra en la k-ésima transición, llamemos a este evento E_k.Los eventos E_k (k= 1,2,.... n) son claramente mutuamente excluyentes. La probabilidad del evento de que el primer retorno al estado i ocurra en la k-ésima transición es por definición $f_{i,i}^{(k)}$ y en las (n-k) transiciones restantes no nos interesa a que estados se mueva el proceso, sino únicamente que en la n-ésima transición se encuentre en el estado i nuevamente.

En términos de probabilidades lo anterior quedaría expresado como:

$$P^n(i \mid i) = P\,(X(n)= i \mid X(0)= i)$$

$$= \sum_{k=1}^{N} P(X(n) = i, X(k) = i, X(r) \neq i \quad r = 1,2,\ldots k-1 \mid X(0) = i)$$

$$= \sum_{k=1}^{N} P(X(n) = i \mid X(k) = i) P(X(k) = i, X(r) \neq i \quad r = 1,2,\ldots k-1 \mid X(0) = i)$$

$$= \sum_{k=1}^{N} P(X(n) = i \mid X(k) = i) P(E_k)$$

$$= \sum_{k=1}^{N} P^{n-k}(i \mid i)\, f_{i,i}^{(k)}$$

$$= \sum_{k=0}^{N} P^{n-k}(i \mid i)\, f_{i,i}^{(k)}$$

Esta última igualdad se da ya que $f_{i,i}^{(0)} = 0$.

Ahora ya tenemos los elementos necesarios para definir $f_{i,i}$ .Sea E_k como antes. Definimos el evento E= $\{X(k)= i$ para algún entero $k\geq 0 \mid X(0)= i\}$, como E es la unión de la sucesión ajena de conjuntos $\{E_k \ k=1,2,....\}$ se sigue que:

$$f_{i,i} = P(E) = \sum_{k=0}^{\infty} P(E_k)$$

$$= \sum_{k=0}^{\infty} P(X(k) = i, X(r) \neq i \quad r = 1,2,\ldots k-1 \mid X(0) = i)$$

$$= \sum_{k=0}^{\infty} f_{i,i}^{(k)}$$

$$\therefore \quad f_{i,i} = \sum_{k=0}^{\infty} f_{i,i}^{(k)} = \lim_{N\to\infty} \sum_{k=0}^{N} f_{i,i}^{(k)}.$$

DEFINICION 1.13 ESTADO RECURRENTE

Se dice que un estado i es recurrente si $f_{i,i}= 1$ es decir, un estado i es recurrente si y sólo si , dado que el proceso comienza en el estado i, la probabilidad de que retorne a dicho estado después de una longitud finita de tiempo es uno.

Un estado que no es recurrente, es decir $f_{i,i} < 1$, se dice que es transitorio.

Consideremos un estado i transitorio. Denotemos nuevamente como E el evento de regresar a i al menos una vez, entonces $P(E)= f_{i,i} < 1$.

Ahora denotemos como F al evento de regresar al estado i al menos dos veces. Claramente se observa que la ocurrencia del evento F implica la ocurrencia del evento E, es decir $F \subset E$ por lo que $P(F)= P(F\cap E)$.

Ahora bien, despejando de la ecuación de probabilidad condicional tenemos que :

$P(F\cap E)= P(F \mid E) \bullet P(E)$

Poniendo en forma explícita los eventos E y F tenemos que:

P(F)= P(regresar al estado i al menos dos veces)

=P(regresar al estado i al menos dos veces | regresamos al estado i al menos una vez) • P(regresar al estado i al menos una vez)

Pero por la propiedad de Markov, tenemos que:

P(regresar al estado i al menos dos veces | regresamos al estado i al menos una vez)

=P(regresar al estado i al menos una vez) = $f_{i,i}$

Por lo que P(F)= $f_{i,i} \bullet f_{i,i} = (f_{i,i})^2$.

Repitiendo este argumento tendríamos que para el caso general, la probabilidad de que el proceso regrese al estado i al menos k veces es $(f_{i,i})^k$ k=1,2,....

Consideremos ahora una variable aleatoria M que cuente el número de veces que el proceso regresa al estado i .De esta forma el evento E es equivalente al evento $\{M \geq 1\}$ y el evento F es equivalente al evento $\{M \geq 2\}$. Por lo expuesto anteriormente M tiene una distribución geométrica con parámetro $(1- f_{i,i})$ .En términos de probabilidad lo anterior se expresa como:

$$P(M \geq k \mid X(0)= i) = (f_{i,i})^k \text{ para } k=1,2,.....$$

Por lo que la esperanza de M está dada por:

$$E(M \mid X(0) = i) = \frac{f_{i,i}}{1 - f_{i,i}} .$$

Lo expuesto en los párrafos anteriores nos permitirá demostrar el siguiente teorema que establece un criterio para determinar si un estado i es recurrente o transitorio en términos de las probabilidades de transición $P^n (i \mid i)$.

TEOREMA 1.2

Un estado i es recurrente si y sólo si:

$$\sum_{n=1}^{\infty} P^n(i \mid i) = \infty$$

En forma equivalente, un estado es transitorio si y sólo si:

$$\sum_{n=1}^{\infty} P^n(i \mid i) < \infty$$

Demostración:

$\Rightarrow$) Supongamos que un estado i es transitorio, por definición $f_{i,i} < 1$.

Sea M una v.a. que cuenta el número de retornos al estado i. M puede ser escrita en términos de una función indicadora como:

$$M = \sum_{n=1}^{\infty} I(X(n) = i)$$

donde:

$$I(X(n) = i) = \begin{cases} 1 & \text{si } X(n) = i \\ 0 & \text{si } X(n) \neq i \end{cases}$$

Ya que M es una v.a. geométrica con parámetro $(1- f_{i,i})$ su valor esperado está dado por:

$$E(M \mid X(0) = i) = \frac{f_{i,i}}{1 - f_{i,i}}$$

Por lo que cuando i es transitorio $E(M \mid X(0) = i) < \infty$.

Por otro lado tenemos que :

$$\infty > E(M \mid X(0) = i) = \sum_{n=1}^{\infty} E(I(X(n) = i) \mid X(0) = i) =$$

$$\sum_{n=1}^{\infty} \left[0 \bullet P(X(n) \neq i \mid X(0) = i) + 1 \bullet P(X(n) = i \mid X(0) = i)\right] =$$

$$\sum_{n=1}^{\infty} P^n(i \mid i) \qquad \therefore \sum_{n=1}^{\infty} P^n(i \mid i) < \infty.$$

$\Leftarrow$) Supongamos que $\sum_{n=1}^{\infty} P^n(i \mid i) < \infty$

Tenemos por hipótesis que M es una v.a. cuya media es finita, así que M debe tomar valores finitos. En otras palabras ésto significa que si el proceso comienza en el estado i, debe regresar al estado i en una longitud de tiempo finita. Entonces tenemos una probabilidad positiva de que empezando en el estado i el proceso nunca retorne a dicho estado, en otras palabras $1- f_{i,i} > 0$ ó $f_{i,i} < 1$, es decir el estado i es transitorio.

$\therefore$ Ha quedado demostrado el teorema. ■

Como corolario al teorema anterior tenemos que:

COROLARIO

Si i $\leftrightarrow$ j y el estado i es recurrente entonces el estado j también es recurrente.

Demostración:

Por hipótesis i $\leftrightarrow$ j, entonces existen m, n $\geq$ 1 tales que:

$P^n(j \mid i) > 0$ y $P^m(i \mid j) > 0$. Sea v > 0, tenemos por la ecuación de Chapman-Kolmogorov que:

$$P^{(m+n+v)}(j \mid j) = \sum_{v=0}^{\infty} P^m(j \mid i) P^v(i \mid i) P^n(j \mid i) \geq P^m(j \mid i) P^v(i \mid i) P^n(j \mid i)$$

Sumando en ambos lados de la desigualdad tenemos que:

$$\sum_{v=0}^{\infty} P^{(m+n+v)}(j \mid j) \geq \sum_{v=0}^{\infty} P^m(j \mid i) P^v(i \mid i) P^n(j \mid i)$$

$$= P^m(j \mid i) P^n(j \mid i) \sum_{v=0}^{\infty} P^v(i \mid i)$$

Como por hipótesis i es recurrente, así que $\sum_{v=0}^{\infty} P^v(i \mid i) = \infty$ en consecuencia:

$$\sum_{v=0}^{\infty} P^{(m+n+v)}(j \mid j) > \infty \quad \Rightarrow j \text{ es recurrente.} \quad \blacksquare$$

Este corolario demuestra que la recurrencia así como la periodicidad son propiedades de clases, es decir, en una clase equivalente todos los estados son recurrentes o son transitorios.

Las probabilidades más usadas para efectos cuantitativos en una cadena de Markov son las llamadas probabilidades estacionarias denotadas por $\pi = (\pi_0, \pi_1, ...)$ donde en términos de probabilidad:

$$\pi_i = \lim_{n \to \infty} P(X(n) = i) \qquad i = 1,2....N.$$

Esta convergencia significa que a largo plazo ($n \to \infty$) la probabilidad de encontrar a la cadena en el estado i es aproximadamente π_i sin importar cual fue el estado inicial. En otras palabras las π_i's representan la proporción del tiempo que la cadena de Markov pasa en el estado i. El siguiente teorema nos da las condiciones necesarias para su existencia y nos muestra un método para calcularlas.

TEOREMA 1.3 TEOREMA BASICO DEL LIMITE EN UNA CADENA DE MARKOV

a) Considérese una cadena de Markov recurrente, irreducible y aperiódica. Sea $P^n(i \mid i)$ la probabilidad de alcanzar el estado i en la n-ésima transición n= 0,1,2... dado que X(0)= i. Por convención $P^0(i \mid i)= 1$ .Sea $f_{i,i}^{(n)}$ la probabilidad del primer retorno al estado i en la n-ésima transición n= 0,1,2.... donde $f_{i,i}^{(0)}= 1$. Entonces:

$$\lim_{n\to\infty} P^n(i \mid i) = \frac{1}{\sum_{n=0}^{\infty} n f_{i,i}^{(n)}}$$

b) Bajo las mismas condiciones del inciso a),

$$\lim_{n\to\infty} P^n(i \mid j) = \lim_{n\to\infty} P^n(i \mid i) \quad \text{para cualquier estado j.}$$

Si $\lim_{n\to\infty} P^n(i \mid i) > 0$ para algún estado i en una clase aperiódica y recurrente, entonces $\pi_j > 0$ para todo j en la clase de i. En este caso decimos que la clase es positiva recurrente o fuertemente ergódica, por el contrario si cada $\pi_j = 0$ y la clase es recurrente decimos que la clase es nula recurrente o débilmente ergódica.

Un método alternativo para determinar la distribución límite para una clase positiva recurrente y aperiódica está dado en el siguiente Teorema.

TEOREMA 1.4

En una clase positiva recurrente aperiódica j= 0,1,2....

$$\lim_{n\to\infty} P^n(j \mid i) = \pi_j = \sum_{i=0}^{\infty} \pi_j P(j \mid i) \quad \text{donde} \quad \sum_{i=0}^{\infty} \pi_j = 1 \ \text{ y}$$

las π's están determinadas en forma única por el conjunto de ecuaciones:

$$\pi_j \geq 0 \qquad \pi_j = \sum_{i=0}^{\infty} \pi_j P(j \mid i) \qquad \sum_{i=0}^{\infty} \pi_j = 1 \quad j= 0,1,2....$$

Cualquier conjunto $(\pi_j)_i^{\infty} = 0$ que satisface el sistema de ecuaciones anterior es llamada una distribución de probabilidad límite de la cadena de Markov.

A continuación veremos un ejemplo de la aplicación y cálculo de las probabilidades estacionarias.

EJEMPLO

Frecuentemente los sociólogos suponen que las clases sociales de generaciones sucesivas en una familia pueden ser modeladas como una cadena de Markov. De este modo la ocupación de un hijo dependerá únicamente de la ocupación de su padre y no de la de su abuelo. Supongamos que la matriz de probabilidades de transición para el modelo que refleja esta cadena de Markov es la siguiente:

		Clase social del hijo		
		ALTA	MEDIA	BAJA
	ALTA	.40	.50	.10
Clase social	MEDIA	.05	.70	.25
del padre	BAJA	.05	.50	.45

¿Qué fracción de esta población pertenecerá a la clase alta a largo plazo?

Usando el resultado que se muestra en el teorema 1.4 tenemos que para determinar la distribución límite(π_0,π_1,π_2) planteamos el siguiente sistema de ecuaciones:

$$.40\,\pi_0 + .05\pi_1 + .05\pi_2 = \pi_0 \quad \ldots\ldots (1)$$
$$.50\,\pi_0 + .70\pi_1 + .50\pi_2 = \pi_1 \quad \ldots\ldots (2)$$
$$.10\,\pi_0 + .25\pi_1 + .45\pi_2 = \pi_2 \quad \ldots\ldots (3)$$
$$\pi_0 + \pi_1 + \pi_2 = 1 \quad \ldots\ldots (4)$$

Una de las ecuaciones 1,2 ó 3 es redundante por la condición de que $\sum_j P(j|i) = 1$ .Eliminemos arbitrariamente la ec. 3 y simplificando el sistema tenemos que:

$$-60\,\pi_0 + 5\,\pi_1 + 5\pi_2 = 0 \quad \ldots.. (5)$$
$$5\,\pi_0 - 3\pi_1 + 5\pi_2 = 0 \quad \ldots.. (6)$$
$$\pi_0 + \pi_1 + \pi_2 = 1 \quad \ldots.. (7)$$

Ahora eliminamos π_2 restando la ecuación 5 primero a la ecuación 6 y después a cinco veces la ecuación 7, con lo que el sistema se reduce a :

$$65\,\pi_0 - 8\,\pi_1 = 0$$
$$65\,\pi_0 \qquad = 5$$

De donde finalmente:

$$\pi_0 = 1/13$$
$$\pi_1 = 65/\,104$$
$$\pi_2 = 1 - \pi_0 - \pi_1 = 31/104.$$

De esta forma sabemos que bajo este modelo 1/13 de la población estará en la clase alta a largo plazo.

CAPITULO II

PROCESOS MARKOVIANOS DE DECISION

INTRODUCCION

En el capítulo I comenzamos hablando de los procesos estocásticos en general, posteriormente nos avocamos a considerar los aspectos más relevantes de un tipo especial de ellos, las cadenas de Markov. Ya hemos hablado de que en un proceso estocástico están involucradas dos características fundamentales: la aleatoriedad y el tiempo. En este capítulo estudiaremos un tipo especial de proceso en el que se encuentra involucrada otra característica adicional: el factor económico.

CARACTERISTICAS DE UN PROCESO MARKOVIANO DE DECISION

Sea $\{X(t), t=0,1,2....\}$ una cadena de Markov cuyo espacio de estados $S=\{0,1,...N\}$ es un conjunto finito. En cada observación del proceso al tiempo t debe tomarse una decisión $d \in D(i)$ donde $D(i)$ es también un conjunto finito. Asociado a cada estado $i \in S$ en el que se encuentre el proceso y a la decisión $d \in D(i)$ que se elija existe una ganancia esperada conocida $r_{i,d}$ .Cuando el proceso pasa en el siguiente tiempo a un cierto estado j debe tomarse una decisión $d' \in D(j)$, ésto renueva un ciclo de eventos que se repite sin límite.

Bajo las condiciones anteriores $\{X(t), t=0,1,2....\}$es llamado un Proceso Markoviano de Decisión.

De la definición de Proceso Markoviano de Decisión (P.M.D.) se desprende que las probabilidades de transición $P(j \mid i)$ dependerán ahora no sólo de los estados i y j, sino también de la decisión elegida.

Formalmente:

$$P(X(t+1)=j \mid X(0)=i_0, X(1)=i_1 \ldots\ldots\ldots X(t)=i,\ d)=$$

$$P(X(t+1)=j \mid X(t)=i,\ d) \quad i_0, i_1, \ldots, i, j \in S \quad d \in D(i).$$

Una vez que un problema se ha modelado como un P.M.D., con lo cual como ya dijimos se encuentra involucrado el aspecto económico, lo que se busca es maximizar las ganancias esperadas, por lo que resulta muy importante determinar un criterio que nos permita elegir para cada periodo y cada estado $i \in S$ la decisión $d \in D(i)$ que nos lleve a lograr ésto.

Una política, denotada por δ, es una regla que nos dice que decisión tomar en cada tiempo t. Si δ es independiente del periodo en el cual se encuentre el proceso entonces decimos que δ es una política estacionaria.

Todos los resultados que veremos en este estudio se referirán a políticas estacionarias, para este caso δ está completamente caracterizada por la sucesión de decisiones siguiente: $\{ \delta(0), \delta(1), \ldots\ldots \delta(N) \}$ donde $\delta(i) \in D(i)$ $i= 1,2,\ldots.N$.

Asociada a cualquier cadena de Markov existe su matriz de probabilidades de transición P. En el caso de un P.M.D. no podemos hablar de una única matriz de probabilidades de transición, ya que ahora éstas están determinadas conjuntamente por la probabilidad con la que el proceso pasa de un estado a otro y por la política elegida. De esta forma tendremos distintas P_δ dependiendo de la política δ de que se trate.

Como en todo proceso estocástico, en un P.M.D. uno de los aspectos relevantes es el tiempo en el cual se desarrolla y se estudia el proceso. Avocandonos a esta característica es interesante hacer notar que el horizonte o periodo sobre el cual se estudia el proceso puede ser finito o infinito. El caso más comúnmente estudiado, que será al que nos referiremos, es que el horizonte sobre el cual se estudie el proceso y sobre el cual se deseen maximizar las ganancias sea un horizonte infinito. Existen dos criterios para evaluar las ganancias sobre dicho horizonte: las ganancias descontadas y las ganancias promedio.

Los métodos de solución de un P.M.D. que estudiaremos en este capítulo están basados en el criterio de las ganancias descontadas, por ello pasaremos a revisar brevemente en que consiste este criterio.

El criterio de las ganancias descontadas consiste en traer a valor presente las ganancias que serán recibidas en los próximos periodos (donde es importante hacer notar que todos los periodos son de igual longitud). En el modelo se encuentra involucrada una cierta tasa de interés de tal forma que representa lo mismo recibir Y unidades de ganancia al tiempo n que β^n Y unidades de ganancia al tiempo 0, donde : $\beta = 1/(1+r)$, $r > 0$ por lo que $0 < \beta < 1$.

Suponiendo que las ganancias $r_{i,d}$ son acotadas, ésto es, existen $m = \min\{r_{i,d}\}$ y $M = \max\{r_{i,d}\}$, es posible obtener una cota superior para la ganancia descontada obtenida en un número infinito de periodos, que estará dada por:

$$M + M\beta + M\beta^2 + \ldots\ldots\ldots = M(1 + \beta + \beta^2 + \ldots\ldots\ldots)$$

y usando la fórmula para una serie geométrica tenemos que:

$$= \frac{M}{1-\beta} < \infty .$$

De igual forma una cota inferior para la ganancia descontada mínima cuando $n \to \infty$ será $\frac{m}{1-\beta}$.

Comenzaremos dando algunas definiciones de ciertos conceptos generales.

Sea δ una política estacionaria arbitraria y Δ el conjunto de todas las políticas estacionarias. Asociada a cada política δ existe un vector V_δ de dimensiones N*x*1 cuya i-ésima componente se denota por $V_\delta(i)$ y se define como la ganancia esperada descontada recibida durante un número infinito de periodos, dado que el proceso comienza en el estado i y la política estacionaria que se sigue es δ. Formalmente:

$$V_\delta(i) = E_\delta\left(\sum_{t=1}^{\infty} \beta^{t-1} r_{X(t),d} \mid X(0) = i\right).$$

Se define a **f** como el vector de ganancias óptimo de dimensiones N*x*1 cuya i-ésima componente es :

$$f(i) = \max_{\delta \in \Delta} \{ V_\delta(i) \} \qquad i = 1,2,\ldots. N.$$

Ahora bien si una política δ^* tiene la propiedad de que $\forall\, i \in S$ se cumple que: $f(i) = V_{\delta^*}(i)$, es decir, δ^* maximiza las ganancias esperadas descontadas durante un número infinito de periodos, independientemente del estado inicial del proceso, entonces δ^* es una política óptima.

De las características vistas al inicio del capítulo sabemos que un P.M.D. tiene un número finito de estados y decisiones, por lo tanto el conjunto de políticas es también un conjunto finito lo que nos lleva a afirmar que el máximo se alcanza.

La existencia de una única política que simultáneamente es óptima para toda $i \in S$ no es obvia. Como vimos anteriormente si las $r_{i,d}$'s son acotadas podemos decir que:

$$\frac{m}{1-\beta} < f(i) < \frac{M}{1-\beta} \quad \text{para } i = 1,2,\ldots N$$

y ésto nos permitirá mostrar que existe una política estacionaria óptima.

A continuación estudiaremos una serie de resultados que nos permitirán afirmar la existencia de la política δ^*.

Por un momento ignoraremos el problema de optimización y estudiaremos las propiedades de una política en particular. Para ello suprimiremos el índice d de $r_{i,d}$ y de $P(j \mid i, d)$.

Sea R el vector de dimensiones **N*x*1** cuyo i-ésimo elemento es r_i. Sea P como antes la matriz de probabilidades de transición cuyo i-ésimo elemento es $P(j \mid i)$ y llamemos Q a la matriz de N*x*N definida por :

$$Q = \beta P$$

Por las propiedades de P vistas en el capítulo anterior sabemos que si el proceso comienza al tiempo 0 en el estado i es decir X(0)= i la ganancia esperada al tiempo 1 está dada por :

$\sum_{j=1}^{N} P(j \mid i)\, r_j$ y el valor presente al tiempo 0 de esta ganancia será :

$$\beta \sum_{j=1}^{N} P(j \mid i)\, r_j$$

Análogamente el valor presente de la ganancia esperada recibida en dos periodos está dada por :

$$\beta^2 \sum_{j=1}^{N} \sum_{k=1}^{N} P(k \mid i) P(j \mid k)\, r_j$$

que es precisamente el i-ésimo elemento del vector $\beta^2 P^2 R$, en otras palabras, el elemento i-ésimo del vector Q^2 **R**.

Siguiendo con el mismo razonamiento, es claro que en general el valor presente al tiempo 0 de la ganancia esperada recibida durante un cierto tiempo m es el i-ésimo elemento del vector $(\alpha P)^m R = Q^m R$.

Sea V^t el vector de dimensiones Nx1 cuyo i-ésimo elemento tiene la siguiente interpretación :

V^t (i) = la ganancia máxima esperada recibida durante los primeros t periodos dado que el estado inicial del proceso es i , es decir:

$$V^t = R + QR + ...+ Q^{t-1} R \qquad t= 0,1,2,........$$

Ahora bien, como nuestro interés se fija en el comportamiento del modelo cuyo horizonte es infinito lo que nos interesa conocer es el comportamiento de V^t cuando $t \to \infty$.

Se tiene que:

$$V^t = R + Q(R + QR+ ....+ Q^{t-2}R)$$

$$= R + QV^{t-1}$$

Pero por otro lado también:

$$V^t = (I + QR + ...+ Q^{t-1})\, R \quad ...... (2.1)$$

donde I es la matriz identidad de **dimensiones NxN.**

De la expresión anterior **es claro que el comportamiento de V^t cuando** $t \to \infty$ está determinado por el **comportamiento límite de la suma de términos Q^t donde** como ya habíamos dicho $Q = \beta P$.

TEOREMA 2.1

Sea P una matriz estocástica y sea **$Q= \beta P$ donde $0 < \beta < 1$ entonces (I-Q)** es invertible y

$$(I-Q)^{-1} = 1 + Q + Q^2 + ...$$

Demostración.

Como $Q= \beta P$ sabemos entonces que sus elementos son no negativos, por lo que también lo son los de Q^t . Del capítulo anterior sabemos que P es una matriz estocástica, ya que la suma de sus elementos por renglones es 1, entonces la suma de los renglones de $Q^t = \beta^t P^t$ suman β^t, consecuentemente los renglones de Q^t tienden a cero cuando $t \to \infty$.

Por otro lado se tiene que:

$(I-Q^t) = (I-Q)(I+Q+.......+Q^{t-1})$ .... (2.2)

Considérese un vector renglón x de dimensiones $1xN$ para el cual se cumple que $x(I-Q) = 0$.

Premultiplicando la ecuación (2.2) por x tenemos que:

$x(I-Q)(I+Q+.......+Q^{t-1}) = x(I-Q^t)$

Usando la hipótesis de que $x(I-Q) = 0$ tenemos que:

$$0 = x\,I - x\,Q^t, \text{ entonces:}$$

$$x = x\;Q^t$$

Cuando $t \rightarrow \infty$ $Q^t \rightarrow 0 \Rightarrow x = 0 \therefore$ (I-Q) es invertible, es decir, existe $(I-Q)^{-1}$.

Ahora premultiplicando la ecuación (2.2) por $(I-Q)^{-1}$ tenemos que:

$(I-Q)^{-1}(I-Q^t) = (I-Q)^{-1}(I-Q)(I+Q+.......+Q^{t-1})$

$(I-Q)^{-1} - (I-Q)^{-1}Q^t = I+Q+.......+Q^{t-1}$

Cuando $t \rightarrow \infty$ $Q^t \rightarrow 0$, por lo que :

$(I-Q)^{-1} = I+Q+Q^2+.......$ con lo cual queda demostrado el teorema. ■

El teorema 2.1 es aplicable a la suma $(I + QR + ...+ Q^{t-1})$ expresada en la ecuación 2.1 , de esta forma V^t converge cuando $t \rightarrow \infty$ lo que nos permite definir el vector V de dimensiones $Nx1$ como:

$$V = \lim_{t\rightarrow\infty}\{V^t\} = (I-Q)^{-1}R \quad ....(2.3)$$

Donde interpretamos a V(i) como la ganancia total descontada recibida en un número infinito de periodos dado que el estado inicial del proceso es i .

Por convención diremos que si w y x son dos vectores de dimensiones $Nx1$ entonces $w \geq x$ si $w_i \geq x_i$ para i= 1,2,....N. Análogamente decimos que $w \leq x$ si $w_i \leq x_i$ para i= 1,2,....N.

El siguiente teorema probará su importancia cuando en el próximo capítulo estudiemos uno de los métodos que se utilizan para determinar una política óptima, en un P.M.D.

TEOREMA 2.2

Sea P una matriz estocástica. Sea $Q = \beta P$ donde $0 < \beta < 1$ y sea $V = (I-Q)^{-1} R$. Considérese cualquier vector w de dimensiones $N \times 1$.

a) Si $R + Qw \geq w$ entonces $V \geq R + Qw$
b) Si $R + Qw \leq w$ entonces $V \leq R + Qw$.

Comentario.

Interpretamos R como un vector de ganancias, $Q = \beta P$ como una matriz de transición y $R+Qw$ como el vector de ganancias para el proceso descontado truncado en una sola transición en el cual se recibe como ganancia terminal w_j si la transición ocurre al estado j. La parte a) asegura que si este proceso truncado resulta ser mejor que w (es decir si $R + Qw \geq w$) entonces puede obtenerse una mejora futura usando el vector de ganancias R y la matriz de probabilidades de transición Q indefinidamente, es decir $V \geq R + Qw$.

Demostración.

Definimos el vector t de dimensiones $N \times 1$ como:

$$
\begin{aligned}
t &= R + Qw - w \\
  &= R - (I - Q)w \qquad \ldots\ldots (2.4)
\end{aligned}
$$

Premultiplicando (2.4) por $(I-Q)^{-1}$ y usando (2.3) se tiene que :

$$(I-Q)^{-1} t = (1-Q)^{-1} R - w$$

$$(I-Q)^{-1} t = V - w \qquad \ldots\ldots (2.5)$$

Por hipótesis se tiene que $t \geq 0$, además sabemos que también $Q \geq 0$, por lo que usando el Teorema 2.1 se tiene que:

$$(I-Q)^{-1} t = \sum_{i=0}^{\infty} Q^i t \geq Q^0 t = t \qquad \ldots (2.6)$$

De la combinación de (2.5) y (2.6) tenemos que:

$V - w \geq t \Rightarrow V \geq w + t$ y usando (2.4) finalmente se concluye que:

$$V \geq R + Qw$$

Con lo cual queda probado el inciso a) del teorema. Para demostrar la parte b) el procedimiento es análogo pero el sentido de las desigualdades será al revés. ■

Con los resultados vistos hasta este momento estamos en posición de retomar nuevamente el problema de la optimización, es decir, el problema de determinar una política δ^* que sea óptima para i=1,...N.

De esta forma el índice d aparece nuevamente en $r_{i,d}$ así como en $P(j \mid i, d)$ y en $Q(j \mid i, d) = \beta P(j \mid i,d)$.

También δ aparece en el vector de dimensiones Nx1 R_δ y en la matriz de dimensiones NxN P_δ, donde el i-ésimo elemento de R_δ es $r_{i,\delta(i)}$ y donde el i-ésimo elemento de P_δ es $P(j \mid i, \delta(i))$.

Similarmente δ aparece como subíndice en los vectores V_δ^t y V_δ y en la matriz de dimensiones NxN $Q_\delta = \beta P_\delta$.

Ya que P es una matriz estocástica y que $0< \beta < 1$ la matriz $(I-Q_\delta)^{-1}$ es invertible y de acuerdo a la ecuación 2.3 tenemos que:

$$V_\delta = (I-Q_\delta)^{-1} R_\delta \quad \ldots\ldots\ldots (2.7)$$

Donde la i-ésima componente de V_δ denotada por $V_\delta(i)$ se interpreta, como antes, como la ganancia esperada descontada recibida durante un número infinito de periodos dado que el estado inicial del proceso es i y la política que se sigue es δ.

Premultiplicando la ecuación (2.7) por $(I-Q_\delta)$ vemos que V_δ es la única solución de la ecuación:

$$V_\delta = R_\delta + Q_\delta V_\delta \quad \ldots\ldots\ldots (2.8)$$

La ecuación (2.8) se conoce como ecuación de evaluación de política.

En el siguiente capítulo estudiaremos los tres métodos más comúnmente usados para determinar una política δ^* que nos permita en cada estado elegir la decisión que maximice las ganancias esperadas.

CAPITULO III

METODOS DE SOLUCION DE PROCESOS MARKOVIANOS DE DECISION

INTRODUCCION

En este capítulo estudiaremos los lineamientos básicos de tres de los métodos más comúnmente usados para obtener una política óptima δ^* en un P.M.D. Estos métodos son: i) mejoramiento de política, ii) aproximaciones sucesivas y iii) programación lineal.

También se presentarán una serie de teoremas que nos permitirán afirmar la existencia de δ^*. En particular, como se verá más adelante, el algoritmo de mejoramiento de política es la demostración constructiva de que es posible encontrar una política δ^* en un número finito de iteraciones.

1) MEJORAMIENTO DE POLITICA

El algoritmo que estudiaremos será el de mejoramiento de política propuesto inicialmente por Howard en 1960, que permite encontrar una política estacionaria óptima en un problema que pueda ser modelado como un P.M.D.

1) INICIALIZACION.
Elegir cualquier política $\delta \in \Delta$.

2) EVALUACION DE LA POLITICA.
Sea $t_i = 0$ para i= 1,2,... N. Calcular $V_\delta = (I-Q_\delta)^{-1} R_\delta$

3) MEJORAMIENTO DE POLITICA
Para cada i= 1,2,... N y para cada $d \in D(i)$ calcular:

$$t_{i,d} = r_{i,d} + \sum_{j=1}^{N} Q(j \mid i,d) V_\delta(j) - V_\delta(i)$$

Si $t_{i,d} > t_i$ reemplazar la decisión $\delta(i)$ por d y el valor de t_i por el de $t_{i,d}$.

4) Si $t_i = 0$ para i=1,2, ... N TERMINA. δ es una política óptima.
En otro caso regresar al paso 2.

Como vimos al inicio del capítulo anterior una política óptima δ^* debe cumplir que $V_{\delta^*}(i) = f(i)$ para i= 1,2,... N. El siguiente teorema demuestra que la política óptima δ^* encontrada usando el método de mejoramiento de política cumple la condición anterior.

TEOREMA 3.1

El método de mejoramiento de política termina en un número finito de iteraciones y la última política δ^* evaluada en el paso 2 cumple que $V_{\delta^*} = \mathbf{f}$.

Demostración.

Por convención diremos que al comparar dos vectores w y x escribimos $w > x$ si $w \geq x$ y se cumple que $w_i > x_i$ para al menos una i. Similarmente $w < x$ si $w \leq x$ pero al menos para una i se cumple que $w_i < x_i$.

Consideremos una iteración no terminal del paso 4 del algoritmo de mejoramiento de política. Sea δ la política evaluada en la ejecución precedente del paso 2 y sea δ' la política que será evaluada en la próxima ejecución del paso 2.

Sea t el vector de dimensiones Nx1 cuya i-ésima componente es el valor actual de t_i .Ya que t fue inicial izado en cero y sólo cambia su valor por el de $t_{i,d}$ cuando $t_{i,d} > 0$, se cumple que:

$$0 < t = R_{\delta'} + Q_{\delta'} V_\delta - V_\delta \Rightarrow V_\delta < R_{\delta'} + Q_{\delta'} V_\delta$$

Por otro lado aplicando el inciso a) del teorema 2.2 con $w = V_\delta$, $R = R_{\delta'}$ y $Q = Q_{\delta'}$, se tiene que:

$$V_{\delta'} \leq R_{\delta'} + Q_{\delta'} V_\delta$$

$$\therefore \quad V_\delta < V_{\delta'}$$

Es decir, cada iteración no terminal del paso 2 del algoritmo mejora las iteraciones anteriores.

Como sabemos existe un número finito de políticas, por lo tanto el algoritmo termina en un número finito de pasos.

Sea δ^* la última política evaluada. Ya que al terminar el algoritmo se cumple que $t_{i,d} \leq 0$ para i= 1,2,... N , usando notación matricial se cumple la siguiente desigualdad:

$$R_\delta + Q_\delta V_{\delta^*} - V_{\delta^*} \leq 0 \Rightarrow V_{\delta^*} \geq R_\delta + Q_\delta V_{\delta^*}$$

Por otro lado aplicando el inciso b) del teorema 2.2 con $w = V_{\delta^*}$, $R = R_\delta$ y $Q = Q_\delta$, se cumple para cualquier política δ que:

$$V_\delta \leq R_\delta + Q_\delta V_{\delta^*}$$

$$\therefore \quad V_\delta \leq V_{\delta^*}$$

Ya que esta última desigualdad es válida para cada política δ, y como recordamos:

$$f = \max_{\delta \in \Delta} \{V_\delta\} \text{ ,entonces } V_{\delta^*} = f \text{ .}$$

$\therefore$ δ^* es una política óptima. ■

El siguiente método de solución de un P.M.D. que estudiaremos, denominado aproximaciones sucesivas, está basado en el uso de la programación dinámica y ya que la ecuación funcional o ecuación de optimalidad es la base de cualquier algoritmo de programación dinámica, nos va a resultar muy importante el siguiente teorema que relaciona a nuestro vector de ganancias óptimas **f** en términos de una ecuación funcional. Se incluye en esta sección porque en su demostración se aplican los resultados que acabamos de revisar.

TEOREMA 3.2

El vector de ganancias óptimas **f** es la única solución a la siguiente ecuación funcional:

$$f(i) = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) f(j) \right\} \quad i = 1,2,\ldots. N \quad \ldots. (3.1)$$

Demostración.

Para demostrar porque **f** satisface la ecuación (3.1) analizaremos la condición bajo la cual el algoritmo de mejoramiento de política termina: el algoritmo termina cuando se encuentra una política δ^* para la cual se cumple que:

$$0 = t_{i,\delta^*(i)} = \max_{d \in D(i)} \{t_{i,d}\} = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) V_\delta(j) - V_\delta(i) \right\} \quad i = 1,2,\ldots. N$$

Por otro lado el teorema 3.1 muestra que $V_{\delta^*} = \mathbf{f}$, usando esta igualdad y la condición anterior tenemos que los valores de las $t_{i,d}$'s cuando el algoritmo termina están dados por:

$$0 = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) f(j) - f(i) \right\} \quad i = 1,2,\ldots. N$$

Sumando f(i) a ambos lados de la ecuación (3.1) se tiene que:

$$f(i) = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) f(j) \right\} \quad i = 1,2,\ldots. N$$

Con lo cual queda demostrado que f cumple la ecuación (3.1).

Para demostrar la unicidad supongamos que existe un vector **g** que también cumple la ecuación (3.1).

Para cada estado i sea δ^* (i) la decisión que maximiza el lado derecho de la ecuación (3.1), entonces, en términos de notación matricial, se cumple que:

$$\mathbf{g} = R_{\delta^*} + Q_{\delta^*}\,\mathbf{g}$$

Por otro lado supongamos que existiera $\delta \in \Delta$ tal que:

$$\mathbf{g} \leq R_{\delta} + Q_{\delta}\,\mathbf{g}$$

Por el inciso a) del teorema 2.2 se cumpliría que:

$$V_{\delta} \geq R_{\delta} + Q_{\delta}\,\mathbf{g} \geq \mathbf{g} \Rightarrow V_{\delta} \geq \mathbf{g}$$

que sería una contradicción porque $\mathbf{g}$ no sería máximo.

$$\Rightarrow \quad \mathbf{g} \geq R_{\delta} + Q_{\delta}\,\mathbf{g}$$

por el inciso b) del teorema 2.2 se tiene que:

$$V_{\delta} \leq R_{\delta} + Q_{\delta}\,\mathbf{g}$$

$$\Rightarrow V_{\delta} \leq g \quad \therefore \; g = \max_{\delta \in \Delta}\{V_{\delta}\} = f.$$

Con lo cual queda completamente **demostrado el teorema.** ■

2)APROXIMACIONES SUCESIVAS

Este método, como ya se dijo, está basado en el uso de la programación dinámica, pero en el campo estocástico, se emplea el principio de la ecuación recursiva de Bellman que permite optimizar cada etapa del problema por separado.

En este caso, el método de aproximaciones sucesivas busca aproximar el valor de f(i), que como sabemos es la ganancia máxima esperada descontada recibida durante un número infinito de periodos dado que el estado inicial del proceso es i , con el valor de V^t (i) que como ya había sido dicho es la ganancia máxima esperada descontada recibida durante los primeros t periodos dado que el estado inicial del proceso es i, para i= 1,2,.... N.

El método consiste en comenzar con un vector arbitrario V^0 de dimensiones **Nx1** y calcular los vectores V^1, V^2 , ... también de dimensiones **Nx1** usando la siguiente ecuación recursiva :

$$V^t(i) = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) V^{t-1}(j) \right\} \qquad i = 1,2,....N \quad ....(3.2)$$

Se probará que V^t aproxima a **f** cuando $n \to \infty$. Para ello introduciremos el concepto de la norma de un vector. Sea u de dimensiones **Nx1**, definimos la norma del vector u como el máximo de los valores absolutos de sus entradas, es decir :

$$\| u \| = \max_i \{ | u_i | \} \quad ..... (3.3)$$

El siguiente teorema muestra que V^t está mas cerca de **f** (en esta norma) de lo que lo está V^{t-1} .

TEOREMA 3.3

Para t=1,2,....

$$\| V^t - f \| \leq \beta \| V^{t-1} - f \|$$

Demostración.

Sea d tal que maximiza el lado derecho de (3.2).

$$V^t(i) = r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) V^{t-1}(j) \quad ....(3.4)$$

Del teorema 3.2 sabemos que se cumple que:

$$f(i)=\max_{d\in D(i)}\left\{r_{i,d}+\beta\sum_{j=1}^{N}P(j\,|\,i,d)f(j)\right\}\qquad i=1,2,\ldots.N.$$

De donde se cumple que:

$$f(i)\geq r_{i,d}+\beta\sum_{j=1}^{N}P(j\,|\,i,d)f(j)\qquad \ldots..(3.5)$$

Restando (3.5) de (3.4) y agrupando términos semejantes tenemos que:

$$V^{t}(i)-f(i)\leq\beta\sum_{j=1}^{N}P(j\,|\,i,d)(V^{t-1}(j)-f(j))$$

De (3.3) sabemos que:

$$V^{t}(i)-f(i)\leq\left|\,V^{t-1}(i)-f(i)\right|\leq\left\|\,V^{t-1}(i)-f(i)\right\|$$

Por lo que:

$$V^{t}(i)-f(i)\leq\beta\sum_{j=1}^{N}P(j\,|\,i,d)\left\|\,V^{t-1}(i)-f(i)\,\right\|$$

$$\Rightarrow V^{t}(i)-f(i)\leq\beta\left\|\,V^{t-1}(i)-f(i)\,\right\|\qquad \ldots..(3.6)$$

Sea d que maximiza el lado derecho de (3.1):

$$f(i)=r_{i,d}+\beta\sum_{j=1}^{N}P(j\,|\,i,d)f(j)\qquad \ldots..(3.7)$$

De (3.2) se cumple que:

$$V^{t}(i)\geq r_{i,d}+\beta\sum_{j=1}^{N}P(j\,|\,i,d)V^{t-1}(j)\qquad \ldots..(3.8)$$

Restando (3.8) de (3.7) y repitiendo el análisis precedente tenemos que:

$$f(i)-V^{t}(i)\leq\beta\left\|\,f-V^{t-1}\,\right\|\qquad \ldots..(3.9)$$

Combinando los resultados obtenidos en 3.6 y 3.9 tenemos que:

$$\left|\,V^{t}(i)-f(i)\,\right|\leq\beta\left\|\,V^{t-1}-f\,\right\|$$

Maximizando sobre i se cumple que:

$$\| V - f \| \leq \beta \| V^{t-1} - f \|$$

Con lo cual queda demostrado el teorema. ■

Aplicando repetidas veces el teorema 3.3 se cumple la siguiente desigualdad:

$$\| V^t - f \| \leq \beta^t \| V^0 - f \| \quad .....(3.10)$$

De donde también se cumple que:

$$| V^t(i) - f(i) | \leq \beta^t \| V^0 - f \| \text{ para } i = 1,2,.....N \quad .....(3.11)$$

Cuando $t \to \infty$ $\beta^t \to 0$, por lo que (3.11) muestra que $V^t(i) \to f(i)$.

Ya que esta convergencia está dada precisamente en términos de **f** que como sabemos es desconocida, veremos el siguiente resultado que nos da una cota que indica en cada iteración que tan lejos nos encontramos de **f**.

COROLARIO

Suponga que $V^0 = 0$ .Sean como antes $m = \min_{i,d}\{r_{i,d}\}$ y $M = \max_{i,d}\{r_{i,d}\}$.

Entonces:

$$\| V^t - f \| \leq \frac{\beta^t}{1-\beta} \max\{ | m |, | M | \}$$

Demostración.

De (3.10) se tiene que :

$$\| V^t - f \| \leq \beta^t \| V^0 - f \|$$

Pero por hipótesis se tiene que $V^0 = 0$ de esta manera:

$$\| V^t - f \| \leq \beta^t \| f \|$$

$$= \beta^t \max_i \{ | f(i) | \} \quad .....(3.12)$$

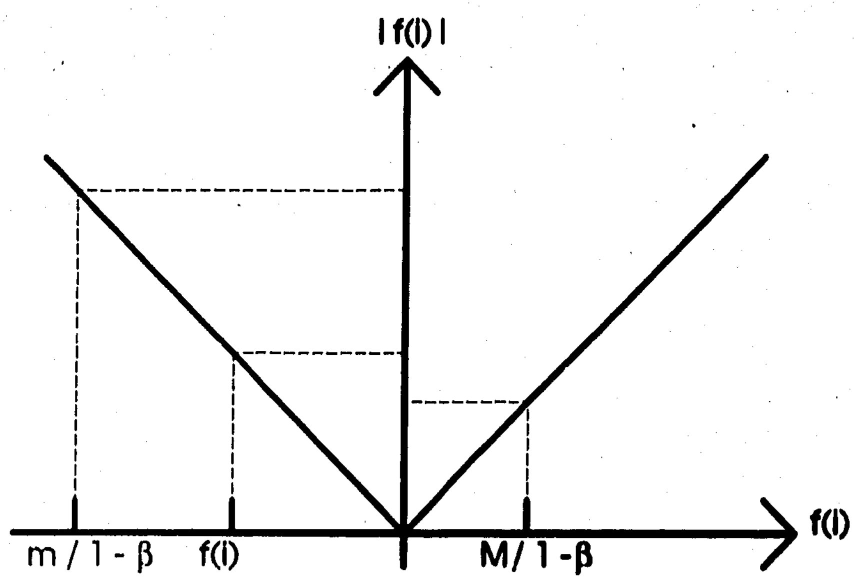
Sea k que maximiza (3.12), entonces:

$$= \beta^t \mid f(k) \mid$$

De acuerdo a lo visto al inicio del capítulo sabemos que:

$$\frac{m}{1-\beta} < f(i) < \frac{M}{1-\beta} \quad \text{para } i = 1,2,\ldots. N.$$

Para entender como se comporta $\mid f(i) \mid$ veremos como es la gráfica de la función valor absoluto:



De esta forma tenemos que se cumple que:

$$\beta^t \mid f(k) \mid \leq \beta^t \max\left\{ \frac{\mid m \mid}{1-\beta}, \frac{\mid M \mid}{1-\beta} \right\}$$

$$= \frac{\beta^t}{1-\beta} \max\{ \mid m \mid, \mid M \mid \}.$$

Con lo que finalmente queda demostrado el corolario. ■

3)PROGRAMACION LINEAL

El uso de un algoritmo de programación lineal para la obtención de una política óptima en un P.M.D. para el caso en que se emplea el criterio de las ganancias descontadas fue propuesto por d'Epenoux.

En esta sección se verá que el vector de ganancias óptimo **f** puede obtenerse al resolver el problema de programación lineal presentado en el siguiente teorema.

La razón del planteamiento de dichas restricciones en particular, surge de la ecuación funcional planteada en el teorema 3.2, y la función objetivo que se usa es minimizar la suma de las componentes del vector de variables de decisión, porque ello resulta ser lo mejor, como se verá en la demostración del teorema.

TEOREMA 3.4

f es una solución factible del siguiente problema de programación lineal:

$$(P) \qquad \min \sum_{i=1}^{N} f^{\wedge}(i)$$

$$\text{s.a.} \quad f^{\wedge}(i) - \beta \sum_{j=1}^{N} P(j \mid i,d) f^{\wedge}(j) \geq r_{i,d} \qquad i = 1,2,\ldots N, \; d \in D(i)$$

$$f^{\wedge}(i) \quad \text{s.r.s.} \qquad\qquad i = 1,2,\ldots N$$

Además la función objetivo se minimiza únicamente cuando $f^{\wedge} = \mathbf{f}$.

Demostración.

Para verificar que **f** es una solución factible de P, debe cumplir las restricciones, lo cual es inmediato del teorema 3.2, ya que:

$$f(i) \geq r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) f(j) \qquad i = 1,2,\ldots N, \; d \in D(i)$$

$$\Rightarrow \quad f(i) - \beta \sum_{j=1}^{N} P(j \mid i,d) f(j) \geq r_{i,d} \qquad i = 1,2,\ldots N, \; d \in D(i)$$

y como sabemos **f** s.r.s. $\therefore$ **f** es una solución factible de P.

Por otro lado, por el teorema 3.1 sabemos que existe una política óptima δ^*.

Consideremos cualquier solución factible $f^\wedge$ del problema P. Entonces, en términos de notación matricial, sabemos que se da la siguiente desigualdad:

$$f^\wedge - Q_{\delta *} f^\wedge \geq R_{\delta *}$$

$$\Rightarrow f^\wedge \geq R_{\delta *} + Q_{\delta *} f^\wedge$$

del inciso b) del teorema 2.2 tenemos que:

$$V_{\delta *} \leq f^\wedge$$

es decir:

$$V_{\delta *}(i) \leq f^\wedge(i) \qquad i = 1,2,\ldots N.$$

Por lo que al realizar la suma de las componentes se cumple que:

$$\sum_{i=1}^{N} V_{\delta *}(i) \leq \sum_{i=1}^{N} f^\wedge(i)$$

$$\Rightarrow \sum_{i=1}^{N} f^\wedge(i) \text{ se minimiza cuando } f^\wedge(i) = V_{\delta *}(i), i = 1,2,\ldots N$$

pero del teorema 3.1 sabemos que $V_{\delta *} = f$,por lo que la función objetivo de P se minimiza cuando $f^\wedge(i) = f(i)$. ■

A continuación explicaremos la relación entre una base factible del problema dual de P y una política δ. Para ello recordaremos algunos conceptos de programación lineal.

Dado el siguiente problema de programación lineal (PPL):

$$\begin{aligned} &\text{Min} \quad cx \\ &\text{s.a.} \quad Ax \geq b \\ &\qquad\quad x \ \text{s.r.s.} \end{aligned}$$

donde x es un vector de dimensiones $Nx1$, c es un vector de dimensiones $1xN$, A es una matriz de dimensiones MxN y b es un vector de dimensiones $Mx1$.

Su problema dual es de la forma:

$$\begin{aligned} &\text{Max} \quad yb \\ &\text{s.a.} \quad yA = c \\ &\qquad\quad y \geq 0 \end{aligned}$$

Donde y es un vector renglón de **1xM** y A,b y c como antes.

En nuestro caso plantearemos explícitamente el problema dual de P.

Teníamos que:

$$(P) \qquad \min \sum_{i=1}^{N} f^{\wedge}(i)$$

$$\text{s.a.} \quad f^{\wedge}(1) - \beta \sum_{j=1}^{N} P(j \mid 1,d) f^{\wedge}(j) \geq r_{1,d} \qquad \forall \; d \in D(1)$$

$$f^{\wedge}(2) - \beta \sum_{j=1}^{N} P(j \mid 2,d) f^{\wedge}(j) \geq r_{2,d} \qquad \forall \; d \in D(2)$$

$$f^{\wedge}(N) - \beta \sum_{j=1}^{N} P(j \mid N,d) f^{\wedge}(j) \geq r_{N,d} \qquad \forall \; d \in D(N)$$

Si nos fijamos en cada una de las restricciones anteriores, por ejemplo, en la primera vemos que para cada $d \in D(1)$ tenemos una restricción, por lo que explícitamente el número total de restricciones de P, que denotaremos por M, está dado por:

$$M = \sum_{i=1}^{N} \# D(i) = \sum_{i=1}^{N} m_i$$

Agrupando términos semejantes tenemos que:

$$\text{Para } i=1 \quad \left[1 - \beta P(1 \mid 1,d) \right] f^{\wedge}(1) - \beta \sum_{j \neq 1} P(j \mid 1,d) f^{\wedge}(j) \geq r_{1,d} \qquad \forall \; d \in D(1)$$

$$\text{Para } i=2 \quad \left[1 - \beta P(2 \mid 2,d) \right] f^{\wedge}(2) - \beta \sum_{j \neq 2} P(j \mid 2,d) f^{\wedge}(j) \geq r_{2,d} \qquad \forall \; d \in D(2)$$

$$\text{Para } i=N \quad \left[1 - \beta P(N \mid N,d) \right] f^{\wedge}(N) - \beta \sum_{j \neq N} P(j \mid N,d) f^{\wedge}(j) \geq r_{N,d} \qquad \forall \; d \in D(N)$$

De la definición de problema dual se desprende que el número de restricciones del problema primal (M) corresponde al número de variables del problema dual y por otro lado el número de variables del problema primal (N) equivale al número de restricciones del problema dual.

Denotemos por $X_{i,d}$ i= 1,2, ... N , d ∈ D(i), a las variables de decisión del problema dual (se observa que tenemos precisamente M variables). Entonces tenemos que la función objetivo del problema dual es de la forma:

$$\max \sum_{i=1}^{N} \sum_{d\in D(i)} r_{i,d} X_{i,d} .$$

Ahora pasemos a plantear las restricciones del problema dual que, como ya hemos dicho, serán precisamente N.

Para cada i , i=1,2,... N tenemos m_i variables. Denotemos por $X_{1,d}$ a la forma general (para un cierto d ∈ D(1)) de las variables del primer grupo, $X_{2,d}$ a la del segundo grupo y así sucesivamente hasta $X_{N,d}$ a la del N-ésimo grupo.

La primera restricción es de la forma:

$$\sum_{d\in D(1)} \left[1-\beta P(1|1,d) \right] X_{1,d} - \sum_{d\in D(2)} \beta P(1|2,d) X_{2,d} \ldots\ldots - \sum_{d\in D(N)} \beta P(1|N,d) X_{N,d} = 1$$

Usando la delta de Kronecker la ecuación anterior puede expresarse como:

$$\sum_{i=1}^{N} \sum_{d\in D(1)} \left(\delta_{i,1} - \beta P(1|i,d)\right) X_{i,d} = 1$$

Así cada una de las N restricciones. Usando notación matricial podemos reescribirla como:

$$\sum_{d\in D(1)} X_{i,d} - \sum_{i=1}^{N} \sum_{d\in D(1)} Q(1|i,d)\, X_{i,d} = 1$$

Por lo que finalmente el problema dual queda expresado como:

$$\text{Max} \sum_{i=1}^{N} \sum_{d\in D(1)} r_{i,d}\, X_{i,d}$$

s.a.

$$\sum_{d\in D(j)} X_{j,d} - \sum_{i=1}^{N} \sum_{d\in D(j)} Q(j|i,d)\, X_{i,d} = 1 \qquad j = 1,2,\ldots.N.$$

Ahora ya tenemos los elementos necesarios para plantear la relación entre una base factible del problema dual de P y una política δ. Dicha relación está dada en el siguiente teorema.

TEOREMA 3.5

Una matriz **B** es una base factible para el problema dual si y sólo sí $B=(I-Q_\delta)$ para alguna política δ.

Demostración.

$\Leftarrow$) Sea δ cualquier política. Definimos el vector X_δ de dimensiones $1xN$ como:

$$X_\delta = 1^t(I-Q_\delta)^{-1},$$

donde 1 es el vector de dimensiones $Nx1$ de 1's y 1^t es su vector transpuesto.

Recordamos del teorema 2.1 que $(I-Q)$ es invertible y que $(I-Q_\delta)^{-1} \geq 0$ por lo que se tiene que:

$$X_\delta\ (I-Q_\delta) = 1^t \quad \text{con} \quad X_\delta \geq 0 \quad ......(3.13)$$

Comparando la ecuación 3.13 con las restricciones del dual y haciendo las (M-N) variables restantes iguales a cero, construimos una solución X factible donde:

$$X = (X_\delta, 0)$$

Ya que $(I-Q_\delta)$ es invertible $B=(I-Q_\delta)$ es de rango completo, por lo tanto B es una base factible.

$\Rightarrow$) Ahora consideremos cualquier base factible para el problema dual. Sea el vector X de dimensiones $1xM$ una solución básica factible y X_B de dimensiones $1xN$ el vector de variables básicas, es decir $X_B B = 1^t$ con $X \geq 0$.

Ya que el problema dual tiene N restricciones hay a lo más N elementos positivos en el vector X.

Pero por otro lado de las restricciones del problema dual se tiene que:

$$\sum_{d \in D(i)} X_{i,d} \geq 1 \quad i=1,2,....N.$$

Con lo cual en el vector X hay al menos N variables estrictamente positivas.

∴ el vector X tiene exactamente N componentes positivas, ésto nos lleva a afirmar que hay exactamente una d para la cual $X_{i,d}$ es positivo. ■

De esta forma al resolver el problema dual por el método simplex y encontrar la solución óptima, aplicando el teorema 3.5 podemos identificar cuál es la política óptima, que como ya vimos estará asociada a los N-elementos positivos del vector óptimo.

Por otro lado tenemos que cada variable del problema dual está asociada a una restricción del primal y aplicando el Teorema de las holguras complementarias sabemos que si la variable dual óptima es positiva la restricción asociada en el primal tendrá holgura cero, es decir, se cumplirá como igualdad.

Como recordamos, las restricciones del problema primal son de la forma:

$$f^\wedge(i) - \beta \sum_{j=1}^{N} P(j \mid i,d) f^\wedge(j) \geq r_{i,d} \quad i = 1,2,\dots N, \ d \in D(i)$$

Por lo tanto si la variable dual óptima X_{i,d^*} es positiva, entonces la restricción asociada en el primal se cumplirá como igualdad, es decir:

$$f(i) - \beta \sum_{j=1}^{N} P(j \mid i,d^*) f(j) = r_{i,d} \qquad d^* \in D(i) \quad i = 1,2,\dots.N.$$

despejando :

$$f(i) = r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d^*) f(j) \qquad d^* \in D(i) \quad i = 1,2,\dots.N.$$

Esto nos da un sistema de ecuaciones de NxN, que al resolverse permite encontrar la solución óptima del primal, es decir $f^\wedge = \mathbf{f}$.

CAPITULO IV

EJEMPLOS

EJEMPLO 4.1 ENVIO DE CATALOGO DE UNA CASA COMERCIAL

En una casa comercial se está pensando en la posibilidad de enviar o no un catálogo de compra a sus clientes.

Los clientes pueden ser clasificados dentro de dos posibles tipos:

Tipo	Característica
1	No han realizado ninguna compra recientemente
2	Sí han hecho compras recientemente(últimos 6 meses)

Existen tres distintas opciones para el envío del catálogo, las cuales tienen diferente costo:

Opción	Descripción	Costo de envío
DE	Enviar catálogo con descuento especial	N$ 0.5
NDE	Enviar catálogo pero sin descuento especial	N$ 0.5
NE	No enviar nada	N$ 0

Si el cliente decide comprar, el monto esperado de la ganancia para la casa comercial es de :

N$ 8 si se le envió catálogo sin descuento especial
N$ 3 si se le envió catálogo con descuento especial

Llamaremos $r_{i,j,d}$ a estas ganancias, por ejemplo $r_{1,2,DE}$ es la ganancia obtenida cuando un cliente pasa de ser tipo 1 a 2, dado que se le envió catálogo con descuento especial.

Este envío de catálogos lo realiza la casa comercial cada 6 meses. Se supone que las ventas ocurren al final de cada 6 meses y que β el factor de descuento para ese periodo es 0.9.

La experiencia de la casa comercial le ha permitido obtener las probabilidades con las que un cliente tipo 1 pasa a ser tipo 2 y viceversa y es posible afirmar que esta transición depende de la decisión tomada por la casa comercial respecto al envío o no de catálogo (es decir, DE, NDE ó NE). Dichas probabilidades están agrupadas en las siguientes matrices de probabilidades de transición:

$$P(DE)= \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.823 & 0.167 \\ 0.444 & 0.556 \end{bmatrix} \end{array}$$

$$P(NDE)= \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.922 & 0.078 \\ 0.722 & 0.278 \end{bmatrix} \end{array}$$

$$P(NE)= \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.999 & 0.011 \\ 0.778 & 0.222 \end{bmatrix} \end{array}$$

La ganancia esperada descontada $r_{i,d}$ está dada por :

$$r_{i,d} = \beta \sum_{j=1}^{2} P(j \mid i,d) r_{i,j,d} - c(i,d)$$

donde c(i,d) es el costo en que incurre la casa comercial al enviar o no el catálogo a un cliente tipo i , dado que ha tomado la decisión d.

Usando la ecuación anterior comenzaremos por determinar las ganancias esperadas descontadas para cada estado y cada decisión.

$i = 1$

d = DE

$$r_{1,DE} = 0.9\left[P(1 \mid 1, DE)*0 + P(2 \mid 1, DE)*3\right] - 0.5$$

$$= -0.05$$

d = NDE

$$r_{1,NDE} = 0.9\left[P(1 \mid 1, NDE)*0 + P(2 \mid 1, NDE)*8\right] - 0.5$$

$$= 0.06$$

d = NE

$$r_{1,NE} = 0.9\left[P(1 \mid 1, NE)*0 + P(2 \mid 1, NE)*8\right] - 0$$

$$= 0.08$$

$i = 2$

d = DE

$$r_{2,DE} = 0.9\left[P(1 \mid 2, DE)*0 + P(2 \mid 2, DE)*3\right] - 0.5$$

$$= 1.0$$

d = NDE

$$r_{2,NDE} = 0.9\left[P(1 \mid 2, NDE)*0 + P(2 \mid 2, NDE)*8\right] - 0.5$$

$$= 1.5$$

d = NE

$$r_{2,NE} = 0.9\left[P(1 \mid 2, NE)*0 + P(2 \mid 2, NE)*8\right] - 0$$

$$= 1.6$$

Ahora resolveremos el problema de determinar una política óptima δ^* usando para ello cada uno de los métodos expuestos en el capítulo 3.

i) MEJORAMIENTO DE POLITICA

1) INICIALIZACION

Elegir cualquier política $\delta = (\delta(1), \delta(2))$. Por ejemplo:

δ= (cliente1= DE, cliente2= NDE).

2) EVALUACION DE LA POLITICA

Sea $t = (t_1, t_2) = (0, 0)$. Calcular $V_\delta = (I-Q_\delta)^{-1} R_\delta$.

En nuestro ejemplo:

$$V_\delta = \begin{bmatrix} 0.25 & -0.15 \\ -0.65 & 0.75 \end{bmatrix}^{-1} \begin{bmatrix} -0.05 \\ 1.5 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 8.33 & 1.67 \\ 7.22 & 2.78 \end{bmatrix} \begin{bmatrix} -0.05 \\ 1.5 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 2.09 \\ 3.81 \end{bmatrix}$$

3) **Para i=1,2 y cada $d \in D(i)$ calcular :**

$$t_{i,d} = r_{i,d} + \sum_{j=1}^{2} Q(j \mid i, d) V_\delta(j) - V_\delta(i)$$

Si $t_{i,d} > t_i$ reemplazar la decisión $\delta(i)$ por d y a $t_{i,d}$ por t_i.

i = 1

d = DE

$$t_{1,DE} = -0.05 + [0.75(2.09) + 0.15(3.81)] - 2.09$$

$t_{1,DE} = 2.09 - 2.09 = 0$

$t_{1,DE} \leq t_1 \quad \Rightarrow$ No hay reemplazo.

d = NDE

$t_{1,NDE} = 0.06 + [0.83(2.09) + 0.07(3.81)] - 2.09$

$t_{1,NDE} = 2.06 - 2.09 = -0.03$

$t_{1,NDE} < t_1 \quad \Rightarrow$ No hay reemplazo.

d = NE

$t_{1,NE} = 0.08 + [0.89(2.09) + 0.01(3.81)] - 2.09$

$t_{1,NE} = 1.98 - 2.09 = -0.11$

$t_{1,NE} < t_1 \quad \Rightarrow$ No hay reemplazo.

i= 2

d = DE

$t_{2,DE} = 1.0 + [0.4(2.09) + 0.5(3.81)] - 3.81$

$t_{2,DE} = 3.74 - 3.81 = -0.07$

$t_{2,DE} < t_2 \quad \Rightarrow$ No hay reemplazo.

d = NDE

$t_{2,NDE} = 1.5 + [0.65(2.09) + 0.25(3.81)] - 3.81$

$t_{2,NDE} = 3.81 - 3.81 = 0$

$t_{2,NDE} \leq t_2 \quad \Rightarrow$ No hay reemplazo.

d = NE

$t_{2,NE} = 1.6 + [0.7(2.09) + 0.2(3.81)] - 3.81$

$t_{2,NE} = 3.83 - 3.81 = 0.02$

$t_{2,NE} > t_2 \quad \Rightarrow$ $\delta(2)$ es ahora reemplazada por NE y $t_2 = 0$ es reemplazado por 0.02.

4) Si $(t_1, \ t_2) = (0,0)$ TERMINA. δ es una política óptima. En otro caso regresar al paso 2.

Como $(t_1, \ t_2) = (0, 0.02)$ regresamos al paso 2.

2) EVALUACION DE LA POLITICA.

Ahora δ= (DE, NE).

Sea $t_1 = 0$ y $t_2 = 0$. Calcular $V_\delta = (I-Q_\delta)^{-1} R_\delta$.

Tenemos que:

$$V_\delta = \begin{bmatrix} 0.25 & -0.15 \\ -0.7 & 0.8 \end{bmatrix}^{-1} \begin{bmatrix} -0.05 \\ 1.6 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 8.42 & 1.58 \\ 7.37 & 2.63 \end{bmatrix} \begin{bmatrix} -0.05 \\ 1.6 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 2.11 \\ 3.84 \end{bmatrix}$$

3) Para i=1,2 y cada $d \in D(i)$ calcular :

$$t_{i,d} = r_{i,d} + \sum_{j=1}^{2} Q(j \mid i,d) V_\delta(j) - V_\delta(i)$$

Si $t_{i,d} > t_i$ reemplazar la decisión $\delta(i)$ por d y a $t_{i,d}$ por t_i .

i = 1

d = DE

$t_{1,DE} = -0.05 + [0.75(2.11) + 0.15(3.84)] - 2.11$

$t_{1,DE} = 2.11 - 2.11 = 0$

$t_{1,DE} \le t_1 \Rightarrow$ No hay reemplazo.

d = NDE

$t_{1,NDE} = 0.06 + [0.83(2.11) + 0.07(3.84)] - 2.11$

$$t_{1,NDE} = 2.08 - 2.11 = -0.03$$

$t_{1,NDE} < t_1 \quad \Rightarrow$ No hay reemplazo.

d = NE

$$t_{1,NE} = 0.08 + [0.89(2.11) + 0.01(3.84)] - 2.11$$

$$t_{1,NE} = 1.99 - 2.11 = -0.12$$

$t_{1,NE} < t_1 \quad \Rightarrow$ No hay reemplazo.

i= 2

d = DE

$$t_{2,DE} = 1.0 + [0.4(2.11) + 0.5(3.84)] - 3.84$$

$$t_{2,DE} = 3.76 - 3.81 = -0.08$$

$t_{2,DE} < t_2 \quad \Rightarrow$ No hay reemplazo.

d = NDE

$$t_{2,NDE} = 1.5 + [0.65(2.11) + 0.25(3.84)] - 3.84$$

$$t_{2,NDE} = 3.83 - 3.84 = -0.01$$

$t_{2,NDE} < t_2 \quad \Rightarrow$ No hay reemplazo.

d = NE

$$t_{2,NE} = 1.6 + [0.7(2.11) + 0.2(3.84)] - 3.84$$

$$t_{2,NE} = 3.84 - 3.84 = 0$$

$t_{2,NE} \leq t_2 \quad \Rightarrow$ No hay reemplazo.

4) Como $(t_1, t_2) = (0,0)$ TERMINA. δ^* = (cliente 1=DE, cliente 2 =NE) es la política óptima .

Como recordamos del teorema 3.1, cuando el algoritmo termina y se ha llegado a la política óptima δ^* se cumple que $V_{\delta^*} = \mathbf{f}$, es decir:

$$\mathbf{f} = \begin{bmatrix} 2.11 \\ 3.84 \end{bmatrix}$$

ii) APROXIMACIONES SUCESIVAS

Como recordamos del capítulo 3, este método consiste en comenzar con un vector arbitrario V^0 en este caso de dimensiones 2x1 y calcular V^t t= 1,2,.... (que como vimos aproxima a **f** cuando $n \to \infty$), usando la siguiente ecuación recursiva:

$$V^t(i) = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=1}^{N} P(j \mid i,d) V^{t-1}(j) \right\} \quad i = 1,2.$$

t = 0

Supongamos que: $V^0(1) = 0$
$V^0(2) = 0.$

t=1

i= 1

$$V^1(1) = \max \begin{Bmatrix} r_{1,DE} + \beta \sum_{j=1}^{2} P(j \mid 1, DE) V^0(j) \\ r_{1,NDE} + \beta \sum_{j=1}^{2} P(j \mid 1, NDE) V^0(j) \\ r_{1,NE} + \beta \sum_{j=1}^{2} P(j \mid 1, NE) V^0(j) \end{Bmatrix}$$

Sustituyendo valores tenemos que:

$$V^1(1) = \max \begin{Bmatrix} -0.05 + 0.9(0.823 * 0 + 0.167 * 0) \\ 0.06 + 0.9(0.922 * 0 + 0.078 * 0) \\ 0.08 + 0.9(0.999 * 0 + 0.011 * 0) \end{Bmatrix}$$

$V^1(1)$= max { -0.05, 0.06, 0.08} $\therefore$ **$V^1(1)$= 0.08 (NE)** *i=2*

$$V^1(2)=\max\begin{Bmatrix} r_{2,DE}+\beta\sum_{j=1}^{2}P(j\,|\,2,DE)V^0(j)\\ r_{2,NDE}+\beta\sum_{j=1}^{2}P(j\,|\,2,NDE)V^0(j)\\ r_{2,NE}+\beta\sum_{j=1}^{2}P(j\,|\,2,NE)V^0(j)\end{Bmatrix}$$

Sustituyendo valores tenemos que:

$$V^1(2)=\max\begin{Bmatrix}1.0+0.9(0.444*0+0.556*0)\\ 1.5+0.9(0.722*0+0.278*0)\\ 1.6+0.9(0.778*0+0.222*0)\end{Bmatrix}$$

$V^1(2)=\max\{1.0,\ 1.5,\ 1.6\}$ $\therefore$ $\mathbf{V^1(2)=1.6}$ **(NE)**

t=2

i= 1

Sustituyendo valores tenemos que:

$$V^2(1)=\max\begin{Bmatrix}-0.05+0.9(0.823*0.08+0.167*1.6)\\ 0.06+0.9(0.922*0.08+0.078*1.6)\\ 0.08+0.9(0.999*0.08+0.011*1.6)\end{Bmatrix}$$

$V^2(1)=\max\{0.250,\ 0.239, 0.168\}$ $\therefore$ $\mathbf{V^2(1)=0.250}$ **(DE)**

i=2

Sustituyendo valores tenemos que:

$$V^2(2)=\max\begin{Bmatrix}1.0+0.9(0.444*0.08+0.556*1.6)\\ 1.5+0.9(0.722*0.08+0.278*1.6)\\ 1.6+0.9(0.778*0.08+0.222*1.6)\end{Bmatrix}$$

$V^2(2)=\max\{1.833,\ 1.952,\ 1.976\}$ $\therefore$ $\mathbf{V^2(2)=1.976}$ **(NE)**

Usando el mismo procedimiento repetidas veces obtenemos los siguientes valores:

T	ESTADO	DECISION OPTIMA	VALOR OPTIMO EN EL PERIODO T
3	1	DE	0.435
	2	NE	2.171
4	1	DE	0.602
	2	NE	2.339
5	1	DE	0.752
	2	NE	2.489
6	1	DE	0.887
	2	NE	2.624
7	1	DE	1.009
	2	NE	2.746
8	1	DE	1.119
	2	NE	2.856
9	1	DE	1.218
	2	NE	2.955
10	1	DE	1.307
	2	NE	3.044
11	1	DE	1.387
	2	NE	3.124
12	1	DE	1.459
	2	NE	3.196
13	1	DE	1.524
	2	NE	3.261
14	1	DE	1.582
	2	NE	3.313

Como observamos la diferencia componente a componente entre V^{13} y V^{14} es (0.058, 0.052) y la política δ= (DE,NE) no ha variado. Por lo que bajo este método obtenemos que la política δ^*= (DE,NE) es una política óptima y que el valor de V^{14} =(1.582,3.313) es una aproximación al valor de **f**.

Cabe hacer notar que por el método de mejoramiento de política obtuvimos que **f**= (2.11, 3.84), pero como recordamos del teorema 3.3 la convergencia al valor de **f** por el método de aproximaciones sucesivas está en términos de β^t, ya que en este caso β es muy cercana a 1 (β = 0.9) esta convergencia se hace cada vez más lenta, por ello consideramos a V^{14} sólo como una aproximación de **f**.

iii) PROGRAMACIÓN LINEAL

Aplicando lo visto en el capítulo 3, en la sección de programación lineal, sabemos que **f** es una solución factible al siguiente PPL :

$$\text{Min} \quad f^{\wedge}(1)+f^{\wedge}(2)$$

s.a.

$$f^{\wedge}(i)-\beta\sum_{j=1}^{2}P(j\mid i,d)f^{\wedge}(j)\geq r_{i,d} \qquad i=1,2 \qquad d=DE,NDE,NE$$

$$f^{\wedge}(1), f^{\wedge}(2) \quad \text{s. r. s.}$$

Sustituyendo valores tenemos que:

$$\text{Min} \quad f^{\wedge}(1)+f^{\wedge}(2)$$

s.a.

$f^{\wedge}(1) - 0.9\,[\,0.823\ f^{\wedge}(1) + 0.167\ f^{\wedge}(2)\ ] \geq -0.05$ (cliente 1, d= DE)

$f^{\wedge}(1) - 0.9\,[\,0.922\ f^{\wedge}(1) + 0.078\ f^{\wedge}(2)\ ] \geq 0.06$ (cliente 1, d=NDE)

$f^{\wedge}(1) - 0.9\,[\,0.999\ f^{\wedge}(1) + 0.011\ f^{\wedge}(2)\ ] \geq 0.08$ (cliente 1, d=NE)

$f^{\wedge}(2) - 0.9\,[\,0.444\ f^{\wedge}(1) + 0.556\ f^{\wedge}(2)\ ] \geq 1$ (cliente 2, d=DE)

$f^{\wedge}(2) - 0.9\,[\,0.722\ f^{\wedge}(1) + 0.278\ f^{\wedge}(2)\ ] \geq 1.5$ (cliente 2, d=NDE)

$f^{\wedge}(2) - 0.9\,[\,0.778\ f^{\wedge}(1) + 0.222\ f^{\wedge}(2)\ ] \geq 1.6$ (cliente 2, d= NE)

$$f^{\wedge}(1), f^{\wedge}(2) \quad \text{s.r.s.}$$

Realizando operaciones:

$$\text{Min} \quad f^{\wedge}(1) + f^{\wedge}(2)$$

s.a.

$$0.2593\ f^{\wedge}(1) - 0.1503\ f^{\wedge}(2) \geq -0.05$$

$$0.1702\ f^{\wedge}(1) - 0.1503\ f^{\wedge}(2) \geq 0.06$$

$$0.1009\ f^{\wedge}(1) - 0.0099\ f^{\wedge}(2) \geq 0.08$$

$$-0.3996\ f^{\wedge}(1) - 0.4996\ f^{\wedge}(2) \geq 1$$

$$-0.6498\ f^{\wedge}(1) + 0.7498\ f^{\wedge}(2) \geq 1.5$$

$$-0.7002\ f^{\wedge}(1) + 0.8002\ f^{\wedge}(2) \geq 1.6$$

$$f^{\wedge}(1),\ f^{\wedge}(2)\ \text{s.r.s.}$$

Como observamos el problema primal tiene 2 variables de decisión y 6 restricciones, de esta forma el problema dual tendrá 6 variables de decisión y 2 restricciones. Si denotamos como X_1, X_2, X_3, X_4, X_5 y X_6 a las variables de decisión del problema dual tenemos que éste nos queda planteado como:

$$\text{Max} \ -0.05\ X_1 + 0.06\ X_2 + 0.08\ X_3 + X_4 + 1.5\ X_5 + 1.6\ X_6$$

s.a.

$$0.2593\ X_1 + 0.1702\ X_2 + 0.1009\ X_3 - 0.3996\ X_4 - 0.6498\ X_5 - 0.7002\ X_6 = 1$$

$$-0.1503\ X_1 - 0.0702\ X_2 - 0.001\ X_3 + 0.4996\ X_4 + 0.7498\ X_5 + 0.8002\ X_6 = 1$$

$$X_i \geq 0 \quad i = 1,2,\ldots 6.$$

Utilizando el LINDO se obtiene que la solución óptima es:

$X_1 = 14.67$

$X_2 = 0$

$X_3 = 0$

$X_4 = 0$

$X_5 = 0$

$X_6 = 4.00$

De donde, recordando lo visto en el capítulo 3, en el óptimo si la variable de decisión del problema dual es positiva, su restricción asociada en el primal tendrá holgura cero, es decir se cumplirá como igualdad. Esto es, que la primera y la sexta restricciones del primal (que están asociadas a la decisión DE para los clientes tipo 1 y NE para los clientes tipo 2 , respectivamente) se cumplen como igualdad. Explícitamente tenemos que:

$$f(1) = r_{1,DE} + \beta \sum_{j=1}^{2} P(j \mid 1, DE)\, f(j)$$

$$f(2) = r_{2,NE} + \beta \sum_{j=1}^{2} P(j \mid 2, NE)\, f(j)$$

Con lo cual se tiene que **δ^* la política óptima es:**

δ^* = (cliente 1= DE, cliente2= NE).

Para determinar el valor de f resolvemos el sistema anterior de dos ecuaciones con dos **incógnitas y obtenemos que** f(1)= 1.95 y f(2)= 3.71.

EJEMPLO 4.2 REEMPLAZO DE MAQUINARIA

Un proceso de producción se lleva a cabo con una máquina que se deteriora rápidamente tanto en la calidad como en la cantidad de producción, de manera que se realiza una inspección periódica cada día.

Al terminar la inspección se observa la condición de la máquina y se clasifica en uno de los siguientes 4 estados posibles:

0 = Tan buena como nueva
1 = Operable (deterioro menor)
2 = Operable (deterioro mayor)
3 = Inoperable (producción de calidad inaceptable)

Sea X(t) el estado observado de la máquina después de la inspección al final del día t.

Es razonable suponer que el sistema pasa de un estado a otro de acuerdo a alguna "ley de movimiento" probabilística, por lo que la secuencia de estados {X(t)} se puede interpretar como un proceso estocástico .Es más, se modela como una cadena de Markov, ya que el que el proceso pasa de un estado a otro sin importar el deterioro de la máquina sufrido tiempo atrás, sino únicamente el estado en que se encuentra en el periodo actual. De esta forma el espacio de estados de esta cadena de Markov será S= {0,1,2,3}.

De acuerdo a un registro estadístico se sabe que las probabilidades de transición con las que el proceso pasa de un estado a otro, y que agruparemos en la siguiente matriz de probabilidades de transición, son:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Con esta matriz de transición es evidente que, una vez que la máquina se vuelve inoperable (entra al estado 3) permanece en esa situación .Por tanto, el análisis de este proceso estocástico es tal ves poco interesante , ya que el estado 3 es un estado absorbente y eventualmente la máquina entrará a él y ahí permanecerá, es decir, X(t) será siempre igual a 3.

Es claro que desde un punto de vista práctico lo que se busca es reemplazar o reparar esta máquina para integrarla nuevamente al proceso de producción. Esta acción de reemplazo altera el comportamiento del sistema que ahora se desarrollará en el tiempo de acuerdo al efecto conjunto de las leyes probabilísticas del movimiento y a la acción de reemplazar la máquina inoperable. Para efectos prácticos se supondrá que el reemplazo de la máquina es instantáneo y que dicho reemplazo se hará por una máquina nueva.

Al reemplazar la máquina inoperable por una nueva se incurre en un costo de 200.

El proceso estocástico que resulta del sistema con la política de reemplazo mencionada (reemplazar la máquina hasta que ésta se vuelve inoperable) tiene ahora asociada una nueva matriz de probabilidades de transición:

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \left[\begin{array}{cccc} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Bajo las condiciones planteadas anteriormente el conjunto de decisiones posibles para cada uno de los estados es el siguiente:

Estado	Conjunto de decisiones
0	{No reemplazar}
1	{Reemplazar, No Reemplazar}
2	{Reemplazar, No Reemplazar}
3	{Reemplazar}

Se sabe que las ganancias obtenidas cuando la máquina se encuentra en cada uno de los cuatro estados posibles son las siguientes:

Estado	Ganancia
0	100
1	80
2	50
3	0

A continuación se resolverá el problema de determinar una política óptima que maximize las ganancias esperadas usando cada uno de los tres métodos planteados en el capítulo 3.

En particular, el factor de descuento β se supondrá igual a 0.9.

i) MEJORAMIENTO DE POLITICA

1) INICIALIZACION

Elegir cualquier política $\delta = (\delta(0), \delta(1)\delta(2), \delta(3))$. Por ejemplo:

$\delta = (NR, NR, R, R)$.

2) EVALUACION DE LA POLITICA

Sea $t = (t_1, t_2, t_3, t_4) = (0, 0, 0, 0)$. Calcular $V_\delta = (I-Q_\delta)^{-1} R_\delta$.

En nuestro ejemplo tenemos que:

$$V_\delta = \begin{bmatrix} 1 & -0.788 & -0.056 & -0.056 \\ 0 & 0.325 & -0.113 & -0.113 \\ -0.9 & 0 & 1 & 0 \\ -0.9 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 80 \\ -100 \\ -100 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 2.463 & 5.971 & 0.813 & 0.813 \\ 1.541 & 6.814 & 0.856 & 0.856 \\ 2.217 & 5.374 & 1.731 & 0.731 \\ 2.217 & 5.374 & 0.731 & 1.731 \end{bmatrix} \begin{bmatrix} 100 \\ 80 \\ -100 \\ -100 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 561.38 \\ 528.02 \\ 405.42 \\ 405.42 \end{bmatrix}$$

3) Para i=0, 1, 2, 3 y cada $d \in D(i)$ calcular :

$$t_{i,d} = r_{i,d} + \sum_{j=0}^{3} Q(j \mid i, d) V_\delta(j) - V_\delta(i)$$

Si $t_{i,d} > t_i$ reemplazar la decisión $\delta(i)$ por d y a $t_{i,d}$ por t_i .

i=0

d= NR

$$t_{0,NR} = 100 + [\,0(561.38) + 0.788(528.02) + 0.056(405.42) + 0.056(405.42)] - 561.38$$

$t_{0,NR}$ = 100 + 461.38 - 531.68 = 561.38 - 561.38 = 0

$t_{0,NR} \leq t_0 \Rightarrow$ No hay reemplazo.

i=1

d= R

$t_{1,R}$ = -100 + [0.9(561.38) + 0(528.02) + 0(405.42)
+ 0(405.42)] - 528.02

$t_{1,R}$ = -100 + 505.24 - 528.02 = 505.24 - 628.02 = -122.78

$t_{1,R} < t_1 \Rightarrow$ No hay reemplazo.

d= NR

$t_{1,NR}$ = 80 + [0(561.38) + 0.675(528.02) + 0.113(405.42)
+ 0.113(405.42)] - 528.02

$t_{1,NR}$ = 80 + 448.02 - 528.02 = 528.02 - 528.02 = 0

$t_{1,NR} \leq t_1 \Rightarrow$ No hay reemplazo.

i=2

d= R

$t_{2,R}$ = -100 + [0.9(561.38) + 0(528.02)+ 0(405.42)
+ 0(405.42)] - 405.42

$t_{2,R}$ = -100 + 505.42 - 405.42 = 505.42 - 505.42 = 0

$t_{2,R} \leq t_2 \Rightarrow$ No hay reemplazo.

d= NR

$t_{2,NR}$ = 50 + [0(561.38) + 0(528.02) + 0.45(405.42)
+ 0.45(405.42)] - 405.42

$t_{2,NR}$ = 50 + 364.88 - 405.42 = 414.88 - 405.42 = 9.46

$t_{2,NR} > t_2 \Rightarrow \delta(2)$ es reemplazada por NR y t_2 es reemplazado por 9.46 .

i=3

d= R

$t_{3,R} = -100 + [\,0.9(561.38) + 0(528.02) + 0(405.42) + 0(405.42)] - 405.42$

$t_{3,R} = -100 + 505.42 - 405.42 = 505.42 - 505.42 = 0$

$t_{3,R} \leq t_3 \Rightarrow$ No hay reemplazo.

4) Si $(t_0,\ t_1, t_2,\ t_3) = (0,0,0,0)$ TERMINA. δ es una política óptima. En otro caso regresar al paso 2.

Como $(t_0,\ t_1, t_2,\ t_3) = (0,0,9.46,0)$ regresamos al paso 2.

2) EVALUACION DE LA POLITICA.

Ahora = (NR, NR, NR, R).

Sea $t = (t_1, t_2, t_3, t_4) = (0, 0, 0, 0)$. Calcular $V_\delta = (I-Q_\delta)^{-1} R_\delta$.

En nuestro ejemplo tenemos que:

$$V_\delta = \begin{bmatrix} 1 & -0.788 & -0.056 & -0.056 \\ 0 & 0.325 & -0.113 & -0.113 \\ 0 & 0 & 0.550 & -0.450 \\ -0.9 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 80 \\ 50 \\ -100 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 2.174 & 5.271 & 1.304 & 1.304 \\ 1.237 & 6.076 & 1.374 & 1.374 \\ 1.601 & 3.881 & 2.778 & 1.778 \\ 1.956 & 4.743 & 1.174 & 2.174 \end{bmatrix} \begin{bmatrix} 100 \\ 80 \\ 50 \\ -100 \end{bmatrix}$$

$$V_\delta = \begin{bmatrix} 573.88 \\ 541.08 \\ 431.68 \\ 416.34 \end{bmatrix}$$

3) Para i=0, 1, 2, 3 y cada $d \in D(i)$ calcular :

$$t_{i,d} = r_{i,d} + \sum_{j=0}^{3} Q(j \mid i,d) V_\delta(j) - V_\delta(i)$$

Si $t_{i,d} > t_i$ reemplazar la decisión $\delta(i)$ por d y a $t_{i,d}$ por t_i .

i=0

d= NR

$$t_{0,NR} = 100 + [\, 0(573.88) + 0.788(541.08) + 0.056(431.68) + 0.056(416.34)] - 573.88$$

$$t_{0,NR} = 100 + 473.88 - 573.88 = 573.88 - 573.88 = 0$$

$t_{0,NR} \leq t_0 \Rightarrow$ No hay reemplazo.

i=1

d= R

$$t_{1,R} = -100 + [\, 0.9(573.88) + 0(541.08) + 0(431.68) + 0(416.34)] - 541.08$$

$$t_{1,R} = -100 + 516.49 - 541.08 = 516.49 - 641.08 = -124.59$$

$t_{1,R} < t_1 \Rightarrow$ No hay reemplazo.

d= NR

$$t_{1,NR} = 80 + [\, 0(573.88) + 0.675(541.08) + 0.113(431.68) + 0.113(416.34)] - 541.08$$

$$t_{1,NR} = 80 + 461.08 - 541.08 = 541.08 - 541.08 = 0$$

$t_{1,NR} \leq t_1 \Rightarrow$ No hay reemplazo.

i=2

d= R

$$t_{2,R} = -100 + [\,0.9(573.88) + 0(541.08) + 0(431.68) + 0(416.34)] - 431.68$$

$$t_{2,R} = -100 + 516.49 - 431.68 = 516.49 - 531.68 = -15.19$$

$t_{2,R} < t_2 \;\Rightarrow$ No hay reemplazo.

d= NR

$$t_{2,NR} = 50 + [\,0(573.88) + 0(541.08) + 0.45(431.68) + 0.45(416.34)] - 431.68$$

$$t_{2,NR} = 50 + 381.68 - 431.68 = 431.68 - 431.68 = 0$$

$t_{2,NR} \leq t_2 \;\Rightarrow$ No hay reemplazo.

i=3

d= R

$$t_{3,R} = -100 + [\,0.9(573.88) + 0(541.08) + 0(431.68) + 0(416.34)] - 416.34$$

$$t_{3,R} = -100 + 516.34 - 416.34 = 516.34 - 516.34 = 0$$

$t_{3,R} \leq t_3 \;\Rightarrow$ No hay reemplazo.

4) Como $(t_1, t_2, t_3, t_4) = (0,0,0,0)$ **TERMINA.**

δ^* = (NR, NR, NR, R) es la política óptima .

Como recordamos del teorema 3.1, cuando el algoritmo termina y se ha llegado a la política óptima δ^* se cumple que $V_{\delta^*} = \mathbf{f}$, es decir:

$$\mathbf{f} = \begin{bmatrix} 573.88 \\ 541.08 \\ 431.68 \\ 416.34 \end{bmatrix}$$

ii)APROXIMACIONES SUCESIVAS

Como recordamos del capítulo anterior, este método consiste en comenzar con un vector arbitrario V^0, para este ejemplo en particular de dimensiones $4x1$ y calcular el vector V^t t= 1,2,.... (que como vimos aproxima a f cuando n→ ∞), usando la siguiente ecuación recursiva:

$$V^t(i) = \max_{d \in D(i)} \left\{ r_{i,d} + \beta \sum_{j=0}^{3} P(j \mid i,d) V^{t-1}(j) \right\} \quad i = 0,1,2,3.$$

t = 0

Supongamos que: $V^0(0) = 0$
$V^0(1) = 0$
$V^0(2) = 0$
$V^0(3) = 0$

t=1

i= 0

$$V^1(0) = \max \left\{ r_{0,NR} + 0.9 \sum_{j=0}^{3} P(j \mid 0, NR) V^0(j) \right\}$$

Sustituyendo valores tenemos que:

$$V^1(1) = \max \{100 + 0.9[\, P(0 \mid 0, NR) V^0(0) + P(1 \mid 0, NR) V^0(1) + P(2 \mid 0, NR) V^0(2) + P(3 \mid 0, NR) V^0(3)]\}$$

$$= \max\{100\} = \mathbf{100\ (NR)}$$

De igual forma, ya que $V^0(i) = 0$ para i= 0, 1, 2, 3, tenemos que:

i= 1
$V^1(1) = \max \{ r_{1,R}, r_{1,NR} \} = \max \{ -100, 80 \} = \mathbf{80\ (NR)}$

i=2
$V^1(2) = \max \{ r_{2,R}, r_{2,NR} \} = \max \{ -100, 50 \} = \mathbf{50\ (NR)}$

i=3
$V^1(3) = \max \{ r_{3,R} \} = \max \{ -100 \} = \mathbf{-100\ (NR)}$

t=2

i= 0

$$V^2(0) = \max\left\{ r_{0,NR} + 0.9\sum_{j=0}^{3} P(j \mid 0, NR)V^1(j) \right\}$$

Sustituyendo valores tenemos que:

$$V^2(0) = \max\ \{100\ + 0.9[0(100)+7/8(80)+1/16(50)+1/16(-100)]$$

$$V^2(0) = \max\ \{100\ + 0.9(70+3.13-6.25)\}$$

$$= \max\ \{100+60.19\} = \mathbf{160.19}\ \ \mathbf{(NR)}$$

i= 1

$$V^2(1) = \max\begin{cases} r_{1,R} + 0.9\sum_{j=0}^{3} P(j \mid 1, R)V^1(j) \\ r_{1,NR} + 0.9\sum_{j=1}^{3} P(j \mid 1, NR)V^1(j) \end{cases}$$

Sustituyendo valores tenemos que:

$$V^2(1) = \max\begin{cases} -100 + 0.9[0(100) + 1(80) + 0(50) + 0(-100)] \\ 80 + 0.9[0(100) + 0.75(80) + 0.125(50) + 0.125(-100)] \end{cases}$$

$$V^2(1) = \max\begin{cases} -100 + 0.9(80) \\ 80 + 0.9(60 + 6.25 - 12.5) \end{cases}$$

$$= \max\ \{-28, 128.38\} = \mathbf{128.38}\ \ \mathbf{(NR)}$$

i= 2

$$V^2(2) = \max\begin{cases} r_{2,R} + 0.9\sum_{j=0}^{3} P(j \mid 2, R)V^1(j) \\ r_{2,NR} + 0.9\sum_{j=1}^{3} P(j \mid 2, NR)V^1(j) \end{cases}$$

Sustituyendo valores tenemos que:

$$V^2(2) = \max\begin{cases}-100 + 0.9[0(100) + 0(80) + 1(50) + 0(-100)] \\ 50 + 0.9[0(100) + 0(80) + 0.5(50) + 0.5(-100)]\end{cases}$$

$$V^2(1) = \max\begin{cases}-100 + 0.9(50) \\ 50 + 0.9(25 - 50)\end{cases}$$

$$= \max \quad \{-55, 27.5\} \ = \mathbf{27.5} \quad \mathbf{(NR)}$$

i= 3

$$V^2(3) = \max\left\{r_{3,R} + 0.9\sum_{j=0}^{3} P(j \mid 3,R)V^1(j)\right\}$$

Sustituyendo valores tenemos que:

$$V^2\ (3) = \max\ \{-100\ + 0.9[1(100)+0(80)+0(50)+0(-100)]\}$$

$$V^2\ (0) = \max\ \{-100\ + 0.9(100)\}$$

$$= \max\ \{100+90\}\ = \mathbf{-10}\ \ \mathbf{(R)}$$

Usando el mismo procedimiento repetidas veces obtenemos los siguientes valores:

T	ESTADO	DECISION OPTIMA	VALOR OPTIMO EN EL PERIODO T
3	0	NR	202.08
	1	NR	168.63
	2	NR	57.88
	3	R	44.17
4	0	NR	253.93
	1	NR	205.31
	2	NR	95.93
	3	R	81.872
5	0	NR	271.68
	1	NR	238.59
	2	NR	130.01

	3	R	128.54
6	0	NR	302.43
	1	NR	270.14
	2	NR	166.35
	3	R	144.51
7	0	NR	330.22
	1	NR	297.32
	2	NR	189.89
	3	R	172.19
8	0	NR	354.51
	1	NR	321.43
	2	NR	212.94
	3	R	197.20
9	0	NR	376.20
	1	NR	343.11
	2	NR	234.56
	3	R	219.06
10	0	NR	395.72
	1	NR	362.63
	2	NR	254.13
	3	R	238.58
11	0	NR	413.29
	1	NR	380.21
	2	NR	271.72
	3	R	256.15
12	0	NR	429.11
	1	NR	396.03
	2	NR	287.54
	3	R	271.96
13	0	NR	443.34
	1	NR	410.26
	2	NR	301.76
	3	R	286.20
14	0	NR	456.15
	1	NR	423.07
	2	NR	314.58
	3	R	299.01
15	0	NR	467.68
	1	NR	434.60
	2	NR	326.12
	3	R	310.54
16	0	NR	478.06
	1	NR	444.98
	2	NR	336.50
	3	R	320.91

17	0	NR	487.40
	1	NR	454.32
	2	NR	345.83
	3	R	330.25
18	0	NR	495.81
	1	NR	462.73
	2	NR	354.24
	3	R	338.66
19	0	NR	503.38
	1	NR	470.29
	2	NR	361.81
	3	R	346.23
20	0	NR	510.18
	1	NR	477.10
	2	NR	368.62
	3	R	353.04
21	0	NR	516.31
	1	NR	483.23
	2	NR	374.75
	3	R	359.16
22	0	NR	521.83
	1	NR	488.75
	2	NR	380.26
	3	R	364.68
23	0	NR	526.80
	1	NR	493.71
	2	NR	385.22
	3	R	369.65
24	0	NR	531.26
	1	NR	498.18
	2	NR	389.69
	3	R	374.12
25	0	NR	535.28
	1	NR	502.20
	2	NR	393.71
	3	R	378.13

Como observamos a pesar de que $t=25$ el valor de V^t (i) $i=0,1,2,3$ todavía se encuentra lejos del valor de f(i) $i=0,1,2,3$ obtenido por el método de mejoramiento de política pero la convergencia se hace cada vez más lenta (ésto debido a que $\beta=0.9$ es muy cercana a 1 y como recordamos del teorema 3.3 dicha convergencia está en términos de β^t).Pero tomando en cuenta que la política $\delta=$ (NR, NR, NR, R) no ha variado, la consideramos como óptima y V^{25} = (535.28, 502.20, 393.71, 378.13) como una aproximación de **f**.

iii)PROGRAMACION LINEAL

De acuerdo a lo visto en el capítulo 3, tenemos que el planteamiento del PPL es el siguiente:

Min $f\hat{}(0) + f\hat{}(1) + f\hat{}(2) + f\hat{}(1) + f\hat{}(2)$

s.a

$$f\hat{}(0) - \beta \sum_{j=0}^{3} P(j \mid 0, NR)\, f\hat{}(j) \geq r_{0,NR}$$

$$f\hat{}(1) - \beta \sum_{j=0}^{3} P(j \mid 1, R)\, f\hat{}(j) \geq r_{1,R}$$

$$f\hat{}(1) - \beta \sum_{j=0}^{3} P(j \mid 1, NR)\, f\hat{}(j) \geq r_{1,NR}$$

$$f\hat{}(2) - \beta \sum_{j=0}^{3} P(j \mid 2, R)\, f\hat{}(j) \geq r_{2,R}$$

$$f\hat{}(2) - \beta \sum_{j=0}^{3} P(j \mid 2, NR)\, f\hat{}(j) \geq r_{2,NR}$$

$$f\hat{}(3) - \beta \sum_{j=0}^{3} P(j \mid 3, R)\, f\hat{}(j) \geq r_{3,R}$$

$f\hat{}(0), f\hat{}(1), f\hat{}(2), f\hat{}(3)$ s.r.s.

Sustituyendo valores tenemos que:

Min $f\hat{}(0) + f\hat{}(1) + f\hat{}(2) + f\hat{}(3)$

s.a.

$$
\begin{aligned}
&f\hat{}(0) - \beta(0 * f\hat{}(0) + 7/8 * f\hat{}(1) + 1/16 * f\hat{}(2) + 1/16 * f\hat{}(3)) \geq 100\\
&f\hat{}(1) - \beta(1 * f\hat{}(0) + 0 * f\hat{}(1) + 0 * f\hat{}(2) + 0 * f\hat{}(3)) \geq -100\\
&f\hat{}(1) - \beta(0 * f\hat{}(0) + 3/4 * f\hat{}(1) + 1/8 * f\hat{}(2) + 1/8 * f\hat{}(3)) \geq 80\\
&f\hat{}(2) - \beta(1 * f\hat{}(0) + 0 * f\hat{}(1) + 0 * f\hat{}(2) + 0 * f\hat{}(3)) \geq -100\\
&f\hat{}(2) - \beta(0 * f\hat{}(0) + 0 * f\hat{}(1) + 1/2 * f\hat{}(2) + 1/2 * f\hat{}(3)) \geq 50\\
&f\hat{}(3) - \beta(1 * f\hat{}(0) + 0 * f\hat{}(1) + 0 * f\hat{}(2) + 0 * f\hat{}(3)) \geq -100
\end{aligned}
$$

$f\hat{}(0), f\hat{}(1), f\hat{}(2), f\hat{}(3)$ s.r.s.

Realizando operaciones tenemos que:

$$\text{Min} \quad f^{\wedge}(0) + f^{\wedge}(1) + f^{\wedge}(2) + f^{\wedge}(3)$$

s.a.

$$f^{\wedge}(0) - 0.788\, f^{\wedge}(1) - 0.056\, f^{\wedge}(2) - 0.056\, f^{\wedge}(3) \geq 100$$

$$-0.9\, f^{\wedge}(0) + f^{\wedge}(1) \geq -100$$

$$0.325\, f^{\wedge}(1) - 0.113\, f^{\wedge}(2) - 0.113\, f^{\wedge}(3) \geq 80$$

$$-0.9\, f^{\wedge}(0) + f^{\wedge}(2) \geq -100$$

$$0.55\, f^{\wedge}(2) - 0.45\, f^{\wedge}(3) \geq 50$$

$$-0.9\, f^{\wedge}(0) + f^{\wedge}(3) \geq -100$$

$$f^{\wedge}(0), f^{\wedge}(1), f^{\wedge}(2), f^{\wedge}(3) \quad \text{s.r.s.}$$

Siguiendo con las ideas expuestas en el capítulo 2 lo que nos interesa realmente es el planteamiento del problema dual para que al resolver éste obtengamos la política óptima. Como observamos, en el problema primal están consideradas 6 restricciones y 4 variables de decisión, por lo que el problema dual tendrá 4 restricciones y 6 variables de decisión. Si denotamos por X_1, X_2, X_3, X_4, X_5 y X_6 las variables de decisión del dual tenemos que éste queda planteado como:

$$\text{Max} \quad 100\, X_1 - 100\, X_2 + 80\, X_3 - 100\, X_4 + 50\, X_5 - 100\, X_6$$

s.a.

$$X_1 - 0.9\, X_2 - 0.9\, X_4 - 0.9\, X_6 = 1$$

$$-0.788\, X_1 + X_2 + 0.325\, X_3 = 1$$

$$-0.056 X_1 - 0.113\, X_3 + X_4 + 0.45 X_5 = 1$$

$$-0.056 X_1 - 0.113\, X_3 - 0.55\, X_5 + X_6 = 1$$

$$X_1, X_2, X_3, X_4, X_5 \text{ y } X_6 \geq 0$$

Ya una vez planteado el dual, utilizando el LINDO obtenemos que la solución óptima es:

X_1 = 10.87

X_2 = 0

X_3 = 29.43

X_4 = 0

X_5 = 10.96

X_6 = 10.96

Por lo que aplicando lo estudiado en el capítulo 3, sabemos que en el óptimo si la variable de decisión del problema dual es positiva, su restricción asociada en el primal tendrá holgura cero, es decir se cumplirá como igualdad.

Para este problema tenemos que la primera, la tercera, la quinta y la sexta de las restricciones del primal (asociadas a las decisiones NR, NR, NR, R para los estados 0,1,2,3 respectivamente) se darán como igualdad, explícitamente se cumple que:

$$f(0) = r_{0,NR} + \beta \sum_{j=0}^{3} P(j \mid 0, NR)\, f(j)$$

$$f(1) = r_{1,NR} + \beta \sum_{j=0}^{3} P(j \mid 1, NR)\, f(j)$$

$$f(2) = r_{2,NR} + \beta \sum_{j=0}^{3} P(j \mid 2, NR)\, f(j)$$

$$f(3) = r_{3,R} + \beta \sum_{j=0}^{3} P(j \mid 3, R)\, f(j)$$

Por lo que δ^*, la política óptima, es : δ^* = (NR, NR, NR, R).

Para determinar el valor de **f** resolvemos el sistema de ecuaciones que aparece arriba y obtenemos que:

$$f(0) = 573.75$$

$$f(1) = 540.98$$

$$f(2) = 431.39$$

$$f(3) = 416.38 \quad .$$

CONCLUSIONES

Al estudiar los Procesos Markovianos de Decisión y en particular como el presente trabajo lo hace, los métodos de solución de un P.M.D., el primer aspecto interesante que salta a la vista es el hecho de que se conjugan el aspecto probabilístico del proceso estocástico en sí con el aspecto determinista de las técnicas de investigación de operaciones que sirven como fundamento de los métodos que permiten obtener la política óptima.

Ahora pasemos a comentar la eficiencia de cada uno de los tres métodos estudiados y revisemos a manera de conclusión algunas de sus principales ventajas y desventajas.

METODO	VENTAJAS	DESVENTAJAS
MEJORAMIENTO DE POLITICA	Se conoce el verdadero valor de f.	Es necesario en cada iteración calcular: $V_\delta = (I - Q_\delta)^{-1}$ por lo que encontrar la inversa de la matriz $(I - Q_\delta)$ de dimensiones NxN resulta ser muy complicado cuando N es muy grande.
APROXIMACIONES SUCESIVAS	Cada iteración se desarrolla en forma simple y rápida.	No se conoce el valor real de f si no sólo una aproximación de él, que será mejor cuando $t \to \infty$.
PROGRAMACION LINEAL	Existen algoritmos para computadora como el LINDO que permiten obtener la solución óptima al PPL en forma relativamente rápida.	

Avocándonos a analizar el aspecto númerico, observamos que aparecen ciertas dificultades cuando el factor de descuento β es muy cercano a 1 (es decir, cuando la tasa de interés r es muy pequeña).

Como ya lo hemos mencionado en el método de mejoramiento de política es necesario calcular $V_\delta = (I - Q_\delta)^{-1}$. Sea **1** el vector de dimensiones Nx1 de 1's. Observemos que:

$$(I - \beta P_\delta)\,\mathbf{1} = (1 - \beta)\,\mathbf{1} \quad \left(\text{Ya que} \sum_{j=1}^{N} P(j \mid i) = 1 \quad i = 1,2, \ldots N\right)$$

En consecuencia $(1-\beta)$ es un eigenvalue de $(I - \beta P_\delta)$, lo cual significa que el determinante de la matriz $(I - \beta P_\delta)$ se aproxima a 0 cuando β tiende a 1, por lo cual está muy cerca de ser singular y en este caso aparecerán dificultades numéricas.

En el caso del método de aproximaciones sucesivas también se presentan dificultades numéricas cuando $\beta \to 1$: el término β^t se aproxima a 0 muy despacio; ésto hace que la convergencia se torne también muy lenta (como pudimos comprobarlo en los ejemplos del capítulo 4) y que los errores numéricos se vean disminuidos muy lentamente.

BIBLIOGRAFIA

1 **A first course in stochastic processes**
Second edition
Samuel Karlin
Howard M. Taylor
Academic Press 1975

2 **Métodos y modelos de investigación de operaciones**
Vol. 2 Modelos estocásticos
Juan Prawda
Limusa 1991

3 **Introducción a la investigación de operaciones**
Frederick S. Hillier
Gerald J. Lieberman
Mc. Graw Hill quinta edición 1991

4 Dynamic Programming. Models **and applications**
Eric V. Denardo
Prentice - Hall 1982