UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
ACATLAN

" PROCESOS MARKOVIANOS DE DECISION "

T E S I S

Que para obtener el Título de :

Licenciatura en Matemáticas Aplicadas y Computación

P R E S E N T A

Martínez Sánchez José Francisco

Abril 1993

PREFACIO

El objetivo del presente trabajo es presentar algunos de los conceptos básicos para el análisis de problemas de decisión bajo incertidumbre. Concretamente, el trabajo se centra en el estudio de los **procesos Markovianos de decisión** para los que la herramienta fundamental es la programación dinámica estocástica.

Los resultados obtenidos se aplican en problemas de control de **cadenas de Markov** con espacio de estados numerable, los cuales son comunes en Investigación de Operaciones y Estadística entre otras áreas.

El trabajo está organizado en tres capítulos, en cada capítulo se considera a los procesos Markovianos de decisión bajo un cierto criterio de óptimalidad.

El capítulo I trata de los procesos Markovianos de decisión utilizando el criterio de los costos descontados, en donde el objetivo es minimizar una función costo con un factor de descuento α [$\in (0,1)$] que equivale al cociente (1/1+i) donde i es la tasa de rendimiento actual. Por lo tanto α^n; representa el valor actual de una unidad monetaria n-períodos en el futuro.

El capítulo II trata de los procesos Markovianos de decisión utilizando el criterio de los costos positivos sin descuento, en los que el objetivo es minimizar una función costo que puede ser infinita.

El capítulo III trata de los procesos Markovianos de decisión utilizando el criterio del costo promedio esperado, el cual, como se verá puede considerarse como un límite de la función costo del capítulo I.

La opinión del asesor de tesis, asi como la de algunos sinodales acerca de que en un trabajo de tesis en el nivel de licenciatura, el alumno conoce los conceptos básicos de alguna área especifica sin tener el dominio de ésta; me lleva a concluir que hablar de "aportación" es pretencioso.

Así, la naturaleza del trabajo es teórica; es decir, el objetivo principal es conocer los elementos básicos de los procesos Markovianos de decisión y presentarlos de una manera sintetizada. También, las aplicaciones consideradas son de la misma naturaleza; puesto que para desarrollar alguna aplicación "real", se requiere de experiencia (práctica) y un conocimiento profundo de ésta área.

Para obtener el mínimo de la función costo que caracteriza a cada uno de los criterios es a través de un proceso iterativo.

La intención no es presentar al proceso iterativo en forma exhaustiva, si no dar los elementos teóricos que lo fundamentan.

Indice de contenidos

Indice de símbolos

I	Conjunto de posibles estados
A	Conjunto de posibles acciones
i	Estado actual
j	Próximo estado
P_{ij}	Probabilidad de transición
a	Acción tomada
α	Factor de descuento
C	Costo
X_t	Variable aleatoria que indica el estado del proceso al tiempo t
f	Política estacionaria
π	Política arbitraria
E_π	Esperanza condicional cuando se usa la política π.
V	Función de costo óptimo.
T_f	Costo esperado cuando se usa la política estacionaria f pero se termina un período después.
$B(I)$	Conjunto de todas las funciones acotadas.
$N(I)$	Conjunto de todas las funciones no negativas.
ϕ_π	Función costo promedio esperado cuando se usa la política π.

INTRODUCCION

Se define un proceso estocástico como un sistema que se desarrolla a través del tiempo de acuerdo a ciertas leyes de probabilidad; los valores tomados por el sistema se llaman estados (que pueden ser discretos o continuos). Los cambios de estados se denominan transiciones.

En el presente trabajo, se considera un proceso estocástico como la familia de variables aleatorias $\{X_t\}$ cuyo contradominio es el conjunto contable $I=\{0,1,2,...\}$, que es observado en puntos de tiempo $t=0,1,2,...$. Dependiendo del estado del proceso, se toma una acción $\alpha \in A$, donde A representa el conjunto de todas las acciones posibles. Se supone que A es un conjunto finito a lo más contable.

Si el proceso se encuentra en el estado i al tiempo t y si además, se toma la acción α, entonces las siguientes situaciones se presentan.

i) se incurre en un costo $C(i,\alpha)$;

$C(i,\alpha)$ es el costo cuando el proceso está en el estado i y se toma la decisión α

ii) el próximo estado del proceso, digamos j, ocurre de acuerdo con una probabilidad de transición $P_{ij}(\alpha)$

$P_{ij}(\alpha)$ es la probabilidad de transición del estado i al j dado que se toma la decisión α.

Si X_t denota el estado del proceso al tiempo t, α_t la acción tomada en t, entoces $ii)$ es equivalente a:

$P\{X_{t+1}=j \mid X_0,\alpha_0,X_1,\alpha_1,\ldots,X_t=i,\alpha_t=\alpha\}=P_{ij}(\alpha)$.

Nótese que tanto los costos $C(i,\alpha)$ como las probabilidades de transición $P_{ij}(\alpha)$ son funciones sólo del último estado y de la acción tomada. Además, se supone que los costos son acotados.

A cualquier regla para escoger acciones dependiendo del estado del proceso se le llamará política (o estrategia de decisión). Así, la acción tomada por una política puede, por ejemplo, depender de la historia del proceso ó ser aleatoria en el sentido de que toma la acción α con alguna probabilidad P_α, $\alpha \in A$. Dentro de la clase de todas las políticas, existe una subclase de políticas estacionarias.

Una política se dice ser <u>estacionaria</u> si no es aleatoria y la acción que se toma en cualquier tiempo t, únicamente depende del estado del proceso; es decir, el tiempo t no influye. En otras palabras, una política estacionaria es una función del espacio de estados al espacio de decisiones. De lo anterior se sigue que si se usa una política estacionaria f, entonces la secuencia de estados $\{X_t, t=0,1,2,\ldots\}$ forma una **cadena de Markov** con probabilidades de transición $P_{ij}=P_{ij}[f(i)]$; es por esta razón que al proceso se le llama **PROCESO MARKOVIANO DE DECISION.**

El propósito de este trabajo es describir la teoría básica para encontrar políticas que seán óptimas de acuerdo con algún criterio de optimalidad.

En el capítulo I se considera el criterio del costo total descontado. Se presenta la teoría concerniente a la existencia de políticas estacionarias óptimas y métodos para su obtención.

Los resultados obtenidos serán aplicados en:

a) determinación del tiempo óptimo para el reemplazo de una máquina;

b) determinación de la política óptima en la venta de una casa.

En el capítulo II se considera el criterio de los costos positivos sin descuento. De manera similar al capítulo I se presentan las condiciones para la existencia de una política óptima y métodos para obtenerla. Estos resultados serán aplicados en la teoría de paro óptimo.

En el capítulo III se utiliza el criterio del costo promedio esperado. Se presentan contraejemplos para proporcionar una idea de las dificultades asociadas al problema de la existencia de políticas estacionarias.

Por ejemplo, en un problema dado puede ocurrir que existan políticas óptimas pero no en la clase de políticas estacionarias.

CAPITULO I

COSTOS DESCONTADOS

1.1 Ecuación Funcional.

En este capítulo se estudia la teoría que permita encontrar políticas que seán óptimas para el criterio de costos descontados. Se supone que los costos son acotados.

Este criterio asume un factor de descuento $\alpha \in (0,1]$, e intenta minimizar el costo total esperado descontado, es decir, busca $\inf_{\pi} V_{\pi}(i)$

Donde

$$V_{\pi}(i) = E_{\pi}\left[\sum_{t=0}^{\infty} \alpha^{t}\, C(X_t, a_t) \mid X_0 = i\right], \quad i \in I, \qquad (1.1)$$

$I = \{0, 1, 2, \ldots\}$

E_{π} esperanza condicional cuando se emplea la política π

X_t estado del proceso al tiempo t

X_0 estado inicial del proceso

a_t acción tomada en t

$C(X_t, a_t)$ costo incurrido en la etapa t.

El factor α produce un descuento en el costo a través del tiempo, de modo que los costos son menos significativos en el futuro que en la actualidad.

Por otra parte, sea

$$V(i) = \inf_{\pi} V_{\pi}(i), \quad i \in I.$$

Es decir, $V(i)$ es la máxima cota inferior de la función costo $V_{\pi}(i)$.

Se dice que una política π^* es <u>óptima</u> si

$$V_{\pi^*}(i) = \inf_{\pi} V_{\pi}(i), \text{ para todo } i \in I.$$

En otras palabras, la política π^* es óptima si el costo esperado en (1.1) es mínimo para todo estado inicial cuando se usa π^*.

El siguiente teorema proporciona la ecuación de Programación Dinámica, también conocida como ecuación de optimalidad para el caso descontado, que caracteriza a la función de costo óptimo V.

Teorema 1.1.

$$V(i)=\min_{\alpha}\{C(i,\alpha)+\alpha\sum_{j=0}^{\infty}P_{ij}(\alpha)V(j)\},\quad i\in I. \tag{1.2}$$

Demostración:

Sea π una política arbitraria, y supongamos que π toma la acción α en la etapa 0 con probabilidad P_{α} , $\alpha \in A$, donde A es el conjunto de todas las acciones posibles. Entonces

$$V_{\pi}(i)=\sum_{\alpha\in A} P_{\alpha}\{C(i,\alpha)+\sum_{j=0}^{\infty} P_{ij}(\alpha)\ W_{\pi}(j)\};$$

dode i es el estado inicial;

$C(i,\alpha)$ costo que se incurre cuando se está en el estado i y se toma la decisión α;

$W_{\pi}(j)$ costo esperado incurrido para una etapa adelante, cuando se usa la política π y el estado siguiente es j.

Por definición de V (costo mínimo) tenemos que

$$W_{\pi}(j)\geq\alpha\ V(j)$$

Por lo que

$$V_{\pi}(i) \geq \sum_{a\in A} P_{a} \{C(i,a)+ \alpha \sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\} \geq$$

$$\sum_{a\in A} P_{a}\ \min_{a}\{C(i,a)+ \alpha \sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\} = \qquad (1.3)$$

$$\min_{a}\{C(i,a)+ \alpha \sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\},$$

donde la igualdad es válida ya que $\sum_{a\in A} P_{a}=1$.

Puesto que π es arbitraria, (1.3) implica que

$$V(i) \geq \min_{a}\{C(i,a)+ \alpha \sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\} \qquad (1.4)$$

Por otro lado, sea a_{0} tal que

$$C(i,a_{0})+ \alpha \sum_{j=0}^{\infty} P_{ij}(a_{0})\ V(j) = \min_{a}\{C(i,a)+\alpha \sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\} \qquad (1.5)$$

Sea π la política que selecciona la decisión a_{0} en la etapa 0; y si el próximo estado es j, entonces se sigue una política π_{j} tal que, dado ε arbitrario,

$$V_{\pi_{j}}(j) \leq V(j)+\varepsilon, \quad \text{para todo } j \in I.$$

Entonces

$$V_{\pi}(i)= C(i,a_{o})+ \alpha \sum_{j=o}^{\infty} P_{ij}(a_{o})\ V_{\pi_{j}}(j)\leq$$

$$C(i,a_{o})+ \alpha \sum_{j=o}^{\infty} P_{ij}(a_{o})\ V(j)+\alpha\ \varepsilon.$$

Puesto que, por definición, $V(i)\leq V_{\pi}(i)$, se tiene

$$V(i)\leq C(i,a_{o})+ \alpha \sum_{j=o}^{\infty} P_{ij}(a_{o})\ V(j)+\alpha\ \varepsilon$$

Por lo tanto, de (1.5)

$$V(i)\leq \min_{a}\{C(i,a)+ \alpha \sum_{j=o}^{\infty} P_{ij}(a)\ V(j)\}+\alpha\ \varepsilon. \qquad (1.6)$$

Luego, como ε era arbitrario, se obtiene

$$V(i)\leq \min_{a}\{C(i,a)+ \alpha \sum_{j=o}^{\infty} P_{ij}(a)\ V(j)\}.$$

Esta desigualdad, en combinación con (1.4), implica (1.2)

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

1.2 Políticas Estacionarias

<u>Definición:</u> Una política es estacionaria si no es aleatoria y el tiempo no influye para tomar una decisión; es decir, la política toma una decisión dependiendo

únicamente del estado presente, independientemente de la etapa en que se encuentra el proceso.

Alternativamente, las políticas estacionarias son funciones del conjunto de estados, I, al conjunto de acciones, A.

Esto es, dado el estado i, una política estacionaria $f:I \longrightarrow A$ selecciona la acción $f(i)$.

Si $B(I)$ denota al conjunto de todas las funciones acotadas (de valor real) definidas sobre el espacio de estados entonces V_π pertenece a $B(I)$. (recuérdese que, por hipótesis, la función de costo $C(i,a)$ es acotada).

Para alguna política estacionaria f se define la función

$$T_f:B(I) \longrightarrow B(I)$$

através de la siguiente expresión:

$$(T_f u)(i)=C[i,f(i)]+ \alpha \sum_{j=0}^{\infty} P_{ij}\ [f(i)]\ u(j) \tag{1.7}$$

Esto es, para cada función $u \in B(I)$, $T_f u$ es la función cuyo valor en i es dado por (1.7). Dado que tanto u como C son funciones acotadas, entonces $T_f u$ es también acotada y por lo tanto pertenece a $B(I)$.

Se interpreta a $T_f u$ evaluada en i, como el costo esperado cuando se usa la política estacionaria f, pero el proceso termina una etapa después y se incurre en un costo final $\alpha u(j)$, cuando el estado final es j.

En lo que resta de la sección, se mostrarán algunas propiedades de la función T_f que serán de suma importancia en las siguientes secciones.

Se define las iteraciones T_f^n de la función T_f mediante las relaciones

$$T_f^1=T_f \quad ; \text{ para } n>1 \qquad T_f^n=T_f(T_f^{n-1})$$

<u>Definición</u>: para cualesquiera dos funciones $u,v \in B(I)$, se dice que $u\leq v$ si $u(i)\leq v(i)$ para todo $i \in I$. Asimismo, se dice que $u=v$ si $u(i)=v(i)$ para todo $i \in B(I)$.

<u>Definición</u>: si u y u_n $(n=1,2,\ldots)$ son funciones en $B(I)$, se dice que $u_n \longrightarrow u$ (u_n tiende a u) si $u_n(i)\longrightarrow u(i)$ para todo $i \in I$.

<u>Lema 1.2.</u>

Para $u,v \in B(I)$ y una política estacionaria f:

i) $u\leq v \Rightarrow T_f u\leq T_f v$

ii) $T_f\ V_f=V_f$

iii) $T_f^n u \longrightarrow V_f$ para todo $u \in B(I)$.

Demostración:

Parte i).

Si $u \leq v$ y se aplica T_f a ambos lados de la desigualdad, entoces

$$T_f u \leq T_f v$$

Parte ii).

Por definición de T_f [ver (1.7)], tenemos

$$(T_f V_f)(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij} [f(i)] V_f(j)$$

Por otro lado

Por definición de costo descontado,

$$V_f(i) = E_f\{ \sum_{t=0}^{\infty} \alpha^t C[X_t, f(X_t)] \mid X_0 = i\}$$

$$= E_f\{C[X_0, f(X_0)] + \alpha \sum_{t=1}^{\infty} \alpha^{t-1} C[X_t, f(X_t)] \mid X_0 = i\}$$

$$= C[i, f(i)] + \alpha\ E_f\{ \sum_{t=1}^{\infty} \alpha^{t-1} C[X_t, f(X_t)] \mid X_0 = i\}$$

$$= C[i, f(i)] + \alpha\ E_f\{E_f[\sum_{t=1}^{\infty} \alpha^{t-1} C[X_t, f(X_t)] \mid$$

$$X_0 = i, a_0 = f(i), X_1 = j] \mid X_0 = i\}$$

$$=C[i,f(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]*$$

$$E_f\{ \sum_{t=1}^{\infty} \alpha^{t-1} C[X_t, f(X_t)] \mid X_1=j\}$$

haciendo el cambio de variable $t'=t-1$

$$=C[i,f(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]*$$

$$E_f\{ \sum_{t'=0}^{\infty} \alpha^{t'} C[X_{t'+1}, f(X_{t'+1})] \mid X_1=j\}$$

$$=C[i,f(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]\ V_f(j)$$

Entonces

$$T_f V_f = V_f \ .$$

Parte iii). Primero se observa que

$T_f^2=T_f(T_f)$ por definición de T_f se tiene

$$(T_f^2 u)(i)=C[i,f(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f(i)](T_f u)(j)$$

Aplicando la definición de T_f a u

$$(T_f^2 u)(i)=C[i,f(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]*$$

$$\{C[j,f(j)]+\alpha \sum_{k=0}^{\infty} P_{jk}[f(j)]u(k)\}$$

Haciendo operaciones

$$C[i,f(i)]+\alpha\sum_{j=0}^{\infty}P_{ij}[f(i)]\{C[j,f(j)]+\alpha^2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}$$

$$P_{ij}[f(i)]P_{jk}[f(j)]u(k)$$

En otras palabras T_f^2u representa el costo esperado en dos etapas, dado que la política estacionaria f es empleada y se incurre en un costo final α^2u.

Así, para la primera etapa, el estado presente es i, las probabilidades de transición al estado j dado que se tomó la decisión $f(i)$ son:

$$P_{ij}[f(i)]$$

y el costo incurrido para el estado j es:

$$C[j,f(j)].$$

Para la segunda etapa, las probabilidades de transición del estado j al estado k dado que se tomó la decisión $f(j)$ son:

$$P_{ij}[f(i)]P_{jk}[f(j)]$$

y el costo incurrido para el estado k es:

$$\alpha^2u(k).$$

Por inducción se tiene que T_f^nu representa el costo esperado para n etapas, dado que se utiliza la política estacionaria f y el costo incurrido para la etapa n es $\alpha^n u$.

Puesto que α pertenece a $(0,1)$, y si el número de etapas, n, tiende a infinito, entonces

$$\alpha^n \longrightarrow 0$$

Esto implica que el costo incurrido para la etapa n, $\alpha^n u$, tiende a cero cuando $n \longrightarrow \infty$, ya que u es acotada.

De lo anterior se sigue que

$$T_f^n u(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)][C[j, f(j)]]$$

$$= C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)] V_f(j)$$

$$= V_f(i)$$

L.Q.Q.D.

El inciso iii) del lema anterior demuestra que para toda función $u \in B(I)$ (conjunto de funciones acotadas definidas en el espacio de estados), si se aplica la función T_f^n a la función u y se permite que n tienda a infinito, entonces, esta función evaluada en el estado presente i, tiende a la función V_f evaluada en i.

1.3 Teorema de Programación Dinámica.

En esta sección se desarrolla la teoría concerniente a la existencia de políticas estacionarias óptimas.

Además, se proporciona un método para obtener la función V; es decir, el costo total descontado esperado óptimo.

Teorema 1.3.

Si f_α es la política estacionaria, la cual, cuando el proceso se encuentra en el estado i, selecciona la acción que minimiza el lado derecho de (1.2); es decir, $f_\alpha(i)$ es tal que

$$C[i,f_\alpha(i)]+\alpha\sum_{j=0}^{\infty}P_{ij}[f_\alpha(i)]V(j)=$$

$$\min_a\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)V(j)\}, \qquad i\in I.$$

Entonces

$$V_{f_\alpha}(i)=V(i) \quad \text{para todo } i\in I$$

y por lo tanto f_α es la política óptima

Demostración:

Aplicando a V la función T_{f_α} definida por (1.7), se tiene:

$$(T_{f_\alpha} V)(i) = C[i, f_\alpha(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f_\alpha(i)] V(j)$$

(por hipótesis)

$$= \min_a \{C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V(j)\}$$

(por teorema 1.1)

$$= V(i)$$

Por lo que

$$T_{f_\alpha} V = V \quad . \tag{1.7'}$$

Esto implica que

$$T^2_{f_\alpha} V = T_{f_\alpha}(T_{f_\alpha} V)$$

$$= T_{f_\alpha}(V) \quad \text{por } (1.7')$$

$$= V \quad \text{por } (1.7')$$

Así, por inducción,

$$T^n_{f_\alpha} V = V \qquad \text{para toda } n$$

Si se permite que n tienda a infinito y haciendo uso del lema 1.2 inciso iii)

$$T^n_{f_\alpha} V \longrightarrow V_{f_\alpha} \quad .$$

Por lo tanto, cuando n tiende a infinito

$V_{f_\alpha}(i)=V(i)$

Es decir, f_α es una política estacionaria óptima.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

Así, una política óptima existe y es estacionaria. Y se encuentra determinada por la ecuación (1.2).

De lo anterior, si se determina la función de costo esperado óptimo V; entonces, una política estacionaria seleccionada como f_α en el teorema 1.3 es óptima.

Algoritmo Iterativo de Políticas.

Supóngase que se evalua la función de costo esperado para alguna política estacionaria f. Se denota a esta función por V_f .

Ahora sea la política estacionaria f^*, la cual, cuando el proceso se encuentra en el estado presente i, selecciona la acción que minimiza la siguiente expresión:

$$C(i,a)+\alpha \sum_{j=0}^{\infty} P_{ij}(a)V_f(j)$$

Es decir, f^* es tal que

$$C[i,f^*(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f^*(i)]V_f(j)=$$

$$\min_a\{C(i,a)+\alpha \sum_{j=0}^{\infty} P_{ij}(a)V_f(j)\}, \quad i \in I \tag{1.8}$$

Si se comparán ambas políticas, se tiene que f^* da un costo menor o igual que el que da f. Esto será demostrado en el siguiente corolario.

Corolario 1.4.

$V_{f^*}(i) \leq V_f(i)$ para todo $i \in I$.

En palabras, se tiene que el costo esperado cuando se usa la política estacionaria f^* es menor o igual, que el costo esperado cuando se usa la política f.

Demostración:

Por (1.8) y la definición de T_{f^*} [ver (1.7)],

$$T_{f^*} V_f(i)=C[i,f^*(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f^*(i)]V_f(j)\leq$$

$$C[i,f(i)]+\alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]V_f(j)= \tag{1.9}$$

(por definición de V_f)

$$V_f(i). \tag{1.10}$$

Por lo que,

$$T_f^* V_f \leq V_f . \qquad (1.10')$$

Aplicando T_f^* en ambos lados de la desigualdad, se tiene

$$T_f^{2*} V_f \leq T_f^* V_f \leq V_f$$

[por (1.10')].

Más generalmente, por inducción se tiene

$$T_f^{n*} V_f \leq V_f \qquad \text{para todo } n=1,2,...$$

Si n tiende a infinito y haciendo uso del lema 1.2 inciso iii), entonces

$$T_f^{n*} \ V_f \longrightarrow V_f^* \quad \text{cuando } n \text{ tiende a infinito;}$$

Por lo tanto

$$V_f^*(i) \leq V_f(i).$$

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

Este corolario es de suma importancia para la busqueda de políticas estacionarias óptimas, por ser la base del algoritmo iterativo de políticas. Este comienza con una política inicial, que se mejora en cada iteración del algoritmo hasta encontrar la política óptima.

El teorema 1.3, demuestra las propiedades que satisface la política estacionaria f_α. En palabras, se tiene que f_α proporciona el mínimo costo; caracterizado por (1.2).

Por otra parte, el corolario 1.4 garantiza que la política f^* que minimiza el lado derecho de (1.2), es mejor que la política f, por lo que la función de costo V para f^* evaluada en i, es menor o igual que si se usa f.

Hasta aquí, se ha desarrollado la teoría para encontrar políticas estacionarias óptimas.

Ahora, se enfocará la atención, a la función de valor óptimo V. De manera similar que en las políticas óptimas, se desarrolla la teoría que proporcione la función de valor óptimo V.

<u>Definición:</u> Para una función $u \in B(I)$. se define la norma

$$\| u \| = \sup_{i \in I} | u(i) | .$$

<u>Definición:</u> Una función $T{:}B(I) \longrightarrow B(I)$ se dice ser de contracción si existe una constante $\beta<1$ tal que

$$\| T_u - T_v \| \leq \beta \| u - v \|$$

para todo $u,v \in B(I)$. [$u-v$ es la función cuyo valor en i es $u(i)-v(i)$].

El siguiente teorema se da sin prueba, ya que sólo se aplica para lograr nuestro propósito.

Teorema 1.5 (caso especial del teorema de punto fijo de Banach)

Si $T:B(I) \longrightarrow B(I)$ es una función de contracción, entonces existe una única función $g \in B(I)$ tal que

$Tg=g$,

y se dice que g es punto fijo de T. Además, para cualquier $u \in B(I)$,

$T^n u \longrightarrow g$ cuando n tiende a infinito.

A fin de aplicar este teorema, se define la función $T_\alpha:B(I) \longrightarrow B(I)$ de la siguiente manera

$$(T_\alpha u)(i)=\min_a\{C(i,a)+\alpha \sum_{j=0}^{\infty} P_{ij}(a)u(j)\} \qquad (1.11)$$

Si se aplica T_α a la función V se tiene

$$(T_\alpha V)(i)=\min_a\{C(i,a)+\alpha \sum_{j=0}^{\infty} P_{ij}(a)V(j)\}$$

(por el teorema 1.1)

$$=V(i).$$

Por lo tanto,

$T_\alpha V = V$.

Así, si se demuestra que T_α es una función de contracción, entoces, por el teorema 1.5, V es la única solución para (1.2); es decir, V es el único punto fijo de T_α.

Además, para cualquier $u \in B(I)$ y haciendo que n tienda a infinito entonces:

$$T_\alpha^n u \longrightarrow V$$

Es decir, V se obtiene aplicando sucesivamente T_α a cualquier función inicial $u \in B(I)$. Este es conocido como el método de <u>aproximaciones sucesivas</u>, o método de <u>iteración de valores</u>.

Para que el desarrollo anterior sea válido, se tiene que demostrar que la función T_α en (1.11) es una función de contracción.

<u>Teorema 1.5.</u>

La función T_α definida por (1.11) es una función de contracción.

<u>Demostración:</u>

Para cualquier $u,v \in B(I)$ por definición se tiene

$$|T_\alpha u(i)-T_\alpha v(i)|=|\min_a\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)u(j)\} -$$

$$\min_a\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)v(j)\}|$$

Si se hace uso de la siguiente propiedad

$$|\inf_x f(x)-\inf_x g(x)|\leq \sup_x|f(x)-g(x)|,$$

Entonces

$$|T_\alpha u(i)-T_\alpha v(i)|\leq \sup_a|\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)u(j)\}-$$

$$\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)v(j)\}|=$$

$$\alpha\sup_a|\sum_{j=0}^{\infty}P_{ij}(a)[u(j)-v(j)]|\leq$$

$$\alpha\sup_a\sum_{j=0}^{\infty}P_{ij}(a)|u(j)-v(j)|\leq$$

$$\alpha\,||u-v||\sum_{j=0}^{\infty}P_{ij}(a)=$$

$$\alpha\,||u-v||$$

Esta última igualdad es válida puesto que $\sum_{j=0}^{\infty}P_{ij}=1$.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

Del teorema de punto fijo y el teorema 1.5, se obtiene el siguiente corolario.

Corolario 1.6.

V es la única solución de la ecuación

$$V(i)=\min_{a}\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)V(j)\},\qquad i\in I.$$

Además, para cualquier $u\in B(I)$

$T_{\alpha}^{n}u \longrightarrow V$ cuando n tiende a infinito.

Ahora, se demuestra que la función T_f definida por (1.7) también es una función de contracción y puesto que $T_fV_f=V_f$ por teorema 1.2, entonces V_f es la única solución.

Para cualquiera dos funciones $u,v\in B(I)$, por definición de T en (1.7) se tiene

$$|(T_fu)(i)-(T_fv)(i)|=$$

$$|C[i,f(i)]+\alpha\sum_{j=0}^{\infty}P_{ij}[f(i)]u(j)-C[i,f(i)]-\alpha\sum_{j=0}^{\infty}P_{ij}[f(i)]v(j)|=$$

$$\alpha|\sum_{j=0}^{\infty}P_{ij}[f(i)][u(j)-v(j)]|\leq$$

$$\alpha\sum_{j=0}^{\infty}P_{ij}[f(i)]|u(j)-v(j)|\leq$$

$$\alpha \, || \, u-v \, || \sum_{j=o}^{\infty} P_{ij}[f(i)] =$$

$$\alpha \, || \, u-v \, || \text{ ,ya que } \sum_{j=o}^{\infty} P_{ij}[f(i)]=1.$$

Por lo que,

$$|| \, T_f u - T_f v \, || \leq \alpha \, || \, u-v \, || \, .$$

$\mathscr{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

1.4 Aplicaciones.

Una vez presentado el desarrollo de la teoría para el criterio de los costos descontados, se presenta alguna aplicación que ilustra los resultados obtenidos.

Modelo para el reemplazo de una máquina.

Considérese una máquina que puede estar en uno de los siguientes estados 0,1 2,... . Supóngase que al comienzo de cada día se ve el estado de la máquina y se toma una decisión sobre reemplazar o no la máquina. Si se tomó la decisión de reemplazar, se supone que la máquina es reemplazada inmediatamente por una nueva, por lo que el estado de la máquina es 0.

El costo de reemplazar una máquina se denota por R. Además se supone que se incurre en un costo de mantenimiento $C(i)$, cada día que la máquina se encuentra en el estado i.

También se tienen las probabilidades de transición del estado i al estado j denotadas por P_{ij}; es decir, al comienzo de un día se encuentra en el estado i y al comienzo del día siguiente se encuentra en el estado j.

Se tiene que lo anterior es un modelo de **decisión Markoviano** con dos acciones. En la cual, la acción 1) es la de reemplazar y la acción 2) es la de no reemplazar.

Los costos para cada etapa y las probabilidades de transición están dadas como sigue:

a) si se toma la decisión 1 (reemplazar), el costo es

$C(i,1)=R+C(0)$

donde

R costo de reemplazo

$C(0)$ costo de estar en el estado 0

$C(i,1)$ costo de estar en el estado i, dado que se tomó la decisión 1

$P_{ij}(1)=P_{oj}$, dado que se tomó la decisión 1 y como consecuencia se regresa al estado 0;

$P_{ij}(1)$, probabilidad de transición del estado i a j, dado que se tomó la decisión 1

b) si se toma la decisión 2 (no reemplazar), se tiene

$C(i,2)=C(i) \qquad i \in I,$

donde

$C(i,2)$ costo de estar en el estado i, dado que se tomó la decisión 2

$C(i)$ costo de mantenimiento, cuando el estado de la máquina es i.

$P_{ij}(2)=P_{ij} \ , \qquad i \in I,$

donde

$P_{ij}(2)$, probabilidad de transición del estado i a j, dado que se tomó la decisión 2.

Además, se tienen las siguientes suposiciones en los costos y probabilidades de transición.

i) $\{C(i), \ i \in I\}$ es una secuencia creciente y acotada, es decir

$C(0) < C(1) < C(2) < \ldots \leq C$

ii) la $\sum_{j=k}^{\infty} P_{ij}$ es una función creciente de i para cada $k \geq 0$

Lema 1.7.

La suposición ii) implica que para cualquier función creciente y no negativa $h(i)$, la función

$$\sum_{j=0}^{\infty} P_{ij} h(j)$$

es también creciente en i

Demostración:

Por hipótesis

$$\sum_{j=0}^{\infty} P_{1j} < \sum_{j=0}^{\infty} P_{2j} < \sum_{j=0}^{\infty} P_{3j} < \ldots < \sum_{j=0}^{\infty} P_{ij} < \ldots$$

Multiplicando por $h(i)$

$$\sum_{j=0}^{\infty} P_{1j} h(i) < \sum_{j=0}^{\infty} P_{2j} h(i) < \sum_{j=0}^{\infty} P_{3j} h(i) < \ldots < \sum_{j=0}^{\infty} P_{ij} h(i) < \ldots$$

De aquí se sigue que

$\sum_{j=0}^{\infty} P_{ij} h(i)$ es una función creciente de i.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

Lema 1.8.

Si se consideran las suposiciones i) y ii), entonces $V(i)$ es creciente en i

Demostración:

Sea $V_o(i) = min\ C(i,\alpha)$; donde $\alpha \in A=\{1,2\}$

$$= min\{C(i,1),C(i,2)\}$$

$$= min\{R+C(0),C(i)\} , \qquad i \in I.$$

Donde $R+C(0)$, es el costo que se incurre dado que se está en el estado i y se toma la decisión 1. Es decir, es el costo de reemplazar más el costo de regresar al estado 0.

$C(i)$, es el costo que se incurre dado que se está en el estado i.

Para $n>1$; es decir, para más de una etapa, recursivamente se define

$$V_n \cong T^n_\alpha V_o = T_\alpha V_{n-1}$$

[Por definición de T_α [Ver (1.11)]]

$$V_n(i) \cong \min_{\alpha \in A}\{C(i,\alpha) + \alpha \sum_{j=o}^{\infty} P_{ij}(\alpha)\ V_{n-1}(j)\}$$

$$= \min\{C(i,1)+\alpha \sum_{j=0}^{\infty} P_{ij}(1)\; V_{n-1}(j)\;,$$

$$C(i,2)+\alpha \sum_{j=0}^{\infty} P_{ij}(2)\; V_{n-1}(j)\}$$

$$= \min\{R+C(0)+\alpha \sum_{j=0}^{\infty} P_{0j}\; V_{n-1}(j),$$

$$C(i)+\alpha \sum_{j=0}^{\infty} P_{ij}\; V_{n-1}(j)\}$$

Donde

$V_{n-1}(j)$, costo óptimo incurrido en el estado j para $(n-1)$ etapas,

$\alpha \sum_{j=0}^{\infty} P_{0j} V_{n-1}$, costo descontado esperado, dado que se está en el estado 0 y existe una probabilidad P_{0j}, de pasar al estado j en la etapa n.

De la suposición i) se tiene que $V_0(i)$ es creciente en i. Supóngase que V_{n-1} es creciente en i, entonces por el lema 1.7 $V_n(i)$ es creciente en i para toda n.

Por otro lado, por el teorema 1.4 (caso especial del teorema de punto fijo de Banach) y el método de aproximaciones sucesivas

$$V_n = T_{\alpha}^{n} V_0 \longrightarrow V$$

Por lo tanto, ya que $V_n(i)$ es creciente en i, $V(i)$ también es creciente en i.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

Para poder determinar, la política óptima del modelo de reemplazo de una máquina, se demuestra el siguiente teorema.

Teorema 1.9.

Bajo las suposiciones i) y ii), existe un entero i^*, $i^*\leq\infty$, tal que una política óptima reemplaza para $i>i^*$ y no reemplaza para $i\leq i^*$.

Demostración:

Por el teorema 1.1 se tiene

$$V(i)=min\{R+C(0)+\alpha\sum_{j=o}^{\infty}P_{oj}\ V(j),$$

$$C(i)+\alpha\sum_{j=o}^{\infty}P_{ij}V(j)\}\ ,\quad i\in I \qquad (1.16)$$

Sea

$$i^*=max\{i:C(i)+\alpha\sum_{j=o}^{\infty}P_{ij}\ V(j)\leq$$

$$R+C(0)+\alpha\sum_{j=o}^{\infty}P_{oj}\ V(j)\}$$

En palabras i^* es el mayor estado i cuyo costo asociado es menor o igual al costo asociado al estado cero. La igualdad se cumple si el estado i es el estado cero.

Ahora, por los dos lemas anteriores

$C(i)+\alpha \sum_{j=0}^{\infty} P_{ij} V(j)$ es creciente en i y, por lo tanto, por (1.16),

$$V(i)=\begin{cases} C(i)+\alpha \sum_{j=0}^{\infty} P_{ij}\, V(j) & \text{para } i\le i^* \Rightarrow \text{no reemplazar} \\ \\ R+C(0)+\alpha \sum_{j=0}^{\infty} P_{0j}\, V(j) & \text{para } i> i^* \Rightarrow \text{reemplazar.} \end{cases}$$

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

CAPITULO II

COSTOS POSITIVOS SIN DESCUENTO

2.1 Ecuación Funcional.

En este capítulo se utiliza el criterio de los costos positivos sin descuento, el cual supone que todos los costos son no negativos; es decir, $C(i,a) \geq 0$ para toda i,a.

No asume, ningún factor de descuento, ni requiere que los costos sean acotados e intenta minimizar el costo total esperado, es decir, busca $\inf_{\pi} V_{\pi}(i)$, donde

$$V_{\pi}(i) = E_{\pi}\left[\sum_{t=0}^{\infty} C(X_t, a_t) \mid X_0 = i\right], \quad i \in I \qquad (2.1)$$

$I = \{0,1,2,\ldots\}$

E_{π} esperanza condicional cuando se emplea la política π

X_t estado del proceso al tiempo t

X_0 estado inicial del proceso

a_t acción tomada en t

$C(X_t, \alpha_t)$ costo incurrido en la etapa t.

Por otra parte, sea

$$V(i) = \inf_{\pi} V_\pi(i), \qquad i \in I.$$

Es decir, $V(i)$ es la máxima cota inferior de la función $V_\pi(i)$.

Debido a la suposición de que $C(i,\alpha) \geq 0$ y no necesariamente acotado, es posible que $V(i)$ sea una función infinita, en cuyo caso el problema es trivial: cualquier política es óptima. Luego, el modelo sólo será de interés si el problema permite tener una función costo $V(i) < \infty$ para al menos algunos valores de i.

Una política π^* se dice que es óptima si

$$V_{\pi^*}(i) = \inf_{\pi} V_\pi(i), \text{ para todo } i \in I.$$

En palabras, la política π^* es óptima si el costo esperado en (2.1) es el mínimo para todo estado inicial cuando se usa π^*.

De manera análoga al capítulo I, se presenta el siguiente teorema (ecuación de programación dinámica).

Teorema 2.1.

$$V(i)=\min_{a}\{C(i,a)+\sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\},\quad i\geq 0 \tag{2.2}$$

Demostración:

Sea π una política arbitraria, y supóngase que π toma la acción a en la etapa 0 con probabilidad P_a, $a \in A$. (recuérdese que A es el conjunto de todas las acciones posibles). Entonces

$$V_\pi(i)=\sum_{a\in A} P_a\{C(i,a)+\sum_{j=0}^{\infty} P_{ij}(a)\ W_\pi(j)\},\quad i\in I,$$

donde

$C(i,a)$ costo que se incurre cuando se está en el estado i y se toma la decisión a;

$W_\pi(j)$ costo esperado incurrido para una etapa adelante, cuando se usa la política π y el estado siguiente es j.

Por definición de V (costo mínimo) tenemos que

$$W_\pi(j)\geq V(j).$$

Por lo que

$$V_\pi(i)\geq \sum_{a\in A} P_a\{C(i,a)+\sum_{j=0}^{\infty} P_{ij}(a)\ V(j)\}\geq$$

$$\sum_{\alpha\in A} P_{\alpha} \min_{\alpha}\{C(i,\alpha)+ \sum_{j=0}^{\infty} P_{ij}(\alpha) V(j)\}= \tag{2.3}$$

$$\min_{\alpha}\{C(i,\alpha)+ \sum_{j=0}^{\infty} P_{ij}(\alpha) V(j)\};$$

La igualdad es válida ya $\sum_{\alpha\in A} P_{\alpha}=1$.

Debido a que π es arbitraria, (2.3) implica que

$$V(i)\geq\min_{\alpha}\{C(i,\alpha)+ \sum_{j=0}^{\infty} P_{ij}(\alpha) V(j)\} \tag{2.4}$$

Por otro lado sea α_o tal que

$$C(i,\alpha_o)+ \sum_{j=0}^{\infty} P_{ij}(\alpha_o) V(j)=\min_{\alpha}\{C(i,\alpha)+ \sum_{j=0}^{\infty} P_{ij}(\alpha) V(j)\} \tag{2.5}$$

Y, dado ε arbitrario, sea π la política que selecciona la decisión α_o en la etapa 0; y si el próximo estado es j, entonces el proceso pasa al estado j y se sigue una política π_j tal que

$$V_{\pi_j}(j)\leq V(j)+\varepsilon, \quad j \in I.$$

Puesto que π escoge α_o en la etapa 0; y se sigue la política π_j para el estado j se tiene

$$V_\pi(i)=C(i,a_o)+\sum_{j=o}^{\infty} P_{ij}(a_o)\, V_{\pi_j}(j)\leq$$

(por la desigualdad anterior)

$$C(i,a_o)+\sum_{j=o}^{\infty} P_{ij}(a_o)\, V(j)+\varepsilon.$$

Por definición, $V(i)\leq V_\pi(i)$, lo cual implica que

$$V(i)\leq C(i,a_o)+\sum_{j=o}^{\infty} P_{ij}(a_o)\, V(j)+\varepsilon$$

del lado derecho de (2.5), se obtiene

$$V(i)\leq \min_a\{C(i,a)+\sum_{j=o}^{\infty} P_{ij}(a)\, V(j)\}+\varepsilon \tag{2.6}$$

La ecuación (2.2) se sigue de (2.4) y (2.6), por ser ε arbitrario.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

2.2 Políticas Estacionarias.

Se denota por $N(I)$, al conjunto de todas las funciones no negativas (posiblemente con valores infinitos) definidas sobre el espacio de estados, I. Para cualquier política estacionaria f se define la función

$$T_f : N(I) \longrightarrow N(I)$$

a través de la siguiente expresión

$$(T_f u)(i) = C[i, f(i)] + \sum_{j=0}^{\infty} P_{ij}\,[f(i)]\; u(j), \qquad (2.7)$$

donde $u \in N(I)$.

Nótese en el lado derecho de la expresión (2.7), que tanto u como C son funciones no negativas, por lo que $T_f u(i)$ es no negativa y posiblemente con valores infinitos.

Se interpreta a $T_f u$ evaluado en i, como el costo esperado dado que se utiliza la política estacionaria f y el proceso termina un período después incurriendo en un costo final $u(j)$, cuando el estado final es j.

Ahora se presenta la teoría concerniente ala función $T_f u$ que será de suma importancia en la sección siguiente.

La teoría que se desarrolla muestra algunas de las propiedades de la función $T_f u$ y de las funciones que pertenecen a $N(I)$.

Lema 2.2.

Para $u, v \in N(I)$, y una política estacionaria f,

i) $u \le v \Rightarrow T_f u \le T_f v$;

ii) $T_f V = V_f$;

iii) $(T_f^n 0)(i) \longrightarrow V_f(i)$ cuando $n \longrightarrow \infty$ para cada i, en donde 0 es la función cero, y $T_f^n \cong T_f(T_f^{n-1})$ para $n=1,2,..$ con con $T_f^0 \cong$ identidad.

Demostración:

Parte i)

Si $u \leq v$ y aplicando T_f en ambos lados de la desigualdad, se tiene

$$T_f u \leq T_f v$$

Parte ii)

Por definición de T_f [ver (2.7)], tenemos

$$(T_f V_f)(i) = C[i, f(i)] + \sum_{j=0}^{\infty} P_{ij}[f(i)]\, V_f(j)$$

$$= V_f(i);$$

Por lo que,

$$T_f V_f = V_f.$$

Parte iii). Primero se observa que

$T_f^2=T_f(T_f)$, por definición de T_f se tiene

$$(T_f^2u)(i)=C[i,f(i)]+\sum_{j=0}^{\infty}P_{ij}\ [f(i)](T_fu)(j)$$

Aplicando la definición de T_f a u

$$(T_f^2u)(i)=C[i,f(i)]+\sum_{j=0}^{\infty}P_{ij}\ [f(i)]*$$

$$\{C[j,f(j)]+\sum_{k=0}^{\infty}P_{jk}\ [f(j)]\ u(k)\}$$

(haciendo operaciones)

$$=C[i,f(i)]+\sum_{j=0}^{\infty}P_{ij}\ [f(i)]\ C[j,f(j)]+$$

$$\sum_{j=0}^{\infty}\ \sum_{k=0}^{\infty}\ P_{ij}\ [f(i)]\ P_{jk}\ [f(j)]u(k).$$

En particular, si u es la función cero, entonces

$$(T_f^2 0)(i)=C[i,f(i)]+\sum_{j=0}^{\infty}P_{ij}\ [f(i)]\ C[j,f(j)]$$

$$=E_f\{\sum_{t=0}^{1}C[X_t,f(X_t)]\ |\ X_0=i\}$$

Por inducción, se tiene que T_f^n representa el costo total esperado en n-etapas, es decir

$$(T_f^n 0)(i)=E_f\{\sum_{t=0}^{n-1}C[X_t,f(X_t)]\ |\ X_0=i\}.$$

Si se toma el límite cuando $n \longrightarrow \infty$

$$\lim_{n \to \infty} (T_f^n 0)(i) = \lim_{n \to \infty} E_f\{ \sum_{t=0}^{n-1} C[X_t, f(X_t)] \mid X_0 = i \}$$

$$= E_f\{ \sum_{t=0}^{\infty} C[X_t, f(X_t)] \mid X_0 = i \}$$

$$= V_f(i).$$

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

Nótese que el resultado del inciso iii) es cierto si u es la función cero, y no necesariamente para cualquier $u \in N(I)$.

2.3 Teorema de Programación Dinámica.

Teorema 2.3.

Si f_1 es la política estacionaria que selecciona la acción que minimiza el lado derecho de (2.2) para cada estado i, es decir, $f_1(i)$ es tal que

$$C[i,f_1(i)]+\sum_{j=0}^{\infty}P_{ij}[f_1(i)]\,V(j)=$$

$$\min_{a}\{C(i,a)+\sum_{j=0}^{\infty}P_{ij}(a)\,V(j)\},\qquad i\in I,$$

entonces

$V_{f_1}(i)=V(i)$, para todo $i\in I$,

y por lo tanto f_1 es óptima.

<u>Demostración</u>:

Aplicando la función T_{f_1} definida por (2.7) a V, se obtiene

$$(T_{f_1}V)(i)=C[i,f_1(i)]+\sum_{j=0}^{\infty}P_{ij}[f_1(i)]\,V(j)$$

(por hipótesis)

$$=\min_{a}\{C(i,a)+\sum_{j=0}^{\infty}P_{ij}(a)\,V(j)\}$$

(por teorema 2.1)

$$=V(i).$$

Por lo que,

$T_{f_1}V=V.$

Como $C(i,a)\geq 0$, se tiene que $V\geq 0$ y puesto que T_{f_1} es una función no decreciente, entonces

$$T_{f_1} 0 \leq T_{f_1} V = V.$$

Aplicando T_{f_1} nuevamente,

$$T^2_{f_1} 0 \leq T_{f_1} (T_{f_1} V) = T_{f_1} V = V$$

y por una aplicación sucesiva de T_{f_1}, se tiene

$$T^n_{f_1} 0 \leq V.$$

Si n tiende a infinito y haciendo uso del lema 2.2 inciso iii) se tiene que

$$T^n_{f_1} 0 \longrightarrow V_{f_1}$$

Luego, cuando n tiende a infinito

$$V_{f_1} \leq V$$

Por otra parte, por definición, $V_{f_1} \geq V$ luego se sigue que

$$V_{f_1} = V.$$

Por lo tanto f_1 es una política óptima.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

La solución de (2.2) no necesariamente es única, debido a que no existe una función de contracción que garantice su unicidad, como sucedió en el criterio de los costos descontados.

2.4 Aplicaciones

En esta sección se presenta alguna aplicación que ilustra la teoría desarrollada en el capítulo.

Modelo de Paro Optimo.

Considérese un proceso con espacio de estados $\{0,1,2,..\}$, el cual, cuando se encuentra en el estado i, se puede ya sea parar (acción 1) y recibir un pago final $R(i)$, o (acción 2) pagar un costo $C(i)$ e ir al próximo estado de acuerdo a las probabilidades de transición P_{ij}; $i,j\geq 0$.

Se dice que el proceso se encuentra en el estado "infinito" (∞) cuando se escoge la acción de parar. Una vez que el proceso pasa al estado ∞, permanece ahí con probabilidad 1. Así, se tiene un proceso Markoviano de decisión con dos acciones, costos y probabilidades de transición dados por:

$$C(i,1)=-R(i) \qquad i=0,1,2,...$$

$$C(i,2)=C(i) \qquad i=0,1,2,...$$

$$C(\infty,.)=0$$

$$P_{i\infty}(1)=1 \qquad i=0,1,2,...$$

$$P_{ij}(2)=P_{ij} \qquad i,j=0,1,2,...$$

$P_{\infty\,\infty}(\cdot)=1$

Se supone que

i) $\inf\limits_{i\geq o} C(i)>0$

ii) $\sup\limits_{i} R(i)<\infty$

Puesto que $R(i)$ es interpretado como un costo negativo, y debido a que los resultados anteriores sólo son válidos para costos no negativos, se transforma el modelo original a uno equivalente en donde se pueda aplicar el criterio de los <u>Costos positivos sin descuento</u>.

Si se hace $R=\sup\limits_{i} R(i)$, y se considerá un proceso relacionado el cual es tal que, cuando se está en el estado i, se puede ya sea parar y pagar un costo final $R-R(i)$, o bién pagar un costo $C(i)$ e ir al próximo estado de acuerdo a las probabilidades de transición P_{ij} ; $i,j\geq 0$.

Para cualquier política π, si V_{π} denota el costo esperado con respecto al proceso original cuando se usa π y V_{π}^{*} denota el costo esperado con respecto al proceso relacionado, entonces para cualquier política π que *para* en un tiempo finito esperado, se tiene

$$V_{\pi}^{*}(i)=E_{\pi}\{\sum_{t=o}^{N-1} C(X_t)+R-R(X_N) \mid X_o=i\}$$

también, puesto que para $t>n$ el proceso se encuentra en el estado ∞ y esto implica $C(\infty,.)=0$, entonces

$$=E_{\pi}\{\sum_{t=0}^{\infty} C(X_t)-R(X_N) \mid X_0=i\}+R=V_{\pi}(i)+R$$

Nótese que sólo estas políticas son factibles, puesto que por la suposición i), cualquier política que no *para* en un tiempo finito esperado, se tiene:

$$V_{\pi}(i)=V'_{\pi}(i)=\infty$$

Es decir, que el costo asociado tiende a infinito; por lo que ésta política no es óptima.

Entonces, cualquier política que es óptima para el proceso original es también óptima para el proceso relacionado y viceversa.

Sin embargo, el proceso relacionado es un proceso de decisión Markoviano con costos no negativos y por los resultados de la sección anterior, existe una política óptima y por (2.2) el costo óptimo $V'(i)$ satisface

$$V'(i)=min\{R-R(i),C(i)+\sum_{j=0}^{\infty} P_{ij}V'(j)\},\quad i=0,1,2,...$$

Además, la política que escoge la acción que minimiza la expresión anterior es óptima.

Poniendo estos resultados en términos de la función de costo óptimo $V(i)=V'(i)-R$ del proceso original, se tiene:

$$V(i)=min\{-R(i),C(i)+\sum_{j=0}^{\infty}P_{ij}V(j)\},\ i=0,1,2,...$$

La política que escoge la decisión que minimiza el lado derecho de la expresión anterior para cada $i\geq 0$ es óptima.

Sea $V_0(i)=\min_{a\in A}\{C(i,a)\}=min\{C(i,2),C(i,1)\}$

$$=min\{C(i),-R(i)\}$$

$$=-R(i),$$

y para $n>0$,

$$V_n(i)\cong TV_{n-1}(i)\cong$$

$$\min_{a\in A}\{C(i,a)+\sum_{j=0}^{\infty}P_{ij}(a)V_{n-1}(j)\}=$$

$$min\{-R(i),C(i)+\sum_{j=0}^{\infty}P_{ij}\ V_{n-1}(j)\},\ i=0,1,2,...$$

En palabras, $V_n(i)$ es el mínimo costo esperado si se comienza en i, y se permite ir a lo más n-etapas antes de parar. De esta interpretación se sigue que, para toda i,n

Por inducción:

$n=1 \quad V_1(i)=min\{-R(i),C(i)+\sum_{j=0}^{\infty}P_{ij}\,V_0(j)\}\leq$

$(min\ \{a,b\}\ \leq a,\ \leq b)$

$-R(i)=$

$V_0(i)$

$n>1$: supóngase (hipótesis de inducción) que

$V_{n-1}(.)\geq V_n(.)$.

Entonces

$$V_n(i)=min\{-R(i),C(i)+\sum_{j=0}^{\infty}P_{ij}\,V_{n-1}(j)\}\geq$$

$$min\{-R(i),C(i)+\sum_{j=0}^{\infty}P_{ij}\,V_n(j)\}=$$

$$V_{n+1}(i)$$

Por definición de V (costo mínimo) se tiene

$$V(j)\leq V_n(j)$$

Por lo que

$$V_{n+1}(i)\geq min\{-R(i),C(i)+\sum_{j=0}^{\infty}P_{ij}V(j)\}=$$

$$V(i)$$

Luego se sigue que,

$$V_n(i) \geq V_{n+1}(i) \geq V(i)$$

Así

$$\lim_{n\to\infty} V_n(i) \geq V(i)$$

Definición: Se dice que el proceso es estable si

$$\lim_{n\to\infty} V_n(i) = V(i) \qquad i=0,1,2,\ldots$$

Ahora, sea

$$B=\{i : -R(i) \leq C(i) - \sum_{j=0}^{\infty} P_{ij}\ R(j)\}$$

$$=\{i : R(i) \geq \sum_{j=0}^{\infty} P_{ij}\ R(j) - C(i)\}$$

es decir, B representa el conjunto de estados para el cual *parar* es al menos tan bueno como continuar un período más y entonces *parar*.

En seguida se presenta un teorema para el cual, bajo ciertas condiciones y haciendo uso del conjunto B, proporciona la política de paro óptimo.

Teorema 2.4.

Si el proceso es estable, y si $P_{ij}=0$ para $i \in B$, $j \notin B$, entonces la política óptima para en i si y solo si $i \in B$.

Demostración:

Se demuestra primero que si $i \in B$, entonces la decisión óptima es parar, para toda n; es decir,

$$V_n(i)=-R(i) \quad \text{si } i \in B.$$

Para $n=0$ se sigue trivialmente que $V_0(i)=-R(i)$.

Supóngase que se cumple para $n-1$; es decir, $V_{n-1}(i)=-R(i)$, $i \in B$.

Entonces, para $i \in B$,

$$V_n(i)=min(-R(i),C(i) + \sum_{j=0}^{\infty} P_{ij} V_{n-1}(j))$$

$$=min(-R(i),C(i) + \sum_{j\in B} P_{ij} V_{n-1}(j))$$

(debido a que $P_{ij}=0$ para $j \notin B$)

$$=min(-R(i),C(i) - \sum_{j\in B} P_{ij} R(j)) \; .$$

[la igualdad es válida debido a la suposición para la etapa $n-1$, cuyo costo generado es $-R(j)$]

$$=-R(i)$$

Esta última igualdad es válida puesto que $i \in B$; es decir,

$$-R(i) \leq C(i) - \sum_{j=o}^{\infty} P_{ij} R(j) \ .$$

Por lo tanto,

$V_n(i) = -R(i)$ para todo $i \in B$, y toda n.

Haciendo que n tienda a infinito y usando la hipótesis de estabilidad, se tiene

$V(i) = -R(i)$ para $i \in B$.

Reciprocamente,

Si $i \notin B$, la política que continua un período más y después *para* tiene un costo esperado, dado por

$$C(i) - \sum_{j=o}^{\infty} P_{ij} R(j)$$

el cual es estrictamente menor que $-R(i)$, puesto que $i \notin B$.

Entonces

$$V(i) \begin{cases} = -R(i) & \text{para } i \in B \\ < -R(i) & \text{para } i \notin B \end{cases}$$

L.Q.Q.D.

A continuación se presenta un ejemplo en el que se aplican los resultados del teorema 2.3.

Supóngase que un individuo desea vender su casa y recibe una oferta al principio de cada día. Supóngase también que las ofertas sucesivas son independientes y que una oferta j es hecha con probabilidad P_j, $j=0,1,2,\ldots,N$. Se supone que si una oferta es rechazada en este momento, puede ser aceptada en cualquier otro día. También, se incurre en un costo de mantenimiento C cada día que la casa no es vendida.

El estado en el tiempo t "es la mayor" oferta recibida hasta este tiempo (incluyendo t). Entonces los valores para P_{ij} serán:

a) 0; $j<i$

Caso en el que la oferta es menor a la actual

b) $\sum_{k=0}^{i} P_k$; $j=i$

Caso en donde la oferta es igual a la actual

c) P_j; $j>i$

Caso en donde la oferta es mayor a la actual

Así, el conjunto B en el teorema 2.3 es

$$B=\{i:\ -i\leq C\ -i\sum_{k=o}^{i}P_k\ -\sum_{j=i+1}^{N}jP_j\}.$$

$$=\{i:\ C\geq\sum_{j=i+1}^{N}jP_j\ +\ i\sum_{k=o}^{i}P_k\ -\ i\sum_{k=o}^{N}P_k\ \}$$

(puesto que $\sum_{k=o}^{N}P_k=1$)

$$=\{i:\ C\geq\sum_{j=i+1}^{N}jP_j\ +\ i\sum_{k=o}^{i}P_k\ -\ i\ [\sum_{k=o}^{i}P_k\ +\sum_{k=i+1}^{N}P_k]\}.$$

$$=\{i:\ C\geq\sum_{j=i+1}^{N}jP_j\ -\ i\sum_{k=i+1}^{N}P_k\}$$

$$=\{i:\ C\geq\sum_{j=i+1}^{N}(j-i)P_j\}\qquad(2.8)$$

Es decir, $i\in B$ si el costo de mantenimiento, C, es mayor o igual al costo esperado de la diferencia entre la oferta actual, i, y la siguiente.

Puesto que el lado derecho de (2.8) es decreciente en i, se sigue que

$B=\{i^*,\ i^*+1,\ \ldots,\ N\}$,

Donde

$$i^*=min\{i : C \geq \sum_{j=i+1}^{N} (j-i)P_j\}.$$

Entonces, $P_{ij}=0$ para $i \in B$, $j \notin B$; y la política óptima es aquella que acepta la primera oferta que sea al menos i^*.

CAPITULO III

COSTO PROMEDIO ESPERADO

3.1 Condiciones para la existencia de políticas estacionarias óptimas.

En este capítulo se utiliza el criterio del costo promedio esperado; en el cual, el objetivo es minimizar el costo promedio esperado a la larga por unidad de tiempo que se define en la ecuación (3.1). Se supone que los costos $C(i,a)$ son acotados.

Para cualquier política π, se define el costo promedio esperado cuando se usa la política π, dado el estado inicial X_0, como:

$$\phi_\pi(i) = \lim_{n \to \infty} E_\pi\left[\frac{\sum_{t=0}^{n} C(X_t, a_t)}{n+1} \;\middle|\; X_0 = i \right] \tag{3.1}$$

Si el límite en (3.1) no existe, se tomará entonces el límite superior.

Se dice que la política π^* es óptima en costo promedio si

$\phi_{\pi^*}(i) = \min_{\pi} \phi_\pi(i)$ para toda i

En el siguiente ejemplo se muestra que una política óptima no necesariamente existe.

Ejemplo 1

El espacio de estados es el conjunto $\{1,1',2,2',3,3',\ldots\}$. Se tienen 2 acciones, y las probabilidades de transición están dadas por

$$\begin{aligned} &P_{i\,i+1}(1)=P_{i\,i'}(2)=1 \\ & \qquad\qquad\qquad\qquad i=1,2,\ldots \\ &P_{i'\,i'}(1)=P_{i'\,i'}(2)=1 \end{aligned}$$

los costos sólo dependen del estado presente y son dados por

$$\begin{aligned} &C(i,.)=1 \\ & \qquad\qquad\qquad\qquad i=1,2,\ldots \\ &C(i',.)=1/i. \end{aligned}$$

Es decir, cuando el proceso se encuentra en el estado i y se toma la decisión 1, el proceso pasa al estado $i+1$ con un costo de 1, mientras que si se toma la decisión 2, el proceso pasa al estado i' con un costo $1/i$.

Supóngase que el proceso comienza en el estado 1; es decir, $X_0=1$, y sea π cualquier política. Se tienen dos casos.

Caso 1)

Con probabilidad 1, π siempre escoge la acción 1. Entonces

$$\phi_{\pi}(1)=1>0;$$

es decir, el costo esperado, dado el estado inicial $X_o=1$, es 1

Caso 2)

Con probabilidad positiva, π escoge la acción 2 en algún tiempo. En este caso, se tiene que para alguna n, existe una probabilidad positiva $P_n>0$ de que π tome la acción 2 cuando se está en el estado n.

Sea $\tilde{n}$ la más pequeña n con esta propiedad, entonces

$$\phi_{\pi}(1)=P_{\tilde{n}}[1\ /\tilde{n}]\ >0$$

por lo que, en ambos casos, $\phi_{\pi}(1)>0$. Sin embargo, si se escoge valores muy grandes de $\tilde{n}$ el costo promedio asociado puede acercarse a cero tanto como se desee.

Esto es, una política óptima no existe.

Ahora la pregunta sería si se puede restringir las políticas a ser estacionarias. La respuesta es dada en el siguiente ejemplo.

Ejemplo 2

Supóngase que el espacio de estados es el conjunto de los enteros positivos $\{1,2,...\}$ y que se tienen 2 acciones. Los costos y las probabilidades de transición están dados por:

$$P_{i\,i+1}(1)=1=P_{ii}(2)$$

$$\left.\begin{array}{l} C(i,1)=1 \\ C(i,2)=1/i \end{array}\right. \qquad i=1,2,...$$

En palabras, si el estado presente es i y se toma la decisión 1, se pasa al estado $i+1$ y se incurre en un costo de 1 unidad; si se toma la decisión 2, se permanece en i y se incurre en un costo de $1/i$ unidades.

Supóngase que el proceso comienza en el estado 1; es decir, $X_o=1$, y sea π cualquier política estacionaria. Se tienen 2 casos

Caso 1

π siempre escoge la acción 1, por lo que

$$\phi_\pi(1)=1$$

Caso 2

π escoge la acción 2 en el primer tiempo que ocurre el estado n.

Es decir, el proceso se desarrolla del estado 1 al estado 2 al estado 3, hasta el estado n. Cuando se encuentra en el esatdo n toma la decisión 2; por lo tanto, permanece en este estado. El correspondiente costo promedio es:

$$\phi_{\pi}(1)=1/n>0.$$

Entonces para cualquier política estacionaria π

$$\phi_{\pi}(1)>0.$$

Si π^* es la política no estacionaria la cual, cuando el proceso alcanza el estado i por vez primera, escoge la acción 2 durante i tiempos consecutivos, y después toma la acción 1. Dado que el estado inicial es $X_o=1$, entonces los costos sucesivos incurridos bajo π^* serán:

1, 1, 1/2, 1/2, 1, 1/3, 1/3, 1/3, 1, 1/4, 1/4, 1/4, 1/4, 1, 1/5,...

El costo promedio de la anterior secuencia converge a cero, así

$$\phi_{\pi}*(1)=0$$

Por lo tanto, la política no estacionaria π^*es mejor que toda política estacionaria.

Nótese que π^* es una política no estacionaria, puesto que está en función del tiempo; es decir, que se toma la decisión dependiendo del tiempo en que se encuentre el proceso.

3.2 Casos Especiales.

Teorema 3.1.

Sea $I=\{0,1,2,...\}$ el espacio de estados. Si existe una función acotada $h(i)$ y una constante g tal que

$$g+h(i)=\min_{a}\{C(i,a)+\sum_{j=0}^{\infty}P_{ij}(a)h(j)\}, \quad i \in I, \qquad (3.2)$$

entonces existe una política estacionaria π^* tal que

$$g=\phi_{\pi^*}(i)=\min_{a}\ \phi_{\pi}(i), \text{ para toda } i \in I,$$

y π^* es cualquier política que, para cada i, preescribe una acción que minimiza el lado derecho de (3.2).

Demostración:

Sea $H_t=(X_0, a_0, ...,X_t, a_t)$; es decir, H_t denota la historia del proceso hasta el tiempo t. Para cualquier política π,

$$E_{\pi}\{\sum_{t=1}^{n}[h(X_t)-E_{\pi}(h(X_t)\mid H_{t-1})]\}=$$

$$E_{\pi}\{\sum_{t=1}^{n}h(X_t)-\sum_{t=1}^{n}E_{\pi}[(h(X_t)\mid H_{t-1}]\}=$$

$$E_{\pi}\{\sum_{t=1}^{n}h(X_t)\}-E_{\pi}\{\sum_{t=1}^{n}E_{\pi}[(h(X_t)\mid H_{t-1}]\}=$$

$$E_{\pi}\{\sum_{t=1}^{n}h(X_t)\}-E_{\pi}\{\sum_{t=1}^{n}h(X_t)\}=0$$

Pero

$$E_{\pi}[h(X_t)\mid H_{t-1}]=\sum_{j=0}^{\infty}h(j)P_{X_{t-1}\,j}(a_{t-1})=$$

$$C(X_{t-1},a_{t-1})+\sum_{j=0}^{\infty}h(j)P_{X_{t-1}\,j}(a_{t-1})-C(X_{t-1},a_{t-1})\geq$$

$$\min_{a}\{C(X_{t-1},a)+\sum_{j=0}^{\infty}h(j)P_{X_{t-1}\,j}(a)\}-C(X_{t-1},a_{t-1})=$$

(por hipótesis)

$$g+h(X_{t-1})-C(X_{t-1},a_{t-1})$$

Entonces

$$E_\pi\{\sum_{t=1}^{n}[h(X_t)-g-h(X_{t-1})+C(X_{t-1},a_{t-1})]\}=$$

$$E_\pi\{h(X_n)-h(X_0)-ng+\sum_{t=1}^{n}C(X_{t-1},a_{t-1})\}=$$

$$E_\pi h(X_n)-E_\pi h(X_0)-ng+E_\pi[\sum_{t=1}^{n}C(X_{t-1},a_{t-1})]\geq 0$$

Haciendo operaciones

$$E_\pi h(X_n)-E_\pi h(X_0)+E_\pi[\sum_{t=1}^{n}C(X_{t-1},a_{t-1})]\geq ng$$

$$E_\pi \frac{h(X_n)}{n}-E_\pi \frac{h(X_0)}{n}+E_\pi \frac{[\sum_{t=1}^{n}C(X_{t-1},a_{t-1})]}{n}\geq g$$

Si se toma el límite cuando n tiende a infinito en ambos lados de la desigualdad y usando el hecho de que h es acotada, se tiene:

$$\lim_{n\to\infty} E_\pi \frac{[\sum_{t=1}^{n}C(X_{t-1},a_{t-1})]}{n}\geq g$$

para cualquier estado inicial $X_0=i$. Por lo que,

$$\lim_{n\to\infty} E_\pi \frac{[\sum_{t=0}^{n}C(X_t,a_t) \mid X_0=i]}{n+1}=$$

$\phi_\pi(i)$,

En donde la última igualdad se sigue de la definición (3.1).

Por lo tanto,

$\phi_\pi(i)\geq g$,

Y la igualdad se cumple para $\pi=\pi^*$; es decir,

$g=\phi_{\pi^*}(i)= \min_\pi \phi_\pi(i)$ para todo i.

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

De lo anterior se tiene que si se cumplen las condiciones del teorema 3.1, entonces existe una política estacionaria que es caracterizada por la ecuación (3.2).

Por otra parte, en el criterio de los costos descontados se tiene un factor de descuento α, mientras que los costos en el criterio de costo promedio tienen todos el mismo peso.

Parece razonable pensar que bajo ciertas condiciones, el caso del costo promedio pudiera ser en algún sentido el límite del costo descontado.

Considérese la función costo $V_\alpha(i)$ caracterizada por:

$$V_\alpha(i)=\min_a\{C(i,a)+ \alpha \sum_{j=0}^{\infty} P_{ij}(a)\ V_\alpha(j)\},\quad i\geq 0 \qquad (3.3)$$

La política óptima minimiza el lado derecho de (3.3). Una posibilidad para obtener la política óptima para el costo promedio, sería aquella que escogiera la acción que minimiza la siguiente expresión:

$$\lim_{\alpha \to 1} [C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)\, V_{\alpha}(j)] .$$

Sin embargo, este límite no necesariamente existe, y si existe, puede ser infinito; por lo que esta interpretación no es conveniente.

Otra interpretación sería la siguiente:

Fíjese algún estado; digamos 0, y defínase la siguiente función

$$h_{\alpha}(i) = V_{\alpha}(i) - V_{\alpha}(0),$$

como el α-costo del estado i relativo al estado 0. De (3.3) se obtiene

$$h_{\alpha}(i) + V_{\alpha}(0) = \min_{a} \{C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)[h_{\alpha}(j) + V_{\alpha}(0)]\}$$

$$= \min_{a} \{C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) h_{\alpha}(j) +$$

$$\alpha \sum_{j=0}^{\infty} P_{ij}(a)\, V_{\alpha}(0)\}$$

$$=\min_{a}\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)h_{\alpha}(j)\}+\alpha V_{\alpha}(0).$$

Entonces

$$(1-\alpha)V_{\alpha}(0)+h_{\alpha}(i)=\min_{a}\{C(i,a)+\alpha\sum_{j=0}^{\infty}P_{ij}(a)h_{\alpha}(j)\} \qquad (3.4)$$

Además, la política que escoge la acción que minimiza el lado derecho de (3.4) es una política óptima.

Si se supone que existe una sucesión $\alpha_n \longrightarrow 1$ tal que $h_{\alpha_n}(j)$ converge a la función $h(j)$ y $(1-\alpha_n)V_{\alpha_n}(0)$ converge a la constante g, entonces de (3.4) obtenemos

$$g+h(i)=\min_{a}\{C(i,a)+\sum_{j=0}^{\infty}P_{ij}(a)\ h(j)\},$$

que es precisamente la ecuación (3.2).

En el siguiente teorema se formaliza el resultado anterior.

Teorema 3.2.

Si existe una constante $N<\infty$ tal que

$|V_{\alpha}(i)-V_{\alpha}(0)|<N$ para toda $\alpha\in(0,1)$, y toda $i\in I$,

Entonces:

i) existe una función acotada $h(i)$ y una constante g que satisfacen (3.2);

ii) para cualquier sucesión $\alpha_n \longrightarrow 1$,

$$h(i) = \lim_{n\to\infty} [V_{\alpha_n}(i) - V_{\alpha_n}(0)];$$

iii) $\lim_{\alpha\to 1^-} (1-\alpha)V_\alpha(0) = g$

<u>Demostración:</u>

Por suposición $h_\alpha(i) = V_\alpha(i) - V_\alpha(0)$ es uniformemente acotado en α e i. Entonces, puesto que el espacio de estados es contable, se puede, por el método de diagonalización de Cauchy, obtener una sucesión $\alpha_n \longrightarrow 1$ tal que el límite

$$\lim_{n\to\infty} h_{\alpha_n}(i) \cong h(i),$$

existe para todo i. También, puesto que los costos son acotados, se sigue que $(1-\alpha_n)V_{\alpha_n}(0)$ es una sucesión acotada; así, existe una subsucesión $\{\alpha_{n'}\} \subset \{\alpha_n\}$ tal que

$$\lim_{n\to\infty} (1-\alpha_{n'})V_{\alpha_{n'}}(0) \cong g,$$

existe. Por otra parte, de (3.4) se tiene

$$(1-\alpha_{n'})V_{\alpha_{n'}}(0)+h_{\alpha_{n'}}(i)=\min_{a}\{C(i,a)+\alpha_{n'}\sum_{j=0}^{\infty}P_{ij}(a)h_{\alpha_{n'}}(j)\}$$

Si $n \longrightarrow \infty$, y puesto que $[h_{\alpha_{n'}}]$ es acotado implica, por teorema de convergencia de Lebesgue que

$$\sum_{j=0}^{\infty} P_{ij}(a)\ h_{\alpha_{n'}}(j) \longrightarrow \sum_{j=0}^{\infty} P_{ij}(a)\ h(j)\},$$

luego se sigue que

$$g+h(i)=\min_{a}\{C(i,a)+\sum_{j=0}^{\infty} P_{ij}(a)\ h(j)\},$$

con lo cual se demuestra las partes i) y ii).

Para la parte iii)

Puesto que $[(1-\alpha)V_{\alpha}(0)]$ $0<\alpha<1$ es acotado. Se sigue que para cualquier sucesión $\alpha_n \longrightarrow 1$ existe una subsucesión $\alpha_{n'}$ de $\{\alpha_n\}$ tal que

$$\lim_{n\to\infty}(1-\alpha_{n'})V_{\alpha_{n'}}(0),$$

existe. Por la prueba de $i)$ tenemos que este límite será g; es decir,

$$g=\lim_{\alpha\to 1^-}(1-\alpha)V_{\alpha}(0).$$

$\mathcal{L}.\mathcal{Q}.\mathcal{Q}.\mathcal{D}.$

3.3 Aplicaciones.

Modelo para el reemplazo de una máquina.

Considérese una máquina que puede estar en cualquiera de los siguientes n-estados, $I=\{1,2,...,n\}$. Se supone que el estado i es mejor que el estado $i+1$, $i=1,2,...,n-1$. El estado 1 corresponde a la máquina en condiciones perfectas.

El costo de operación por unidad de tiempo si la máquina se encuentra en el estado i se denota por $C(i)$. Se asume que

$$0\le C(1)\le C(2)\le C(3)\le ...\le C(n) \qquad (3.5)$$

y que las probabilidades de transición están dadas por

$$P_{ij}=0 \qquad \text{si } j<i \qquad (3.6)$$

$$P_{ii}<1 \qquad i=1,2,...,n \qquad (3.7)$$

es decir, que el proceso no puede pasar a un estado mejor que el actual. También se asume que

$$i\le i' \Rightarrow \sum_{j=k}^{n} P_{ij} \le \sum_{j=k}^{n} P_{i'j} \quad \text{para toda } k=1,2,..,n \qquad (3.8)$$

Hay dos acciones posibles: En el comienzo de cada etapa se observa el estado de la máquina y se toma la decisión de reemplazar la máquina por una nueva (que trae como consecuencia regresar al estado 1) y se incurre en un costo $R>0$; o bién, se decide continuar en operación con un costo $C(i)$.

El problema es encontrar una política que minimize el costo promedio.

Considérese la política que nunca reemplaza. Entonces por las suposiciones (3.6) y (3.7), para esta política el único estado que puede ser alcanzado desde cualquier otro es el n; es decir, si nunca reemplazamos nuestra máquina llegará en algún momento hasta el peor estado n, sin importar el estado actual.

Ahora, si se considerá la política que reemplaza en todo estado, entonces $P_{in}=0$; es decir, que la máquina nunca estará en el peor estado n.

Enfoquese el problema hacia el correspondiente problema descontado con un factor de descuento $\alpha<1$. Se tiene

$$V_\alpha(i)=min\{R+C(1)+\alpha\sum_{j=1}^{n}P_{1j}V_\alpha(j);C(i)+\alpha\sum_{j=1}^{n}P_{ij}V_\alpha(j)\},$$

$$i=2,\ldots,n$$

$$V_\alpha(1)=min\{C(1)+\alpha\sum_{j=1}^{n}P_{1j}V_\alpha(j);C(1)+\alpha\sum_{j=1}^{n}P_{1j}V_\alpha(j)\},$$

Por lo que

$$V_{\alpha}(i)-V_{\alpha}(1)=min\{R,[C(i)-C(1)]+\alpha\sum_{j=1}^{n}[P_{ij}-P_{1j}]V_{\alpha}(j)\}\leq R$$

Tomando en cuenta las suposiciones (3.5) a la (3.8) se tiene

$$V_{\alpha}(i)-V_{\alpha}(1)\geq 0 \quad , \quad i=1,2,\ldots,n$$

En efecto para $i=1$ la igualdad se cumple trivialmente y para $i>1$, $V_{\alpha}(i) > V_{\alpha}(1)$ por (3.5) y (3.8).

Por lo tanto, $h_{\alpha}(i)=V_{\alpha}(i)-V_{\alpha}(1)$ es una función acotada y por el teorema 3.2, existe una función acotada $h(i)$ y una constante g, tal que

$$g+h(i)=min\{R+C(1)+\sum_{j=1}^{n}P_{ij}h\ (j);C(i)+\sum_{j=1}^{n}P_{ij}h\ (j)\},$$

$$i=1,2,\ldots,n$$

Además, la política que escoge la acción que minimiza la expresión anterior es óptima en promedio. Más concretamente, si

$$i^{*}=max\{i:\ C(i)+\sum_{j=1}^{n}P_{ij}\ h(j)\ \leq\ R+C(1)+\sum_{j=1}^{n}P_{1j}\ h(j)\},$$

entonces, la política que reemplaza si el estado actual i es mayor que i^{*} y no reemplaza en otro caso, es óptima.

CONCLUSIONES

Este trabajo presenta una herramienta muy poderosa para formular modelos y encontrar políticas óptimas para controlar procesos Markovianos de decisión. Los resultados obtenidos son aplicables a diversas áreas, tales como teoría de colas, inventarios, mantenimiento y programación dinámica probabilística en general.

En el capítulo I se estudió el criterio de los costos descontados con un factor de descuento $\alpha \in (0,1)$. El objetivo es minimizar el costo total descontado mediante decisiones que son tomadas teniendo en cuenta únicamente el estado presente del proceso, sin importar el tiempo t en que se encuentre. A este conjunto de decisiones les llamamos políticas estacionarias óptimas. También puede considerarse a estas políticas como funciones del estado presente X_t; es decir, que toman elementos del conjunto de estados y los relacionan con elementos del conjunto de decisiones. Una característica importante es la retroalimentación de la información; esto es, una vez tomada una decisión a, existe una probabilidad P_{ij} de que el proceso pasa al estado j, el cual será el nuevo estado del proceso y en el cual se volverá a tomar la decisión más conveniente.

Una política estacionaria óptima existe, y se encuentra determinada por la ecuación (1.2) [ecuación de programación dinámica]. En base a la definición (1.11) y al teorema 1.4 (caso especial del teorema de punto fijo de Banach) se tiene que la función de costo óptimo V es única y se obtiene mediante el método de aproximaciones sucesivas.

En el capítulo II, los costos son positivos y no se requiere que sean acotados, por lo que se puede tener costos infinitos, en cuyo caso toda política será óptima (caso trivial). Por lo anterior únicamente son de interés aquellos problemas que generen un costo finito.

Debido a la estructura (2.2) [ecuación de programación dinámica], no existe una función de contracción que garantice que exista una solución única.

En el capítulo III, se hicieron consideraciones para determinar una estructura en la cual se pudiera representar a la función de costo mínimo cuando no es finita.

La ecuación (3.2) proporciona condiciones para que una política estacionaria sea óptima ; es decir, que genera el costo promedio mínimo. En este caso el límite en (3.1) converge a una constante g.

Así, tenemos que las estructuras definidas para los tres criterios, tienen los elementos necesarios para aplicar programación dinámica probabilística (o estocástica).

En el presente trabajo se presentarón procesos de decisión en el cual las decisiones son hechas en períodos de tiempo iguales. Es decir, que los modelos estan basados en Cadenas de Markov en Tiempo Discreto. Sin embargo, si el tiempo de estancia del proceso en un estado es aleatorio, se harán decisiones con un período diferente de tiempo; en este caso los modelos estarán basados en Cadenas de Markov de Tiempo Continuo.

En el mundo "real" tenemos que la mayoría de los sistemas tienen un comportamiento aleatorio; es decir, no sabemos con certeza su comportamiento futuro, por lo que para poder entenderlos y tener un control de ellos es necesario realizar un análisis donde se tome en cuenta el factor incertidumbre.

El análisis que se presentó en este trabajo utiliza precisamente el factor de incertidumbre, a través de modelos matemáticos dinámicos; dando origen a los procesos Markovianos de decisión cuya finalidad es controlar procesos de decisión bajo incertidumbre.

La utilidad que considero pudiese tener este desarrollo de tesis, es que: una vez teniendo los elementos teóricos de los procesos Markovianos de decisión, una asesoria especializada , y una experiencia laboral se apliquen en algún sistema "real".

En lo particular, confirmé que en Matemáticas Aplicadas las bases teóricas son indispensables; de tal forma que no es posible aplicar Matemáticas si no se tiene el fundamento teórico (matemático).

Por otra parte, tuve la oportunidad de conocer gente dedicada a la investigación.

BIBLIOGRAFIA

1) **Ross, S.M. (1970).**

Applied Probability Models with Optimization Aplications. Holden-Day, San Francisco.

2) **Ross, S.M. (1983).**

Introduction to Stochastic Dynamic Programming Academic Press, New York.

3) **Bertsekas, D.P. (1987).**

Dynamic Programing: Deterministic and Stochastic Models. Prentice-Hall, Englewood Cliffs,N.J.

4) **Bertsekas, D.P. and Shreve, S.E (1987).**

Stochastic Optimal Control: The Discrete Time Case.

5) **Heyman D.P. and Sobel, M.J. (1984).**

Stochastic Models in Operations Research, Vol II: Stochastic Optimization Mc Graw-Hill, New-York.

Por último deseo agradecer la paciencia y colaboración del **Dr. Onésimo Hernández Lerma** (Investigador del Centro de Investigación y de Estudios Avanzados del I.P.N.) en la elaboración de esta tesis. También, a las personas que de alguna forma participarón en la elaboración de este trabajo.