DIVISION DE ESTUDIOS DE POSGRADO
FACULTAD DE INGENIERIA

CADENAS DE DESICION MARKOVIANAS CON COSTOS ESPECIALES

ZHANG JIE

TESIS

Presentada a la División de Estudios de
Posgrado de la

FACULTAD DE INGENIERIA

de la

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

como requisito para obtener
el grado de

MAESTRO EN INGENIERIA

(Investigación de Operaciones)

CIUDAD UNIVERSITARIA Febrero, 1986

UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

CONTENIDO

INTRODUCCION.

Uno de los sistemas dinámicos que cambia probabilísticamente en el tiempo es representado por los denominados cadenas de Markov. La importancia de dicho sistema radica en las propiedades analíticas y resultados asintóticos que presenta así como en la simplicidad de su interpretación probabilística y la aplicación que tiene a diferentes ramas de la ciencia.

En los sistemas de Cadenas de Markov, el estado en un instante o período n, se denota x(n), y representa la información relevante del sistema. Por ejemplo, en una presa, el estado del sistema podría ser el nivel de agua al inicio de un período de producción agrícola o bien, en un sistema financiero, el estado es la cantidad de dinero en efectivo que se dispone más el valor de los bienes al principio del periodo lectivo.

En sistemas dinámicos que cambian probabilísticamente con el tiempo se desea determinar el comportamiento de la sucesión de estados {x(n)} cuando se aplica una política de operación sobre el desarrollo del sistema. Idealmente, lo que se desea es asociarlo con cada decisión efectuada en el sistema, en el periodo n, un costo o beneficio determinar aquella política que minimice costos o maximice beneficios.

Una clase importante de sistemas dinámicos denominados de decisión markovianos, por las reglas de cambio probabilístico, han sido estudiados en la literatura desde la década de los sesenta por Blackwell y Howard. El primer autor formuló matemáticamente el problema mientras que el segundo propuso un método de solución que resulta sencillo y elegante. El problema resuelto se denomina programación markoviana con descuento. Posteriormente, se ilustró la relación de los métodos propuestos por Howard y el concepto de mapeo de contracción, con lo que métodos alternativos, computacionalmente fueron propuestos y aplicados a una extensa variedad de problemas reales.

Una de las dificultades del proceso de decisión markoviana con descuento es la justificación del factor de descuento. Es por ello que se propusieron nuevos modelos que rescataran las propiedades y métodos de solución del primero, sin tener la restricción de dicho factor. Esto dió lugar a procesos de decisión markoviana con costos especiales, esto es, costos negativos o bien costos positivos. En ese formato se formulan una clase importante de problemas que habían sido estudiadas de manera aislada en la literatura como problemas de paro óptimo, teoría de apuesta, reemplazo de equipo y otros.

En este trabajo se formulan y analizan los procesos de decisión markoviana con costos especiales. El propósito del trabajo es la justificación del análisis de tales procesos bajo un mismo método de análisis, mapeos y de contracción y programación dinámica. En particular, la existencia y caracterización de solución se efectúa mediante el nuevo enfoque de latices. Los métodos disponibles de busqueda de la solución, tales como método de aproximación sucesiva, mejoramiento de políticas y programación lineal se discuten y analizan bajo los marcos teóricos de la existencia y caracterización establecidos. Para cada caso de costos especiales se indican las restricciones de aplicación para los métodos considerados. Ejemplos ilustrativos se anexan.

Este trabajo se desarrolla como sigue: el primer capítulo describe las bases metodológicas de análisis de una cadena de Markov con especial énfasis en las propiedades asintóticas de la matriz de transición. El capítulo dos describe el clásico problema de cadenas de decisión markoviana con descuento y caracteriza las políticas óptimas así como los correspondientes métodos y ejemplos ilustrativos. El capítulo tres, formula las cadenas de decisión markovianas con costos negativos y caracteriza las políticas óptimas. Asimismo, se describen los métodos de solución clásicos: mejoramiento de políticas de Howard, aproximaciones sucesivas y programación lineal, especificando las ventajas y desventajas de tales métodos. La aplicación a problemas clásicos de reemplazo de equipo se desarrolla. En el capítulo cuarto se analizan el caso de costos positivos y una vez caracterizadas las políticas óptimas y métodos de solución se aplica al problema de teoría de apuestas. Las conclusiones de este trabajo se tienen en el capítulo cinco. Asimismo se anexan dos apéndices técnicos, uno sobre matrices y sus propiedades espectrales aplicadas a matrices no-negativas y otro sobre las bases y resultados fundamentales de latices usadas en este trabajo.

CAPITULO I

CADENAS DE MARKOV

Uno de los procesos estocásticos más estudiados en la literatura es el proceso de cadenas de Markov. La importancia del proceso radica en las propiedades analíticas y resultados asintóticos que presentan así como la simplicidad de su interpretación probabilística y la aplicación que tiene a diferentes ramas de la ciencia. Las cadenas de Markov que nos ocupan son las relacionadas con números finitos de estados y parámetro discreto aunque las propiedades y resultados que analizaremos se extienden al caso de números contables de estados y parámetro contínuo. La idea del análisis es establecer el marco teórico básico para las cadenas de decisión markoviana que presentamos en los siguientes capítulos.

Este capítulo se desarrolla como sigue: La primera sección define el concepto de cadena de Markov mientras que en la segunda se presentan las características de la distribución límite de una matriz regular. En la tercera sección se describe la clasificación de estados de una cadena, y en la cuarta se efectúa un análisis de los estados transitorios. El concepto de periodicidad se analiza en la quinta sección y, finalmente se presentan ejemplos ilustrativos.

1.1 Cadenas de Markov simples.

Uno de los sistemas dinámicos que cambian probabilísticamente en el tiempo es el representado por las cadenas de Markov. Dicha clase de sistemas dinámicos tiene aplicaciones a diversas ramas de la ingeniería, biología, finanzas, etc. Entre los conceptos básicos de un sistema dinámico se distingue el concepto de estado del sistema, esto es, la información relevante al sistema en un instante de tiempo dado. Por ejemplo, en una presa, el estado del sistema podría ser el nivel de agua al inicio de un período de producción agrícola o bien, en un sistema financiero, el estado es la cantidad de dinero en efectivo que se dispone más el valor de los bienes al principio del año lectivo.

El estado de un sistema en el instante o período n se denota como $x(n)$ o bien x_n y lo que se desea es conocer el comportamiento de los valores $\{x(n)\}$ cuando se aplica una política de operación sobre el desarrollo del sistema. Idealmente, desearíamos determinar una política que maximizara beneficios o minimizara costos en un periodo de planeación especificado. Conviene señalar que los valores que puede adquirir $x(n)$ son, en general finitos. Una forma esquemática de representar los valores $x(n)$ se muestra en la figura 1.

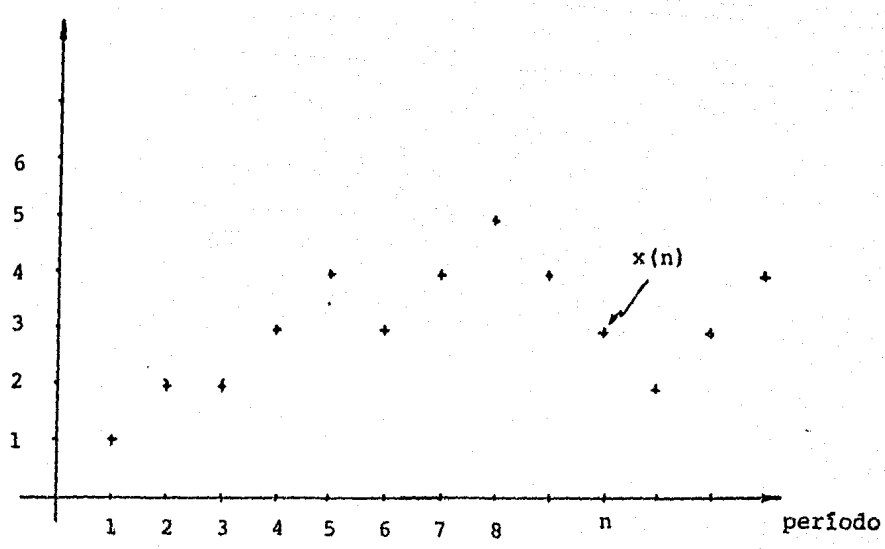


Fig. 1. Desarrollo del sistema.

Considere un proceso estocástico $X = \{x_n \mid n \in N\}$ donde los valores que puede tomar una variable x_n forman un conjunto finito E.

Definición. Un proceso estocástico $X = \{x_n \mid n \in N\}$ es una cadena de Markov si cumple que

$$P\{x_{n+1} = j \mid x_0, \ldots, x_n\} = P\{x_{n+1} = j \mid x_n\}$$

para todo $j \in E$ y toda etapa n.

En otras palabras, una cadena de Markov es una sucesión de variables aleatorias tal que en la etapa n, el evento x_{n+1} depende únicamente de x_n (el presente) y no de la historia pasada $x_0, x_1, \ldots, x_{n-1}$. Otro aspecto importante es que

$$P\,[x_{n+1} = j \mid x_n = i] = P_{ij} \qquad i,j \in E$$

esto es, la probabilidad condicional es independiente de la etapa n. Este tipo de cadenas de Markov se dice que es homogénea, pues no depende del tiempo. Asimismo si $E=\{1,2,\ldots,n\}$, la matriz de probabilidades de transición es

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ - & - & - & - \\ - & - & - & - \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

Definición. Una matriz $P = [P_{ij}]$ de orden nxn se denomina matriz de transición si sus elementos son no-negativos y la suma de los elementos de cada hilera es igual a uno, esto es

a. $P_{ij} \geq 0 \qquad i,j = 1,\ldots,n$

b. $\sum_{j=1}^{n} P_{ij} = 1 \qquad i = 1,\ldots,n.$

Es común decir que la matriz de transición P tiene n estados y que cada elemento P_{ij} representa la probabilidad de pasar del estado i al j en una transición. Por otra parte, es conveniente puntualizar que si P es una matriz de transición, P^2 es también una matriz de transición y cada uno de sus elementos (i,j) representa la probabilidad de pasar de un estado

i a uno j con dos etapas. Una interpretación semejante se tiene para la matriz P^m donde m es un entero positivo.

1.2 Cadenas de Markov regulares y distribuciones límite.

Un aspecto básico de una matriz de transición P es caracterizar el comportamiento de sus potencias P^m cuando m tiende a infinito.

Definición. Una cadena de Markov se dice regular si existe un entero positivo m tal que P^m es estrictamente positiva.

La regularidad de una cadena de Markov significa que después de un número suficientemente grande de transiciones o pasos, la cadena es estrictamente positiva, esto es, existe una probabilidad positiva de ir de un estado i a otro cualesquiera.

El principal resultado asociado con el concepto de cadenas de Markov regular se tiene en el teorema siguiente, cuya interpretación es como sigue. El límite de P^m existe y la probabilidad de ir de un estado a otro es independiente del estado inicial.

Teorema. (Propiedades límites de una cadena de Markov). Sea P matriz de transición de una cadena de Markov regular. Entonces

a. Existe vector $\Pi_0 > 0$ único tal que $\Pi_0 P = \Pi_0$

b. Dado un estado i cualesquiera y e_i vector hilera con elemento igual a uno en la posición i y cero en el resto. Entonces

$$\Pi_0 = \lim e_i P^m$$

c. $\underline{P} = \lim P^m$ donde $\underline{P}$ es una matriz con vectores hilera igual a Π_0.

Prueba. El resultado a es inmediato del Teorema de Frobenius-Perron (Teorema 4. Apéndice A). En tal caso $\lambda_0 = 1$ pues la suma de cada hilera de P es igual a uno (Proposición 1. Apéndice A). Por otra parte, cualquier otro vector característico de P, digamos λ, es tal que $|\lambda| < 1$. Aplicando el teorema de Jordan se tiene que podemos escribir $P = TJT^{-1}$ donde T es matriz no-singular y J es la matriz Jordan, que sin perdida de generalidad, puede expresarse como

$$J = \begin{bmatrix} 1 & & & \\ & J_1 & & \\ & & J_2 & \\ & & & \ddots \\ & & & & J_p \end{bmatrix}$$

donde J_i es la forma Jordan con $|\lambda_i| < 1$ $i=1,\ldots,P$.

Es inmediato que $P^m = TJ^mT^{-1}$ y que (Lema 1, Apéndice A)

$$\lim P^m = T \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} T^{-1} = \underline{P}$$

De donde $e_i\underline{P} = \lim e_iP^m = (\lim e_i P^{m-1})\, P = (e_i\underline{P})P$ y se tiene que $\Pi_0 = e_i\underline{P}$ pues $e_i\underline{P}$ es un vector de probabilidades. Dado que i es arbitrario se concluye que todas las hileras de $\underline{P}$ son iguales a Π_0 y la prueba termina.

Una pregunta inmediata asociada con este teorema es la verificación a priori de la regularidad de una cadena de Markov, pues bajo tal suposición es posible garantizar que se tiene una distribución límite Π_0 que inclusive puede calcularse - resolviendo el sistema lineal homogéneo $\Pi_0 P = \Pi_0$ cuya solución es no-trivial.

1.3 Clasificación de Estados.

En el desarrollo de una teoría general de cadenas markovianas un primer paso es sistemáticamente estudiar la estructura de intercorrelaciones de los estados. En esta sección, se presenta la clasificación de estados lo cual ayuda en el análisis de cadenas markovianas.

Se dice que el estado j es accesible del estado i si existe algún $m > 0$ tal que la probabilidad de transición $p_{ij}^{(m)} > 0$. Nótese que la propiedad de accesibilidad no es simétrica. Si los dos estados i y j son accesibles mutuamente, se dice que i y j son comunicantes. Nótese que la propiedad de comunicación es una relación de equivalencia. Esto es, se cumple:

a). $i \leftrightarrow i$

b). si $i \leftrightarrow j$, $j \leftrightarrow i$

c). si $i \leftrightarrow j$ y $j \leftrightarrow k$, $k \leftrightarrow i$.

los postulados a y b son inmediatos de la definición de comunicación entre estados. El postulado c se puede comprobar por la ecuación de Chapman-Kolmogorov pues:

$$p_{ik}^{m+n} = \sum_{r=0}^{N} p_{ir}^{m} \, p_{rk}^{n} \geq p_{ij}^{m} \, p_{jk}^{n} > 0$$

El concepto de estados comunicantes divide los estados de una cadena markoviana en clases ajenas, como se muestra a continuación.

Proposición 1. El conjunto de estados de una cadena de markov se divide en clases comunicantes ajenas. Cada estado comunica con los estados de la misma clase y no con otros.

Prueba. Sea C_i una clase comunicante. Suponga que $k \notin C_i$ y $k \leftrightarrow i \in C_i$. Por la propiedad de comunicación, k se comunica con todos los estados de C_i y viceversa; esto implica que $k \in Ci$ y es una contradicción. Por lo tanto, las diferentes clases no deben contener estados en común; cualquier estado de la cadena pertenece a una y solo una clase.

Las clases comunicantes pueden clasificarse en varias clases específicas las cuales se definen como sigue. Una clase comunicante es cerrada si cualquier estado de esta clase no puede accesar otros estados que no pertenecen a ella. Si un estado es el único elemento de una clase cerrada se llama estado absorbente. Una clase cerrada es irreducible si todos los subconjuntos no propios de esta clase no son cerrados. Una cadena de Markov es irreducible si sólo contiene una clase cerrada.

Una clase comunicante es una clase transitoria si los estados de esta clase pueden accesar algunos estados fuera de ella. Los estados de una clase transitoria se denominan estados transitorios.

Resumiendo la discusión anterior, es interesante observar que la clasificación de los estados indica que una cadena de Markov tiene al menos una clase cerrada. En particular, no importa que el proceso parta inicialmente de una clase transitoria siempre existe una probabilidad positiva de entrar una clase cerrada dentro de un número finito de pasos. En otras palabras, el proceso deja los estados de clase transitoria y permanece en los estados de clase cerrada. Podemos expresar la matriz de transición de una cadena de Markov en una forma canónica, esto es, expresar P como sigue:

$$P = \begin{bmatrix} P_1 & 0 \\ R & Q \end{bmatrix}$$

donde P_1 es una matriz estocástica de orden rxr que representa la matriz de probabilidades dentro de las clases cerradas (se suponen que r estados pertenecen a las clases cerradas); Q es una matriz subestocástica de orden (n-r)x(n-r), esto es, al menos un renglón de Q tiene una suma menor que uno. Esta matriz representa las probabilidades de transición entre los estados transitorios. R es una matriz de orden (n-r)xr que representa las probabilidades entre estados transitorios y estados de la clase cerrada. En este caso, el vector característico correspondiente al valor característico 1 debe tener la forma $P = [p_1, 0]$ donde p_1 es un vector r-dimensional, porque sólo los estados en las cla-

ses cerradas pueden tener probabilidad positiva de equilibrio.

Conviene señalar que si una cadena regular de Markov es irreducible, entonces todos los estados de ella se comunican. Sin embargo, no todas cadenas irreducibles de Markov son regulares. Por ejemplo, considere la matriz perfodica

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

En este caso, el estado 1 y el 2 pueden comunicarse mutuamente, sin embargo, cualquier potencia de P contiene dos elementos igual a cero y no existe entero n tal que P^n sea una matriz estrictamente positiva.

1.4 Análisis de estados transitorios.

Considere la forma canónica de la matriz de transición

$$P = \begin{bmatrix} P_1 & 0 \\ R & Q \end{bmatrix}$$

donde Q es una matriz subestocástica, esto es, alguna hilera de Q suma menor que la unidad y los estados que forman Q son una clase comunicante. Se define la matriz $M = [I - Q]^{-1}$ como la matriz fundamental de la cadena de Markov.

Proposición. $M = [I-Q]^{-1}$ existe y es no-negativa.

Prueba. Observe que $|\sigma(Q)| < 1$ y que $\lim Q^m = 0$. Por lo que la inversa de I - Q es dada por

$$[I-Q]^{-1} = \sum_{i=0}^{\infty} Q^i \geq 0$$

y la prueba termina.

Se observa que los elementos de la matriz fundamental M representan el número promedio de visitas de varios estados transitorios. Formalmente:

Teorema 1. El elemento m_{ij} de M de una cadena de Markov con estados transitorios es igual al número promedio de veces que el proceso visita j dado que partió de i.

Teorema 2. La suma del renglón i de M es igual al número promedio de pasos de la cadena antes de entrar una clase cerrada cuando el proceso parte del estado transitorio i.

Si una cadena inicia de un estado transitorio, va a llegar (con probabilidad 1) a algún estado dentro de una clase cerrada. A veces nos interesa saber la probabilidad de que dado un estado transitorio inicial, la cadena entre por primera vez a una clase cerrada por un estado particular de esta clase. El siguiente teorema nos dice cómo calcular estas probabilidades.

Teorema 3. Sea b_{ij} la probabilidad de que, una cadena de Markov que parte del estado i entre por primera vez a una clase cerrada por el estado j. Sea B la matriz (n-r)xr con b_{ij}, entonces B = MR.

Prueba. Suponga que i pertenece a una clase transitoria y j está en una clase cerrada. entonces

$$b_{ij} = P_{ij} + \sum_{\forall k} P_{ik} b_{kj}$$

donde k es un estado transitorio. Entonces, matricialmente:

$$B = R + QB$$

cuya solución es $B = [I - Q]^{-1} R = MR$.

1.5 Periodicidad.

Definición: Se dice que un estado j de una cadena de Markov tiene periodo d(j) si d(j) es el máximo común divisor del conjunto

$$Q_j = \{n \mid p_{ij}^n > 0\}$$

El estado que tiene d=1 es referido como estado aperiódico.

Teorema 4. Suponga que $i \leftrightarrow j$ entonces $d(i) = d(j)$.

Prueba: Note que d(i) es el máximo común divisor de

$$Q_i = \{n \mid p_{ii}^n > 0\}$$

y existen m_1 y m_2 tales que

$$p_{ji}^{m1} > 0 \quad \text{por } j \to i \; ; \; p_{ij}^{m2} > 0 \quad \text{por } i \to j$$

entonces para algún $n \in Q_i$ se tiene $p_{ii}^n > 0$ y

$$p_{jj}^{m1+n+m2} \geq p_{ji}^{m_1} p_{ii}^n p_{ij}^{m2} > 0$$

de donde $m1 + n + m2 \in Q_j$. Si $p_{ii}^n > 0$, implica $p_{ii}^{2n} > 0$. Por la misma razón $p_{jj}^{m1+2n+m2} > 0$ es decir, $m1+2n+m2 \in Qj$. Observe que d(j) divide ambos m1+n+m2 y m1+2n+m2 de donde d(j) divide a n; es decir, d(j) divide a d(i). Por el mismo argumento tenemos que d(i) divide a d(j) y concluimos que $d(i) = d(j)$. Equivalentemente todos los estados de una clase tienen el mismo número de períodos.

Teorema: Si P es una matriz de transición irreducible entonces P estará en uno y solo uno de los siguientes dos casos:

a). P es aperíodica y existe un M tal que $P^m > 0$ para toda $m \geq M$.

b). P es periódica y puede expresarse como

$$P = \begin{bmatrix} 0 & P_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & P_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & P_{n-1} \\ P_n & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

con período $n \geq 2$.

En algunas aplicaciones, es natural que formulen modelos de cadena de Markov de un número infinito (contable) de estados. Frecuentemente una cadena de Markov tiene gran simetría en este caso y por ello nos conduce a simplificar la formulación. En realidad, muchos de los conceptos y la esencia de la teoría de cadena finita son extendibles para el caso de cadena infinita, como los conceptos de accesibilidad, clases de comunicación e irreducibilidad que son útiles para el caso infinito sin hacer ningún cambio.

1.6 Ejemplos ilustrativos.

A continuación se presentarán unos ejemplos de aplicaciones de la cadena de Markov.

Ejemplo 1. El tiempo en una cierta ciudad puede ser caracterizado como soleado, nublado o lluvioso. Si el día es soleado, existe la misma probabilidad de que el día siguiente sea soleado o nublado. Si el día es nublado, entonces hay un 50% de probabilidad de que el siguiente día sea soleado, 25% de probabilidad que continúa nublado y 25% de probabilidad que cambie a lluvioso. Si el día es lluvioso, continuará lluvioso o nublado con una probabilidad igual para cada estado.

Suponga que los estados del tiempo, soleado, nublado y lluvioso se denotan por S, N y LL respectivamente. Un modelo de cadenas de Markov que represente las transiciones probabilísticas de un estado a otro es

$$P = \begin{array}{c} \\ S \\ N \\ LL \end{array} \begin{array}{c} \begin{array}{ccc} S & N & LL \end{array} \\ \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix} \end{array}$$

donde los valores de las hileras correspondiente a cada estado representan la probabilidad de pasar de ese estado a otro. Una forma esquemática de los cambios de un estado a otro se muestra en la figura 1.

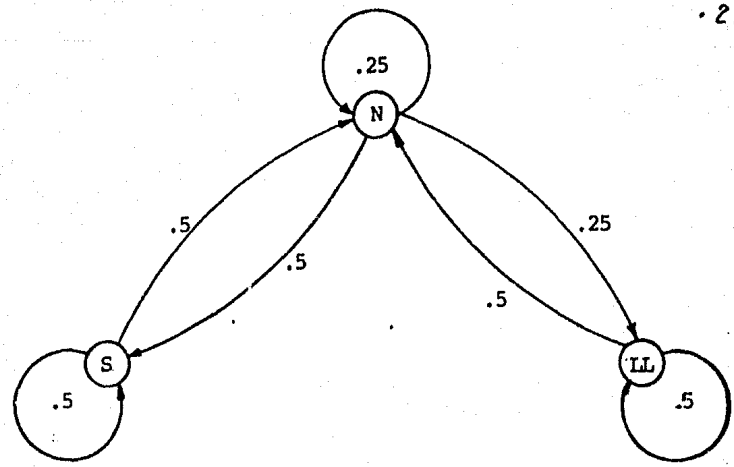


Fig. 1. Diagrama de transiciones.

Las potencias sucesivas de la matriz de transición se muestran a continuación y se observa que el vector de probabilidades estacionarias es $\Pi=[0.4,\ 0.4,\ 0.2]$, esto es, 40% de las veces está soleado, 40% de las veces está nublado mientras que 20% está lluvioso. Conviene observar que las potencias de la matriz de transición convergen rápidamente a la matriz límite cuyos vectores hileras son idénticos al vector de probabilidades estacionarias.

$$P^2 = \begin{bmatrix} .500 & .375 & .125 \\ .375 & .438 & .187 \\ .250 & .375 & .375 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} .422 & .399 & .179 \\ .398 & .403 & .199 \\ .359 & .399 & .242 \end{bmatrix} \qquad P^6 = \begin{bmatrix} .405 & .401 & .194 \\ .400 & .401 & .199 \\ .390 & .400 & .210 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} .401 & .401 & .198 \\ .400 & .401 & .199 \\ .397 & .401 & .202 \end{bmatrix} \qquad P^{16} = \begin{bmatrix} .400 & .400 & .200 \\ .400 & .400 & .200 \\ .400 & .400 & .200 \end{bmatrix}$$

Ejemplo 2. (Juego de monopolio). Un juego clásico de riesgo e incertidumbre para varios jugadores se muestra en la figura 2, donde cada jugador posee una ficha que generalmente mueve en el sentido de las manecillas del reloj de una casilla a otra. En este juego, el jugador en turno lanza una moneda al aire; si el resultado es "cara" se desplaza una casilla y si es "águila" se desplaza dos. Un jugador que cae en la casilla "Ir a la carcel", debera ir a dicha casilla y esperar su próximo turno. Durante el desarrollo del juego, los jugadores pueden apropiarse de las casillas, excepto las correspondientes a "Ir a la carcel" y "carcel". Si un jugador cae en un casillero propiedad del otro, éste deberá pagar una renta al propietario, de acuerdo a la cantidad marcada en la casilla. Con el propósito de formular la estrategia de juego resulta útil conocer qué casilleros son más valiosos en términos de la renta que se espera generar a lo largo del juego. Sin análisis previo, es díficil conocer los valores relativos asociados a cada casilla.

El movimiento de las fichas a lo largo del tablero puede ser considerado como una cadena de Markov con siete estados, correspondientes a las siete casillas. Observe que el casillero "Ir a la carcel" es una forma equivalente de decir vete a la casilla cuatro.

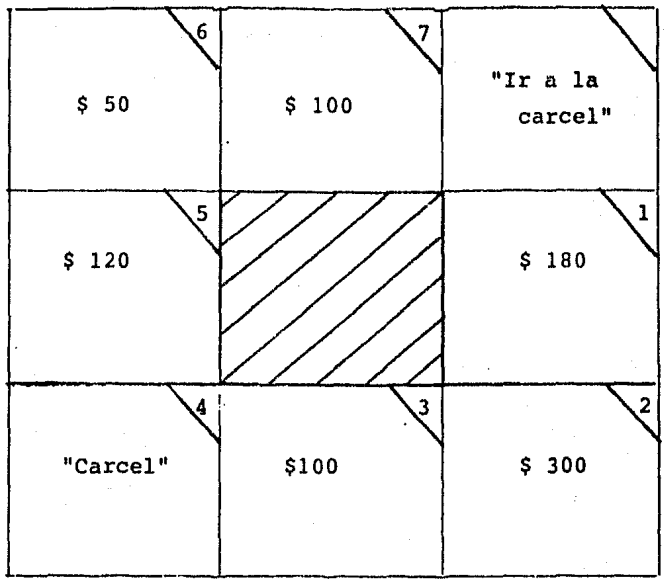


Fig. 3. Tablero de juego.

La matriz de transición asociada al problema es:

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \end{bmatrix}$$

y es sencillo verificar que existe una potencia N tal que la matriz P^N es estrictamente positiva. Por lo tanto, podemos calcular el vector de probabilidades estacionarias de la cadena por medio del límite de las potencias de P. En realidad la matriz P^{64} es suficiente para obtener

$$\Pi = [.0909 \quad .0455 \quad .0682 \quad .2500 \quad .1591 \quad .2045 \quad .1818]$$

cuya interpretación es como sigue: 0.09% de las veces se visita el estado 2; y así sucesivamente hasta tener que 18.18% de las veces se visita el estado 7. No es sorprendente observar que la "carcel" es la casilla más visitada y que las casillas 1, 2 y 3 se visitan con menos frecuencia. Es por ello que aunque tales casillas son las de mayor renta, no resultan ser las más atractivas debido a la frecuencia con que se visitan, como puede verificarse de la tabla siguiente, en donde los ingresos relativos están normalizados a que el estado 7 reciba $100.

Casilla	Renta	Ingreso relativo	Jerarquía
1	180	90.00	3
2	300	75.00	4
3	100	35.00	6
4	—	—	-
5	120	105.00	1
6	50	56.25	5
7	100	100.00	2

Tabla 1. Jerarquía de las casillas.

Ejemplo 3. Considere dos personas A y B que se involucran en una serie de juegos donde A gana con probabilidad p y B con probabilidad q = 1 - p donde $0 \leq p \leq 1$. Se supone que los resultados de juegos sucesivos son independientes entre sí y que el perdedor de un juego paga una unidad monetaria al ganador. Suponga que en el inicio del juego A posee i unidades monetarias y B unicamente N - i. El juego termina cuando alguno de ellos pierde todo su capital disponible. Se desea conocer la probabilidad de que A gane el juego o equivalentemente, que A logre tener un capital de N unidades monetarias antes de quedarse sin capital.

Este es un problema clásico conocido como "caminata aleatoria", pues cada vez que A pierde o gana es equivalente a decir que se desplaza un paso a la izquierda o derecha. Si logra un capital N o pierde todo su capital, es equivalente a decir que llegó a uno de los dos extremos de un camino de longitud N + 1.

Con el propósito de resolver el problema conviene definir el término u(i) como la probabilidad de que A logre una fortuna N antes de perder todo su capital, dado que el capital inicial es de i unidades monetarias. Observe que con esta definición es posible escribir:

$$u(i) = p\,u(i+1) + q\,u(i-1)$$

donde $1 < i < N-1$ y $u(N) = 1$; $u(0) = 0$, son las condiciones de frontera. Sin embargo, dado que $q = 1 - p$ podemos reformular la ecuación anterior como sigue

$$q[u(i) - u(i-1)] = p\,[u(i+1) - u(i)]$$

Donde, si hacemos $r = q/p$ y $v(i) = u(i+1) - u(i)$ la ecuación puede reducirse a la fórmula recursiva $v(i) = r\,v(i-1)$ para toda $1 < i < N-1$.

Sin embargo, la solución de esta ecuación es dada por $v(i) = r^i v(0)$ para toda $i = 0,1,\ldots,N$. Por otra parte, se observa que si $r \neq 1$,

$$\begin{aligned} 1 &= u(N) - u(0) \\ &= \sum_{i=0}^{N-1} [u(i+1) - u(i)] \\ &= \sum_{i=0}^{N-1} v(i) \\ &= \sum_{i=0}^{N-1} r^i\, v(0) \\ &= \left[\frac{1-r^N}{1-r}\right] \cdot v(0) \end{aligned}$$

y se implica que $v(0) = (1-r)/(1-r^N)$. Un argumento semejante para $u(i) = u(i) - u(0)$ pues $u(0) = 0$ demuestra que

$$u(i) = \left[\frac{1 - r^i}{1 - r}\right] v(0)$$

Sustituyendo el valor de v(0) y recordando que $r = q/p$ podemos concluir que si $r \neq 1$

$$u(i) = \frac{1 - (q/p)^i}{1 - (q/p)^N} \qquad 0 \leq i \leq N .$$

En el caso particular de $r = 1$ se tiene que

$$\begin{aligned} 1 &= u(N) - u(0) \\ &= \sum_{i=0}^{N-1} [u(i+1) - u(i)] \\ &= \sum_{i=0}^{N-1} v(i) \\ &= \sum_{i=0}^{N-1} v(0) \\ &= N\, V(0) \end{aligned}$$

o bien $v(0) = N^{-1}$ y es sencillo verificar que $u(i) = \frac{i}{N}$ para toda $i = 0,1,\ldots,N$. En resumen

$$u(i) = \begin{cases} \dfrac{1 - (q/p)^i}{1 - (q/p)^N} & \text{si } q \neq p \\ \dfrac{i}{N} & \text{si } q = p = \frac{1}{2} \end{cases}$$

CAPITULO II

CADENAS DE DECISION MARKOVIANA

Uno de los modelos más adecuados para representar sistemas dinámicos estocásticos con decisiones secuenciales es dado por una cadena de decisión markoviana. En dichos sistemas, existe un beneficio asociado con cada decisión y la manera estocástica en que se desenvuelve el sistema de una etapa a otra depende de tal decisión. Una clasificación importante de estas cadenas de decisión es de acuerdo al tamaño del periodo de planeación: finito o infinito.

El análisis de las cadenas de decisión markoviana con periodo finito es sencillo y una política óptima se obtiene de la aplicación directa de la programación dinámica. Sin embargo, una cadena con periodo infinito no puede resolverse directamente por esta técnica. Una manera de resolverla es introducir un factor llamado factor de descuento β $(0 \leq \beta \leq 1)$ que permite una interpretación económica y evita el problema de divergencia del beneficio total esperado.

Este capítulo se desarrolla como sigue: primero se establecen los conceptos básicos de cadenas de decisión markoviana y se analiza el caso de periodo finito. Luego se concentra en el análisis del caso con periodo infinito y factor de descuento.

2.1 Descripción del problema.

Considere un sistema dinámico que es observado en cada uno de los instantes de una sucesión de puntos del tiempo, denotados por 1, 2, 3,.... En cada instante el sistema se encuentra en uno de los estados del conjunto $S = \{1,2,3,...s\}$ o bien el sistema "para". Si el sistema es observado en el estado $s \in S$ una acción, denotada a, es ejecutada. Dicha acción a pertenece a un conjunto A_s de posibles acciones; y un beneficio r(s,a) (o un costo c(s,a)) es recibido. Se supone que la probabilidad condicional de que el sistema se observe en el estado s, en el tiempo N+1, dado que se encuentra en el estado i en el tiempo N, que la acción a ha sido ejecutada y que los estados observados y las acciones ejecutadas en los tiempos 1, 2,... N-1, son conocidos, es una función no-negativa p(s|i,a) que depende únicamente de s,i,a,. Equivalentemente

$$p(s \mid i,a;\ i_{N-1},a_{N-1}\ ;\ldots\ i_1\ ;\ a_1) = p(s \mid i,a)$$

la propiedad markoviana se satisface. La probabilidad condicional de que el sistema "pare" en el tiempo N+1 es dado por

$$1 = \sum_s p(s \mid i,a)$$

una vez que el sistema "para" permanece en ese estado y no se obtiene ningún beneficio.

Se observa que los beneficios y las probabilidades de transición son funciones sólo del último estado y la acción subsecuente. Usualmente se le llama tomar decisión al proceso de decidir cúal de las posibles acciones se va a ejecutar en un estado i, el cual es observado en el instante (o período) n.

Para tomar una buena decisión, debe seguirse alguna política o estrategia. Una política es una regla para tomar decisiones. Denotaremos una política por el símbolo π. En realidad una política π es una sucesión de las acciones $\{a_1, a_2, \ldots,\}$ donde a_i representa la acción tomada en el período i y pertenece al conjunto A de posibles acciones.

Un subconjunto importante de todas políticas es el de políticas estacionarias. Una política se dice estacionaria si la decisión que se especifique en el período n sólo depende del estado presente. Denotamos una política estacionaria por $f^\infty = (f, f, \ldots,)$. Algebraicamente, una política estacionaria es una función f que mapea el espacio de estados en el espacio de acciones. Esto puede interpretarse como: para cada estado i, f(i) denota la acción que se toma por la política cuando esté en el estado i. Una vez que la política estacionaria f es empleada, la secuencia de estados $\{x_n, n=0,1,2,\ldots,\}$ forma una cadena de Markov con probabilidades de transición $p_{ij} = p_{ij}\ (f(i))$. Por ello, la cadena

es llamada cadena de decisión markoviana.

Sea π una política cualquiera. El beneficio o costo total esperado dado que se emplea π y esté en el estado i inicialmente, puede representarse por

$$V_\pi(i) = E_\pi \left[\sum_{n=0}^{\infty} r(x_n, a_n) \mid x_0 = i \right]$$

donde $r(x_n, a_n)$ es el beneficio obtenido en el instante n.

Para el sistema dinámico que se ha descrito anteriormente, se desea determinar una política óptima tal que se maximiza el beneficio total esperado o equivalentemente, se minimiza el costo total esperado.

Sea $V_N(s)$ el beneficio máximo esperado de un sistema, dado que se esté en el estado s de un problema conN períodos (se supone que el máximo existe). Suponga que se ejecuta la acción $a \in A_s$ inicialmente y que después, se saque una política "óptima" (esto es una política que está asociada al máximo beneficio esperado). Entonces se cumple que

$$V_N(s) = \max_{a \in A_s} \{ r(s,a) + \Sigma\, p(j \mid s,a)\, V_{N-1}(j) \}$$

que es equivalente al principio de optimalidad de la programación dinámica.

2.2 El problema con horizonte finito.

Considere una cadena de decisión markoviana con período finito, como se ha mencionado anteriormente. Este tipo de cadenas puede resolverse directamente por la técnica de programación dinámica, la cual, esencialmente, es un procedimiento analítico para resolver problemas discretos secuenciales. La idea central de la programación dinamica es el principio de optimalidad propuesto por Richard Bellman en 1957. El principio de optimalidad establece que la política óptima tiene la propiedad de que no importa el estado inicial ni la decisión previamente tomada, las decisiones restantes deben constituir una política óptima con respecto al estado resultante por la primera transición". La esencia de programación dinámica es la optimalidad de los procesos de decisión secuenciales. Esto es, en base del principio de optimalidad se convierte un problema dado, en n subproblemas relacionados entre si, por la ecuación recursiva formulada, las cuales son mucho más simple de resolver que el problema original, y cada uno de ellos forma un paso para resolver el problema original. Además, cada solución de subproblema constituye, junto con las soluciones de periodos anteriores, una política óptima. La solución del último subproblema planteado es la solución óptima del problema original.

En el problema con N períodos, el objetivo es encontrar una política óptima π tal que satisface el principio de optimalidad. Este tipo de problemas se puede resolver por técnicas estándares de programación dinámica en forma como sigue:

La ecuación recursiva está dado por:

$$V_k(i) = \max_a \{r(i,a) + \sum_{j \in S} P_{ij}(a) V_{k-1}(j)\}, \quad k = 0,1,2,\ldots N$$

donde $V_k(i)$ es el beneficio máximo esperado para k períodos dado que inicie en el estado i y la condición de frontera es

$$V_{N+1}(i) = u(i) \qquad \forall\, i \in S.$$

donde generalmente se supone que $u(i) = 0$. Sea $V_\pi(i)$ el beneficio total esperado usando la política π, entonces

$$V_\pi(i) = E_\pi \left[\sum_{t=0}^{N} r(x_t, A_t) \mid x_0 = i \right].$$

La siguiente proposición nos asegura que mediante la técnica estándar de programación dinámica puede lograrse el objetivo del caso con período finito mencionado anteriormente.

Proposición 1. La política de decisión y el valor óptimo asociados al problema markoviano con horizonte finito pueden obtenerse mediante la solución de las ecuaciones.

$$1) \quad V_k(i) = \max_a \{r(i,a) + \sum_j p_{ij}(a)\, V_{k+1}\} \qquad k=1,\ldots,N$$

$$2) \quad V_{N+1} = u(i),$$

donde (1) y (2) se denominan las ecuaciones recursivas y condición de frontera, respectivamente.

Prueba. En un problema de N períodos, el procedimiento de programación dinámica es como sigue: En el N+1-ésimo período, suponga que $V_{N+1}(i)$ es óptimo y que

$$V_{N+1}(i) = u(i) \qquad i \varepsilon S$$

donde u(i) son funciones acotadas; que es equivalente a la condición de frontera de la proposición. En el periodo $k = 1,2,\ldots,N$, las decisiones a realizar, denotados $(a_k, a_{k-1},\ldots,a_N)$ dependen únicamente del conocimiento del vector de estados x_k. Dichas soluciones son óptimas si resuelven el problema

$$V_k(x_k) = \max \{E[r(x_k,a) + \sum_{j=k+1}^{N} r(x_j, a_j)] \mid a_k \varepsilon A_k\}$$

Sin embargo, esto es equivalente a

$$V_k(x_k) = \max_{a_k \varepsilon A_k} E\{r(x_k,a)\} + \sum_{j=k+1}^{N} \max_{a_j \varepsilon A} E\{r(x_j,a_j)\}$$

$$= \max_{a_k \varepsilon A_k} \{r(x_k,a) + \sum P_{x_k}(a_k) V_{k+1}(j)\} \qquad \forall\, k.$$

2.3 El problema con horizonte infinito.

Considere una cadena de decisión markoviana, como se describe al principio de la sección uno. Si $n \to \infty$ se dice que es una cadena con período infinito. Para una política π dada:

$$V_\pi(i) = E_\pi\left[\sum_{t=0}^{\infty} r(x_t, A_t) \mid x_0 = i\right], \qquad \forall\, i \in S$$

donde $V_\pi(i)$ representa el beneficio total esperado. El objetivo en un problema de horizonte infinito es encontrar una política óptima π^* tal que

$$V_{\pi^*}(i) = V^*(i) = \sup_\pi V_\pi(i) \qquad i \in S$$

Debido a que existe un número infinito de políticas, la existencia de una política óptima no es obvia.

Las cadenas de decisiones markovianas con período infinito y descuento han sido bien estudiado desde la década de los sesenta. El uso de un factor de descuento es motivado por el efecto económico de actualización del dinero a valor presente.

Se dice que una política π^* es α-óptima si su valor esperado con factor de descuento $0 < \alpha < 1$ es máximo para cada estado inicial, esto es $V_{\pi^*}(i) = V^*(i)$ para todo estado i en S.

Teorema 1. La ecuación de optimalidad se cumple:

$$V^*(i) = \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a) V^*(j)\}, \quad \forall\, i \in S.$$

Prueba: Sea π una política arbitraria. Entonces, el beneficio total esperado es dado como

$$V_\pi(i) = \sum_{a \in A} Qa \,[r(i,a) + \alpha \sum_j P_{ij}(a) V_\pi(j)]$$

que implica que la política π toma la acción a en el instante 0 con probabilidad Qa, $a \in A$. Sin embargo, ya que $V_\pi(j) \leq V^*(j)$,

$$\begin{aligned} V_\pi(i) &\leq \sum_{a \in A} Qa \,[r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)] \\ &\leq \sum_{a \in A} Qa \max_a [r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)] \\ &= \max_a [r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)]. \end{aligned}$$

Debido a que π es arbitraria, se tiene

$$V^*(i) \leq \max_a [r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)].$$

En otro lado, por definición de V* se observa que

$$V^*(i) \geq \max_a [r(i,a) + \alpha \sum P_{ij}(a)\, V^*(j)]$$

por lo tanto

$$V^*(i) = \max_a [r(i,a) + \alpha \sum_j (a)\, V^*(j)]$$

y la prueba termina.

Teorema 2. Sea f política estacionaria tal que para todo $i \in S$

$$r(i, f(i)) + \alpha \sum_j P_{ij}(f(i)) V^*(j) = \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a) V^*(j)\}$$

entonces $V_f(i) = V^*(i)$ para toda $i \in S$ y desde luego, f es α-óptima.

Para demostrar el teorema 2 se necesita el siguiente lema.

Lema 1. El mapeo

$$(T_f u)(i) = r(i,f(i)) + \alpha \sum_j P_{ij}(f(i)) u(j)$$

tiene las siguientes propiedades:

1). $u \leq v$ implica que $T_f u \leq T_f v$,

2). $T_f V_f = V_f$,

3). $\lim_{n\to\infty} T_f^n u = V_f$.

Prueba 1). Por la definición de T_f, obviamente que

$$T_f u = r(f) + \alpha P(f) u \leq r(f) + \alpha P(f) v = T_f v.$$

2). Dado que $(T_f V_f)(i) = r(i,f(i)) + \alpha \sum_j P_{ij}(f(i)) V_f(j)$ es equivalente a

$$V_f(i) = r(i,f(i)) + \alpha \sum_j P_{ij}(f(i)) V_f(j).$$

3). Observe que si $T_f u = r(f) + \alpha P(f)u$. Entonces:

$$T_f^N u = r(f) + \alpha\, P(f)\, T_f^{n-1} u$$

$$= \sum_{n=0}^{N-1} [\alpha P(f)]^n\, r(f) + [\alpha P(f)]^N u$$

Por ser u una función acotada y $0 \leq \alpha < 1$, podemos hacer $N \to \infty$ y obtener que $\lim_{N\to\infty} T_f^N u = V_f$.

Prueba. (Teorema 2). Observe que

$$(T_f V^*)(i) = r(i,f(i)) + \alpha \sum_j P_{ij}(f(i))\, V^*(j)$$

$$= \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)\} = V^*(i)$$

y se obtiene que $T_f^1 V^* = V^*$. Suponga que $T_f^{n-1} V^* = V^*$. Entonces $T_f^n V^* = T_f(T_f^{n-1} V^*) = T_f V^* = V^*$. Cuando $n \to \infty$ y usando el lema 1, se tiene $V_f(i) = V^*(i)$ $\forall$ $i \in S$. Esto es, f es α óptima. ▲

Ahora, se define el mapeo T como sigue

$$(Tu)(i) = \max_a \{r(i,a) + \alpha \sum_j p_{ij}(a) u(j)\} \qquad \forall\ i \in S.$$

Lema 2. $\lim_{n\to\infty} T^n u = V^*, \qquad \forall\ i \in S.$

Prueba. Suponga que a_0 maximiza $r(i,a) + \alpha \sum_j P_{ij}(a) V^*(j)$ y que a_1 maximiza $r(i,a) + \alpha \sum_j P_{ij}(a)\, u(j)$.

Entonces

$$(Tu)(i) - V^*(i) = [r(i,a_1) + \alpha \sum_j P_{ij}(a_1)u(j)]$$

$$- [r(i,a_0) + \alpha \sum_j P_{ij}(a_0)\, V^*(j)]$$

$$\leq [r(i,a_1) + \alpha \sum_j P_{ij}(a_1)u(j)]$$

$$- [r(i,a_1) + \alpha \sum_j P_{ij}(a_1)\, V^*(j)]$$

$$= \alpha \sum_j P_{ij}(a_1)[u(j) - V^*(j)]$$

$$< \alpha \sum_j P_{ij}(a_1)B = \alpha B$$

donde $B = 2 \max \{\sup u(i), \sup V(i)\}$ y

$$V^*(i) - (Tu)(i) \leq \alpha \sum_j P_{ij}(a_0)[V^*(j) - u(j)]$$

$$\leq \alpha \sum_j P_{ij}(a_0)\, B = \alpha\, B.$$

Entonces $|(Tu)(i) - V^*(i)| < \alpha B$. Suponga que $|T^{n-1}u(i) - V^*(i)| < \alpha^{n-1} B$. Sea a_1^n la acción que maximiza $r(i,a) + \alpha \sum_j P_{ij}(a)\ (T^{n-1}u)(j)$. Entonces

$$(T^n u)(i) - V^*(i) \leq \alpha \sum_j P_{ij}(a_1^n)\ [T^{n-1}u(j) - V^*(j)]$$

$$< \alpha \sum_j P_{ij}(a_1^n)\alpha^{n-1} B = \alpha^n B$$

y

$$V^*(i) - (T^n u)(i) \leq \alpha \sum_j P_{ij}(a_0)[V^*(j) - (T^{n-1}u)(j)]$$

$$< \alpha \sum_j P_{ij}(a_0)\ \alpha^{n-1} B = \alpha^n B.$$

Esto es $|(T^n u)(i) - V^*(i)| < \alpha^n B \quad \forall\, n \geq 1$ cuando $n \to \infty$ se tiene el resultado deseado.

Se observa que u es una función arbitraria y que el resultado del lema 2 nos indica la manera de buscar una política óptima. Más bien, la aplicación del mapeo T a la función u es conocido como el método de aproximación sucesiva que se discutirá en más detalle en adelante.

El siguiente teorema muestra que la unicidad de la solución acotada V^* de la ecuación de optimalidad.

Teorema 3. V^* es la única solución acotada de la ecuación de optimalidad.

Prueba. De manera semejante del teorema 2, se puede demostrar que

$$T^n V^* = V^* \qquad \forall\, n \geq 1.$$

Cuando $n \to \infty$ y usando el lema 2 se tiene el resultado requerido que es lo que V^* es la solución única de la ecuación de optimalidad.▲

2.4 Métodos de solución.

Se discuten tres métodos para calcular el valor de la función óptima V^* y determinar la política óptima π^*.

A. Método de Aproximación Sucesiva.

Sea $V_0(i)$ una función acotada arbitraria y se defina V_1 por

$$V_1(i) = \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a)\, V_0(j)\} \qquad \forall\, i \in S$$

En general, para $n > 1$,

$$V_n(i) = \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a)\, V_{n-1}(j)\}, \qquad \forall\, i \in S$$

y por el lema 2, cuando $n \to \infty$ se tiene que

$$\lim V_n(i) = V^*(i) \qquad \forall\, i \in S$$

Como se implica en el teorema 2, se podría determinar la función de valor óptimo V^*. Es posible conocer la política óptima si existe una política estacionaria f, que satisface

$$r(i,f(i)) + \alpha \sum_j P_{ij}(f(i))\, V^*(j)$$

$$= \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)\} \qquad \forall\ i \in S$$

En la práctica, el método de aproximación sucesiva podría ser usado cuando una aproximación a una política óptima es disponible, pues en pocas iteraciones es posible aproximar a V^*.

B. Método de Mejoramiento de Política.

El método de mejoramiento de política fué propuesto por Ronald A. Howard en 1960. El método es simple y suficientemente valioso para encontrar efectivamente una política óptima de un problema con período infinito y factor de descuento dado.

Como en el teorema 2 del capítulo presente se indica, una vez que V* está determinado, la política óptima es aquella acción a que maximiza $r(i,a) + \alpha \sum_j P_{ij}(a)\, V^*(j)$, cuando esté en el estado i. Es interesante observar el siguiente caso. Se suponen que para alguna política estacionaria g, se ha obtenido V_g, el valor esperado bajo la ejecución de esta política; y se encuentre una nueva política h tal que h(i) es la acción que maximiza $r(i,a) + \alpha \sum_j P_{ij}(a)\, V_g(j)$, cuando esté en i. Entonces de allí surge una pregunta ¿Qué tan buena es la política h comparandola con la g?. En el siguiente teorema se demostrará que la política h es al menos tan buena como la g, y si $V_h(i) = V_g(i)$ para todo i, entonces, las g y h ambas son óptimas. Como se ve, el siguiente teorema forma base teórica del método de mejoramiento de política.

Teorema 4. Sea g una política estacionaria con valor esperado V_g y sea h la política tal que para todo $i \in S$

$$r(i,h(i)) + \alpha \sum_j p_{ij}(h(i))V_g(j) = \max_a \{r(i,a) + \alpha \sum_j p_{ij}(a)V_g(j)\}$$

entonces $V_h(i) \geq V_g(i)$, $\forall\, i \in S$. Si $V_h(i) = V_g(i)$ para todo $i \in S$, entonces $V_g = V_h = V^*$.

Prueba: Observe que

$$\begin{aligned} T_h V_g(i) &= r(i,h(i)) + \alpha \sum_j p_{ij}(h(i))\, V_g(j) \\ &\geq r(i,g(i)) + \alpha \sum_j p_{ij}(g(i))\, V_g(j) \\ &= V_g(i) \qquad\qquad \forall\, i \in S. \end{aligned}$$

De aqui es sencillo verificar que

$$T_h^n V_g = T_h\,(T_h^{n-1} V_g) \geq T_h V_g \geq V_g.$$

Cuando $n \to \infty$ el lema 2 implica que $V_h \geq V_g$.

Si $V_h = V_g$, entonces

$$\begin{aligned} V_g(i) = V_h(i) &= r(i,h(i)) + \alpha \sum_j P_{ij}(h(i))\, V_g(j) \\ &= r(i,h(i)) + \alpha \sum_j P_{ij}(h(i))\, V_h(j) \\ &= \max_a \{r(i,a) + \alpha \sum_j P_{ij}(a)\, V_g(j)\} \end{aligned}$$

y $V_g = V_h$; h y g satisfacen la ecuación de optimalidad. Por el teorema 3, $V_g = V_h = V^*$.▲

Corolario: El algoritmo de mejoramiento de política converge a una política óptima en un número finito de iteraciones.

Prueba. Dado que sólo existe un número finito de políticas, cuando el conjunto de estados es finito, por el teorema 4, cada iteración resulta un mejoramiento estricto, entonces no se repiten políticas. Cuando no es posible mejorar la V, entonces h es igual que g, la cual es óptima por el resultado del teorema 3 o la unicidad de la solución acotada. ▲

Lema 3: Si $u \geq T_f u$ $\forall$ $f \in A$, entonces $u \geq V^*$. Si en particular $u = V_{\pi^*}$, entonces π^* es la política óptima.

Prueba: Si $u \geq T_f u$, por lema 1, $u \geq T_f^n u$, cuando $n \to \infty$ y usando el lema 2 $u \geq V^*$. ▲

Algoritmo de mejoramiento de políticas (Howard 1960).

Sea $f_0 \in A$ una regla de decisión (arbitraria). Calcule

$$V_{f_0} = \sum_{t=0}^{\infty} \alpha^t p^t(f_0) r(f_0)$$

$$= [I - \alpha p(f_0)]^{-1} r(f_0)$$

Sea $k = 0$.

Paso 1: Si $V_{f_k} \geq T_f V_{f_k}$, $\forall f \in A$, entonces

$$V_{f_k} = \max_{\pi} \{V(\pi) \mid \pi \in A^{\infty}\} \quad \text{por el lema 3.}$$

Paso 2: Suponga que $T_g V_{f_k} \geq V_{f_k}$ para algún $g \in A$, entonces $V_g \geq V_f$ por el teorema 4. Hacer $k := k+1$, $f_k = g$ y regresar al paso 1.

El procedimiento de mejorar política suministra una secuencia convergente monótona de políticas y logra la política óptima en un número finito de iteraciones. La desventaja del método es la dificultad para calcular $[I-\alpha p(f)]^{-1}$ cuando el orden de la matriz crece.

C. Programación Lineal.

Debido a que la función de valor óptima V^* es la función de valor mínimo que satisface, por el lema 3,

$$u(i) \geq \max_{a} \{r(i,a) + \alpha \sum_{j} p_{ij}(a)\, u(j)\}, \qquad \forall\, i \in S,$$

entonces, V^* es la única solución del problema de optimización

$$\min_{u} \sum_{t \in S} u(t)$$

S. A.

$$u(i) \geq \max_{a} \{r(i,a) + \alpha \sum_{j} p_{ij}(a)\, u(j)\} \qquad \forall\, i \in S,$$

o bien

$$u(i) \geq r(i,a) + \alpha \sum_{j} p_{ij}(a)\, u(j) \qquad \forall\, i \in S,\ a \in A.$$

Sin embargo, esto es un problema de programación lineal y en el caso de que el conjunto de estados es finito, se puede resolver por una técnica bien conocida como el método de Simplex.

A continuación, se discute el algoritmo de programación lineal.

Primero, para cualquiera etapa n, se define $\lambda_n(i,f)$ la probabilidad de la coyuntura de estar en estado $i \in S$ y efectuar la decisión $f \in A_i$. Se considera el problema de maximización bajo la distribución inicial $q = [q_1, \ldots, q_n]$ debido a que se logra simultáneamente una política óptima para cada estado inicial. Como $\lambda_n(i,f)$ obedece la ley de probabilidad p_{ij}^{f}, se podría escribir

(1)

$$\sum_{f\in A_j} \lambda_n(j,f) = \begin{cases} \lambda_0(j) = q_j & \text{si } n = 0 \\ \sum_{i\in S} \sum_{f\in A_i} p_{ij}^f \lambda_{n-1}(i,f) & n=1,2,\ldots,j\in S \end{cases}$$

donde q_j = probabilidad de que el sistema está en estado j en el tiempo 0.

Lema 4. Cualquiera solución no negativa $\lambda_n(i,f)$ de (1) es una distribución de probabilidad y el valor esperado total del caso con descuento correspondiente es acotado.

Prueba: Observe que $\sum_{j\in S} p_{ij}^f = 1$. De donde

$$\sum_{j\in S} \sum_{f\in A_j} \lambda_n(j,f) = \begin{cases} \sum_{j\in S} q_j & \text{si } n = 0 \\ \sum_{i\in S} \sum_{f\in A_i} \lambda_{n-1}(i,f) & n = 1,2,\ldots,. \end{cases}$$

la suposición de que q es una distribución de probabilidad, implica que

$$\sum_{i\in S} \sum_{f\in A_i} \lambda_n(i,f) = 1 \qquad n = 0,1,2,\ldots,.$$

la no negatividad de $\lambda_n(i,f)$ prueba la primera parte del lema. Defina:

$$V\alpha(\pi) = \sum_{n=0}^{\infty} \alpha^n p^n(\pi) r(f_{n+1})$$

y sean $\bar{r} = \max_{i,f} r(i,f)$ y $\underline{r} = \min_{i,f} r(i,f)$. Entonces se tiene que

$$\frac{\underline{r}}{1-\alpha} e \leq V\alpha(\pi) \leq \frac{\bar{r}}{1-\alpha} e$$

donde $e = [1,1,\ldots,1]^{-1}$. ▲

Como un resultado del lema 4, se obtiene la siguiente función objetivo

$$\max \sum_{n=0}^{\infty} \alpha^n \sum_{i \in S} \sum_{f \in A_i} r(i,f) \lambda_n(i,f)$$

bajo la restricción (1) y $\lambda_n(i,f) \geq 0$ $\forall n$, $i \in S$, y $f \in A_i$.

Se observa que este problema es similar a un problema de programación lineal estándar. Sin embargo, debido a que se contiene un número infinito de restricciones y variables, no se puede analizar con la teoría clásica de programación lineal. Entonces se define un conjunto de nuevas variables $x(j,f)$ como:

$$x(j,f) = \sum_{n=0}^{\infty} \alpha^n \lambda_n(j,f) \qquad \forall\, j \in S,\ f \in A_j,$$

usando $x(j,f)$, se obtiene el siguiente problema de programación lineal estándar:

$$\max \sum_{j \in S} \sum_{f \in A_j} r(j,f) x(j,f)$$

S.A.

$$(2) \qquad \sum_{f \in A_j} x(j,f) - \alpha \sum_{i \in S} \sum_{f \in A_i} p_{ij}(f) x(i,f) = q_j$$

$$x(j,f) \geq 0 \quad \forall\, j \in S, \quad f \in A_i.$$

Teorema 5. Si se restringe el sistema (2) a las variables seleccionadas $x(i,f)$ por cualquiera política estacionaria, entonces:

a). El correspondiente subsistema tiene una única solución;

b). Si $q_j \geq 0$ $(j \in S)$, $x(j,f) \geq 0$ $(j \in S)$;

c). Si $q_j > 0$ $(j \in S)$, $x(j,f) > 0$ $(j \in S)$.

Prueba: Considere el sistema homogéneo asociado con

$$x(j) - \alpha \sum_{i \in S} p_{ij}\, x(i) = q_j \qquad j \in S$$

por suponer $q_j = 0$ $(j \in S)$. Este sistema debe tener al menos una solución nula. Si x fuera otra solución no cero, se define $I^- = \{i \mid x(i) < 0\}$ Sumando

$$V\alpha(\pi) = \sum_{n=0}^{\infty} \alpha^n p^n(\pi)\, r(f_{n+1}) \quad \text{un vector Nx1}$$

sobre todos $j \in I^-$, se obtiene

$$\sum_{i \in I^-} [1 - \alpha \sum_{j \in I^-} p_{ij}]\, x(i) - \alpha \sum_{i \notin I^-} \sum_{j \in I^-} p_{ij}\, x(i) = \sum_{j \in I^-} q_j$$

ya que $q_j = 0$, y $\alpha < 1$ implica que $1 - \alpha \sum_{j \in I^-} p_{ij} > 0$, el cual implica que la primera sumatoria debería ser estrictamente negativa. Esto es una contradicción excepto cuando $I^- = \emptyset$. Por el mismo argumento, se concluye que $I^+ = \emptyset$ donde $I^+ = \{i \mid x(i) > 0\}$. Esto es, la solución nula es la única del sistema. Además, es cierto que existe una única solución para cualquier valor de q_j.

b). Si $q_j = 0$, esto implica que $I^- = \emptyset$, => $x(j) \geq 0$ $(j \in S)$.

c). Si $q_j > 0$, se define $I^- = \{i \mid x(i) \leq 0\}$. Esto implica que $I^- = \emptyset$ o $x(j) > 0$. ▲

Usando el teorema 5, en el siguiente se demuestra una relación importante entre políticas estacionarias y soluciones básicas factíbles del sistema (2) cuando $q_j > 0$ $(j \in S)$.

Teorema 6: Cuando $q_j > 0$ $(j \in S)$, existe una correspondencia una a una entre políticas estacionarias y soluciones básicas factíbles del (2). Además cualquiera solución básica factíble es no degenerada.

Prueba: Por c) del teorema 5, una política estacionaria tiene una única solución de

$$x(j) - \alpha \sum_{i \in S} p_{ij} x(i) = q_j, \qquad j \in S$$

que tiene N variables positivas. Esto es, por definición una solución básica no degenerada factíble del (2). Por otro lado, si x(i,f) es una solución factíble del (2), se tiene

$$\sum_{f\in A_j} x(i,f) = q_j + \alpha \sum_{i\in S} \sum_{f\in A_i} p_{ij}(f)x(i,f) \geq q_j > 0 \quad \forall\, j\in S$$

desde entonces, al menos existe una variable x(j,f) que es positiva. Entonces, existe exactamente una x(i,f) > 0 para cada estado j, esto define únicamente una política escionaria.

Ahora, si se considera que

$$\max \sum_{j\in S} \sum_{f\in A_j} r(j,f)x(j,f)$$

S.A.

$$\sum_{f\in F_j} x(j,f) - \alpha \sum_{i\in S} \sum_{f\in A_i} p_{ij}(f)\, x(i,f) = q_j \qquad \forall\, j\in S$$

$$x(j,f) \geq 0 \qquad \forall\, j\in S, \quad f\in F_j$$

como problema primal, entonces su dual es representado por

$$\min \sum_{i\in S} q_i\, u_i$$

S.A.

$$(3) \qquad u_i \geq r(i,f) + \alpha \sum_{j\in S} p_{ij}(f)\, u_j \qquad \forall\, i\in S,\ f\in A_i.$$

Esto se puede derivar inmediatamente de la siguiente manera: para una política estacionaria óptima $\pi^* = f^\infty$, se tiene

$$U_\alpha\ (f^\infty) \geq L_g U_\alpha(f^\infty) = r(g) + \alpha\ p(g)\ U_\alpha(f^\infty) \qquad \forall\ g \in A.$$

si se escribe la ecuación anterior en elementos uno por uno, se obtiene (3). La función objetivo en la distribución inicial q es

$$q\ U_\alpha(f^\infty) = \sum_{i \in S} q_i\ u_i$$

esto es el problema dual.

Teorema 7: Cuando $q_j > 0$, el (2) tiene una solución básica óptima y su dual tiene una única solución óptima. Cualesquiera política estacionaria óptima asociada con él mantiene óptima para cualquier $q_j \geq 0$.

Prueba: Por el teorema 5, obviamente existen soluciones factibles y por el lema 4 la función objetivo es acotada. Esto nos garantiza la existencia de una solución óptima para ambos problemas primal y dual.

Por teorema 6, cualesquiera solución básica va a ser no degenerada, y por estas condiciones inactivas complementarias, cualquiera solución óptima del dual debería satisfacer el sistema correspondiente de N igualdades duales, entonces, la solución dual es única.

Se nota que la optimalidad de una solución básica factible dada de un problema de programación lineal depende sólo de

la función objetivo y no de q_j. El cambio del segundo sólo afecta la factibilidad de cualquier valor no negativo de q_j. ▲

Debido a la discusión el algoritmo de mejoramiento de políticas es sólo una extensión especial de programación lineal el cual tiene la propiedad de que se ejecutan simultáneamente las operaciones de pivote para muchas variables.

Para problemas con conjunto de estados finito, se puede calcular V^* por método de programación lineal o de mejoramiento de políticas.

2.5 Ejemplos ilustrativos.

Ejemplo 1. (Problema de producción-inventario). Considere un sistema de producción-inventario que desea determinar su política óptima de producción-inventario para los próximos N períodos. En este sistema puede ordenarse una cantidad u_k de artículos al principio de cada período (k=1,2,...,N). Cada artículo tiene un costo c. Asimismo, puede almacenarse una cantidad x_k de artículos en cada periodo a un costo unitario h de inventario. La demanda de artículos en cada período denotado w_k, es una variable aleatoria que sigue una función de distribución conocida. Se supone que las demandas de artículos de un período con respecto a otro son estocásticamente independientes y que la demanda insatisfecha en cada período se pierde con un costo unitario c_p de insatisfacción.

Sea $V_k(x_k)$ el mínimo valor obtenido al aplicar una política óptima de producción-inventario del período k al N dado que se tiene un nivel de inventario x_k al principio del período k. Se observa que lo que se desea es $V_1(x_1)$ y que para el caso especial en que k=N se tiene que

$$V_N(x_N) = \min \left[\underset{w_N}{E} \{q(x_N, u_N, w_N)\} \mid u_N \geq 0 \right]$$

donde

$$q(x_N, u_N, w_N) = cu_N + h \max\{0,\ x_N + u_N - w_N\} + c_p \max\{0,\ w_N - w_N - u_N\}.$$

Aquí se supone que los artículos al principio del período N+1 no tienen valor o bien $V_{N+1}(x_{N+1}) = 0$.

En general, para k=1,2,...,N-1 se tiene

$$V_k(x_k) = \min_{W_k} E\left[q(x_k, u_k, w_k) + V_{k+1}(x_{k+1}) \mid u_k \geq 0\right]$$

donde

$$q(x_k,u_k,w_k) = cu_k + h\max\{0,x_k+u_k-w_k\} + c_p \max\{0,w_k-x_k-u_k\}$$

y

$$x_{k+1} = \max\{0, x_k + u_k - w_k\}.$$

que son las ecuaciones recursivas de la programación dinámica.

Suponga que las demandas de artículos y los niveles de inventario son variables enteras no-negativas. El costo unitario c_p es igual a tres y los costos unitarios de almacenamiento y de producción son iguales a uno. Suponga que la capacidad máxima de almacenamiento es dos y la función de distribución de probabilidades de las demandas es la misma para todos los períodos e igual a

w	0	1	2
p(w)	0.1	0.7	0.2

Se desea determinar la política óptima para tres periodos dado que se tiene un nivel de inventario inicial cero al inicio del primer periodo.

Para los datos del problema la condición de frontera es

$V_4(x_4) = 0$ y las ecuaciones recursivas son

$$V_k(x_k) = \min_{W_k} E\left[q(x_k,u_k,q_k) + V_{k+1}(x_{k+1}) \mid 0 \leq u_k \leq 2-x_k\right]$$

donde $q(x_k,u_k,w_k) = u_k+\max\{0,\ x_k+u_k-w_k\} + 3\max\{0,\ w_k-x_k-u_k\}$ y $x_{k+1} = \max\{0,\ x_k + u_k - w_k\}$. Específicamente:

$V_3 = \underset{w_3}{E}[q(x_3,u_3,w_3)] + V_4(x_4)$					
x_3 \ u_3	0	1	2	V_3^*	u_3^*
0	3.3	1.7	2.9	1.7	1
1	0.7	1.9	-	0.7	0
2	0.9	-	-	0.9	0

$V_2 = \underset{w_2}{E}[q(x_2,u_2,w_2)] + V_3(x_3)$					
x_2 \ u_2	0	1	2	V_2^*	u_2^*
0	5.0	3.3	3.82	3.3	1
1	2.3	2.82	-	2.3	0
2	1.82	-	-	1.82	0

Por lo tanto la política óptima es ordenar un artículo si el nivel de inventario es cero y nada si ocurre lo contrario debido a que $x_1 = 0$ y:

V_1	6.6	4.9	5.3	4.9 (óptimo)
U_1	0	1	2	1 (óptimo)

Ejemplo 2. (Problema de mantenimiento). Considere una máquina que opera sincrónicamente, es decir, una vez por hora. En cada periodo hay dos estados, uno es en operación (estado 1); y otro es en condición de falla (estado 2). Si la máquina funciona bien, hay una ganancia de $3.00 por periodo y la probabilidad de quedar en estado 1 en el siguiente periodo es 0.7; la probabilidad de ir al estado 2 es 0.3. Si la máquina está en condición de falla, se tienen dos acciones para reparar la máquina, una es una reparación rápida que requiere un costo de $2.00, con la probabilidad de ir al estado 1 igual a 0.6; otra es una reparación normal que requiere un costo $1.00 con la probabilidad de ir al estado 1 igual a 0.4. Se desea determinar la política óptima y el valor óptimo. El factor de descuento es 0.9.

Se definen $F = \{f_1, f_2\}$ y $\alpha = 0.9$. Donde f_1 representa la política de reparación rápida y f_2 representa la política de reparación normal.

$$P(f_1) = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} \qquad r(f_1) = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

$$P(f_2) = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \qquad r(f_2) = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

a. Se aplica el método de mejoramiento de política. Suponga que la política inicial es $a_0 = f_1$. Entonces

$$V = V(a_0) = [I - \alpha P(a_0)]^{-1} r(a_0)$$

$$= \begin{bmatrix} 0.37 & -0.27 \\ -0.54 & 0.64 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 1380/91 \\ 880/91 \end{bmatrix}$$

Considere ahora la relación

$$V(2) = r(2,f_2) + \alpha \sum_j p\ (f_2)\ V(j)$$

$$= -1 + 0.9\ [0.4 \quad 0.6] \begin{bmatrix} 1380/91 \\ 880/91 \end{bmatrix}$$

$$= 881/91 > 880/91$$

entonces la política f_2 es mejor que f_1. Sea $a_1 = f_2$ y note que

$$V = V(a_1) = [I - \alpha P(a_1)]^{-1} r(a_1)$$

$$= \begin{bmatrix} 1-0.9x0.7 & -0.9x0.3 \\ -0.9x0.4 & 1-0.9x0.6 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.46 & 0.27 \\ 0.36 & 0.37 \end{bmatrix} \frac{1}{0.073} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 1110/73 \\ 710/73 \end{bmatrix}$$

Por lo tanto la política óptima es f_2 y el valor óptimo es

$$\begin{bmatrix} 1110/73 \\ 710/73 \end{bmatrix}$$

b. Se aplica el método de aproximación sucesiva con $e = 0.001$. Sea $V_0 = 0$.

$$TV_0 = \max\left\{\begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}\right\}$$

$$T^2V_0 = \max\left\{\begin{bmatrix} 3 \\ -2 \end{bmatrix} + 0.9 \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 4.62 \\ -0.74 \end{bmatrix}\right\}$$

$$\left\{\begin{bmatrix} 3 \\ -1 \end{bmatrix} + 0.9 \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 4.62 \\ -0.46 \end{bmatrix}\right\}$$

$$T^3V_0 = \begin{bmatrix} 5.7864 \\ 0.4148 \end{bmatrix} \quad \cdots \quad T^{10}V_0 = \begin{bmatrix} 10.63733545 \\ 5.157454821 \end{bmatrix}$$

$$T^{20}V_0 = \begin{bmatrix} 13.6126019 \\ 8.133149849 \end{bmatrix} \cdots T^{30}V_0 = \begin{bmatrix} 14.65029881 \\ 9.170846667 \end{bmatrix}$$

$$T^{40}V_0 = \begin{bmatrix} 15.01172343 \\ 9.532274802 \end{bmatrix} \cdots T^{50}V_0 = \begin{bmatrix} 15.13723453 \\ 9.65778729 \end{bmatrix}$$

$$T^{60}V_0 = \begin{bmatrix} 15.18168464 \\ 9.70223258 \end{bmatrix} \quad \ldots \; T^{70}V_0 = \begin{bmatrix} 15.19645194 \\ 9.7169436 \end{bmatrix}$$

por lo tanto, el valor aproximado al óptimo con un error de 0.001 es igual al $T^{70}V_0 \doteq \begin{bmatrix} 15.19645194 \\ 9.7169436 \end{bmatrix}$

c. Se aplica el método de programación lineal.

$$\min \; x_1 + x_2$$

S.A.

$$\begin{aligned} 0.27x_1(f_1) - 0.27x_2(f_1) &\geq 3 \\ -0.54x_1(f_1) + 0.64x_2(f_1) &\geq -2 \\ 0.37x_1(f_2) - 0.27x_2(f_2) &\geq 3 \\ -0.36x_1(f_2) + 0.46x_2(f_2) &\geq -1 \\ x_1(f_i) \; , \; x_2(f_i) &\geq 0 \qquad i=1,2 \end{aligned}$$

la solución óptima es $x_1 = 1110/73$, $x_2 = 710/73$ y se observe que . coincide con la obtenida por el método de mejoramiento de política.

$$T^{60}V_0 = \begin{bmatrix} 15.18168464 \\ 9.70223258 \end{bmatrix} \quad \ldots \quad T^{70}V_0 = \begin{bmatrix} 15.19645194 \\ 9.7169436 \end{bmatrix}$$

por lo tanto, el valor aproximado al óptimo con un error de 0.001 es igual al $T^{70}V_0 = \begin{bmatrix} 15.19645194 \\ 9.7169436 \end{bmatrix}$

c. Se aplica el método de programación lineal.

$$\min \; x_1 + x_2$$

S.A.

$$\begin{aligned} 0.27x_1(f_1) - 0.27x_2(f_1) &\geq 3 \\ -0.54x_1(f_1) + 0.64x_2(f_1) &\geq -2 \\ 0.37x_1(f_2) - 0.27x_2(f_2) &\geq 3 \\ -0.36x_1(f_2) + 0.46x_2(f_2) &\geq -1 \\ x_1(f_i) \; , \; x_2(f_i) &\geq 0 \qquad i=1,2 \end{aligned}$$

la solución óptima es $x_1 = 1110/73$, $x_2 = 710/73$ y se observe que . coincide con la obtenida por el método de mejoramiento de política.

CAPITULO III

CADENAS DE DECISION MARKOVIANA CON COSTOS NEGATIVOS

Una variedad importante de problemas dinámicos reales están clasificados de acuerdo a su estructura como problemas de decisión markoviana con costo negativo. Casos típicos de estos problemas son: inventario, reemplazo de equipo, confiabilidad, etc. Los métodos desarrollados para obtener la correspondiente solución óptima son llamados métodos de Programación Dinámica Negativa.

En este capítulo se describe la forma general de los problemas de decisión markoviana con costo negativo así como algunos problemas reales simplificados. Se analizan las características estructurales de este tipo de problemas y se establece la existencia y forma de las políticas estacionarias óptimas así como los métodos de solución disponibles. Finalmente, se aplican los métodos desarrollados para resolver problemas de paro óptimo, reemplazo de equipo y otros.

3.1 Descripción del problema con costo negativo.

Considere el problema de cadenas de decisión markoviana y suponga que el conjunto de estados S es finito o infinito contable y que el conjunto de decisiones disponibles A es finito. Si se toma una decisión $a \in A$ cuando estamos en el estado $i \in S$, se obtiene un beneficio esperado, denotado como $r(i,a) \in R_-$. El objetivo es determinar la sucesión de decisiones que maximiza el beneficio total esperado.

Considere una cadena de decisión markoviana con período infinito denotada $\{x_n, A_n\}$ donde x_n es el estado de la cadena en la etapa n y A_n el conjunto de decisión ejercida en la misma etapa para una política π. Entonces, el valor:

$$V_\pi(i) = E_\pi\left[\sum_{n=0}^{\infty} r(x_n, A_n) \mid x_0 = i\right]$$

representa el beneficio total esperado. Observe que por ser $r(i,a) \in R_-$ se tiene que $V_\pi(i)$ está bien definida (podría ser $-\infty$). Sea el valor

$$V^*(i) = \sup_\pi V_\pi(i)$$

y si π^* la política tal que

$$V_{\pi^*}(i) = V^*(i), \qquad \forall\, i \in S$$

Entonces se dice que π^* es la política óptima.

Ejemplo 1. (Paro óptimo). Considere un proceso dinámico con estados posibles 0,1,2,..., en donde para cada estado i, se puede tomar la decisión de paro (acción 1) y recibir una cierta ganancia terminal R(i); o bien, se puede tomar la decisión de pagar un cierto costo C(i) e ir al estado j (acción 2) de acuerdo con las probabilidades de transición. ($P_{ij} \geq 0$). Se proponen dos condiciones para analizar el problema.

i. $\sup_i R(i) < \infty$

ii. $\sup_i C(i) < 0.$

Se desea determinar una política óptima de este proceso de manera tal que la ganancia sea máxima, suponiendo que el proceso va al estado ∞ cuando se tome la decisión de paro y que una vez en ese estado permanece indefinidamente.

Sean

$$C(i,f_1) = R(i) > 0; \qquad C(i,f_2) = C(i) \in R_-;$$

$$C(\infty, f_1 \text{ ó } f_2) = 0; \qquad P_{i,\infty}(f_1) = 1$$

$$P_{ij}(f_2) = P_{ij}; \qquad P_{\infty,\infty}(f_1 \text{ ó } f_2) = 1$$

donde, f_1 y f_2 representan la regla de decisión de paro y de continuación, respectivamente.

Nótese que este proceso no cumple las condiciones de la programación dinámica negativa, debido a que no todos los costos son negativos ($C(i,f_1) > 0$). Es posible transformar el proceso a uno que satisfaga todas las características de programación dinámica negativa de la siguiente manera. Sea $R = \sup_i R(i)$, considere que cuando esté en el estado i, se puede parar y recibir una ganancia terminal $R(i) - R$ (o equivalentemente, pagar un costo terminal $R - R(i) \geq 0$); o bien ir al estado j con $P_{ij} \geq 0$ y pagar un costo de continuación $C(i)$. Se denomina este proceso, el proceso modificado.

Suponga que para cualquiera política π se denota por V_π y $\bar{V}_\pi$ las funciones de beneficio esperado del proceso original y modificado, respectivamente, entonces, para políticas que paran con probabilidad 1, se tiene que

$$\bar{V}_\pi(i) = V_\pi(i) - R, \qquad \forall\ i \in S.$$

Sin embargo, por la condición (ii), cualquiera política π que no para con probabilidad 1 tiene $V_\pi(i) = \bar{V}_\pi(i) = -\infty$. Por lo tanto, cualquiera política que es óptima para el proceso original es óptima para el proceso modificado, y viceversa. Por otra parte el proceso modificado ya es un proceso de decisión markoviana con costo negativo, de manera tal que, su ecuación de optimalidad está dada por

$$\tilde{V}(i) = \max \begin{cases} R(i) - R \\ \\ C(i) + \sum_{j=0}^{S} p_{ij} \tilde{V}(j) \end{cases} \qquad \forall\, i \in S.$$

esto es, el beneficio esperado total del proceso modificado cuando se inicie en el estado i. Equivalentemente

$$V(i) = \max \begin{cases} R(i) \\ \\ C(i) + \sum_{j}^{S} p_{ij} V(j) \end{cases} \qquad \forall\, i \in S.$$

Ejemplo 2. (Venta de un inmueble). Considere una persona que desea vender su casa. Una oferta se recibe al principio de cada día y la persona debe decidir inmediatamente si acepta o rechaza la oferta. Se supone que las ofertas sucesivas son independientes entre sí y tomen el valor i con probabilidad P_i, para toda $i \geq 0$. Se supone que se tiene un costo de mantenimiento C por cada día que la casa no es vendida. La persona desea determinar la política óptima de aceptar o rechazar ofertas de manera de tener máximo beneficio.

El problema puede formularse como cadena de decisión markoviana con dos acciones y aunque no se cumple la condición de no-pasitividad de los costos es posible efectuar una transformación para convertir el problema a un problema de paro óptimo cuya función de beneficio óptimo asociado al estado i satisface la ecuación de optimalidad dada como

$$V(i) = \max \{i, -c + \sum_{j=0}^{S} P_j V(i)\}$$

para todo estado $i \in S$.

Ejemplo 3. (Confiabilidad de un equipo). Considere una máquina que puede estar en uno de los dos estados; bueno o malo. Se supone que la máquina produce un artículo al principio de cada día. El artículo producido es bueno o malo dependiendo del estado de la máquina. Se supone que si la máquina está en el estado bueno va al estado malo con una probabilidad q; si está en el estado malo permanece en ese estado hasta que se reemplaza por una nueva. Se supone también que después de la producción del artículo, se puede elegir una de las dos opciones, inspeccionar y no inspeccionar el artículo. Si el artículo se encuentra mal, después de la inspección, se reemplaza inmediatamente la máquina con un costo adicional R. El costo de inspeccionar es I y el costo de un artículo mal producido es C. Se desea determinar la política óptima de inspección de manera tal que el costo total de operación sea mínimo.

Se puede formular el problema como sigue. Sean

$$r(a_1) = \begin{bmatrix} -(I+R+C) \\ -I \end{bmatrix} \qquad r(a_2) = \begin{bmatrix} -C \\ 0 \end{bmatrix}.$$

donde $-(I+R+C)$, $-I$ y $-C \in R_-$, y

$$P(a_1) = \begin{bmatrix} 0 & 1 \\ q & 1-q \end{bmatrix} \qquad P(a_2) = \begin{bmatrix} 1 & 0 \\ q & 1-q \end{bmatrix}$$

en donde a_1 y a_2 representan las acciones de examinar, no examinar el artículo, respectivamente, mientras que los estados 2 y 1 son de bueno o malo, respectivamente.

3.2 Caracterización de soluciones óptimas.

En esta sección se estudian las propiedades de la función de beneficios esperados óptimos de una cadena de decisión markoviana con costos negativos. Siguiendo el mismo camino del caso con descuento, se observa que dicha función satisface la ecuación de optimalidad de la programación dinámica como se presenta por la siguiente proposición:

Proposición 1. La función de beneficios esperados óptimos, denotada por V*, satisface

$$V^*(i) = \max_a \{r(i,a) + \sum_{j=0}^{S} P_{ij}(a)\ V^*(j)\}.$$

Prueba. Es similar a la efectuada en el caso con descuento y no se reproduce.

El teorema que se presenta a continuación caracteriza la condición suficiente de existir una política estacionaria óptima en el caso con costos negativos, la cual es establecida en bases del mapeo de contracción y la teoría de latiz.

Teorema 1. Considere el mapeo $T:R^S \to R^S$ dado como

$$Tv = \max \{r(f) + P(f)\, v \mid f \in A\}$$

y suponga que tiene un punto deficiente, esto es, existe $v \in R^S$ tal que $Tv \geq v$. Entonces existe una política estacionaria $\underset{\sim}{f}^{\infty}$ tal que

$$V(\underset{\sim}{f}^{\infty}) = V^* = \max \{V(\Pi) \mid \Pi \in A^{\infty}\}$$

Asimismo $V(\underset{\sim}{f}^{\infty}) = V^*$ si y solo si $V^* = r(\underset{\sim}{f}) + P(\underset{\sim}{f})V^*$.

Prueba. Sea v el punto deficiente de T. Considere el intervalo $[v,0]$ y observe que es una sublatiz compacta de R^S. Sea $V^* = \sup L_D$. Debido al teorema 2 apéndice B se tiene que $V^* = \sup L_F$ y V^* pertenece a L_F. De donde $V^* = TV^*$ y existe $\underset{\sim}{f}$ tal que

$$TV^* = r(\underset{\sim}{f}) + P(\underset{\sim}{f})V^*$$

Pues A es un conjunto finito. Por otra parte (Lema 1).

$$T^N 0 = \max \{V^N(\Pi) + P^N(\Pi)\, 0 \mid \Pi \in A^{\infty}\}$$

De donde $T^N 0 \geq V^N(\Pi)$ para todo $\Pi \in A^{\infty}$. Sin embargo, dado que $0 = \sup [v,0]$ se tiene (teorema 3, Apéndice B) que $V^* = \lim_n T^N(0) \geq V(\Pi)$ para todo $\Pi \in A^{\infty}$. Asimismo, para cada f tal que

$$V^* = r(f) + P(f)\, V^*$$

$$= \sum_{n=0}^{N-1} P^n(f) r\ (f) + P^N(f) V^*$$

$$\leq\ V^N(f^\infty)$$

Haciendo N tender a infinito se tiene $V^* \leq V(f^\infty)$. Pero, en particular, $V^* \geq V(f^\infty)$ y se implica $V^* = V(f^\infty)$. Esto demuestra la existencia de una política estacionaria tal que $V^* = V(\underset{\sim}{f}^\infty)$.

Finalmente, si $V(f^\infty) = V^*$ se observa que si hacemos

$$T_f(\cdot) = r(f) + P(f)(\cdot)$$

entonces $T_f^N 0$ converge a $V(f^\infty) = V^*$ cuando N tiende a infito. Dado que $T_f(\cdot)$ es continuo se tiene que

$$V^* = V(f^\infty) = \lim T_f(T_f^N 0) = T_f(V(f^\infty)) = T_f\, V^*$$

y la demostración se termina.

3.3 Métodos de solución.

Los métodos disponibles para encontrar la política estacionaria óptima y el valor óptimo de la función de beneficios esperados óptimos se desarrollan en esta sección.

El método de aproximación sucesiva es un método sencillo y válido para encontrar el valor óptimo de la función V^* en el caso con costos negativos. Sin embargo, por la naturaleza de sí mismo, el método no puede encontrar generalmente una política estacionaria óptima en este caso. El método de mejoramiento de política no es simplemente funcionado como en el caso con descuento sino tiene sus propias características del caso negativo. Las discuciones adelante manifestarán sus características. El método de mejoramiento de política nos ayuda a concluir que una política estacionaria sea óptima si y solo si cuando aquella política satisfaga estrictamente la ecuación de optimalidad. El método de programación lineal no fue considerado debido a que no se puede garantizar siempre que la adquiere el valor óptimo de V^* mediante este método. La proposición 2 que presenta a continuación declara este punto.

A. Método de Aproximación Sucesiva.

Sean $V_0(i) = 0$, y para $n > 0$

$$V_n(i) = \max_{a \in A_i} \{r(i,a) + \sum_{j=0}^{s} p_{ij}(a) \cdot V_{n-1}(j)\}, \quad \forall i \in s.$$

en donde $V_n(i)$ representa el beneficio máximo esperado para un problema de n periodos cuando se inicie en el estado i. Para demostrar que V* es la mínima solución negativa de la ecuación óptima conviene tener el resultado siguiente:

Proposición 2. Si la función $u(i) \in R_-$ es tal que

$$u(i) = \max_{a} \{r(i,a) + \sum_{j=0}^{s} p_{ij}(a)\ u(j)\} \qquad \ldots(4)$$

entonces $u(i) \leq V^*(i)$.

Prueba. Sea f la política determinada por la ecuación de optimalidad, esto es, u(i) igual a

$$r(i,f(i)) + \sum_{j=0}^{s} p_{ij}(f(i))u(j) = \max_{a \in A_i} \{r(i,a) + \sum_{j=0}^{s} p_{ij}(a)u(j)\}$$

Considere el problema usual con la edición de una decisión de paro que cuesta u(i), si se ejecuta en el estado i. Entonces la expresión anterior establece que el paro inmediato es equivalente a usar f en un período y luego parar. Repitiendo este argumento, se demuestra que el paro inmediato es equivalente a usar f en n períodos y luego paro.

Entonces,

$$E_f[\text{costo de n periodo} \mid x_0 = i] + E_f[u(x_n) \mid x_0 = i] = u(i).$$

Ya que $u \in R_-^S$, esto implica que

$$E_f[\text{costo de n periodo} \mid x_0 = i] \geq u(i),$$

y cuando $n \to \infty$, $V_f(i) \geq u(i)$. Por ser $V_f(i) \leq V^*(i)$ se implica que $u(i) \leq V^*(i)$. ▲

La proposición ha asegurado que V^* es el máximo punto fijo del mapeo T en R_-^S. El siguiente teorema demuestra la validez del método de Aproximaciones Sucesivas.

Teorema 2. Si existe un punto deficiente de $T:R_-^S \to R_-^S$ en R_-^S, entonces $T^n 0 \to V^*$ por abajo, cuando $n \to \infty$, en donde V^* es el máximo punto fijo y el punto deficiente de T en R_-^S.

Prueba. Sea v un punto deficiente de T en R_-^S. Entonces, $v \leq Tv$ y sea $v \leq u \leq 0$. Por isotonicidad de T, $v \leq Tv \leq Tu \leq T0 \leq 0$. Esto es, $T:[v,0] \to [v,0]$. Dado que T es un mapeo continuo, el resultado se obtiene por el teorema 2 en el apéndice (el teorema de punto fijo de Latiz). ▲

B. Método de Mejoramiento de Política.

Sea g una política estacionaria con función de beneficio esperado V_g. Si se define h tal que

$$T_h\, V_g(i) = \max_{a \in A_i} \{r(i,a) + \sum_{j=0}^{s} P_{ij}(a)\, V_g(j)\}$$

entonces h es al menos tan buena como g. Específicamente,

Proposición 3. Una política h que satisface la expresión anterior es tal que $V_h(i) \geq V_g(i)$, para todo $i \in S$.

Prueba: Observe que

$$E_h\,[\text{benef. total de n periodos} \mid x_0 = i] + E_h\,[V_g(x_n) \mid x_0 = i] \geq V_g(i)$$

Ya que $V_g(i) \in R_-$. Por lo tanto

$$E_h\,[\text{benef. total de n periodos} \mid x_0 = i] \geq V_g(i)$$

y cuando $n \to \infty$, se tiene $V_h(i) \geq V_g(i)$. ▲

Note que h satisface la ecuación óptima, pero por la proposición 1, $V_h \leq V^*$ y no se puede concluir que $V_h = V^*$. Por ello, es posible que el algoritmo para en una política no óptima que satisface la ecuación óptima. En el siguiente teorema se representan las condiciones suficientes para concluir una política óptima.

Teorema 3. Sea $r(i,a) \in R_-$. Si el conjunto de estados y de decisiones son finitos y si f es una política estacionaria tal que, para cada i,

$$V_f(i) > r(i,a) + \sum_{j=0}^{s} P_{ij}(a)V_f(j) \qquad \forall\, a \neq f(i) \quad \ldots (7)$$

entonces $V_f = V^*$.

Prueba: La expresión anterior implica que tomar la decisión f es mejor que hacer cualquiera otra cosa en un periodo y luego conmutar a f. Además, si se hace cualquiera otra decisión en un periodo, del siguiente periodo conmutarse a f es mejor que hacer cualquiera otra cosa en un periodo más y luego conmutarse a f. Repitiendo este argumento, se ve que tomar f es mejor que hacer cualquiera otra decisión en n periodos y luego conmutarse a f. Esto es,

$$V_f(i) > E_\pi[\sum_{t=0}^{n-1} r(x_t,A_t) \mid x_0 = i] + E_\pi[V_f(x_n \mid x_0 = i)]$$

ya que $r(i,a) \in R_-$, $V_f(i) \leq 0$, $\forall\, i \in S$. Si hacemos $n \to \infty$.

$$V_f(i) \geq V_\pi(i) + E_\pi\,[V_f(x_\infty) \mid x_0 = i]$$

Pero por $E_\pi\,[V_f(x_\infty) \mid x_0 = i] = 0$, y $V_f(i) \geq V_\pi(i)$ para toda π ya que π es arbitraria,

$$V_f(i) \geq V^*(i) \qquad \forall\, i \in S.$$

Por otro lado, de la proposición 1, se sabe que

$$V_f(i) \leq V^*(i), \qquad \forall\, i \in S.$$

y esto implica que $V_f = V^*$. ▲

El siguiente teorema afirma que el método de mejoramiento de política funciona en el caso estrictamente negativo. En particular, si el método inicia con una política estacionaria transitoria, automáticamente, se itera dentro de esa clase.

Teorema 4: Si $r(i,a) < 0$ y el conjunto de políticas transitorias no está vacio, esto es $A_t \neq 0$. Entonces

(i) V^* es el único punto fijo de T en R_-^S, y cada $a \in A$ tal que se cumple $V^* = r(a) + P(a)\,V^*$, es transitoria;

(ii) si $f^\infty \in A_t^\infty$ y $V(g,f^\infty) > 0$ para alguna $g \in A$, entonces $g^\infty \in A_t^\infty$ y $V(g^\infty) > V(f^\infty)$.

Prueba: Sea $L_f : R_-^S \to R_-^S$ tal que

$$L_f V = r(f) + P(f)V, \qquad \forall\, f \in A,\ L_f \text{ es isótono.}$$

Se define que

$$T\,V = \max_{f \in A} L_f V,$$

y

$$L_f^n\, V = V^n(f^\infty) + P^n(f)\, V.$$

Si $f^\infty \in A_t^\infty$, cuando $n \to \infty$

$$L_f^n\, V \to V(f^\infty), \qquad \forall\, V \in R^S.$$

Dado que $A_t \neq 0$, sea $f_0^\infty \in A_t^\infty$, entonces

$$V(f_0^\infty) = L_{f_0} V(f_0^\infty) \geq \max_{f \in A} L_f V(f_0^\infty) \equiv T\,V(f_0).$$

Esto es, $V(f_0^\infty)$ es un punto excesivo de T en R_-^S.

Por el teorema 2 de esta sección, existe un máximo punto fijo V de T en $R_{\underline{}}^{s}$, además, por el teorema 1 de este capítulo existe una política estacionaria $g \in A$ tal que

$$V(g^{\infty}) \equiv V^{*}.$$

Obviamente $g^{\infty} \in A_t^{\infty}$, en otro caso $V(g^{\infty}) \to -\infty$ en al menos una componente ya que $r(g) < 0$.

Suponga que V^{**} es otro punto fijo de T en $R_{\underline{}}^{s}$, entonces $V^{**} \leq V^{*}$. Sin embargo, $V^{**} = T\,V^{**} \geq L_g V^{**} \geq L_g^{n} V^{**}$, $\forall\, n \geq 1$.

Por $g^{\infty} \in A_t^{\infty}$, $n \to \infty$

$$V^{**} \geq V(g^{\infty}) \equiv V^{*}$$

esto es, $V^{**} = V^{*}$, que el único punto fijo de T en $R_{\underline{}}^{s}$. Ahora, si $V^{*} = r(f) + P(f)\,V^{*} \equiv L_f V^{*}$ para alguna $f \in A$, entonces

$$V^{*} = L_f^{n} V^{*} = V^{n}(f^{\infty}) + P^{n}(f) V^{*} \qquad \forall\, n \geq 1.$$

Esto es, $f^{\infty} \in A_t^{\infty}$. En otro caso, $V^{n}(f^{\infty}) \to -\infty$ al menos en una componente.

Si $f^{\infty} \in A_t^{\infty}$, y $v(g,f^{\infty}) = V(g,f^{\infty}) - V(f^{\infty}) \equiv L_g V(f^{\infty}) - V(f^{\infty}) > 0$, entonces,

$$L_g^{n} V(f^{\infty}) \geq L_g V(f^{\infty}) > V(f^{\infty}), \qquad \forall\, n \geq 1,$$

o bien, $V^{n}(g^{\infty}) + P^{n}(g) V(f^{\infty}) \geq V(g,f^{\infty}) > V(f^{\infty})$.

Lo mismo como antes, $g^{\infty} \in A_t^{\infty}$, $n \to \infty$: $V(g^{\infty}) \geq V(g,f^{\infty}) > V(f^{\infty})$. ▲

3.4 Ejemplos ilustrativos.

A continuación, se estudia el problema de paro óptimo con más detalle. Con respecto a las características especiales de este tipo de problemas, se proporciona una manera de clasificar aquellos estados, los cuales forman un subconjunto de estados en donde la acción de parar es, al menos mejor que la de continuar una etapa más y después parar; y se demuestra que la política óptima establece que para en un estado i si y solo si i pertenece a aquel conjunto, dado que las probabilidades de transición de i a otros estados que no estén en este conjunto son ceros. Asimismo, se demuestra que bajo las condiciones

a. $R = \sup_i \{R(i) \mid i = 1,2,\ldots\} < \infty$, y

b. $C = \sup_i \{C(i) \mid i = 1,2,\ldots\} \leq 0$,

el problema de paro óptimo puede resolverse por el método de aproximación sucesiva ya que la diferencia entre el valor $V_n(i)$ y el valor óptimo $V^*(i)$ tiene una cota superior. Otro problema que se estudia con más detalle en esta sección es el problema de reemplazo de un equipo, las condiciones de existencia de una política óptima para este tipo de problemas se describen en adelante.

Ejemplo. Considere un problema de paro óptimo donde el espacio de estados es el conjunto de los números enteros.

$$C(i) = 0$$

$$R(i) = i$$

$$P_{i,i+1} = \tfrac{1}{2} = P_{i,i-1}$$

Entonces, se observa que para cualquier estado i, $V_0(i)=i$ y para cualquier n la ecuación recursiva

$$V_n(i) = \max [R(i), C(i) + P_{i,i+1} V_{n-1}(i+1) + P_{i,i-1} V_{n-1}(i-1)]$$

implica que $V_n(i) = i$. Sin embargo el sistema se comporta de acuerdo a una cadena de una caminata aleatoria simétrica hasta antes de parar. Dicha caminata es una cadena de Markov recurrente nula, y todos los estados se comunican. Entonces, con probabilidad uno, cualquier estado N es, eventualmente visitado. Por lo tanto, si partimos del estado i se garantiza que podemos llegar al estado N y parar ahí obteniendo un beneficio final N. Sin embargo, dado que N es arbitraria se concluye que $V^*(i) = \infty$ para toda i. Por lo tanto dicho proceso no es estable.

La discusión anterior demuestra que si el proceso es estable se tiene que $\lim V_n(i) = V^*(i)$ para todo i. Sin embargo, no especifica condiciones bajo las cuales podamos determinar la política estacionaria óptima. Para obtener re

sultados en esta dirección defina el conjunto.

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

y observe que B representa el conjunto de estados en donde la acción de parar es, al menos mejor que la de continuar una etapa más y después parar. La política que para la primera vez que el proceso entra en B se denomina la política miope (pues sólo considera una etapa más).

Teorema. Suponga que el proceso de paro es estable y $P_{ij}=0$ si $i\in B$, $j\notin B$. Entonces la política óptima establece que para en el estado i si y solo si $i\in B$.

Prueba. Es inmediato que $V_0(i) = R(i)$ cuando $i\in B$. Supongamos que $V_{n-1}(i) = R(i)$ si $i\in B$. Entonces

$$V_n(i) = \max \{ R(i); C(i) + \sum_{j=0}^{s} P_{ij} V_{n-1}(i)\}$$

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

$$= \max \{ R(i), C(i) + \sum_{j\in B} P_{ij} R(j)\}$$

$$= R(i)$$

De donde $V_n(i) = R(i)$ si $i\in B$ y cualquier n. Haciendo n tender a infinito y recordando la hipótesis de estabilidad se tiene que $V^*(i) = R(i)$ para toda $i\in B$. Si $i\notin B$, podemos observar que

$$V(i) \geq C(i) + \sum_{j=0}^{s} P_{ij}R(j) > R(i)$$

donde la primera desigualdad es inmediata, y corresponde a la comparación con la política de continuar una etapa más y después parar mientras que la segunda desigualdad es obvia de la definición del conjunto B.

Proposición 4. En problemas de paro óptimo bajo las condiciones:

a. $R = \sup \{R(i) \mid i = 1,2,\ldots\} < \infty$

b. $C = \sup \{C(i) \quad i = 1,2,\ldots\} \leq 0$

se cumple, para toda n y toda i que

$$V_n(i) - V^*(i) \geq \frac{(R+C)[R-R(i)]}{(n+1)C}$$

Prueba. Sea f la política óptima y denote por T el tiempo, variable aleatoria, en que f para. Sea f_n la política que elige las mismas acciones que f para los tiempos 0, 1,...,n-1, pero que para en el tiempo n (si previamente no lo ha hecho). Si hacemos que X denote el costo total, se tiene que

$$V^*(i) = V_f(i) = E_f(X|T\leq n)P(T\leq n) + E_f(X|T>n)P(T>n)$$

y

$$V_n(i) \geq V_{f_n}(i) = E_f(X|T\leq n)P(T\leq n) + E_{f_n}(X|T>n)P(T>n)$$

donde las esperanzas condicionales consideran $x_0=i$. Entonces

$$V_n(i) - V^*(i) \geq [E_{f_n}(X|T>n) - E_f(X|T>n)]\ P(T>n)$$

$$\geq -[R + C]\ (P(T>n)$$

Sin embargo, se observa de las expresiones anteriores que

$$R(i) \leq V^*(i) \leq RP(T \leq n) + [R + (n+1)C]\ P(T > n)$$

$$= R + (n+1)\ C\ P\ (T > n)$$

De donde se obtiene la cota superior para P(T>n) y la prueba termina.

Ejemplo . Considere el problema de venta de un inmueble. Suponga que se permite aceptar cualquier oferta pasada. En ese caso, el estado en cualquier periodo, es la máxima oferta recibida hasta ese periodo. Las probabilidades de transición P_{ij} de un estado a otro son:

$$P_{ij} = \begin{cases} 0 & \text{si } j < i \\ \sum_{k=0}^{i} P_k & \text{si } j = i \\ P_j & \text{si } j > i. \end{cases}$$

El conjunto que caracteriza los estados de paro asociados a la política miope en un periodo está definido como

$$B = \{ i \mid i \leq \sum_{j=i+1}^{S} j\, P_j - C + i \sum_{j=0}^{i} P_j \}$$

$$= \{ i \mid C \leq \sum_{j=i+1}^{S} (j-i)\, P_j \}$$

$$= \{ i \mid C \leq E\,[(x - i)^+] \}$$

donde x es una variable aleatoria que representa la oferta en un día dado. Debido a que $(x-i)^+$ decrece en i, es sencillo demostrar que B es conjunto cerrado. Suponiendo que existe estabilidad del proceso se puede demostrar cuando $E(x^2) < \infty$, la política óptima acepta la primera oferta que es al menos i*, donde

$$i^* = \max \{ i \mid -C \leq \sum_{j=i+1}^{S} (j-i)\ p_j \}.$$

Como la política óptima nunca retira una oferta pasada, la óptima también es una política legítima para el problema original que no permite retirar ninguna oferta pasada. Por lo tanto, la política anterior debe ser óptima para el problema original.

Ejemplo 4: (Reemplazo de un equipo). Considere una máquina que puede estar en cualquier uno de los estados 0,1,2,... . Se supone que al inicio de cada día, se observa el estado de la máquina y se tome una de las dos decisiones: reemplazar o no la máquina. Una vez de reemplazar la máquina, se supone que la máquina se reemplaza por una nueva cuyo estado es 0. El costo de reemplazo es R y el de mantenimiento por día es C(i), dependiendo del estado. Sea P_{ij} la probabilidad de que una máquina esté en el estado i al inicio del día y pasa al estado j al inicio del siguiente día. Las siguientes suposiciones sobre los costos y las probabilidades de transición se cumplen:

(i) $\{C(i), \forall\, i \in S\}$ es una sucesión acotada y creciente;

(ii) $\sum_{j=k}^{S} p_{ij}$ es una función creciente de i, $\forall\, k \geq 0$.

La suposición (i) es obvia pues el costo de mantenimiento es una función creciente de estado. La segunda suposición indica que la probabilidad de una transición dentro de cualquier bloque de estados {k, k+1,...,} es una función creciente del estado presente.

Se desea determinar la política óptima de reemplazo y mantenimiento de equipo de tal manera se maximiza el beneficio total esperado.

Este problema es uno de cadena de decisión markoviana de dos acciones. Sean

$$C(i,f_1) = R; \qquad C(i,f_2) = C(i), \qquad \forall\, i \in S;$$

$$P_{ij}(f_1) = P_{0j}; \qquad P_{ij}(f_2) = P_{ij}, \qquad \forall\, i \in S.$$

En donde R y $C(i) \in R_-$. Las f_1 y f_2 representan las acciones de reemplazo y mantenimiento, respectivamente. Entonces el modelo matemático se puede presentar como sigue:

$$V_1(i) = \max \begin{cases} R + C(0) \\ \\ C(i) \end{cases}$$

que representa el beneficio esperado para el primer día cuando se inicie en el estado i y para n > 1

$$V_n(i) = \max \begin{cases} R + \sum_{j=0}^{S} P_{0j} V_{n-1}(j) \\ \\ C(i) + \sum_{j=0}^{S} P_{ij} V_{n-1}(j) \end{cases}$$

que representa el beneficio esperado para n periodos cuando se inicie en el estado i.

Para determinar la estructura de la política óptima de tal problema se necesitan los siguientes dos lemas necesarios.

Lema 7.1: La suposición (ii) del problema (ejemplo 4) implica que para cualquier función decreciente g(i), la función

$$\sum_{j=0}^{S} P_{ij}\, g(j)$$

es decreciente en i.

Prueba: Por la suposición (ii)

$$\sum_{j=k}^{S} P_{i+1,j} \geq \sum_{j=k}^{S} P_{i,j} \qquad \forall\, k \geq 0$$

esto implica que

$$P_{i+1,j} \geq P_{i,j}. \qquad \forall\, i,j$$

ya que g(i) es decreciente en i, entonces es obvio que:

$P_{i+1,s}\, g(s) \leq P_{i,s}\, g(s)$

sea $$\sum_{j=k}^{S} P_{i+1,j}\, g(j) = \sum_{j=k}^{S} P_{i,j}\, g(j), \qquad j \geq 1$$

entonces

$$\sum_{\ell=0}^{k-1} P_{i+1,}\, g(\ell) + \sum_{j=k}^{S} P_{i+1,j}\, g(j) = \sum_{j=0}^{S} P_{i+1,j} g(j) \leq \sum_{j=0}^{S} P_{ij}\, g(j)$$

Esto es, $\sum_{j=0}^{S} P_{ij}\, g(j)$ es una función decreciente en i.

Lema 7.2: Bajo suposiciones (i) e (ii), V*(i) es función decreciente en i.

Prueba: Sean

$$V_1(i) = \max \begin{cases} R + C(0) \\ \\ C(i) \end{cases} \qquad \forall\, i \in S,$$

y para $n > 1$

$$V_n(i) = \max \begin{cases} R + C(0) + \sum_{j=0}^{S} P_{0j}\, V_{n-1}(j) \\ \\ C(i) + \sum_{j=0}^{S} P_{ij}\, V_{n-1}(j) \end{cases}$$

por la suposición (i), $V_1(i)$ es decreciente en i. Suponga que $V_{n-1}(i)$ es decreciente en i, para todo i. Entonces, por el lema 3.1, $V_n(i)$ también es decreciente en i, para todo i. Como

$$V^*(i) = \lim_n V_n(i)$$

$V^*(i)$ es decreciente en i. ▲

Entonces, la estructura de la política óptima está dada por el siguiente teorema.

Teorema 3.5. Bajo las dos suposiciones, existe un entero i^*, $i^* \leq \infty$, tal que la política óptima es reemplazar equipo si $i > i^*$, o no reemplazar el equipo si $i \leq i^*$.

Prueba: por el teorema 3.1.

$$V^*(i) = \max \begin{cases} R + C(0) + \sum_{j=0}^{S} P_{0j}\, V^*(j) \\ \\ C(i) + \sum_{j=0}^{S} P_{ij}\, V^*(j) \end{cases} \qquad \forall\, i \in S, \ldots, (\gamma)$$

sea

$$i^* = \max \{ i : C(i) + \sum_{j=0}^{S} P_{ij}\, V^*(j) \leq C(0) + R + \sum_{j=0}^{S} P_{0j}\, V^*(j) \},$$

Por los dos lemas anteriores, $C(i) + \sum_{j=0}^{S} P_{ij}\, V^*(j)$ es decreciente en i, y por (γ):

$$V^*(i) = \begin{cases} C(i) + \sum_{j=0}^{S} P_{ij}\, V^*(j), & i \leq i^* \\ \\ R + C(0) + \sum_{j=0}^{S} P_{0j}\, V^*(j) & i > i^*. \end{cases}$$

CAPITULO IV

CADENAS DE DECISION MARKOVIANA CON COSTOS POSITIVOS

Si se consideran sólo aquellos problemas de decisión markoviana con costos positivos llamado beneficios, los métodos desarrollados de solución basados en la técnica de programación dinámica configuran una rama importante que es denominada Programación Dinámica Positiva. La programación dinámica positiva tiene notable aplicación en el área de la teoría de apuestas. También es aplicable para resolver muchos problemas especiales reales tales como problema de inversión, selección de localidad, etc.

En el presente capítulo, se presenta la estructura general de problemas con costo positivo con algunos problemas simplificados. A continuación, se caracteriza la forma más general de las políticas estacionarias óptimas y sus correspondientes valores óptimos. Asimismo se analizan y discuten los métodos disponibles de solución y su implantación en los problemas que se consideran.

4.1 Descripción de problemas de costos positivos.

Una vez más, se suponen que el conjunto de las decisiones $a \in A$ es finito y el conjunto de los estados S es finito o infinito contable. También se supone que $r(i,a) \in R_+$ es el beneficio obtenido por tomar la decisión a cuando esté en el estado i. Entonces, para cualesquiera política, se define el término

$$V_\pi(i) = E_\pi \left[\sum_{t=0} r(X_t , A_t) \mid X_o = i \right] \qquad i \in S$$

esto es el beneficio total esperado del proceso bajo la política π cuando se inicie en el estado i. Como $r(i,a) \in R_+$, para toda $i \in S$, entonces $V_\pi(i)$ está bien definido (y podría ser $+\infty$). Sea

$$V^*(i) = \sup_\pi V_\pi(i)$$

Se dice que π^* es una política óptima si

$$V_{\pi^*}(i) = V^*(i), \qquad \text{para toda} \quad i \in S .$$

A continuación, se consideran un par de problemas interesantes con costos positivos. Y posteriormente se analiza, las características de la política óptima y los métodos de solución.

Ejemplo 1. Considere un individuo que posea i monedas y entra a un casino de juegos. El juego permite cualesquiera apuesta de la siguiente forma: si se posee i monedas entonces puede apostarse cualesquiera cantidad entera positiva menor que ó igual a i. Si se apuesta j monedas entonces: (a) se ganaría con una probabilidad p o (b) se perdería con una probabilidad 1-p. Se desea saber cuál política de apuesta que maximiza la probabilidad de que el individuo obtendrá un fortuna de N monedas.

Se puede formular el problema como sigue:
Sea el conjunto $S = \{0,1,2,\ldots,N\}$ de estados, donde se dice que está en el estado i cuando la fortuna presente es i.
Se define

$$r(i,a) = 0, \quad i \neq N \text{ y toda a}$$

$$r(N,a) = 1; \quad P_{N0}(a) = P_{00}(a) = 1, \quad \forall\, a.$$

esto es, se gana 1 si y solo si su fortuna alguna vez logra N. El modelo matemático se presenta como sigue: Sea $V_0(i) = 0$; $i \in S$, el beneficio esperado del periodo 0 cuando se inicie en el estado i; y para $n > 0$.

$$V_n(i) = \max \{r(i,a) + \sum_{j=1}^{S} P_{ij}(a)\, V_{n-1}(j)\}$$

que es el beneficio total esperado del problema cuando se inicie en el estado i y en el n-ésimo periodo.

Ejemplo 2. Considere un conductor de taxi que trabaja en un área circunda por tres zonas. Si el conductor está en la zona 1, puede hacer una de las siguientes tres acciones:

(i) pasando las calles, tal vez encuentre un pasajero por casualidad;

(ii) esperar en la caja de sitio más cercana;

(iii) esperar la llamada de control por radio, en el lado de la acera.

En la zona 3, se puede escoger una de las mismas tres acciones. Pero en la zona 2, como no existe el servicio de radio en esa zona, la última acción no es disponible. Para cualesquiera zona dada i y una acción dada a asociada con la zona, existen una probabilidad P_{ij} de ir a cada una de las zonas 1, 2 y 3 para el siguiente viaje, y un beneficio correspondiente conocido r(i,a) asociado con el viaje. Se desea determinar una política de selección de acciones para los siguientes viajes de tal manera que el beneficio total esperado sea máximo.

Se puede formular el problema como sigue: Sean a_i, i = 1, 2, 3 denotada las acciones i, ii, iii, respectivamente, y los estados 1, 2 y 3 corresponden a las zonas 1, 2 y 3 respectivamente. El modelo matemático es presentado como sigue:

$$V_0(i) = 0, \qquad \forall\, i \in S$$

que es el beneficio total esperado en el periodo 0 cuando se inicie en el estado i. Y para $n > 0$

$$V_n(i) = \max_a \{r(i,a) + \sum_{j=1}^{S} p_{ij}(a)\, V_{n-1}\} \qquad \forall\, i \in S.$$

que representa el beneficio total esperado en el n-ésimo periodo cuando se inicie en el estado i.

4.2 Caracterización de soluciones óptimas.

En esta sección se analizan las propiedades de la función de beneficios esperados óptimos de una cadena de decisión markoviana con costos positivos. De manera semejante al caso analizado en la sección 3.2, se cumple que dicha función, denotada por V^*, satisface la ecuación de optimalidad de la programación dinámica.

Propsición 1. La función V^* satisface

$$V^*(i) = \max_a \{r(i,a) + \sum_{j=0}^{S} P_{ij}(a)\ V^*(j)\}$$

Prueba. Es similar a la efectuada en el caso con descuento del capítulo 2, y no se reproduce.

El resultado principal de la programación dinámica positiva se presenta por el siguiente teorema. En el cual la condición suficiente de existir una política estacionaria óptima es establecida en bases del mapeo de contracción y la teoría de latiz.

Teorema 1. Si existe un punto excesivo de $T:R_+^S \to R_+^S$ en R_+^S, esto es, existe un $v \in R_+^S$ que $v \geq Tv \in R_+^S$, donde el mapeo T está definido como sigue:

$$Tv = \max_a \{r(a) + P(a)\, v\}.$$

entonces,

(1) existe una política estacionaria f^∞ que maximiza $V(\cdot)$;

(2) $V(f^\infty) = V^*$ si y solo si $V^* = r(f) + P(f)\, V^*$.

Prueba. Observe que

$$T^N 0 = \max \{V^N(\pi) \mid \pi \in A^\infty\} \geq V^N(\pi) \qquad \forall\, \pi$$

haciendo $N \to \infty$ se tiene que

$$V^* \geq \overline{\lim_{N\to\infty}} V^N(\pi) = V(\pi).$$

Nótese que V^* es el mínimo punto excesivo y fijo de T en R_+^S, entonces se define $T_f : L_E \to L_E$ y se observa que

$$V^* \geq T_f V^* \geq T_f^N V^*$$

por V^* es un punto excesivo y T_f es isótono. Como L_E es compacta (ver teorema 2 de apéndice B).

$$T_f^N V^* \geq V^*$$

Cuando $N \to \infty$ $V(f^\infty) = \overline{\lim_{N\to\infty}} T_f^N V^* = V^*$, esto es, f^∞ existe y es óptima. Si $V(f^\infty) = V^*$, se puede hacer $T_f V = r(f) + P(f)V$.

Ya que $T_f^N\, 0$ converge a $V(f^\infty)$ que es exactamente V^* por suposición. De otro lado, debido a que T_f es continuo,

$$V^* = \lim_{N\to\infty} T_f(T^N 0) = T_f\, V^* = r(f) + P(f)\, V^*.$$

Esto se demuestra el inciso (2). ▲

4.3 Métodos de solución.

Los métodos disponibles para encontrar la política estacionaria óptima y el valor óptimo se descubren en esta sección.

A. Método de Aproximación Sucesivas. Considere el proceso

$$V_0(i) = 0 \qquad \text{para todo} \quad i \in S$$

y para $n > 0$

$$V_n(i) = \max_a \{r(i,a) + \sum_{j=1}^{S} P_{ij}(a)\, V_{n-1}(j)\}, \qquad \forall \ i \in S.$$

En donde $V_n(i)$ representa el máximo beneficio esperado para un problema de n periodos cuando se inicie en el estado i. El siguiente teorema demuestra la validéz del método de aproximación sucesiva.

Teorema 2. En el caso positivo, si existe un punto excesivo de T en R_+^S donde T es un mapeo definido como

$$Tv = \max_a \{r(a) + P(a)\ v\}.$$

Entonces, $T^N 0 \to V^*$ por abajo cuando $N \to \infty$. En donde, V^* es el mínimo punto fijo y excesivo de T en R_+^S.

Prueba. Ver apéndice sobre latices.

Nótese que este método obtiene sólo el valor óptimo del problema pero no la política óptima.

B. Método de Programación lineal. Sea $u(i) \in R_+$, para todo $i \in S$ tal que

$$u(i) \geq \max_a \{r(i,a) + \sum_{j=0}^{S} p_{ij}(a)\, u(j)\} \ldots\ldots (B).$$

Proposición 2. Una función no negativa u que se satisface (B) es tal que

$$u(i) \geq V^*(i), \qquad \forall\, i \in S.$$

donde $V^*(i)$ es el valor óptimo del problema con costos positivos.

Prueba: Si define que T es un mapeo tal que

$$Tv = \max_a \{r(a) + P(a)\, v\}$$

entonces la expresión (B) nos dice que u es un punto excesivo del mapeo T. Por los incisos c) y d) del teorema 2 en el apéndice B, V^* es el mínimo punto fijo y excesivo de T en R_+^S. Por lo tanto, se cumple que $u(i) \geq V^*(i)$ para todo $i \in S$.

Suponga que $V^*(i) < \infty$, para todo $i \in S$. Entonces, debido a que V^* satisface la expresión (B) (más bien con igualdad), podemos usar la proposición 2, se puede obtener V^* como una solución de un problema de programación lineal.

Se pueden representar los problemas primal y dual en las siguientes formas:

Primal	Dual
$\max \sum_{f \in A} x(f)\, r(f)$	$\min\ q\, V$
s.a.	s.a.
$\sum_{f \in A} x(f)[I - P(f)] = q$	$[I - P(f)]V \geq r(f)$
$x(f) \geq 0, \quad f \in A$	$f \in A$

Donde x(f) es una solución básica factible del problema primal, si y solo si, para alguna f, $x(f)\,[I - P(f)] = q$ y $x(g) = 0$, para toda $g \neq f$. Pero existe una solución x(f) no negativa si y solo si $f^{\infty} \in A_T^{\infty}$.

Por ser V* el mínimo punto fijo y excesivo de T en R_+^S, V* es óptimo para el problema dual. Entonces, existe una solución básica factible óptima correspondiente para el problema primal. Observe que el problema dual es idéntico al problema

$$\min q\, u$$

s.a.

$$u - P(a)\, u \geq r(a),$$

$$u \geq 0$$

Este problema se puede resolver por el método estándar de programación lineal (SIMPLEX, etc).

C. Discución del método de Mejoramiento de Políticas.

Sea g una política estacionaria con función de beneficio esperado V_g. Si se define la política h tal que

$$T_h V_g(i) = \max \{r(i,a) + \sum_{j=\varepsilon}^{S} P_{ij}(a)\, V_g(i)\} \qquad \forall\, i \in S$$

entonces, después de n periodos se tiene:

$$E_h \text{ [beneficio total en n periodos } \mid x_0 = i]$$

$$+ E_h [V_g(x_n) \mid x_0 = i] \geq V_g(i)$$

Debido a que $E_h[V_g(x_n) \mid x_0 = i] \geq 0$, no se puede concluir que

$$E_h \text{ (beneficio total de n periodos } \mid x_0 = i] \geq V_g(i).$$

Esto es, en el caso positivo, es muy probable que no exista una política estacionaria óptima. Entonces, no puede asegurarse que por el método de mejoramiento de política se pueda encontrar una política estacionaria óptima. Sin embargo, se puede demostrar que si la función de beneficio satisface la ecuación de optimalidad para una política dada, entonces esta política es óptima.

Teorema 3. Sea V_f la función de beneficio esperado para la política estacionaria f. Si

$$V_f(i) = \max_a \{r(i,a) + \sum_{j=0}^{S} p_{ij}(a)\, V_{f(j)}\}, \qquad i \in S,$$

Entonces

$$V_f(i) = V^*(i), \qquad i \in S.$$

Prueba: La hipótesis del teorema establece que para toda a

$$V_f(i) \geq r(i,a) + \sum_{J}^{S} p_{ij}(a)\, V_f(j),$$

La ecuación implica que tomar f es mejor que hacer cualquiera otra cosa en el primer periodo y luego conmutar f. Además, si se hace cualquiera otra cosa en el primer periodo, del siguiente periodo conmutarse a f es mejor que hacer cualquiera otra cosa en un periodo más y luego conmutarse a f. Repitiendo este argumento, se ve que tomar f es mejor que hacer cualquiera otra cosa en n periodos y luego conmutarse a f. Esto es,

$$V_f(i) = E_\pi[\sum_{t=0}^{n-1} r(x_t, a_t) \mid x_0 = i] + E_\pi[V_f(x_n)].$$

Como $r(i,a) \in R_+$, se obtiene

$$V_f(i) \geq E_\pi[\sum_{t=0}^{n-1} (x_t, a_t) \mid x_0 = i].$$

pues $E_\pi[V_f(x_n)] \geq 0$. Por lo tanto $V_f(i) \geq V_\pi(i)$. Por otro lado $V_f(i) \leq V^*(i) = \sup_\pi V_\pi(i)$, y se puede concluir que f es la política estacionaria óptima.

4.4 Aplicaciones a teoría de apuestas.

Considere un individuo cuyo capital es i unidades monetarias y participa las apuestas de un casino. El casino permite apostar de la siguiente manera: si se dispone de un capital i se puede apostar cualquier cantidad $j \leq i$ y se ganan j unidades monetarias con probabilidad p o se pierde j unidades monetarias con probabilidad 1-p. Se desea establecer la estrategia que maximice la probabilidad que el individuo obtenga una fortuna N antes de perder el juego. La formulación de este problema en términos de la programación dinámica positiva demuestra que si $p \geq \frac{1}{2}$ (que equivale a tener una apuesta favorable) la estrategia óptima es apostar siempre una unidad monetaria hasta que el capital sea N o bien cero. Dicha estrategia se denomina la estrategia tímida. Por otra parte, si $p < \frac{1}{2}$ (que equivale a tener apuestas desfavorables) entonces la estrategia óptima es apostar todo el capital en cada jugada hasta tener un capital N o bien perder. Esta estrategia se denomina la estrategia agresiva o total.

Con el propósito de establecer este problema eh el marco de la programación dinámica positiva considere el conjunto de estados $S = \{0,1,2,\ldots,N\}$ y diremos que estamos en el estado i si el capital disponible es i. Defina la estructura de beneficios asociados con cada decisión como

$$r(i,a) = 0 \quad \text{si } i \neq N \text{ y } a \text{ arbitraria}$$

$$r(N,a) = 1$$

y probabilidades $P_{N0}(a) = P_{00}(a) = 1$. Equivalentemente un beneficio igual a uno se obtiene si y solo si el capital es N o más y el beneficio total esperado es simplemente la probabilidad que el capital sea N.

Demostraremos la optimalidad de la política descrita inicialmente aplicando el teorema 5 que equivale a proponer una política estacionaria tal que cumpla los postulados de dicho teorema. Específicamente, observe que si nuestro capital es i no conviene aportar más de N-i y podemos limitar las apuestas a 1,2,..., min (i, N-i). Sea V(i) el beneficio esperado asociado con una política estacionaria de apuestas. Si V(i) satisface

$$V(i) \geq p\, V(i+k) + (1-p)\, V(i-k)$$

para toda $0 < i < N$ y $1 \leq k \leq \min(i, N-i)$, sabemos (Teorema 2) que dicha política es óptima.

Considere la estrategia tímida, esto es, la estrategia que apuesta una unidad monetaria en cada jugada. Bajo esta estrategia el juego de apuestas se convierte en un problema de caminata aleatoria con fronteras finitas o simplemente, el problema clásico del jugador.

Si V(i) denota la probabilidad de llegar a N antes que a cero cuando se emplieza con un capital i < N sabemos (Ejemplo 3 capítulo 1) que

$$V(i) = \begin{cases} \dfrac{1-(q/p)^{i}}{1-(q/p)^{N}} & p \neq \frac{1}{2} \\ \\ \dfrac{i}{N} & p = \frac{1}{2} \end{cases}$$

donde q = 1 - p. La conclusión del resultado propuesto en esta sección se tiene en las proposiciones siguientes.

Proposición 3. Si $p \geq \frac{1}{2}$ la estrategia tímida maximiza la probabilidad de lograr una fortuna de N unidades.

Prueba. Si $p = \frac{1}{2}$ entonces $V(i) = i/N$ satisface

$$V(i) = p\,V(i+k) + q\,V(i-k)$$

para toda $0 < i < N$ y $k = \min\{i, N-i\}$. Por lo tanto la política tímida es óptima (Teorema 2). De manera semejante, si $p > \frac{1}{2}$ debemos demostrar que

$$\frac{1 - (q/p)^i}{1 - (q/p)^N} \geq p\left[\frac{1 - (q/p)^{i+k}}{1 - (q/p)^N}\right] + q\left[\frac{1 - (q/p)^{i-k}}{1 - (q/p)^N}\right]$$

o bien

$$(q/p)^i \leq p(q/p)^{i+k} + q(q/p)^{i-k}$$

Sin embargo, esto también equivale a que

$$1 \leq p[(q/p)^k + (p/q)^{k-1}]$$

que es cierto si $k = 1$. Sea $f(k)$ la expresión entre paréntesis de la desigualdad anterior y note que $f(k)$ es creciente pues

$$f'(k) = (q/p)^k \log(q/p) + (p/q)^{k-1} \log(p/q)$$

$$= [(p/q)^{k-1} - (q/p)^k] \log(p/q) \geq 0$$

es decir si $p > \frac{1}{2}$ estrategia tímida también es óptima.

Proposición 4. Si $p \leq \frac{1}{2}$ la estrategia total maximiza la probabilidad de lograr una fortuna N en un tiempo $n > 0$.

Prueba. Debido al teorema 2 es suficiente demostrar que

$$V_{n+1}(r) \geq p\, V_n(r+s) + q\, V_n(r-s)$$

o bien

$$V_{n+1}(r) - p\, V_n(r+s) - q\, V_n(r-s) \geq 0$$

para todo $0 < s \leq \min\{r, N-r\}$. Aplicaremos un proceso de inducción para obtener el resultado. Observe que si $n = 0$ el resultado es inmediato. Supongamos entonces que

$$V_n(i) - p\, V_{n-1}(i+k) - q\, V_{n-1}(i-k) \geq 0$$

para todo $i = 0,1,\ldots,N-1$ y $0 < k \leq \min\{i, N-i\}$. Para el caso $n + 1$ cuya desigualdad se muestra al principio consideraremos cuatro casos: a) $r+s \leq N/2$; b) $r - s \geq N/2$; c) $r \leq N/2 \leq r + s$; y d) $r - s \leq N/2 \leq r$. En el caso a se observa que

$$\alpha = V_{n+1}(r) - p\, V_n(r+s) - q\, V_n(r-s)$$

$$= p[V_n(2r) - p\, V_{n-1}(2r+2s) - q\, V_{n-1}(2r-rs)] \geq 0$$

debido a la hipótesis de inducción con $i = 2r$ y $k = 2s$.

Caso b. Es semejante al a pues

$$\alpha = V_{n+1}(r) - p\,V_n(r+s) - q\,V_n(r-s) =$$

$$= p+q\,V_n(2r-N) - p[p+qV_{n-1}(2r+2s-N)] - q[p+qV_{n-1}(2r-2s-N)]$$

$$= q\,[V_n(2r-N) - pV_{n-1}(2r-N+2s) - q\,V_{n-1}(2r-N-2r)] \geq 0$$

debido a la hipótesis de inducción con $i = 2r-N$ y $k = 2s$.

Caso c. Observe que $r - s \leq r \leq N/2 \leq r + s$ y que

$$\alpha = V_{n+1}(r) - p\,V_n(r+s) - q\,V_n(r-s)$$

$$= p[V_n(2r) - p - qV_{n-1}(2r+2s-N) - q\,V_{n-1}(2r-2s)].$$

Sin embargo, $2r \geq r + s \geq N/2$ y $V_n(2r)$ puede reemplazarse

$$\alpha = p[p+qV_{n-1}(4r-N) - p - qV_{n-1}(2r+2s-N) - q\,V_{n-1}(2r-2s)]$$

$$= q[pV_{n-1}(4r-N) - pV_{n-1}(2r+2s-N) - p\,V_{n-1}(2r-2s)]$$

$$= q[V_n(2r-N/2) - p\,V_{n-1}(2r+2s-N) - p\,V_{n-1}(2r-2s)]$$

donde la última igualdad se sigue de que $0 \leq 2r - N/2 \leq N/2$. Analizaremos dos alternativas sobre s. Supongamos que $s \geq N/4$. Dado que $p \leq q$ se tiene que α es mayor o igual a

$$q[V_n(2r-N/2) - pV_{n-1}(2r+2s-N) - qV_{n-1}(2r-2s)] \geq 0,$$

debido a la hipótesis de inducción con $i = 2r - N/2$ y $k = 2s - N/2$.

Por otra parte si $s < N/4$ se tiene que α es mayor o igual a

$$q[V_n(2r-N/2) + q\, V_{n-1}(2r+2s-N) - p\, V_n(2r-2s)] \geq 0$$

por la misma hipótesis con $i = 2r - N/2$ y $k = N/2 - 2s$. El último caso: $r - s \leq N/2 \leq r \leq r + s$ es semejante y no se reproduce.

Corolario. Con tiempo ilimitado; la estrategia total maximiza la probabilidad de llegar a N cuando $p \leq \frac{1}{2}$.

Prueba. Sea $U(r)$ la probabilidad de llegar a N antes que a 0 si empezamos con un capital r y usamos la política total. Dado que $U(r) = \lim_n U_n(r)$ se tiene que

$$U(r) \geq p\, U(r+s) + q\, U(r-s) \qquad s \leq \min(r, N, r)$$

debido a la proposición anterior. El teorema 2 termina la prueba.

CAPITULO V

CONCLUSIONES.

En este trabajo se han estudiado los problemas dinámicos que evolucionan probabilísticamente y conforman las propiedades de cadenas markovianas bajo una estructura de costos. El correspondiente análisis de existencia y caracterización de políticas óptimas se ha sido realizado en bases del mapeo de contracción y la teoría de latíz. Los métodos disponibles de soluciones tales como métodos de aproximación sucesiva, mejoramiento de políticas y programación lineal fueron considerados. Y las condiciones bajo las cuales dichos métodos son aplicables fueron indicados para cada caso de problemas dinámicos mencionados. En resumen los tres métodos son aplicables en el caso con descuento, el método de programación lineal no es aplicable en el caso con costos negativos; y el método de mejoramiento de políticas no es aplicable en el caso con costos positivos. Los ejemplos ilustrativos han sido anexado permitiendo la unificación de los análisis correspondientes.

APENDICE.- Matrices especiales y conceptos básicos de Latices.

En el análisis de modelos probabilísticos y económicos se utiliza con frecuencia cierto tipo de matrices - matrices no-negativas, matrices cuya inversa es no-negativa y funciones de matrices expresadas como series. Debido a la necesidad de manipular de manera conveniente estas matrices se requiere estudiar sus propiedades y caracterizaciones equivalentes. Por ello, en la primera parte del apéndice, el análisis se concentra en las propiedades de interés de matrices. Como usual, se expresan las matrices en su forma Jordan y se analizan sus propiedades en términos de los valores característicos.

En el análisis de modelos probabilísticos y económicos, la aplicación de la teoría de latíz también ocupa un lugar muy importante. Debido a que la necesidad de caracterizar en una manera concisa las soluciones de problemas de programación dinámica en casos especiales, se introducen los conceptos generales de latices en la segunda parte del apéndice. Asimismo, las propiedades básicas y los resultados principales son desarrollados.

A. Matrices especiales.

Sea A una matriz compleja de orden nxn. El conjunto de valores característicos asociados con A se denota por $\sigma(A)$ y se dice que el radio espectral de A es $|\sigma(A)|$, esto es el máximo valor absoluto de los elementos de $\sigma(A)$. Asimismo diremos que J es una matriz Jordan de orden sxs con valor característico λ si

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}$$

o bien $J = \lambda I + E$ donde I es la matriz identidad de orden sxs; y E es una matriz con elementos igual a uno arriba de la diagonal principal y cero en otras posiciones.

La propiedad más importante que nos interesa, de la matriz Jordan se presenta mediante el lema 1 en adelante. Luego, se presentan las propiedades importantes de matrices generales, las cuales incluyen matrices cuadradas complejas, cuadradas no-negativas y positivas en una manera progresiva.

Lema 1. Sea J matriz de Jordan asociada al valor característico λ. Entonces $|\lambda| < 1$ si y solo si $0 = \lim_n J^n$.

Prueba. La matriz J de orden sxs puede expresarse como $J = \lambda I + E$ donde la matriz E satisface $E^s = 0$. Asimismo,

$$J^n = \sum_{k=0}^{n} \binom{n}{k} \lambda^{n-k} E^k$$

sin embargo, si $n \geq s$ la fórmula anterior se reduce a

$$J^n = \sum_{k=0}^{s-1} \binom{n}{k} \lambda^{n-k} E^k$$

Si fijamos el índice k ($0 \leq k \leq s-1$) y comparamos los coeficientes de E^k en las matrices sucesivas J^n y J^{n+1} se tiene que el cociente de estos coeficientes es igual a:

$$q_n = \frac{\binom{n+1}{k} \lambda^{n+1-k}}{\binom{n}{k} \lambda^{n-k}} = \frac{n+1}{n+1-k} \lambda$$

que implica $|q_n| < 1$ si y solo si $|\lambda| < 1$ y n es suficiente grande. De esta relación podemos concluir que cada uno de los coeficientes de E^k converge a cero cuando n tiende infinito si y solo si $|\lambda| < 1$. Equivalentemente $0 = \lim_n J^n$ si y solo si $|\lambda| < 1$.

Teorema 1. Suponga que B es una matriz cuadrada compleja. Entonces, los postulados siguientes son equivalentes:

a. $|\sigma(B)| < 1$

b. $\lim_{N} B^N = 0$

c. $||B^N|| < 1$ para algún $N \geq 1$

d. $\sum_{N=0}^{\infty} B^N$ es absolutamente convergente

Prueba. El teorema de Jordan implica que podemos escribir $B = QJQ^{-1}$ donde Q es una matriz no-singular y J es la forma

$$J = \begin{bmatrix} J_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & J_n \end{bmatrix}$$

donde cada J_i es una matriz de Jordan, esto es, podemos expresar J_i como $J_i = \lambda_i I + E$ donde λ_i es un elemento de $\sigma(B)$ y cada una de las matrices descritas es de orden $n_i x n_i$ (i=1, ...,n). Puesto que $B^N = QJ^NQ^{-1}$ para toda N se observa que

$$\begin{aligned} ||B^N|| &= ||QJ^NQ^{-1}|| \\ &\leq ||Q|| \cdot ||J^N|| \cdot ||Q^{-1}|| \\ &= ||Q|| \cdot ||Q^{-1}|| \cdot ||J^N|| \end{aligned}$$

De aquí se concluye que

$$\lim_N B^N = \lim_N J^N$$

pues también $J^N = Q^{-1}B^NQ$. Por otra parte, $0 = \lim_N J^N$ sí y solo si $0 = \lim_N J_i^N$ para toda i=1,...,n. Sin embargo, debido al lema 1 esto es cierto sí y solo si $|\lambda_i| < 1$ para todo $i = 1,...,n$ ó equivalentemente $|\sigma(B)| < 1$. Por lo tanto $0 = \lim B^N$ sí y solo si $|\sigma(B)| < 1$.

La prueba de b implica c es inmediata. Ahora bien, por definición $\sum_{k=0}^{\infty} B^k$ es absolutamente convergente sí y solo si

$$\sum_{k=0}^{\infty} ||B^k|| < \infty$$

Sin embargo, podemos escribir

$$\sum_{k=0}^{\infty} ||B^k|| = \sum_{j=0}^{\infty} \sum_{i=0}^{n-1} ||B^{jN+i}||$$

$$\leq \sum_{j=0}^{\infty} \sum_{i=0}^{n-1} ||B^N||^j \, ||B^i||$$

$$= \sum_{i=0}^{n-1} ||B^i|| \sum_{j=0}^{\infty} ||B^N||^j < \infty$$

si $||B^N|| < 1$ para algún $N > 1$. Esto prueba que c implica d. Finalmente, se observa que d implica b pues la convergencia absoluta implica que $||B^N||$ converge a cero cuando N tiende a infinito. Equivalente a $0 = \lim B^N$ y la prueba termina.

Teorema 2. Sea B una matriz cuadrada y no-negativa. Entonces, cualquier postulado del teorema 1 es equivalente a cada uno de los siguientes:

e. $(I-B)^{-1}$ existe y es no-negativa

f. Existe un vector $x \geq 0$ tal que $(I-B)x > 0$

Prueba. Usando la identidad

$$I-B^{N+1} = (I-B)\left(\sum_{i=0}^{N} B^i\right) = \left(\sum_{i=0}^{N} B^i\right)(I-B)$$

y la condición b se implica que la inversa de I-B existe y es dada por la serie $\sum_{i=0}^{\infty} B^i$. Dicha inversa es no-negativa, pues la suma de matrices no-negativas. Esto demuestra que b implica e. Por otra parte e implica f, pues la solución de la ecuación $(I-B)x = U = [i,\ldots,1]^t$ es

$$x = [\ I-B\ ]^{-1}U$$

y $x \geq 0$ debido a la no-negatividad de $[I-B]^{-1}$. El caso f implica b es como sigue: La condición $[I-B]\ x > 0$ donde $x \geq 0$ puede escribirse como $x > Bx$ e implica que $x > 0$. Asimismo, existe un escalar $0 < \lambda < 1$ tal que $\lambda x > Bx$. sin embargo esto implica que $\lambda^N x > B^N x \geq 0$ y $0 = \lim B^N x$. De donde $0 = \lim B^N$ pues $x > 0$ y la prueba termina.

Corolario 1. Suponga que B es una matriz sub-estocástica, esto es, B es no-negativa y $||B|| < 1$. Entonces cualquiera de los postulados del teorema 2 es equivalente a cada uno de los siguientes:

g. $||B^S|| < 1$. Donde S es el orden de B.

h. I - B es no-singular.

Prueba. Observe que g implica el postulado c del Teorema 1. La demostración del resultado recíproco es como sigue: Suponga que B es la matriz de transición de un proceso de Markov con S estados. Para cada hilera i defina la probabilidad de transición del estado i a un estado absorbente externo y denotela por $1 - \sum_{j=1}^{n} b_{ij}$. Sea i el conjunto de estados tales que $\sum_{j=1}^{n} b_{ij} < 1$. Si c es cierto, se implica que el proceso tiene como única clase comunicante cerrada el estado absorbente (pues de otra manera B tendría una submatríz tal que $||B^N|| = 1$ para toda N). Considerando que B^S es la matriz de transición de S pasos se tiene que $\sum_{J=1}^{N} b_{ij}^S < 1$ donde b_{ij}^S es el elemento (i,j) de B^S. Esto implica que $||B^S|| < 1$ y se prueba g. Finalmente observe que e implica h. Probaremos que h implica d. Usando la identidad $\sum_{i=0}^{N} B^i = (I-B)^{-1}(I-B^{N+1})$ se tiene que:

$$||\sum_{i=0}^{N} B^i|| \leq ||(I-B)^{-1}||\,[||I||+||B||^{N+1}] \leq 2\,||(I-B)^{-1}||$$

y se tiene que $\sum_{i=0}^{N} B_i$ es acotada superiormente y converge (absolutamente) cuando N tiende a infinito.

Teorema 3. (mapeo Espectral). Sea B una matriz cuadrada compleja con conjunto de vectores característicos denotado por $\sigma(B)$. Suponga que

$$f(B) = \sum_{N=0}^{\infty} a_N B^N$$

es absolutamente convergente. Entonces la función $f(\lambda)$ donde $\lambda\varepsilon\sigma(B) = \{f(\lambda);\ \lambda\varepsilon\sigma(B)\}$.

Prueba. El teorema de Jordan implica que podemos expresar $B = QJQ^{-1}$ donde J es de la forma

$$J = \begin{bmatrix} J_1 & & & \\ & \ddots & & \\ & & & J_n \end{bmatrix}$$

y cada submatriz $J(i=1,\ldots,n)$ es una matriz de Jordan asociada con un escalar $\lambda_i\varepsilon\sigma(B)$. Note que los valores característicos λ_i y λ_j asociados con las matrices de Jordan J_i y J_j, no son necesariamente distintos. Asimismo,

$$\sigma(B) = \sigma(J) = \{\lambda_1,\ldots,\lambda_n\}$$

pues B y J son matrices similares. Usando el hecho que $B^N = QJ^NQ^{-1}$ para toda N podemos implicar que $f(B)=Qf(J)Q^{-1}$ y

$$f(J) = \sum_{N=0}^{\infty} a_N J^N$$

Sin embargo, el conjunto de elementos en la diagonal principal de f(J) son dados por el conjunto

$$\left\{ \sum_{N=0}^{\infty} a_N \lambda_1^N , \ldots , \sum_{N=0}^{\infty} a_N \lambda_n^N \right\}$$

que coincide con los elementos del conjunto $f(\sigma(B))$. Si λ_i es valor característico de J sabemos que $|\lambda_i| \leq ||J||$ y podemos implicar

$$\sum_{N=0}^{\infty} |a_N \lambda_i^N| \leq \sum_{N=0}^{\infty} |a_N| \; ||J^N||$$

$$= \sum_{N=0}^{\infty} |a_N| \; ||Q^{-1} B^N Q||$$

$$\leq ||Q^{-1}|| \sum_{N=0}^{\infty} |a_N| \; ||B^N|| \cdot ||Q||$$

Entonces, f(B) absolutamente convergente implica la misma propiedad para $f(\lambda_i)$ $(i=1,\ldots,n)$, i.e., $f(\lambda)$ absolutamente convergente para todo $\lambda \varepsilon \sigma(B)$. Usando el hecho que f(B) y f(J) son matrices similares se tiene que $\sigma(f(B)) = \sigma(f(J))$. Sin embargo, $\sigma(f(J)) = f(\sigma(B))$ pues J es una matriz triangular. Por lo tanto, $\sigma(f(B)) = f(\sigma(B))$ y la prueba termina.

Teorema 1. (Frobenius-Perron). Sea A matriz cuadrada estrictamente positiva. Entonces:

a. Existen escalar $\lambda_0 > 0$ y vector $x_0 > 0$ tales que $Ax_0 = \lambda_0 x_0$

b. Si $Ax = \lambda_0 x$ se tiene que x es múltiplo de x_0

c. Si $Ax_0 = \mu x_0$ se tiene que $\mu = \lambda_0$

d. Si λ es otro valor característico de A entonces $|\lambda| < \lambda_0$

e. Existe vector $y_0 > 0$ tal que $y_0^t A = \lambda_0 y_0^t$.

Prueba. Considere el conjunto

$$S = \{\lambda \geq 0 \mid \text{Existe } x \geq 0 \text{ tal que } Ax \geq \lambda x\}$$

y observe que siendo A estrictamente positiva entonces $Ax > 0$ para toda $0 \neq x > 0$. Asimismo, para cada $0 \neq x \geq 0$ existe $\lambda > 0$ tal que $Ax > \lambda x$. Si en la definición de S suponemos que $ex = 1$ donde $e = [1,\ldots,1]$ se tiene que S es cerrado y acotado en R. Equivalentemente, S es compacto y su máximo, denotado λ_0, es positivo. La existencia del vector $x_0 \geq 0$ tal que $ex_0 = 1$ y $Ax_0 \geq \lambda_0 x_0$ es sencilla de establecer usando un argumento de continuidad. Si $Ax_0 = \lambda_0 x_0$ es inmediato que x_0 y se demuestra. a. Si $0 \neq Ax_0 - \lambda_0 x_0 \geq 0$ entonces $A[Ax_0 - \lambda_0 x_0] > 0$ que equivale a $Aw > \lambda_0 w$ donde $w = Ax_0$. Sin embargo esto contradice la definición de λ_0.

Sea $Ax = \lambda_0 x$ donde $x = u + iv$. Si x no es múltiplo de x_0 entonces u ó v no es múltiplo de x_0. Suponga que u no es múltiplo de $x_0 > 0$. Observe que existe escalar ε tal que $0 \neq x_0 + \varepsilon u \geq 0$ con alguna componente igual a cero, pero

$\lambda_0(x_0 + \epsilon u) = A(x_0 + \epsilon u) > 0$ que es falso y se demuestra b. Sea $Ax_0 = \mu x_0$. Usando el vector $y_0^t > 0$ de e se tiene:

$$\lambda_0 y_0^t x_0 = y_0^t A x_0 = \mu y_0^t x_0$$

donde $y_0^t x_0 > 0$. Por lo tanto $\lambda_0 = \mu$ y se demuestra c.

Sea $Ay = \lambda y$ donde $y \neq 0$. Denote por $|y|$ el vector cuyas componentes son los valores absolutos de y. Puesto que A es estrictamente positiva se observa $A|y| \geq |Ay| = |\lambda|\,|y|$. Debido a la definición de λ_0 se tiene que $|\lambda| \leq \lambda_0$. Suponga $|\lambda| = \lambda_0$ y $\lambda \neq \lambda_0$. Sea $\delta > 0$ tal que $A_\delta = A - \delta I$ es matriz estrictamente positiva y observe que si $\sigma(A)$ denota el conjunto de valores característicos de A se cumple que la relación $\sigma(A) - \delta = \sigma(A-\delta)$. Repitiendo los argumentos anteriores con A_δ en lugar de A se tiene $|\lambda-\delta| \leq \lambda_0 - \delta$ y

$$|\lambda| = |\lambda-\delta+\delta| \leq |\lambda-\delta| + \delta \leq \lambda_0$$

De donde $|\lambda| = \lambda_0$ implica que $|\lambda-\delta| + \delta = \lambda_0$. Sin embargo, desarrollando la igualdad $|\lambda-\delta| = \lambda_0 - \delta$ se concluye que λ es real y positivo. Entonces $\lambda = |\lambda| = \lambda_0$ que es una contradicción. Por lo tanto $|\lambda| < \lambda_0$ y se demuestra d. La parte e es inmediata si usamos A^t en lugar de A en la parte a y recordamos que $\sigma(A) = \sigma(A)^t$.

Teorema 5. Sea A matriz cuadrada no-negativa. Entonces

a. Existen escalar $\lambda_0 \geq 0$ y vector $x_0 \geq 0$ tales que $Ax_0 = \lambda_0 x_0$

b. Si $Ax_0 = \mu x_0$ donde $x_0 > 0$ entonces $\mu = \lambda_0$

c. Si $\lambda \neq \lambda_0$ es valor característico de A entonces $|\lambda| \leq \lambda_0$

d. Existe vector $y_0 \geq 0$ tal que $y_0^t A = \lambda_0 y_0^t$.

Prueba. Sea E matriz cuadrada cuyos elementos son todos iguales a uno y defina la matriz estrictamente positiva A_n como $A_N = A + (1/N)E$. Observe que la sucesión de matrices $\{A_N\}$ es tal que $A = \lim A_N$. Para $N \geq 1$ fija, se tiene del teorema 4. (Frobenius-Perron), que existen escalar $\lambda_N > 0$ y vector $x_N > 0$ tales que $A_N x_N = \lambda_N x_N$ y $\lambda_N = |\sigma(A_N)|$. Es inmediato que existen $\lambda_0 \geq 0$ tal que $\lambda_0 = \lim \lambda_N$ pues $\lambda_1 > \lambda_2 > \lambda_3 > \ldots\ldots$, dado que $A_1 > A_2 > A_3 > \ldots$. En particular se tiene que $\lambda_0 \geq |\sigma(A)|$. Sin perdida de generalidad podemos suponer que cada vector x_N está normalizado, esto es, $ex_N = 1$ y la sucesión de vectores $\{x_N\}$ tiene una subsucesión convergente a un vector $x_0 \geq 0$ tal que $e x_0 = 1$. El limite de $A_N x_N = \lambda_N x_N$ $Ax_0 = \lambda_0 x_0$ donde $\lambda_0 \geq 0$ y $0 \neq x_0 \geq 0$. Esto demuestra que $\lambda_0 \in \sigma(A)$. El resto de la prueba es sencilla y no se reproduce.

Corolario 1. Sea A matriz cuadrada no-negativa tal que A^m es estrictamente positiva para algun m. Entonces

a. Existe escalar $\lambda_0 > 0$ y vector $x_0 >$ tales que $Ax_0 = \lambda_0 x_0$
b. Si $Ax = \lambda_0 x$ se tiene que x es multiplo de x_0
c. Si $Ax_0 = \mu x_0$ se tiene que $\mu = \lambda_0$
d. Si $\lambda \neq \lambda_0$ es valor característico de A entonces $|\lambda| < \lambda_0$
e. Existe vector $y_0 > 0$ tal que $y_0^t A = \lambda_0 y_0^t$.

Prueba. Sea $A^m > 0$. Aplicando el teorema 4 se implica que existen $\alpha_0 > 0$ y $x_0 > 0$ tales que $A^m x_0 = \alpha_0 x_0$. Por otra parte se observa que $\alpha_0(Ax_0) = A(\alpha_0 x_0) = A(A^m x_0) = A^m(Ax_0)$ y se implica que $Ax_0 > 0$ es múltiplo de $x_0 > 0$. De donde $Ax_0 = \lambda_0 x_0$ para algún $\lambda_0^m > 0$. La igualdad $\alpha_0 x_0 = A^m x_0 = \lambda_0^m x_0$ implica $\alpha_0 = \lambda_0$. Si λ es otro valor característico de A con vector característico x se tiene que $Ax = \lambda x$ y $A^m x = \lambda^m x$. De aqui se implica que $|\lambda^m| < \alpha_0 = \lambda_0^m$ que equivale a $|\lambda| < \lambda_0$. Si $Ax = \lambda_0 x$ se tiene que $A^m x = \lambda_0^m x$ y se implica que x es múltiplo de x_0. Si $Ax_0 = \mu x_0$ y $Ax_0 = \lambda_0 x_0$ podemos premultiplicar por $q \neq 0$ ambos sistemas e implicar que $\mu = \lambda_0$. La última parte es inmediata si recordamos que $\sigma(A) = \sigma(A^t)$.

Nota. La matriz A en el corolario es regular.

Proposición 1. Sea A matriz no negativa de orden nxn y λ_0 su correspondiente valor característico de Frobenius-Perron. Entonces

$$\min \{\delta_i \mid i=1,\ldots,n\} \leq \lambda_0 \leq \max \{\delta_i \mid i=1,\ldots,n\}$$

donde δ_i es la suma de los elementos de la hilera (columna) i de A.

Prueba. Sea $x_0^t = [x_1, x_2, \ldots, x_n]$ al correspondiente vector característico asociado a λ_0 y suponga por conveniencia que la suma de los elementos de x_0 es igual a uno. Observe que $Ax_0 = \lambda_0 x_0$ y que si e es un vector hilera con n componentes igual a uno entonces $eAx_0 = \lambda_0 e\, x_0$ equivale a

$$\Delta_1 x_1 + \Delta_2 x_2 + \ldots + \Delta_n x_n = \lambda_0 (x_1 + x_2 + \ldots + x_n) = \lambda_0$$

donde Δ_i es la suma de los elementos en la i-ésima columna de A. Equivalentemente λ_0 es un promedio pesado de la suma de las columnas. Puesto que el promedio pesado está localizado en los valores extremos de las sumas de las columnas se tiene que

$$\min \{\delta_i \mid i=1,\ldots,n\} \leq \lambda_0 \leq \max \{\delta_i \mid i=1,\ldots,n\}$$

Un argumento semejante aplicado a A^t termina la prueba.

Lema 2. Sea A matriz cuadrada no-negativa e irreducible de orden n. Entonces $[I+A]^{n-1}$ es estrictamente positiva.

Prueba. El resultado es inmediato si $[I+A]^{n-1} w > 0$ para toda $0 \neq w \geq 0$. El postulado es cierto si dado $0 \neq w \geq 0$ con algún elemento igual a cero, el vector $z = [I+A]w$ tiene menos ceros que w. Suponga que esto último es falso. Note que elementos positivos de w originan elementos positivos en z, pues $z = w + Aw$. Suponga que

$$w = \begin{bmatrix} 0 \\ u \end{bmatrix} \quad ; \quad z = \begin{bmatrix} 0 \\ v \end{bmatrix}$$

donde $u > 0$ y $v > 0$ son vectores de la misma dimensión. Procediendo a particionar la matriz A de manera de expresar convenientemente la relación $z = w + Ax$ se tiene

$$\begin{bmatrix} 0 \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ u \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ u \end{bmatrix}$$

que implica $A_2 u = 0$. De donde $A_2 = 0$ pues $u > 0$. Sin embargo demuestra que A es reducible, pues

$$A = \begin{bmatrix} A_1 & 0 \\ A_3 & A_4 \end{bmatrix}$$

donde A_1 y A_2 son matrices cuadradas, que es una contradicción.

Repite este argumento n-1 veces, se observa que para cualquier vector $0 \neq w \geq 0$, se cumple

$$(I + A)^{n-1} w > 0$$

esto es, $(I+A)^{n-1} > 0$.

El lema 2 es interesante porque nos indica una manera relativamente simple de verificar si una matriz es irreducible o no.

B1. Conceptos generales de Latices.

Considere un conjunto no-vacío A con una relación binaria T definida en AxA que por simplicidad notacional la expresaremos por medio del símbolo "$\leq$" y diremos que $(a,b) \in T$ si $a \leq b$.

Un sistema $(A,\leq)$ se dice espacio parcialmente ordenado si la relación "$\leq$" cumple con los postulados

i.	$a \leq a$ para todo $a \in A$	(reflexividad)
ii.	$a \leq b$ y $b \leq c$ implican $a \leq c$	(transitividad)
iii.	$a \leq b$ y $b \leq a$ implican $a = b$	(antisimetría)

Asimismo si B es un subconjunto de A y a es un elemento de A tal que $b \leq a$ para toda $b \in B$ se dice que a es una cota superior de B. Análogamente, si $a \leq b$ para toda $b \in B$ se dice que es una cota inferior de B. Si c es una cota superior de B y cada cota superior a de B satisface $c \leq a$ entonces se dice que c es la mínima cota superior de B o bien el supremo de B y se denota como $c = \sup B$. Análogamente, si d es una cota inferior de B y cada cota inferior a de B satisface $a \leq d$, se dice que d es la máxima cota inferior de B o bien el infimo de B y se denota como $d = \inf B$.

Una latíz, denotada $(A,\leq)$, es un conjunto parcialmente ordenado tal que cada subconjunto de A que consista de dos elementos tiene un infimo y supremo. Específicamente, si a, b están en A entonces, existe elemento c en A tal que

$c = \sup\{a,b\}$ y se denota como $c = a \vee b$. Análogamente, existe elemento d en A tal que $d = a \wedge b$.

Una latíz $(A, \leq)$ se dice completa si cada subconjunto B de A tiene supremo e infimo. Una latiz de este tipo tiene siempre dos elementos $\underline{0}$ y $\underline{1}$ definidos por las fórmulas $\underline{0} = \inf A$ y $\underline{1} = \sup A$. Dados dos elementos cualesquiera a,b en A tales que $a \leq b$ se define el conjunto intervalo como

$$[a,b] = \{x \in A \mid a \leq x \leq b\}$$

y se observa que $([a,b], \leq)$ es una latiz completa si A es completa.

Una función $f:A \to A$ donde A es una latíz se dice isótona si dados x, y en A tales que $x \leq y$ se tiene que $f(x) \leq f(y)$. Se dice que la función f tiene un punto fijo x en A si se tiene que $f(x) = x$. El punto se dice excesivo si $x \geq f(x)$ y deficiente si $x \leq f(x)$.

Un resultado muy interesante de la teoría de latíz para nuestro estudio es el resultado de Tarski. El teorema de Tarski (Pacific J. Math, 1955) indica que un conjunto de puntos fijos de una latíz completa asociada a una función isótona es una latíz completa. Y se presenta este teorema a continuación.

Teorema 1. (Tarski). Sea $(A, \leq)$ una latíz completa y $f:A \to A$ una función isótona. Sea P el conjunto de puntos fijos de f en A. Entonces P es no-vacio y el sistema $(P, \leq)$ es una latíz completa. Asimismo se tiene que

$$VP = V\{x \in A \mid x \leq f(x)\} \in P$$
$$\Lambda P = \Lambda\{x \in A \mid x \geq f(x)\} \in P$$

Prueba. Sea $Q = \{x \in A \mid x \leq f(x)\}$ y defina $u = VQ$. Es inmediato que $x \leq u$ para toda $x \in A$ tal que $x \leq f(x)$. Asimismo $f(x) \leq f(u)$ por ser f isótona. Por lo tanto $x \leq f(u)$ para toda x en Q. Esto implica $u \leq f(u)$ debido a la definición de u. Por otra parte $u \in Q$ y $f(u) \leq f(f(u))$. Esto implica que $f(u) \leq u$ y se demuestra que $u = f(u)$. Asimismo $u = VP = VQ$ pues $v \in P \subset Q$. De manera análoga sea $S = \{x \in A \mid x \geq f(x)\}$ y defina $v = \Lambda S$. Observe que $x \geq v$ para toda $x \in A$ tal que $x \geq f(x)$. Asimismo, debido a la propiedad isótona de f se tiene que $f(x) \geq f(v)$. De aquí que $x \geq f(v)$ para toda x en S. La definición de v implica que $v \geq f(v)$. Sin embargo $f(v) \geq f(f(v))$ demuestra que $f(v)$ pertenece a S y por lo tanto $f(v) \geq v$. De donde $v = f(v)$ y se concluye que $v = \Lambda P = \Lambda S$ pues $v \in P \subset S$.

Sea Y un subconjunto arbitrario de P. Entonces el sistema $([vy, 1], \leq)$ es una latíz completa. Si $x \in Y$ se tiene que $x \leq vy$ y $x = f(x) \leq f(vy)$. Por definición de vy se implica $vy \leq f(vy)$. Asimismo, $vy \leq z$ implica $vy \leq f(z)$. De donde

la restricción del dominio de f al intervalo [VY,1] equivale a una función isótona f' de [VY,1] a [VY,1]. Aplicando el resultado de la primera parte de la prueba al conjunto

$$S' = \{x \in [VY, 1] \mid f(x) \leq x\}$$

se demuestra que el infimo de S' es igual al infimo de todos los puntos fijos de f' y que tal elemento es un punto fijo de f' que denotaremos por r. Obviamente, r es el punto fijo mínimo de f que es una cota superior de todos los elementos de Y. Equivalentemente r es el supremo de Y en el sistema [P, $\leq$].

Un argumento semejante demuestra la existencia del infimo de Y en el sistema [P, $\leq$]. Específicamente, se observa que el intervalo [0, AY] es una latíz completa y que la restricción de f a dicho intervalo equivale a una función isótona f" de [0, AY] a [0, AY]. Aplicando el resultado del principio de la prueba al conjunto

$$Q' = \{x \in [0, AY] \mid x \geq f(x)\}$$

se demuestra que el supremo de Q es igual al supremo de todos los puntos fijos de f" y que tal elemento es un punto fijo de f" que denotaremos por q. Dicho elemento es obviamente el infimo de Y en el sistema [P, $\leq$]. De aqui se concluye que [P, $\leq$] es una latíz completa.

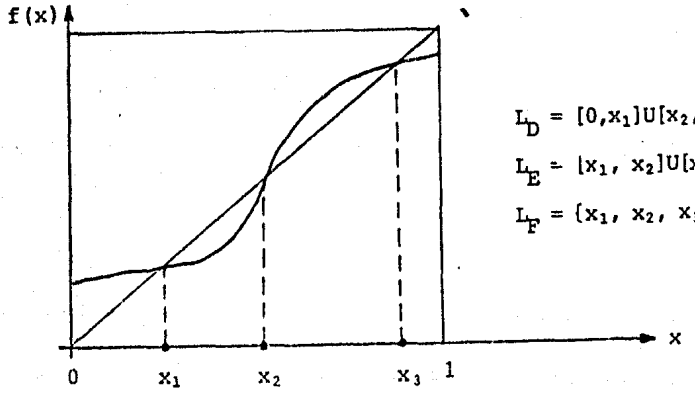
B2. Latices y funciones isótonas en R^n.

Sean u y v vectores en R^n. Se dice que $u \leq v$ si se cumple que $u_i \leq v_i$ para todo $i = 1,\ldots,n$. Asimismo $u < v$ si $u \leq v$ y $u \neq v$. Si $L \subset R^n$ se dice que $T:L \to L$ es isótono si $u \leq v$ implica $Tu \leq Tv$ y estrictamente isótono si $u < v$ implica $Tu < Tv$. Un punto u en L se dice excesivo, deficiente o fijo respecto al mapeo T si $u \geq Tu$, $\leq Tu$ y $u = Tu$, respectivamente. El conjunto de puntos excesivos, deficientes y fijos de f en L se denotan por L_E, L_D y L_F, respectivamente.

Ejemplo 1. Sea $0 \in L$ y $T0 < 0$. Entonces el punto 0 es un punto excesivo.

Ejemplo 2. Sea $T:L \to L$ definido como $Tu = r + Pu$ donde $r \in R^n$ y P es matriz no-negativa de orden nxn. Entonces T es un mapeo isótono, pues si $u \leq v$ se tiene que $Tu \leq Tv$.

Ejemplo 3. La función real definida en [0,1] mostrada en la figura tiene tanto puntos fijos como excesivos y deficientes.



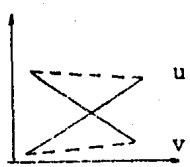
$L_D = [0,x_1]U[x_2,x_3]$

$L_E = [x_1, x_2]U[x_3,1]$

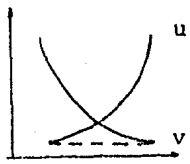
$L_F = \{x_1, x_2, x_3\}$

Un conjunto L de R^n se dice una inf-semilatíz si cada par de puntos u_1, u_2 en L tienen una máxima cota inferior que denotaremos por $u_1 \wedge u_2$ y se denomina el infimo de u_1, u_2. Análogamente, L es una sup-semilatíz si cada par de puntos u_1 y u_2 en L tienen una máxima cota inferior que denotaremos por $u_1 \vee u_2$ y se denomina el supremo de u_1, u_2. Un subconjunto J de L se dice un sup-subsemilatíz si J es una sub-semilatíz y el supremo de cada par de elementos en J coincide con el supremo de ese mismo par considerados como elementos de L. La definición de inf-subsemilatiz es semejante. Observe que L es una latíz si es tanto inf-semilatíz como sup-semilatíz.

Como ejemplo, sea LCR^n es una latíz. Suponga que L_F es un conjunto de todas las sucesiones afines en L. Entonces, L_F es una latíz no-vacia pero no es una inf ni sup - subsemilatíz de L. Suponga que L_E es un conjunto de todas las sucesiones convexas en L, entonces L_E es una sup-subsemilatiz de L. Suponga que L_D es un conjunto de todas las sucesiones cóncavas de L, entonces, L_D es una inf-subsemilatiz de L. Gráficamente:

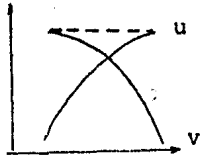


$u,v \in L_F$



$u,v \in L_E$

$u \vee v = \max\{u,v\}$



$u,v \in L_D$

$u \wedge v = \min\{u,v\}$

Teorema 1. Sea $T : L \to L$ tal que

$$T v = \max_{f \in A} \{r(f) + p(f)v\}$$

en donde A es finito, $v \in L$. Entonces T es un mapeo continuo.

Prueba: Sean $u, v \in L$. Supongan que a_0 es la decisión que maximiza Tu y a_1 es la que maximiza Tv. Entonces,

$$\begin{aligned}
Tu(i) - Tv(i) &= \max_{f \in A} \{r(i,f) + \sum_{j=0}^{S} p_{ij}(f)u(j)\} - \\
&\quad \max_{f \in A} \{r(i,f) + \sum_{j=0}^{S} p_{ij}(f)v(j)\} \\
&= r(i,a_0) + \sum_{j=0}^{S} p_{ij}(a_0)u(j) - \{r(i,a_1) + \sum_{j=0}^{S} p_{ij}(a_1)v(j)\} \\
&\leq r(i,a_0) + \sum_{j=0}^{S} p_{ij}(a_0)u(j) - r(i,a_0) - \sum_{j=0}^{S} p_{ij}(a_0)v(j) \\
&= \sum_{j=0}^{S} p_{ij}(a_0)\,[u(j) - v(j)] \\
&\leq p_{ij}(a_0)\, C \\
&= C. \qquad \forall\, i \in S,\ f \in A
\end{aligned}$$

donde $C = 2 \max \{\sup_i u(i), \sup_i v(i)\} \geq 0$. Por otro lado se obtiene que

$$Tv(i) - Tu(i) \leq \sum_{j=0}^{S} p_{ij}(a_1)\,[v(j) - u(j)] \leq \sum_{j=0}^{S} p_{ij}(a_1)\, C = C, \quad \forall\, i \in S, \ f \in A$$

esto implica que $||Tu - Tv|| \leq C$, dado que $||u-v|| \leq \delta$.

Teorema 2. Sea L una latiz compacta y T:L→L un mapeo isótono y continuo. Entonces

a. El conjunto de puntos excesivos de T es una inf-semilatiz no-vacia y compacta.
b. El conjunto de puntos deficientes de T es una sup-semilatíz no-vacía y compacta.
c. Los conjuntos de puntos excesivos, deficientes y fijos de T son una latiz no-vacia y compacta.

Prueba. a. Es inmediato que sup L es un elemento de la latíz por ser compacta y que dicho elemento pertenece al conjunto de puntos excesivos de T, denotado L_E. Sean u_1 y u_2 elementos de L_E. Entonces $u_1 \geq Tu_1$ y $u_2 \geq Tu_2$. Por otra parte, $u_1 \geq u_1 \wedge u_2$ y $u_2 \geq u_1 \wedge u_2$ implican que

$$u_1 \wedge u_2 \geq Tu_1 \wedge Tu_2 \geq T(u_1 \wedge u_2)$$

pues $u_i \geq Tu_i \geq Tu_1 \wedge Tu_2$ para i=1,2. Entonces L_E es una inf-semilatiz. La compacidad de L_E es inmediata.

b. Es inmediato que inf L es un elemento de la latíz y que pertenece al conjunto de puntos deficientes de T, denotado L_D. El resto de la prueba es semejante a la descrita en a.

c. Sean u_1 y u_2 elementos de L_E. Entonces

$$a = \inf \{w \in L_E \mid w \geq u_1 , w \geq u_2\}$$

$$= \inf L_E \cap [u_1, \sup L] \cap [u_2, \sup L]$$

que equivale a la determinación del ínfimo de un conjunto no-vacio y compacto. Por lo tanto existe $w \in L_E$ tal que $w = u_1 \vee u_2$. Esto demuestra que L_E es una sup-semilatíz y debido a la parte a se tiene que es una latiz no-vacia y compacta. La demostración para L_D es semejante.

Finalmente, sean u_1, u_2 elementos del comjunto de puntos fijos de T, denotado L_F. Observe que $L_F \subset L_E$ y que

$$u_1 \wedge u_2 \geq Tu_1 \wedge Tu_2 \geq T(u_1 \wedge u_2)$$

pues $u_1 \wedge$ y u_2 son puntos excesivos. Entonces

$$u_1 \wedge u_2 \geq T(u_1 \wedge u_2) \geq T^2(u_1 \wedge u_2) \geq \ldots \geq T^n(u_1 \wedge u_2)$$

para toda $n \geq 1$. Sin embargo, la sucesión $\{T^n(u_1 \wedge u_2)\}$ es acotada inferiormente por inf L y existe $w \in L$ tal que

$$w = \lim_n T^n(u_1 \wedge u_2) = \lim_n T\,(T^{n-1}(u_1 \wedge u_2)) = Tw$$

debido a la continuidad de T. Asimismo $w = u_1 \wedge u_2$ pues para todo punto fijo $\underline{w}$ tal que $u_1 \wedge u_2 \geq \underline{w}$ se cumple que

$$w = \lim_n T^n(u_1 \wedge u_2) \geq \underline{w} = T(\underline{w})$$

y se concluye que w es la máxima cota inferior. Un argumento semejante demuestra la existencia de una mínima cota inferior y L_F es una latíz no-vacia y compacta.

Teorema 3. Sea L una latiz compacta y $T:L \to L$ un mapeo isótono y continuo. Sean L_E, L_D y L_F los conjuntos de puntos excesivos, deficientes y fijos de T. Entonces

$$\inf L_E = \inf L_F$$

$$\sup L_D = \sup L_F$$

y ambos elementos pertenecen a L_F. Asmimismo se tiene

$$\lim T^n (\inf L) = \inf L_F \quad (\uparrow)$$

$$\lim T^n (\sup L) = \sup L_F \quad (\downarrow)$$

donde $(\uparrow)$ indica que la sucesión $\{T^n(\inf L)\}$ es creciente y converge a $\inf L_F$. La interpretación de $(\downarrow)$ es semejante.

Prueba. Es inmediato que $\mu = \inf L_E$ pertenece a L_E. Entonces $\mu \geq T\mu \geq T(T\mu)$ debido a la propiedad isótona de T y $T\mu$ pertenece a L_E. De donde $T\mu \geq \mu$ y se demuestra que $\mu = T\mu$ o bien $\mu \in L_F$. La demostración $\sup L_D = \sup L_F$ es semejante. Por otra parte, si $\mu = \inf L$ se tiene que $\mu \leq T\mu$ y se cumple que $T^n \mu \geq T^{n-1}\mu$ para toda $n=2,3,\ldots$. Entonces la sucesión no-decresiente $\{T^n\mu\}$ converge a un $v \in L$ tal que $v \in L_F$ pues

$$v = \lim_n T^{n+1}\mu = \lim_n T\,(T^n\mu)$$

debido a la continuidad de T. Note que si $w \in L_F$ entonces

$\mu \leq w$ y $T^n\mu \leq T^n w = w$. Tomando limites se implica que $v \leq w$ y se demuestra que $v = \inf L_F$. El otro caso es semejante.

B I B L I O G R A F I A

1. Bellman Richard E. & Dreyfus S.E. "Applied dynamic programming" Princeton University Press, 1962.

2. Blackwell David. "Discounted dynamic programmings" Ann. Math. Statistic 36, 226-235, 1965.

3. Blackwell David, "Discrete dynamic programming" Ann. Math. Statistic 33, 719-726, 1962.

4. Blackwell David, "Positive dynamic programming" Proc. 5th. Berkeley Symp. Math. Statistic Probability 1, 1967.

5. Bradley & Magnanti, "Applied mathematical programming" Addson Wesley, 1978.

6. Derman C., "Finite state markovian decision processes" Academic Press, 1970.

7. Dubins, L. & Savage L. J., "How to gamble if you must". McGraw-Hill, 1965.

8. Fuentes Maya Sergio. "Apuntes de clase de Programación Dinámica"., 1985.

9. Grätzer, G., "General lattice theory" Freeman, 1978.

.....

10. Howard Ronald A., "Dynamic programming and markov processes". The M.I.T. Press, 1960.

11. Karlin Samuel, "A first course in stochastic processes". Academic Press, 1966.

12. Lancaster Peter, "Theory of Matrices" Academic Press, 1969.

13. Luenberger David, "Introduction to dynamic systems model methods and applications". Jonhn Wiley & Sons. 1979.

14. Mine Hisashi & Osaki Shunji, "Markovian decision processes". American Elsevier Publishing Company, Inc., 1970.

15. Seneta Eugene, "Non-negative matrices: an Introduction to theory and applications". London, George Allen & Unwin Ltd, 1973.

16. Sheldon Ross M., "Applied probability models with optimization applications". Holden-Day, 1970.

17. Sheldon Ross M., "Introduction to stochastic dynamic programming", Academic Press, 1983.

18. Strauch Ralph E., "Negative dynamic programming" Ann. Math. Statistic 37, 871-890, 1966.

.....

19. Tarski Alfredo. "A lattice theoretical Fixpoint theorem and its applications". Pacific J. Math. 5, 285-309, 1955.

20. Veinott Arthur F. Jr., "Discrete dynamic programming with sensitive discount optimality criteria", Ann. Math. Statistic 40-5, 1635-1650, 1969.

21. White D. J. "Finite dynamic programming: an approach to Finite markov decision processes", John Wiley & Sons, 1978.