Modelos de Decisión en Ambientes Inciertos

(Apunte de Clases para el Curso Investigación Operativa IN44A)

Departamento de Ingeniería Industrial - Universidad de Chile

René A. Caldentey Susana V. Mondschein [1]

Enero, 1999

[1]La presente es una versión preliminar de este apunte docente, el cual se encuentra en construcción. Los autores agradecen los comentarios y correcciones de eventuales errores que aún permanezcan en el texto, los cuales pueden ser comunicados a smondsch@dii.uchile.cl, rcaldent@mit.edu o hawad@dii.uchile.cl

Capítulo 5

Cadenas de Markov con Beneficio

5.1 Introducción

Hasta aquí nos hemos dedicado a describir la evolución en probabilidad de un sistema que puede modelarse como una cadena de Markov. ¿Por qué querríamos, sin embargo, efectuar dicha descripción?. Presumiblemente porque no nos resulta indiferente que el sistema tome en uno u otro estado: la evolución del sistema tiene consecuencias, las cuales estamos interesados en predecir.

Supongamos, por ejemplo que una persona posee acciones de cierta empresa, cuyas utilidades anuales constituyen un proceso estocástico (y por ende el dividendo que recibe por sus acciones también). Esa persona puede estar interesada en cuantificar el flujo acumulado de dividendos a lo largo de varios años. Con ese propósito en mente le resultará necesario conocer la ley de probabilidades para los resultados de la empresa en cada año, que es el tipo de preguntas que nos hacíamos en el capítulo anterior.

En este capítulo nos interesaremos en determinar el beneficio esperado asociado a la evolución del sistema a través de tiempo, específicamente en el caso en que los beneficios dependen de un proceso estocástico subyacente el cual puede modelarse como una cadena de Markov.

5.2 Formulación del Modelo

Consideremos un sistema cuya evolución en el tiempo puede ser descrita a través de una cadena de Markov finita y homogénea con matriz de transición de un período P.[1] Supongamos que para cada par de estados E_i y E_j es posible asignar un beneficio r_{ij} asociado a la transición desde E_i a E_j, y que estamos interesados en cuantificar el beneficio total acumulado a lo largo de un número dado de períodos (i.e. la suma de los beneficios asociados

[1]En la práctica muchos sistemas pueden modelarse como una cadena de Markov haciendo el número de estados suficientemente grande y la duración de los períodos suficientemente pequeña

a las transiciones realizadas durante esos períodos). En tal caso diremos que el sistema es susceptible de ser modelado como una cadena de Markov con beneficios. Los valores r_{ij} no necesariamente corresponden a beneficios o costos económicos, sino cualquier función real de las transiciones, con tal que tenga sentido acumularlos de un período a otro.

Muchas veces seremos capaces de identificar beneficios que están asociados al hecho de alcanzar un estado en particular, más que a la transición de un estado a otro. En esas situaciones denotaremos por r_i al beneficio asociado a ocupar el estado E_i. En rigor esta notación adicional es innecesaria, pues ese beneficio asociado al estado E_i podría haber sido incorporado en todos los beneficios r_{ij} asociados a transiciones partiendo desde E_i. Sin embargo el identificarlos separadamente muchas veces aporta claridad a la modelización.

Si sabemos que en un período dado el sistema alcanza el estado E_i, el beneficio que se percibirá durante ese período es todavía probabilístico, pues no sabemos hacia qué estado evolucionará el sistema (y existen beneficios asociados a las transiciones). Podemos definir $\hat{r}_i$ como el beneficio esperado en un período asociado al estado E_i. Si p_{ij} es la probabilidad que el sistema evolucione de E_i a E_j, se tiene que

$$\hat{r}_i = r_i + \sum_j p_{ij} \cdot r_{ij},$$

es decir $\hat{r}_i$ es la suma del beneficio percibido por alcanzar dicho estado (r_i) más el valor esperado de los beneficios asociados a las posibles transiciones desde E_i. Denotaremos por $\hat{r}$ al vector de beneficios esperados en un período asociados a los distintos estados del sistema.

Supongamos que nos interesa describir el valor esperado del beneficio acumulado asociado a la evolución del sistema en un horizonte de n períodos. Sea $V_k(i)$ el beneficio total esperado si faltan k períodos para el final del horizonte y el sistema se encuentra actualmente en el estado E_i.[2] Denotaremos por $V_0(i)$ al beneficio de terminar en el estado E_i, al que se denomina valor residual del estado E_i (usualmente los valores residuales no guardan directa relación con la estructura de beneficios propios de la evolución del sistema los que vienen dados por $\hat{r}$).

5.3 Ejemplo

Imagine una empresa que al comenzar cada año invierte todo su patrimonio en su operación. Al finalizar el año observa sus resultados. Si obtuvo ganancias entonces capitaliza la mitad de ellas y el resto lo reparte como dividendos entre sus accionistas. Si hubo pérdidas entonces no reparte dividendos. Las ganancias o pérdidas de un año son una variable aleatoria cuya distribución depende sólo del patrimonio con que comienza ese año (mientras mayor es el patrimonio aumenta la probabilidad de obtener mayores ganancias). Además, por simplicidad supondremos que dicha variable aleatoria sólo puede tomar valores enteros múltiplos de 2.

[2] V_k corresponde al vector de beneficios acumulados esperados si faltan k períodos para el final.

Bajo los supuestos realizados la evolución del patrimonio de un año a otro puede modelarse como una cadena de Markov. Llamando X_n el patrimonio en el año n, la ley de probabilidades para X_{n+1} depende sólo de X_n. Las probabilidades de transición vienen dadas por

$$\Pr[X_{n+1} = j | X_n = i] \equiv p_{ij} = \begin{cases} \Pr[\text{ganancia } = 2k | \text{patrimonio } = i] & \text{Si } j = i + k \\ \Pr[\text{ganancia } = -2k | \text{patrimonio } = i] & \text{Si } j = i - 2k \\ 0 & \text{Si } j = i - 2k - 1 \end{cases} \quad k \in I\!N$$

Suponga ahora que ud. posee acciones de dicha empresa, y tiene intenciones de mantenerlas en su poder durante 10 años. Pasado ese lapso venderá las acciones, y con el dinero obtenido por la venta y por los dividendos acumulados durante los 10 años se retirará a descansar en algún agradable rincón de este planeta (¿si todavía quedan?). El precio de venta de las acciones dentro de 10 años es una función (determinística) del patrimonio de la empresa en ese momento. Si el patrimonio de la empresa es j ud. podrá vender las acciones que posee en $z(j)$.

¿Cual es el valor esperado del capital con el que ud. se retirará? Para responder esa pregunta construiremos un modelo de Markov con beneficios que dé cuenta de sus ingresos. Ya hemos descrito como la evolución del patrimonio puede modelarse como una cadena de Markov. Toca ahora identificar los beneficios asociados a estados y transiciones.

Resulta natural asociar beneficios a las transiciones (cambios de valor del patrimonio). Si el patrimonio aumenta de un año a otro es porque la empresa obtuvo utilidades, y parte de ellas serán percibidas por usted como dividendos. Si el patrimonio cae, ud. no recibirá nada ese año (la empresa tuvo pérdidas). Definimos entonces

$$r_{ij} = \begin{cases} \alpha \frac{j-i}{2} & \text{Si } j > i \\ 0 & \text{Si no} \end{cases}$$

donde α representa la fracción de la propiedad de la empresa que ud. posee. En este caso no es necesario asociar beneficios directamente a los estados ($r_i = 0 \, \forall \, i$). Así

$$\hat{r}_i = 0 + \sum_j p_{ij} \cdot r_{ij}.$$

Finalmente, la venta de las acciones (al cumplirse los 10 años) la incorporamos como un beneficio residual:

$$V_0(i) = z(i) \quad \forall \, i.$$

5.4 Expresión General para el Beneficio Acumulado

Debido a la propiedad markoviana el cálculo de $V_k(i)$ es independiente del comportamiento del sistema en los períodos anteriores y sólo depende del los k períodos que restan. Teniendo en cuenta este hecho y suponiendo que los valores residuales son conocidos podemos calcular el beneficio acumulado esperado si falta 1 período para el final, V_1.

Si faltando un período para el final el sistema se encuentra en el estado E_i entonces el beneficio acumulado esperado se calcula como el beneficio esperado de estar en el estado E_i más el beneficio residual esperado, es decir

$$V_1(i) = \hat{r}_i + \sum_j p_{ij} \cdot V_0(j) \tag{5.1}$$

De igual forma si faltan k períodos para el final, el beneficio acumulado esperado de encontrarse en el estado E_i viene dado por el beneficio esperado en el período presente $(\hat{r}_i)$ más el beneficio esperado desde el período siguiente en adelante $(V_{k-1}(j))$ el cual depende del estado que el sistema alcance en el período siguiente, por lo que debemos condicionar y sumar sobre los estados posibles (ponderados por sus respectivas probabilidades):

$$V_k(i) = \hat{r}_i + \sum_j p_{ij} \cdot V_{k-1}(j) \tag{5.2}$$

Vectorialmente (5.2) queda como

$$V_k = \hat{r} + P\, V_{k-1} \tag{5.3}$$

donde $\hat{r}, V_k, V_{k-1}$ son vectores columnas.

Ahora bien, sustituyendo en (5.3) k por $k-1$ y reemplazando en (5.3) se tiene

$$V_k = \hat{r} + P\,\hat{r} + P^2\, V_{k-2} \tag{5.4}$$

si el procedimiento anterior se aplica en forma recursiva se tiene:

$$V_k = \hat{r} + P\,\hat{r} + P^2\,\hat{r} + ... + P^k\, V_0 \tag{5.5}$$

La ecuación (5.5) permite calcular el valor de V_k conocidos P, $\hat{r}$ y V_0, sin embargo, la expresión es aún incómoda de manejar sobre todo para valores grandes de k. Para obtener una expresión más simplificada para V_k es necesario exigir una condición más a la cadena de Markov.

5.5 Caso Cadena de Markov Ergódica

Supongamos que la cadena de Markov con beneficios que estudiamos es ergódica (o ergódica + transiente), es decir, admite una única clase recurrente la que es aperiódica. Bajo estas condiciones la sucesión ${P^n}_{n\in N}$ converge a una matriz Π con todas sus filas iguales al vector de probabilidades estacionarias π. Ahora bien, reemplazando k por $k-1$ en (5.5) y restándola a (5.5) se obtiene:

$$V_k - V_{k-1} = P^{k-1}\,\hat{r} + [P^k - P^{k-1}]\, V_0 \tag{5.6}$$

Tomando límite cuando k tiende a infinito en (5.6) se tiene que

$$\lim_{k\to\infty}(V_k - V_{k-1}) = \Pi\,\hat{r} \tag{5.7}$$

Llamando $g = \sum_i \pi_i \cdot \hat{r}_i$ y e al vector columna de 1's se tiene:

$$\lim_{k\to\infty}(V_k - V_{k-1}) = g \cdot e \tag{5.8}$$

Observamos que el beneficio adicional que se obtiene por aumentar en un período el horizonte de planificación converge a un valor constante , el cual no depende del estado inicial. Este resultado es intuitivo, pues el beneficio asociado a ese período extra es el beneficio esperado asociado a una transición que será realizada dentro de mucho tiempo; ese beneficio esperado depende del estado en que se encuentre el sistema, y la probabilidad de estar en cada estado converge a la ley de probabilidades estacionaria independiente del estado inicial. Es decir, g es el incremento límite de V_k por unidad de crecimiento de k. De este resultado es razonable pensar que $V_k - kge$ converge a un cierto vector límite u, lo cual por el momento lo asumiremos como cierto (este resultado se desprende de la Proposición 5.1, que enunciaremos más abajo). Resulta útil escribir $u = W + \alpha e$ con $W = (W1, W2, ..., Wr)$ y $W1 = 0$.[3]

$$\lim_{k\to\infty} V_k - kge = W + \alpha e \qquad W_1 = 0 \tag{5.9}$$

Ahora bien, como P es una matriz de probabilidad se tiene que $Pe = e$ con lo cual (5.3) puede reescribirse como:

$$\begin{aligned} V_k - kge &= \hat{r} + P\,V_{k-1} - kge \\ &= \hat{r} - ge + P\,[V_{k-1} - (k-1)ge] \end{aligned} \tag{5.10}$$

por lo tanto tomando límite en la expresión anterior se tiene por (5.9)

$$W + \alpha e = \hat{r} - ge + P\,[W + \alpha e] \qquad W_1 = 0 \tag{5.11}$$

la que puede simplificarse a

$$W + ge = \hat{r} + P\,W \qquad W_1 = 0 \tag{5.12}$$

El sistema anterior admite solución única bajo el supuesto de una cadena ergódica. El vector W solución del sistema anterior es independiente de V_0 y es este el motivo que originó la separación en W y α. W se conoce como el vector asintótico de beneficios relativos debido a que satisface $W = \lim_{k\to\infty}[V_k - V_k(1)e]$. En efecto,

[3] La elección de $W1$ es arbitraria considerando que siempre es posible reordenar los estados de la cadena pudiendo ser cualquiera E_1.

$$\begin{aligned}
\lim_{k\to\infty}[V_k - V_k(1)e] &= \lim_{k\to\infty}[V_k - kge - (V_k(1)e - kge)] \\
&= W + \alpha e - (W_1 e + \alpha e)] \\
&= W - W_1 e \\
&= W \qquad (5.13)
\end{aligned}$$

Es decir W_i representa el beneficio adicional de largo plazo de empezar la evolución en el estado E_i en lugar de E_1.

Proposición 5.1 *Sea una cadena de Markov finita, homogénea y ergódica con matriz de transición de un período P. Sea W la única solución de (5.12). Entonces el beneficio esperado acumulado durante k períodos (V_k) viene dado por:*

$$V_k = kge + W + [P]^n (V_0 - W)$$

donde $g = \sum_j \pi_j \cdot \hat{r}_j$ es el beneficio esperado por transición en régimen estacionario. Además, en el límite se tiene:

$$\lim_{k\to\infty}[V_k - kge] = W + \alpha e \quad \alpha = \sum_j \pi_j [V_0(j) - W_j]$$

La demostración se encuentra en el Apéndice 5-A.

5.6 Aplicación: Tiempo Esperado en el Transiente

Consideremos una cadena de Markov ergódica con al menos una clase transiente. Mediante una formulación de cadena de Markov con beneficios seremos capaces de determinar el valor esperado del tiempo (N^o de períodos) necesario para llegar desde un estado transiente cualquiera hasta la única clase recurrente del sistema.

Para ello, definamos $\hat{r}_i = 1$ si E_i es un estado transiente y $\hat{r}_j = 0$ si E_j es recurrente. De esta forma el tiempo promedio para llegar desde E_i a la clase recurrente viene dado por:

$$\lim_{k\to\infty} V_k(i)$$

Para calcular el límite anterior debemos primero determinar los valores para W a partir de la ecuación (5.12). Ahora bien para este caso particular el valor de g es 0 pues los estados recurrentes tienen un beneficio $\hat{r} = 0$ y los estados no recurrentes tienen una probabilidad estacionaria $\pi = 0$. Por lo tanto el sistema a resolver es en este caso:

$$W = \hat{r} + PW$$

Si se ordenan adecuadamente los estados el sistema anterior toma la forma:

$$\begin{pmatrix} W_R \\ W_T \end{pmatrix} = \begin{pmatrix} 0 \\ e_T \end{pmatrix} + \begin{bmatrix} P_{RR} & P_{RT} \\ P_{TR} & P_{TT} \end{bmatrix} \cdot \begin{pmatrix} W_R \\ W_T \end{pmatrix} \qquad P_{RT} = 0$$

en donde los subíndices T y R representan los estados transientes y recurrentes respectivamente. Para el sistema anterior se obtienen dos ecuaciones vectoriales

$$W_R = P_{RR} W_{RR}$$

$$W_T = e_T + P_{TR} W_R + P_{TT} W_T$$

En la primera ecuación la matriz P_{RR} es un matriz de probabilidad (todos sus elementos son no negativos y la suma de sus elementos por fila es igual a 1) asociada a una cadena ergódica por lo tanto la solución W_R toma la forma:

$$W_R = \lambda \cdot e_R \quad \lambda = cte$$

Además, el sistema de ecuaciones anterior tiene un grado de libertad, luego escogiendo $\lambda = 0$ se tiene que $W_R = 0$. Reemplazando este resultado en la segunda ecuación y despejando W_T se obtiene:

$$W_T = (I_T - P_{TT})^{-1} e_T$$

La matriz $(I_T - P_{TT})$ es invertible, pues P_{TT} es la matriz de probabilidades de transición entre los estados transientes (el razonamiento para verificarlo se expuso en la Sección **??** 4.3.4).

Finalmente, el cálculo del tiempo promedio de evolución desde un estado transiente a uno recurrente viene dado por:

$$\lim_{k \to \infty} V_k = \lim_{k \to \infty} W - [P]^k W = W - \Pi W$$

Como $W_i = 0$ si E_i es recurrente y $\pi_j = 0$ si E_j es transiente se tiene que $\Pi W = 0$ con lo cual se tiene

$$\lim_{k \to \infty} V_k = W$$

con

$$W = \begin{pmatrix} W_R \\ W_T \end{pmatrix} = \begin{pmatrix} 0 \\ (I_T - P_{TT})^{-1} e_T \end{pmatrix}$$

Note que este resultado es consecuente con la interpretación de W como vector de beneficios relativos: para cualquier estado transiente E_i y cualquier estado recurrente E_j se tiene que $W_i - W_j = W_i$ representa cuánto tiempo más, en promedio, se está en el transiente por partir desde E_i en lugar de hacerlo desde E_j (caso en el que no se pasa por el transiente).

5.7 EJERCICIOS

1. Considere un juego de azar con las siguiente características.

    - En cada jugada el jugador tiene una probabilidad p de ganar.
    - Si el jugador gana en una jugada recibe el pozo de $\$P$ y se termina el juego.
    - El jugador no se puede retirar del juego mientras no haya ganado.
    - La primera jugada vale \$1000, la segunda vale \$2000, la tercera vale \$4000, la k-ésima jugada vale $\$1000^*\ 2^{k-1}$.

    (a) Determine la cantidad esperada de dinero que un jugador debe gastar antes de ganar el juego.

    (b) Determine el número esperado de jugadas a realizar.

Apéndice

5-A Demostración Propiedades cadenas de Markov con Beneficio

Demostración de la Proposición 5.1

Sea una cadena de Markov finita, homogénea y ergódica con matriz de transición de un período P. Sea W la única solución de (5.12). Entonces el beneficio esperado acumulado durante k períodos (V_k) viene dado por:

$$V_k = kge + W + [P]^n\,(V_0 - W)$$

donde $g = \sum_j \pi_j \cdot \hat{r}_j$ es el beneficio esperado por transición en regimen estacionario. Además, en el límite se tiene:

$$\lim_{k\to\infty}[V_k - kge] = W + \alpha e \quad \alpha = \sum_j \pi_j [V_0(j) - W_j]$$

Demostración

Consideremos otra cadena de Markov con beneficios idéntica a la anterior salvo por su vector de beneficios residuales, el cual consideraremos igual al vector de beneficios relativos, W. Denotemos por V'_k el beneficio esperado acumulado en esta nueva situación. Hemos impuesto $V'_0 = W$. Se tiene entonces que

$$V'_1 = \hat{r} + P\,W = W + ge \qquad \text{(por la condición (5.12))} \tag{5.14}$$

De igual forma

$$V'_2 = \hat{r} + P\,V'_1 = \hat{r} + P\,(W + ge) = W + 2ge \tag{5.15}$$

razonando inductivamente es fácil ver que en general se cumple

$$V'_k = kge + W \tag{5.16}$$

además, usando (5.5) se tiene

$$V'_k = \hat{r} + P\,\hat{r} + ... + P^k\,W \tag{5.17}$$

Luego considerando V_k como el beneficio esperado acumulado para un vector de valores residuales arbitrario dado por la ecuación (5.5) y restando la expresión (5.17) se tiene:

$$V_k - V'_k = P^k[V_0 - W] \tag{5.18}$$

con lo cual reemplazando (5.16) en la ecuación anterior se tiene

$$V_k = kge + W + P^k[V_0 - W] \tag{5.19}$$

Finalmente de (5.19) se tiene que $\lim_{k \to \infty} V_k - kge = W + \alpha e$ con $\alpha = \sum_j \pi_j (V_0(j) - W_j)$ (por la convergencia de P^k cuando $k \longrightarrow \infty$). ∎