CONTROL DE LÍNEAS DE ESPERA A TIEMPO DISCRETO CON COSTO DESCONTADO TOTAL ESPERADO

Mirelda Dionicio Arevalo*, Heliodoro Daniel Cruz Súarez

Universidad Juárez Autánoma de Tabasco, México

Recibido enero 15 de 2015 y aceptado abril 30 de 2015

Abstract

Un problema de líneas de espera controladas a tiempo discreto, con espacio finito de estados y con espacio finito de acciones se aproxima por medio de aproximaciones sucesivas. El método de programación dinámica nos proporciona la solución a los problemas de cadenas de decisión de Markov a tiempo discreto y con espacio finito de estados. En particular, sea el criterio de costo descontado total esperado para resolver el problema de interés por medio de programación dinámica utilizando iteración de valores y haciendo uso de la herramienta de excel para realizar los calculos.

A problem of waiting lines controlled to discrete time, with finite state space and finite action space is approximated by successive approximations. The dynamic programming method provides the solution to the problems of Markov chains decision to discrete time and finite state space. In particular, it is deal with the criteria of total discounted cost expected, to solve the problem of interest by using dynamic programming interection of values and using excel tool to perform calculations.

Keywords: Programación dinámica, cadenas de decisión de Markov, costo descontado total esperado.
dynamic programming, Markov chains decision, total discounted cost expected

1. Introducción

En este artículo se presenta una problema de líneas de espera controladas, el cual se modela y se resuelve como un proceso de decisión de Markov (PDM).

Una línea de espera o cola se forma en sistemas que ofrecen servicios con cierta capacidad de atención, al llegar los clientes si el servidor no está disponible, y el cliente decide esperar, se forma dicha cola.

Un proceso de decisión de Markov o problema de control estocástico, consiste de un modelo mediante el cual se representa la dinámica de un sistema cuya evolución es aleatoria, pero que su comportamiento puede ser influenciado de tal forma que se logren obtener ciertas metas. Este modelo está conformado por estados, acciones, costos y probabilidades de transición. Las políticas de control son reglas bajo las cuales se eligen los controles (acciones) aplicados para cualquier futura eventualidad. Así, un PDM consiste, además, del modelo de decisión markoviano, de un índice de

*__Dirección postal__: Carr. Cunduacán-Jalpa Km 1, Cunduacán, Tabasco, México. A.P. 24, C.P. 86690 Tel.(+52)914 336-0928. **Correo electrónico**: meilyd a@yahoo_com.mx

funcionamiento (o criterio de optimalidad) mediante el cual se mide la respuesta del sistema a los controles aplicados. Así que, el objetivo de un problema de decisión markoviano es encontrar políticas que minimicen al índice de funcionamiento.

La solución del modelo se presenta por programación dinámica.

En la segunda sección de preliminares se presenta la teoría básica de líneas de espera, en la tercera sección se presenta la teoría de PDM. Finalmente en la sección 4 se resuelve un problema.

2. PRELIMINARES

La estructura matemática que se considera se conoce como: **Cadena de decisión de Markov** o **Proceso de control estocástico a tiempo discreto**. Esta construcción es útil para estudiar el control de fenómenos estocásticos a tiempo discreto (ver[5]). Un modelo de control estocástico a tiempo discreto con espacio de estados numerables es una quíntupla $(X, A, \{A(x)|x \in X\}, p, c)$, que consta del espacio de estados finito X, el espacio de acciones finito A, un subconjunto de acciones admisibles $A(x)$ para cada estado $x \in X$, una ley de transición $p_{xy}(a)$ entre los estados y la función de costos medible c. Para cualquier política $\pi \in \Pi$ y estado inicial $x \in X$, se define el siguiente **índice de funcionamiento** de horizonte finito conocido también como **criterio de optimalidad** el cual es de utilidad para plantear el problema de control estocástico, este criterio proporciona una medida del rendimiento de cada política de control aplicada, (ver [5]). El cual se define de la siguiente manera:

$$J_{\alpha,n}(\pi, x) := E_x^{\pi}\left[\sum_{t=0}^{n-1} \alpha^t c_t(x_t, a_t) + \alpha^n c(x_n)\right], \tag{1}$$

$x \in X$, llamado el *costo descontado total esperado*, donde $\alpha \in (0, 1)$ es factor de descuento.

Entonces, el problema de control es encontrar una política π^* tal que:

$$J_{\alpha,n}^*(x) = J_{\alpha,n}(\pi^*, x) = \min_{\pi \in \Pi} J_{\alpha,n}(\pi, x). \tag{2}$$

A la ecuación (2) se le conoce como *función de valor óptimo* y a la política π^* se le llama *política óptima*.

$J_{\alpha,n}^*(x)$ satisface la siguiente ecuación

$$J_{\alpha,n}^*(x) = \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in X} p_{xy}(a) J_{\alpha,n}^*(y) \right\},$$

llamada *ecuación de programación dinámica*.

Suponemos que el mínimo en dicha ecuación se alcanza en una política estacionaria óptima a^*, es decir,

$$J_{\alpha,n}^*(x) = c(x, a^*) + \alpha \sum_{y \in X} p_{xy}(a^*) J_{\alpha,n}^*(y).$$

Para obtener la solución de la ecuación de programación dinámica se utiliza el *método de iteración de valores.*

$$J_{\alpha,t}(x) = \min_{a\in A(x)} \left\{ c(x,a) + \alpha \sum_{y\in X} p_{xy}(a) J_{\alpha,t+1}(y) \right\}.$$

3. Procesos de Decisión de Markov

La teoría básica para solucionar este tipo de problemas de control estocástico a tiempo discreto es la programación dinámica estocástica. De manera general se puede decir que el enfoque de la programación dinámica para la solución del problema de control estocástico, con horizonte finito n, consiste en descomponer el problema de optimización de n períodos en n problemas de optimización de una etapa, es decir, el problema se reduce a encontrar una sucesión, $\{a_t^*\}_{t=0}^{n-1}$, de reglas de decisión con la característica de que cada a_t^* haga mínimo el costo descontado total esperado al tiempo t.

PROPOSICIÓN: Sea $\{q(a)\}_{a\in A}$ una distribución de probabilidad en el conjunto finito $A(\neq \emptyset)$. Sea $\bar{u} : A \to (-\infty, \infty)$ una función. Entonces,

1. $\sum_{a\in A} q(a)\bar{u}(a) \geq \min_{a\in A}\{\bar{u}(a)\}$,
2. la igualdad ocurre si y sólo si la distribución de probabilidad está concentrada sobre el subconjunto $B_x(\alpha, n)$

$$B = \{b \in A | \bar{u}(b) = \min_{a\in A}\{\bar{u}(a)\}\}.$$

Prueba. Por demostrar la primera parte de la proposición:

$$\sum_{a\in A} q(a)\bar{u}(a) \geq \min_{a\in A}\{\bar{u}(a)\}.$$

La distribución está concentrada en $B := B_x(\alpha, n)$ si $q(a) = 0$ para $a \notin B$. Suponemos que $\min_{a\in A}\{\bar{u}(a)\} = w < \infty$. Entonces,

$$\begin{aligned} &\bar{u}(a) \geq w, \\ \Rightarrow &\sum_a q(a)\bar{u}(a) \geq w \sum_a q(a) = w, \end{aligned}$$

i.e.

$$\sum_a q(a)\bar{u}(a) \geq \min_{a\in A}\{\bar{u}(a)\}.$$

Para demostrar la segunda parte de la proposición se debe probar que la distribución de probabilidad está concentrada en B si y sólo si la igualdad ocurre. La minimización es sobre un conjunto finito B no vacío.

Si q está concentrada en B, entonces se cumple la igualdad, de lo contrario suponer

que existe $a^* \in A - B$ tal que $q(a^*) > 0$. Sea $\bar{u}(a^*) = w + \delta$, donde $\delta > 0$. Entonces,

$$\begin{aligned}\sum_a q(a)\bar{u}(a) &= \sum_{a \neq a^*} q(a)\bar{u}(a) + q(a^*)\bar{u}(a^*) \\ &\geq w \sum_{a \neq a^*} q(a) + (w+\delta)q(a^*) \\ &= w\left(\sum_{a \neq a^*} q(a) + q(a^*)\right) + \delta q(a^*) \\ &= w + q(a^*)\delta \\ &> w.\end{aligned}$$

■

TEOREMA: La función del valor óptimo de horizonte finito satisface la ecuación de programación dinámica (ver [4]),

$$J^*_{\alpha,n}(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in X} p_{xy}(a) J^*_{\alpha,n-1}(y) \right\}, \tag{3}$$

$x \in X, n \geq 1$.

1. Una política π es óptima para el horizonte $n = 1$ si y sólo si dado el estado inicial x, la distribución $p(a|x)$ se concentra en el conjunto $B_x(\alpha, n)$ (igual a cero fuera de este conjunto).

2. Una política π es óptima para el horizonte $n \geq 2$ si y sólo si
(a). Dado el estado inicial x, la distribución $p(a|x)$ se concentra en el conjunto $B_x(\alpha, t)$.
(b). Dado los movimientos del proceso para el estado y en $t = 1$ entonces, π sigue una política óptima de horizonte $n - 1$ con estado inicial y.

Prueba. La demostración de este teorema se hace por inducción sobre el horizonte.

Primero suponemos para $n = 1$. El caso para el primer horizonte, el decisor (o controlador) actúa en $t = 0$. Entonces, observa el estado en $t = 1$ e incurre en un costo terminal. Sean el estado inicial x y π una política arbitraria para el horizonte $n = 1$, entonces

$$\begin{aligned}J_{\alpha,1}(\pi, x) &= \sum_{a \in A(x)} p(a|x) E^{\pi}\left\{c(x_0, a_0) + \alpha c(x_1) | x_0 = x, a_0 = a\right\} \\ &= \sum_{a \in A(x)} p(a|x) \left\{ c(x,a) + \alpha \sum_{y \in X} p_{xy}(a) J^*_{\alpha,1}(y) \right\} \\ &= \sum_{a \in A(x)} p(a|x) u_{\alpha,1}(x,a) \\ &\geq \min_{a \in A(x)} \{u_{\alpha,1}(x,a)\}. \end{aligned} \tag{4}$$

Como (4) es válido para todo π, se sigue que

$$\min_{\pi \in \Pi} J_{\alpha,1}(\pi, x) \geq \min_{a \in A(x)} \{u_{\alpha,1}(x,a)\}.$$

Entonces, se tiene que

$$J_{\alpha,1}(\pi,x) \geq J^*_{\alpha,1}(x) \geq \min_{a\in A(x)}\{u_{\alpha,1}(x,a)\}. \tag{5}$$

De la última expresión de la ecuación (5) se observa que se cumple con la ecuación (3).

Luego, sea π una política con $p(a|x)$ concentrada en $B_x(\alpha,1)$. La primera parte de la proposición anterior indica que cualquier política satisface

$$J_{\alpha,1}(\pi,x) = \min_{a\in A(x)}\{u_{\alpha,1}(x,a)\}.$$

Esto significa que los términos en (5) cumplen con la igualdad e implica que (3) es válido para $n=1$ y que π es óptimo. Esto prueba la primera parte del inciso (1) del teorema.

Para demostrar la segunda parte de (1), sea π una política arbitraria, de la segunda parte de la proposición y de la ecuación (5) tenemos que π es óptima sólo si $p(a|x)$ se concentra en $B_x(\alpha,1)$. Por lo tanto, esto completa la prueba para $n=1$.

Ahora, suponemos las declaraciones para $n\geq 2$. La suposición que se usará para la inducción es la existencia de un horizonte $n-1$ y la correspondiente política óptima π^*.

Sean el estado inicial x y π una política arbitraria para el problema de horizonte n. Si la acción inicial es a y el estado al tiempo $t=1$ es y, sea $\psi(x,a,y)$ la regla de decisión para el horizonte $n-1$ bajo π, iniciando al tiempo $t=1$. Entonces,

$$\begin{aligned}
J_{\alpha,n}(\pi,x) &= \sum_{a\in A(x)} p(a|x)E_\pi[c(x_0,a_0) + \alpha\sum_{t=1}^{n-1}\alpha^{t-1}c(x_t,a_t) \\
&\quad + \alpha^n c(x_n)|x_0=x, a_0=a] \\
&= \sum_{a\in A(x)} p(a|x)\{c(x,a) + \alpha\sum_{y\in X} p_{xy}(a)J_{\alpha,n-1}(\psi(x,a,y),y)\} \\
&\geq \sum_{a\in A(x)} p(a|x)\left\{c(x,a) + \alpha\sum_{y\in X} p_{xy}(a)J^*_{\alpha,n-1}(y)\right\} \\
&= \sum_{a\in A(x)} p(a|x)u_{\alpha,n}(x,a) \\
&\geq \min_{a\in A(x)}\{u_{\alpha,n}(x,a)\}.
\end{aligned} \tag{6}$$

donde $x\in X$, $n=0,1,\cdots$.

De (6) se sigue que

$$J_{\alpha,n}(\pi,x) \geq J^*_{\alpha,n}(x) \geq \min_{a\in A(x)}\{u_{\alpha,n}(x,a)\}.$$

Luego, suponemos que $p(a|x)$ está concentrada en $B_x(\alpha,n)$ y entonces sigue la política π^*. De la primera parte de la proposición se sigue que la última línea de (6) es una igualdad. Así, $\psi(x,a,y)=\pi^*$ nos dice que la tercera línea es una igualdad. De esta

manera, (3) es válida para n y existe una política óptima que satisface la igualdad. Por lo tanto esto demuestra el inciso (a) de (2).

Para el resto de la demostración de (2), suponemos que $p(a|x)$ > 0 para algún $a \notin B_x(\alpha, n)$. De la segunda parte de la proposición se sigue que la última desigualdad es estricta en (6) y de esta manera π no puede ser óptima.

Ahora, se supone que $p(a|x)$ está concentrada en $B_x(\alpha, n)$. Si existe $b \in B_x(\alpha, n)$ tal que $p(b|x) > 0$ y y tal que $p_{xy}(b) > 0$, entonces esto significa que el estado y puede ser alcanzado al tiempo $t = 1$ bajo la política π. Suponemos que π no actúa óptimamente para el horizonte $n-1$ en y. Esto implica que

$$J_{\alpha,n-1}(\psi(x,a,y),y) > J^*_{\alpha,n-1}(y).$$

Esto significa que la primera desigualdad en (6) es estricta, y por lo tanto π no es óptima. ■

TEOREMA: Sean $\{J_{\alpha,t}\}_{t=0}^n$ funciones definidas (recursivamente). Si para cada $t = 0, 1, \cdots, n$ existe una regla de decisión a_t^* tal que

$$J_{\alpha,t}(x) = c(x_t, a_t^*(x)) + \alpha \sum_{y \in X} p_{xy}(a_t^*(x)) J_{\alpha,t+1}(y),$$

$x \in X$, entonces la política de decisión markoviana $\pi^* = \{a_0^*, a_1^*, \cdots, a_n^*\}$ es óptima y $J_{\alpha,0}(x)$ es la función de valor óptimo, esto es,

$$J_{\alpha,0}(x) = J^*_{\alpha,n}(x).$$

Prueba. Sea $\pi = \{a_t\}_{t=0}^n$ una política arbitraria y se define

$$c_{\alpha,l}(\pi, h_l) := E_x^\pi \left[\sum_{t=l}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n c(x_n) | h_l \right], l = 0, 1, ..., n-1, \tag{7}$$

$$c_{\alpha,n}(\pi, h_n) := \alpha^n E^\pi [c(x_n)|h_n].$$

Nótese que $c_{\alpha,l}(\pi, h_l)$ representa el costo descontado total esperado bajo la política π del tiempo l al tiempo n, dada la historia h_l.

En particular, nótese que

$$\begin{aligned} J_{\alpha,n}(\pi, x) &= E_x^\pi \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n c(x_n) \right] = E_x^\pi \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n c(x_n) | h_0 \right] \\ &= c_{\alpha,0}(\pi, h_0), \end{aligned}$$

$x \in X$, por lo tanto,

$$J_{\alpha,n}(\pi, x) = c_{\alpha,0}(\pi, h_0) \tag{8}$$

cuando $h_0 = x$.

Para demostrar el teorema, se debe probar que para todo $x \in X$ y $l = 0, 1, \cdots, n$,

$$c_{\alpha,l}(\pi, h_l) \geq J_{\alpha,l}(x), \tag{9}$$

cuando $h_l = (h_{l-1}, a_{t-1}, x)$, si $\pi = \pi^*$ se proporciona la igualdad, es decir,

$$c_{\alpha,l}(\pi^*, h_l) = J_{\alpha,l}(x), \tag{10}$$

cuando $h_l = (h_{l-1}, a_{l-1}, x)$.

En efecto, para $l = 0$, se observa de (8), (9) y (10) que

$$J_{\alpha,n}(\pi, x) = c_{\alpha,0}(\pi, h_0) \geq J_{\alpha,0}(x) = c_{\alpha,0}(\pi^*, h_0) = J^*_{\alpha,n}(x),$$

cuando $h_0 = x$.

Luego,

$$J_{\alpha,n}(\pi, x) \geq J_{\alpha,0}(x) = J^*_{\alpha,n}(x) \geq \min_{\pi\in\Pi} J_{\alpha,n}(\pi, x),$$

para todo $x \in X$.

Como el ínfimo de un conjunto $B \subset \mathbb{R}$ es la máxima de las cotas inferiores, se sigue de lo anterior que

$$\min_{\pi\in\Pi} J_{\alpha,n}(\pi, x) \geq J_{\alpha,0}(x) \geq \min_{\pi\in\Pi} J_{\alpha,n}(\pi, x),$$

es decir,

$$J^*_{\alpha,n}(x) \geq J_{\alpha,0}(x) \geq J^*_{\alpha,n}(x),$$

por lo tanto

$$J^*_{\alpha,n}(x) = J^*_{\alpha,0}(x). \tag{11}$$

Se demuestra (9) y (10) por inducción (recursivamente). Por definición se tiene que

$$c_{\alpha,n}(\pi, h_n) = c(x_n) = J^*_{\alpha,n}(x),$$

cuando $h_n = (h_{n-1}, a_{n-1}, x)$.

Suponemos ahora, que para $x \in X$

$$c_{\alpha,l+1}(\pi, h_{l+1}) \geq J_{\alpha,l+1}(x), \tag{12}$$

cuando $h_{l+1} = (h_l, a_l, x)$.

Se debe demostrar que para toda $x \in X$

$$c_{\alpha,l}(\pi, h_l) \geq J_{\alpha,l}(x), \tag{13}$$

cuando $h_l = (h_{l-1}, a_{l-1}, x)$.

Por (7), se tiene

$$\begin{aligned} c_{\alpha,l}(\pi, h_l) &= E^\pi \left[\sum_{t=l}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n c(x_n) | h_l \right] \\ &= E^\pi \left[c(x_l, a_l) + \alpha \sum_{t=l+1}^{n-1} \alpha^{t-1} c(x_t, a_t) + \alpha^n c(x_n) | h_l \right] \\ &= E^\pi [c(x_l, a_l) | h_l] + \alpha E^\pi \left[\sum_{t=l+1}^{n-1} \alpha^{t-1} c(x_t, a_t) + \alpha^n c(x_n) | h_l \right]. \end{aligned}$$

Ahora, se obtiene

$$\begin{aligned} c_{\alpha,l}(\pi, h_l) &= c(x,a) + \alpha E^{\pi}\left[E^{\pi}\left[\sum_{t=l+1}^{n-1} \alpha^{t-1} c(x_t, a_t) + \alpha^n c(x_n)|h_{l+1}\right]|h_l\right] \\ &= c(x,a) + \alpha E^{\pi}[c_{\alpha,l+1}(\pi, h_{l+1})|h_l]. \end{aligned}$$

Por la hipótesis de inducción

$$\begin{aligned} c_{\alpha,l}(\pi, h_l) &\geq c(x,a) + \alpha E^{\pi}[J_{\alpha,l+1}(x)|h_l] \\ &= c(x,a) + \alpha \sum_{y \in X} p_{xy}(a_l) J_{\alpha,l+1}(y) \\ &\geq \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in X} p_{xy}(a_l) J_{\alpha,l+1}(y) \right\} \\ &= J_{\alpha,l}(x). \end{aligned} \tag{14}$$

Así,

$$c_{\alpha,l}(\pi, h_l) \geq J_{\alpha,l}(x),$$

de esta manera queda demostrado (9).

Por otra parte, si se cumple la igualdad en (13) y $\pi = \pi^*$ entonces, se cumple la igualdad en (10).

■

3.1 Aplicación

En este trabajo se proporciona un ejemplo el cual ilustra la teoría de control la cual resuelve el problema de la espera que consiste en minimizar los costos totales de aquellos que solicitan el servicio y de aquellos quienes lo prestan, y no solo en minimizar el tiempo que los clientes pasan en el sistema. Para resolver dicho problema de control se hace uso del método de programación dinámica y del criterio de costo descontado total esperado, utilizando la técnica de iteración de valores. El objetivo del problema de control es encontrar la política óptima que minimice el criterio de costo costo descontado total esperado con horizonte de planeación finito.

3.2 Ejemplo

Una sala de espera, con capacidad para dos clientes, es atendida por un solo servidor. Cuenta con dos tasas de servicio una de 0.4 y otra de 0.7 y por utilizar cualquiera de estas tasas de servicio se genera un costo de 1 y 2 unidades, respectivamente. Además, existe una probabilidad $p = 0.5$ de que un nuevo cliente arribe en cualquier etapa.

La sala de espera puede almacenar como máximo dos clientes, uno en servicio y uno en espera de servicio, debido a esto se genera un costo de una unidad por número de cliente. Ya que, si un cliente llega y encuentra la sala de espera llena, este se retira, o si el sistema esta vacio puede entrar al servicio inmediatamente.

Con los datos que se proporcionaron anteriormente se puede modelar el problema de control en una cadena de decisión de Markov de la siguiente manera. El modelo de control Markoviano es $(S, A, \{A(i)|i \in S\}, p, C)$, donde,

- $S = \{0, 1, 2\}$ es el espacio de estados.
- $A = A(i) = \{a_1, a_2\} = \{0.4, 0.7\}$ es el espacio de acciones (que en este ejemplo son las razones de servicio) donde $i \in S$.
- Las probabilidades de transición se van a obtener de la siguiente manera:

$$p_{ij}(\rho) = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix}$$
$$= \begin{pmatrix} 1-p+ap & (1-a)p & 0 \\ a(1-p) & ap+(1-a)(1-p) & (1-a)p \\ 0 & a(1-p) & 1-a(1-p) \end{pmatrix}$$

por ejemplo: $p_{00} = 1 - p + ap$ ya que el sistema permanecerá en el estado 0 sino existen arribos, o si existe un arribo y el servicio fue completado.

Entonces las probabilidades de transición por elegir la acción $a_1 = 0.4$ (es decir, la razón de servicio más lenta) están dadas por:

$$p_{ij}(a_1) = \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

y por elegir la acción $a_2 = 0.7$ (el servicio mas rápido) están dadas por:

$$p_{ij}(a_2) = \begin{pmatrix} 0.85 & 0.15 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0 & 0.35 & 0.65 \end{pmatrix}.$$

- Los costos que se generan por elegir cada una de las acciones se obtiene como sigue:

$$C(i, a) = i + c(a)$$

Entonces por tomar la acción a_1 son:

$$C(0, a_1) = 0 + 1 = 1,$$
$$C(1, a_1) = 1 + 1 = 2,$$
$$C(2, a_1) = 2 + 1 = 3.$$

y los costos que se generan por tomar la acción a_2 son los siguientes:

$$C(0, a_2) = 0 + 2 = 2,$$
$$C(1, a_2) = 1 + 2 = 3,$$
$$C(2, a_2) = 2 + 2 = 4.$$

Se utiliza el criterio de costo descontado total esperado para resolver el problema anterior, el objetivo será encontar la política óptima que minimice el costo descontado total esperado en un horizonte de planeación finito n, cuando el sistema se encuentre

en el estado x.
El problema de control es ver ecuación (2):

$$J^*_{\alpha,n}(x) = \min_{\pi\in\Pi} J_{\alpha,n}(\pi, x).$$

Utilizando el método de iteración de valores,

$$J_{\alpha,t}(x) = \min_{a\in A(x)} \left\{ c(x,a) + \alpha \sum_{y\in X} p_{xy}(a) J_{\alpha,t+1}(y) \right\}.$$

Para $t = n-1$,

$$\begin{aligned} J_{\alpha,1}(x) &= \min_{a\in A(x)} \{c(x,a) + \alpha \sum_{y\in X} p_{xy}(a) J_{\alpha,n}(y)\} \\ &= \min_{a\in A(x)} \{c(x,a_1) + \alpha \sum_{y=0}^{2} p_{xy}(a_1) J_{\alpha,0}(y), c(x,a_2) + \alpha \sum_{y=0}^{2} p_{xy}(a_2) J_{\alpha,0}(y)\}, \end{aligned}$$

como el costo inicial $J_{\alpha,0}(y) = c(x_0) = 0$, entonces se tiene que:

$$\begin{aligned} J_{\alpha,1}(0) &= \min_{a\in A(x)} \{c(0,a_1), c(0,a_2)\} = \min_{a\in A(0)} \{1,2\} = 1, B_0(n-1) = \{a_1\}, \\ J_{\alpha,1}(1) &= \min_{a\in A(x)} \{c(1,a_1), c(1,a_2)\} = \min_{a\in A(1)} \{2,3\} = 2, B_1(n-1) = \{a_1\}, \\ J_{\alpha,1}(2) &= \min_{a\in A(x)} \{c(2,a_1), c(2,a_2)\} = \min_{a\in A(2)} \{3,4\} = 3, B_2(n-1) = \{a_1\}, \end{aligned}$$

Para $t = n-2$,

$$\begin{aligned} J_{\alpha,n-2}(x) &= \min_{a\in A(x)} \{c(x,a) + \alpha \sum_{y\in X} p_{xy}(a) J_{\alpha,n-1}(y)\} \\ &= \min_{a\in A(x)} \{c(x,\rho) + \alpha \sum_{y=0}^{3} p_{xy}(\rho) J_{\alpha,n-1}(y), c(x,r) + \alpha \sum_{y=0}^{3} p_{xy}(r) J_{\alpha,n-1}(y)\} \\ &= \min_{a\in A(x)} \{c(x,\rho) + \alpha[p_{x0}(\rho) J_{\alpha,n-1}(0) + p_{x1}(\rho) J_{\alpha,n-1}(1) + p_{x2}(\rho) J_{\alpha,n-1}(2) \\ &\quad + p_{x3}(\rho) J_{\alpha,n-1}(3)], c(x,r) + \alpha[p_{x0}(r) J_{\alpha,n-1}(0) + p_{x1}(r) J_{\alpha,n-1}(1) \\ &\quad + p_{x2}(r) J_{\alpha,n-1}(2) + p_{x3}(r) J_{\alpha,n-1}(3)]\}, \end{aligned}$$

de lo anterior se tiene que:

$$\begin{aligned} J_{\alpha,n-2}(0) &= \min_{a\in A(0)} \{c(0,\rho) + \alpha[p_{00}(\rho) J_{\alpha,n-1}(0) + p_{01}(\rho) J_{\alpha,n-1}(1) + p_{02}(\rho) J_{\alpha,n-1}(2) \\ &\quad + p_{03}(\rho) J_{\alpha,n-1}(3)], c(0,r) + \alpha[p_{00}(r) J_{\alpha,n-1}(0) + p_{01}(r) J_{\alpha,n-1}(1) \\ &\quad + p_{02}(r) J_{\alpha,n-1}(2) + p_{03}(r) J_{\alpha,n-1}(3)]\} \\ &= \min_{a\in A(0)} \{1 + \frac{1}{9}[\frac{5}{12}(1.231) + \frac{1}{3}(1.212) + \frac{1}{4}(2.259) + 0], 1.5 + \frac{1}{9}[1(1.231) + 0]\} \\ &= \min_{a\in A(0)} \{1.164, 1.636\} = 1.164, \end{aligned}$$

$$B_0(2) = \{a_1\}.$$

$$
\begin{aligned}
J_{\alpha,n-2}(1) = & \min_{a\in A(1)}\{c(1,\rho)+\alpha[p_{10}(\rho)J_{\alpha,n-1}(0)+p_{11}(\rho)J_{\alpha,n-1}(1)+p_{12}(\rho)J_{\alpha,n-1}(2)\\
& +p_{13}(\rho)J_{\alpha,n-1}(3)], c(1,r)+\alpha[p_{10}(r)J_{\alpha,n-1}(0)+p_{11}(r)J_{\alpha,n-1}(1)\\
& +p_{12}(r)J_{\alpha,n-1}(2)+p_{13}(r)J_{\alpha,n-1}(3)]\}\\
= & \min_{a\in A(1)}\{1+\frac{1}{9}[\frac{1}{6}(1.231)+\frac{5}{12}(1.212)+\frac{1}{3}(2.259)+\frac{1}{12}(1.074)],\\
& 1.25+\frac{1}{9}[\frac{2}{3}(1.231)+\frac{1}{3}(1.212)+0]\}\\
= & \min_{a\in A(1)}\{1.172, 1.386\} = 1.172,
\end{aligned}
$$

$$B_1(n-2) = \{\rho\}.$$

$$
\begin{aligned}
J_{\alpha,n-2}(2) = & \min_{a\in A(2)}\{c(2,\rho)+\alpha[p_{20}(\rho)J_{\alpha,n-1}(0)+p_{21}(\rho)J_{\alpha,n-1}(1)+p_{22}(\rho)J_{\alpha,n-1}(2)\\
& +p_{23}(\rho)J_{\alpha,n-1}(3)], c(2,r)+\alpha[p_{20}(r)J_{\alpha,n-1}(0)+p_{21}(r)J_{\alpha,n-1}(1)\\
& +p_{22}(r)J_{\alpha,n-1}(2)+p_{23}(r)J_{\alpha,n-1}(3)]\}\\
= & \min_{a\in A(2)}\{2.25+\frac{1}{9}[0+\frac{1}{6}(1.212)+\frac{5}{12}(2.259)+\frac{5}{12}(1.074)],\\
& 2+\frac{1}{9}[0+\frac{2}{3}(1.212)+\frac{1}{3}(2.259)+0]\}\\
= & \min_{a\in A(2)}\{2.426, 2.173\} = 2.173,
\end{aligned}
$$

$$B_2(n-2) = \{r\}.$$

$$
\begin{aligned}
J_{\alpha,n-2}(3) = & \min_{a\in A(3)}\{c(3,\rho)+\alpha[p_{30}(\rho)J_{\alpha,n-1}(0)+p_{31}(\rho)J_{\alpha,n-1}(1)+p_{32}(\rho)J_{\alpha,n-1}(2)\\
& +p_{33}(\rho)J_{\alpha,n-1}(3)], c(3,r)+\alpha[p_{30}(r)J_{\alpha,n-1}(0)+p_{31}(r)J_{\alpha,n-1}(1)\\
& +p_{32}(r)J_{\alpha,n-1}(2)+p_{33}(r)J_{\alpha,n-1}(3)]\}\\
= & \min_{a\in A(3)}\{2+\frac{1}{9}[0+\frac{2}{3}(2.259)+\frac{1}{3}(1.074)], 1+\frac{1}{9}[0+\frac{2}{3}(2.259)+\frac{1}{3}(1.074)]\}\\
= & \min_{a\in A(3)}\{2.207, 1.207\} = 1.207,
\end{aligned}
$$

$$B_3(n-2) = \{r\}.$$

Para $t = n-3$,

$$
\begin{aligned}
J_{\alpha,n-3}(x) = & \min_{a\in A(x)}\{c(x,a)+\alpha\sum_{y\in X}p_{xy}(a)J_{\alpha,n-2}(y)\}\\
= & \min_{a\in A(x)}\{c(x,\rho)+\alpha\sum_{y=0}^{3}p_{xy}(\rho)J_{\alpha,n-2}(y), c(x,r)+\alpha\sum_{y=0}^{3}p_{xy}(r)J_{\alpha,n-2}(y)\}\\
= & \min_{a\in A(x)}\{c(x,\rho)+\alpha[p_{x0}(\rho)J_{\alpha,n-2}(0)+p_{x1}(\rho)J_{\alpha,n-2}(1)+p_{x2}(\rho)J_{\alpha,n-2}(2)\\
& +p_{x3}(\rho)J_{\alpha,n-2}(3)], c(x,r)+\alpha[p_{x0}(r)J_{\alpha,n-2}(0)+p_{x1}(r)J_{\alpha,n-2}(1)\\
& +p_{x2}(r)J_{\alpha,n-2}(2)+p_{x3}(r)J_{\alpha,n-2}(3)]\},
\end{aligned}
$$

de lo anterior se tiene que:

$$\begin{aligned}
J_{\alpha,n-3}(0) &= \min_{a\in A(0)}\{c(0,\rho)+\alpha[p_{00}(\rho)J_{\alpha,n-2}(0)+p_{01}(\rho)J_{\alpha,n-2}(1)+p_{02}(\rho)J_{\alpha,n-2}(2)\\
&\quad +p_{03}(\rho)J_{\alpha,n-2}(3)], c(0,r)+\alpha[p_{00}(r)J_{\alpha,n-2}(0)+p_{01}(r)J_{\alpha,n-2}(1)\\
&\quad +p_{02}(r)J_{\alpha,n-2}(2)+p_{03}(r)J_{\alpha,n-2}(3)]\}\\
&= \min_{a\in A(0)}\{1+\frac{1}{9}[\frac{5}{12}(1.164)+\frac{1}{3}(1.172)+\frac{1}{4}(2.173)+0],\\
&\quad 1.5+\frac{1}{9}[1(1.164)+0]\}\\
&= \min_{a\in A(0)}\{1.157, 1.629\} = 1.157,\\
&\qquad B_0(n-3)=\{\rho\}.
\end{aligned}$$

$$\begin{aligned}
J_{\alpha,n-3}(1) &= \min_{a\in A(1)}\{c(1,\rho)+\alpha[p_{10}(\rho)J_{\alpha,n-2}(0)+p_{11}(\rho)J_{\alpha,n-2}(1)+p_{12}(\rho)J_{\alpha,n-2}(2)\\
&\quad +p_{13}(\rho)J_{\alpha,n-2}(3)], c(1,r)+\alpha[p_{10}(r)J_{\alpha,n-2}(0)+p_{11}(r)J_{\alpha,n-2}(1)\\
&\quad +p_{12}(r)J_{\alpha,n-2}(2)+p_{13}(r)J_{\alpha,n-2}(3)]\}\\
&= \min_{a\in A(1)}\{1+\frac{1}{9}[\frac{1}{6}(1.164)+\frac{5}{12}(1.172)+\frac{1}{3}(2.173)+\frac{1}{12}(1.207)],\\
&\quad 1.25+\frac{1}{9}[\frac{2}{3}(1.164)+\frac{1}{3}(1.172)+0]\}\\
&= \min_{a\in A(1)}\{1.167, 1.379\} = 1.167,\\
&\qquad B_1(n-3)=\{\rho\}.
\end{aligned}$$

$$\begin{aligned}
J_{\alpha,n-3}(2) &= \min_{a\in A(2)}\{c(2,\rho)+\alpha[p_{20}(\rho)J_{\alpha,n-2}(0)+p_{21}(\rho)J_{\alpha,n-2}(1)+p_{22}(\rho)J_{\alpha,n-2}(2)\\
&\quad +p_{23}(\rho)J_{\alpha,n-2}(3)], c(2,r)+\alpha[p_{20}(r)J_{\alpha,n-2}(0)+p_{21}(r)J_{\alpha,n-2}(1)\\
&\quad +p_{22}(r)J_{\alpha,n-2}(2)+p_{23}(r)J_{\alpha,n-2}(3)]\}\\
&= \min_{a\in A(2)}\{2.25+\frac{1}{9}[0+\frac{1}{6}(1.172)+\frac{5}{12}(2.173)+\frac{5}{12}(1.207)],\\
&\quad 2+\frac{1}{9}[0+\frac{2}{3}(1.172)+\frac{1}{3}(2.173)+0]\}\\
&= \min_{a\in A(2)}\{2.428, 2.167\} = 2.167,\\
&\qquad B_2(n-3)=\{r\}.
\end{aligned}$$

$$\begin{aligned}
J_{\alpha,n-3}(3) &= \min_{a\in A(3)}\{c(3,\rho)+\alpha[p_{30}(\rho)J_{\alpha,n-2}(0)+p_{31}(\rho)J_{\alpha,n-2}(1)+p_{32}(\rho)J_{\alpha,n-2}(2)\\
&\quad +p_{33}(\rho)J_{\alpha,n-2}(3)], c(3,r)+\alpha[p_{30}(r)J_{\alpha,n-2}(0)+p_{31}(r)J_{\alpha,n-2}(1)\\
&\quad +p_{32}(r)J_{\alpha,n-2}(2)+p_{33}(r)J_{\alpha,n-2}(3)]\}\\
&= \min_{a\in A(3)}\{2+\frac{1}{9}[0+\frac{2}{3}(2.173)+\frac{1}{3}(1.207)], 1+\frac{1}{9}[0+\frac{2}{3}(2.173)+\frac{1}{3}(1.207)]\}\\
&= \min_{a\in A(3)}\{2.205, 1.205\} = 1.205,\\
&\qquad B_2(n-3)=\{r\}.
\end{aligned}$$

La siguiente tabla muestra los resultados de la etapa n hasta la etapa $n-10$ para cada estado, la acción que minimizó el costo en cada etapa para los estados $x\in\{0,1\}$ es ρ y para los estados $x\in\{2,3\}$ es r,

costos \ estados	0	1	2	3
$J_{\alpha,n}(x)$	2	3	1	0
$J_{\alpha,n-1}(x)$	1.231	1.212	2.259	1.074
$J_{\alpha,n-2}(x)$	1.164	1.172	2.173	1.207
$J_{\alpha,n-3}(x)$	1.157	1.167	2.167	1.205
$J_{\alpha,n-4}(x)$	1.156	1.166	2.166	1.205
$J_{\alpha,n-5}(x)$	1.156	1.166	2.166	1.205
$J_{\alpha,n-6}(x)$	1.156	1.166	2.166	1.205
$J_{\alpha,n-7}(x)$	1.156	1.166	2.166	1.205
$J_{\alpha,n-8}(x)$	1.156	1.166	2.166	1.205
$J_{\alpha,n-9}(x)$	1.156	1.166	2.166	1.205

Entonces, para cualquier etapa $t = n - l$ se tiene que,

$$J_{\alpha,n-l}(0) = 1.156,$$
$$J_{\alpha,n-l}(1) = 1.166,$$
$$J_{\alpha,n-l}(2) = 2.166,$$
$$J_{\alpha,n-l}(3) = 1.205.$$

La política óptima es

$$\pi^* = \begin{cases} \rho & \text{si } x \in \{0,1\} \\ r & \text{si } x \in \{2,3\}. \end{cases}$$

Por demostrar que vale para la etapa $t = n - (l+1)$,

$$\begin{aligned} J_{\alpha,n-(l+1)}(0) &= c(0,\rho) + \alpha[p_{00}(\rho)J_{\alpha,n-l}(0) + p_{01}(\rho)J_{\alpha,n-l}(1) \\ &\quad + p_{02}(\rho)J_{\alpha,n-l}(2) + p_{03}(\rho)J_{\alpha,n-l}(3)] \\ &= 1 + \frac{1}{9}[\frac{5}{12}(1.156) + \frac{1}{3}(1.166) + \frac{1}{4}(2.166)] \\ &= 1.156. \end{aligned}$$

$$\begin{aligned} J_{\alpha,n-(l+1)}(1) &= c(1,\rho) + \alpha[p_{10}(\rho)J_{\alpha,n-l}(0) + p_{11}(\rho)J_{\alpha,n-l}(1) \\ &\quad + p_{12}(\rho)J_{\alpha,n-l}(2) + p_{13}(\rho)J_{\alpha,n-l}(3)] \\ &= 1 + \frac{1}{9}[\frac{1}{6}(1.156) + \frac{5}{12}(1.166) + \frac{1}{3}(2.166) + \frac{1}{12}(1.205)] \\ &= 1.166. \end{aligned}$$

$$\begin{aligned} J_{\alpha,n-(l+1)}(2) &= c(2,\rho) + \alpha[p_{20}(\rho)J_{\alpha,n-l}(0) + p_{21}(\rho)J_{\alpha,n-l}(1) \\ &\quad + p_{22}(\rho)J_{\alpha,n-l}(2) + p_{23}(\rho)J_{\alpha,n-l}(3)] \\ &= 2 + \frac{1}{9}[\frac{2}{3}(1.166) + \frac{1}{3}(2.166) \\ &= 2.166. \end{aligned}$$

$$\begin{aligned} J_{\alpha,n-(l+1)}(3) &= c(3,\rho) + \alpha[p_{30}(\rho)J_{\alpha,n-l}(0) + p_{31}(\rho)J_{\alpha,n-l}(1) \\ &\quad + p_{32}(\rho)J_{\alpha,n-l}(2) + p_{33}(\rho)J_{\alpha,n-l}(3)] \\ &= 1 + \frac{1}{9}[\frac{2}{3}(2.166) + \frac{1}{3}(1.205)] \\ &= 1.205. \end{aligned}$$

Entonces, en este problema las funciones de valor óptimo para cada $x \in X$ son:

$$J^*(0) = J_{\alpha,0}(0) = 1.156,$$
$$J^*(1) = J_{\alpha,0}(1) = 1.166,$$
$$J^*(2) = J_{\alpha,0}(2) = 2.166,$$
$$J^*(3) = J_{\alpha,0}(3) = 1.205.$$

Por lo tanto, se han encontrado las funciones de valor óptimo para cada estado y la política óptima.

4. Conclusión

En este artículo, se presenta la solución a un problema de control estocástico. Para resolver dicho problema de control es importante sealar que se hizo uso del método de programación dinámica y del criterio de costo descontado total esperado, utilizando la técnica de iteración de valores. El objetivo del problema de control es encontrar la política óptima que minimice el criterio de costo costo descontado total esperado con horizonte de planeación finito.

Referencias

[1] Hernández-Lerma O., Procesos Estocásticos: Introducción a la Teoría de colas, 2 Coloquio del Departamento de Matemáticas del Centro de Investigación y de Estudios Avanzados del IPN, Oaxtepec, 1981.

[2] Lipschutz S., Probabilidad, McGraw-Hill, México, 1991.

[3] Puterman M. L., Markov Decision Processes, Wiley, New York, 1994.

[4] Sennott L. I., Constrained discounted Markov decision chains, Probability in the Engineering and Informational Sciences, 5, 1991, 463-475.

[5] Sennott L. I., Stochastic Dynamic Programming and the Control of Queueing Systems, Wiley Canada, 1999.